Лазо Олег Иванович

ФИЛЬТРАЦИЯ НЕЖЕЛАТЕЛЬНОЙ ЭЛЕКТРОННОЙ КОРРЕСПОНДЕНЦИИ

ВВЕДЕНИЕ

Нежелательная электронная корреспонденция (другие названия СПАМ, UCE, Unsolicited Commercial Email) приобрела популярность еще около десяти лет назад, и с тех пор темпы роста лишь увеличиваются. Это объясняется несовершенством законодательства, крайне низкими расходами на рассылку спама и отсутствием 100% эффективных методов борьбы со спамом. Сейчас трудно найти хоть одного человека, использующего электронную почту и не встречавшего среди нее спам. Вред от спама трудно недооценить: это и увеличение загруженности каналов, и возросшие нагрузки на почтовый сервер для обработки писем, и затрата времени на сортировку почты получателями, и пропадание полезных писем, некорректно определенных антиспам-программами или не замеченными среди спама и вместе с ним удаленных.

СПОСОБЫ ФИЛЬТРАЦИИ СПАМА

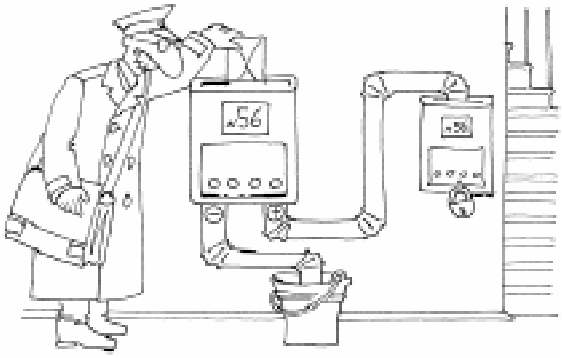
Фильтрация перед приемом почты (во время SMTP сессии)

Примечание. SMTP (Simple Mail Transfer Protocol) – протокол передачи почтовых сообщений, с помощью которого функционирует в Интернет электронная почта. *SMTP сессия* – соединение, устанавливающееся между MTA (Message Transfer Agent, почтовый сервер) и MTA или MUA (Mail User Agent, пользовательская почтовая программа) и MTA для отправления/доставки почтовых сообщений. Во время SMTP сессии принимающей стороне отправляются SMTP команды (см. далее по тексту), идентифицирующие отправляющую сторону (HELO/EHLO), отправителя (MAIL FROM), получателя (RCPT TO), само письмо (DATA) и т.д. На каждую команду принимающая сторона выдает код возврата команды: 25* (команда завершена успешно), 45* или 55* (команда не принята, возможные причины: письмо определено как спам, получателя не существует или его ящик переполнен, неправильная команда SMTP и т. д. В зависимости от причины, вместо последней * подставляется определенное число, чтобы пользователь мог однозначно определить причину, по которой его письмо не было доставлено).

Обычно это проверки, выполняемые непосредственно MTA.

Достоинства: отправитель уведомляется об ошибке доставки, снижается загруженность канала, не требуется дополнительного места на сервере.

Вред от спама трудно недооценить...



Фильтрация перед приемом почты

Недостатки: недоступно для обработки тело письма, при ошибочном срабатывании антиспам-фильтра получатель так и не узнает о том, что ему приходило письмо.

Фильтрация после приема почты (после SMTP сессии).

В этом случае письма благополучно доставляются, а отфильтровываются позже, обычно одной или несколькими отдельными внешними программами.

Достоинства: доступно максимум информации о письме для обработки.

Недостатки: требует дополнительных ресурсов сервера, увеличивается загруженность канала.

МЕТОДЫ ФИЛЬТРАЦИИ СПАМА

1. Перечни адресов для блокировки (блоклисты реального времени, черные списки). RBL (Realti-me Blackhole Lists)

Блоклисты реального времени позволяют с помощью запросов ДНС определить, использовался ли компьютер (хост) отправителя ранее для отправки спама, принадлежит или нет адрес модемному пулу, является или нет открытым почтовым резервным сервером (релеем) и многое другое.

Фильтрация после приема почты

Примечание. Хост (*host*) – так называется любой компьютер, находящийся в Интернет, отправивший письмо (в том случае, если это не ваш релей). *Релей* (*mail relay*) – резервный сервер, который получает и хранит у себя почту в случае недоступности вашего основного сервера.

