НАУЧНОЕ ИЗДАНИЕ МГТУ ИМ. Н. Э. БАУМАНА
НАУКА и ОБРАЗОВАНИЕ
Эл № ФС77 - 48211. Государственная регистрация №0421200025. ISSN 1994-0408
электронный научно-технический журнал

Что такое СПАМ?

77-48211/551869
02, февраль 2013

Афонин А. И.

УДК 004.7:347.4

Россия, МГТУ им. Н.Э. Баумана
sa853@bmstu.ru

Одним из самых популярных сервисов сети Интернет является электронная почта. По приблизительным оценкам ее ежедневный объем превышает 200 млрд. единиц. Если учесть трафик, передаваемый по внутренним каналам больших корпоративных сетей и сетей Интранет, то общее число электронных сообщений значительно превзойдет эту громадную величину. Значительную долю этого информационного массива составляют нежелательные сообщения или послания, опасные для информационной безопасности адресата – spam (спам). По некоторым оценкам [5] доля спама может достигать 85% от совокупного информационного трафика для определенных категорий пользователей электронной почты. Этот феномен представляет значительную опасность для целостности и конфиденциальности ключевой информации предприятия или персональных данных пользователя. Авторы несанкционированных рассылок должны нести юридическую ответственность за последствия своей противоправной деятельности, а пользователей почтовых сервисов следует надежно оградить от спама современными программно-техническими системами защиты.

Если правовые аспекты регулирования спама не пока получили исчерпывающих формулировок, то технические методы и средства борьбы с ним развиты достаточно глубоко. Целью данной работы является анализ возможных подходов к точному юридическому определению понятия «спам» рассылки и обзор технических методов фильтрации спама.

Определение спама

По поводу специфических характеристик спама имеется широкий диапазон мнений [1-4]. Известно большое количество различных, вплоть до «бытовых» определений, многие из которых не раскрывают сути термина, весьма поверхностны или расплывчаты и поэтому не могут быть использованы в юридической практике.

В российском законодательстве была предпринята попытка правового регулирования распространения спама в ст. 18 Федерального закона от 13 марта 2006 г *№* 38-ФЗ «О рекламе». Согласно указанной статье распространение рекламы по сетям электросвязи, в том числе посредством использования телефонной, факсимильной, подвижной радиотелефонной связи, допускается только при условии предварительного согласия абонента или адресата на получение рекламы. При этом реклама признается распространенной без предварительного согласия абонента или адресата, если рекламораспространитель не докажет, что такое согласие было получено. Практика показала, что это определение не может быть совершенно корректным. Так, например, переход по некоторым ссылкам – совершенно обычная и легальная операция для подавляющего числа пользователей Интернет – может быть квалифицирована как нелегальная.

Сейчас каждый адресат электронной почты (конечный пользователь или почтовый сервер) самостоятельно принимает решение об отнесении поступившего сообщения к категории «спам», руководствуясь при этом некоторой собственной системой критериев или признаков. В простейших случаях это может быть однокритериальная пользовательская оценка. Приведем несколько вариантов определения спама такого типа:

- сообщения от лиц, с которыми адресат не состоит в регулярной переписке;
- любые сообщения рекламного характера;
- рекламные сообщения, в которых используются грубые или вульгарные приемы продвижения предлагаемого товара или услуги
- реклама определенной категории товаров или услуг (например, предложения о быстром обогащении или обучении иностранному языку за несколько дней и пр.);
- массовые сообщения;
- регулярные сообщения, приходящие с некоторой постоянной периодичностью, например каждый день;

- сообщения, полученные с адресов, которые занесены в так называемые черные списки;
- сообщения, рассылаемые посредством специального программного обеспечения, предназначенного для email-рекламы. Например, SpamPro, Massbulk Mailer, ePochta Mailer и пр.

Однако целесообразно использование многокритериального подхода, в соответствии с которым сообщение может быть квалифицировано как спам при удовлетворении его нескольким критериям (наличии нескольких признаков). Например, «Спам – это анонимная, массовая, несанкционированная рассылка почтовых сообщений».

В общем случае в качестве критериев могут использоваться технические, информационные, технологические, коммуникационные и другие признаки, свойства и характеристики сообщения. Практически во всех известных определениях спама обращается внимание на:

- массовость, незапрошенность спам-сообщений;
- способы получения адресов для рассылки;
- формальные признаки сообщений;
- используемое программное обеспечение и некоторые другие факторы.

