Математические структуры и моделирование
2010, вып. 21, с. 133–140

УДК 519.688

МЕТОД РАСПОЗНАВАНИЯ СПАМ-СООБЩЕНИЙ НА ОСНОВЕ АНАЛИЗА ЗАГОЛОВКА ПИСЬМА

А.Н. Мироненко

Разработана и реализована технология борьбы с массовыми рассылками на основе временной задержки сообщения и запроса о его повторной отправке. Проведено исследование влияния величины времени задержки сообщения на эффективность метода распознавания спам-сообщений.

Введение

Все современные методики борьбы со спамом можно условно разделить на следующие категории:

1. Методы, основанные на анализе письма. Их задача состоит в изучении письма с целью его классификации. Подобный метод решает две задачи: обнаруживает спам и выявляет письма, которые гарантированно не являются спамом. Известно несколько наиболее распространенных методов анализа:
   - По формальным признакам.
   - По содержимому с использованием сигнатурного анализа. Данный метод основан на поиске в тексте письма определенных сигнатур, описанных в обновляемой базе данных.
   - По содержимому с применением статистических методик. Большинство современных методов данного класса основано на теореме Байеса.
   - По содержимому с использованием SURBL (Spam URL Realtime Block Lists — списка блокировки спамерских URL). Идея метода состоит в поиске расположенных в теле письма ссылок и их проверке по базе SURBL. Этот метод эффективен против спама, в котором для обхода фильтров вместо рекламы применяется ссылка на сайт с рекламой.
2. Методы, основанные на признании отправителя письма в качестве спамера. Они опираются на различные «черные», «белые» списки IP- и почтовых адресов.

Омский государственный университет им. Ф.М. Достоевского.
E-mail: mironim84@mail.ru

3. Детекторы массовой рассылки. Как следует из названия, их задачей является обнаружение рассылки похожего письма большому количеству абонентов.
4. Методы, основанные на верификации обратного адреса отправителя и его домена. Простейшим методом защиты является обычный DNS-запрос по имени домена отправителя письма. Если выясняется, что домен отправителя не существует, то, вероятнее всего, адрес отправителя является поддельным. Однако этот метод малоэффективен, поскольку в качестве адреса отправителя спамеры могут использовать реальные адреса, случайным образом выбранные из базы рассылки.

Рассмотрим способ фильтрации спама, разработанный на основе анализа заголовка сообщения с запросом повтора отправки.

1. Описание разработанного метода фильтрации

Принцип, реализованный в данной работе, можно отнести сразу к нескольким категориям методик борьбы со спамом. Для распознавания спам-писем в нем используется комбинация проверки по «белому», «черному» списку и временной задержки сообщения. Данный метод обладает двумя ключевыми качествами, характеризующими эффективность работы любого метода фильтрации электронной почты. Это полнота и точность фильтрации. Под полнотой подразумевается процент обнаруженного спама, точность — это количество ложных срабатываний.

Пусть есть папка «карантин» (для задержки сообщений), список адресов электронной почты доверенных пользователей («белый» список), который на начальном этапе формируется самостоятельно, и «черный» список, изначально он пуст. Множество сообщений электронной почты (M). Время хранения сообщения в «карантине» (N) будет константой системы и выбирается на основе эксперимента. Первоначально при увеличении N эффективность метода возрастает, но при достижении некоторого порогового значения меняется слабо. В качестве рабочей величины N было выбрано 5 дней.

Первичная обработка входящего сообщения состоит в выделении заголовка письма. Он содержит следующие ключевые поля:

1. Return-Path — обратный адрес.

2. Received — строчка журналирования прохождения письма. Каждый почтовый сервер (MTA) помечает процесс обработки этим сообщением. Если сообщение проходит через несколько почтовых серверов, то новые сообщения дописываются над предыдущими (и журнал перемещения читается в обратном порядке, от ближайшего узла к самому дальнему).

3. From — имя и адрес отправителя. Может не совпадать с Return-Path.

4. Sender — отправитель письма. Добавлено для возможности указать, что письмо от чьего-то имени (from) отправлено другой персоной (например, секретарем от имени начальника).

