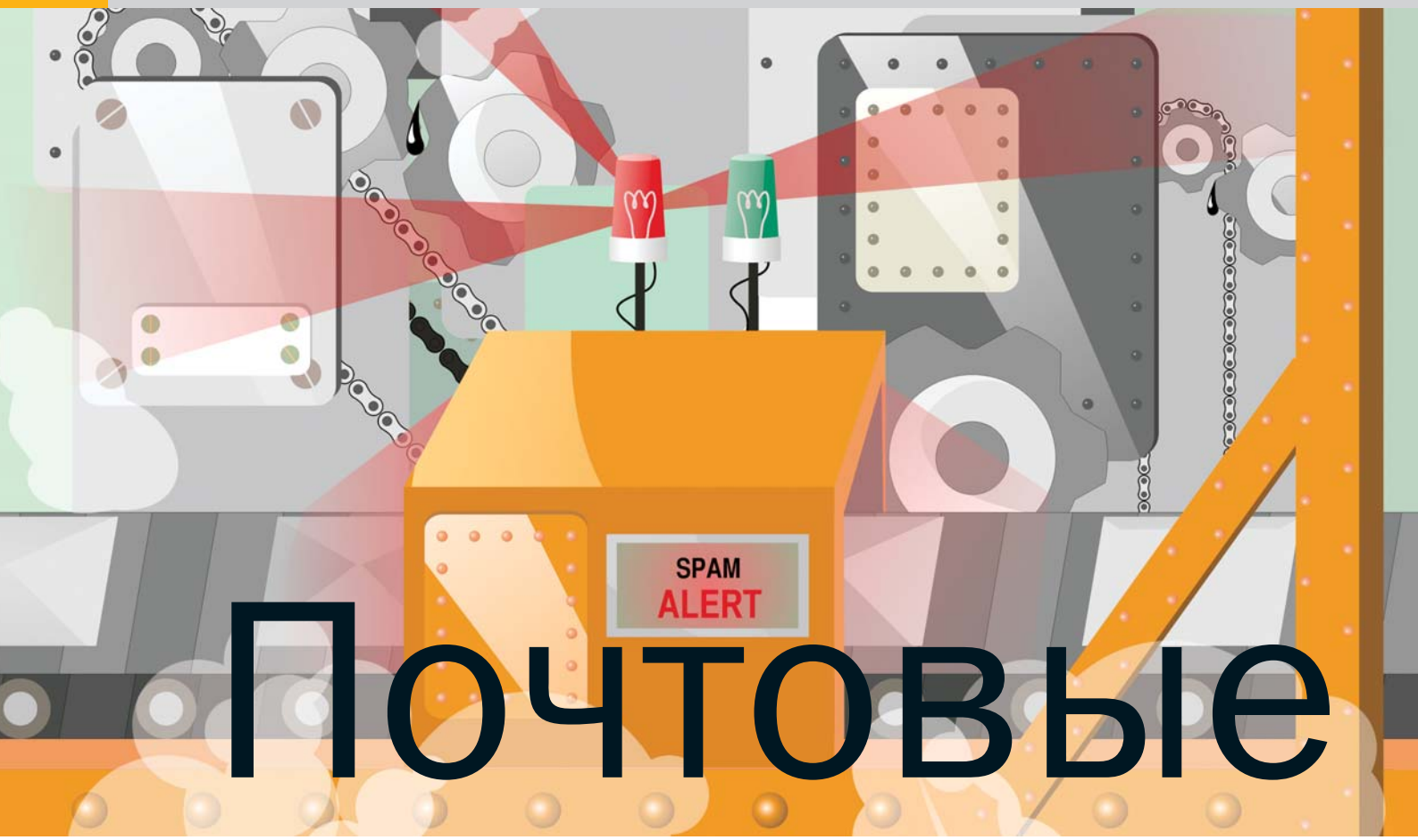


Почтовые санитары

Установка фильтров на клиентской машине

Если на почтовом сервере провайдера нет антиспамового фильтра, то отсеять нежелательную корреспонденцию помогут специальные программы, которые устанавливаются на клиентскую машину. Но и отсеивать они будут уже в почтовой программе.

Каждый человек, активно использующий электронную почту, уже узнал, что такое спам. Число нежелательных писем различно для каждого конкретного адреса, но в большинстве случаев оно приблизительно в три раза превышает количество нужной корреспонденции.

