СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТОВ НЕПРЕРЫВНОЙ ОНЛАЙН-АУТЕНТИФИКАЦИИ И СИСТЕМ ОБНАРУЖЕНИЯ АНОМАЛИЙ ДЛЯ ПОСТОЯННОГО ПОДТВЕРЖДЕНИЯ ЛИЧНОСТИ ПОЛЬЗОВАТЕЛЯ

DOI: 10.36724/2072-8735-2022-16-5-48-55

Уймин Антон Григорьевич,
Владимирский Государственный Университет имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ), г. Владимир, Россия, ***uimin.vlgu@bk.ru***

Морозов Илья Михайлович,
ГБПОУ МИК, г. Владимир, Россия

Manuscript received 25 April 2022;
Accepted 12 May 2022

Ключевые слова: *компьютерная мышь, биологические данные, пользователь, данные, аутентификация, модель, анализ, машинное обучение, дерево решений, метод k-ближайших соседей, алгоритм случайных лесов, сверхточные нейросети*

Постановка задачи: **увеличение безопасности системы актуализирует вопросы связанные с подтверждением личности пользователя различными методами. Известные способы повышения безопасности, непрерывной аутентификации по биометрическим параметрам, таких как движение компьютерной мышью, скорость нажатия кнопок мыши является сложной и не дает стопроцентную точность.** ***Целью работы*****: является изучение существующих решение в области обнаружения аномалий и систем непрерывной аутентификации. Предлагается выработать параметры для снятия данных с компьютерной мыши, требования для создания DataSet, эффективность анализа и извлечения основных признаков из необработанных данных. Рассмотреть модели DL для CA и AD, позволяющие верифицировать пользователя, произвести их качественное сравнение.** ***Используемые методы*****: в исследовании применяется сравнительный анализ технологий. Таких как, дифференцированного машинного обучения (ML), включая классификатор дерева решений (DT), метод k-ближайших соседей (k-NN), алгоритм случайных лесов (RF) и сверхточных нейросетей (CNN).** ***Результат*****: на основе сравнительный анализа инструментов непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя. В результате работы определено, что существующие методы, связанные с CA и AD описывают в основном лабораторные исследования, и носят характер подтверждения гипотез. В рассмотренных исследованиях описаны использованные параметры, по разработке программного обеспечение для снятия данных о действиях мыши. Исследования показывают, что инструменты сбора незначительно влияют на качество получаемых данных, при условии предобработки данных с использованием алгоритмов выделения ключевых признаков. Большинство исследований рассматривает наборы данных полученных, либо из свободных источников, либо из групп добровольцев в диапазоне от 10 до 60 человек. Необходимо отметить что количество превышающее 10 можно считать достаточным для определения репрезенторотивной выборки. Определено, что наибольшее значение на результат дали признаки описывающие описывает перемещение мыши между двумя местоположениями экрана. Рассмотрены модели DL для CA и AD, позволяющих верифицировать пользователя, произведено их качественное сравнение.** ***Практическая значимость*****: проведенное исследование позволяет судить о возможности реализации решения по непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя в рамках разработанной ранее системы проведения дистанционных чемпионатов RemoteTopology.**

Уймин Антон Григорьевич*, соискатель ученой степени кандидата технических наук. Кафедра радиотехники и радиосистем. Владимирский Государственный Университет имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ). Область научных интересов: информационная безопасность, мониторинг информационных ресурсов, сбор и обработка информации, г. Владимир, Россия*
Морозов Илья Михайлович, *технический специалист ГБПОУ МИК. Сертифицированный эксперт демонстрационного экзамена, сертифицированный инструктор академии Huawei, с Область научных интересов: информационная безопасность, мониторинг информационных ресурсов, сбор и обработка информации, г. Владимир, Россия*

Для цитирования:
Уймин А.Г., Морозов И.М. Сравнительный анализ инструментов непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя // T-Comm: Телекоммуникации и транспорт. 2022. Том 16. №5. С. 48-55.

For citation:
Uymin A.G., Morozov I.M. (2022) Comparative analysis of continuous online authentication tools and anomaly detection systems for permanent confirmation of the user's identity. *T-Comm*, vol. 16, no. 5, pp. 48-55. (*in Russian*)

В современную эпоху цифровизации и повсеместного внедрения дистанционной работы можно особо выделить проблемы информационной безопасности, связанные с идентификацией и аутентификацией пользователей. Большинство систем решают данную проблему путем применения простых решений на основе пар логин-пароль или более сложных, на основе двухфакторной аутентификации с применением дополнительных средств, как правило, мобильных устройств. Возможно, применение биометрических средств защиты.

