К. П. Рудик
Московский инженерно-физический институт (государственный университет)

ВЫЯВЛЕНИЕ СЕТЕВЫХ УЗЛОВ, УЧАСТВУЮЩИХ В РАСПРОСТРАНЕНИИ ВРЕДОНОСНЫХ ПРОГРАММ И СПАМА В СИСТЕМАХ ЭЛЕКТРОННОЙ ПОЧТЫ

Приводится описание технологий, применяемых для рассылки вредоносных программ и спама в системах электронной почты, а также проводится анализ этих технологий; анализируются методы, используемые для сокрытия источника рассылки; предлагается метод определения источника несанкционированной рассылки и описание алгоритма определения присутствия фиктивных заголовков.

В связи с развитием и распространением сети Интернет растет и число способов распространения вредоносных программ, к которым можно отнести электронную почту, файлообменные сети, различные интернет-пейджеры (типа ICQ и Windows Messenger), интернет-чаты, непосредственную передачу от одного сетевого узла другому путем использования уязвимостей в применяемых программах.

Из возможных способов распространения в сети Интернет вирусы активно используют каналы электронной почты. Об этом свидетельствуют отчеты, которые предоставляются компаниями, занимающимися антивирусной защитой. Так, при анализе данных по вирусной активности, предоставленных компанией «Лаборатория Касперского», выяснилось, что практически все самые распространенные в 2006 г. вирусы оказались способны рассылать свои копии с использованием электронной почты.

Следует также отметить тенденцию к объединению вирусных технологий с технологиями анонимной незапрошенной массовой рассылки электронной почты [1] (далее спама). По оценкам экспертов по информационной безопасности [2], основным способом рассылки спама является использование сетей компьютеров обычных пользователей, превращенных посредством использования уязвимостей в ПО или посредством использования вирусных технологий в компьютеры-зомби.

По данным различных компаний, занимающихся проблемами спамовых рассылок, процент спамовых сообщений, рассылаемых с использованием компьютеров-зомби, превышает 60 % (по данным экспертов компании SophosLabs [3]), а по другим оценкам, может превышать 80 % [4] от всего объема спама.

Компьютеры-зомби дают возможность авторам спама не только рассылать миллионы писем, но и скрывать истинные источники рассылок, уходя от судебного преследования. Отсюда можно сделать вывод, что правильно определенные источники рассылки спама могут с высокой достоверностью указывать на инфицированные вредоносным ПО компьютеры.

Одним из направлений борьбы с распространением вредоносных программ, а также спама является локализация источников распространения с последующим воздействием на них (отключение от сети, удаление вредоносных программ и т. п.) для прекращения дальнейшего распространения ими несанкционированной рассылки. Важным вопросом, встающим при рассмотрении задачи локализации, является задача сбора и хранения данных, полученных от подсистем антивирусной и антиспамовой фильтрации почтовых сообщений. Собранные данные могут быть использованы как с целью локализации источников несанкционированной рассылки, так и для последующего анализа конфликтных ситуаций, например, в случаях, когда антиспамовая система ошибочно приняла сообщение за спам и важное для пользователя письмо было удалено.

Для определения источника рассылки электронной почты может использоваться IP-адрес сетевого узла, с которого происходит распространение вредоносных программ или спама. Однако в некоторых случаях определение IP-адреса источника затруднительно.

Так, при решении задачи определения источника, участвующего в распространении вредоносного ПО в системах электронной почты, основной проблемой является то, что из-за специфики технологии электронной почты IP-адрес источника почтового сообщения, т. е. IP-адрес узла, установившего соединение с почтовым сервером получателя, не всегда является адресом узла, сгенерировавшего почтовое

сообщение: в основном это адрес почтового сервера отправителя или промежуточный почтовый сервер, а не адрес узла, сгенерировавшего почтовое сообщение.

1. Анализ технологий, используемых для рассылки вредоносных программ и спама

В связи с тем, что в настоящее время подавляющее большинство рассылаемого спама и практически все вредоносные программы, за исключением, возможно, первых копий, которые запускаются злоумышленниками в Сеть «вручную», рассылаются с использованием специализированных программ, можно определить основные технологии, используемые при рассылке вредоносных программ и спама, путем сравнительного анализа функциональности данного вида программ.

Рассмотрим программы и их функциональные особенности, используемые для рассылки незапрошенной электронной почтовой корреспонденции, а также вредоносные программы, используемые для аналогичных целей (Таблица 1).

Таблица 1. Функциональные особенности программ для рассылки спама.

