УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2022 Том 22 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
Jule–August 2022 Vol. 22 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2022-22-4-760-768
УДК 004.056.53

Метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя

Дмитрий Андреевич Есипов[1]✉, Наргиз Асланова[2], Егор Евгеньевич Шабала[3], Даниил Сергеевич Щетинин[4], Илья Юрьевич Попов[5]

[1,2,3,4,5] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] some1else.d.ma@gmail.com✉, https://orcid.org/0000-0003-4467-5117
[2] aslanova.077@gmail.com, https://orcid.org/0000-0003-3650-7412
[3] e.shabala@altdisasm.ru, https://orcid.org/0000-0001-6659-9878
[4] 1nta4r@gmail.com, https://orcid.org/0000-0003-3913-3997
[5] ilyapopov27@gmail.com, iupopov@itmo.ru, https://orcid.org/0000-0002-6407-7934

Аннотация

Предмет исследования. В настоящее время значительный объем атак на информационные системы составляют многоэтапные целевые атаки. Зачастую ключевыми субъектами атаки становятся внутренние нарушители — инсайдеры. Действия инсайдера отличаются от активности легитимного пользователя. Тогда возможно формирование модели поведения пользователя, отличия от которой могут быть классифицированы как события или инциденты информационной безопасности. Существующие подходы к обнаружению аномалий в активности пользователя предполагают использование отдельных характеристик его поведения, без учета их взаимозависимостей и зависимостей от различных факторов. Задача исследования состоит в формировании комплексной характеристики поведения пользователя при использовании компьютера — «цифровой метрики», для обнаружения событий и инцидентов информационной безопасности. **Метод.** Предложен метод обнаружения инцидентов информационной безопасности посредством формирования цифровой метрики пользователя за счет анализа его поведенческих характеристик и их зависимостей, выбранных в качестве предикторов. Разработанный метод предполагает формирование модели посредством машинного обучения без учителя. Рассмотрены алгоритмы: опорных векторов для одного класса, изолирующего леса и эллипсоидальной аппроксимации данных. Основной метрикой качества моделей выбран коэффициент корреляции Мэтьюса, однако были рассмотрены и другие показатели. Выполнен сравнительный анализ моделей, обученных выбранными алгоритмами с различными параметрами по метрикам качества. **Основные результаты.** Выполнен эксперимент с целью получения оценки разработанного метода и сравнения его эффективности с ближайшим аналогом. Для обучения и оценки моделей в рамках исследуемых методов использованы реальные данные о поведении 138 пользователей. По результатам сравнительного анализа, разработанный метод продемонстрировал отличные показатели по всем рассмотренным метрикам, в том числе повышение коэффициента корреляции Мэтьюса на 0,6125. **Практическая значимость.** Разработанный метод может быть использован для непрерывной аутентификации пользователя в средствах защиты информации от несанкционированного доступа и выявления инцидентов информационной безопасности, связанных с действиями инсайдеров.

Ключевые слова

обнаружение и реагирование на угрозы на конечных точках, клавиатурный почерк, цифровая метрика, машинное обучение, обнаружение аномалий, метод опорных векторов, алгоритм изолирующего леса, алгоритм эллипсоидальной аппроксимации данных

Ссылка для цитирования: Есипов Д.А., Асланова Н., Шабала Е.Е., Щетинин Д.С., Попов И.Ю. Метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 760–768. doi: 10.17586/2226-1494-2022-22-4-760-768

A method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics

Dmitry A. Esipov[1]✉, Nargiz Aslanova[2], Egor E. Shabala[3], Daniil S. Shchetinin[4], Ilya Yu. Popov[5]

[1,2,3,4,5] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] some1else.d.ma@gmail.com✉, https://orcid.org/0000-0003-4467-5117
[2] aslanova.077@gmail.com, https://orcid.org/0000-0003-3650-7412
[3] e.shabala@altdisasm.ru, https://orcid.org/0000-0001-6659-9878
[4] 1nta4r@gmail.com, https://orcid.org/0000-0003-3913-3997
[5] ilyapopov27@gmail.com, iupopov@itmo.ru, https://orcid.org/0000-0002-6407-7934

Abstract

Nowadays a significant amount of attacks on information systems are multi-stage attacks. In many cases the key subjects of attacks are insiders. The actions of an insider differ from the activity of a legitimate user, so it is possible for the latter to form a model of his behavior. Then the differences from the specified model can be classified as information security events or incidents. Existing approaches to anomaly detection in user activity use separate characteristics of user behavior, without taking into account their interdependencies and dependencies on various factors. The task of the study is to form a comprehensive characteristic of the user`s behavior when using a computer — a "digital pattern" for detecting information security events and incidents. The essence of the method is in the formation of a digital pattern of the user's activity by analyzing his behavioral characteristics and their dependencies selected as predictors. The developed method involves the formation of a model through unsupervised machine learning. The following algorithms were considered: one-class support vector machine, isolating forest and elliptic envelope. The Matthews correlation coefficient was chosen as the main metric for the quality of the models, but other indicators were also taken into consideration. According to the selected quality metrics, a comparative analysis of algorithms with different parameters was conducted. An experiment was carried out to evaluate the developed method and compare its effectiveness with the closest analogue. Real data on the behavior of 138 users was used to train and evaluate models within the studied methods. According to the results of the comparative analysis, the proposed method showed great performance for all the considered metrics, including an increase in the Matthews correlation coefficient by 0.6125 compared to the anomaly detection method by keystroke dynamics. The proposed method can be used for continuous user authentication from unauthorized access and identifying information security incidents related to the actions of insiders.