5. To — имя и адрес получателя. Может содержаться несколько раз (если письмо адресовано нескольким получателям).

6. cc — (от англ. carbon copy) содержит имена и адреса вторичных получателей письма, к которым направляется копия.

7. bcc — (от англ. blind carbon copy) содержит имена и адреса получателей письма, чьи адреса не следует показывать другим получателям.

8. Reply-To — имя и адрес, куда следует адресовать ответы на это письмо.

9. Message-ID — уникальный идентификатор сообщения. Состоит из адреса узла-отправителя и номера (уникального в пределах узла). Алгоритм генерации уникального номера зависит от сервера. Выглядит примерно так: E1Nfsty-0001Ux-00. example-bk-ru@f118.mail.ru. Вместе с другими идентификаторами используется для поиска прохождения конкретного сообщения по журналам почтовой системы (почтовые системы фиксируют прохождение письма по его Message-ID) и для указания на письмо из других писем (используется для группировки и построения цепочек писем). Обычно создаётся первым почтовым сервером (MTA) в момент принятия почты от пользователя.

10. In-Reply-To — Указывает на Message-ID, для которого это письмо является ответом (с помощью этого почтовые клиенты могут легко выстраивать цепочку переписки — каждый новый ответ содержит Message-ID для предыдущего сообщения).

11. Subject —тема письма.

12. Date — дата написания письма.

Для данного метода распознавания спама важны поля Return-Path, Message-ID и Subject. Сообщение копируется по протоколу POP3. Затем проверяем вхождение адреса отправителя в «черный» список. Если он найден, то сообщение – спам, и оно удаляется, если нет, то проверяется по «белому» списку. Если адреса нет в списке, то сообщение копируется в папку «карантин» и определяется его Message-ID, указанный в заголовке. По шаблону формируется ответное сообщение, содержащее просьбу о подтверждении отправки с указанием в поле Subject ранее выявленного Message-ID. У каждого входящего сообщения Mi, попавшего в «карантин», адрес отправителя сравнивается с множеством адресов отправителей письма, от которых уже находятся в этой папке M-Mi. Если обнаруживается пара писем Mi и Mi+1 с одинаковыми адресами, то проверяется поле Subject письма Mi, содержащее некоторый набор символов, и сопоставляется с Message-ID Mi+1. В случае совпадения сообщение Mi удаляется, а Mi+1 перемещается в папку «входящие», и адрес отправителя добавляется в «белый» список. В случае не обнаружения пары M и M1, сообщение остается в «карантине»

и по истечении срока хранения удаляется с занесением адреса отправителя в «черный» список.

Если необходимо осуществить переписку с кем-либо без подтверждения отправки (добавив адресата в «белый» список автоматически), при составлении сообщения дописывается код в поле письма Subject, вся исходящая корреспонденция проверяется на наличие в ней этого кода. Если он обнаружен, то адрес автоматически добавляется в список доверенных, все ответы поступают без задержки.

2. Описание критериев оценки спам-фильтров

Для оценки качества работы антиспам-сервисов следует одновременно использовать два следующих критерия:

1. Точность (ложные срабатывания, или false positive) — доля нормальных (не являющихся спамом) сообщений, ошибочно классифицированных как спам в общем потоке нормальной почты.

2. Пропуск спама (false negative) — доля пропущенного спама в общем потоке спама. Полнота – доля отфильтрованного спама.

Обе характеристики нужно рассчитывать корректно, а именно:

1. Процент ложных срабатываний — это отношение числа нормальных писем, ошибочно признанных спамом, к количеству всей нормальной почты (пропущенных и заблокированных нормальных писем), а не от всего потока, включающего и спам тоже. Таким образом, 0,3% ложных тревог могут означать, например, что всего пришло 10 000 нормальных писем, и из них 30 было ошибочно признано спамом.

2. Процент пропусков — отношение количества пропущенного спама к объему всего спама (как пропущенного, так и распознанного). Таким образом, 15% пропусков (уровень фильтрации – 85%) означают, например, что всего пришло 10 000 спам писем, из которых 1 500 не было распознано как спам.

Ложные срабатывания можно разделить на критические и некритические.

