Мельник Эдуард Всеволодович – Южный федеральный университет; e-mail: evmelnik@sfedu.ru; г. Ростов-на-Дону, Россия; д.т.н.; профессор кафедры вычислительной техники.

Блох Денис Евгеньевич – Юго-Западный государственный университет; e-mail: den5553@yandex.ru; г. Курск, Россия; кафедра вычислительной техники; аспирант.

Безмельцев Александр Игоревич – Юго-Западный государственный университет; e-mail: a.i.bezmeltsev@yandex.ru; г. Курск Россия; кафедра вычислительной техники; аспирант.

Панищев Владимир Славиевич – Юго-Западный государственный университет; e-mail: gskunk@yandex.ru; г. Курск, Россия; к.т.н.; доцент; доцент кафедры вычислительной техники.

Полторацкий Сергей Николаевич – Юго-Западный государственный университет; e-mail: merlinserg@list.ru; г. Курск, Россия; к.т.н.; доцент кафедры вычислительной техники.

Melnik Eduard Vsevolodovich – Southern Federal University; e-mail: evmelnik@sfedu.ru; Rostov-on-Don, Russia; dr. of eng. sc.; professor, Department of Computer Engineering.

Blokh Denis Evgenievich – Southwestern State University; e-mail: den5553@yandex.ru; Kursk, Russia; the Department of Computer Engineering; postgraduate student.

Bezmeltsev Alexander Igorevich – Southwestern State University; e-mail: a.i.bezmeltsev@yandex.ru; Kursk, Russia; the Department of Computer Engineering; postgraduate student.

Panishchev Vladimir Slavievich – Southwestern State University; e-mail: gskunk@yandex.ru; Kursk, Russia; cand. of eng. sc.; associate professor; associate professor, Department of Computer Engineering.

Poltoratsky Sergey Nikolaevich – Southwestern State University; e-mail: merlinserg@list.ru; Kursk, Russia; cand. of eng. sc.; associate professor, Department of Computer Engineering.

УДК 004.89 DOI 10.18522/2311-3103-2025-5-229-243

В.А. Частикова, К.В. Козачёк, Е.С. Коробская, В.П. Кравцов

ОБНАРУЖЕНИЕ КИБЕРВТОРЖЕНИЙ НА ОСНОВЕ СЕТЕВОГО ТРАФИКА И ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЯ С ИСПОЛЬЗОВАНИЕМ ДАТАСЕТА UNSW-NB15

В статье основное внимание уделяется исследованию поведения пользователей и созданию поведенческих моделей. Это помогает улучшить точность определения аномалий и оперативно выявлять нестандартную активность в сети. Целью данного исследования является сравнительный анализ эффективности двух моделей машинного обучения – многослойного персептрона (MLP) и алгоритма Random Forest – для обнаружения кибервторжений на основе анализа сетевого трафика и поведения пользователей. Поведенческие модели позволяют выявлять отклонения от нормальной активности пользователей и сетевых взаимодействий, что значительно повышает полноту обнаружения кибервторжений. При проведении исследования использовался набор данных UNSW-NB15, который включает актуальные типы атак и характеристики как сетевого трафика, так и пользовательской активности. Перед реализацией моделей была проведена предварительная обработка данных, выбор признаков, нормализация и кодирование категориальных признаков. Оценка моделей осуществлялась с использованием различных метрик, таких как точность (accuracy), полнота (recall), AUC-ROC, precision, F1-score и другие. Результаты исследования показали, что алгоритм Random Forest обеспечивает высокую точность классификации (95%), а многослойный персептрон (MLP), в свою очередь, достиг выдающихся результатов по AUC (0.9830) и точности (precision, 0.9869). В работе представлен анализ и характеристика методов анализа поведения пользователей и классификации сетевого трафика, показано сравнение наборов данных для систем обнаружения вторжений (IDS), а также даны практические рекомендации по выбору моделей в зависимости от условий эксплуатации. Результаты исследования могут быть полезны при разработке адаптивных систем защиты, которые сочетают высокую точность и скорость работы.

Обнаружение кибервторжений; машинное обучение; UNSW-NB15; многослойный персептрон (mlp); random forest; классификация сетевого трафика; AUC-ROC.

V.A. Chastikova, K.V. Kozachek, E.S. Korobskaya, V.P. Kravtsov

DETECTION OF CYBER INTRUSIONS BASED ON NETWORK TRAFFIC AND USER BEHAVIOR USING THE UNSW-NB15 DATASET

The article focuses on the study of user behavior and the creation of behavioral models. This helps to improve the accuracy of anomaly detection and quickly identify non-standard network activity. The purpose of this study is to compare the effectiveness of two machine learning models – the multilayer perceptron (MLP) and the Random Forest algorithm – for detecting cyber intrusions based on the analysis of network traffic and user behavior. Behavioral models make it possible to detect deviations from normal user activity and network interactions, which significantly increases the completeness of cyber intrusion detection. The study used the UNSW-NB15 dataset, which includes current types of attacks and characteristics of both network traffic and user activity. Prior to the implementation of the models, preliminary data processing, feature selection, normalization and coding of categorical features were carried out. The models were evaluated using various metrics such as accuracy, recall, AUC-ROC, precision, F1-score, and others. The results of the study showed that the Random Forest algorithm provides high classification accuracy (95%), and the multilayer perceptron (MLP), in turn, achieved outstanding results in AUC (0.9830) and accuracy (precision, 0.9869). The paper presents an analysis and characterization of methods for analyzing user behavior and classifying network traffic, a comparison of data sets for intrusion detection systems (IDS), and practical recommendations for choosing models depending on operating conditions. The results of the study can be useful in the development of adaptive protection systems that combine high accuracy and speed.

Cyber intrusion detection; machine learning; UNSW-NB15; multilayer perceptron (MLP); random forest; network traffic classification; AUC-ROC.

Введение. В современном цифровом мире вопросы информационной безопасности приобретают всё большую актуальность. С ростом числа подключённых к интернету устройств, объёмов передаваемых данных и усложнением сетевых инфраструктур возрастает и количество потенциальных угроз. Кибервторжения представляют серьёзную опасность как для организаций, так и для частных пользователей, поскольку могут привести к утечке конфиденциальной информации, финансовым потерям и нарушению функционирования критически важных систем.

Одним из ключевых направлений в обеспечении информационной безопасности является обнаружение кибервторжений на основе анализа сетевого трафика и поведения пользователей, что позволяет выявлять не только уже известные угрозы, но и новые, ранее не регистрировавшиеся атаки. Анализ технических характеристик сетевого взаимодействия в совокупности с моделированием поведенческих шаблонов пользователей повышает точность выявления аномалий и способствует более эффективной защите информационных систем.

