ТЕХНИКАЛЫҚ ҒЫЛЫМДАР
ТЕХНИЧЕСКИЕ НАУКИ
TECHNICAL SCIENCE

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. №4 (32), Том 2. С. 299–303
DOI: https://doi.org/10.5281/zenodo.24200406
Сайт журнала: https://cajournal.kz

УДК 37

ГИБРИДНЫЙ ПОДХОД К ОБНАРУЖЕНИЮ АНОМАЛЬНОЙ СЕТЕВОЙ АКТИВНОСТИ НА ОСНОВЕ АНСАМБЛЕВЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ, ОПТИМИЗИРОВАННОГО ОТБОРА ПРИЗНАКОВ И МЕТОДОВ XAI

П.А. Рурак, В.А. Ситников, И.А. Сарсенгалиев
КазНИТУ им. К. И. Сатпаева, Алматы, Казахстан

Рурак Полина Андреевна — Студент 4 курса, кафедра Кибербезовасности, КазНИТУ им. К. И. Сатпаева, Алматы, Казахстан.

Ситников Виталий Александрович — Студент 4 курса, кафедра Кибербезовасности, КазНИТУ им. К. И. Сатпаева, Алматы, Казахстан.

Сарсенгалиев Илья Айтбекович — Студент 4 курса, кафедра Кибербезовасности, КазНИТУ им. К. И. Сатпаева, Алматы, Казахстан.

© П.А. Рурак, В.А. Ситников, И.А. Сарсенгалиев, 2026

Аннотация: В статье рассматривается гибридный подход к автоматизированному обнаружению аномальной сетевой активности в корпоративных сетях. Предложена концептуальная модель программного модуля AI-IDS Engine, объединяющего модуль предварительной обработки и отбора признаков, ядро обнаружения угроз на основе ансамблевых алгоритмов машинного обучения (Random Forest / XGBoost) и модуль интерпретации результатов на базе методов объяснимого искусственного интеллекта (SHAP). Проведён многофакторный эксперимент по плану 2^3 на наборе данных UNSW-NB15. Исследовано влияние архитектуры детектора, состава признаков и профиля трафика на показатель F1-score. Экспериментально показано преимущество Random Forest над простым правилом ($\Delta F1 = 0{,}1739$), а также влияние расширения набора признаков на баланс Precision и Recall. Полученные результаты подтверждают целесообразность гибридной архитектуры, сочетающей высокую полноту обнаружения атак с возможностью интерпретации решений модели.

Ключевые слова: обнаружение аномалий, машинное обучение, Random Forest, отбор признаков, XAI, SHAP, UNSW-NB15, факторный эксперимент, SIEM, кибербезопасность.

1. Введение

Современные корпоративные информационные системы генерируют значительные объёмы событий безопасности, поступающих от сетевого оборудования, серверов, рабочих станций и средств защиты. Ручная обработка такого потока затруднена, а традиционные сигнатурные методы ограничены в обнаружении новых и нестандартных угроз. В связи с этим активно развиваются технологии искусственного интеллекта и машинного обучения, позволяющие автоматизировать анализ событий, выявлять отклонения от нормального поведения и оценивать вероятность угрозы.

Несмотря на высокую точность современных ML- и DL-моделей, их практическое внедрение в контур SOC сталкивается с рядом проблем: компромиссом между точностью и вычислительной эффективностью, низкой интерпретируемостью «чёрных ящиков», чувствительностью к дисбалансу классов, концептуальным дрейфом данных и уязвимостью

к состязательным воздействиям. Кроме того, эффективность детекции существенно зависит от состава входных признаков.

Целью настоящей работы является разработка и экспериментальная оценка гибридного подхода к обнаружению аномальной сетевой активности, объединяющего ансамблевые модели машинного обучения, адаптивный отбор информативных признаков и методы объяснимого ИИ (XAI). В рамках исследования проведён полный факторный эксперимент 2^3, позволивший оценить влияние архитектуры детектора, набора признаков и профиля трафика на качество классификации.

