

РАСПОЗНАВАНИЕ УНИКАЛЬНОСТИ ОБРАЩЕНИЯ К ИНТЕРНЕТ-ИНФОРМАЦИИ

З.И. Маслова, канд. техн наук, доцент;

Л.Д. Назаренко, доцент;

К.Н. Панасенко, студент

Сумский государственный университет, г. Сумы

Проведен анализ существующих методов учета уникальных обращений к Интернет-информации, выявивший их недостатки и ограниченность возможностей. Предлагается модифицированный алгоритм. Разработан программный фильтр, распознающий уникальность пользователей.

Проведено аналіз існуючих методів обліку унікальних звернень до Інтернет-інформації, що виявив їхні недоліки й обмеженість можливостей. Пропонується модифікований алгоритм. Розроблено програмний фільтр, що розпізнає унікальність користувачів.

ВВЕДЕНИЕ

Большинство ресурсов современной сети Интернет существует исключительно за счет рекламы. Владельцы сайтов стремятся увеличить количество показов баннеров всеми возможными – и честными, и нечестными средствами. Рекламодатель не заинтересован оплачивать показы баннеров, которых никто не видел. В договоре рекламодателю обещают некоторое количество гарантируемых посещений его сайта или просмотра некоторой информации, при этом рекламодатель не может подтвердить или опровергнуть действительно ли то гарантируемое число уникальных пользователей просмотрело требуемую информацию. Существует два подхода к продаже рекламы: либо оплачивается каждый показ (клик) баннера, либо рекламодатель выплачивает держателю сайта некую фиксированную сумму, зависящую от его посещаемости. Последний способ считается более предпочтительным, поскольку приблизительную посещаемость ресурса можно определить и визуально безо всяких счетчиков. Однако такой подход крайне непрактичен и неудобен. И хуже всего – он подходит только для крупных сайтов. Владельцам же домашних страничек никто оптом за рекламу платить не будет. И не потому, что их посещаемость невозможно оценить. Просто вручную контролировать огромное количество ресурсов физически невозможно, поэтому придется либо прибегать к автоматизированным средствам (а их часто легко обмануть), либо иметь дело только с несколькими тысячами держателей крупнейших WEB-проектов, которые заведомо не будут жульничать в силу своего авторитета. Последнее не приемлемо в первую очередь для самих рекламодателей, стремящихся своей рекламой охватить как можно большую аудиторию и готовых ради этого идти на неизбежный риск.

Многие из рекламных агентств используют автоматическое прокликивание баннеров и «накрутку» счётчиков посещений, которых не столь велико как было обещано рекламным агентством рекламодателю. Для этих целей используются разные методики и алгоритмы автоматического «накручивания» счётчиков посещений. Владельцы сайтов, заинтересованные в увеличении числа посетителей, организовали добровольное сообщество для использования технологий клик-клубов [1].

Если злоумышленник, накручивая счётчик баннера, не будет при каждом запросе менять свой IP-адрес, его легко можно идентифицировать. Традиционное оружие обмана – анонимные Проху-

сервера, выполняющие запрос злоумышленника от своего имени. Существует не столь уж много таких серверов – какой-то десяток тысяч, а то и меньше. Вполне реально собрать их в один список и проверять IP-адрес клиента на совпадение. Такой метод прост в реализации, но очень неэффективен.

Как бы злоумышленник не пытался изменить свой IP-адрес, автоматизированные «крутилки» легко распознаются статистическим анализом. В простейшем случае - это: слишком высокая интенсивность и строгая периодичность заходов, неестественно большое отношение количества нажатий к числу показов баннера, слишком короткий интервал времени между запросом баннера и нажатием на него (а некоторые «крутилки» и вовсе выполняют клик по баннеру без предварительного запроса). Подозрение вызывает пользователь, который через регулярные промежутки времени заходит на сайт и тут же тянется мышкой к еще не успевшему загрузиться баннеру.

Отсутствие надежных средств идентификации пользователя – одна из центральных проблем современного Интернета. Именно анонимность позволяет злоумышленникам совершать атаки, оставаясь ненаказанными. Если рекламодатели действительно хотят донести информацию до клиента, им необходимо настоять на доработке http-протокола и внесению в него команд, предназначенных специально для загрузки баннеров.

Актуальной и важной является задача распознавания уникальности пользователей. Ее решением занимаются многие разработчики. Система AdRiver [2] реализует одну из методик подсчета показов, кликов и посетителей сайта. На сегодняшний день самым авторитетным и популярным счетчиком и рейтинговой системой в Рунете является Рамблер [3]. Недостатком этого счетчика является то, что идентификация уникальных пользователей происходит по IP-адресам. Наиболее продвинутый счетчик SPILOG [4] определяет уникальность посетителей по cookies. Для того чтобы разработать фильтр авторизации кликов, необходимо проанализировать как можно большее количество данных о пользователе.

ПОСТАНОВКА ЗАДАЧИ

Необходимо разработать программный модуль для определения уникальности посетителя сайта по следующим данным:

- параметры соединения клиент-сервер – IP-адрес и порт, с которого пришел запрос серверу, порт сервера, на который пришел запрос. Получив данные о подключении – исключить возможность злоумышленнику генерировать успешные запросы с одного и того же хоста. Также выявить злоумышленника, который подменил свой IP-адрес путем использования FTP-сервера;

- пользовательские данные: версия браузера, с которого был выполнен запрос, версия операционной системы;

- временные параметры входа и выхода со странички рекламодателя. Эти данные необходимы для определения продолжительности пребывания пользователя на страничке и распознавания автопрокликивания баннеров и загрузки страничек без ведома пользователя;

- признаки повторного посещения пользователя некоторого ресурса с использованием метода cookies (при условии, что поддержка cookies включена).