Название программы	Функциональные особенности
Mega Mailer 3.1	возможность отправлять письма через почтовые серверы, требующие SMTP- или POP3-авторизацию; может осуществлять рассылки одновременно через несколько smtp-серверов; имеет возможность производить рассылку через несколько цепочек Socks прокси-серверов, переключаясь между ними;
Advanced Direct Remailer 2.12	подключение напрямую к почтовому серверу получателя; позволяет выполнять анонимные рассылки через прокси-сервера Socks5, скрывая IP-адрес отправителя;
Advanced Mass Sender 4.3	подключение напрямую к почтовому серверу получателя; отправка сообщений через заданный набор почтовых серверов;
Dark Mailer 3.12	рассылает письма через встроенный или внешний smtp-сервер; рассылка через прокси-сервера;
SocksChain 3.0	программа, позволяющая работать через цепочку SOCKS или HTTP-proxy серверов для того, чтобы скрыть истинный IP-адрес;
Socks Cap	программа позволяет любой программе работать через прокси socks5 для того, чтобы скрыть истинный IP-адрес;
SpamTool.Win32.Delf.h (источник www.viruslist.com)	программа может быть использована для рассылки спама на адреса электронной почты, найденные на зараженном компьютере; подключение напрямую к почтовому серверу получателя; информацию для рассылки получает с определенных web-ресурсов; вредоносная программа имеет функцию загрузки других файлов из Интернета и запуска их на зараженном компьютере;
SpamTool.Win32.Maniac.b (источник www.viruslist.com)	рассылает сообщения через почтовые серверы, указанные в клиентских почтовых программах, установленных на зараженном компьютере; информацию для рассылки получает с определенных web-ресурсов;
Spam-Maxy (источник www.mcafee.com)	подключение напрямую к почтовому серверу получателя; информацию для рассылки получает с определенных web-ресурсов;
Trojan.Spamlia (источник www.symantec.com)	подключение напрямую к почтовому серверу получателя; рассылает спам по адресам из адресной книги Windows;

Trojan.Ascetic.C/Email-Worm.Win32.Sober.q (источник www.symantec.com)	подключение напрямую к почтовому серверу получателя; рассылает спам по адресам, найденным на зараженном компьютере;
Spam-Mespam (источник www.mcafee.com)	рассылает сообщения через почтовые системы с Web-интерфейсом; информацию для рассылки получает с определенного web-ресурса;
Spam-Skull.dll (источник www.mcafee.com)	подключение напрямую к почтовому серверу получателя; информацию для рассылки получает с определенных web-ресурсов;
TROJ_MITGLIED.EO (источник www.trendmicro.com)	открывает произвольный порт и позволяет пересылать через него почтовые сообщения, т. е. работает в качестве прокси-сервера;
Spam-Loot (источник www.mcafee.com)	подключение напрямую к почтовому серверу получателя; информацию для рассылки получает с определенных web-ресурсов;
Proxy-Agent.d (источник www.mcafee.com)	позволяет использовать зараженный компьютер как открытый релей для пересылки почтовых сообщений;
Backdoor.Win32.VanBot.bj (источник www.viruslist.com)	позволяет использовать зараженный компьютер в качестве прокси-сервера;
TROJ_REGATE.A (источник www.trendmicro.com)	позволяет использовать зараженный компьютер как открытый релей для пересылки почтовых сообщений;
TROJ_AGENT.CYO (источник www.trendmicro.com)	позволяет использовать зараженный компьютер в качестве прокси-сервера;
TROJ_SPABOT.AG (источник www.trendmicro.com)	подключение напрямую к почтовому серверу получателя; информацию для рассылки получает с определенных web-ресурсов.

Из данных, приведенных в таблице 1, видно, что программное обеспечение для рассылки вредоносного ПО или спама можно разделить на следующие категории в зависимости от того, как осуществляется рассылка:

· рассылка посредством подключения непосредственно к почтовому серверу получателя с использованием встроенного почтового клиента;

· рассылка через почтовый сервер, определенный политикой, действующей в рамках узла, с которого ведется рассылка, или сети, которой он принадлежит;

· рассылка через заданный почтовый сервер или шлюз;

· рассылка через посредника с применением протоколов, отличных от SMTP, например прокси-сервера, или не полностью соблюдающих его, например без добавления заголовка «Received» или добавления ложных заголовков.

Подключение к почтовому серверу получателя с использованием встроенного почтового клиента является на сегодняшний день самым распространенным способом рассылки вирусов и спама. Анализ, проведенный на примере тестовой выборки сообщений, содержащих спам и вредоносные программы, показал, что такая тактика характерна для большинства вирусов, распространяемых с использованием электронной почты, а также спама. Широкое применение данного способа рассылки можно объяснить, во-первых, большой скоростью рассылки, а во-вторых, отсутствием единого канала пересылки, который можно своевременно обнаружить и «перекрыть».

Рассылка через почтовый сервер, определенный политикой, действующей в рамках узла, с которого ведется рассылка, или сети, которой он принадлежит, в основном используется, когда прямой доступ к серверам получателей по каким-либо причинам закрыт. Например, провайдер или администратор сетевого узла, с которого ведется рассылка, с целью пресечения возможности своих пользователей рассылать спам или вирусы блокировал доступ пользователей ко всем внешним почтовым

серверам, оставив возможность отправки электронной почты только через свои почтовые серверы. При этом ПО, осуществляющее рассылку, должно быть соответствующим образом настроено на отправку сообщений через почтовый сервер провайдера. Если рассылку осуществляет вредоносное ПО, то данные настройки оно может получить, проанализировав настройки почтовых клиентов, которые используются на зараженном компьютере.

