НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2025 Том 25 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
Jule–August 2025 Vol. 25 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2025-25-4-663-675
УДК 004.056.5

Применение машинного обучения для профилирования устройств Интернета вещей с целью обнаружения вредоносной активности

Даниил Михайлович Легкодымов[1], Дмитрий Сергеевич Левшун[2], Игорь Витальевич Котенко[3]✉

[1] Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
[2,3] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] danillegk65@gmail.com, https://orcid.org/0009-0002-2874-6632
[2] levshun.d@iias.spb.su, https://orcid.org/0000-0003-1898-6624
[3] ivkote@comsec.spb.ru✉, https://orcid.org/0000-0001-6859-7120

Аннотация

Введение. Защита устройств Интернета вещей представляет собой актуальную и значимую задачу в условиях постоянного увеличения числа подключенных к сети устройств и нарастающей угрозы кибератак. Одним из ключевых решений данной проблемы является профилирование таких устройств с целью повышения уровня безопасности систем, в которых они функционируют. Применение методов машинного обучения является перспективным подходом к решению подобной задачи. В настоящем исследовании представлен подход к профилированию устройств Интернета вещей, направленный на обнаружение вредоносной активности. Представленное решение позволяет обнаруживать сетевые события, которые могут свидетельствовать о наличии кибератак. **Метод.** Сущность метода заключается в создании индивидуальных профилей поведения для каждого устройства Интернета вещей с использованием алгоритмов машинного обучения. Профили создаются на основе анализа сетевого трафика. Модели машинного обучения выполняют задачи классификации и обнаружения аномалий. В работе подробно описаны основные этапы предложенного подхода, которые включают процессы сбора и предварительной обработки данных, выбора и обучения моделей, тестирования и оценки эффективности разработанного решения. **Основные результаты.** В ходе исследования построено 26 профилей устройств на основе набора данных CIC IoT 2022. В исходный набор данных добавлен 21 новый признак. Обновленный набор сбалансирован методами оверсемплинга и андерсемплинга. Для каждого устройства получены сравнительные оценки эффективности моделей Random Forest, XGBoost, CatBoost для задачи обнаружения атак, а также Isolation Forest, Elliptic Envelope, One-Class Support Vector Machine для задачи обнаружения аномалий. Показано, что предложенные в исследовании новые признаки входят в число наиболее информативных. **Обсуждение.** Сравнение полученных результатов с релевантными исследованиями подтвердило применимость предложенного подхода для обеспечения безопасности устройств Интернета вещей и снижения рисков, связанных с их эксплуатацией.

Ключевые слова

поведенческое профилирование устройств, безопасность Интернета вещей, информационная безопасность, алгоритмы машинного обучения, выявление аномальной активности, детектирование сетевых атак

Благодарности

Работа выполнена при поддержке гранта Российского научного фонда № 24-71-10095, https://rscf.ru/project/24-71-10095/.

Ссылка для цитирования: Легкодымов Д.М., Левшун Д.С., Котенко И.В. Применение машинного обучения для профилирования устройств Интернета вещей с целью обнаружения вредоносной активности // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 4. С. 663–675. doi: 10.17586/2226-1494-2025-25-4-663-675

Leveraging machine learning for profiling IoT devices to identify malicious activities

Daniil M. Legkodymov[1], Dmitry S. Levshun[2], Igor V. Kotenko[3]✉

[1] The Bonch-Bruevich Saint Petersburg State University of Telecommunications (SPbSUT), Saint Petersburg, 193232, Russian Federation

[2,3] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation

[1] danillegk65@gmail.com, https://orcid.org/0009-0002-2874-6632

[2] levshun.d@iias.spb.su, https://orcid.org/0000-0003-1898-6624

[3] ivkote@comsec.spb.ru✉, https://orcid.org/0000-0001-6859-7120

Abstract

Protecting IoT devices is a relevant and important task in the context of a constantly increasing number of devices connected to the network and a growing threat of cyberattacks. One of the key solutions to this problem is profiling such devices to increase the security level of the systems in which they operate. The application of machine learning methods represents a promising approach to solving this problem. This study presents a method for profiling Internet of Things (IoT) devices aimed at detecting malicious activity. The proposed solution enables the identification of network events that may indicate the presence of cyberattacks. The essence of the method lies in the creation of individualized behavioral profiles for each IoT device using machine learning algorithms. Profiles are constructed based on the analysis of network traffic. The machine learning models are employed to perform classification and anomaly detection tasks. The study provides a detailed description of the main stages of the proposed approach, including data collection and preprocessing, model selection and training, testing, and evaluation of the effectiveness of the developed solution. In the course of the study, 26 device profiles were constructed using the CIC IoT 2022 dataset. An additional 21 new features were incorporated into the original dataset. The augmented dataset was balanced using oversampling and undersampling techniques. For each device comparative performance evaluations were conducted for Random Forest, XGBoost, and CatBoost models in the context of attack detection as well as for Isolation Forest, Elliptic Envelope, and One-Class Support Vector Machine for anomaly detection. It was demonstrated that the newly proposed features are among the most informative. A comparison of the obtained results with relevant studies confirmed the applicability of the proposed approach for ensuring the security of IoT devices and reducing the risks associated with their operation.

Keywords

device behavioral profiling, IoT security, information security, artificial intelligence, machine learning algorithms, abnormal activity detection, network attack detection

Acknowledgements

The study was supported by the grant of the Russian Science Foundation No. 24-71-10095, https://rscf.ru/en/project/24-71-10095/.

For citation: Legkodymov D.M., Levshun D.S., Kotenko I.V. Leveraging machine learning for profiling IoT devices to identify malicious activities. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 4, pp. 663–675 (in Russian). doi: 10.17586/2226-1494-2025-25-4-663-675

Введение

Стремительное развитие сети Интернета вещей (IoT) характеризуется неуклонным увеличением количества подключенных устройств, что открывает существенные перспективы для повышения уровня комфорта и эффективности в разнообразных сферах человеческой деятельности. Согласно экспертным оценкам, к 2025 году их число может превысить 37 млрд[1]. Вместе с тем данная тенденция сопряжена с возрастанием рисков кибербезопасности, поскольку устройства IoT обладают повышенной уязвимостью к атакам. Это обусловлено такими факторами, как ограниченность их вычислительных и энергетических ресурсов, значительная гетерогенность аппаратных и программных платформ, а также сложности в обеспечении своевременного обновления встроенного программного обеспечения [1, 2]. Следовательно, обеспечение адекватного уровня защищенности требует разработки инновационных решений, учитывающих специфические характеристики данных устройств [3, 4]. Применение методов машинного обучения (Machine Learning, ML) для создания систем профилирования выступает одним из перспективных подходов в рассматриваемой области [5].

Профилирование в контексте данной работы — процесс отбора и предобработки трафика IoT-устройств для создания характерного поведенческого сетевого профиля, который используется для обнаружения аномалий и атак с помощью ML-моделей.

Основным недостатком существующих решений в области профилирования IoT-устройств является их сосредоточенность на задаче идентификации типа устройств, в то время как задаче обнаружения вредоносной активности уделяется недостаточно внимания.

Цель работы заключается в разработке и исследовании полноценного подхода к поведенческому профилированию IoT-устройств с использованием методов ML. Подход направлен на эффективное обнаружение вредоносной активности. Для достижения этой цели решается задача создания индивидуальных профилей поведения для различных типов устройств IoT на основе анализа их сетевого трафика, задача разработки и внедрения новых извлекаемых из сетевого трафика

[1] [Электронный ресурс]. Режим доступа: https://www.ericsson.com/en/reports-and-papers/mobility-report/mobility-visualizer, свободный. Яз. англ. (дата обращения: 10.03.2025).

