УДК 303.734

Системы обнаружения вторжений на основе искусственного интеллекта: обнаружение аномалий для современной киберзащиты

Ляпунцова Елена Вячеславовна[1] **lev86@bmstu.ru**
SPIN-код: 3500-4160

Арм Ажи Азиз Салих[2] **arm.azhi@yandex.com**

[1] *МГТУ им Н.Э. Баумана, Москва, Россия*

[2] *НИТУ МИСиС, Москва, Россия*

Аннотация. *Растущая сложность кибератак выявила ограничения традиционных систем обнаружения вторжений* (IDS), *основанных на правилах, которые часто не справляются с эксплойтами нулевого дня,* DDoS-*атаками и развивающимся вредоносным ПО. В данном исследовании представлена усовершенствованная на основе ИИ система обнаружения вторжений* (IDS), *сочетающая в себе методы машинного и глубокого обучения для обнаружения аномалий в сетевом трафике. Оценка с использованием набора данных* CICIDS2017 *показала, что предлагаемая система демонстрирует более высокую точность обнаружения и более низкий уровень ложных срабатываний по сравнению с традиционными подходами к* IDS. *Результаты исследования подчеркивают практические преимущества обнаружения аномалий на основе ИИ для повышения кибербезопасности в телекоммуникационных сетях, государственных инфраструктурах и корпоративных средах.*

***Ключевые слова*:** *сетевая безопасность, система обнаружения вторжений, искусственный интеллект, обнаружение аномалий, глубокое обучение, машинное обучение*

AI-Enhanced intrusion detection systems: anomaly detection for modern cyber defense

Lyapuntsova Elena Vyacheslavovna[1] **lev86@bmstu.ru**
SPIN-code: 3500-4160

Arm Azhi Aziz Salih[2] **arm.azhi@yandex.com**

[1] *BMSTU, Moscow, Russia*

[2] *NUST MISIS, Moscow, Russia*

Abstract. The growing sophistication of cyberattacks has exposed the limitations of traditional rule-based intrusion detection systems (IDS), which often fail against zero-day exploits, DDoS attacks, and evolving malware. This study presents an AI-enhanced IDS that combines machine learning and deep learning techniques to detect anomalies in network traffic. Evaluated using the CICIDS2017 dataset, the proposed system demonstrates higher detection accuracy and lower false alarm rates compared to conventional IDS approaches. The findings highlight the practical benefits of AI-driven anomaly detection for enhancing cybersecurity in telecom networks, government infrastructures, and enterprise environments.

Keywords: network security, intrusion detection system, artificial intelligence, anomaly detection, deep learning, machine learning

Введение. Угрозы кибербезопасности стремительно усложняются — от низкоскоростных DDoS и ботнетов до инсайдерских атак, обходящих сигнатурные IDS (Snort, Suricata). Для повышения эффективности все чаще применяются ИИ-системы, способные изучать сложные паттерны трафика, выявлять скрытые аномалии и адаптироваться к новым угрозам. Исследования показывают, что алгоритмы МО (случайный лес, SVM) и глубокие модели (LSTM, автоэнкодеры) эффективно фиксируют временные и нелинейные зависимости с высокой точностью [1].

В данном исследовании предлагается гибридная IDS, объединяющая методы машинного и глубокого обучения для обнаружения аномалий в реальном времени с минимальными ложными срабатываниями. Система ориентирована на применение в критически важных средах — телеком-сегменте, корпоративных сетях и государственных структурах.

Методы и материалы; результаты. В исследовании использовался набор данных CICIDS2017, содержащий маркированный сетевой трафик с различными типами атак, включая сканирование портов, подбор паролей, DDoS-атаки, ботнеты и инфильтрацию, который служил эталоном для обнаружения аномалий [2]. Предварительная обработка данных включала нормализацию, категориальное кодирование и снижение размерности с помощью PCA.

