UDC 004.942:004.93'12

Y. BABYCH
N. BABYCH
M. HOLOFIEIEVA
Odessa national polytechnical university

SIMULATION ALGORITHM FOR IMPLICIT AUTHENTICATION OF THE USER OF THE MOBILE DEVICE BASED ON FUZZY INFERENCE SYSTEM

Abstract – In connection with frequent cases of violations of user privacy of mobile devices in this article the algorithm of an implicit authentication of the user based on the fuzzy inference system. The proposed algorithm that allows to calculate the total score based on certain actions of the user of the mobile device and the limit value in real time; runs in the background and ensures high accuracy of the system with timely detection of unauthorized access to your mobile device. The data for the experiment were collected from different users of mobile devices with Android operating system.

Keywords: mobile device, mobile user, the algorithm implicit authentication, the system of fuzzy logical inference.

Ю. І. БАБИЧ
М.І. БАБИЧ
М.О. ГОЛОФЄЄВА
Одеський національний політехнічний університет

МОДЕЛЮВАННЯ АЛГОРИТМУ НЕЯВНОЇ АУТЕНТИФІКАЦІЇ КОРИСТУВАЧА МОБІЛЬНОГО ПРИСТРОЮ НА ОСНОВІ СИСТЕМИ НЕЧІТКОГО ЛОГІЧНОГО ВИВОДУ

Анотація – У зв'язку з частими випадками виникнення порушень конфіденційності даних користувачів мобільних пристроїв, в даній статті розглядається алгоритм неявної аутентифікації таких користувачів на основі системи нечіткого логічного виводу. Запропоновано алгоритм, який дозволяє розраховувати загальний бал на основі певних дій користувача мобільного пристрою і граничне значення в режимі реального часу; працює у фоновому режимі та забезпечує високу точність роботи системи при своєчасному виявленні несанкціонованого доступу до мобільного пристрою. Дані для експерименту було зібрано від різних користувачів мобільних пристроїв з операційною системою Android.

Introduction. In modern society, with the development of information technologies the use of mobile devices (MD) becomes an integral part thereof. MD is widely used by users for communication, everyday tasks, increase workflow efficiency, conducting financial transactions, storage or transmission of confidential information etc. As a result, increasing daily the number of malicious programs that threaten the security and privacy of user data. With the aim of preventing such threats of the modern MD is equipped with a system of user authentication. Traditionally used for this purpose is a PIN or password. Simple passwords are easy to break-and difficult – inconvenient to use. Statistics research [1] showed that most users choose simple passwords. In addition, a high risk of unauthorized access to the MD is provided by the fact that users leave their accounts in the public domain.

Problem. Popular in recent year's biometric user authentication MD [2] on the basis of his fingerprints, voice, iris recognition eyes or face in General is gaining popularity in Europe and America [3]. However, it can be violated. As an alternative, the algorithms for implicit user authentication MD, which will protect the MD against unauthorized use and at the same time ensure maximum convenience to its user.

Analysis of recent researches and publications. For example, in [4-6] proposed algorithms based on learning patterns of user behavior or environment. The results of MD with the use of such algorithms are high, but the maintenance of the balance between accuracy, adaptability and practicality remains a problem. In [7] investigated the algorithm for the authentication of the user, given its daily interaction with the MD.

The purpose of the research. To improve the accuracy of the algorithm due to its modernization through the use of fuzzy inference system, which will allow to make rational decisions in conditions of uncertainty and incompleteness of information (fig.1).

