RΣVISTΛ ΣJΣF
ISSN: 2965-1395

KEYSTROKE DYNAMICS FEATURES IN FORENSIC IDENTIFICATION: THEORETICAL AND EXPERIMENTAL APPROACHES

RECURSOS DE DINÂMICA DE DIGITAÇÃO NA IDENTIFICAÇÃO FORENSE: ABORDAGENS TEÓRICAS E EXPERIMENTAIS

Anna D. Tsvetkova*
Dmitriy V. Bakhteev**

ABSTRACT

This article discusses the use of typing dynamics, or individual typing style, as a form of biometric identification. It is found that keystroke dynamics has great forensic identification potential, particularly in the detection and investigation of cybercrimes. The study examines different characteristics of keystroke dynamics, such as key press time and release latency, through an experiment that collects typing data from 9 students. The results suggest that, although keystroke dynamics can be influenced by factors such as keyboard type, a set of typing characteristics can be used to accurately identify individuals, paving the way for advancements in the forensic application of this technology.

Keywords: keystroke dynamics; keystroke dynamics features; free-text systems; digital criminalistics; cybercrime; forensic identification; behavioral biometrics; keylogger software.

RESUMO

* Junior Researcher of the Associate of the Autonomous non-commercial organization “Centre for Assistance to the Development of Criminalistics ‘CrimLib’”. *E-mail*: at@crimlib.info.
** Doctor in Law, Associate Professor, Professor of the Department of Criminalistics, Head of the Laboratory of the Digital Technologies in Criminalistics, Ural State Law University named after V. F. Yakovlev. *E-mail*: dmitry.bakhteev@gmail.com.

Este artigo discute o uso da dinâmica de digitação, ou estilo de digitação individual, como uma forma de identificação biométrica. Foi descoberto que a dinâmica de digitação tem grande potencial de identificação forense, particularmente na detecção e investigação de crimes cibernéticos. O estudo examina diferentes características da dinâmica de digitação, como tempo de pressionamento de tecla e latência de liberação, por meio de um experimento que coleta dados de digitação de nove alunos. Os resultados sugerem que, embora a dinâmica de digitação possa ser influenciada por fatores como o tipo de teclado, um conjunto de características de digitação pode ser usado para identificar indivíduos com precisão, abrindo caminho para avanços na aplicação forense dessa tecnologia.

Palavras-chave: dinâmica de digitação; recursos de dinâmica de digitação; sistemas de texto livre; criminalística digital; crime cibernético; identificação forense; biometria comportamental; software keylogger.

1 INTRODUCTION

Biometric technologies are becoming more and more discussed every year among scientists and in political circles (Vacca, 2007). This trend is caused, on the one hand, by the significant identification potential of biometric data, i.e. biological and physiological characteristics of a person, and, on the other hand, by the emergence of widely available technologies for their recording and subsequent establishment of their belonging to a particular individual.

At the same time, in scientific research and at the level of applied use are used mainly the so-called physiological (static) biometric data, which are stably present in a person and characterize his appearance or internal structure. Such biometric data include: face, iris, ear pattern, vein pattern, papillary patterns, DNA, etc. However, no less promising for solving identification problems are behavioral (dynamic) biometric characteristics, which are manifested exclusively in the process of any human activity. They include: gait, voice, characteristic of movements when performing a handwritten signature, keystroke dynamics, etc.

We will not dwell on the description of all promising biometric studies and will consider in detail only keystroke dynamics, which seems important in connection with the increasing practice of committing computer crimes, including through the creation and distribution of texts (extremist materials, defamatory publications, destructive content, fraudulent emails, etc.) (Kente, Ishaku, 2024, p. 117-119). Today the practice follows the path of selecting alternatives in the form of author's identification or interrogation of suspects, which cannot guarantee an accurate quantitative result.

In business processes it may also be important to accurately determine who entered specific information into electronic reporting documents, prepared a plan-project for the development of the corporation, or to control the compliance of employees with the labor and rest regime.

In order to solve these problems, as well as a wide range of other unspoken ones (Wyciślik, Wylężek, Momot, 2024) we consider it necessary to refer to the study of keystroke dynamics.

