World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:12, 2008

# Using the Keystrokes Dynamic for Systems of Personal Security

Gláucya C. Boechat, Jeneffer C. Ferreira, and Edson C. B. Carvalho, Filho



***Abstract*—**This paper presents a boarding on biometric authentication through the Keystrokes Dynamics that it intends to identify a person from its habitual rhythm to type in conventional keyboard. Seven done experiments: verifying amount of prototypes, threshold, features and the variation of the choice of the times of the features vector. The results show that the use of the Keystroke Dynamics is simple and efficient for personal authentication, getting optimum resulted using 90% of the features with 4.44% FRR and 0% FAR.

***Keywords*—**Biometrics techniques, Keystroke Dynamics, pattern recognition.

## I. Introduction

NOWADAYS, to protect information of a system, two conditions are necessary to assure that only authorized people can access or to modify the data: identification and personal authentication, assuring the control and the access legitimacy to the information [1]. The identification establishes who the person is. This process happens during the initial login of the system, while the authentication confirms or denies the personal identity, demanding of the same a proof of her identity and obtaining the certainty that to people is really who is affirming to be.

Exist three methods different from authentication [2]: the first method of authentication is based on something that the person knows, as password and personal document, also call as Proof of Knowledge. The second method of authentication is based on something that the person possesses, as magnetic card or smart cards, also call as Proof of Posse. And third method of authentication is based in that the person is, as your physical or behavioral characteristic, that distinguishes a person of the others, also call as Proof of Biometric. The two first methods are more used, however, very vulnerable. In the first method the person can forget, can share their data, in the second method can lose or be stolen, however, in the third method the person presents a characteristic its, cannot be forged and nor be forgotten. Among the physical characteristics exist, for example, the geometry of the hand, face, iris and the features considered behavioral as digital signature, voice and the Keystroke Dynamics [3].

Biometric treated in this paper is the Keystrokes Dynamics, related with the way or habitual rhythm of as a person it types a password, words/phrases or text in a terminal [4]. Each person possesses a different rhythm of typing of the other, even an imposter having knowledge of the password of a person, which if it tries to pass, difficultly will go to be authenticated [5]. The Keystroke Dynamics is relatively a cockroach technique; it needs only a keyboard and software for authentication, different of the others biometrics techniques that possess one high cost of the captation devices and analysis of the necessary data in the authentication, and can also be used with or without the knowledge of the person.

Some features can be extracted of the keystroke rhythm as: the time that a key is pressed (keystroke duration), the time between successive keys (keystroke latency), speed of the keystroke, placement of the fingers and pressure that the person applies when pressing a key (pressure keystroke) [6].

The remaining of this paper is organized in four sections. In the section 2 are presented works published in the area. In the section 3 the proposed methodology is discussed: the extraction of features and the used classifier. The experiments are presented and discussed in the section 4, and finally the conclusions are found in the section 5.

## II. Related Works

The first study using the Keystroke Dynamic for identification it happened at the beginning of the decade of 80 for Gaines et al. [7], in their experiments made with seven would secrete using statistical method T-Tests for classification and authentication. For the composition of the pattern the keystroke latency was used among digraphs of an English text, words and sentences random, but just for digraphs that happened more than 10 times. The work resulted in a 4% FRR and 0% FAR.

In the work of Bleha et al. [8] they used password and phrase. The analyzed characteristic was keystroke latency between keys with Bayes Classifier and Minimum Distance. First experiment accomplished with nine volunteers in a period of nine weeks, second with ten volunteers, in a period of five weeks and 26 volunteers in a period of eight weeks. The tests were accomplished with ten worth users and 22 impostors, resulting in 3.1% FRR and 0.5% FAR.

Joyce and Gupta [5] in its experiments with 33 persons used the following data: username, password, firstname, lastname. The extracted feature was keystroke latency between two consecutive keys and using statistical classification. Were

Manuscript received October 15, 2006; revised November 18, 2006.

G. C. Boechat, J. C. Ferreira and E. C. B. Carvalho, Jr. are with the Computer Science Center, State University of Pernambuco (UFPE) cx.7851 50732-970 Recife, PE – Brazil (email: {gcb,jcf,ecdbcf}@cin. ufpe.br).

collected 8 patterns for formation of the set of training and 5 patterns for the set of test having as resulted 0.17% FAR and 13.3% FRR.