Исходя из вышеизложенного, целесообразно дать характеристику и оценку известным признакам спама с целью последующего формулирования юридически значимого определения понятий «спама» и «спаминга».

Известен подход, в соответствии с которым предлагается во множестве признаков, которые могут характеризовать электронное сообщение как спам, выделить три группы – формальные, лингвистические и технологические [2]. Данный подход представляется весьма полезным, поскольку структурирует признаки по их природе. Однако значительное их число, которое у разных авторов существенно отличается, создает трудности определения набора признаков, необходимых и достаточных для отнесения сообщения к категории спам.

Поэтому определение основных и дополнительных (квалифицирующих) признаков спама является актуальной задачей, решение которой позволит определить содержание термина «спам». Для этого, прежде всего, необходимо сформулировать основной принцип спам-категорирования электронных сообщений, который, по нашему мнению, должен

отвечать требованию «прямого применения», то есть использования его в реальном масштабе времени без подготовки и навыков адресата, а также без применения им специальной аппаратуры.

Трудности с юридически корректным и исчерпывающим определением термина «спам» можно преодолеть, если воспользоваться громадным опытом, накопленный инженерами и программистами, разрабатывающими и эксплуатирующими программно-технические системы фильтрации спама.

Методы фильтрации электронных сообщений

Существует несколько сетевых протоколов для обмена электронными сообщениями, но в наше время самым распространенным из них является протокол SMTP (simple mail transfer protocol). Это стандарт де-факто для пересылки электронной почты в любых сетях TCP/IP. Согласно этому протоколу любое электронное сообщение состоит из двух частей: заголовка и тела письма. В заголовок входят служебные данные, идентифицирующие сообщение. Тело содержит текстовый или графический контент сообщения.

Заголовок разбивается на поля, главными из которых являются: Received, Return-Path, To, Reply-To, From, Date и некоторые другие. На основе сведений, содержащихся в этих полях, может быть выполнена предварительная фильтрация входящих электронных сообщений.

Поле заголовка Received используется для идентификации цепочки почтовых серверов, которые принимали участие в процессе доставки сообщения от отправителя получателю. Каждый сервер добавляет к почтовому сообщению свое поле Received, с указанием специфических сведений о себе. Поле Return-Path хранит адрес на который должно быть переслано почтовое сообщение в случае, если оно не может быть доставлено адресату. Поле Reply-To содержит адрес электронной почты, на который будет послан ответ на сообщение. Поле From хранит адрес отправителя почтового сообщения. На основе информации данного поля реализуются сортировка сообщений по белым и черным спискам (см далее). Поле Date содержит информацию о дате и времени отправления письма. Поле To содержит адрес получателя письма.

Перечисленные поля являются обязательными для любого почтового сообщения. Кроме того в заголовке могут присутствовать несколько необязательных полей, дающих более детальную информацию о сообщении для сервера SMTP. Самым полезным с точки зрения фильтрации спама является поле Subject (Тема). Согласно протоколу RFC 822, оно

должно содержать краткую информацию о теме письма. По содержанию этого поля адресат должен получать предварительное общее представление о смысле послания. Существую эффективные алгоритмы фильтрации спама, основанные на лексическом и смысловом анализе поля Subject и соответствии его содержания тексту, приведенному в теле письма.

Процедуру получения и обработки нового электронного сообщения можно разделить на следующие основные этапы:

1. получение заголовка сообщения почтовым сервером;
2. анализ сведений, извлеченных из заголовка;
3. прием тела письма или отказ от его получения;
4. анализ тела письма. На основе глубокого анализа контента письма можно принять окончательное решение о фильтрации сообщения как спама или подтверждении его валидности;
5. доставка письма адресата или его удаление.

Необходимо отметить важность второго этапа процедуры обработки почтовых сообщений. Многие распространённые методы фильтрации спама принимают решение на основе анализа данных заголовка. Сортировка писем на данном этапе не требует больших вычислительных затрат и использования сложных алгоритмов интеллектуального анализа текста.

В общем случае анализ заголовочной части может включать в себя следующие мероприятия.

1. Проверка подлинности домена, с которого отправлено электронное сообщение Если почтовый сервер отправителя не существует, то письмо классифицируется как спам.
2. Проверка сетевого узла, который передал почту на соответствие MX-записям DNS.
3. Проверка почтового адреса отправителя.
4. Анализ темы письма с помощью лексического анализатора.

Примерная классификация способов фильтрации спама приведена на рис. 1. Первая группа методов основывается на некоторых организационных мероприятиях, позволяющих отсечь нежелательные сообщения или свести их количество к приемлемому минимуму.

