УДК 004.93’4

В. И. Васильев, М. Ф. Калямов, Л. Ф. Калямова

# ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ С ПРИМЕНЕНИЕМ АЛГОРИТМА РЕГИСТРАЦИИ ЧАСТЫХ БИГРАММ

*Уфимский государственный авиационный технический университет, Уфа, Россия*

*В статье рассматривается разработанный алгоритм для идентификации пользователей по клавиатурному почерку с регистрацией использования наиболее часто встречающихся биграмм. С помощью данного алгоритма можно проводить идентификацию пользователей в постоянном, непрерывном режиме при работе за компьютером. Предлагаемый алгоритм устраняет недостатки существующих методов идентификации пользователя в системе, которые используются только во время входа в систему и тем самым не защищают систему от вторжения после авторизации пользователя. В качестве временных показателей клавиатурного почерка используются следующие характеристики: время нажатия клавиши, пауза между нажатиями клавиш. Временные характеристики собираются по каждой биграмме в отдельности, это необходимо, потому что клавиши расположены на разных расстояниях друг от друга и получается, что временные задержки нажатия одной и той же клавиши будут отличаться в разных биграммах. Чем чаще будут использоваться биграммы в ходе работы пользователя в системе, тем более точными будут временные характеристики и соответственно возрастет эффективность идентификации пользователя. После проведения идентификации, если данные будут отличаться от эталонного, то возможна последующая блокировка выполнения действий пользователем. Для подтверждения эффективности работы алгоритма представлены результаты проверки с применением метода нахождения евклидово расстояния.*

**Ключевые слова**: информационная безопасность, защита информации, идентификация, аутентификация, биометрия, клавиатурный почерк, биграммы.

**Введение.** В современном мире вопросы обеспечения защиты информации являются актуальными. Такие средства аутентификации пользователей в компьютерных системах, как проверка пароля, использование аппаратного идентификатора или доступ по отпечатку пальцев, применяются только во время входа пользователя в систему. Тем самым возникает проблема, когда необходимо следить за достоверностью того, что после авторизации за компьютерной системой продолжает работать законный пользователь, что на его место не сел посторонний человек. Для решения данной проблемы в последние годы многие предлагают использовать клавиатурный мониторинг пользователей вычислительных систем [1]. Так как такой метод является экономичным и не требует использования дополнительных аппаратных средств, например, видео камер, которые используются для идентификации личности по

биометрии лица. Предлагаются различные методы проведения клавиатурного мониторинга, основными из которых являются сбор временных показателей при использовании определенных букв или при наборе ключевых слов [2]. Под временными показателями, которые собираются в ходе клавиатурного мониторинга, обычно понимаются такие значения, как время удержания клавиши, пауза между нажатиями и время последовательных нажатий клавиш [3]. При этом возникают определенные недостатки использования приведенных выше методов клавиатурного мониторинга, поскольку при использовании определенных букв не учитываются положения букв на клавиатуре, ведь каждая определенная буква в сочетании с другой может давать различные временные показатели, т.к. пользователь тратит разное время при переходе с ближайшей и с дальней буквы на нужную. При использовании ключевых слов хоть и устраняется проблема непоследовательности нажатий, но главным недостатком является то, что вероятность использования ключевого слова при работе за компьютером крайне мала [4].

**Существующие подходы.** Алгоритмы распознавания клавиатурного почерка можно разделить на три группы:

- алгоритмы, которые анализируют почерк в ходе авторизации пользователя в системе;
- алгоритмы, которые анализируют почерк после входа в систему при вводе дополнительного текстового фрагмента или фразы;
- алгоритмы, которые проводят непрерывный скрытый мониторинг клавиатурного почерка пользователя.

Алгоритмы первой группы обеспечивают наибольшее быстродействие, пользователю необходимо только ввести свой логин и пароль. Однако точность в этом случае невысока, особенно в случае короткого пароля. Вход может осуществляться оператором, а далее возможна подмена на другого человека. Также выявлено, что логин и пароль иногда могут вводиться одной свободной рукой, из чего следует, что почерк оператора не будет распознан [5].

Алгоритмы второй группы могут обеспечить более высокую точность по сравнению с первой группой. Однако на ввод дополнительного фрагмента текста требуется время, что может вызывать негативные эмоции у пользователя, особенно в случае, если ему часто приходится проходить процедуру аутентификации. Изменение психоэмоционального состояния пользователя может повлиять на его скорость печати, а также на корректность ввода необходимой фразы, что в свою очередь негативно повлияет на точность распознавания [6,7].