Monrose and Rubin [4] collected sample of 63 users, in a period of 11 months, extracted features: keystroke duration and keytroke latency, using Bayes Classifier with 92.14% of success.

Cavalcanti et. al. [9] Used statistical classifier, analyzing the features: keystroke duration and keystroke latency, from 24 volunteers, resulting in a 6.04% FRR and 0% FAR.

In the work of Costa et. al. [6] used the features: keystroke duration and 3 keystroke latencies between two keys, interval of time to the next key to be pressed, interval of time to press two consecutive keys and the interval of time to liberate two consecutive keys. Used classifier using Occult Models of Markov (HMM) they Obtained 4.5% EER. Other works were developed using the Keystroke Dynamics as in [10] for recognition in virtual keyboard or in [11] that used correlation between keys as measured feature. A marketed product using the Keystroke Dynamics is the *biopassword* [14] could be adapted to the system of login of Windows NT/2000/XP.

## III. Methodology

In the moment of the authentication of the user in a system some features can be obtained of the Keystroke Dynamic as, the keystroke duration and keystroke latency between successive keys, and the considered important aspects to have a good authentication are presented to follow them.

### A. Base of Data

The formation of the pattern of the Keystroke Dynamics is obtained through the capture in the way as the user types your full name, containing 40 characters in the maximum. The used base was acquired of the work of Cavalcanti [9], in the process of acquisition the user informed your name 20 times in each moment that entered in the system, in the total of three sections and five times in the other users' name for formation of the impostors' data. As result this acquisition a group of 24 classes, tends each class on mean 60 patterns of the user legitimize and 50 patterns of user impostors.

### B. Features

The features are extracted from the user's keystroke for formation of template and later for verification. Two features were extracted during the keystroke: keystroke duration and keystroke latency. Keystroke duration is the interval of time that a key is pressed and liberated. Keystroke latency is the interval of time the pressed of between two consecutive keys [12] interval of time to liberate a key and press the key successor. In the Fig. 1 shows the extracted features: keystroke duration (duration) e keystroke latency (interval) of the word "IVAN".

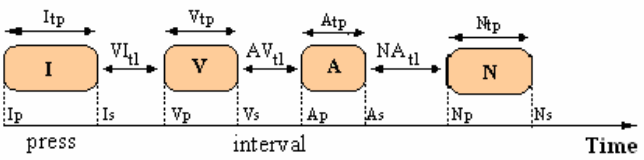


Fig. 1 Extracted features of the Keystroke Dynamics

The features extracted for formation of the pattern form the Features Vector (1) possessing keystroke duration and keystroke latency. Example in the Fig. 1:

$$\text{Features Vector} = [Itp, Vitl, Vtp, Avtl, Atp, Natl, Ntp] \quad (1)$$

$I_{tp}$ : Keystroke Duration of the key (I), that is, time that the user leads for press and liberate the key (I).

$VI_{tl}$ : Keystroke Latency between of the keys (V) and (I), that is, interval of time that the user leads for liberate the key (I) and press the key (V).

The Keystroke Duration is just composed by positive whole values, however, The Keystroke Duration is just composed by positive whole values, however, and the Keystroke Latency can contain positive values as negative. The negative value happens when the user before of liberate the key press the key successor. This usually happens with users that it possesses practice of typing.

### C. Prototypes

Is the prototype generated of the Mean ($\mu$), Minimum (*Min*) or Maximum (*Max*) and standard deviation ($\sigma$) that are calculated for each feature ($x_i$) of the pattern with size *n*, done compose by N pattern of the class, in agreement with the following equations:

$$\text{Maximum} = Max\,(x_1, x_2, \ldots, x_n) \quad (2)$$

$$\text{Minimum} = Min\,(x_1, x_2, \ldots, x_n) \quad (3)$$

$$\text{Mean}\,(\mu) = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (4)$$

$$\text{Standard Deviation}\,(\sigma) = \frac{1}{N-1}\sum_{i=1}^{N} |\, x_i - \mu_i \,| \quad (5)$$

