УДК 004.42
doi:10.21685/2227-8486-2021-4-9

ТЕХНОЛОГИИ ЗАЩИТЫ САЙТОВ ОНЛАЙН-АУКЦИОНОВ

Н. С. Рассказов[1], М. А. Кривцов[2], М. А. Митрохин[3]

[1, 2, 3, 4] Пензенский государственный университет, Пенза, Россия
[1] nikita_serdobsk12@mail.ru, [2] mikkri58@gmail.com, [3] mmax83@mail.ru

Аннотация. *Актуальность и цели*. Рассматриваются способы защиты площадок интернет-аукционов от автоматизированного программного обеспечения (ПО), использование которого приводит к недобросовестной конкуренции. *Методы.* Проводится анализ существующих методов защиты интернет-площадок от ботов, акцентируется внимание на инструментах, основанных на анализе поведения пользователя на сайте, таких как reCAPTCHA v3, и сторонних сервисах, предоставляющих услуги по защите. Также рассматривается существующий подход к анализу сетевой активности, базирующийся на основе техник машинного обучения. Описываются основные инструменты, использующиеся разработчиками ПО, предназначенного для автоматизации действий на сайтах, в том числе веб-драйвер Selenium, а также приемы, позволяющие обнаружить использование подобных инструментов. *Результаты и выводы.* В результате исследования был разработан программный модуль автоматизации, предназначенный для проверки защищенности выбранных сайтов двумя способами. При тестировании площадки с установленной reCAPTCHA v3 степень уязвимости сайта к работе автоматизированного ПО составила 75 %. Были выработаны и предложены наиболее эффективные способы обнаружения и блокировки ботов на сайтах, позволяющие значительно снизить уровень недобросовестной конкуренции на площадках интернет-аукционов.

Ключевые слова: интернет-аукцион, reCAPTCHA, WAF, защита от ботов, Selenium, Python

Для цитирования: Рассказов Н. С., Кривцов М. А., Митрохин М. А. Технологии защиты сайтов онлайн-аукционов // Модели, системы, сети в экономике, технике, природе и обществе. 2021. № 4. С. 96–105. doi:10.21685/2227-8486-2021-4-9

SITE PROTECTION TECHNOLOGIES ON THE EXAMPLE OF ONLINE AUCTION SITES

N.S. Rasskazov[1], M.A. Krivtsov[2], M.A. Mitrokhin[3]

[1, 2, 3, 4] Penza State University, Penza, Russia
[1] nikita_serdobsk12@mail.ru, [2] mikkri58@gmail.com, [3] mmax83@mail.ru

Abstract. *Background.* The methods of protecting Internet auction sites from automated software, the use of which leads to unfair competition, are considered. *Materials and methods*. The analysis of existing methods of protecting Internet sites from bots is carried out, the attention is focused on tools based on the analysis of user behavior on the site, such as reCAPTCHA v3 and third-party services that provide protection services. The existing approach to the analysis of network activity based on machine learning techniques is also

considered. It describes the main tools used by developers of software designed to automate actions on websites, including the Selenium web driver, as well as techniques to detect the use of such tools. *Results and conclusions.* As a result of the study, an automation software module was developed designed to check the security of selected sites in two ways, when testing a site with reCAPTCHA v3 installed, the vulnerability of the site to the work of automated software was 75 %. The most effective ways of detecting and blocking bots on websites have been developed and proposed, which significantly reduce the level of unfair competition on Internet auction sites.

Keywords: online auction, reCAPTCHA, WAF, protection from bots, Selenium, Python

For citation: Rasskazov N.S., Krivtsov M.A., Mitrokhin M.A. Site protection technologies on the example of online auction sites. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, systems, networks in economics, technology, nature and society*. 2021;(4):96–105. (In Russ.). doi:10.21685/2227-8486-2021-4-9

Введение

Электронный аукцион – аукцион, проводящийся посредством интернет-ресурсов, где ставки производятся через сайт или компьютерную программу аукциона. Последнее время наблюдается бурный рост электронной коммерции в виде интернет-аукционов, этому способствует рост числа пользователей глобальной сети, простота организации и возможность использования разнообразных бизнес-моделей, в том числе стандартной, голландской, двойной и т.д. [1].

