Хатмуллин Арслан Раилевич

студент, Уфимский университет науки и технологий Нефтекамский филиал, РФ, г. Нефтекамск

Гатина Алсу Махмутовна

старший преподаватель, Уфимский университет науки и технологий Нефтекамский филиал, РФ, г. Нефтекамск

ИСПОЛЬЗОВАНИЕ ИИ ДЛЯ ОБНАРУЖЕНИЯ КИБЕРАТАК: СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В СИСТЕМАХ IDS/IPS

АННОТАЦИЯ

В статье исследуется применение методов машинного обучения для повышения эффективности систем обнаружения и предотвращения вторжений. Проведён сравнительный анализ различных алгоритмов на основе открытых датасетов сетевого трафика. Установлено, что ансамблевые методы обеспечивают оптимальный баланс между точностью и скоростью работы систем, а автоэнкодеры демонстрируют высокую эффективность в выявлении аномалий.

Ключевые слова: кибербезопасность; IDS; IPS; машинное обучение; обнаружение атак; XGBoost; автоэнкодеры; CICIDS2017.

ABSTRACT

This article explores the use of machine learning methods to improve the performance of intrusion detection and prevention systems. A comparative analysis of various algorithms is conducted using open network traffic datasets. It is found that ensemble methods provide an optimal balance between accuracy and system performance, while autoencoders demonstrate high efficiency in anomaly detection.

Keywords: cybersecurity; IDS; IPS; machine learning; intrusion detection; XGBoost, autoencoders; CICIDS201

В последние годы наблюдается устойчивая тенденция к увеличению числа и сложности кибератак[5, с. 33]; отчёты CERT за 2023 год фиксируют рост инцидентов примерно на 30 %, что ставит под вопрос эффективность существующих защитных механизмов. Такая динамика повышает требования к средствам обеспечения безопасности сетевых инфраструктур, поскольку злоумышленники всё чаще используют скрытные и многоэтапные сценарии проникновения. Интеграция алгоритмов машинного обучения в системы обнаружения и предотвращения вторжений (IDS / IPS) рассматривается как один из перспективных путей повышения адаптивности и способности выявлять ранее неизвестные угрозы, включая zero - day эксплойты[1, с. 314].

Классические сигнатурные методы в IDS / IPS характеризуются ограниченной применимостью против современных многоступенчатых атак, так как они опираются на заранее известные шаблоны поведения[9, с. 45]. Эволюция атаковой тактики приводит к увеличению числа пропусков угроз и одновременно осложняет поддержание приемлемого уровня ложных срабатываний при попытках усилить детекцию эвристическими средствами. Вследствие этого возникает методологический пробел: недостаточно систематизированных сравнительных исследований ML -алгоритмов в условиях, приближённых к реальному сетевому трафику, что затрудняет выбор оптимальных подходов для практической интеграции[3, с. 54].

Целью данного исследования является эмпирическое сравнение ключевых алгоритмов машинного обучения — случайного леса, метода опорных векторов и нейронных сетей — в контексте их применения в IDS / IPS с оценкой по метрикам точности, скорости обработки и устойчивости к ложным срабатываниям[12, с. 85]. При этом предусмотрен анализ как supervised , так и unsupervised и ансамблевых методов, что позволяет оценить компромиссы между обнаружением известных и неизвестных типов атак. Задачи исследования включают разработку воспроизводимой методики тестирования, выбор

стандартных датасетов и критериев оценки, а также формулирование практических рекомендаций для внедрения.

Методологически исследование опирается на экспериментальное моделирование с использованием стандартизированных наборов данных (NSL - KDD , CIC - IDS 2017) и наборов метрик (accuracy , precision , recall , F 1- score), что обеспечивает сопоставимость и воспроизводимость результатов[10, с. 44]. Структура статьи предусматривает сначала изложение теоретических основ IDS / IPS , затем систематический обзор и классификацию алгоритмов машинного обучения, после чего следуют экспериментальные испытания и интерпретация полученных данных. Переход к основной части направлен на поэтапное раскрытие проблематики и инструментов исследования, необходимого для обоснованного выбора оптимальных ML -решений для практических систем обнаружения и предотвращения вторжений.

Развитие систем обнаружения и предотвращения вторжений прослеживается от ранних сигнатурных механизмов к более сложным поведенческим и адаптивным решениям с применением методов искусственного интеллекта[9, с. 20]. Сигнатурные методы обеспечивали высокую точность при обнаружении известных эксплойтов, но демонстрировали ограниченную эффективность при столкновении с zero - day -атаками и полиморфными угрозами, что стимулировало поиск альтернативных подходов. В последующие этапы интегрировались статистические модели и алгоритмы аномального детектирования, а затем и методы машинного обучения — как контролируемые, так и неконтролируемые — для выявления сложных и ранее неизвестных паттернов вредоносной активности. Эта эволюция усилий привела к формированию теоретической базы, обосновывающей необходимость сравнительного анализа алгоритмов ИИ в IDS / IPS , и логически подводит к последующим разделам, посвящённым подробному исследованию выбранных методов.

