Algorithm for personal identification in distance learning system based on registration of keyboard rhythm

P V Nikitin¹ A N Savinov², R I Bazhenov³ S V Sivandaev⁴
¹ Mari State University, 1, Lenin Square, Yoshkar-Ola, 425400, Russia
² Volga State University of Technology, 3, Lenin Square, Yoshkar-Ola, 425400, Russia
³ Sholom-Aleichem Priamursky State University, Shirokaya 70a, Birobidzhan 679015, Russia
⁴ Interregional Open Social Institute, 28, Prokhorova St. Yoshkar-Ola, 424007, Russia

E-mail: petrvlni@rambler.ru

Abstract. The article describes the method of identifying a person in distance learning systems based on a keyboard rhythm. An algorithm for the organization of access control is proposed, which implements authentication, identification and verification of a person using the keyboard rhythm. Authentication methods based on biometric personal parameters, including those based on the keyboard rhythm, due to the inexistence of biometric characteristics without a particular person, are able to provide an advanced accuracy and inability to refuse authorship and convenience for operators of automated systems, in comparison with other methods of conformity checking. Methods of permanent hidden keyboard monitoring allow detecting the substitution of a student and blocking the key system.

1. Introduction
In the process of all levels of education, distance learning systems are frequently introduced, with the help of which the levels of the formation of the corresponding competences are determined [1]. But, in most cases, in these systems, there is no definition of the student's personality, which reduces the quality of training as a whole.

Authentication methods based on biometric personal parameters, including those based on the keyboard rhythm (KR), due to the inexistence of biometric characteristics without a particular person, are able to provide an advanced accuracy, inability to refuse authorship and convenience for operators of automated systems, in comparison with other methods of checking conformity. Methods of permanent hidden keyboard monitoring can detect the substitution of a legitimate operator and block the key system (KS) from intrusion by an attacker. Thus, the task for researching models, methods and algorithms of recognizing the keyboard rhythm of key system operators is relevant at the moment. The main characteristics of the KR are the time of holding keys and the time between keystrokes. In the process of developing a system for analyzing the KR of key system operators, it is necessary to develop algorithms for registering the operator (training), authentication and authorization of the operator, as well as verification and identification.

2. Materials and methods
Keyboard rhythm is a set of dynamic characteristics of keyboard operation. The standard keyboard
allows one to measure the following time characteristics: time of holding the key pressed and the time interval between keystrokes. As it was revealed in the work [2], using the algorithms for recognizing the time of holding a key (THK) and the methods of recognizing KR using a floating text, it is possible to build a system of permanent hidden monitoring that allows verifying the legitimate operator.

The time of holding a key is the period during which the key is clicked. The software measures this indicator from the moment the key is clicked (the "onkeydown" event) until it is released ("onkeyup" event). This parameter is usually expressed in milliseconds. The average time of holding a key is the mathematical expectation of a selection of samples of the periods of time holding a particular key collected over the period of typing the text. Empirical studies of the base of standards of KR operators showed that the time of holding the key also depends on the overlap, rhythm and accuracy. The average time of holding the key can vary significantly for different people at close typing rates because of the difference in the technique used by the typist.

Overlapping - simultaneous movements of several fingers of the typists who can type efficiently. The overlapping of clicking keys occurs when one key is not released yet, and the other is already clicked. There is a tendency to increase the number of overlaps with an increase in the speed of typing. The overwhelming majority of overlapping occurs when the keys of neighboring letters in the word are clicked by different fingers. However, with a very fast typing with sliding movements, overlapping is also possible.

Overlapping occurs for the following reasons:

- high speed of typing, when the typist does not have time to release the previous keys before clicking the next;
- long time of holding keys;
- combination of the first and second factors.

It has been established that with the same typing speed, as a rule, there are more overlaps on setting "SDFV NJKL," than on "ASDF JKL;" and with dynamic typing - more than when typing by zones. The increase in the number of overlaps is due to the increase in the number of combinations that are pressed by different fingers. It was revealed that the more abruptly the click is and the more rhythmic the typing is, the less the number of overlaps is. When the keys are clicked more than they hit, and the rhythm is low, the number of overlaps is more. If the typing is very jerky, the time of holding the key can be 65 ms or less, and typing with a large number of overlaps - 120 ms or more. The average time, as a rule, is 80-100 ms. Figure 1 shows an example of typing with overlaps.

Three types of overlaps can be distinguished:

1. At the moment, the first key is held, the second key is clicked. The "K1" key is clicked. Next, one clicks the "K2" key, but "K1" has not yet been released. Then, the "K1" key is released, then
the "K2" key is released.
2. At the moment of holding one key, another key is released, i.e. the first key was clicked while the second key was held. "K2" is clicked. "K1" is being clicked. After this, the key "K2" and then "K1" are released.
3. Clicking and releasing the key occurs while holding another key. The "K2" key is clicked, "K1" is being clicked and released and then "K2" is released.

Empirical studies of the keyboard rhythm of operators showed that the THK during typing is bimodal (the intersection of two normal), rather than the Gaussian distribution. Thus, the methods and models proposed for determining the THK as a mathematical expectation of the normal distribution have a significant error and do not reflect the actual dependencies of the typing process.

