УДК 004.8 DOI: 10.18698/0236-3933-2020-1-124-139

МЕТОДЫ РЕШЕНИЯ ЗАДАЧ АУТЕНТИФИКАЦИИ И ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ НА ОСНОВЕ АНАЛИЗА КЛАВИАТУРНОГО ПОЧЕРКА

Ю.В. Ямченко yamchenko.y.v@yandex.com

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

В настоящее время информационные технологии применяют во всех сферах нашей жизни. Большие объемы личной и корпоративной информации хранятся в оцифрованном виде. В связи с этим возникает проблема защиты этой информации от несанкционированного доступа. Важную роль играют подсистемы управления доступом и технологии, используемые в этих подсистемах, в частности методы аутентификации и идентификации пользователей. Приведен обзор методов решения задач аутентификации и идентификации пользователей информационных систем на основе анализа их клавиатурного почерка. Рассмотрены особенности этих задач, описана краткая история возникновения методов аутентификации и идентификации пользователей по клавиатурному почерку, приведены их достоинства и недостатки по сравнению с другими методами, обоснована актуальность такого направления исследований. Перечислены основные стадии создания математических моделей для решения задач аутентификации и идентификации пользователей. Приведено описание методов, используемых на каждой стадии аутентификации и идентификации пользователей

Ключевые слова

Клавиатурный почерк, пользователь, аутентификация, идентификация, машинное обучение, методы классификации

Поступила 26.04.2019
Принята 29.10.2019

Введение. Важнейшее место в современном мире цифровых технологий и всеобщей компьютеризации занимают вопросы защиты конфиденциальной информации пользователей информационных систем. Угрозы информационной безопасности могут возникать как вследствие халатности этих пользователей, так и использования некачественного программного обеспечения. Нарушение политики безопасности может привести к утрате конфиденциальности, нарушению целостности и доступности защищаемой информации, что может повлечь за собой значительные мо-

ральные и материальные затраты как отдельного человека, так и целой организации.

Одним из главных факторов, которые определяют защищенность любой информационной системы от несанкционированного доступа, является эффективность работы подсистемы управления доступом в эту систему и защиты информации в процессе работы с ней. Подобная защита обеспечивается с использованием процедур аутентификации и идентификации (АИ) пользователей.

В зависимости от типа информации, на основе которой осуществляется АИ пользователей, методы АИ относят к одной из трех групп: 1) парольные; 2) атрибутные; 3) биометрические. Парольные методы в качестве информации для АИ используют парольную фразу или ПИН-код; атрибутные методы — уникальные предметы (ключ или смарт-карты); биометрические методы — уникальную информацию, характеризующую конкретного пользователя информационной системы. Примерами такой информации могут служить отпечаток пальца, снимок сетчатки глаза, голосовые характеристики, особенности почерка и геометрии руки [1].

В современных системах в основном применяют *парольные методы* аутентификации. Однако эти методы обладают существенными недостатками. Один из таких недостатков — полная незащищенность информации в случае нарушения конфиденциальности пароля. Другой недостаток — отсутствие средств защиты от подмены пользователя после его авторизации в системе. Если прошедший авторизацию пользователь оставит систему без наблюдения, то любой злоумышленник, находящийся в непосредственной близости к этой системе, может получить к ней доступ.

Атрибутные методы также обладают несколькими недостатками, в частности, возможность обмана системы защиты посредством кражи или имитации атрибута доступа, а также относительно высокие стоимость атрибутов доступа и средств для взаимодействия с ними.

Биометрические методы аутентификации ввиду неотъемлемости биометрических характеристик конкретного человека могут обеспечить повышенную точность и более высокую надежность АИ пользователей. Однако большая часть биометрических методов требует специализированного дорогостоящего оборудования, что ограничивает возможность их повсеместного использования.

Одним из наименее затратных методов с позиции финансов и используемой аппаратуры является метод АИ по клавиатурному почерку (КП). Такой метод относится к числу поведенческих биометрических методов и основан на выделении особенностей, характерных для подсозна-

тельных движений пользователя, осуществляемых им в процессе работы с устройствами ввода информации. Метод АИ по КП обладает следующими преимуществами:

– относительная простота алгоритмов;

– небольшие затраты на оборудование — в процессе АИ используется только клавиатура и мышь;

– возможность использования незаметно для пользователя;

– относительно высокая эффективность [2].

Отметим, что решение задачи аутентификации предполагает решение задачи классификации пользователей, поэтому далее, когда речь будет идти об аутентификации пользователей, будет также иметься в виду решение задачи их классификации.

Проблема АИ людей по данным их КП изучена во многих работах, например [3–9] и др.