### D. Classifier

The Classifier is responsible for the process of decision of the authentication. In this paper the authentication of the type verification is used, classifying accepts or it rejects the user, based on Criterion of Separation (Threshold). The Classifier verifies the similarity between the pattern to be verified and the template of the prototypes, using the Distance Pattern between the vector of feature of the pattern and the prototype. The distance is calculated from the equation (6):

$$\text{D (pattern, prototype)} = \frac{1}{n}\sum_{i=1}^{n} \frac{|\, pattern_i - prototype_i \,|}{\sigma_i} \quad (6)$$

### E. Criterion of Separation

A pattern is only authenticated, if the calculated distance between features vector of the pattern and the template of the prototypes to be inside of the value of the Threshold adopted. The separation criterion or Threshold defines two areas: users and impostors.

$$D(\text{pattern}, \text{prototype}) \leq \text{Threshold} \quad (7)$$

The value of the threshold can admit two forms: assumed the same value for all of the classes and values of independent thresholds for each class, each class is treated individually, depending on the class, can reach better results and more trustful the measure that the number of patterns of prototypes increases.

## IV. Experiments

The experiments were accomplished with 24 classes; containing patterns of the user legitimize and user impostors, these divided in patterns of the Prototypes set and tests, where the Prototypes set is composed by 30 patterns of the class and the test set is composed by 30 class patterns and 50 impostors patterns.

The performance of the system of authentication biometrics is measured through two error rates:

*False Rejection Rate* (FRR), also called Error of Type I, represents the percentage to reject incorrectly a legitimate user owed some variation in your normal type of typing. This error cause frustration, the user will have to type the pattern again.

*False Acceptance Rate* (FAR), also called Error of Type II, represents the percentage of incorrect acceptance the user impostors as a legitimate user. This type of error is caused by fraud.

The authentication systems are configured in accordance with the type of application could have a weak detection (low FRR and high FAR) or a sensitive detection (low FAR and high FRR).

In Fig. 2 it show an example of the relationship between FRR and FAR, can observe three important points: the point ZeroFRR indicates the value of FAR when the FRR is equal the zero, the point Equal Error Rate (ERR) indicates the value when FAR and FRR are equal and the point ZeroFAR indicates the value of FRR when the FAR is equal the zero.

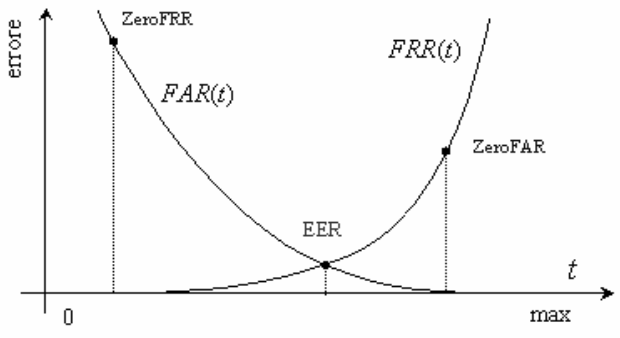


Fig. 2 Evaluation of performance

Initially, combinations of the features vector were analyzed and the measure of choices of the features of the pattern for composition of the Prototypes set. The used characteristics are: only keystroke duration, only a keystroke latency and the combination of keystroke duration and keystroke latency. The choice of the time of the features for formation of the Prototypes set is the mean, the minimum and the maximum of the times of the features vector.

Two experiments were accomplished to know the amount of patterns used in the Prototypes set and other with the choice of the threshold. It is important to inform that all the experiments were 30 times accomplished. in all the graphs and tables are shown the averages of the iteration.

### A. Amount of Patterns

Made experiment to analyze the impact of the amount of patterns for formation of the Prototypes set, using the features of keystroke duration and the keystroke latency, where are using the mean, the minimum and the maximum of the times of the features of the pattern for formation of the Prototypes set. The amount of patterns chosen for formation of the Prototypes set is 1 and 2 the 30 with variation of 4 patterns.

In Fig. 3 is shown the behavior of the FRR with increase of the amount of patterns. It is observed that the tests using the mean of the times of the features obtained performance better. The measure that the amount of pattern is increased; the FRR is diminished improving the performance of the system. The resulted priors obtained when using the minimum and the maximum of the times of the features, limiting the classification, as the Keystroke Dynamics is a characteristic behavior, for example, in the choice of the maximum time of the features, once that the user the delay to type as custom it can harm system performance.