This biometric characteristic today can be presented in two ways: as a set of certain habits and skills of a person and as a digital trace. It was first described in the 1970s (Spillane, 1975). Even then, the high identification significance of this characteristic was already established, to the point that by the end of the last century there were suggestions that its research should be prioritized over facial image identification (Rastorguev, 1993, p. 64). As we can see, these predictions did not come true, but the research on keystroke dynamics has not stopped and continues to this day. However, the vast majority of research projects are located in the field of computer science, while the legal and criminalistic aspects of the phenomenon in question have not been developed. In recent years, there have been some legal works on this topic, but keystroke dynamics is not disclosed in them with a sufficient degree of completeness (Foygel, 2023; Smushkin, 2022; Fedorov, 2019). In addition, they lack a real interdisciplinarity, which we tried to achieve in this study by combining the efforts and competencies of criminalists and specialists in the field of computer technologies.

We developed technical tool to record keystroke dynamics of computer device user, and on the basis of the data collected we wanted to verify the actual identification significance of the considered personal characteristic, and the effect of the keyboard change factor on the stability of keystroke dynamics features.

It seems that the results of the experiment described in this paper not only can be used as a basis for further developments on this topic, but also will be very useful for a qualitative change in the activities of law enforcement agencies or other stakeholders to determine the typist.

Before proceeding to the substantive part of this study, we consider it necessary to precisely define its subject, revealing the content of the concept of keystroke dynamics. In the present study it will be understood as: 1) in the subjective sense: behavioral biometric characteristic of a person, which combines a set of skills and habits of interaction of a user with a keyboard while creating a text; 2) in the objective sense (also "information about keystroke dynamics"): an external form of expression of skills and habits of interaction of a user with an keyboard while creating a text, which is manifested in the corresponding records both directly on the user's device and by specialized software or hardware devices (if it is present in particular case) in form of digital traces.

2 MATERIALS AND METHODS

The main method in this paper was an experiment in which developed keylogger was used to record keystroke dynamics. This keylogger was run on nine computer devices used in the course of the experiment and recorded all keyboard events, saving them in the human-readable form of tabular files in .csv format. The main analytical work was performed in these spreadsheets run in Microsoft Excel. All calculations were based on basic formulas of mathematical statistics. The texts typed by the participants of the experiment were taken from open sources of the Internet and modified for experimental purposes.

Interdisciplinary approach made it possible to combine principles of law and legal forms with narrow applied technology and computer-technical works devoted to its detailed study.

The experiment described in the main part of this paper involved adult students who expressed interest in the study and were willing to participate in the project. All participants were notified in advance that keystroke dynamics refers to personal data. Respondents gave informed voluntary consent for collection and research of this data for the purposes of the scientific non-commercial experiment. Personal data of the participants in depersonalized form are stored by the Organizers of the experiment.

3 RESULTS

3.1 Features of keystroke dynamics: preliminary theoretical remarks

In order to carry out an experimental study of keystroke dynamics, we studied relevant works on this topic, in which various features of this phenomenon were considered. In total, we analyzed more than 600 scientific papers, the results of which are summarized in Table 1.

Table 1 - Frequency of occurrence of keystroke dynamics features in scientific papers

Feature	Frequency of mentioning (%)
Key hold (dwell) time	100
Release-to-press latencies (flight time)	100
Typing speed	64
Characteristic of overlaps	62
Frequency and content of errors made by the user	19
Use of service keys and special characters	15
Intervals between presses of dual and strobed sequentially arranged groups of characters (n-graph duration and latency)	12
Vibro-sound characteristics of typing	8
Arrhythmicity of typing	6
Typing forces	3
Spatial orientation of presses	2

It should be noted that the table does not reflect those features that occur in less than 1% of the studied works, in particular, such as the model of editing patterns (Villani, Tappert, Ngo, Simone, Fort, Cha, 2006), the fact of using the main or additional keyboard, typical techniques and methods of working on the keyboard (Gunetti, Ruffo, 1999). Such a rare consideration of them, as it seems, may be due to insufficient scientific elaboration, as well as possible low identification significance of these features, which, however, is not a reason to completely ignore them, because together with other characteristics, they are able to form a forensically significant set.

