Proceedings of the 6th WSEAS Int. Conf. on FUZZY SYSTEMS, Lisbon, Portugal, June 16-18, 2005 (pp17-22)

Fuzzy aproach in biometric authentication by keystroke dynamics

JAN CAPEK, MILOSLAV HUB
Institut of System Engineering and Informatics
Faculty of Economic and Administration
University of Pardubice
Studentska 84, 530 09 Pardubice
THE CZECH REPUBLIC
http://fes.upce.cz

Abstract: A person's identity verification became very important in information society. This article presents some results of our research of biometric authentication by keystroke dynamics. The comparison of the stochastic approach and using fuzzy numbers is presented as well.

Key-Words: authentication, biometric, keystroke dynamics, password, fuzzy numbers

1 Introduction

Authentication as a data security instrument in our information society is very important for keeping our data as safe as possible. The aim of authentication is to decide whether some subject is really claimed subject. There are 3 types of authentication: authentication by knowledge, authentication by ownership of something, and authentication by attribute. Each one has both advantages and disadvantages. They can be combined to increase the security of our information as well. It is well known that password do not prove high security level. From previous works for example [7, 14].

The one possible way to increase security level of access to information systems is the combination of authentication by knowledge and authentication by attribute ie parallel usage of passwords and keystroke dynamics. Everyone has different keyboard typing [12]. It is similar to one's own signature.

In keystroke dynamics it is possible to recognize various kinds of identifiably characteristics which are measurable: duration times (times between key press and key release of the same key), latency times (times between key release of the first key and key press of the second key), key typing speed, position of the finger on the key, pressure on the key and so on.

A template records the specific rhythm and touch, which is used to verify the user's identity at next logon. Research on keystroke dynamics—the study of individual typing patterns—shows that an individual's typing style is unique. That means that even if someone knows your password, he won't be able to authentificate as you.

The most identifiable characteristics for this biometric authentication system appear to be keystroke duration and latency times because no special hardware is necessary.

But using of keystroke dynamics together with passwords entails trouble. Passwords are usually very short in practice. In spite of all recommendations, and they provide very little amount of identifiable characteristics [7]. For this reason, it is necessary to find a good algorithm for user and impostor recognition. This means to find an algorithm which minimizes potential errors.

2 Related works

Begining of keystroke dynamics is dated in 19th century when telegraph operators where able to recognize each other only on their keying dynamics [15]. Using of keystroke dynamics in authentication was suggested by Spillan in 1975 [12]. However, the first scientific study in this field was carried out till in 1980 by Gaines and his research group [5]. In this research was studied keyboard typing of 300 – 400 word's long text by 7 secretaries. Conclusion of this work is proved statistical dependence of times of big-rams typing by the same user. This experiment had a lot of imperfections, especially very small number of testing persons. In thr 90's a lot of researches continued in Gaines work [10, 11, 15]. They were oriented to long texts too.

Radical changes were in Garica's patent of 1986 [6]. His idea was that the best data for this authentication is an individual username. The username is several times written and by means is created template of latency times. When a person wants to access computer resource, he is required to type his name. The latency vector of the keystroke times is compared with template stored in the computer. If this claimant's latency vector and template are similar, the user is granted access to the system. As a criterion of similarity he used so called Mahalanobis distance function.

The next important approach was patented by Young and Hammon [16]. They are using not only keystroke latency, but keystroke latency and duration at the same time. As criterion they selected Euclid distance function d_E :

$$d_E(\vec{x}_1,\vec{x}_2) = \sqrt{(\vec{x}_1 - \vec{x}_2)^T \times (\vec{x}_1 - \vec{x}_2)} \qquad (1)$$

$\vec{x}_1, \vec{x}_2$ .... N-dimension vectors

After the success of this work, a lot of work followed. The most important and the most cited are Joice [8] and Bleha [2] researches. Joice had suggested using user's name, surname and password together as data for this biometric authentication. As a criterion of similarity between template and access vectors, he recommended sum of absolute differences between template and access times d_A:

$$d_A(\vec{x}_1,\vec{x}_2) = \sum_i \left| x_{1,i} - x_{2,i} \right| \qquad (2)$$

