A Multimodal Biometric User Identification System Based on Keystroke Dynamics and Mouse Movements

Piotr Panasiuk, Maciej Szymkowski, Marcin Dąbrowski, Khalid Saeed

A Multimodal Biometric User Identification System Based on Keystroke Dynamics and Mouse Movements

Piotr Panasiuk[1], Maciej Szymkowski[2], Marcin Dąbrowski[2], Khalid Saeed[1,2]

[1]Faculty of Mathematics and Information Science,
Warsaw University of Technology, Warsaw, Poland

p.panasiuk@mini.pw.edu.pl
k.saeed@mini.pw.edu.pl

[2]Faculty of Computer Science,
Bialystok University of Technology, Bialystok, Poland

szymkowskimack@gmail.com
marcin.dabrowski@poczta.fm

Abstract. In this work it is shown how the behavioral biometrics allows to strengthen security of a personal computer during casual use. The user does not have to be even aware of verification system running in the background. Unfortunately, short passwords do not supply enough data for keystroke dynamics algorithms to be precise enough to keep the way and level the biometrics system requires. Behavioral biometrics cannot grant such authentication level as the other physiological biometric methods, e.g. fingerprint or retina scan. However, their transparency in analyzing data allows to merge methods into multimodal systems with a minimal cost. The benefit of keystroke dynamics is that it can be easily connected with some other biometric methods, especially with other human input interface devices. In this paper an approach to analyze keystroke dynamics along with mouse movement is presented. Even though both of the features are of behavioral character and hence with low repeatability, the results are good and promising for further research and modification.

Keywords: keystroke dynamics, mouse, biometrics, behavioral biometrics, authentication, systems security, multibiometrics, fusion, multimodal system.

1 Introduction

Today data safety is one of the most discussed terms. People need to prove who they really are at every turn. This includes banking, healthcare, communication and much more. *Something you know* and *something you have* - these are the most common methods used to prove your identity. You know your password, but you may forget it if you are not using it often or have too many of them to remember. Things such as tokens or cards can be used instead and they let you free of remembering sophisticated sequences of various letters, numbers and other special characters. The thing is,

tokens and cards can be lost, stolen or even destroyed quite easily. Thus, another way for authentication is needed. Here comes biometrics. *Something you are* cannot be lost or forgotten. These are based on human behavioral and physical characteristics that can be measured and cannot be easy to imitate. Physiological features may include fingerprint, DNA, hand geometry, retinal scan and others that come from how organisms are built. Behavioral features on the other hand are based on how people do things, for example voice (the way one talks), gait (the way one walks) or keystroke dynamics (the way one strikes or touches the keys on the keyboard), and so on.

This paper focuses on behavioral methods. They are cheaper in implementation, usually do not require specialized hardware and often work without bothering or notifying the user. On the other hand these features are often hard to repeat in exactly the same way. This introduces information noise. Both valid users may find it difficult to repeat the activity in the same manner to fit into their patterns and from the other side impostors may also be close to imitate valid data and be falsely accepted. The goal is to find a method that despite those difficulties will make the right decision with the highest possible accuracy.

One of the ways to make biometric algorithms more robust is to join multiple biometric features. This way the algorithm gets more data to analyze what helps it in correct classification. Thus this paper proposes a method to combine multiple behavioral features to provide greater reliability and safety in computer systems. In this particular case mouse movement and keystroke dynamics were chosen as they are not very involving to the user and are often naturally used together. What is more they use standard equipment of every personal computer nowadays.

2 Known approaches

Lately more interest in the field of behavioral biometrics has been observed. To account this, the authors of this paper decided to present some of the recent approaches in mouse and keystroke dynamics.

In [1] an interesting approach regarding mouse movement has been described. Authors analyzed user online activities by tracking mouse movements across web interfaces in certain areas of interest. One of three user activities was being recognized. Hidden Markov Models and Conditional Random Fields were used in the process. 51 students performed one of three tasks twice. Tasks were based on memorizing graphical representation of a given quadratic equation or discovering and memorizing intersection coordinates of two given quadratic equations. HMM and CRF models performance was evaluated using ratio of ground truth matches number of observation sequences to whole number of test sequences. Distance in pixel of the vicinity extent has been found accuracy determining as its higher value generally resulted in improved accuracy. For distance equal 0, HMM and PCRF models using classical observation sequences gave recognition rate of 88.24%.

