User authentication by information source using fuzzy approach in biometric keystroke dynamics

MILOSLAV HUB
Institute of System Engineering and Informatics
Faculty of Economics and Administration, University of Pardubice
Studentska 84, 532 10 Pardubice
CZECH REPUBLIC

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 modeling is presented as well.

Key-Words: Authentication, identification, biometric, keystroke dynamics, data security, fuzzy.

1 Introduction

The term information security is based on the definition safety information so those confidentiality, integrity and accessibility are preserved [13]. The term confidentiality meaning that information are accessible only for those who are authorised to access, the term integrity meaning indemnity rightness and absoluteness of information and methods their processing and finally accessibility of the information is the same as their usability for authorised users in the time of needs.

Information security is connected with enterprise managements at least by two linkages.

The first is marketing connection. As far as the enterprise increase one's credibility on the market by certificate one's quality system they have to expect questions on the level of information security form auditors. The second connection is inseparability of the information from the management and firm processes. In that case the information is sourcing likewise money and/or human. Non-ensure enterprises own sources threaten of the production, so the customer's input and lead to risk increasing for enterprise itself and avalanche-like wide spreading of the threats into commerce environment [15].

Information source is operation system, information system, data warehouse, intranet, different applications, remote access into this sources etc. The users of this information sources are both, own employees of institution, its business partners and/or e-business customers or suppliers within the just in time systems. Managing users' access to the information sources meaning techniques connecting user register so called authentication to the information sources with authorisation mechanism which on the information sources side verify access user's certification.

2 Related works

Authentication as a data security measurement is very important for keeping data as safe as possible in the framework of information society [2], [10]. The aim of authentication is to decide whether some subject is really the claimed one [18]. There are three 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 information as well [8].

One possible way how to increase security level of access to information systems, is a combination of authentication by knowledge and authentication by attribute, i.e. parallel usage of passwords and keystroke dynamics. Everyone has different style of keyboard typing [11]. It is quite similar to man's own signature. It is a reason why this way of biometric authentication was selected for a research which results are discussed in this article. But, the principles of keyboard typing authentication can be used as the other ways of authentications, e.g. hand geometric [1].

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

3 Our fuzzy approach

Our idea is seemingly simply – to create a fuzzy inference system for every template that is saved in a template database. Because we hope our authentification system could be used in a real security system this fuzzy

inference system has to be generated automatically without any intervention of a staff. We decided to model this way of authentification with help of Takagi-Sugeno-Kang fuzzy inference [14]. Although it is very similar to the most commonly methodology - Mamdani method in many respects, it offers a lot of advantages, expecially:

- It is computationally efficient.
- It works well with optimization and adaptive analysis.
- It has guaranteed continuity of the output surface.
- It is well suited to mathematical analysis.

Because we had only little vision about biometric data and we wanted to create a reasonably fast security system the first two advantages was the most important for us.

We analyzed a fuzzy inference of every template will have n inputs and only 1 output. The inputs will be created by n biometric characteristics (both keystroke durations and keystroke latencies) and output will be created by a measure of similarity.

The number of biometric characteristics you can calculate with the help of Eq. (1).

$$n = 2l - 1 \tag{1}$$

n ..... Number of biometric characteristics
l ..... Length of string in characters

Somebody can say at this moment that it is possible extract some others biometric characteristics from key typing, for example key press intensity, finger position and so on. But we used latency and duration times only because our goal is to suggest authentification systems that need not special hardware components.

An ideal situation will arise when valid user will acquire this measure of similarity 1 and an impostor will acquire 0. Because output level z (see Eq. X) is a constant (see Eq. Y) we ca say it is a zero-order Sugeno model.

$$z = a \cdot input_1 + b \cdot input_2 + \ldots \\ \ldots + c \cdot input_n + const. \tag{2}$$

$$a = b = \ldots = c = 0 \tag{3}$$

As it is said we had only a little vision about biometric data from previous researches. Sure we tried to analyze these data wit help of statistic, especially with help of exploration analysis and correlation analysis but we got only overall image about it. But in real world biometric characteristics of every user can have very different features, for example correlation between individual characteristics. It is reason why we hoped that ability of generalization of fuzzy models could help us.

