Чернов Владислав Владимирович,
курсант, ФГКВОУ ВО ВАЗИ
Chernov Vladislav Vladimirovich

Шаповалов Сергей Леонидович,
к.в.н, доцент, ФГКВОУ ВО ВАЗИ
Shapovalov Sergey Leonidovich

СОЗДАНИЕ СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ, ПРИМЕНЯЮЩИХ ТЕХНОЛОГИИ ИИ: НЕЙРОСЕТЕВАЯ IDS С ОНЛАЙН-ОБУЧЕНИЕМ
DEVELOPMENT OF AI-BASED INFORMATION SECURITY TOOLS: NEURAL NETWORK IDS WITH ONLINE LEARNING

Аннотация. В статье рассмотрены теоретические и практические аспекты создания программной модели SOC с нейросетевой IDS и функцией онлайн-обучения. Обоснован переход от сигнатурных методов к адаптивным ИИ-системам. Представлена архитектура системы (модуль захвата пакетов, нейросетевой классификатор 5→32→16→1, механизм онлайн-обучения, веб-дашборд). Сравнение с аналогами показало преимущества в адаптивности, вычислительной эффективности и применимости.

Abstract. This article examines the theoretical and practical aspects of developing a software model for SOC featuring a neural network-based IDS with online learning capability. The necessity of transitioning from static signature-based methods to adaptive AI-driven systems is justified. The architecture of the proposed system is presented, including a packet capture module, a neural network classifier (5→32→16→1), an online learning mechanism, and a web-based visualization dashboard. A comparative analysis with existing solutions demonstrates advantages in adaptability, computational efficiency, and practical applicability.

Ключевые слова: Искусственный интеллект, системы обнаружения вторжений, онлайн-обучение, нейронные сети, кибербезопасность, SOC.

Keywords: Artificial intelligence, intrusion detection systems, online learning, neural networks, cybersecurity, SOC.

В условиях стремительного роста числа и сложности кибератак, которые, по оценкам экспертов, происходят каждые 11 секунд, а глобальный ущерб от киберпреступности к 2025 году достигнет $10,5 трлн ежегодно, проблема эффективной защиты информационных систем приобретает критическое значение [1, 2].

Традиционные сигнатурные системы обнаружения вторжений (S-IDS) обеспечивают высокую точность при выявлении известных атак, однако малоэффективны против новых, модифицированных и полиморфных угроз: точность обнаружения новых атак не превышает 17–20% [8]. Аномалийные системы (A-IDS), напротив, характеризуются высоким уровнем ложных срабатываний (False Positive Rate), особенно в динамичных сетевых средах, что снижает эффективность и доверие к системе [3]. В связи с этим актуальной научно-практической задачей является создание адаптивных средств защиты информации, способных непрерывно обучаться на потоковых данных и оперативно выявлять новые угрозы. Перспективным решением является применение технологий искусственного интеллекта, в частности нейронных сетей с механизмами онлайн-обучения [8, 10].

1. Парадигма онлайн-обучения в задачах IDS

Классические нейросетевые IDS обучаются на статических наборах данных, что снижает их эффективность при появлении новых атак и изменении паттернов легитимного трафика. Онлайн-обучение позволяет инкрементально обновлять модель по мере поступления данных, обеспечивая адаптацию к изменяющимся условиям в реальном времени без периодического переобучения на собранных датасетах. Работы [1] показывают высокую эффективность такого подхода в динамичных IoT-средах.

В последнее время активно исследуются гибридные архитектуры, сочетающие нейронные сети с адаптивными алгоритмами, например ALMA (Approximate Large Margin Algorithm) с усреднением весов SWA, обеспечивающие ROC AUC до 98,58% при потреблении всего 14,64 КБ памяти [1]. При этом DNN в задачах обнаружения вторжений превосходят классические методы (Naïve Bayes, SVM), достигая точности 96,2% [5]. Однако сравнительный анализ LLM и традиционных ML/DL-моделей показал, что трансформеры не дают существенного преимущества при классификации сетевых потоков по сравнению с оптимизированными моделями, такими как XGBoost, достигающими F1-score 96,96% при загрузке CPU всего 4%, что подчеркивает необходимость выбора архитектуры с учетом баланса точности и вычислительных затрат.

