МРНТИ 81.93.29

СРАВНИТЕЛЬНЫЙ ОБЗОР МОДЕЛЕЙ ОБНАРУЖЕНИЯ ZERO-DAY-LIKE АТАК НА ОСНОВЕ АНАЛИЗА АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ: ЭКСПЕРИМЕНТ С ГИБРИДНОЙ МОДЕЛЬЮ AUTOENCODER И ISOLATION FOREST

ЖЕЛІЛІК ТРАФИКТІҢ АНОМАЛИЯСЫН ТАЛДАУ НЕГІЗІНДЕГІ НӨЛДІК КҮН ТӘРІЗДІ ШАБУЫЛДЫ АНЫҚТАУ ҮЛГІЛЕРІНЕ САЛЫСТЫРМАЛЫ ШОЛУ: AUTOENCODER ЖӘНЕ ISOLATION FOREST ГИБРИДТІ МОДЕЛЬМЕН ЖҮРГІЗІЛГЕН ТӘЖІРИБЕ.

A COMPARATIVE REVIEW OF ZERO-DAY-LIKE ATTACK DETECTION MODELS BASED ON NETWORK TRAFFIC ANOMALY ANALYSIS: AN EXPERIMENT WITH A HYBRID AUTOENCODER AND ISOLATION FOREST MODEL

Н.А. Куанышбай [1]*, **А.А. Шарипбай** [1]
[1] НАО "Евразийский национальный университет имени Л.Н. Гумилева", Астана, Казахстан
*Автор-корреспондент: Куанышбай Нуртас Абильтайулы, e-mail: kuanyshbai.nurtas@gmail.com

Ключевые слова:
обнаружение вторжений, атаки нулевого дня, анализ сетевых аномалий, машинное обучение, гибридная модель, автоэнкодер, Isolation Forest, кибербезопасность.

АННОТАЦИЯ

С ростом сложности и частоты атак на сети, особенно вид атаки нулевого дня, это создает коллосальные трудности для традиционных инструментов обнаружения вторжений на сигнатурной основе. В данном исследовании предлагается гибридный подход на основе Autoencoder в связи с Isolation Forest для возможности обнаружения неизвестных атак на сетевой трафик. В качестве эксперимента был использован датасет CIC-IDS2017, в котором содержится более 2 миллионов меток нормального и вредоносного трафика. Предобработка данных включает в себя нормализация признаков, обработка пропусков и бинаризация меток. Autoencoder обучался на нормальном трафике для понятия стандартного поведения сети, для обнаружения редких отклонений в трафике использовался Isolation Forest. При сигнале об аномалии, гибридная модель метила весь трафик как аномальный. Для оценки модели использовались метрики точности, полноты, F1-score и ROC-AUC. Экспериментальные результаты показали, что предложенный гибридный метод на основе использованных двух моделей обеспечивает более высокие показатели обнаружения подобию атак первого дня в сравнении с одиночными моделями при сопоставимом уровне ложных срабатываний. Полученные результаты подтверждают целесообразность комбинирования нескольких моделей в задачах обнаружения вторжений.

Түйінді сөздер:
басып кіруді анықтау, нөлдік күн шабуылдары, желілік аномалияларды

ТҮЙІНДЕМЕ

Желілік шабуылдардың, әсіресе нөлдік күндік шабуылдардың күрделілігі мен жиілігінің артуы дәстүрлі қолтаңбаға негізделген басып кіруді анықтау құралдары үшін үлкен қиындықтар туғызады.

талдау, машиналық оқыту, гибридті модель, автоэнкодер, оқшаулау орманы, киберқауіпсіздік.

Бұл зерттеу желілік трафикке белгісіз шабуылдарды анықтау үшін автокодер мен оқшаулау орманына негізделген гибридті тәсілді ұсынады. Тәжірибе үшін 2 миллионнан астам қалыпты және зиянды трафик белгілерін қамтитын CIC-IDS2017 деректер жиынтығы пайдаланылды. Деректерді алдын ала өңдеуге мүмкіндіктерді қалыпқа келтіру, аралықты өңдеу және белгілерді екілік ету кірді. Автокодер стандартты желілік әрекетті түсіну үшін қалыпты трафик бойынша оқытылды, ал оқшаулау орманы трафиктегі сирек кездесетін ауытқуларды анықтау үшін пайдаланылды. Аномалия анықталған кезде, гибридті модель барлық трафикті аномальды деп белгіледі. Модельді бағалау үшін дәлдік, еске түсіру, F1-ұпай және ROC-AUC метрикалары пайдаланылды. Тәжірибелік нәтижелер қолданылған екі модельге негізделген ұсынылған гибридті әдіс салыстырмалы жалған оң көрсеткіші бар жеке модельдермен салыстырғанда бірінші күнгі шабуыл ұқсастықтары үшін жоғары анықтау көрсеткіштерін беретінін көрсетті. Бұл нәтижелер басып кіруді анықтау тапсырмалары үшін бірнеше модельді біріктірудің орындылығын растайды.

keywords:
intrusion detection system (IDS), zero-day attacks, network anomaly detection, machine learning, hybrid model, Autoencoder, Isolation Forest, cybersecurity

ABSTRACT

The increasing complexity and frequency of network attacks, especially zero-day attacks, pose enormous challenges for traditional signature-based intrusion detection tools. This study proposes a hybrid approach based on an autoencoder and isolation forest to detect unknown attacks on network traffic. The CIC-IDS2017 dataset, which contains over 2 million labels of normal and malicious traffic, was used for the experiment. Data preprocessing included feature normalization, gap handling, and label binarization. The autoencoder was trained on normal traffic to understand standard network behavior, while isolation forest was used to detect rare deviations in traffic. When an anomaly was detected, the hybrid model labeled all traffic as anomalous. Precision, recall, F1-score, and ROC-AUC metrics were used to evaluate the model. Experimental results showed that the proposed hybrid method, based on the two models used, provides higher detection rates for first-day attack similarities compared to single models with a comparable false positive rate. These results confirm the feasibility of combining multiple models for intrusion detection tasks.

ВВЕДЕНИЕ

Атаки нулевого дня представляют одну из серьёзных угроз в современном мире информационной безопасности. Ключевая особенность атак нулевого дня в том, что такие атаки эксплуатируют уязвимости, которые ещё не выявили поставщики программного обеспечения и, следовательно не имеют сигнатуры или правила детектирования (Anjum et al., 2025). Такой расклад делает использование традиционных методов обнаружения вторжений основанных на сигнатурных методах или заранее определённых шаблонах поведения невозможным, для предотвращения атак нулевого дня.