Следует отметить, что одновременно с ростом аудитории возрастает количество программных решений, помогающих получать конкурентное преимущество на данных площадках. Так, согласно исследованию Bad Bot Report 2021, сегодня до 50 % всего онлайн-трафика составляют не люди, а боты [2].

Причиной, по которой все больше людей используют ботов, является то, что в подобных программах по сравнению с реальным оператором сведено к минимуму время на принятие решения и совершение ставки. Таким образом, автоматизация принятия решений и совершения действий на торговой площадке способна принести значительное преимущество относительно остальных участников аукциона.

Работа ботов основывается на имитации действий реального человека, при этом используются HTTP-запросы, автоматизировано отправляемые от имени пользователя, или драйвер для веб-браузеров, позволяющий осуществить практически полное мимикрирование под клиента сайта. Наиболее популярные – Selenium Webdriver, Phantomjs, Puppeteer. Они позволяют программно взаимодействовать с браузером, управлять его поведением, получать со страницы нужные данные и выполнять различные действия: переход по ссылкам, извлечение текста страницы, поиск и нажатие кнопок и др.

Многие площадки онлайн-аукционов во избежание нечестной борьбы проводят политику запрета программного обеспечения, позволяющего автоматически размещать ставки. В такой ситуации администраторы площадок сталкиваются с проблемой идентификации нечестных пользователей и защи-

ты сайта от ботов. Ниже будут рассмотрены наиболее эффективные методы защиты.

Обновление html-структуры сайта

Временно приостановить работу ботов на сайте возможно частичной или полной заменой структуры html-кода страниц, а также изменением логики работы отдельных элементов или добавлением новых функций (окно подтверждения, деактивация блоков, находящихся вне экранного пространства). При жесткой привязке ПО к структуре документа в данном случае его работа будет парализована до полной реорганизации программного кода. Эффективность данного способа будет зависеть от соотношения потраченных ресурсов на изменение верстки (или логики) сайта и возможного причиненного ущерба за период времени, в течение которого активность ботов будет восстановлена.

Стоимость переработки сайта определяется временем работы программиста и размером оплаты его труда. Оценка ущерба оценивается индивидуально для каждой площадки и может складываться из многих факторов: отток новых аукционистов, уход текущих игроков, повышенная нагрузка на сервер и т.д.

Строгий отбор и мониторинг активности

Противостоять нечестной игре на аукционах возможно путем мониторинга игроков, причем как в автоматическом режиме, так и в ручном. Встает проблема организации достаточной степени отслеживания и анализа поведения участников, что требует при большом охвате соответствующих значительных вычислительных мощностей или объема человеко-часов.

Ограничение трафика без существенного ущерба компании возможно путем повышения порога входа на площадку, который позволит остаться на последней только фактически заинтересованным лицам. Доступ в данном случае осуществляется только после прохождения определенных шагов верификации на этапе регистрации. Уровень проверок определяется организаторами. Это может быть документ, удостоверяющий личность, собеседование и т.д. Так, например, на портале «РРТ-Аукцион» перед участием в торгах производится обязательная проверка по паспорту.

Анализ сетевой активности

Существуют различные методы обнаружения веб-ботов в сетевом трафике, в том числе ограничение частоты запросов к узлу, занесение IP-адресов в черный список, анализ значения HTTP-заголовка User-Agent, идентификация устройства путем снятия отпечатка.

Для выявления и классификации ботов также возможно использовать техники машинного обучения, включающие в себя, в общем, три этапа: обучение и тестирование, предсказание, анализ результатов. Так, в исследовании SecurityLab рассматривались HTTP-сессии (последовательность запросов от одного узла в определенном интервале времени, в данном случае 30 минут). Концептуально подход следует классической схеме обучения и применения моделей машинного обучения. Сначала определяют метрики

качества и признаки для классификации. После формируют вектор признаков и проводят серии экспериментов (различные перекрестные проверки) для валидации модели и подбора гиперпараметров. На последнем этапе выбирают наилучшую модель и проверяют качество модели на отложенной выборке (рис. 1) [3].

