Пресс-релиз №101202

НОВАЯ ТЕХНОЛОГИЯ БОРЬБЫ СО СПАМОМ

В октябре 2002 года компания «Инфосистемы Джет» объявила о скором выходе следующей версии системы мониторинга и архивирования электронной почты "Дозор-Джет". В данную версию включен модуль категоризации сообщений, разработанный специалистами компании на основе алгоритмов фильтрации спама, предложенных Полом Грэмом (Paul Graham). Новый модуль позволит на основе предварительно отобранной администратором базы образцов писем автоматически категоризировать сообщения и применять к ним соответствующие действия.

Одной из наиболее серьезных проблем Интернет является рассылка по электронной почте спама. Как правило, это сообщения рекламного характера, содержащие навязчивые предложения самых разнообразных услуг, товаров и т.п., от порносайтов и виагры до быстрых способов заработать деньги. В принципе спам безвреден, если не считать, что такого рода почта является "группой риска" с точки зрения переноса вирусов. Однако большое количество ненужной почты загружает каналы, "замусоривает" почтовые ящики, отнимает время на удаление ненужных писем и повышает вероятность случайного удаления важной информации. Конечно, рассылка подобных сообщений напрямую не преследует цели "засорить" почтовую систему, однако косвенно приводит к крайне негативным последствиям. Использование списков рассылки, в которую могут входить все пользователи одной корпоративной сети, и получение одновременно всеми этими пользователями сообщений рекламного характера грозит компании снижением производительности ее сетевых ресурсов.

Рассмотрим современные методы борьбы со спамом.

Во-первых, спам можно выявить по наличию в письме определенных признаков. К таким признакам относятся:

- наличие в письме определенных слов (или словосочетаний);
- характерные написания тем письма (например, все заглавные буквы и большое количество восклицательных знаков);
- специфическая адресная информация и т.п.

Во-вторых, поскольку спам по своей природе является массовой рассылкой, его можно определять по идентичному содержанию большой серии писем, направленных в разные адреса и по малым промежуткам времени между отправкой таких писем (они, как правило, посылаются роботами). Такой метод определения спама применим только в условиях, когда имеется доступ к большой выборке писем (например, у провайдера).

В-третьих, косвенно со спамом можно бороться с помощью так называемых "черных списков" почтовых серверов. В эти списки заносятся те серверы, которые замечены в массовых рассылках спама и идея состоит в том, чтобы вообще не принимать и не транслировать почту, исходящую с этих серверов.

Наиболее распространенными являются антиспамные фильтры, реализующие первый метод. В самом деле, далеко не всегда имеется возможность эффективно анализировать большие потоки писем с целью выявить большие серии повторяющихся сообщений. Автоматический же отказ от приема почты с определенных серверов может привести к невозможности получить действительно важную информацию. В то же время очевидно, что именно содержание письма является единственным критерием, по которому его можно отнести к спаму.

Однако, традиционные фильтры, на поиске в компонентах письма определенных признаков, обладают довольно серьезными недостатками.

Начнем с того, что решать, является ли конкретное письмо спамом, в конечном итоге должен тот, кому оно отправлено. В самом деле, для социолога, занимающегося исследованием механизмов распространения финансовых пирамид, массовые рассылки, предлагающие заработать миллион долларов в неделю, представляют профессиональный интерес, в то время как для других людей это злостный спам. Поэтому конкретный набор признаков (ключевых слов и т.п.), по которым отбраковываются письма, должен быть индивидуальным. Однако настройка таких фильтров требует времени и незаурядной изобретательности. Плохой фильтр либо пропускает много спама, либо отбраковывает много нужных писем. Типичные результаты тестирования даже хорошо настроенного фильтра таковы: обнаружено 79,7% спама и имеется 1,2% ложных срабатываний, то есть к спаму были отнесены обычные письма.

