

ЗАЩИТА ОТ СПАМА В ИНТЕРНЕТ ПРОСТРАНСТВЕ

АБДУЛЛАЕВ В.Г.

Рассматриваются основные методы и правила для эффективной борьбы с рекламными сообщениями. Описываются методы фильтрации почтовых сообщений на стороне сервера. Предлагаются варианты противодействия способам обхода методов фильтрации и защиты от рекламных сообщений.

Ключевые слова: спам, рассылка рекламных сообщений, методы фильтрации, использование e-mail адресов в Интернете.

Введение

СПАМ – это сообщения рекламного характера, которые доставляются без согласия пользователя, а также сообщения для мошеннических целей. Количество пользователей в сети Интернет с каждым годом увеличивается, в 2011 году – 718 миллионов, а в 2012 уже 1,3 миллиарда пользователей Интернет [1]. Это связано со многими факторами, такими как удешевление оборудования (как результат – цена доступа к сети снижается), активное распространение планшетных компьютеров и смартфонов (пользователи мобильного интернета).

Доля спама в мировом трафике составляет почти 80% за 2011 год [2]. В 2012 году общее количество обнаруженных и удаленных сообщений составляет 65 миллиардов комментариев [3]. Как показывает статистика, уровень спама к концу 2012 года значительно возрос. На данный момент уровень спамовских комментариев превысил показатель 120 миллионов в день.

На рост спама повлияло массовое появление незащищенных сайтов, вики и форумов. Также люди, которые рассылают спам, начали активно использовать нелегальные методы, такие, как взлом сайтов, зара-

жение компьютеров пользователей для создания бот-нетов (сеть зараженных компьютеров и серверов, которые используются злоумышленниками в целях нанесения вреда) для рассылки спама. Значительный рост спровоцировало также увеличение трафика из Китая – это связано с продвижением подделок известных брендов.

С ростом популярности сервисов мгновенных сообщений и социальных сетей растет количество спама в данных сервисах. Также отмечается рост спама в смс-сообщениях.

Убытки, которые приносит спам, колоссальные. Это – время, перегруженность оборудования и как результат – снижение срока эксплуатации. Все это приводит к финансовым потерям как организаций, так и государственных учреждений.

Постановка задачи

Цель - рассмотреть правила и разработать методы защиты e-mail от спама. Учитывая спам, который с каждым годом увеличивается в объеме и количестве, а также то, что совершенствуются приемы обхода разных методов защиты, необходимо постоянно совершенствовать методы защиты. Соблюдение простых правил обеспечит уменьшение количества спама. Также необходимо учесть спам, который публикуется на сайтах и форумах в комментариях. Задача – разработать противодействия способов обхода спам-фильтров.

Правила использование e-mail адресов в Интернете

Один из надежных методов, который понизит количество спама на электронный адрес и затруднит получить адрес злоумышленниками - это соблюдение простых правил:

– Не следует публиковать свой электронный адрес на общедоступных сайтах в открытой форме. Можно использовать специальные символы, что затруднит его распознавание. Также можно применять адрес в виде картинки.