признаков, задача обучения и сравнительной оценки производительности моделей ML с целью классификации и обнаружения аномалий, а также задача определения наиболее значимых признаков для каждого профиля устройства, в том числе оценка вклада предложенных новых признаков.

Ожидается, что полученные результаты могут быть использованы для повышения безопасности информационных систем с IoT-устройствами. В свою очередь, это позволит снизить риски, связанные с киберугрозами, что определяет практическую значимость данной работы.

Анализ существующих работ

Согласно анализу научных работ, основные направления исследований в области профилирования и безопасности IoT-устройств включают: применение ML-методов для обнаружения угроз безопасности [6, 7]; профилирование IoT-устройств в режиме реального времени [6]; улучшение методов аутентификации и контроля доступа в IoT-системах [7]; обеспечение конфиденциальности данных IoT-устройств [8]; применение технологии блокчейн [6]; защита от распределенных атак типа «отказ в обслуживании» и ботнетов [9]; повышение безопасности протоколов связи и внедрение новых [10].

Как правило, при профилировании IoT-устройств исследователи ставят перед собой целый ряд задач. *Идентификация типа устройства* важна для применения настроек безопасности, соответствующих устройству, например, разные настройки необходимы для камер и датчика температуры. *Идентификация экземпляра устройства* позволяет различать экземпляры устройств для возможности применения отдельных механизмов безопасности к каждому конкретному устройству. *Обнаружение новых устройств* выявляет новые устройства, для которых еще не построен поведенческий профиль. *Обнаружение аномального поведения* направлено на применение поведенческих профилей для обнаружения подозрительной активности устройств. *Обнаружение атак* связано с применением поведенческих профилей устройств для обнаружения известной вредоносной активности. При этом в области профилирования и обеспечения безопасности IoT-устройств существует ряд сложностей.

— *Разнообразие устройств*. Данное обстоятельство затрудняет разработку универсальных методов защиты. Учитывая огромное количество IoT-устройств, затруднительно подобрать универсальные наборы признаков для их профилирования [5, 6, 11].
— *Обновление программного обеспечения*. IoT-устройства от небольших производителей зачастую имеют проблемы с обновлением программного обеспечения, что влечет за собой наличие уязвимостей, которые никогда не будут исправлены [6].
— *Обучающие данные*. Сбор обучающих данных для профилирования необходимо осуществлять на протяжении значимого промежутка времени. Это создает дополнительные ограничения для сетей, в которых часто появляются новые IoT-устройства [5].
— *Динамическое поведение устройств*. Обновления IoT-устройств могут привести к потере актуальности их профиля [5, 6].
— *Высокая степень взаимосвязанности*. IoT-устройства взаимодействуют друг с другом и другими системами, создавая высокую степень взаимозависимости [11].
— *Ограниченные возможности защиты*. IoT-устройства часто не имеют базовых средств защиты, таких как межсетевые экраны или системы обнаружения вторжений. Это связано с ограниченностью их ресурсов и особенностями использования таких устройств [5, 6, 11].

Рассмотрим некоторые работы в области обеспечения защиты и профилирования IoT-устройств более подробно.

В [12] разработана система обнаружения вторжений в сетях IoT с использованием ML. Система основана на профилировании IoT-устройств без применения ML и последующей классификации атак с применением данного метода. Разработанная система достигла оценки метрик аккуратности (Accuracy) в 98,35 % и точности (Precision) — 99,31 %.

Многомерный набор данных для профилирования IoT-устройств представлен в работе [13]. Особое внимание авторов уделено созданию ML-модели для определения типа устройства. Эксперименты показали способность ML-моделей классифицировать трафик при решении данной задачи с Accuracy до 98,70 %.

Работа [14] посвящена исследованию применения ML-методов для классификации сетевого трафика. Всего выполнен анализ и сравнение 18 ML-алгоритмов. При этом проведено обучение трех типов ML-моделей, которые определяют: является ли устройство IoT-устройством; тип устройства; экземпляр устройства. Экспериментальная оценка показала преимущество алгоритмов Random Forest (RF) и Decision Tree, Accuracy которых при решении данных задач достигает 99,60 %.

В [15] приведено сравнение системы обнаружения вторжений, основанной на ML-методах, с системой, основанной на сигнатурах. В качестве основного преимущества первой системы выделяется возможность обнаружения ранее неизвестных атак, а также возможность работы на зашифрованном трафике. При этом в работе делается вывод о целесообразности построения гибридных систем обнаружения, сочетающих в себе возможности сигнатурного и эвристических методов анализа.

Подход к идентификации устройств в IoT-сетях на основе методов обучения без учителя предложен в работе [16]. Экспериментальная оценка показала, что предлагаемый подход способен классифицировать IoT-устройства с Accuracy до 96,50 %. В качестве преимуществ подхода выделяются большая гибкость и возможность применения для динамических сетей, где размеченные данные могут быть недоступны.

Анализ работ показал, что профилирование и обеспечение безопасности IoT-устройств — сложная и многоуровневая задача, требующая учета множества факторов. Современные подходы, использующие ML, демонстрируют высокую эффективность. Однако в

области безопасности IoT на основе ML все еще существуют направления, требующие дальнейших исследований и разработок.

Предлагаемый подход

Рассмотрим подход к профилированию IoT-устройств. Предложенный подход охватывает полный цикл работы с данными и моделями ML для решения задач обнаружения аномалий и классификации трафика, начиная со сбора и предобработки подготовки исходных данных, переходя к этапам обучения и последующей валидации ML-моделей. Схема подхода показана на рис. 1, где представлены форматы передаваемых между этапами данных. Между этапами передаются данные в формате библиотеки Packet Capture (PCAP) и Comma Separated Values (CSV).

Этап 1. *Сбор данных*. На данном этапе осуществляется сбор сетевого трафика IoT-устройств в формате библиотеки Packet Capture (PCAP). Такие данные представляют собой детализированное описание сетевого взаимодействия анализируемых устройств. Задачей этапа является запись всех пакетов данных, передаваемых через сеть, что позволяет получить полный набор данных о сетевой активности IoT-устройств.

Этап 2. *Предобработка данных*. Данный этап состоит из четырех основных шагов — извлечение признаков, расстановка меток, форматирование данных, а также оверсемплинг и андерсемплинг.

Шаг 2.1. *Извлечение признаков*. Различные признаки сетевого трафика извлекаются из PCAP-файлов. С полным списком признаков, использованных в процессе экспериментов можно ознакомиться в таблице[1]. Были использованы признаки, приведенные в наборе данных CIC IoT 2022, а также 21 дополнительный признак. Признаки, добавленные в рамках настоящего исследования, представлены в табл. 1.

Шаг 2.2. *Расстановка меток*. Каждому экземпляру трафика присваивается метка, указывающая на имя устройства и характер трафика (легитимный или вредоносный). Эти метки необходимы для обучения ML-моделей и оценки их эффективности. Информация для разметки берется из исходного набора данных — CIC IoT 2022.

Шаг 2.3. *Форматирование данных*. Извлеченные признаки и их метки агрегируются и сохраняются в CSV-файл. Каждая запись итогового файла содержит полный набор признаков, метку названия устройства и типа его трафика. Такой формат удобен для последующего использования в ML-алгоритмах. Пример строки из полученного набора данных можно увидеть на сайте[2].