Keywords

endpoint detection and response, EDR, keystroke dynamics, digital pattern, machine learning, anomaly detection, support vector machine, SVM, isolation forest, elliptic envelope

For citation: Esipov D.A., Aslanova N., Shabala E.E., Shchetinin D.S., Popov I.Yu. A method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2022, vol. 22, no. 4, pp. 760–768 (in Russian). doi: 10.17586/2226-1494-2022-22-4-760-768

Введение

В связи с ростом эффективности и распространением средств защиты информации, массовые атаки становятся все менее эффективными. Сохраняется тенденция роста сложности кибератак [1–6], кроме того, происходит их эволюция — появление новых, более сложных в обнаружении. Значительную опасность представляют целевые атаки (Advanced Persistent Threat, APT)[1]. APT-атаки отличаются от массовых атак по многим параметрам, ключевыми их отличиями являются: длительность в подготовке и проведении, а также в обнаружении; нацеленность злоумышленника на достижение конкретной цели.

Первый этап противодействия атакам — обнаружение. Согласно данным компании «Gartner», для обнаружения атак наиболее эффективным является использование: Network Traffic Analysis (NTA), Endpoint Detection and Response (EDR) и User and Entity Behavior Analytics (UEBA)[2,3]. Отметим, что UEBA часто служит компонентом других средств защиты информации, в том числе NTA и EDR. Многие специалисты отмечают, что механизмы обнаружения, ориентируемые на конечные устройства, предоставляют больше данных, позволяющих детектировать целевые атаки на наиболее ранних этапах, так как чаще всего точкой входа и конечной целью является конечное устройство[1].

Заметим, что более половины инцидентов безопасности в различных организациях связаны с деятель-

[1] Anti-Malware. Защита от целевых атак. 2021 [Электронный ресурс]. URL: https://www.anti-malware.ru/event/2021/03/30 (дата обращения: 01.02.2022).

[2] BTB Security. Common and best practices for SOCs: results of the 2019 SANS SOC survey. 2019 [Электронный ресурс]. URL: https://www.btbsecurity.com/blog/sans-soc-2019 (дата обращения: 01.02.2022).

[3] SecurityLab.ru by Positive Technologies. Модель зрелости SOC от Gartner. 2019 [Электронный ресурс]. URL: https://www.securitylab.ru/blog/personal/Business_without_danger/346702.php (дата обращения: 01.02.2022).

ностью нелояльных сотрудников [7][1,2]. По данным Ponemon Institute, за 2020 год 62 % инцидентов информационной безопасности связаны с действиями так называемых инсайдеров[3]. При этом ущерб от таких инцидентов превысил 11 млн долларов, причем более 4 млн из них приходится на умышленные преступные действия[1].

Деятельность инсайдера — инцидент информационной безопасности, при котором действия сотрудника организации не соответствуют его нормам обычного поведения. Пользователи, будучи сотрудниками организации, следуют определенному распорядку рабочего дня, характер их работы может иметь некоторые особенности в определенные дни недели, в работе они используют конкретные приложения. Каждый пользователь имеет уникальный клавиатурный почерк [8, 9] и свойственный ему набор привычек. Отметим, что клавиатурный почерк пользователя может меняться в зависимости от времени суток в связи с хронотипом владельца [10].

В совокупности перечисленные особенности формируют «цифровую метрику» пользователя, характеризующую его поведение. Тогда аномалии в поведении могут свидетельствовать о потенциальном инциденте безопасности.

Связанные работы

В работах [3–6] отмечено, что в условиях роста сложности кибератак необходимы методы и системы обнаружения вторжений «нового поколения» (next generation IDS, advanced IDS), использующие контекстную высокоуровневую информацию об объекте защиты. В качестве такой информации рекомендовано использование «образа жизни» (Pattern-of-Life) объекта защиты. «Образ жизни» — совокупность информации о числе пользователей (конечных устройств), использования ими сетевых ресурсов и времени суток.

В [8] рассмотрены сложности и значимость методов непрерывной аутентификации по клавиатурному почерку. Подобные методы способны существенно повысить общий уровень безопасности и на данный момент находятся на стадии формирования. В [9] предложен метод непрерывной аутентификации пользователя по клавиатурному почерку, основанный на комбинации сверточной и рекуррентной нейронных сетей, а в работе [11] — на методах математической статистики.

В [12] отмечена значимость автоматизации решений по защите конечных устройств, в том числе по обнаружению инцидентов информационной безопасности, в первую очередь при помощи методов машинного обучения.

В работах [8, 9, 11] описаны методы непрерывной аутентификации по клавиатурному почерку, однако указанные методы предполагают его использование без учета других компонентов метрики и не учитывают зависимости клавиатурного почерка от различных факторов, что потенциально приводит к снижению эффективности методов. В [3–6] предложено использование контекстной высокоуровневой информации о локальной сети для обнаружения вторжений посредством выявления аномальной сетевой активности, и ее применение в системах обнаружения вторжений уровня сети, что не позволяет обнаруживать инциденты на конечных устройствах.

