БОРЬБА СО СПАМОМ СРЕДСТВАМИ CAPTCHA

Автор: Александра МАРКИНА
Научный руководитель: ст. пр. Татьяна СКОРОХОДОВА

Технический Университет Молдовы

В статье рассматривается проблема спама и распространения спам-ботов на электронных ресурсах такого типа как форумы, социальные сети, доски объявлений. Для более эффективной защиты предлагается использование такого механизма как CAPTCHA. В статье объясняется смысл этого понятия, способы его использования. Помимо общего описания данного способа защиты описывается также способы его оптимизации, приведен пример графической CAPTCHA, которая является достаточно устойчивой к взлому.

Ключевые слова: спам, CAPTCHA, спам-бот, рассылка, регистрация, Automated Turing Tests.

Наверняка многим из нас приходилось сталкиваться с письмами и сообщениями рекламного направления, которые бы совсем не хотелось получать. Чаще всего эта реклама носит коммерческий характер, является назойливой и бесполезной. Подобные сообщения создают проблему спама.

По определению лаборатории Касперского, спам – это массовая анонимная не запрошенная рассылка сообщений пользователям, причем нет разницы, коммерческая ли это реклама или просто полезная, по мнению отправителя, информация.

Обычно принято считать, что спам распространяется посредством почты. Но есть и другие каналы, по которым распространяются спамовые сообщения. Это системы мгновенных сообщений (различные чаты, ICQ и пр.), форумы, блоги, социальные сети, доски объявлений.

На почтовых серверах установлено различное программное обеспечение, которое фильтрует сообщения, в результате чего до конечного пользователя, если и доходит спам, то в достаточно малом количестве, несравнимым с тем, что изначально приходит на сервера. Иная же ситуация обстоит с форумами или блогами, они просто напичканы спамом, при том не сразу можно определить, кем отправлено сообщение: человеком или роботом. В связи с этим на формах регистрации и отправки сообщений многих сайтов разработчики и администраторы располагают механизмы, предотвращающие автоматическую отправку данных. Подобные механизмы называются CAPTCHA.

CAPTCHA – это аббревиатура от следующего выражения на английском языке «***C****ompletely* ***A****utomated* ***P****ublic* ***T****uring test to tell* ***C****omputers and* ***H****umans* ***A****part*», в переводе на русский язык это звучит следующим образом: «полностью автоматический публичный тест Тьюринга для различия компьютеров и людей». CAPTCHA была создана для того, чтобы убедиться, что введенные данные не были сгенерированы компьютером. На рисунке 1 представлен пример нескольких графических CAPTCHA.

Чтобы понять, зачем нужен этот тест, следует понять цели, которые побуждают спамеров создавать и использовать автоматические система ввода информации. Целями спама являются следующие пункты:

- массовая рассылка;
- манипуляция системами голосования;
- публикация неуместных ссылок для повышения рейтинга в поисковых системах;
- получение доступа к личной информации;
- распространение вредоносного кода.

Автоматизация процесса рассылки спама приводит к очень большой скорости его распространения и, соответственно, к большей эффективности.

Рисунок 1 – Пример графических CAPTCHA

В такой ситуации CAPTCHA является очень привлекательным решением проблемы, так как время, необходимое для постоянного контроля пользовательского контента несопоставимо со временем, необходимым для введения данного теста на сайте – именно это толкает разработчиков на его использование. Данный механизм решает проблему прямо: его предназначение заключается исключительно в остановке спамеров. У разных механизмов CAPTCHA разная степень эффективности. Основные параметры успешной защиты таковы:

- пользователь должен решить задачу в любых условиях, но компьютер не сможет этого сделать;
- ввод данных должен быть минимален;
- время на решение должно быть также минимальным;
- задача должна быть легко выполнимой для всех пользователей, в том числе и тех, кто страдает различными специфическими заболеваниями.

Одно из наиболее заметных превосходств человека над компьютером выражается в умении различать визуальные образы и модели. Наиболее распространенные и популярные CAPTCHA отталкиваются именно от этого факта. [1]

В первую очередь при генерации графической CAPTCHA необходимо использовать фон, на котором есть какой-либо шум. При этом следуем учесть, что если символы отличается по цвету от фона, то такое изображение возможно взломать. Фон и шум на нем по цвету должны минимально отличаться от символов, но нужно помнить, что визуально символы все-таки должны выделяться. Помимо шума фон может содержать сторонние элементы, которые также затруднят процесс отделения фона и текста, но не скажется на читабельности.