Первые идеи по исследованию «почерка» пользователя при работе с устройством ввода информации возникли в задачах идентификации операторов телеграфов при отправке закодированных сообщений с использованием азбуки Морзе. Идентификация операторов происходила по характеристикам ритма и скорости набора [8].

В 1980-х годах была выдвинута гипотеза о том, что характеристики КП можно использовать в качестве средства аутентификации пользователей. В это же время начались эксперименты по определению наиболее значимых характеристик КП для решения задачи аутентификации при наборе фиксированных слов (паролей).

В 1995 г. Шеперд С. (S. Shepherd) впервые проявил интерес к проблеме АИ пользователей на текстах произвольной длины [3]. В 1997 г. Ф. Монрос (F. Monrose) и А. Рубин (A. Rubin) предприняли попытку аутентификации пользователей на таких текстах [4]. При этом итоговая точность классификации пользователей составила лишь 23 %, в то время как точность классификации на текстах фиксированной длины составляла около 90 %. В дальнейших исследованиях прослеживается разделение методов решения задачи АИ на методы АИ для текстов фиксированной и произвольной длины.

Методы аутентификации по КП на текстах фиксированной длины используют исключительно в целях аутентификации пользователя в информационной системе по заранее известной парольной фразе. В таких методах обучение классификаторов происходит посредством аккумулирования и постобработки различных временных и частотных характеристик при многократном наборе пользователем парольной фразы. Таким

образом, в результате обучения классификатора для каждого пользователя составляют вектор характерных признаков (ВХП), определяющий стиль набора пользователем этой фразы.

Методы аутентификации пользователей на текстах произвольной длины используют информацию о работе пользователя с устройствами ввода, получаемую на протяжении всей рабочей сессии. Обучение классификатора при использовании таких методов осуществляют за счет сбора данных о работе пользователей на протяжении некоторого промежутка времени, например нескольких дней, недель или бóльших временных промежутков. На основании собранной информации для каждого пользователя формируют ВХП, отражающий стиль его работы с устройствами ввода информации.

Принято выделять следующие стадии обучения классификаторов КП:

1) создание выборки;

2) сбор «сырых» (необработанных) данных, полученных в результате взаимодействия пользователя с информационной системой;

3) выделение ВХП из «сырых» данных;

4) разбиение полученных данных на тестовую и тренировочную выборки;

5) построение классификатора на тренировочной выборке;

6) валидация классификатора на тестовой выборке.

Далее последовательно рассмотрим каждую указанную стадию.

Создание выборки. Больший размер выборки, используемой при сборе данных, создании и валидации модели классификатора, обеспечивает большую статистическую значимость получаемых результатов и более высокую точность классификации [11, 12]. Однако подавляющее число исследований в этой области ограничивается небольшим размером выборки [13].

Важной характеристикой используемой выборки является ее репрезентативность. В большинстве работ по КП исследуют почерк людей, имеющих опыт работы с устройствами ввода информации, и практически нет исследований с участием людей, не имеющих такого опыта. Однако, учитывая современное развитие технологий, людей, не имеющих такого опыта, становится все меньше. Кроме того, подавляющее большинство исследователей предполагает использование разрабатываемых ими методов на предприятиях, где сотрудники находятся в постоянном взаимодействии с устройствами ввода информации. Поэтому факт превалирования в выборках людей, имеющих опыт работы с этими устройствами, соответствует целям исследований. Создание репрезентативной

выборки большого размера — сложная задача, поэтому хорошей практикой является использование данных других исследователей [14, 15].

Сбор «сырых» данных, полученных в результате взаимодействия пользователя с информационной системой. Перед передачей данных классификатору их необходимо предварительно обработать. Для удобства обработки данные накапливают в определенном заранее установленном формате, который должен поддерживать возможность хранения всей необходимой информации для последующего выделения ВХП.

Для сбора «сырых» данных используют специальные программы или устройства — кейлоггеры (*Keylogger*), регистрирующие различные действия пользователя, такие как нажатия клавиш на клавиатуре, движения и нажатия клавиш мыши и др.

Нажатия клавиш на устройствах ввода информации генерируют прерывания, которые перехватываются и обрабатываются кейлоггером. При этом важной характеристикой является разрешение хронометра, используемого для фиксации прерываний. В первых работах, связанных с исследованием КП, использованы таймеры с разрешением 10 мс [16]. Очевидно, что точность данных, полученных с таймеров такого низкого разрешения, не может быть высокой. Современные компьютеры оснащены таймерами, разрешение которых может составлять несколько наносекунд. Вследствие этого возрастает точность данных и качество создаваемых на их основе моделей классификаторов.

