На правах рукописи

РУДИК КИРИЛЛ ПЕТРОВИЧ

ИССЛЕДОВАНИЕ СПОСОБОВ ВЫЯВЛЕНИЯ СЕТЕВЫХ УЗЛОВ, УЧАСТВУЮЩИХ В НЕСАНКЦИОНИРОВАННОЙ РАССЫЛКЕ СООБЩЕНИЙ ЭЛЕКТРОННОЙ ПОЧТЫ

Специальность 05.13.19 – методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Автор:

Москва — 2009

Работа выполнена в Московском инженерно-физическом институте (государственном университете).

Научный руководитель кандидат технических наук, доцент
Петров Вячеслав Александрович

Официальные оппоненты: доктор технических наук, профессор
Дворянкин Сергей Владимирович

кандидат технических наук
Шарков Анатолий Евгеньевич

Ведущая организация: ОАО "Центральный научно-исследовательский институт радиоэлектронных систем"

Защита состоится «13» мая 2009 г. в 16:30 на заседании диссертационного совета ДМ 212.130.08 при Московском инженерно-физическом институте (государственном университете) по адресу: 123557, Москва, Пресненский Вал, 17, Центр информационных технологий и систем органов исполнительной власти (ЦИТиС).

С диссертацией можно ознакомиться в библиотеке Московского инженерно - физического института (государственного университета).

Автореферат разослан "___" апреля 2009 г.

Ученый секретарь
диссертационного совета Горбатов В.С.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

В настоящее время информационные технологии во многом определяют успех в деятельности организаций любого уровня от небольших предприятий до общенациональных государственных и коммерческих структур. Информация становится все более критичным ресурсом, а ее изменение, хищение или уничтожение могут привести к большим потерям. По данным специалистов в области компьютерной экономики и информационной безопасности, основную угрозу в настоящее время представляют компьютерные вирусы.

Наиболее благоприятной средой для распространения вредоносного программного обеспечения является сеть Интернет. Из возможных путей распространения в сети Интернет вирусы активно используют каналы электронной почты. Об этом свидетельствуют отчёты, предоставляемые компаниями, занимающимися антивирусной защитой.

Существует ещё один важный связующий элемент, объединяющий вредоносные программы и электронную почту, – спам (незапрошенная массовая рассылка электронной почты). Тенденция к объединению вирусных технологий с технологиями анонимной незапрошенной массовой рассылки электронной почты сегодня приобретает всё большую значимость. По оценкам экспертов, в рассылке спама, в основном, участвуют сети компьютеров обычных пользователей, превращённых в компьютеры-зомби посредством использования уязвимостей или других вирусных технологий. По данным различных компаний, занимающихся проблемами спамовых рассылок, процент спамовых сообщений, рассылаемых с использованием компьютеров-зомби, превышает 80% от всего объёма рассылаемого спама.

Следовательно, правильно определённые источники рассылки спама могут с высокой достоверностью указывать на инфицированные вредоносным ПО компьютеры.

Одним из способов 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.

Для выявления источника несанкционированной рассылки необходимы детальный анализ и проверка информации, находящейся в служебных заголовках сообщений электронной почты. В связи с большими и постоянно возрастающими объёмами почтового трафика *ручной анализ и проверка такой информации оказываются крайне неэффективными*.

Резюмируя вышеизложенное, можно сделать заключение, что *одним из возможных способов локализации инфицированных вредоносным ПО компьютеров является определение источников рассылки почтовых*

сообщений (о которых априорно известно, что они являются спамом или содержат вредоносный код) путём анализа их служебных заголовков. Следовательно, разработка специализированных моделей, методов и алгоритмов, предназначенных для автоматизации процесса определения источников рассылки почтовых сообщений, является **актуальной задачей**.

Информация о сетевых узлах, участвующих в распространении вредоносных программ или спама по электронной почте, несомненно, будет важна как при решении задачи нейтрализации вирусных эпидемий, так и для решения 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 систем информационной безопасности.

Объектом исследования данной работы являются проблемы безопасности информации, связанные с распространением вредоносных программ и спама, а также механизмы рассылки электронных сообщений, использующие сокрытие сетевых узлов отправителя.

Предметом исследования являются методы определения источников несанкционированной рассылки вредоносных программ и спама с учётом возможного присутствия в теле сообщения фиктивных данных об отправителе.