Применение для рассылки заданного почтового сервера или шлюза может использоваться для повышения скрытности реального узла рассылки. Так, например, при использовании WEB-шлюзов для рассылки почты (www.mail.ru, www.yandex.ru и т. п.) совместно с общедоступным анонимным прокси-сервером может быть достигнута практически полная анонимность, и задача определения узла, с которого ведется рассылка без привлечения администраторов WEB-шлюзов и прокси-серверов, становится неразрешимой.

Дополнительным поводом к использованию данного метода рассылки является то, что почта может рассылаться через серверы известных почтовых провайдеров (mail.ru, gmail.com, yahoo.com и т. п.), а блокировать такие рассылки путем занесения серверов в «черные списки» (blacklist) проблематично из-за неизбежной потери легальной корреспонденции.

Рассылка почтовых сообщений через посредника с применением протоколов, отличных от SMTP или не полностью соблюдающих его, направлена в первую очередь на сокрытие реального источника рассылки. В таких случаях исходным адресом рассылки на основании информации из служебных заголовков сообщения будет промежуточный узел, являющийся прокси-сервером (протокол, отличный от SMTP) или почтовым ретранслятором (неполное соблюдение SMTP).

Анализ способов рассылки сообщений, описанных выше, позволяет сделать вывод о том, что, например, при рассылке через промежуточный узел с применением протоколов, отличных от SMTP, определение источника, сгенерировавшего почтовое сообщение, в общем случае невозможно. Единственное, что в таких случаях можно определить, — IP-адрес узла, участвующего в рассылке.

2. Анализ методов, используемых для сокрытия источника рассылки вредоносных программ и спама

Проанализируем основные методы, используемые для сокрытия источника, с которого отсылается электронное почтовое сообщение.

Методы сокрытия источника рассылки можно условно разделить на два типа:

· добавление фиктивной информации;

· использование сторонних сетевых узлов для передачи почтового сообщения.

Кроме того, возможны комбинации этих типов.

Самым распространенным методом, основанным на добавлении фиктивной информации, без которого не обходится практически ни одна рассылка спама и вредоносных программ, является *использование фиктивной информации в поле «From», указывающей на электронный адрес отправителя*. Однако на процесс определения по служебным заголовкам сетевого узла, с которого пришло письмо, этот метод не оказывает влияния.

Еще одним распространенным способом сокрытия источника рассылки является *Добавление фиктивных заголовков Received*. Этот способ широко применяется при рассылке спама и в меньшей степени — при распространении вредоносного ПО. Добавление фиктивных заголовков усложняет задачу поиска источника рассылки, а в некоторых случаях делает ее невозможной.

Среди методов, использующих сторонние сетевые узлы, можно отметить: использование общедоступных шлюзов (Open Relay); прокси-серверов или шлюзов с изменением среды передачи, а также использование компьютеров-зомби.

Метод, предполагающий *использование общедоступных шлюзов (Open Relay)*, позволяющих передавать через себя почтовые сообщения для любых адресатов, в настоящий момент используется

сравнительно редко, так как число таких шлюзов, а это, как правило, неправильно сконфигурированные почтовые серверы, постоянно уменьшается.

В случае использования нарушителем *общедоступных прокси-серверов или шлюзов с изменением среды* передачи (например, почтовые серверы, работающие через WEB-интерфейс) достоверность определения IP-адреса, с которым он подключается к сети Интернет, практически исчезает. Это связано с тем, что если промежуточный сервер (прокси-сервер или шлюз) не выдает информацию о работающем через него клиенте (нарушителе, использующем данный сервер для передачи почтовых сообщений) каким-либо образом, то необходимым, но недостаточным условием для определения IP-адреса нарушителя является содействие со стороны владельца данного прокси-сервера или шлюза.

Однако использование нарушителем для сокрытия источника рассылки прокси-серверов или шлюзов с изменением среды передачи ведет к усложнению процесса рассылки, так как появляются следующие дополнительные особенности:

· поиск и обновление списков доступных прокси-серверов или шлюзов;

· усложнение программного обеспечения, используемого для рассылки, и, как следствие, увеличение его размера и «заметности» (если оно используется на «зомби» машинах).

В связи с усложнением процесса рассылки применение данного метода при распространении вредоносного программного обеспечения является маловероятным, однако при рассылке спама, судя по функциональности программного обеспечения для рассылки спама, предлагаемого на специфических форумах и сайтах, метод широко используется.