Система обнаружения вторжений (рис. 1) включает четыре уровня: сбор данных, предварительная обработка, ИИ-ядро и принятие решений [3]. Ядро объединяет методы МО (случайный лес, XGBoost) с ГО (LSTM, автокодировщики). Гибридный ансамбль использует взвешенное голосование, сочетая сильные стороны моделей: случайный лес — для многомерных данных, LSTM — для последовательностей, автокодировщики — для редких аномалий.

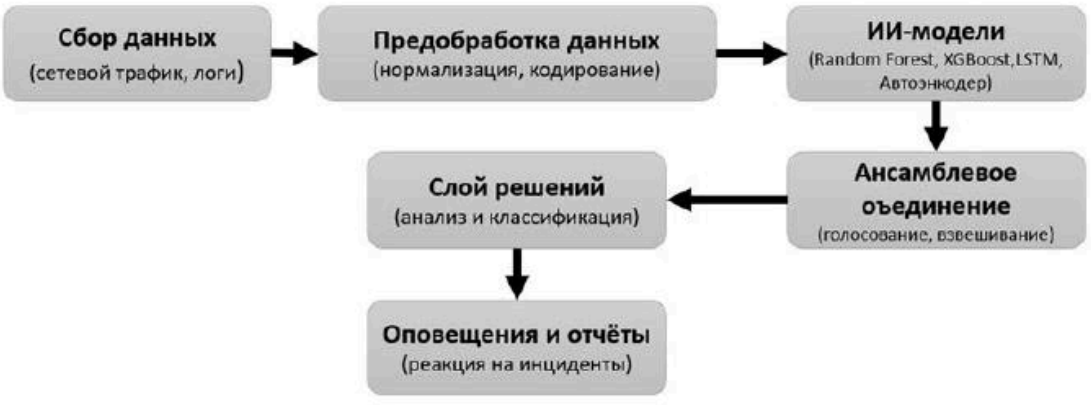

Рис. 1. Архитектура ИИ-усиленной системы IDS

Алгоритм работы ИИ-усиленной IDS (упрощенная версия):

– сбор данных и извлечение сетевых потоков, пакетов и журналов с маршрутизаторов, серверов и рабочих процессов;

– предобработка и очистка ошибок, нормализация числовых признаков, кодирование категориальных результатов, уменьшение размерности (PCA);

– обучающие модели и тренировка, случайная Forest и XGBoost (ML) и LSTM, автоэнкодера (DL);

– ансамблирование и объединение моделей предсказаний методом голосования; большинство → «атака»;

– принятие решения и расчет вероятности аномалии; верхние порога → предупреждение;

– оповещения и администратора и формирование отчета с типом воздействия и определением времени.

Традиционная сигнатурная IDS достигла 86,2 % точности при 12,4 % ложных срабатываний, что подтвердило ограничения статических правил. Модели машинного обучения улучшили результаты [4, 5]: случайный лес — 94,6 %, XGBoost — 95,4 %. Глубокие модели показали еще более высокие показатели: LSTM — 96,1 %, автоэнкодеры — 92,3 %. Наилучших результатов достигла гибридная ансамблевая модель, объединившая МО и ГО: точность — 97,5 %, прецизионность — 96,9 %, полнота — 96,2 %, F1 — 96,5 %. Уровень ложных срабатываний снизился до 2,8 %, значительно превзойдя как традиционные IDS, так и отдельные алгоритмы (см. таблицу).

Сравнение ИИ-усиленных и базовых систем эффективности

Модель	Точность (Accuracy), %	Точность (Precision), %	Отзывать (Recall), %	F1-оценка (F1-Score), %	ФПР (FPR), %
Фырканье (базовый уровень)	86,2	84,1	82,5	83,3	12,4
Случайный лес	94,6	93,8	92,7	93,2	5,2
XGBoost	95,4	94,9	93,1	94,0	4,9
LSTM	96,1	95,7	94,6	95,1	4,1
Автоэнкодер	92,3	90,5	89,4	89,9	6,7
Ансамбль (Гибрид)	97,5	96,9	96,2	96,5	2,8