Существующие методы обнаружения кибервторжений можно разделить на три основных категории: сигнатурные методы, методы анализа аномалий и поведенческие модели. [1–3] Сигнатурный анализ предполагает сравнение входящего трафика с известными шаблонами атак, что эффективно против ранее известных угроз, но малоэффективно против новых, изменённых или модифицированных атак. Методы анализа аномалий направлены на выявление отклонений от типичного поведения системы и обладают большей гибкостью, но требуют тщательной настройки и отбора признаков. Поведенческие модели, в свою очередь, строят профиль нормального функционирования пользователей и систем, что позволяет своевременно фиксировать нетипичную активность.

Целью настоящего исследования является сравнительный анализ эффективности алгоритмов машинного обучения – случайного леса (Random Forest) и нейронной сети – в задаче обнаружения кибервторжений на основе реального и широко используемого в научных работах датасета UNSW-NB15. В рамках исследования рассматривается способность указанных алгоритмов точно классифицировать сетевые атаки различных типов и адаптироваться к разнообразным характеристикам сетевого трафика. Полученные результаты могут способствовать развитию более точных и адаптивных систем обнаружения угроз в реальных условиях.

Обзор существующих подходов. В последние годы исследователи и специалисты в области информационной безопасности всё больше внимания уделяют проблеме своевременного и эффективного обнаружения кибервторжений. Количество атак растёт, они усложняются, и методы их проведения становятся разнообразнее. Всё это делает защиту информационных систем крайне важной задачей. Для повышения эффективности выявления киберугроз разрабатываются разные подходы. Среди них – классические методы анализа и современные решения на базе машинного обучения и искусственного интеллекта.

Учитывая актуальность этой темы и разнообразие подходов, стоит рассмотреть существующие работы, посвящённые обнаружению кибервторжений, которые могут послужить основой дальнейших исследований.

В работе [4] проанализированы популярные наборы данных NSL-KDD и UNSW-NB15. Долгое время NSL-KDD был ключевым датасетом для исследований в области обнаружения вторжений, и большинство исследований опирались именно него. Однако сегодня анализ поведения пользователей и сетевой активности играет настолько важную роль, что этот набор данных является недостаточным из-за ограниченного охвата современных угроз и отсутствия детализированных данных о действиях злоумышленников. В исследовании были удалены избыточные признаки и выбраны наиболее информативные для обнаружения вторжений. Затем был создан новый набор данных с моделированием современных атак. Алгоритмы машинного обучения протестировали на исходных и новом наборах. Наилучшие результаты в многоклассовой классификации показал XGBoost. Однако вопросы применимости моделей в реальных сетях остались без внимания.

В статьях [5, 6] исследованы возможности применения нейронных сетей для обнаружения аномального трафика в сетях Интернета вещей (IoT). Был проведён сравнительный анализ современных архитектур нейросетей на примере набора данных CIC IoT Dataset 2023. Замечено, что нейронные сети демонстрируют высокую точность в задачах обнаружения атак, однако требуют значительных вычислительных ресурсов и больших объёмов обучающих данных.

В работах [7, 8] исследуются методы многоклассовой классификации сетевых атак с применением алгоритмов машинного обучения. Проведен анализ эффективности логистической регрессии, наивного байесовского классификатора, дерева решений и метода опорных векторов на основе наборов данных CICIDS2018 и CICDDoS2019. Выявлено, что отбор признаков способствует повышению точности классификации. Однако в статье не обсуждаются вопросы устойчивости моделей к изменениям характеристик трафика и их адаптации к новым типам атак.

В исследовании [9] проводится анализ эффективности систем обнаружения и предотвращения вторжений (IDS/IPS), основанных на Suricata. Сравниваются различные решения с точки зрения производительности, точности обнаружения и интеграции в сетевую инфраструктуру. Отмечаются преимущества Suricata, такие как многопоточность и аппаратное ускорение. Однако вопросы адаптации к новым угрозам и автоматизации обновлений в статье не рассматриваются.

В статье [10] предложено использовать большие языковые модели (LLM) для уменьшения числа ложных срабатываний в системах обнаружения аномалий в сетевом трафике. Рассмотрен подход с применением трансформеров для анализа сетевых данных. Эксперименты показали эффективность метода. Однако не обсуждены вычислительные затраты и ограниченная применимость в реальных условиях эксплуатации.

В работах [11, 12] изучаются методы формирования обучающих наборов данных для моделей обнаружения компьютерных атак с использованием машинного обучения. Проведён анализ общедоступных датасетов и инструментов, применяемых для их обработки. В результате выявлены их недостатки и ошибки. В статье [11] разработана объектно-ориентированная библиотека на основе многослойного персептрона с применением датасета KDD Cup 1999 Data и модель на базе LSTM и эмбеддинговой сети с обучени-

ем по алгоритму Adam с использованием датасета CSE-CIC-IDS2018. А в [12] предложена методика сбора собственных данных, учитывающая параметры защищаемой сети. Также разработана и апробирована система генерации трафика и разметки данных. Показано, что обучение моделей на основе собственных данных даёт значительно лучшие результаты по сравнению с общедоступными наборами.

В исследованиях [13, 14] рассматривается разработка гибридных нейросетевых систем для анализа сетевого трафика и выявления атак. Основное внимание уделяется гетерогенной архитектуре, включающей LSTM-сети, которые оказались наиболее ресурсоёмкими. Для повышения эффективности обнаружения атак предложен нейроиммунный подход, объединяющий методы глубокого обучения и искусственные иммунные системы. Разработанный программный комплекс подтвердил перспективность гибридных моделей. Однако требуется дальнейшая оптимизация для снижения вычислительных затрат. Исследование сочетает теоретические разработки и практические решения в сфере кибербезопасности.

На основе статей можно выделить несколько алгоритмов:

1. Для классификации атак (на основе датасетов NSL-KDD, CICIDS, UNSW-NB15):
   - Random Forest (XGBoost): учитывают множество признаков для многоклассовой классификации и устойчивы к шуму.
   - Support Vector Machine (SVM): используются с ядрами (RBF, линейный) для нелинейного разделения и бинарной классификации.
   - Логистическая регрессия: базовый метод для оценки вероятности атаки, обычно комбинируется с другими алгоритмами.
2. Для анализа трафика и обнаружения аномалий:
   - LSTM/GRU: разновидность рекуррентных нейронных сетей с долгосрочной краткосрочной памятью, хорошо работающая с последовательными данными. Особенно полезна для анализа временных зависимостей в сетевом трафике.
   - Автокодировщики (Autoencoders): нейронные сети, которые изучают нормальное поведение и определяют аномалии по высокой ошибке реконструкции. Эффективны для Zero-Day атак.
   - Isolation Forest: нейронные сети, которые, в отличие от других алгоритмов, обнаруживают аномалии через «изоляцию» выбросов.
3. Гибридные системы:
   - Suricata (правила) + ML (SVM, LSTM): сочетают в себе сигнатурный анализ и машинное обучение для снижения ложных срабатываний, что позволяет снизить нагрузку на аналитиков.
4. Снижение ложных срабатываний через семантический анализ (LLM, NPL):
   - Применение языковых моделей (BERT, GPT, RoBERTa): LLM способны классифицировать события как "угроза" или "ложное срабатывание" с учетом контекста, для повышения точности работы систем кибербезопасности.