$\vec{x}_1, \vec{x}_2$ .... N-dimension vectors

$x_{1,i}$ .... i-th element of the vector $\vec{x}_1$

$x_{2,i}$ .... i-th element of the vector $\vec{x}_2$

Bleha had chosen different metrics of similarity. He suggested applying of two different metrics. The first one was normalized distance function d_N, the second one was normalized Bayes classificator d_B.

$$d_N(\vec{x}_1,\vec{x}_2) = \frac{(\vec{x}_1 - \vec{x}_2)^T \times (\vec{x}_1 - \vec{x}_2)}{\|\vec{x}_1\| \cdot \|\vec{x}_2\|} \qquad (3)$$

$\vec{x}_1, \vec{x}_2$ ....N-dimension vector

$\|\vec{x}_1\|, \|\vec{x}_2\|$ ....Norms of vectors are defined in following eq. (4)

$$\|\vec{x}\| = \sqrt{\vec{x}^T \times \vec{x}} \qquad (4)$$

$\vec{x}$ ....N-dimension vector

As at present, the American firm Bionet Systems offers BioPassword. BioPassword is patented authentication software based on keystroke dynamics with undisclosed algorithm [17]. However, Bragg [3] after her own experience gave following result of BioPassword. She says: "In my test, logon worked as documented. I could keep that post-it note with my password on the monitor, but as long as I logged off, no one could log on as me. Similarly, I couldn't logon if I purposefully changed my typing style."

3 Our fuzzy approach

The first idea of this approach, as we know, was published by Capek in 2004 [4]. This idea is seemingly simply – to create a fuzzy inference system for every template. Suppose the subject whitch is going through authentication is posing as a subject S_j and he is presenting n identifiable features $\{i_i\}_{i=1}^{n}$. Then the fuzzy inference system for j-th template is defined as:

$$if\ i_1\ is\ I_1\ and\ i_2\ is\ I_2\ and \ldots i_n\ is\ I_n\ than\ access \qquad (5)$$

A task is to find suitable membership functions of $I_1, I_2 \ldots I_n$ for every template. It means to find an algorithm capable of creating this membership functions only from template data.

Certainly, the center of a membership function is localized at a mean of the concrete feature $\bar{x}_{i,j}$. However, some features are more typical for j-th user than others. Suitable characteristic of "typicality" of some features is his standard deviation $s_{i,j}$. Thus, possible membership function $\mu(x_{i,j})$ of i-th feature of j-th template can look as follows:

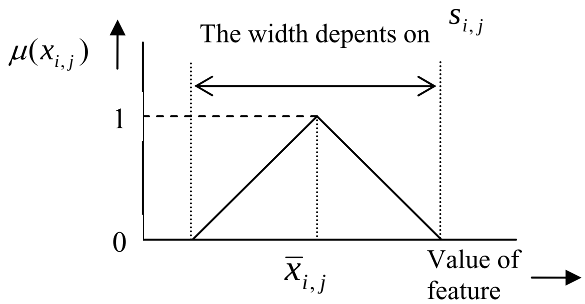

Fig. 1 A possible membership function for $I_{i,j}$ (6)

$$\mu(i,j)=\begin{cases}0 & x\le \bar{x}_{i,j}-2\cdot\sqrt{6}\cdot s_{i,j}\\ \dfrac{x-\bar{x}_{i,j}}{2\cdot\sqrt{6}\cdot s_{i,j}}+1 & \bar{x}_{i,j}-2\cdot\sqrt{6}\cdot s_{i,j}<x\le\bar{x}_{i,j}\\ \dfrac{\bar{x}_{i,j}-x}{2\cdot\sqrt{6}\cdot s_{i,j}}+1 & \bar{x}_{i,j}<x<\bar{x}_{i,j}+2\cdot\sqrt{6}\cdot s_{i,j}\\ 0 & x\ge\bar{x}_{i,j}+2\cdot\sqrt{6}\cdot s_{i,j}\end{cases} \tag{6}$$

$$\bar{x}_{i,j}=\frac{1}{n_j}\cdot\sum_{k=1}^{n_j}x_{i,j,k} \tag{7}$$

$x_{i,j,k}$ ....k-th value of i-th feature during creating of j-th template

n_j ....Number of measures for creating of j-th template

$$s_{i,j}=\sqrt{\frac{1}{n_j-1}\cdot\sum_{k=1}^{n_j}(x_{i,j,k}-\bar{x}_{i,j})^2} \tag{8}$$