Борьба со спамом ведется уже давно, за это время разработано немало методов и концепций. Участие в этой борьбе с разными результатами принимают провайдеры, крупные производители программного обеспечения, публичные почтовые сервисы и другие организации. Но, несмотря на все усилия, спам продолжает поступать к конечному пользователю. Изничтожить его трудно не только потому, что спамеры постоянно развивают свои методы и придумывают новые способы противодействия. Проблема еще и в том, что иногда человек не сразу может отличить спам от обычного письма. К примеру, реклама туристического агентства может выглядеть как рассказ какого-то знакомого об отдыхе на море. И в конце письма очень естественно смотрится ссылка на сайт фирмы, которая организовала такой замечательный отдых. »

Можно вести себя осторожно — не оставлять свой адрес на общедоступных страницах, а если и оставлять, то в измененном виде. Но ситуации бывают всякие, и рано или поздно ваш адрес может попасть в список спам-рассылки. Что же делать, если на ящик приходят нежелательные письма?

Существует множество программ, которые могут выяснить, является ли определенное письмо спамом, но ни один метод не дает гарантии, что все письма, объявленные спамом, являются таковыми. Равно как и нет гарантии, что все письма, содержащие спам, будут выявлены. Поэтому задача таких программ — помечать подозрительные письма и отделять их от нужных. Тогда в случае если письмо было определено как спам ошибочно, его можно будет все-таки найти. А читать корреспонденцию, в которой уже нет мусора, гораздо удобнее.

Черные списки

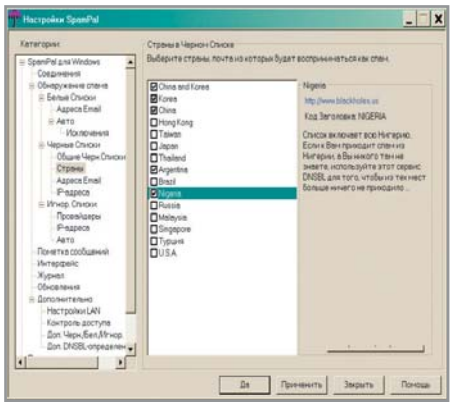
Как же программа может определить, является письмо спамом или нет? Один из довольно простых и эффективных методов — проверка IP-адресов узлов, через которые прошло письмо. Все эти адреса содержатся в заголовке письма. В Интернете существует множество служб, содержащих списки IP-адресов узлов — источников спама. Сюда попадают открытые релеи (серверы, разрешающие отправку писем всем, даже анонимным пользователям), открытые прокси, зомбированные сети и другие потенциальные источники спама. Конечно же, сюда заносятся и адреса, для которых уже выявлен факт массовой рассылки. Списки пополняются не только по жалобам от людей, но и по результатам работы роботов, которые проверяют различные узлы. Поэтому периодически в таких списках оказываются «нормальные» серверы, которые по каким-либо причинам не совсем корректно настроены.

Анализ содержимого

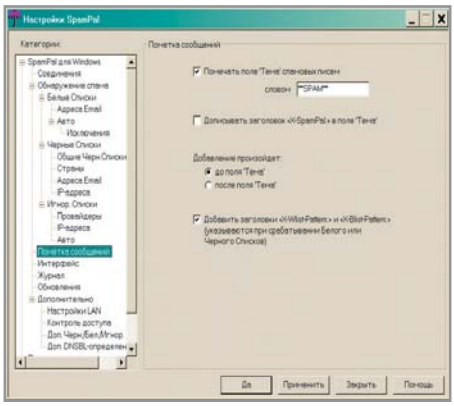
Принципиально другой метод определения спама — просмотр содержимого письма. Здесь есть несколько подходов.

Первый — это метод Байеса. Фильтр использует специально подготовленную базу слов, в которой вероятность того, что конкретное слово содержится в спам-письме, выражена числом. На основе слов, имеющих ярко выраженную принадлежность к спаму или, наоборот, к обычным письмам, вычисляется общая характеристика письма. Получившееся число отражает вероятность того, что проверенное письмо — спам. Далее, в зависимости от этой вероятности можно установить нужный порог, при котором письмо будет отфильтровано. При этом такой фильтр можно обучать, указывая ему на ошибки. Таким образом он будет приспосабливаться к конкретному пользователю и к его типу корреспонденции.

Второй подход — это анализ «оформления» письма. Спам-письма часто обладают характерными особенностями. Например, если это письмо в формате HTML, то в нем используются шрифты разных размеров и цветов. Или же текст письма представлен в виде графического файла. Есть также множество других деталей, не присущих «нормальным» письмам, которые можно обнаружить в заголовках спама.