Алгоритмы третьей группы позволяют обеспечить более высокую точность, но требуют при этом больше ресурсов. Достоинством этой группы является возможность распознать злоумышленника, который использует компьютер, на котором ранее авторизовался пользователь [8].

**Алгоритм.** Для уменьшения ошибок при распознавании пользователей предложен новый алгоритм сбора статистических данных для клавиатурного мониторинга. Особенностью алгоритма является то, что запись временных параметров осуществляется не во время ввода определенных заданных слов (в большинстве алгоритмов по клавиатурному почерку используется такой метод), а при использовании пользователем определенных биграмм (пар букв).

Чтобы выявить наиболее часто встречающиеся биграммы в русском языке, был проведен частотный анализ самых популярных буквосочетаний. Была разработана программа, с помощью которой можно проверить любой текст на частоту использования биграмм. Демонстрация работы программы показана на Рисунке 1.

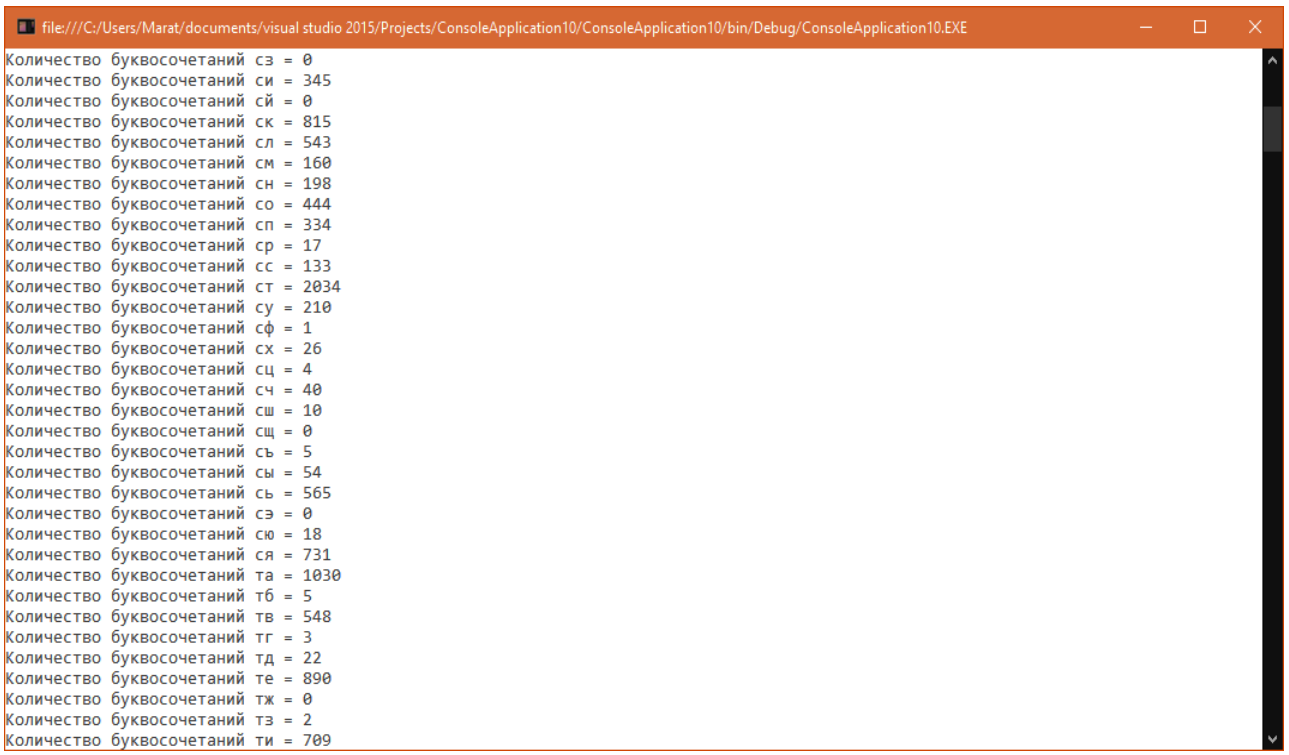


Рисунок 1 – Скриншот программы для вычисления частоты биграмм в тексте

С целью получения более точной информации о частоте биграмм проверялись различные тексты. Было выявлено, что хотя есть небольшая разница в частоте биграмм при анализе научно-технической документации и художественной литературы, но все же в обоих случаях наиболее

частыми являются одни и те же биграммы. На Рисунке 2 приведена диаграмма наиболее часто встречающихся биграмм в русском языке.