Наиболее известные публичные блоклисты можно найти по адресам:
http://www.ordb.org/
http://openrbl.org/
http://dnsbl.org/
http://moensted.dk/spam/

2. Фильтры на основе регулярных выражений. Regexp (Regular Expressions)

Примечание. Регулярные выражения – это язык шаблонов, с помощью которого удобно описывать последовательности символов.

Сайт одного из распространенных фильтров, основывающихся на регулярных выражениях, находится по адресу:
http://www.procmail.org/
http://www.postfix.org/ – сайт бесплатного MTA, поддерживающего regexp-фильтры.

3. Ограничение на HTML-файлы. HTML-restrict

Отказ получения писем в формате HTML и автоматическое создание ответа с просьбой прислать письмо в формате plain-text. Реализуется с помощью различных модулей к MTA и MUA.

4. Создание временных почтовых ящиков. Time-stamping

Создание временных (ограниченных по времени жизни или по количеству пришедших писем) почтовых ящиков:
http://www.spamgourmet.com/ – здесь вы можете создать себе один или несколько подобных ящиков.

5. Ловушка для спама (горшочки с медом). Spam traps (honeypot)

«Поддельные» e-mail – «ловушки» для обнаружения спама. Суть этого метода зак-

лючатся в создании специальных, никем не используемых e-mail и рекламирование этих ящиков для спамеров. Например, размещая эти e-mail на html-страницах своего сайта, предварительно сделав их невидимыми для человеческого глаза, используя цвет текста, идентичный с цветом фона. Это помешает человеку узнать об этом e-mail'е, но не помешает специальной спамерской программе, собирающей e-mail'ы с сайтов, поместить в свою базу данных. Таким образом, если после этого на «поддельные» e-mail придет письмо, можно со 100% уверенностью знать, что это спам, и немедленно заблокировать все остальные подобные письма, также прием писем с хоста.

http://www.stalker.com/CommuniGatePro/ – сайт MTA, имеющий встроенную возможность использования spam-traps.

6. Присвоение штрафных баллов. Bouncing

Письмо разбирается по баллам в зависимости от произвольных условий. Чем больше баллов, тем вероятней, что письмо является спамом. Этот метод может успешно объединить в себе фактически любые методы фильтрации спама. Администратору (или пользователю) остается только расставить приоритеты, означающие, насколько он доверяет одному методу или не доверяет другому, а также определить уровень (в условных баллах, с помощью которых задаются приоритеты), при достижении которого письмо считается спамом. После этого надо будет только лишь решить куда отправлять подобные письма: в отдельную папку, в специальный почтовый ящик для спама или сразу удалять. Обычно этот вопрос оставляют на усмотрение пользователя.

http://spamassassin.org/ – сайт разработчиков наиболее известного бесплатного bouncing фильтра Spamassassin. Это одна из самых эффективных на сегодняшний день систем фильтрации спама.

http://spamtest.ru/ , http://so.yandex.ru/ – сайты отечественных разработчиков в этой области.

«Поддельные» e-mail – «ловушки»...

7. Прием писем только от известных адресатов (занести в белые списки). White-listing

Принимаются только те письма, которые есть в авторизованном списке отправителей (например, только из вашей адресной книги), остальным отправляется ссылка, по которой можно добавить отправителя в *white-list* и, соответственно, получить его письмо.

http://www.paganini.net/ask/ , http://tmda.net/ – сайты программ, реализующих подобную возможность.

8. Временная задержка писем (занести в серые списки). Grey-listing

На письма от неизвестных отправителей во время SMTP сессии выдается ошибка 45*, что означает «Временные проблемы». В этом случае все правильно настроенные MTA (в отличие от спамерских программ) через некоторое время (обычно пол-

Письмо разбирается по баллам...

Когда словари заполнены, вычисление вероятности принадлежности конкретного нового письма к тому или иному типу производится по формуле Байеса.

часа, час) повторяют попытку доставить почту и на этот раз почта принимается.
http://projects.puremagic.com/greylisting/ – сайт разработчиков этого проекта.

9. Вероятностная оценка гипотезы о принадлежности к спаму (байесовский метод). Bayesian

Фильтрация спама на основе теоремы Байеса, применение частотного фильтра спам-сигнатур для вычисления вероятности принадлежности к спаму.