Questions about amount of fuzzy sets, rules and shapes of membership function we decided to solve with help of adaptive techniques especially with help of subtractive clustering [10] and combination of the least-squares method and the backpropagation gradient descent method for training FIS membership function parameters.

4 Data collecting

Data collecting started two years ago when the special software was created for these purposes. The three-layer architecture was used. One part of this software is represented by a client (programmed in Java programming language). This client consist of graphical user interface (GUI) that is used for measuring of relevant duration and latency times when an user is typping a desired string (password). When the user finishes a typing the data are coded and sent to server. The server part of this software (programmed in PHP programming language) reads the data that are sent by client, decode them, check them, re-process them and saves it into database (MySQL database). So all information about simulated password typing (for example duration and latency times, ID of user, sequence keys that were used, time, ID of computer…) are stored in database for future treating.

The persons that were participated it this research were students of our university. It is necessary to note all participants were volunteers that attended in the experiment willingly. They were a young people from nineteen to twenty-for and their expediencies with computers were different – from slightly practiced to skilled experts.

In the first test every volunteer simulated template creating – he typed a required string several times. In next experiments volunteers simulated authentication process – they typed the required string only one times.

5 Authentification simulations

In authenification process there are two kinds of authentificied entities:

- valid users
- impostors

A valid user is authentificated by identity of his own. An impostor is authentificated by identity of some other subject.

The situations where a valid user and an impostor is accepted or rejected in authentification process are expressed in Table 1.

Table 1: Potential situations of the authentication decision

		Decision	
		Acceptance	**Rejection**
Situation	**Valid access**	Suitable situation	False rejection
	Impostor access	False acceptance	Suitable situation

Authentification simulation for us purpose has two different stages:

(1) Stage: authentification simulation for model training.
(2) Stage: authentification simulation for model evaluating.

For purposes of model learning we used only data that were acquired at the template creation simulating phase. As it is said, in this phase the participants were asked to type the desired string a several times (t times). The data from participant's template were used for valid access simulation and the data from other participant's templates were used for impostor access simulations. It is necessary to note we did not use all data from the whole template of others participants but we used only m first measures because we suppose that using of whole templates of others participants for impostors access simulation would break a balance in model learning process. The overall model consists of models for every registered user (template). The numbers of valid users and impostors used for this model you can calculate on base (4) and (5) respectively.

$$N_V^{(1)} = N \cdot t \qquad (4)$$

$$N_I^{(1)} = \left(N^2 - N\right) \cdot m \qquad (5)$$

$N_V^{(1)}$ .... Number of valid user access in model learning stage
N ..... Number of templates
t ..... Number of measures for every template
$N_I^{(1)}$ .... Number of impostor access in model training stage
m ..... Number of measures in one template used for impostor access simulations

When the model of authentification is sufficiently learned it is necessary to evaluate this model. Of course, we could not use the same data the model was learning with the help of and for this reason we had to use some different data. At this evaluating stage we used data from authentification simulations. Participant accesses were tested both against his own partial fuzzy model and against other participant's fuzzy model. By this technique it has been obtained the following numbers of valid accesses $N_V^{(2)}$ (6) and impostor accesses $N_I^{(2)}$ (7) that were used for our model evaluation.

$$N_V^{(2)} = N \cdot N_a \qquad (6)$$

$$N_I^{(2)} = \left(N^2 - N\right) \cdot N_a \qquad (7)$$

$N_V^{(2)}$ ... Number of valid user access in model evaluating stage
N ..... Number of templates
N_a ..... Number of simulated access by one person
$N_I^{(2)}$ ... Number of impostor access in model evaluation stage

The data of simulated access of valid users and impostors in second stage we used for evaluating of learned fuzzy model. As inputs to our partial models we used this data and we investigated the output data. As a criterion for comparison of our fuzzy model and contemporary models we have chosen false acceptance ratio FAR and false rejection ratio FRR represented with help of Detection Error Tradeoff [20].