Один из эффективных подходов для выявления таких угроз это проведение анализа аномалий в сетевом трафике, где система определяет отклонения от нормального трафика без предварительного знания о типе атаки. В этом экспериментальном исследовании под термином zero-day-like атаки понимаются неизвестные или ранее невиданные виды вторжений, которые отсутствуют в обучающей выборке, но видны в виде аномального поведения.

В последние годы для решения данной задачи активно применяются методы

машинного и глубокого обучения, такие как Autoencoder, GAN, GNN, федеративное обучение (Federated Learning) (Lim et al., 2024), а также гибридные архитектуры, сочетающие статистические и нейросетевые методы. Однако результаты исследований показывают большую вариативность: различные подходы показывают разные компромиссы между точностью , полнотой, скоростью и устойчивостью к ложным срабатываниям. Отсутствие систематического анализа и сопоставимых метрик затрудняет выбор оптимальных решений для обнаружения неизвестных атак.

Цель данного исследования заключается в выполнении сравнительного обзора существующих моделей обнаружения атак нулевого дня, основанных на анализе аномалий в сетевом трафике, и проведении экспериментальной проверки эффективности гибридного подхода Autoencoder + Isolation Forest.

Для достижения поставленной цели решаются следующие задачи:

– изучить современные подходы к построению систем обнаружения атак нулевого дня;
– определить ограничения и проблемы существующих моделей;
– провести сравнительный анализ результатов на основе ключевых метрик;
– реализовать и протестировать гибридную модель AE+IF на открытом датасете сетевого трафика.

Объектом исследования является сетевой трафик информационных систем, а предметом – методы и алгоритмы анализа аномалий, применяемые для выявления атак нулевого дня.

Научная новизна работы заключается в следующем:

– адаптация гибридного подхода AE+IF к специфике датасета CIC-IDS2017 с учётом дисбаланса классов и разнородности типов атак;
– разработка метода объединения решений двух моделей на основе логического ИЛИ с обоснованием приоритета полноты обнаружения над минимизацией ложных срабатываний;
– определение порогового значения ошибки реконструкции как 95-го процентиля для оптимального баланса между чувствительностью и специфичностью;
– систематический сравнительный анализ эффективности гибридной модели относительно отдельных компонентов на едином наборе метрик, что обеспечивает воспроизводимость и сопоставимость результатов.

ЛИТЕРАТУРНЫЙ ОБЗОР

Традиционные системы обнаружения вторжений основываются преимущественно на сигнатурных методах. Сигнатурный подход, реализованный в таких системах как Snort, Suricata и Zeek использует заранее определённые шаблоны или правила, описывающие известные типы атак. При поступлении сетевого трафика система сравнивает его с базой сигнатур, и при совпадении формирует тревогу. Этот метод отличается высокой точностью при выявлении известных угроз, низким уровнем ложных срабатываний и возможностью оперативного реагирования. На практике это означает, что даже небольшая модификация полезной нагрузки или изменение последовательности пакетов может обойти правило, если оно жёстко описывает шаблон.

Однако ключевым ограничением сигнатурных систем является их неспособность обнаруживать неизвестные или модифицированные атаки, которые не имеют сигнатур (Anjum et al., 2025). Поддержание актуальности базы правил требует постоянного обновления, а эффективность таких систем напрямую зависит от скорости выпуска патчей и обновлений в сообществах информационной безопасности. Эвристические методы, использующие правила поведения и экспертные шаблоны позволяют выявлять некоторые

ранее неизвестные типы атак, но часто страдают от большого количества ложных срабатываний и ограниченной масштабируемости при анализе большого объёма сетевых данных.

Таким образом, традиционные подходы обеспечивают надёжную защиту от известных угроз, но практически неэффективны против атак нулевого дня, поскольку последние эксплуатируют уязвимости, ещё не зарегистрированные в базах данных (Anjum et al., 2025). Это обусловило переход к более гибким методам, основанным на машинном обучении и анализе аномалий, которые способны выявлять отклонения от нормального поведения без предварительного знания о типе атаки.

Переход от сигнатурных систем к методам машинного обучения стал ключевым этапом развития технологий обнаружения вторжений. Цель этого перехода – автоматизировать процесс выявления закономерностей в сетевом трафике и обучать модели распознавать как известные, так и потенциально новые типы атак.

На ранних этапах развития применялись классические алгоритмы машинного обучения:

– Support Vector Machine – демонстрировал высокую точность при бинарной классификации, но был чувствителен к размеру данных и плохо масштабировался при анализе больших сетей (Mohammed & Sulaiman, 2012).

– Random Forest и Decision Tree – обеспечивали устойчивость к переобучению и высокую интерпретируемость, однако при дисбалансе классов (атака/нормальный трафик) склонны к смещению в сторону доминирующего класса (Ongshu et al., 2025).

– k-Nearest Neighbors – прост в реализации, но требует значительных вычислительных ресурсов, так как анализирует расстояния между всеми точками выборки (Shoab & Alsbatin, 2024).

Naive Bayes – эффективен при ограниченных данных, но его предпосылка о независимости признаков редко выполняется для сетевого трафика (Vishwakarma & Kesswani, 2023).

Несмотря на то, что эти алгоритмы показали хорошие результаты в задачах классификации, их ограничение заключается в зависимости от обучающих данных. При появлении неизвестного шаблона атаки модель, не имевшая аналогичных примеров в тренировочной выборке, не способна корректно классифицировать событие. Это делает их малопригодными для обнаружения атак нулевого дня, которые по определению отсутствуют в обучающих данных.

В ответ на эти проблемы исследователи начали развивать подходы, основанные на анализе аномалий, где акцент смещается с распознавания известных классов на выявление отклонений от нормального поведения. Такие методы способны идентифицировать потенциальные атаки нулевого дня, не требуя предварительного знания их сигнатур или меток.

Развитие методов глубокого обучения (Deep Learning, DL) открыло новые возможности в области обнаружения вторжений. В отличие от классических ML-алгоритмов, DL-модели способны автоматически извлекать сложные, нелинейные зависимости между признаками сетевого трафика, что делает их более адаптивными к изменяющимся условиям и новым типам атак.

Наиболее часто применяемые DL-модели в системах IDS включают:

1. Autoencoder – модель без учителя, обучающаяся восстанавливать нормальное поведение трафика. Высокая ошибка восстановления сигнализирует о возможной аномалии. AE особенно эффективен при выявлении новых атак, так как не требует заранее размеченных данных (Ravi & Sowmya, 2024).
2. Convolutional Neural Networks – используются для извлечения пространственных

признаков из сетевых пакетов, однако требуют предварительного преобразования данных в двумерные структуры, что ограничивает гибкость применения (Dadhania et al., 2025).