Примечание. Теорема Байеса (в теории вероятностей) – одна из основных теорем элементарной теории вероятностей. Названа по имени установившего ее английского математика Т. Байеса (Th. Bayes, XVII в.). Эта теорема имеет дело с расчетом вероятности верности гипотезы в условиях, когда на основе наблюдений известна лишь некоторая частичная информация о событиях. Другими словами, по формуле Байеса можно более точно пересчитывать вероятность, беря в учет как ранее известную информацию, так и данные новых наблюдений.

Применение теоремы Байеса для фильтрации спама нашел американский программист и предприниматель Пол Грэм (Paul Graham). В разработанном Грэмом прототипе фильтра для вычисления вероятности спама используются таблицы вероятностей (принадлежности слов из письма, относящегося к категории «спам»), созданные в процессе обучения фильтра. А именно: берутся как минимум два списка слов различных категорий писем (например, «разрешенных» и «запрещенных») и передаются на обработку программе обучения. Она вычисляет частотные словари для каждой категории сообщений – сколько раз какое слово встречалось в письмах этой категории (в данном случае спама). Когда словари заполнены, вычисление вероятности принадлежности конкретного нового письма к тому или иному типу производится по формуле Байеса для каждого слова этого нового письма. Суммированием и нормализацией вероятностей слов получают вероятности для всего письма. Как правило, вероятность принадлежности электронного письма к одной из категорий на порядок выше, чем к другим. К данной категории и следует относить сообщение.

Высокая вероятность присваивается как излюбленным спамерами словам, вроде sexy или promotion, так и таким, неожиданным, на первый взгляд, сочетаниям как ff0000 – код ярко-красного цвета в HTML. Соответственно, низкая вероятность соответствует профессиональным терминам или просто редко использующимся в рекламе словам вроде standartization или mandatory.

В процессе испытания системы фильтрации спама Грэм пропустил через нее 8000 писем, половина из которых являлась спамом. В результате, через фильтры смогли просочиться лишь 0,5% рекламных сообщений, а количество ошибочных срабатываний фильтра на основе байесовского подхода оказалось нулевым.

Более подробно о разработанном Грэмом прототипе фильтра можно почитать по адресу http://www.paulgraham.com/spam.html.

Для того чтобы подобная система была действительно эффективной, она должна поддерживать возможность индивидуальной настройки, поскольку терминология, использующаяся в электронной переписке разными людьми, отличается. Если же пользователь будет регулярно помечать рекламные письма как спам, то программа сможет накопить достаточно информации для эффективной фильтрации электронной почты.

Сайты программ, основывающихся на этом методе фильтрации можно найти по адресам:

http://www.tuxedo.org/~esr/bogofilter/
http://popfile.sourceforge.net/
http://spambayes.sourceforge.net/

10. Метод контрольных сумм (бритва). Razor

Фильтрация спама на основе определения контрольных сумм писем, являющихся спамом. Метод заключается в том, что для каждого письма, являющегося спамом, определяется контрольная сумма (сигнатура), и сигнатуры всех последующих писем, приходящих на почтовый сервер, сверяются с сигнатурой письма-спама. В «чистом» виде сейчас этот метод фактически не применяется по следующим причинам:

– необходима постоянно обновляемая большая база данных спам сигнатур, в то время как однозначно классифицировать спам может лишь сам пользователь (организация публичных почтовых ящиков или почтовых папок для классификации спама или виртуального «голосования» с помощью веб-интерфейса к почтовой системе) или *Spam traps*;

– в данный момент почти в любое письмо-спам вставляется уникальная для каждого отдельного письма последовательность случайных символов или цитат из обычных писем, художественных произведений и др. с целью обмануть такие фильтры.

Тем не менее, в связке с другими типами фильтрации этот метод активно применяется, особенно на MTA с большим почтовым трафиком и веб-интерфейсом к почте.
http://razor.sf.net/, http://pyzor.sourceforge.net/ – сайты проектов, занимающихся этим направлением.

11. Прочие, менее эффективные методы

• Проверка FQDN, Fully Qualified Domain Name (полностью определенное доменное имя) В соответствии с RFC степень стандартности тоже может быть разной – от общепринятых и обязательных технических спецификаций до аккуратно сформулированных предложений группы разработчиков.