Another interesting approach regarding mouse movement has been described in [2]. Authors presented a method for user emotional state prediction basing on mouse dynamics. To collect data, authors created a simple computer game that required users

to click differently sized and colored rectangles in correct order since they were placed randomly. Samples from 262 users were gathered of which 44 users were asked after session, about their emotions during task. Two different states were distinguishable. Features including distance and direction were extracted from mouse logs among others. Classification was performed using: Logistic regression, Support Vector Machine, Random Forest, and C4.5. Authors used 10 fold cross validation for evaluation, dividing whole data into 10 random parts of same size, next using 9 for training and remaining one for validation. This setup allowed authors to obtain accuracy of 94.61% (for SVM) in prediction of user emotional state basing on mouse movement and knowing target. Without knowledge of user emotional state during collecting data, accuracy dropped to 82.38%.

Authors of [3] proposed a novel multimodal biometrics user verification technique based on keystroke dynamics and mouse movement. They focused on several layers of mouse events. This aspect seems interesting and worth further research. Authors claim that they have very good accuracy. The technique seems to be advanced however not all tests were performed and authors do not present the FAR level of the solution. FRR is quite low at range of 3.2% but in case of verification system, this value may be adjusted to any level by the cost of increasing the opposite metric - FAR. The classification method of the presented solution is not precisely described. Moreover, the results were calculated on unknown database that is not publicly available (state for April 23, 2016). Thus the results and accuracy cannot be confirmed.

A fuzzy approach based method on commands typed by users was considered by the authors in [4]. Authors presented a way to detect impostor by creating two different user activity profiles, local one based on recent activity of the user and one combining multiple local profiles representing user general computer behavior. Authors used publicly available SEA data set consisting of system calls made by 70 users, giving a total of 15000 recorded commands. 50 randomly chosen users were considered legitimate, and 20 were taken as impostors. The lowest FRR ratio of 0.8% paired with FAR equal to 70.1 was obtained by the mentioned method. Taking low computational complexity into account, this method can successfully be used in real-time.

Another keystroke dynamics based approach was presented by authors of [5]. In that paper user password typing dynamics was observed. Fuzzy sets were used to construct user model. Data for 51 users, each one typing the same password 400 times in 8 sessions of 50 tries, contained in publicly available database *Keystroke Dynamics Benchmark Data Set* by Kevin Killourhy and Roy Maxion [6] was used to perform experiments. Using proposed method allowed to obtain EER value of 9.2% - an improvement over original [6] methods giving best EER of 9.6%. According to authors, results improved greatly due to data normalization.

The authors of this paper also have achievements in the field of biometric methods. Apart from the algorithms suggested on the basis of physiological features, which is not the subject of this paper, many other behavioral approaches were introduced [7-14]. The mouse movement was considered for the first time in 2005 [7]. The work introduced then comprised a new method for analyzing biometric features for human authentication. The rhythm of the movement is individual and characteristic for each person, so it can be used for identification in small defined groups or verification for

larger groups of users. The method analyzes the dynamics of the mouse cursor movement. The processed signal is the cursor changing speed obtained during the random movement of the mouse. This signal is transformed into frequency domain with Discrete Fourier Transform and then analyzed by Toeplitz matrix minimal eigenvalues method [7]. The resulting feature vector is used for classification performed by two methods: *k*-nearest neighbors and NN - artificial neural networks. The obtained results were promising and showed a large possibility of integrating the method with other features in multimodal biometrics systems.

The authors' team performed other multiple approaches on keystroke dynamics user identification over the past few years. Algorithms take into account many features including: dwell and flight times, average keystrokes per minute, overlapping specific keys, typing errors, the way of error correction and others. Using simple 1-NN classifier resulted in accuracy of 75.68% on 37 users [8]. Later approach used improved algorithm based on *k*-NN classifier on authors' database consisting of samples left by over 250 individuals. Gathered samples included one-word phrase and longer sentences in Polish and English. It was proven that even small number of samples may be enough for successful recognition with the high user amount and right choice of phrase. Best classification accuracy of 90.83% was obtained with 21 users [9]. Other algorithm modifications were conducted in the next works. On fixed-text approach high accuracy of 98.78% was obtained for 16 users, although decreasing with greater number of users (e.g. 72.3% for 79 individuals) [10]. Database quality, however, was taken into account in the following approaches.