Thus, using a simple mechanism for determining the presence of overlaps of keys while typing, it is possible to obtain two independent samples and two values of the mathematical expectation of the THK for each key, which characterize the KR of the operator of the KSII, which in its turn increases the accuracy of identifying the operator, i.e. will reduce the possibility of errors of 1 and 2 kind. In this case, each of the samples obeys the normal distribution law, and therefore the mathematical model of the THK developed in is applicable to it.

3. Algorithm for registering keyboard rhythm
The construction of biometric identification systems is based on the creation of sample representations of the persons identified. The sample is created, when the system is in training mode. It represents a stored in-memory access control system, a person's biometric characteristics, and is used to compare with the biometric parameters of persons claiming access to resources. In the case when parameter values of the user defined by the system differ from the sample more than allowed by the detection threshold, it receives a denial of access to resources.

The operation of the algorithm can be represented by an activity diagram (Figure 2). A probability-statistical method is used to detect the averaged values of the time events of the keyboard, so it is necessary to collect statistics consisting of a sample of time values, where the sampling time is the time of holding the key. The algorithm is based on the mathematical model developed in [3].

In this algorithm, the process of obtaining an information file is performed - obtaining a reference sample of the operator's KR. To do this, at the beginning the operator is identified by his unique identifier, for example, the login.

The dynamic three-dimensional array KeyEventsArr is initialized, in which keyboard events are stored. The array is filled in until the sufficient number of keys (specified by the administrator or by default) is clicked.

The next step is to count the time of holding the keys. There is an event of clicking a particular key. Then there is an event of releasing this key. The click event is subtracted from the release event click and it is divided by the high-resolution counter frequency in order to get the THK value in milliseconds. Clicks of events are determined by the QueryPerformanceCounter function, the counter frequency by the QueryPerformanceFrequency function [4]. On multi-core systems, the SetThreadAffinityMask function is used to indicate the relationship of the processor to the system.

The algorithm involves a filter of long keystrokes used by the typist to enter the n-gram of the same letters, for example "mm" in the word "program" or "SSS" in the "USSR". Also, in the algorithm there is a filter for pressing system keys (for example, BACKSPACE or ENTER), the pressing of which is not saved in the sample.

Depending on the presence or absence of overlapping when the key is held, the THK value is entered in the sample of the first (without overlaps) or the second (with overlaps) normal distribution. Then the mathematical expectation of each sample is calculated, and the reference sample of KR is stored in the operator account.

Since the floating text, which is used for training or further authentication and identification procedures, has a different probability of occurrence of different letters and symbols [5], the collection of a large enough THK sample for carrying out the probability-statistical analysis requires the input of
a very large text (from 5000 symbols and more). Therefore, it is suggested to enter the "quantity" and "sample" fields in the information file.

Figure 2. A diagram of the algorithm for registering KR.
In the field "sample", the THK, separated by the ";" sign, is stored in the "quantity" field, the number of elements in the sample is entered. When a sufficient number of elements are reached in the sample, the THK is counted, the sample is cleared, and the "!" symbol is entered in the "quantity" field, which is used to indicate the fact of the end of counting the THK of this key. Further elements will not be entered in this sample. If a sufficient number of sample elements is not collected during the training, the temporary value of the THK will be calculated, and the process of the collection and calculation of the THK will continue with the operation of the authorization and monitoring algorithms, upon confirmation that the text is typed by the legitimate operator to which this sample corresponds. Completed THK of the keys will have the advantage for authentication and identification; they will be given the large influence factors proposed in [6], the unfinished ones - the smaller ones. For the separation, it is suggested to divide the alphabet of the system into several groups, depending on the probability of occurrence in the texts.

4. Conclusion
It is suggested to use the bimodal distribution as a representation of the THK instead of the normal distribution in the analysis of the keyboard rhythm. Bimodal distribution is proposed to be divided into two normal ones, which will allow us to apply a mathematical model of the keyboard rhythm based on the Gaussian distribution.

An algorithm for recognizing and registering the operator's keyboard rhythm has been developed. The training algorithm allows preserving the biometric characteristics of the keyboard rhythm as a bimodal distribution.

The introduction of this algorithm into the distance learning process will allow determining the personality of the student and monitoring its learning path, which will significantly improve the quality of training.

References
[1] Nikitin P V 2017 Proceedings of the 30th International Business Information Management Association Conference (IBIMA) 5533-5539
[2] Savinov A N 2011 Vestnik of Volzhsky University after V. N. Tatishchev 18 115-119
[3] Vasilev V I 2016 Scientific and Technical Information Processing 3 21-37
[4] Gatchin Yu A 2012 Sci. Tech. J. Inf. Technol. Mech. Opt. 2 92-95
[5] Korobeinikov A G 2010 Bulletin of the Mari State Technical University. Series: Radio-technical and infocommunication systems. 2(9) 91-98
[6] Gorbunova E S 2016 Cybernetics and Programming 4 65-72