Спам-контроль

Для того чтобы пользователям не приходилось скачивать килобайты рекламной шелухи, провайдеры устанавливают спам-фильтры на свои почтовые серверы. Основная задача фильтра — заблокировать максимум спама и при этом не удалить ни одного важного письма.

В борьбе со спамом пользователи Сети постоянно находятся в роли догоняющих. Технологии спама совершенствуются каждый день, а для того чтобы найти на спамеров управу, необходимо время. Во многом поэтому никоим образом, даже теоретически, нельзя считать, что спам-фильтр гарантированно блокирует всю нежелательную корреспонденцию. Системы фильтрации допускают ошибки, и наиболее наглядным доказательством этого грустного факта является каждодневное пополнение вашего почтового ящика рекламным мусором.

В статистике принято различать ошибки первого и второго рода. Не вдаваясь в тонкости, будем говорить, что ошибкой первого рода является успешный проход спама через фильтр, а ошибкой второго рода — блокирование фильтром «полезного» письма. Конечно же, практически все спам-фильтры настроены таким образом, чтобы минимизировать вероятность ошибки второго рода, резонно полагая, что пусть уж лучше пользователь прочитает лишний десяток нежелательных писем, чем потеряет, к примеру, важное деловое письмо.

Положение усугубляется тем неочевидным фактом, что, вообще говоря, различие между спамом и неспамом достаточно размыто. Согласно исследованию, проведенному Yandex, при неперсонали- »

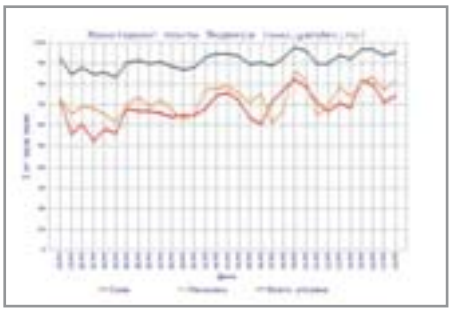
» зированной ручной фильтрации к категории полуспама участники отнесли до 40% всей корреспонденции! Действительно, большая часть информационных рассылок и почтовых уведомлений имеет явные признаки спама, и часто лишь получатель может дать окончательный ответ, который можно считать верным.

Методы фильтрации

Информацией для анализа спам-фильтром может служить любая часть электронного письма, начиная от IP-адреса отправителя и заканчивая собственно самим текстом и его оформлением. При превышении определенного критического количества совпавших признаков спам-фильтр отмечает письмо как спам. Нетрудно догадаться, что признаков оформления текста может быть очень много, от сотен до тысяч, при этом все они могут как присутствовать, так и не присутствовать в данном конкретном письме. Статистика слов тоже требует работы с тысячами записей, полученных путем анализа спама. IP-адрес же, напротив, всего один, и его проверка, по сути, сводится к поиску его в различных «черных списках».

«Черных списков» в Интернете существует достаточно много. Большинство из них поддерживается множеством энтузиастов, вносящих туда новые данные о провайдерах, лояльных к спамерам, или о некорректно настроенных почтовых серверах — открытых релеях, через которые спамеры легко проделывают свои темные дела. Существуют, конечно же, и черные списки на платной основе.

▲ «Спамтест» о фильтрации спама компании «Ашманов и партнеры»



▲ Почта Yandex отсеивает в виде спама по 95% почты ежедневно

DNSBL

Чтобы не возникало проблемы постоянного обновления списков, а также организации эффективного поиска в них, проверка осуществляется уже имеющимися и давно отработанными средствами, а именно при помощи поиска в DNS. Такой принцип получил название DNSBL, от аббревиатуры DNS и первых букв слов black list.

Самым главным недостатком этой системы является то, что фактически администраторы серверов перекладывают всю работу и ответственность по фильтрации спама на посторонних людей, которые, как правило, работают бесплатно и не отчитываются ни перед кем за эффективность своей деятельности. Случалось, что организаторы популярного «черного списка», решив забросить его поддержку, начинали подтверждать вообще все запрашиваемые адреса, видимо, с целью отпугнуть всех, кто этим списком пользовался. Хорошо, если администратор вовремя замечал подобную диверсию и исключал такую DNSBL-систему из конфигурационного файла своего спам-фильтра.

Однако и платные DNSBL тоже не станут панацеей. Ярким примером является DNSBL SORBS (dnsbl.sorbs.net). Попавший в ее «черные списки» IP-адрес может быть изъят из них только за плату. Конечно, деньги они просят небольшие, однако провайдеры и хостинговые компании активно протестуют против такой практики и настроены достаточно агрессивно. Их тоже можно понять. В конце концов, именно они и не совершали никакого преступления, тем более что попасть в общий список можно и за вполне невинную промашку, а то и просто по прихоти адми- »

Фильтры

Все средства хороши

Вероятно, самым популярным из всех серверных спам-фильтров стоит считать SpamAssassin (www.spamassassin.org). Этот свободно распространяемый программный продукт с открытым кодом также является ярким примером интегрированной системы фильтрации. Он использует методы фильтрации по IP, по «лезвию Вайпула» (одна из реализаций «нечетких» контрольных сумм), эвристические методы анализа заголовка и генетические алгоритмы для анализа самого текста — то есть большинство известных сегодня методов, за исключением, пожалуй, greylisting.