In [11] the authors used their own and Maxion-Killourhy's [6] databases with self-developed improved algorithm allowing to discard samples with errors. Said method allowed accuracy increase of 3.6% for Maxion's data and increase of 5.6% for authors' database in comparison to initial values. Next [12] authors further analyzed database impact on results. Data gathering precision and conditions along with various algorithm modifications allowed authors to deeply compare mentioned databases and classification methods. This research lead to conclusion that databases with longer samples are more suitable for user identification than authentication and inclusion of user-specific imperfect samples can improve FRR. Additionally, updating training set over time is believed to affect classification accuracy in a positive way. Authentication by non-fixed text of various length was also taken into consideration by the authors [13]. Data were gathered over the Web using browser application and also locally with the use of dedicated applications. Sample length provided to be vital on recognition accuracy as longer texts generally allowed to obtain better results. Using statistical characteristics of the sample gave better outcome than using raw sample data, eventually leading to EER value of 6.1% for 200-keystroke long samples. Comparison of the Keystroke Dynamics databases was conducted by the authors in [14]. In said work newly-gathered database was presented and compared with existing one which is publicly available [6]. Authors collected it in the way it can be directly compared with *Keystroke Dynamics Benchmark Data Set*. That led to two databases being almost identical. It was possible by using the same phrase typed equal number of times by every user of both databases, i.e., 400 valid samples in 8 sessions. Main difference introduced with authors' database was that its data were collected in unsu-

pervised, less restrictive conditions with the use of commonly available devices when Maxion's database was supervised and used specific high-precision devices. The use of the same algorithms on both databases resulted in differences in the outcome ranging up to about 30% in some situations, which led to the conclusion that new algorithms should be tested on multiple databases, including publicly available ones and not limiting to the ones gathered by the authors' for the specific research purpose. Moreover, less restrictive method of collecting data allowed to obtain generally higher recognition accuracy.

3 Proposed approach

In order to get data a web application has been created. It is located under the Internet address [15]. Users have to register in order to create their unique account. During the registration process no biometric data are being collected. Then after logging in they can leave their biometric samples. Authors encourage everyone to visit our system and contribute to the database. User details like email address are stored only to remember the user and allow him to reset the password. The database is meant to be published online. More details will be available on the mentioned website in the near future as the samples set grows rapidly.

A sample in our database consists of two phrases that a user has to type and the mouse data recorded as an interaction with the user interface. The first phrase is a fixed text. It imitates fairly strong password "_Y9u3elike22". It is common for all users. The second phrase is a free-text phrase that a user comes up with spontaneously while typing it. Its only limitation is that it has to be not shorter than 80 characters and not longer than 4000. When it comes to mouse movement data are gathered in a raw form - events like button press, button release and move. Each data has a timestamp from the beginning of the sample and coordinates of a cursor on the web page. Keystroke events are recorded within each text field; however mouse events are being recorded for whole duration of leaving a complete sample. This means since pressing the first button, through selecting each text field and pressing additional button, until submitting a form by clicking the last button on the data acquisition web page.

Data examined during this research were only fixed phrase and mouse activity. From these data authors had to extract the most valuable features. Samples were gathered in unsupervised conditions so an algorithm for various corner-cases was applied. When it comes to mouse data single mouse moves are being extracted. Mouse move is considered as cursor position change from the beginning of a move until button press. Due to unwanted cursor movements during releasing the mouse by the user, authors decided to ignore move events after mouse click and before typing a text. Finally, the authors examined few mouse movement features - move time, move speed (in pixels per second) and move distance. As a separate feature mouse button dwell times while clicking buttons or text fields are accounted. When it comes to keyboard data we have tried to extract flight times and dwell times as they are the most common keystroke features, but to our surprise they did not perform well with success rate at about 11% using approach and setup presented in the following para-

graph. It might be the case that the sample was really difficult, especially the first half of it. Some additional difficulties came up with shift key being pressed different amount of times. Authors tried to implement an algorithm that would deal with those differences however the overall recognition rate was really affected by those artifacts. So the dwell times of specific key presses were analyzed instead. They have proven to give really good accuracy which is presented in the next chapter.

