УДК 004.85:004.056.5

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ И ПРЕДОТВРАЩЕНИЯ КИБЕРАТАК В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Фролов К.В.

Студент

Поволжский государственный университет телекоммуникаций и информатики

Самара, Россия

Куваева Е.Н.

ст. преподаватель кафедры ИСТ

ПГУТИ

Самара, Россия

Аннотация. Актуальность. Число фиксируемых инцидентов информационной безопасности в корпоративных сетях ежегодно увеличивается, при этом классические средства защиты не справляются с обнаружением ранее неизвестных вредоносных сценариев, что делает востребованным построение адаптивных систем обнаружения вторжений на базе машинного обучения. Цель исследования — выполнить сравнительную оценку групп алгоритмов машинного обучения, применяемых для выявления нестандартного сетевого поведения, и определить условия, при которых каждая из групп даёт наибольший практический эффект. Методика. Использованы анализ и обобщение публикаций по теме исследования, а также сопоставление трёх групп методов — классификаторов с учителем, алгоритмов поиска выбросов без учителя и моделей глубокого обучения — по пяти параметрам: точности, полноте обнаружения, доле ложных срабатываний, вычислительным затратам и требованиям к разметке обучающих данных. Результаты. Установлено, что ни одна из групп методов не является универсальной: классификаторы с учителем

точны на известных классах атак, но не распознают новые угрозы; эффективность методов unsupervised learning существенно зависит от стабильности исходных характеристик сетевого трафика; глубокие нейронные сети точнее прочих улавливают многошаговые атаки ценой роста вычислительной нагрузки. Наилучшие показатели полноты обнаружения дают двухступенчатые схемы, объединяющие предварительную фильтрацию признаков с последующей классификацией. Выводы. Сформулированы критерии выбора метода машинного обучения исходя из масштаба защищаемой инфраструктуры, допустимой задержки анализа и доступности размеченных данных об инцидентах.

Ключевые слова: машинное обучение, информационная безопасность, обнаружение аномалий, сетевой трафик, кибератаки, системы обнаружения вторжений, нейронные сети, защита информационных систем.

MACHINE LEARNING METHODS FOR ANOMALY DETECTION AND CYBERATTACK PREVENTION IN CORPORATE INFORMATION SYSTEMS

Frolov K.V.

Student

Volga Region State University of Telecommunications and Informatics

Samara, Russia

Kuvaeva E.N.

Senior Lecturer, Department of Information Technologies and Communications

PSUTI

Samara, Russia

Abstract. Relevance. The annual number of recorded information security incidents in corporate networks continues to grow, while classical signature-based protection tools fail to detect previously undescribed malicious scenarios, which creates demand for adaptive intrusion detection systems built on machine learning. Purpose. The study performs a comparative assessment of machine learning algorithm groups used to identify anomalous network behavior and determines the conditions under which each group yields the greatest practical effect. Methods. Analysis and generalization of relevant publications were used, together with a comparison of three method groups — supervised classifiers, unsupervised outlier-detection algorithms, and deep learning models — across five parameters: accuracy, detection recall, false-positive rate, computational cost, and requirements for labeled training data. Results. It was established that no single group of methods is universal: supervised classifiers are accurate on known attack classes but fail to recognize new threats; unsupervised methods are sensitive to drift in the normal traffic profile; deep neural networks capture multi-stage attacks more accurately than the alternatives at the cost of higher computational load. The best recall figures are achieved by two-stage schemes combining preliminary feature filtering with subsequent classification. Conclusions. Criteria for selecting a machine learning method are formulated based on the scale of the protected infrastructure, acceptable analysis latency, and the availability of labeled incident data.

Keywords: machine learning, information security, anomaly detection, network traffic, cyberattacks, intrusion detection systems, neural networks, information systems protection.