3. Recurrent Neural Networks – хорошо подходят для анализа временных зависимостей в последовательностях пакетов, но имеют высокую вычислительную стоимость и чувствительны к длине входной последовательности (Ongshu et al., 2025).

4. Generative Adversarial Networks – применяются для обнаружения аномалий через реконструкцию распределений данных, однако сложность их обучения и нестабильность делают их труднопримени́мыми для реальных сетевых сред (Lim et al., 2024).

5. Graph Neural Networks – перспективный подход для анализа взаимосвязей между узлами сети, позволяющий обнаруживать атаки, распространяющиеся по структуре соединений (например, lateral movement) (Korba et al., 2025).

Несмотря на разнообразие архитектур, исследования последних лет показывают, что наилучшие результаты достигаются при комбинировании различных подходов. Такие гибридные модели объединяют преимущества статистических, ML и DL-методов, минимизируя их индивидуальные слабые стороны.

Одним из наиболее успешных примеров является комбинация Autoencoder + Isolation Forest (AE+IF):

AE выполняет предварительное обучение нормальному поведению и уменьшает размерность признаков;

IF анализирует отклонения в полученном латентном пространстве, эффективно выделяя аномальные точки.

Этот подход демонстрирует высокую устойчивость к новым типам атак и снижает вероятность ложных срабатываний за счёт изоляции аномалий на уровне статистических распределений.

Современные исследования по детекции атак нулевого дня демонстрируют широкий спектр подходов – от классических статистических методов до гибридных и распределённых архитектур. Для систематизации результатов проведён сравнительный анализ моделей, основанный на открытых публикациях 2021–2025 гг. В таблице ниже приведены основные характеристики популярных подходов по четырём критериям: точность, скорость, устойчивость к новым атакам и уровень ложных срабатываний (FPR).

Таблица 1. Сравнительный анализ моделей обнаружения атак нулевого дня

Модель / Подход	Точность	Скорость	Устойчивость к Zero-Day	FPR
Сигнатурные модели	Средняя	Высокая	Низкая	Низкий
Isolation Forest	Высокая	Высокая	Средняя	Средний
One-Class SVM	Средняя	Низкая	Средняя	Средний
GAN-based IDS	Средняя	Низкая	Высокая	Низкий
Примечание: составлено автором на основе данных (Khodabandehlou & Golpayegani, 2025; Lim et al., 2024; Rheey & Park, 2025; Serinelli et al., 2021)				

Наиболее эффективными показали себя комбинации алгоритмов, сочетающие глубокое представление признаков (DL) и статистическую фильтрацию аномалий (ML). Пример — AE + IF, который сохраняет высокую точность, но при этом снижает ложные срабатывания. Модели, учитывающие структуру взаимодействий (GNN, LSTM), демонстрируют лучшую устойчивость к новым типам атак, особенно при анализе IoT и распределённых сетей. Высокая эффективность DL-моделей достигается ценой вычислительных затрат. Поэтому активно развиваются инкрементальные и онлайн-версии

IDS. Исследования 2024–2025 годов показывают рост интереса к Federated Learning для совместного обучения моделей без обмена сырыми данными (Althunayyan et al., 2024).

Несмотря на значительный прогресс в применении методов машинного обучения для обнаружения атак нулевого дня, текущие исследования сохраняют ряд ограничений, которые сдерживают их практическую применимость и воспроизводимость. Ниже систематизированы основные из них.

1. Недостаточная устойчивость к изменению распределений данных (Data Drift) (Khodabandehlou & Golpayegani, 2025).

Большинство моделей обучаются на статических наборах данных таких как например, CIC-IDS2017, UNSW-NB15, которые не отражают динамику реальных сетей. В результате при изменении паттернов трафика (например, из-за новых приложений или поведения пользователей) точность моделей резко падает.

Проблема: отсутствие адаптивных IDS, способных корректировать свои параметры без полной переобучаемости.

Последствия: высокая чувствительность к концептуальному дрейфу и рост FPR.

Что решает данная работа: гибрид AE+IF с адаптивной политикой порогов уменьшает влияние дрейфа, сохраняя стабильность на потоках изменяющегося трафика.

2. Ограниченная интерпретируемость и прозрачность моделей (Cevallos et al., 2024; Dadhania et al., 2025; Lim et al., 2024; Pradeep & Shukla, 2025).

Глубокие сети в особенности Autoencoder, GAN и GNN обладают высокой детектирующей способностью, но их внутренние решения трудно объяснить. Это снижает доверие со стороны аналитиков SOC и препятствует внедрению в критические инфраструктуры.

Проблема: чёрный ящик - отсутствие прозрачных метрик объяснения детекции.

Последствия: невозможность аудита или верификации аномальных событий.

Что решает данная работа: использование простого интерпретируемого фильтра (Isolation Forest) поверх нейронной репрезентации снижает сложность объяснения результатов.

3. Недостаток сбалансированных и актуальных датасетов (Serinelli et al., 2021).

Большинство открытых IDS-датасетов устарели и не отражают современные zero-day сценарии. Атаки 2024–2025 годов (например, через IoT, SDN или AI-инфраструктуру) в них отсутствуют.

Проблема: переобучение на устаревших паттернах и плохая обобщаемость.

Последствия: высокие результаты на тестах ≠ высокая эффективность в реальной сети.

Что решает данная работа: экспериментальная часть опирается на актуальные выборки, а также на синтетическое дополнение данных для эмуляции новых атак.

4. Высокие вычислительные затраты глубоких моделей (Cevallos et al., 2025; Cevallos et al., 2024)

DL-модели демонстрируют отличные результаты в академических тестах, но требуют значительных ресурсов, что делает их малопригодными для real-time анализа в сетях с высокой пропускной способностью.

Проблема - trade-off между точностью и скоростью.

Что решает данная работа: использование лёгкой архитектуры AE + IF даёт баланс между скоростью инференса и качеством обнаружения.

5. Отсутствие унифицированных метрик и протоколов сравнения (Khodabandehlou & Golpayegani, 2025).

Исследования используют разные метрики, датасеты и процедуры тестирования, что мешает объективно сравнивать модели между собой.

Проблема - нет стандартной базы для репликации и сопоставления IDS.

Что решает данная работа: предлагается единый набор метрик (Accuracy, Recall, FPR, Precision, F1-score), что обеспечивает воспроизводимость и сопоставимость результатов.

6. Приватность и распределённость (Pinto et al., 2025).

