

Kaspersky® Anti-Spam 3.0

Как это работает

- Спамоприемники
- Лингвистическая лаборатория
- Обновления антиспам-баз
- Сервер антиспам-фильтрации

Фильтрация спама – это не программа, а технологический процесс, человекомашинная цепочка. Она состоит из следующих звеньев:

- 1** Спамоприемники
- 2** Лингвистическая лаборатория
- 3** Регулярные обновления
- 4** Сервер фильтрации

1

Как работают спамоприемники

Спам-лаборатория получает спам из трех основных источников.

- Ловушки. Наши ловушки, расставленные по всему Интернету, включают зарегистрированные на множестве порносайтов и в базах спамеров адреса, ящики бесплатных почтовых систем и пр. Чем дольше существует ловушка, тем больше ее улов.
- «Добровольцы». Письма от добровольных поставщиков спама. В частности, жалобы на спам, поступающие от пользователей бесплатных почтовых веб-сервисов.
- Обратная связь. Примеры нераспознанного спама от клиентов «Лаборатории Касперского» и бета-тестеров.

Интересно отметить, что зачастую разные источники – это разный жанр спама. Это касается как географии (страны) источников, так и расположения их на почтовых серверах разного характера. Например, спам, приходящий на бесплатные почтовые сервера наподобие Hotmail (куда сыплется масса предложений купить университетский диплом), значительно отличается от спама на корпоративных адресах типа info@company.com, куда приходит большое количество приглашений на семинары и конференции.

Наши спамоприемники перенаправляют свои потоки на единый адрес лингвистической лаборатории.

В настоящий момент мы получаем несколько тысяч разных спамерских писем в день (в числе нескольких сотен тысяч неуникальных, повторяющихся сообщений) и постоянно расширяем набор своих источников. На ближайшее будущее, с дальнейшим развитием сети спамоприемников, можно прогнозировать получение до 10 000 уникальных рассылок в день.

2 Как работает лингвистическая лаборатория

Лингвистическая лаборатория сейчас включает 12 человек (коллектив постоянно расширяется за счет специалистов по новым языкам), которые непрерывно занимаются анализом спама. Это специалисты с высшим лингвистическим образованием, с опытом работы в области прикладной лингвистики и искусственного интеллекта. Для лингвистического анализа и обработки писем используется специальное ПО.

Процесс работы лингвиста

В лингвистическую лабораторию поступает спам из расставленных в Интернете ловушек для спама. Лингвисты проверяют всю эту массу модулем фильтрации и отделяют новые письма, то есть не распознанные по текущей базе. Затем они классифицируют письма, пропущенные фильтром: отделяют «нормальные» письма (такие в потоке присланного спама иногда попадают), затем раскладывают спам по рубрикам.

Первый этап работы: в базу сигнатур сразу же добавляются сигнатуры (образцы) всех нераспознанных писем. Это помогает распознаванию «повторных» писем или писем, присланных второй раз, но несколько модифицированных.

В продукте Kaspersky Anti-Spam 3.0, установленном у клиента, реализована технология UDS – Urgent Detection System. Она позволяет получать информацию о самых последних рассылках в режиме реального времени. Сразу же после обнаружения новой рассылки на UDS-сервере появляется ее сигнатура, которая отличается от стандартной сигнатуры форматом и меньшим объемом (подробнее о технологии UDS см. раздел «Как работает сервер фильтрации спама»).

Затем лингвисты начинают тонкий анализ: выделяют в письмах новые термины, назначают им веса и добавляют их в семантические образы. Это подготовка данных для следующего этапа – работы эвристического анализатора.

ПО лингвиста позволяет делать все это очень быстро и эффективно: выделил мышкой, перетаскил в рубрику, кликнул по стрелочке – назначил вес. Встроенные средства контроля позволяют сразу проверить качество распознавания:

- а) на новых письмах (улучшилось ли),
- б) на эталонной базе спамерских писем (не ухудшилось ли),
- б) на эталонной базе обычных писем (чтобы не было ложных срабатываний).