Оптимальным для злоумышленника с точки зрения сокрытия информации о своем сетевом местонахождении является *использование компьютеров-зомби* (зомби — компьютеры с функционирующим на них вредоносным программным обеспечением, дающим злоумышленнику возможность удаленного управления), так как в этом случае почтовая рассылка ведется от лица (с сетевого узла, который имеет определенный IP-адрес) зомби. Компьютеры-зомби дают возможность авторам спама не только рассылать миллионы писем, но и скрывать истинные источники рассылок, уходя от судебного преследования.

3. Метод выявления источника несанкционированной рассылки сообщений электронной почты

Ранее были рассмотрены технологии, применяемые для рассылки вредоносных программ или спама, а также методы, используемые для сокрытия истинных источников рассылки. Рассмотрим теперь предложенный автором метод определения источника несанкционированной рассылки.

Основное требования к решению задачи определения источника несанкционированной рассылки состоит в том, что решением не должен оказаться узел, не принимавший участия в несанкционированной рассылке (т. е. подставной узел из фиктивного заголовка). Иными словами, в качестве узла отправителя допустимо указать промежуточный узел, принимающий участие в рассылке, и недопустимо принимать за источник рассылки узел, заявленный в поддельном заголовке.

Спецификация протокола SMTP обязывает каждого почтового транспортного агента (MTA), через которого проходит сообщение, добавлять к сообщению свой заголовок «Received»; в заголовке «Received», как правило, указывается, кем было получено почтовое сообщение и от кого. Поэтому наличие в почтовом сообщении нескольких заголовков «Received» означает, что сообщение проходило через несколько почтовых серверов, однако это может также означать, что часть заголовков фиктивна.

Анализ последовательности заголовков «Received» электронных почтовых сообщений обеспечивает решение задачи определения источника несанкционированной рассылки, т. е. задачи поиска сетевого узла, которым было первоначально сформировано сообщение. Кроме того, разбор служебных заголовков почтового сообщения является единственным способом определения источника рассылки, поскольку только в заголовке «Received» содержатся данные о сетевых узлах, участвующих в передаче сообщения.

При проведении анализа предполагается, что заголовок, сформированный на последнем (в соответствии с маршрутом следования почтового сообщения) почтовом сервере, является истинным, так как его сгенерировал доверенный почтовый сервер.

В общем случае, с учетом возможности присутствия ложных заголовков, набор служебных заголовков «Received» можно представить в виде, показанном на рис.1:

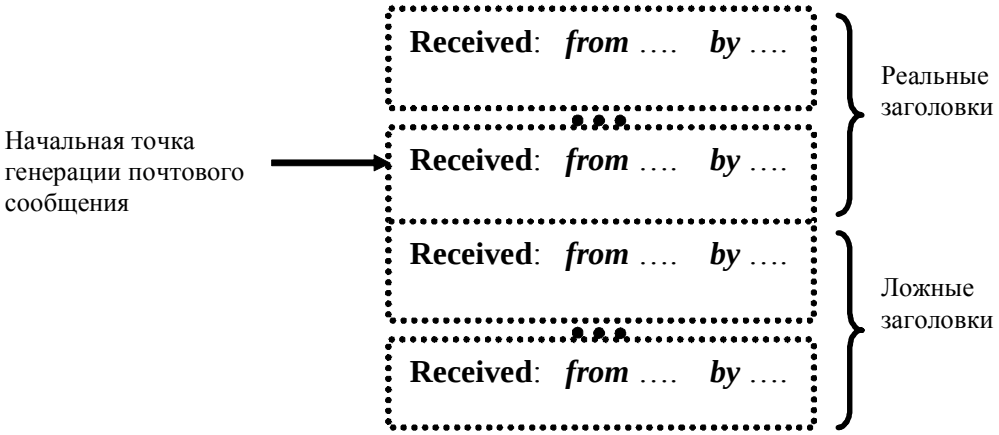

Рис. 1. Набор служебных заголовков «Received».

Разбор последовательности заголовков «Received» можно представить в виде конечного автомата $M=(Q,\Sigma,\delta,q_0,L)$, где Q — конечное непустое множество состояний; Σ— конечный входной алфавит; δ— отображение типа $Q\times\Sigma\to Q$; $q_0\in Q$ — начальное состояние; $L\subseteq Q$ — множество конечных состояний.

Однако в связи с тем, что множество заголовков $R=\{r_i\}$ не является конечным алфавитом, введем дополнительно функцию F, анализирующую информацию из заголовков, такую, что

$$F=\begin{cases}0, \text{ если текущий заголовок является последним в цепочке или поддельным;}\\ 1, \text{ если текущий заголовок соответствует промежуточной точке генерации почтового сообщения.}\end{cases}$$

Тогда входной алфавит $\Sigma=\{w_i\}$ — конечное множество значений функции F, такое, что $\forall\, r_i\in R$ $w_i = F(r_i)$.