Выделение векторов характерных признаков из «сырых» данных. В большинстве работ при анализе КП пользователей используют временные и частотные признаки нажатий клавиш на клавиатуре как для отдельной клавиши, так и для комбинаций клавиш. Последовательность, содержащую $n \geq 2$ клавиш, принято называть n-граммой.

Временные признаки часто описывают с использованием терминов *Down* и *Up*, обозначающих нажатие и отпускание клавиши клавиатуры или кнопки мыши соответственно. Например, процесс нажатия и отпускания определенной клавиши *key* обозначают как $Down_{key} - Up_{key}$ или сокращенно $D_{key}U_{key}$. В качестве *key* может использоваться название либо номер клавиши.

Опишем несколько наиболее часто используемых временных признаков.

Время нажатия (*Dwell Time*, *DT*) рассчитывают как время между нажатием и отпусканием клавиши. В случае вычисления величины *Dwell Time* для n-грамм используют временной интервал между нажатием первой клавиши и отпусканием последней клавиши n-граммы, т. е. интервал D_1U_n.

Время между нажатиями (*Flight Time, FT*) характеризует время между отпусканием одной клавиши и нажатием следующей, т. е. время U_1D_2 [17, 18]. Для n-грамм значение *Flight Time* может быть вычислено как сумма *FT* для каждого интервала, входящего в n-грамму:

$$FT_n = \sum_{i=1}^{n-1} U_i D_{i+1}.$$

Продолжительность (*Duration, Dur*) — временной промежуток между нажатиями двух клавиш: промежуток D_1D_2. Для n-грамм значение *Duration* вычисляют как сумму *Duration* для каждой пары клавиш, входящих в n-грамму:

$$Dur_n = \sum_{i=1}^{n-1} D_i D_{i+1}.$$

Частотные характеристики нередко используют в качестве характерных признаков. Их рассчитывают для различных групп символов, таких как символы алфавита, цифры, знаки пунктуации, специальные символы (например, *, ?, %, №, $, #), функциональные символы (*NumLock*, *CapsLock*, *Cntr*, *Alt*, *PageUp*, *PageDown*).

Частоту ошибок вычисляют на основе частоты нажатий клавиш *Delete*, *Backspace*, стрелочки, *Home*, *End*, которые пользователи обычно используют для редактирования введенных текстов.

На основе временных характеристик обычно вычисляют статистические признаки, такие как среднее значение этих характеристик и значение стандартного отклонения. Так, в работах [4, 19] в качестве ВХП используют значения средних и стандартных отклонений для би-грамм и для каждой клавиши отдельно. В работе [3] в качестве ВХП использованы только средние значения и значения стандартных отклонений характеристики *Latency* для редко встречающихся би-грамм. Статистические характеристики в работе [20] определены для би/три-грамм и последовательностей большей размерности (n-грамм).

Статистические признаки в работе [21] вычислены для следующих групп би-грамм: клавиши правой половины клавиатуры; клавиши левой половины клавиатуры; пробел и символ *Backspace*. На основе комбинаций клавиш из этих групп составлено 16 пар клавиш. Далее для каждой пары определены статистические временные характеристики КП. Такой подход позволил сократить время составления ВХП и повысить стабильность значений средних и стандартных отклонений.

Более сложная техника группировки клавиш использована в работе [22]. Здесь клавиши объединены в восемь локаций в зависимости от их расположения на клавиатуре. Клавиатуру делят пополам, затем каждую половину разбивают на четыре группы по расположению клавиш в строках.

Другой подход к отбору символов и би-грамм применен в работах [23, 24]. Здесь для анализа и построения ВХП выбирают наиболее часто встречающиеся в английском языке символы и би-граммы. Затем для каждого символа и би-граммы вычисляют значения величин *Duration* и *Latency* и рассчитывают значения средних и стандартных отклонений этих характеристик. Применение такого подхода также позволяет повысить стабильность значений указанных статистических признаков.

Разбиение данных. Получение данных достаточного размера для высококачественного обучения классификатора является сложной задачей для любого исследователя. Кроме того, необходимо обеспечить наличие набора данных для валидации построенной модели классификатора. Для этого данные каждого пользователя разделяют на две части: одну часть используют в процессе обучения классификатора, другую — для его валидации. Способ разделения исходной выборки на части зависит от выбранной стратегии валидации, которые рассмотрены далее.

Построение классификатора. При построении классификаторов пользователей по КП применяют различные методы (метрические, деревья решений, методы на основе построения искусственных нейронных сетей).

Метрические методы являются наиболее простыми методами, базирующимися на использовании некоторой метрики, характеризующей степень схожести/различия между исследуемым и размеченными ВХП. На основании сходства между этими векторами предполагают наличие принадлежности исследуемого ВХП к одному из целевых классов. В качестве такой метрики могут выступать евклидово расстояние [22], взвешенное евклидово расстояние [19], манхэттенское расстояние [25], расстояние Бхэттэчарья [26].