We should also mention that at the early stages of the study of this phenomenon, it was a question of fixing and studying only temporal indicators (typing speed). Later, this characteristic became more detailed: key holding time and release-to-press latencies were recorded. This approach today can be called basic (classical) – as reflected in the table by the 100% occurrence rate – because the vast majority of fundamental and applied studies are limited to the consideration of only these features in pure or slightly modified form (Banerjee, Woodard, 2012, p. 120), and the works that propose a more extensive system of features necessarily contain the above-mentioned ones, because, as it will be shown in the description of the experiment, almost all identifiably significant features can be deduced from those. From the standpoint of forensic science, the study of only two basic characteristics may be insufficient to obtain an identificationally significant set of features, based on which it is possible to make a categorical conclusion. On the other hand, the identification and analysis of the entire set of characteristics proposed in science may create «information noise», which will also complicate identification. We believe that it is necessary to find a balance between the two positions. To do this, in the experiment, the initial stages of which will be described below, we plan to study all possible features of keyboard handwriting, then divide them into two groups: original (those that are directly recorded by the keylogger – the duration of holding the key, intervals between presses, frequency of use of individual keys, etc.) and derivatives (those that are calculated from the original – the nature of errors and typos, the model of making corrections, etc.). Then it is supposed to rank the features by the degree of their identification significance, taking as a basis the characteristics of their individuality, stability and reproducibility. It is assumed that the developed system will help in conducting expert studies of keyboard handwriting: the expert will have an idea of all the features of this phenomenon, but will study only the most situationally conditioned ones in order to answer the questions posed to him.

It should be taken into account that not all features can be obtained using keyloggers. For example, to capture the typing force and the spatial orientation of keystrokes, it is necessary to modernize the keyboard by equipping it with an additional intermediate layer, which will track the depth of pressing (through which the force is derived (Hughes, Aulck, Johnson, 2011)) and the point of application. In the framework of investigative activities, the suspect cannot be relied upon to have such a keyboard in his possession, and therefore these features cannot be taken into account today. A similar situation arises with respect to vibro-sound characteristics of typing – they are obtained by means external to the computer and keyboard (Harrison, Toreini, Mehrnezhad, 2023), moreover, they are not a sign of keystroke dynamics as such, but only a form of recording information about it, used as an alternative to keylogger records for the subsequent calculation of the same temporal and content characteristics of typing.

So we come to the conclusion that it is necessary to conduct an experiment that will show what features of keystroke dynamics can be recorded by means of modern technical means, subjected to a detailed study and effectively used to identify the typist.

3.2 Experiment on studying the features of keystroke dynamics

In this paper, we will focus on the description of the experimental verification of only a small number of initial hypotheses, although their list is quite extensive, which predetermines for us the directions of further research. At these stages of the experiment we will try to verify the following theoretical statements:

1) Features of keystroke dynamics are unique for individual users and are constantly reproduced in the typing process;
2) These features are relatively stable, but changing the type of keyboard on which the text is typed can lead to distortions that exclude identification or significantly reduce its accuracy;
3) Profiles of different users differ significantly from one another even under the same printing conditions;

4) With the help of a software keylogger and mathematical analysis methods, it is possible to establish a sufficient set of keystroke dynamics features for identification;

5) Software keylogger can be configured so that it does not explicitly record information about what a person is typing, thus proving to be safe for work with various kinds of sensitive information.

6) Regardless of the type of computer device, operating system settings, keyboard configuration and other technical characteristics, the keylogger collects accurate information about the pressed key and the characteristics of its pressing.

At the preparatory stage of the experiment, in order to verify the formulated hypotheses, we developed a software keylogger, with a polling frequency of keyboard events 1/100000 sec. This keylogger records the facts of pressing and releasing each key with time characteristics of these events in the format, for example: "2024-02-19_10:39:42.937302 69 released".

These data are generated in an .csv spreadsheet. The first column of the table contains the year, month, date and time of day in the format "Hours":"Minutes":"Seconds"."Milliseconds". This information coincides with the system settings of user's device. The second column specifies the ASCII-code of the key, if it provides display of any character, or the name of the function key (for example, "SHIFT_R", where R – right). After that, if the key is pressed it displays "pressed", if released it displays "released".

We prepared 3 texts differing in complexity, style and variety of elements so that they included various letters of the Russian (native for the participants of the experiment) and Latin alphabets, abbreviations, numbers, additional symbols (quotation marks, dashes and hyphens, colons, various punctuation marks and arithmetic signs). For the first text, an excerpt from a dictation for adults was used, written in a journalistic style and devoted to the analysis of the behavior of characters from a classic literature. This text was written in the style closest to law students. Next, an excerpt from a dissertation for the MD thesis was selected, containing a large amount of specialized terminology and written in a scientific style. It was assumed to be the most difficult for the participants in the experiment to perceive. The last text was an excerpt from a children's fairy tale with repetitive fragments, written in simple language and artistic style. We provided the texts with additional formatting by emphasizing certain places with bold or italics. These texts were offered to the participants of the experiment for typing on their keyboards during the experiment.

