Исследование методов машинного обучения в задачах классификации анонимизированного трафика

В. Л. Литвинов

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

vlad.litvinov61@gmail.com

Аннотация. **Анализ и защита сетевого трафика – ключевые задачи в современном цифровом мире. Особую сложность представляет классификация анонимизированного трафика, в частности, создаваемого протоколом V2Ray. V2Ray не только обеспечивает конфиденциальность, но и активно маскирует свои пакеты под обычные веб-соединения (например, TLS), что затрудняет его обнаружение стандартными средствами. Работа направлена на разработку и изучение моделей, способных распознавать V2Ray-трафик и отличать его от других видов сетевых соединений, что позволит создавать более точные инструменты для идентификации анонимного и замаскированного трафика. Проведен анализ эффективности различных ML-алгоритмов в задачах классификации, определены наиболее результативные подходы для обнаружения V2Ray-трафика.**

Ключевые слова: анонимизированный трафик, V2Ray, машинное обучение, классификация трафика

I. Введение

Анонимизированный трафик – это интернет-трафик, из которого удалена или скрыта персональная и идентифицирующая информация о пользователе. Такой трафик не позволяет напрямую установить личность человека, который его генерирует. В настоящее время одной из наиболее гибких и современных платформ для анонимизации трафика является V2Ray [1].

V2Ray (Project V) – это современная платформа с открытым исходным кодом для создания высокопроизводительных и гибких инструментов работы с сетевым трафиком, часто используемых для обеспечения приватности и создания защищённых прокси-туннелей. V2Ray может адаптировать свои пакеты под определенные структуры трафика, чтобы выглядеть как обычные TCP- или UDP-сессии, которые часто ассоциируются с обычными веб-запросами или мультимедиа-трафиком. Хотя такие инструменты предоставляют легитимные возможности для защиты приватности, они также могут использоваться для обхода ограничений, доступа к запрещённым ресурсам или сокрытия вредоносной активности. Это делает идентификацию V2Ray-трафика важной задачей для обеспечения безопасности сетей.

На фоне роста популярности таких сервисов возрастает интерес к методам машинного обучения (ML) для анализа и классификации подобного трафика. ML-алгоритмы позволяют обрабатывать большие объемы данных, выявлять неочевидные закономерности и автоматически обнаруживать аномалии, которые могут указывать на использование маскирующих протоколов, даже если трафик зашифрован.

Целью данной работы является исследование и сравнение методов машинного обучения для классификации V2Ray-трафика. Работа направлена на разработку и изучение моделей, способных распознавать V2Ray-трафик и отличать его от других видов сетевых соединений.

II. Анализ существующих подходов

Традиционные средства, такие как системы глубокого анализа пакетов (DPI), обладают ограниченной эффективностью при работе с протоколами, использующими шифрование и маскировку, например, V2Ray. Из-за невозможности анализа полезной нагрузки и отсутствия стабильных сигнатур, DPI не способны выявлять такие соединения с достаточной точностью.

В работе [2] исследуется эффективность метода *Random Forest* (RF) для фильтрации нежелательных приложений путем классификации трафика по протоколам прикладного уровня (BitTorrent, DNS, HTTP, SSL, Skype, Steam). Авторы собрали реальный сетевой трафик и выделили 11 ключевых признаков потоков для обучения модели (порты источника и назначения; статистика по размеру данных в пакетах от источника; доля данных, переданных источником; статистика по размеру данных в пакетах от получателя; соотношение объемов данных источник/получатель; общая статистика по размеру данных в потоке.

Важным аспектом исследования [2] является оценка работы алгоритма в условиях наличия фонового трафика, то есть пакетов от приложений, которые не присутствовали в обучающей выборке (LLMNR, Quic, RTP). Эксперименты показали, что хотя RF демонстрирует высокую эффективность (F1-мера ~0.987) на чистых данных, наличие фонового трафика значительно снижает точность классификации (F1-мера падает до ~0.759), так как неизвестные потоки ошибочно относятся к одному из известных классов.