Почти такой же обширной функциональностью обладает интернет-сервис онлайн-проверки на спам spamtest.ru. Несмотря на то что его идеология состоит в работе с конечными пользователями, ничто не мешает веб-серверам электронной почты рекомендовать его использование своим клиентам и даже интегрировать нужные функции в свой интерфейс. Не лишним будет так же упомянуть, что пока spamtest.ru работает бесплатно, и зарегистрироваться в этой системе может любой желающий.

Разумеется, не осталась в стороне от такого злободневного дела, как фильтрация спама, и Лаборатория Касперского. Их решение, конечно же, небесплатно, однако и свои плюсы тоже имеет. Реализуя самые разнообразные способы фильтрации — DNSBL, «нечеткие» контрольные суммы, эвристические методы, — Лаборатория берет на себя большую часть работы по поддержке и обновлению баз данных своей системы. То есть фактически действует по привычной для себя схеме, давно отработанной в борьбе с вирусами, с той лишь разницей, что обновления антиспамовых баз выходят в полтора раза чаще, чем антивирусных.

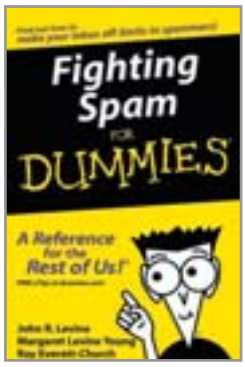
▲ SpamAssassin — почтовый фильтр, хорошо распознающий спам

нистраторов SORBS. У SORBS есть множество различных списков, часть из них содержит IP-адреса некорректно настроенных почтовых и прокси-серверов, то есть является вполне объективной и полезной информацией, а другая часть, в частности spam.dnsbl.sorbs.net, далеко не всегда располагает верной информацией. Многие администраторы, к сожалению, не считают нужным разбираться в том, какие именно списки стоит использовать, а какие — нет, подключая сразу все.

Пользоваться фильтрацией по IP-адресу нужно осторожно и ни в коем случае не доверяться только одному DNSBL. Самым безопасным вариантом будет не блокировать письма, а помещать их, например, в специальную папку почтового ящика, если это технически возможно.

Серые списки

Сравнительно новым методом борьбы со спамом являются так называемые «серые списки» (greylisting). Суть метода сводится к тому, что почтовый сервер запоминает триады, характеризующие переписку двух конкретных людей: IP-адрес сервера, который пытается осуществить доставку, и почтовые адреса отправителя и получателя. В случае, если такая триада встречается ему впервые, сервер блокирует письмо. Сервер отправителя, если он реально существует (а это уже высокая гарантия благонадежности), через некоторое время повторит доставку, триада будет



О спаме уже пишут книги: Джон Левайн, «Борьба со спамом для чайников»

внесена в базу, и дальнейшая переписка будет идти без проблем.

Минус такого подхода — первые письма с адресов, с которыми вы ни разу не переписывались, приходят с задержкой как минимум в полчаса. Именно такое время в среднем проходит, прежде чем почтовый сервер отправителя повторит попытку отослать электронное письмо. Второй минус состоит в том, что многие крупные бесплатные сервисы, например Mail.ru, предпочитают не повторять попытку отправки после первой же неудачи, чтобы не тратить ресурсы на письма, отправленные уже несуществующим адресатам. Если это необходимо, большую часть таких серверов также можно внести в «белый список», потому что вероятность получить спам от них крайне низка. В будущем, если «серые списки» получат широкое распространение, возможно, именно эту потенциальную прореху в методе будут использовать спамеры для своих рассылок.

Пока, однако, «серые списки» работают достаточно эффективно и лишены вышеупомянутых недостатков, свойственных DNSBL. К слову сказать, в последней версии OpenBSD система спам-фильтрации greylisting интегрирована в службу отправки почты по умолчанию.

Kaspersky Anti-Spam защищает от спама пользователей корпоративной почты

Анализ содержимого

Пожалуй, единственным исключением среди методов контент-анализа, вполне применимым на серверах, можно считать подсчет, накопление и последующий обмен контрольных сумм. Действительно, любая спам-рассылка по определению является массовой, следовательно, если какая-то конкретная контрольная сумма встретилась на нескольких почтовых серверах очень большое количество раз, то с высокой степенью вероятности можно утверждать, что любое письмо с такой контрольной суммой — спам, и его следует заблокировать. Этот простой и, казалось бы, надежный барьер был с легкостью преодолен спамерами. Для этого они вставляют в письма имя адресата, иногда случайное число, либо вообще рассылают в каждом письме различные анекдоты. Контрольные суммы таких писем, конечно же, не совпадают. Эффективнее считать контрольную сумму не всего письма, а каких-либо его частей. Например, выделять отдельно контрольные суммы одного или нескольких предложений и считать их все отдельными признаками, по которым и принимать решение. Эти суммы называют «нечеткими» контрольными суммами.