▲ SpamPal: отметьте страны, письма откуда будут считаться спамом



▲ SpamPal: определите, как будут помечаться нежелательные письма

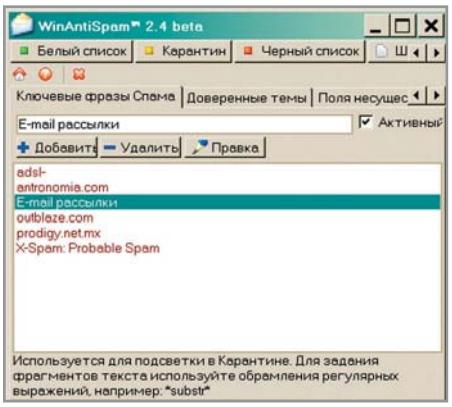
Мы отобрали несколько программ, предназначенных для фильтрации спама на компьютере конечного пользователя. Эти программы используют как вышеописанные способы обнаружения нежелательной корреспонденции, так и некоторые другие.

Большинство таких программ требует изменения настроек почтового клиента. В результате этих изменений почтовый клиент соединяется не с самим почтовым сервером, а уже с программой-фильтром на этом же компьютере, после чего фильтр проводит почтовую сессию с почтовым сервером и возвращает результат клиенту. Для этого в настройках почтовой программы адрес POP3-сервера заменяется на 127.0.0.1 (или localhost, что, по сути, одно и то же), а после имени пользователя через специальный символ-разделитель (как правило, это @) добавляется имя (адрес) настоящего сервера. Иногда требуется изменить в большую сторону и время ожидания данных от сервера, так как процесс проверки письма антиспамовым фильтром замедляет доставку почты.

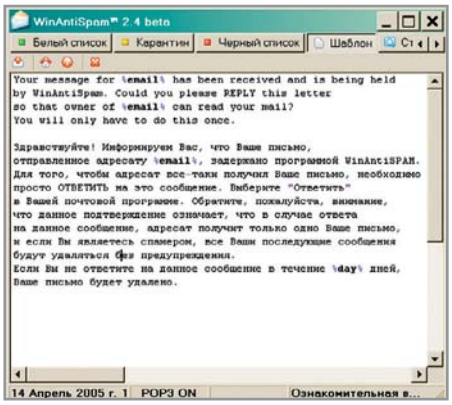
Плагины для почтовых программ

Защита от спама для программ Microsoft

Как известно, в продуктах Microsoft тоже есть встроенные механизмы борьбы со спамом. И если, например, в почтовой программе Outlook XP это еще простенький статический фильтр, находящий выражения вроде «for free!» и «SPECIAL PROMOTION», то в версии 2003 применяется уже более продвинутая технология — Microsoft SmartScreen Technology. По сути, это фильтр, подобный байесовскому, с применением анализа частоты слов. Отличие от других программ заключается в том, что обучением фильтра занимаетесь не вы, а специальное подразделение Microsoft — Microsoft Anti-spam Technology & Strategy Group. Сюда поступают миллионы писем, и на их основе формируется база данных «спамных» и «неспамных» слов и выражений. База регулярно обновляется и доступна для скачивания через Интернет.



▲ WinAntiSpam: введите слова, характерные для рекламных писем



▲ WinAntiSpam: вот такое письмо-запрос нужно подтвердить

» SpamPal

SpamPal — бесплатная программа, которая делает то, что, по идее, должно происходить не на компьютере пользователя, а на почтовом сервере: находит в заголовке письма IP-адреса узлов, участвовавших в его доставке, и запрашивает различные службы черных списков (DNSBL). Если хотя бы одна служба черного списка вернула положительный ответ, то письмо помечается как спам и передается в почтовую программу. Помечается письмо строкой «X-SpamPal» в служебном заголовке письма с указанием службы черного списка, которая выдала положительный ответ. Кроме того, изменяется и тема письма — туда добавляется строка «*** SPAM ***». Для более удобной работы со спамом в почтовой программе следует настроить фильтр, отправляющий все письма с такой темой в особую папку. Ответы от DNSBL кешируются, чтобы уменьшить количество отправляемых запросов.

Основа работы программы — список DNSBL — обновляется автоматически. Из списка всех доступных DNSBL можно выбрать, какие именно стоит использовать. Есть предустановленные наборы активных DNSBL (называемые стратегиями), различающиеся «агрессивностью» фильтрации. Для каждого DNSBL из списка доступно его краткое описание. Кроме того, в черный список можно добавить принадлежность открытого релея определенной стране, почтовый или IP-адрес.