В настоящей работе предложен метод использования цифровой метрики как высокоуровневой информации об активности пользователя на конечном устройстве. Метод включает анализ клавиатурного почерка и влияющих на него факторов для определения аномальной активности и обнаружения инцидентов информационной безопасности. В результате возможно повысить эффективность обнаружения инцидентов информационной безопасности, обусловленных активностью инсайдеров, а также автоматизировать этот процесс на конечных устройствах.

Предлагаемый метод

Сущность предлагаемого метода состоит в формировании цифровой метрики пользователя за счет анализа его поведенческих характеристик и их зависимостей, выбранных в качестве предикторов для построения эффективной модели. Диаграмма метода в нотации IDEF0 приведена на рис. 1.

В качестве предикторов модели выбраны следующие компоненты цифровой метрики пользователя: время суток; день недели; скорость и динамика нажатия на клавиши клавиатуры и мыши. Ввиду специфики взаимодействия с различными процессами, для каждого используемого пользователем процесса необходима разработка отдельной модели. Тогда осуществляется включение в набор данных имени активного процесса, которому соответствует информация о поведении пользователя.

Для сбора данных в соответствии с указанным набором предикторов выполняется отслеживание следующих событий: переключение активного процесса, перевод клавиш клавиатуры и мыши в верхнее и нижнее положения. Формирование паттерна текущего поведения пользователя должно выполняться до переключения процесса или завершаться по истечении времени формирования паттерна. Для дальнейшего применения собранных данных следует нормализовать все численные предикторы. Блок-схема алгоритма сбора данных представлена на рис. 2.

Отметим, что при работе пользователя возможны случайные переключения активного процесса, что приводит к возникновению выбросов, а именно паттернов с нулевыми значениями численных предикторов, описы-

[1] Ponemon Institute. 2020 Cost of insider threats global report. 2020 [Электронный ресурс]. URL: https://www.exclusive-networks.com/uk/wp-content/uploads/sites/28/2020/12/UK-VR-Proofpoint-Report-2020-Cost-of-Insider-Threats.pdf (дата обращения: 01.02.2022).

[2] IBM Security. The Cost of Insider Threats. 2020 [Электронный ресурс]. Режим доступа: https://www.ibm.com/security/digital-assets/services/cost-of-insider-threats/#/ (дата обращения: 01.02.2022).

[3] Anti-Malware. Защита от целевых атак. 2021 [Электронный ресурс]. URL: https://www.anti-malware.ru/event/2021/03/30 (дата обращения: 01.02.2022).

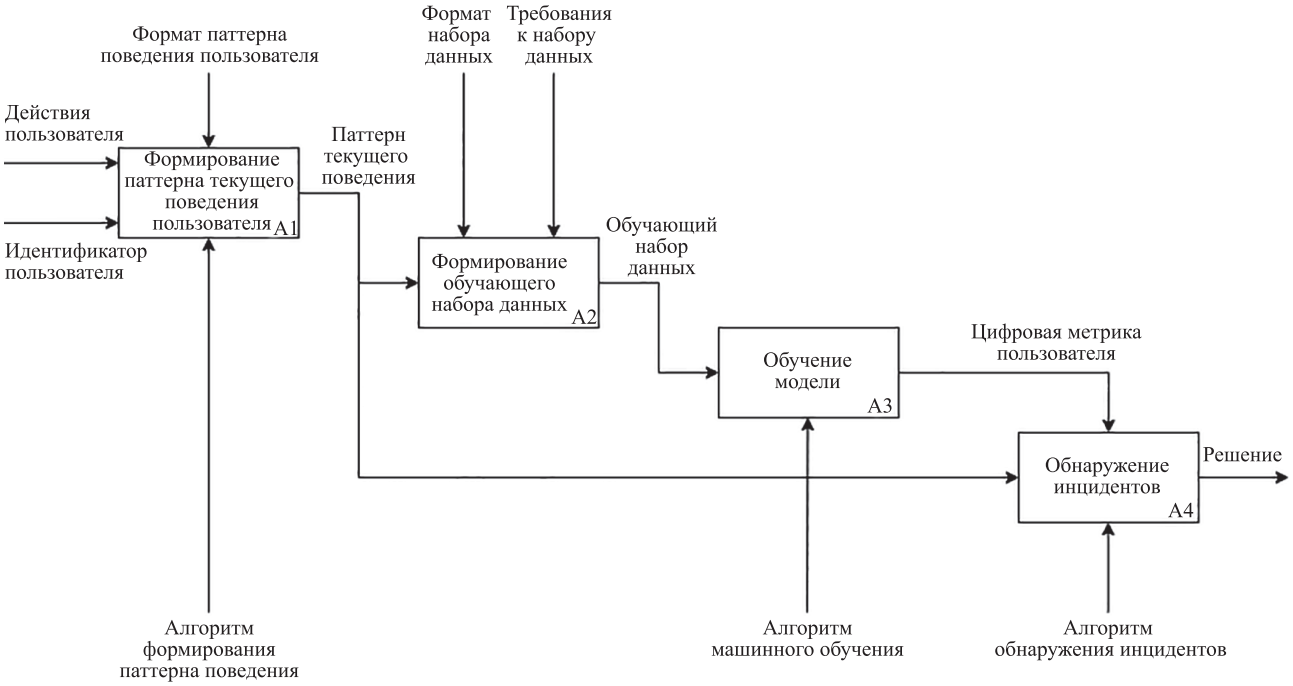

Рис. 1. Диаграмма метода обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя

Fig. 1. Diagram of the method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics

вающих взаимодействие пользователя с контроллерами. Такие паттерны должны быть исключены из обучающего набора данных перед обучением модели.

