ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2004 (995-1005)
VOLUME : 4
NUMBER : 1

USER IDENTIFICATION AT LOGON VIA KEYSTROKE DYNAMICS

İbrahim SOĞUKPINAR[1] **Levent YALÇIN**[2]

[1] *Gebze Institute of Technology, Dept. of Computer Engineering, 41400 Gebze-Kocaeli-Turkey*
[2] *Turkish Navy, Golcuk Naval Shipyard, Dept. of Planning and Design, Golcuk-Kocaeli-Turkey*

[1]E-mail: ispinar@bilmuh.gyte.edu.tr [2]E-mail: levyalcin@hotmail.com

ABSTRACT

Main goal of information system security is to protect the information by preventing unauthorized users to log on system and provide to log on system only authorized users. For this goal, it is necessary to distinguish the authorized and unauthorized users. In this study, a security method, which doesn't require an extra hardware, was designed according to this goal. Also this system makes difficult to system logon even if the password is known by the hypothesis of distinguishing users according to their typing rhythms is possible.

Keywords: *Computer Security, Keystroke Dynamics, Keystroke Latency, Keystroke Duration, Typing Rhythm, Biometrics*

1. INTRODUCTION

Identification of computer users is important problem for authentication during the system logon. Conventional systems use the user-identity and password to recognize the authorized system users. But these methods do not provide enough security due to password weakness. Therefore biometric features of users are used to recognize the users to provide right users access.

Keystroke recognition systems are designed for recognizing the users continuously or during system logon. It is possible to design keystroke recognition systems in both ways but in this study it is aimed to distinguish the users only at logon. These systems are based on distinguishing of users according to their keystroke pattern features. These features are keystroke latency or flight time, which represents the time between successive keystrokes and keystroke duration or dwell time that represents length of time keys are depressed.

Biometric systems have some errors even if uses of these systems are expected widespread in near future. Because there are two important kinds of errors that biometric systems do. These are False Acceptance Error (FAR) and False Rejection Error (FRR). FRR that the authorized user is rejected and FAR that the unauthorized user is accepted are inversely proportional. It must be expressed that the main goal of biometric systems is to decrease FAR as much as possible unless FRR is very high or in other words recently used metric for the evaluation of biometric systems is FRR when FAR =0.

Many researchers have proposed to use the duration of keystrokes and latencies between keystrokes as a biometric for user authentication. The first researches for proving of everyone's typing rhythm is different, started lately in 70's and in 1980 R.Stockton Gaines, William Lisowski, S.James Press, Norman Shapiro on behalf of Rand Corporation sponsored by National Science Foundation carried out some

Received Date : 25.4.2003
Accepted Date:27.10.2003

experiments with seven secretaries [1]. In this study, secretaries was asked to type three kinds of texts, which every text is about one page length with the interval of four month. Only the diagraphs was investigated which repeats more than 10 or more and T-Test statistics method was used by the hypothesis of the mean and variance of diagraph times is the same in the studies conducted with the interval of four month.

In 1989, Minimum Distance (Euclidean) and Mahalanobis Distance (Bayes) classification algorithms were tested by Saleh Bleha, Charles Slivinsky and Bassam Hussein and two kinds of password mechanisms which is the names of people or fixed phrases was used. [2, 3,] Fixed phrases in identification phase and names of the people in the verification and overall recognition phase was used and in identification phase 10 users was tested and % 1.2 made a decision wrong. In verification phase, 26 volunteer user was tested and %8.1 FAR and % 2.8 FRR, in overall recognition phase with 32 volunteer user %3.1 FAR and % 0.5 FRR was achieved.

Between 1985 and 1989, similar experiments to the studies of R.Stockton Gaines [1] was conducted by Glen Williams, John Leggett and David Umphress [4., 5, 6]. In first study conducted by 17 programmers with the varying typing speeds it was necessary for users to type approximately 1400 words in training phase and 300 words in test phase for the recognition of the users by the system. In the second experiment [5] 36 participants typed a 537 characters passage at two different times separated by over a month and diagraph latencies was stored in a 26 x 26 reference latency matrix whose rows correspond to the first letter of a diagraph and columns correspond to the second letter. In this study the users was accepted if the test signature (for every diagraph) is within the 0.5 standard deviation of mean diagraph time and the similarity with the reference signature is % 60 or more. In 1989 John Leggett, Glen Williams and David Umphress investigated 12 different diagraph latencies and used 300,500,750 msec filters for the recognition of users [6]. But more than 1000 characters were needed for the registration of the users to the system and FRR was very high. So a practicable system was not being put forward.