При сравнении различных методов стоит отметить, что классические алгоритмы машинного обучения, такие как логистическая регрессия, SVM и деревья решений, обладают высокой интерпретируемостью и устойчивостью, но ограничены при работе с большими объёмами данных и сложными взаимосвязями между признаками. В то же время нейросетевые методы, включая LSTM, CNN и Autoencoders, способны учитывать временные и контекстные особенности сетевого трафика, однако они требуют значительных вычислительных ресурсов и менее объяснимы. В табл. 1 представлено сравнение характеристик различных подходов к реализации систем IDS.

Таблица 1

Сравнение классических и нейросетевых моделей IDS

Критерий	Классические методы (SVM, RF, DT)	Нейросетевые методы (CNN, LSTM, Autoencoders)
Точность обнаружения	Средняя и высокая	Высокая и очень высокая
Скорость обучения	Высокая	Низкая и средняя
Потребление ресурсов	Низкое	Высокое
Интерпретируемость	Хорошая	Ограниченная
Адаптивность к новым атакам	Средняя	Высокая
Применимость в реальном времени	Высокая	Ограниченная

Для обучения и тестирования IDS-моделей используются различные датасеты. Основные виды датасетов [15–17] и их сравнительные характеристики приведены в табл. 2.

Таблица 2

Сравнение набора данных для задач IDS

Датасет	Год разработки	Преимущества	Недостатки	Типы атак	Применимые модели
KDD'99	1999	Первый массовый IDS-датасет; большой объём; простая структура	Устаревшие атаки; много дубликатов; не сбалансирован	DoS, Probe, R2L, U2R	Decision Tree, Naive Bayes, SVM
NSL-KDD	2009	Улучшенная версия KDD'99; удалены дубликаты; более сбалансирован	Основан на старом трафике	DoS, Probe, R2L, U2R	Random Forest, SVM, DNN
CICIDS2018	2018	Реалистичный сетевой трафик; PCAP+NetFlow; много метрик	Высокая ресурсоемкость; сложная структура	DDoS, Botnet, Brute Force, Heartbleed, Infiltration и др.	CNN, LSTM, Autoencoders
UNSW-NB15	2015	Современные типы атак; 49 атрибутов; сбалансированный; поведенческие и сетевые данные	Меньший объём, чем у CICIDS2017	Fuzzers, Analysis, Backdoors, Exploits, Generic, Reconnaissance, Shellcode, Worms	DNN, CNN, LSTM, Ensemble methods

Материалы и методы исследования. Опираясь на исследования, проводимые в других работах, и сравнительную таблицу наборов данных, для обнаружения кибервторжений будем использовать алгоритм Random Forest и нейронную сеть в виде многослойного персептрона на основе датасета UNSW-NB15.

Random Forest (RF, случайный лес) – ансамблевый метод на основе дерева решений, который объединяет несколько деревьев для повышения точности и устойчивости модели, предприимчивый к переобучению и способный обрабатывать данные с различной структурой. При обработке данных из датасета UNSW-NB15, Random Forest позволяет эффективно выявлять различные типы атак, особенно в условиях мультиклассовой классификации. Дополнительным преимуществом является возможность анализа важности признаков, что облегчает интерпретацию результатов и отбор наиболее значимых параметров сетевого трафика.

MLP способен обрабатывать как количественные, так и категориальные признаки из набора UNSW-NB15 после предварительной нормализации и кодирования. Использование функции активации ReLU и механизма регуляризации (Dropout, L2-регуляризация)

позволяет избежать переобучения, а оптимизация с помощью алгоритма Adam обеспечивает быструю сходимость. Модель MLP демонстрирует высокие результаты при бинарной и мультиклассовой классификации, эффективно различая как нормальный трафик, так и конкретные виды атак.

Набор данных UNSW-NB15 разработан в Университете Нового Южного Уэльса (UNSW) и предназначен для создания современного набора данных для тестирования сетей. Он улучшает существующие наборы данных, такие как KDD98, за счёт включения новых атак, которые реализуются с помощью инструмента IXIA PerfectStorm, обеспечивая как нормальное, так и вредоносное поведение. Датасет содержит около 2,5 миллионов записей сетевых соединений, каждая из которых характеризуется 49 различными признаками.

Данные представлены в формате CSV и включают разнообразные характеристики сетевого трафика:

- Характеристики потока (IP-адреса, порты, используемые протоколы).
- Основные функции (количество пакетов, байты от источника к получателю).
- Характеристики контента (флаги TCP, размер полезной нагрузки).
- Временные функции (интервалы между пакетами, запись времени начала).

Каждая запись имеет метку класса, которая указывает на принадлежность к одному из двух классов:

- label = 0 – нормальный (легитимный) сетевой трафик;
- label = 1 – вредоносная активность (атака).

Особенно важно, что данные детально размечены: каждое соединение отнесено к одной из девяти категорий атак или классифицировано как нормальный трафик.

Категории атак, которые определяются на основе используемых признаков набора данных UNSW-NB15:

- Fuzzers – атаки с использованием фаззеров (программ для тестирования на уязвимости).
- Analysis – атаки, направленные на сбор информации о системе (например, порт-сканирование).
- Backdoor – скрытый доступ к системе, минуя стандартные механизмы аутентификации.
- DoS (Denial of Service) – атаки отказа в обслуживании.
- Exploits – использование уязвимостей в ПО или протоколах.
- Generic – универсальные атаки, такие как взлом шифрования.
- Reconnaissance – разведывательная активность для выявления уязвимостей.
- Shellcode – внедрение и исполнение вредоносного кода.
- Worms – самораспространяющиеся вредоносные программы [18].

Для улучшения качества обучения моделей был проведён предварительный анализ корреляции признаков в датасете UNSW-NB15. Расчёт коэффициентов корреляции Пирсона выявил умеренную зависимость между временными и транспортными характеристиками трафика, такими как dur, Sload, Dload и sttl. Эти связи указывают на наличие дублирующей информации в некоторых признаках, что подчёркивает важность отбора признаков для повышения точности классификации.

При предобработке данных использовались стандартные методы: нормализация значений, кодирование категориальных переменных с помощью one-hot encoding и удаление выбросов с помощью межквартильного размаха (IQR). Для решения проблемы дисбаланса классов применялся метод синтетического увеличения выборки (SMOTE). Этот подход позволил улучшить полноту (recall) на 3-5 %, не ухудшая общую точность модели.

