УДК 004.056

Д.А. МАРГАРЯН, Д.С. ЮДИНА

Национальный исследовательский ядерный университет «МИФИ», Москва

ВЫБОР ОПТИМАЛЬНОЙ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ СИСТЕМЫ ОБНАРУЖЕНИЯ DDOS-АТАК

Распределённые атаки типа «отказ в обслуживании» (DDoS) остаются одной из ключевых угроз сетевой безопасности. В работе проведено сравнение двух алгоритмов машинного обучения для обнаружения аномалий – Isolation Forest и One-Class SVM. На основе набора данных CIC-DDoS2019 установлено, что Isolation Forest обеспечивает более высокие показатели точности и полноты, демонстрируя устойчивость к несбалансированным данным. Полученные результаты позволяют рекомендовать данный алгоритм как базовый инструмент для построения систем раннего обнаружения DDoS-атак.

Современная цифровая инфраструктура критически зависит от бесперебойной работы сетевых сервисов [1]. Распределённые атаки типа DDoS направлены на исчерпание ресурсов сервера или сети. Традиционные методы, основанные на сигнатурах, неэффективны против новых гибридных атак, поэтому применяются алгоритмы машинного обучения, способные выявлять аномалии в сетевом трафике [2].

В исследовании проведено сравнение моделей Isolation Forest и One-Class SVM. Isolation Forest изолирует редкие наблюдения посредством случайных деревьев, тогда как One-Class SVM строит границу, отделяющую нормальные данные от начала координат [3, 4]. Метрики оценки включали точность, полноту, F1-меру и AUC-ROC.

Эксперименты на наборе CIC-DDoS2019 показали, что Isolation Forest обеспечивает точность 0.98, полноту 0.96 и AUC-ROC 0.99, тогда как One-Class SVM – 0.94, 0.91 и 0.95 соответственно. Таким образом, Isolation Forest демонстрирует лучшие результаты и устойчивость к несбалансированным данным.

Дополнительно было проведено сравнение стабильности работы моделей при изменении объёма обучающей выборки и степени шумности данных. Результаты показали, что алгоритм Isolation Forest демонстрирует меньшую чувствительность к выбросам и способен сохранять высокие показатели точности даже при снижении доли обучающих данных на 30%. В то время как One-Class SVM в подобных условиях терял до 7% точности и демонстрировал рост ложноположительных срабатываний.

Анализ вычислительной эффективности показал, что Isolation Forest требует меньше ресурсов по сравнению с One-Class SVM, особенно при обработке больших объёмов потоковых данных. Среднее время предсказания для одной записи оказалось на 25–30% меньше, что делает возможным его применение в системах обнаружения атак в режиме реального времени.

Алгоритм Isolation Forest можно рекомендовать для реализации систем раннего обнаружения DDoS-атак. Перспективы дальнейших исследований включают использование ансамблей и нейронных сетей для анализа временных рядов сетевого трафика с целью повышения точности и адаптивности обнаружения. Традиционные сигнатурные подходы, применяемые в системах обнаружения вторжений, не способны распознавать новые или модифицированные типы атак, поскольку они основаны на заранее известных шаблонах вредоносного поведения. В отличие от них, методы машинного обучения позволяют выявлять ранее неизвестные аномалии за счёт анализа статистических закономерностей в сетевом трафике, что делает их особенно актуальными для современных систем защиты. Датасет CIC-DDoS2019 содержит более 80 типов сетевых атак, включая UDP Flood, SYN Flood, HTTP Flood и другие, что делает его эталонным для тестирования моделей обнаружения DDoS.

Список литературы

1. Назаров А.Ш., Ли И.Т. DDoS-атаки и средства защиты от них // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2023. – № 1(61). – С. 42–45.
2. Терновой О.С. Раннее обнаружение DDOS атак на основе статистического анализа // Перспективы развития информационных технологий. – 2011. – № 6. – С. 212–215.
3. Шелухин О.И., Полковников М.В. Исследование алгоритма Isolation Forest при бинарной классификации сетевых аномалий // Безопасные информационные технологии: Сборник трудов Десятой международной научно-технической конференции, Москва, 03–04 декабря 2019 года. – Москва: Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), 2019. – С. 387–393
4. Ляпин А.Д. Сравнительный анализ алгоритмов SVM (Support Vector Machine) и RF (Random Forest) // Вестник науки. – 2025. – Т. 3, № 6(87). – С. 1816–1824.