Интеллектуальное управление доступом на основе скрытой идентификации пользователей по биометрическим признакам

М.Ф. Калямов
Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: yadeforz@gmail.com

В.И. Васильев
Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: vasilyev@ugatu.ac.ru

Аннотация[1]

Работа посвящена разработке интеллектуальной системы управления доступом на основе скрытой идентификации пользователей по биометрическим признакам. Предлагается 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. По результатам идентификации предоставляется определенный уровень доступа к рабочей системе или среде. 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.

1. 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, а не злоумышленник, который намерен осуществить несанкционированный доступ к компьютерной системе и овладеть находящимися в нем информационными данными. Для решения данной проблемы в последние годы многие предлагают использовать биометрическую идентификацию пользователей по клавиатурному почерку и по особенностям работы с мышью при работе на вычислительных системах. Так как такой метод является экономичным и не требует использования дополнительных аппаратных средств, например, видео камер, которые используются для идентификации личности по биометрии лица [1-3].

Предлагаются различные методы проведения клавиатурного мониторинга, основными из которых являются сбор временных показателей при использовании определенных букв или при наборе ключевых слов. 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. При этом возникают определенные недостатки использования приведенных выше методов клавиатурного мониторинга, поскольку при использовании определенных букв не учитываются положения букв на клавиатуре, использование каждой определенной клавиши в сочетании с другой могут выдавать различные временные показатели, т.к. пользователь тратит разное время при переходе с ближайшей и с дальней клавиши на нужную. Использование ключевых слов устраняет проблему непоследовательности нажатий, но главным недостатком является то, что вероятность использования ключевого слова при работе за компьютером крайне мала [4].

Основной биометрической характеристикой при работе с мышью, которая позволяет правильно идентифицировать пользователя, является траектория движения мыши. Установлено, что пользователи перемещают курсор мыши к объектам на экране по разной траектории, тем самым данная биометрическая характеристика выделяет почерк каждого пользователя [5].

Труды Шестой всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28-31 мая, Уфа-Ставрополь, Россия, 2018

Используя совместно данные по клавиатурному почерку и особенностям работы с мышью, можно добиться существенного повышения точности распознавания пользователей.

Анализ биометрических показателей позволяет выявить изменение психоэмоционального и психофизического состояния законного пользователя. В зависимости от настроения или усталости почерк пользователя может немного изменятся, временные характеристики клавиатурного почерка при этом могут повышаться, также повышается количество ошибок совершаемых пользователем. Для малых компаний польза от данной особенности может быть несущественной, но в стратегически важных предприятиях, в крупных финансовых организациях, и в государственных учреждениях, где очень высока ценность информации или, если работа осуществляется с важными и опасными объектами, где ошибки пользователей могут повлечь серьезные последствия, своевременное выявление изменений состояния пользователей могут помочь устранить проблему до его появления [6].

2. Обоснование проведения скрытой идентификации

При проведении идентификации по биометрическим признакам от пользователей в большинстве случаев требуется предоставление/ввод уникальных индивидуальных характеристик, например, когда происходит идентификация пользователей по отпечатку пальца, с использованием специального сканера, пользователь должен приложить палец, по рисунку которого, ранее предоставлен доступ. Следует отметить, что не всегда данная процедура проходит успешно, и возможно пользователю придется повторить свои действия. Такая же картина наблюдается и при распознавании по сетчатке глаз, на сей раз эта процедура даже более докучательна для пользователей, т.к. устройство статично прикрепляется к стене и его высота не соответствуют уровню глаз каждого человека [7].

Другой же особенностью таких систем биометрической идентификации является их открытость и доступность для посторонних пользователей, соответственно, если кто-то будет иметь злой умысел по получению негласного доступа к системе, обойти систему защиты не составит труда.

Для устранения данных недостатков применяются системы скрытой идентификации пользователей по биометрическим признакам, которые в свою очередь не требуют от пользователей выполнения конкретных действий для получения доступа к системе, наоборот, такие системы проектируют таким образом, чтобы система сама получала необходимую информацию для идентификации [8].

3. Алгоритмы идентификации пользователей по клавиатурному почерку

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.

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 [9].

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.

4. Разработка алгоритма клавиатурного мониторинга по биграммам

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, а при использовании пользователем определенных биграмм (пара 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 (рис. 1).