Рисунок 2 – Диаграмма частот использования биграмм в русском языке

Для анализа клавиатурного почерка было принято решение использовать только 10 наиболее часто встречающихся биграмм, в числе которых: «ен», «ст», «ра», «ни», «но», «ро», «ов», «пр», «ан», «по». Обоснованием такого выбора служит то, что вероятность встретить другие биграммы в тексте очень мала (около половины из всех биграмм вообще не встречаются), кроме того использование большого количества биграмм уменьшит производительность вычислений.

В целом можно сделать вывод, что для каждой биграммы необходимо записывать 3 параметра:

– время удержания первой клавиши;
– время удержания второй клавиши;
– пауза между нажатиями двух клавиш.

Таблица 1 – Пример выборки по клавиатурному мониторингу

| **Биграммы** | **Время удержания первой клавиши (*t*, мс)** | **Время удержания второй клавиши (*t*, мс)** | **Пауза между нажатиями (*t*, мс)** |
|---|---|---|---|
| ен | 65 | 71 | 185 |
| ст | 79 | 69 | 133 |
| ра | 68 | 67 | 161 |
| ни | 63 | 81 | 140 |
| но | 65 | 74 | 201 |
| ро | 64 | 75 | 171 |
| ов | 68 | 65 | 150 |

| | | | |
|---|---|---|---|
| пр | 61 | 65 | 87 |
| ан | 64 | 70 | 120 |
| по | 61 | 72 | 92 |

**Результаты.** С целью проведения экспериментальной проверки работоспособности данного алгоритма была разработана программа на языке *C#* для проведения клавиатурного мониторинга, которая осуществляет сбор временных показателей [9].

При вычислении времени удержания клавиши на клавиатуре фиксируется системное время при срабатывании события нажатия клавиши *OnKeyDown* и затем вычисляется разница с временем, полученным после события отжатия клавиши *OnKeyUp*.

При вычислении времени между нажатиями последующих клавиш на клавиатуре находится разность во времени срабатывания события нажатия предыдущей клавиши со временем срабатывания события нажатия со следующей клавишей *OnKeyDown* [10].

Эксперимент проводился с участием 4-х человек. При этом для точности эксперимента все участники работали за одним рабочим местом.

Изначально были собраны биометрические данные по клавиатурному почерку по всем четырем пользователям. Исходная выборка содержит в себе 30 строк и 11 столбцов на каждого пользователя. Первые 10 столбцов являются эталонной выборкой, а оставшийся столбец используется в качестве тестовых данных.

На Рисунке 3 изображена диаграмма, которая построена на основе собранных данных в ходе клавиатурного мониторинга. На диаграмме по горизонтали расположены анализируемые параметры: первые 10 значений– это временные значения по удержанию первой клавиши, вторые 10 значений – это временные значения по удержанию второй клавиши, последние 10 значений – это пауза между нажатиями первой и второй клавиши. По вертикали указаны соответствующие временные значения в миллисекундах.

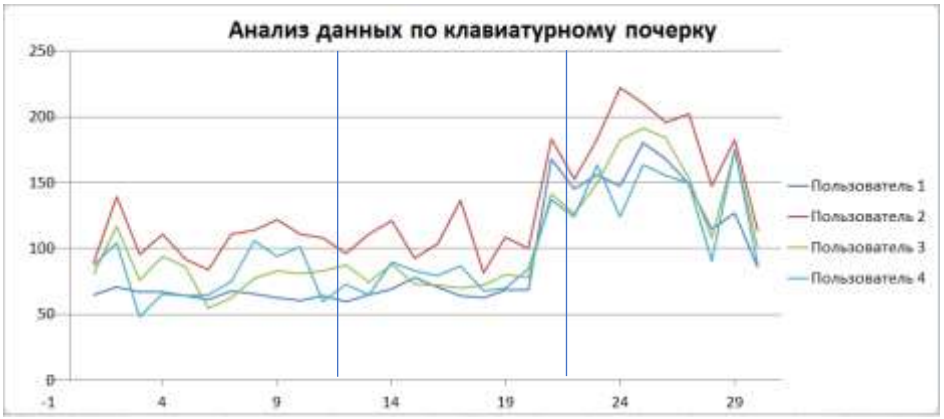


Рисунок 3 – Диаграмма анализа данных по клавиатурному почерку

Затем собранные данные проверялись в среде математического моделирования *MATLAB*, где использовалась функция *pdist*(), которая вычисляет евклидово расстояние для выборки числовых параметров. На Рисунке 4 приведена иллюстрация использования метода нахождения евклидово расстояния для заранее собранных данных.