Характеристика признаков. Признаки делятся на количественные, которые могут быть измерены в числовых значениях (например, длительность соединения, количество байт, объем персональных данных, количество пакетов) и категориальные, которые представляют собой категории или группы (например, тип протокола (TCP, UDP), направление трафика (входящий/исходящий), статус соединения (открыто/закрыто) и другое.

Для обучения моделей использовались 47 признаков, представленных в табл. 3.

Таблица 3

Список используемых признаков набора данных UNSW-NB15

№	Признак	Описание	№	Признак	Описание
1	srcip	IP-адрес отправителя	25	trans_depth	Глубина транзакции (например, для HTTP)
2	sport	Порт отправителя	26	res_bdy_len	Длина тела ответа
3	dstip	IP-адрес получателя	27	Sjit	Джиттер (вариация задержки) отправителя
4	dsport	Порт получателя	28	Djit	Джиттер (вариация задержки) получателя
5	proto	Протокол (TCP, UDP, ICMP и др.)	29	Stime	Время начала соединения
6	state	Состояние соединения (например, EST, FIN, RST)	30	Ltime	Время окончания соединения
7	dur	Длительность потока (в секундах)	31	Sintpkt	Среднее время между пакетами отправителя
8	sbytes	Количество байт от отправителя	32	Dintpkt	Среднее время между пакетами получателя
9	dbytes	Количество байт от получателя	33	tcprtt	Время установки TCP-соединения (RTT)
10	sttl	Время жизни (TTL) пакета отправителя	34	synack	Время между SYN и SYN-ACK
11	dttl	Время жизни (TTL) пакета получателя	35	ackdat	Время между ACK и данными
12	sloss	Потеря пакетов от отправителя	36	is_sm_ips_ports	Флаг (0/1), указывает, совпадают ли IP и порт отправителя и получателя
13	dloss	Потеря пакетов от получателя	37	ct_state_ttl	Количество уникальных состояний TTL
14	service	Сетевой сервис (HTTP, FTP, SSH и др.)	38	ct_flw_http_mthd	Количество HTTP-методов в потоке
15	Sload	Скорость передачи данных отправителя (бит/с)	39	is_ftp_login	Флаг (0/1), указывает на FTP-логин
16	Dload	Скорость передачи данных получателя (бит/с)	40	ct_ftp_cmd	Количество FTP-команд в потоке
17	Spkts	Количество пакетов от отправителя	41	ct_srv_src	Количество соединений от одного источника к одному сервису
18	Dpkts	Количество пакетов от получателя	42	ct_srv_dst	Количество соединений к одному сервису от разных источников
19	swin	Размер окна отправителя	43	ct_dst_ltm	Количество соединений к одному получателю
20	dwin	Размер окна получателя	44	ct_src_ltm	Количество соединений от одного отправителя
21	stcpb	Размер буфера TCP отправителя	45	ct_src_dport_ltm	Количество соединений от одного отправителя на один порт
22	dtcpb	Размер буфера TCP получателя	46	ct_dst_sport_ltm	Количество соединений к одному получателю с одного порта
23	smeansz	Средний размер пакета отправителя	47	ct_dst_src_ltm	Количество соединений между парой (отправитель-получатель)
24	dmeansz	Средний размер пакета получателя			

Данные для экспериментов в UNSW-NB15 представляются в обучающей выборке (training set) – 175,341 записей и тестовой выборке (testing set) – 82,332 записей.

Реализация и обучение модели с помощью ансамблевого алгоритма Random Forest. В процессе обучения модели были задействованы сведения из двух наборов UNSW-NB15: UNSW_NB15_training_set и UNSW_NB15_testing_set. Это позволило увеличить объём доступных данных и обеспечить более репрезентативную тренировку модели.

Для начала все категориальные признаки были преобразованы в количественные при помощи one-hot кодирования. Затем объединённый набор данных был случайным образом разделён на обучающую и тестовую выборки в соотношении 80:20. Для построения модели использовался базовый классификатор RandomForestClassifier из библиотеки scikit-learn, настроенный на 100 деревьев решений.

Обучение модели заняло 50.48 секунд, что демонстрирует её высокую вычислительную эффективность по сравнению с более ресурсоёмкими нейросетевыми подходами. После завершения обучения модель была протестирована на ранее отложенной тестовой выборке. Результаты классификации приведены в табл. 4. Они показали высокие значения метрик: точность модели составила 95%, точность положительных срабатываний (precision) для атак – 96%, а полнота (recall) – также 96%. Это говорит о высокой способности модели точно определять вредоносную активность и минимизировать как ложноотрицательные, так и ложноположительные срабатывания. Матрица ошибок для алгоритма Random Forest представлена на рис. 1.

Таблица 4

Результат работы алгоритма Random Forest

	Precision	Recall	F1-score	Support
0	0.93	0.93	0.93	18675
1	0.96	0.96	0.96	32860
Accuracy	–	–	0.95	51535
Macro avg	0.95	0.95	0.95	51535
Weighted avg	0.95	0.95	0.95	51535

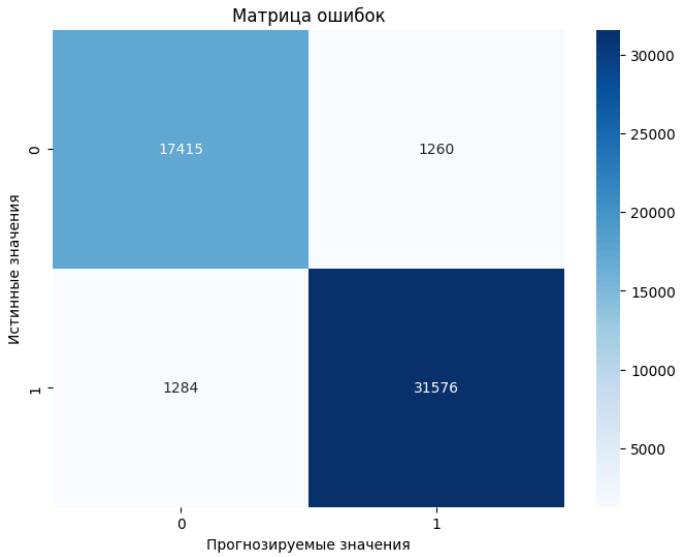

Рис. 1. Матрица ошибок для алгоритма Random Forest

Главным преимуществом Random Forest является возможность анализа важности признаков. Модель автоматически определяет, какие переменные оказывают наибольшее влияние на результат классификации. Анализ feature_importances_ показал, что среди наиболее значимых признаков выявлены:

- Время жизни (TTL) пакета отправителя.
- Количество уникальных состояний TTL.
- Скорость передачи данных отправителя.
- Количество байт от отправителя.
- Количество соединений к одному сервису от разных источников.
- Время установки TCP-соединения (RTT).
- Количество байт от получателя.
- Среднее время передачи пакетов.
- Скорость передачи данных получателя.
- Время между SYN и SYN-ACK.