In 1990, one of the promising researches about keystroke dynamics realized by Rick Joyce and Gopal Gupta. [7] Even though the algorithm used is very simple, very impressive results were obtained. In this study, for the registration of the users to the system, users were asked to type their names, surnames, usernames and password 8 times. From these entries mean reference vector was extracted and then reference vector and test vector was compared. If the difference between reference vector and test vector is under a predefined threshold value, user was accepted, otherwise rejected. In this study % 0.25 FAR and % 16.36 FRR was achieved.

In 1997, taking into the consideration keystroke duration in addition to keystroke latencies and a larger set of users of varying ages, nationalities, background and examining the effect of using non-structured free texts, research spectrum was tried to be extended by Fabian Monrose and Aviel D. Rubin [8]. Clustering according to the typing speed of users was used in this study and with Euclidian distance, non-weighted probability, weighted probability and Bayesian classification algorithms users was recognized with the rates of %83.22, % 85.63, %87.13 and % 92.14 respectively. In the non-weighted probability, an algorithm was used by the hypothesis of the distribution is Gaussian distribution and merging K-Nearest Neighbour algorithm. In weighted probability, since some diagraphs is more frequently used in English, calculations was made by taking into account the weights of diagraphs.

In 1997, Dawn Song, Peter Venable, Adrian Perrig tried to verify the identity of users not only at logon but also continuously, contrary to other studies. [9] In this study, it was determined that histograms of the diagraph times is similar and very close to Gaussian distribution and tried to be distinguished by an algorithm similar to Markov chain which uses the mean and variance of diagraph times.

In 2000, users were tried to be recognized and distinguished with the neural network approach by Sungzoon Cho, Chigeun Han, Dae Hee Han, Hyung_II Kim [10]. While the back-propagation models used yield favorable performance results on small databases, neural networks have a fundamental limitation in that each time a new user is introduced to the database, the network must be retrained. For applications such as access control, the training requirements are prohibitively expensive and time consuming.

Furthermore, in situations where there is a high turnover of users, the down time can be significant.

In the study of A.Peacock [11], users were asked to type name, surname, password and fixed phrase (ionlyread) 20 times and recognized with % 92 success rate (% 8 FAR) by the modified K-Nearest Neighbour (K-NN) algorithm. % 4.2 FRR (4 out of 95) was achieved by testing 11 users and this application developed as Web-based. All successful imposter attempts were made by the same user in this study.

In 2000, in the study conducted by Aykut Guven and Ibrahim Sogukpinar an algorithm based on calculating the angle between vectors was used and developed by the hypothesis of keystroke latency and keystroke duration effectiveness is % 90 and % 10 respectively [12].

The studies conducted by us showed that especially in expert typists keystroke latencies might be shorter than keystroke durations and the main distinctive feature is keystroke latencies not keystroke durations. Furthermore, especially keystroke durations might be fluctuating and depending upon keystroke durations more weighty can cause system unstable. But giving a specific weight for keystroke latency (% 90) and keystroke duration (% 10) must be determined after a long period of experiments. Furthermore, it is stated in this study that it is necessary to continue training entrances until the program informs the user that he/she was recognized. On the other hand, the threshold that is used for the recognition of users and password length was used as constant and no result was declared about FAR in the experimental studies.

In this study, the mean and standard deviation of the keystroke latencies was calculated and the best threshold for the recognition was tried to be determined with the experimental studies between 2 and 3 standard deviation of the mean. This paper is organizes in five sections. Introduction and related studies are explained in section 1. Some different mathematical models of key press recognition are introduced in section 2. Then the model that was developed in this study is introduced in section 3. Experimental results and comparisons with another algorithm are given in section 4. Last section includes the ideas about results and future studies.

2. MATHEMATICAL MODELS OF THE KEYSTROKE DYNAMICS BASED ALGORITHMS

In this section, since mathematical models was not expressed explicitly in some papers or was explained more detailed with the sophisticated formulas, the algorithms and mathematical models of the previous studies was tried to be explained as much as possible according to the expression in their papers.

In 1989, Minimum-Distance (Euclidean) and Bayes (Mahalanobis) Classification Algorithms were applied by S.Bleha, C.Slivinsky and B.Hussein [3] and in the verification stage.

$$D_i(X) = \frac{(X - m_i)^t (X - m_i)}{\|X\| \|m_i\|} < T_1 \qquad (1)$$

For the normalized Minimum- Distance Classification stage, verification was made calculating the distance between test vector and mean vector from (1) and (2).