На каждом шаге автомат рассматривает значение функции F, которое зависит от содержания очередного заголовка, и принимает решение о переходе к рассмотрению значения функции от следующего заголовка либо решение об остановке. Следовательно, основной проблемой является выбор параметров, от которых зависит функция F, т. е. необходимо определить параметры, анализируя которые можно сделать предположение о присутствии в конкретном сообщении фиктивных заголовков. При этом если заголовок r_i является первым, встретившимся фиктивным, то r_{i-1} соответствует начальной точке генерации почтового сообщения.

В ходе анализа заголовков почтовых сообщений разных категорий (спам, вредоносные программы, обычные почтовые сообщения) эмпирическим путем был введен ряд условий, проверка которых в определенных случаях позволит определить, является ли рассматриваемый заголовок «Received» истинным или поддельным. Эти условия следующие:

- соответствие заголовка формату [5] (формат заголовков «Received» должен удовлетворять требованиям протокола — RFC 2821 и RFC 2822);
- для доменного имени, указанного в заголовке в качестве MTA, должна быть соответствующая запись «MX» или «A»;
- доменное имя, указанное в заголовке в качестве MTA, должно быть не выше третьего уровня;
- связность заголовков «Received»;

· связность полей *by* и *from* заголовка «Received»;

· проверка доступности из сети Интернет почтовой службы (TCP/25, TCP/110, TCP/465) на сервере, указанном в качестве MTA.

Однако опыт анализа заголовков почтовых сообщений показал, что существуют ситуации, когда сведений, которые можно почерпнуть из служебных заголовков почтового сообщения (или из сообщения в целом), оказывается недостаточно для определения источника рассылки даже для эксперта, не говоря уже об автоматизированной системе.

Рассмотрим довольно распространенный случай отправки сообщения с узла через почтовую систему своего провайдера. При этом возможны три варианта:

1) почтовая система провайдера при получении письма от клиента добавляет заголовок «Received» с истинным адресом;

2) почтовая система провайдера добавляет неполный по формату заголовок «Received» или заголовок с измененным адресом отправителя;

3) почтовая система провайдера не добавляет заголовок «Received», например, с целью защиты информации о внутренней структуре сети.

В варианте 3, частично в варианте 2, при добавлении фиктивного заголовка на стороне узла отправителя, при условии соблюдения всех формальных зависимостей, определение факта добавления не представляется возможным при отсутствии дополнительной информации о конкретных настройках почтовой системы провайдера.

Рассмотрим приведенные выше варианты на примере. Допустим, что узел с адресом «A» послал сообщение через сервер своего провайдера «B» для пользователя, находящегося «за» почтовым сервером «C»; тогда в упрощенной форме заголовки почтового сообщения на стороне получателя могут быть такими:

для случая 1:

Received: from B ([**B**]) by **C** with ESMTP for test@C; Fri, 04 May 2005 14:46:18 +0400 (MSD)

Received: from A ([A]) by B with SMTP for test@C; Fri, 04 May 2005 14:43:10 +0400 (MSD)

для второго случая:

Received: from B ([**B**]) by **C** with ESMTP for test@C; Fri, 04 May 2005 14:46:18 +0400 (MSD)

Received: by B with SMTP for test@C; Fri, 04 May 2005 14:43:10 +0400 (MSD)

для третьего случая:

Received: from B ([**B**]) by **C** with ESMTP for test@C; Fri, 04 May 2005 14:46:18 +0400 (MSD)

Или, при добавлении фиктивного заголовка, таким:

Received: from B ([**B**]) by **C** with ESMTP for test@C; Fri, 04 May 2005 14:46:18 +0400 (MSD)

Received: from dummy.domen ([X.Y.Z.W]) by **B** with ESMTP for <test@C>; Fri, 04 May 2005 14:46:22 +0400 (MSD)

В последнем примере отличить фиктивный заголовок от настоящего без дополнительных знаний о системе отправителя невозможно. Т. е. констатировать факт добавления фиктивного заголовка возможно в случае, когда известно, что почтовая система не добавляет свой заголовок или что система может принимать почту для пересылки только от абонентов внутренней сети.

Анализ тестовой выборки почтовых сообщений, содержащих спам и вредоносные программы, собранной за период с 01.03.2007 по 01.04.2007 и содержащей 1256 сообщений, дал следующие результаты:

· процент сообщений, отправленных через почтовые системы провайдеров, не добавляющих заголовок «Received» (3-й вариант), по нижней оценке составляет 18 %;

· 9 % сообщений были отправлены через почтовые системы провайдеров, и в них присутствовал соответствующий заголовок «Received», указывающий на то, от какого узла почтовый сервер провайдера получил сообщение. И только для 33 % таких сообщений эксперт мог с определенной достоверностью определить, было ли действительно сообщение послано с узла, находящегося в заголовке.

Результаты, полученные на основании тестовой выборки, подтверждают тот факт, что существует класс случаев, для которых отличить фиктивный заголовок от настоящего без дополнительных знаний о системе отправителя невозможно.