Метод предполагает обучение модели для каждого используемого пользователем процесса. Для формирования модели предполагается применение алгоритма машинного обучения без учителя, направленного на выявление аномалий [13, 14].

Для обнаружения инцидентов информационной безопасности осуществляется выбор модели в соответствии с именем процесса в текущем паттерне поведения пользователя. Компоненты цифровой метрики, выбранные в качестве предикторов, являются входными параметрами модели для классификации паттерна поведения.

Тогда могут быть обнаружены следующие аномалии в поведении пользователя:

— активность в нетипичные для пользователя время суток и день недели;
— использование нетипичного для пользователя программного обеспечения;
— нетипичный для пользователя активного процесса и времени суток клавиатурный почерк.

Исследуемые методы машинного обучения

Для обучения модели рассмотрены методы машинного обучения: опорных векторов; методы на основах плотности распределения и тензора данных.

Метод опорных векторов включает преимущественно алгоритмы обучения с учителем. Исключениями являются алгоритмы опорных векторов для одного класса (One-Class Support Vector Machine, OCSVM) [15, 16] и для описания данных (Support Vector Data Description, SVDD) [15, 17]. Заметим, что алгоритм SVDD рассчитывает гиперсферу минимального радиуса, внутри которой располагаются объекты, соответствующие норме, снаружи — аномалии [15]. Тогда SVDD можно отнести к методам на основе тензора данных. Алгоритм OCSVM определяет гиперповерхность, разделяющую нормальные и аномальные объекты [15].

Методы машинного обучения на основе плотности распределения включают множество алгоритмов, из которых были рассмотрены наиболее популярные, в том числе алгоритмы *k*-ближайших соседей (*k*-Nearest Neighbors, *k*NN), локальный уровень выброса (Local Outlier Factor, LOF) и изолирующего леса (Isolation Forest). Согласно сравнительному анализу алгоритмов [18], алгоритм Isolation Forest демонстрирует более высокие значения по метрикам качества и достигает высоких показателей при работе с многомерными данными.

Из методов машинного обучения на основе тензора данных [19] рассмотрены алгоритмы SVDD и эллипсоидальной аппроксимации данных (Elliptic Envelope, EE) [20, 21]. Оба алгоритма имеют схожий принцип обнаружения аномалий, однако если алгоритм SVDD рассчитывает гиперсферу, то Elliptic Envelope вычисляет многомерный эллипсоид. Так как сфера — частный случай эллипсоида, то выбран алгоритм Elliptic Envelope.

В итоге, для формирования цифровой метрики пользователя выбраны следующие алгоритмы машинного обучения: OCSVM, Isolation Forest и Elliptic Envelope.

