УДК 004.942

ПОДХОД ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ НА ОСНОВЕ АНСАМБЛЕЙ ДЕРЕВЬЕВ РЕШЕНИЙ

Горбачев В.Д. (ВКА), **Моховиков В.А.** (ВКА), **Тельбух В.В.** (ВКА)
Научный руководитель – кандидат технических наук, преподаватель Тельбух В. В.
(Военно-космическая академия имени А.Ф.Можайского)

Введение. В условиях глобальной цифровизации проблема обеспечения информационной безопасности становится одной из наиболее актуальных. Согласно ежегодному отчету компании Curator (ранее Qrator Labs) за 2024 год, наблюдается значительный рост числа сетевых аномалий, включающих DDoS-атаки, активность вредоносных ботов и инциденты, связанные с протоколом BGP. В частности, количество DDoS-атак увеличилось на 53% по сравнению с 2023 годом, причем наиболее уязвимыми оказались сегменты FinTech (25,8%), электронной коммерции (20,5%) и медиа (13,5%). Рекордная по мощности DDoS-атака достигла 1,14 Тбит/с, что на 65% превышает показатель предыдущего года [1].

В связи с этим все больше научных исследований и их практическая реализация посвящается разработки подходов автоматического анализа сетевого трафика и выявления отклонений от номинального состояния. Одним из перспективных подходов является использование ансамблей деревьев решений, которые способны эффективно обрабатывать многомерные данные и выявлять сложные паттерны аномальной активности.

Основная часть. Метод машинного обучения ансамбли деревьев решений включает в себя различные алгоритмы, одним из которых является алгоритм случайных лесов (Random Forests). Это алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев. Данный алгоритм создает несколько деревьев решений и объединяет их прогнозы. В задачах классификации каждое дерево формирует предсказание для класса, и выбирается класс, набравший наибольшее количество предсказаний. В задачах регрессии итоговый прогноз представляет собой среднее значение результатов отдельных деревьев [2].

Объединяя результаты нескольких деревьев, случайные леса часто обобщают данные лучше, чем одно дерево решений, сокращая переобучение и обеспечивая стабильную производительность даже в многомерных пространствах объектов.

Случайный лес обучается на данных, отражающих нормальные условия, и сравнивает новые точки данных с изученным поведением. Отклонения с низкой достоверностью прогнозов помечаются как аномалии, что делает метод эффективным для выявления подозрительного сетевого трафика.

На основание проведенного анализа предметной области и сделанных выводов выдвигается гипотеза, что применение ансамблей деревьев решений позволит обнаруживать аномалии за счёт построения поведенческих моделей сетевого взаимодействия и динамического анализа изменений в характеристиках трафика.

В качестве исходных данных использовались набор данных NSL-KDD, который улучшает исходный набор данных KDD Cup 1999 за счёт удаления избыточных записей и исправления несбалансированного распределения классов. Исследователи обычно используют его в качестве эталона для оценки эффективности различных моделей обнаружения вторжений.

Провелась предварительная обработка данных NSL-KDD, включающая в себя создание набора предопределённых функций и меток. Чтобы работать с ними напрямую, мы должны сопоставить эти функции с осмысленными именами столбцов. Мы определяем список имён столбцов, соответствующих различным наблюдаемым характеристикам сетевых подключений и атак. Названия столбцов обеспечивают правильную

идентификацию каждого признака и метки. Они включают в себя общую сетевую статистику (например, duration, src_bytes, dst_bytes), категориальные поля (protocol_type, service), а также метки (attack, level), которые классифицируют наблюдаемый тип трафика, что позволило выделить нормальные и аномальные состояния трафика.

В качестве модели для обнаружения аномалий в трафике использовалась ансамблевая методика, основанная на деревьях решений. Эффективность модели оценивалась на основе метрик Accuracy, Precision, Recall и F1-Score на валидационном и тестовом наборах данных [3].

В результате экспериментов установлено, что модель демонстрирует высокую точность в обнаружении аномалий. На валидационном наборе данных Accuracy составила 99.50%, F1-Score – 99.49%, а на тестовом наборе – 99.49% и 99.47% соответственно. Наибольшую сложность в классификации вызвал класс Privilege Access, для которого F1-Score составил 0.51 на валидационном и 0.34 на тестовом наборе. В то же время для других классов атак (Normal, DoS, Probe, Access) точность и полнота модели находятся на уровне 99-100%, что свидетельствует о высокой эффективности используемого подхода.

Таким образом, использованный ансамблевый метод демонстрирует высокую точность классификации, обеспечивая надежное обнаружение аномалий в сетевом трафике. Высокие показатели Accuracy и F1-Score для большинства классов подтверждают его эффективность.

Выводы. В ходе исследования была реализована методика выявления сетевых аномалий с использованием ансамблевых методов на основе деревьев решений. Проведенный анализ подтвердил высокую точность предложенного подхода в задаче обнаружения атак, особенно для классов Normal, DoS, Probe и Access, где точность и полнота достигли 99-100%.

Дальнейшее развитие предложенного метода позволит повысить точность детекции сложных атак и снизить уровень ложных срабатываний. Кроме того, использование данной методики может стать основой для создания адаптивных систем кибербезопасности, способных выявлять и предотвращать угрозы в режиме реального времени. Представленный подход может быть применен в системах поддержки принятия решений для защиты государственных и корпоративных сетей от кибератак.

Список использованных источников:

1. DDoS-атаки, боты и BGP-инциденты в 2024 году: статистика и тренды // Смотрим: [сайт]. – 2025. – URL: https://blog.qrator.net/ru/ddos-boty-i-bgp-incidenty-v-2024-godu-statistika-i_210/ (дата обращения: 29.01.2025).
2. Метод случайного леса // Смотрим: [сайт]. – 2025. – URL: https://ru.wikipedia.org/wiki/Метод_случайного_леса (дата обращения: 29.01.2025).
3. Обучение и оценка (обнаружение сетевых аномалий) // Смотрим: [сайт]. – 2025. – https://academy.hackthebox.com/module/292/section/3304 (дата обращения: 30.01.2025).