SELECTION OF HANDWRITTEN SIGNATURE DYNAMIC INDICATORS FOR USER AUTHENTICATION

Selection and evaluation of handwritten signature indicators in user authentication systems are considered. An important and unresolved issue of information security is the effective identification of the user who accesses confidential information. Traditional password protection has a number of disadvantages. As an alternative to the password system or its addition, user identification by biometric characteristics is considered. An identifier that uses biometric characteristics is inextricably linked to the user and it is almost impossible to use it without authorization. The decision about the truth of the user of biometric authentication systems is made on the basis of comparison of his template with the data entered when trying to authenticate. The template is formed on the basis of studying the selected individual characteristics of the user. To do this, we use the biometric characteristics of users, which reflect the dynamic, behavioral characteristics of the person. It is proposed to use a handwritten signature as a biometric characteristic of the user. A handwritten signature is a socially and legally recognized biometric characteristic used for human authentication. It has a rather complex structure and high detail - all this makes solving the problem of user identification by mathematical methods quite complex and requires high computational costs. Another significant disadvantage is that handwritten authentication systems require the installation of additional specialized equipment, which makes the use of such systems as an ordinary means of authentication very expensive. To solve this problem, a scheme for implementing a computer data protection system against unauthorized access based on handwritten signatures using mobile devices based on the Android operating system as signature input devices is proposed. A method of biometric vector generation for use in dynamic biometric user authentication systems is proposed. The optimal characteristics are investigated and the efficiency of using the proposed form biometric characteristics vector is estimated. Certificates of copyrights registration for software applications developed during the work were received.

Keywords: biometric user authentication, biometric indicator, dynamic indicator, biometric vector, handwritten signature, biometric authentication system.

Problem statement. One of the urgent tasks of information technology development at the present stage is to ensure reliable protection of information. Existing methods today are divided into: hardware, software, mixed. The last one combines both hardware and software. The task of information protection is especially important in information and telecommunication networks, where confidential information circulates. An important and unresolved issue is the effective identification of the user who accesses it. Traditional password protection has several disadvantages [1], [2]. For example, in case of violation of the password confidentiality, it can often go unnoticed by its owner, immediately violating the protection of all information to which he has access. As an alternative to the password system or its addition, user identification by biometric characteristics can be considered [3], [4]. Biometric identification and authentication technologies have several advantages over traditional ones and are increasingly used in computer systems.

Methods of biometric identification are divided into two large groups: static methods, which are based on physiological characteristics of humans; dynamic methods that are based on the peculiarities of human behavior - subconscious movements in the process of performing any action.

© V. Yevetskyi, I. Horniichuk, 2020
Static and dynamic biometric identification methods are two interrelated and complementary areas of research. The main advantage of static biometric identification methods is their relative independence from the psychological state, low cost of user effort, and, as a consequence, the ability to organize biometric identification of large flows of people [4]. Dynamic methods of authentication are based on the behavioral (dynamic) characteristics of a person, i.e., based on features characteristic of subconscious movements in the process of reproduction of any action, have greater protection than static. The disadvantages of these methods include a fairly high probability of authentication error and false positives, as well as the need for a long, compared to static methods, the learning process. They are based on the study of the human voice, the dynamics of writing text using handwritten or keyboard handwriting of the user.

It is advisable to use a handwritten signature as a biometric characteristic. It is socially and legally recognized for human authentication. Therefore, its use in computer user authentication systems is relevant. The advantage of dynamic signature recognition systems is that due to the presence of a dynamic component, it is almost impossible for an attacker to forge the signature of the “victim” [5], [6].

**Analysis of recent research and publications.** Today’s handwritten signature is more often used for signing electronic documents by authorized persons than in authentication systems. That is why there is a wide range of software that allows you to perform the function of signing. Such software is focused on the use of specialized graphics tablets. The most famous manufacturers of specialized graphics tablets for handwritten signatures: Wacom, Huion, Signotec. They all have a wide range of devices, most of which are equipped with a specialized electronic pen. And also have their tools for developing applications for processing data obtained in the painting process [7], [8], [9].