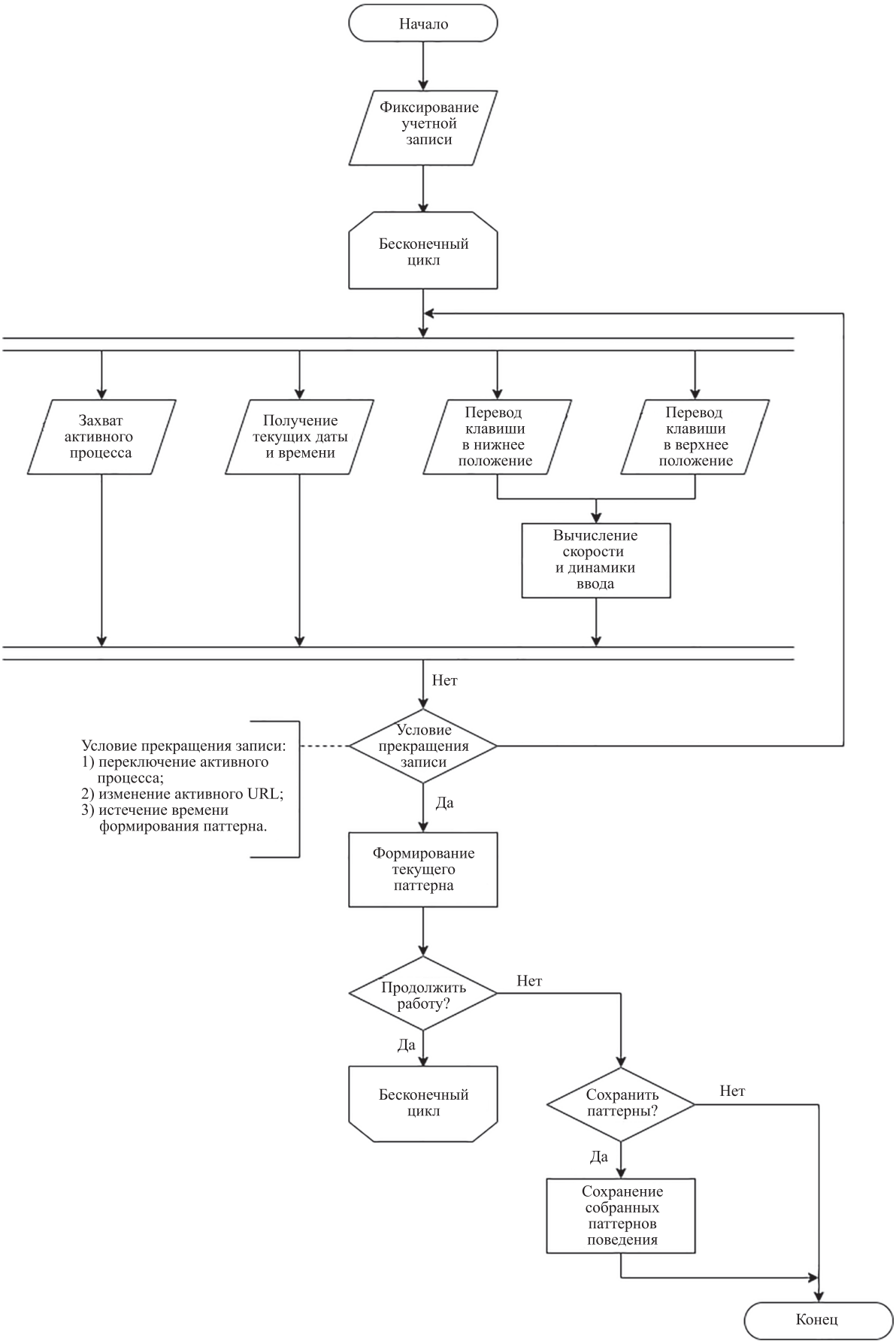

Рис. 2. Блок-схема алгоритма сбора данных

Fig. 2. Data collection algorithm flowchart

Алгоритмы были реализованы на языке программирования Python при помощи библиотеки sklearn [22]. Определение алгоритма и его параметров, демонстрирующих лучший результат, было выполнено в ходе проведения эксперимента.

Выбор метрик качества

Для оценки качества предлагаемого метода выбраны метрики:

— корректность (accuracy) [23];
— точность (precision) [23];
— полнота (recall) [23];
— F-мера (F-score) [23];
— коэффициент корреляции Мэтьюса (Matthews Correlation Coefficient, MCC) [24].

При выборе метрики учтены значения временных характеристик, таких как время обучения и классификация. Рассмотрим метрики подробнее.

Метрика accuracy показывает долю верно классифицированных объектов. Отметим, что accuracy не является показательной при значительно различающимися по мощности классами объектов. Другие метрики лишены подобного недостатка.

Recall определяет возможность идентификации объектов некоторого класса алгоритмом. Precision показывает способность отличать объекты некоторого класса от объектов остальных классов. F-score представляет собой единую метрику качества, учитывающую precision и recall. Однако F-score учитывает метрики precision и recall только для «положительного» класса.

MCC является еще одной метрикой качества двоичной классификации, не подверженной влиянию несбалансированных по мощности классов. При этом указанная метрика дает высокий результат только в том случае, когда классификатор смог корректно предсказать большинство «положительных» и «отрицательных» объектов [24].

Для оценки моделей рассчитаем перечисленные метрики при условии, что в качестве решающей метрики для оценки будет использована метрика MCC.

Формирование набора данных

На момент исследования не существовало открытых наборов данных, содержащих необходимые сведения в соответствии с выбранными параметрами цифровой метрики. Как следствие, в силу специфичности необходимых для построения модели данных необходимо формирование собственного набора данных.

В формировании набора данных принимали участие 138 человек. Для эксперимента выбраны три приложения различных типов. Сбор данных выполнен в соответствии с алгоритмом, приведенным на рис. 2.

Мощность набора данных составила 388 168 патернов (строк) реального поведения пользователя в течении сессии работы. Для проведения эксперимента выполнено разделение набора данных следующим образом: 310 534 (80 %) строк для обучения и 77 634 (20 %) — для тестирования. Заметим, что все строки, соответствующие аномальному поведению, отнесены к тестовому набору данных.

Выбор алгоритма и его параметров

Для определения наиболее эффективной модели выполнено обучение модели посредством применения различных алгоритмов машинного обучения и варьирования их параметров. Наибольшие показатели по метрикам качества моделей по каждому алгоритму представлены в табл. 1.

Согласно табл. 1, лучший результат по выбранным метрикам качества достигает алгоритм OCSVM. Выполним сравнение времени обучения и классификации алгоритмов (табл. 2). Указанные характеристики рассчитаем на 100 итераций указанных операций.

Таблица 1. Сравнительный анализ метрик качества алгоритмов машинного обучения
Table 1. Comparative analysis of quality metrics for machine learning algorithms

Алгоритмы	Параметры	Метрики				
		Accuracy	Precision	Recall	F-score	MCC
OCSVM	функция ядра (kernel): RBF ν (nu): 0,012 γ: 0,197	0,9218	0,9143	0,9951	0,9530	0,7464
Isolation Forest	число деревьев (n_estimators): 66 зашумление (contamination): 0,230	0,7805	0,9092	0,8049	0,8539	0,4321
Elliptic Envelope	зашумление: 0,287	0,7106	0,9011	0,7154	0,7976	0,3400

Таблица 2. Сравнительный анализ временных характеристик алгоритмов
Table 2. Comparative analysis of time characteristics of algorithms

Алгоритм	Время обучения, с	Время классификации, с
OCSVM	22,5131	0,0510
Isolation Forest	95,6583	10,3003
Elliptic Envelope	213,3738	0,1050

Таблица 3. Сравнительный анализ методов
Table 3. Comparative analysis of methods

Параметр		Метод	
		по цифровой метрике	по клавиатурному почерку
Метрики	Accuracy	0,9218	0,5689
	Precision	0,9143	0,8480
	Recall	0,9951	0,5592
	F-score	0,9530	0,6740
	MCC	0,7464	0,1339
Время, с	обучения	22,5131	83,3031
	классификации	0,0510	5,4050

В результате анализа получено, что минимальных значений затраченного времени по обеим операциям достигает алгоритм OCSVM, выбранный по итогу сравнительного анализа для обучения моделей нормального поведения.