Системы с запросом на подтверждение (challenge response systems, системы вызов-ответ). В этих системах используется очень простой прием. По электронному адресу отправителя автоматически высылается запрос на валидацию почтового сообщения. Отсутствие подтверждения служит надежным свидетельством спама. Если, как это часто бывает в спам рассылках, адрес отправителя фиктивный, то обратная связь не будет установлена и сообщение автоматически удаляется. Если адрес действительный и инициатором спам кампании является человек или некоторый коллектив, то, предполагается, что им будет не выгодно отвечать на запросы подобного типа. Понятно, что справедливость обоих предположений относительна, а защита данного типа может быть нарушена.

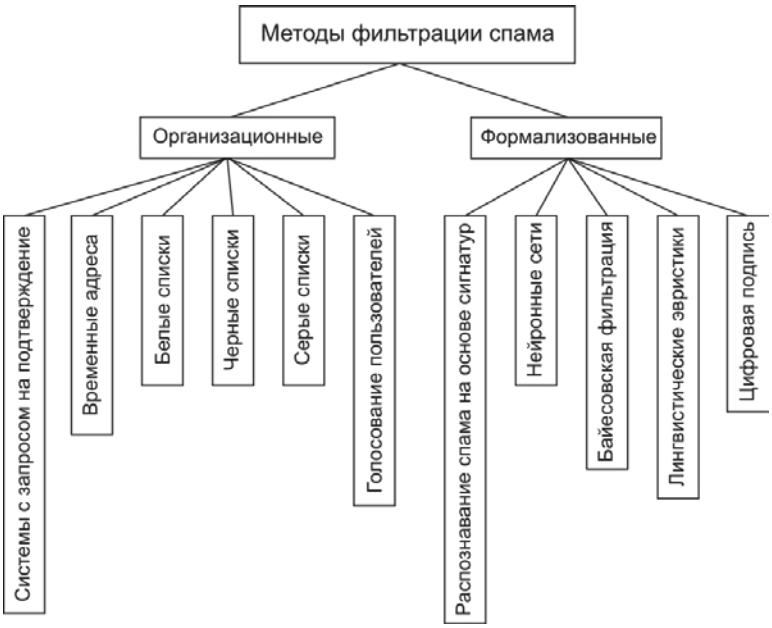

Рис. 1. Классификация методов фильтрации спам-сообщений

В современной практике используются различные приемы валидации по запросу. Перечислим некоторые из них:

- запрос, содержащий указатель на сетевой ресурс (URL). Ответом на такой запрос будет переход по указанной ссылке;
- запрос, требующий от отправителя выполнения некоторой процедуры, например, распознавание кода, цитирование короткого фрагмента из заведомо известного текста, ответ на вопрос об отправителе или получателе и др.;
- в некоторых случаях для тщательной фильтрации входного потока используется запрос на перечисление небольшой платы на счет получателя. Если полученное письмо не является спамом, то деньги возвращаются отправителю.

В общем, основная идея систем с запросом подтверждения состоит в повышении почтовых издержек (временных, интеллектуальных, материальных и пр.). Для генераторов спама, чей доход напрямую зависит от массовости рассылки, увеличение трудоемкости может оказаться довольно действенной запретительной мерой.

Эффективность таких систем в значительной степени зависит от компромисса между сложностью запроса и затратами на ответ. С одной стороны, проверка не должна отпугнуть реальных пользователей от коммуникации. С другой стороны, стоимость систем, способных автоматически преодолевать защиту данного типа, обязана быть настолько высока, чтобы рассылка спама стала экономически невыгодной.

Способ временных адресов (time-stamping) – это очень простой прием, основная идея которого – использование временных почтовых адресов. Если на зарегистрированный адрес начинает приходить большое количество несанкционированных сообщений, то пользователь просто приостанавливает работу с ним или даже аннулирует данный адрес.

Черные списки (blackhole lists). Это один из самых простых и достаточно эффективных методов борьбы со спамом. Все входящие сообщения проходят предварительную фильтрацию, когда проверяются адреса отправителей. Адресаты, замеченные в несанкционированной рассылке, заносятся в черные списки. Дальнейшая коммуникация с ними автоматически блокируется. Черные списки могут формироваться как почтовым сервером, так и почтовым агентом на персональном компьютере пользователя. В списки могут заноситься не только уникальные почтовые адреса, которые легко изменить, но и группы адресов и даже доменные имена (например *@spam.com), IP-адреса серверов и клиентов.