В основе принципа биометрической аутентификации лежат как физические, так и поведенческие свойства индивида. Примером физических свойств является работа с радужной оболочкой, лица, отпечатков пальцев, и голоса. Физическая биометрия дает высокую точность аутентификации, при этом обладая рядом потенциальных проблем конфиденциальности [1,2,3]. Альтернативой является поведенческая биометрия, которая позволяет обеспечить дополнительную систему защиты от мошеннических действий пользователей [4,5]. Принцип поведенческой биометрии основан на деятельности субъекта и чертах его моторных действий, таких как образцы почерка человека, образцы клавиатурного почерка человека, походка, динамика мыши, нажатие клавиш и т.д. [6,7]. Биометрия поведения представляет собой количественный метод, обеспечивающий генерацию (составление) поведенческих профилей пользователя. Компьютерная безопасность сегодня важна как для отдельных пользователей, так и для общества, бизнеса, производства, ввиду непрекращающихся процессов цифровизации. Для работы, как правило, применяются системы непрерывной аутентификации (continuous authentication systems, далее CAs) и системы обнаружения аномалий (anomaly detection systems, далее ADs), входящие в состав части специализированных систем контроля работы пользователей. Интеграция биометрической аутентификации в такие решения позволит уникально определять личность пользователя, а также обеспечит более сильную аутентификацию.

Поведенческая биометрия дает преимущества по сравнению с традиционными методами аутентификации, так снятие данных поведения мыши и клавиатуры является незаметным для пользователя, но при этом, позволяет обеспечивать его непрерывную аутентификацию. Кроме этого, поведенческий биометрический анализ уникального пользователя при работе мыши и клавиатуры не требует доступа к конфиденциальным данным пользователя, что позволяет увеличить область использования решения [8]. Анализ динамики трехпозиционного графического манипулятора, типа «мышь», является примером поведенческой биометрии, которая может сохранять и анализировать действия от устройства ввода, такие как: перемещение, перетаскивание, щелчок. Динамика действий мыши показывает факторы взаимодействия человека с определённым графическим интерфейсом [9,10].

Методы CA и AD позволяют избежать проблем стандартной однофакторной аутентификации, при этом, они не могут использоваться для основного входа в систему, но могут применяться для обеспечения дополнительной верификации пользователя в процессе работы: обнаружения процедур передачи учетной записи, процедур аномальных действий [11,12].

В исследовании применяется сравнительный анализ технологий: дифференцированного машинного обучения (ML), включая классификатор дерева решений (DT), метод k-ближайших соседей (k-NN), алгоритм случайных лесов (RF) и сверхточных нейросетей (CNN) с их классификацией. Целью работы является решение научно-технической задачи анализа разработанных ранее моделей, алгоритмов и процедур непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя. В соответствии с целью были поставлены и решены следующие научные задачи:

1) проведен обзор существующих методов, связанных с CA и AD. рассмотрены способы, в которых они были использованы;
2) определены параметры, по разработке программного обеспечение для снятия данных о действиях мыши;
3) подготовлены требования, для сбора dataset;
4) актуализирована эффективность анализа и извлечения основных признаков из необработанных данных пользователей;
5) рассмотрены модели DL для CA и AD, позволяющих верифицировать пользователя, произведено их качественное сравнение;

Как обсуждали ранее, сеть подвержена некоторым типам угроз безопасности. Неправильное поведение в сети может привести к угрозе безопасности или иным противоправным действиям пользователей. Биометрические технологии могут повысить безопасность системы путём анализа биологических данных о действиях пользователя. Извлечения полученных данных пользователя и сравнение их с образом данных аутентифицированного пользователя. Биометрические системы применяются для обеспечения безопасности в различных областях, например правоохранительные органы, промышленность, бизнес, безопасность информационных систем.

В исследовании [13] описано исследование индикации пользователей через онлайн покупки. Первоначально авторы собрали массив многообразных данных от 28 участников. Затем они провели эксперимент, используя классификатор K ближайших соседей. Авторам удалось достигнуть точности 87. 5% используя исключительно 20 признаков, извлекаемых из данных компьютерной мыши. Авторы в данном исследовании использовали традиционный ML для классификации, используя отдельные функции, которые задавали вручную. В нашем исследовании проводится сравнение продуктивности DL с разнообразными методами ML. Определим различные исследования поведенческих компонентов движения компьютерной мыши термином динамика мыши.

В исследовании [14] авторы изучали аутентификацию пользователей в веб-ресурсах. Исследователи концентрируются исключительно на нажатиях клавиши мыши. В исследовании было использовано сопоставление кумулятивных функций распределения. Из 20 участников опыта выделяли три параметра. Это функции нажатия клавиш, продолжительность нажатия и задержка после нажатия. Они получили достоверность аутентификации пользователя (Будем называть точность аутентификации пользователя термином ACC)(ACC) 44% при применении функций, извлеченных из 100 движений мыши, и они получили точность 96% ACC при извлечении из 100 нажатий клавиши мыши.

В исследовании [15] авторы разработали систему для сбора биометрических данных пользователя с использованием нажатий кнопок мыши. Система разработана для захвата координат x и y компьютерной мыши с отсчетами по времени в миллисекундах в момент события. Пользователям необходимо нажимать на кнопки, которые появляются на экране, перемещая компьютерную мышь. Пользователю необходимо нажать 20 кнопок для того, чтобы завершить эксперимент. После того как пользователь нажимает одну кнопку, появляется следующая кнопка в случайном месте. Данные действия необходимо повторить шесть раз для создания частного отпечатка пользователя в контролируемой среде. В исследовательском подходе использовались сведения пяти пользователей с одним и тем же ноутбуком и мышью. Авторы собрали 30 разнообразных файлов для всех пользователей по 6 файлов на каждого пользователя. Исследователи применяли евклидово расстояние для алгоритма идентификации.