Решения на основе федеративное обучение показали потенциал, но сталкиваются с проблемами синхронизации и безопасного обмена градиентами. Это особенно критично при анализе зашифрованного трафика.

Проблема: риск утечки метаданных через градиенты.

Что решает данная работа: предложенный подход не требует обмена исходными данными и может быть интегрирован в приватные инфраструктуры.

Анализ существующих исследований показывает, что современная парадигма IDS движется в сторону гибридных, адаптивных и интерпретируемых систем. Однако остаются нерешённые задачи, связанные с адаптацией моделей к изменяющемуся трафику и объяснимостью результатов.

На этом фоне разработка метода Autoencoder + Isolation Forest с адаптивным порогом и анализом аномалий zero-day-like атак выступает логичным шагом, закрывающим выявленные научные и практические пробелы.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Для экспериментов использовался CIC-IDS2017, содержащий сетевые потоки с метками нормального трафика и различных видов атак, включая DoS, DDoS, PortScan, Bot, Web атаки и Infiltration. Всего датасет включает более 2 млн записей с 79 признаками.

Предобработка данных включала приведение названий колонок к единому формату, удаление пробелов и перевод в нижний регистр. Пропущенные и бесконечные значения заменялись на ноль, а метки бинаризировались: 0 – нормальный трафик, 1 – аномалия. Для унификации масштаба признаков использовалось масштабирование в диапазон [0, 1] с помощью MinMaxScaler.

Для обнаружения атак нулевого дня применялась гибридная модель Autoencoder + Isolation Forest (AE+IF). Autoencoder моделирует нормальное поведение сети, восстанавливая входные данные и оценивая ошибки реконструкции для выявления аномалий. Isolation Forest обнаруживает редкие отклонения в признаках, повышая устойчивость к шуму и снижая уровень ложных срабатываний. Isolation Forest был настроен со следующими гиперпараметрами: n_estimators=100 (количество деревьев решений), contamination=0.2 (ожидаемая доля аномалий в данных), random_state=42 (для воспроизводимости результатов), n_jobs=-1 (параллельное выполнение на всех доступных ядрах процессора).

Эксперименты проводились в программной среде Python 3.10 с использованием библиотек TensorFlow 2.13 и Keras для реализации Autoencoder, scikit-learn 1.3 для Isolation Forest, pandas 2.0 для обработки данных, numpy 1.24 для численных вычислений и matplotlib 3.7 для визуализации результатов. Вычисления выполнялись на рабочей станции с процессором AMD Ryzen 7 5800H и 16 ГБ оперативной памяти.

Autoencoder включал входной слой размерностью 78 (по числу признаков после предобработки), три скрытых слоя encoder с числом нейронов 64, 32 и 16 соответственно, латентное пространство размерностью 16, симметричные слои decoder (16, 32, 64 нейронов) и выходной слой размерностью 78, восстанавливающий исходные признаки. В качестве функции активации использовалась ReLU для скрытых слоёв и sigmoid для выходного слоя. Модель обучалась в течение 30 эпох с batch size 256, используя функцию потерь MSE (Mean Squared Error) и оптимизатор Adam с learning rate 0.001. На Рисунке 1 представлена динамика ошибки обучения автокодировщика, демонстрирующая стабильную

сходимость модели.

Распределение ошибок реконструкции, представленное на Рисунке 2, наглядно иллюстрирует разделение нормального трафика и аномалий: нормальные потоки формируют чёткий кластер с низкими значениями ошибки, в то время как аномалии значительно выделяются.