Единственным возможным решением данной проблемы, по мнению автора, является сужение задачи поиска источника рассылки путем введения дополнительных ограничений, при выполнении которых можно с большой достоверностью утверждать, что найденный узел, участвующий в рассылке, не является фиктивным.

Одним из таких ограничений может быть проверка того, что узлы, указанные в полях *by* и *from* заголовка «Received», имеют одного владельца. Говоря другими словами, проверка связности полей by_i и $from_i$ для *i*-го заголовка «Received» $(i>1)$ может установить, принадлежат ли IP-адреса узлов из by_i и $from_i$ одной организации. Такое ограничение можно мотивировать тем, что основной целью добавления нарушителем фиктивного заголовка является сокрытие своего местоположения и, следовательно, указание в качестве ложного источника узла из своей же подсети нецелесообразно, так как это обстоятельство легко проверяется администратором этой сети.

На основании описанного выше определим набор параметров $\boldsymbol{Q}=\{\boldsymbol{R},\boldsymbol{C}_R,\boldsymbol{C}_{fb},\boldsymbol{D}_{MX},\boldsymbol{D}_A,\boldsymbol{L},\boldsymbol{S}\}$ и их возможные значения (Таблица 2).

Таблица 2. Описание параметров.

Обозначение	Описание
$\boldsymbol{R}$	Показатель соответствия заголовка формату (формат заголовков «Received» должен удовлетворять требованиям протокола — RFC 2821 и RFC 2822). Случаи несоответствия формату, как правило, связаны со следующими причинами: · добавление фиктивного заголовка вредоносным ПО для усложнения поиска источника рассылки; · несоблюдение протокола ПО почтового сервера.
$\boldsymbol{C}_R$	Показатель корреляции (связности) заголовков «Received». Сравниваются текущий и предыдущий заголовки. Условие связности заголовков следует из соответствия протоколу RFC 2821. Отсутствие связности заголовков может быть следствием того, что: · добавлен фиктивный заголовок «Received» без соблюдения зависимостей; · не соблюден протокол программным обеспечением почтового сервера.
$\boldsymbol{C}_{fb}$	Показатель связности полей *by* и *from* для заголовка «Received». Связность имеет место, если: · соблюдается протокол; · узел, который отправил сообщение, и сервер, получивший его и добавивший данный заголовок, находятся в одной подсети (с одинаковым владельцем IP-адресов). Условие выполняется, в основном если узел ведет рассылку через почтовый сервер своего провайдера. Условие не выполняется для почтовых систем типа www.mail.ru, www.gmail.ru и т. п. Для таких систем необходимо предусмотреть список исключений.

D_{MX}, D_A	Показатели наличия для доменного имени, указанного в заголовке в качестве MTA, соответствующей записи «MX» (D_{MX}) или «A» (D_A) (RFC 2821: «…в транзакциях SMTP появление локальных псевдонимов недопустимо…»). Как правило, отсутствие доменной записи «A» для узла является признаком того, что IP-адрес выдавался узлу динамически, следовательно, возможность того, что данный узел является MTA, — небольшая. Присутствие записи «MX» говорит о том, что данный узел является получателем почтовых сообщений для данного домена, а для небольших организаций эти функции, как правило, выполняет один и тот же сервер (т. е. при наличии записи «MX» для узла, указанного в поле *by* заголовка, вероятность того, что данный заголовок поддельный, меньше, чем при ее отсутствии).
L	Доменное имя, указанное в заголовке в качестве MTA, должно быть не выше третьего уровня.
S	Показатель доступности из сети Интернет почтовой службы (TCP/25, TCP/110, TCP/465) на сервере, указанном в качестве MTA. При выполнении условия для узла, указанного в поле *by* заголовка, возможность того, что данный заголовок поддельный, меньше, чем при невыполнении условия.

Каждый из параметров, перечисленных в таблице 2, может принимать значения {0;1} в зависимости от выполняемости конкретного условия.

В связи с тем, что в ряде случаев определение источника рассылки не представляется возможным, было решено ввести параметр N, который зависит от значений параметров из множества Q, перечисленных в таблице 2, и который характеризует степень достоверности правильного определения начальной точки движения почтового сообщения, представимого в виде натуральных чисел в определенном диапазоне. При этом любому набору параметров из множества Q ставится в соответствие натуральное число. Чем больше значение этого параметра, тем больше достоверность правильного определения начальной точки движения почтового сообщения.

Определим множество всех возможных наборов параметров из множества Q как W. Мощность множества, т. е. число его элементов, равна $|W| = 2^{|Q|}$, где $|Q|$ — мощность множества Q. В нашем случае для $Q=\{R, C_R, C_{fb}, D_{MX}, D_A, L, S\}$, теоретически, мощность $|W| = 2^7 = 128$. Однако, исходя из того, что без выполнения условия $R=1$, $C_R=1$ определение источника рассылки является невозможным (т. е. случаи, когда $R=0$ или $C_R=0$ нас не интересуют и принимаются за один случай), на практике $|W| = 2^5+1=33$.