It should also be noted that the preparation of texts in terms of determining the optimal volume took into account the experience of other scientific teams involved in the field of keystroke dynamics research. There is no doubt that the larger the volume of texts created, the more accurate it is possible to establish the characteristics of keystroke dynamics of the typist, and the easier it is to apply statistical methods of analysis. At the same time, a significant part of research is aimed at creating multifactor authentication systems, in connection with which they study the manifestations of keystroke dynamics while typing password phrases, and the average volume of data from one user is up to 800-1000 characters (Antal, Nemes, 2016). The volumes obtained for analyzing free texts are not significantly different as well – a little more than 2000 characters from one user (Iapa, Cretu, 2021). For the purposes of our study, it is assumed that the subject of forensic activity will work with texts of different volumes, including quite multi-character texts, so we focused on 700-800 characters (approximately 150 words) in each of the three texts, differing in style and complexity.

It was also stipulated that participants in the experiment not only retyped the texts offered to them, but also created 2 small essays on a keyboard they found more comfortable (mechanical or membrane), in a free style on a given topic with a suggestion to target 200 words in each. We tried to select topics that law students would hardly have to think about for long, so that writing an essay would be as close as possible to situations of free communication on the Internet. As a result, we settled on the following topics: «My profession after university» and «How my week goes». This was important to understand how person types when he or she does not need to be distracted by a sheet of pre-prepared material, but can simply follow the flow of thoughts.

So, at this stage of the experiment, we collected an average of about 24,000 keyboard events from each user and 218,458 events in total.

Given that the experiment was intended to utilize proprietary software and to capture and further process personal data, we prepared a set of legal

ISSN: 2965-1395

documents: Privacy Policy, Non-Disclosure Agreement, Consent to Personal Data Processing and Consent for Processing Personal and Personal Biometric Data – which were read and signed by all participants of the experiment.

After completing all preparatory activities, we invited senior students of the Law University to participate in the experiment. Taking into account the non-compliance of some participants with the recommendations of the experiment organizers, the first stage of the experiment resulted in 9 copies of materials suitable for research from 9 participants, respectively. Each copy included:

3 text documents: three texts proposed by the Organizers, typed on a membrane keyboard; the same three texts, typed on a mechanical keyboard; essays typed on a keyboard that seemed more convenient to the user. It should be noted that most of the participants in the first stage of the experiment were accustomed to working on membrane keyboards on laptops in their daily lives, so they chose them when writing essays.

3 csv files in which keylogger performance was stored; each such file corresponded to one text document.

The experiment involved 2 men and 7 women, of whom 2 participants were 22 years old, 6 were 21 years old, and 1 was 20 years old. All participants were 4th-year full-time students at the Law University.

The main work within the experiment was carried out in the spreadsheet files with the data collected by the keylogger. As it was mentioned earlier, these files recorded information about the pressed (released) key and the time of its pressing (releasing). For further calculations, the time of each event was converted into the number of seconds elapsed since the beginning of the day. As a result, the working document took the following form (see Figure 1).

Figure 1 - Fragment of the table worksheet after recalculating the event time

ISSN: 2965-1395

	A	B	C
1	DT	ST	TZ
2	2024-02-19_11:19:02.599250	SHIFT pressed	40742599,00
3	2024-02-19_11:19:02.739765	78 pressed	40742739,00
4	2024-02-19_11:19:02.786662	SHIFT released	40742786,00
5	2024-02-19_11:19:02.802282	78 released	40742802,00
6	2024-02-19_11:19:02.911274	84 pressed	40742911,00
7	2024-02-19_11:19:03.005332	84 released	40743005,00
8	2024-02-19_11:19:03.989497	82 pressed	40743989,00
9	2024-02-19_11:19:04.051958	82 released	40744051,00
10	2024-02-19_11:19:04.176994	67 pressed	40744176,00
11	2024-02-19_11:19:04.255045	67 released	40744255,00
12	2024-02-19_11:19:04.332864	78 pressed	40744332,00
13	2024-02-19_11:19:04.379934	78 released	40744379,00
14	2024-02-19_11:19:04.473734	SPACE pressed	40744473,00
15	2024-02-19_11:19:04.536257	SPACE released	40744536,00
16	2024-02-19_11:19:04.708076	SHIFT pressed	40744708,00
17	2024-02-19_11:19:05.223570	SHIFT pressed	40745223,00
18	2024-02-19_11:19:05.254794	SHIFT pressed	40745254,00
19	2024-02-19_11:19:05.286041	SHIFT pressed	40745286,00
20	2024-02-19_11:19:05.317335	SHIFT pressed	40745317,00
21	2024-02-19_11:19:05.348522	SHIFT pressed	40745348,00
22	2024-02-19_11:19:05.379807	SHIFT pressed	40745379,00
23	2024-02-19_11:19:05.379807	51 pressed	40745379,00
24	2024-02-19_11:19:05.489108	51 released	40745489,00
25	2024-02-19_11:19:05.489108	SHIFT released	40745489,00
26	2024-02-19_11:19:05.551600	SPACE pressed	40745551,00
27	2024-02-19_11:19:05.645231	SPACE released	40745645,00

