УДК 004.93'11
DOI: 10.46548/21vek-2021-1055-0018

МЕТОД ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ ДОВЕРИЯ

Пащенко Дмитрий Владимирович, доктор технических наук, профессор, ректор
Бальзанникова Елена Алексеевна, младший научный сотрудник
центра сопровождения научных исследований
Пензенский государственный технологический университет
(440039, Пенза, проезд Байдукова/ул. Гагарина, 1а/11
e-mails: dmitry.pashchenko@gmail.com, elenabalzannikova@gmal.com)

Аннотация. В этой статье рассматривается метод динамической идентификации пользователя по динамике нажатия клавиш на основе ранее предложенного представления в виде контекстов состояний. Предложенный подход позволяет использовать широкий набор алгоритмов и подходов, применимых как к статическому анализу, так и динамическому анализу клавиатурного почерка. Рассмотрен алгоритм идентификации с использованием подхода классификатора ансамблей, состоящего из трех отдельных классификаторов. Для повышения точности идентификации было решено применять метод адаптивного усиления. Кроме того, поскольку результат работы ансамблья классификаторов отражает решение по одному контесту состояния, для объединения последовательности результатов работы классификатора и вынесения идентификационного решения была рассмотрена возможность применения модели доверия. Результаты применения данного подхода показаны на примере динамики изменения уровня доверия предоставляемого биометрического образа в процессе непрерывной идентификации как авторизованного пользователя, так и пользовтеля, который идентифицируется в системе как «чужой».

Ключевые слова: клавиатурный почерк, биометрия, системное программное обеспечение, поведенческая биометрия, динамическая идентификация, клавиатура, анализ данных, машинное обучение.

A METHOD FOR IDENTIFYING A USER BY KEYBOARD HANDWRITING USING A TRUST MODEL

Pashchenko Dmitriy Vladimirovich, doctor of Technical Sciences, professor, rector
Balzannikova Elena Alekseevna, junior researcher of the scientific research department
Penza State Technological University
(440039, Penza, Baydukov passage / Gagarin street, 1a / 11,
e-mails: dmitry.pashchenko@gmail.com, elenabalzannikova@gmal.com)

Abstract. This article discusses a method for dynamically identifying a user based on the dynamics of keystrokes based on the previously proposed representation in the form of state contexts. The proposed approach allows you to use a wide range of algorithms and methods, applicable to both static analysis and dynamic analysis of keystroke dynamics. An identification algorithm is considered using the ensemble classifier approach, which consists of three separate classifiers. To improve the identification accuracy, it was decided to use the adaptive amplification method. In addition, since the result of the work of the ensemble classifiers reflects the decision of only one state context, to combine the sequence of the results of the classifier's work and make an identification decision, the possibility of using a trust model was considered. The results of applying this approach are shown on the example of the example of the dynamics of change in the level of confidence for a biometric image, while continuous identification of both an authorized user and a user who is identified in the system as a "stranger".

Keywords: keyboard handwriting, biometrics, system software, behavioral biometry, dynamics identification, keyboard, data analysis, machine learning.

Введение. На сегодняшний день проблемы предотвращения несанкционированного доступа к конфиденциальной и личной информации, ее незаконного распространения, а также предотвращения противоправных действий от имени другого пользователя являются актуальными задачами в области информационной безопасности. Традиционно для защиты информационных систем используются криптографические средства и средства аутентификации на основе знаний (кодовая фраза или ответ на определенный вопрос) или атрибута (смарт-карта, ключ). Кроме того, существует класс методов биометрической идентификации и аутентификации, которые используются как в качестве автономных решений, так и в качестве усиления традиционных инструментов безопасности. Биометрические методы имеют несколько преимуществ по сравнению с другими методами, основанными на знаниях или атрибутах: источник биометрических данных всегда с пользователем, их нельзя потерять, а взлом и копирование биометрического изображения произвести часто чрезвычайно сложно.

Для этих целей могут использоваться стандартные устройства ввода персонального компьютера: клавиатура и мышь. Эти средства идентификации дают дополнительные преимущества: отсутствие дополнительного оборудования и возможность проведения

скрытой процедуры идентификации.

Существует множество методов как динамического, так и статического анализа почерка на клавиатуре. Как описано ранее, большинство методов предназначены для парольной фразы фиксированной длины. Для непрерывной идентификации в большинстве случаев используется статистический метод, для которого доверительный интервал определяется для каждого параметра времени удержания и интервала между щелчками[1-6].

Однако к основному недостатку биометрических методов можно отнести изменения исходного биометрического образа пользователя из-за возрастных изменений, психофизического состояния человека или наличия травм. Этот аспект может затруднить процедуру идентификации или сделать ее полностью невозможной. Кроме того, биометрический анализ часто требует дополнительного и зачастую дорогостоящего оборудования.