2. Анализ предметной области и смежных исследований

2.1. Современные SIEM- и NDR-платформы

Анализ существующих коммерческих решений (Splunk Enterprise Security, Microsoft Sentinel, Google Security Operations, Elastic Security, Vectra AI, Exabeam, CrowdStrike Falcon) показывает, что большинство флагманов рынка проектируются как облачные (Cloud-Native/SaaS) платформы. Google SecOps демонстрирует наивысшие показатели производительности: до 70 ТБ/сутки и порядка 35 000 EPS на базовый юнит. Microsoft Sentinel и Splunk обеспечивают 50 и 45 ТБ/сутки соответственно. При этом все ведущие платформы интегрируют генеративный ИИ (Security Copilot, Gemini, Agent Builder) для автоматизации расследований и написания детектов на естественном языке.

Ключевыми ограничениями коммерческих решений являются высокая стоимость, зависимость от облачной инфраструктуры конкретного вендора, недостаточная прозрачность алгоритмов принятия решений и сложность адаптации под локальные требования организаций. Это обосновывает разработку гибридной системы с возможностью локального развёртывания и встроенными механизмами объяснения результатов.

2.2. Классическое машинное обучение

Классические алгоритмы (Random Forest, Decision Tree, SVM, Gradient Boosting) остаются востребованными благодаря высокой скорости работы и относительно низкой вычислительной сложности. В работе Damanik и Anggraeni Random Forest на агрегированном сетевом трафике достиг Accuracy = 94,1 % и F1-score ≈ 94 %. Однако декомпозиция по классам атак показывает снижение эффективности на скрытых и распределённых угрозах (APT, low-and-slow brute-force). Основные ограничения: предположение i.i.d., чувствительность к дисбалансу классов, неспособность обнаруживать zero-day атаки и уязвимость к adversarial-воздействиям.

2.3. Глубокое обучение и генеративный ИИ

Глубокие нейросетевые архитектуры (LSTM, GRU, CNN, Autoencoders, Transformers) способны автоматически извлекать иерархические признаки и моделировать временные зависимости. LLM применяются для формирования естественно-языковых описаний инцидентов и сопоставления с матрицей MITRE ATT&CK. Главные барьеры внедрения — высокие вычислительные затраты, задержки инференса и проблема «чёрного ящика».

2.4. Объяснимый искусственный интеллект (XAI)

Методы SHAP и LIME позволяют ранжировать вклад признаков в решение модели. Структурные подходы (XNN) и LLM-объяснения расширяют возможности интерпретации. Для SOC-оператора критически важно получать не только вердикт, но и обоснование, какие именно характеристики сетевого события привели к классификации его как аномалии.

3. Предлагаемый подход

3.1. Объект, предмет и гипотеза исследования

Объектом исследования являются процессы автоматизированного мониторинга, выявления и анализа событий информационной безопасности в корпоративных сетях.

Предметом — разработка моделей, алгоритмов и программного обеспечения для автоматизированного анализа сетевых событий с использованием гибридных подходов машинного обучения и методов XAI.

Гипотеза исследования: применение гибридного подхода, объединяющего алгоритмы машинного обучения для обнаружения аномальной сетевой активности с методами выбора информативных признаков и средствами объяснимого искусственного интеллекта, позволит повысить качество обнаружения угроз и одновременно обеспечить интерпретируемость результатов при приемлемых вычислительных затратах.

3.2. Концептуальная модель AI-IDS Engine

Предлагаемый программный модуль гибридного обнаружения и локальной интерпретации аномальной сетевой активности (AI-IDS Engine) включает три функциональных блока:

Модуль предварительной обработки данных и отбора признаков (Preprocessing and Feature Selection). Выполняет подготовку сетевых событий, выделение характеристик соединений и формирование набора признаков.

Модуль обнаружения угроз (Detection Core). Выявляет подозрительную сетевую активность с использованием сигнатурных правил или ансамблевых ML-моделей (Random Forest / XGBoost + Autoencoder).