Рисунок 4 – Иллюстрация проверки методом нахождения евклидово расстояния

Принцип проверки методом нахождения евклидово расстояния заключается в поиске минимального расстояния от тестового примера к обучающим примерам каждого пользователя.

В качестве тестового примера использовался пример пользователя 1, который не содержится в обучающей выборке (Рисунок 5).

| Пользователи | Расстояния от тестового примера до всех обучающих примеров каждого пользователя | | | | | | | | | | Среднее расстояние |
|---|---|---|---|---|---|---|---|---|---|---|---|
| L1 | 220.35 | 299.90 | 223.06 | 362.52 | 310.27 | 541.01 | 219.64 | 419.87 | 194.50 | 335.32 | 312.64 |
| L2 | 1006.50 | 715.00 | 423.70 | 421.60 | 370.20 | 403.20 | 368.40 | 433.90 | 345.00 | 468.70 | 495.62 |
| L3 | 1030.00 | 274.20 | 499.60 | 353.50 | 522.30 | 365.10 | 296.20 | 268.00 | 404.30 | 936.20 | 494.94 |
| L4 | 355.20 | 283.00 | 9464.00 | 392.50 | 347.30 | 311.10 | 350.50 | 233.90 | 242.10 | 281.80 | 1226.14 |

Рисунок 5 – Результаты проверки методом нахождения евклидово расстояния

По итогам проверки получены средние значения расстояний от тестового примера ко всем обучающим примерам каждого пользователя. Минимальное значение равно 312,64 и оно принадлежит пользователю *L*1 с номером один. Проверка прошла успешно.

**Заключение.** Экспериментальная проверка доказывает эффективность работы предложенного алгоритма клавиатурного мониторинга. Преимуществом данного алгоритма является непрерывность проведения клавиатурного мониторинга, его независимость от

определенных ключевых слов, а также возможность использования при работе с иностранным текстом. При этом нужно будет предварительно провести анализ частоты биграмм того языка, который будет использован. Имеется возможность дополнить данный алгоритм для фиксирования нажатий специальных клавиш, например, клавиши «*Ctrl*», которая часто используется в сочетании с клавишами «*V*» и «*C*» для копирования данных в буфер и вставки из буфера.

## ЛИТЕРАТУРА

1. Биометрические технологии [Электронный ресурс] / М.: ID Expert. – Режим доступа: http://www.idexpert.ru/technology/119/, свободный
2. Бочкарев, С.Л. Унификация биометрических технологий: интерфейс BioAPI [Текст] / С.Л. Бочкарев. – М. :Конфидент, 2002. – 174 с.
3. Брюхомицкий, Ю.А. Иммунологический подход к организации клавиатурного мониторинга [Текст] / Известия ЮФУ. Технические науки. Тематический выпуск «Информационная безопасность». – Таганрог: Изд-во ТТИ ЮФУ, 2014. – №2 (151). – С.33-41.
4. Брюхомицкий, Ю.А. Система скрытного клавиатурного мониторинга [Текст] / Ю.А. Брюхомицкий, М.Н. Казарин / Известия ТРТУ. – 2006. – № 9 (64). –С. 153-154.
5. Брюхомицкий, Ю.А. Цепочный метод клавиатурного мониторинга [Текст] / Известия ЮФУ. Технические науки. Тематический выпуск «Информационная безопасность». – Таганрог: Изд-во ТТИ ЮФУ, 2009. – №11. – С.135-145.
6. Васильев В.И. Распознавание психофизиологических состояний пользователей на основе скрытого мониторинга действий в компьютерных системах [Текст] / В.И. Васильев, А.Е. Сулавко, Р.В. Борисов, С.С. Жумажанова / Искусственный интеллект и принятие решений, 2017. – № 3. – С.21-37.
7. Концепция и модели менеджмента безопасности информационных и телекоммуникационных технологий [Электронный ресурс] / ГОСТ Р ИСО/МЭК 13335-1-2006 Методы и средства обеспечения безопасности – Режим доступа: http://vsegost.com/Catalog/27/271.shtml, свободный.
8. Макаревич, О.Б. Актуальные аспекты информационной безопасности [Текст] / О.Б. Макаревич.(ред.) – Таганрог: Изд-во ТТИ ЮФУ, 2011. – 448 с.
9. Пилецкий, И.И. Методы и технологии программирования [Текст] / И.И. Пилецкий. – Минск: БГУИР, 2007. – 238 с.
10. Троелсен, Э. Язык программирования C# 2010 и платформа .NET4. 5-е издание [Текст] / Э. Троелсен. – М. : Вильямс, 2011. – 1392 с.