In normalized Bayes classification

$$d_i(X) = \frac{(X - m_i)^t C_i^{-1} (X - m_i)}{\|X\| \|m_i\|} < T_2 \quad (2)$$

where,

X	: Test Vector
m_i	: Average Vector
C_i	: Covariance Matrix
T_1 ,T_2	: Threshold Level

At this study, each user was forced to make 30 entries in order to determine user average vector and in Minimum-Distance classification, 0.030 was used as Threshold Level at first trial, 0.029 at second trial, in Bayes classification 0.000030 and 0.000029 values were used respectively and if the distance between test vector and average vector is little than threshold levels stated above, it was decided that test trial was real or belonged to authorized user. If it is paid attention to this study, threshold level was determined as constant.

At the studies carried out by F. Monrose and A.Rubin [8, 13] in 1997 and 2000, the trials were made according to Eucludian distance measure, non-weighted probability, weighted probability and Bayes classification algorithms respectively. In Euclidean distance measure algorithm

$$D(R,U)=\left[\sum_{i=1}^{N}(r_i - u_i)^2\right]^{1/2} \qquad (3)$$

In non-weighted probability algorithm

$$Score(R,U)=\sum_{i=1}^{N}S_{u_i} \qquad (4)$$

where,

$$S_{u_i}=\frac{1}{o_{u_i}}\left[\sum_{j=1}^{o_{u_i}}\Pr ob(\frac{X_{ij}^{(u)}-\mu_{r_i}}{\sigma_{r_i}})\right] \qquad (5)$$

In weighted probability algorithm

$$Score(R,U)=\sum_{i=1}^{N}(S_{u_i} * weight_{u_i}) \qquad (6)$$

$$weight \quad t_{u_i}=\begin{cases} 0 & \text{if } U_i \text{ or } R_i \text{ is empty} \\ \dfrac{o_{u_i}}{\sum_{k=1}^{N}(o_{u_k})} & \text{Other} \end{cases} \qquad (7)$$

In Bayes Classification Algorithm

$$\Delta^{\alpha}(x,x')=\sum_{i=1}^{n}w_i(\frac{|x_i - x_i'|}{\sigma_i})^{\alpha} \qquad (8)$$

Formulas used where

U	:	Test Vector
R	:	Reference (Average) Vector
$X_{ij}^{(u)}$	:	j th repetition of i th property of U test vector
o_{u_i}	:	Repetition Number
x_i	:	Property Vector
w_i	:	Weight Vector

Each member of vectors consisted of mean, standard deviation, repetition number and quantity values. After the calculations were made according to the non-weighted and weighted probability algorithms, they were related with the nearest neighbor in the database. In Bayes Classification, property vectors were calculated with Factor Analysis (FA). For all algorithms the studies were made with the hypothesis of that distributions seems like normal (Gaussian) distribution.

Another approach that has developed by R.Joyce and G.Gupta is based on the comparing two signatures via computing l_1 norm [7]. In this method, each vector consists of the set of four latency values. The mean reference signature vector, M includes the following values;

$$M = \{ M_{\text{username}}, M_{\text{password}}, M_{\text{firstname}}, M_{\text{lastname}} \} \quad (9)$$

M reference vector is compared with the T test vector via calculating the magnitude of difference between them. For discussion the method, let M={m_1, m_2,......., m_n} and T={t_1,t_2,.....,t_n }, where n is the total number of latencies in the signature. The present verifier computes the magnitude of the difference between M and T as the l_1 norm;

$$\|M\text{-}T\|_1 \qquad (10)$$

Given by;

$$\sum_{i=1}^{i=n}| m_i - t_i | \qquad (11)$$

Training signature is obtained as follows. A measure of variation is computed 8 reference signatures, and the mean reference signature obtained from them. The 8 training signatures are, S_1, S_2,............ S_n . The value of $\|M\text{-} S_i\|_1$ is calculated for i=1 to 8. A threshold value is found between a given T and M by using the mean and standard deviation of these norms.

The verification algorithm works as follows. Test signature T, is obtained while login attempts. The norm $\|M\text{-}T\|_1$ is computed and if this norm is less than the threshold for the user. Otherwise, system decides that the attempts is imposter [7]

There are two parameters as an input considered by the system at the study [12] that is made by Aykut GÜVEN in 2000. These are B(I) array that is stated as keystroke hold time and G(I) array, which is stated as keystroke delay time. Other parameters and functions that are needed by the system's other stages are produced by these two arrays.

The below notations were used at the system design when mathematical model was made.

B(i) : The hold time for I th keystroke is pressed.

G(i) : The delay time that passes from (I-1) th key release to I th key is pressed.

θb : The angle difference between reference B vector and measured B vector.