Сравнение предложенного метода с обнаружением аномалий по клавиатурному почерку

Для сравнения предложенного метода с методом обнаружения аномалий по клавиатурному почерку проведено обучение моделей для обоих указанных методов. Выбор алгоритма обучения модели и его параметров для обнаружения аномалий по клавиатурному почерку выполнен с использованием тех же алгоритмов машинного обучения и наборов данных для обучения и тестирования. В качестве алгоритма обучения выбран Isolation Forest с числом деревьев, равным 86, и параметром зашумления — 0,469, так как такая модель продемонстрировала наибольшие показатели качества.

Сравнительный анализ методов обнаружения аномалий по цифровой метрике и по клавиатурному почерку по метрикам качества и временным характеристикам приведен в табл. 3.

Согласно табл. 3, метод обнаружения по цифровой метрике демонстрирует более высокие показатели по всем метрикам качества. Причем для метода обнаружения аномалий по клавиатурному почерку наблюдается значительная доля ошибок первого рода — более 37 %. Минимальных значений затраченного времени на указанные операции также достигает метод обнаружения инцидентов по цифровой метрике.

Заключение

Предложенный метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах (цифровой метрике) пользователя достигает коэффициента корреляции Мэтьюса в 0,746363 и показателя F-меры в 0,953008, что превышает аналогичные показатели метода обнаружения аномалий, основанного на клавиатурном почерке. Поскольку при проведении эксперимента были использованы данные, соответствующие активности реальных пользователей при работе с различными приложениями, построенная модель является устойчивой для тестовых и реальных данных.

Метод может быть использован для непрерывной аутентификации пользователя в средствах защиты информации от несанкционированного доступа, что позволит контролировать не только предоставление доступа в начале, но и в течение всего времени поддержания сессии. Указанная модификация позволит обнаруживать факты вмешательства в сессию легитимного пользователя, а также факты неавторизованной передачи доступа, тем самым расширив функционал указанных средств защиты информации.

В дальнейшей работе возможно теоретическое обоснование эффективности предложенного метода, а также исследование применимости других алгоритмов машинного обучения. Возможно расширение цифровой метрики пользователя активностью процессов и сетевых портов, взаимодействием пользователя и процессов с файловой системой и событиями операционной системы, что позволит сформировать цифрового двойника пользователя и сместить фокус с аутентификации на обнаружение инцидентов и определение аномалий в его поведении, а также в работе его автоматизированного рабочего места.

Литература

1. Siddiqi M.A., Mugheri A., Oad K. Advance persistent threat defense techniques: A review // Pakistan Journal of Computer and Information Systems. 2016. V. 1. N 2. P. 53–65.
2. Al-Zewairi M., Almajali S., Ayyash M. Unknown security attack detection using shallow and deep ANN classifiers // Electronics. 2020. V. 9. N 12. P. 2006. https://doi.org/10.3390/electronics9122006
3. Aparicio-Navarro F.J., Kyriakopoulos K.G., Gong Y., Parish D.J., Chambers J.A. Using pattern-of-life as contextual information for anomaly-based intrusion detection systems // IEEE Access.

References

1. Siddiqi M.A., Mugheri A., Oad K. Advance persistent threat defense techniques: A review. *Pakistan Journal of Computer and Information Systems*, 2016, vol. 1, no. 2, pp. 53–65.
2. Al-Zewairi M., Almajali S., Ayyash M. Unknown security attack detection using shallow and deep ANN classifiers. *Electronics*, 2020, vol. 9, no. 12, pp. 2006. https://doi.org/10.3390/electronics9122006
3. Aparicio-Navarro F.J., Kyriakopoulos K.G., Gong Y., Parish D.J., Chambers J.A. Using pattern-of-life as contextual information for anomaly-based intrusion detection systems. *IEEE Access*, 2017,

