

ОБНАРУЖЕНИЯ ВЕБ-РОБОТОВ НА ОСНОВЕ АНАЛИЗА КЛИЕНТСКОЙ АКТИВНОСТИ В ВЕБ-ПРИЛОЖЕНИИ

Сигети П.И. (Университет ИТМО)

Научный руководитель – доцент, кандидат педагогических наук, Государев И.Б.
(Университет ИТМО)

Введение. При разработке веб-приложений, программисты используют различные фреймворки или CMS. Данные технологии во многом следуют современным трендам по обеспечению безопасности, но не гарантируют полноценную защиту. Сегодня злоумышленники применяют автоматизированные инструменты (веб-роботов), которые анализируют веб-ресурсы и пытаются найти уязвимости. Согласно исследованиям компаний, занимающихся защитой веб-сайтов, каждый пятый запрос осуществляется одним из таких веб-роботов [1].

Основная часть. В рамках текущей работы были выполнены следующие задачи:

- 1) Осуществлена реализация алгоритма по сбору клиентских данных на основе их запросов к веб-приложению. На стороне сервера собираются следующие данные из полученных запросов: IP-адрес, user-agent, accept language, cookie, время запроса, метод запроса и запрашиваемый адрес.
- 2) На основе полученных данных, запросы сгруппированы по сессиям. С помощью различных алгоритмов и библиотек собран датасет с признаками, позволяющие классифицировать сессии клиентов. Как результат, датасет включают в себя следующие характеристики: расположение IP-адреса, Интернет-провайдер клиента (ASN), используется ли IP-адреса как прокси, общее количество запросов, среднее время между запросами в рамках одной сессии, количество сессий, виды запросов к серверу (get, post и др.), поддержка cookie, посещение «запрещенных» или скрытых адресов (например, ограничения в robots.txt), используемые языки на устройстве клиента, изменение user-agent, классификация user-agent [2].
- 3) С помощью различных алгоритмов машинного обучения (дерево принятия решений, случайный лес, метод опорных векторов, алгоритм на основе нейронной сети и др.) решена задача классификации для собранного датасета. Рассчитаны метрики для каждого метода классификации, а именно Precision, Recall, F-мера и Accuracy.
- 4) Полученные метрики проанализированы, что позволило прийти к заключению об эффективности подходов к обнаружению веб-роботов [3].

Выводы. Проведен анализ работы различных алгоритмов, позволяющих определить роботов среди посетителей веб-сайта. Выбраны модели машинного обучения, показавшие лучшие метрики.

Список использованных источников:

1. Менщиков А. А., Гатчин Ю. А. Метод обнаружения веб-роботов на основе анализа графа пользовательского поведения // Программные продукты и системы. 2019. №4. URL: <https://cyberleninka.ru/article/n/metod-obnaruzheniya-veb-robotov-na-osnove-analiza-grafa-polzovatelskogo-povedeniya> (дата обращения: 15.02.2023).
2. Мельников Г.А., Карпук А.А. Обнаружение попыток несанкционированного доступа на основе выявления нестандартных поведенческих факторов // Universum: технические науки. 2021. №3-1 (84). URL: <https://cyberleninka.ru/article/n/obnaruzhenie-popytok-nesanktsionirovannogo-dostupa-na-osnove-vyyavleniya-nestandartnyh-povedencheskih-faktorov> (дата обращения: 15.02.2023).

3. Hanlin Chen, Hongmei He, Andrew Starr. An Overview of Web Robots Detection Techniques. 2020. URL: <https://ieeexplore.ieee.org/document/9138856> (дата обращения: 15.02.2023).