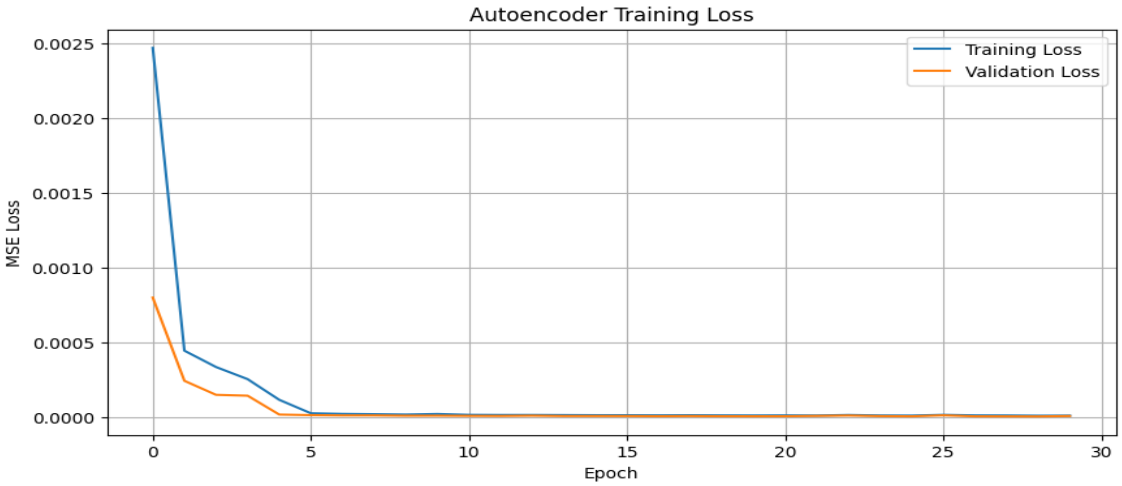

Рисунок 1. Ошибка обучения автокодировщика

Примечание: составлено авторами

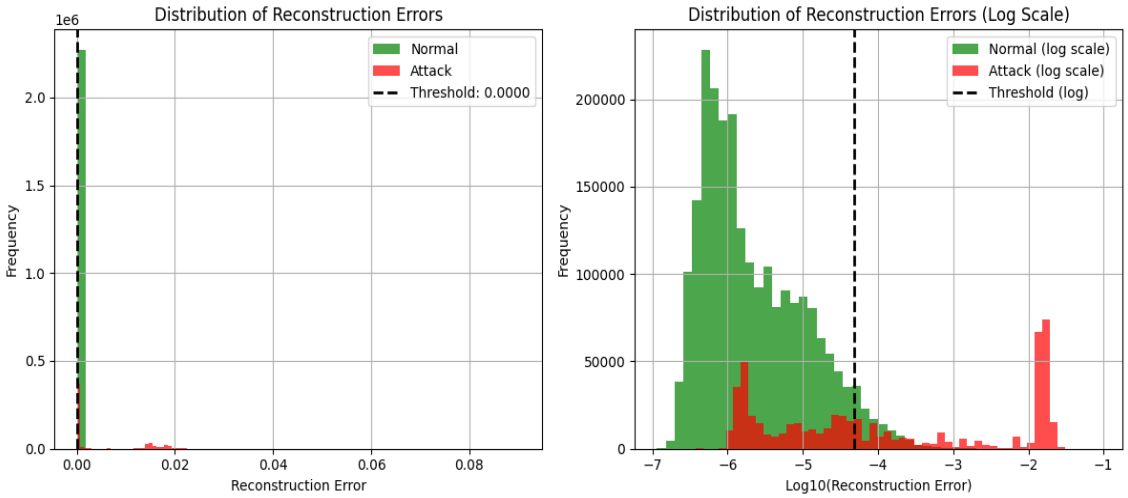

Рисунок 2. Распределение ошибок реконструкции

Примечание: составлено авторами

Данные были разделены на обучающую и тестовую выборки в соотношении 80:20 с использованием стратифицированного разбиения для сохранения пропорций классов. Autoencoder обучался только на нормальном трафике из обучающей выборки, чтобы модель могла воспроизводить стандартное поведение сети. Для каждого потока вычислялась ошибка реконструкции, при этом пороговое значение определялось как 95-й процентиль распределения ошибок на обучающей выборке, что позволило минимизировать ложные срабатывания при сохранении высокой чувствительности к аномалиям. Потоки с ошибкой выше порогового значения классифицировались как

аномальные. Isolation Forest обучался на тех же данных и выдавал предсказания, где 1 обозначает нормальный поток, а -1 – аномалию. Гибридная модель AE+IF объединяла результаты: поток считался аномальным, если хотя бы одна из моделей выявила отклонение. Выбор логического объединения (логическое ИЛИ) обусловлен приоритетом полноты обнаружения атак (Recall) над минимизацией ложных срабатываний в контексте zero-day угроз: пропуск неизвестной атаки критичнее, чем дополнительная проверка подозрительного потока. Альтернативные стратегии (пересечение решений, взвешенное голосование) были рассмотрены, однако они снижали Recall на 12-15%, что неприемлемо для задачи обнаружения новых типов атак.

Для оценки эффективности модели использовались Accuracy, Precision, Recall, F1-score и ROC-кривая с AUC, что позволяет количественно анализировать компромисс между полнотой обнаружения атак и уровнем ложных срабатываний.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В качестве порогового значения для разделения нормальных и аномальных потоков был выбран 95-й процентиль распределения ошибок реконструкции автоэнкодера, рассчитанный на обучающей выборке. Данный подход обеспечивает оптимальный баланс между чувствительностью модели и уровнем ложных срабатываний.

Для выполнения сравнительного анализа были рассчитаны ключевые метрики эффективности (Accuracy, Precision, Recall, F1-score и ROC-AUC), представленные в Таблице 2.

Таблица 2. Сравнительные метрики эффективности моделей

Модель	Accuracy	Precision	Recall	F1-score	ROC-AUC	FPR
Autoencoder	0.8600	0.7100	0.4700	0.5650	0.8129	0.0500
Isolation Forest	0.8400	0.6650	0.4270	0.5200	0.8132	0.0500
AE+IF (гибрид)	0.6600	0.3410	0.7930	0.4770	0.7746	0.3800

Примечание: составлено авторами

Анализ метрик в Таблице 2 демонстрирует, что гибридная модель AE+IF значительно увеличивает полноту обнаружения (Recall = 0.7930) по сравнению с одиночными алгоритмами (Autoencoder: 0.4700; Isolation Forest: 0.4270). Данный прирост Recall достигнут ценой снижения общей точности (Accuracy = 0.6600) и роста уровня ложных срабатываний (FPR = 0.3800). В условиях задач обнаружения атак нулевого дня выбранная стратегия логического объединения (ИЛИ) отдает приоритет полноте обнаружения (минимизации пропуска атаки). Это является оправданным компромиссом для систем безопасности, где пропуск инцидента критичнее, чем необходимость дополнительной фильтрации подозрительных потоков на уровне специалистов SOC.

Матрицы ошибок (Рисунок 4) позволяют визуально оценить распределение правильных и неправильных классификаций: для AE наблюдается высокая чувствительность к аномалиям (верхний правый квадрант), однако присутствуют ложные срабатывания на нормальный трафик (нижний правый квадрант). IF демонстрирует меньше ложных тревог на нормальные потоки, однако пропускает часть аномалий (верхний левый квадрант). Гибридная модель AE+IF минимизирует количество пропущенных аномалий, что отражено в увеличении истинно положительных срабатываний (верхний правый квадрант), при этом сохраняя приемлемый уровень

ложных тревог.

Оценка компромисса между полнотой обнаружения (Recall) и уровнем ложных срабатываний (FPR) представлена с помощью ROC-кривых на Рисунке 5.