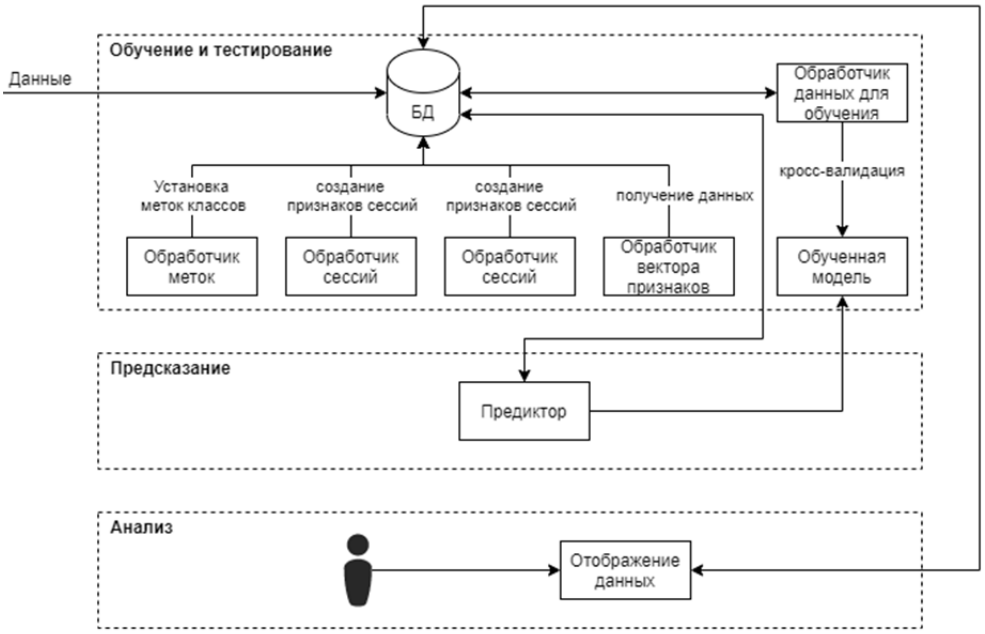

Рис. 1. Концептуальная схема подхода анализа HTTP-сессии, основанного на машинном обучении

Данное решение было проверено на трафике портала SecurityLab.ru. Объем данных – более 15 ГБ, более 130 часов. Количество сессий – более 10 000. На полученных данных средняя точность и полнота для бинарной классификации – более 95 % [3].

Введение reCAPTCHA

К концу января 2021 г. более 6 млн веб-сайтов используют виджет reCAPTCHA, и более 1,3 млн из них используют последнюю невидимую версию – reCAPTCHA v3 [4]. Третья версия данной технологии была представлена в мае 2018 г., в ней используется система поведенческого анализа пользователей «advanced risk analysis». Данная система работает в фоне и оценивает действия пользователя по шкале от 0.0 (вероятный бот) до 1.0 (вероятный человек), в том числе движение курсором мыши, нажатия мыши, но никакой информации о внутренней работе системы Google на дает, что снижает вероятность опасности.

Однако существует уязвимость, продемонстрированная в исследовании [5], заключающаяся в обходе reCAPTCHA v3 с помощью сервиса по рас-

познаванию Google «Text-to-Speech» альтернативного способа прохождения «капчи» через аудиофайл. В данном исследовании представлен скрипт, который перехватывает аудиофайл с голосовым сообщением и отправляет его в сервис «Text-to-Speech», после чего результат распознавания вставляется обратно в сервис reCAPTCHA. На версии reCAPTCHA v2 код распознается должным образом более чем в 90 % случаев [5], с reCAPTCHA v3 скрипт также справляется.