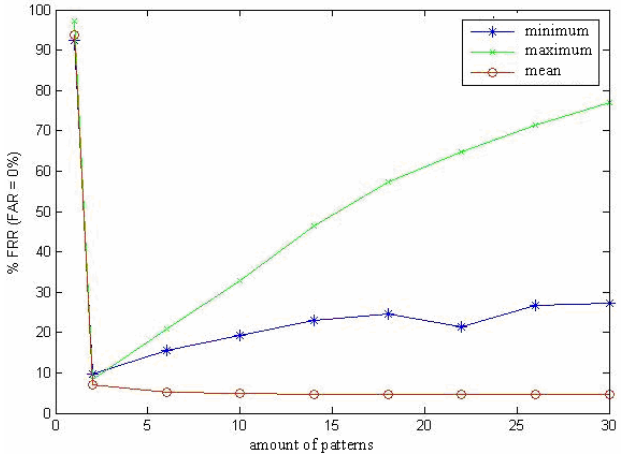


Fig. 3 Behavior of the FRR (when FAR = 0%) with the amount of patterns

### B. Threshold

A second experiment is to verify the effectiveness of the Threshold, using a Local Threshold attributing thresholds independent for each class, with the features of keystroke

duration and the keystroke latency, and using the mean of 30 pattern for the formation of the Prototypes set varying the values of the Threshold at three moments: in the first moment the values are between zero and one, with a variation of 0.1. From this result can meet the concentration of the patterns, between an inferior threshold (where FRR = 100% and FAR = 0%) and a superior threshold (FRR = 0% and FAR = 100%), in second moment with a variation of 0.01, of these also new thresholds are found inferior and superior for accomplishment of the last test varying the threshold 0.001.

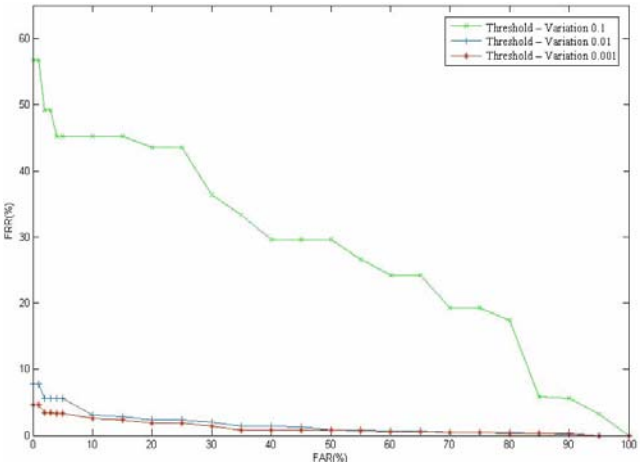


Fig. 4 Behavior of the FRR in relation the different FAR with variations of the Threshold (Combination of keystroke duration and keystroke latency)

It is observed in Fig. 4 that the tax FRR is decreased of the Threshold; the worse results were obtained varying 0.1 having a sensitive detection. It is noticed in Fig. 4, the tests with variation 0.01 and 0.001 had an enormous fall of the% FRR, improving the performance of the system, and difference between results with variation 0.01 and 0.001 obtained a fall of almost the half the FRR for FAR < 10 %. The results of the experiments with the variation of the Threshold for FRR when FAR = 0% can be seen in Table I.

TABLE I
BEHAVIOR OF THE FRR IN RELATION THE DIFFERENT FAR WITH VARIATIONS OF THE THRESHOLD

| Variations of the Threshold | %FRR (FAR = 0%) |
|---|---|
| 0.1 | 56.81 |
| 0.01 | 7.78 |
| 0.001 | 4.58 |

### C. Features Keystroke Duration

Experiments with only the feature Keystroke Duration; varying the mean, minimum and maximum of the times of the features of the patterns for formation of the Prototypes set.

It is observed in Fig. 5, the choice of the mean of the times of the features it reached best the rate of FRR for FAR < 35% of what the results obtained for minimum o and maximum of the times of the features, had resulted similar.

