AUTHENTICATION OF INFORMATION SYSTEMS USERS, BASED ON THE ANALYSIS OF THEIR HANDWRITING

Vysotska Olena[1], Davydenko Anatolii[2]
National Aviation University[1]
Pukhov Institute for modeling in energy engineering of NAS of Ukraine[2]

ABSTRACT. In this paper there was analyzed a relevance of the problem of an authentication of information systems users. There was also reasoned a choice of dynamic biometric authentication methods, namely the methods based on an analysis of a person's handwriting. Based on the results of the performed experiments, there were selected the handwriting characteristics, which were analyzed for recognition further. There were defined the requirements to training items and the stages of its selection and adjustment. On the basis of the created algorithm, there was written a program to perform the authentication of information systems users, with the help of which a number of experiments were carried out. According to the results of the experiments, it was concluded that it is advisable to use the handwriting recognition systems for the implementation of the authentication of information systems users.

Keywords: authentication, recognition, biometrics, handwriting, information systems.

With the increasing degree of computerization of most human activity areas, there is an increasing need of the authentication of information systems users. In different cases, it is advisable to use one or another authentication method. All the authentication methods can be divided into the following three types:

1. Password protection. The user presents a secret data (for example, password or PIN-code).

2. Key usage. The user presents his / her personal identifier, which is the physical carrier of a private key. For example, plastic cards with a magnetic stripe, key chains and other devices.

3. Biometric authentication, i.e. the usage of human biometric characteristics. The user presents a parameter that is a part of himself. With such authentication, the person's identity is exposed to recognition – his individual characteristics (fingerprints, face thermogram, retina, voice, handwriting, etc.).

In recent times, there increasingly began to use the biometric authentication [1]. The biometric authentication systems are very user-friendly. Passwords and storage media can be lost, stolen, copied. The biometric authentication systems are based on human parameters, that always remain with a person, and the problem of their safekeeping doesn't appear. It is almost impossible to lose them. It is also impossible to transfer the identifier to third parties.

Thus, we can say that the development of the biometric authentication systems is now one of the relevant problems.

The objective of this paper was to develop a biometric authentication system of information systems users, which could be used in various fields of human activity, not always associated with the computer.

To solve this problem, the following was done:

1. There were analyzed the existing methods of biometric authentication [2,3]. On the basis of the conducted analysis and taking into account the fact that biometric authentication, in this case, should be used for the user accessing the information system in cases which are not always associated with the user's work at the computer, as the analyzed characteristics of the person, it is proposed to use his handwriting in this work.
2. The varieties of handwriting authentication methods were analyzed and its characteristics are determined for their further use in the authentication process.
3. The analysis of the selected handwriting characteristics was conducted to determine their validation for further recognition [4,5].
4. The selection of training handwriting items of information systems users was made to increase the probability of correct recognition [6].
5. The adjustment of training handwriting items of information systems users was performed to increase the probability of correct recognition [6].
6. The program for handwriting authentication of information systems users was written. The neural network was used as a recognition mechanism [7].
7. On the basis of the conducted analysis, with the help of the written program, it was concluded that it is advisable to use handwriting recognition systems for the implementation of authentication of information systems users.

All methods of the biometric authentication are divided into two groups:

1. Statistical methods.
2. Dynamic methods.

Statistical methods are based on the measurement of physical (static) characteristics of a person, which must be unique for each person, or at least for most people. These characteristics aren't supposed to change significantly over a long period of time and be influenced by any external factors, such as cosmetics, weather events, etc. Statistical biometric authentication methods include methods that use the fingerprint, the shape of the palm, the location of the veins on the front side of the palm, the retina, the iris (each eye has its own picture), the shape of the face (full face, profile, volumetric geometry), the thermogram of the face, DNA, etc.