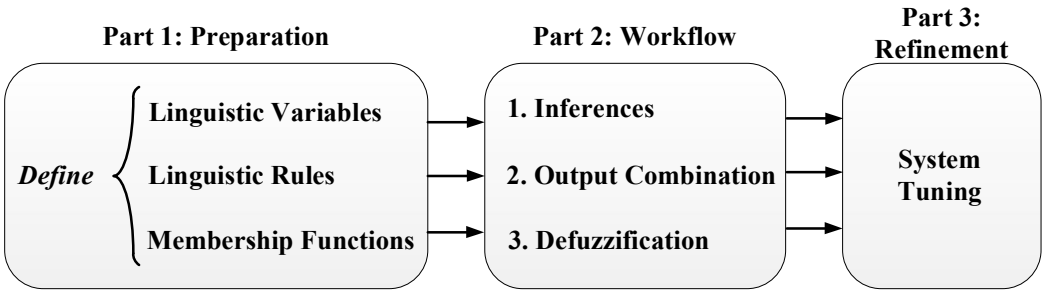

Fig 1. The scheme of the process of implementation of fuzzy inference

The results of the research. For the implementation of the algorithm implicit user authentication MD creates a model of its information profile, which takes into account the list of actions that execute them with MD: receiving incoming and outgoing SMS messages, incoming and outgoing calls, history of visited web pages and history of WI-FI network. Based on this, formed the membership function. The process of the algorithm implicit authentication is shown in fig. 2.

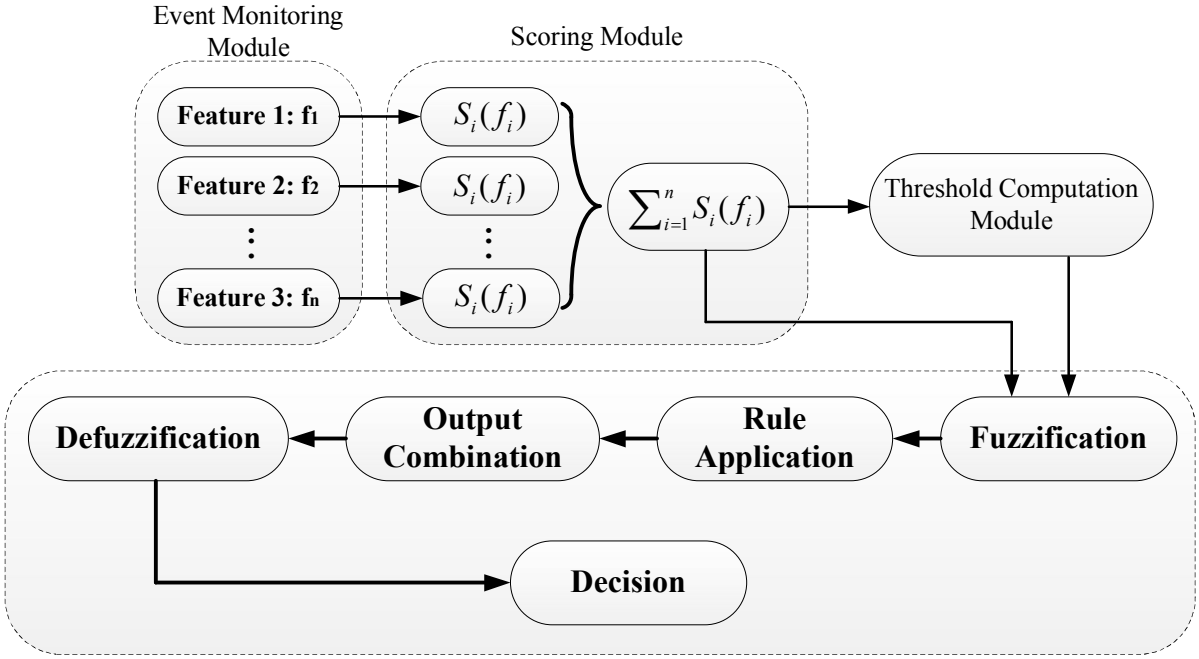

Fig. 2. Block diagram of the algorithm implicit authentication

Special module monitoring of user actions fixes the membership functions and calculates the result for the corresponding action (for example, occurrence of an incoming or outgoing SMS messages). The result is formed thus: the distance between the current and previous result, and the distance between the current result and the threshold values of the input variables. Processing the input data in the system allows you to obtain crisp output value indicating the level of full rights of the user of the MD. Knowing this level and the threshold value, the system determines whether the user is to be tested for implicit authentication.