Further research was based on the investigation of one integral index – the Press-to-press latencies (PPL). For this purpose, we sorted the collected data by selecting only the press events and calculated the intervals between the time indices of each two consecutive presses. After that, we performed another sorting, excluding noise, for which we took latencies greater than 2000 ms, as they indicate distraction from the typing process and at these stages of the study can significantly distort the data. Such a value was chosen based on the previous experience of other teams (Allen, 2010). The final array with which the work was carried out took the following form (Figure 2).

Figure 2 - Fragment of the table worksheet after calculating the Press-to-press latencies

ISSN: 2965-1395

	A	B	C	D
1	DT	ST	TZ	L
2	2024-02-19_11:19:02.739765	78 pressed	40742739,00	140,00
3	2024-02-19_11:19:02.911274	84 pressed	40742911,00	172,00
4	2024-02-19_11:19:03.989497	82 pressed	40743989,00	1078,00
5	2024-02-19_11:19:04.176994	67 pressed	40744176,00	187,00
6	2024-02-19_11:19:04.332864	78 pressed	40744332,00	156,00
7	2024-02-19_11:19:04.473734	SPACE pressed	40744473,00	141,00
8	2024-02-19_11:19:04.708076	SHIFT pressed	40744708,00	235,00
9	2024-02-19_11:19:05.223570	SHIFT pressed	40745223,00	515,00
10	2024-02-19_11:19:05.254794	SHIFT pressed	40745254,00	31,00
11	2024-02-19_11:19:05.286041	SHIFT pressed	40745286,00	32,00
12	2024-02-19_11:19:05.317335	SHIFT pressed	40745317,00	31,00
13	2024-02-19_11:19:05.348522	SHIFT pressed	40745348,00	31,00
14	2024-02-19_11:19:05.379807	SHIFT pressed	40745379,00	31,00
15	2024-02-19_11:19:05.379807	51 pressed	40745379,00	0,00
16	2024-02-19_11:19:05.551600	SPACE pressed	40745551,00	172,00
17	2024-02-19_11:19:05.723454	49 pressed	40745723,00	172,00
18	2024-02-19_11:19:07.569419	ENTER pressed	40747569,00	1846,00
19	2024-02-19_11:19:20.753499	66 pressed	40760753,00	156,00
20	2024-02-19_11:19:20.909993	67 pressed	40760909,00	156,00
21	2024-02-19_11:19:21.019033	78 pressed	40761019,00	110,00
22	2024-02-19_11:19:21.206657	74 pressed	40761206,00	187,00
23	2024-02-19_11:19:21.378634	72 pressed	40761378,00	172,00
24	2024-02-19_11:19:21.534851	66 pressed	40761534,00	156,00
25	2024-02-19_11:19:21.644196	90 pressed	40761644,00	110,00
26	2024-02-19_11:19:21.753547	SPACE pressed	40761753,00	109,00
27	2024-02-19_11:19:22.081608	71 pressed	40762081,00	328,00

Next, we calculated for each participant in the experiment for each spreadsheet the average and standard deviations of the described PPL. The overall results of these calculations are presented in Table 4.

Table 4 - Indicators of the average and standard deviations of Press-to-press latencies among experimental participants when printing in different situations

Participant No	Type of text	μ_{PPL}	σ_{PPL}
1	3 texts on membrane keyboard	237,75	214,93
	3 texts on mechanical keyboard	374,09	251,53

	Essay on membrane keyboard	367,30	191,67
2	3 texts on membrane keyboard	370,05	267,56
	3 texts on mechanical keyboard	295,45	241,76
	Essay on membrane keyboard	242,26	161,95
3	3 texts on membrane keyboard	269,00	203,64
	3 texts on mechanical keyboard	306,56	197,79
	Essay on mechanical keyboard	234,83	113,82
4	3 texts on membrane keyboard	275,56	191,96
	3 texts on mechanical keyboard	260,57	182,83
	Essay on mechanical keyboard	195,95	122,45
5	3 texts on membrane keyboard	312,06	245,79
	3 texts on mechanical keyboard	313,40	245,58
	Essay on membrane keyboard	240,20	162,24
6	3 texts on membrane keyboard	254,36	180,07
	3 texts on mechanical keyboard	346,38	265,37