Dynamic methods of biometric authentication are based on the behavioral (dynamic) characteristics of a person, that is, which are built on the features, which are typical for the subconscious person's movements in the process of presentation of any action. Unlike physical distinctive characteristic, in this case, the biometric system doesn't necessarily have to measure the same phenomenon each time: a person may be asked to say, write, or walk in a certain way to reduce the risk of reproducing the characteristics by the violator. Dynamic biometric authentication methods are methods that use for recognition the handwriting, the keyboard pattern, the painting with a mouse, the voice (speech), the acoustic signal from the human body, the movement of the lips when reproduction a keyword, the gait, etc.

Let's consider in more detail the method of dynamic biometric authentication of information systems users, namely the method based on the analysis of handwriting. This method doesn't require an expensive equipment (only a graphics tablet or similar device is required), it can be used in cases unrelated to the work at computer of the authenticated one and, at the same time, it is a fairly reliable authentication method.

Handwriting recognition systems relate to dynamic person's identification systems based on the analysis of the dynamics of fast subconscious movements reproduction. Handwriting is an individual and quite stable characteristic of a person, when analyzing of which it is possible to identify the person who wrote the keyword. This fact is confirmed by the physiological maker of a person. Let's consider what determines the handwriting of a person more closely.

During the writing, the muscles of most fingers and forearm muscles are involved. In total, more than 50 muscles can be involved, but the most significant influence have about 10 muscles. That is, when writing some text, a person controls about ten muscles. This is quite a complex task and it can't be solved in real time. Therefore, during the writing the control of a person by muscles is based on standard solutions that are developed during a long enough writing training and are individual for each person. These well-established standard solutions are kept throughout life and stay almost unchanged.

Dynamic characteristics of a person, including handwriting, can be used not only for recognition of people, but also for determining the character of the person, his mental and physical condition. For example, if a person is nervous or in a hurry, then the speed of writing increases and the handwriting becomes more boldly; the weak, sickly, and mentally ill persons have the letters that are non-uniform in size and with a variable inclination; the incompleteness of the words, simplified writing of the letters, bouncing up and down lines indicate the negligence of a person; etc. Analysis of these and other similar features is very useful when hiring (especially to the critical job) and while monitoring the work of computer users. For example, if the air traffic controller is tired or annoyed, he can make a mistake that will lead to serious consequences. A wide field of handwriting recognition algorithms usage determines the relevance of its analysis and improvement.

Let’s consider the handwriting authentication more closely.

To perform this type of authentication, the user must depict (write) his password (signature or some keyword). As a password there can be used:

1. Some keyword, word or a letter combination.
2. Signature of a user.
3. Some figure.

In this paper, as a password there is proposed to use any common keyword or different words, but provided that all these words have the same fragment (letter combination).

In addition to that, in this paper it is proposed to divide the functioning of the authentication system into the following two stages:

1. The recognition of the written keyword. *At* this stage, it is checked whether the correct word was written by the authenticated one. If an incorrect word is entered, i.e. that it doesn’t correspond to the pattern stored in the system for the given user, then the authenticated one is recognized as a violator, if the correct word is entered, then the second stage of recognition is performed [8].

2. Recognition of the keyword writing style. *At* this stage, the features of writing the analyzed word are checked for correspondence to the features stored in the system for this user. If the features correspond, then the authenticated one is recognized as a legal system user, otherwise he is recognized as a violator.

Mathematically the authentication (recognition) function $R=f(par_1,par_2,\ldots,par_g,\ldots,par_q)$ (where $1 \le g \le q$) can be expressed in the following way:

$$R = \begin{cases} 1, \text{ if } & Usx \in Usl, \ \ Usx = Usl_d; \\ 0, \text{ if } & Usx \in Usl, \ \ Usx \neq Usl_d; \\ 0, \text{ if } & Usx \in Usb; \end{cases}$$

where $Usx \in Us$ is the authenticated person;

$Us=\{Us_1,Us_2,\dots,Us_\infty\}$ is a set of people who can try to access the protected system;

$Usl=\{Usl_1,Usl_2,\dots,Usl_t,\dots,Usl_d,\dots,Usl_l\}$; $1 \le t \le l$; Usl is a set of legal users of this system;

Usl_d is a legal user impersonated by an authenticated person ($Usl_d \in Us$ и $Usl_d \in Usl$);

Usb is the violator, i.e. the unregistered person in this authentication system ($Usb \in Us$, но $Usb \notin Usl$);

$par_1,par_2,\dots,par_g,\dots,par_q$ is the list of parameters on which the result of evaluation of the recognition function R depends.