Model information user profile of MD is:

$$Features : (f_1, f_2, f_3, ..., f_n) \tag{1}$$

Membership function:

$$Functions : (S_1, S_2, S_3, ..., S_n) \tag{2}$$

Thus, the total number of points received according to the model information of the user profile may be represented as:

$$AS = \sum_{i=1}^{n} S_i(f_i) \tag{3}$$

Since the algorithm works on the basis of processing user actions, the statistical score is calculated for each activity separately. After that, the overall result is used to calculate the threshold value.

$$Threshold_i = F_{threshold}(AS_i, Threshold_{i-1}), \tag{4}$$

where $F_{threshold}$ specifies the algorithm used to calculate an appropriate threshold value, which then starts processing module evaluation, which gives the result of implicit user authentication MD.

Information, the user profile includes a list of actions [8], which describe its current behavior and are used to generate the total score.

Table 1. The distribution of scores for each user action MD

	Contact list	Top 5	Duration, min.
Incoming call	4	6	10
Outgoing call	5	7	8
Incoming SMS	4	6	–
Outgoing SMS	6	6	–
History WI-FI	–	12	–
Browser history	–	5	–

As shown in table 1, the score of each object is calculated on the basis of various conditions. Top 5 of the table is obtained taking into account the total number of occurrences of "and". Long talk time user of MD is likely the basis of authorized use, and this is taken into account when calculating points for incoming/outgoing calls. Score 5 is assigned to each actions, if it is in the Top 5.

The calculation of the limit values is based taking into account a standard deviation and an exponential moving average.

$$\begin{cases} Threshold_1 = mean - (standart\_deviation) \\ Threshold_t = \alpha \cdot ASBA_{t-1} + (1-\alpha) \cdot Threshold_{t-1}, t > 1 \end{cases}, \quad (5)$$

where $ASBA$ is the average number of points per block.

The system uses the average score of the current block as an input parameter. The new threshold value used for decision at the next chunk, and is calculated as follows:

$$ASBA = \frac{\left(\sum_{i=k}^{i=k+b} AS_i\right)}{b} \quad (6)$$

The new threshold value is used to determine the threshold of the next block from AS_{k+b+1} to AS_{k+2b+1}. The optimal values of the parameters (size of block – b, factor – α and the length of experiment in days – l) are taken as $b = 6, \alpha = 0,2, l = 5$.

As shown in figure 2, the system fuzzy logic is integrated into the algorithm implicit authentication as a critical component. First, membership functions of input values to determine the distance between the current and previous result and a distance between the current result and threshold are three kinds of "negative", "medium" and "positive" as shown in figure 3.

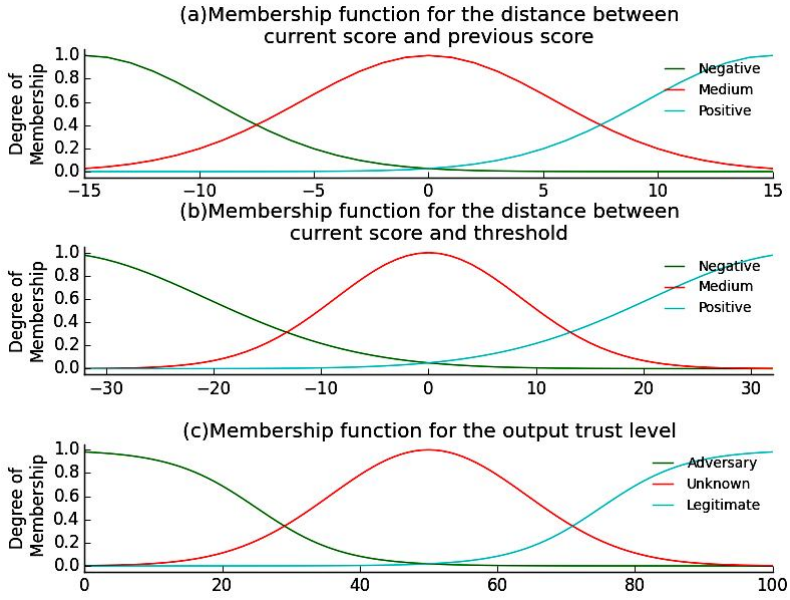

Fig. 3. Membership functions for determining the distance between the current and previous result, and the distance between the current result threshold value and output function

The initial value of the function is "breaking", "unknown" and "valid". In addition, five types of input features – "highly negative", "negative", "average", "positive" and "highly positive" are shown in figures 4 and 5. Based on figure 4, figure 6 added another two types of membership function, the result of which is shown in figure 7.