Целью работы является повышение эффективности методов борьбы с вредоносными программами на основе своевременного автоматизированного определения очагов вирусной активности, которая проявляется в виде рассылки ими сообщений электронной почты, содержащих известное (антивирусным системам и системам антиспамовой защиты) вредоносное ПО или спам, а также создание автоматизированной системы, реализующей эти методы.

Для достижения поставленной цели в ходе работы над диссертацией были проведены исследования существующих методов несанкционированной рассылки почтовых сообщений и анализ методов, используемых для сокрытия источника несанкционированной рассылки, а также **решались следующие научные задачи**:

1. Построение математической модели выявления фальсификации служебных заголовков почтового сообщения, включающей формализацию условий присутствия фиктивных заголовков.
2. Разработка метода выявления источника несанкционированной рассылки сообщений электронной почты с учётом возможного наличия фиктивной информации в служебных заголовках и его алгоритмического обеспечения.
3. Разработка способов сбора информации о несанкционированной рассылке почтовых сообщений от различных источников и её анализа.
4. Разработка алгоритма принятия решений о действиях, имеющих целью противодействие распространению вредоносных программ.

Методы исследования

Для решения поставленных задач использовались системный анализ, теория алгоритмов, теория множеств, методы математической логики, теория конечных автоматов.

Научная новизна результатов, полученных в диссертации, состоит в следующем:

1. Предложен, проанализирован и реализован новый подход к обеспечению информационной безопасности компьютерных систем, основанный на своевременном определении очагов активности вредоносных программ, проявляемой в виде рассылки электронных почтовых сообщений, содержащих спам или вирусы, с целью последующего воздействия на них.
2. Показано, что в общем случае источник почтовой рассылки в рамках протокола SMTP определить невозможно. Определены и обоснованы условия, при которых возможно решение данной задачи.
3. Впервые разработана математическая модель выявления фальсифицированных (фиктивных) служебных заголовков почтового сообщения. Модель основана на выполненной математической формализации условий, указывающих на присутствие фиктивных заголовков. Показано, что с помощью введённого в модели критерия устанавливается, является заголовок истинным или фиктивным.
4. Впервые предложен метод анализа заголовка электронного почтового сообщения с целью локализации сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама, с учётом возможного присутствия в нём фальсифицированных данных.
5. Впервые разработаны концептуальная модель и функциональная структура программного обеспечения автоматизированной системы выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты.

Практическую ценность представляют разработанное алгоритмическое обеспечение автоматизированной системы, а также полученные в диссертационной работе рекомендации по выбору численного значения степени достоверности заголовка, от которого зависит точность выявления сетевых узлов, участвующих в несанкционированной рассылке сообщений электронной почты.

Автоматизированная система выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты (далее АСВСУ), была разработана на основе комплексного подхода с учетом всей доступной информации, связанной с решением данной задачи. Основными возможностями автоматизированной системы являются:

- сбор и хранение электронных почтовых сообщений, попавших в категорию вирусов и спама, для последующего анализа конфликтных ситуаций (например, «антиспамовая» система ошибочно приняла сообщение за спам, и важное для пользователя письмо было удалено);
- обнаружение сетевых узлов, инфицированных как известными, так и не известными ранее вредоносными программами, путём анализа активности этих программ, которая проявляется в виде рассылки сообщений электронной почты, содержащих спам или известные (антивирусным системам) вредоносные программы;
- сбор информации о сетевом узле, участвующем в распространении вредоносных программ, которая необходима для идентификации владельца скомпрометированного сетевого узла;
- уведомление пользователя/владельца скомпрометированного сетевого узла об имеющей место несанкционированной деятельности с его компьютера;
- информирование провайдера о нелегальных действиях с идентифицированного сетевого узла (компьютера пользователя);
- формирование на основе IP-адресов источников рассылки «чёрных списков» для систем фильтрации электронной почты;
- противодействие путём изменения прав доступа или пропускной способности на сетевых устройствах (межсетевых экранах, маршрутизаторах).