Белые списки (safe lists, wtite lists). Белые списки формируются из почтовых адресов, доменных имен и IP-адресов, которые не замечены в несанкционированной рассылке сообщений. Такие списки могут использоваться почтовыми серверами и составляться конечными пользователями при помощи средств почтового клиента.

Серые списки (grey lists). Во время почтовой сессии на письма от неизвестных отправителей выдается код ошибки 45*, что означает наличие временных проблем. Программы, настроенные на отправление обычной почты, повторяют попытку через небольшой промежуток времени (обычно час или полтора). Сообщения, прошедшие «проверку временем», принимаются. Спамерские почтовые агенты, как правило, отказываются от повторной коммуникации.

Метод голосования пользователей основан на сборе статистической информации о нежелательной корреспонденции. На почтовом сервере накапливаются и обрабатываются данные о подозрительных сообщениях, полученных пользователями. Если количество уведомлений по определенному источнику превысит установленный порог, то его сигнатура заносится в специальную базу, все сообщения с такой сигнатурой (см. далее) считаются в дальнейшем спамом. Понятно, что эффективность этого метода всецело зависит от активности и добросовестности пользователей.

Распознавание спама на основе сигнатур. Сигнатура – это образ или характеристика электронного сообщения, предназначенные для идентификации оригинала. В качестве сигнатур могут быть использованы короткие текстовые фрагменты, ключевые слова или их наборы, текстовые свертки, контрольные суммы и пр. Важно, чтобы такой образ полноценно идентифицировал свой носитель и, по сравнению с ним, имел существенно меньший размер. Для каждого нового письма вычисляется его сигнатура и сравнивается с базой, в которой хранятся характеристики сообщений ранее классифицированных как спам. При совпадении сигнатуры письма с одной из записей базы сообщение считается нелегальным. Данный метод отличается низким процентом ложных срабатываний, поскольку редко блокирует легитимные письма. Однако он не реагирует на новые спам-сообщения, сигнатура которых не представлена в базе. Кроме того, данную проверку можно обойти небольшим изменением текста письма.

Лингвистические эвристики. Многочисленные методы этой группы основаны на поиске в теле письма ключевых слов и словосочетаний, которые позволяют отнести данное сообщение спаму. Например, это могут быть термины из фармацевтики, предложения о выгодной покупке, информация о крупном денежном выигрыше, анонсы методик быстрого обогащения, реклама способов оздоровления и пр. Если количество подозрительных лексических единиц в тексте письма превышает некоторый установленный порог, то сообщение классифицируется как спам. В некоторых случаях, методы этой группы могут давать высокий процент ложных срабатываний, когда легальные письма не проходит спам-проверки. При помощи языка регулярных выражений можно сформулировать эффективные поисковые образцы для нахождения в больших текстах ключевых слов и фраз. По этой причине методы этой группы иногда называются методами фильтрации с использованием регулярных выражений.

Системы фильтрации с использованием электронной цифровой подписи. Электронная цифровая подпись представляет собой реквизит документа, полученный в результате

криптографического преобразования информации с использованием так называемого закрытого ключа. Цифровая подпись предназначена для защиты электронного документа от подделки. Ключ позволяет идентифицировать владельца, а также определить наличие фальсифицированных данных или искажений оригинала. Цифровой подписью могут быть защищены как отдельные поля документа, так и весь контент. При любом случайном или преднамеренном редактировании документа или его части изменится значения вычисленного хеша, следовательно, подпись станет недействительной.

Самообучающиеся фильтры Байеса. В наше время эта техника фильтрации спама очень распространена. Она основана на классической теореме Байеса для вычисления вероятностей условных событий. Успешное распознавание спама требует предварительного обучения фильтра на значительном массиве сообщений. Обучающая выборка представляет собой отсортированный вручную массив писем, в котором указаны нормальные и нежелательные сообщения.

Пусть получено сообщение, в тексте которого встречается слово *W*, которое может сигнализировать о нежелательности сообщения. Обозначим через $P(S|W)$ – вероятность события «письмо, содержащее слово *W* является спамом». Согласно теореме Байеса о вычислении вероятности условного события имеем $P(S|W) = \frac{P(S \cap W)}{P(W)}$, где $P(W)$ – априорная вероятность появления слова *W* в любых сообщениях, а $P(S \cap W)$ – вероятность совместного события появления спама и контрольного слова. Числитель этого выражения можно записать в виде $P(W|S)P(S)$, тогда $P(S|W) = \frac{P(W|S)P(S)}{P(W)}$.