θg : The angle difference between reference G vector and measured G vector.
Γb : The weight factor for B vector.
Γg :The weight factor for G vector
ξ : The decision function
Ψb : B vector
Φg : G vector
Ω : The weight factor for ξ

Supposing Bi is one of the dimensions, which is perpendicular and orthogonal to each other in N Dimension space, it is being defined that N, which is the length of the password, is equal to N dimension space and in other words each character of the password describes a dimension of the N dimension space. B and G arrays are defined by the parameters, which are produced by the system user's logon,

$$B = B(1)+B(2)+\ldots\ldots\ldots+B(N) \qquad (12)$$

$$G = (1)+G(2)+\ldots\ldots+G(N-1) \qquad (13)$$

Each member of B array, which is defined above, produces an n dimension vector.

$\Psi b = \Sigma B(i) Ii$; N dimension Ψb vector(14)

$\Phi g = \Sigma G(j) Ij$; N-1 dimension Φg vector (15)

Ψb , Φg vectors are produced by the system in real time. At the same time two same vectors that are used by the system as a reference and produced by the system ago are stated like that:

$\Psi br = \Sigma Br(i) Ii$; N dimension Ψbr vector (Reference Vector) (16)

$\Phi gr = \Sigma Gr(j) Ij$; N-1 dimension Φgr vector (Reference Vector) (17)

A and B are the similar vectors in the same dimension; angle between two vector in n dimension space is
$\mathrm{Cos}\ \theta = AB / [(\Sigma A^2)(\Sigma B^2)]^{1/2}$ stated like that and (18)
$\mathrm{Cos}\theta b = \Psi b \Psi br / [(\Sigma \Psi b^2)(\Sigma \Psi br^2)]^{1/2}$ (19)
$\mathrm{Cos}\theta g = \Phi g \Phi gr / [(\Sigma \Phi g^2)(\Sigma \Phi gr^2)]^{1/2}$ (20)
Equations (19) and (20) are derived from the above equations. It is stated that Cosθb's value approaches to 1 when Ψb and Ψbr vectors seems like each other and Cosθb's theoretical value is 1.

By stating the same situation is true for Φg and Φgr vectors and weight factor for Cosθb value is Γb ; weight factor for Cosθg value is Γg . Relation between Γb and Γg terms are defined like below
$\Gamma b + \Gamma g = 1$ (21)
$\Gamma g = 1 - \Gamma b$ (22)
and ξ decision function are derived from above expressions.

$$\xi = \Gamma g \mathrm{Cos}\theta g + \Gamma b\ \mathrm{Cos}\theta b \qquad (23)$$

where i denotes the situation of system in any time and $\xi i\text{-}1$ is the decision function that belongs to entrance before i th situation taken from the database and
Ω is weight factor between 0 and 1 ($0 < \Omega < 1$)

$$\xi i-1 - (\Omega\ \xi i-1) < \xi i < \xi i-1 + (\Omega\ \xi i-1) \qquad (24)$$

Equation (24) is derived from above formulas and if the value is between the interval in (24) the user is accepted to the biometric system.

3. MATHEMATICAL MODEL OF THE DESIGNED SYSTEM AND ALGORITHM

In this study, the mean and standard deviation of the keystroke latencies was calculated and the best threshold for the recognition was tried to be determined with the experimental studies between 2 and 3 standard deviation of the mean.

The origin of the our algorithm is based on the fact that the probabilities of being within 1, 2 and 3 standard deviation of the mean of the samples are % 68.26, % 95.4 and %99.7 respectively. Likewise in the studies realized up to now, it was determined that when the size of the samples is big enough, distribution of the keystroke latencies is very close to Gaussian Distribution [9, 14] and even a series of 15-20 characters is enough for the distinction of the users.

The mathematical formula of the system can be defined as follows.
Let $i = 1,2, \ldots.., n$ and $j = 1,2, \ldots., t$

where n is the number of characters and t is number of trial entry for the training phase.
Then,

$$m_i = \frac{\sum_{i=1}^{n-1} R_{i_j}}{n-1} \qquad (25)$$

$$\sigma_i = (\frac{1}{n}\sum_{i=1}^{n-1}(R_i - m_i)^2)^{1/2} \qquad (26)$$

$$m_i - SL * \sigma_i \leq X_i \leq m_i + SL * \sigma_i \qquad (27)$$

Where;

X_i = Test Vector
R_i = Reference Vector
m_i = Mean Vector
σ_i = Standard Deviation
SL = Security Level

According to the assumption of us, for every attempt the latency time for some diagraphs (especially some diagraphs in the name and surname of every person) in many of the typists (not unexperienced typists) is very close to each other and this is the main distinctive characteristic of a people's keystroke rhythm. That is, the standard deviation of some diagraph latencies for most of the people is very small value. The algorithm does not permit the unauthorized users who cannot imitate these diagraphs especially. Let's try to explain this by giving an example. the values and graphics for a user are given in Table -1 and Figure 1.