В ходе опыта было получено 14 схожих файлов из 30, что соответствует точности в 46,67%. Они попытались идентифицировать по движениям ладоней или рук пользователей, используя компьютерную мышь, и получили (Будем называть вероятность того, что пользователь, который должен быть принят, будет отклонен системой термином FRR) FRR 53,55%. Разрабатываемая биометрическая система применяется для идентификации или для проверки, как часть аутентификации, в соответствии с подходом, описанным Hamid et al. (2011). Решения ориентируются исключительно на идентификации пользователей.

В исследовании [16] авторы оценивают движение мыши как биометрический показатель. Они предложили два метода аутентификации: первый из них предназначен для первоначального входа пользователя в систему (регистрация), второй метод – это отслеживание компьютера на наличие сомнительных действий (проверка). Это требует от пользователя 20 секунд для выполнения каждого из двух методов. На этапе регистрации пользователь должен использовать мышь и следить за серией точек, которые отображаются по одной на экране. Цель этого шага – заносить координаты мыши каждые 50 мс, а затем вычислять скорость, отклонения от прямой и угол. Авторы использовали данные, собранные на этапе регистрации, для этапа проверки путем сопоставления данных пользователя и данных слепка. Исследование было протестировано на 15 участниках разного возраста от 22-30 лет. Было достигнуто 20% погрешности при применении 1,5 стандартных отклонений среднего значения от соответствующего набора значений и уровень погрешности 15% с использованием 1 стандартного отклонения среднего от соответствующего значение регистрации. На этапе регистрации для пассивной аутентификации они запускали программу в фоновом режиме, чтобы записать координаты мыши для более кратковременного периода времени. На каждого участника было выделено 15 минут тестового времени.

В исследовании [17] авторы разработали систему сбора данных для сбора информации о пользователях по использовании компьютерной мыши. Система записывает все взаимодействия пользователей в сети интернет. Набор данных был собран у 50 участников; у каждого пользователя было 400 ходов. Ход определяется как группа точек между двумя действиями. Авторы предложили 58 поведенческих признаков, извлеченных из необработанных данных с помощью некоторых математических операций. Эти особенности использовались для идентификации пользователя на основе того, как они взаимодействует с системой. Авторы разработали последовательный классификатор с использованием методов статистического распознавания образов, чтобы различать пользователей. Авторы добились одинакового уровня ошибок 0,7%, на 100 ходов мыши. Система использовала только характеристики взаимодействия пользователя, а не производительность пользователя.

В исследовании [18] авторы предложили подход к повторной аутентификации с использованием действий мыши пользователя. Они собрали исходные данные от 11 студентов-добровольцев, которые провели около двух часов за своими персональными компьютерами в неконтролируемой среде. Добровольцы использовали Internet Explorer на компьютере с Windows для сбора данных. Эксперимент был сосредоточен на использовании только приложений Internet Explorer, чтобы уменьшить сложность распознавания поведения пользователей. Исследователи использовали метод контролируемого обучения – C5.0 дерево принятия решений. Авторы разделили набор данных на 70% для обучения и 30% только для тестирования. Они получили средний показатель ложного принятия (FAR) 1,75% и средний показатель ложного отклонения (FRR) 0,43%. Авторы не собирали данные о движениях мыши за пределами приложения Internet Explorer, поскольку их частота была высокой.

В исследовании [19] авторы разделили типы действий мыши на три категории: перемещение мыши (MM), клик мышью (PK) и перетаскивание (DD). В последующем исследовании [19], основанном на том же наборе данных и использующем все три типа действий мыши, они собрали свои данные от 22 участников в течение 998 сеансов и провели эксперименты по аутентификации пользователей. Авторы предложили новую форму поведенческой биометрии с помощью динамики компьютерной мыши: метод обнаружения с использованием нейронной сети. Они достигли относительно высокого уровня ложного принятия (FAR) в 2,4649% и уровня ложного отклонения (FRR) в 2,4614%.

В исследовании [20] авторы описали результаты измерений на расширенном наборе данных из 48 пользователей. Они предложили систему биометрического распознавания динамики мыши для идентификации «коммерческих» пользователей. Для объединения соответствующих биометрических показателей был использован метод нечеткого классификатора. Авторы сообщили о результатах с коэффициентом ложного принятия 0% и коэффициентом ложного отклонения 0,36%. В части исследований использовались два типа действий мыши из трёх определённых ранее. Zheng et al. (2011) разработали надежный и эффективный механизм непрерывной аутентификации, используя только действия с щелчком мыши (PC) и движением мыши (MM), как определено в исследовании [21] Ahmed and Traore (Ahmed Awad E. Ahmed & Traore, 2007) авторы использовали классификатор опорных векторов (SVM) для проверки пользователей. Результаты показали, что их новая система проверки пользователя достигла одинаковой частоты ошибок (Будем называть метод опорных векторов термином EER) (EER) в 1,3% при использовании щелчков мыши и 1,9% при использовании

движений мыши. Такая производительность не соответствует европейскому стандарту контроля доступа, который требует, чтобы «коммерческая» биометрическая система достигала FAR менее 0,001% и FRR менее 1%.