Внедрение результатов работы

Результаты диссертационной работы внедрены в:

- ФГУП «Ситуационно-Кризисный Центр Государственной корпорации по атомной энергии «РОСАТОМ». Результаты диссертационной работы использовались при создании и внедрении в информационно-вычислительную сеть Росатома автоматизированной системы сбора и анализа данных аудита для систем «антивирусной» и «антиспамовой» фильтрации электронной почтовой корреспонденции.
- Московском инженерно-физическом институте (государственном университете). Результаты диссертационной работы использовались при создании и внедрении в информационно-вычислительную сеть факультета «Информационная безопасность» автоматизированной системы сбора и анализа данных аудита для систем «антивирусной» и «антиспамовой» фильтрации электронной почтовой корреспонденции – «СЛЕДОПЫТ».

Апробация работы

Основные методические и практические результаты исследований докладывались на следующих конференциях и выставках:

1. XI Всероссийская научно-техническая конференция "Проблемы информационной безопасности в системе высшей школы" – Москва, 2004г.
2. XIV Общероссийская научно-техническая конференция. Методы и технические средства обеспечения безопасности информации» С-Петербург СПбГПУ 2005г.
3. XIII Всероссийская научно-техническая конференция "Проблемы информационной безопасности в системе высшей школы" – Москва, 2006г.
4. XIV Всероссийская научно-техническая конференция "Проблемы информационной безопасности в системе высшей школы" – Москва, 2007г.
5. XI Выставка-конференция "Телекоммуникации и новые информационные технологии в образовании" – Москва, 2007г.
6. XVII Общероссийская научно-техническая конференция. Методы и технические средства обеспечения безопасности информации» С-Петербург СПбГПУ 2008г.

За разработанную в процессе работы над диссертацией автоматизированную систему сбора и анализа данных аудита для систем «антивирусной» и «антиспамовой» фильтрации электронной почты автор был награждён дипломом XI-ой Выставки-конференции "Телекоммуникации и новые информационные технологии в образовании".

Публикации

По теме диссертации опубликованы 6 научных работ.

Основные положения, выносимые на защиту:

1. Математическая модель выявления фальсифицированных (фиктивных) служебных заголовков почтового сообщения.
2. Классификация ключевых признаков, указывающих на присутствие фиктивных заголовков.
3. Алгоритм определения источников несанкционированной рассылки.
4. Метод определения заражённых вредоносными программами сетевых узлов путём анализа заголовков распространяемых ими электронных почтовых сообщений, содержащих инфицированные вложения.
5. Методы противодействия несанкционированной рассылке.
6. Концептуальная модель автоматизированной системы выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты.

7. АСВСУ как средство: активного информационного противодействия угрозам нарушения информационной безопасности; обеспечения внутреннего аудита и мониторинга каналов электронной почты с целью обнаружения инфицированных вредоносными программами сетевых узлов; защиты от потери информации в связи с некорректной обработкой почтовых сообщений системами «антивирусной» и «антиспамовой» фильтрации электронной почтовой корреспонденции (удаление писем при ложных срабатываниях).

Объем и структура

Диссертация состоит из введения, пяти глав, заключения, списка использованной литературы из 103 наименований и приложений. Основная работа диссертации содержит 133 страницы текста, включая 22 рисунка и 8 таблиц. Объем диссертационной работы с учетом приложений составляет 180 стр.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования, определяются и формируются цели и задачи исследования, научная новизна и практическая ценность, положения, выносимые на защиту, а также результаты внедрения.

В первой главе «Несанкционированная рассылка сообщений электронной почты» проводится анализ классификации вредоносных программ, способов их распространения и обнаружения, а также определяются угрозы, связанные с их распространением; оценивается роль электронной почты в распространении вредоносных программ; даётся обзор методов сбора информации о сетевых узлах, участвующих в несанкционированной рассылке, и способов противодействия таким рассылкам.

Проведённый анализ существующих путей распространения вредоносного программного обеспечения показал, что наиболее благоприятной средой для их распространения является сеть Интернет. При этом из всех возможных путей распространения в сети Интернет вредоносные программы для своего распространения в большинстве случаев используют каналы электронной почты.

Важным связующим элементом, объединяющим вредоносные программы и электронную почту, является спам (незапрошенная массовая рассылка электронной почты). Тенденция к объединению вирусных технологий с технологиями анонимной незапрошенной массовой рассылки электронной почты сегодня приобретает всё большую значимость. По оценкам экспертов по информационной безопасности, в большинстве случаев рассылка спама осуществляется с использованием сетей компьютеров обычных пользователей, превращённых посредством

использования уязвимостей в ПО или использования вирусных технологий в компьютеры-зомби. По данным различных компаний, занимающихся проблемами спамовых рассылок, процент спамовых сообщений, рассылаемых с использованием компьютеров-зомби, может превышать 80% от всего объёма спама. Компьютеры-зомби дают возможность авторам спама не только рассылать миллионы писем, но и скрывать истинные источники рассылок, оставаясь безнаказанными.