2. Архитектура нейросетевой IDS

Предлагаемая система основана на многоуровневой архитектуре (рис. 1), включающей:

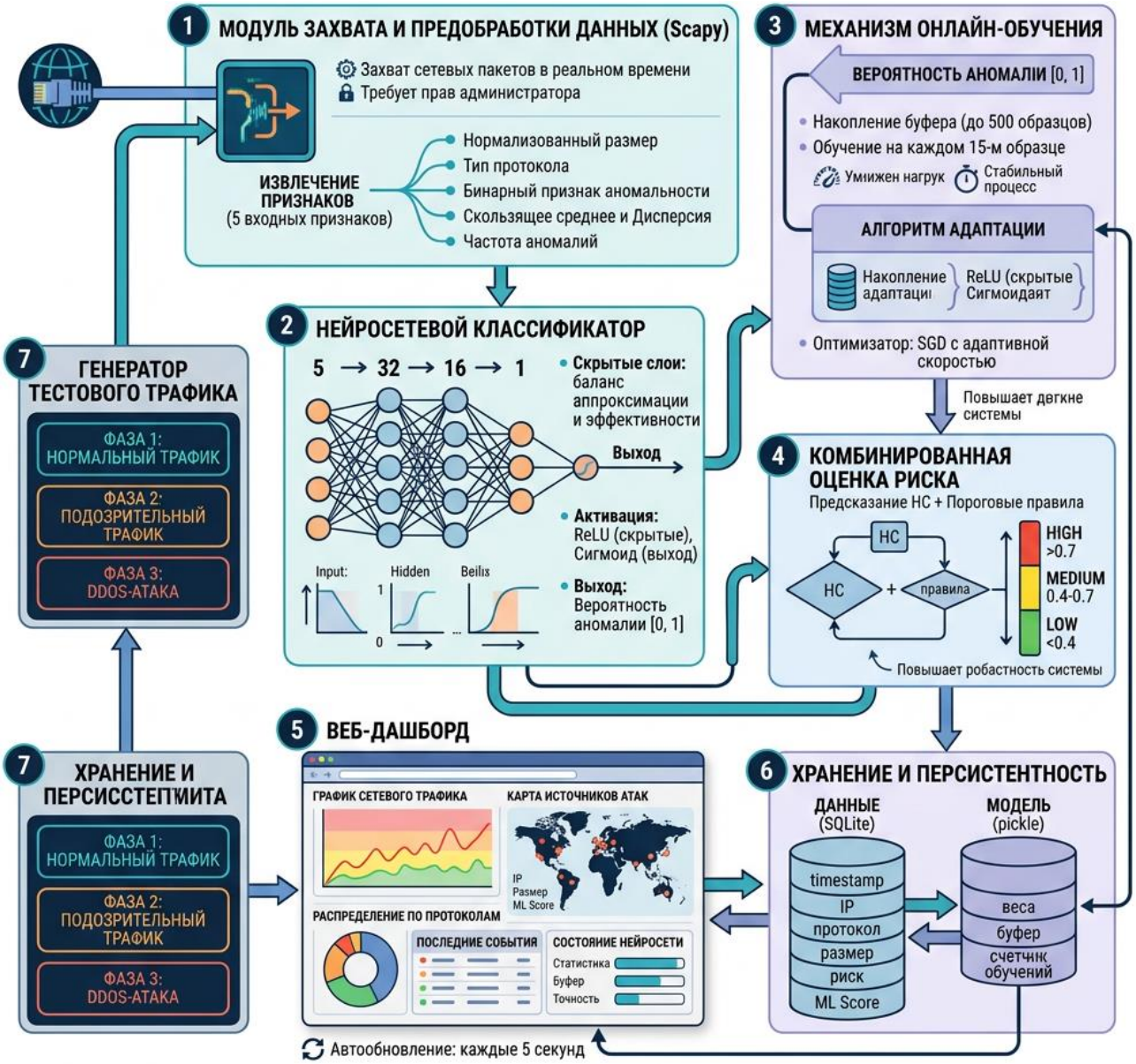

Рисунок 1 – Многоуровневая архитектура системы обнаружения аномалий

1. Модуль захвата и предобработки данных. Использует библиотеку Scapy для захвата сетевых пакетов в реальном времени (требует прав администратора). На этом этапе осуществляется извлечение признаков: нормализованный размер пакета, тип протокола, бинарный признак аномальности размера, статистические показатели (скользящее среднее и дисперсия), частота аномалий. Всего – 5 входных признаков.

2. Нейросетевой классификатор. Архитектура: 5→32→16→1. Выбор двух скрытых слоев (32 и 16 нейронов) обусловлен необходимостью баланса между аппроксимирующей способностью и вычислительной эффективностью. Функция активации – ReLU для скрытых слоев, сигмоид для выходного (вероятность аномалии в диапазоне [0,1]).

3. Механизм онлайн-обучения. Реализован следующий алгоритм адаптации: накопление буфера образцов до 500, запуск процедуры обучения на каждом 15-м образце. Такой подход позволяет снизить вычислительную нагрузку и стабилизировать процесс обучения. Используется оптимизатор SGD с адаптивной скоростью обучения.

4. Комбинированная оценка риска. На основе предсказания нейросети и пороговых правил формируется трехуровневая оценка: HIGH (>0.7), MEDIUM (0.4-0.7), LOW (<0.4). Комбинированный подход повышает робастность системы.

5. Веб-дашборд. Обеспечивает визуализацию: график сетевого трафика с цветовой индикацией риска (красный/желтый/зеленый), карта источников атак с отображением IP-адреса, объема пакета и ML Score, круговая диаграмма распределения по протоколам, таблица последних событий, панель состояния нейросети (статистика обучения, буфер, точность). Автообновление каждые 5 секунд.