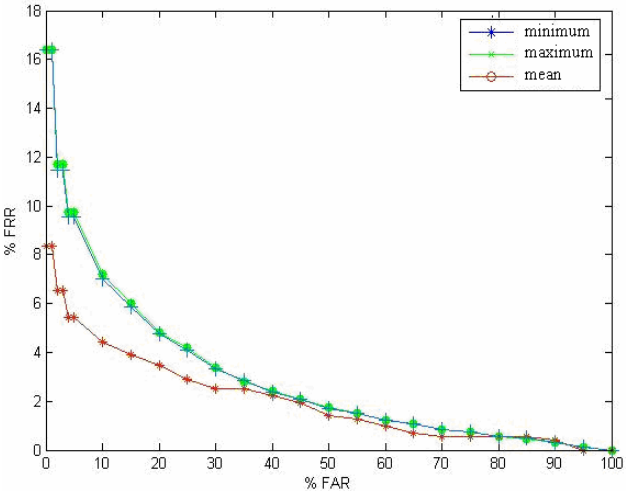


Fig. 5 Behavior of the FRR in relation the FAR (Feature: Keystroke Duration)

A summary contend the best ones resulted of the three configurations, using feature Keystroke Duration can be observed in Table II. Obtained best result 8.33% FRR and 0% FAR.

TABLE II
FRR(%)WHEN FAR = 0% (FEATURES AND CHOICE OF TIMES)

| Choice of Time | Keystroke Duration | Keystroke Latency | Keystroke Duration and Latency |
|---|---|---|---|
| Minimum | 16.41 | 21.06 | 9.18 |
| Mean | 8.33 | 10.42 | 4.58 |
| Maximum | 16.38 | 19.30 | 9.87 |

### D. Features Keystroke Latency

The second characteristic to be observed is the keystroke latency, varying the mean, the minimum and the maximum of the times of the features of the patterns for formation of the Prototypes set.

It is shown in Fig. 6, that the best resulted were obtained for choice of the mean of the times of the features with 10.42% FRR and 0 % FAR. Again the worse results found with the minimum and the maximum of the times of features, having resulted similar. It can be observed that the results only using the feature Keystroke Duration Were a little better that in the use only of the Keystroke Latency as it is shown in Table II.

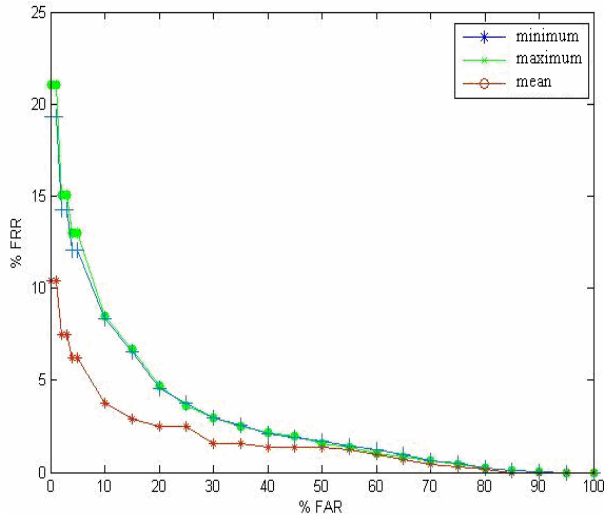


Fig. 6 Behavior of the FRR in relation the FAR
(Feature: Keystroke Latency)

### *E. Combination of the Feature Keystrokes Duration and Keystroke Latency*

The third experiment was made using the combination of the features Keystroke Duration and Keystroke Latency, varying the mean, the minimum and the maximum of the times of the features of the patterns for formation of the Prototypes set. Purpose of the combination is to diminish the rates and to improve the security of the system.

As in the previous experiments the choice of the mean of the times of the features, obtained better rates FRR for FAR < 50%, of what the results obtained for minimum o and maximum of the times of the features, had resulted similar, shown in Fig. 7. Can be observed in Table II the results using the combination of the features were superior to results only using one of the features.