Шаг 2.4. *Оверсемплинг и андерсемплинг.* Соответствующие методы используются для балансировки представленности классов трафика в наборе данных. Для выбранного набора данных было решено увеличить количество примеров данных в наименее представленных классах до 20 000 с помощью метода Adaptive Synthetic Sampling (ADASYN). В свою очередь, количество примеров данных в других классах уменьшалось с помощью метода Random Under Sampler по следующим правилам: если количество примеров превышало 400 000, то оно уменьшалось до 400 000; если более 200 000 — до 200 000; если более 100 000 — до 100 000; если более 75 000 — до 75 000; если более 40 000 — до 40 000; классы с количеством примеров от 20 000 до 40 000 оставались без изменений. Такой подход позволяет сохранить природу данных, где одни классы трафика представлены больше, чем другие, и в то же время сократить разрыв между наименее и наиболее представленными классами. В свою очередь, это дает возможность ML-моделям более эффективно идентифицировать и обнаруживать все классы трафика.

Этап 3. *Подготовка моделей*. На данном этапе ML-модели обучаются на извлеченных признаках. Обучение реализуется через ключевые шаги: разделение данных, кросс-валидацию, оптимизацию гиперпараметров и финальное обучение моделей.

Шаг 3.1. *Разделение данных*. Из CSV-файлов, созданных на этапе 2, извлекаются данные. При этом в предлагаемом подходе для каждого устройства создается отдельный набор данных, что позволяет учитывать уникальные характеристики трафика каждого IoT-устройства в отдельности.

Шаг 3.2. *Кросс-валидация*. Данные разделяются в соотношении 60/20/20: 60 % данных использовались для обучения, 20 % для валидации и 20 % для финального тестирования моделей. Тестовые данные не известны моделям, а обучающие и валидационные

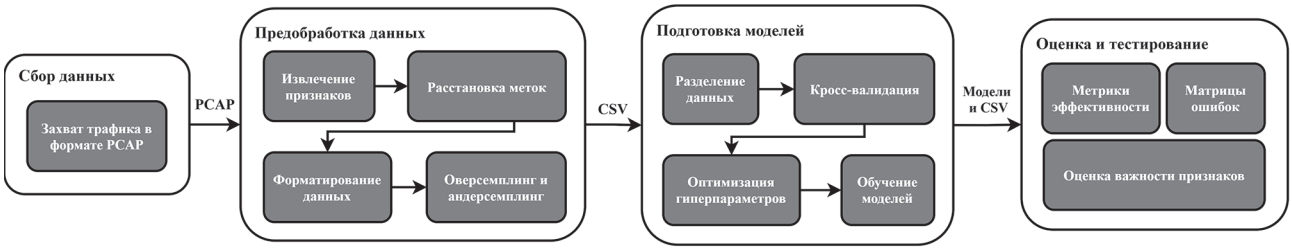

Рис. 1. Подход к обучению ML-моделей для профилирования IoT-устройств

Fig. 1. Approach for training of ML models for profiling IoT devices

[1] [Электронный ресурс]. Режим доступа: https://github.com/levshun/IoT_profiling/blob/main/Listing.%20Example%20of%20the%20dataset%20row.pdf/ (дата обращения: 20.03.2025).

[2] [Электронный ресурс]. Режим доступа: https://github.com/levshun/IoT_profiling/blob/main/Table%201.%20List%20of%20features.pdf (дата обращения: 20.03.2025).

Таблица 1. Список добавленных признаков
Table 1. List of the added features

Признак	Описание	Признак	Описание
unique_ip_dst_count	Количество уникальных адресов IP (Internet Protocol) за последние 20 пакетов (как назначения, так и источника)	is_icmp	Является ли пакет ICMP (1 или 0)
unique_ip_src_count	Количество уникальных IP за последние 20 пакетов (как назначения, так и источника)	icmp_type	Тип ICMP
L3_ip	Является ли IP-пакетом (1 или 0)	is_eapol	Является ли пакет EAPOL (1 или 0)
packet_rate	Количество пакетов в секунду	eapol_type	Тип ICMP
number_of_servers	Подсчет количества уникальных IP-адресов серверов для 20 ближайших пакетов	wifi	Тип EAPOL
tcp_window	Размер буфера	wifi_sub_type	Тип Wi-FI
tcp_data_offset	Смещение данных	zigbee	Является ли пакет ZigBee (1 или 0)
NTP (Network Time Protocol)	Является ли пакет NTP (1 или 0)	zigbee_type	Тип ZigBee
ntp_interval	Является ли пакет NTP текущим (1 или 0)	tcp_payload_size	Полезная нагрузка TCP
Domain Name System (DNS)	Является ли пакет DNS (1 или 0)	total_length	Размер всего IP-пакета, включая заголовок и данные
dns_interval	Является ли пакет DNS текущим (1 или 0)		

данные используются для оптимизации их гиперпараметров. Такой подход обеспечивает надежную оценку эффективности моделей, а также проверяет однородность данных.

Шаг 3.3. *Оптимизация гиперпараметров*. Параметры ML-моделей, такие как количество деревьев в RF или выбросов в Isolation Forest (IF), выбираются с помощью метода Random Search.

Шаг 3.4. *Обучение моделей*. ML-модели обучаются с наилучшими значениями гиперпараметров, полученными на шаге 3.3. Для каждого устройства создаются модели, решающие следующие задачи:

— *обнаружение аномалий*. Модель обучается только на легитимных данных, чтобы выявить аномальное поведение устройства на основе его отличий от нормальной активности. Для этой задачи тестировались такие методы, как IF, Elliptic Envelope (EE) и One-Class Support Vector Machine (1-SVM);
— *обнаружение атак*. Модели обучаются на размеченных данных, чтобы идентифицировать известные сценарии нормального и аномального поведения устройств. Для этой задачи тестировались RF, XGBoost (XGB) и CatBoost (CB).

Этап 4. *Оценка и тестирование*. Accuracy, полнота (Recall), Precision и F-мера (F-score) — основные метрики, используемые при оценке эффективности моделей.

Для задачи *обнаружения атак* это означает правильную или неправильную классификацию сетевых событий. В случае задачи обнаружения аномалий, True Positive и False Positive представляют собой корректное и некорректное обнаружение аномального поведения.

Кроме того, для выявления сценариев поведения, в которых ML-модели не справляются с поставленными задачами, в рамках предлагаемого подхода используются расширенные отчеты классификации (Classification Report) и матрицы ошибок (Confusion Matrix). Подобные отчеты и матрицы предоставляют расширенную информацию о производительности ML-моделей на каждом отдельном классе трафика.

Также для объяснения решений ML-моделей используется метод Local Interpretable Model-agnostic Explanations (LIME). Данный метод позволяет определить признаки сетевого трафика, оказывающие наибольшее влияние на отнесение сетевых событий к тому или иному классу, и решает проблему интерпретируемости ML-моделей.

Экспериментальная оценка

Приведем дополнительные сведения об используемом наборе данных, полученных экспериментальных результатах и их анализе.

Набор данных. В данной работе для профилирования IoT-устройств и обнаружения вредоносной активности была использована улучшенная версия набора данных CIC IoT 2022. Этот набор данных включает в себя PCAP-файлы, содержащие как записи трафика атак на устройства, так и сценарии их легитимной работы. Отметим, что перед проведением экспериментов данные были сбалансированы, так как исходный набор данных имеет значительный перевес вредоносного трафика (рис. 2). С полным описанием используемого набора данных можно ознакомиться на сайте[1].

[1] [Электронный ресурс]. Режим доступа: https://github.com/levshun/IoT_profiling/blob/main/Table%202.%20Description%20of%20the%20dataset.pdf (дата обращения: 20.03.2025).

Рис. 2. Соотношение примеров легитимного и вредоносного трафика

Fig. 2. Ratio of normal and abnormal traffic examples in the dataset before and after balancing

План эксперимента. Для каждого IoT-устройства была произведена оценка производительности 6 ML-моделей для решения следующих задач: *обнаружение атак* — RF, XGB и CB; *обнаружение аномалий* — IF, EE и 1-SVM.