Для определения значения параметра N для каждого набора $w_i \in W$ будем использовать методы ранжирования объектов по важности, применяемые при анализе экспертных оценок.

При решении задачи упорядочивания большого числа объектов возникают трудности психологического характера, обусловленные восприятием экспертами множества свойств объектов. В этом случае целесообразно использовать метод парного сравнения объектов.

Предположим, что необходимо расположить в определенной последовательности n объектов по какому-либо фактору (критерию). Пусть m экспертов производят оценку всех пар объектов, давая следующую числовую оценку:

$$r_{ij} = \begin{cases} 1, \text{ если } w_i > w_j \ (w_i \text{ предпочтительнее } w_j) \\ 0{,}5, \text{ если } w_i \approx w_j \ (w_i \text{ и } w_j \text{ эквивалентны}) \\ 0, \text{ если } w_i < w_j \ (w_j \text{ предпочтительнее } w_i) \end{cases} .$$

При этом вектор коэффициентов относительной важности объектов $k=(k_1, k_2, k_3, \ldots, k_n,)$ стремится к собственному вектору матрицы $X = \|x_{ij}\|$, соответствующему максимальному собственному числу

матрицы, где $x_{ij} = 0.5 + \frac{m_i - m_j}{2m}$, а m_i — число экспертов, высказавшихся в пользу предпочтения w_i над w_j; m_j — число экспертов, высказавшихся в пользу предпочтения w_j над w_i.

На практике вычисление коэффициентов относительной важности объектов проще производить последовательной процедурой по следующим формулам:

$k^t = \frac{1}{\lambda^t} Xk^{t-1}$ — вектор коэффициентов относительной важности объектов порядка t,

$\lambda^t = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} k_j^{\,t-1}$.

Вернемся к задаче определения значений параметра $\boldsymbol{N}$ (степени достоверности). Пусть имеются упорядоченные по степени важности входящих в них параметров наборы $\boldsymbol{w}$, тогда в качестве значения $\boldsymbol{N}$ можно взять порядковый номер набора.

Для определения значения $\boldsymbol{N}$ с использованием метода парного сравнения была построена матрица математических ожиданий оценок пар объектов, на основании которой определен вектор коэффициентов относительной важности объектов. После упорядочивания коэффициентов относительной важности объектов по возрастанию были получены искомые значения степени достоверности $\boldsymbol{N}$ для заданных наборов условий (Таблица 3):

Таблица 3. Соответствие степени достоверности набору условий.

$\boldsymbol{N}$	Набор условий
0	$R=0$ *или* $C_R=0$
1	$R=1,\ C_R=1$
2	$R=1,\ C_R=1,\ D_A=1$
3	$R=1,\ C_R=1,\ L=1$
3	$R=1,\ C_R=1,\ D_A=1\ L=1$
4	$R=1,\ C_R=1,\ S=1$
5	$R=1,\ C_R=1,\ D_{MX}=1$
6	$R=1,\ C_R=1,\ D_A=1\ S=1$
7	$R=1,\ C_R=1,\ D_A=1\ D_{MX}=1$
8	$R=1,\ C_R=1,\ L=1,\ S=1$
9	$R=1,\ C_R=1,\ D_A=1,\ L=1,\ S=1$
10	$R=1,\ C_R=1,\ D_{MX}=1,\ L=1$
10	$R=1,\ C_R=1,\ D_{MX}=1,\ D_A=1,\ L=1$
11	$R=1,\ C_R=1,\ D_{MX}=1,\ S=1$
11	$R=1,\ C_R=1,\ D_{MX}=1,\ D_A=1,\ S=1$
12	$R=1,\ C_R=1,\ C_{fb}=1$
13	$R=1,\ C_R=1,\ C_{fb}=1,\ D_A=1$

$\boldsymbol{N}$	Набор условий
14	$R=1,\ C_R=1,\ D_{MX}=1,\ L=1,\ S=1$
14	$R=1,\ C_R=1,\ D_{MX}=1,\ D_A=1,\ L=1,\ S=1$
15	$R=1,\ C_R=1,\ C_{fb}=1,\ L=1$
15	$R=1,\ C_R=1,\ C_{fb}=1,\ D_A=1,\ L=1$
16	$R=1,\ C_R=1,\ C_{fb}=1,\ S=1$
16	$R=1,\ C_R=1,\ C_{fb}=1,\ D_A=1,\ S=1,$
17	$R=1,\ C_R=1,\ C_{fb}=1,D_{MX}=1$
17	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,\ D_A=1$
18	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,\ L=1$
18	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,\ D_A=1,\ L=1$
19	$R=1,\ C_R=1,\ C_{fb}=1,\ D_A=1,\ L=1,\ S=1$
19	$R=1,\ C_R=1,\ C_{fb}=1,\ L=1,\ S=1$
20	$R=1,\ C_R=1,\ C_{fb}\ D_{MX}\ S=1$
21	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,\ D_A=1,\ S=1$
22	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,\ L=1,\ S=1$
22	$R=1,\ C_R=1,\ C_{fb}=1,\ D_{MX}=1,D_A=1,L=1,\ S=1$