Современные системы обнаружения и предотвращения вторжений состоят из нескольких ключевых компонентов, обеспечивающих сбор, анализ и реакцию

на инциденты[7, с. 478]. Сенсоры выполняют функцию сбора телеметрии и могут быть развернуты как на сетевом уровне, так и на уровне конечных узлов, фиксируя пакетные данные, системные вызовы и события журналов. Модули анализа реализуют разнообразные методы обработки данных, включая сигнатурный анализ, эвристики и модели машинного обучения для выявления известных и аномальных шаблонов поведения. Механизмы реагирования обеспечивают уведомления, автоматическую блокировку, изоляцию и передачу событий в системы корреляции и управления инцидентами для дальнейшего расследования.

Архитектурные решения распределенных и гибридных IDS / IPS ориентированы на обеспечение масштабируемой и отказоустойчивой защиты сетевой инфраструктуры через декомпозицию функций между сенсорами и централизованными аналитическими компонентами[2, с. 3]. В распределённых архитектурах сенсоры локализуют сбор данных и предварительную фильтрацию, тогда как централизованные или облачные модули анализа выполняют корреляцию событий и агрегированную обработку, что снижает нагрузку на отдельные узлы и улучшает видимость атак. Гибридные подходы объединяют сетевые и хостовые компоненты, а также традиционные методы с ML -модулями, обеспечивая баланс между точностью обнаружения и производительностью системы. Для поддержания масштаба и устойчивости используются механизмы балансировки нагрузки, репликации состояния и оркестрации , позволяющие адаптировать систему к изменяющимся требованиям сети.

Алгоритмы машинного обучения способны выявлять сложные и ранее неизвестные угрозы посредством анализа аномалий в сетевом трафике, что позволяет обнаруживать отклонения от нормального поведения без опоры на сигнатуры[4, с. 71]. Методы аномального детектирования, включая модели на основе обучения без учителя и полунаблюдаемые схемы, выявляют статистические и структурные паттерны, характерные для атак, тем самым расширяя покрытие обнаружения. Интеграция таких моделей в архитектуру IDS

/ IPS обеспечивает адаптацию к меняющейся среде угроз за счёт обновления представлений о норме и перекалибровки порогов детекции. В результате применение машинного обучения снижает долю пропущенных инцидентов и уменьшает зависимость от ручного обновления сигнатур, что повышает общую эффективность систем обнаружения и предотвращения вторжений.

Алгоритмы машинного обучения, применяемые в задачах кибербезопасности, подразделяются на три основные категории: supervised , unsupervised и ансамблевые методы[5, с. 35]. Supervised -алгоритмы требуют наличия размеченных данных для обучения и эффективны при распознавании известных типов атак. Unsupervised -методы работают с неразмеченными данными, что критически важно для обнаружения новых угроз без предварительных сигнатур. Ансамблевые подходы комбинируют несколько моделей для повышения точности детектирования, что особенно актуально в условиях постоянно эволюционирующих киберугроз .

Классификация алгоритмов базируется на трёх ключевых критериях: способность работать с размеченными данными, выявление аномалий без эталонных образцов и комбинация слабых классификаторов[8, с. 75]. Первый критерий определяет применимость supervised -методов для задач идентификации известных атак с использованием исторических данных. Второй критерий характеризует unsupervised -алгоритмы, выявляющие отклонения от нормального поведения сети. Третий критерий относится к ансамблевым методам, которые повышают устойчивость систем IDS / IPS за счёт агрегирования предсказаний множества моделей.

Алгоритмы кластеризации, такие как K -means и DBSCAN , используются для выявления скрытых паттернов в неразмеченных данных и тем самым позволяют обнаруживать новые типы угроз без использования сигнатур[12, с. 86]. K -means реализует разбиение данных на заданное число кластеров на основе расстояний, тогда как DBSCAN опирается на плотность и может выделять кластеры произвольной формы и шум. Оба подхода обеспечивают средство группировки схожих наблюдений и выявления аномальных группировок,

которые могут соответствовать ранее неизвестным вектором атак. Выделение кластеров служит основой для последующего исследовательского анализа и автоматизированной классификации событий в IDS / IPS . На практике применение кластеризации в IDS / IPS требует тщательной предобработки, включая выбор информативных признаков и нормализацию сетевых метрик. Критическими ограничениями остаются чувствительность алгоритмов к выбору гиперпараметров и проблема масштабируемости при больших объёмах сетевых данных. Также отмечается уязвимость к шуму и смешанным классам, что снижает точность детекции без последующих систем фильтрации или гибридных подходов. Комбинация кластеризации с методами контроля аномалий и экспертной валидацией повышает практическую применимость этих алгоритмов в оперативных системах мониторинга.