Исходя из вышесказанного, можно сделать вывод, что правильно определённые источники рассылки спама могут с высокой достоверностью указывать на инфицированные вредоносным ПО компьютеры.

В силу специфики протокола передачи электронных почтовых сообщений использование систем обнаружения/предотвращения вторжений (IDS/IPS), а также систем антивирусной фильтрации почтового трафика и анти-спамовых фильтров для определения источников несанкционированной рассылки может оказаться неэффективным. При этом, чем больше процент сообщений, пришедших в систему не напрямую (через промежуточные почтовые сервера), тем меньше эффективность традиционных способов обнаружения источника нарушений.

Для выявления источника несанкционированной рассылки необходимы детальный анализ и проверка информации, находящейся в служебных заголовках сообщений электронной почты, причём, из-за весьма больших объемов почтового трафика, «ручные» анализ и проверка такой информации оказываются крайне неэффективными.

Представленные на рынке средств информационной безопасности системы, способные выполнять функции сбора и анализа информации о несанкционированных почтовых рассылках с целью определения источников рассылки, имеют определённые недостатки, а именно:

- в качестве адреса отправителя может быть определён только узел, с которого было получено сообщение. Отсутствует процедура анализа служебных заголовков Received с целью более «глубокого» поиска источника, если сообщение прошло через несколько почтовых систем;
- не предусмотрена возможность ответной реакции системы на инциденты безопасности;
- тяжеловесность системы (необходимость для обеспечения работы системы отдельного почтового сервера организации);
- закрытость системы для модификации и относительно высокая стоимость.

Разработка автоматизированной системы, предназначенной для сбора и анализа информации о несанкционированных почтовых рассылках с целью определения источников рассылки, и совершения каких-либо ответных действий по отношению к источнику рассылки, является перспективным направлением развития средств сетевой защиты,

направленных на решение задач, связанных с распространением вредоносного программного обеспечения.

Следует отметить, что в открытых источниках отсутствует информация об автоматизированных системах, которые имеют своим назначением сбор и анализ информации о несанкционированных почтовых рассылках с целью определения источников рассылки, **с учётом возможного присутствия в теле сообщения фиктивных данных об отправителе**, и принятия каких-либо действий по отношению к источнику рассылки.

Во второй главе «Исследование методов несанкционированной рассылки сообщений электронной почты» рассмотрены методы рассылки вредоносных программ и спама; проведён сравнительный анализ возможностей программного обеспечения, используемого для рассылки незапрошенной электронной почтовой корреспонденции, на предмет выявления особенностей, связанных с передачей сообщений.

С целью определения основных технологий, используемых при рассылке вредоносных программ и спама, было проведено исследование возможностей программного обеспечения, используемого для рассылки незапрошенной электронной почтовой корреспонденции. Исследование показало, что такое программное обеспечение может обладать следующей возможностями:

- рассылка посредством подключения непосредственно к почтовому серверу получателя с использованием встроенного почтового клиента;
- рассылка через почтовый сервер, определённый политикой, действующей в рамках узла, с которого ведётся рассылка, или сети, которой он принадлежит;
- рассылка через заданный почтовый сервер или шлюз;
- рассылка через посредника с применением протоколов, отличных от SMTP (прокси-сервера) или не полностью соблюдающих его (например, без добавления заголовка Received или добавления ложных заголовков).

Установлено, что основным способом распространения спама и вредоносного ПО (более 86%) является рассылка путём непосредственного подключения к почтовому серверу получателя с использованием встроенного почтового клиента. Результат получен путём анализа способов рассылки вредоносных программ и спама на тестовой выборке сообщений, содержащих спам и вредоносное ПО.

Выполнен анализ действий, совершаемых для сокрытия источника рассылки вредоносных программ и спама, который показал, что основными методами, используемыми для того, чтобы скрыть сетевой адрес, с которого отсылается электронное почтовое сообщение, являются:

- добавление фиктивных заголовков Received;

- использование общедоступных шлюзов (Open Relay);
- использование общедоступных PROXY- серверов или шлюзов с изменением среды передачи (например, почтовые серверы, работающие через Web интерфейс);
- использование «компьютеров-зомби».