Примечание. RFC (Request For Comments) – это ряд документов, являющихся стандартами де-факто сети Интернет. Сейчас этих документов уже более 3500, и все они пронумерованы, например, стандарты на электронное сообщение описаны в RFC822 (1991 г.), RFC1341 (1992 г.), RFC1521 (1993 г.), RFC2045–RFC2049 (1996 г.).

Фильтрация спама на основе определения контрольных сумм писем, являющихся спамом.

Прочитать любой из документов RFC можно по адресу: http://www.faqs.org.

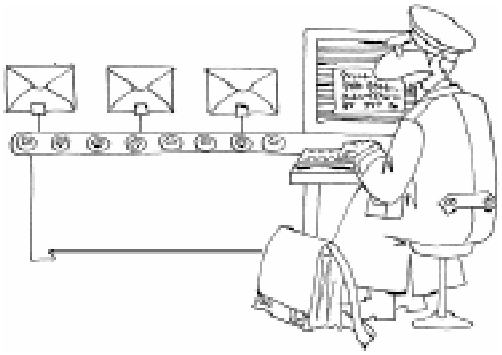
• Проверка домена отправителя (проверка в ДНС MX-записи, которая указывает на то, что домен отправителя имеет почтовый сервис).

• Черные списки e-mail (поле *From*).

• Проверка PTR-записи в DNS (проверка на то, чтобы хост отправителя был прописан в обратной зоне, – тоже рекомендация RFC).

• Проверка на валидность SMTP команд HELO/EHLO, MAIL FROM, RCPT TO.

• Ограничение количества получателей одного письма.



Прочие, менее эффективные методы.

Современная рассылка спама...

• *Throttling.* Задержка во время SMTP-сессии при попытке послать письмо более чем *n* адресатам.

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ РАССЫЛКИ СПАМА

Рассылка спама через открытые сервера (*open mail relays*) или с модемного пула (*dialup*) уже в прошлом. Современная рассылка спама, как правило, ведется через зараженные специальными вирусами (троянами) компьютеры обычных пользователей, причем часто перед отправлением спама каждый из этих компьютеров проверяется на присутствие в различных RBL. Формат писем, содержащих спам, чаще всего HTML, но нередко встречается и plain text. Поле *from* генерируется из списка реальных e-mail'ов, не имеющих к этому спаму ни малейшего отношения. Рекламируемые домены обычно не несут функциональной нагрузки и служат лишь перенаправителями на реально рекламируемые сайты. Такие домены часто регистрируются под каждую новую рассылку. IP-адреса используются каждый раз новые. Каждое письмо-спам из одной и той же рассылки имеет несколько уникальных словосочетаний, чтобы обойти razor- и bayes-фильтры. Каждая новая рассылка носит уникальный характер и каждое слово (или словосочетание), по которому можно сделать соответствующие выводы (типа SAVE MONEY, UNSUBSCRIBE), заменено синонимами или надежно зашифровано (например, ANABKRIBE) Таким образом, грамотная организация спам-рассылки позволяет обойти фактически любую антиспам-систему.

К счастью для нас, соблюдение всех этих условий делает рассылку спама значительно более дорогостоящей, а значит, менее рентабельной. На рынке спам-услуг также имеется конкуренция, а непосредственные заказчики спама редко интересуются тонкостями: в расчет обычно принимаются лишь общее количество e-mail'ов в базе рассылки и тематическая направленность. Таким образом, обычно в спам-письме все же существуют признаки, за которые можно «ухватиться». Есть еще одна особенность спама, на которую можно обратить внимание. При всех своих ухищрениях спамеры должны оставить письмо читабельным, и одновременно с этим им нужно продать вам виагру, доступ к XXX-сайту, отправить вас на семинар или продать туристическую путевку. Словом, то, что вы ежедневно видите в своих почтовых ящиках. Число вариантов написания предлагаемых товаров и услуг ограничено, а значит, поддается полному или частичному распознаванию. Лучше всего тут помогут программы, основанные на байесовском методе фильтрации или использующие «Хи-квадрат вероятности» (*popfile*, *spambayes*) Серьезная проблема возникает при обслуживании большого почтового сервера (от тысячи пользователей). Она заключается в том, что письма, востребованные одними пользователями, другими однозначно определяются как спам. Решить такую проблему могут лишь персонифицированные фильтры, но сложность реализации, отсутствие контроля системного администратора за такими фильтрами, резко возрастающие нагрузки на сервер и, как правило, низкая компьютерная грамотность у конечных пользователей привели к тому, что широкого распространения эти фильтры не получили.