В общем же случае метод обмена контрольными суммами для обнаружения массовости рассылки ограничен производительностью сетей. В периоды пиковой нагрузки серверы не успевают обмениваться «свежей» информацией, и их производительность резко падает.

Большинство других алгоритмов, реализующих методы спам-фильтрации, основанные на анализе содержимого письма, малопривлекательны для использования на серверах в условиях массовой почтовой службы. Использование функции Байеса, генетические и эвристические методы могут быть очень эффективны в условиях некрупной локальной сети либо в качестве конечных спам-фильтров на клиентской машине. Но, так или иначе, именно массовость является камнем преткновения и естественным ограничением большинства из них.

▲ Проверка адреса в более чем ста DNS-базах «черных списков»

▲ Поиск адреса 213.33.152.252 сразу в нескольких «черных списках»

▲ Спам вывозится на природу и уничтожается собственноручно

» Важным следствием массовости является необходимость неперсонализированности фильтра. Однако, как уже было упомянуто выше, даже человек при ручной фильтрации и при неперсонализированном подходе не сможет обеспечить приемлемый уровень вероятности ошибки первого рода, не говоря уже об алгоритме конечной сложности. Выход, таким образом, видится один — позволить пользователю самостоятельно заниматься настройкой правил (или обучением) спам-фильтра для всей почты, приходящей на его почтовый адрес. В компании «Караван-Интернет» (www.caravan.ru) такая система внедрена в качестве эксперимента (используются рассмотренные ниже SpamAssassin и spamtest.ru) в дополнение к стандартной DNSBL-фильтрации. Судя по их отзывам, предварительный результат не самый впечатляющий: большинство пользователей не желают постоянно заниматься обновлением и обучением персональных фильтров. Результат — довольно высокий процент ложных срабатываний. Это, конечно же, не является каким-либо серьезным аргументом против дальнейших попыток применения методов контент-анализа. Несомненно, при наличии качественной обратной связи в совокупности с фильтрацией по IP-адресу эти методы дадут хорошие результаты и смогут снизить вероятность ошибки первого рода практически до теоретического минимума.

Заключение

Каждый пользователь хотел бы навсегда избавиться от навязчивой рекламы, ежедневно скапливающейся в его электронном почтовом ящике. Что же нужно сделать для того, чтобы рассылки перестали быть эффективными, перестали оправдывать вложенные в них деньги, погасив тем самым растущую активность спамеров? В первую очередь необходимо понять, что постоянные звонки в Центр американского английского, сопровождаемые нецензурной бранью, общему делу не помогут. Как и в любом другом деле, не стоит пытаться в одиночку решать глобальные задачи.

Современное общество устроено так, чтобы реагировать на конкретный спрос. Дайте понять, что проблема спама действительно вас волнует — сообщайте своему провайдеру об ошибках его системы фильтрации, если она обладает средствами обратной связи. Если таких средств не имеется, интересуйтесь, когда планируется их появление и планируется ли вообще. Смените провайдера или почтовый сервис, если его администрация не желает совершенствовать свои спам-фильтры. Жесткая конкуренция сделает свое дело.

Спам — это тяжелая болезнь современного Интернета, едва ли не такая же серьезная, как вирусы и почтовые черви. Из-за нашего с вами снисходительного отношения к спаму российский Интернет потерял четверть миллиарда долларов в прошлом году. Давайте постараемся в нынешнем году отнестись к этой болезни серьезнее.

■ ■ ■ **Анатолий Юдов**

Комментарий специалиста

Выбор спам-фильтра для провайдера

Сергей Хрипко
руководитель отдела хостинга компании «Караван»

Специалисты «Каравана» протестировали Spamtest, систему антиспама для провайдеров от компании «Ашманов и партнеры». В ходе испытаний было установлено, что система вполне может успешно фильтровать спам в мелких и средних организациях, а вот для крупных провайдеров возможностей Spamtest явно недостаточно. Также программа компании «Ашманов и партнеры» не позволяет настраивать фильтры для каждого отдельного ящика. Единственная возможность улучшить работу Spamtest — скачивать для нее обновления. Поэтому мы решили отказаться от ее использования. Качество работы Spamtest было вполне удовлетворительным.

Другой протестированной системой антиспама была программа CRM114. Эта разработка имеет другую «философию» настройки фильтров. Изначально пользователь сам указывает программе, какие письма спам, а какие — нет. В ходе работы программы можно помечать письма, в которых она ошиблась. Это дает нам гибко настроенную систему фильтров под каждого конкретного пользователя. Возможности CRM114 не ограничиваются фильтрацией спама. Система позволяет настроить фильтрацию на различные категории — «Работа», «Личное», «Спам» и т. п.

В будущем мы планируем перевести систему фильтрации спама для наших клиентов на основу CRM114. Мы не станем ставить для всех пользователей единственный спам-фильтр, созданный нами. При индивидуальном подходе наши клиенты получат возможность настраивать фильтры лично под себя.