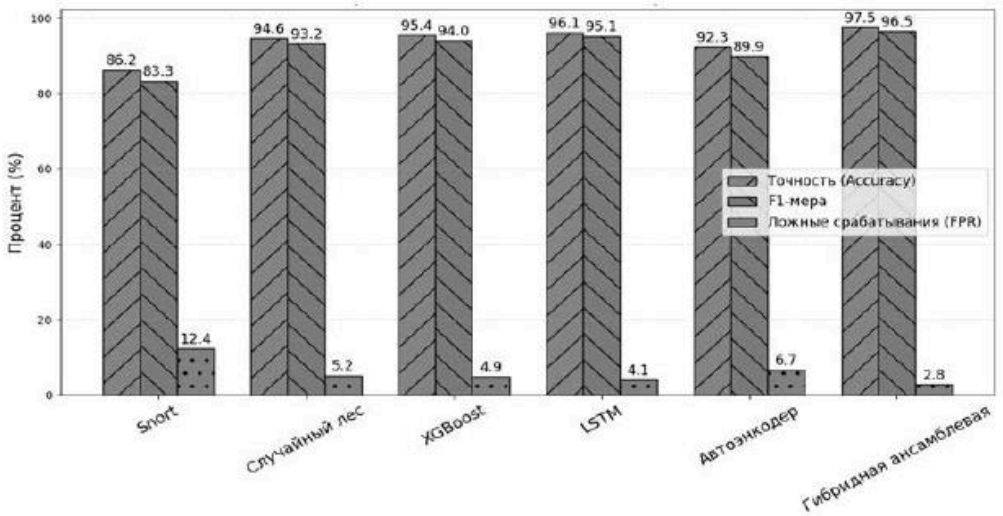

Рис. 2. Сравнение моделей IDS по метрикам

Сравнение моделей IDS по метрикам эффективности визуализировано на рис. 2, где представлена столбчатая диаграмма точности и F1-оценка для различных моделей. Важным наблюдением также является снижение количества ложных срабатываний, что демонстрирует, как ансамблевый метод минимизировал количество ложных тревог, создающих нагрузку на аналитиков безопасности.

Заключение. Системы обнаружения вторжений (IDS) на базе искусственного интеллекта (ИИ) превосходят традиционные системы, основанные на сигнатурах, обнаруживая эксплойты нулевого дня и трудноуловимые аномалии. Гибридные ансамбли адаптируются к меняющемуся трафику, обеспечивая обнаружение телекоммуникационных атак в режиме реального времени, защиту правительственных сетей и мониторинг корпоративного трафика. Среди проблем — интерпретируемость, вычислительные затраты и состязательные атаки. В будущем планируется внедрение объяснимого ИИ, федеративного обучения и безопасного протоколирования на основе блокчейна.

Список источников

[1] Wang W., Zhu M., Wang J., Zeng X., Yang Z. End-to-end encrypted traffic classification with one-dimensional convolution neural networks. IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, pp. 43–48. https://doi.org/10.1109/ISI.2017.8004872

[2] Sharafaldin I., Habibi Lashkari A., Ghorbani A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018, pp. 108–116. https://doi.org/10.5220/0006639801080116

[3] Shone N., Ngoc T. N., Phai V. D., Shi Q. A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, vol. 2, no. 1, pp. 41–50. https://doi.org/10.1109/tetci.2017.2772792

[4] Kim J., Kim J., Thu H.L.T., Kim H. Long short term memory recurrent neural network classifier for intrusion detection. International Conference on Platform Technology and Service (PlatCon), 2016, pp. 1–5. https://doi.org/10.1109/PlatCon.2016.7456805

[5] Liu H., Lang B., Liu M., Yan H. CNN and RNN based payload classification methods for attack detection. Knowledge-Based Systems, 2019, vol. 163, pp. 332–341. https://doi.org/10.1016/j.knosys.2018.10.003