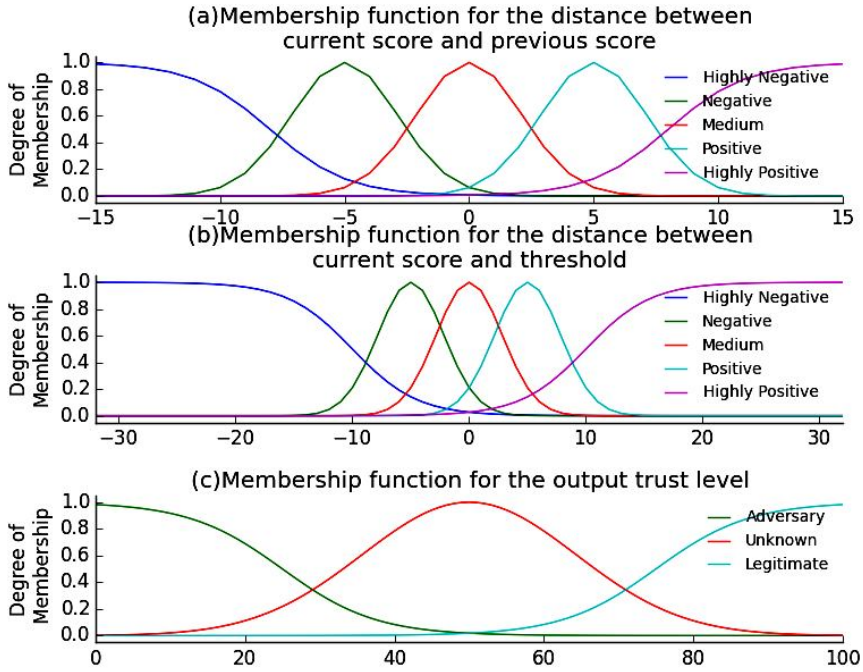

Fig. 4. Membership functions for determining the distance between the current and previous result, the distance between the current result and the threshold, and source function

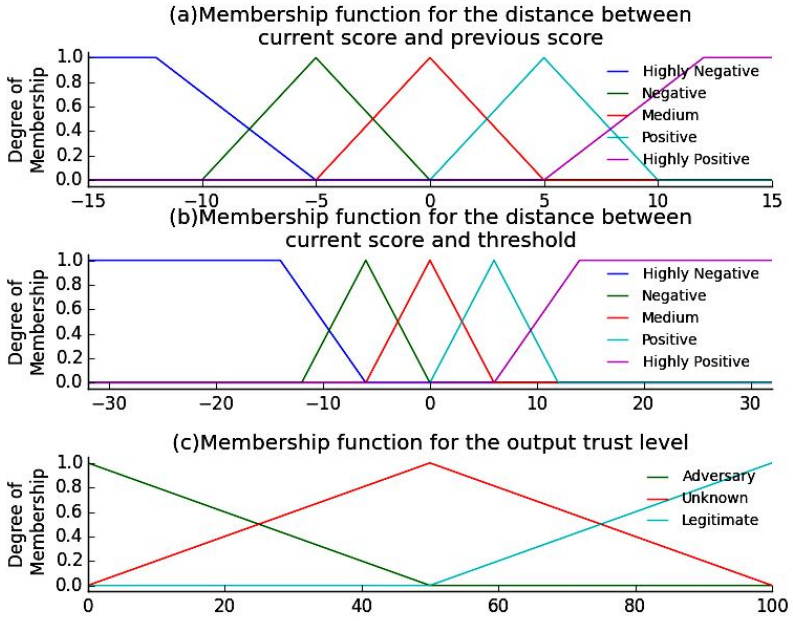

Fig. 5. Membership functions for determining the distance between the current and previous result, the distance between the current result and the threshold, and source function