If it is paid attention to the table-1, in spite of the small size of data it can easily understood that the distribution of diagraph latency times is very close to Gaussian Distribution. Besides this, if the bold cells is taken into account, one can understood that 4 bold cells of the cells among the 5 bold cells (one in name, two in surname, two in password) is not between + 2 and –2 standard deviation of the mean. On the other hand, it can be easily seen that 4 of the cells is
It is necessary for the unauthorized users to adjust their keystroke rhythms to be in the band of unauthorized users for every diagraph as shown in Figure1 to be able to enter the system instead of the unauthorized users. It is very little possibility for the unauthorized users to imitate especially the bold diagraphs, which is shown in the above paragraphs.

Table 1. The Diagraph Latency Values for a Person after the Training Phase

NAME	**l**	**e**	**v**	**e**	**n**	**t**
1. Attempt	1	0	1	2	2	
2. Attempt	2	1	0	1	3	
3. Attempt	1	0	2	0	3	
4. Attempt	1	1	0	1	3	
5. Attempt	1	0	0	1	2	
6. Attempt	2	0	0	2	2	
7. Attempt	1	0	0	1	3	
8. Attempt	0	1	0	2	1	
MEAN	1,125	0,375	0,375	1,25	2,375	
ST DEV	0,64	0,517	0,744	0,707	0,744	
MEAN + 2*ST	2,305	1,409	1,863	2,664	3,863	
MEAN - 2*ST	-0,15	-0,65	-1,11	-0,16	0,887	
SURNAME	**y**	**a**	**l**	**c**	**i**	**n**
1. Attempt	1	5	2	2	4	
2. Attempt	2	4	1	2	2	
3. Attempt	1	5	1	4	3	
4. Attempt	1	4	1	3	3	
5. Attempt	1	3	1	3	2	
6. Attempt	1	4	1	3	1	
7. Attempt	1	4	1	4	3	
8. Attempt	1	4	1	3	2	
MEAN	1.125	4,125	1.125	3	2,5	
ST DEV	0,353	0,64	0,353	0,755	0,925	
MEAN + 2*ST	1,831	5,405	1,831	4,51	4,35	
MEAN - 2*ST	0,419	2,845	0,419	1,49	0,65	

PASSWORD l	e	v	y	a	l	c	i	n
1. Attempt	1	3	2	2	6	1	4	3
2. Attempt	1	4	1	2	5	2	3	3
3. Attempt	0	3	1	3	4	2	3	3
4. Attempt	1	4	2	2	6	1	4	2
5. Attempt	1	4	2	2	6	2	4	2
6. Attempt	1	3	2	2	5	1	2	2
7. Attempt	1	4	2	2	5	1	4	2
8. Attempt	1	3	2	2	6	1	4	2
MEAN	0,875	3,5	1,75	2,125	5,375	1,375	3,5	2,375
ST DEV	0,353	0,534	0,462	0,353	0,744	0,517	0,755	0,517
MEAN + 2*ST	1,581	4,568	2,674	2,831	6,863	2,409	5,1	3,409
MEAN - 2*ST	0,169	2,432	0,826	1,419	3,887	0,341	1,99	1,341