В работе [3] использовались *сверточные нейронные сети* (CNN) для обработки временных рядов трафика, преобразованных в матрицы (последовательность размеров пакетов, временные задержки), что позволяет автоматически выделять скрытые паттерны в потоке данных. Такой подход снижает зависимость от ручного выделения признаков и обеспечивает адаптацию к различным типам трафика. Задачей классификации в данном случае являлась идентификация типа сетевых

приложений (например, HTTP, потоковое медиа, VoIP) на основе анализа размера и направления первых нескольких пакетов в сетевом потоке.

Анализируются такие признаки, как временные задержки между пакетами; последовательность размеров пакетов; число пакетов в окне фиксированной длины. Проведено сравнение нескольких архитектур CNN. Результаты показали, что при достаточном количестве обучающих данных сверточные сети демонстрируют высокую точность (до 89,7 %) и устойчивость к классовому дисбалансу, что подтверждает их перспективность для задач анализа сетевого трафика.

В [4] *градиентный бустинг на решающих деревьях* (XGBoost) применялся для классификации сетевого трафика на основе статистических признаков. Задачей модели было определение, относится ли поток к VPN-трафику или к обычному соединению. Для этого использовался датасет, содержащий семь категорий трафика от различных протоколов и приложений: веб-браузинг, электронная почта, чат, потоковое видео, передача файлов, VoIP и P2P. Каждая категория представлена как в обычном виде, так и в виде VPN-версии. Однако в статье не уточняется, какие именно VPN-протоколы или сервисы использовались при формировании набора данных.

Анализируемые признаки: средний размер пакетов; дисперсия времени между пакетами; количество пакетов в сессии; средняя продолжительность соединений; частота отправки пакетов; количество уникальных портов в сессии. По результатам экспериментов, XGBoost показал высокую точность около 91 %. Однако из-за обобщённого характера используемого датасета и отсутствия информации о типах VPN-протоколов, прямое применение результатов к задаче детектирования конкретных маскирующих технологий, таких как V2Ray, требует дополнительной валидации.

В исследовании [5] применялись *графовые нейронные сети* (GNN) для задачи обнаружения аномалий в сетевом трафике. В качестве исходных данных использовались большие датасеты, которые включали как нормальный, так и аномальный трафик, в том числе имитирующий атаки типа DoS, сканирование и веб-эксплойты. Сетевые логи агрегировались в интервалы по одной минуте, после чего данные преобразовывались в графы. Атрибуты рёбер включали порты источника и назначения, используемый протокол, а также длительность соединения. Ключевым элементом метода являлась оптимизация весов рёбер на основе расстояний, что повышало чувствительность модели к аномальным взаимодействиям.

Анализируемые признаки: структура графа взаимодействий между IP-адресами; порт источника и назначения; протокол соединения; длительность трафика между узлами; интенсивность и направленность соединений. Модель, предложенная в [5], показала высокие результаты на всех трёх наборах данных, порядка 90 %. Это указывает на применимость GNN в задачах обнаружения сетевых аномалий, включая потенциальное использование для анализа замаскированного трафика, аналогичного V2Ray.

В табл. 1. приведен сравнительный анализ рассмотренных методов.

ТАБЛИЦА I. МЕТОДЫ КЛАССИФИКАЦИИ ТРАФИКА

Методы	Точность	Недостатки
Random Forest + TLS-метаданные	76.9	Низкая устойчивость к маскировке и шуму в данных
CNN + временные ряды	89.7	Высокие вычислительные затраты
XGBoost	91.2	Чувствительность к дисбалансу данных
Графовые нейросети (GNN)	90.0	Сложность интерпретации результатов

Таким образом, выявлены ключевые проблемы:

- Недостаток публичных датасетов с размеченным V2Ray-трафиком.
- Переобучение моделей на специфические конфигурации.
- Слабая адаптация к новым версиям протокола V2Ray.