Технологии передачи почтовых сообщений, в совокупности с методами, используемыми при рассылке вредоносных программ и спама, не дают гарантий однозначного определения источника, сгенерировавшего почтовое сообщение. Таким образом, определение источника в общем случае становится невозможным. Однако, в ряде случаев, определить источник рассылки возможно.

Исходя из текущих тенденций развития методов несанкционированной рассылки, можно сделать следующие заключения:

- почти весь объём спама исходит от компьютеров «зомби»;
- наиболее распространенным способом сокрытия источника рассылки является добавление фиктивных заголовков «Received».

<u>В третьей главе</u> «Метод выявления источника несанкционированной рассылки сообщений электронной почты» предложен метод выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама посредством протокола SMTP; разработан алгоритм анализа SMTP заголовка электронного почтового сообщения на предмет поиска в нём фиктивных данных; определены параметры, анализируя которые можно сделать предположение о присутствие в конкретном сообщении фиктивных заголовков.

Метод заключается в последовательном анализе цепочки служебных заголовков почтового сообщения. Истинность или фиктивность каждого заголовка устанавливается в соответствии с разработанной математической моделью выявления фальсифицированных служебных заголовков. Математическая модель основана на выполненной формализации условий, указывающих на присутствие фиктивных заголовков, и применении, введённого в модели, параметра, который назван степенью достоверности заголовка. Численное значение степени достоверности заголовка зависит от того, какие из имеющихся условий присутствия фиктивных заголовков выполнены.

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.

Анализ этой 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.

В общем случае, с учётом возможности присутствия ложных заголовков, набор служебных заголовков «Received» можно представить следующим образом (см. рис. 1):

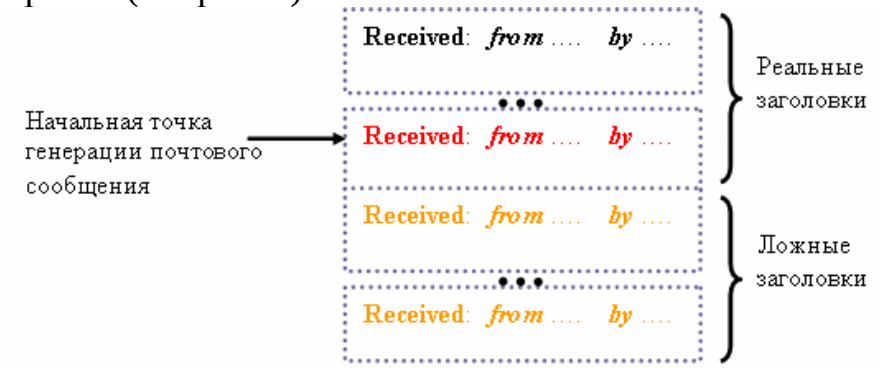

Рис. 1. Набор служебных заголовков «Received».

Для описания решения задачи анализа последовательности заголовков «Received», с целью поиска фиктивных заголовков, используется теория конечных автоматов.

Конечный автомат представим в виде формальной системы $M=(U,\sum,\delta,u_0,L)$, где U — конечное непустое множество состояний; $\sum$ — конечный входной алфавит; δ — отображение типа $U \times \sum \to U$; $u_0 \in U$ — начальное состояние; $L \subseteq U$ — множество конечных состояний.

Входной алфавит $\sum = \{w_i\}$ может принимать следующие условные значения:

0 – найден поддельный заголовок;

1 – текущий заголовок соответствует промежуточной точке генерации почтового сообщения;

2 – текущий заголовок является последним в последовательности заголовков.

Следовательно, входной алфавит является конечным множеством и $\sum=\{0,1,2\}$.

Для анализа двух последовательных заголовков из множества заголовков $R=\{r_i\}$ введём дополнительно функцию F, анализирующую информацию, находящуюся в заголовках, такую, что:

$$F(r_i, r_{i+1}) = \begin{cases} 0, \text{ если следующий заголовок является поддельным;} \\ 1, \text{ если следующий заголовок не является поддельным;} \\ \text{текущий заголовок соответствует промежуточной точке генерации почтового сообщения.} \end{cases}$$

Рассмотрим приведенную на рисунке 2 диаграмму состояний конечного автомата, анализирующего последовательность заголовков

«Received». Входной алфавит конечного автомата $M=(U,\sum,\delta,u_0,L)$ равен $\sum=\{0,1,2\}$, $U=\{u_0,u_1\}$, $L=\{u_0\}$, $\delta(u_0,1)=u_0$, $\delta(u_0,0)=u_1$, $\delta(u_0,2)=u_1$.