Использование сторонних сервисов

Существует множество готовых сервисов, предоставляющих услуги по защите сайта от подозрительной активности, такие как BotGuard, STRUST. Многие из них предоставляют услуги по комплексному анализу и сбору статистики пользователей. В их основе лежат различные методы поведенческого анализа, а также мониторинг через Web Application Firewall. На российском рынке подобные решения обойдутся владельцу площадки от одной копейки за пользователя. Данные сервисы имеют доступную процедуру внедрения и предоставляют множество возможностей, в том числе журналирование, анализ, исследование трафика, поведенческий анализ конкретных пользователей. Так, сервис «STRUST» используется более чем на 10 000 сайтах по всему миру. В среднем после установки данного модуля нагрузка на сервер падает на 30–50 %, а количество спама сокращается на 90 % [6].

Обнаружение ботов, использующих веб-драйвер

Симуляторы браузера, основанные на использовании веб-драйвера, позволяют маскироваться под реального оператора путем гибкой настройки параметров клиента (user-agent, ip, размер экрана и т.д.), а также возможности эмуляции поведения человека – постепенное заполнение форм, намеренное создание ошибок, нажатие кнопок по определенным координатам и т.д., поэтому простые методы защиты («ловушки», скрытые поля, время на ввод формы) могут не работать против данных ботов.

Однако существуют методы определения Selenium-ботов, в основном заключающиеся в проверке заранее определенных переменных JavaScript, которые появляются при работе веб-драйвера. Сценарии обнаружения ботов обычно ищут все, что содержит слово «selenium» / «webdriver» в любой из переменных (на объекте окна), а также переменные документа, называемые $cdc_ и $wdc_.

Экспериментальная проверка защищенности некоторых онлайн-площадок

Для проверки защищенности существующих сайтов был реализован модуль автоматизации, реализующий покупку товара на двух онлайн-площадках с аукционами двумя различными способами. Модуль реализует запросы к удаленному серверу для выявления механизмов защиты.

На первом сайте для автоматизации действий оператора была использована библиотека Selenium языка Python. На данном сайте не установлено ни одного из видов «капчи», поэтому весь функционал бота заключен в перехо-

дах на страницу активного аукциона, выполнении JavaScript кода на странице (введение цен, нажатие кнопки подтверждения) и анализа текущего состояния аукциона путем разбора HTML-кода страницы с помощью библиотеки BeautifulSoup. Единственным ограничением для автоматизированного ПО на данном сайте является блокировка слишком частых запросов, в связи с чем можно сделать вывод, что на данном сайте реализован самый минимальный уровень защиты от ботов.

В ходе анализа второго сайта было выявлено, что в запросе на покупку товара используется reCaptcha v3. Для ее прохождения было принято решение воспользоваться сторонними сервисами, такими как RuCaptcha, Xevil, Captcha.Guru. Так как взаимодействие с данными сервисами происходит посредством HTTP-запросов, покупку товара было решено реализовать также с помощью запросов, для этого применялась библиотека requests языка Python. На вход данные сервисы принимают следующие необходимые параметры:

– **sitekey** – находящийся в html в значении аргумента render при загрузке api.js, в параметре k в URI iframe, в который грузится reCAPTCHA, либо в javascript в функции grecaptcha.execute или в конфигурационном объекте ___grecaptcha_cfg;

– **action** – событие, находящееся в функции grecaptcha.execute;

– **pageurl** – полный URL страницы с reCAPTCHA V3.

На выходе был получен результат решения, который необходимо корректно использовать на сайте, в данном случае он был вставлен в заголовки HTTP-запроса на покупку товара. Так, запрос за решение reCAPTCHA будет выглядеть следующим образом:

```
body = {
       «key»: apiKey,
       «method»: 'userrecaptcha',
       «version»: «v3»,
       «action»: «submit»,
       «googlekey»: sitekey,
       «pageurl»: urlProduct,
       }
r = requests.post('https://rucaptcha.com/in.php', data = body).text
requestID = json.loads(r)['request']
```

Запрос на получение результата, соответственно, выглядит так:

```
urlRes = 'https://rucaptcha.com/res.php'
dataRes = {
       «key»: apiKey,
       «action»: 'get',
       «id»: requestID,
       «header_acao»: 1,
       «domain»: 'recaptcha.ru'
       }
r = json.loads(requests.post(urlRes, data=dataRes).text)
```

После получения результата был отправлен тестовый запрос на покупку товара, если этот запрос блокировался, то решение «капчи» не удовлетворяло

требованиям, установленным на сайте, и отправлялся отчет о неудовлетворительном решении, иначе отправлялся отчет о положительном результате. В данном эксперименте успешность решения составила 75 % (рис. 2).