n_j ....Number of measures for creating of j-th template

$x_{i,j,k}$ ....k-th value of i-th feature during creating of j-th template

$\bar{x}_{i,j}$ ....Mean of i-th feature of j-th template

4 Experiments

For purpose of testing of our approach there has been created a special software which is able to measure latency and duration times during typing appointed text works with precision microseconds. This software consists of two parts – client part (programmed in Java) which measures relevant times and server part (programmed in PHP) which saves measured times into database. Three different passwords have been selected: “biometrika”, “informatika” and “koleje” (“biometrics”, “informatics” and “hostel” in English) for testing. These words have been selected for their different lengths. Students with age range 19-25 were asked to write these words fifteen times. For precision it is necessary to add that chosen students are common users of computer and they have been familiarized of this experiment. The next week they were asked to write this words one time and the next week again. Measured data has been continuously saved into database for the future processing. The following number of students has complied to the criterion of fifteen measures for template and consecutive two measures for an access:

- 80 students for the word “koleje”
- 42 students for the word “biometrika”
- 45 students for the word “informatika”

Individual accesses were tested both against student’s own template and against other student’s templates. It has been obtained by the following number of valid accesses N_V and number of impostor accesses N_I:

$$N_V = N_a \cdot N \tag{9}$$

$$N_I = N_a \cdot N \cdot (N-1) \tag{10}$$

N_a .... Number of access by one person

N .... Number of templates

The following table describes numbers of valid accesses and impostors for the selected words:

Table 1 Numbers of users and impostors

Word	Valid accesses	Impostor accesses
koleje	160	12640
biometrika	84	3444
informatika	90	3960

The goal of the authentication process is to decide whether a subject is claimed subject, in other words “if you are really you”. This decision may not always be valid. The possible situations are shown in the next table.

Table 2 Potential situations of the authentication decision

	Acceptance	Rejection
Valid access	Suitable situation	False rejection
Impostor access	False acceptance	Suitable situation

This is a reason why as criterion for comparison of our model and contemporary models we have chosen false acceptance ratio FAR and false rejection ratio FRR, which are defined by following formulas:

$$FAR = \frac{NFA}{NIA} \qquad (11)$$

$$FRR = \frac{NFR}{NVA} \qquad (12)$$

NFA ....Number of false acceptances
NIA ....Number of impostor accesses
NFR ....Number of false rejections
NVA ....Number of valid accesses

It is necessary to note that FAR is dependent on FRR and it is not possible to decrease FAR without increasing FRR. Reversing statement is valid too. Various membership functions for i-th template and j-th feature were tested. Testing was done both for duration and latency times and for duration together with latency times.

In the following graphs the negative exponentially-weighted smoothing is used [13].