Модуль интерпретации результатов (XAI Engine). Объясняет решения модели путём определения признаков, оказавших влияние на результат классификации (метод SHAP).

3.3. Факторы эксперимента

Для проведения многофакторного эксперимента (полный факторный план 2^3) выбраны три управляемых входных фактора:

Фактор A (X_1) — Архитектура детектора: нижний уровень (−1) — сигнатурно-правиловый детектор; верхний уровень (+1) — ансамблевая ML-модель (Random Forest / XGBoost).

Фактор B (X_2) — Набор входных признаков: нижний уровень (−1) — базовый набор (адреса, порты, объём данных, продолжительность); верхний уровень (+1) — оптимизированный набор, дополненный статистическими и поведенческими характеристиками (частота событий, статистика TCP-флагов, энтропия).

Фактор C (X_3) — Условия сетевого трафика: нижний уровень (−1) — нормальный профиль; верхний уровень (+1) — зашумлённый / пиковый трафик.

В качестве основного выходного параметра выбран F1-score. Дополнительно фиксируются FPR, Latency и индекс интерпретируемости XAI.

4. Экспериментальная часть

4.1. Данные и методика

Эксперименты выполнены на открытом наборе данных UNSW-NB15. Обучение моделей проводилось на 175 341 записи, оценка качества — на тестовой выборке из 82 332 записей. Целевой столбец label обозначает нормальный трафик (0) и атаку (1). В качестве базового алгоритма классификации использован Random Forest.

4.2. Сравнение наборов признаков

Базовый набор включал 10 признаков, расширенный — 42 признака. Результаты сравнения представлены в таблице 1.

Таблица 1 – Результаты сравнения двух наборов признаков

Метрика	Базовый набор (10 признаков)	Расширенный набор (42 признака)
Accuracy	0,8772	0,8712

Precision	0,8364	0,8183
Recall	0,9659	0,9848
F1-score	0,8965	0,8939

Базовый набор обеспечил более высокие значения Accuracy, Precision и F1-score. Расширенный набор показал более высокий Recall (0,9848 против 0,9659), что указывает на способность обнаруживать большую долю атакующих соединений, однако сопровождался снижением Precision и ростом ложных срабатываний.

4.3. Сравнение правил и машинного обучения

Сравнение простого правила (на основе признака ct_state_ttl) и Random Forest на одной тестовой выборке дало следующие результаты (таблица 2).

Таблица 2 – Сравнение качества обнаружения

Метрика	Простое правило	Random Forest	Разница
Accuracy	0,7164	0,8712	0,1548
Precision	0,7888	0,8183	0,0295
Recall	0,6622	0,9848	0,3226
F1-score	0,7200	0,8939	0,1739

Random Forest превзошёл простое правило по всем метрикам. Наиболее значительное улучшение наблюдалось по полноте обнаружения (ΔRecall = 0,3226).

4.4. Анализ ошибок классификации

В тестовой выборке содержалось 37 000 записей нормального трафика и 45 332 записи атак. Модель правильно классифицировала 27 085 нормальных соединений и 44 644 атаки. Доля ложных срабатываний (FPR) составила 26,80 %. Доля обнаруженных атак (Recall) — 98,48 %. Результаты эксперимента показывают, что модель обладает высокой способностью обнаруживать атаки, однако допускает значительное количество ложных срабатываний на нормальном трафике.

4.5. Регрессионная модель полного факторного эксперимента

Для описания зависимости выходного параметра от факторов использована модель:

$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3 + \varepsilon$, где X_1, X_2, $X_3 \in \{-1, +1\}$ — кодированные значения факторов; b_0 — среднее значение отклика; b_1, b_2, b_3 — линейные эффекты; b_{12}, b_{13}, b_{23}, b_{123} — эффекты взаимодействий; ε — случайная ошибка. Выбор модели основан на методических рекомендациях NIST/SEMATECH.