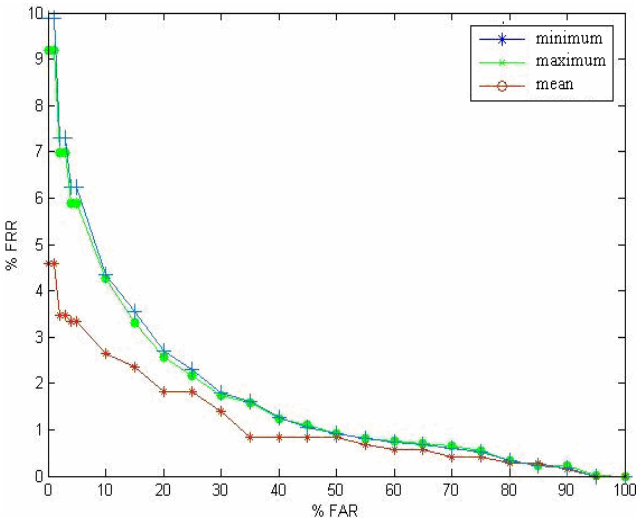


Fig. 7 Behavior of the FRR in relation the FAR
(Features: Keystroke Duration and Keystroke Latency)

The next experiments were made to verify some aspects of the methodology to verify the impacts in the rates obtained of Combination of the feature Keystrokes Duration and Keystroke Latency using the mean of the times of the features of the patterns for formation of the Prototypes set.

### *F. Select of Features*

With intention to find a subset from the set of features that it can reduce the errors rates. An experiment was made selecting N features of the features vector with the minors of standard deviation, eliminating the features less significant.

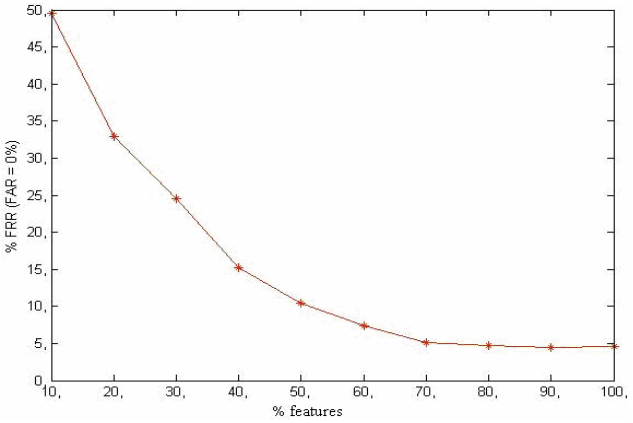


Fig. 8 Evaluation FRR adopting (FAR=0%) when the selection of feature varies

As observed in Fig. 8, FRR diminish with the increase of the amount of selected features, between 70% and 90% of selected features the performance profit is small, after this point it had a small increase in the FRR. It is observed that selecting 90% of the characteristics FAR obtained a reduction of the FRR tax passing the 4.44% which = 0%.

### *G. Global Threshold*

As explained previously the determination of the value of Threshold he is independent for each class. To verify the impact of the value of the Threshold, an experiment was made with a Global Threshold. In Fig. 9 is showed the behavior of FRR in relation FAR, is observed that the determination of the Threshold for Class, have better resulted compared with resulted obtained with Global Threshold reaching 57.56% FRR for 0% FAR. The use of the Global Threshold harms the performance of the class, losing its inherent features.

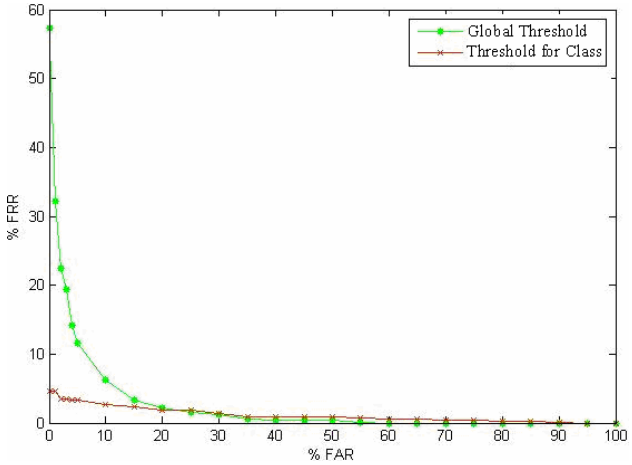

Fig. 9 Behavior FRR in relation FAR (Global Threshold)

## V. Conclusion

In this work was presented a methodology using the Keystroke Dynamics as one biometrics technique for authentication. The Keystroke Dynamics is the process to analyze the way that person types monitoring a keyboard, and the authentication if she bases on its habitual rhythm to type. Moreover, it is a form of not intruder of recognition and can be applied the security of systems.