Как видно на рис. 2, визуальное представление важности этих признаков подтверждает данный список, где такие метрики, как sttl, ct_state_ttl и sload, являются одними из наиболее влиятельных.

Таким образом, модель Random Forest показала высокую точность и устойчивость, оставаясь интерпретируемой и относительно простой в обучении. Её применение целесообразно в условиях, когда важно добиться высокой скорости классификации при минимальных вычислительных затратах.

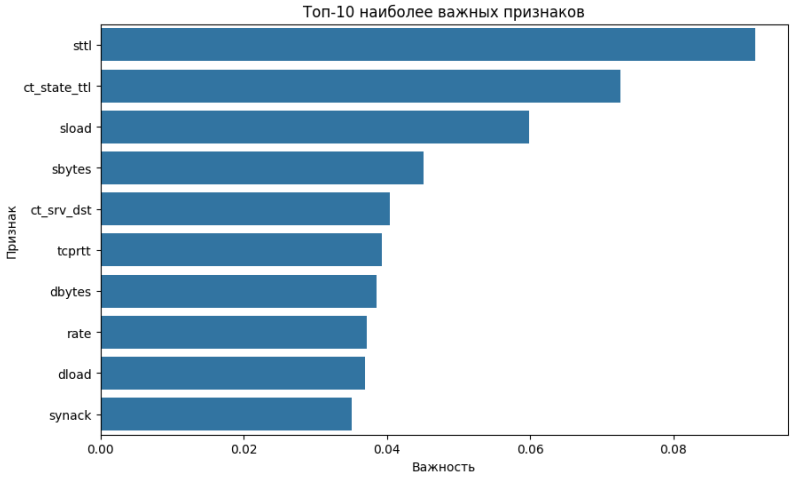

Рис. 2. Диаграмма наиболее важных признаков алгоритма

Реализация и обучение нейронной сети для обнаружения кибервторжений. В рамках данного исследования была разработана модель многослойного персептрона (MLP), предназначенная для бинарной классификации сетевого трафика с целью обнаружения вредоносной активности.

Аналогично, в процессе обучения модели были использованы UNSW_NB15_training_set и UNSW_NB15_testing_set, и все категориальные признаки были преобразованы в количественные при помощи one-hot кодирования. Далее все атрибуты были приведены к единому масштабу с использованием стандартизации. А для борьбы с дисбалансом классов, характерным для задач информационной безопасности, автоматически рассчитаны веса классов.

Нейросеть была реализована с использованием библиотеки Keras через API Sequential. Архитектура включает четыре последовательно соединённых полносвязных слоя.

Входной слой содержит 64 нейрона и использует функцию активации ReLU (Rectified Linear Unit), которая передаёт на выход только положительные значения, эффективно устраняя проблему затухающего градиента и ускоряя обучение. После него применяются пакетная нормализация и механизм Dropout с вероятностью 0.5 – это регуляризационная техника, при которой на каждом шаге обучения случайным образом "отключается" часть нейронов, что предотвращает переобучение модели.

Во втором скрытом слое используется 32 нейрона, ReLU-активация, L2-регуляризация с коэффициентом 0.01 для ограничения роста весов и Dropout с вероятностью 0.3.

Третий слой включает 16 нейронов и Dropout с вероятностью 0.2.

Завершается архитектура выходным слоем с одним нейроном, который преобразует выход в значение от 0 до 1, интерпретируемое как вероятность принадлежности к одному из двух классов – нормальному трафику или атаке. Визуальное представление архитектуры нейронной сети изображено на рис. 3.

Для оптимизации модели использовался метод эффективной стохастической оптимизации Adam, с адаптивной скоростью обучения 0.0001. Он разработан таким образом, что объединяет преимущества методов AdaGrad (Duchietal., 2011), который хорошо работает с разреженными диапазонами, и RMSProp (Tieleman & Hinton, 2012), действующий в сетевых и нестационарных условиях. Этот оптимизатор автоматически адаптирует момент и размер шага для каждого параметра, так как он ограничен гиперпараметром stepsize, что особенно эффективно при работе с зашумленными или разреженными данными [19].

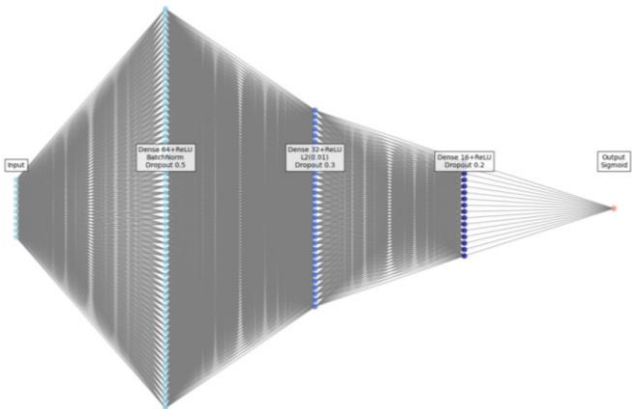

Рис. 3. Детализированная архитектура нейронной сети

Для обнаружения кибератак, распознавания сетевых угроз, выявления хакерских проникновений, идентификации попыток взлома и определения признаков несанкционированного доступа применялась бинарная кросс-энтропия. Эта функция потерь, оптимальная для задач бинарной классификации, вычисляется как среднее значение логарифмических потерь между предсказанными вероятностями и истинными метками. Она строго штрафует модель за уверенные, но ошибочные предсказания.

Для комплексной оценки качества модели использовались четыре ключевые метрики. Accuracy (общая точность), показывающая долю правильных предсказаний среди всех примеров, Precision (точность положительных срабатываний), Recall (полнота), оценивающая способность модели обнаруживать реальные атаки, помогая избежать опасных пропусков угроз. Метрика AUC-ROC, отражающая площадь под ROC-кривой, демонстрирующая общую способность модели различать классы независимо от выбранного порога классификации, где значение 1 соответствует идеальному разделению, а 0.5 – случайным догадкам.

Для предотвращения переобучения была реализована стратегия ранней остановки, основанная на отслеживании функции потерь на валидационной выборке. Если в течение пяти эпох подряд не наблюдалось улучшения значения val_loss, обучение прерывалось, а модель возвращалась к наилучшим весам.

Обучение выполнялось на протяжении не более 30 эпох при размере пакета 32 батча. Благодаря использованию весов классов и стратегии EarlyStopping, модель демонстрировала устойчивую сходимость и хорошую обобщающую способность.

Обучение производилось в течение 20 эпох и заняло 564.55 секунд, что является высоким показателем эффективности модели. На обучающей выборке были достигнуты следующие результаты: точность – 94.29%, полнота – 93.28%, точность положительных предсказаний – 96.27%, площадь под кривой (AUC) – 0.9871. На датасете UNSW_NB15_testing-set модель продемонстрировала точность 88.73%, AUC – 0.9830, точность (precision) – 0.9869, полноту (recall) – 0.8456, и значение функции потерь val_loss – 0.3935. Более наглядно результаты представлены в табл. 5.