There are also some commercial solutions, most of which support all of the above devices. The most famous software solutions include the following:

- **SignToLogin.** Software product of Russian origin. The solution has a dedicated server and is designed for iOS [10];
- **DocuSign.** A software product from the USA, designed mainly for signing electronic documents. Works with both mobile devices and specialized graphics tablets [11];
- **3M™ Electronic Signature Authentication Software.** Software product made in the UK, intended for use in banking and medical services, as well as in the field of electronic document management [12].

In such systems, information about various components of the painting process is collected during the signing process. Such information may include the following characteristics [13], [14]:

- spatial coordinates of the end of the pen;
- the pressure of the end of the pen on the tablet;
- azimuthal angle of the pen;
- the angle of the pen.

With a range of such dynamic features, it is possible to form a variety of biometric vectors and decision-making algorithms, based on both statistical methods and methods using neural networks [14], [15]. The advantage of dynamic signature recognition systems is that due to the presence of a dynamic component, it is almost impossible for an attacker to forge the victim's signature [6], [9], [10]. However, the main disadvantage is that such systems require the installation of additional specialized equipment for the implementation of the signature, which makes their use as an ordinary means of authentication very expensive.

The presence today of mobile devices of almost all users prompted the idea of using them in authentication systems. This may allow replacing specialized hardware with mobile devices. Previous works have proposed schemes for establishing a communication channel between a mobile device and a computer, as well as a scheme for implementing a users authentication system by their handwritten signature using mobile devices [16]. For the proposed system, it was decided to use only the values of the pen end spatial coordinates at different points in time. Knowledge of the
signature coordinates at specific points in time allows you to calculate the various features of the signature in biometric vector standard formation.

The aim of this paper is, to select dynamic indicators of handwritten signature and evaluate the effectiveness of their use in dynamic biometric authentication systems of users. It is achieved by addressing the following specific tasks:

1. Develop a method of forming a biometric vector.
2. Investigate the optimal parameters of the biometric vector.
3. Evaluate the effectiveness of the proposed biometric vector.

The main material research. In biometric identification/authentication systems based on the recognition of a handwritten signature by an applicant for admission, a fixed handwritten signature (presented at the training stage) is entered using a specialized graphics tablet or other input devices. For the proposed system, the values of the x and y coordinates of the end of the pen at time t are used. In this case, the period of time \( \Delta t \), after which the coordinates will be obtained must be constant and small enough, for the accuracy of calculations, in the system under development \( \Delta t = 10^{-3} \text{ second} \).

It makes it possible to get these characteristics using a touch screen of any smartphone or tablet. Thus, after entering the signature we get the following time characteristics [16]:

\[
\{(x_1; y_1), (x_2; y_2), ..., (x_N; y_N)\}, \quad N = T / \Delta t,
\]

where \( N \) – the total number of points received during the signature;

\( T \) – total time of signature entry.

When forming a biometric vector, the entire signature is divided into a fixed number of intervals \( n \) of equal length \( k = N / n \). It is proposed for each investigated interval to determine the average speed of its introduction \( s_i \) and the angle of vector inclination of beginning and end for the studied interval \( d_i \). It is proposed to calculate these values according to the following formulas:

\[
s_i = \frac{\sum_{i=1}^{k} l_j}{k}, \quad i = 0, n,
\]

\[
l_j = \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2},
\]

where \( s_i \) – the average speed of interval introduction \( i \);

\( l_j \) – Euclidean distance between adjacent points in the interval.

\[
d_i = \begin{cases} 
\arccos(\cos \alpha_i), & \Delta y_i > 0, \\
360^\circ - \arccos(\cos \alpha_i), & \Delta y_i < 0,
\end{cases}
\]

where \( d_i \) – the angle of vector inclination of interval’s beginning and end.