При построении классификатора в работе [21] использована совокупность двух метрик. Первая из них отражает степень различия между двумя ВХП, вторая является мерой абсолютного расстояния между этими векторами. Несмотря на то что в данном исследовании учитываются исключительно временные характеристики n-грамм, данный подход позволил добиться более высокой точности классификации по сравнению с точностью, полученной другими авторами. Впоследствии данный метод получил развитие в работах [27–29].

Деревья решений часто используют для построения классификаторов [30]. Существует несколько алгоритмов построения деревьев решений, например *Iterative Dichotomiser* 3 (*ID*3), *C*4.5, *Classification and Regression Tree* (*CART*), *Chi-squared Automatic Interaction Detector* (*CHAID*), *Multivariate Adaptive Regression Splines* (*MARS*). Принцип построения деревьев решений основан на концепции рекурсивного деления обучающего множества ВХП на ряд подмножеств. Разделение множеств на основании установленного критерия осуществляется в узлах дерева. Выбор атрибута разбиения может осуществляться различными способами. В алгоритмах *ID*3, *C*4.5 выбор атрибута происходит так, чтобы прирост информации, полученный в результате уточнения его значения, был максимальным.

Методы на основе построения искусственных нейронных сетей являются одними из наиболее перспективных и быстро развивающихся направлений исследований в различных областях науки и техники. В связи с этим многие ученые предпринимают попытки использования этих методов, в том числе и для решения задачи классификации по КП.

Для классификации по КП часто используют такие архитектуры нейронных сетей, как *Multilayer Perceptron* (*MP*), *Probabilistic Neural Network* (*PNN*), *Radial Basis Function Neural Network* (*RBFNN*), *Learning Vector Quantization* (*LVQ*), самоорганизующиеся карты (*SOM*) [31, 32]. Точность и скорость прогнозирования с использованием этих моделей могут быть высокими, однако для построения высококачественной модели требуется наличие большого объема данных в обучающей выборке. Кроме того, результаты работы таких моделей часто сложно интерпретировать.

В исследованиях часто применяют такие хорошо известные методы классификации, как *KNN*, *SVM* и «наивный» классификатор Байеса (*Naive Bayes Classifier*), точность классификации которых может превосходить точность, полученную с использованием перечисленных выше методов.

Для отбора наиболее значимых признаков применяют различные популяционные методы оптимизации: оптимизация методом роя частиц (*Particle Swarm Optimization*); оптимизация колонией муравьев (*Ant Colony Optimization*); варианты генетического алгоритма (*Genetic Algorithm*). В некоторых случаях применение этих методов позволяет существенно повысить точность классификации [33].

Валидация модели. Заключительным этапом создания любого классификатора является его валидация и тестирование. Введем вспомогательные обозначения.

True Positive (*TP*) — число предоставлений доступа в информационную систему пользователям, имеющим на это право.

True Negative (*TN*) — число отказов в доступе в информационную систему пользователям, не имеющим на это права.

False Positive (*FP*) — число ошибочных предоставлений доступа в информационную систему пользователям, не имеющим на это права.

False Negative (*FN*) — число ошибочных отказов в доступе в информационную систему пользователям, имеющим на это право.

Качество построенного классификатора определяют исходя из значений следующих показателей.

False Rejection Rate (*FRR*) отражает вероятность ошибочного отказа в предоставлении доступа к информационной системе:

$$FRR = FN / (TP + FN).$$

False Acceptance Rate (FAR) характеризует вероятность ошибочного доступа к информационной системе пользователя, не имеющего на это права:

$$FAR = FP / (FP + TN).$$

Equal Error Rate (*EER*) или *Cross-over Error Rate* (*CER*) определяет точку, в которой значения *FAR* и *FRR* равны при изменении порога срабатывания модели. Чем ниже значение показателя, тем более высокой надежностью обладает построенная модель.

Дополнительно могут использоваться значения следующих показателей.

Accuracy (*Acc*) — процент положительных срабатываний модели:

$$Acc = (TP + TN) / (TP + TN + FP + FN).$$

Error Rate — процент ошибочных срабатываний модели:

$$ErrRate = 1 - Acc = (FP + FN) / (TP + TN + FP + FN).$$

Precision (точность) — процент объектов, верно отнесенных к целевому классу, от общего числа объектов, отнесенных к данному классу: $Precision = TP / (TP + FP)$.

Recall (полнота) — процент объектов, верно отнесенных к целевому классу, от общего числа объектов, принадлежащих к данному классу: $Recall = TP / (TP + FN)$.