По формуле полной вероятности имеем $P(W) = P(W|S)P(S) + P(W|\overline{S})P(\overline{S})$, где $\overline{S}$ событие «письмо не является спамом», а $P(\overline{S})$ – его вероятность. В итоге получим $P(S|W) = \frac{P(W|S)P(S)}{P(W|S)P(S) + P(W|\overline{S})P(\overline{S})}$.

Это базовое выражения, которое можно использовать для оценки вероятности спама в различных ситуациях. Рассмотрим самый простой вариант применения данной формулы – расчет вероятности спама по одному слову без самообучения. Иногда этот вариант называется «наивной байесовской фильтрацией» [5] Для вычисления условных вероятностей $P(W|S)$ и $P(W|\overline{S})$ используется обучающая выборка. Оценкой первой

вероятности служит частота слова W в спаме $P(W|S) \approx N_{sw} / N_s$, а второго – частота появления в легальных письмах $P(W|\overline{S}) \approx N_{\overline{s}w} / N_{\overline{s}}$. Значения $P(S)$ $P(\overline{S})$ определяются на основе некоторой гипотезы, достоверной в конкретных условиях функционирования системы защиты. Например, часто принимается предположение о равной вероятности обоих событий S и $\overline{S}$, то есть $P(S) = P(\overline{S}) =$ 0,5. Такие системы часто называют «фильтрами без предубеждений». Фильтры почтовых адресов, которые требуют тщательной защиты, основываются на более осторожных оценках $P(S)$ = 0,2, а $P(\overline{S})$ = 0,8. Полученное сообщение, содержащее слово W, считается спамом, если рассчитанная величина $P(S|W)$ превосходит некоторый установленный порог. Обычно пороговое значение выбирается в диапазоне 0,6 – 0,8.

Метод байесовской фильтрации прост с вычислительной точки зрения. Даже самые продвинутые методики этого типа (с оценкой по набору слов, списками исключений и самообучением) используют элементарные формулы и простые операции. Опыт эксплуатации показал, что при достаточно большой обучающей выборке эффективность байесовских фильтров может достигать 95-97%. Большая часть современных систем защиты от спама используют технику байесовской фильтрации.

Отметим некоторые особенности байесовских систем защиты.

1. Редкие слова. Редким считается слово, у которого мало вхождений в обучающую выборку. В этом случае числитель и знаменатель основной формулы вычисления вероятности спама становятся близкими к нулю, что делает невозможным ее использование. Обычно редкие слова игнорируются.
2. Нейтральные слова. Нейтральными называются такие слова, которые могут встретиться в тексте любого содержания. В русском языке это многие глаголы, в английском – такие лексические единицы, как «some», «is и пр. Обычно при обработке текста такие слова просто пропускаются.
3. Некоторые развитые системы байесовской защиты вместе с отдельными словами анализируют целые словосочетания заданной длины (patterns). Этот способ более чувствителен к контексту и устраняет шум точнее, поскольку основан на базе данных большего размера.
4. Метод основывается на предположении о том, что словарная частота зависит от содержания и тематики текста. Если это предположение не выполняется, то попытка фильтрации может оказаться неуспешной. На этом основана известная

спамерская уловка, когда подлинный смысл сообщения маскируется фоновым текстом. Понять истинный смысл подобного может только человек, вчитавшись в него. Этот прием называется «байесовским отравлением».

Значительные усилия исследователей направлены на разработку контентных фильтров на основе современных методов нейронных сетей и автоматической классификации [1]. В настоящее время перечисленные подходы не вышли из стадии лабораторных исследований и не используются в составе прикладных систем защиты. Если проблема спама имеет исчерпывающее решение, то, по всей видимости, оно будет получено при помощи интеллектуального машинного анализа контента сообщения.

Список литературы

1. Доля А. Тенденции развития спама и средства борьбы с ним // Компьютер Пресс. –2006. – № 10. – с. 4–7.
2. Калятин В.О. Право в сфере Интернета. – М.: Норма. –2004. – 480.
3. Кеглер Т., Доуминг П., Тейлор Б., Тестерман Д. Реклама и маркетинг в Интернете. – М.: Альпина Паблишер, 2003. – 360.
4. Левин М. Антиспам без секретов. Практические рекомендации по борьбе с нелегальной рассылкой по электронной почте. – М.: Пресс, 2006. – 320 с.
5. Спам. Статья из свободной энциклопедии. http://ru.wikipedia.org.