6 Results of experiments

During training stage we used various strategies, we tried to use a part of learning data as checking data and when we tried to change m - the number of measures in one template used for impostor access simulations. More, we had tried to change initial state of our model before the machine learning was used for setting of model parameters.

As we expected, this modifications did not have a big effect on model quality. The main reasons are two:

Fuzzy model has property to generalize.

The number of measures for template creation (fifteen) for model training is too small.

Although the second reason is not very propitious for us, our model d_F that was suggested and trained has better values of criteria characteristics that we choose then classical models – d_E [21] and [2] (see Fig. 1).

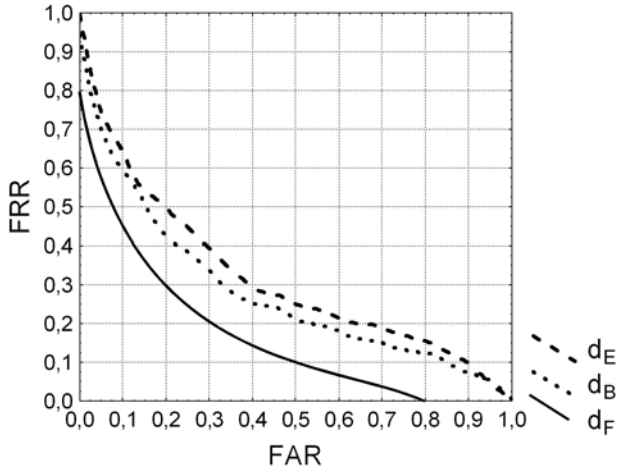

Fig. 1: Comparison of fuzzy and traditional models

The following figure (see Fig. 2) represents an example of surface of fuzzy model of the first user. It is possible to see a dependence of "similarity" on the first two inputs in this figure.

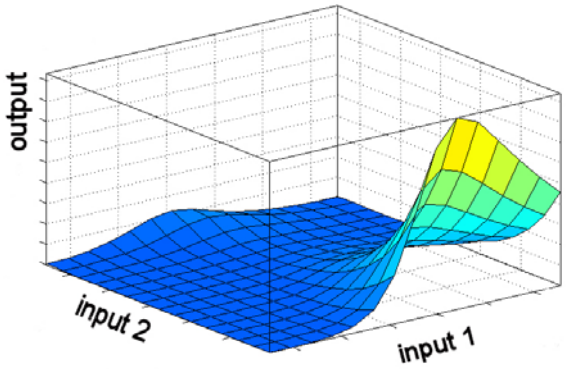

Fig. 2: Example of a part of a fuzzy model

7 Conclusion and future work

It is evident from previous graph (see **Chyba! Nenalezen zdroj odkazů.**) that our fuzzy approach of authentication process modelling gives better results than conventional algorithms when measure of dissimilarity are used. DET curve of our model d_F lies below curves d_E and d_B and therefore the chosen quantitative characteristics FAR and FRR are smaller in our model than in other models.

This work proves the fuzzy approach of authentification modeling works and even reaches better results than classical models. In future work it is necessary to continue in this research and try to get better results with more precise setting of initialize parameters and with enhancing of the amount of data for model training stage. We believe that on the end of this research we will find general relations that could be used for template creation without any training. Finally the implementation process will be start soon.

References:

[1] Artazi, P., R., at all: Hand geometric and hand print texture based prototype for identity authentication. In *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] Čapek, 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.

[4] Čapek, J., Hub, M.: Fuzzy Approach in Biometric Authentication by Keystroke Dynamics. In *WSEAS Transactions on Systems*, 2005, Issue 4, Volume 4. ISSN 1109-2777

[5] Bayes, T.: *An essay towards solving a problem in the doctrine of chance*. Philosophical Transactionsnof the Royal Society, London, 1763. Reprinted in Biometrika, Vol. 45, 1958, s. 298-315.