	Essay on membrane keyboard	250,98	165,92
	3 texts on membrane keyboard	254,36	180,07
7	3 texts on mechanical keyboard	290,07	218,64
	Essay on membrane keyboard	231,25	139,57
	3 texts on membrane keyboard	328,14	252,02
8	3 texts on mechanical keyboard	351,16	266,33
	Essay on membrane keyboard	231,20	135,72
	3 texts on membrane keyboard	185,66	180,52
9	3 texts on mechanical keyboard	222,93	174,65
	Essay on membrane keyboard	249,52	186,49

The ratios of the average and standard deviations of Press-to-press latencies of different participants are visualized in Figure 3 and Figure 4.

Figure 3 - A visual representation of the average of PPL for each participant

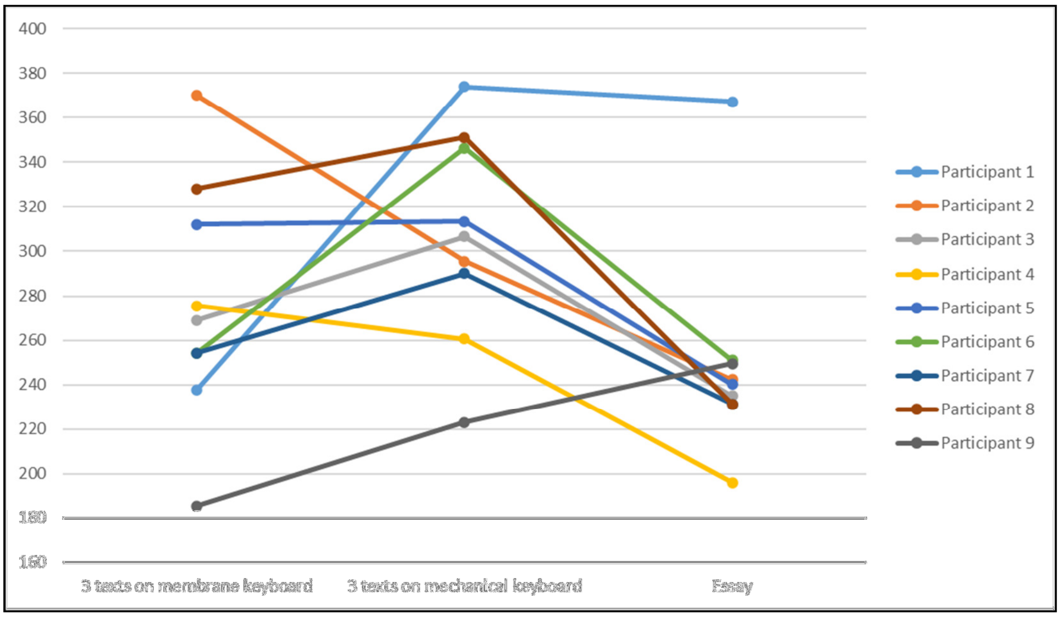
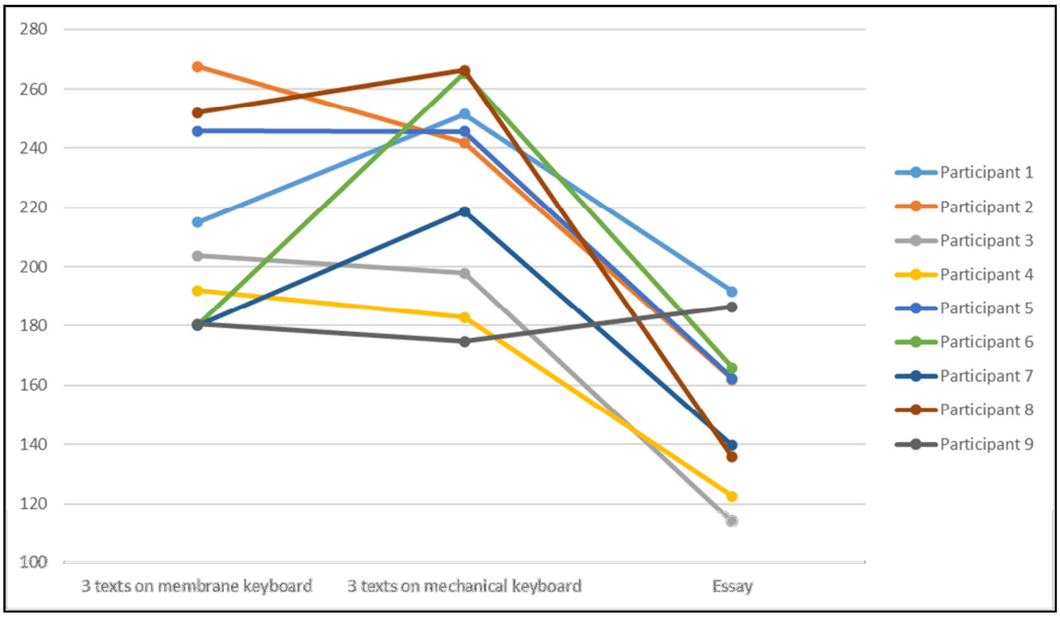