F-measure — показатель, учитывающий значения величин *Precision* и *Recall* в совокупности:

$$F\text{-}measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}.$$

Во многих работах используется соотношение

$$FMeasure = \frac{2}{\frac{\beta^2}{1+\beta^2}\frac{1}{Precision} + \frac{1}{1+\beta^2}\frac{1}{Recall}}.$$

Значения величин *Precision*, *Recall*, *F-measure* характеризуют надежность и качество построенного классификатора.

Существуют различные стратегии валидации модели: *k-fold cross validation, leave-one-out, random sub-sampling* и др. Каждая стратегия фиксирует некоторое множество разбиений всей выборки на обучающую и валидационную. Затем для каждого элемента этого множества выполняются обучение модели на обучающей выборке и оценка средней ошибки классификации на объектах валидационной выборки. Завершающий этап — вычисление значения средней ошибки на валидационной выборке среди всех разбиений. Такой подход позволяет получить несмещенную оценку вероятности ошибки и снизить влияние эффекта переобучения.

Заключение. Высокий интерес научного сообщества к исследованиям АИ по КП объясняется весомыми преимуществами такой технологии по сравнению с существующими аналогами. Особый интерес представляют методы АИ по КП на текстах произвольной длины. В настоящее время достигнутая точность классификации с использованием этих методов не позволяет применять их в качестве основного метода аутентификации. Тем не менее известны успешные примеры их комбинации с другими методами, например парольными методами аутентификации. Таким образом, улучшение точности классификации методов такого типа является актуальной задачей.

Очевидно, идея использования данных о взаимодействии пользователя с устройствами ввода информации для АИ не ограничивается лишь анализом данных, собранных с клавиатуры персонального компьютера. Аналогичные исследования ведутся применительно к данным, собранным с мобильных устройств и терминалов банкоматов.

ЛИТЕРАТУРА

[1] Брагина Е.К., Соколов С.С. Современные методы биометрической аутентификации: обзор, анализ и определение перспектив развития. *Вестник АГТУ*, 2016, № 1 (61), с. 40–44.

[2] Alsultan A., Warwick K. Keystroke dynamics authentication: a survey of free-text methods. *IJCSI*, 2013, vol. 10, no. 4, p. 10.

[3] Shepherd S.J. Continuous authentication by analysis of keyboard typing characteristics. *European Convention on Security and Detection*, 1995, pp. 111–114.
DOI: https://doi.org/10.1049/cp:19950480

[4] Monrose F., Rubin A. Authentication via keystroke dynamics. *Proc. 4th ACM CCCS*, 1997, pp. 48–56. DOI: https://doi.org/10.1145/266420.266434

[5] Umphress D., Williams K. Identity verification through keyboard characteristics. *Int. J. Man Mach. Stud.,* 1985, vol. 23, iss. 3, pp. 263–273.
DOI: https://doi.org/10.1016/S0020-7373(85)80036-5

[6] Leggett J., Williams G. Verifying identity via keystroke characteristics. *Int. J. Man Mach. Stud.,* 1988, vol. 28, iss. 1, pp. 67–76.

DOI: https://doi.org/10.1016/S0020-7373(88)80053-1

[7] Leggett J., Williams G., Umphress D. Verification of user identity via keystroke characteristics. In: *Human Factors in Management Information Systems,* 1988, pp. 29–41.

[8] Obaidat M.S., Sadoun B. Verification of computer users using keystroke dynamics. *IEEE Trans. Syst., Man. Cybern. B*, 1997, vol. 27, iss. 2, pp. 261–269.
DOI: https://doi.org/10.1109/3477.558812

[9] Сидоркина И.Г., Савинов А.Н. Три алгоритма управления доступом к КСИИ на основе распознавания клавиатурного почерка оператора. *Вестник Чувашского университета*, 2013, № 3, с. 293–301.

[10] Karnan M., Akila M., Krishnaraj N. Biometric personal authentication using keystroke dynamics: a review. *Appl. Soft. Comput.*, 2011, vol. 11, iss. 2, pp. 1565–1573.
DOI: https://doi.org/10.1016/j.asoc.2010.08.003

[11] Chandrasekar V., Kumar S. Biometric based keystroke dynamics authentication — a review. *Asian J. Res. Soc. Sci. Human.*, 2016, vol. 6, no. 9, pp. 698–718.

[12] Zaidan D., Salem A., Swidan A., et. al. Factors affecting keystroke dynamics for verification data collecting and analysis. *Proc. 8th ICIT*, 2017, pp. 392–398.