В белый список (whitelist) добавляются почтовые адреса, с которых заведомо не придет спам, даже если информация из черных списков утверждает обратное. Есть режим автоматического занесения адресов в белый список, если с этих адресов в течение определенного времени приходит не спам. Для исключения ложного срабатывания используется также игнорирующий список. Его содержимое (одиночные IP-адреса и диапазоны IP-адресов) тоже пропускается фильтром, то есть письма не считаются спамом, даже если DNSBL вернет положительный ответ.

В SpamPal предусмотрена установка плагинов. Есть плагин, реализующий фильтрацию по алгоритму Байеса, плагин для проверки URL, находящихся в теле письма, а также плагин для фильтрации с использованием регулярных выражений (regexp).

Надо заметить, что даже при установке стратегии средней агрессивности фильтрации в спам попали письма, написанные с публичных почтовых серверов, рассылки Subscribe.ru и даже письма, отправленные через провайдера «МТУ-Интел». Поэтому нужно устанавливать самую щадящую стратегию или разбираться, какой из DNSBL блокирует нужное, и вносить изменения в их список.

Сейчас ведется разработка SpamPal 2. В этой версии будет добавлена возможность работы программы в виде службы Windows, параллельная обработка пи- »

Плагины для почтовых программ

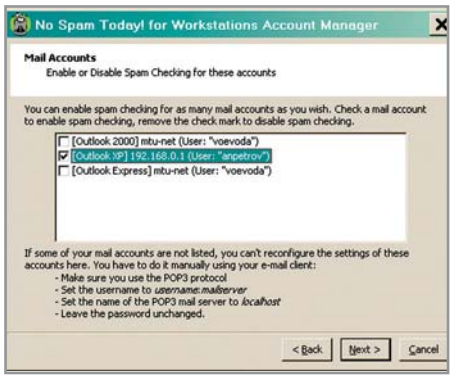
Байесовский фильтр, встроенный в The Bat!

Если вы используете для работы с почтой программу The Bat!, то в этом случае можно воспользоваться плагином для фильтрации спама. Плагин занимается лишь оценкой писем, а уже в настройках The Bat! вы выбираете, удалять эти письма или помещать их в папку «Спам». Одновременно можно использовать несколько плагинов, и тогда The Bat! будет принимать решение о действии на основании максимальной, минимальной или средней оценки, выставленной всеми плагинами. На официальном сайте программы (www.ritlabs.com) указаны два таких плагина: Bayes Filter Plugin и BayesIt! BayesIt! уже входит в дистрибутив The Bat!. Для того чтобы включить его, нужно добавить файл bayesit.tbp (он располагается в каталоге Bayesit) в список плагинов в разделе настроек «Свойства → Настройка → Предупреждения → Защита от спама». Там же, в разделе «Защита от спама», когда фильтр будет достаточно натренирован, нужно включить удаление или перемещение в особую папку для писем, являющихся спамом. Обучается фильтр Байеса, когда вы помечаете письмо как спам или как не спам. Помимо улучшенного байесовского фильтра, BayesIt! содержит белый, черный и игнорирующий списки для выражений в теле письма или любого из полей заголовка. Кроме того, используя функцию The Bat! «Выборочное скачивание», можно добиться удаления некоторого количества спама прямо на сервере до проверки спам-фильтром. Для этого BayesIt! умеет экспортировать часто встречающиеся в спаме слова в текстовый файл с регулярными выражениями. Этот файл нужно указать как источник сигнальных строк в разделе «Выборочное скачивание» сортировщика писем.
Второй плагин — Bayes Filter Plugin — как следует из его названия, тоже работает по алгоритму Байеса. Но, в отличие от первого плагина, он еще делает запросы к DNSBL (черным спискам IP-адресов узлов, через которые проходило письмо), причем письмо определится как спам, только если несколько DNSBL дадут положительный ответ. Количество DNSBL вы можете настроить по собственному усмотрению.

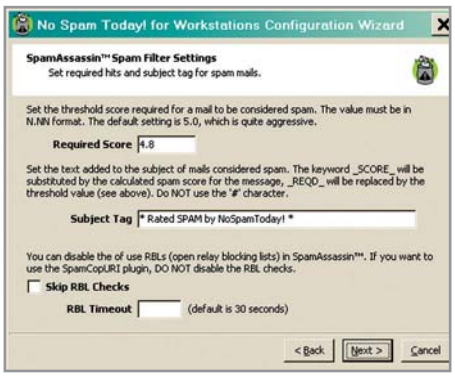
» сем, а самое главное — работа в режиме прозрачного прокси (transparent proxy). То есть уже не понадобится изменять настройки почтового клиента: SpamPal будет сама ловить и обрабатывать соединения почтового клиента с серверами. Описание и бета-версии SpamPal 2 доступны по адресу www.spampal2.org.