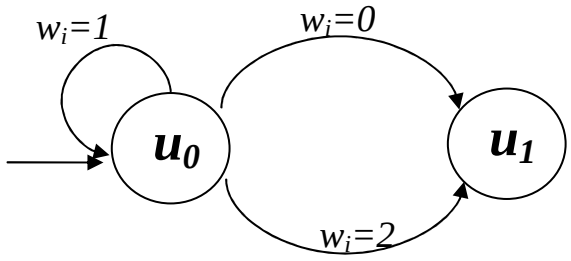

Рис. 2. Диаграмма состояний конечного автомата.

На каждом шаге автомат рассматривает значение функции F (если текущий заголовок не является последним), 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 r_{i+1} – заголовок является первым, встретившимся фиктивным, то заголовок r_i соответствует начальной точке генерации почтового сообщения.

В ходе анализа заголовков почтовых сообщений разных категорий (спам, рассылка вредоносных программ, обычные почтовые сообщения) эмпирическим путём был введён ряд параметров, проверка значений которых, в определённых случаях, позволит определить, является рассматриваемый заголовок «Received» истинным или поддельным. Этими параметрами являются:

- R – соответствие заголовка формату (формат заголовков «Received» должен удовлетворять требованиям протокола – RFC 2821 и RFC 2822);
- D_{MX}, D_A – для доменного имени, указанного в заголовке в качестве MTA, должна быть соответствующая запись «MX» или «A»;
- L – доменное имя, указанное в заголовке в качестве MTA, должна быть не выше третьего уровня;
- C_R – связность двух последовательных заголовков «Received»;
- C_{fb} – связность полей *by* и *from* одного заголовка «Received»;
- S – проверка доступности из сети Интернет почтовой службы (TCP/25, TCP/110, TCP/143, TCP/465) на сервере, указанном в качестве MTA.

Определим следующий набор параметров $Y=\{R,C_R,C_{fb},D_{MX},D_A,L,S\}$, где каждый из параметров может принимать значение {0,1} в зависимости от выполняемости соответствующего условия.

В связи с тем, что в ряде случаев однозначное определение источника рассылки не представляется возможным, было решено ввести

некоторый параметр N, значение которого от значений параметров из множества $\boldsymbol{Y}=\{\boldsymbol{R},\boldsymbol{C_R},\boldsymbol{C_{fb}},\boldsymbol{D_{MX}},\boldsymbol{D_A},\boldsymbol{L},\boldsymbol{S}\}$ и который характеризует степень достоверности указанных в заголовке служебных данных. При этом любому набору параметров из множества Y ставится в соответствие натуральное число. Чем больше значение этого параметра, тем больше достоверность указанных в заголовке данных. Следует отметить, что термин «степени достоверности» используемый в работе не связан с термином «статистическая значимость».

Для задания значения параметра $\boldsymbol{N}$ для каждого набора $\boldsymbol{Y}=\{\boldsymbol{R},\boldsymbol{C_R},\boldsymbol{C_{fb}},\boldsymbol{D_{MX}},\boldsymbol{D_A},\boldsymbol{L},\boldsymbol{S}\}$ упорядочим наборы входящих в них параметров по степени влияния на достоверность данных, указанных в заголовке. И после упорядочивания пронумеруем наборы. Номера наборов будем использовать как значения параметра N.

Для упорядочивания наборов использовался метод ранжирования объектов по важности путём парного сравнения (метод парных сравнений), применяемый при анализе экспертных оценок.

После опроса экспертов и упорядочивания коэффициентов относительной важности объектов по возрастанию были получены искомые значения степени достоверности $\boldsymbol{N}$ для заданных наборов условий (таблица 1):

Таблица 1.