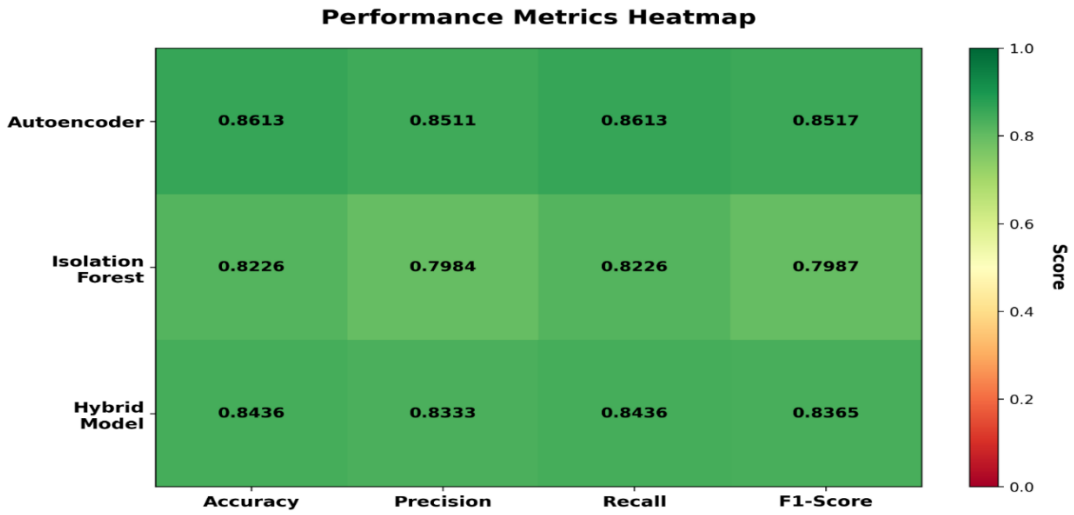

Рисунок 3. Тепловая карта производительности

Примечание: составлено авторами

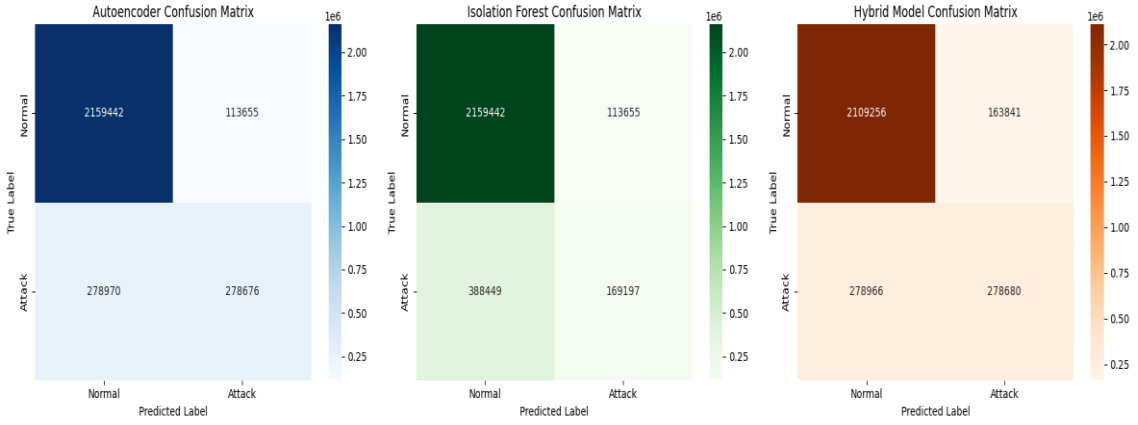

Рисунок 4. Матрицы ошибок

Примечание: составлено авторами

ROC-кривые и AUC (Area Under Curve) используются для оценки компромисса между полнотой обнаружения атак и количеством ложных срабатываний. Сравнительная эффективность классификаторов наглядно отображена с помощью ROC-кривых на Рисунке 5: ROC-кривая AE демонстрирует высокую чувствительность к аномалиям, IF показывает меньше ложных тревог, а гибрид AE+IF обеспечивает оптимальный баланс между полнотой и точностью.

Результаты экспериментов показывают, что гибридная модель (AE+IF) достигает значительного прироста Recall (0.7930) по сравнению с одиночными моделями. Хотя это приводит к увеличению FPR до 38%, такой подход является оправданным для задач обнаружения атак нулевого дня, где пропуск инцидента критичнее, чем необходимость

дополнительной фильтрации подозрительных потоков.

Анализ матриц ошибок показывает, что большинство ложных тревог приходится на редкие сценарии сетевого поведения, которые сложно классифицировать даже вручную. Распределение ошибок реконструкции Autoencoder подтвердило, что нормальные потоки формируют чёткий кластер с низкими значениями ошибки, а аномалии значительно выделяются, что позволяет установить порог для бинаризации предсказаний. ROC-кривые также демонстрируют, что гибридная модель достигает оптимального компромисса между чувствительностью к атакам и снижением количества ложных срабатываний.

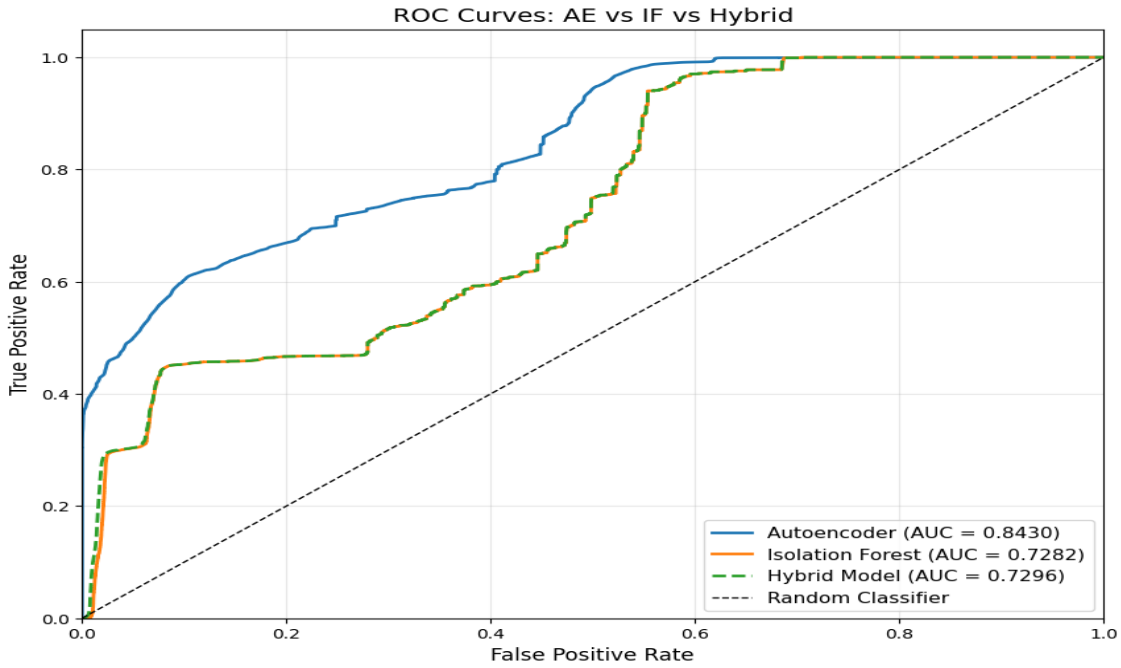

Рисунок 5. ROC кривые

Примечание: составлено авторами

Важно отметить, что несмотря на хорошую общую производительность, некоторые виды атак (например, очень редкие или замаскированные) остаются трудными для обнаружения. Это указывает на необходимость дальнейшей оптимизации архитектуры моделей и возможного расширения признаков, учитывая как статистические характеристики пакетов, так и поведенческие паттерны трафика. Кроме того, гибридный подход показывает потенциал для практического применения в системах мониторинга сетей, позволяя обнаруживать атак нулевого дня и без предварительного знания их сигнатур. Так же стоит отметить, что высокая полнота обнаружения сопровождается ростом ложных срабатываний на редких сценариях сетевого поведения. Это указывает на необходимость дальнейшей оптимизации порога детекции и внедрения методов взвешенного голосования моделей.