Гиперпараметры ML-моделей были оптимизированы с помощью метода случайного поиска (Random Search). Диапазоны значений проанализированных гиперпараметров представлены в табл. 2.

В процессе экспериментов данные разделялись в соотношении 60/20/20. Это означает, что 60 % данных использовалось для обучения ML-моделей, 20 % для оптимизации их гиперпараметров и валидации, а оставшиеся 20 % не были известны моделям и использовались только для тестирования. Более подробная информация представлена в разделе «Предлагаемый подход».

Результаты. Результаты проведенных экспериментов для пяти устройств представлены в табл. 3. С полным обзором результатов экспериментов для всех 26 устройств можно ознакомиться на сайте[1]. Отметим, что задача обнаружения аномалий решалась только для IoT-устройств с примерами вредоносного трафика.

Значения метрик производительности приведены для ML-моделей с оптимальными значениями гиперпараметров. При этом важно отметить, что каждая из исследованных моделей показала наилучшую производительность по крайней мере для одной задачи одного IoT-устройства: *обнаружение атак*: RF — 8, CB — 14, XGB — 4; обнаружение аномалий: IF — 4, EE — 6, 1-SVM — 1.

В целом, разработанные ML-модели продемонстрировали значительный потенциал, хотя для их адаптации к индустриальным требованиям необходимы дальнейшие усовершенствования и тестирование. Это особенно актуально для задачи *обнаружения аномалий*, где достижение лучших результатов возможно на основе применения моделей глубокого обучения (Deep Learning, DL). По этой причине в рамках дальнейших исследований планируется оценить эффективность DL для той же задачи. При этом основное внимание планируется обратить на тех моделях, которые наиболее эффективны в прогнозировании сетевых событий и обнаружении аномалий в трафике.

[1] [Электронный ресурс]. Режим доступа: https://github.com/levshun/IoT_profiling/blob/main/Table%204.%20Results%20of%20experiments.pdf (дата обращения: 10.03.2025).

Таблица 2. Анализируемые значения гиперпараметров

Table 2. Analyzed values of the models hyperparameters

Модель	Параметр	Значения	Описание
IF	n_estimators max_features	100, 200, 300, 400, 500 1,0; 0,9; 0,8; 0,7; 0,6; 0,5	Количество деревьев решений, образующих лес. Количество признаков, выбираемых из X для обучения каждого разбиения дерева
EE	support_fraction	None; 0,1; 0,3; 0,5; 0,7; 0,9	Доля точек, включаемых в поддержку оценки Minimum Covariance Determinant
	contamination	0,1; 0,2; 0,3; 0,4; 0,5	Доля выбросов в наборе данных
1-SVM	kernel	linear, poly, rbf, sigmoid	Тип ядра, используемого в алгоритме
	gamma nu	scale, auto 0,1; 0,2; 0,3; 0,4; 0,5	Коэффициент ядра для rbf, poly и sigmoid. Верхняя граница доли ошибок обучения и нижняя граница доли опорных векторов
RF	n_estimators	100, 200, 300, 400, 500	Количество деревьев, образующих лес
	criterion max_features	gini, entropy, log_loss sqrt, log2	Функция для измерения качества разбиения. Количество признаков, рассматриваемых при поиске лучшего разбиения
CB	iterations	1000, 1500, 2000, 2500, 3000	Максимальное количество деревьев
	learning_rate	0,001; 0,03; 0,1	Скорость обучения
	grow_policy	SymmetricTree, Lossguide	Способ построения деревьев
XGB	n_estimators	100, 200, 300, 400, 500	Максимальное количество деревьев
	learning_rate	0,1; 0,01; 0,001	Скорость обучения
	booster	gbtree, gblinear	Тип бустера

Таблица 3. Результаты экспериментальной оценки
Table 3. Results of the experiments

Устройство	Количество сценариев	Модели	Метрики			
			Accuracy	Precision	Recall	F-score
Amazon Echo Dot	8	RF	0,985	0,985	0,985	0,985
		CB	0,981	0,981	0,981	0,981
		XGB	0,986	0,986	0,986	0,986
Amazon Echo Studio	8	**RF**	0,991	0,991	0,991	0,991
		CB	0,990	0,990	0,990	0,990
		XGB	0,989	0,989	0,989	0,989
ArloQ Camera	9	RF	0,999	0,999	0,999	0,999
		CB	0,999	0,999	0,999	0,999
		XGB	0,999	0,999	0,999	0,999
	2	IF	0,977	0,978	0,977	0,976
		EE	0,704	0,892	0,704	0,753
		1-SVM	0,988	0,988	0,988	0,988
Atomi Coffee Maker	13	**RF**	0,994	0,994	0,994	0,994
		CB	0,993	0,993	0,993	0,993
		XGB	0,993	0,993	0,993	0,993
	2	IF	0,960	0,962	0,960	0,959
		EE	0,976	0,977	0,976	0,976
		1-SVM	0,673	0,821	0,673	0,696
DLink Camera	3	RF	1,000	1,000	1,000	1,000
		CB	1,000	1,000	1,000	1,000
		XGB	1,000	1,000	1,000	1,000

Примечание. Полужирным шрифтом выделены лучшие ML-модели.

Анализ результатов

Для более детального изучения полученных результатов рассмотрим одно из устройств — Philips Hue Bridge. Для данного устройства всего представлено 11 сценариев, два из которых являются вредоносными, а остальные представляют собой легитимную активность. Набор данных для Philips Hue Bridge содержит 980 000 сетевых событий.

Отметим, что подробно рассмотрим результаты только для тех ML-моделей, которые показали лучшие результаты. Согласно табл. 3, для Philips Hue Bridge это CB в задаче обнаружения атак и EE в задаче обнаружения аномалий. Результаты, полученные для каждого класса трафика Philips Hue Bridge, представлены в табл. 4. Матрица ошибок для детектора атак Philips Hue Bridge показана на рис. 3.

Рис. 3 и табл. 4 показывают, что каждый класс трафика устройства идентифицируется эффективно — наименьшее значение F-score показателя составляет 0,997 для сценария легитимной активности ALEXAON. Более того, False Positive и False Negative встречаются только для сценариев нормального трафика, а вредоносных событий, которые были неверно интерпретированы, нет.

Результаты анализа важности признаков представлены на рис. 4. Важность признаков выражена в безразмерных единицах, полученных путем нормализации, что позволяет сравнивать вклад различных признаков. Отметим, что среди 10 лучших признаков присутствуют tcp_window, DNS, packet_rate и NTP, которые были добавлены в данном исследовании при улучшении набора данных CIC IoT 2022.

В табл. 5 представлены результаты обнаружения аномалий для модели EE. Данная ML-модель обучалась только на легитимном трафике IoT-устройства Philips Hue Bridge. Матрица ошибок для задачи обнаружения аномалий показана на рис. 5.

Табл. 5 и рис. 5 показывают, что одно легитимное сетевое событие идентифицируется как вредоносное, а для вредоносных данных 3655 событий были ошибочно определены как легитимные.

Результаты анализа важности признаков с помощью LIME для задачи обнаружения аномалий представлены на рис. 6.

Среди 10 наиболее значимых признаков присутствуют такие, как *total_length*, *DNS*, *NTP*, *dns_interval*, *ntp_interval* и *icmp_type*. Они также были добавлены в настоящей работе при улучшении набора данных CIC IoT 2022. Этот и предыдущий примеры подтверждают, что расширение набора данных признаками, представленными в табл. 3, может улучшить качество профилирования IoT-устройств.

