УДК 004.89

Березницкий А.С., к.э.н., доцент кафедры вычислительной математики и кибернетики факультета математики и компьютерных наук имени профессора Н.И. Червякова,
ФГАОУ ВО «Северо-Кавказский федеральный университет»
Россия, г. Ставрополь
Водской Д. О.
студент
6 курс специалитета, направление «Компьютерная безопасность»
Факультет математики и компьютерных наук имени профессора Н.И.Червякова
ФГАОУ ВО «Северо-Кавказский федеральный университет»
Россия, г. Ставрополь

ГИБРИДНЫЙ МЕТОД ОБНАРУЖЕНИЯ АНОМАЛИЙ В КОРПОРАТИВНОМ СЕТЕВОМ ТРАФИКЕ НА ОСНОВЕ ИНТЕГРАЦИИ СИГНАТУРНОГО АНАЛИЗА И МАШИННОГО ОБУЧЕНИЯ

Аннотация: Статья посвящена решению проблемы низкой эффективности классических сигнатурных систем IDS/IPS против современных целевых атак (APT) и угроз «нулевого дня». Предлагается гибридный метод анализа сетевого трафика, интегрирующий сигнатурный подход и поведенческий анализ на базе машинного обучения. Проведен анализ ограничений фреймворков Snort и Suricata. Основное внимание уделено описанию разработанного программного модуля на Python, реализующего ансамбль моделей ML (Random Forest, Isolation Forest, K-Means) для анализа PCAP-файлов и генерации структурированных оповещений. Результатом является работоспособный прототип, преодолевающий ключевые недостатки традиционных решений за счет комбинированного подхода.

Ключевые слова: Кибербезопасность, сетевой трафик, обнаружение аномалий, машинное обучение, гибридный метод, сигнатурный анализ, Suricata.

Abstract: This article addresses the low effectiveness of traditional signature-based IDS/IPS systems against advanced targeted attacks (APTs) and zero-day threats. A hybrid network traffic analysis method is proposed, integrating a signature-based approach and behavioral analysis based on machine learning. The limitations of the Snort and Suricata frameworks are analyzed. The focus is on describing the developed Python software module, which implements an ensemble of ML models (Random Forest, Isolation Forest, and K-Means) for analyzing PCAP files and generating structured alerts. The result is a viable prototype that overcomes the key shortcomings of traditional solutions through a combined approach.

Keywords: Cybersecurity, network traffic, anomaly detection, machine learning, hybrid method, signature analysis, Suricata.

Эволюция киберугроз в сторону сложных целевых атак (Advanced Persistent Threats, APT), использующих легитимные инструменты (Living-off-the-Land) и методы обфускации, обусловила кризис классических средств защиты, основанных на сигнатурном анализе. Межсетевые экраны и системы обнаружения вторжений (IDS) демонстрируют низкую эффективность против неизвестных (zero-day) угроз и генерируют высокий уровень ложных срабатываний [1]. В качестве ответа на этот вызов индустрия развивает технологии анализа сетевого трафика (Network Traffic Analysis, NTA) и сетевого обнаружения и реагирования (Network Detection and Response, NDR), основанные на поведенческом анализе и машинном обучении (МО) [2].

Целью данной работы является разработка и верификация гибридного метода обнаружения аномалий, который интегрирует надежность сигнатурного подхода для известных угроз и проактивные возможности МО для выявления новых аномалий.

Нормальное функционирование сети характеризуется предсказуемыми статистическими паттернами (цикличность, стабильный состав протоколов, устойчивая матрица коммуникаций), которые формируют поведенческий базис для обнаружения отклонений [3]. Архитектурные уязвимости сетевой инфраструктуры на всех уровнях OSI создают обширную поверхность для атак, особенно в гибридных и облачных средах.

Сравнительный анализ ведущих открытых фреймворков Snort и Suricata выявил фундаментальные ограничения сигнатурной парадигмы:

– Слепота к неизвестным угрозам: эффективность напрямую зависит от актуальности базы правил;

– Высокий уровень ложных срабатываний (False Positives): приводит к «усталости алертов» у аналитиков SOC;

– Отсутствие поведенческого анализа и корреляции: неспособность выявлять сложные многоэтапные атаки;

– Ресурсоемкость глубокой инспекции (DPI) в реальном времени.

При этом Suricata, благодаря многопоточной архитектуре и структурированному формату вывода EVE-JSON, представляет собой оптимальную основу для интеграции с системами МО, выступая в роли источника качественной сетевой телеметрии [4].