Помимо этого, спам-аналитики параллельно занимаются анализом «конверта» письма, то есть его формальных признаков (отправитель, получатель, путь следования и т.п.) и созданием новых правил для распознавания по этим признакам.

Лингвистическая лаборатория работает круглосуточно (24x7x365), обновления антиспам-баз выкладываются строго каждые 20 минут, а UDS-сигнатуры выкладываются в режиме реального времени.

3 Как происходит обновление антиспам-баз

Каждые 20 минут обновленная база фильтрации – семантические образы, образцы писем и новые формальные правила – выкладывается на сервер обновлений.

Сервер фильтрации (Kaspersky Anti-Spam 3.0) скачивает эти обновления и начинает распознавать самые свежие спамерские письма – как за счет новых семантических правил, так и по внесенным новым образцам.

Нужно заметить, что регулярное обновление баз чрезвычайно важно по трем причинам.

1. Большая подвижность лексикона спамеров. Хотя цели спамеров практически неизменны – они хотят что-то продать пользователю, заманить его на сайт или заставить ответить по электронной почте, – используемые ими выражения постоянно меняются.

По нашему опыту, качество распознавания спама при использовании старой, «замороженной» базы может снижаться на несколько процентов в неделю, падая от 90-95% до 40-60% в пределе (этот предел определяется стандартным спамерским лексиконом и типичными признаками рекламных писем).

2. Повторяемость писем. Повторяемость спамерских писем довольно велика, она может достигать до 10-15% за месяц. Иногда приходит пять-шесть копий одного и того же письма в неделю, а порой и в день. Таким образом, большая часть повторных писем может быть отфильтрована за счет «свежести» базы.

3. Конечная скорость распространения. Чтобы разослать 10 миллионов писем, нужно затратить определенное время. Сама по себе работа почтового сервера, рассылающего миллионы писем, может занять много часов. Кроме того, электронная почта – это средство с негарантированным временем доставки. Это означает, что последние спамерские письма из большой партии могут доходить до получателя через несколько часов после начала рассылки. А следовательно, при регулярном обновлении большая часть клиентов успеет получить новую версию базы с сигнатурой нового письма раньше прихода к ним этого же письма.

4 Как работает сервер фильтрации спама

Фильтр спама – это серверная программа, которая устанавливается на входе в сеть и фильтрует входящий поток почты. Kaspersky Anti-Spam 3.0 предназначен для использования как в небольшой компании, имеющей собственный почтовый сервер, так и в крупных организациях с почтовыми потоками, достигающими сотен гигабайт в день.

Сервер фильтрации предназначен для распознавания и фильтрации нежелательных почтовых сообщений (спама) в процессе приема электронной почты по протоколу SMTP, т. е. до того, как сообщения будут доставлены в почтовый ящик конечного получателя.

Основные достоинства Kaspersky Anti-Spam 3.0:

- **эвристические лингвистические методы** анализа содержания почтовых сообщений (контентная фильтрация);
- **обновление антиспам-базы каждые 20 минут**;
- **реакция на новый вид спама в режиме реального времени** – технология UDS;
- **объединение всех методов фильтрации** (по формальным признакам и по содержанию) в едином модуле; их комплексное использование;
- **централизованное управление** всеми параметрами фильтрации через единый интерфейс;
- **масштабируемость нагрузки**: от уровня небольших предприятий до крупнейших почтовых систем.

Kaspersky Anti-Spam 3.0 является полнофункциональным серверным почтовым продуктом, работающим на *nix-платформах (Linux и FreeBSD).

Используемые методы фильтрации спама

Сервер фильтрации объединяет известные формальные методы с методами контентной фильтрации, осуществляющими распознавание сообщений по их содержанию на основе эвристического поиска ключевых терминов и нечеткого сравнения с письмами-образцами.