Выполнено сравнение полученных результатов работы с результатами, полученными в других рабо-

Таблица 4. Philips Hue Bridge: результаты классификации
Table 4. Philips Hue Bridge: Classification Report

Сценарий	Количество экземпляров трафика	Метрики			
		Precision	Recall	F-score	Accuracy
Flood_TCP	80 000	1,000	1,000	1,000	0,999
Flood_UDP	80 000	1,000	1,000	1,000	
ALEXAOFF	4000	0,995	0,999	0,997	
ALEXAON	4000	0,999	0,995	0,997	
GOOGLEOFF	4000	0,999	0,999	0,999	
GOOGLEON	4000	0,999	1,000	0,999	
LANOFF	4000	1,000	1,000	1,000	
LANON	4000	1,000	1,000	1,000	
LOCALBUTTON	4000	1,000	1,000	1,000	
WANOFF	4000	0,997	0,999	0,998	
WANON	4000	0,999	0,997	0,998	
macro avg	196 000	0,999	0,999	0,999	
weighted avg	196 000	0,999	0,999	0,999	

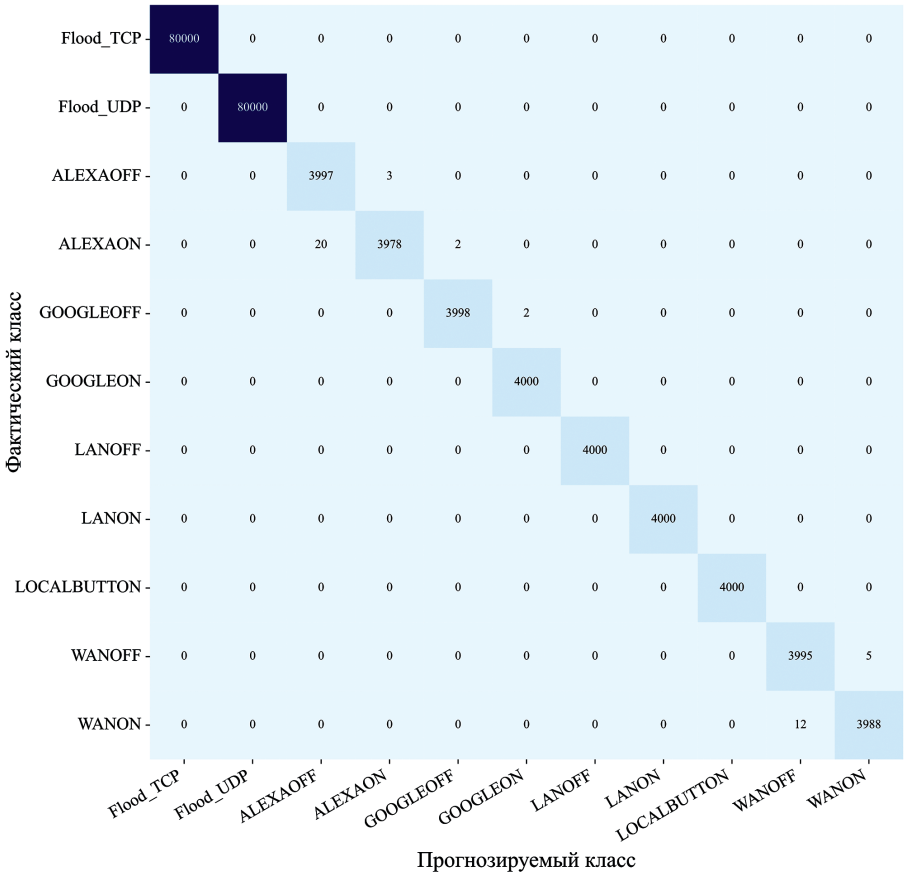

Рис. 3. Philips Hue Bridge: матрица ошибок
Fig. 3. Philips Hue Bridge: Confusion Matrix

тах (табл. 6). При этом важно отметить, что напрямую сравнить результаты, полученные в данной работе, с результатами других исследователей, затруднительно, поскольку использовались различные версии набора данных CIC IoT 2022 [14].

Заметим, что многими исследователями решались задачи, которые не совпадают с теми, что решались в данной работе. В целом, сравнительный анализ показал, что полученные результаты не уступают результатам других исследований. Это подтверждает при-

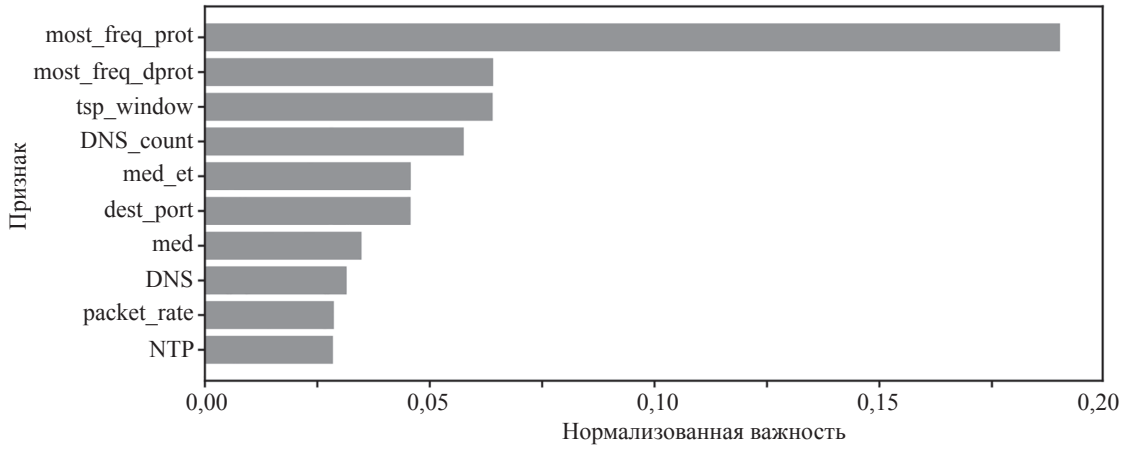

Рис. 4. Philips Hue Bridge: 10 наиболее значимых признаков по 10 000 случайным наблюдениям для модели CatBoost
Fig. 4. Philips Hue Bridge: 10 most significant features for 10,000 random observations for the CatBoost model

Таблица 5. Philips Hue Bridge: результаты реконструкции
Table 5. Philips Hue Bridge: Results of the Reconstruction

Сценарий	Количество экземпляров трафика	Метрики			
		Precision	Recall	F-score	Accuracy
Легитимный	160 000	0,978	0,999	0,989	0,981
Вредоносный	36 000	0,999	0,898	0,947	
macro avg	196 000	0,989	0,949	0,968	
weighted avg	196 000	0,982	0,981	0,981	

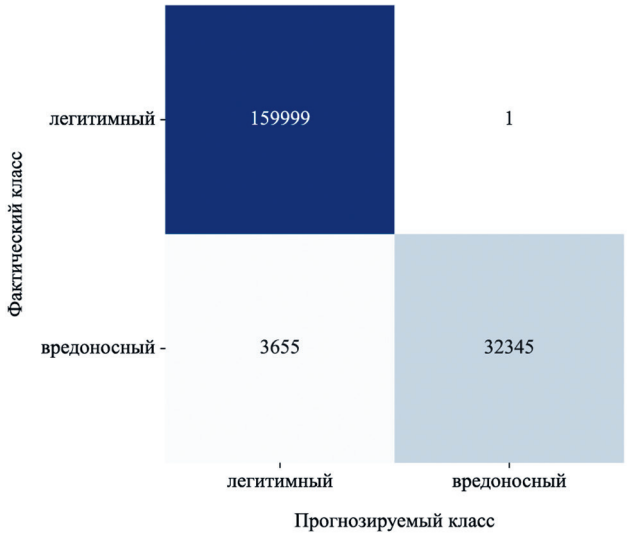

Рис. 5. Philips Hue Bridge: матрица ошибок для модели Elliptic Envelope
Fig. 5. Philips Hue Bridge: error matrix for the Elliptic Envelope model

менимость предлагаемого подхода для обеспечения безопасности IoT-устройств.