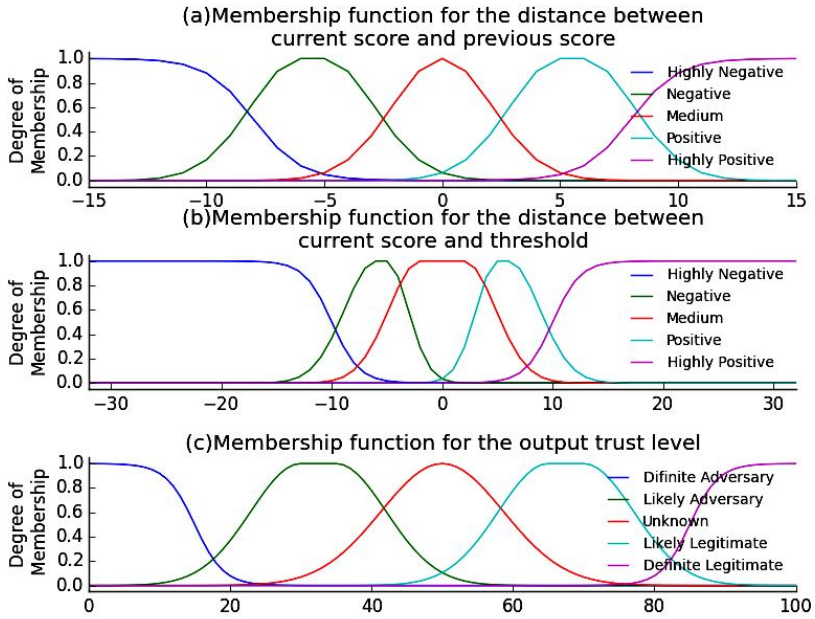

Fig. 6. Membership functions for determining the distance between the current and previous result, the distance between the current result and the threshold, and source function

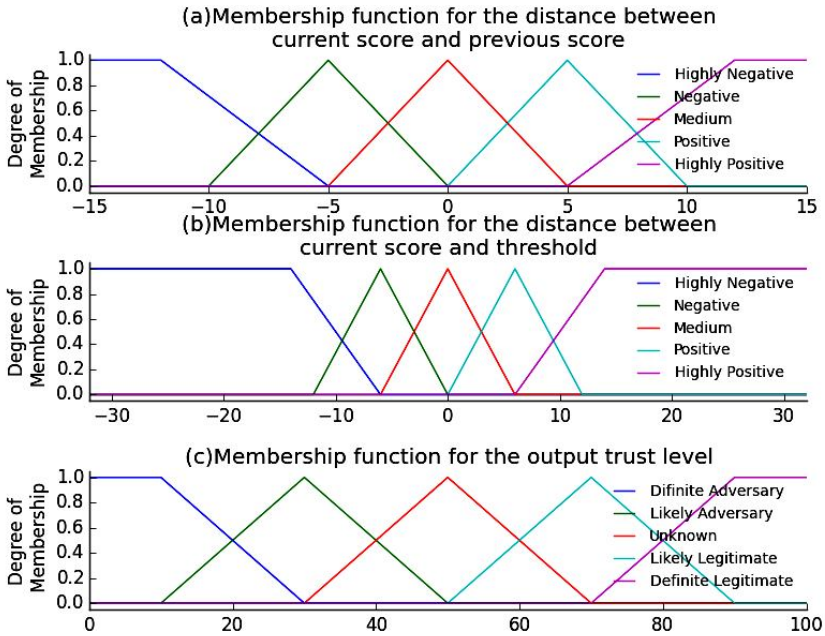

Fig.7. Membership functions for determining the distance between the current and previous result, the distance between the current result and the threshold, and the source function with the changes

The rule base formed on the basis of fuzzy outputs in the form of "if-then" [9] as shown in table 2, 3 and 4.

$$Grade\_strenght = \sqrt{\sum_i Rule_i^2} \tag{7}$$

Defuzzify process is carried out using the method of center of gravity.

$$Centroid = \frac{\int_s x \cdot y(x)dx}{\int_s y(x)dx}, \tag{8}$$

where S denotes the support of $y(x)$, which in turn is a result of changes in the source function.

Table 2. Matrix of rules for figure 3

The distance between the current result and the previous / the distance between the current result and the threshold value	Negative	Average	Positive
Negative	Violation	Unknown	Positive

Average	Unknown	Unknown	Positive
Positive	Unknown	Positive	Positive

Table 3. Matrix of rules for figures 4 and 5

The distance between the current result and the previous and the distance between the current result and the threshold value	HN	N	Avg	P	HP
HN	Vn	По	U	U	U
N	Vn	U	U	V	V
Avg	H	U	V	V	V
P	U	U	V	V	V
HP	U	U	V	V	V

where Vn – «violation», U – «unknown», V – «valid», HN – «highly negative», N – «negative», Avg – «average», P – «positive» and HP – «highly positive».