Введение

Корпоративные информационные системы ежедневно генерируют десятки и сотни тысяч сетевых событий, и лишь малая доля этого потока связана с реальными угрозами. Классический подход к защите — межсетевые экраны и

сигнатурные системы обнаружения вторжений — строится на сопоставлении трафика с базой известных шаблонов атак; он быстр и предсказуем, однако бессилен перед вариациями уже известных атак и тем более перед атаками ранее не известными, сигнатуры которых ещё не описаны [3]. Практика показывает, что промежуток между появлением новой техники атаки и её включением в базу сигнатур может составлять от нескольких дней до нескольких месяцев — именно в этом окне корпоративная сеть остаётся уязвимой. Альтернативой служит поведенческий анализ на основе машинного обучения: вместо сравнения с шаблонами модель строит представление о «нормальном» состоянии сети и помечает как подозрительные любые существенные отклонения от него [1, 5]. Настоящая статья посвящена сопоставлению групп методов машинного обучения, пригодных для такой задачи, и выработке практических критериев их выбора для информационных систем разного масштаба.

Обзор литературы

Исследования по теме условно делятся на три направления. Первое опирается на алгоритмы классификации с учителем — случайный лес, метод опорных векторов, градиентный бустинг, — которые обучаются на размеченных выборках вредоносного и легитимного трафика и показывают высокую точность на известных типах атак [1]. Второе направление образуют методы обнаружения выбросов без учителя: изоляционный лес, одноклассовый SVM, кластеризация плотности, — не требующие заранее размеченных примеров атак и потому способные реагировать на ранее не встречавшиеся аномалии, хотя и ценой более высокой доли ложноположительных срабатываний [2]. Третье, наиболее активно развивающееся направление связано с глубоким обучением: рекуррентные сети (в том числе LSTM) применяются для анализа временной динамики трафика и выявления растянутых во времени, многошаговых сценариев вторжения, а свёрточные сети — для распознавания локальных паттернов в признаковом представлении пакетов [5, 7]. Отдельная линия работ посвящена гибридным схемам: комбинация иммунных, нейросетевых и нейронечётких

классификаторов на последовательных этапах обработки трафика позволяет сначала быстро отсеять заведомо нормальные соединения, а затем подробно анализировать только подозрительный остаток, что снижает вычислительную нагрузку без потери качества обнаружения [6]. Чтобы защитить систему от мощных DDoS-атак, лучше использовать легкие нейросети. Они быстрее анализируют общую статистику NetFlow, чем тяжелые модели, которые разбирают каждый пакет данных по отдельности. [4]. Общим ограничением для всех трёх направлений исследователи называют дисбаланс классов в обучающих данных: доля примеров атак в реальном трафике обычно составляет проценты или доли процента от общего объёма, что смещает модели в сторону переобучения на «нормальном» классе и повышает риск пропуска редких, но опасных инцидентов [2, 5].

Постановка задачи

Исходя из выявленного в литературе разрыва между отдельными сравнениями конкретных алгоритмов и практической задачей выбора метода для конкретной инфраструктуры, в статье решаются следующие задачи: во-первых, свести рассмотренные методы в единую систему координат по пяти сопоставимым параметрам; во-вторых, оценить, при каких условиях эксплуатации (объём трафика, допустимая задержка реакции, доступность размеченных инцидентов) каждая группа методов является предпочтительной; в-третьих, описать ограничения, общие для всех групп, и указать способы их частичной компенсации; в-четвёртых, предложить прикладной алгоритм выбора метода, которым может руководствоваться проектировщик системы защиты корпоративной сети.

Основная часть

Методика исследования.

Для сопоставления методов использован анализ восьми публикаций, посвящённых практическому применению машинного обучения в задачах сетевой безопасности [1–8], дополненный сведением рассматриваемых в них

подходов к трём укрупнённым группам: (а) классификаторы с учителем, (б) детекторы выбросов без учителя, (в) модели глубокого обучения. Для каждой группы фиксировались пять параметров сравнения: точность на известных классах атак, полнота обнаружения новых аномалий, характерная доля ложных срабатываний, вычислительная сложность обучения и применения модели, а также требования к объёму и качеству размеченных данных. Дополнительно рассмотрен сценарий эксплуатации: типичные наборы данных для обучения подобных моделей (в частности, NSL-KDD как усовершенствованная версия KDD Cup 99, устраняющая избыточность исходного набора) содержат десятки тысяч размеченных сетевых соединений и позволяют оценить обобщающую способность модели на данных, не встречавшихся при обучении [5].

Результаты сравнения сведены в таблицу 1.