Целью данной работы является изучение метода непрерывной идентификации пользователя клавиатурным почерком на основе представления в виде контекстов состояний.

Материалы и результаты исследования. *Метод динамической идентификации пользователей на основе контекстов состояния.* В статье [7] предложен метод представления на основе контекста состояния, который позволяет использовать множество методов, применимых к парольной фразе, путем интерпретации каждого контекста состояния как парольной фразы фиксированной длины. В этом случае необходимо построить классификатор для каждого контекста, но такой подход позволяет использовать наиболее точные методы статической идентификации для динамического анализа почерка клавиатуры пользователя. Однако сегодня в области машинного обучения существуют технологии повышения точности классификации путем объединения набора «слабых» классификаторов в ансамбль.

Основная цель методов ансамблевого классификатора – объединить разные классификаторы в один метаклассификатор, который имеет лучшую обобщаемость, чем каждый классификатор. Есть много способов объединить классификаторы в ансамбль [8].

Самым простым вариантом объединения классификаторов является простое голосование большинством голосов или среднее значение результатов классификации. Для повышения точности данного подхода, возможно взвешенного голосование или взвешенное среднее, где весовой коэффициент классификатора подбирается в зависимости от точности каждого отдельного классификатора.

В качестве развития модели ансамблей классификаторов на основе большинства голосов является технология бэггинга. Данная технология заключается в том, что вместо того, чтобы использовать один и тот же набор данных для обучения отдельных классификаторов в ансамбле, мы выберем исходные образцы (случайные образцы с возвратом) из первоначального набора данных, вот почему бэггинг еще называют агрегацией начальной загрузки. Алгоритм бэггинга впервые предложил Лео Брейман (Leo Breiman) в статье 1994 г. Он также показал, что данный подход может повысить точность нестабильных моделей и снизить степень переобучения [9].

Другим методом улучшения предсказаний является бустинг (*boosting*), идея которого заключается в итеративном процессе последовательного построения частных моделей. В алгоритме усиления ансамбль состоит из очень простых классификаторов, часто называемых слабыми моделями, которые имеют лишь незначительное преимущество по сравнению со случайным угадыванием. Основная идея бустинга заключается в обучении трудноклассифицируемых объектов, позволяя тем самым слабым моделям обучаться на неверно классифицированных обучающих данных с тем, чтобы повысить производительность ансамбля. Отличие от алгоритма бэггинга состоит в изначальной формулировке; определения бустинга таково: алгоритм, использующий случайные подмножества обучающих данных, извлекающийся из обучающей выборки без возврата. Исходный алгоритм бустинга можно представить следующим образом в виде четырех шагов:

1. Извлечь случайное подмножество обучающих данных *d1* без возврата из обучающей выборки *D* для обучения слабого классификатора *C1*.

2. Извлечь случайное подмножество обучающих данных *d2* без возврата из обучающей выборки *b* добавить 50 процентов образцов, которые ранее были неверно классифицированы и обучить слабую модель *C2*.

3. Найти обучающие образцы *d3* в обучающей выборке, для которых результаты *C1* и *C2* расходятся, чтобы обучить слабую модель *C3*.

4. Объединить модель *C1, C2* и *C3* через метод большинства голосов.

Развитием идеи метода «усиления» ансамбля классификаторов является метод «Адаптивного усиления» – *AdaBoost*. В отличие от рассмотренного исходного алгоритма бустинга, *AdaBoost* использует весь набор данных для обучения слабых моделей, где обучающие образцы используются повторно для построения сильного классификатора, который учится на ошибках предыдущих слабых классификаторах в ансамбле.

Одним из недостатков алгоритма адаптивного усиления является необходимость достаточно длинных обучающих выборок. Другие методы линейной коррекции, в частности, бэггинг, способны строить алгоритмы сопоставимого качества по меньшим выборкам данных.

В рамках проводимого исследования применяется алгоритм бустинга, поскольку позволяет строить ансамбль простой ансамбль из небольшого набора различных классификаторов для анализа клавиатурного почерка пользователя согласно рассмотренным ранее методам. Для реализации алгоритма идентификации пользователей согласно биометрическому был ис-

пользован ансамбль из трех классификаторов: идентификатор на основе вероятностного алгоритма, метода опорных векторов и случайного леса [10-13].

Модель доверия. Решение, которое выносит ансамбль классификаторов, относится к каждому анализируемому контексту состояния. Поскольку каждый контекст отражает четыре последовательных событий клавиатуры и интервалы между ними, вынесение ауентификационного решения по одному контексту будут приводить к большому количеству ошибок, как первого, так и второго рода. Следовательно, для проведения корректной идентификации пользователя необходим анализ значений классификатора на основании последовательности контекстов состояний.