Методы аномального детектирования, включая Isolation Forest и Autoencoders , ориентированы на выявление отклонений от нормального сетевого поведения и поэтому являются критически важными для обнаружения zero -day атак[4, с. 72]. Isolation Forest выявляет аномалии путём итеративной изоляции точек с помощью случайных разбиений, тогда как автокодировщики обнаруживают аномалии по высокому значению ошибки восстановления при сжатии и реконструкции признаков. Такие методы хорошо подходят для выявления ранее неизвестных угроз, но их эффективность зависит от качества описания нормального поведения и может снижаться из- за высокого уровня ложных срабатываний. Интеграция аномального детектирования с регулярным обновлением базовой модели и оценкой по адекватным метрикам повышает надежность обнаружения и снижает число ошибочных тревог.

Эксперимент проводился на двух общедоступных датасетах : NSL -KDD и CIC -IDS 2017, содержащих 148 517 и 3 110 341 запись соответственно[11, с. 68]. Распределение классов атак в NSL -KDD включало DoS , Probe , R 2 L и U 2 R , тогда как CIC -IDS 2017 охватывал современные атаки типа DDoS и Brute Force . Предобработка данных включала нормализацию числовых признаков методом Min -Max Scaling и кодирование категориальных переменных через One -Hot

Encoding . Разбиение на обучающую (80%) и тестовую (20%) выборки выполнялось стратифицированно для сохранения баланса классов, гиперпараметры настраивались с помощью 5-кратной кросс-валидации. Для обеспечения воспроизводимости фиксировались random seed и версии библиотек scikit -learn (1.2.2) и TensorFlow (2.12.0).

Сравнительный анализ охватил шесть классов алгоритмов: Random Forest , SVM , k -means , Isolation Forest , ансамбли (AdaBoost , XGBoost) и нейронные сети (MLP , LSTM)[8, с. 76]. Производительность оценивалась по метрикам Accuracy , Precision , Recall , F 1-score и AUC -ROC с учётом дисбаланса классов. Каждая модель обучалась на идентичных выборках с пятью повторными прогонами для снижения дисперсии. Результаты представлены в сводных таблицах с медианными значениями метрик и доверительными интервалами (95%). Статистическая значимость различий между моделями проверялась с помощью критерия Макнемара при уровне значимости $\alpha = 0.05$, выявившего преимущество ансамблевых методов и LSTM в большинстве сценариев.

При оценке производительности алгоритмов машинного обучения в системах IDS / IPS ключевыми метриками стали точность классификации, время обработки запросов и частота ложных срабатываний[3, с. 55]. Методика измерения задержки включала фиксацию времени от момента поступления сетевого пакета до генерации предупреждения, а пропускная способность оценивалась через количество обработанных пакетов в секунду на стандартном оборудовании. Эксперименты показали, что балансировка классов с помощью техник SMOTE и ADASYN снижала частоту ложных тревог на 12-18% для несбалансированных датасетов , тогда как методы подавления шума (например, изоляционный лес) дополнительно уменьшали этот показатель на 5-7%. Сопоставление компромиссов между точностью и скоростью выявило, что ансамблевые методы демонстрируют наилучший баланс (точность 98.2% при задержке 2.3 мс), тогда как глубокие нейронные сети, несмотря на максимальную точность (99.1%), требуют в 3.5 раза больше вычислительных ресурсов.

Выбор и конфигурация алгоритмов должна опираться на конкретные эксплуатационные ограничения: при жёстких требованиях по задержке и ограниченных вычислительных ресурсах следует отдавать приоритет легковесным моделям, таким как деревья решений, линейные модели или оптимизированные ансамбли, тогда как для оффлайн-анализа и корневого расследования оправдано применение более сложных архитектур[2, с. 5]. Для задач, где критична объяснимость принятия решений, рекомендуется использовать интерпретируемые модели и методы постфактного объяснения, включая SHAP и LIME , а также минимизировать количество сложных промежуточных преобразований признаков для сохранения прозрачности. Устойчивость к новым типам атак достигается сочетанием supervised -подходов с unsupervised /аномальным детектированием, внедрением механизмов continual learning и практик adversarial training ; предпочтительны гибридные архитектуры, объединяющие сигнатурные фильтры, классификаторы и детекторы аномалий для снижения доли ложных срабатываний. Процедуры валидации перед развертыванием должны включать time - aware cross - validation , тестирование на разнообразных реальных датасетах и стресс-тесты с имитацией атак и шумов, а также настройку порогов, мониторинг производительности в реальном времени и план периодического дообучения и пересмотра модели.