Таблица 5

Результат обучения модели многослойного персептрона (MLP)

	Precision	Recall	F1-score	Support
0	0.75	0.98	0.85	56000
1	0.99	0.85	0.91	119341
Accuracy	–	–	0.89	175341
Macro avg	0.87	0.91	0.88	175341
Weighted avg	0.91	0.89	0.89	175341

Оценка модели проводилась как с использованием стандартного отчёта classification_report, так и с расчётом площади под ROC-кривой (AUC). ROC-кривая (Receiver Operating Characteristic) представляет собой двумерный график, отображающий соотношение между долей истинно положительных результатов изображенной на оси Y и долей ложноположительных результатов – по оси X, при различных порогах классификации. Чтобы сравнить классификаторы, мы можем свести производительность ROC к одному скалярному значению, представляющему ожидаемую производительность. Распространённым методом является вычисление площади под кривой ROC, сокращённо AUC. Поскольку AUC – это часть площади единичного квадрата, её значение всегда будет находиться в диапазоне от 0 до 1,0, и чем оно выше, тем лучше модель различает классы [20].

В ходе исследования, значение ROC-кривой у данной модели составляет – 0.9826, что подтверждает высокую способность модели отличать нормальный трафик от вредоносного.

Данные результаты можно представить в виде матрицы ошибок, а также графиков изменения точности и функции потерь на обучающей и тестовой выборках, в зависимости от количества эпох, которые представлены на рис. 4 и 5 соответственно, что наглядно демонстрирует стабильное поведение модели и ее применимость в задачах обнаружения сетевых атак.

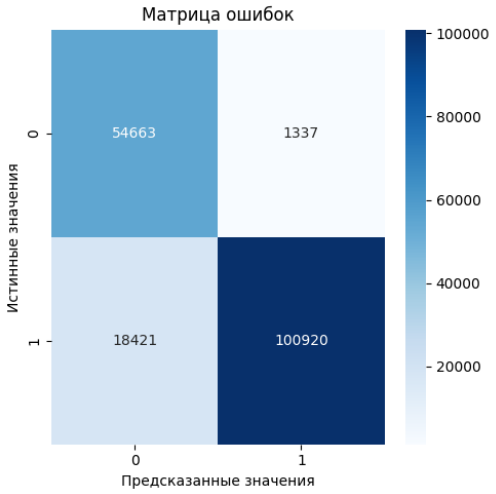

Рис. 4. Матрица ошибок реализуемого метода

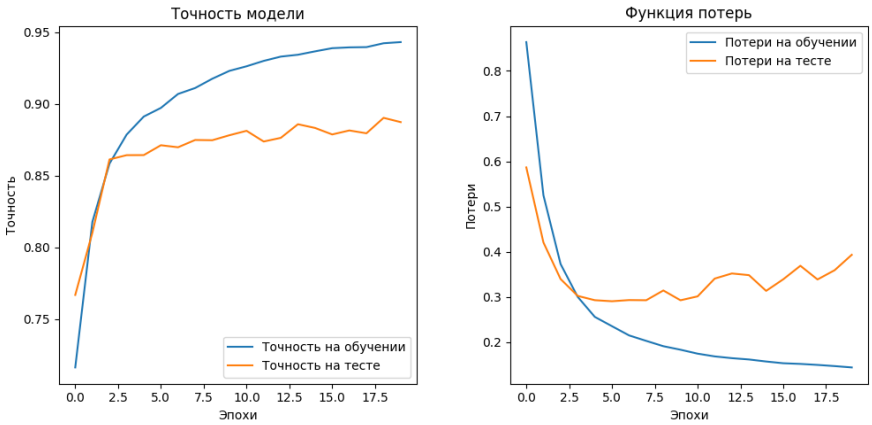

Рис. 5. Графики точности модели и функции потерь

Результаты и обсуждение. Проведенное исследование показало высокую эффективность двух реализованных моделей – многослойного персептрона (MLP) и Random Forest – в обнаружении сетевых атак на основе набора данных UNSW-NB15.

Многослойный персептрон продемонстрировал отличные результаты в задачах бинарной классификации. Значение AUC составило 0,9830, что говорит о его отличной способности различать нормальный трафик и атаки. Также модель показала высокую точность предсказаний (precision – 0,9869), что означает минимальное количество ложных срабатываний. Однако у этой модели есть и недостатки: значительные вычислительные затраты, поэтому обучение нейронной сети требует много ресурсов и времени, что может ограничить её применение в системах реального времени.

Модель Random Forest, в отличие от MLP, показала сопоставимое качество классификации (accuracy – 95%) при значительно меньших вычислительных затратах. Более того, Random Forest отличается лучшей интерпретируемостью: можно оценить важность признаков, что помогает аналитикам понять, какие параметры сетевого трафика наиболее важны для обнаружения атак.

Сравнение результатов с данными из других исследований подтверждает, что разработанные модели показывают сопоставимое или более высокое качество классификации. Например, в [8] точность Random Forest составила 92,7 %, а в [13] нейронная сеть LSTM достигла AUC = 0,975 на датасете CICIDS2018. В нашем исследовании показатели Random Forest и MLP оказались выше на 2-3 %, что указывает на высокую эффективность использованной методики предварительной обработки и отбора признаков.

Предложенные решения могут быть интегрированы в системы мониторинга безопасности SIEM. В таких системах алгоритмы машинного обучения автоматически выявляют подозрительные события и коррелируют сетевые журналы. Модели, применяемые в модулях предиктивного анализа, не только фиксируют факты вторжений, но и прогнозируют потенциальные угрозы, анализируя паттерны поведения пользователей и сетевые взаимодействия.

Заключение. Таким образом, MLP стоит использовать, когда критически важна максимальная чувствительность к атакам (например, для защиты высоконагруженных критических инфраструктур). При этом доступные вычислительные ресурсы могут компенсировать длительное время обучения.

Random Forest – оптимальный выбор для быстрого развёртывания и работы в условиях ограниченных ресурсов. Также его стоит использовать, когда требуется объяснимость результатов для последующего анализа и принятия решений.

Выбор между этими моделями нужно делать, исходя из конкретных требований к задаче в сфере кибербезопасности. Необходимо найти баланс между точностью, скоростью работы и интерпретируемостью. Обе модели показали свою эффективность, но для разных сценариев их применения.

Разработанные модели можно адаптировать для работы в реальном времени при интеграции с потоковыми платформами анализа данных, такими как Apache Kafka или Flink. Это обеспечит оперативное обнаружение вторжений с минимальной задержкой и позволит системе гибко реагировать на новые типы угроз.