Figure 4 - A visual representation of the standard deviations of PPL for each participant



4 DISCUSSION

Let us consider how the obtained data affect the hypotheses formulated at the beginning of the paper.

1-3. Not all features of keystroke dynamics are individualized: the characteristic of Press-to-press latencies that we studied can be attributed to the general features of the phenomenon, for which it is normal to reproduce in different objects or people (other general features in people, for example, include height, gender, age, and many others). In this regard, it is incorrect to talk about the uniqueness of one feature – this is a very rare phenomenon, it is much more correct to assess the uniqueness of a skill consisting of many features.

The characteristic we have studied reaches very close values in different users typing the same or similar texts (if we talk about essays), while it differs significantly in one user creating different texts on different keyboards.

The type of keyboard, as well as the fact of the presence of confounding factors (in the form of the need to distract from the typing process to remember the next part of the retyped text) significantly affect the stability of the Press-to-press latencies, and much more significantly from the existence of pre-prepared texts to be typed. So, if analyzing a single feature of keystroke dynamics, it is not possible to carry out identification in such cases, because the profile of the performer is significantly distorted.

4. At the same time, from the data collected by the, it is possible to calculate the following features:

- duration of key holding (by grouping the events of pressing and releasing a key having a specific code);
- release-to-press latencies (by grouping consecutive events of releasing one key and pressing the next one);
- characteristic of overlaps (set if the duration of the interval described in the previous point takes a negative value);
- typing speed (by calculating the average number of characters typed per minute for the performer);
- frequency of corrected errors (by recording the events of using the backspace and delete keys);

- the fact of using service keys and special characters (by recording their codes);
- the fact of using additional keyboard blocks (by recording key codes).

Note that these are only the simplest features linearly derived from the collected data. If we use also the analysis of text materials, it is possible to establish typical for the user errors and misprints.

5. The keylogger developed by us records, as it was mentioned earlier, the ASCII codes of keyboard keys or the names of service and function keys. This approach limits the possibility of determining what the performer is typing, because, as scientists correctly note: "keystroke dynamics implies that we do not care about what is being typed, but rather how it is being typed" (Zeid, ElKamar, Hassan, 2022, p. 95).

To compare the result of keylogger's work with the final text, we need to 1) determine the typing language, 2) decipher each character, 3) connect characters between SPACE presses into words, 4) clean the data from "noise" (typos, errors, key combinations, etc.). Undoubtedly, all these actions can be performed and even partially automated, but it will require additional efforts, the use of unnecessary software and time resources, which allows us to conclude, on the one hand, about the relative security of personal information, and on the other hand, about the possibility, if necessary, to compare the information collected by the keylogger with the final typed text, which is critically necessary in the detection and investigation of crimes.

6. The developed keylogger works equally on different devices (laptops, desktop computers), with keyboards of different types, brands and models, however, as it was shown earlier, this is not true in all cases: different readings are collected from the block of function keys, sometimes a computer mouse with programmable additional structural elements is taken as a keyboard key. This is directly related to the principle of keylogger functioning, which is to record events that are read in parallel by the system. Its various settings may not support some functions provided on the keyboard, so if the operating system does not "see" an event, the keylogger cannot record it either. At the same time, such a limitation can be solved by changing the configuration of the keylogger and using, for example, a hardware module connected directly to the keyboard controller, which is technically possible, but was not realized by us at the experiment stages described in the paper, because the level of information losses is not significant.

5 CONCLUSIONS

Summarizing the above, we emphasize that the presented experiment is only the initial stage of research and will be further improved: the team is already working on the calculation of other features of keystroke dynamics, which, as it seems, may contribute to the correction of hypotheses and conclusions obtained so far.