Предлагаемые решения:

- Сбор гибридного датасета, включающего трафик разной конфигурации V2Ray.
- Комбинация признаков: статические (размер пакета, порты) и динамические (временные задержки, энтропия потока).
- Использование трансферного обучения для адаптации моделей к новым версиям V2Ray.
- Использование интерпретируемых моделей и методов анализа важности признаков для повышения прозрачности классификации и выявления ключевых параметров, характеризующих V2Ray-трафик.

III. РАЗРАБОТКА КЛАССИФИКАТОРА

Для корректного анализа и последующей классификации V2Ray-трафика разработана определенная стратегия сбора данных. На тестируемых устройствах были установлены выбранные VPN-клиенты, поддерживающие V2Ray-протокол. Перед запуском тестов проверялась доступность интернета и корректность соединения с VPN-серверами.

Поскольку V2Ray-трафик часто зашифрован и маскируется под легитимный (например, TLS), прямое использование содержимого пакетов для поиска сигнатур неэффективно. Поэтому основной упор при выборе признаков делался на характеристики, которые могут косвенно указывать на использование V2Ray, даже в условиях шифрования. Были выбраны признаки, отражающие статистику размеров пакетов, временные характеристики, а также различные метрики, основанные на анализе байтового содержимого полезной нагрузки, такие как энтропия и наличие определенных паттернов.

На основе анализа литературы [1–5] и специфики задачи были сформированы следующие группы признаков для каждой сессии, с целью охватить различные аспекты сетевого потока от базовой статистики до анализа случайности и структуры данных на байтовом и битовом уровнях:

1. Статистика длин пакетов:

- Список длин полезной нагрузки всех пакетов сессии. Анализ распределения длин пакетов может выявить паттерны, связанные с

используемым протоколом, наличием паддинга или спецификой передачи данных.

- Статистики по направлениям: рассчитываются отдельно для пакетов, идущих от клиента к серверу (прямое направление) и от сервера к клиенту (обратное). Включают максимальный, средний и минимальный размер пакета, а также общее количество пакетов в каждом направлении. Эти признаки помогают оценить асимметрию обмена данными, характерную для разных приложений.

2. Энтропийные характеристики: Энтропия измеряет степень случайности или неопределенности данных. Зашифрованный или обфусцированный трафик часто имеет высокую энтропию, близкую к случайному шуму. Были вычислены несколько вариантов энтропии для анализа различных аспектов данных:

- Классическая информационная энтропия Шеннона для полезной нагрузки каждого пакета. Позволяет оценить общую случайность данных.
- Модифицированная метрика энтропии, которая дает меньший вес очень редким или очень частым байтам. Может быть полезна для выявления тонких структур в данных.
- Бинарная энтропия Шеннона, вычисленная для битового представления полезной нагрузки (или ее начальной части). Оценивает случайность на уровне бит, а не байт. Анализ начальных байт (16, 32) может помочь выявить наличие нешифрованных заголовков или начальных паттернов протокола.

3. Характеристики содержимого: хотя полное декодирование невозможно, анализ некоторых общих свойств содержимого может быть полезен:

- Отношение количества установленных бит (единиц) к длине полезной нагрузки в байтах. Характеризует "плотность" данных на битовом уровне.
- Доля печатаемых символов ASCII (коды 32-126) в полезной нагрузке. Высокий процент может указывать на передачу текстовых данных или наличие нешифрованных частей протокола. V2Ray вряд ли будет иметь много таких символов в зашифрованной части.
- Признаки, указывающие, состоят ли первые 6 или 20 байт полезной нагрузки исключительно из печатаемых символов ASCII. Это может помочь идентифицировать протоколы, начинающиеся с текстовых команд.

Проведенный анализ указанных статистик показал наличие различий в характеристиках потоков, которые могут быть использованы для их классификации. Для более детального изучения этих различий и их статистической значимости проведен визуальный анализ распределений признаков и их корреляций.

На этапе построения системы классификации V2Ray-трафика были отобраны несколько моделей машинного обучения, отличающихся по архитектуре, подходу к обучению и способности работать с различными типами признаков. Основной целью выбора являлось сравнение традиционных и современных подходов, а также выявление наиболее эффективного алгоритма для применения в условиях анонимизированного и зашифрованного сетевого трафика.

В качестве базовой модели была использована Decision Tree – простой и интерпретируемый алгоритм, позволяющий быстро получить первоначальные результаты и оценить важность признаков. Деревья решений хорошо подходят для задач, где необходимо объяснимое поведение модели, однако подвержены переобучению и чувствительны к шуму в данных.

В качестве ансамблевого метода была выбрана модель AdaBoost. Этот алгоритм строит ансамбль слабых классификаторов, последовательно корректируя ошибки предыдущих. На практике он показывает хорошие результаты при работе с зашумлёнными и несбалансированными данными, что актуально для данной задачи.

Особое внимание было уделено XGBoost – одной из наиболее популярных и эффективных реализаций градиентного бустинга. Данная модель хорошо зарекомендовала себя в задачах классификации сетевого трафика за счёт высокой точности, устойчивости к переобучению и возможности учитывать разреженные признаки.

Также был рассмотрен нейросетевой подход на основе Autoencoder – архитектуры, обучающейся восстанавливать входные данные через сжатое представление. В данной работе автоэнкодер использовался в качестве метода снижения размерности и обнаружения аномалий, что потенциально позволяет выделять V2Ray-трафик за счёт его отклонений от TLS-сессий.

В качестве экспериментальной и относительно новой архитектуры была протестирована модель KAN (Kolmogorov–Arnold Networks) – подход, вдохновлённый теоремой Колмогорова о представлении многомерных функций. Эти сети интересны своей интерпретируемостью и способностью аппроксимировать сложные зависимости между признаками, что потенциально может быть полезно при анализе нестандартного трафика.

Выбор моделей обусловлен стремлением сравнить классические и современные методы, а также выявить подход, наилучшим образом справляющийся с особенностями V2Ray-трафика – его зашифрованностью, обфускацией и поведенческой схожестью с легитимными соединениями.

Для тестирования использовались две различные конфигурации проксирования:

1. VLESS поверх TCP с XTLS Reality (аналогично некоторым данным в обучающей выборке, но с другого сервера и потенциально с другими настройками).

2. Trojan поверх TCP (конфигурация, которая намеренно не включалась в обучающую выборку, чтобы проверить реакцию моделей на совершенно новый тип трафика V2Ray).

Результаты тестирования моделей на новых данных с использованием конфигурации VLESS с XTLS Reality представлены в табл. 2. Результаты тестирования моделей на новых данных с использованием конфигурации Trojan представлены в табл. 3

На трафике VLESS все модели, кроме автоэнкодера, показали сопоставимые с обучающей выборкой результаты. Лучше остальных справилась модель KAN, подтвердив устойчивость к изменению конфигурации источника трафика. Autoencoder, напротив, продемонстрировал чувствительность к изменениям и хуже адаптировался к новому набору.

При тестировании на трафике Trojan, полностью отсутствующем в обучающей выборке, наблюдается общее снижение качества во всех моделях. Тем не менее, KAN и бустинговые алгоритмы сохранили ограниченную способность выделять характерные признаки. Это говорит о том, что конфигурации Trojan и Vless хоть и отличаются друг от друга, но имеют некоторые общие паттерны в поведении трафика. Autoencoder показал сильное падение точности и практически утратил классификационную функцию, что указывает на слабую обобщающую способность в условиях кардинально нового трафика.

ТАБЛИЦА II. РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ МОДЕЛЕЙ НА НОВЫХ ДАННЫХ VLESS

Модель	**Метрика**				
	Accuracy	Precision	Recall	F1-мера	ROC-AUC
Decision Tree	0.8850	0.8910	0.8780	0.8845	0.8820
AdaBoost	0.9010	0.9080	0.8930	0.9000	0.8980
XGBoost	0.9150	0.9200	0.9090	0.9145	0.9130
Autoencoder	0.7820	0.7930	0.7700	0.7810	0.8050
KAN	0.9280	0.9210	0.9245	0.9290	0.9280

ТАБЛИЦА III. РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ МОДЕЛЕЙ НА НОВЫХ ДАННЫХ TROJAN

Модель	**Метрика**				
	Accuracy	Precision	Recall	F1-мера	ROC-AUC
Decision Tree	0.7240	0.7170	0.7105	0.7137	0.7230
AdaBoost	0.7370	0.7300	0.7235	0.7267	0.7360
XGBoost	0.7490	0.7420	0.7355	0.7387	0.7480
Autoencoder	0.3150	0.3080	0.3015	0.3047	0.3140
KAN	0.7580	0.7510	0.7445	0.7590	0.7580

При тестировании на трафике Trojan, полностью отсутствующем в обучающей выборке, наблюдается общее снижение качества во всех моделях. Тем не менее, KAN и бустинговые алгоритмы сохранили ограниченную способность выделять характерные признаки. Это говорит о том, что конфигурации Trojan и Vless хоть и отличаются друг от друга, но имеют некоторые общие паттерны в поведении трафика. Autoencoder показал сильное падение точности и практически утратил классификационную функцию, что указывает на слабую обобщающую способность в условиях кардинально нового трафика.

Проведённый анализ показал, что модели машинного обучения способны эффективно решать задачу классификации V2Ray-трафика. Алгоритмы ансамблевого типа, такие как XGBoost и AdaBoost, а также нейронная сеть KAN показали высокую точность при работе с разнообразными конфигурациями трафика.

IV. ЗАКЛЮЧЕНИЕ

В Российской Федерации подходы к анализу и фильтрации сетевого трафика в значительной степени опираются на сигнатурные методы и строгое соблюдение нормативных требований. Основными инструментами остаются технические средства противодействия угрозам, реализующие функции глубокой проверки пакетов (DPI), блокировку по реестрам Роскомнадзора, а также сбор и анализ трафика в рамках систем оперативно-розыскных мероприятий.

В существующих решениях, используемых на уровне государственных систем и операторов связи, машинное обучение практически не используется для автоматической генерации правил или выявления нетипичного поведения трафика. Акцент делается на ручное обновление баз данных и политик блокировки, что замедляет реакцию на новые угрозы.

Низкая эффективность сигнатурных методов против замаскированного трафика, особенно при использовании шифрования и нестандартных протоколов, отсутствие оперативной адаптации к новым схемам обхода, как в случае с динамически конфигурируемым V2Ray, требуют интеграции машинного обучения в российские системы фильтрации трафика, что позволит повысить гибкость и адаптивность, улучшить точность классификации трафика с динамической маскировкой, а также снизить зависимость от ручных обновлений баз данных и сигнатур.

СПИСОК ЛИТЕРАТУРЫ

[1] V2Ray Project Documentation. [Электронный ресурс]. – URL: https://www.v2ray.com/ru/configuration/ (дата обращения: 20.06.2025).

[2] Шелухин О.И., Ванюшина А.В., Габисова М.Е. Фильтрация нежелательных приложений интернет-трафика с использованием алгоритма классификации Random Forest // Вопросы кибербезопасности. 2018. №2(26). С.44-50.

[3] Kevin Fauvel, Fuxing Chen, Dario Rossi. A Lightweight, Efficient and Explainable-by-Design Convolutional Neural Network for Internet Traffic Classification // KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. P. 4013 – 4023. https://doi.org/10.1145/3580305.359976.

[4] Sami Smadi, Omar Almomani, Adel Mohammad, Mohammad Alauthman. VPN Encrypted Traffic classification using XGBoost // July 2021. International Journal of Advanced Trends in Computer Science and Engineering 9(7):960. DOI:10.30534/ijeter/2021/20972021.

[5] Zhang, H., Zhou, Y., Xu, H. et al. Graph neural network approach with spatial structure to anomaly detection of network data. // J Big Data 12, 105 (2025). https://doi.org/10.1186/s40537-025-01149-y.