After defining feature vectors our next step was the identification process. In this purpose authors used fast and simple *k*-NN algorithm. The most important part is the distance calculation. There are three features that are being extracted from both keystroke and mouse data. A *keyboardDistance* (1) is a distance calculated using Manhattan metrics between corresponding dwell times of a training and a testing sample. In a case of mouse data represented by *mouseDistance* (2) there are mouse key dwell times, the metric used is Manhattan metric as well. The last feature is *moveDistance* (3) which proved to be the best metric to calculate the moves and is defined by the Euclidean distance from the move start point to the move end point in a two-dimensional space according to the move definition in a foregoing paragraph.

$$keyboardDistance = \frac{1}{k}\sum_{i=0}^{k}|dwellA_i - dwellB_i| \tag{1}$$

$$mouseDistance = \frac{1}{m}\sum_{i=0}^{m}|mouseDwellA_i - mouseDwellB_i| \tag{2}$$

$$moveDistance = \frac{1}{n}\sum_{i=0}^{n} dist(moveA_i, moveB_i) \tag{3}$$

As one can see each distance is normalized. To calculate the final distance between two samples authors use equation (4). What should be explained is that *keyboardWeight* is the importance factor of keyboard event. Mouse events importance factor is *mouseWeight* and the *moveWeight* is the importance factor of the mouse moves.

$$distance = \frac{kD \cdot kW + mD \cdot mW + moveD \cdot moveW}{kW + mW + moveW} \tag{4}$$

where:

kD - keyboardDistance
kW - keyboardWeight
mD - mouseDistance
mW - mouseWeight
moveD - moveDistance
moveW - moveWeight

Mentioned weights have been selected empirically. The detailed information about the experiment setup and preparation is described in the following chapter. After defining the distance the classic *k*-NN algorithm is being followed.

4 Results of the experiment

Authors' database has irregular amount of samples per user. This is due to constant growth of the dataset. Because of this the users that does not get sufficient amount of samples have been removed from the experimental setup. In our *k*-NN-based method the minimum number of samples per user is *k*. This number allows to prepare a valid training set having exactly *k* samples for each user. The rest of remaining user samples are used for a testing purpose. Finally for a $k = 10$ there are 50 valid classes in each of the experiments. This is why all setups has been limited to those 50 classes. So the users set among all experiment runs are the same. To get reliable results each experiment was repeated 100 times for each setup, every time using randomly selected training and testing samples.

At the beginning authors had to select proper feature weights to get the best results. This way the algorithm has been run for each of three features separately and authors got the information about accuracy of each of the methods. The weakest method was mouse dwell times alone with average success rate of about 12%. The second one that was mouse move distances gave the success rate around 31%. It is worth mentioning that while analyzing move speeds in pixels per second instead of move distances the accuracy has dropped to the level of 8%. Finally the best was keystroke dynamics with around 44% of accuracy. In order to obtain the best weight values authors picked the two weakest features and joined them using different weight ratio.

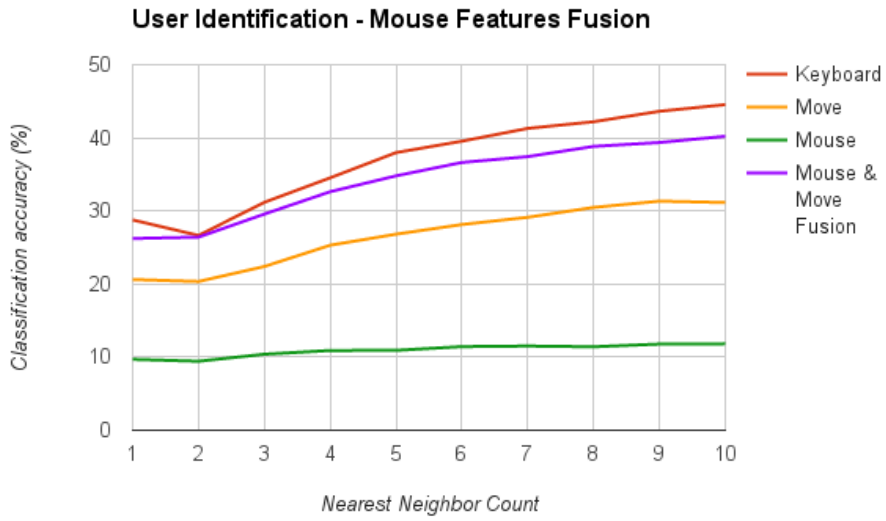

Fig. 1. Influence of joining mouse button dwell times and mouse moves on algorithm accuracy.