2017. V. 5. P. 22177–22193. https://doi.org/10.1109/ACCESS.2017.2762162
4. Aparicio-Navarro F.J., Chambers J.A., Kyriakopoulos K., Gong Y., Parish D. Using the pattern-of-life in networks to improve the effectiveness of intrusion detection systems // Proc. of the 2017 IEEE International Conference on Communications (ICC). 2017. P. 7997374. https://doi.org/10.1109/ICC.2017.7997374
5. Aparicio-Navarro F.J., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Chambers J.A. Multi-stage attack detection using contextual information // Proc. of the IEEE Military Communications Conference (MILCOM). 2018. P. 920–925. https://doi.org/10.1109/MILCOM.2018.8599708
6. Aparicio-Navarro F.J., Chadza T.A., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Assadhan B. Addressing multi-stage attacks using expert knowledge and contextual information // Proc. of the 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN). 2019. P. 188–194. https://doi.org/10.1109/ICIN.2019.8685841
7. Budiarto R., Alqarni A.A., Alzahrani M.Y., Pasha M.F., Firdhous M.F.M., Stiawan D. User behavior traffic analysis using a simplified memory-prediction framework // Computers, Materials and Continua. 2022. V. 70. N 2. P. 2679–2698. https://doi.org/10.32604/cmc.2022.019847
8. Quraishi S.J., Bedi S.S. Keystroke dynamics biometrics, a tool for user authentication–review // Proc. of the 7th International Conference on System Modeling and Advancement in Research Trends (SMART). 2018. P. 248–254. https://doi.org/10.1109/SYSMART.2018.8746932
9. Xiaofeng L., Shengfei Z., Shengwei Y. Continuous authentication by free-text keystroke based on CNN plus RNN // Procedia Computer Science. 2019. V. 147. P. 314–318. https://doi.org/10.1016/j.procs.2019.01.270
10. Druijff-van de Woestijne G.B., McConchie H., de Kort Y., Licitra G., Zhang C., Overeem S., Smolders K.C.H.J. Behavioural biometrics: Using smartphone keyboard activity as a proxy for rest–activity patterns // Journal of Sleep Research. 2021. V. 30. N 5. P. e13285. https://doi.org/10.1111/jsr.13285
11. Крутохвостов Д.С., Хиценко В.Е. Парольная и непрерывная аутентификация по клавиатурному почерку средствами математической статистики // Вопросы кибербезопасности. 2017. № 5(24). С. 91–99. https://doi.org/10.21681/2311-3456-2017-5-91-99
12. Sjarif N.N.A., Chuprat S., Mahrin M.N., Ahmad N.A., Senan F.M., Zamani N.A., Saupi A. Endpoint detection and response: Why use machine learning? // Proc. of the 10th International Conference on Information and Communication Technology Convergence (ICTC). 2019. P. 283–288. https://doi.org/10.1109/ICTC46691.2019.8939836
13. Kumar Singh Gautam R., Doegar E.A. An ensemble approach for intrusion detection system using machine learning algorithms // Proc. of the 8th Confluence International Conference on Cloud Computing, Data Science and Engineering. 2018. P. 14–15. https://doi.org/10.1109/CONFLUENCE.2018.8442693
14. Alqudah N., Yaseen Q. Machine learning for traffic analysis: a review // Procedia Computer Science. 2020. V. 170. P. 911–916. https://doi.org/10.1016/j.procs.2020.03.111
15. Lampert C.H. Kernel methods in computer vision // Foundations and Trends in Computer Graphics and Vision. 2009. V. 4. N 3. P. 193–285. http://dx.doi.org/10.1561/0600000027
16. Bounsiar A., Madden M.G. One-class support vector machines revisited // Proc. of the 5th International Conference on Information Science & Applications (ICISA). 2014. P. 6847442. https://doi.org/10.1109/ICISA.2014.6847442
17. Tax D.M.J., Duin R.P.W. Support vector data description // Machine Learning. 2004. V. 54. N 1. P. 45–66. https://doi.org/10.1023/B:MACH.0000008084.60811.49
18. Liu F.T., Ting K.M., Zhou Z.H. Isolation forest // Proc. of the 8th IEEE International Conference on Data Mining (ICDM). 2008. P. 413–422. https://doi.org/10.1109/ICDM.2008.17
19. Ji Y., Wang Q., Li X., Liu J. A survey on tensor techniques and applications in machine learning // IEEE Access. 2019. V. 7. P. 162950–162990. https://doi.org/10.1109/ACCESS.2019.2949814
20. Howard S. The Elliptical Envelope // arXiv. 2007. arXiv:math/0703048. https://doi.org/10.48550/arXiv.math/0703048
21. Ashrafuzzaman M., Das S., Jillepalli A.A., Chakhchoukh Y., Sheldon F.T. Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems // Proc. of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI). 2020. P. 1131–1137. https://doi.org/10.1109/SSCI47803.2020.9308523