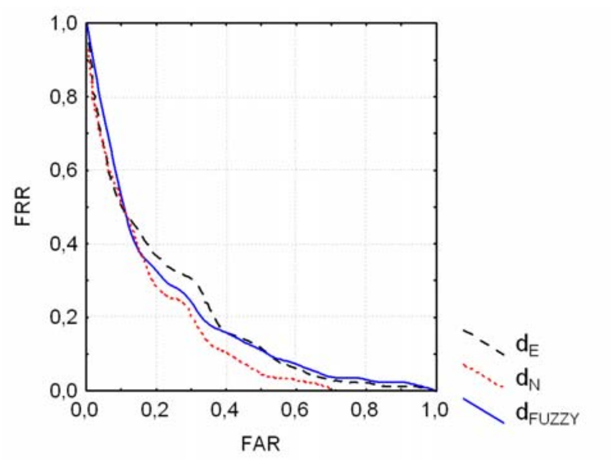

Fig. 2 Word "biometrika", duration times

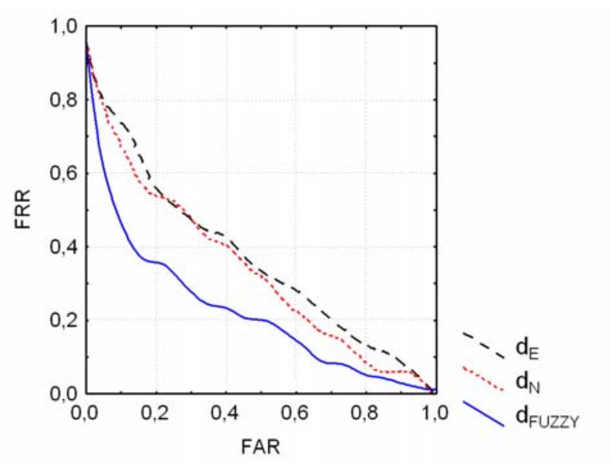

Fig. 3 Word "biometrika", latency times

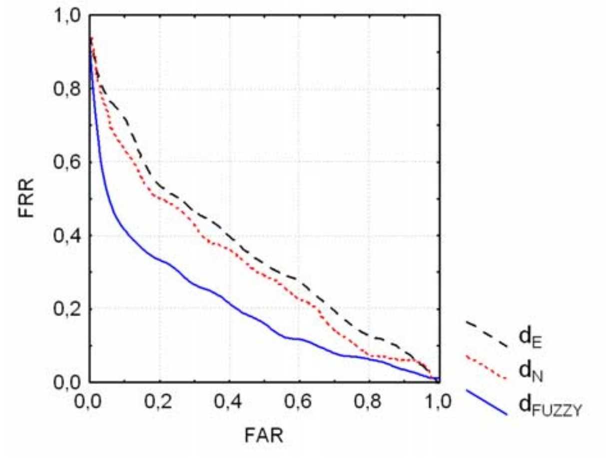

Fig. 4 Word "biometrika", both duration and latency times

Apperance of curves for both word “biometrika” and “koleje” is very simmilar to curves for the word “informatika”. For the lack of space we present only figures for using duration and latency times together.

