



НАУЧНЫЙ ЖУРНАЛ НАУКА И МИРОВОЗЗРЕНИЕ

МЕТОДЫ ОБНАРУЖЕНИЯ И СПОСОБЫ ПРОТИВОДЕЙСТВИЯ АВТОМАТИЗИРОВАННОМУ СБОРУ ДАННЫХ С ВЕБ РЕСУРСОВ

Иванов Александр Дмитриевич

студент кафедры Информационная безопасность, Московский Политехнический Институт, РФ, Москва

Кесель Сергей Александрович

научный руководитель, Московский Политехнический Институт, РФ, Москва

Аннотация. В современном мире одним из ценнейших ресурсов является информация. Люди стали чаще использовать веб ресурсы, такие как маркетплейсы, сервисы бронирования, агрегаторы объявлений. Появляются новые компании и растет конкуренция. Современные технологии позволяют использовать автоматизированные средства для сбора данных у конкурентов и автоматически предлагать более привлекательные услуги для клиента. Для противодействия автоматизированным средствам сбора информации необходим комплексный подход к безопасности веб ресурсов.

Противодействие автоматизированным средствам по сбору данных

Автоматизированные средства сбора данных с веб ресурсов (веб скраперы) могут нанести большой ущерб владельцу объекта защиты. Можно выделить несколько векторов атак:

Сбор информации с веб ресурса. Основная цель автоматизированных средств это получение информации с сайта. Собранные данные служат для конкурентов неким маркерами, способные изменять цены на своих площадках, агрегировать данные и собирать их на одном ресурсе, переманивая аудиторию; Заполнение форм; Замедление работоспособности объекта защиты; Поиск уязвимостей. Автоматизированные средства при анализе сайта могут проводить сканирование для поиска веб уязвимостей. Эти данные могут быть использованы для внедрения вредоносного кода для последующих атак. Бронирование мест в залах. Автоматические средства или физическое лицо может злонамеренно производить действия по бронированию или аренде помещения;

Для противодействия выделенным векторам атак необходим комплексный подход для обнаружения и противодействия веб скраперам. Разработанный модуль по обнаружению и противодействию ботам имеет несколько способов обнаружения автоматического сканирования. Одним из действенных способов обнаружения ботов, являются ловушки. Ловушки представляют из себя невидимые для человеческого глаза формы или ссылки. Ловушка скрывается при помощи определенных тегов или специальных правил верстки, чтобы пользователь не мог в нее попасть. При парсинге страницы боты попадут в ловушку, о чем получит сигнал средство блокировки.

Модуль обнаружения веб скраперов предполагает усовершенствование устоявшегося подхода. Многие разработчики встраивают статичные ловушки, но данный метод имеет ряд недостатков. К таким недостаткам относится статичность, это позволяет четко

идентифицировать ловушку и передавать информации о ее наличии. Чтобы устранить вышеуказанные недостатки, был разработан модуль безопасности для обнаружения и блокирования автоматизированных сборщиков. Модуль представляет из себя библиотеку, написанную на языке Python. Библиотека позволяет автоматически встраивать в страницу фронтенда ловушки разных типов. При новом обращении библиотека встраивает новые ловушки. Стоит отметить что в параметрах, заданы условия, которые позволяют увеличить шанс обнаружения веб скрапера.

Для более точного детектирования существует возможность встраивать несколько ловушек разного типа на одну страницу. Данная функция была разработана с целью подстраховки системы защиты. При возможном обнаружении ловушки, веб скраперы могут обходить ее. При встраивании двух ловушек разного типа шанс обнаружения веб скрапера повышается, что снижает риск выделенных угроз. После детектирования веб скрапера необходимо принять решение что дальше делать с запросом. В модуле предусмотрено два режима. Первый режим – режим наблюдения, он позволяет детектировать и записывать в лог аудиты произошедшее событие. Данный режим может быть полезен при отладке. Второй режим – это режим блокировки подключения. Он позволяет незамедлительно производить блокировка подключения бота. Блокировка производится по средствам блокировки ip адреса.

К минусу данного метода можно отнести возможность автоматизированных средств использовать прокси сервера для новых подключений. Блокировка подключения является одним из наиболее эффективных методов борьбы. Для обеспечения наибольшей безопасности стоит заблокировать подключения на своем веб сервер с известных прокси серверов. Злоумышленники часто используют внешние сервера, т.к. иметь несколько своих достаточно финансово затратная задача. К еще одному методу защиты можно отнести капчу. Данный метод эффективно применять при подозрительном трафике. Использование капчи не является эффективным методом борьбы, т.к. большинство умных веб скраперов умеют обходить все типы капч: звуковые, видео капчи, капчи с картинками. Наиболее вероятный сценарий — это вывод капчи легитимному пользователю что приведет к потере времени на ее решение и как следствие к негативным эмоциям.

После разработки модуля детектирования и блокировка, был создан тестовый стенд. Стенд представлял несколько страниц схожих по контенту с онлайн кинотеатром. Приложение было развёрнуто в двух докер контейнерах: база данных и само приложение. Данный стенд был создан для тестирования эффективности работоспособности модуля безопасности. В качестве средств автоматизированного сборщика данных были взяты бесплатные, доступные в сети интернет веб скраперы. К ним относится Octoparse и Outwit. В ходе испытаний модуль показал хороший результат. Около 78 процентов вредоносных подключений было обнаружено и заблокировано. Данный показатель указывает на высокую эффективность модуля защиты, что позволяет применять его в пром средах. Среди недостатков можно выделить умные веб скраперы, которые используют в своей основе искусственный интеллект и машинное обучение. Они начинают достаточно быстро понимать и выделывать похожие части кода и стараются их избегать. При построении защиты необходим комплексный подход, нельзя останавливаться на одном модуле защиты. Необходимо использовать программное обеспечение, работающее по разным принципам.