One of the most important factors when constructing the recognition system is the optimal choice of the recognition mechanism that, accordingly, means the recognition algorithm constructing. There are various mechanisms for solving such problems. One of the most effective mechanisms for recognition is neural networks, and it is proposed to use it in this work.

For all types of biometric recognition technologies, an initial user template should be created first. To create it, it is necessary to collect (make a fixation of) a number of measurements from any device used (in this case, from a graphics tablet). Then, it is needed to pick out the specialties (characteristics) inherent to the user from the measurements, and use the extraction results for template creation. The creation of this initial template is called a registry. This initial template is then saved by the system and plays a part of a kind of password further. After registration, when the user attempts to pass the authentication procedure, the measurement data from the reader device is collected, processed into a usable form and checked for correspondence to the template that was previously registered. In case of confirmation, the user is recognized as the impersonated person. Thus, one can say that the handwriting authentication system should work in the following two modes:

1. The registration or accumulation of the training items database.
2. The user recognition (authentication).

Depending on the implementation of each specific biometric system, there may be some additional actions. For example, if necessary, it is possible to continue to accumulate a database of training items, even at the recognition stage. Depending on the biometric authentication method used, the minimum size of such database will vary. For example, if the recognition method uses a fingerprint, there should be several items for each user (for several fingers and taking into account that the finger on the scanner can be placed at different angles) in the database. Conversely, if the method for recognition uses the handwriting, as in this work, then the database size is many times larger. This is explained by the fact that the fingerprint is a static parameter and does not change during a person's life, and the handwriting is a dynamic parameter and, depending on various factors, though not significantly, but can change. The need for a high volume of the training items database is also explained by the specifics of the functioning of the neural network, which is proposed to be used as a recognition mechanism in this work. Such a database usually reaches several hundred, and sometimes thousands of items for each user.

The key indicators of the efficiency of any authentication system are the error of the first kind (the denial of access to a legal user) and the error of the second kind (the omission of the violator). Therefore, when creating a handwriting authentication system, it is necessary to solve the problem of constructing the recognition function *R*, in which the error of the first kind (R=0, if $Usx \in Usl, Usx = Usl_d$) and the error of the second kind (R=1, if $Usx \in Usl, Usx \neq Usl_d$ or R=1, if $Usx \in Usb$) would be minimal, that is, to minimize the probability of false reject of a legal user (in case if he isn't impersonating another legitimate user) and the probability of omission the violator.

The result of evaluation the authentication function *R* depends on the list of parameters $par_1, par_2, \ldots, par_g, \ldots, par_q$. These parameters (features) selection is a primary target when constructing any recognition system. In this paper, there is used a set of features $At = \{At_1, At_2, \ldots, At_i, \ldots, At_n\}$; $1 \le i \le n$; as a set of features for recognition; where *At* is a set of features of person's handwriting; At_i is the *i*-th characteristic of handwriting.