WinAntiSpam

Эта программа работает по совершенно другому принципу. Она даже не пытается анализировать, находится спам в почтовом ящике или нет. Когда WinAntiSpam видит письмо с незнакомого адреса, она создает специальное письмо с просьбой к отправителю подтвердить, что он не робот-спамер. Если письмо отправил кто-то, кто подтвердит это в течение определенного времени, то его адрес помещается в белый список, а первое письмо доставляется получателю. Если же в течение обозначенного времени ответа не поступит, то, скорее всего, его отправил спамер, и такой адрес будет занесен в черный список, а все дальнейшие письма с этого адреса будут блокироваться. До получения подтверждения (или же истечения срока) письма находятся в особом хранилище — карантине. Разумеется, далеко не все письма, отправку которых не может подтвердить человек вручную, являются нежелательными. Это могут быть уведомления с форумов, рассылки и другие письма, посылаемые автоматически. Для такого рода корреспонденции придется сразу создавать соответствующие записи в белом списке.



▲ No Spam Today! for Workstations: выбор учетной записи



▲ No Spam Today! for Workstations: укажите число «очков» спам-письма

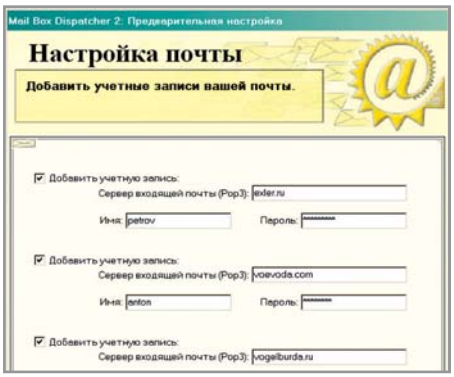
Помимо такого способа блокировки нежелательной корреспонденции можно добавить еще в настройки фильтра темы писем, которые точно не являются спамом. Или же задать строку, присутствие которой в заголовке письма приведет к его удалению без занесения в черный список.

В связи с тем что WinAntiSpam не только принимает письма, но и рассылает запросы, его придется настроить и для работы с SMTP-сервером.

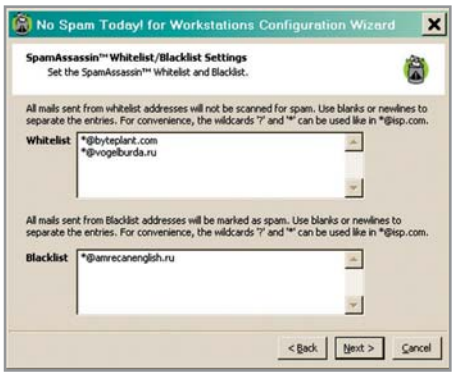
No Spam Today! for Workstations

Эта программа-фильтр работает на основе SpamAssassin — популярного набора скриптов на языке Perl. Вообще SpamAssassin создан как универсальный спам-фильтр, работающий в связке с любыми почтовыми службами в UNIX-системах. Поэтому его установка в Windows-систему чревата ручными сборками, скачиванием дополнительных компиляторов и библиотек и другими «занимательными» вещами. Программа же No Spam Today! for Workstations предоставляет всю мощь SpamAssassin в уже готовом виде, причем требует минимума настроек.

No Spam Today, как и SpamPal, действует между почтовым клиентом и почтовым сервером. Но кроме опроса нескольких DNSBL он еще и проводит комплексный анализ содержимого писем. В результате каждого из этапов этого анализа письмо получает очки, которые в конце суммируются. Вы можете указать в настройках, при каком итоговом значении письмо будет помечено как спам. На этих этапах проверяется содержимое письма байесовским методом, проверяются поля исходящего адреса, темы и URL внутри тела — не содержат ли они слишком много цифр и подозрительных символов. Дополни- »



▲ Mail Box Dispatcher: укажите нужные учетные записи



▲ Mail Box Dispatcher: укажите домены для белого и черного списков

» тельные очки получает письмо, если в нем содержатся буквы, близкие по цвету к фону, большие или яркие шрифты, попытка подделки под письмо, созданное обычным почтовым клиентом. Кроме того, делаются запросы к DNSBL для серверов, ссылки на которые содержатся в теле письма. Для более точного контроля за отправителями действует система AWL (Auto-Whitelist). Для каждого отправителя, который определяется не только адресом почты, но и IP-адресом, хранится усредненное значение веса его писем. При каждом новом письме его вычисленный вес изменяется в зависимости от этого усредненного значения. Так можно избежать попадания в спам письма от известного отправителя, если вдруг то, что он напишет, будет очень похоже на спам.