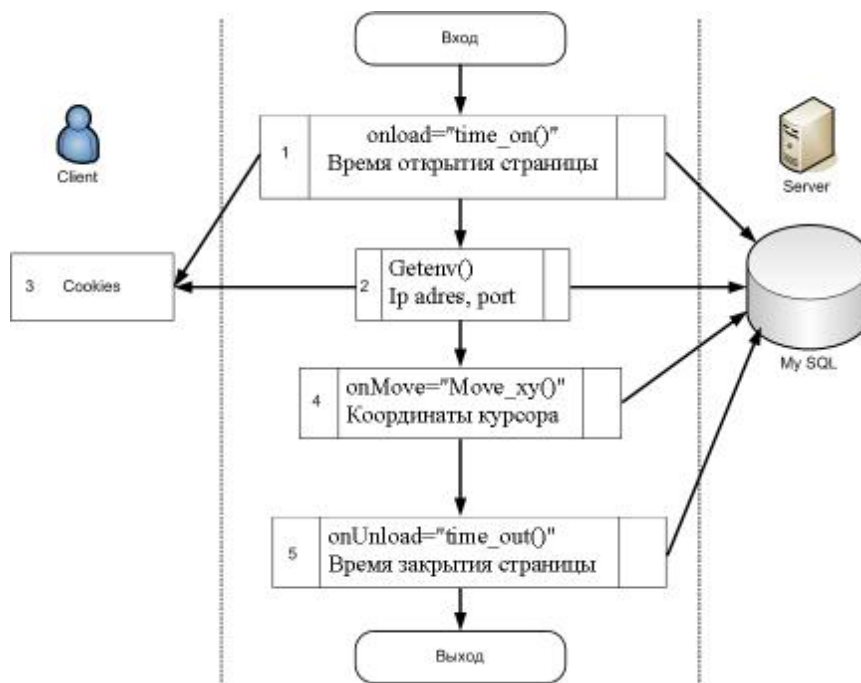
После получения всех перечисленных параметров их необходимо сохранить в базе данных на сервере для последующего анализа и обработки.

ОПИСАНИЕ ПРОГРАММНОГО МОДУЛЯ

Программа реализована на скриптовых языках программирования JavaScript и PHP [5,6,7,8]. Укрупненная схема программного модуля представлена на рисунке.

Разработанный программный модуль выполняет следующие функции:

- как только пользователь начал загружать страничку, запоминаем время начала открытия страницы (блок 1);
- получить данные о пользователе: уникальный IP-адрес в глобальной сети Интернет, порт, с которого поступил запрос, порт на который поступил серверу запрос, тип браузера и его версию, данные о операционной системе. После получения данных сохраняем их в базе данных (блок 2);
- следующий шаг для получения информации идентификации клиента заключается в использовании cookies: сервер вместе с кликом по баннеру передает браузеру специальную метку, которую тот при условии, что поддержка cookies включена, сохраняет на диске. При последующих запросах сервер проверяет наличие метки в cookies у клиента и если его нет, считает, что запрос выполнен не браузером, а скриптом злоумышленника (блок 3);
- если пользователь проявит активность и курсор изменит свое положение на страничке, то выполнится скрипт, который вычисляет значения координат курсора мышки в данный момент времени и сохранит изменения в базе данных (блок 4);
- после того как пользователь нажмет на кнопку закрытия окна, выполнится процедура `time_out()`, сохранятся все данные в базе данных, и окно закроется (блок 5).



ВЫВОДЫ

1 Достоинством предложенного подхода является то, что при его использовании достаточно трудно сгенерировать большое количество запросов, которые могли бы рассматриваться как уникальные обращения.

2 Использование учета портов при подключении практически исключает возможность использования злоумышленником FTP-сервера как инструмента скрытия IP-адреса.

3 Учет временных параметров исключает возможность использования автокликеров и фреймов нулевого размера, так как учитывается активность пользователя на страничке, а именно передвижение курсора по страничке.

4 Полученные данные после обработки и анализа могут служить параметрами ограничения в фильтре при подсчете количества уникальных просмотров и обращений к некоторой Интернет-информации.

SUMMARY

RECOGNITION OF UNIQUENESS OF THE REFERENCE TO THE INTERNET INFORMATION

Zoya Maslova, Lyudmila Nazarenko, Konstantin Panasenko
Sumy State University

The analysis of existing methods to account unique Internet searches demonstrated their weaknesses and limited capabilities. A modified algorithm was proposed. Program filter to recognize unique users was developed.

СПИСОК ЛИТЕРАТУРЫ

1. Что такое клик-клуб? http://www.antula.ru/rang-click_clubs.htm
2. Методика подсчета в системе AdRiver <http://www.adriver.ru/doc.adv-agency/audience/>
3. Рамблер <http://top100.rambler.ru>
4. Spylog <http://www.spilog.ru>
5. Баррет Д. JavaScript. Web-профессионалам. – Киев: БХВ – Киев, 2001.
6. Бранденбау Д. JavaScript: сборник рецептов. – СПб.: Питер, 2000.
7. Вайк А. JavaScript в примерах. – Киев: ДияСофт, 2000.
8. Дмитриева М. Самоучитель JavaScript. – СПб.: БХВ Санкт-Петербург, 2001.

Поступила в редакцию 19 декабря 2008 г.