[6] Devijer, P. A., Kittler, J.: *Pattern Recognition: A Statistical Approach*.1st printing. New York: Prentice-Hall, Englewood Cliffs, 1982.

[7] Duda, R. O., Hart, P. E.: *Pattern Classification and Scene Analysis*. 1st printing. New York: John Willey and Sons, 1973.

[8] Hendl, J: Overview of statistical data analysis (original in czech). 1. ed. Praha: Portál, 2004. 583 p. ISBN 80-7178-820-1.

[9] Hub, M.: Strategy of the choice identification signs within multifactorial authentication. (originally in Czech). *E+M Economics and Management*. pp 147-150, Liberec 2003. ISSN 1212-3609

[10] Hub, M. *Data security – authentication* (original in Czech). 1. vyd. Pardubice: Univerzita Pardubice, 2005. ISBN 80-7194-825-X

[11] Chiu, S.: Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent & Fuzzy Systems*, Vol 2, No. 3. Sept. 1994

[12] Komárková, J., Šimonová, S., Dušek, V. *Geographic Information on the Web*. WSEAS TRANSACTIONS on INFORMATION SCIENCE AND APPLICATIONS, November 2004, vol. 1, issue 5, s. 1185 – 1188, ISSN 1790-0832

[13] Kostiha, F.: Data security (originnaly in Czech). Ikaros [online]. 2006, roč. 10, č. 5 [cit. 2007-03-18]. Dostupný na World Wide Web: <http://www.ikaros.cz/node/3332>. URN-NBN:cz-ik3332. ISSN 1212-5075.
[14] Legget. J, Williams, G., Usink, M.: Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, v36, s. 859-870, Sept. 1990
[15] Linda, B.: Interval reliability constructing by bootstrap method (originally in Czech). STATISTIKA 4/2003, s. 59-65, ISSN 0322-788x
[16] Linda, B., Kubanová, J.: Bias reduction with bootstrap method. In: Sborník 12. medzinárodný seminár Výpočtová štatistika. Bratislava 2004, s.62-64, ISBN 80-88946-29-8, EAN 9788088946298
[17] Neyman, J., Pearson, E. S.: On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrica*, 1928, 20A, p. 175-240.
[18] Neyman, J., Pearson, E. S.: *On the problem of the most efficient tests of statistical hypothese*. Phil. Trans. Royal Soc. London, 1933, 231, p.289-337.
[19] Sugeno, M., *Industrial applications of fuzzy control*, Elsevier Science Pub. Co., 1985
[20] Schlesinger, M. I., Hlaváč, V.: *Deset přednášek z teorie statistického a strukturálního rozpoznávání*. 1. vyd. Praha: Vydavatelství ČVUT, 1999. 521 s. ISBN 80-01-01998-5.
[21] Wald, A.: *Sequential Analysis*. New York: John Wiley, 1974.
[22] Wald, A.: *Basic idea sof a general Tudory of statisticalmdecision rules*. Proceeding of the International Congess of Mathematicians, 1950, vol. I.
[23] Wald, A., Wolfowitz, J.: *Optimum charakter of the sequential ratio test*. Ann. Math. Stat., 1948,19(3), s. 326-339.
[24] Young, J., Hammon, R.W.: *Method and apparatus for verifying an individual's identity*. Patent Number 4.805.222. U.S. Patent and Trademark Office. Washington, D.C., 1989.
[25] *Best Practices in Testing and Reporting Performance of Biometric Device*. U.K. Biometric Working Group. In Collected works 1997-2000. U.S. National Biometric Test Center. San José State University. August, 2003. p. 221-236.
[26] *Guide Understanding I&A*. National Computer Security Center. NCSC-TG-017 Library No. 5-235,479. Version 1.