11. *Шумилин А. В.* Основные элементы мандатной сущностно-ролевой ДП-модели управления доступом и информационными потоками в СУБД PostgreSQL ОС специального назначения Astra Linux Special Edition // Прикладная дискретная математика. 2013. № 3(21). С. 52–67.

12. *Смольянинов В. Ю.* Анализ условий предоставления и получения прав доступа в модели управления доступом MS SQL Server // Прикладная дискретная математика. 2014. № 2(24). С. 48–78.

УДК 004.934 DOI 10.17223/2226308X/11/30

АВТОМАТИЗИРОВАННОЕ ПРОХОЖДЕНИЕ GOOGLE RECAPTCHA V2

И. Н. Манашев

Показана неактуальность текущего подхода к решению задачи разграничения реальных пользователей и компьютерных ботов. Приводится способ автоматизированного прохождения теста Google reCAPTCHA v2.

Ключевые слова: *распознавание captcha, автоматизация, reCAPTCHA v2.*

1. Общие сведения

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart, далее — капча) — один из вариантов теста Тьюринга, который позволяет различить реальных пользователей и компьютерных ботов. В его основе лежит задача, лёгкая для человека, но трудоёмкая для компьютера [1]. Этот механизм защиты должен оградить сайты от спама, автоматических регистраций, накруток, DDoS-атак и прочих дел, которыми обычно занимаются боты. На данный момент самой популярной реализацией капчи является Google reCAPTCHA v2.

Классическая (текстовая) капча представляет собой картинку с последовательностью искажённых символов (букв, цифр и спецсимволов). Решением текстовой капчи считается получение изображённых на картинке символов в текстовом виде [2]. Технология Google отходит от стандартной концепции автоматизированного теста Тьюринга и оценивает поведение пользователя, а не его способность разгадывания слов.

Пользователю нужно выполнить простейшее действие — отметить галочкой утверждение «Я не робот». В этот момент капча оценивает косвенные параметры, указывающие на возможного бота: время, проведённое на странице, траекторию движения курсора, IP-адрес и пр. Если у капчи закрадываются сомнения в том, что пользователь — человек, то она предложит выполнить одно из двух заданий: образный или аудиотест.

Образная капча — это тест, для прохождения которого требуется решить задачу классификации образов: нужно выбрать из нескольких изображений те, которые соответствуют заранее объявленному критерию (например, выбрать изображения, на которых есть автомобиль).

Аудиокапча представляет собой аудиозапись, в которой проговаривается какая-либо фраза, содержащая последовательность слов или цифр. Как правило, в аудиозаписи присутствуют различные искажения: варьируемая тональность, фоновый шум, паузы. Решением аудиокапчи считается получение фразы в текстовом виде.

Одна из главных проблем любой капчи — её исполнение. Боты — проблема не для пользователей, а для администраторов сайта. Перекладывать её решение на обычных

людей некорректно, тем более что при вводе очередной капчи пользователи испытывают лишь раздражение [3]. С развитием алгоритмов и искусственного интеллекта многие механизмы защиты стали практически бесполезными. Такая судьба постигла не только текстовую капчу, но и аудиокапчу.

2. Автоматизированное прохождение аудиокапчи

Основная идея метода заключается в прохождении аудиотеста вместо стандартного образного, который является более сложным. Для взаимодействия с капчей также понадобится имитировать поведение реального пользователя. Разделим автоматизированное прохождение Google reCAPTCHA v2 на три этапа:

1) получение задания — выбор аудиокапчи и скачивание аудиозаписи;
2) распознавание речи — отправка аудиозаписи в сервис распознавания речи и получение решения аудиокапчи;
3) подтверждение ответа — отправка решения и прохождение теста.

2.1. Получение задания

Главная цель этого этапа — получение аудиокапчи. Требуется найти виджет reCAPTCHA на странице, отметить галочкой утверждение «Я не робот» и, если reCAPTCHA предложит решить задание, переключиться на аудиокапчу и скачать её.

Для автоматизации всех действий на этапах 1 и 3 будем использовать библиотеку Selenium для языка Python. Selenium WebDriver — это инструмент для автоматизации действий веб-браузера. Все нужные элементы на странице находим по их локаторам. Локатор — это строка, уникально идентифицирующая UI-элемент. Локаторы элементов известны заранее.

Для переключения между элементами, а также с целью имитации «человеческого» поведения будем использовать задержки и непрямолинейное движение курсора. Для имитации «человеческой» траектории курсора воспользуемся квадратичными кривыми Безье. Квадратичная кривая Безье $B(t)$ задаётся тремя опорными точками P_0, P_1, P_2, где P_0 — источник (текущее положение курсора); P_2 — назначение (куда нужно переместить курсор); P_1 задаётся случайно, она характеризует кривизну линии между P_0 и P_2:

$$B(t) = (1-t)^2 P_0 + 2t(1-t)P_1 + t^2 P_2, \quad t \in [0,1].$$

Перемещение курсора по таким кривым позволяет обмануть reCAPTCHA и продолжить прохождение теста.

2.2. Распознавание речи

Для решения капчи воспользуемся сторонним сервисом распознавания речи IBM Watson Speech to Text. С помощью API отправим полученную на первом шаге аудиозапись и, получив ответ от сервиса в формате json, извлечём из него поле с решением капчи.

2.3. Подтверждение ответа

На этом этапе требуется ввести полученный ответ в текстовое поле и кликнуть «Подтвердить». Все действия выполняются аналогично шагу 1. Иногда может потребоваться решить несколько аудиокапч. В этом случае требуется повторить шаги 1–3.

Заключение

Таким образом, возможно автоматизированное прохождение Google reCAPTCHA v2, что говорит о том, что капча в текущем её понимании не является эффективным и

точным способом разграничения реальных пользователей и компьютерных ботов. Следовательно, возникает необходимость в создании новых подходов к решению данной задачи. Один из возможных способов — это более глубокий и незаметный для пользователя анализ его поведения и взаимодействия с ресурсом. В данный момент ведётся бета-тестирование Google reCAPTCHA v3, в которой и реализован данный подход [4].

ЛИТЕРАТУРА

1. *Von Ahn L.* Massive-Scale Online Collaboration. TED, 2011. `https://www.youtube.com/watch?v=-Ht4qiDRZE8`
2. *Von Ahn L., Maurer B., McMillen D., et al.* reCAPTCHA: human-based character recognition via web security measures // Science. 2008. V. 321. No. 5895. P. 1465–1468.
3. *Allen T.* Having a CAPTCHA is Killing Your Conversion Rate. [Электронный ресурс]. `https://moz.com/blog/having-a-captcha-is-killing-your-conversion-rate` (дата обращения: 09.06.2018).
4. reCAPTCHA: Easy on Humans, Hard on Bots. [Электронный ресурс]. `https://www.google.com/recaptcha/intro/index.html` (дата обращения: 09.06.2018).