V.I. Vasilyev, M.F. Kaliamov, L.F. Kaliamova

# IDENTIFICATION OF USERS BY KEYBOARD HANDWRITING USING THE ALGORITHM OF FREQUENT BIGRAMS REGISTRATION

*Ufa State Aviation Technical University, Ufa, Russia*

*In this article the developed algorithm for identification of users by keyboard handwriting with registration of frequent bigrams is considered. By means of this algorithm it is possible to carry out identification of users in a constant, continuous operation in computer system. The offered algorithm eliminates defects of existing user identification methods which are used only during login and by that do not protect system from invasion after authorization of the user. As temporal indices of keyboard handwriting the following characteristics are used: key press time, pauses between clicking of keys. Time response characteristics are gathered according to each bigram separately, it is necessary because keys are located at different distances from each other and it turns out that temporal time delays of clicking of the same key will differ in different bigrams. That will use more often bigrams during operation of the user in system, especially time response characteristics will be exact and respectively the efficiency of user identification will increase. After carrying out identification if data differ from reference, then the subsequent lock of execution of actions by the user is possible. For confirmation of overall performance of an algorithm results of check using a finding method Euclidean distances are provided.*

**Keywords**: information security, information protection, identification, authentication, biometry, keyboard handwriting, bigrams.

## REFERENCES

1. Biometricheskie tekhnologii [Digital resource] / M.: ID Expert. – Access mode: http://www.idexpert.ru/technology/119/, free access
2. Bochkarev, S.L. Unifikatsiya biometricheskikh tekhnologiy: interfeis BioAPI [Text] / S.L. Bochkarev. – M.: Konfident, 2002. – 174 p.
3. Bryokhomitskiy, Y.A. Immunologicheskiy podhod k organizatsii klaviaturnogo monitoringa [Text] / Izvestiya YFU. Tekhnicheskie nauki. Tematicheskiy vypisk «Informatsionnaya bezopasnost». – Taganrog: Izd-vo TTI YFU, 2014. – № 2 (151). pp.33-41.
4. Bryokhomitskiy, Y.A. Sistema skrytogo klaviaturnogo monitoringa [Text] / Y.A. Bryokhomitskiy, M.N. Kazarin / Izvestiya TRTU. – 2006. pp.153-154.
5. Bryokhomitskiy, Y.A. Tsepochniy metod klaviaturnogo monitoringa [Text] / «Izvestiya YFU. Tekhnicheskie nauki». Tematicheskiy vypusk «Informatsionnaya bezopasnost». – Taganrog: Izd-vo TTI YFU, 2009. – № 11. pp.135-145.
6. Vasilyev V.I. Raspoznavanie psikhofiziologicheskikh sostoyaniy polzovateley na osnove skrytogo monitoringa deistviy v kompyoternikh sistemakh [Text] / V.I. Vasilyev, A.E. Sulavko, R.V. Borisov, S.S. Zhumazhanova / «Iskusstvenniy intellekt i prinyatie resheniy», 2017. – № 3. pp. 21-37.

7. Kontsepsiya I modeli menedzhmenta bezopasnosti informatsionnikh I telekommunikatsionnikh tekhnologiy [Digital resourse] / GOST R ISO/MEK 13335-1-2006 Metody I sredstva obespecheniya bezopasnosti. – Acess mode: http://vsegost.com/Catalog/27/271.shtml, free access.
8. Makarevich, O.B. (Ed.) «Aktualnie aspekty informatsionnoy bezopasnosti» [Text] / O.B. Makarevich. – Taganrog: Izd-vo TTI YFU, 2011. – 448 p.
9. Piletskiy, I.I. Metody I tekhnologii programmirovaniya [Text] / I.I. Piletskiy. – Minsk: BGUIR, 2007. – 238 p.
10. Troelsen, E. Yazik programmirovaniya C# 2010 I platforma .NET4. 5-e izdanie [Text] / E. Troelsen. – M.: Vilyams, 2011. – 1392 p.