

**ИССЛЕДОВАНИЕ ВЛИЯНИЯ ОБЪЕМА ОБУЧАЮЩЕЙ ВЫБОРКИ НА ТОЧНОСТЬ
АУТЕНТИФИКАЦИИ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ НА МОБИЛЬНЫХ УСТРОЙСТВАХ**

Е.В. Рассказов, И.В. Горбунов, М.О. Калмыков

Научный руководитель: к.т.н. И.В. Горбунов

Томский государственный университет систем управления и радиоэлектроники,

Россия, г. Томск, пр. Ленина, 40, 634050

E-mail: rev7.azure@gmail.com

**RESEARCH OF THE INFLUENCE OF LEARNING SAMPLE SIZE ON THE ACCURACY OF
KEYSTROKE DYNAMICS AUTHENTICATION ON MOBILE DEVICES**

E.V. Rasskazov, I.V. Gorbunov, M.O. Kalmykov

Scientific Supervisor: PhD in Engineering sciences I.V. Gorbunov

Tomsk State University of Control Systems and Radioelectronics,

Russia, Tomsk, Lenin str., 40, 634050

E-mail: rev7.azure@gmail.com

Abstract. Keystroke dynamics is a well-investigated behavioral biometric based on the way and rhythm. In this paper, I present a survey of the stability of keystroke dynamics on mobile devices. It contains an analysis of the statistical features users' keystroke dynamics to show how different they are. I also show how a classification accuracy depends on the size of the learning sample.

Введение. В настоящее время повсеместно используются мобильные устройства. Они являются незаменимым средством и важной частью нашей жизни, поэтому существует необходимость в обеспечении их защищенности, т.к. люди хранят на них не только персональную, но и критически важную информацию [1].

В актуальности проблемы эффективности клавиатурного почерка на мобильных устройствах нет сомнения. На данный момент существуют и постоянно появляются научные статьи, проводятся исследования с использованием различных видов классификаторов. Однако, чтобы добиться максимальной эффективности системы распознавания клавиатурного почерка, (рис. 1) на начальном этапе очень важно представить информацию с клавиатурным почерком в правильном виде, а также определить оптимальный объем обучающей выборки.

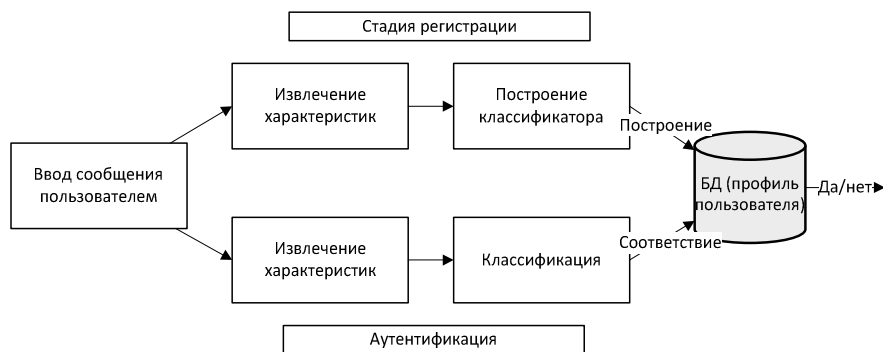


Рис.1. Принцип работы системы распознавания клавиатурного почерка

В данной работе исследуется стабильность клавиатурного почерка пользователей путем анализа статистических показателей клавиатурного почерка, определения различимости профилей, а также определяется влияние объема обучающей выборки на величину ошибок 1-го и 2-го рода.

Оценка эффективности системы распознавания по клавиатурному почерку. После сбора информации для каждого пользователя формируется свой эталон, на каждой последующей аутентификации с которым сравниваются характеристики пользователя посредством алгоритма сравнения.

Для определения эффективности системы используется два важных показателя ошибок [2]:

– FRR (False Reject Rate), или ошибка первого рода – вероятность ошибочных отказов авторизованному пользователю (ошибочный отказ «своему»):

$$FRR = \frac{\text{Количество_ошибочных_отказов}}{\text{Общее_количество_совершенных_попыток}} * 100\% \quad (1)$$

– FAR (False Acceptance Rate), или ошибка второго рода – вероятность допуска злоумышленника (ошибочный пропуск «чужого»):

$$FAR = \frac{\text{Количество_ошибочных_разрешений}}{\text{Общее_количество_совершенных_попыток}} * 100\% \quad (2)$$

Исследование статистических характеристик клавиатурного почерка. В качестве характеристики клавиатурного почерка использовалось время между нажатиями на клавиши. Была сформирована база данных из 11 человек, в которую записывались данные по каждой авторизации. В исследовании рассматривались случаи, когда объем обучающей выборки составлял 5, 10, 15 и 20 авторизаций, т.к. большинство участников авторизовались в системе на момент написания статьи не более 20-23 раз. Все участники вводили одинаковое кодовое слово – “*keystroke*”. В результате каждой авторизации получался вектор вида: $[t_{12}, t_{23}, t_{34}, t_{45}, t_{56}, t_{67}, t_{78}, t_{89}]$, где t_{ij} – время между i -ой и j -ой клавишами.