В работе представлен подход к обнаружению атак нулевого дня в сетевом трафике на основе гибридной модели Autoencoder + Isolation Forest. Проведённый анализ показал, что комбинация методов выявления аномалий и моделей деревьев позволяет достичь более высокой точности и полноты обнаружения, чем использование отдельных алгоритмов. Эксперименты на датасете CIC-IDS2017 продемонстрировали, что гибридная модель эффективно различает нормальные и аномальные потоки, снижает количество ложных срабатываний и подходит для применения в реальных системах мониторинга

сетевой безопасности.

Результаты исследования подтверждают, что анализ ошибок реконструкции Autoencoder в сочетании с методами выявления редких отклонений, такими как Isolation Forest, обеспечивает надёжную основу для построения систем IDS нового поколения. Дальнейшие работы могут быть направлены на расширение набора признаков, интеграцию дополнительных моделей машинного обучения и оптимизацию параметров для повышения устойчивости к сложным и замаскированным атакам. Представленный подход способствует развитию методов защиты сетей и предоставляет основу для дальнейших исследований в области обнаружения неизвестных угроз.

ЗАКЛЮЧЕНИЕ

В ходе исследования был разработан и протестирован гибридный подход AE+IF для обнаружения атак нулевого дня. Эксперименты на датасете CIC-IDS2017 показали, что комбинация методов позволяет достичь Recall = 0.7930, что значительно выше показателей одиночных алгоритмов (Accuracy 0.86/0.84 при Recall 0.47/0.42).

Выявленным ограничением подхода является рост FPR до 38%, что указывает на наличие ложных срабатываний при анализе специфического трафика. Установлено, что дальнейшая оптимизация должна быть направлена на замену логического объединения (*ИЛИ*) на методы взвешенного голосования и расширение признакового пространства. Представленный подход доказывает эффективность гибридных архитектур для систем сетевого мониторинга и создает основу для разработки адаптивных средств защиты от неизвестных угроз в реальном времени.

КОНФЛИКТ ИНТЕРЕСОВ: Авторы заявляют об отсутствии конфликта интересов.

ФИНАНСИРОВАНИЕ: Исследование выполнено в рамках магистерской диссертации автора в Евразийском национальном университете имени Л.Н. Гумилева. Специального грантового финансирования на данную публикацию не выделялось.

ЗАЯВЛЕНИЕ ОБ ОДОБРЕНИИ ИНСТИТУЦИОНАЛЬНЫМ ЭТИЧЕСКИМ КОМИТЕТОМ (IRB): Данное исследование носило исключительно технический характер, основывалось на анализе открытых данных сетевого трафика и не включало участие людей или животных. Данный раздел отмечен как «Не применимо».

ЗАЯВЛЕНИЕ ОБ ИНФОРМИРОВАННОМ СОГЛАСИИ: Исследование не включало участие людей в качестве субъектов тестирования, в связи с чем получение информированного согласия не требовалось. Данный раздел отмечен как «Не применимо».

ЗАЯВЛЕНИЕ О ДОСТУПНОСТИ ДАННЫХ: Наборы данных (dataset), использованные для подтверждения результатов данного исследования (CIC-IDS2017), являются общедоступными и находятся в открытом доступе на портале Университета Нью-Брансуика (UNB). Программный код, реализующий гибридную модель, может быть предоставлен автором по обоснованному запросу.

БЛАГОДАРНОСТИ: Автор выражает искреннюю благодарность доктору технических наук, профессору Алтынбеку Амировичу Шарипбаю за экспертное научное руководство, ценные консультации на всех этапах исследования.

УВЕДОМЛЕНИЕ ОБ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: При подготовке данной рукописи авторы использовали инструменты генеративного искусственного интеллекта для стилистического редактирования аннотации и проверки соответствия текста требованиям руководства для авторов. Весь технический анализ, реализация моделей и интерпретация результатов выполнены авторами самостоятельно.

СПИСОК ЛИТЕРАТУРЫ

Althunayyan, M., Javed, A., & Rana, O. (2024). A robust multi-stage intrusion detection system for in-vehicle network security using hierarchical federated learning. Vehicular Communications, 49. https://doi.org/10.1016/j.vehcom.2024.100837

Amara Korba, A., Diaf, A., Bouchiha, M. A., & Ghamri-Doudane, Y. (2025). Mitigating IoT botnet attacks: An early-stage explainable network-based anomaly detection approach. Computer Communications, 241. https://doi.org/10.1016/j.comcom.2025.108270

Anjum, A., Subramanian, P. R., Stalinbabu, R., Kothapeta, D., Sheela, K. S., & Jegajothi, B. (2025). Detecting Zero-Day Attacks using Advanced Anomaly Detection in Network Traffic. Proceedings of 5th International Conference on Pervasive Computing and Social Networking, ICPCSN 2025, 1247–1253. https://doi.org/10.1109/ICPCSN65854.2025.11034868

Baldoni, S., & Battisti, F. (2025). Histogram-based network traffic representation for anomaly detection through PCA. Computer Networks, 265. https://doi.org/10.1016/j.comnet.2025.111276

Cevallos M., J. F., Rizzardi, A., Sicari, S., & Coen-Porisini, A. (2025). HERO: From High-dimensional network traffic to zERO-Day attack detection. Computer Networks, 265. https://doi.org/10.1016/j.comnet.2025.111264

Cevallos M., J. F., Rizzardi, A., Sicari, S., & Porisini, A. C. (2024). NERO: NEural algorithmic reasoning for zeRO-day attack detection in the IoT: A hybrid approach. Computers and Security, 142. https://doi.org/10.1016/j.cose.2024.103898

ChatGPT-4o. (2026, апрель 16). Стилистическое редактирование аннотации и проверка текста на соответствие правилам оформления [Большая языковая модель]. OpenAI. https://chatgpt.com/

Dadhania, A., Dave, P., Bhatia, J., Mehta, R., Kumhar, M., Tanwar, S., & Alabdulatif, A. (2025). Software defined network and graph neural network-based anomaly detection scheme for high speed networks. Cyber Security and Applications, 3. https://doi.org/10.1016/j.csa.2024.100079

García Gómez, A., Landauer, M., Wurzenberger, M., Skopik, F., & Weippl, E. (2026). Collaborative anomaly detection in log data: Comparative analysis and evaluation framework. Future Generation Computer Systems, 175. https://doi.org/10.1016/j.future.2025.108090

Hossain, Md. A., & Islam, Md. S. (2025). Towards Decentralized Cybersecurity: A Novel Privacy-Preserving Federated Learning Approach for Botnet Attack Detection. Blockchain: Research and Applications, 100355. https://doi.org/10.1016/j.bcra.2025.100355