ОПЫТ ФИЛЬТРАЦИИ

В связи с тем, что ни один из методов фильтрации спама не исключает случаев ложного срабатывания (*false positive*), мы были вынуждены отказаться от фильтрации перед приемом почты. Опыт показывает, что после приема почты с наибольшей эффективностью работают bouncing фильтры с параллельным использованием наиболее действенных методов фильтрации спама: RBL, Bayes, Regex, Razor, White-listing, Spam traps, FQDN, PTR etc.

Все эти проверки могут выполняться на сервере при использовании специализированного программного пакета spamassassin, который может интегрироваться с любым из известных MTA. Этот продукт бесплатен (*freeware*) и очень динамично развивается.

Почтовая система mail.runnet.ru обслуживает пользователей всех школ Санкт-Петербурга и других образовательных учреждений (несколько тысяч пользователей), для чего используется программное обеспечение CommunigatePro (официальный сайт – http://www.stalker.com/CommuniGatePro/), Spamassassin и бесплатный антивирус Clamav (официальный сайт – http://www.clamav.net/). Использование такого решения позволило фильтровать до 98% спама при чрезвычайно низком (<0.1%) коэффициенте ложных срабатываний.

Кроме использования персональных фильтров, каждый из пользователей может отключить серверные антиспам-фильтры или сам решить, что ему делать с письмами, определенными сервером как спам. Он может положить их в отдельный почтовый ящик или IMAP-папку специально для спама с тем, чтобы периодически просматривать такие письма, а может, сразу их удалять. Также всем пользователям на почтовом сервере доступны многие другие возможности (доступ по протоколам POP3, IMAP4, HTTP/HTTPS, FTP, антивирус, SMTP-Auth, APOP, AIMAP, возможность добавления событий, заметок, задач, веб-интерфейс, Shared directory, изменение пароля, RPOP, SSL/TLS, «автоответчик», LDAP и т. д.), описание которых выходит за рамки данной статьи. Полное описание

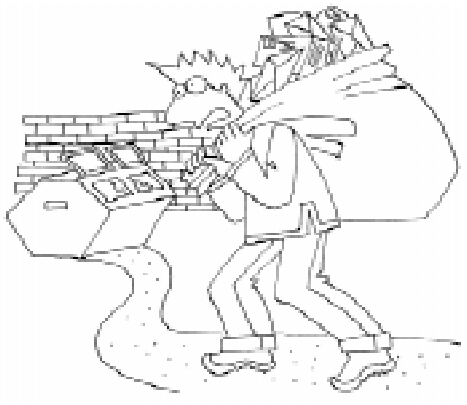
...ни один из методов фильтрации спама не исключает случаев ложного срабатывания...

этих сервисов доступно по адресу: http://mail.runnet.ru/advmailfaq.wssp.

Подобный сервис может быть предоставлен любым пользователям образовательных учреждений, а их администраторам предоставляется веб-интерфейс для управления своими пользователями и оказывается помощь в создании и настройке антиспам-фильтров.

ПЕРСПЕКТИВЫ ФИЛЬТРАЦИИ СПАМА

На данный момент уже существуют технологии, которые могут позволить в будущем прекратить такое явление, как спам, или сделать его долю незначительной. Перечислим основные из них:



...каждый из пользователей может отключить серверные антиспам-фильтры или сам решить, что ему делать с письмами, определенными сервером как спам...

1. Протокол X.400

Этот протокол был принят еще в 1984 году и с тех пор широко используется в банковском секторе, в телеграфных сетях, в системах корпоративного электронного документооборота и др. X.400 исключает возможность подстановки поддельного обратного адреса, а письмо, отправленное по этому протоколу, в принципе не может потеряться по дороге; при технической невозможности доставки, а также доставке письма в почтовый ящик адресата и его прочтении отправитель гарантированно получит извещение с указанием точного времени и т. п.
http://www.x400.org/ – сайт, на котором детально описывается этот протокол и его современное применение.

2. Белый список WWW. DRBL (World Wide Web white list)

Глобальный распределенный список доверенных релеев по технологии RBL.

3. Электронные почтовые марки (наличность для мелочевки). Hash cash

Применение электронных почтовых марок для расчетов за отправленные Email процессорным временем.
http://hashcash.org/ – сайт разработчиков этой технологии.