Для решения данной задачи могут быть применены несколько подходов. Самым простым вариантом является установка порогового значения для доли отрицательных решений классификатора для окна контекстов определенного размера. В рамках данного подхода определяется размер окна, который означает количество анализируемых контекстов. Результатом анализа данного контекста будет либо «совпадает», что означает, что данный контекст соответствует эталонному образу, либо «не совпадает» в обратном случае. Если в рамках анализируемого окна доля контекстов, результат классификатора отрицательный выше заранее определенного предельного значения. Окно может быть как фиксированного размера, так и плавающего. Размер и пороговое значение подбираются в процессе обучения модели так, чтобы обеспечивать наименьший уровень ошибок первого и второго рода [14].

Еще одним вариантом вынесения идентификационного решения путем анализа значения классификатора является модель доверия. Концепция модели доверия была впервые предложена Борусом [15]. Она заключается в том, что на оценку подлинности текущего пользователя влияет каждое действие, производимое пользователем. Если параметры действия пользователя соответствуют параметрам действий сохраненного образа, уровень подлинности или уровень доверия пользователю повышается. Величина повышения уровня доверия называется награда. Если наблюдается значительно отклонение от эталонных параметров, уровень доверия пропорционально понижается. Данный показатель называется штраф. Величина изменения уровня доверия может быть фиксирована или зависеть от величины отклонения от эталонного образа.

Очевидно, что временные характеристики пользователя не всегда полностью соответствую своему эталонному образцу, и даже в процессе штатной работы могут наблюдаться некоторые отклонения ввиду различных причин, что в свою очередь будет приводить к снижению уровня доверия. Однако большинство параметров будут совпадать с сохраненным образом пользователя, что в совокупности будет поддерживать высокий уровень доверия. В случае подмены оператора, параметры которого, очевидно отличаются, приведет к быстрому снижению уровня доверия. Как только уровень снизится до порогового значения, пользователь будет идентифицирован как чужой. В идеальном случае в подобной системе уровень доверия авторизованного пользователя не опускается ниже предельного значения, и для нелегального пользователя уровень быстро понижается, позволяя оперативно его обнаружить [16].

Согласно предложенной модели доверия с вариативным изменением уровня доверия величина изменения Δ_{Trust} будет вычисляться по следующей формуле:

$$\Delta_{Trust}(sc_i) = min\left\{-D + D \times \left(\frac{1+\frac{1}{C}}{\frac{1}{C}+\exp\left(-\frac{sc_i - A}{B}\right)}\right), C\right\}$$

где sc_i – результат оценки текущего параметра, A – граничное значение между наградой и штрафом, B – ширина сигмоиды данной функции, C и D – максимальные значения награды и штрафа соответственно. Параметры A, B и C подбираются эмпирически.

В рамках данного исследования анализируемые действия будут параметрами контекстов состояний, а величина изменения уровня доверия будет зависеть от результата классификационного решения ансамбля классификаторов, описанного ранее.

На рисунке 1 показан пример динамики изменения уровня доверия для авторизованного пользователя. Как видно на рисунке, уровень доверия колеблется в районе 100%.

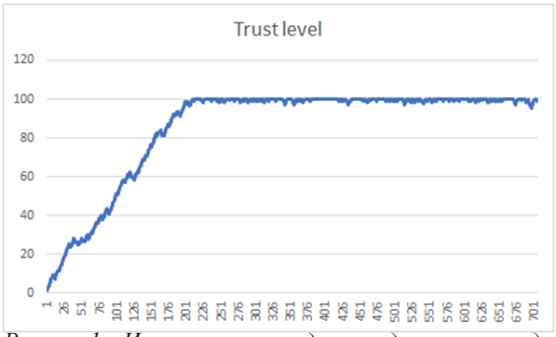

Рисунок 1 – Изменение уровня доверия для случая, когда текущий пользователь соответствует авторизованному

На рисунке 2 показан пример изменения уровня доверия для «чужого» пользователя. На данном примере видно, что, если предоставляемый образ не соответствует зарегистрированному уровень доверия не поднимается выше 40% и в зависимости от порога доступа система идентифицирует подмену оператора.

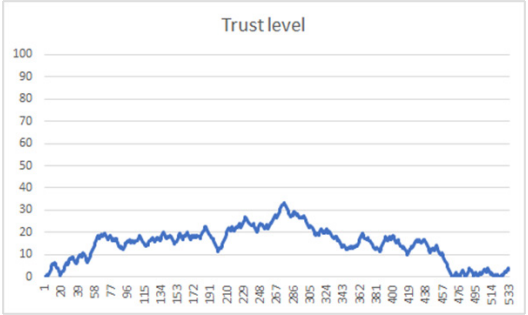

Рисунок 2 – Изменение уровня доверия для случая, когда текущий пользователь не соответствует авторизованному