It is determined from the cosine of the angle obtained as the scalar product of the unit vector \( \bar{e}(0; 1) \) and the vector of the interval \( \bar{z}_i(x_{i+1} - x_i, y_{i+1} - y_i) \):

\[
\cos \alpha_i = \frac{\bar{e} \cdot \bar{z}_i}{|\bar{e}| \cdot |\bar{z}_i|}
\]

(1)

where \( \bar{e} \cdot \bar{z}_i \) – scalar product of vectors calculated using (2);

\[|\bar{e}|, |\bar{z}_i|\] – the absolute lengths of the vectors calculated using (3).

\[
\bar{e} \cdot \bar{z}_i = x_e \cdot x_i + y_e \cdot y_i = y_{z_i}
\]

(2)

\[|\bar{e}| = 1,
\]

(3)

\[
|\bar{z}_i| = \sqrt{x_i^2 + y_i^2}.
\]

Substituting (2) and (3) in (1) we obtain its full form:
Consider Fig. 1. It shows an example of a handwritten signature, the color indicates the points at which the characteristics are obtained. The difference between the points in time is $\Delta t$. The endpoints of the interval are highlighted in blue. The variables $l_j,l_{j+1},l_{j+2},\ldots$ schematically denote the Euclidean distances between adjacent points based on the sum of such distances and the number of intervals, the value of $s_i$ for this interval is calculated. Variable $d_i$ indicates the angle between the interval vector and the $OX$ axis.

![Figure 1 – Example of determining signature characteristics](image)

Before using the biometric vector in further calculations, it must be normalized by the maximum value in the interval:

$$s'_i = \frac{s_i}{\max(s)},$$

$$d'_i = \frac{d_i}{360°},$$

where $s_i, d_i$ – non-normalized value of the speed and direction parameter;

$s'_i, d'_i$ – normalized value of the speed and direction parameter;

$\max(s)$ – the maximum value of the average speed.

Thus, the general form of the biometric vector will be as follows:

$$v = (s'_1, s'_2, \ldots, s'_n, d'_1, d'_2, \ldots, d'_n).$$ (4)

Hamming distance is used to make decisions about the user's truth. It is the number of positions in which the corresponding characters of the two same length words are different. In a more general case, it is used for rows (vectors) of the same length and serves as a difference metric (a function that allows you to determine the distance in the metric space) of objects with the same dimension [17].

In training mode, the authorized user provides $L$ of their signatures ($L$ times enter the passphrase). This will correspond to $L$ implementations of the vector of biometric parameters $V = \{v_1,v_2,v_3,\ldots,v_L\}$. 
By analyzing the obtained matrix of $L$ implementations of the vector of time characteristics of the user $v$, we can obtain its characteristic change interval of each specific time parameter $[\min(P_i), \max(P_i)], i = 1, N$, which in the future will become the basis for forming a standard of biometric characteristics of the user.

At the authentication stage, the user provides his signature (enters a passphrase), which will correspond to a certain vector of biometric characteristics:

$$v = (P_1, P_2, ..., P_n, P_{n+1}, P_{n+2}, ..., P_{2n}) .$$

$$P_i = \begin{cases} s_i, & i = (1, n), \\ d_{i/2}, & i = (n+1, 2n), \end{cases}$$

(5)

where $P_i$ – biometric vector parameter.

Next, it is necessary to analyze the provided parameters of the biometric characteristics vector to get into the intervals set by the user's biometric standard, on whose behalf the login is carried out. During this analysis, the system creates a vector $E = (e_1, e_2, ..., e_n)$. Vector $E$ is formed as follows:

$$e_i = \begin{cases} 0, & t_i \in [\min(P_i), \max(P_i)], \\ 1, & t_i \notin [\min(P_i), \max(P_i)]. \end{cases}$$

The obtained $E$ – Hamming vector of the applicant for access to the system. For a user registered in the system, the vector $E$ should consist of almost only 0, and for an unregistered user who provides inaccurate biometric parameters, it will have many 1. The absolute value of the Hamming distance $E_v$ from the presented vector of biometric characteristics $v_p$ to the biometric standard $V_v$ is determined as the number of discrepancies with the biometric standard of the provided parameters, ie the number 1 in the Hamming vector. The distance $E_v$ is always positive and can vary from 0 to $N$.