4. Методы защиты E-mail-адресов от подделки отправителем

• *SenderPolicyFramework (SPF).* На сегодняшний день SPF-записи включены в DNS уже более 15000 доменов. При получении SMTP команды MAIL FROM: <e-mail>, SMTP-сервер проверяет, разрешил ли владелец домена, указанного в e-mail (или в HELO, если MAIL FROM:<>) отправлять почту со своим доменом в From тому IP-адресу, с которого произведено подключение этой почтовой сессии. Эта проверка производится по DNS-записям этого домена – запросом TXT-записи SPF-формата.
http://spf.pobox.com/ – сайт разработчика этой технологии.

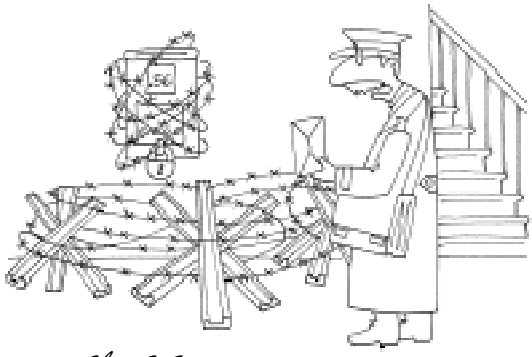
• *MS CallerID.* DNS TXT-записи аналогичного назначения, но XML-формата по предложенной Microsoft спецификации.
http://spf.pobox.com/caller-id/ – ссылка, описывающая эту технологию.

• *SenderID.* В 2004 г. спецификации SPF и MS CallerID объединились в SenderID. Этот метод позволяет проверять право отправителя использовать обратный e-mail адрес с заданным доменом. Для этого администраторы почтовых доменов заносят в DNS этих доменов TXT-записи особого формата, где указывается список хостов, которые могут отправлять почту с этого домена. http://www.microsoft.com/mscorp/twc/privacy/spam/senderid/default.mspx – описание этой технологии на сайте microsoft.

• *Системы цифровой подписи для e-mail Open-PGP и S/MIME* (поддерживает сертификаты X.509). Подпись вставляется в S/MIME-секцию. Open-PGP и S/MIME, к сожалению, не совместимы друг с другом.
http://www.ietf.org/html.charters/smime-charter.html – спецификация подписи S/MIME.
http://www.ietf.org/html.charters/openpgp-charter.html – спецификация цифровой подписи Open-PGP.

• *Подпись доменным ключом. Yahoo Domain Keys.* Эта технология предусматривает автоматическую цифровую подпись исходящей почты «доменным ключем» на сервере-отправителе и проверку принимающим SMTP-сервером этих подписей на базе открытых ключей, опубликованных в DNS домена-отправителя.

Письма подписываются 384-битным ключем, подпись вставляется в заголовок письма «DomainKey-Signature». Открытый ключ



...существуют технологии, которые могут позволить в будущем прекратить такое явление как спам...

домена публикуется в DNS в TXT-записи субдомена _domainkey. Никаких привязок к IP отправителя не производится – этим YDK превосходит методы SenderPolicyFramework и CallerID. Но это и его недостаток: письмо с YDK-подписью придётся принимать полностью, а SPF позволяет отвергать письмо еще до его приема – на стадии SMTP-команды MAIL FROM.

Главное отличие – в YDK предполагается использование не персональных пользовательских ключей, а по-доменных. И YDK предлагает использовать инфраструктуру DNS вместо традиционных для криптографии PKI (Public Key Infrastucture). Видимо, для того чтобы не привязываться к платным сертификатам и владельцам PKI/CA. Или просто для уменьшения нагрузки на серверы, проверяющие подписи.

Вывод: При всех своих преимуществах все указанные выше технологии будут эффективны только в том случае, если все пользователи электронной почты (или хотя бы их большая часть) выберут одну из них и начнут ее использовать, что в ближайшем будущем малоосуществимо. А пока вполне неплохие результаты могут дать и уже имеющиеся методы борьбы со спамом.

...все указанные выше технологии будут эффективны только в том случае, если все пользователи электронной почты ... выберут одну из них и начнут ее использовать

Лазо Олег Иванович,
ГосНИИ ИТТ «Информика»
СПб филиал.