6. Хранение и персистентность. Данные сохраняются в SQLite (timestamp, IP, протокол, размер, риск, ML Score). Модель нейросети (веса, буфер, счетчик обучений) сохраняется в формате pickle для возможности восстановления состояния.

7. Генератор тестового трафика. Для валидации системы реализован режим симуляции, включающий три фазы: нормальный трафик, подозрительный трафик, DDoS-атака.

В Таблице 1 представлен сравнительный анализ разработанной системы с современными аналогами, выделенными на основе обзора литературы [1, 2, 3].

Таблица 1

Сравнительная характеристика IDS-решений

Критерий	**Разработанная система**	**ALMANET**	**LLM-based IDS**	**AI-E-DiD**	**DNN IDS**	**Hybrid DL**
Архитектура	5→32→16→1	ALMA+SWA+NET	GPT-2, LLaMA	GAN-LSTM-AE	DNN	XGBoost +FL
Тип обучения	Онлайн	Онлайн	Fine-tuning (офлайн)	Офлайн	Офлайн	Офлайн
Адаптация к новым атакам	Высокая	Высокая	Высокая	Высокая	Высокая	Высокая
Вычисли-тельные затраты	Низкие	Низкие	Очень высокие	Высокие	Средние	Средние
Точность (F1/AUC)	~95% (оценка)	98.58% AUC	95%+ F1	95.6% (DoS)	96.2% Acc	96.2% F1
Обнаружение zero-day	Да	Да	Огранич-енно	Ограни-ченно	Нет	Да
Реальное время	Да	Да	Ограни-ченно	Да	Нет	Да
Визуализация (SOC)	Встроенная	Нет	Нет	Нет	Нет	Нет

Из анализа видно, что разработанная система занимает уникальную нишу: она сочетает онлайн-обучение с низкими вычислительными затратами и встроенной визуализацией, что делает её готовым к применению решением для центров мониторинга безопасности (SOC). В отличие от громоздких LLM-решений, требующих значительных ресурсов, система эффективна для развертывания на стандартных рабочих станциях (Intel Core i3+). В отличие от офлайн-моделей, она способна адаптироваться к новым угрозам в реальном времени.