[13] Ali M., Monaco J., Tappert C., et. al. Keystroke biometric systems for user authentication. *J. Sign. Process Syst.*, 2017, vol. 86, iss. 2-3, pp. 175–190.
DOI: https://doi.org/10.1007/s11265-016-1114-9

[14] Salem A., Sharieh A., Sleit A., et.al. Enhanced authentication system performance based on keystroke dynamics using classification algorithms. *KSII TIIS*, 2019, vol. 13, no. 8, pp. 4076–4092. DOI: https://doi.org/10.3837/tiis.2019.08.014

[15] Pisani P.H., Lorena A.C., de Carvalho A.C. Adaptive approaches for keystroke dynamics. *IJCNN*, 2015, pp. 1–8. DOI: https://doi.org/10.1109/IJCNN.2015.7280467

[16] Bergadano F., Gunetti D., Picardi C. User authentication through keystroke dynamics. *TISSEC*, 2002, vol. 5, no. 4, pp. 367–397. DOI: https://doi.org/10.1145/581271.581272

[17] Monaco J.V., Ali M.L., Tappert C.C. Spoofing key-press latencies with a generative keystroke dynamics model. *Proc. IEEE 7th BTAS*, 2015, pp. 1–8.

DOI: https://doi.org/10.1109/BTAS.2015.7358795

[18] Al-Obaidi N., Al-Jarrah M. Statistical median-based classifier model for keystroke dynamics on mobile devices. *Proc. 6th ICDIPC*, 2016, pp. 186–191.
DOI: https://doi.org/10.1109/ICDIPC.2016.7470816

[19] Samura T., Nishimura H. Keystroke timing analysis for individual identification in Japanese free text typing. *ICCAS-SICE*, 2009, pp. 3166–3170.

[20] Park S., Park J., Cho S. User authentication based on keystroke analysis of long free texts with a reduced number of features. *Proc. 2nd ICCSNA*, 2010.
DOI: https://doi.org/10.1109/ICCSNA.2010.5588979

[21] Gunetti D., Picardi C. Keystroke analysis of free text. *TISSEC*, 2005, vol. 8, no. 3, pp. 312–347. DOI: https://doi.org/10.1145/1085126.1085129

[22] Sing S., Arya K.V. Key classification: a new approach in free text keystroke authentication system. *Proc. 3rd PACCS*, 2011.
DOI: https://doi.org/10.1109/PACCS.2011.5990168

[23] Bours P. Continuous keystroke dynamics: a different perspective towards biometric evaluation. *Inf. Secur. Tech. Report*, 2012, vol. 17, iss. 1-2, pp. 36–43.
DOI: https://doi.org/10.1016/j.istr.2012.02.001

[24] Curtin M., Tappert C., Villani M. Keystroke biometric recognition on long-text input: a feasibility study. *IMECS*, 2006, p. 5.

[25] Gunetti D., Ruffo G. Intrusion detection through behavioral data. In: Hand D.J., Kok J.N., Berthold M.R. (eds). *Advances in Intelligent Data Analysis. IDA 1999. Lecture Notes in Computer Science*, vol. 1642. Berlin, Heidelberg, Springer, pp. 383–394.
DOI: https://doi.org/10.1007/3-540-48412-4_32

[26] Bours P., Barghouthi H. Continuous authentication using biometric keystroke dynamics. *NISK*, 2009, pp. 1–7.

[27] Janakiraman R., Sim T. Dynamics in a general setting. In: *Advances in Biometrics*, 2007, vol. 4642, pp. 584–593.

[28] Hu J., Gingrich D., Sentosa A. A *k*-nearest neighbor approach for user authentication through biometric keystroke dynamics. *IEEE ICC*, 2008, pp. 1556–1560.
DOI: https://doi.org/10.1109/ICC.2008.301

[29] Davoudi H., Kabir E. A new distance measure for free text keystroke authentication. *Proc. 14th Int. CSICC*, 2009, pp. 570–575.
DOI: https://doi.org/10.1109/CSICC.2009.5349640

[30] Zhong Y., Deng Y. A survey on keystroke dynamics biometrics: approaches, advances and evaluations. In: *Recent Advances in User Authentication Using Keystroke Dynamics Biometrics.* Science Gate Publ., 2015, pp. 1–22.

[31] Lin C.-H., Liu J.-C., Lee K.-Y. On neural networks for biometric authentication based on keystroke dynamics. *Sens. Mater.*, 2018, vol. 30, no. 3 (1), pp. 385–396.
DOI: https://doi.org/10.18494/SAM.2018.1757

[32] Wankhede S., Verma S. Keystroke dynamics authentication system using neural network. *IJIRD*, 2014, vol. 3, no. 1, pp. 157–164.

[33] Vinayak R., Arora K. A survey of user authentication using keystroke dynamics. *IJSRET*, 2015, vol. 4, no. 4, pp. 378–384.