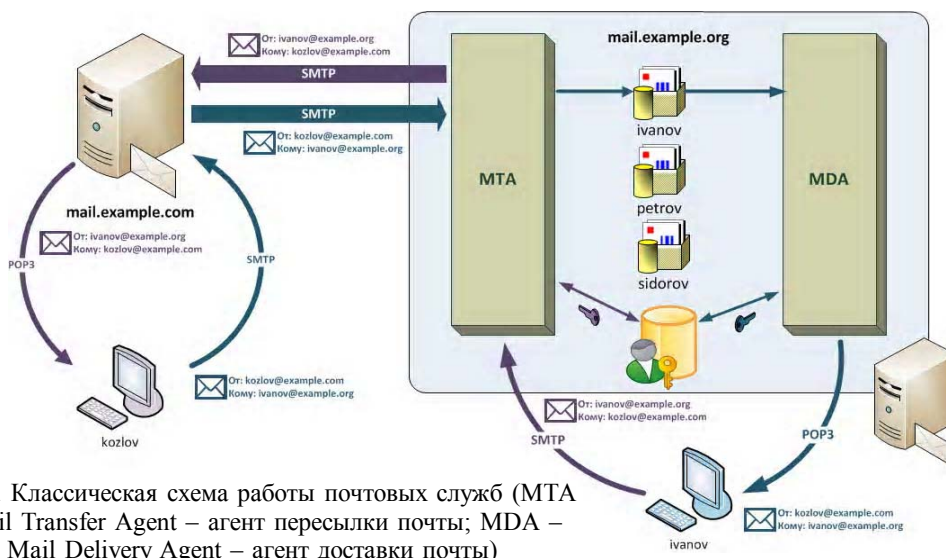


Рис. 1. Классическая схема работы почтовых служб (MTA – Mail Transfer Agent – агент пересылки почты; MDA – Mail Delivery Agent – агент доставки почты)

– Большинство пользователей имеют аккаунты в социальных сетях. А учитывая последнюю тенденцию, что на сайтах очень часто используется авторизация через социальную сеть, в таком случае лучше воспользоваться аккаунтом социальной сети.

– Не стоит использовать адрес при регистрации на малоизвестных форумах и сайтах.

– Можно завести специальный адрес для регистрации на форумах и сайтах или использовать службы для получения одноразовых адресов, к примеру сервис - <http://mailinator.com/>.

– Не следует отвечать на спам или переходить по ссылкам из письма, в том числе на ссылку, что обеспечивает отписку от рассылки. Как правило, данными действиями подтверждается реальный адрес. Также не следует открывать вложения.

– Желательно отключить загрузку картинок, которые в письме, чтобы не отслеживалась активность пользователя.

– При получении ссылки от неизвестного человека или от контакта, который находится у вас в списке, но ссылка вызывает подозрение, лучше отказаться от перехода по ней.

– Необходимо создавать длинные имена адресов, к примеру, который состоит из фамилии и имени, чтобы исключить метод подбора адреса случайными генераторами.

Выполнение простых правил и осторожность обеспечивают частичную защиту от спама, а также от персонального компьютера, который может стать одним из узлов мошеннической сети.

Способы защиты e-mail адресов и Веб-страниц

Для того чтобы защитить e-mail адреса, которые размещены на сайтах, есть разные способы и правила, которые реализуются исключительно технически. Самое главное - это защита от ботов (автоматических программ (скриптов), предназначенных для сбора информации или отправки сообщений).

На большинстве Веб-страниц есть формы обратной связи для написания обращений, жалоб и так далее, или в блогах, комментариях к статьям. Самым распространённым и относительно надёжным является размещение CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart* – автоматизированный текст Тьюринга для определения различий компьютер или человек) [4]. В основе данного метода лежит такая задача, которую может решить только человек. К примеру, могут быть изображения искривлённых символов или изображения математической задачи, как правило сложение. Автоматические скрипты не могут распознать искаженный текст, но это может сделать человек. Также есть варианты модифицированных примеров, таких как посчитать количество рисунков с изображением определенного образа.

На данный момент существует достаточно способов обхода CAPTCHA:

– Создание базы данных ответов или получение базы данных нелегальным путем. Некоторые разработчики используют готовую базу данных заданий и ответов, т.е. задачи не генерируются случайным образом. Поэтому злоумышленники могут в ручном режиме сформировать базу данных или скачать ее. Противодействие – не создавать базы данных, а задания генерировать случайным образом.

– Неквалифицированное написание скрипта, что приведет при определенном формировании запроса к странице получать правильный ответ. Решение – использовать готовые скрипты, которые себя зарекомендовали как надежные, или обращаться к профессиональным исполнителям.

– Автоматическое распознавание возможно в том случае, если имеется неискаженная или не завуалированная картинка с заданиями. Также использование только одного метода маскировки задания. Это все способствует разработке метода автоматического распознавания. Решение – применять разные методы маскировки, которые будут меняться в произвольном порядке.

– Автоматический подбор вариантов ответов. Ответ на задание автоматически подбирается. Для защиты от данного метода служит блокировка IP-адреса или аккаунта на определенное время, при повторе – полная блокировка без восстановления.

– Ручное распознавание. Специальные сервисы, которые используют человеческий ресурс для распознавания. Данный способ обойти практически не реально. Если отслеживать с какого адреса или аккаунта приходят эти сообщения нужно временно блокировать, а также ввести определенное время для ввода ответа. Как правило, используются разные данные, поэтому необходимо в ручном режиме модерировать сообщения.