After conducting the experiment with weights ranging from 0.0 to 1.0 with step of 0.05, the best ratio proved to be 0.8 for *mouseWeight* vs. 0.2 for *moveWeight* that returned the accuracy of around 40% for mouse data only.
Figure 1 presents the results of classification accuracy after joining mouse dwell times and mouse movement features in comparison to keystroke dynamics alone. Having these results authors tested different *moveWeight* values and the results shown that the big increase can be gained. The accuracy of the algorithm in a setup where *keyboardWeight* = 2, *mouseWeight* = 0.4, and *moveWeight* = 0.1 returned the identification accuracy of 68.8% for 50 classes. This result has exceeded authors' expectations for such little of data. In Figure 2 one can see results of our experiments. Different values of parameter *k* (from 1 to 10) were taken into account whereas count of classes was stable and equal to 50. Each method accuracy has been marked separately and the optimal fusion method results are also presented.

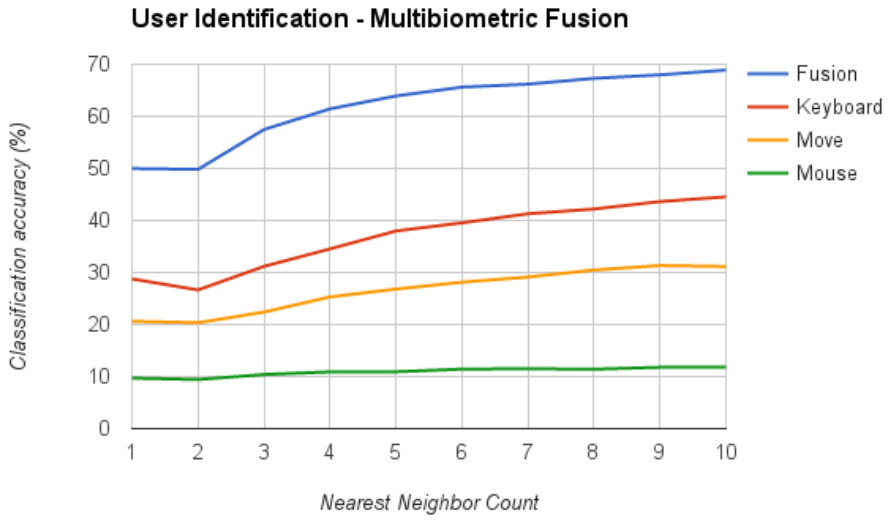

Fig. 2. Chart presenting classification accuracy using different approaches along with different *k*-nearest neighbor value used.

5 Conclusions and Future Work

Biometric methods present a very convenient way to harden the computer system security. Even if the user knows the password it is really hard for him to repeat it in the same manner to breach the security. If it comes to identification, it is even a harder task because a user does not claim his identity. Thus the whole database has to be searched for matching the user pattern. Identification algorithms have to be fast and robust. Presented in this paper *k*-NN algorithm perfectly matches these requirements.

As it was presented, user identification by using mouse button dwell is not very reliable. When mouse moves are analyzed, the accuracy increases significantly. When both mouse button press and mouse move features are used together they are almost as effective as keystroke dynamics using dwell times. What was surprising in this experiment is the fact that flight times and dwell times which usually make the user typing features more persistent in time, resulted in a huge accuracy drop (from 44% to merely 11%). This is possible due to the short sample length and few ways to type the phrase correctly (using the shift key).

As expected, combining both keystroke dynamics and mouse features allowed to obtain much better recognition ratio than relying on them separately. Accuracy of 68.8% as a result of fusion is quite impressive, taking into account the high number of users in relation to quite short samples. Fusion of different biometric algorithms gives us a great advantage at low values of *k* which significantly decreases time required to prepare a user profile (training set). When the authors examined the speed of mouse movements, the success rate dropped dramatically (level of 5-8%, depending on the setup). There might be some inconsistencies in the mouse samples (users made mistakes, missed the button, etc.). However, unexpectedly strong feature turned out to be the position where user parks the mouse cursor and the position where he clicks items on the user interface.

Our database is continuously expanding as the existing users are leaving more of their samples, and new users are willing to help in the research. The authors encourage everyone to participate. The database will be publicly available online. For more information the reader can track the information given in the authors' system website [15]. In the near future the authors are planning to take into consideration more mouse-specific characteristics in addition to clicks and moves distances currently used in our algorithm. Authors believe that the analysis of other behavioral aspects will definitely improve the accuracy. Examples that are worth examining are: rapid mouse movements during mouse button press, cursor fixation on a target, mouse movement when user releases the device, and other issues that may come up during the research. Additionally, as an extension to this research authors would like to introduce some decision algorithm and understand user mistakes for better handling of unusual users and data anomalies.