Преимуществами выполненного исследования являются несколько ключевых факторов. Реализованная предобработка данных эффективно решает задачи извлечения признаков и нормализации, что обеспечивает подготовку высококачественных наборов. Обучение моделей в большинстве случаев демонстрирует высокую эффективность при решении задач обнаружения атак и аномалий. Кроме того, предложенный подход обеспечивает наглядное представление результатов, включая расширенные отчеты о классификации, матрицы ошибок, визуализацию значимости признаков, а также определение ключевых показателей эффективности.

Отметим, что наличие таких данных позволяет упростить интерпретацию данных от ML-моделей и принимать более обоснованные решения по управ-

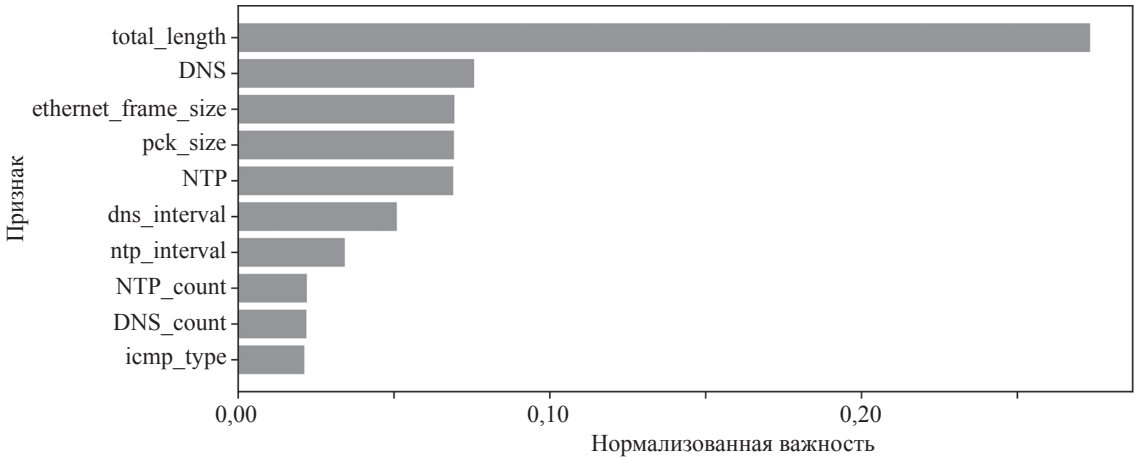

Рис. 6. Philips Hue Bridge: 10 наиболее значимых признаков по 10 000 случайным наблюдениям для модели Elliptic Envelope
Fig. 6. Philips Hue Bridge: 10 most significant features for 10,000 random observations for the Elliptic Envelope

Таблица 6. Сравнение с аналогами
Table 6. Comparison

Ссылка на источник	Данные	Подход	Модель	Метрики				Тип
				Accuracy	Precision	Recall	F-score	
[17]	CIC IoT 2022	Одна модель для всех устройств	FFNN	0,9993	0,9993	0,9993	0,9993	Max
			LSTM	0,9989	0,9989	0,9989	0,9989	Max
			RandNN	0,9642	0,9642	0,9642	0,9642	Max
[18]	CIC IoT 2022	Одна модель для всех устройств	YaTC	0,9658	—	—	0,9658	Max
	ISCXTor2016			0,9972	—	—	0,9972	Max
[19]	CIC IoT 2023	Одна модель для всех устройств	non-ML (IoT-PRIDS)	0,9874	0,9384	0,9971	0,9529	Max
[20]	CIC IoT 2023	Одна модель для всех устройств	EnsAdp CIDS	0,9893	0,9950	0,9940	0,9945	Max
	CICIDS-2017			0,9977	0,9982	0,9986	0,9978	Max
[21]	CIC IoT 2023	Одна модель для всех устройств	2-class RF	0,9955	0,9955	0,9955	0,9955	Max
			34-class RF	0,9633	0,9628	0,9633	0,9626	Max
[22]	CIC IoT 2023	Одна модель для всех устройств	Ensemble Learning Boosting (LR, NB, SVM, KNN, MLP)	0,9319	0,9353	0,9319	0,9324	Max
	Edge-IIoTset 2023			0,9601	0,9606	0,9601	0,9594	Max
[23]	IoTID20	Одна модель для всех устройств	RF	0,9868	—	—	—	Max
			XGB	0,9867	—	—	—	Max
			Extra Tree	0,9845	—	—	—	Max
Настоящая работа	CIC IoT 2022 (расширенная версия)	Отдельные модели для каждого устройства	Anomaly detection (IF, EE, 1-SVM)	0,9940	0,9940	0,9940	0,9930	Max
				0,9549	0,9568	0,9549	0,9547	Avg
				0,7720	0,7740	0,7720	0,7720	Min
			Attack detection (RF, XGB, CB)	1,0000	1,0000	1,0000	1,0000	Max
				0,9880	0,9880	0,9880	0,9880	Avg
				0,9270	0,9270	0,9270	0,9270	Min

лению безопасностью сети с IoT-устройствами. Что касается проблем, которые необходимо решить для улучшения предлагаемого подхода, то можно выделить следующие: эффективность моделей сильно зависит от качества и объема исходных данных, что может потребовать дополнительных усилий по сбору и предварительной обработке информации; существует необходимость дальнейшей оптимизации и тестирования для адаптации системы к индустриальным требованиям и повышения ее производительности.

Заключение

В работе представлен подход к профилированию устройств Интернета вещей (IoT) с целью обнаружения вредоносной активности. Данный подход работает с сетевой активностью устройств и анализирует их поведение с помощью моделей машинного обучения (Machine Learning, ML). Основой анализа служит сетевой трафик, используемый для выявления аномального поведения и обнаружения кибератак.

В процессе исследования был улучшен и расширен открытый набор данных сетевого трафика IoT-устройств. В процессе расширения были обработаны PCAP-файлы, соответствующие легитимным и вредоносным сценариям IoT-устройства. Также были использованы как признаки трафика из оригинального набора данных, так и добавлены новые признаки. Кроме того, в набор были добавлены синтетические данные, чтобы решить проблему дисбаланса для недостаточно представленных сценариев поведения каждого устройства.

Задача идентификации вредоносной активности была разделена на две подзадачи: обнаружение аномалий и обнаружение атак. Для этого ML-реконструкторы обучались только на легитимном трафике каждого устройства (профиль нормального поведения) и затем использовались для прогнозирования аномалий. ML-классификаторы обучались для каждого устройства отдельно как на легитимном, так и вредоносном трафике (общий профиль поведения). Задача данных моделей — идентифицировать, какое легитимное (тип сценария работы устройства) или вредоносное (тип атаки на устройство) поведение представляет собой трафик.

В ходе экспериментов были созданы профили для 26 IoT-устройств. Для каждого устройства, где имеется трафик вредоносной активности, были отобраны наиболее эффективные ML-реконструктор (обнаружение аномалий) и ML-классификатор (обнаружение атак). Для остальных устройств был отобран только ML-классификатор.

Для каждого устройства выполнено сравнение эффективности Random Forest, XGBoost и CatBoost в задаче обнаружения атак, а также Isolation Forest, Elliptic Envelope и One Class Support Vector Machine в задаче обнаружения аномалий (где это было возможно).

На основании полученных результатов можно дать следующие рекомендации по дальнейшему развитию и применению полученных результатов: расширение набора данных (больше устройств и сценариев); разработка дополнительных методов предобработки данных; расширение наборов гиперпараметров для оптимизации; интеграция с другими моделями и алгоритмами; расширение функциональности системы; адаптация к новым угрозам. В ходе дальнейших исследований планируется сосредоточиться на разработке адаптивных методов защиты, многофакторном профилировании, повышении отказоустойчивости системы и других аспектах, связанных с обеспечением безопасности систем с IoT-устройствами. Это позволит улучшить защиту от киберугроз, минимизировать риски несанкционированного доступа, повысить эффективность управления и мониторинга в условиях динамически меняющихся угроз и требований.