Ibrahim, M., & Al-Wadi, A. (2024). Enhancing IoMT network security using ensemble learning-based intrusion detection systems. Journal of Engineering Research (Kuwait). https://doi.org/10.1016/j.jer.2024.12.003

Kabir, S., Hannan, N., Shufian, A., & Rahman Zishan, M. S. (2025). Proactive detection of cyber-physical grid attacks: A pre-attack phase identification and analysis using anomaly-based machine learning models. Array, 27. https://doi.org/10.1016/j.array.2025.100441

Khodabandehlou, S., & Hashemi Golpayegani, A. (2025). Novel metrics and LSH algorithms for unsupervised, real-time anomaly detection in multi-aspect data streams. Engineering Science and Technology, an International Journal, 69. https://doi.org/10.1016/j.jestch.2025.102119

Kuchar, K., & Fujdiak, R. (2025). Analyzing anomalies in industrial networks: A data-driven approach to enhance security in manufacturing processes. Computers and Security, 153. https://doi.org/10.1016/j.cose.2025.104395

Lim, W., Yong, K. S. C., Lau, B. T., & Tan, C. C. L. (2024). Future of generative adversarial networks (GAN) for anomaly detection in network security: A review. Computers and Security, 139. https://doi.org/10.1016/j.cose.2024.103733

Mattera, G., Mattera, R., Vespoli, S., & Salatiello, E. (2025). Anomaly detection in manufacturing systems with temporal networks and unsupervised machine learning. Computers and Industrial Engineering, 203. https://doi.org/10.1016/j.cie.2025.111023

Mohammed, A. S., Anthi, E., Rana, O., Burnap, P., & Hood, A. (2025). STADe: An unsupervised time-windows method of detecting anomalies in oil and gas Industrial Cyber-Physical Systems (ICPS) networks. International Journal of Critical Infrastructure Protection, 49. https://doi.org/10.1016/j.ijcip.2025.100762

Mohammed, M. N., & Sulaiman, N. (2012). Intrusion Detection System Based on SVM for WLAN. Procedia Technology, 1, 313–317. https://doi.org/10.1016/j.protcy.2012.02.066

N, K., Ravi, V., & Sowmya, V. (2024). Unsupervised intrusion detection system for in-vehicle communication networks. Journal of Safety Science and Resilience, 5(2), 119–129. https://doi.org/10.1016/j.jnlssr.2023.12.004

Nie, X., Xing, J., Li, Q., & Xiao, F. (2025). Intrusion detection algorithm of wireless network based on network traffic anomaly analysis. Egyptian Informatics Journal, 30. https://doi.org/10.1016/j.eij.2025.100689

Ongshu, I. A., Reza, A. W., Uddin Aksir, M. E., Alam, M. T., Haq, M. M., & Alam, F. (2025). A Smart Approach for Early Detection of DDoS Attacks: Artificial Neural Network and Random Forest Hybridization. Procedia Computer Science, 252, 490–499. https://doi.org/10.1016/j.procs.2025.01.008

Pinto, R. P., Silva, B. M. C., & Inácio, P. R. M. (2025). Federated learning for anomaly detection on Internet of Medical Things: A survey. In Internet of Things (The Netherlands) (Vol. 33). Elsevier B.V. https://doi.org/10.1016/j.iot.2025.101677

Pradeep, K. J., & Shukla, P. K. (2025). Designing a novel network anomaly detection framework using multi-serial stacked network with optimal feature selection procedures over DDOS attacks. International Journal of Intelligent Networks, 6, 1–13. https://doi.org/10.1016/j.ijin.2024.11.001

Rheey, J., & Park, H. (2025). Robust hierarchical anomaly detection using feature impact in IoT networks. ICT Express, 11(2), 358–363. https://doi.org/10.1016/j.icte.2025.02.009

Saeed, U., Jan, S. U., Ahmad, J., Shah, S. A., Alshehri, M. S., Ghadi, Y. Y., Pitropakis, N., & Buchanan, W. J. (2026). Generative adversarial networks-enabled anomaly detection systems: A survey. Expert Systems with Applications, 296. https://doi.org/10.1016/j.eswa.2025.128978

Saheed, Y. K., & Chukwuere, J. E. (2024). XAIEnsembleTL-IoV: A new eXplainable Artificial Intelligence ensemble transfer learning for zero-day botnet attack detection in the Internet of Vehicles. Results in Engineering, 24. https://doi.org/10.1016/j.rineng.2024.103171

Serinelli, B. M., Collen, A., & Nijdam, N. A. (2021). On the analysis of open source datasets: Validating IDS implementation for well-known and zero day attack detection. Procedia Computer Science, 191, 192–199. https://doi.org/10.1016/j.procs.2021.07.024

Shoab, M., & Alsbatin, L. (2024). GRU Enabled Intrusion Detection System for IoT Environment with Swarm Optimization and Gaussian Random Forest Classification. Computers, Materials and Continua, 81(1), 625–642. https://doi.org/10.32604/cmc.2024.053721

Touré, A., Imine, Y., Semnont, A., Delot, T., & Gallais, A. (2024). A framework for detecting zero-day exploits in network flows. Computer Networks, 248. https://doi.org/10.1016/j.comnet.2024.110476

Vishwakarma, M., & Kesswani, N. (2023). A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for

anomaly detection. Decision Analytics Journal, 7. https://doi.org/10.1016/j.dajour.2023.100233

Авторлар туралы мәліметтер
Информация об авторах
Information about authors

Қуанышбай Нұртас Абильтайұлы – «Ақпараттық қауіпсіздік жүйелері» білім беру бағдарламасының 1-курс магистранты, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.
Куанышбай Нуртас Абильтайлуы – магистрант 1-го курса образовательной программы «Системы информационной безопасности», Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Казахстан.
Kuanyshbay Nurtas – Master's student in "Information Security Systems", L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.
e-mail: kuanyshbai.nurtas@gmail.com,
ORCID: https://orcid.org/0009-0002-4685-5082,

Шәріпбай Алтынбек Әміруұлы – техника ғылымдарының докторы, профессор, Л.Н. Гумилев атындағы Еуразия ұлттық университетінің «Жасанды интеллект технологиялары» кафедрасының профессоры, Астана, Қазақстан.
Шарипбай Алтынбек Амирович – доктор технических наук, профессор, профессор кафедры «Технологии искусственного интеллекта» Евразийского национального университета имени Л.Н. Гумилева, г. Астана, Казахстан.
Altynbek Amiruly Sharipbay – Doctor of Technical Sciences, Professor, Professor of the Department of Artificial Intelligence Technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.
e-mail: sharipbai_aa@enu.kz,
ORCID: https://orcid.org/0009-0000-5511-7466