Для повышения устойчивости против взлома картинки CAPTCHA стоит данный шум также наложить на символы, это еще больше затруднит автоматическое распознавание текста, так как одним из шагов автоматического распознавания изображений является отделение фона и символов.

Отдельной задачей при генерации изображения CAPTCHA является генерация символов, поскольку обычные недеформированные символы очень легко распознать. Деформация буквенных символов может быть различной. При генерации каждого символа рекомендуется поворачивать его на угол, величина которого генерируется случайным образом.

Но даже при повороте символ можно легко распознать при помощи робота. Поэтому к символу следует применить произвольное геометрическое искажение, желательно нелинейное. Помимо всего прочего имеет смысл сдвигать символы друг относительно друга, использовать различный шрифт.

Согласно вышеописанным параметрам, наиболее сложным к взлому будет изображение наподобие рисунка 2.

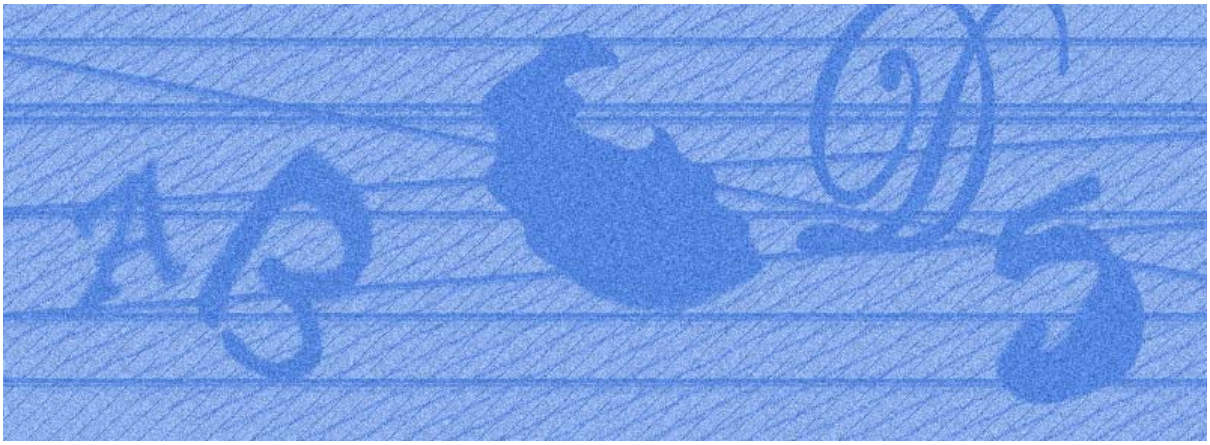

Рисунок 2 – Пример устойчивой графической CAPTCHA

Данное изображение было протестировано на нескольких сервисах распознавания символов, ни один из них не справился со своей задачей.

Помимо использования обычного текста имеет смысл брать за основу несложные арифметические примеры или логические задачки, это достаточно высоко повысит эффективность защиты. К примеру, можно задавать такие вопросы: «Сколько будет два плюс два?», «Как называется последняя буква в слове 'университет'?», «Сколько вершин у треугольника?». Правда следует отметить, что у подобных вопросов есть недостаток – зависимость от языковых познаний пользователя. Но если у проекта есть несколько языковых версий, то под каждую версию можно использовать локализованные вопросы.

Несмотря на то, что при разумном подходе к использованию CAPTCHA может быть очень эффективным механизмом, данный механизм не дает 100% результат. Меры против спама изобретают и внедряют люди, противостоят этим мерам и внедряют ботов также люди. Поэтому одержать полную победу над спамом невозможно. Но для того, чтобы быть впереди спамеров следует всегда что-нибудь изменять и модернизировать в системах защиты, которые должны включать в себя несколько различных инструментов.

Список использованных источников

1. Понкин Д. *В поисках идеальной CAPTCHA* [Электронный ресурс]. – Режим доступа: http://habrahabr.ru/blogs/infosecurity/120851/ . – Загл. с экрана.
2. May M. *Inaccessibility of CAPTCHA* [Электронный ресурс]. – Режим доступа: http://www.w3.org/TR/turingtest/. – Загл. с экрана.