To form confidence intervals, it is necessary to calculate the value of the mathematical expectation of $m(P_i)$ parameters and their standard deviation $\sigma(P_i)$. Then the minimum and maximum limit of the interval will be calculated as follows:

$$\min(P_i) = m(P_i) - T[L_e(1-p)] \cdot \sigma(P_i),$$

$$\max(P_i) = m(P_i) + T[L_e(1-p)] \cdot \sigma(P_i),$$

where $L$ – the number of vectors used in training; $p$ – set value of errors of the first kind (probability of refusal of authentication to “authentic” user); $T[L_e(1-p)]$ – Student’s coefficient.

Thus, we form a confidence interval for the mathematical expectation of the biometric standard’s parameter.

Mean and variance calculated by the following formulas:

$$m(P_i) = \frac{1}{L} \sum_{j=1}^{L} P_{ij},$$

$$\sigma^2(P_i) = \frac{1}{L} \sum_{j=1}^{L} [m(P_i) - P_{ij}]^2.$$
each user it is necessary to determine the threshold of the measure of Hamming $E_v$, at which it will be authenticated in the system, and it will be impossible to authenticate unregistered users.

The $E_p$ threshold can be determined using the mathematical expectation and variance of the Hamming measure values for the registered user:

$$E_p = m(E_v) + C[L,(1 - p)] \cdot \sigma(E_v),$$

where $C[L,(1 - p)]$ – Student’s coefficient, given based on the number of examples used $L$ and the value of the first kind error probability $p$.

We form the final biometric standard of user which will have the following look:

$$V_e = (\min(P_1), \max(P_1), \min(P_2), \max(P_2), \ldots, \min(P_n), \max(P_n), E_p).$$

**Determination of the biometric vector optimal length.** According to (4), the length of the biometric vector depends on the number of intervals into which the signature is divided when the parameters are obtained. The length of the vector is twice the number of intervals. To build a system based on the selected recognition algorithm and biometric vector, it is advisable to investigate the dependence of the biometric vector length and the efficiency of user recognition. A practical experiment was conducted for this purpose. Its essence is to determine the influence of the biometric vector length on the number of errors of the first and second kind [18]. For this purpose, a software application was created, which based on 10 signature entries created biometric standards for vectors of different lengths, in particular, for the division into 15, 20, 25, 30, 35, 40, and 45 intervals. At the subsequent input of the signature on its basis, 7 vectors of the specified lengths are formed, and based on Hamming’s distance the decision on the truth of the user is made. It was attended by two people who first taught the system their signature, and then entered it 100 times. After that, they swapped places and tried to forge the signature of another participant in the experiment for 100 times. The results are shown in Table 1.

FRR (False Reject Rate) or error of the first kind – the probability of erroneous failures to the registered user;

FAR (False Accept Rate) or error of the second kind – the probability of granting access to an unregistered user;

$$FRR = \frac{n_1}{n}; \quad FAR = \frac{n_2}{n},$$

where $n_1$ – the number of failures to the true user;

$n_2$ – the number of cases of incorrect access allow;

$n$ – total number of attempts.

