Средняя оценка качества работы моделей МО после обучения на наборе, состоящем из 3222 наблюдений, в котором 50% наблюдений являются аномалиями, достигает в F1-score 0,85 (табл. 1).

Сравнение результатов обнаружения вредоносного трафика моделью Isolation Forest и WAF показывает, что доли ложноположительных предсказаний (FPR) составили 8% и 0%, а истинно положительных (TPR) – 82% и 29% соответственно.

В результате вычислительного эксперимента наилучший результат в классификации аномального трафика среди моделей машинного обучения показала модель на основе алгоритма Isolation Forest (0,86 F1), так как модель LOF (0,86 F1) характеризуется высокой долей ложноположительных срабатываний (FP). Кроме того, по сравнению с Mod_Security, Isolation Forest выявила аномалии на 53% эффективнее.

Результаты экспериментов подтверждают, что интеграция методов машинного обучения в процесс анализа журналов WAF способствует более эффективному обнаружению аномалий. Разработанную систему предлагается интегрировать с межсетевым экраном веб-приложений Дальнейшее развитие этого направления связано с использованием более сложных моделей анализа и интеграцией с другими источниками событий, связанных с безопасностью. Это повысит устойчивость и адаптивность систем безопасности к современным киберугрозам.

Список использованных источников

1. Shaheed A., Kurdy M.H.D.B. Web application firewall using machine learning and features engineering / Security and Communication Networks, 2022, Т. 2022, № 1, С. 5280158.
2. Applebaum S., Gaber T., Ahmed A. Signature-based and machine-learning-based web application firewalls: A short survey / Procedia Computer Science, 2021, Т. 189, С. 359-367.

Атарская Елена Андреевна, аспирант каф. вычислительной техники и защиты информации, atarskaya.ea@ugatu.su.
Вульфин Алексей Михайлович, д.т.н., профессор каф. вычислительной техники и защиты информации, vulfin.alexey@gmail.com.
Кириллова Анастасия Дмитриевна, к.т.н., доцент каф. вычислительной техники и защиты информации, kirillova.ad@ugatu.su.

УДК 004.056, 004.75:004.8

СИСТЕМА ОБНАРУЖЕНИЯ СЕТЕВЫХ АТАК В ПРОМЫШЛЕННОЙ СЕТИ

А.М. Вульфин, А.Д. Кириллова, Э.Е. Садыкова
Уфимский университет науки и технологий, г. Уфа

Ключевые слова: сетевой трафик, алгоритмы объяснимого искусственного интеллекта, большие языковые модели

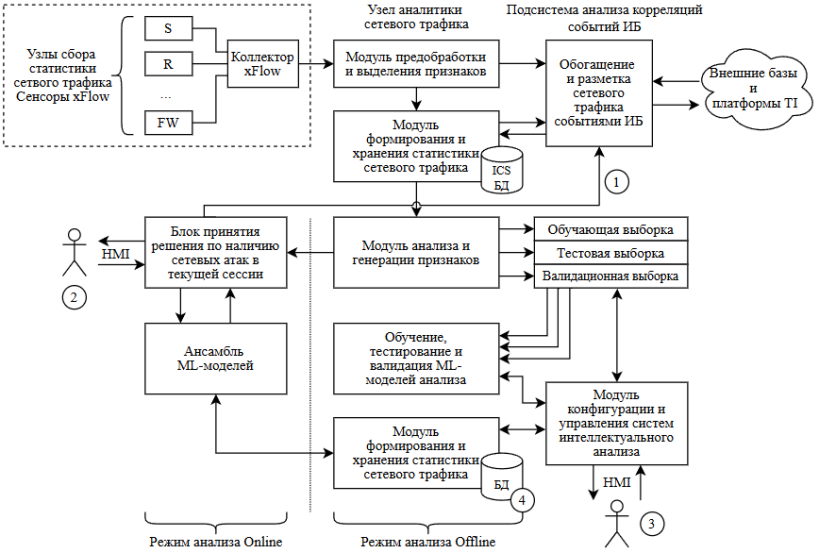
Разработка систем безопасности промышленных сетей и мониторинга

сетевого трафика является приоритетной задачей. Для промышленного оборудования, пограничных систем и точек входа в промышленную сеть необходимо обеспечить эффективный анализ трафика на наличие признаков сетевых атак и сетевых аномалий с поддержкой анализа промышленных протоколов. Совершенствование средств защиты сетевой инфраструктуры направлено на развитие инструментов интеллектуального мониторинга трафика и состояния объектов и узлов промышленной сети.

Целью работы является повышение эффективности обнаружения сетевых атак в промышленной сети АСУ ТП с помощью алгоритмов объяснимого искусственного интеллекта.

Структурная схема предлагаемой системы анализа сетевого трафика в промышленной сети представлена на рисунке 1.

В работе используется набор данных WUSTL-IIOT-2021 [1], который содержит сетевые сессии с 46 числовыми и категориальными признаками (метаданные).