The methodology proposal uses two characteristics of the Keystroke Dynamic: Keystroke duration and keystroke latency. Some experiments were made observing the changes in different aspects boarded in the methodology. Comparison of the extracted features, the best resulted were found when combined the Keystroke Duration and Keystroke Latency, and when analyzed the separate features the Keystroke Duration better were resulted that the Keystroke Latency.

In the choice of the time of the features for composition of the set of archetypes, the best resulted were obtained with the use of the mean the times of the features and e the worse were obtain by the minimum and maximum of the features. Experiment with selection of features that aim at obtain minors errors rates from with the elimination of certain features, Experiment with selection of features that aim at obtain minors errors rates from with the elimination of certain features, was found when selecting 90% of the characteristics. E finally the determination of the Threshold of decision in function of a Global Threshold, obtain worse results comparing with the determination of the Threshold for Class, that finishes for harming the performance of classification of the class.

Comparing with the work of Cavalcanti et. al. [9] it was possible to improve the rates for 4.44% FRR and 0% FAR. For future works, it intends to combine Keystroke Dynamic with techniques physical biometrics, where the process of personal authentication from physical or behavioral characteristic.



## Acknowledgment

The authors would like to express their sincere gratitude to the Cavalcanti et. al for making available the base of data, without her would not be possible make the work.

## References

[1] P.S. Magalhães, H. D Santos. Biometria e Autenticação. *IV Conferência Associação Portuguesa de Sistemas de Informação*. Porto, Portugal. 2003. (ed em CD-ROM : ISBN 97 2-9354-42-1).

[2] S.M Matyas, J Stapleton. A Biometric Standard for Information Management and Security. *Computers & Security* Vol.19 Issue: 2. 2000.pp.428-441

[3] L.C.F. Araújo. *Uma Metodologia para Autenticação Pessoal Baseada em Dinâmica da Digitação*. Universidade Estadual de Campinas, 2004.

[4] F. Monrose, A. D. Rubin. Keystroke Dynamics as a Biometric for Authentication. *Journal: Security on the Web (special issue) Future Generation Computing Systems (FGCS).* March 2000, pp.351–359.

[5] R. Joyce, G. Gupta. Identity Authentication Based on Keystroke Latencies. *Communications of the ACM*. Vol 33. Issue: 2. Feb 1990. pp 168-175.

[6] C.R.N. Costa, G.F.G. Yared, R.N. Rodrigues, J.B.T. Yabu-Uti, F. Violaro, L. L. Ling. Autenticação Biométrica via Dinâmica da Digitação em Teclados Numéricos *XXII Simpósio Brasileiro de Telecomunicações – SBrT'05.* Campinas, SP, Set, 2005.

[7] R. Gaines, W. Lisowski, S. Press, N. Shapiro. *Authentication by keystroke timing: some Preliminary Results*. Report R-256-NSF. Rand Corporation. 1980.

[8] S. Bleha, C. Slivinsky, B. Hussien. Computer-Access Security Systems using Keystroke Dynamics, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol.12 Issue: 12 Dec 1990, pp.1217-1222.

[9] G. D. C. Cavalcanti, E. H. F. Pinheiro, E. C. B. Carvalho Filho. Um Sistema de Verificação de Identidade Pessoal Através de Dinâmica da Digitação. *XXV Congresso da Sociedade Brasileira de Computação*. São Leopoldo-RS. Jul. 2005.

[10] J.Mäntyjärvi, J. Koivumäki, P. Vuori. Keystroke Recognition for Virtual Keyboard. *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, Aug. 2002. pp. 429-432.

[11] F. Bergadano, D. Gunetti, C. Picardi. User Authentication through Keystroke Dynamics. ACM Transactions on Information and System Security, Vol. 5, No. 4, November 2002, pp. 367–397.

[12] S.R.L. Silva Filho. *Autenticação Contínua pela Keystroke Dynamic usando Máquinas de Comitê.* Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. Florianópolis, SC. Nov 2005.

[13] M. Brown, S. J. Rogers. A Practical Approach to User Authentication. In *Proceedings of 10th Computer Security Applications Conference*. Dec. 1994, pp.108-116.

[14] BioPassword® Available in http://www.biopassword.com. Accessed in October 2, 2006