Herewith, it should be noted that the *At* set depends on the level of computer technique development. The higher the level of technique, the larger the *At* set and the higher the maximum possible authentication quality. At earlier stages of the technique development it was possible to analyze only static characteristics of handwriting (the X and Y points coordinates), that's why the recognition quality wasn't high enough. And only with the advent of graphics tablets (and other similar devices) it became possible to determine and, therefore, to analyze the dynamic parameters of handwriting (the pen pressure on the tablet, the angle of the pen inclination, the speed and trajectory of writing, etc.), which significantly increased the quality of recognition. Most of the features of the *At* set are the characteristics of the set of points presented for recognition of the keyword. These characteristics, as mentioned earlier, are such parameters as the X and Y points coordinates, the pressure (P) with which the user presses the pen on the tablet when drawing the next point, the writing speed, the angle of the pen the angle of the pen inclination, the angle of the characters inclination, the trajectory of writing, etc. (different systems can use a different set of characteristics). Thus, it may be said that the *At* set of features looks like this: $At=\{X, Y, P, ...\}$. However, to achieve the highest efficiency of the authentication system, f it is necessary to select the set of the most significant (control) points $Kt=\{Kt_1,Kt_2,...,Kt_{nkt},...,Kt_b\};$ $1 \le nkt \le b$; whose characteristics will be analyzed during recognition process, rom the set of these points $T=\{T_1, T_2,..., T_a,..., T_v\}$;. That is, from the set of data about the conveyed points $Pac=\{Pac_1, Pac_2, ..., Pac_a,..., Pac_v\}$ (where $1 \le a \le v$;), it’s needed to create (to pick out) a set of data about control points $Pac\_kt=\{Pac\_kt_1, Pac\_kt_2, ..., Pac\_kt_{nkt},..., Pac\_kt_b\}$ (where $1 \le nkt \le b$;). This is quite an important stage of the authentication system operation, so we consider in detail.

When selecting control points, there appears the problem of the correctness of the automatic selection of control points. It is also important to determine the required number of control points. These two factors have a great impact on the efficiency of a particular recognition system, because if too few control points are selected or they are not correctly placed, the percentage of false recognition will be inadmissible large, and if there are set too many control points, then too much resources (time and memory) will be spent, and the recognition quality will increase slightly (or not increase) at the same time.

In this paper, there is proposed to emphasize control points of the following three types (Fig. 1):

1. The start and end points of each line. In figure 1, these points are shown in red color (points 1 and 15).

2. The angular points, that is, the points on the line bend. In figure 1, these points are shown in blue (points 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13).

3. The points of intercrossing of lines. In figure 1, these points are shown in green (points 8 and 14).

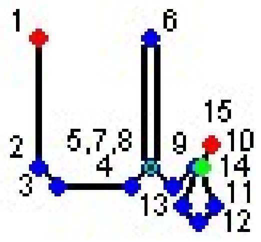

Fig.1. An example of control points placing

It is also rational to use the control point type as the point parameter to be analyzed.

In this work, it is recommended to divide the image of the whole keyword into images of individual letters and later work with images of each symbol separately. There are two reasons for this.

First, when checking the correctness of the written password (at the first stage of authentication), it is preferable to recognize the password not entirely, but symbolically. This is caused by the following reasons:

1. It is easier to collect items of writing N symbols (where N – is the number of characters in the used alphabet, taking into account the fact that not only lowercase letters can be used, but also uppercase letters and numbers, and various punctuation marks), than items of writing all possible passwords, which can be any combination of these N symbols, in the database. The number of possible passwords (the combinations of symbols), in this case is equal to $\sum_{Ks=1}^{Mks} Ks^N$ (where Ks is the password length, MKs is the maximum possible password length) or is equal to the Ks^N, if the Ks is known. The collecting of this amount of data is much more difficult than in the case of character-oriented recognition.

2. It is easier to check the classified object for belonging to one of the N classes than for belonging to one of the $\sum_{Ks=1}^{Mks} Ks^N$ classes. That is, the recognition process is easier in case of the character-oriented recognition analysis.

Secondly, data processing on all the password at once takes excessively large resources. When performing handwriting authentication with a help of the graphics tablet while writing a single symbol, the system receives the data on a one hundred points (sometimes on several hundred points) at an average. If the password, for example, has 6 characters, then it is necessary to process an average of 600 data points to validate of the entered password. That is, at the first stage of authentication the neural network will need to process 1800 features (three parameters for each point: the X and Y coordinates and the type of control point), and it will take a very long time (and at the second stage there are more features and, therefore, there is needed even more time). Therefore, when performing each authentication step, it is recommended to analyze not all the password at once, but symbol by symbol, and if one of the symbols is incorrect, then the other symbols can be not analyzed (at least at the first authentication step). However, using such a strict condition, it is necessary to provide a very low probability of false failure when recognizing each character, so that false non-recognition of one character leads to false non-recognition of the entire password, and, consequently, the denial of accessing the protected system for the legal user.