<table>
<thead>
<tr>
<th>The intervals number of biometric vector</th>
<th>User 1</th>
<th></th>
<th></th>
<th>User 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>FRR</td>
<td>n2</td>
<td>FAR</td>
<td>n1</td>
<td>FRR</td>
<td>n2</td>
</tr>
<tr>
<td>15</td>
<td>44</td>
<td>0,44</td>
<td>2</td>
<td>0,02</td>
<td>40</td>
<td>0,40</td>
</tr>
<tr>
<td>20</td>
<td>51</td>
<td>0,51</td>
<td>0</td>
<td>0</td>
<td>55</td>
<td>0,55</td>
</tr>
<tr>
<td>25</td>
<td>54</td>
<td>0,54</td>
<td>1</td>
<td>0,01</td>
<td>51</td>
<td>0,51</td>
</tr>
<tr>
<td>30</td>
<td>25</td>
<td>0,25</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0,23</td>
</tr>
<tr>
<td>35</td>
<td>32</td>
<td>0,32</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0,27</td>
</tr>
<tr>
<td>40</td>
<td>7</td>
<td>0,7</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0,4</td>
</tr>
<tr>
<td>45</td>
<td>48</td>
<td>0,48</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0,40</td>
</tr>
</tbody>
</table>

Using table. 1 and fig. 2 it can be concluded that the number of errors of the first and second kind depends on the biometric vector length. It is proved experimentally that the optimal length value of the vector is 40 biometric parameters.
Evaluating the effectiveness of the selected indicators use. Let us estimate the probability \( p \) of correct recognition of the user by its frequency in \( n \) independent experiments [19]. Using the developed software application we will carry out experiment. To do this, 5 users successively entered their signature for 100 times. The number of access denials is noted and shown in table 2.

Table 2 – Data on the number of denials of access to the true user

<table>
<thead>
<tr>
<th>User №</th>
<th>Number of false authentication failures</th>
<th>Frequency of false access deny</th>
<th>Access allow`s frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.03</td>
<td>0.97</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The average value of the correct user recognition frequency in a series of 100 experiments is 0.954. Determine the 90% confidence interval for the probability.

To check the applicability of the normal distribution law, the values of \( np \) and \( nq \) are estimated [19]. Assuming approximately \( p \approx p^* \) we obtain:

\[
np \approx np^* = 95,
\]

\[
nq \approx n(1 - p^*) = 5,
\]

where \( p^* \) – average frequency of access.

The obtained values give reason to believe that the normal distribution law can be applied in this case. According to the tables given in [19], we find \( t_\beta = 1.643 \) for \( \beta = 0.9 \). Next, calculate \( p_1 \) and \( p_2 \) according to the following formulas:

\[
p_1 = \frac{p^* + \frac{1}{2} t_\beta^2 - t_\beta \sqrt{p^*(1 - p^*) + \frac{1}{4} t_\beta^2}}{n + \frac{1}{4} t_\beta^2},
\]

\[
p_2 = \frac{p^* + \frac{1}{2} t_\beta^2 + t_\beta \sqrt{p^*(1 - p^*) + \frac{1}{4} t_\beta^2}}{n + \frac{1}{4} t_\beta^2},
\]
\[ p_1 = \frac{0.954 + 0.0135 - 0.0366}{1.027} = 0.91, \]

\[ p_2 = \frac{0.954 + 0.0135 + 0.0366}{1.027} = 0.98. \]

Thus, the probability of correct user recognition at 90\% confidence interval is in the range from 0.91 to 0.98.

**Conclusions.** Summarizing the obtained results, it is worth noting that authentication using users biometric characteristics is an effective way to protect information. One of the best methods must be considered systems based on handwritten signature recognition. For systems based on handwritten signature recognition, there are two main methods of implementation - statistical and dynamic. Among them, the more advantages have the dynamic method, provided that there are additional means of entering the signature. The decision about the authenticity of such systems users is made based on a comparison of their template with the data entered during the attempt to authenticate. The template is formed based on studying selected individual characteristics. To implement the user authentication system by his handwritten signature, a method of biometric vector formation is proposed. The speed of movement at certain intervals and the inclination angle of the interval’s vector were chosen as indicators of the handwritten signature. These indicators reflect the dynamic component of a handwritten signature and can be obtained without the use of specialized hardware.