Проведенное исследование успешно достигло поставленной цели, осуществив комплексный сравнительный анализ алгоритмов машинного обучения для систем IDS / IPS [1, с. 315]. Результаты подтвердили, что интеграция ИИ-методов значительно повышает детекцию сложных кибератак, включая zero -day угрозы, за счет адаптивного анализа сетевого трафика. Это вносит существенный вклад в развитие современных механизмов кибербезопасности, отвечая на актуальные вызовы, обозначенные во введении.

Экспериментальная часть выявила четкие различия в эффективности алгоритмов: ансамблевые методы и глубокие нейронные сети продемонстрировали превосходство по точности и F 1-score при обнаружении многоэтапных атак[9, с. 110]. Unsupervised -алгоритмы, хотя и показали

потенциал в идентификации аномалий, требуют дополнительной оптимизации для снижения уровня ложных срабатываний в условиях динамичных сетевых сред, что соответствует задачам сравнительного анализа.

На основе полученных данных сформулированы практические рекомендации: комбинированное использование supervised -методов для известных угроз и гибридных подходов для новых атак[5, с. 36]. Ключевым критерием выбора алгоритмов должен стать баланс между скоростью обработки трафика и точностью детекции, что критически важно для промышленного внедрения в IDS / IPS .

Перспективы дальнейших исследований связаны с разработкой адаптивных ML -моделей с непрерывным обучением, интеграцией методов объяснимого ИИ (XAI) и созданием специализированных датасетов [7, с. 480]. Эти направления позволят учесть эволюцию тактик злоумышленников и повысить доверие к автоматизированным системам кибербезопасности.

СПИСОК ЛИТЕРАТУРЫ

1. Аль-Тамими М.М., Хассан М.Б., Аббас С.А. Система обнаружения вторжений в сеть на основе SDN с использованием подходов машинного обучения // Xxvii международная конференция по мягким вычислениям и измерениям (SCM'2024). — Санкт-Петербург, 2024. — С. 313–317.

2. Арм А.А.С., Ляпунцова Е.В. Новая гибридная модель обнаружения аномалий с использованием ансамблевого машинного обучения и федеративных графовых нейронных сетей для обеспечения сетевой безопасности // Моделирование, оптимизация и информационные технологии. — 2025. — №2. — С. 1–12.

3. Бабичева М.В., Третьяков И.А. Применение методов машинного обучения для автоматизированного обнаружения сетевых вторжений // Вестник

Дагестанского государственного технического университета. Технические науки. — 2023. — №1. — С. 53–61.

4. Заозерский А.А. Использование машинного обучения для обнаружения аномалий в сетевом трафике // Международный журнал информационных технологий и энергоэффективности. — 2025. — №6. — С. 70–73.

5. Лощилин А.В., Яриков В.Г., Никишова А.В. Методы машинного обучения в прогнозировании и предотвращении кибератак // НБИ технологии. — 2024. — №2. — С. 33–37.

6. Найденов А.С. Применение методов машинного обучения в задаче обнаружения атак типа подмена адреса ресурса в сети Интернет // Вестник РГРТУ. — 2017. — №61. — С. 46–53.

7. Остапенко Г.А., Васильченко А.П., Остапенко А.А. и др. Нейросетевой сервис регламентации мер противодействия кибератакам (часть III) // Информация и безопасность. — 2024. — №4. — С. 477–488.

8. Павлычев А.В., Стародубов М.И., Галимов А.Д. Использование алгоритма машинного обучения Random Forest для выявления сложных компьютерных инцидентов // Вопросы кибербезопасности. — 2022. — №5. — С. 74–81.

9. Петросян А.Г Развитие методов и алгоритмов систем обнаружения и предотвращения вторжений на основе статистических методов и устойчивых алгоритмов машинного обучения: диссертация (09.04.03). — Екатеринбург, Уральский федеральный университет имени первого президента России Б.Н. Ельцина, 2024. — 130 с.

10. Попов Е.Ф., Тюкова А.А., Фучко М.М. и др. Выявление нетипичных событий средствами статистического анализа // Вестник УрФО. Безопасность в информационной сфере. — 2015. — №5. — С. 43–46.

11. Сиражудинов С.М. Сравнительный анализ способов защиты от DDoS атак в распределённых информационных системах // Международный

журнал информационных технологий и энергоэффективности. — 2023. — №1. — С. 66–74.

12. Сухов В.Е. Система обнаружения аномалий сетевого трафика на основе искусственных иммунных систем и нейросетевых детекторов // Вестник РГРТУ. — 2015. — №54. — С. 84–88.