Ямченко Юрий Владимирович — аспирант кафедры «Системы автоматизированного проектирования» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Ямченко Ю.В. Методы решения задач аутентификации и идентификации пользователя на основе анализа клавиатурного почерка. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2020, № 1 (130), с. 124–139.
DOI: https://doi.org/10.18698/0236-3933-2020-1-124-139

METHODS FOR SOLVING USER AUTHENTICATION AND IDENTIFICATION PROBLEMS BASED ON KEYSTROKE DYNAMICS ANALYSIS

Yu.V. Yamchenko yamchenko.y.v@yandex.com

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

Nowadays, information technologies are used in almost every sphere of our life. Large amounts of personal and corporate data are stored in digital form. Thus, the problem of protecting this data from unauthorized access is raised. Access control subsystems and technologies used in these subsystems, in particular, methods of user authentication and identification, play an important role in this case. The paper provides an overview of methods for solving information system user authentication and identification problems based on the analysis of their keystroke dynamics. These tasks are considered in detail, a brief history of methods for keystroke dynamics authentication and identification is described, their advantages and disadvantages compared with other known methods are indicated and the relevance of this line of research is substantiated. The main stages of creating mathematical models for solving these problems are listed. The description of the methods used in each of these stages is also presented

Keywords

Keystroke dynamics, user, authentication, identification, machine learning, classification methods

Received 26.04.2019
Accepted 29.10.2019

REFERENCES

[1] Bragina E.K., Sokolov S.S. Modern methods of biometric authentication: review, analysis and determination of outlook. *Vestnik AGTU* [Vestnik of ASTU], 2016, no. 1 (61), pp. 40–44 (in Russ.).

[2] Alsultan A., Warwick K. Keystroke dynamics authentication: a survey of free-text methods. *IJCSI*, 2013, vol. 10, no. 4, p. 10.

[3] Shepherd S.J. Continuous authentication by analysis of keyboard typing characteristics. *European Convention on Security and Detection*, 1995, pp. 111–114.
DOI: https://doi.org/10.1049/cp:19950480

[4] Monrose F., Rubin A. Authentication via keystroke dynamics. *Proc. 4th ACM CCCS*, 1997, pp. 48–56. DOI: https://doi.org/10.1145/266420.266434

[5] Umphress D., Williams K. Identity verification through keyboard characteristics. *Int. J. Man Mach. Stud.*, 1985, vol. 23, iss. 3, pp. 263–273.
DOI: https://doi.org/10.1016/S0020-7373(85)80036-5

[6] Leggett J., Williams G. Verifying identity via keystroke characteristics. *Int. J. Man Mach. Stud.*, 1988, vol. 28, iss. 1, pp. 67–76.
DOI: https://doi.org/10.1016/S0020-7373(88)80053-1

[7] Leggett J., Williams G., Umphress D. Verification of user Identity via keystroke characteristics. In: *Human Factors in Management Information Systems*, 1988, pp. 29–41.

[8] Obaidat M.S., Sadoun B. Verification of computer users using keystroke dynamics. *IEEE Trans. Syst., Man. Cybern. B*, 1997, vol. 27, iss. 2, pp. 261–269.
DOI: https://doi.org/10.1109/3477.558812

[9] Sidorkina I.G., Savinov A.N. Three algorithms of control access to the KSII on the basis of recognition of keystroke dynamics. *Vestnik Chuvashskogo Universiteta*, 2013, no. 3, pp. 293–301 (in Russ.).

[10] Karnan M., Akila M., Krishnaraj N. Biometric personal authentication using keystroke dynamics: a review. *Appl. Soft Comput.*, 2011, vol. 11, iss. 2, pp. 1565–1573.
DOI: https://doi.org/10.1016/j.asoc.2010.08.003

[11] Chandrasekar V., Kumar S. Biometric based keystroke dynamics authentication — a review. *Asian J. Res. Soc. Sci. Human.*, 2016, vol. 6, no. 9, pp. 698–718.

[12] Zaidan D., Salem A., Swidan A., et. al. Factors affecting keystroke dynamics for verification data collecting and analysis. *Proc. 8th ICIT*, 2017, pp. 392–398.