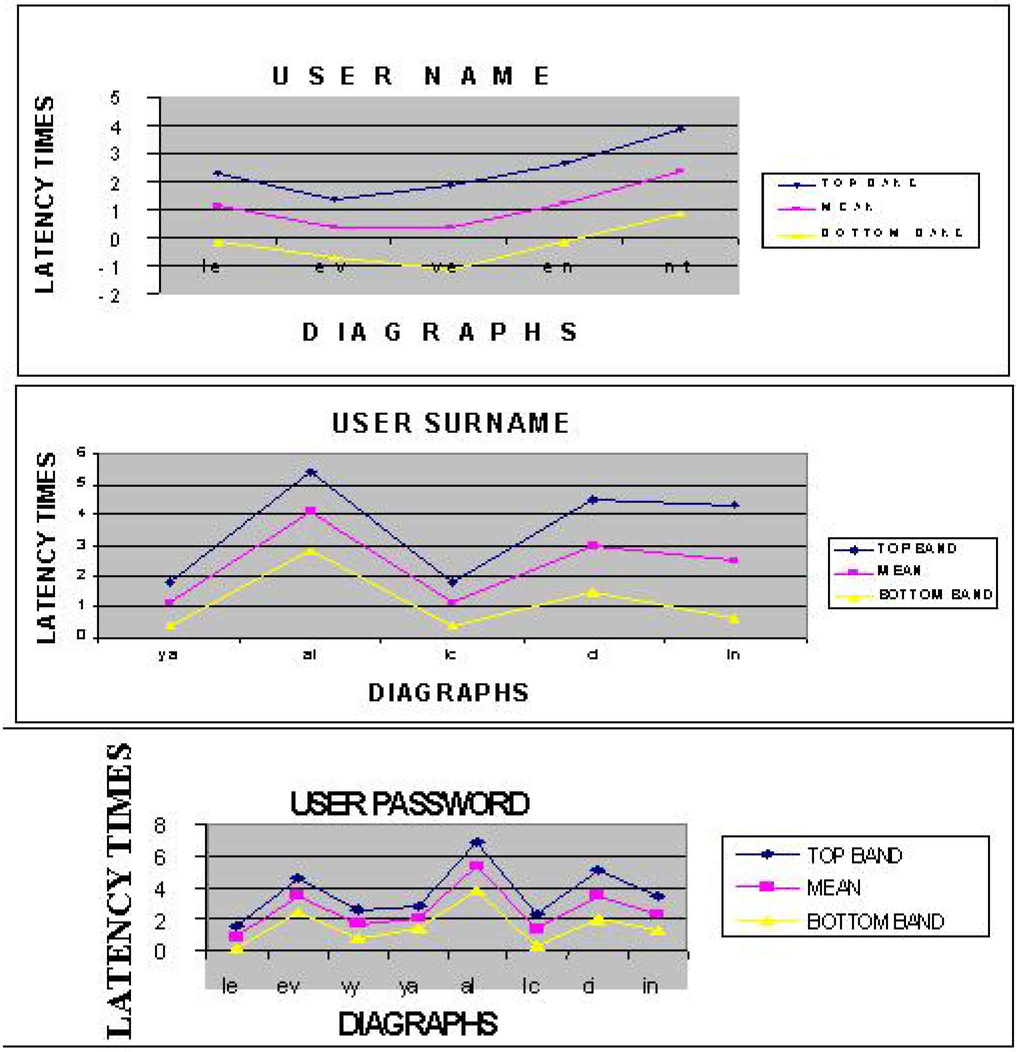

Figure 1. The Graphics of a User after Training Phase

In the Verification Phase, by the hypothesis of distribution of the keystroke latencies is similar to Gaussian Distribution and according to the security level, which can changeable by the system administrator the user is permitted to logon if the values for every diagraph in the test entrances are within the limits or interval of the formula **(27).**

If we try to describe the designed system according to the system phases **in the training phase** keystroke latency and duration times for every diagraph and their standard deviation and means is registered to the related tables.

In the Identification Phase, the name, surname and password of the users who wants to try to logon is compared with the related fields and if the entry information is same with the information registered system translates to the verification phase, otherwise the user is warned with a message which tells the user he/she is not defined in the system.

In the Verification Phase, decision is made according to standard deviation and the security or threshold level that is variable by system administrator contrary to most of previous work that is done according to the fixed threshold level for the distinction of the users. To assist to the system administrator in the determination of the best threshold level for the distinction of the users, the information such as date, time, success entry count, success entry percent, security level, system registration count, user account lock count and real name and surname of the person who tries to enter to the system for the correct determination of FAR and FRR is logged.

In the Update Phase, the user is accepted to the system according to the threshold level and if the success entry count reaches the system registration count or its multiples, training phase information are updated by the assumption of keystroke patterns of the users can change as the time passes.

The algorithm which is developed in this study is said to be implemented with different threshold level and seems to the algorithm which is implemented by John Leggett, Glen Williams and David Umphress [4., 5, 6] and Rick Joyce and Gupta Gopal [7]. The algorithm doesn't resemble the K-NN and Neural Network algorithms at all but it is said to resemble the Minimum Distance Algorithm and Mahalanobis Distance Algorithm since the information of mean and standard deviation are used in the Minimum Distance and Mahalanobis Distance Algorithms respectively. But on the other hand in Mahalanobis Distance Algorithm covariance matrix, in the Minimum Distance Algorithm fixed threshold level is considered in calculations.

To be considered good sides of the developed algorithm are the usage of the variable threshold level, free selection of the password length unless it is less than 6, having a small FTR (Failure to Enroll Rate) which is the ability of the biometric to enroll a biometric user or in other words rapid enrollment of the users to the system within in a few minutes, determining the feature vectors of the users more accurate by using the backspace and delete keys of the keyboard, updating the user feature vectors as the time passes and being a practicable system as the most important feature.

4. EXPERIMENTAL RESULTS AND THE COMMENTS

25 users that were evaluated as the representative of the whole sample space were introduced to the system. But totally 40 users (15 extra users together with 25 users) were used during the test entrances. It was tried to be chosen not only experienced computer users such as the secretaries in the office but also the users that are not very familiar with the computers.