Несмотря на простоту интерфейса, настроек в программе великое множество. Как и принято в мире UNIX, процесс конфигурирования происходит путем правки текстовых файлов.

Mail Box Dispatcher

Эта программа не требует перенастройки клиента, так как работает независимо от него. Mail Box Dispatcher периодически проверяет почтовый ящик на сервере и удаляет оттуда спам. После этого вы можете читать свою почту почтовым клиентом или через веб-интерфейс. Выяснить, является ли письмо спамом, и при необходимости удалить его Mail Box Dispatcher может не загружая тело письма полностью.

Основной механизм фильтрации в этой программе основан на улучшенном алгоритме Байеса. Здесь также учитываются неестественно длинные слова, бессмысленные сочетания букв и цифр и другие признаки того, что письмо вам вряд ли нужно.

В настройках Mail Box Dispatcher разделены события (вкладки «Сервис → Настройки → Общие события» и пункт меню «Сервис → Редактор событий») и действия (пункт меню «Сервис → Редактор действий»).

Событиями называются различные состояния писем. Например: письмо получило 60 очков от байесовского фильтра, или адрес отправителя совпадает с адресом получателя, или несколько писем имеют одно и то же вложение и т. п.

Для каждого события можно назначить одно действие. Такое действие подразумевает пометку письма на удаление, сохранение его в файл, выделение особым цветом в окне Mail Box Dispatcher или какое-либо другое оповещение пользователя.

После того как Mail Box Dispatcher загрузит заголовки и первые строки писем (количество строк указывается в настройках), обрабатываются события, и вы можете просматривать письма и помечать их для удаления или, наоборот, снимать эту пометку, ошибочно поставленную фильтром. Тут же можно добавить адрес отправителя в белый или черный список. Каждый раз, когда вы устанавливаете или снимаете пометку для удаления, фильтр делает выводы, самообучается. Далее вы нажимаете кнопку «Process», и программа очищает ящик от нежелательных писем.

После обучения фильтра и настройки белого и черного списков можно поручить программе выполнять все эти действия самостоятельно, включив «Сервис → Настройки → Программа → Настройки автопроверки».

Выводы

Натренированный фильтр способен распознавать спам с вероятностью, очень близкой к 100%, но для этого необходимо скачать часть письма. Причем чем больше будет часть письма, тем выше вероятность верного распознавания. Так что, если вы хотите сэкономить трафик при помощи спам-фильтра, то это не самый действенный способ. Кроме того, удаляя письма, не скачивая их на компьютер, при любом методе фильтрации вы все-таки рискуете однажды потерять нужное письмо, а может, даже и не одно. Поэтому истребить спам таким образом у вас не выйдет. Зато вы сможете отделить спам от нужных писем, сделать так, чтобы он не мешал. А процесс фильтрации должен происходить на SMTP-серверах, которые принимают почту от внешних отправителей.

Дмитрий Солошенко

ПРОГРАММЫ	
SpamPal	
Версия	1.591
Разработчик	James Farmer
Сайт	www.spampal.org
ОС	Windows
Объем дистрибутива	744 кбайт
Условия распространения	freeware
Интерфейс	русскоязычный
WinAntiSpam	
Версия	2.4 beta
Разработчик	WinAntiSpam
Сайт	www.winantispam.ru
ОС	Windows
Объем дистрибутива	1,35 Мбайт
Условия распространения	shareware ($5)
Интерфейс	русскоязычный
No Spam Today! for Workstations	
Версия	2.0.3.1
Разработчик	byteplant GmbH
Сайт	www.byteplant.com
ОС	Windows
Объем дистрибутива	4,63 Мбайт
Условия распространения	shareware ($40)
Интерфейс	англоязычный
Mail Box Dispatcher	
Версия	2.20
Разработчик	Alex Kaul
Сайт	www.maximumsoft.com
ОС	Windows
Объем дистрибутива	912 кбайт
Условия распространения	freeware
Интерфейс	русскоязычный