Для определения того, насколько различимыми являются пользователи между собой, попарно был применен статистический критерий Манна-Уитни [3]. При объеме выборки $n = 5$ количество неразличимых пар пользователей $m = 15$; при $n = 10$: $m = 9$; при $n = 15$: $m = 4$; при $n = 20$: $m = 3$.

Для исследования влияния объема обучающей выборки на величину ошибки 2-го рода, 1-го рода, а также на точность классификации, использовался нечеткий классификатор (нечеткая система типа питтсбургский классификатор) [4].

Результаты вычислений представлены в таблице 1. Значения указаны в средних величинах среди 11 пользователей, участвовавших в эксперименте. График зависимости ошибок 1-го и 2-го рода от объема обучающей выборки представлен на рис. 2.

Из результатов в таблице видно, что по мере увеличения объема обучающей выборки величина ошибки 2-го рода в целом становится меньше. Кроме того, меньше становится ошибка 1-го рода и повышается точность классификации. Лучшие показатели получились при максимальном объеме обучающей выборки ($n = 20$) из рассматриваемых в рамках эксперимента.

Таблица 1

Результаты вычислений

	5 авторизаций	10 авторизаций	15 авторизаций	20 авторизаций
Точность классификации, %	92,05	93,31	93,77	94,36
Ошибка 1-го рода, %	6,43	5,70	5,64	5,23
Ошибка 2-го рода, %	1,52	0,99	0,58	0,41

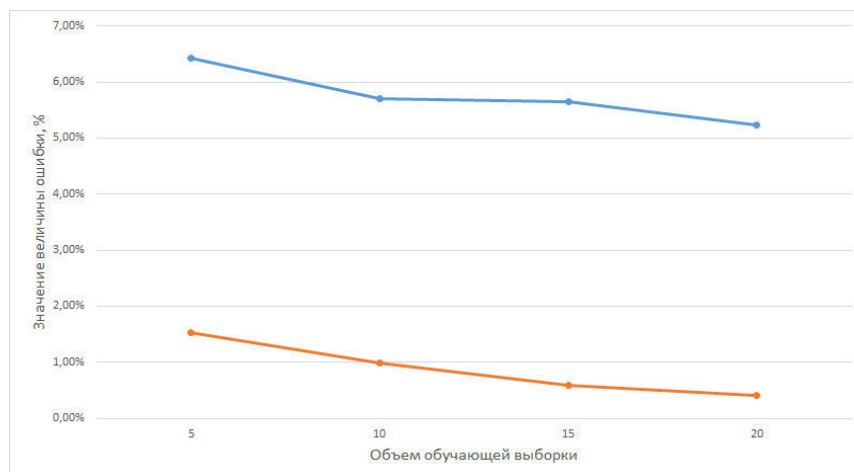


Рис.2. Зависимость ошибок 1-го и 2-го рода от объема обучающей выборки

Выводы. В результате работы при различных объемах обучающей выборки были проанализированы данные по клавиатурному почерку пользователей, проверена различимость пользователей, вычислены значения ошибок 1-го, 2-го рода и точности классификации и сделаны соответствующие выводы о том, что:

1. по мере увеличения объема обучающей выборки пользователи становятся более различимыми;
2. по мере увеличения объема обучающей выборки величина ошибки 2-го рода в целом становится меньше (к тому же уменьшается ошибка 1-го рода и повышается точность классификации);
3. при наибольшем объеме обучающей выборки среди рассматриваемых, равном ($n = 20$) величина ошибки второго рода = 0,41%, что является лучшим результатом.

СПИСОК ЛИТЕРАТУРЫ

1. Antal M, Szabó L, László I. Keystroke dynamics on android platform // 8th International Conference Interdisciplinarity in Engineering, INTER-ENG. - Tirgu Mures, Romania, 2014. – P. 1-2.
2. Campisi P, Maiorana E, Lo Bosco M, Neri A. User authentication using keystroke dynamics for cellular phones. // Signal Processing, IET. – 2009. – Vol. 3(4). – P. 333–341.
3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
4. Горбунов И.В. Алгоритмы и программные средства идентификации Парето-оптимальных нечетких систем на основе метаэвристических методов: диссертация ... кандидата технических наук. Том. гос. университет систем управления и радиоэлектроники. - Томск, 2014. – 192 с.