Таким образом, сервер фильтрации использует следующие методы фильтрации:

1. **Списки.** Проверяется наличие почтового адреса и IP-адреса отправителя в черных списках, которые ведут провайдеры и различные общественные организации (так называемые RBL – Real-time Black Lists). Администратор системы может также вести свои белые списки («списки друзей»), от которых почта принимается всегда, минуя этапы анализа.
2. **SPF и SURBL.** В процессе фильтрации может учитываться авторизация отправителя по технологии SPF (Sender Policy Framework). В дополнение к DNSBL, блокирующим спамерские IP-адреса, используется технология SURBL (Spam URI Realtime Block List), выявляющая спамерские URL в теле сообщения.
3. **Формальные признаки письма.** Отсутствие адреса отправителя, отсутствие или слишком много получателей, отсутствие IP-адреса в системе интернет-адресов DNS и т.п. – все это является признаками спамерского сообщения. Также осуществляется фильтрация по размеру, формату сообщения.
4. **Содержание письма (лингвистические эвристики).** Проверяется наличие в письме признаков спамерского содержания: определенного набора и распределения по письму специфических слово-сочетаний. Причем сервер фильтрации анализирует не только текст самого письма, но и вложения.
5. **Сигнатуры (образцы).** По каждому спамерскому письму может быть автоматически создана так называемая лексическая сигнатура, позволяющая распознать это письмо даже с небольшими модификациями. Такие сигнатуры добавляются в базы лингвистической лабораторией.
6. **Графические сигнатуры.** Процесс создания и добавления в антиспам-базы графических сигнатур схож с процессом создания обычных сигнатур спама, но в данном случае работа ведется с изображениями, которые используются спамерами как в теле письма, так и в виде вложений.
7. **UDS-запросы в режиме реального времени.** Если некое письмо не получило однозначной оценки, выполняется запрос к UDS-серверу. Он содержит данные о самых последних рассылках, информация о новом спаме добавляется в тот же момент, когда он обнаружен спам-аналитиком. На основании полученной информации письму присваивается окончательный статус (спам / не-спам).

Существенной особенностью фильтра является возможность распознавания нежелательных сообщений путем анализа их содержания. Фильтр осуществляет автоматическую рубрикацию сообщений, то есть отнесение входящих сообщений к одной или нескольким категориям на основе смыслового анализа их текста.

В результате работы сервера фильтрации конкретное сообщение должно быть отнесено к одной из следующих категорий: спам, нецензурная лексика, «возможно спам», не-спам.

Качество работы и ложные срабатывания

Главным показателем качества работы фильтра является не распознавание максимального количества спамерских писем, а отсутствие ложных срабатываний, то есть писем, ошибочно занесенных в категорию спамерских.

Письмо может быть отнесено к той или иной категории с очень высокой, но не 100%-ной вероятностью. Поэтому системному администратору компании, осуществляющему настройку продукта, рекомендуется никогда не уничтожать входящую почту, отфильтрованную на основе контентного анализа. Такая почта должна архивироваться (например, путем перенаправления ее на специальный адрес) и храниться в течение определенного срока.

В настоящее время сервер фильтрации позволяет отсеивать до 95-98% спамерских писем, при уровне ложных обнаружений в 0,001% (1 письмо на 100 000). Увеличение процента распознавания спамерских писем относительно указанного нежелательно именно в связи с недопустимостью ложных обнаружений.

Нужно сказать, что ложные срабатывания обычно связаны не с деловыми письмами, а с пресс-релизами и рассылками с преобладанием рекламной лексики.

Значительно снизить риск ложных срабатываний позволяет так называемый белый список, или «список друзей», в который администратор продукта может добавить всю адресную книгу компании, в том числе всех сотрудников, деловых партнеров, прессу и проч.

База фильтрации

Для анализа почтовых сообщений по содержанию сервер фильтрации использует специализированные лингвистические данные (базу контентной фильтрации), которые каждые 20 минут автоматически обновляются через Интернет. База составляется лингвистической лабораторией.