The dependence of the first and second kind errors number on the length of the biometric characteristics vector is investigated. Due to this, the optimal biometric vector’s length was experimentally obtained, which was 40x2 of biometric parameters. It is under such conditions the probabilities of the first and second kind errors were the lowest. It is also estimated that the probability of correct user recognition by the developed software application at 90\% confidence interval is in the range from 0.91 to 0.98. These values are calculated based on experimental data obtained from 5 users when entering a signature for 100 times. Based on them, it can be argued that in this formation of the biometric standard, in the proposed scheme, the users dynamic biometric authentication system implementation based on their handwritten signature is quite effective.

According to the work results, copyright registration certificates for the developed software applications were obtained [20], [21], [22].

In the prospects of further research, it is planned to evaluate the permanence of the characteristics of handwriting over a long period of time with further adjustment of the proposed methods.

**REFERENCE**


The article was received 30.03.2020.
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ


ВИБІР ДИНАМІЧНИХ ПОКАЗНИКІВ РУКОПИСНОГО ПІДПИСУ ДЛЯ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ

Розглянуто обирання та оцінювання показників рукописного підпису в системах автентифікації користувача. Важливою та ще не вирішеною проблемою захисту інформації є ефективна ідентифікація користувача, який отримує доступ до конфіденційної інформації. Традиційний парольний захист має ряд недоліків. Як альтернатива парольній системі або її доповнення розглядається ідентифікація користувачів за біометричними характеристиками. Ідентифікатор, що використовує біометричні характеристики, нерозривно пов’язаний з користувачем, і скористатися ним несанкціоновано практично неможливо. Рішення про істинність користувача біометричних систем автентифікації приймається на основі порівняння його шаблону з даними введенями при спробі автентифікуватися. Шаблон формується на основі вивчення обраних індивідуальних характеристик користувача. Для цього використовуються біометричні характеристики користувачів, які відображають динамічні, поведінкові характеристики особи. Як біометричну характеристику користувача пропонується застосовувати рукописний підпис. Рукописний підпис є суспільно і законно визнаною біометричною характеристикою, що використовується для автентифікації людини. Він має достатньо складну структуру і високу деталізацію, – все це робить вирішення проблеми ідентифікації користувача математичними методами досить складним і потребує великих обчислювальних затрат. Також суттєвим недоліком є те, що системи автентифікації з використанням рукописного підпису вимагають встановлення додаткового спеціалізованого обладнання, що робить використання таких систем як рядового засобу автентифікації дуже дорогою. Для вирішення цієї проблеми запропоновано схему реалізації системи захисту комп’ютерних даних від несанкціонованого доступу на основі рукописного підпису з використанням мобільних пристроїв на основі операційної системи Android як пристроїв введення підпису. Запропоновано метод формування біометричного вектору для використання в динамічних біометричних системах автентифікації користувачів. Досліджено оптимальні характеристики та оцінено ефективність використання вектору біометричних характеристик запропонованої форми. Отримано свідоцтва про реєстрацію авторських прав на розроблені протягом виконання роботи програмні застосунки.

Ключові слова: біометрична автентифікація користувача, біометричний показник, динамічний показник, біометричний вектор, рукописний підпис, система біометричної автентифікації.
Yevetskyi Viktor, candidate of technical sciences, associate professor, associate professor at the cybersecurity and application of information systems and technologies academic department, Institute of special communication and information protection of National technical university of Ukraine “Igor Sikorsky Kyiv polytechnic institute”, Kyiv, Ukraine.
  ORCID: 0000-0002-5364-8076.
  E-mail: viktorevetskv@gmail.com.

Horniichuk Ivan, postgraduate student, Institute of special communication and information protection of National technical university of Ukraine “Igor Sikorsky Kyiv polytechnic institute”, Kyiv, Ukraine.
  ORCID: 0000-0001-6754-4764.
  E-mail: horniychuk.ivan@gmail.com.

Євецький Віктор Леонідович, кандидат технічних наук, доцент, доцент кафедри кібербезпеки і застосування інформаційних систем і технологій, Інститут спеціального зв’язку та захисту інформації Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна.

Горнійчук Іван Вікторович, аспірант, Інститут спеціального зв’язку та захисту інформації Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна.