УДК 004.85 https://doi.org/10.26160/2541-8637-2026-17-79-81

СИСТЕМА ОБНАРУЖЕНИЯ ВРЕДОНОСНОГО C2-ТРАФИКА НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Вульфин А.М., Кириллова А.Д., Башмаков Н.М.
Уфимский университет науки и технологий, Уфа

Ключевые слова: C2-трафик, машинное обучение, CatBoost, XGBoost, RandomForest, k-ближайшие соседи.
Аннотация. В работе решается задача обнаружения C2-трафика в компьютерных сетях с использованием методов машинного обучения. Создан стенд для исследования трафика фреймворков управления ботнетами Sliver и Metasploit. Проведен эксперимент по построению моделей бинарной классификации с использованием CatBoost, XGBoost, RandomForest и kNN с оптимизацией гиперпараметров. Показана высокая эффективность алгоритмов градиентного бустинга в случае одного источника данных и критическое снижение обобщающей способности при переносе модели на сетевой трафик иных реализаций C2-каналов.

A MACHINE LEARNING-BASED SYSTEM FOR DETECTING MALICIOUS C2 TRAFFIC

Vulfin A.M., Kirillova A.D., Bashmakov N.M.
Ufa University of Science and Technology, Ufa

Keywords: C2 traffic, machine learning, CatBoost, XGBoost, RandomForest, k-nearest neighbors.
Abstract. This paper examines the problem of detecting C2 traffic in computer networks using machine learning methods. A setup for analyzing traffic using the Sliver and Metasploit frameworks was deployed. A binary classification experiment was conducted using CatBoost, XGBoost, RandomForest, and kNN with hyperparameter optimization. High efficiency of gradient boosting within a single data source and a critical reduction in generalization ability when transferring to other C2 channel implementations were identified.

Современные кибератаки предполагают долговременное присутствие злоумышленника в инфраструктуре через сервера управления и контроля (Command & Control, C2). Обнаружение C2-трафика является приоритетной задачей для систем мониторинга сетевой безопасности, однако маскировка вредоносного взаимодействия под HTTP/HTTPS и использование обратных подключений снижают эффективность сигнатурных методов. Перспективным направлением решения задачи является применение методов машинного обучения для выявления аномалий в сетевых потоках [1]. Однако многие существующие исследования основаны на использовании устаревших наборов данных, не отражающих современные тактики C2-фреймворков [1, 2].

Цель работы – экспериментальная оценка обнаружения C2-трафика фреймворков Sliver и Metasploit методами машинного обучения и оценка обобщающей способности обученных моделей на независимых данных.

Для формирования репрезентативной выборки реализован виртуальный стенд, имитирующий сегмент локальной вычислительной сети. Стенд включал виртуальный маршрутизатор для захвата трафика (с помощью tcpdump) и несколько подсетей с рабочими станциями под управлением ОС Windows. В

каждой подсети одна станция выступала в роли зараженного узла: Sliver использовал HTTP-канал (beacon-режим), Metasploit – reverse TCP.

За неделю собрано 2,5 млн сессий Sliver и 405 тыс. сессий Metasploit (с сильным дисбалансом классов: 404 тыс. против 664). Для Metasploit применена балансировка методом Random Under Sampling.

Каждая сессия описывалась 79 признаками. На этапе подготовки данных были исключены служебные поля. Применены медианное заполнение пропусков, робастное масштабирование для числовых признаков и кодирование категориальных (ordinal, one-hot). Разделение на обучающую и тестовую выборки – 80/20 со стратификацией.

Для решения задачи бинарной классификации использованы алгоритмы: CatBoost, XGBoost, RandomForest и k-ближайших соседей (kNN). Гиперпараметры всех моделей оптимизировались автоматически с помощью библиотеки Optuna (100 итераций). Оценка качества производилась по метрикам precision, recall, F1-measure (взвешенная), ROC-AUC и G-mean.

Классификация трафика Sliver. Результаты экспериментов на датасете Sliver показали высокую эффективность всех ансамблевых методов (табл. 1).

Табл. 1. Сравнение моделей на тестовой выборке Sliver

Модель	Точность	Полнота	F1-мера	Средняя F1	G-mean
CatBoost	0,95	0,99	0,97	0,98	0,98
XGBoost	0,95	0,99	0,97	0,98	0,98
RandomForest	0,95	0,99	0,97	0,97	0,98
kNN	0,93	0,98	0,95	0,97	0,97

Модели CatBoost, XGBoost продемонстрировали наилучшие и наиболее стабильные результаты, незначительно опережая RandomForest. Метод kNN показал некоторое снижение точности на тестовой выборке, что объясняется его чувствительностью к локальной плотности данных. Визуализация признакового пространства подтвердила наличие кластерной структуры, где вредоносный трафик Sliver формирует обособленные группы.

Классификация трафика Metasploit. Из-за специфики данных (малое количество и однотипность вредоносных сессий) модель CatBoost показала практически идеальные результаты на сбалансированной выборке (F1-мера = 1,0). Однако это свидетельствует не столько о сложности задачи, сколько о высокой степени детерминированности и уникальности трафика Metasploit [3] в рамках данного стенда.

Ключевым этапом исследования стала проверка моделей на независимых данных. Модель, обученная на трафике Sliver, была протестирована на наборах данных Metasploit и Mythic (сторонний фреймворк). Результаты представлены в таблице 2.

Аналогичная ситуация наблюдается и при обратной проверке: модель, обученная на Metasploit, не обнаруживает трафик Sliver и Mythic (полнота 0%). Это свидетельствует о фундаментальных различиях в характере сетевого взаимодействия исследуемых фреймворков. Sliver генерирует регулярные, статистически устойчивые паттерны, в то время как Metasploit формирует

короткие, событийно-ориентированные сессии, статистически близкие к легитимному трафику.

Табл. 2. Тестирование модели (обученной на Sliver) на других датасетах

Тестовый датасет	Точность	Полнота	F1-мера	Вывод
Mythic	0,97	0,95	0,96	Высокая обобщающая способность
Metasploit	0,00	0,00	0,00	Полная потеря детектирующей способности.

Расширение обучающей выборки. Для преодоления выявленного ограничения был сформирован объединенный датасет (Sliver + Metasploit). Модель CatBoost, обученная на объединенных данных, показала:

1. На трафике Metasploit. Полнота обнаружения достигла 1.0, однако точность упала до 0,15 из-за резкого роста числа ложных срабатываний.

2. На трафике Mythic. Сохранение высокого качества (F1 = 0,99), что подтверждает эффективность градиентного бустинга при наличии репрезентативных данных.

Исследование подтвердило высокую эффективность методов градиентного бустинга (CatBoost, XGBoost) для обнаружения C2-трафика в условиях, когда обучающая и тестовая выборки принадлежат к одному типу источников. Однако эксперимент выявил критическую проблему обобщения: модели, обученные на паттернах одного фреймворка (например, Sliver), не могут обнаруживать трафик другого (Metasploit) из-за различий в логике каналов управления. Расширение датасета разнородными примерами повышает полноту обнаружения, но может приводить к недопустимому уровню ложных срабатываний. Полученные результаты подчеркивают, что ключевым фактором успешного применения машинного обучения в задачах сетевой безопасности является не столько сложность алгоритма, сколько полнота и разнообразие сценариев, представленных в обучающей выборке.

Список литературы

1. Wang Z., Fok K.W., Thing V.L.L. Machine learning for encrypted malicious traffic detection: Approaches, datasets and comparative study // Computers & Security. 2022, vol. 113, p. 102542.
2. Бондарев М.В., Гузев О.Ю. Возможности и ограничения классификации атак в зашифрованном трафике методами машинного обучения // International Journal of Open Information Technologies. – 2025. – Т. 13, №.7. – С. 27-36.
3. Xavier G., Novo C., Morla R. Tweaking Metasploit to Evade Encrypted C2 Traffic Detection // arXiv preprint arXiv:2209.00943. 2022.

Сведения об авторах:

Вульфин Алексей Михайлович – д.т.н., профессор кафедры вычислительной техники и защиты информации (ВТиЗИ);

Кириллова Анастасия Дмитриевна – к.т.н., доцент кафедры ВТиЗИ.

Башмаков Наиль Маратович – ассистент кафедры ВТиЗИ.