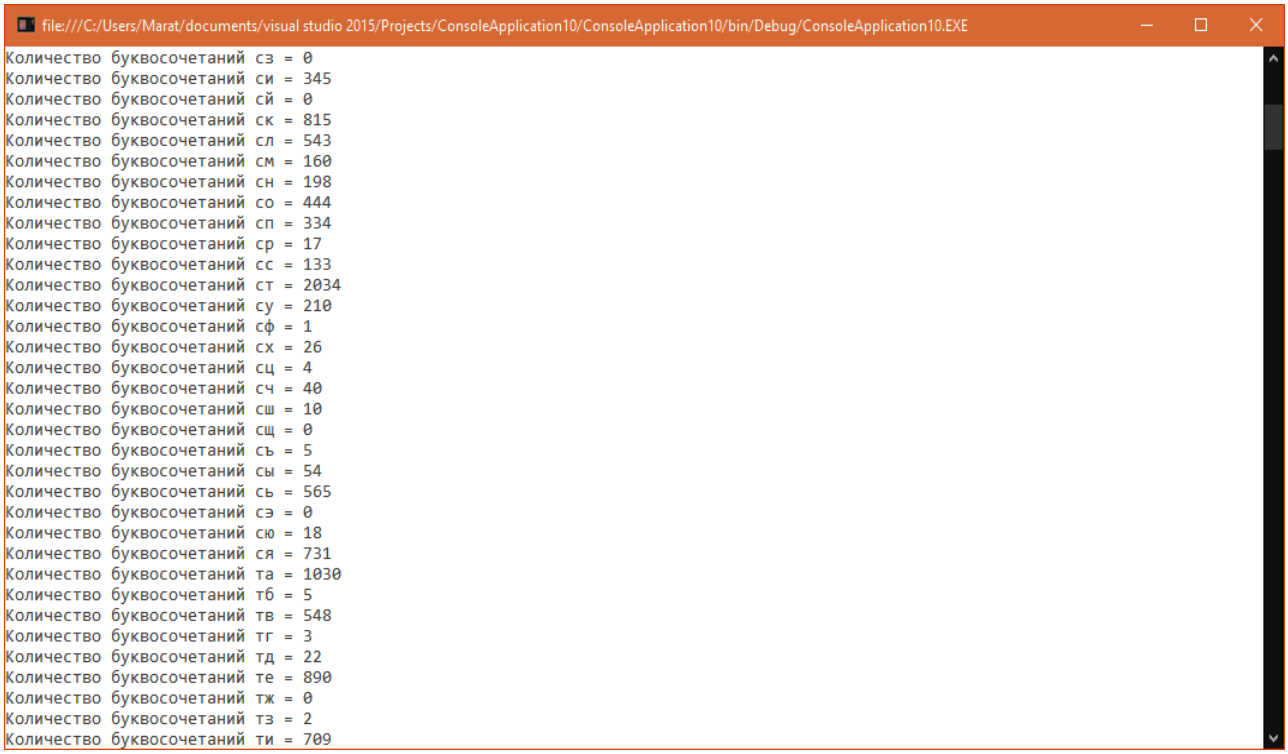

Рис. 1. Скриншот программы для вычисления частоты биграмм в тексте

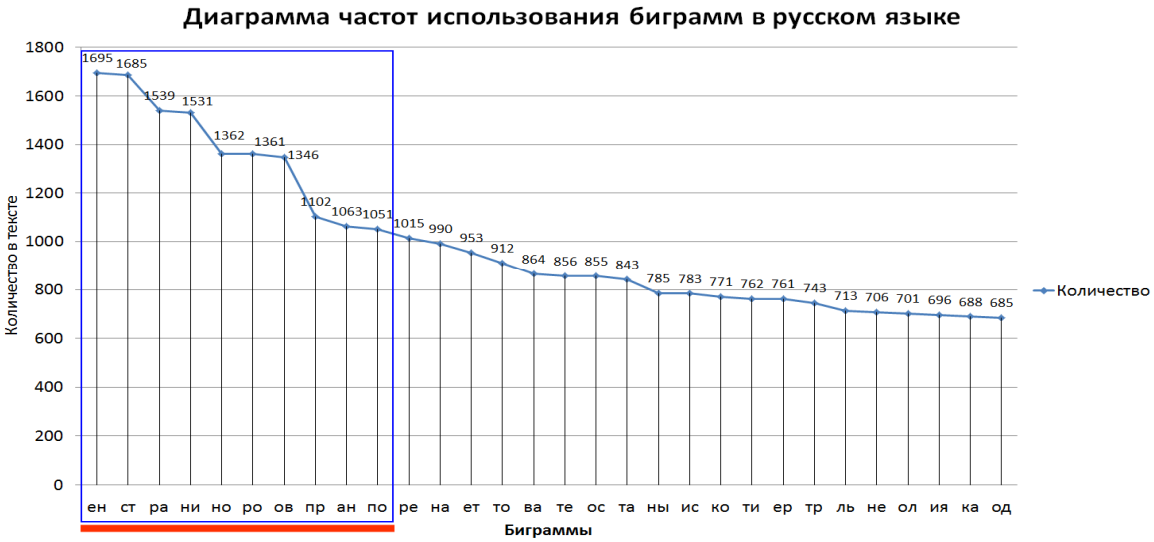

Рис. 2. Диаграмма частот использования биграмм в русском языке

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.