Разработка модуля осуществлялась на языке Python 3.8+, что обусловлено его доминированием в области Data Science и Machine Learning, богатой экосистемой библиотек и высокой скоростью прототипирования. Для реализации алгоритмов МО был выбран фреймворк scikit-learn, предоставляющий оптимизированные, проверенные реализации необходимых алгоритмов: RandomForestClassifier, IsolationForest и MiniBatchKMeans. Исследования подтверждают высокую эффективность Random Forest (точность до 98.5%) и Isolation Forest в задачах классификации и обнаружения аномалий трафика соответственно [5].

Для обработки сетевых пакетов и анализа PCAP-файлов использовалась библиотека Scapy, которая предоставляет беспрецедентную гибкость для

низкоуровневого парсинга и создания пакетов. Обработка данных осуществлялась с помощью Pandas и NumPy. Такой технологический стек обеспечивает баланс между производительностью, зрелостью решений и соответствием современным исследовательским и индустриальным трендам в области NDR.

Ключевые компоненты и их взаимодействие разработанного модуля отражены на Рисунке 1.

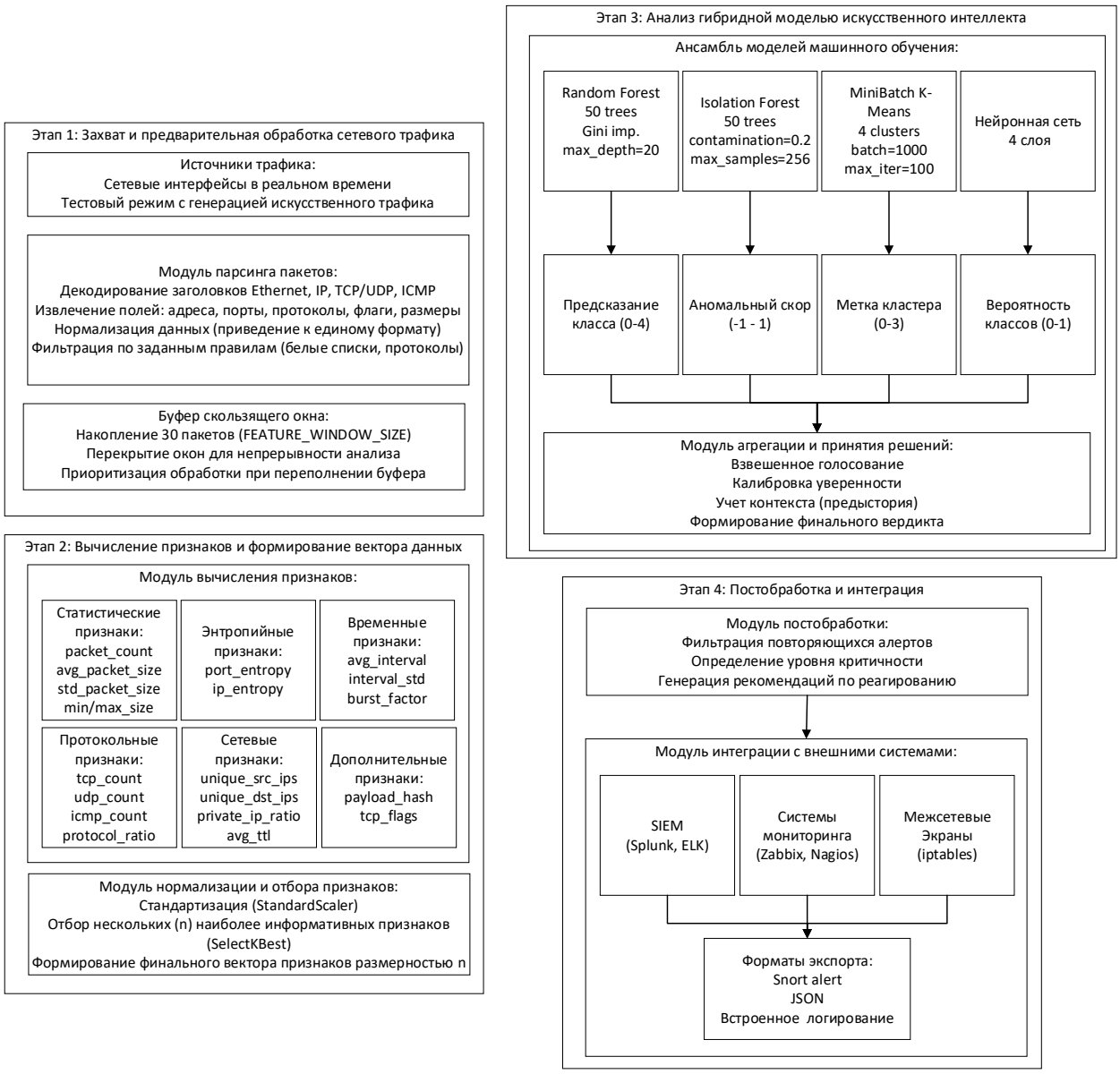

Рисунок 1 – Ключевые компоненты и их взаимодействие

Система работает по конвейерному принципу:

– Конфигурация (Config): централизованное управление параметрами системы (пути, пороги, настройки моделей) через загрузку из JSON-файла.