В исследовании [22] авторы предложили новую одномерную архитектуру сверхточной сети с использованием двух наборов данных: общедоступного набора данных Balabit для оценки производительности [23] и набора данных (Набор данных DFL использовался для инициализации весов наших моделей) DFL для обучения передаче. Чтобы избежать переобучения, они использовали функцию активации сигмовидной мышцы и слой отсева с вероятностью 0,15. Более того, их модель 1D-CNN была обучена в Keras с использованием оптимизатора Adam (скорость обучения: 0,002, затухание: 0,0001, функция потерь: двоичная перекрестная энтропия). Они разделили данные о динамике мыши на блоки фиксированного размера и выполнили два типа измерений: измерения с использованием 300 блоков от каждого пользователя (измерение с балансом классов) и измерения с использованием всех блоков данных от каждого пользователя (измерение с несбалансированным классом). Они оценили модель, используя три сценария: (i) ПРОСТЫЕ модели, обученные с нуля с использованием обучающих данных из набора данных Balabit; (ii) модели TRANSFER1, использующие обучение передаче, где модели были предварительно обучены на наборе данных DFL; и (iii) модели TRANSFER2, которые были инициализированы с помощью обучения передаче, а затем веса были обновлены с использованием обучающих данных из набора данных Balabit. Это были результаты для количества блоков (300): PLAIN = 0,63, TRANSFER1 = 0,50 и TRANSFER2 = 0,66. Это были результаты для количества блоков (всех): PLAIN = 0,55, TRANSFER1 = 0,34 и TRANSFER2 = 0,62. В этом исследовании использовался общедоступный набор данных Balabit; он содержит данные мыши только от 10 пользователей, чего может быть недостаточно для создания надежной и безопасной модели пользователя, основанной на динамике мыши.

В исследовании [24] авторы предложили различные стратегии, которые потенциальный злоумышленник может использовать для выполнения синтетически сгенерированных состязательных выборок, используя подходы, основанные на имитации, суррогате или статистике. На основании результатов своих экспериментов они пришли к выводу, что атаки, основанные на нейронных сетях, работают лучше, чем атаки, основанные на статистике. Авторы показали, что генерация последовательностей мыши является сложной задачей для решения/реализации, и, следовательно, авторы предположили, что состязательные атаки имеют свои недостатки при выполнении. Авторы также подробно остановились на способах, с помощью которых надежность этих моделей аутентификации может отрицательно сказаться, даже при реалистичном тестировании: не синтетические тесты. Во второй половине статьи авторы показали механизм получения результатов различных экспериментов, обсуждаемых в этой статье. В конце этой статьи они предоставляют обзор расширения своего подхода к атаке на основе суррогатов. Численных результатов не переведено.

В исследовании [25] авторы предложили систему эмпирического биометрического исследования для идентификации пользователей с использованием различных нейронных сетей в онлайн-игре League of Legends «https://www.leagueoflegends.com». Результаты их экспериментов показали, как различные нейронные сети ведут себя с биометрическими данными и базами данных League of Legends. RBF и байесовские сети показали, что можно улучшить результаты, собирая образцы чаще, несмотря на стоимость обработки. В конце авторы заявляют, что в будущей работе может использоваться алгоритм, в котором анализируются различия между выборками ранней игры, середины игры и поздней игры; этот подход исследует профиль пользователя на каждом уровне, поскольку один и тот же игрок выполняет игру от начала до конца.

Подводя итог этой статье, авторы отметили, что традиционный способ получения подтверждения пользователя простым использованием электронной почты неудобен для пользователя, поскольку многие функции на основе искусственного интеллекта могут быть скомпрометированы, если используется такой подход. Объем собранных данных был недостаточен для того, чтобы эксперименты позволили более точно проверить, не атакует ли пользователь сам проблему совместного использования учетных записей.

В таблице 1 представлено краткое изложение 10 наиболее важных исследований по аутентификации пользователей с использованием динамики мыши.

Таблица 1

Характеристики наиболее важных существующих работ

Количество пользователей	**Окружающая среда**	**Действие мыши**	**Тип исследования**	**Используемые данные**	**ERR**	**F**	**EER**
30	Контролируемая	MM-PC	Непрерывная аутентификация	Собранные данные	0.86%	2.96%	1.3%
25	Неконтролируемая	MM-PC-DD	Непрерывная аутентификация	Собранные данные	17.66%	Not given	8.53%
10	Неконтролируемая	MM-PC-DD	Обнаружение вторжений	Babbit	N/A	N/A	0.04%
39	Контролируемая	MM	Статическая аутентификация	Собранные данные	5.26%	4.59%	N/A
58	Контролируемая	MM-PC	Обнаружение аномалий	Собранные данные	N/A	N/A	11.63%
30	Неконтролируемая	MM-PC	Непрерывная аутентификация	Собранные данные	1.3%	1.3%	1.3%
28	Неконтролируемая	MM-PC-DD	Непрерывная аутентификация	Собранные данные	7.78-2.75%	9.45-3.39%	N/A
31	Контролируемая	MM-PC-DD	Непрерывная аутентификация	Собранные данные	2.10%	2.24%	N/A
52	Неконтролируемая	MM-PC-DD	Непрерывная аутентификация	Собранные данные	N/A	N/A	N/A
50	Контролируемая	MM-PC	Статическая аутентификация	Собранные данные	**2%**	2	0.020%

Столбцы содержат дополнительную информацию, как указано ниже:

Количество пользователей: количество пользователей, принявших участие.