Перспективным направлением развития является использование технологий федеративного обучения (Federated Learning), которые позволяют обучать модели на распределённых данных без их передачи. Это гарантирует защиту конфиденциальной информации, сохраняя при этом высокую эффективность обучения.

Также в дальнейших исследованиях можно рассмотреть другие модели, такие как LSTM, Дерево решений (Decision Tree), Extra Trees и Градиентный бустинг. На основе разработанных моделей по обнаружению кибервторжений можно реализовать самостоятельное программное обеспечение, которое будет анализировать существующий трафик и помогать пользователям узнать о появившихся атаках.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Чипига А.Ф., Пелешенко В.С.* Формализация процедур обнаружения и предотвращения сетевых атак // Информационное противодействие угрозам терроризма. – 2006. – № 8. – С. 156-163.
2. *Усков Е.Д., Корепанова Н.Л.* Анализ информативных признаков аномалий сетевого трафика корпоративных сетей // Современные инновации. – 2019. – № 3 (31). – С. 13-16.
3. *Аброськина Е.С.* Анализ методов выявления сетевых вторжений и аномалий // Экономика и социум. – 2021. – № 3-2 (82). – С. 688-698.
4. *Чаругин В.В., Чесалин* А.Н. Анализ и формирование наборов данных сетевого трафика для обнаружения компьютерных атак // International Journal of Open Information Technologies. – 2023. – С. 100-105.
5. *Исратова Е.Е.* Применение нейронных сетей для обнаружения аномального трафика в сетях Интернета вещей // International Journal of Open Information Technologies. – 2024. – С. 65-69.
6. *Гайфулина Д.А., Котенко И.В.* Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий интернета вещей // Информационно-управляющие системы. – 2021. – № 1 (110). – С. 28-37.
7. *Chastikova V.A., Sotnikov V.V.* Method of analyzing computer traffic based on recurrent neural networks // Journal of Physics: Conference Series. International Conference "High-Tech and Innovations in Research and Manufacturing," HIRM 2019. – 2019. – P. 012133.
8. *Кажемский М.А., Шелухин О.И.* Многоклассовая классификация сетевых атак на информационные ресурсы методами машинного обучения // Тр. учебных заведений связи. – 2019. – Т. 5, № 1. – С. 107-115. – DOI: 10.31854/1813-324X-2019-5-1-107-115.
9. *Саматов М.А.* Анализ эффективности IDS/IPS систем на базе Suricata в обеспечении сетевой кибербезопасности // Вестник науки. – 2024. – Т. 2, № 12. – С. 1352-1363.
10. *Болодурина И.П., Нефедов Д.А.* Применение большой языковой модели для уменьшения ложнопозитивных срабатываний в задачах выявления аномалий в сетевом трафике // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2024. – Т. 24, № 4. – С. 5-15. – DOI: 10.14529/ctcr240401.
11. *Частикова В.А., Жерлицын С.А., Воля Я.И., Сотников В.В.* Нейросетевая технология обнаружения аномального сетевого трафика // Прикаспийский журнал: управление и высокие технологии. – 2020. – № 1 (49). – С. 20-32.
12. *Гетьман А.И., Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А.* Методика сбора обучающего набора данных для модели обнаружения компьютерных атак // Тр. ИСП РАН. – 2021. – Т. 33, № 5. – С. 83-104. – DOI: 10.15514/ISPRAS-2021-33(5)-5.
13. *Chastikova V.A., Zherlitsyn S.A., Volya Y.I., Sotnikov V.V.* Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic // IOP Conference Series: Materials Science and Engineering. Krasnoyarsk Science and Technology City Hall., Krasnoyarsk, Russian Federation. – 2021. – P. 12135.
14. *Chastikova V.A., Mitugov A.I.* The method for detecting network attacks based on the neuroimmune approach // Journal of Physics: Conference Series. Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. Krasnoyarsk, Russia. – 2021. – P. 32035.
15. KDD Cup 1999 Data. – URL: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (дата обращения: 10.04.2025).
16. NSL-KDD. – URL: https://www.kaggle.com/datasets/hassan06/nslkdd (дата обращения: 10.04.2025).

17. CSE-CIC-IDS2018. – URL: https://fkie-cad.github.io/COMIDDS/content/datasets/cse_cic_ids2018/ (дата обращения: 10.04.2025).
18. UNSW-NB15 Network Intrusion Detection Dataset. // Fraunhofer FKIE. – URL: https://fkie-cad.github.io/COMIDDS/content/datasets/unsw_nb15/ (дата обращения: 14.04.2025).
19. *Kingma D.P., Ba J. Adam.* A Method for Stochastic Optimization // The 3rd International Conference for Learning Representations. – San Diego, 2015. – P. 1-15.
20. *Fawcett T.* An Introduction to ROC Analysis // Pattern Recognition Letters. – 2006. – Vol. 27, No. 8. – P. 861-874. – DOI: 10.1016/j.patrec.2005.10.010.