Теперь, возвращаясь к постановке задачи разбора служебных заголовков, можно определить функцию $\boldsymbol{F}(r_i)$ перехода для конечного автомата, анализирующего заголовки, следующим образом:

$$\boldsymbol{F}(r_i)=\begin{cases} 0, \text{ если } \boldsymbol{n} <= \alpha \text{ (текущий заголовок является последним в цепочке или поддельным;} \\ 1, \text{ если } \boldsymbol{n} > \alpha \text{ (текущий заголовок соответствует промежуточной точке генерации почтового сообщения).} \end{cases}$$

где α — параметр, задаваемый оператором системы и влияющий на точность определения источника рассылки.

Для усовершенствования метода поиска источника и обхода некоторых ограничений введем дополнительно список узлов, назовем их *доверенными узлами*, о которых априорно известно, что они могут использоваться другими узлами для пересылки почты. Примером таких узлов могут служить серверы почтовых систем типа MAIL.RU, GMAIL, РОСНТА.RU. При этом если заголовок был добавлен доверенным узлом, то функция перехода для конечного автомата, анализирующего заголовки, — $\boldsymbol{F}(r_i)=1$.

Контроль работы метода определения источника несанкционированной рассылки проводился на двух тестовых выборках, состоявших из 104 почтовых сообщений, отнесенных к категории «спам» и 95 сообщений категории «вирусы» без предварительного обучения системы. Параметр α для функции перехода был выбран равным 4. Для более полного охвата возможных комбинаций каждый набор заголовков присутствовал в выборке в единственном экземпляре. Для тестовой выборки использовались сообщения, состоящие из двух заголовков «Received», что могло означать, что либо сообщение проходило через два почтовых сервера, либо первый заголовок фиктивный. Все сообщения тестовых выборок были проверены на наличие фиктивных заголовков оператором «вручную». Результаты работы данного метода изложены в таблице 4.

Таблица 4. Результаты работы метода определения источника несанкционированной выборки.

Тип сообщения	**«Ручной» анализ**	**Работа метода**
Сообщения, попавшие в категорию «спам»	в 84 сообщениях найдены фиктивные заголовки	в 94 сообщениях найдены фиктивные заголовки; 10 заголовков (в 10 сообщениях), определенные алгоритмом как фиктивные, таковыми не являлись
Сообщения, попавшие в категорию «вирусы»	фиктивных заголовков не найдено	47 заголовков (в 47 сообщениях), определенные алгоритмом как фиктивные, не являлись таковыми

Для данной тестовой выборки все найденные по приведенному выше методу источники рассылки участвовали в ней, т. е. среди них не было узлов, не принимавших участия в несанкционированной рассылке (подставных узлов из фиктивных заголовков).

Заключение

Описанные в статье методы выявления источника несанкционированной рассылки сообщений электронной почты можно использовать при построении автоматизированных систем, предназначенных для решения следующих задач:

- определение вероятного маршрута следования почтового сообщения от предполагаемого нарушителя;
- определение сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама посредством протокола SMTP;
- сбор информации об участвующем в распространении вредоносных программ сетевом узле, которая может быть использована для идентификации (разоблачения) нарушителя;
- противодействие путем информирования провайдера о нелегальных действиях с идентифицированного сетевого узла (компьютера пользователя);
- противодействие путем изменения прав доступа или пропускной способности на сетевых устройствах (межсетевых экранах, маршрутизаторах);
- противодействие посредством разрыва TCP-соединения.

Информация о сетевых узлах, участвующих в распространении вредоносного программного обеспечения или спама посредством протокола SMTP, несомненно, будет важна при решении как задачи нейтрализации вирусных эпидемий, так и других задач, связанных с распространением вирусов, а системы, предоставляющие такую информацию, могут с успехом применяться как в глобальных, так и в локальных сетях для обеспечения более эффективной работы системы информационной безопасности.

СПИСОК ЛИТЕРАТУРЫ

1. Почтовые вирусы становятся все умнее // http://www.ivnet.ru/modules/news/article.php?storyid=412.
2. http://www.viruslist.com/ru/viruses/analysis.
3. Suspected zombie kings who ran botnet of 100,000 PCs arrested, reports Sophos // http://www.sophos.com/pressoffice/news/articles/2005/10/va_dutchbotarrests.html.
4. Spam Slayer: Slaying Spam-Spewing Zombie PCs // http://www.pcworld.com/article/id,121381-page,1/article.html.
5. RFC 2821 // ftp://ftp.rfc-editor.org/in-notes/rfc2821.txt.