Заключение. Таким образом, в данной статье рассматривается метод динамической идентификации клавиатурного почерка на основе предложенного представления на основе контекстов состояний. Преимущество данной презентации – широкий спектр применимых методов анализа. В рамках данной работы использовался метод адаптивного усиления, основанный на ансамбле из трех классификаторов: статического метода, нейронной сети и машины опорных векторов. Чтобы принять решение об идентификации и агрегировать результаты классификации для каждого контекста состояния, используется оценка доверия.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-90127

СПИСОК ЛИТЕРАТУРЫ:

1. И. Г. Сидоркина, А. Н. Савинов Три алгоритма управления доступом к КСИИ на основе распознавания клавиатурного почерка оператора // Вестник ЧГУ. 2013. №3. URL: https://cyberleninka.ru/article/n/tri-algoritma-upravleniya-dostupom-k-ksii-na-osnove-raspoznavaniya-klaviaturnogo-pocherka-operatora (дата обращения: 03.09.2020).

2. Р.Р. Шарипов, А.С. Катасёв, А.П. Кирпичников Методы анализа клавиатурного почерка пользователей с использованием эталонных гауссовских сигналов // Вестник Казанского технологического университета. 2016. №13. URL: https://cyberleninka.ru/article/n/metody-analiza-klaviaturnogo-pocherka-polzovateley-s-ispolzovaniem-etalonnyh-gaussovskih-signalov (дата обращения: 03.09.2020).

3. А. И. Аверин, Д. П. Сидоров Аутентификация пользователей по клавиатурному почерку // Огарёв-Online. 2015. №20 (61). URL: https://cyberleninka.ru/article/n/autentifikatsiya-polzovateley-po-klaviaturnomu-pocherku (дата обращения: 03.09.2020).

4. Shen Teh P., Beng Jin Teoh A, Yue S. “A Survey of Keystroke Dynamics Biometrics,” The Scientific World Journal, –Hindawi Publishing Corporation, 24 c, 2013.

5. Vuyyuru, Sampath K. et al. “Computer User Authentication using Hidden Markov Model through Keystroke Dynamics.” (2006).

6. Syed Idrus, Syed Zulkarnain & Cherrier, Estelle & Rosenberger, Christophe & Bours, Patrick. (2014). Soft Biometrics for Keystroke Dynamics.

7. «Imitation model for keystroke dynamics base on state contexts representation» Dmitry V. Pashchenko et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 994 012001. (https://iopscience.iop.org/article/10.1088/1757-899X/994/1/012001).

8. Д. В. Пащенко, Е. А. Бальзанникова , И. Г. Сергина, ”Метод идентификации пользователей по биометрическому образу клавиатурного почерка с использованием двусвязного представления,” Вопросы радиоэлектроники, № 12, С 83–89, 2018, DOI 10.21778/2218-5453-2018-12-83-89.

9. Ю.А. Брюхомицкий “Статистические методы распознавания клавиатурного почерка,” Известия Южного федерального университета. Технические науки, Тематический выпуск, С 139 – 147, 2010.

10. Ю.А. Брюхомицкий “Гистограммный метод распознавания клавиатурного почерка,” Известия Южного федерального университета. Технические науки, т.112, №11, 8 с, 2010.

11. И.А. Ходашинский, М.В.Савчук, И.В.Горбунов, Р.В.Мещеряков “Технология усиленной аутентификации пользователей информационных процессов” Управление, вычислительная техника и информатика. Доклады ТУСУРа, № 2 (24), часть 3,. С 236 – 248, 2011.

12. Д.С Крутохвостов, В.Е. Хиценко, “Парольная и непрерывная аутентификация по клавиатурному почерку средствами математической статистики,” Вопросы кибербезопасности, №5(24), 2017, С 91 – 99, DOI: 10.21681/2311-3456-2017-5-91-99.

13. A Bhatia, Hanmandlu, M. “Keystroke Dynamics Based Authentication Using Information Sets,” Journal of Modern Physics, Vol.8 No.9, August 2017, DOI: 10.4236/jmp.2017.89094.

14. P. Bours, S. Mondal “Continuous Authentication with Keystroke Dynamics,” 2015, DOI:10.13140/2.1.2642.5125.

15. P. Bours, H. Barghouthi “Continuous authentication using keystroke dynamics,” Proceedings of the Norsk Informasjonssikk erhetskonferanse (NISK'09). 1-11, 2009.

16. P. Pinto, B. Patrão, H. Santos “Free Typed Text Using Keystroke Dynamics for Continuous Authentication”, Communications and Multimedia Security : 15th IFIP TC 6/TC 11 International Conference, CMS 2014, Aveiro, Portugal, September 25-26, 2014 pp.33-45, DOI:10.1007/978-3-662-44885-4_3.

Статья поступила в редакцию 16.08.2021
Статья принята к публикации 15.09.2021