Table 4. Matrix of rules for figures 6 and 7

The distance between the current result and the previous and the distance between the current result and the threshold	HN	N	Avg	P	HP
HN	DV	DV	PV	U	U
N	DV	PV	U	PVa	PVa
Avg	PV	U	PVa	PVa	PVa
P	U	U	PVa	A	A
HP	U	PVa	PVa	A	A

where DV – «definitely a violation», PV – «possible violation», U – «unknown», PVa – «possibly valid», A – «acceptable».

To verify the effectiveness of the proposed algorithm, implicit authentication, studies were conducted of two users of MD in a few weeks. Data about user *A* behavior and was obtained from MD Samsung Galaxy A5 with Android operating system version 4.4.4, the data of the user *B* from the MD Samsung Galaxy S3 with Android operating system version 4.2. The essence of the experiment lay in the fact that the user *A* and perform authentication in the normal mode, and a user *B* in a simulated unauthorized access. The example is shown in figure 8 (membership functions used from figure 6).

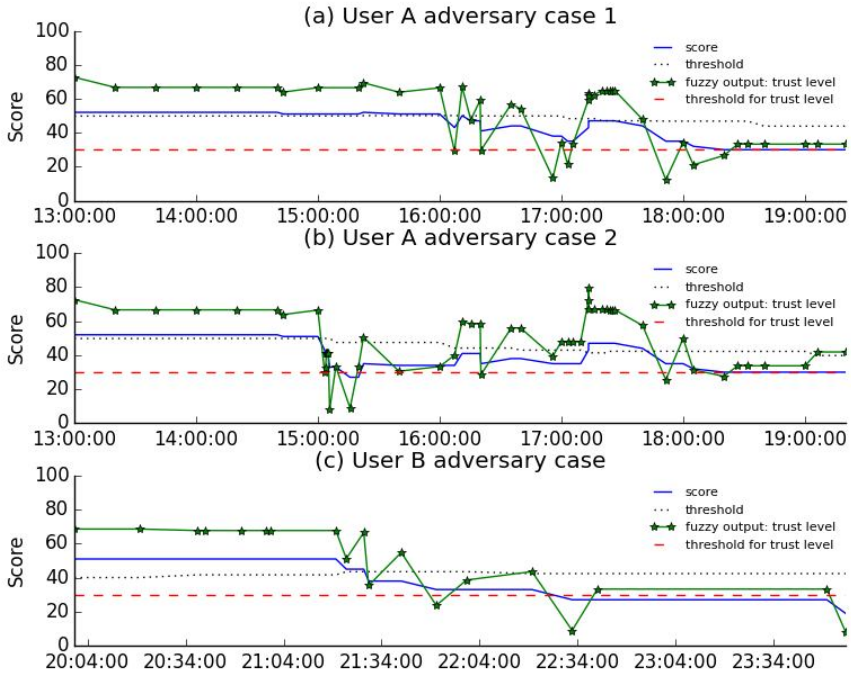

Fig. 8. Example of a standard authentication method,
user a and unauthorized user access

The threshold value for confidence level is set to 30 seconds. The constant change of the result and the threshold value reflects the user behavior of the attacker [11]. Thus, as soon as we introduce a new distance between the current and previous result, and the distance between the current result and the threshold value, the system fuzzy inference calculates the threshold level of confidence [10]. When the system evaluates the trust level below the

threshold, the user is perceived by the operating system of MD as an attacker. The efficiency of identifying the attacker is evaluated by examining two factors – the number of calculations before the first unauthorized access was detected and the actual duration in minutes (to the first detection of cracking).

For user A, event 1. A malicious user uses MD, but inactive (fig. 8 (a)): connects to a previously unused WI-FI network, browsing unknown browser sites and calling on previously unused and not listed in phone book numbers.

For user A, event 2. More active use of MD after was subjected to standard authentication by the owner (fig. 8 (b)).