[13] Ali M., Monaco J., Tappert C., et. al. Keystroke biometric systems for user authentication. *J. Sign. Process Syst.*, 2017, vol. 86, iss. 2-3, pp. 175–190.
DOI: https://doi.org/10.1007/s11265-016-1114-9

[14] Salem A., Sharieh A., Sleit A., et. al. Enhanced authentication system performance based on keystroke dynamics using classification algorithms. *KSII TIIS*, 2019, vol. 13, no. 8, pp. 4076–4092. DOI: https://doi.org/10.3837/tiis.2019.08.014

[15] Pisani P.H., Lorena A.C., de Carvalho A.C. Adaptive approaches for keystroke dynamics. *IJCNN*, 2015, pp. 1–8. DOI: https://doi.org/10.1109/IJCNN.2015.7280467

[16] Bergadano F., Gunetti D., Picardi C. User authentication through keystroke dynamics. *TISSEC*, 2002, vol. 5, no. 4, pp. 367–397. DOI: https://doi.org/10.1145/581271.581272

[17] Monaco J.V., Ali M.L., Tappert C.C. Spoofing key-press latencies with a generative keystroke dynamics model. *Proc. IEEE 7th BTAS*, 2015, pp. 1–8.
DOI: https://doi.org/10.1109/BTAS.2015.7358795

[18] Al-Obaidi N., Al-Jarrah M. Statistical median-based classifier model for keystroke dynamics on mobile devices. *Proc. 6th ICDIPC*, 2016, pp. 186–191.
DOI: https://doi.org/10.1109/ICDIPC.2016.7470816

[19] Samura T., Nishimura H. Keystroke timing analysis for individual identification in Japanese free text typing. *ICCAS-SICE*, 2009, pp. 3166–3170.

[20] Park S., Park J., Cho S. User authentication based on keystroke analysis of long free texts with a reduced number of features. *Proc. 2nd ICCSNA*, 2010.
DOI: https://doi.org/10.1109/ICCSNA.2010.5588979

[21] Gunetti D., Picardi C. Keystroke analysis of free text. *TISSEC*, 2005, vol. 8, no. 3, pp. 312–347. DOI: https://doi.org/10.1145/1085126.1085129

[22] Sing S., Arya K.V. Key classification: a new approach in free text keystroke authentication system. *Proc. 3rd PACCS*, 2011.
DOI: https://doi.org/10.1109/PACCS.2011.5990168

[23] Bours P. Continuous keystroke dynamics: a different perspective towards biometric evaluation. *Inf. Secur. Tech. Report*, 2012, vol. 17, iss. 1-2, pp. 36–43.
DOI: https://doi.org/10.1016/j.istr.2012.02.001

[24] Curtin M., Tappert C., Villani M. Keystroke biometric recognition on long-text input: a feasibility study. *IMECS*, 2006, p. 5.

[25] Gunetti D., Ruffo G. Intrusion detection through behavioral data. In: Hand D.J., Kok J.N., Berthold M.R. (eds). *Advances in Intelligent Data Analysis. IDA 1999. Lecture Notes in Computer Science*, vol. 1642. Berlin, Heidelberg, Springer, pp. 383–394.
DOI: https://doi.org/10.1007/3-540-48412-4_32

[26] Bours P., Barghouthi H. Continuous authentication using biometric keystroke dynamics. *NISK*, 2009, pp. 1–7.

[27] Janakiraman R., Sim T. Dynamics in a general setting. In: *Advances in Biometrics*, 2007, vol. 4642, pp. 584–593.

[28] Hu J., Gingrich D., Sentosa A. A *k*-nearest neighbor approach for user authentication through biometric keystroke dynamics. *IEEE ICC*, 2008, pp. 1556–1560.
DOI: https://doi.org/10.1109/ICC.2008.301

[29] Davoudi H., Kabir E. A new distance measure for free text keystroke authentication. *Proc. 14th Int. CSICC*, 2009, pp. 570–575.
DOI: https://doi.org/10.1109/CSICC.2009.5349640

[30] Zhong Y., Deng Y. A survey on keystroke dynamics biometrics: approaches, advances and evaluations. In: *Recent Advances in User Authentication Using Keystroke Dynamics Biometrics*. Science Gate Publ., 2015, pp. 1–22.

[31] Lin C.-H., Liu J.-C., Lee K.-Y. On neural networks for biometric authentication based on keystroke dynamics. *Sens. Mater.*, 2018, vol. 30, no. 3 (1), pp. 385–396.
DOI: https://doi.org/10.18494/SAM.2018.1757

[32] Wankhede S., Verma S. Keystroke dynamics authentication system using neural network. *IJIRD*, 2014, vol. 3, no. 1, pp. 157–164.

[33] Vinayak R., Arora K. A survey of user authentication using keystroke dynamics. *IJSRET*, 2015, vol. 4, no. 4, pp. 378–384.

Yamchenko Yu.V. — Post-Graduate Student, Department of Computer Aided Design, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Yamchenko Yu.V. Methods for solving user authentication and identification problems based on keystroke dynamics analysis. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2020, no. 1 (130), pp. 124–139 (in Russ.).
DOI: https://doi.org/10.18698/0236-3933-2020-1-124-139