В 25 % случаев значение параметра score в полученном токене было меньше порога, установленного на сайте (на текущий момент с помощью сервисов по решению reCaptcha сложно получить значение выше 0.3 [7]).

Таким образом, можно сделать вывод, что вполне современная защита от ботов, такая как Google reCAPTCHA v3, обладает уязвимостью к решению при помощи сторонних сервисов с успешностью решения 75 %.

Рис. 2. Статистика успешности решения reCaptcha v3
при помощи сторонних сервисов

Рекомендации владельцам сайтов

На данный момент защита сайтов от ботов, маскирующихся под реальных пользователей, обретает все большую актуальность. Автоматизация действий в сети позволяет ускорять рабочие процессы и увеличивать прибыль, поэтому становится все больше ботов, а их логика работы и паттерны поведения в сети все больше становятся похожими на реальных людей. О защите своих порталов владельцы должны задумываться на этапе разработки. Таким образом, были сформулированы следующие рекомендации организаторам сайтов онлайн-аукционов:

– в форме регистрации/авторизации и других критичных местах следует добавить reCAPTCHA v3, ее использование для разработчика является бесплатным и позволит изрядно сократить объем спама;

– внедрить систему блокировки пользователей, использующих подозрительные ip-адреса и user-agent, а также систему проверки трафика с эмуляторов браузера, использование которых свидетельствует о автоматизированном ПО, что позволит установить начальный уровень защиты от ботов;

– установить систему сбора и отображения статистики запросов для определения в поведении пользователей аномалий (слишком быстрое осуществление ставок, слишком быстрый ввод в формы и т.д.), которые не характерны для реального человека с последующей блокировкой данных пользователей на площадке в ручном режиме, что позволит при оптимальных затратах избавиться от большей части ботов;

– внедрить систему полного анализа запросов пользователей, например на основе техник машинного обучения, которая позволит на уже имеющейся статистике классифицировать ботов и блокировать их. Сбор статистики возможно осуществлять также с помощью сторонних WAF сервисов или хостингов (Cloudflare), предоставляющих такие возможности.

Наиболее эффективным будет являться комбинирование данных техник.

Заключение

Таким образом, защита от автоматизированного ПО на площадках интернет-аукционов является важной составляющей в борьбе с недобросовестной конкуренцией. В данной статье было продемонстрировано два способа обхода блокировок на сайтах: с использованием библиотеки Selenium и использованием сторонних сервисов для решения reCAPTCHA v3. При тестировании защиты существующей площадки вторым способом эффективность решения reCAPTCHA составила 75 %. Использование сформулированных в данной работе рекомендаций по защите сайта позволит снизить уровень нечестной конкуренции, вызванной применением ботов.

Список литературы

1. Аровина М. П. Интернет-аукционы в электронном бизнесе Украины // Теоретичні і практичні аспекти економіки та інтелектуальної власності http://tpa.pstu.edu/article/view/74640 : збірник наукових праць : у 2 вип. / ПДТУ. Маріуполь, 2015. Вип. 1, Т. 1. С. 34–40. (дата обращения: 10.09.2021).
2. Bad Bot Report 2021: The Pandemic of the Internet. URL: https://www.imperva.com/resources/resource-library/reports/bad-bot-report/ (дата обращения: 17.09.2021).
3. Один подход к обнаружению веб-ботов – SecurityLab. 19.08.2020. URL: https://www.securitylab.ru/blog/company/pt/349258.php (дата обращения: 15.09.2021).
4. reCAPTCHA Usage Statistics – BUILTWITH. 15.09.2021. URL: https://trends.builtwith.com/widgets/reCAPTCHA (дата обращения: 15.09.2021).
5. Breaking the Google Audio reCAPTCHA with Google's own Speech to Text API – incolumitas. 01.09.2021. URL: https://incolumitas.com/2021/01/02/breaking-audio-recaptcha-with-googles-own-speech-to-text-api/ (дата обращения: 15.09.2021).
6. Защита сайта от ботов. Защита от спама и DDOS атак, защита от ботов – STRUST.ru. 10.10.2021. URL: https://strust.ru/services/bad-bot-protection/ (дата обращения: 10.10.2021).
7. Документация API сервиса rucaptcha. URL: https://rucaptcha.com/api-rucaptcha#solving_recaptchav3 (дата обращения: 15.09.2021).