1 – передача результатов анализа в систему SIEM/SOC; 2 – специалист по сетевой безопасности, DevOps инженер; 3 – специалист по интеллектуальному анализу данных; 4 – база обученных ML-модулей анализа сетевого трафика

Рисунок 1 – Структурная схема предлагаемой системы анализа сетевого трафика

В ходе анализа метаданные сетевого трафика загружаются и проходят предварительную обработку, включающую обработку категориальных и числовых признаков, а также устранение пропусков и шумов. Далее набор данных разделяется на обучающую и тестовую выборки в соотношении

80% на 20%. Обучающая выборка содержит 764456 примеров нормального трафика и 191115 записей атак. Тестовая выборка включает 238893 записей нормального и аномального трафика. На тестовой выборке производится оптимизация гиперпараметров с помощью Optuna [2], качество обучения классификатора CatBoost [3] оценивается по набору метрик качества. Результаты интерпретируются с использованием метода LIME [4], с помощью которого признаки ранжируются по степени значимости. Для выбранного сегмента данных вычисляются локальные веса LIME, после чего формируется текстовый запрос в большую языковую модель (БЯМ), которая на основе полученных значений признаков и их весов генерирует текстовое объяснение результатов.

Вычислительный эксперимент показывает, что подобранные гиперпараметры обеспечивают высокую эффективность классификатора на тестовой выборке (~1,0 F1, ~1,0 Precision, ~1,0 Recall).

Первичный анализ значимости признаков с помощью классификатора показывает наибольшее влияние на решение модели параметров DIntPkt, SrcAddr, num_sTtl и num_DstPkts. Метод SHAP выделяет признак SrcAddr, заметное влияние DIntPkt и sTtl, остальные признаки оказывают значительно меньшее воздействие.

Локальная интерпретация результатов с использованием метода LIME показывает, что наибольший вклад в решение модели вносят признаки sTtl и SrcAddr. При внесении небольших возмущений в исходные данные наблюдается незначительное перераспределение весов признаков, однако общий характер их влияния остаётся схожим.

Для формирования текстового объяснения решения модели используется БЯМ Qwen3-14B. На основе значений признаков и их весов по методу LIME модель отнесла выбранную сессию к классу «Normal».

Новизна предложенного решения заключается в повышении интерпретируемости результатов анализа сетевых сессий в промышленной сети за счет сочетания обученного классификатора на базе комитета деревьев решений с оптимизацией гиперпараметров и алгоритмов объяснимого искусственного интеллекта LIME и SHAP с последующей трансляцией кортежа параметров в человеко-читаемый формат с помощью БЯМ.

Список использованных источников

1.WUSTL-IIOT-2021 Dataset for IIoT Cybersecurity Research [Электронный ресурс], режим доступа: https://www.cse.wustl.edu/ ~jain/iiot2/index.html (дата обращения: 10.02.2026).

2.Akiba T. et al. Optuna: A next-generation hyperparameter optimization framework / Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, С. 2623-2631.

3.Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features / Advances in neural information processing systems, 2018, Т. 31, С. 1-11.

4.Zhang Y., Chen L., Tian Y. A Method for Evaluating the Interpretability of

Machine Learning Models in Predicting Bond Default Risk Based on LIME and SHAP [Электронный документ] / arXiv preprint arXiv:2502.19615, 2025, режим доступа: https://arxiv.org/abs/2502.19615 (дата обращения: 10.02.2026).

Вульфин Алексей Михайлович, д.т.н., проф. кафедры вычислительной техники и защиты информации, vulfin.alexey@gmail.com.
Кириллова Анастасия Дмитриевна, к.т.н., доцент каф. ВТиЗИ, kirillova.ad@ugatu.su.
Садыкова Эльвина Евгеньевна, аспирант каф. вычислительной техники и защиты информации, elvina.sadykova.01@mail.ru.

УДК 004.7.056; 004.7:004.8

МОДЕЛЬ АНАЛИЗА СЛАБОСТРУКТУРИРОВАННЫХ ДАННЫХ ОБ УГРОЗАХ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ ГРАФОВ ЗНАНИЙ

А.И. Луцкович, А.М. Вульфин, А.Д. Кириллова
Уфимский университет науки и технологий, г. Уфа

Ключевые слова: граф знаний, расширенная дополненная генерация, тактики и техники, анализ угроз информационной безопасности.

Для решения задач обработки слабоструктурированных данных о техниках, тактиках и процедурах (TTP) сегодня все чаще привлекаются технологии искусственного интеллекта (ИИ), связанные с использованием Large Language Models (LLMs) [1].

Одним из эффективных подходов к уменьшению эпизодов «галлюцинаций» и устаревших данных является предложенная в 2020 году технология Retrieval Augmented Generation (RAG) – генерация ответов с расширенным поиском [2]. Архитектура RAG-системы (рисунок 1) в общем виде состоит их двух основных частей – поисковика и генератора. Поисковик на основании получаемого запроса осуществляет поиск наиболее близких к смыслу запроса релевантных документов в векторной базе данных (БД), где хранятся документы компании, относящиеся к вопросам обеспечения ее информационной безопасности (ИБ). Полученные таким образом документы содержат максимально полезную контекстную информацию, необходимую для последующей обработки запроса модулем генератора. Генератор, используя полученную от поисковика информацию (запрос + найденный контекст), генерирует с помощью LLM ответ на поступивший в систему запрос пользователя.

Примеры успешного применения RAG-систем в задачах обработки и анализа слабоструктурированных данных в домене ИБ приводятся в достаточно большом числе публикаций (например, [3, 4]).