Dynamic characteristics of a person are characterized by some instability. Neural networks cope with a small instability well enough (that is another reason for choosing this mechanism for solving the problem), but gross errors in the original data must be discarded, i.e. it is necessary to make the original data selection. For example, in the resulting training items there may be random deviations (errors), which characterize nothing and will only worsen the recognition quality, for this reason these deviations must be removed from the items, and if there are too many of such deviations in one item or they are too gross, then the item must be removed from the database (or simply not considered during recognizing). Most of these erroneous and atypical data are caused by the specifics of using a graphics tablet for the handwriting authentication of users. After analyzing the results of the performed experiments, the following five types of errors can be emphasized, which are caused by the specifics of the use of a graphics tablet for the handwriting authentication of users:

1. A sequence of points with zero pressure (except the first such point in each sequence).
2. Random points (in small amount).
3. Repeats, i.e. a sequence of consecutive points with unchanged coordinates along the X and Y axes (except for the case when one of the points has zero pressure).
4. Random small bends at the beginning of the lines.
5. Poor-quality item, discarded due to the inability to split the image of the password word for a given number of images of symbols (more often occurs as a result of lack of skills with the graphics tablet pen).

It should be noted that the algorithm for excluding erroneous data should be different at different stages of the authentication system. That is, some data that are erroneous at the stage of recognition of a written keyword can be distinctive characteristics of the style of writing a keyword by a specific user and, accordingly, can be useful at the second stage of the authentication system operation.

In addition to that, to improve the efficiency of the performed recognition in this work, the original data is corrected, plus this correction should be different at different stages of recognition. The need for this adjustment is caused by the specifics of the graphics tablet usage for the handwriting authentication of users. The data adjustment should be performed after nominal splitting of the whole image of the password word into images of separate symbols, but before the control points placing. The necessity to adjust the data is explained by the following facts. The recognition is complicated by the fact that authentication can be performed using different graphics tablets. Herewith the characteristics of the tablet can vary (the size of the workspace, the press sensitivity, etc.), that affect the values of the analyzed parameters. For example, if the size of the tablet workspace is various, then the size of the analyzed image of the password word and its location may be various too. Therefore, to provide the correct recognition, it is necessary to recalculate the image parameters in accordance with some neutral (without reference to any particular tablet) workspace of the selected size. To do this, in any handwriting authentication system, with the usage of the graphics tablet, there are needed the image scaling and shifting (moving) functions on each axis. Therewith, even if the same graphics tablet is used, some writings may be written not horizontally, but at an angle (for example, if the tablet is rotated relative to the user), some may be written in large letters, some – in small, some labels may not be centered, but shifted in some direction. Such writing features interfere with the recognition of the password. Therefore, the recognition system should have the image scaling, shifting (moving) functions on each of the axes and the image rotation function. At the same time, it should be noted that we need not just the image shifting, scaling and rotation functions, but such its varieties as shifting the image to the center of the screen, rotating the image to a horizontal position and scaling (stretching) the image to the full screen. The presence of these functions is necessary to be able to recognize exactly the symbols of the word-password, not its location. Otherwise, the probability of correct recognition of the password word symbols will be very small.

All these functions are necessary as for the recognition of the whole password word and so for the character-oriented recognition, but in the second case, all these functions should be applied not to the image of the whole password word, but to the images of particular symbols. That is, the image of each symbol first should be rotated to a horizontal position, then placed in the center of the used workspace of the selected size, and then proportionally stretch the image to the entire used workspace of the selected size. Otherwise, the location of the symbol will be recognized instead of itself.