It was reached to the best threshold level that is changeable by the users at 2,75 in the experiments that is done by the intervals 0,25 from 2 to 3. But since it is possible to change the threshold level with the intervals 0,05 by system administrator, the best threshold level may come true at another level between 2,5 and 3. Six of the users out of 40 total test users succeeded to enter to the system instead of 1 user from 25 users that are introduced to system. When this user was excluded from the system, the users could make successful entries at the rate of % 0,1 or in other words 0,001. It is evaluated that this situation appears to be come true since the name, surname and password of that person is not long enough and spreads out well arranged on the keyboard and have a password easily imitable or reproducable like 757575 and the most important one from not having a stable keystroke pattern or the reasons such as hesitation, carelessness or abstraction during the training entries that cause wrong typing pattern. When the security level was 2.75 the users could make successful entry at the rate of % 40 or in other words 4 in 10 entries. Comparisons with the previous studies are summarized in the Table 2 according to the references.

Table 2. Comparisons with the Previous Studies

Reference	Brief of Implemented Algorithm	Sample Size	FAR %	FRR %	Disadvantages/Limitations
[1]	T-Test	7	0	4	1. Small Sample Size 2. Being the first study related to subject
[3]	Minimum and Mahalanobis Distance Classification	32	3.1	0.5	Big FAR
[6]	Acceptance of the users If the test signature fall within 0.5 Standard Deviation of the mean reference diagraph Latency and if the comparison between the test signature and the mean reference diagraph Latency is more than % 60	36	5.5	5	1. Long Training Time 2. Big FAR 3. No Distinction of Keystroke Latency and Duration
[7]	Comparison of Test and Reference vectors according to a threshold level	33	0.25	16.36	
[10]	Neural Network	21	0	1	1. Long time for the enrollment 2. Continuous Retraining
[13]	Minimum Distance Classification	63	16.78	?	Big FAR
[13]	Non-weighted Probability /K-NN	63	14.37	?	Big FAR
[13]	Weighted Probability / K-NN	63	12.82	?	Big FAR
[13]	Bayes Classification	63	7.86	?	Big FAR
[11]	K-NN Classification	11	8	4.2	1. Small Sample Size 2. No update of user features
[12]	Calculation of angle between vectors	12	~1	?	1. Small Sample Size 2. Mandatoriness of 8 characters password 3. Long Training Time 4. Fixed Threshold Level 5. No result about FRR
This Study	Determination of whether test vector is between +/- 2.75 standard deviation of the mean of diagraph time for every diagraph or not	40	0.6	60	Big FRR * Results of This Study

If it is paid attention to the table, we reached the best FAR value with the exception of study made by S.Cho, C. Han, D.H. Han and H. Kim with the approach neural network [10]. As mentioned in section 1, in the applications implemented by neural network approach there are some limitations to put into practice. The bigness of FRR of our study can be tolerable and will not be system administrator adjusts a problem if user account lock counts to 4-6 since a user can enter the system in 2.5 attempts on the average.

5. CONCLUSION

Nowadays Biometric Security Systems develops continuously, intensive studies on it goes on and in a short time it seems to be a candidate for the systems widespread also in our country.

The greatest advantage of the Biometric Security Systems is the thing that each person use their individual features. So these features and their password become difficult to get.

The greatest disadvantages of the Biometric Security Systems need extra hardware and the result of this it has certain error rate in general with needing high cost for now. But nowadays it is possible to say that these error rates become better by means of new algorithms, which develops as a result of increasing studies, or by means of developing existing algorithms.

In this study, the distinguishing of users according to keyboard's writing rhythm which doesn't need extra hardware is done and very attractive results are obtained. Therefore its cost is lower than other bio-metric security systems. As a result, it is better than sound and signature recognition systems and also it is nearer to hand-geometry sensitivity and it is more usable than any aforementioned systems without needing any extra hardware.

The studies about subject continue more than 20 years and nowadays it still go on rapidly. At this study, with the parallel of the algorithms in literature, it doesn't seem to be as same as any algorithms; very attractive results are obtained from the algorithm, which is developed by inspiring from existing algorithms. In parallel with developed algorithms, not only the time between character pairs is considered but also it is evaluated 3 or 4 characters which are written by people and at the same time it is estimated to gain new dimension with making time analysis of the groups' of 3 or 4 character pairs.

The developed algorithm is estimated to be very well by making analysis about what the experimental studies are tested with the same text by the large groups and also it gives good results by means of the small improvement. Other advised subject is also that make people test with intervals instead of making them test with a long time in front of keyboard. So it is considered to affect correctness of data in positive way. Because, it is said that normal rhythm can change in positive or negative way after writing with the keyboard 5-10 minutes.