However, we can already conclude unequivocally: keystroke dynamics studies have a significant potential for determining the typist, including for the purposes of crime detection and investigation, but at the same time should be based on a large number of diverse features that can form an identifiably significant set, sufficient for the formation of a stable unique profile of the user of a computer device.

Acknowledgements

This research was funded by Russian Science Foundation, grant number 23-78-10011, https://rscf.ru/project/23-78-10011/.

REFERENCES

ALLEN, Jeffrey D. An Analysis of pressure-based keystroke dynamics algorithms. Master's thesis, Southern Methodist University. Dallas, TX: 2010.

ANTAL, Margit; NEMES, Lehel. The MOBIKEY Keystroke Dynamics Password Database: Benchmark Results. Software Engineering Perspectives and Application in Intelligent Systems; Silhavy, R., Senkerik, R., Oplatkova, Z. K., Silhavy, P., Prokopova Z., Eds. Switzerland: Springer International Publishing, v. 2, p. 35-46, 2016.

BANERJEE, Salil P.; WOODARD, D. L. Biometric Authentication and Identification Using Keystroke Dynamics: A Survey. Journal of Pattern Recognition Research, [s. l.], v. 7, p. 116-139, 2012.

FEDOROV, Igor Z. To the issue of establishing the executor of electronic text by keyboard handwriting in the detection and investigation of crimes. Bulletin of Barnaul Law Institute of the Ministry of Internal Affairs of Russia, [s. l.], v. 2 (37), p. 113-116, 2019.

FOYGEL, Elena I. Modern Trends and Prospects for the Development of Criminalistics Doctrine on the Personality of Participants in Criminal Proceedings. Academic Law Journal, [s. l.], v. 24. n. 1 (91), p. 101-108, 2023.

GUNETTI, Daniele; RUFFO, Giancarlo. Intrusion Detection through Behavioral Data. In Advances in Intelligent Data Analysis. IDA 1999. Lecture Notes in Computer Science; Hand, D. J., Kok, J. N., Berthold, M. R., Eds. Springer: Berlin, Heidelberg, German, 1999. v. 1642, p. 383-394.

HARRISON, Joshua; TOREINI, Ehsan; MEHRNEZHAD, Maryam. A Practical Deep Learning-Based Acoustic Side Channel Attack on Keyboards. Proceedings of the 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Delft, p. 270-280, 3-7 July 2023.

HUGHES, Margaret; AULCK, Lovenoor; JOHNSON, Peter W. Are there Differences in Typing Performance and Typing Forces between Short and Long travel Keyboards? Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Las Vegas, v. 55, n. 1, p. 954-957, 19-23 September 2011.

IAPA, Augustin-Catalin; CRETU, Vladimir-Ioan. Shared Data Set for Free-Text Keystroke Dynamics Authentication Algorithms. Preprints. 2021.

KENTE, Josiah S.; ISHAKU, Jesse. Issues and Perspectives on Cyberterrorism and Information Security in Nigeria. ALSYSTECH Journal of Education Technology, [s. l.], v. 2, p. 174-190, 2024.

RASTORGUEV, Sergey P. Program methods of information protection in computers and networks. Moscow: Yachtsman Agency Publishing House, 1993.

SMUSHKIN, Alexander B. The Digital Transformation of Forensic Science. Part I (Digital Transformation of General Forensic Theory and Forensic Technology). Bulletin of the Saratov State Law Academy, [s. l.], v. 6 (149), p. 250-256, 2022.

SPILLANE, R. Keyboard Apparatus for Personal Identification. IBM Technical Disclosure Bulletin, [s. l.], v. 17, n. 3346, 1975.

VACCA, John R. Biometric technologies and verification systems. Oxford: USA Linacre House, 2007.

VILLANI, Mary et al. Keystroke Biometric Recognition Studies on Long-text Input Under Ideal and Application-oriented Conditions. Proceeding of the International Workshop Scientific Computing and Computational Statistics (IWSCCS 2006), [s. l.], p. C3.1–C3.8, 2006.

WYCIŚLIK, Łukasz; WYLĘŻEK, Przemysław; MOMOT, Alina. The Improved Biometric Identification of Keystroke Dynamics Based on Deep Learning Approaches. Sensors, [s. l.], v. 24 (12), 2024.

ZEID, Shimaa S.; ELKAMAR, Raafat A.; HASSAN, Shimaa I. Fixed-Text vs. Free-Text Keystroke Dynamics for User Authentication. Engineering Research Journal – Faculty of Engineering (Shoubra), [s. l.], v. 51 (1), p. 95-104, 2022