REFERENCES

1. *Chipiga A.F., Peleshenko V.S.* Formalizatsiya protsedur obnaruzheniya i predotvrashcheniya setevykh atak [Formalization of procedures for detecting and preventing network attacks], *Informatsionnoe protivodeystvie ugrozam terrorizma* [Information counteraction to terrorist threats], 2006, No. 8, pp. 156-163.
2. *Uskov E.D., Korepanova N.L.* Analiz informativnykh priznakov anomaliy setevogo trafika korporativnykh setey [Analysis of informative signs of anomalies in corporate network traffic], *Sovremennye innovatsii* [Modern innovations], 2019, No. 3 (31), pp. 13-16.
3. *Abros'kina E.S.* Analiz metodov vyyavleniya setevykh vtorzheniy i anomaliy [Analysis of methods for detecting network intrusions and anomalies], *Ekonomika i sotsium* [Economics and society], 2021, No. 3-2 (82), pp. 688-698.
4. *Charugin V.V., Chesalin A.N.* Analiz i formirovanie naborov dannykh setevogo trafika dlya obnaruzheniya komp'yuternykh atak [Analysis and formation of network traffic data sets for detecting computer attacks], *International Journal of Open Information Technologies*, 2023, pp. 100-105.
5. *Isratova E.E.* Primenenie neyronnykh setey dlya obnaruzheniya anomal'nogo trafika v setyakh Interneta veshchey [Application of neural networks to detect abnormal traffic in Internet of Things networks], *International Journal of Open Information Technologies*, 2024, pp. 65-69.
6. *Gayfulina D.A., Kotenko I.V.* Analiz modeley glubokogo obucheniya dlya zadach obnaruzheniya setevykh anomaliy interneta veshchey [Analysis of deep learning models for the detection of network anomalies of the Internet of Things], *Informatsionno-upravlyayushchie sistemy* [Information and Control Systems], 2021, No. 1 (110), pp. 28-37.
7. *Chastikova V.A., Sotnikov V.V.* Method of analyzing computer traffic based on recurrent neural networks, *Journal of Physics: Conference Series. International Conference "High-Tech and Innovations in Research and Manufacturing," HIRM 2019*, 2019, pp. 012133.
8. *Kazhemskiy M.A., Shelukhin O.I.* Mnogoklassovaya klassifikatsiya setevykh atak na informatsionnye resursy metodami mashinnogo obucheniya [Multiclass classification of network attacks on information resources by machine learning methods], *Tr. uchebnykh zavedeniy svyazi* [Proceedings of educational institutions of communication], 2019, Vol. 5, No. 1, pp. 107-115. DOI: 10.31854/1813-324X-2019-5-1-107-115.
9. *Samatov M.A.* Analiz effektivnosti IDS/IPS sistem na baze Suricata v obespechenii setevoy kiberbezopasnosti [Analysis of the effectiveness of IDS/IPS systems based on Suricata in ensuring network cybersecurity], *Vestnik nauki* [Bulletin of Science], 2024, Vol. 2, No. 12, pp. 1352-1363.
10. *Bolodurina I.P., Nefedov D.A.* Primenenie bol'shoy yazykovoy modeli dlya umen'sheniya lozhnopozitivnykh srabatyvaniy v zadachakh vyyavleniya anomaliy v setevom trafike [The use of a large language model to reduce false positives in problems of detecting anomalies in network traffic], *Vestnik YuUrGU. Seriya «Komp'yuternye tekhnologii, upravlenie, radioelektronika»* [Bulletin of SUSU. The series "Computer technology, control, radio electronics"], 2024, Vol. 24, No. 4, pp. 5-15. DOI: 10.14529/ctcr240401.
11. *Chastikova V.A., Zherlitsyn S.A., Volya Ya.I., Sotnikov V.V.* Neyrosetevaya tekhnologiya obnaruzheniya anomal'nogo setevogo trafika [Neural network technology for detecting abnormal network traffic], *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Management and High Technologies], 2020, No. 1 (49), pp. 20-32.
12. *Get'man A.I., Goryunov M.N., Matskevich A.G., Rybolovlev D.A.* Metodika sbora obuchayushchego nabora dannykh dlya modeli obnaruzheniya komp'yuternykh atak [A methodology for collecting a training dataset for a computer attack detection model], *Tr. ISP RAN* [Proceedings of the ISP RAS], 2021, Vol. 33, No. 5, pp. 83-104. DOI: 10.15514/ISPRAS-2021-33(5)-5.
13. *Chastikova V.A., Zherlitsyn S.A., Volya Y.I., Sotnikov V.V.* Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic, *IOP Conference Series: Materials Science and Engineering. Krasnoyarsk Science and Technology City Hall., Krasnoyarsk, Russian Federation*, 2021, pp. 12135.

14. *Chastikova V.A., Mitugov A.I.* The method for detecting network attacks based on the neuroimmune approach, *Journal of Physics: Conference Series. Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. Krasnoyarsk, Russia*, 2021, pp. 32035.
15. KDD Cup 1999 Data. Available at: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed 10 April 2025).
16. NSL-KDD. Available at: https://www.kaggle.com/datasets/hassan06/nslkdd (accessed 10 April 2025).
17. CSE-CIC-IDS2018. Available at: https://fkie-cad.github.io/COMIDDS/content/datasets/ cse_cic_ids2018/ (accessed 10 April 2025).
18. UNSW-NB15 Network Intrusion Detection Dataset. // Fraunhofer FKIE. Available at: https://fkie-cad.github.io/COMIDDS/content/datasets/unsw_nb15/ (accessed 10 April 2025).
19. *Kingma D.P., Ba J. Adam.* A Method for Stochastic Optimization, *The 3rd International Conference for Learning Representations*. San Diego, 2015, pp. 1-15.
20. *Fawcett T.* An Introduction to ROC Analysis, *Pattern Recognition Letters,* 2006, Vol. 27, No. 8, pp. 861-874. DOI: 10.1016/j.patrec.2005.10.010.

Частикова Вера Аркадьевна – Кубанский государственный технологический университет; e-mail: chastikova_va@mail.ru; г. Краснодар, Россия; тел.: +79184635536; к.т.н.; доцент.

Козачёк Константин Валериевич – Кубанский государственный технологический университет; e-mail: Koza4ek.Konstantin@yandex.ru; г. Краснодар, Россия; тел.: +79182345367; аспирант.

Коробская Екатерина Сергеевна – Кубанский государственный технологический университет; e-mail: kate9.korobskaya@mail.ru; г. Краснодар, Россия; тел.: +79286059807; студент.

Кравцов Владислав Павлович – Кубанский государственный технологический университет; e-mail: vlad.kravtsov.1980@mail.ru; г. Краснодар, Россия; тел.: +79121585302; студент.

Chastikova Vera Arkadyevna – Kuban State Technological University; e-mail: chastikova_va@mail.ru; Krasnodar, Russia; phone: +79184635536; cand. of eng. sc.; associate professor.

Kozachek Konstantin Valerievich – Kuban State Technological University; e-mail: Koza4ek.Konstantin@yandex.ru; Krasnodar, Russia; phone: +79182345367; postgraduate student.

Korobskaya Ekaterina Sergeevna – Kuban State Technological University; e-mail: kate9.korobskaya@mail.ru; Krasnodar, Russia; phone: +79286059807; student.

Kravtsov Vladislav Pavlovich – Kuban State Technological University; e-mail: vlad.kravtsov.1980@mail.ru; Krasnodar, Russia; phone: +79121585302; student.

УДК 004.032.26 DOI 10.18522/2311-3103-2025-5-243-254

А.С. Коваленко, Я.М. Демяненко

МЕТОД ГЕНЕРАЦИИ ШУМА ПО НАБОРУ ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЙ БЕЗ ЧИСТЫХ ПРИМЕРОВ

Предлагается новый метод генерации шума по зашумленным изображениям без необходимости использования выровненных пар чистых и зашумленных данных. В отличие от традиционных подходов, требующих наличия согласованных наборов изображений или априорных моделей шума, разрабатываемый метод позволяет моделировать сложные характеристики шума, присущие конкретным КМОП-сенсорам, основываясь исключительно на наблюдаемых зашумленных данных. Для синтеза шума используется генеративно-состязательная архитектура U-Net-подобного типа, построенная на базе StyleGANv2 с модифицированным дискриминатором, учитывающим параметры камеры и исходных изображений. Основное внимание уделяется сохранению пространственно-цветовой структуры изображения при генерации шума, что достигается введением специализированной функции потерь, сохраняющей характеристики цветопередачи и текстурных деталей. Предлагаемый подход позволяет обучать генератор шума в условиях полного отсутствия пар чистых и зашумленных изображений, что особенно актуально при работе с реальными данными, полученными с различных камер и в различных условиях освещения. В экспе-