5. Обсуждение результатов

Экспериментальные данные подтверждают, что ансамблевые ML-модели существенно превосходят простые сигнатурные правила по качеству обнаружения атак на наборе UNSW-NB15. При этом расширение набора признаков не всегда приводит к улучшению всех метрик одновременно: рост Recall сопровождается снижением Precision. Выбор набора признаков должен определяться приоритетами системы. Если критично не пропускать атаки, предпочтительным выступает расширенный набор; если важным является баланс и снижение ложных срабатываний — базовый набор показывает лучший F1-score.

Высокий FPR (26,80 %) указывает на необходимость дальнейшей настройки порога классификации и, возможно, применения методов балансировки классов или калибровки вероятностей. Модуль XAI (SHAP) позволяет компенсировать проблему интерпретируемости, предоставляя оператору SOC объяснение вклада отдельных признаков в вердикт модели.

Гибридная архитектура (быстрый первичный фильтр на Random Forest + активируемый по событию XAI-модуль) обеспечивает компромисс между скоростью обработки и прозрачностью решений, что соответствует требованиям нормативных документов РК (СТ РК ISO/IEC 27001, Приказ МИК РК № 112) к минимальному времени реакции на инциденты.

6. Заключение

В работе предложен и экспериментально исследован гибридный подход к обнаружению аномальной сетевой активности, объединяющий ансамблевые модели машинного обучения, оптимизированный отбор признаков и методы объяснимого искусственного интеллекта. На данных UNSW-NB15 Random Forest продемонстрировал F1-score = 0,8939, превзойдя простое правило на 0,1739. Базовый набор из 10 признаков обеспечил F1-score = 0,8965, расширенный (42 признака) — 0,8939 при более высоком Recall.

Полученные результаты подтверждают гипотезу исследования: гибридный подход позволяет повысить качество обнаружения угроз при сохранении возможности интерпретации результатов. Дальнейшие направления работы включают полный расчёт коэффициентов регрессионной модели 2^3, исследование влияния порога классификации, оценку вычислительной сложности и Latency в условиях потоковой обработки, а также интеграцию механизмов защиты от состязательных атак.

Список литературы:

1. Козлова Н. Ш., Довгаль В. А. Анализ применения искусственного интеллекта и машинного обучения в кибербезопасности // Вестник Адыгейского государственного университета. – 2023. – Вып. 3 (326). – С. 65–72.

2. Damanik H. A., Anggraeni M. Experimental Analysis of Random Forest-Based Classification for Adaptive Network Attack Detection // ICAR-T1 2025. – Kuala Lumpur, 2026. – P. 128–145.

3. Бекматов А. К. Глубокое обучение для повышения эффективности обнаружения вторжений в сетевых системах // Экономика и социум. – 2024. – № 5(120). – С. 1977–1982.

4. Ичетовкин Е. А. Исследование устойчивости систем обнаружения вторжений с компонентами машинного обучения к состязательным атакам // Вестник Астраханского государственного технического университета. – 2025. – С. 76–87.

5. Ankalaki S. et al. Cyber Attack Prediction: From Traditional Machine Learning to Generative Artificial Intelligence // IEEE Access. – 2025. – Vol. 13. – P. 44662–44706.

6. Сымагулов А. и др. Методы интерпретации черных ящиков машинного обучения и их применение для создания систем поддержки принятия решений // Academic Scientific Journal of Computer Science. – 2021. – № 5. – С. 91–99.

7. Karim M. R. et al. Interpreting Black-box Machine Learning Models for High Dimensional Datasets // IEEE DSAA. – Thessaloniki, 2023. – P. 1–10.

8. Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. – Munich, 2019. – 317 p.

9. Machine Learning for Cybersecurity: Threat Detection and Mitigation / ed. by A. Elhanashi, P. Dini. – Basel, 2024. – 308 p.

10. Kucharavy A. et al. Large Language Models in Cybersecurity. Threats, Exposure and Mitigation. – Zurich, 2024. – 247 p.

11. Приказ Министра информации и коммуникаций Республики Казахстан от 20 марта 2018 года № 112 «Об утверждении Правил проведения мониторинга обеспечения информационной безопасности...».

12. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems // Military Communications and Information Systems Conference. – 2015.