Научная новизна работы заключается в следующем:

1. Разработана и обоснована архитектура легковесной нейросетевой IDS с онлайн-обучением, оптимизированная для работы на ограниченных вычислительных ресурсах, что критически важно для массового внедрения.

2. Предложен алгоритм инкрементального обучения, основанный на накоплении буфера и периодическом обновлении весов, который позволяет минимизировать вычислительную нагрузку при сохранении адаптивности.

3. Реализован комбинированный механизм оценки риска, сочетающий нейросетевое предсказание и пороговые правила, что снижает вероятность ложных срабатываний.

Практическая значимость подтверждена реализацией полного программного комплекса, включающего не только детектор, но и полнофункциональный веб-дашборд, систему хранения данных и генератор тестового трафика. Система может применяться в автоматизированных системах электронного документооборота, корпоративных сетях и учебных центрах по подготовке специалистов по информационной безопасности.

В статье представлена разработка программной модели центра мониторинга безопасности с нейросетевой IDS и онлайн-обучением. Показано, что переход от статических сигнатурных методов к адаптивным нейросетевым системам является необходимым шагом в условиях современной динамики киберугроз. Сравнительный анализ подтверждает конкурентоспособность предложенного решения по ключевым параметрам: адаптивность, вычислительная эффективность и полнота функционала.

Дальнейшие исследования планируется направить на расширение набора признаков, внедрение механизмов защиты от дрифта распределения данных и интеграцию с существующими SIEM-системами. Важным направлением также является исследование робастности моделей к атакам на этапе обучения («отравление» данных) [6, 8] и разработка методов объяснимого ИИ (XAI) для повышения прозрачности принимаемых решений [2, 9].

Список литературы:

1. Agbedanu, P. R. ALMANET: A hybrid online learning IDS for real-time IoT security / P. R. Agbedanu, [et al.]. – Текст: непосредственный // Egyptian Informatics Journal. – 2025.

2. Evaluating large language models effectiveness for flow-based intrusion detection. – Текст: непосредственный // Artificial Intelligence Review. – 2026. – Vol. 59. – P. 50.

3. Applying the defense model to strengthen information security with AI. – Текст: непосредственный // Scientific Reports. – 2025. – Vol. 15. – P. 30292.

4. Comparative summary of recent IDS literature. – Текст: непосредственный // Scientific Reports. – 2025.

5. Nigam, N. Comparative Evaluation of Cloud Security Algorithms / N. Nigam, B. Sireesha. – Текст: непосредственный // IEEE ICTBIG. – 2025.

6. Andreas, A. Ethical AI in Cybersecurity / A. Andreas, [et al.]. – Текст: непосредственный // AI for Cybersecurity. – Wiley, 2025.

7. XGBoost + LightGBM + Federated learning. – Текст: непосредственный // Scientific Reports. – 2025.

8. Neural Network IDS/IPS with Adaptive Online Training. – Текст: непосредственный // MDPI. – 2025. – Vol. 9, No. 11. – P. 267.

9. Косоруков, А. А. Технологии искусственного интеллекта в сфере цифровой безопасности / А. А. Косоруков. – Текст: непосредственный // Эффективное управление. – 2025.References.

10. Искусственный интеллект в кибербезопасности: роль машинного обучения в обеспечении защиты данных. – Текст: электронный // Zenodo. – 2025. – 02 мая. – URL: https://zenodo.org/records/15321663 (дата обращения: 14.07.2026)