Литература

1. Левшун Д.С., Гайфулина Д.А., Чечулин А.А., Котенко И.В. Проблемные вопросы информационной безопасности киберфизических систем // Информатика и автоматизация. 2020. Т. 19. № 5. С. 1050–1088. https://doi.org/10.15622/ia.2020.19.5.6
2. Levshun D.S., Chechulin A.A., Kotenko I.V. Design lifecycle for secure cyber-physical systems based on embedded devices // Proc. of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). 2017. P. 277–282. https://doi.org/10.1109/IDAACS.2017.8095090
3. Levshun D., Chechulin A., Kotenko I., Chevalier Y. Design and verification methodology for secure and distributed cyber-physical systems // Proc. of the 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). 2019. P. 1–5. https://doi.org/10.1109/NTMS.2019.8763814
4. Levshun D., Chechulin A., Kotenko I. A technique for design of secure data transfer environment: Application for I2C protocol // Proc. of the IEEE Industrial Cyber-Physical Systems (ICPS). 2018. P. 789–794. https://doi.org/10.1109/ICPHYS.2018.8390807
5. Rose J.R., Swann M., Bendiab G., Shiaeles S., Kolokotronis N. Intrusion detection using network traffic profiling and machine learning for IoT // Proc. of the. 7th International Conference on Network Softwarization (NetSoft). 2021. P. 409–415. https://doi.org/10.1109/NetSoft51509.2021.9492685
6. Safi M., Dadkhah S., Shoeleh F., Mahdikhani H., Molyneaux H., Ghorbani A.A. A survey on IoT profiling, fingerprinting, and identification // ACM Transactions on Internet of Things. 2022. V. 3. N 4. P. 1–39. https://doi.org/10.1145/3539736
7. Ahmed K.I., Tahir M., Habaebi M.H., Lau S.L., Ahad A. Machine learning for authentication and authorization in iot: Taxonomy,

References

1. Levshun D., Gaifulina D., Chechulin A., Kotenko I. Problematic issues of information security of cyber-physical systems. *Informatics and Automation*, 2020, vol. 19, no. 5, pp. 1050–1088. (in Russian). https://doi.org/10.15622/ia.2020.19.5.6
2. Levshun D.S., Chechulin A.A., Kotenko I.V. Design lifecycle for secure cyber-physical systems based on embedded devices. *Proc. of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2017, pp. 77–282. https://doi.org/10.1109/IDAACS.2017.8095090
3. Levshun D., Chechulin A., Kotenko I., Chevalier Y. Design and verification methodology for secure and distributed cyber-physical systems. *Proc. of the 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2019, pp. 1–5. https://doi.org/10.1109/NTMS.2019.8763814
4. Levshun D., Chechulin A., Kotenko I. A technique for design of secure data transfer environment: Application for I2C protocol. *Proc. of the IEEE Industrial Cyber-Physical Systems (ICPS)*, 2018, pp. 789–794. https://doi.org/10.1109/ICPHYS.2018.8390807
5. Rose J.R., Swann M., Bendiab G., Shiaeles S., Kolokotronis N. Intrusion detection using network traffic profiling and machine learning for IoT. *Proc. of the. 7th International Conference on Network Softwarization (NetSoft)*, 2021, pp. 409–415. https://doi.org/10.1109/NetSoft51509.2021.9492685
6. Safi M., Dadkhah S., Shoeleh F., Mahdikhani H., Molyneaux H., Ghorbani A.A. A survey on IoT profiling, fingerprinting, and identification. *ACM Transactions on Internet of Things*, 2022, vol. 3, no. 4, pp. 1–39. https://doi.org/10.1145/3539736
7. Ahmed K.I., Tahir M., Habaebi M.H., Lau S.L., Ahad A. Machine learning for authentication and authorization in iot: Taxonomy,

challenges and future research direction // Sensors. 2021. V. 21. N 15. P. 5122. https://doi.org/10.3390/s21155122
8. Wójcicki K., Biegańska M., Paliwoda B., Górna J. Internet of things in industry: research profiling, application, challenges and opportunities — a review // Energies. 2022. V. 15. N 5. P. 1806. https://doi.org/10.3390/en15051806
9. Nguyen G.L., Dumba B., Ngo Q.D., Le H.V., Nguyen T.N. A collaborative approach to early detection of IoT Botnet // Computers & Electrical Engineering. 2022. V. 97. P. 107525. https://doi.org/10.1016/j.compeleceng.2021.107525
10. Bansal M., Priya. Performance comparison of MQTT and CoAP protocols in different simulation environments // Lecture Notes in Networks and Systems. 2021. V. 145. P. 549–560. https://doi.org/10.1007/978-981-15-7345-3_47
11. Canavese D., Mannella L., Regano L., Basile C. Security at the edge for resource-limited IoT devices // Sensors. 2024. V. 24. N 2. P. 590. https://doi.org/10.3390/s24020590
12. Rose J.R., Swann M., Bendiab G., Shiaeles S., Kolokotronis N. Intrusion detection using network traffic profiling and machine learning for IoT // Proc. of the 7th International Conference on Network Softwarization (NetSoft). 2021. P. 409–415. https://doi.org/10.1109/NetSoft51509.2021.9492685
13. Dadkhah S., Mahdikhani H., Danso P.K., Zohourian A., Truong K.A., Ghorbani A.A. Towards the development of a realistic multidimensional IoT profiling dataset // Proc. of the 19th Annual International Conference on Privacy, Security & Trust (PST). 2022. P. 1–11. https://doi.org/10.1109/PST55820.2022.9851966
14. Safi M., Kaur B., Dadkhah S., Shoeleh F., Lashkari A.H., Molyneaux H., Ghorbani A.A. Behavioural monitoring and security profiling in the internet of things (IoT) // Proc. of the IEEE 23rd International Conference on High Performance Computing and Communications 7th International Conference on Data Science and Systems 19th International Conference on Smart City and 7th International Conference on Dependability in Sensor Cloud and Big Data Systems and Applications HPCC/DSS/Smartcity/Dependsys. 2021. P. 1203–1210. https://doi.org/10.1109/HPCC-DSS-SMARTCITY-DEPENDSYS53884.2021.00185
15. Гетьман А.И., Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А. Сравнение системы обнаружения вторжений на основе машинного обучения с сигнатурными средствами защиты информации // Труды Института системного программирования РАН. 2022. Т. 34. № 5. С. 111–126. https://doi.org/10.15514/ISPRAS-2022-34(5)-7
16. Koball C., Rimal B.P., Wang Y., Salmen T., Ford C. IoT device identification using unsupervised machine learning // Information. 2023. V. 14. N 6. P. 320. https://doi.org/10.3390/info14060320
17. Bakhsh S.A., Khan M.A., Ahmed F., Alshehri M.S., Ali H., Ahmad J. Enhancing IoT network security through deep learning-powered Intrusion Detection System // Internet of Things. 2023. V. 24. P. 100936. https://doi.org/10.1016/j.iot.2023.100936
18. Zhao R., Zhan M., Deng X., Wang Y., Wang Y., Gui G., Xue Z. Yet another traffic classifier: A masked autoencoder based traffic transformer with multi-level flow representation // Proc. of the 37th AAAI Conference on Artificial Intelligence. 2023. V. 37. N 4. P. 5420–5427. https://doi.org/10.1609/aaai.v37i4.25674
19. Zohourian A., Dadkhah S., Molyneaux H., Neto E.C.P., Ghorbani A.A. IoT-PRIDS: Leveraging packet representations for intrusion detection in IoT networks // Computers & Security. 2024. V. 146. P. 104034. https://doi.org/10.1016/j.cose.2024.104034
20. Roshan K., Zafar A. Ensemble adaptive online machine learning in data stream: a case study in cyber intrusion detection system // International Journal of Information Technology. 2024. V. 16. N 8. P. 5099–5112. https://doi.org/10.1007/s41870-024-01727-y
21. Khan M.M., Alkhathami M. Anomaly detection in IoT-based healthcare: machine learning for enhanced security // Scientific Reports. 2024. V. 14. N 1. P. 5872. https://doi.org/10.1038/s41598-024-56126-x
22. Jeffrey N., Tan Q., Villar J.R. Using ensemble learning for anomaly detection in cyber–physical systems // Electronics. 2024. V. 13. N 7. P. 1391. https://doi.org/10.3390/electronics13071391
23. Bajpai S., Sharma K., Chaurasia B.K. Intrusion detection framework in IoT networks // SN Computer Science. 2023. V. 4. N 4. P. 350. https://doi.org/10.1007/s42979-023-01770-9