vol. 5, pp. 22177–22193. https://doi.org/10.1109/ACCESS.2017.2762162
4. Aparicio-Navarro F.J., Chambers J.A., Kyriakopoulos K., Gong Y., Parish D. Using the pattern-of-life in networks to improve the effectiveness of intrusion detection systems. *Proc. of the 2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 7997374. https://doi.org/10.1109/ICC.2017.7997374
5. Aparicio-Navarro F.J., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Chambers J.A. Multi-stage attack detection using contextual information. *Proc. of the IEEE Military Communications Conference (MILCOM)*, 2018, pp. 920–925. https://doi.org/10.1109/MILCOM.2018.8599708
6. Aparicio-Navarro F.J., Chadza T.A., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Assadhan B. Addressing multi-stage attacks using expert knowledge and contextual information. *Proc. of the 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, 2019, pp. 188–194. https://doi.org/10.1109/ICIN.2019.8685841
7. Budiarto R., Alqarni A.A., Alzahrani M.Y., Pasha M.F., Firdhous M.F.M., Stiawan D. User behavior traffic analysis using a simplified memory-prediction framework. *Computers, Materials and Continua*, 2022, vol. 70, no. 2, pp. 2679–2698. https://doi.org/10.32604/cmc.2022.019847
8. Quraishi S.J., Bedi S.S. Keystroke dynamics biometrics, a tool for user authentication–review. *Proc. of the 7th International Conference on System Modeling and Advancement in Research Trends (SMART)*, 2018, pp. 248–254. https://doi.org/10.1109/SYSMART.2018.8746932
9. Xiaofeng L., Shengfei Z., Shengwei Y. Continuous authentication by free-text keystroke based on CNN plus RNN. *Procedia Computer Science*, 2019, vol. 147, pp. 314–318. https://doi.org/10.1016/j.procs.2019.01.270
10. Druijff-van de Woestijne G.B., McConchie H., de Kort Y., Licitra G., Zhang C., Overeem S., Smolders K.C.H.J. Behavioural biometrics: Using smartphone keyboard activity as a proxy for rest–activity patterns. *Journal of Sleep Research*, 2021, vol. 30, no. 5, pp. e13285. https://doi.org/10.1111/jsr.13285
11. Krutohvostov D., Khitsenko V. Password authentication and continuous authentication by keystroke dynamics using mathematical statistics. *Voprosy kiberbezopasnosti*, no. 5(24), pp. 91–99. (in Russian). https://doi.org/10.21681/2311-3456-2017-5-91-99
12. Sjarif N.N.A., Chuprat S., Mahrin M.N., Ahmad N.A., Senan F.M., Zamani N.A., Saupi A. Endpoint detection and response: Why use machine learning? *Proc. of the 10th International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 283–288. https://doi.org/10.1109/ICTC46691.2019.8939836
13. Kumar Singh Gautam R., Doegar E.A. An ensemble approach for intrusion detection system using machine learning algorithms. *Proc. of the 8th Confluence International Conference on Cloud Computing, Data Science and Engineering*, 2018, pp. 61–64. https://doi.org/10.1109/CONFLUENCE.2018.8442693
14. Alqudah N., Yaseen Q. Machine learning for traffic analysis: a review. *Procedia Computer Science*, 2020, vol. 170, pp. 911–916. https://doi.org/10.1016/j.procs.2020.03.111
15. Lampert C.H. Kernel methods in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 2009, vol. 4, no. 3, pp. 193–285. http://dx.doi.org/10.1561/0600000027
16. Bounsiar A., Madden M.G. One-class support vector machines revisited. *Proc. of the 5th International Conference on Information Science & Applications (ICISA)*, 2014, pp. 6847442. https://doi.org/10.1109/ICISA.2014.6847442
17. Tax D.M.J., Duin R.P.W. Support vector data description. *Machine Learning*, 2004, vol. 54, no. 1, pp. 45–66. https://doi.org/10.1023/B:MACH.0000008084.60811.49
18. Liu F.T., Ting K.M., Zhou Z.H. Isolation forest. *Proc. of the 8th IEEE International Conference on Data Mining (ICDM)*, 2008, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17
19. Ji Y., Wang Q., Li X., Liu J. A survey on tensor techniques and applications in machine learning. *IEEE Access*, 2019, vol. 7, pp. 162950–162990. https://doi.org/10.1109/ACCESS.2019.2949814
20. Howard S. The Elliptical Envelope. *arXiv*, 2007, arXiv:math/0703048. https://doi.org/10.48550/arXiv.math/0703048
21. Ashrafuzzaman M., Das S., Jillepalli A.A., Chakhchoukh Y., Sheldon F.T. Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems. *Proc. of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 1131–1137. https://doi.org/10.1109/SSCI47803.2020.9308523

22. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python // Journal of Machine Learning Research. 2011. V. 12. P. 2825–2830.
23. Saranya T., Sridevi S., Deisy C., Chung T.D., Khane M.K.A.A. Performance analysis of machine learning algorithms in intrusion detection system: A review // Procedia Computer Science. 2020. V. 171. P. 1251–1260. https://doi.org/10.1016/j.procs.2020.04.133
24. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation // BMC Genomics. 2020. V. 21. N 1. P. 1–13. https://doi.org/10.1186/s12864-019-6413-7

22. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, vol. 12, pp. 2825–2830.
23. Saranya T., Sridevi S., Deisy C., Chung T.D., Khane M.K.A.A. Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, 2020, vol. 171, pp. 1251–1260. https://doi.org/10.1016/j.procs.2020.04.133
24. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 2020, vol. 21, no. 1, pp. 1–13. https://doi.org/10.1186/s12864-019-6413-7

Авторы

Есипов Дмитрий Андреевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0003-4467-5117, some1else.d.ma@gmail.com
Асланова Наргиз — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0003-3650-7412, aslanova.077@gmail.com
Шабала Егор Евгеньевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0001-6659-9878, e.shabala@altdisasm.ru
Щетинин Даниил Сергеевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0003-3913-3997, 1nta4r@gmail.com
Попов Илья Юрьевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57202195632, https://orcid.org/0000-0002-6407-7934, ilyapopov27@gmail.com, iupopov@itmo.ru

Authors

Dmitry A. Esipov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0003-4467-5117, some1else.d.ma@gmail.com
Nargiz Aslanova — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0003-3650-7412, aslanova.077@gmail.com
Egor E. Shabala — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0001-6659-9878, e.shabala@altdisasm.ru
Daniil S. Shchetinin — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0003-3913-3997, 1nta4r@gmail.com
Ilya Yu. Popov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57202195632, https://orcid.org/0000-0002-6407-7934, ilyapopov27@gmail.com, iupopov@itmo.ru

Статья поступила в редакцию 03.02.2022
Одобрена после рецензирования 06.06.2022
Принята к печати 20.07.2022

Received 03.02.2022
Approved after reviewing 06.06.2022
Accepted 20.07.2022