For the user B. Unauthorized access made at 21:15:00 MD was stolen in the subway (fig. 8 (c)).

Figures 9-11 present the results of experiments conducted to assess the accuracy and the detection of the user-the attacker with the help of the developed authentication algorithm based on fuzzy inference system.

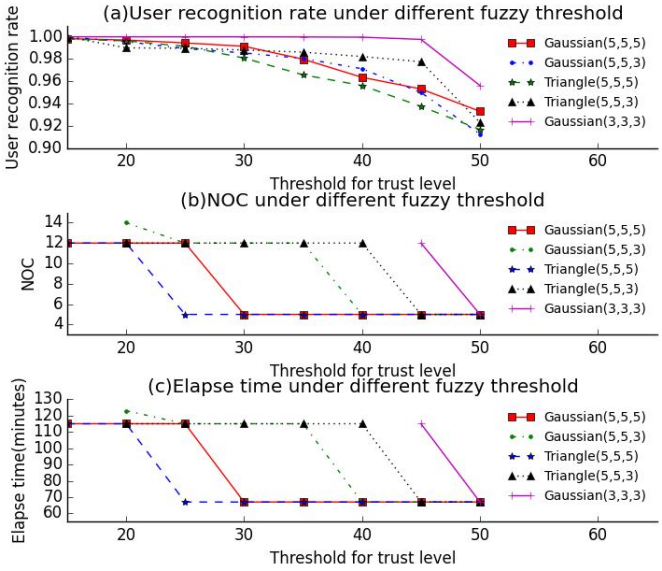

Fig. 9. The results of the experiment for user *A*, event *1*

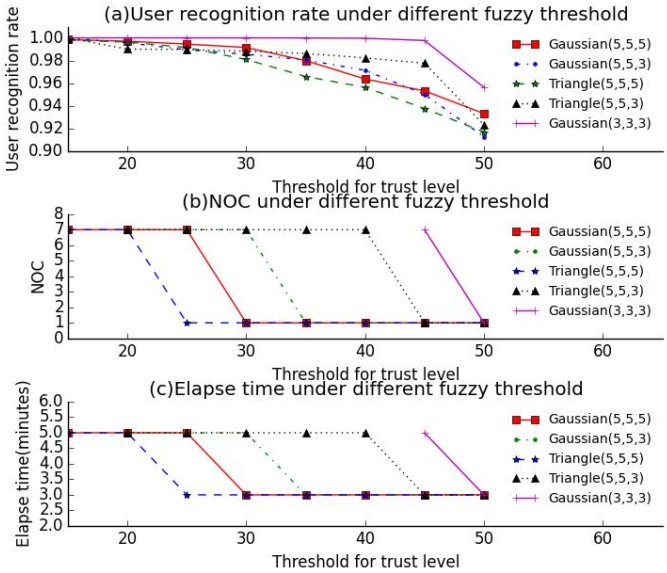

Fig. 10. The results of the experiment for user *A*, event *2*

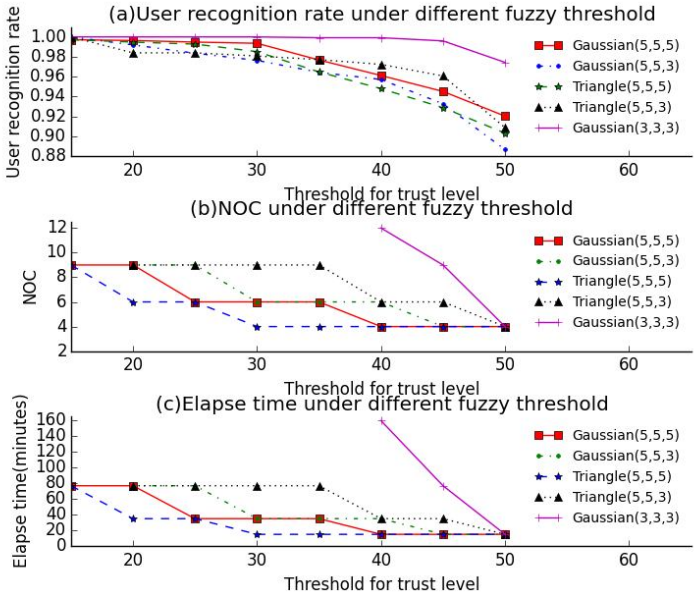

Fig. 11. The results of the experiment for user *B*

The recognition rate of user A and user B is in the range from 90% to 100% for Housiana and triangular membership functions. The optimal threshold value is taken 35.