Окружающая среда: место сбора данных о поведении мыши.

Действие мыши: характеристики действий, полученных от устройства ввода мыши для конкретного пользователя при взаимодействии с определенным графическим пользовательским интерфейсом.

Тип исследования: непрерывная аутентификация, обнаружение вторжений или статическая аутентификация.

Используемые данные: набор данных, который был использован для этого исследования.

FAR: коэффициент ложного принятия.

FRR: частота ложных отказов.

EER: равная частота ошибок.

На основе сравнительный анализа инструментов непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя можно определить, что:

1) Существующие методы, связанные с CA и AD описывают в основном лабораторные исследования, и носят характер подтверждения гипотез. Большая часть работ рассматривает ситуации в рамках синтетических тестов, с применением ограничений выборок, что не позволяет судить о применимости методов в реальных (производственных задачах). При этом исследования позволяют сделать вывод о результативности применения алгоритмов;

2) В рассмотренных исследованиях описаны использованные параметры, по разработке программного обеспечение для снятия данных о действиях мыши. Основным условием выделяется сбор параметров операций мыши каждого отдельного пользователя как в контролируемой, так и не контролируемой среде. Исследования показывают, что инструменты сбора незначительно влияют на качество получаемых данных, при условии предобработки данных с использованием алгоритмов выделения ключевых признаков;

3) Большинство исследований рассматривает наборы данных полученных, либо из свободных источников, либо из групп добровольцев в диапазоне от 10 до 60 человек. Необходимо отметить что количество превышающее 10 можно считать достаточным для определения репрезенторативной выборки. Как описывалось выше, полноту исследования обеспечит не менее 20 участников эксперимента. Влияние понимания сути эксперимента значительно при сборе dataset;

4) Произведен анализ извлечения основных признаков из необработанных данных пользователей. Определено, что наибольшее значение на результат дали признаки описывающие описывает перемещение мыши между двумя местоположениями экрана; признаки описывающие наведение и щелчок мыши, перемещение и наведение мыши на точку (определение объекта средствами операционной системы - фокус на объекте) и затем нажатие одной из кнопок мыши; признаки описывающие движение перетаскивания мыши, инициируемое нажатием основной кнопки мыши и завершающееся ее отпусканием;

5) Рассмотрены модели DL для CA и AD, позволяющих верифицировать пользователя, произведено их качественное сравнение. Что демонстрирует таблица 1

Проведенное исследование позволяет судить о возможности внедрения решения по непрерывной онлайн-аутентификации и систем обнаружения аномалий для постоянного подтверждения личности пользователя в рамках разработанной ранее системы проведения дистанционных чемпионатов [25] RemoteTopology [26, 27]. В рамках системы RemoteTopology задача системы непрерывной аутентификации будет заключатся в сборе информации о поведении пользователя в реальном времени посредством анализа динамики мыши. Поведение текущего пользователя будет сравниваться с данными, хранящимися в базе данных системы о поведении верифицированного пользователя на основе собранных ранее данных. По результатам этого сравнения система делает вывод доверять пользователю и продолжать работу на устройстве, либо не доверять и исключить пользователя из системы с последующей статической аутентификации пользователя.

Литература

1. *Абзалов А.Р., Кашапов И.И., Орлов А.Ю., Мамлеев И.Р.* Аутентификация пользователей на основе трехступенчатой модели клавиатурного почерка // Вестник ДГТУ. Технические науки. 2020. № 3. URL: https://cyberleninka.ru/article/n/autentifikatsiya-polzovateley-na-osnove-trehstupenchatoy-modeli-klaviaturnogo-pocherka (дата обращения: 17.01.2022).

2. Государственная регистрация программы для ЭВМ Remote Topology-Интерфейс пользователя [Электронный ресурс] / Антон Григорьевич Уймин, Сергей Васильевич Любкин. URL: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2021661218&TypeFile=html

3. Государственная регистрация программы для ЭВМ Remote Topology-модуль авторизации [Электронный ресурс] / Антон Григорьевич Уймин, Сергей Васильевич Любкин. URL: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2021619990&TypeFile=html

4. *Зудинов А.С.* Внедрение биометрии в системы контроля доступа на объектах критической информационной инфраструктуры // StudNet. 2021. № 5. URL: https://cyberleninka.ru/article/n/vnedrenie-biometrii-v-sistemy-kontrolya-dostupa-na-obektah-kriticheskoy-informatsionnoy-infrastruktury (дата обращения: 17.01.2022).

5. *Косарев В.Е., Русило Э.С.* Биометрия в банках и факторы, сдерживающие ее развитие // Финансовые рынки и банки. 2021. № 3. URL: https://cyberleninka.ru/article/n/biometriya-v-bankah-i-faktory-sderzhivayuschie-ee-razvitie (дата обращения: 17.01.2022).

6. *Линдигрин А.Н.* Анализ специфики и проблематики процессов поиска аномалий в сетевых данных // Известия ТулГУ. Технические науки. 2021. № 5. URL: https://cyberleninka.ru/article/n/analiz-spetsifiki-i-problematiki-protsessov-poiska-anomaliy-v-setevyh-dannyh (дата обращения: 17.01.2022).