База фильтрации содержит данные трех типов:

- рубрикатор (иерархический список категорий спама);
- семантические образы категорий;
- сигнатуры сообщений-образцов.

Рубрикатор спама включает **около 500 рубрик**, соответствующих различным категориям спама, в том числе «Зайди на сайт», «Для взрослых», «Купи виагру», «Купи софт», «Увеличь то или это», «Горящие путевки», «Посетите семинар», «Обучение английскому», «Заработок в Интернете», «Обеспечь себе финансовую независимость», «Снизь налоги» и т. п.

Каждая рубрика содержит свой семантический образ – набор терминов (словосочетаний) с заданным определенным весом. В базе есть три вида словосочетаний:

- а) **Точные термины**, имеющие вероятность 100% («это не спам», «ваш адрес получен из открытых источников», «чтобы больше не получать этого письма»).
- б) **Вероятностные термины** – словосочетания, которые с определенной вероятностью указывают на то, что письмо является спамом, например, посетите наш сайт, Nigeria, unsubscribe.
- в) **Дополнительные термины**, которые сами по себе не указывают на спам, так как могут встречаться в обычных письмах, но добавляют вес остальным терминам (Ericsson, туры в Турцию, Анталья, Шамои, софт).

В настоящее время база фильтрации содержит **около 50 000 терминов**.

Работа с разными языками

Для работы с письмами на разных языках сервер фильтрации использует встроенные модули лингвистической поддержки. В настоящее время технология «Спамтест» обеспечивает лингвистическую поддержку для русского, английского, немецкого, французского и испанского языков.

Для других европейских языков алгоритмы распознавания тоже будут работать, но с чуть меньшей точностью. В частности, метод сигнатур и термины, добавленные пользователем, могут работать с любыми европейскими языками.

Бизнес-логика фильтрации

После того как письмо признано спамерским, возникает необходимость что-либо с ним делать: удалить, уведомить отправителя, что письмо не принято сервером, добавить специальную метку в тему письма, заархивировать в указанных папках, отослать уведомление администратору или пользователю и т. п. Настройка бизнес-логики – это и есть определение дальнейших действий с письмом.

Настройка бизнес-логики фильтрации производится администратором почтового сервера. Наряду с общими правилами фильтрации, которые применяются для всех писем, проходящих через Фильтр, существуют индивидуальные правила – их можно задать для конкретного получателя (группы получателей).

К письмам, в которых обнаружены признаки спама, могут применяться, в частности, следующие схемы бизнес-логики:

- Непринятие почты («reject»). Спамерская почта данной категории не принимается почтовым сервером, как если бы такого адреса вообще не существовало (отправитель получает об этом уведомление – то есть происходит обман спамера);
- Удаление. Спамерская почта данной категории уничтожается (отправитель не получает никаких уведомлений);
- Архивирование. Спамерская почта данной категории перенаправляется на некоторый архивный адрес и не доставляется адресату (отправитель при этом может либо получать соответствующее уведомление, либо не получать никаких уведомлений);
- Пересылка с разметкой. Спамерская почта данной категории пересылается адресату, при этом каждому сообщению приписывается метка в теме (например, [!SPAM]) на основании которого производится сортировка почты на уровне клиентской почтовой программы (например, правилами MS Outlook).

Перечисленные схемы обработки письма — наиболее типичные, но далеко не единственно возможные. Администратор, используя возможности настройки продукта, может задавать практически любые схемы и варианты бизнес-логики.

Форматы вложений

Сервер фильтрации работает не только с текстом письма, но и с вложениями. Обрабатываются вложения в почтовые сообщения в следующих форматах:

- обычный текст (ASCII),
- HTML,
- Microsoft Word (версии 6.0, 95/98/2000/XP),
- RTF,
- GIF, JPEG и PNG.