В ситуациях, когда электронная почта является важным каналом коммуникации для компании, необходимо поддержание максимально низкой доли ложных срабатываний, особенно для важных деловых писем. Ущерб от потерянного делового письма может быть несопоставим с потерями рабочего времени от спама (это не означает, естественно, что спам вообще не нужно фильтровать).

При анализе ложных срабатываний недостаточно ограничиться только подсчетом их количества. В современных спам-фильтрах используются эвристические алгоритмы, которые могут распознать как спам (или «возможно спам») сообщения, «похожие на спам» (например, письмо всем пользователям интернет-магазина о скидках, написанное с использованием маркетинговой лексики), но при этом спамом с точки зрения получателей не являющееся. Целесообразно

при тестировании разделить ложные срабатывания на критические ложные срабатывания (ложные срабатывания на важной деловой или личной почте) и некритические (ошибочная классификация массовых новостных и маркетинговых рассылок и тому подобной почты) и подсчитывать процент тех и других отдельно.

Критерий на основе доли пропущенного спама является наиболее очевидным. Если один антиспам-фильтр распознает 70%, а второй — 85% спама, то второй фильтр можно считать лучшим. В то же время необходимо понимать, что повышение уровня распознавания может с большой вероятностью дать одновременный рост количества ложных срабатываний.

Поэтому оба критерия нужно рассматривать совместно, причем оценка количества ложных срабатываний должна иметь приоритет при составлении суммарной оценки фильтра.

3. Описание методик тестирования

Требования к эксперименту для получения наиболее достоверного результата следующие:

1. Реальная эксплуатация на реальном потоке почты в реальном времени. Наиболее достоверные результаты тестирования антиспам-систем можно получить только на реальном потоке почты и только при фильтрации немедленно, в реальном времени. Только в этом случае:
   - Распределение почты по типам (спам, неспам и так далее) соответствует реальному.
   - Техническая информация в письмах (IP-адрес посылающей стороны, SMTP envelope, технические заголовки) соответствует реальному положению дел. Содержимое баз данных фильтров (лингвистических, статистических, RBL-списков, «черных»,«белых» списков отправителей) является актуальным. Тексты писем не искажены за счет пересылки, вставки дополнительной информации или подобных действий.
   - Решается задача реальной фильтрации.
2. Тестирование должно продолжаться, как минимум, 2–3 недели. Поток, как спама, так и нормальной почты, сильно меняется во времени, обычно изменения тематики и оформления писем происходят ежедневно. Продолжительный тестовый период должен усреднить эти колебания. Полезно, если часть тестового периода может включить в себя сезонные изменения маркетинговой активности (предпраздничные распродажи, например). Это позволит оценить качество реакции антиспам-системы на пике спама.
3. При тестировании через систему должны пройти несколько десятков тысяч сообщений. В противном случае достоверно оценить уровень ложных срабатываний невозможно (так как приемлемый уровень некритических ложных срабатываний — не выше 0,01%, то есть одна ложная тревога на 10 тысяч писем или меньше).

4. В тестировании должны принимать участие, как минимум, несколько десятков почтовых ящиков. Это требование определяется тем, что вариативность потока спама у разных пользователей очень велика. Например, на ящики с именами info@, sales@ или alex@ приходит много мусорной почты, так как подобные имена легко подбираются методом словарной атаки, а на ящики со сложными именами наподобие Joe.V.User@ спама приходит во много раз меньше. Использование при тестировании большого числа почтовых ящиков позволяет усреднить вариации в потоках спама между различными типами почтовых ящиков.
5. Анализ результатов необходимо проводить с использованием единого определения спама и критичности, некритичности ложных срабатываний. Как пропуски спама, так и (в особенности) ложные срабатывания должны быть тщательно проанализированы. При оценке доли пропусков необходимо использовать корректное и единое для всех тестов определение спама. При оценке ложных срабатываний следует учитывать их критичность, поскольку это принципиально для оценки рисков использования конкретного фильтра.
6. Равные условия тестирования. При сравнении нескольких решений от разных производителей антиспам-фильтры должны быть поставлены в равные условия. Это включает в себя следующие требования:
   - Одинаковый поток почты, приходящий на разные фильтры в реальном времени.
   - При использовании RBL-сервисов — одинаковый набор списков RBL для всех тестируемых систем.
   - При использовании локальных «черных», «белых» списков — использование одинаковых списков.
   - При использовании обучаемых фильтров — обучение на одинаковых выборках. Если в процессе тестирования используется дообучение, дообучение должно быть синхронным по одним и тем же выборкам.
   - При использовании фильтров с получениями обновлений — синхронное получение обновлений.