7. *Ложкин Л.Д., Анкина К.П.* Проверка подписи на основе информации об уровне серого с использованием нейронной сети // StudNet. 2021. № 5. URL: https://cyberleninka.ru/article/n/proverka-podpisi-na-osnove-informatsii-ob-urovne-serogo-s-ispolzovaniem-neyronnoy-seti (дата обращения: 17.01.2022).

8. *Поляничко М.А.* Методика обнаружения аномального взаимодействия пользователей с информационными активами для выявления инсайдерской деятельности // Труды учебных заведений связи. 2020. № 1. URL: https://cyberleninka.ru/article/n/metodika-obnaruzheniya-anomalnogo-vzaimodeystviya-polzovateley-s-informatsionnymi-aktivami-dlya-vyyavleniya-insayderskoy (дата обращения: 17.01.2022).

8. *Савенков П.А., Трегубов П.С.* Поиск поведенческих аномалий в деятельности сотрудников при помощи методов пространственно кластеризации, основанных на плотности // Известия ТулГУ. Технические науки. 2020. № 9. URL: https://cyberleninka.ru/article/n/poisk-povedencheskih-anomaliy-v-deyatelnosti-sotrudnikov-pri-pomoschi-metodov-prostranstvennoy-klasterizatsii-osnovannyh-na (дата обращения: 17.01.2022).

10. *Терёхин С.Н., Вострых А.В., Семёнов А.В.* Оценка графических пользовательских интерфейсов посредством алгоритма поиска последовательных шаблонов // Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России. 2020. № 2. URL: https://cyberleninka.ru/article/n/otsenka-graficheskih-polzovatelskih-interfeysov-posredstvom-algoritma-poiska-posledovatelnyh-shablonov (дата обращения: 17.01.2022).

11. *Уймин А.Г.* Развитие профессиональной подготовки с учетом стандартов "Worldskills Russia" // Вестник педагогических наук. 2021. № 1. С. 42–51.

12. *Фуфаев М.Д., Криворучко С.В.* Биометрическая идентификация: сущность и риски применения технологии в платежной индустрии // Международный журнал гуманитарных и естественных наук. 2021. № 2-3. URL: https://cyberleninka.ru/article/n/biometricheskaya-identifikatsiya-suschnost-i-riski-primeneniya-tehnologii-v-platyozhnoy-industrii (дата обращения: 17.01.2022).

13. *Чемирисов В.В.* Методика оценки оперативности типовых действий оператора при вводе данных [Электронный ресурс] // Военная мысль. 2021. № 10. URL: https://cyberleninka.ru/article/n/metodika-otsenki-operativnosti-tipovyh-deystviy-operatora-pri-vvode-dannyh (дата обращения: 17.01.2022).

14. *Ahmed A.A.E., Traore I.* A New Biometric Technology Based on Mouse Dynamics. IEEE Trans. Dependable Secur. Comput, 2007. No.4, pp. 165-179.

15. *Ahmed A.A.E., Traore I.* Dynamic sample size detection in continuous authentication using sequential sampling. In Proceedings of the 27th Annual Computer Security Applications Conference, Orlando, FL, USA, 5-9 December 2011, pp. 169-176.

16. *Antal M., Egyed-Zsigmond E.* (2019) Intrusion detection using mouse dynamics. IET Biom, no. 8, pp. 285–294.

17. *Chudá D., Krátky P.* (2014) Usage of computer mouse characteristics for identification in web browsing // Proceedings of the 15th International Conference on Computer Systems and Technologies-CompSysTech'14, Ruse, Bulgaria, 27–28 June 2014, pp. 218-225. https://dl.acm.org/doi/10.1145/2659532.2659645

18. *Chuda D., Kratky P., Tvarozek J.* (2015) Mouse Clicks Can Recognize Web Page Visitors! In Proceedings of the 24th International Conference onWorldWide Web-WWW '15 Companion, Florence, Italy, 18-22 May 2015, pp. 21-22.

19. *da Silva V.R., Costa-Abreu M.D.* (2018) An empirical biometric-based study for user identification with different neural networks in the online game League of Legends. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8-13 July 2018, pp. 1-6.

20. *Fülöp, Á., Kovács L., Kurics T., Windhager-Pokol E.* (2016) Balabit Mouse Dynamics Challenge Data Set. 2016. Available at: https://medium.com/balabit-unsupervised/releasing-the-balabit-mouse-dynamics-challenge-data-set-a15a016fba6c (accessed on 8 May 2021).

21. *Gamboa H., Fred A.* (2004) A Behavioural Biometric System Based on Human Computer Interaction. Proc. SPIE 2004, 5404, pp. 381-392.

22. *Hamid N.A., Safei S., Satar S.D.M., Chuprat S., Ahmad R.* (2011) Mouse movement behavioral biometric systems. In Proceedings of the International Conference on User Science and Engineering (i-USEr), Selangor, Malaysia, 29 November–1 December 2011, pp. 206-211.

23. *Hashia S., Pollett C., Stamp M.* (2005) On Using Mouse Movements as a Biometric. 5. In Proceedings of the International Conference on Computer Science and its Applications, Singapore, 9-12 May 2005.