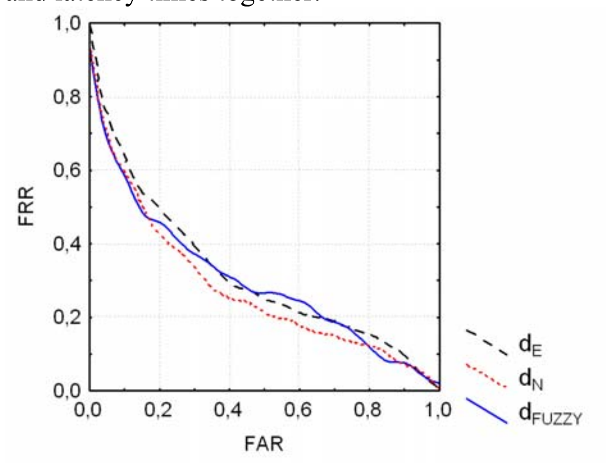

Fig. 5 Word "informatika", both dutation and latency times

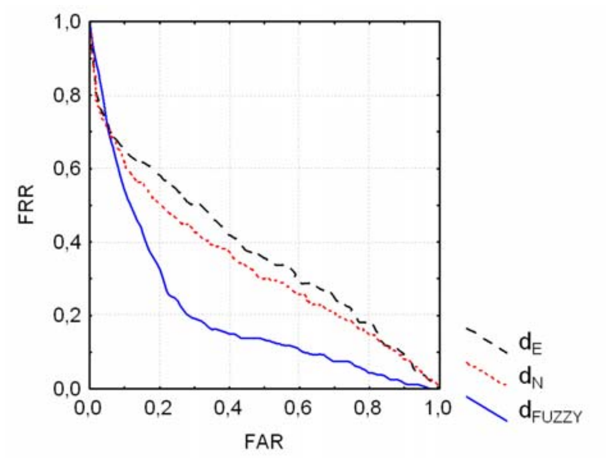

Fig. 6 Word "koleje", both duration and latency times

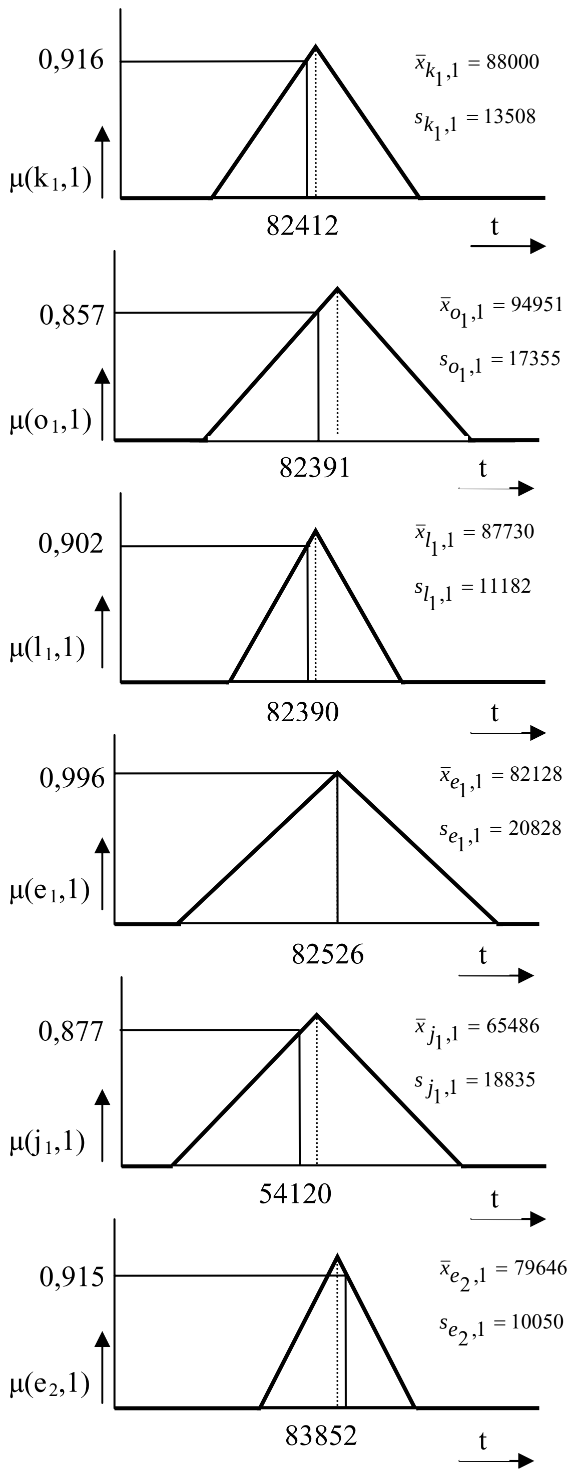

Fig. 7 Fuzzy numbers for accepted password “koleje”

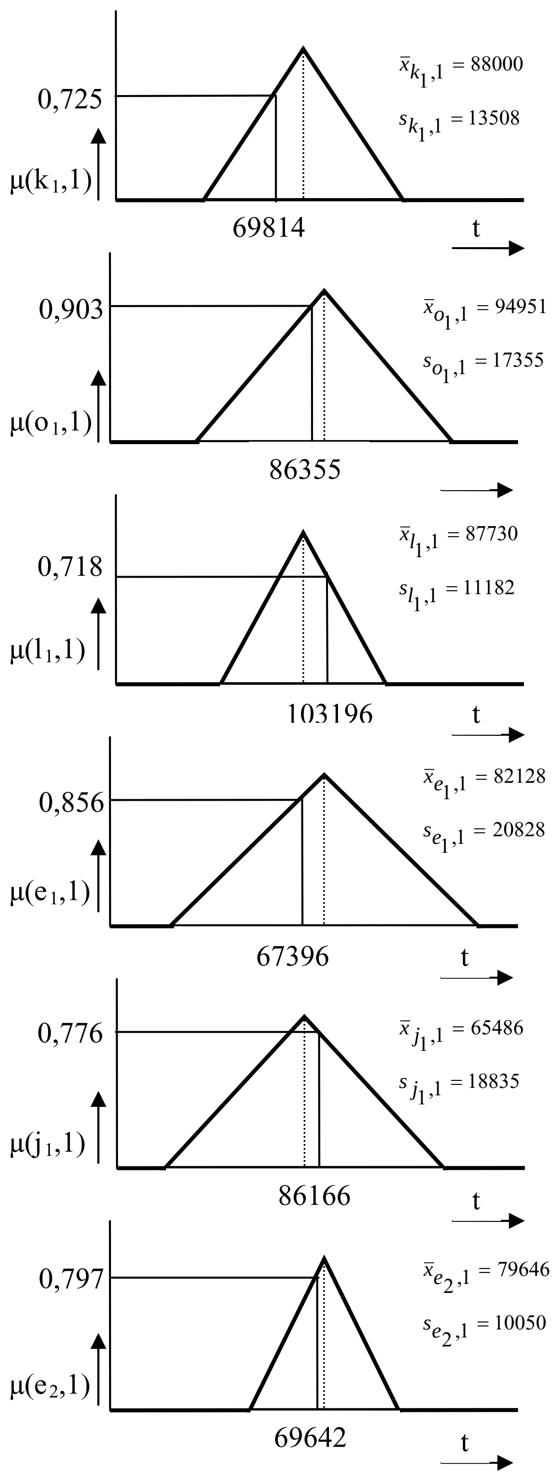

Fig. 8 Fuzzy numbers for rejected password “koleje”