Группа методов	**Точность на известных атаках**	**Обнаружение новых угроз**	**Ложные срабатывания**	**Вычислительные затраты**	**Требования к разметке**
Классификаторы с учителем	Высокая	Низкое	Низкие–средние	Средние	Высокие
Детекторы выбросов без учителя	Средняя	Высокое	Средние–высокие	Низкие–средние	Низкие
Модели глубокого обучения (LSTM, CNN)	Высокая	Среднее–высокое	Средние	Высокие	Высокие
Гибридные двухступенчатые схемы	Высокая	Высокое	Низкие–средние	Средние–высокие	Средние

Прим.: оценки носят качественный, обобщённый характер и основаны на сопоставлении выводов рассмотренных публикаций [1–8]; для конкретной инфраструктуры они уточняются на этапе пилотного тестирования на реальных данных.

Как видно из таблицы, ни одна из групп не доминирует по всем параметрам одновременно, что подтверждает необходимость осознанного выбора метода под конкретную задачу, а не механического переноса «лучшего» алгоритма из одной работы в другую. Классификаторы с учителем оправданы там, где спектр угроз относительно стабилен и накоплена история инцидентов — например, в системах, ориентированных на защиту от массовых, ранее фиксировавшихся атак. Детекторы выбросов уместны на начальном этапе построения системы защиты, когда размеченных инцидентов ещё недостаточно, либо как первый рубеж фильтрации в комбинированных схемах [2, 6]. Модели глубокого обучения целесообразны при защите инфраструктуры с высокой ценой пропуска атаки (финансовый сектор, критически важные объекты), где оправданы дополнительные вычислительные затраты ради выявления сложных, растянутых во времени сценариев [5, 7]. Наконец, для крупных операторов с высокоскоростными каналами связи, где важна обработка потока в реальном времени, часто выбирают компромиссный вариант — облегчённый ИИ для анализа статистики NetFlow вместо детального разбора пакетов [4].

Практический пример.

Для иллюстрации применимости описанных подходов рассмотрим условную схему защиты информационной системы среднего предприятия с распределённой филиальной сетью. На первом рубеже, обрабатывающем весь входящий поток, целесообразно разместить детектор выбросов без учителя (например, на основе изоляционного леса), задача которого — за минимальное время отсеять заведомо нормальные соединения и передать на второй рубеж только подозрительный остаток, составляющий, как правило, единицы процентов от исходного трафика. На втором рубеже подозрительные соединения анализируются моделью глубокого обучения, способной учитывать последовательность событий и распознавать многошаговые сценарии — попытки разведки, последующего закрепления в сети и горизонтального

перемещения между узлами. Такая двухступенчатая архитектура соответствует наблюдению из работы [6] о снижении вычислительной нагрузки за счёт предварительной фильтрации и одновременно обеспечивает полноту обнаружения, характерную для глубоких моделей.

Обсуждение результатов.

Сопоставление результатов подтверждает тезис об отсутствии универсального решения и одновременно указывает на общее для всей области ограничение — несбалансированность обучающих выборок, при которой доля примеров атак значительно уступает доле нормального трафика [2, 5]. Прямое обучение на таких данных смещает модель в сторону мажоритарного класса и провоцирует пропуск редких, но потенциально опасных инцидентов. Среди практикуемых способов компенсации — искусственное увеличение представленности минорного класса, взвешивание ошибок классификации по классам и переход к постановке задачи как обнаружения выбросов, где количество примеров нормального поведения естественным образом преобладает. Ограничением проведённого анализа является его опора на опубликованные результаты сравнения методов, полученные на разных наборах данных и в разных условиях эксперимента, что не позволяет напрямую сопоставлять количественные метрики между работами; тем не менее качественная картина сильных и слабых сторон каждой группы методов оказывается устойчивой в разных источниках. Перспективным направлением дальнейших исследований является разработка схем непрерывного дообучения моделей на потоковых данных, позволяющих системе адаптироваться к постепенному изменению нормального профиля сети без полного переобучения «с нуля».