24. *Pusara M., Brodley C.E.* (2004) User re-authentication via mouse movements. In Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security-VizSEC/DMSEC '04, Washington, DC, USA, 29 October 2004.

25. *Tan Y.X.M., Iacovazzi A., Homoliak I., Elovici Y., Binder A.* (2019) Adversarial attacks on remote user authentication using behavioural mouse dynamics. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14-19 July 2019, pp. 1-10.

COMPARATIVE ANALYSIS OF CONTINUOUS ONLINE AUTHENTICATION TOOLS AND ANOMALY DETECTION SYSTEMS FOR PERMANENT CONFIRMATION OF THE USER'S IDENTITY

Anton G. Uymin, *Vladimir State University named after Alexander Grigoryevich and Nikolai Grigoryevich Stoletov (VISU), Vladimir, Russia,* ***uimin.vlgu@bk.ru***

Ilya M. Morozov, *GBOU MIC, Vladimir, Russia*

Abstract

Problem statement: increasing the security of the system actualizes issues related to the confirmation of the user's identity by various methods. Known ways to increase security, continuous authentication by biometric parameters, such as the movement of a computer mouse, the speed of clicking mouse buttons is complex and does not give one hundred percent accuracy. *The purpose of the work:* is to study the existing solutions in the field of anomaly detection and continuous authentication systems. It is proposed to develop parameters for removing data from a computer mouse, requirements for creating a DataSet, the effectiveness of analysis and extraction of the main features from the raw data. Consider the DL models for CA and AD, allowing to verify the user, to make their qualitative comparison. *Methods used*: the study uses a comparative analysis of technologies. Such as differentiated machine learning (ML), including the decision tree classifier (DT), the k-nearest neighbor method (k-NN), the random forests algorithm (RF) and ultra-precise neural networks (CNN). *Result*: based on a comparative analysis of continuous online authentication tools and anomaly detection systems for permanent confirmation of the user's identity. As a result of the work, it was determined that the existing methods related to CA and AD describe mainly laboratory studies, and are in the nature of confirming hypotheses. The considered studies describe the parameters used to develop software for taking data about mouse actions. Studies show that the collection tools have little effect on the quality of the data obtained, provided that the data is preprocessed using algorithms for identifying key features. Most studies consider data sets obtained either from free sources or from groups of volunteers ranging from 10 to 60 people. It should be noted that the number exceeding 10 can be considered sufficient to determine a representative sample. It was determined that the greatest value on the result was given by signs describing the movement of the mouse between two locations of the screen. The DL models for CA and AD that allow user verification are considered, and their qualitative comparison is made. *Practical significance*: the conducted research makes it possible to judge the possibility of implementing a solution for continuous online authentication and anomaly detection systems for permanent confirmation of the user's identity within the framework of the RemoteTopology remote championship system developed earlier.

Keywords: *computer mouse, biological data, user, data, authentication, model, analysis, machine learning, decision tree, k-nearest neighbor method, random forest algorithm, ultra-precise neural networks*

References

1. A.R. Abzalov, I.I. Kashapov, A.Yu. Orlov, I.R. Mamleev (2020). User Authentication Based on a Three-Level Keystroke Model. Bulletin of DSTU. *Technical science*. https://cyberleninka.ru/article/n/autentifikatsiya-polzovateley-na-osnove-trehstupenchatoy-modeli-klaviaturnogo-pocherka.

2. A.G. Uimin, S.V. Lyubkin. State registration of the computer program Remote Topology-User interface. Available at: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=202166121 8&TypeFile=html.

3. A.G. Uimin, S.V. Lyubkin. State registration of the computer program Remote Topology-authorization module. https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=202161 9990&TypeFile=html.

4. A.S. Zudinov (2021). Implementation of biometrics in access control systems at critical information infrastructure facilities. *StudNet.* https://cyberleninka.ru/article/n/vnedrenie-biometrii-v-sistemy-kontrolya-dostupa-na-obektah-kriticheskoy-informatsionnoy-infrastruktury.

5. V.E. Kosarev, E.S. Rusilo (2021). Biometrics in banks and factors hindering its development. *Financial markets and banks*. No. 3. https://cyberleninka.ru/article/n/biometriya-v-bankah-i-faktory-sderzhivayuschie-ee-razvitie.

6. A.N. Lindigrin. Analysis of the specifics and problems of the processes of searching for anomalies in network data. *News of TulGU*. https://cyberleninka.ru/article/n/analiz-spetsifiki-i-problematiki-protsessov-poiska-anomaliy-v-setevyh-dannyh.

7. L.D. Lozhkin, K.P. Ankina (2021). Verification of a signature based on gray level information using a neural network. *StudNet.* https://cyberleninka.ru/article/n/proverka-podpisi-na-osnove-informatsii-ob-urovne-serogo-s-ispolzovaniem-neyronnoy-seti.

8. M.A. Polyanichko (2020). Technique for detecting anomalous user interaction with information assets to detect insider activity. *Proceedings of educational institutions of communication*. https://cyberleninka.ru/article/n/metodika-obnaruzheniya-anomalnogo-vzaimodeystviya-polzovateley-s-informatsionnymi-aktivami-dlya-vyyavleniya-insayderskoy.