8. Полтавец А. И., Петров И. П., Федотова А. С., Девицкий Н. Е. Проблемы безопасности RECAPTCHA'S. URL: https://elar.urfu.ru/bitstream/10995/65576/1/978-5-7996-2404-0_2018-14.pdf (дата обращения: 15.09.2021).

References

1. Arovina M.P. Online auctions in electronic business of Ukraine. *Teoretichni i praktichni aspekti ekonomiki ta intelektual'noï vlasnosti* http://tpa.pstu.edu/article/view/74640: zbirnik naukovikh prats': u 2 vip. PDTU. Mariupol', 2015;1(1):34–40. (accessed 10.09.2021).
2. *Bad Bot Report 2021: The Pandemic of the Internet*. Available at: https://www.imperva.com/resources/resource-library/reports/bad-bot-report/ (accessed 17.09.2021).
3. *Odin podkhod k obnaruzheniyu veb-botov – SecurityLab. 19.08.2020 = One approach to detecting web bots is SecurityLab. 19.08.2020*. (In Russ.). Available at: https://www.securitylab.ru/blog/company/pt/349258.php (accessed 15.09.2021).
4. *reCAPTCHA Usage Statistics – BUILTWITH*. 15.09.2021. Available at: https://trends.builtwith.com/widgets/reCAPTCHA (accessed 15.09.2021).
5. *Breaking the Google Audio reCAPTCHA with Google's own Speech to Text API – incolumitas*. 01.09.2021. Available at: https://incolumitas.com/2021/01/02/breaking-audio-recaptcha-with-googles-own-speech-to-text-api/ (accessed 15.09.2021).
6. *Zashchita sayta ot botov. Zashchita ot spama i DDOS atak, zashchita ot botov – STRUST.ru. 10.10.2021 = Protection of the site from bots. Protection from spam and DDOS attacks, protection from bots – STRUST.ru . 10.10.2021*. (In Russ.). Available at: https://strust.ru/services/bad-bot-protection/ (accessed 10.10.2021).
7. *Dokumentatsiya API servisa rucaptcha = Documentation of the API of the rucaptcha service*. (In Russ.). Available at: https://rucaptcha.com/api-rucaptcha#solving_recaptchav3 (accessed 15.09.2021).
8. Poltavets A.I., Petrov I.P., Fedotova A.S., Devitskiy N.E. *Problemy bezopasnosti RECAPTCHA'S = RECAPTCHA's security problems*. (In Russ.). Available at: https://elar.urfu.ru/bitstream/10995/65576/1/978-5-7996-2404-0_2018-14.pdf (accessed 15.09.2021).

Информация об авторах / Information about the authors

Никита Сергеевич Рассказов
студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: nikita_serdobsk12@mail.ru

Nikita S. Rasskazov
Student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Михаил Андреевич Кривцов
студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: mikkri58@gmail.com

Mikhail A. Krivtsov
Student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Максим Александрович Митрохин
доктор технических наук, профессор,
заведующий кафедрой
вычислительной техники,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: mmax83@mail.ru

Maxim A. Mitrokhin
Doctor of technical sciences, professor,
head of the sub-department
of computer engineering,
Penza State University
(40 Krasnaya street, Penza, Russia)

Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.

Поступила в редакцию/Received 16.09.2021
Поступила после рецензирования/Revised 17.11.2021
Принята к публикации/Accepted 10.12.2021