Таким образом, достоверные результаты тестирования можно получить при выполнении следующих необходимых условий:

1. Тестирование в реальном окружении (установка антиспам-фильтра на тот же поток почты, где его предполагается в дальнейшем использовать) с достаточной продолжительностью тестирования — 2–3 недели.
2. Достаточный объем тестирующей выборки, как минимум, несколько тысяч сообщений в день.
3. Достаточная выборка почтовых ящиков, как минимум, несколько десятков.
4. Анализ результатов с использованием корректного определения спама и категорий критичных/некритичных ложных срабатываний.
5. Тестируемое ПО должно быть поставлено в максимально одинаковые условия.

4. Тестирование разработанного метода

Для корректного определения уровня качества фильтрации потока сообщений дадим определение понятию спам. Спам — это любые анонимные незапрошенные массовые рассылки электронной почты, как правило, имеющие рекламный характер.

На этапе подготовки к эксперименту по оценке эффективности метода было создано 12 ящиков электронной почты на разных сайтах: gmail.com, rambler.ru, mail.ru и yandex.ru. Для обеспечения потока почты на эти ящики они были использованы при регистрациях на различных сайтах, форумах за несколько недель до начала эксперимента. При этом на них была организована пересылка почты с почтового ящика, который используется для деловой переписки. Это сделано с целью обеспечения потока нормальных сообщений, что необходимо для оценки такого важного критерия эффективности метода фильтрации, как количество ложных срабатываний.

В среднем за 1 день на каждый ящик приходило около 25 писем (300 всего за 1 день), из них не спам – 5 (всего 60 сообщений за 1 день, 20% от общего числа), соответственно спам – 20 (240 всего за 1 день, 80% от общего числа). За все время эксперимента (32 дня) пришло 9 600 сообщений, из них не спам – 1920, соответственно спам – 7680. Изначально «белый» список содержал 2 адреса. «Черный» список был пуст. Время задержки сообщения в «карантине» 5 дней.

Программа, осуществляющая фильтрацию потока электронной почты, запускалась три раза в день: утром, днем и вечером. Перед запуском программы проверка содержимого ящиков производилась вручную, что позволило контролировать работу метода. За время тестирования метода за спам было принято 7 легитимных сообщений, что соответствует уровню ложных срабатываний – 0,07%. Так как общепринятого стандарта не существует, полученное значение можно считать допустимым при норме 0,01%. Допущенные ложные срабатывания не являлись критичными. Критичных ложных срабатываний не было допущено ни одного. Уровень фильтрации – 100%, это означает, что все 7680 спам-сообщения, пришедшие за время эксперимента, были отфильтрованы.

Выводы

В работе были рассмотрены существующие методы фильтрации потока электронной почты. Представлены методы и рекомендации по их тестированию.

В ходе работы был реализован и протестирован собственный метод фильтрации. Разработанный метод показал следующие результаты: полнота — 100%, весь спам был отфильтрован; точность, уровень некритичных ложных срабатываний – 0,07% (7 из 9 600), некритичных – 0%.

Таким образом, можно говорить, что разработанный метод может достаточно эффективно использоваться для фильтрации потока электронной почты.

Литература

1. Сергеев Д. «Черные» и «белые» списки как мера защиты от спама. URL: http://spam.knowledgebase.ru (дата обращения: 24.02.2010).
2. RFC 2076 (rfc2076) – Common Internet Message Headers. URL: http://www.faqs.org/rfcs/rfc2076.html (дата обращения: 20.02.2010).
3. Русских В. А сегодня вот – что почтальон и почта // Наука и жизнь. 1998. N. 3. URL: http://www.nkj.ru/archive/articles/10397/ (дата обращения: 20.02.2010).