challenges and future research direction. *Sensors*, 2021, vol. 21, no. 15, pp. 5122. https://doi.org/10.3390/s21155122
8. Wójcicki K., Biegańska M., Paliwoda B., Górna J. Internet of things in industry: research profiling, application, challenges and opportunities — a review. *Energies*, 2022, vol. 15, no. 5, pp. 1806. https://doi.org/10.3390/en15051806
9. Nguyen G.L., Dumba B., Ngo Q.D., Le H.V., Nguyen T.N. A collaborative approach to early detection of IoT Botnet. *Computers & Electrical Engineering*, 2022, vol. 97, pp. 107525. https://doi.org/10.1016/j.compeleceng.2021.107525
10. Bansal M., Priya. Performance comparison of MQTT and CoAP protocols in different simulation environments. *Lecture Notes in Networks and Systems*, 2021, vol. 145, pp. 549–560. https://doi.org/10.1007/978-981-15-7345-3_47
11. Canavese D., Mannella L., Regano L., Basile C. Security at the edge for resource-limited IoT devices. *Sensors*, 2024, vol. 24, no. 2, pp. 590. https://doi.org/10.3390/s24020590
12. Rose J.R., Swann M., Bendiab G., Shiaeles S., Kolokotronis N. Intrusion detection using network traffic profiling and machine learning for IoT. *Proc. of the 7th International Conference on Network Softwarization (NetSoft)*, 2021, pp. 409–415. https://doi.org/10.1109/NetSoft51509.2021.9492685
13. Dadkhah S., Mahdikhani H., Danso P.K., Zohourian A., Truong K.A., Ghorbani A.A. Towards the development of a realistic multidimensional IoT profiling dataset. *Proc. of the 19th Annual International Conference on Privacy, Security & Trust (PST)*, 2022, pp. 1–11. https://doi.org/10.1109/PST55820.2022.9851966
14. Safi M., Kaur B., Dadkhah S., Shoeleh F., Lashkari A.H., Molyneaux H., Ghorbani A.A. Behavioural monitoring and security profiling in the internet of things (IoT). *Proc. of the IEEE 23rd International Conference on High Performance Computing and Communications 7th International Conference on Data Science and Systems 19th International Conference on Smart City and 7th International Conference on Dependability in Sensor Cloud and Big Data Systems and Applications HPCC/DSS/Smartcity/Dependsys*, 2021, pp. 1203–1210. https://doi.org/10.1109/HPCC-DSS-SMARTCITY-DEPENDSYS53884.2021.00185
15. Getman A.I., Goryunov M.N., Matskevich A.G., Rybolovlev D.A. A comparison of a machine learning-based intrusion detection system and signature-based systems. *Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS)*, 2022, vol. 34, no. 5, pp. 111–126. (in Russian). https://doi.org/10.15514/ISPRAS-2022-34(5)-7
16. Koball C., Rimal B.P., Wang Y., Salmen T., Ford C. IoT device identification using unsupervised machine learning. *Information*, 2023, vol. 14, no. 6, pp. 320. https://doi.org/10.3390/info14060320
17. Bakhsh S.A., Khan M.A., Ahmed F., Alshehri M.S., Ali H., Ahmad J. Enhancing IoT network security through deep learning-powered Intrusion Detection System. *Internet of Things*, 2023, vol. 24, pp. 100936. https://doi.org/10.1016/j.iot.2023.100936
18. Zhao R., Zhan M., Deng X., Wang Y., Wang Y., Gui G., Xue Z. Yet another traffic classifier: A masked autoencoder based traffic transformer with multi-level flow representation. *Proc. of the 37th AAAI Conference on Artificial Intelligence*, 2023, vol. 37, no. 4, pp. 5420–5427. https://doi.org/10.1609/aaai.v37i4.25674
19. Zohourian A., Dadkhah S., Molyneaux H., Neto E.C.P., Ghorbani A.A. IoT-PRIDS: Leveraging packet representations for intrusion detection in IoT networks. *Computers & Security*, 2024, vol. 146, pp. 104034. https://doi.org/10.1016/j.cose.2024.104034
20. Roshan K., Zafar A. Ensemble adaptive online machine learning in data stream: a case study in cyber intrusion detection system. *International Journal of Information Technology*, 2024, vol. 16, no. 8, pp. 5099–5112. https://doi.org/10.1007/s41870-024-01727-y
21. Khan M.M., Alkhathami M. Anomaly detection in IoT-based healthcare: machine learning for enhanced security. *Scientific Reports*, 2024, vol. 14, no. 1, pp. 5872. https://doi.org/10.1038/s41598-024-56126-x
22. Jeffrey N., Tan Q., Villar J.R. Using ensemble learning for anomaly detection in cyber–physical systems. *Electronics*, 2024, vol. 13, no. 7, pp. 1391. https://doi.org/10.3390/electronics13071391
23. Bajpai S., Sharma K., Chaurasia B.K. Intrusion detection framework in IoT networks. *SN Computer Science*, 2023, vol. 4, no. 4, pp. 350. https://doi.org/10.1007/s42979-023-01770-9

Авторы

Легкодымов Даниил Михайлович — студент, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация, SC 59721499100, https://orcid.org/0009-0002-2874-6632, danillegk65@gmail.com

Левшун Дмитрий Сергеевич — кандидат технических наук (Россия), PhD (Франция), старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 57189306576, https://orcid.org/0000-0003-1898-6624, levshun.d@iias.spb.su

Котенко Игорь Витальевич — доктор технических наук, профессор, заслуженный деятель науки Российской Федерации, главный научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 15925268000, https://orcid.org/0000-0001-6859-7120, ivkote@comsec.spb.ru

Authors

Daniil M. Legkodymov — Student, The Bonch-Bruevich Saint Petersburg State University of Telecommunications (SPbSUT), Saint Petersburg, 193232, Russian Federation, SC 59721499100, https://orcid.org/0009-0002-2874-6632, danillegk65@gmail.com

Dmitry S. Levshun — PhD (Russia), PhD (France), Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, SC 57189306576, https://orcid.org/0000-0003-1898-6624, levshun.d@iias.spb.su

Igor V. Kotenko — D.Sc., Professor, Honored Scientist of the Russian Federation, Chief Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, SC 15925268000, https://orcid.org/0000-0001-6859-7120, ivkote@comsec.spb.ru

Статья поступила в редакцию 19.04.2025
Одобрена после рецензирования 28.05.2025
Принята к печати 17.07.2025

Received 19.04.2025
Approved after reviewing 28.05.2025
Accepted 17.07.2025