– Извлечение признаков (PCAPFeatureExtractor): загрузка PCAP-файлов с поддержкой семплирования (параметр SAMPLE_RATE) для работы с большими дампами. На основе захваченных пакетов идентифицируются сетевые потоки (flow), для каждого из которых вычисляется вектор из более чем 55 признаков. Признаки включают временные характеристики (длительность, межпакетные интервалы), объемные метрики (байты/пакеты), статистические показатели (среднее, дисперсия размеров) и флаговую информацию (соотношение SYN, ACK и др.).

– Гибридная модель ИИ (HybridAIModel): ядро системы. Реализует ансамблевый подход:

– Random Forest: контролируемая классификация трафика на категории (нормальный, DDoS, сканирование и т.д.). Обучается на размеченных датасетах (CIC-IDS2017, UNSW-NB15).

– Isolation Forest: бесконтрольное обнаружение аномалий путем изоляции выбросов в пространстве признаков. Эффективен против неизвестных угроз.

– MiniBatch K-Means: кластеризация потоков для выявления скрытых паттернов и группировки схожих аномалий.

Модели работают параллельно. Их предсказания агрегируются с помощью взвешенного голосования и эвристических правил, что позволяет получить итоговую классификацию и оценку уверенности.

Принятие решений и генерация алертов (NetworkAIDetector): Оркестратор системы. На основе результатов гибридной модели и дополнительных эвристических проверок (например, на аномально высокую скорость пакетов или долю SYN-флагов) генерирует структурированные оповещения. Каждый алерт содержит уникальный ID, временную метку, уровень угрозы (Info, Low, Medium, High, Critical), тип трафика,

предположение о характере атаки, IP-адреса и порты, а также рекомендации по реагированию.

Модуль поддерживает два основных режима работы через командный интерфейс (CLI):

Режим обучения: python detector.py --train-csv cicids2017.csv,unsw-nb15.csv --model-name my_model.joblib

Модуль загружает указанные CSV-файлы, автоматически определяет столбец с метками, выполняет предобработку, обучение ансамбля моделей и сохраняет артефакты (scaler, feature selector, модели) в файл.

Режим анализа: python detector.py --load-model models/my_model.joblib --analyze-pcap suspicious_traffic.pcap --output results.csv --show-alerts 10

Модуль загружает обученную модель, извлекает признаки из PCAP-файла, выполняет детектирование и сохраняет подробные результаты в CSV. В консоль выводится статистика и список наиболее критичных алертов.

Ключевые особенности реализации: обработка больших данных, устойчивость к ошибкам, интегрируемость, интерпретируемость.

В статье представлен рабочий прототип гибридной системы обнаружения аномалий сетевого трафика, который реализует комбинированный подход к кибербезопасности. Интеграция сигнатурных принципов (через использование обученного Random Forest на исторических данных) и методов бесконтрольного обучения (Isolation Forest, кластеризация) позволяет адекватно реагировать как на известные, так и на новые, ранее не встречавшиеся угрозы.

Использованные источники:

1. What is Network Detection and Response (NDR)? [Электронный ресурс] // IBM. - 2023. - URL: https://www.ibm.com/think/topics/ndr (дата обращения: 25.01.2026).

2. What Is Network Traffic Analysis (NTA)? [Электронный ресурс] // Cisco. - URL: https://www.cisco.com/site/us/en/learn/topics/security/what-is-network-traffic-analysis.html (дата обращения: 25.01.2026).

3. What is Network Traffic Analysis (NTA)? [Электронный ресурс] / Mesh Flinders, Ian Smalley // IBM Think. - 2025. - URL: https://www.ibm.com/think/topics/network-traffic-analysis (дата обращения: 25.01.2026).

4. 15.1.2. Eve JSON Format [Электронный ресурс] // Suricata User Guide. - URL: https://docs.suricata.io/en/latest/output/eve/eve-json-format.html (дата обращения: 25.01.2026).

5. Machine Learning for Network Intrusion Detection - A Comparative Study // Future Internet. - 2023. - Vol. 15, No. 7. - Art. 243. - DOI: 10.3390/fi15070243.