Дополнительными методами борьбы с нежелательными сообщениями может быть запрет анонимных комментариев, разрешение комментариев только со стороны зарегистрированных пользователей или авторизация через социальные сети; использование надёжных скриптов, что обеспечивает минимальное количество проблем в безопасности системы; запрет прямых публикаций сообщений, разрешение публикаций только после одобрения администратором.

Можно также использовать фильтры для фильтрации сообщений по ключевым словам. Принцип действия данного скрипта в следующем: после того, как пользователь отправил сообщение, скрипт анализирует его, и если сообщение содержит ключевые слова или URL, которые есть в базе данных скрипта, сообщение не публикуется, а отправляется администратору на проверку.

Методы фильтрации почты на почтовых серверах

Для фильтрации почтовых сообщений на сервере можно использовать черные списки, серые списки, выставлять рейтинг пользователю, фильтровать сообщения по домену отправителя, а также по ключевым словам в теме и теле письма (рис. 1).

Существуют черные списки e-mail адресов и IP – это списки, где включены данные пользователей и систем, которые были замечены в рассылке спама или вредоносных сообщений. Сервер при получении нового сообщения автоматически определяет, внесен ли пользователь в эти базы данных. Если в рейтинге они существуют, такие сообщения автоматически отправляются в папку спама или удаляются, в зависимости от настройки системы. Данные списки получили название DNSBL (DNS BlackList). Этот метод утратил свою актуальность из-за того, что злоумышленники, как правило, применяют пользовательские адреса и компьютеры. И после внесения их в данные списки пользователи долгое время не могут отправлять сообщения.

Также возможна автоматическая фильтрация (СПАМ-фильтры) как на стороне сервера, так и на стороне клиента. Один из подходов заключается в анализе содержимого письма (по ключевым словам), и в зависимости от содержания делается вывод. Один из применяемых методов – статистический анализ содержимого письма; как правило, используется байесовский метод фильтрации спама. Суть этого метода заключается в анализе слов в теле письма на возможность (вероятность) его отношения к спаму [5], после чего делается вывод. В начале использования данного метода необходимо провести обучение фильтра на выявление спама. При активном и правильном обучении эффективность фильтрации спама можно довести до 95%, но для этого необходимо постоянно проводить обучение.

В последнее время популярным стал метод серых списков. Суть его заключается в том, что при первой попытке отправить письмо сервер отвечает ошибкой. Программы для рассылки спама не умеют оперативно обрабатывать такие ситуации. Поэтому очень высокая вероятность того, что они находятся в черных списках. Данный метод дает эффективность до 90%.

Дополнительной мерой защиты от спама является запрос PTR – записи (рис. 2). PTR – запись связывает IP – адрес с именем домена. Запросив PTR, МТА примет почту только в случае совпадения IP с именем домена. Учитывая, что спам, как правило, приходит с IP, который не совпадает с именем домена, защита в данном случае достаточно эффективна.

Дополнительные методы защиты

В почтовые сервера встраиваются модули для определения признаков массовой рассылки. Эти модули подсчитывают контрольную сумму письма и проверяют на серверах служб Razon и DCC. Если встреча-

ются с аналогичными контрольными суммами письма, большая вероятность, что это спам.

Ужесточаются требования к письмам отправителя: проверка доменного имени, обратного адреса, IP-адреса компьютера, с которого идет рассылка.

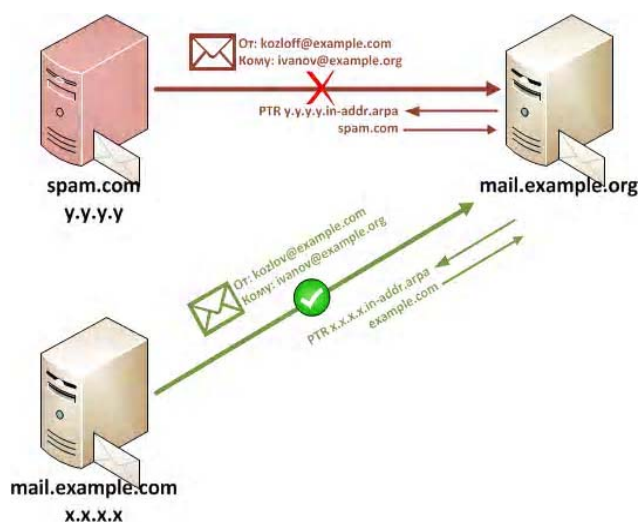


Рис. 2. Пример запроса PTR

Альтернативные методы защиты

Каждое рекламное сообщение имеет в себе слова и слоганы, которые составляют части рекламного сообщения. К примеру: «купите», «лучшее предложение» и так далее.