Table 5. The comparison of the effectiveness of the algorithm for user *A*, event *1*

	Housian (5,5,3)	Triangular membership function (5,5,3)	Housian (5,5,5)	Triangular membership function (5,5,5)
The recognition rate of the user	98,05%	98,62%	97,99%	96,58%
The number of calculations before the first unauthorized access	12	12	5	5
Spent time	115 min	115 min	67 min	67 min

Table 6. The comparison of the effectiveness of the algorithm for the user *B*

	Housian (5,5,3)	Triangular membership function (5,5,3)	Housian (5,5,5)	Triangular membership function (5,5,5)
The recognition rate of the user	96,44%	97,72%	97,68%	96,48%
The number of calculations before the first unauthorized access	6	9	6	4
Spent time	35 min	77 min	35 min	15 min

Conclusions. Thus, the proposed algorithm for implicit user authentication keeps the protection of his MD from unauthorized access. System fuzzy inference is transparent, adaptive and requires no user interaction. Experiments were conducted to study the effectiveness of the proposed authentication algorithm using real data collected from two users of different MD.

Literature.

1. Confident Technologies, «Survey Shows Smartphone Users Choose Convenience over Security»,http://confidenttechnologies.com/news_event/survey-shows-smartphone-users-choose-convenience-security.

2. T. Stockinger, «Implicit authentication on mobile devices». In The Media Informatics Advanced Seminar on Ubiquitous Computing. 2011.

3. M. El-Abed, R. Giot, B. Hemery, C. Rosenberger, «A study of users' acceptance and satisfaction of biometric systems». In Security Technology (ICCST), 2010 IEEE International CamaHan Conference on, IEEE, pp. 170-178, Oct. 2010.

4. E. Shi, N. Yuan, M. Jacobson, and R. Chow, "Implicit authentication through learning user behavior." In Information Security, Springer Berlin Heidelberg, pp. 99 –113, 2011.

5. O. Riva, C. Qin, K. Strauss, and D. Lymberopoulos, «Progressive Authentication: Deciding When to Authenticate on Mobile Phones». In USENIX Security Symposium, pp. 301-316, 2012. [6] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley, «Casa: context-aware scalable authentication». In Proceedings of the Ninth Symposium on Usable Privacy and Security, ACM, pp. 3 – 13, 2013.

6. A. Gupta, M. Miettinen, N. Asokan, and M. Nagy, «Intuitive security policy configuration in mobile devices using context profiling». In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), IEEE, pp. 471– 480, 2012.
7. Kosenko, Yu. A. System dentipes funkcjonalno entrop sub CTA kriticno infrastructure [Text] / Kosenko Y. , Roslyakova S., Nosov P.// Collection of scientific works on materials of the international scientific-practical conference Modern trends in theoretical and applied research. – VIP. 2. – Odessa, 2013. – P. 50 – 54.
8. Kosenko, Yu. A. Neck model I methods dentify the forecast will informatino model student [Text] / P. S. Nosov, Yu. I. Kosenko // Automation. Automation. Electrotechnical complexes and systems. – VIP. 1 (25), Kherson: kntu, 2010. – P. 26 – 30.
9. Gruzdev, O. V. Investigation of the Google Voice service as a means of speech recognition with the use of software technologies.//Youth scientific and technical Bulletin. M., 2013.
10. Geppener, V. V. Wavelet transform in problems of digital signal processing: a tutorial / V. V. Geppener, D. A. Chernichenko, S. A. Athens // SPb.: Publishing house of Etu, 2002. 78 c.
11. Barannikov, V. A. a program Package for building speech recognition / V. A. Barannikov, A. A. Kibkalo // Proceedings of III all-Russian conference "Theory and practice of speech investigations" ARSO-2003. Moscow, Moscow state University. M. V. Lomonosov, September 2003, pp. 7 – 12.