For the last word “koleje” we show pictures (7) and (8) the individual fuzzy numbers constructions for both the accepted (valid) access and the rejected (impostor) access.

If we compare both figures (7) and (8), we can see that for letter o, both numbers are inside core of the fuzzy numbers.

This situation is for one letter only, so it is not enough for accepting both persons. Theoretically it is possible to create a table for how many alphabetic characters from password must be inside core of the fuzzy number for accepting user.

5 Conclusion and future work

It is evident from previous graphs fuzzy approach gives better results than conventional algorithms, based on the stochastic approach. This attitude offers a smaller danger of false acceptance errors and false rejection errors. A special hardware is not necessary for implementation of this algorithm and hence the costs will not increase.

References:

[1] Artazi, P., R., at all: Hand geometric and hand print texture based prototype for identity authentication. *WSEAS Transactions on Systems*. Issue 2 Vol.3 April 2004, pp 526-532, ISSN 1109-2777.

[2] Bleha, S., Slivinsky, CH., Hussein, B.: Computer-Access Security Systems Using Keystroke Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, No. 12, December 1990.

[3] Bragg, R.: *Biometric security products* [on-line]. [Cit. 24-10-2004]. Avaiable from: http://www.mcpmag.com/Features/article.asp?EditorialsID=270

[4] Capek, J: User identification by information system (original in Czech). *Scientific papers of the University of Pardubice* Ser. D. 21-25. Pardubice 2004. ISSN 1211-555X, ISBN 80-7194-716-4.

[5] Gaines, R., Lisowski, W., Press S., Shapiro, N.: *Authentication by keystroke timing: Some preliminary results.* Rand Report R-256-NFS. Rand Corporation, Santa Monica, CA. 1980

[6] Garcia J.: *Personal identification apparatus*. Patent Number 4.621.334. U.S. Patent and Trademark Office. Washington, D.C., 1986

[7] Hub, M.: *Knowledge security authentication* (original in Czech) In. Proceedings of the 3rd International PhD students conference Participation of PhD students on research programmes pp. 53-57. Bratislava 2003, Slovakia, ISBN 80-225-1700-3.

[8] Joice, R., Gupta, G.: *Identity Authentication Based on Keystroke Latencies*. Technical report #5, Department of Computer Science, James Cook University, Australia. 1989.

[9] Kosinski, W.: On normed space of ordered fuzzy numbers. *WSEAS Transactions on Systems*. Issue 2 Vol.3 April 2004, pp 900-905, ISSN 1109-2777.

[10] Legget, J., Williams, G., Umphress, D.: *Verification of user identity via keyboard characteristics*. J.M. Carey, Ed. Ablex Publishing, Norwood, NJ., 1986

[11] Legget, J., Williams, G.: Verifying identity via keyboard characteristics. *International Journal of Man-Machine Studies* 23. 1 (Jan. 1988). 67-76 .

[12] Legget. J, Williams, G., Usink, M.: Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, v36, s. 859-870, Sept. 1990

[13] McLain, D. H. (1974). Drawing contours from arbitrary data points. *The Computer Journal*, 17, 318-324.

[14] Stallings, W.: *Network security essentials: Applications and standards*. Prentice Hall 2000 ISBN 0-13-016093-8

[15] Umphress, D, Williams, G.: Identity verification through keyboard characteristics. *International Journal of Man-Machine Studies* 23, 3 (Sept. 1985). 263-273

[16] Young, J., Hammon, R.W.: *Method and apparatus for verifyng an individual's identity*. Patent Number 4.805.222. U.S. Patent and Trademark Office. Washington, D.C., 1989.

[17] Zilberman, A. G.: *Security method and apparatus empoying authentication by keystroke dynamics* (1998) United States Patent 6.442.692