9. P.A. Savenkov, P.S. Tregubov (2020). Search for behavioral anomalies in the activities of employees using spatial clustering methods based on density. News of TulGU. *Technical science.* https://cyberleninka.ru/article/n/poisk-povedencheskih-anomaliy-v-deyatelnosti-sotrudnikov-pri-pomoschi-metodov-prostranstvennoy-klasterizatsii-osnovannyh-na.

10. S.N. Terekhin, A.V. Vostrykh, A.V. Semenov (2020). Evaluation of graphical user interfaces through a consistent pattern search algorithm. Bulletin of St. Petersburg University of the State Fire Service EMERCOM of Russia. https://cyberleninka.ru/article/n/otsenka-graficheskih-polzovatelskih-interfeysov-posredstvom-algoritma-poiska-posledovatelnyh-shablonov.

11. A.G. Uimin (2021). Development of professional training taking into account the standards of "Worldskills Russia". *Bulletin of Pedagogical Sciences*, no. 1, pp. 42-51.

12. M.D. Fufaev, S.V. Krivoruchko (2021). Biometric identification: the essence and risks of using technology in the payment industry. International. *Journal of the Humanities and Natural Sciences.* No. 2-3. Available at: https://cyberleninka.ru/article/n/biometricheskaya-identifikatsiya-suschnost-i-riski-primeneniya-tehnologii-v-platyozhnoy-industrii.

13. V.V. Chemirisov (2021). Methodology for assessing the efficiency of typical operator actions when entering data. *Military thought.* No. 10. Available at: https://cyberleninka.ru/article/n/metodika-otsenki-operativnosti-tipovyh-deystviy-operatora-pri-vvode-dannyh.

14. A.A.E. Ahmed, I. Traore (2007). A New Biometric Technology Based on Mouse Dynamics. *IEEE Trans. Dependable Secur. Comput*, no. 4, pp. 165-179.

15. A.A.E. Ahmed, I. Traore (2011). Dynamic sample size detection in continuous authentication using sequential sampling. *Proceedings of the 27th Annual Computer Security Applications Conference*, Orlando, FL, USA, 5-9 December 2011, pp. 169-176.

16. M. Antal, E. Egyed-Zsigmond (2019) Intrusion detection using mouse dynamics. *IET Biom*, no. 8, pp. 285-294.

17. D. Chuda, P. Kratky (2014) Usage of computer mouse characteristics for identification in web browsing. *Proceedings of the 15th International Conference on Computer Systems and Technologies-CompSysTech'14,* Ruse, Bulgaria, 27-28 June 2014, pp. 218-225. https://dl.acm.org/doi/10.1145/2659532.2659645

18. D. Chuda, P. Kratky, J. Tvarozek (2015). Mouse Clicks Can Recognize Web Page Visitors! *Proceedings of the 24th International Conference onWorldWide Web-WWW '15 Companion*, Florence, Italy, 18-22 May 2015, pp. 21-22.

19. V.R. da Silva, M.D. Costa-Abreu (2018). An empirical biometric-based study for user identification with different neural networks in the online game League of Legends. *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 8-13 July 2018, pp. 1-6.

20. A. Fulop, L. Kovacs, T. Kurics, E. Windhager-Pokol (2016). Balabit Mouse Dynamics Challenge Data Set. 2016. Available at: https://medium.com/balabit-unsupervised/releasing-the-balabit-mouse-dynamics-challenge-data-set-a15a016fba6c (accessed on 8 May 2021).

21. H. Gamboa, A. Fred (2004). A Behavioural Biometric System Based on Human Computer Interaction. *Proc. SPIE 2004*, 5404, pp. 381-392.

22. N.A. Hamid, S. Safei, S.D.M. Satar, S. Chuprat, R. Ahmad (2011). Mouse movement behavioral biometric systems. *Proceedings of the International Conference on User Science and Engineering (i-USEr),* Selangor, Malaysia, 29 November-1 December 2011, pp. 206-211.

23. S. Hashia, C. Pollett, M. Stamp (2005). On Using Mouse Movements as a Biometric. 5. *Proceedings of the International Conference on Computer Science and its Applications,* Singapore, 9-12 May 2005.

24. M. Pusara, C.E. Brodley (2004). User re-authentication via mouse movements. *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security-VizSEC/DMSEC '04,* Washington, DC, USA, 29 October 2004.

25. Y.X.M. Tan, A. Iacovazzi, I. Homoliak, Y. Elovici, A. Binder (2019). Adversarial attacks on remote user authentication using behavioural mouse dynamics. *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 14-19 July 2019, pp. 1-10.

Information about author:

Anton G. Uymin, *candidate for the degree of Candidate of Technical Sciences. Department of Radio Engineering and Radio Systems. Vladimir State University named after Alexander Grigoryevich and Nikolai Grigoryevich Stoletov (VISU). Research interests: information security, monitoring of information resources, collection and processing of information, Vladimir, Russia*

Ilya M. Morozov, *technical specialist of GBOU MIC. Certified expert of the demonstration exam, certified instructor of the Huawei Academy, with research interests: information security, monitoring of information resources, information collection and processing, Vladimir, Russia*