Заключение

Проведённое сопоставление трёх групп методов машинного обучения — классификаторов с учителем, детекторов выбросов без учителя и моделей глубокого обучения — показало, что выбор конкретного подхода должен

определяться не абстрактным превосходством одного алгоритма над другим, а условиями эксплуатации защищаемой инфраструктуры: объёмом и скоростью трафика, допустимой задержкой реакции, стоимостью пропуска атаки и доступностью размеченных данных об инцидентах. Наиболее устойчивые к разнородным угрозам результаты демонстрируют двухступенчатые гибридные схемы, в которых быстрый детектор выбросов выполняет предварительную фильтрацию трафика, а более ресурсоёмкая модель глубокого обучения проводит уточняющий анализ подозрительного остатка. Сформулированные в статье критерии выбора метода и предложенная схема двухступенчатой архитектуры могут быть использованы при проектировании систем обнаружения вторжений для информационных систем разного масштаба. Направлением дальнейшей работы является экспериментальная проверка предложенной схемы на реальном трафике конкретной инфраструктуры и разработка механизма непрерывной адаптации модели к изменяющемуся профилю сети.

Библиографический список

1. Киселев Р.Г. Модели предсказания кибератак с использованием машинного обучения [Электронный ресурс] // Professional Bulletin: Information Technology and Security. – 2024. – URL: https://cyberleninka.ru/article/n/modeli-predskazaniya-kiberatak-s-ispolzovaniem-mashinnogo-obucheniya (дата обращения: 09.08.2026).
2. Власенко А.В., Дзьобан П.И., Жук Р.В. Обзор инструментов машинного обучения и их применения в области кибербезопасности [Электронный ресурс] // Прикаспийский журнал: управление и высокие технологии. – 2020. – URL: https://cyberleninka.ru/article/n/obzor-instrumentov-mashinnogo-obucheniya-i-ih-primeneniya-v-oblasti-kiberbezopasnosti (дата обращения: 09.08.2026).

3. Линдигрин А.Н. Сравнительный анализ методов машинного обучения в задачах обнаружения сетевых аномалий [Электронный ресурс] // Известия Тульского государственного университета. Технические науки. – 2019. – URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-metodov-mashinnogo-obucheniya-v-zadachah-obnaruzheniya-setevyh-anomaliy (дата обращения: 09.08.2026).
4. Голованов А.А., Мельникова О.И. Сравнительный анализ систем обнаружения аномалий с использованием потока сетевого трафика и протокола NetFlow [Электронный ресурс] // Международный научно-исследовательский журнал. – 2023. URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-sistem-obnaruzheniy-anomaliy-c-ispolzovaniem-potoka-setevogo-trafika-i-protokola-netflow (дата обращения: 09.08.2026).
5. Зуев В.Н. Обнаружение аномалий сетевого трафика методом глубокого обучения [Электронный ресурс] // Программные продукты и системы. – 2021. – URL: https://cyberleninka.ru/article/n/obnaruzhenie-anomaliy-setevogo-trafika-metodom-glubokogo-obucheniya (дата обращения: 09.08.2026).
6. Браницкий А.А., Котенко И.В. Обнаружение сетевых атак на основе комплексирования нейронных, иммунных и нейронечетких классификаторов [Электронный ресурс] // Информационно-управляющие системы. – 2015. – URL: https://cyberleninka.ru/article/n/obnaruzhenie-setevyh-atak-na-osnove-kompleksirovaniya-neyronnyh-immunnyh-i-neyronechetkih-klassifikatorov (дата обращения: 09.08.2026).
7. Федорова В.С., Стригунов В.В. Решение одной задачи обнаружения аномалий сетевого трафика с помощью сверточной нейронной сети [Электронный ресурс] // Вестник Тихоокеанского государственного университета. – 2024. – URL: https://cyberleninka.ru/article/n/reshenie-odnoy-zadachi-obnaruzheniya-anomaliy-setevogo-trafika-s-pomoschyu-svertochnoy-neyronnoy-seti (дата обращения: 09.08.2026).

8. Тимочкина Т.В., Татарникова Т.М., Пойманова Е.Д. Применение нейронных сетей для обнаружения сетевых атак [Электронный ресурс] // Известия высших учебных заведений. Приборостроение. – 2021. – URL: https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-dlya-obnaruzheniya-setevyh-atak (дата обращения: 09.08.2026).