Summing up, it can be said that this system requires the following three types of the character-oriented correction:

1. The character-oriented rotating of the symbols images for the normalization the angle of its axes inclination.
2. The character-oriented shift of the each symbol image to the center of workspace of selected size.
3. The character-oriented proportional scaling (stretching/compression) of each symbol image over the entire workspace of the selected size.

Summing up all the before-mentioned, it can be said that for the reliable recognition function *R* constructing it is necessary:

1. To make an optimal choice of individual biometric characteristics of a person to be used for authentication (the voice, the handwriting or the keyboard handwriting, the fingerprint, the retina, etc.).
2. To select the item recognition mechanism.
3. To determine the lists of *At* features that will be used for recognition (the number of consecutively performed recognition procedures can be one or more) during authentication, in accordance with the selected option.
4. To make a split of the entire password word image to the images of the particular letters.
5. To determine the criteria for the selection of analyzed samples.
6. To determine the necessary adjustments of the analyzed features.
7. To create an algorithm for placing control points.
8. To develop algorithms for solving the tasks of recognition to perform the authentication procedure.

To implement the above, there was written a program for performing the handwriting authentication of the information systems users, at the same time the neural network was used as a recognition mechanism. The results of experiments performed with the help of the written program showed the effectiveness of the proposed algorithm.

Conclusion

In this work, on the basis of the analysis, and taking into account that the developed biometric authentication system should be used for user accessing to the information system in cases not always associated with his work at the computer, and as an analyzed characteristic of a person his handwriting was chosen. There were analyzed the varieties of handwriting authentication methods and its characteristics were determined for their further use in the authentication process. Then the analysis of the selected handwriting characteristics was performed, in order to determine their validity for further recognition. After that, the selection of training handwritings items of information systems users was made to increase the probability of correct recognition. The adjustment of training handwriting items of information systems users was performed to increase the probability of correct recognition, too. To implement the above, a program was written to perform the handwriting authentication of information systems users, while the neural network was used as a recognition mechanism. A number of experiments were carried out with the help of the written program. Based on the results of the performed experiments, there were given the recommendations (requirements) on the data pre-processing for recognition and on the configuration the most critical system parameters.

Thus, it can be concluded that the usage of handwriting recognition systems is advisable for the implementation of authentication of information systems users.

References:

1. Arthur Galeev.: Almost all companies in the U.S. and Europe will use biometrics in two years., 30.03.2018. http://safe.cnews.ru/news/top/2018-03-26_pochti_vse_kompanii_v_ssha_i_evrope_budut_ispolzovat
2. Vysotska O., Davydenko A., “Classification of biometric authentication systems”, Collection of scientific works of the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol. 27, pp. 108-114, 2004.
3. Vysotska O., Davydenko A., “Determination of critical parameters when choosing a biometric authentication system”, Modeling and information technologies. Collection of scientific works of the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol. 27., pp. 80-86, 2004.
4. Vysotska O., “Assessment of quality of biometric authentication methods and the ways of its improvement”, Modeling and information technologies. Collection of scientific works of the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol. 28, pp. 94-102, 2004.
5. Vysotska O., “Selection of analyzed characteristics when handwriting authenticating of computer systems users at different stages of computer technology development”, Modeling and information technologies. Collection of scientific works of

the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol.56. – pp. 31-39, 2010.

6. Vysotska O., Davydenko A., "Analysis of data pre-processing technology when authentication of computer systems users by the keystroke pattern and handwriting", Modeling and information technologies. Collection of scientific works of the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol.55. – pp. 34-41, 2010.

7. Kallan R.: Basic concepts of neural networks. Translate from English. Publishing house "Williams", 2001.

8. Vysotska O., "The problem of recognition of the written keyword as one of the problems solved when performing handwriting authentication of of computer systems users", Modeling and information technologies. Collection of scientific works of the Pukhov Institute for modeling in energy engineering of NAS of Ukraine, Vol.36. – pp. 67-76, 2006.