Плохое разделение спама и обычных писем обусловлено в том числе и некоторой "однобокостью" стандартных фильтров. При отбраковке писем учитываются "плохие" признаки и не учитываются "хорошие", характерные для полезной переписки.

Этих недостатков лишен метод построения антиспамных фильтров, предложенный американским программистом и предпринимателем Полом Грэмом, одним из разработчиков электронного магазина Viaweb Store, известного в настоящее время как Yahoo! Store. Метод Грэма позволяет автоматически настроить фильтры согласно особенностям индивидуальной переписки, а при обработке учитывает признаки как "плохих", так и "хороших" фильтров.

Метод основывается на теории вероятностей и использует для фильтрации спама статистический алгоритм Байеса. По имеющимся оценкам, этот метод борьбы со спамом является весьма эффективным. Так, в процессе испытания через фильтр были пропущены 8000 писем, половина из которых являлась спамом. В результате система не смогла распознать лишь 0,5% спам-сообщений, а количество ошибочных срабатываний фильтра оказалось нулевым.

В фильтре, основанном на статистической технологии фильтрации, каждому встречающемуся в электронной переписке слову или тэгу присваивается два значения: вероятность его наличия в спаме и вероятность его присутствия в письмах, разрешенных для прохождения. Баланс этих двух значений и определяет вероятность того, что письмо, в котором встречаются данные слова и теги, является спамом.

Отличия статистической технологии фильтрации от технологии фильтрации на основе признаков:

1. Особенностью статистической технологии является возможность индивидуальной автоматической настройки фильтра, что является важным преимуществом, поскольку разные люди или же компании (если фильтр устанавливается на корпоративном почтовом сервере) используют в электронной переписке разную лексику. Настройка фильтра производится по результатам статистического анализа имеющегося архива электронной почты или выборки, полученной за определенный период времени. Такой анализ дает возможность накопить достаточно информации для эффективной фильтрации электронной почты.

2. И в том и другом случае результатом оценки является так называемый «вес» письма. Однако при использовании первого метода "вес" письма вычисляется только на основе "плохих" признаков, что приводит к "обвинительному уклону" фильтра – появляются ложные срабатывания.

3. В алгоритме Байеса наборы признаков определяются не субъективно, а в результате статистического анализа реальных подборок писем. Получающиеся наборы признаков оказываются весьма нетривиальными и эффективными. Например, в качестве "плохого" признака может появиться строка "0Xffffff" – ярко красный цвет; а в качестве "хорошего" признака – Ваш номер телефона. И действительно, письмо, содержащее Ваши персональные данные, в любом случае следует прочесть.

На основе статистической технологии фильтрации специалистами нашей компании был разработан модуль категоризации электронных писем. Новый модуль может использоваться и как компонент системы мониторинга и архивирования электронной почты «Дозор-Джет», и отдельно, как самостоятельный фильтр электронной почты. Он предназначен для отфильтровывания электронных писем определенной категории. Письма автоматически относятся к той или иной категории на основании ранее выполненного анализа выбранной администратором базы образцов писем.

В отличие от других фильтров, использующих статистическую технологию, данный модуль может применяться не только для борьбы со спамом, но и для фильтрации любых других категорий писем в зависимости от желания пользователя. Кроме того, как было отмечено выше, особенностью данного модуля является возможность индивидуальной настройки фильтра под условия заказчика.

Другое серьезное преимущество нашего модуля заключается в том, что если он применяется в качестве компонента системы «Дозор-Джет», то можно воспользоваться статистикой архива, входящего в состав системы, а это позволяет автоматически анализировать почтовый поток и периодически корректировать работу уже созданного фильтра. Этот факт позволяет назвать данную систему самообучающейся. Благодаря этому свойству системы практически исключены ошибочные срабатывания фильтра и, следовательно, потери важной информации. Кроме того, автоматическая самокорректировка значительно облегчает задачу администратора системы по ее контролю и настройке и сокращает время на ее обслуживание.