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

5. Математическое описание процесса предоставления уровней доступа

Математическое описание является отражением физической сущности процесса со свойственными ему особенностями и ограничениями. Эти особенности и ограничения должны учитываться как при формулировании задачи, так и при составлении описания и выборе численного метода.

В рамках математического описания модели управления доступом приводятся все множества и функции, которые в нем используются.

$U = \{1..m\}$ – данное множество является обозначением существующих аккаунтов / пользователей зарегистрированных в системе.

$S = \{1..n\}$ – это множество компьютерных систем, с которыми осуществляется работа и к которым предоставляется доступ пользователям.

На этапе идентификации пользователя определяется точность распознавания от 1 до 100, но это является избыточной информацией, необходимо значение точности разделить на категории.

$K = \{0..2\}$ – категории распознавания, при этом значение «0» означает то, что точность распознавания была меньше 70 и это значит то, что законный оператор не может быть определен, значение «1» означает то, что точность распознавания находится в интервале от 70 до 90, в этом случае пользователь считается распознанным, но с отклонениями – это может быть в том случае, если пользователь находится в ненормальном физическом или эмоциональном состоянии и значение «2» говорит о том, что пользователь распознан с точностью распознавания больше 90.

$R = \{0..5\}$ – возможные права доступа, которые могут быть предоставлены пользователям при работе на компьютерных системах. Значения этого множества понимаются следующим образом:

- «0» – доступ запрещен;
- «1» – разрешено чтение при работе со своего аккаунта;
- «2» – разрешено чтение/запись при работе со своего аккаунта;
- «3» – разрешено чтение при работе со своего аккаунта и чтение с использованием чужого аккаунта;
- «4» – разрешено чтение/запись при работе со своего аккаунта и чтение с использованием чужого аккаунта;
- «5» – разрешено чтение/запись при работе со своего аккаунта и чтение/запись с использованием чужого аккаунта;

На основе перечисленных множеств формируется матрица доступа (2.3):

$$accessMatrix[U,S]=\begin{pmatrix} a_{11} & a_{12} & .. & a_{1n} \\ a_{21} & a_{22} & .. & a_{2n} \\ : & : & a_{ij} & : \\ a_{m1} & a_{m2} & .. & a_{mn} \end{pmatrix}, \quad (2.3)$$

где $i = 1..m$, $j = 1..n$, $a_{ij} \in R$.

Далее приводится описание функции, в котором осуществляется предоставление доступа в зависимости от входных параметров.

f(*in*, *out*), где *in* = {*compSystem*, *currentAccount*, *recognizedUser*, *recognizeCategory*) – множество входных параметров, при этом *compSystem* – это компьютерная система на котором необходимо осуществить предоставление доступа, *currentAccount* – текущий аккаунт с которого происходит работа в компьютерной системе, *recognizedUser* – распознанный пользователь на этапе идентификации, *recognizeCategory* – категория распознавания, зависимая от точности распознавания в ходе идентификации пользователя, out – выходной параметр в котором указывается, какой уровень доступа предоставляется пользователю.

Внутренняя реализация данной функции приводится при построении программной имитационной модели, а также при построении модели средствами Simulink. По входным параметрам определяется, работает ли распознанный пользователь под своим аккаунтом или под чужим, исходя из этого в зависимости от установленного значения в матрице доступа будет приниматься решение – какой уровень доступа необходимо предоставить пользователю, при этом, конечно же, учитывается и категория распознавания.

6. Эксперимент

С целью проведения экспериментальной проверки работоспособности данного алгоритма была разработана программа на языке *C#* для проведения клавиатурного мониторинга, которая осуществляет сбор временных показателей [10].

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*.

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.

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.

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

На рисунке 5 представлены результаты проверки методом нахождения евклидового расстояния.

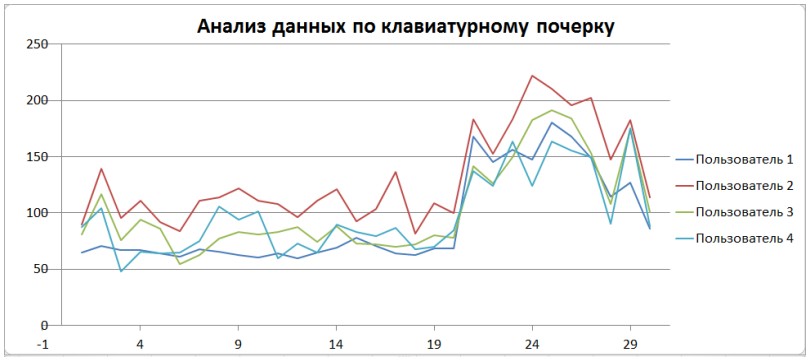

Рис. 3. Диаграмма анализа данных по клавиатурному почерку

Рис. 4. Иллюстрация проверки методом нахождения евклидово расстояния

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

Рис. 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.

7. 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.

Список используемых источников

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 психо-физиологических 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.