Acknowledgments

This work was partially supported by grant number S/WI/1/2013 from Bialystok University of Technology and funded from the resources for research by Ministry of Science and Higher Education. It was also partially supported by Neitec company.

References

1. Elbahi, A., Omri, M.N., Mahjoub, M.A., Garrouch, K.: Mouse Movement and Probabilistic Graphical Models Based E-Learning Activity Recognition Improvement Possibilistic

Model. In: Arabian Journal for Science and Engineering, pp. 1-16. Springer Berlin Heidelberg (2016)
2. Pentel, A.: Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State. In: 5th International Workshop on Personalization Approaches in Learning Environments in conjunction with 23rd Conference on User Modelling, Adaptation and Personalization, CEUR Workshop Proceedings, vol. 1388, pp. 40-45. Dublin, Ireland (2015)
3. Motwani, A., Jain, R., Sondhi, J.: A Multimodal Behavioral Biometric Technique for User Identification using Mouse and Keystroke Dynamics. In: International Journal of Computer Applications, vol. 111, no. 8, pp. 15-20 (2015)
4. Kudłacik, P., Porwik, P., Wesołowski, T.: Fuzzy approach for intrusion detection based on user's commands. In: Soft Computing, pp. 1-15. Springer Berlin Heidelberg (2015)
5. Kaganov, V.Yu., Korolev, A.K., Krylov, M.N., Mashechkin, I.V., Petrovskii, M.I.: Machine Learning Methods in Authentication Problems Using Password Keystroke Dynamics. In: Computational Mathematics and Modeling, vol. 26, no. 3, pp. 398-407. Springer US (2015)
6. Killourhy, K.S., Maxion, R.A.: Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In: Dependable Systems & Networks, Lisbon, Portugal, pp. 125-134. IEEE (2009)
7. Tabędzki, M., Saeed, K.: Nowa metoda do badania dynamiki ruchów myszy komputerowej do celów identyfikacji. Krajowa Konferencja Naukowa - KBIB'05, tom I, Systemy Informatyczne i Telemedyczne, Wydawnictwa Politechniki Częstochowskiej, Częstochowa, Poland, 2005, 467-472 (in Polish)
8. Rybnik, M., Tabedzki, M., Saeed, K.: A Keystroke Dynamics Based System for User Identification. In: Computer Information Systems and Industrial Management Applications, pp. 225-230. IEEE (2008)
9. Rybnik, M., Panasiuk, P., Saeed, K.: User Authentication with Keystroke Dynamics Using Fixed Text. In: International Conference on Biometrics and Kansei Engineering, Cieszyn, Poland, pp. 70-75. IEEE (2009)
10. Panasiuk, P., Saeed, K.: A Modified Algorithm for User Identification by His Typing on the Keyboard. In: Image Processing and Communications Challenges 2. Advances in Intelligent and Soft Computing, vol. 84, pp. 113-120. Springer, Heidelberg (2010)
11. Panasiuk, P., Saeed, K.: Influence of Database Quality on the Results of Keystroke Dynamics Algorithms. In: Computer Information Systems – Analysis and Technologies. Communications in Computer and Information Science, vol. 245, pp. 105-112. Springer Berlin Heidelberg (2011)
12. Rybnik, M., Panasiuk, P., Saeed, K., Rogowski, M.: Advances in the Keystroke Dynamics: The Practical Impact of Database Quality. In: Computer Information Systems and Industrial Management. Lecture Notes in Computer Science, vol. 7564, pp. 203-214. Springer Berlin Heidelberg (2012)
13. Rybnik, M., Tabedzki, M., Adamski, M., Saeed, K.: An Exploration of Keystroke Dynamics Authentication using Non-fixed Text of Various Length. In: International Conference on Biometrics and Kansei Engineering, pp. 245-250. IEEE (2013)
14. Panasiuk, P., Dąbrowski, M., Saeed, K., Bocheńska-Włostowska, K.: On the Comparison of the Keystroke Dynamics Databases. In: Computer Information Systems and Industrial Management, Lecture Notes in Computer Science, vol. 8838, pp. 122-129. Springer Berlin Heidelberg (2014)
15. Authors' system website: http://www.ikds.metna.net. Accessed April 26, 2016