Other researchable subject which, appears from this study is that system manager is warned when the change happens on keystroke rhythm from PC/Terminal by recording keystroke time not only the user entries the system but also after the user entries the system. But as we stated earlier, this study needs more investigations, some researches about it are being made by firms and research groups and so on, and it is not declared that important improvements about this study are being done.

After that, at the related studies from the perspective of improvement it needs to research the effect how writing rhythm can change when the people are tiredness, sleeplessness, nervousness, absentmindedness etc and to research the physical conditions, (sitting shape, height and distance from the place where keyboard is placed, the way how the hands are put on keyboard), how make changes about pattern vector of people when make writings and to research that the person writes with left or right hand and it is useful to make studies about the subject that is not investigated so on and such applications can be designed web based and works with a lot operating systems.

This study is considered to be used as a extra security spread in addition to 128 bit encoding, which can be seen in password based verifying systems, banking, e-commerce etc on internet and also military systems that needs high priority security.

REFERENCES

[1] GAINES, R.S.& LISOWSKI, W. & PRESS. S. J.& SHAPIRO, N. "Authentication by Keystroke Timing: Some Preliminary Results", *Rand Report R-256-NSF, Rand Corporation.* 1980.

[2] BLEHA, S. "Recognition Systems Based on Keystroke Dynamics", *Ph.D. Dissertation, University of Missouri-Colombia,* 1988..

[3] BLEHA, S. & SLIVINSKY, C.& HUSSEIN, B. "Computer-Access Security Systems using Keystroke Dynamics". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol.12, No.12, 1990.

[4] UMPHRESS, D. & WILLIAMS, G. "Identity Verification Through Keyboard Characteristics" *International Journal of Man-Machine Studies,* 23(3): pp:263-273. ,1985.

[5] LEGGETT, J. & WILLIAMS, G. "Verifying Identity Keystroke Characteristics", *International Journal of Man-Machine Studies,* 28(1): pp:67-76, .1978.

[6] LEGGETT, J. & WILLIAMS, G. & UMPHRESS, D, "Verification of User Identity via Keystroke Characteristics", *Human Factors in Management Information Systems*. 1989,.

[7] JOYCE, R. & GUPTA, G. "Identity Authorization Based on Keystroke Latencies", *Communications of the ACM,* 33(2), pp:168-176, 1990.

[8] MONROSE, F. & RUBIN, A. D., "Authentication via Keystroke Dynamics", *Fourth ACM Conference on Computer and Communications Security*, pp:48-56, 1997.

[9] SONG, D. X. & WAGNER, D. & TIAN, X., "Timing Analysis of Keystrokes and Timing Attacks on SSH",*http://paris.cs.berkeley.edu/~dawnsong/papers/ssh-timing.pdf*, 2001.

[10] CHO, S. & HAN, C. & HAN, D. H. & KIM, H., "Web Based Keystroke Dynamics Identity Verification using Neural Network", *Journal of Organizational Computing and Electronic Commerce,* Vol. 10, No. 4, pp. 295-307, 2000.

[11] PEACOCK, A., "Learning User Keystroke Latency Patterns", (Preliminary Report), *http://pel.cs.byu.edu/~alen/personal/CourseWork/cs572/KeystrokePaper.ps*, 2000.

[12] GUVEN,A. & SOĞUKPINAR, İ. "Computer and Network Security System Design via Keystroke Dynamics",. BAS2000, Ankara., 2000.

[13] MONROSE, F. & RUBIN, A. D., "Keystroke Dynamics as a Biometric for Authentication", *Future Generation Computing Systems (FGCS) Journal: Security on the Web (special issue).* March 2000

[14] SONG, D. & VENABLE, P. & PERRIG, A. "User Recognition by Keystroke Latency Pattern Analysis", http://paris.cs.berkeley.edu/~perrig/projects/keystroke,1997.

Dr. İbrahim Soğukpınar. Received his B.Sc.degree in Electronic and Communications Engineering from Technical University of İstanbul in 1982,and his M.Sc. degree in Computer and Control Engineering from Technical University of İstanbul in 1985. He received his Ph.D. Degree in Computer and Control Eng. from Technical University of İstanbul in 1995. Currently he is the chairman of Computer Engineering Department in Gebze Institute of Technology. His interested areas are Networking, Information Systems Applications,Computer Vision and Information Security.

Levent YALÇIN: He was born in Turkey in 1968. After completing elemantary school in Ankara in 1982, he attended Naval High School in İstanbul. Later he graduated from Naval Academy as a Naval Officer in 1990. He worked in battleships for six years. Later he took a computer science training in Middle East Technical Univercity for one year. He is working as a Administrator of Data Processing Center of Naval Shipyard since 1997. During this period he took a Master of Science degree in Computer Science from Gebze High Technology Institiue in Turkey.