Проведя анализ почтового сообщения, кроме темы сообщения, можно с точностью определить является оно рекламным или нет. Аналогично данному методу работают системы распознавания плагиата, когда ищутся части сообщения, которые взяты из других источников.

В данном методе есть ряд недостатков. Прежде всего, во время обучения системы оператор будет видеть чужие сообщения, проверяя соответствия оценки системы. Другая проблема – это защита конфиденциальной информации, но в случае корпоративного сектора это не является ключевой проблемой. Третья — затраты дополнительных ресурсов на обработку информации, увеличение затрат.

Но несмотря на ряд недостатков, данный метод может быть эффективным при борьбе со спамом.

Выводы

Учитывая постоянный рост спама, можно сделать вывод, что данный метод продвижения товаров и услуг приносит выгоду заказчикам. Поэтому не стоит ожидать понижения количества спама в ближайшее время.

Выполнение простых правил поведения в сети Интернет значительно снизит объем получаемой почты. Все зависит от поведения пользователя. И за ним стоит выбор, следовать этим правилам или нет.

Достаточно эффективным и распространенным методом борьбы со спамом есть CAPTCHA. Как показывает статистика, в 2011 году ежедневно вводится приблизительно 200 миллионов CAPTCHA во всем мире [6]. Данный метод хорошо зарекомендовал себя в формах обратной связи – комментариях, но не дает гарантии полной фильтрации рекламных сообщений.

Использование специального программного обеспечения и методов повышает эффективность фильтрации сообщений. Но, как правило, все методы требуют постоянного совершенствования, потому что практически все, кто занимается рассылкой нежелательных сообщений, совершенствуют методы обхода систем защиты.

Нужно понимать, что спрос формирует предложение. Если предприниматели будут дальше заказывать рассылку спама, а пользователи откликаться на рекламные сообщения, спрос будет расти и объем спама также будет увеличиваться.

Не стоит забывать о юридических аспектах - в некоторых странах приняты законы об уголовной ответственности за рассылку спама.

Но простые правила поведения и использование эффективных методов защиты в совокупности принесут результат, и количество получаемого спама уменьшится. Также желательно постоянно проводить обучение систем на повышение их эффективности.

Комбинация перечисленных выше методов защиты даст значительное снижение нежелательных сообщений, а также снизит нагрузку на сервера и сетевой трафик, и не будет отнимать время у пользователей.

Литература: 1. *The statistics portal* <http://www.statista.com/statistics/74592/quarterly-worldwide-smartphone-sales-by-operating-system-since-2009/> 2. *Компьютеры и огротехника* <http://www.computery.ru/news/news2010.php?nid=8302> 3. *Spammy A .Year in Review* <http://blog.akismet.com/2012/12/21/a-spammy-year-in-review/> 4. *Completely Automated Public Turing test to tell Computers and Humans Apart* <http://ru.wikipedia.org/wiki/CAPTCHA> 5. *Байесовская фильтрация спама* http://ru.wikipedia.org/wiki/Байесовская_фильтрация_спама 6. *Welcome to the new TED.com* http://www.ted.com/talks/lang/ru/luis_von_ahn_massive_scale_online_collaboration.html

Рецензент: д-р техн. наук, проф. Юсифов С.И.

Поступила в редколлегию 21.04.2014

Абдуллаев Вугар Гаджимахмудович, канд.техн. наук, доцент кафедры «Компьютерные технологии и программирование», Азербайджанская Государственная Нефтяная Академия (АГНА), Институт систем управления НАНА. Научные интересы: информационные технологии, Веб-программирования, мобильные приложения. Увлечения и хобби: электронная коммерция, B2B, B2C проекты, научные книги, спорт. Адрес: AZ1129, Баку, ул. М.Гади, 53, кв. 81 тел: (99412)571 2428, (050)332 54 83 (моб). E-mail: abdulvugar@mail.com