$\boldsymbol{N}$	**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	$R=1, C_R=1, D_{MX}=1$
5	$R=1, C_R=1, D_A=1\ S=1$
5	$R=1, C_R=1, D_A=1\ D_{MX}=1$
6	$R=1, C_R=1, L=1, S=1$
6	$R=1, C_R=1, D_A=1, L=1, S=1$
7	$R=1, C_R=1, D_{MX}=1, L=1$
8	$R=1, C_R=1, D_{MX}=1, D_A=1, L=1$
9	$R=1, C_R=1, D_{MX}=1, S=1$
9	$R=1, C_R=1, D_{MX}=1, D_A=1, S=1$
10	$R=1, C_R=1, D_{MX}=1, L=1, S=1$
10	$R=1, C_R=1, D_{MX}=1, D_A=1, L=1, S=1$

$\boldsymbol{N}$	**Набор условий**
11	$R=1, C_R=1, C_{fb}=1$
12	$R=1, C_R=1, C_{fb}=1, D_A=1$
13	$R=1, C_R=1, C_{fb}=1, L=1$
13	$R=1, C_R=1, C_{fb}=1, D_A=1, L=1$
14	$R=1, C_R=1, C_{fb}=1, S=1$
15	$R=1, C_R=1, C_{fb}=1, D_{MX}=1$
15	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, D_A=1$
15	$R=1, C_R=1, C_{fb}=1, D_A=1, S=1,$
16	$R=1, C_R=1, C_{fb}=1, L=1, S=1$
16	$R=1, C_R=1, C_{fb}=1, D_A=1, L=1, S=1$
17	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, L=1$
18	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, D_A=1, L=1$
19	$R=1, C_R=1, C_{fb}, D_{MX}, S=1$
19	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, D_A=1, S=1$
20	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, L=1, S=1$
20	$R=1, C_R=1, C_{fb}=1, D_{MX}=1, D_A=1, L=1, S=1$

Теперь, возвращаясь к постановке задачи разбора служебных заголовков, можно определить функцию, анализирующую информацию из заголовков – $\boldsymbol{F}(r_i, r_{i+1})$:

$$\boldsymbol{F}(r_i,r_{i+1})=\begin{cases} 0, \text{ если } \boldsymbol{N} < \boldsymbol{Q} \text{ (следующий заголовок является поддельным);} \\ 1, \text{ если } \boldsymbol{N} \geq \boldsymbol{Q} \text{ (следующий заголовок не является поддельным; текущий заголовок соответствует промежуточной точке генерации почтового сообщения)} , \end{cases}$$

где Q – введённый параметр, названный критерием степени достоверности данных, указанных в заголовке. Критерий степени достоверности Q выполняет функцию порогового значения, определяющего минимальное значение параметра N при котором заголовок считается не поддельным. Критерий задаётся оператором системы и влияет на точность определения источника рассылки.

Для усовершенствования метода поиска источника и обхода некоторых ограничений введём дополнительно список узлов, которые назовём *доверенными узлами* и о которых априорно известно, что они могут использоваться другими узлами для пересылки почты. Вероятность подделки заголовков этими системами считается незначительной, так как может являться только следствием взлома этих систем. Примером таких узлов могут служить серверы почтовых систем типа MAIL.RU, GMAIL, POCHTA.RU. При этом, если заголовок был добавлен доверенным узлом, то функция перехода для конечного автомата, анализирующего заголовки, принимается равной 1, т.е. $\boldsymbol{F}(r_i,r_{i+1}) = 1$.

Рассмотрим алгоритм поиска источника несанкционированной рассылки.

С целью упрощения процедуры определения начальной точки формирования почтового сообщения необходимо провести подготовительные операции, заключающиеся в предварительной очистке служебных заголовков путём удаления избыточной информациии и не участвующих в анализе заголовков (заголовки Return-Path, Received-SPF, Date, From, To, и т.п.).

Для определения узла, который участвует в рассылке и находится наиболее «близко» к источнику рассылки (или является источником), используются следующие правила:

1. Первый заголовок Received (добавляемый почтовой системой на стороне получателя) является истинным и IP-адрес, указанный в поле «from» этого заголовка, становится решением по умолчанию – $\mathrm{IP}(from_1)$.
2. Если в сообщении присутствует больше одного заголовка, то их список просматривается до конца или до появления первого поддельного заголовка. IP-адрес, указанный в поле «from»

последнего не поддельного заголовка, объявляется решением – IP($from_i$), где i – номер последнего, не поддельного заголовка.

Рисунок 3 иллюстрируют принцип определения источника несанкционированной рассылки или узла, наиболее близкого к нему в соответствии с предложенным алгоритмом.

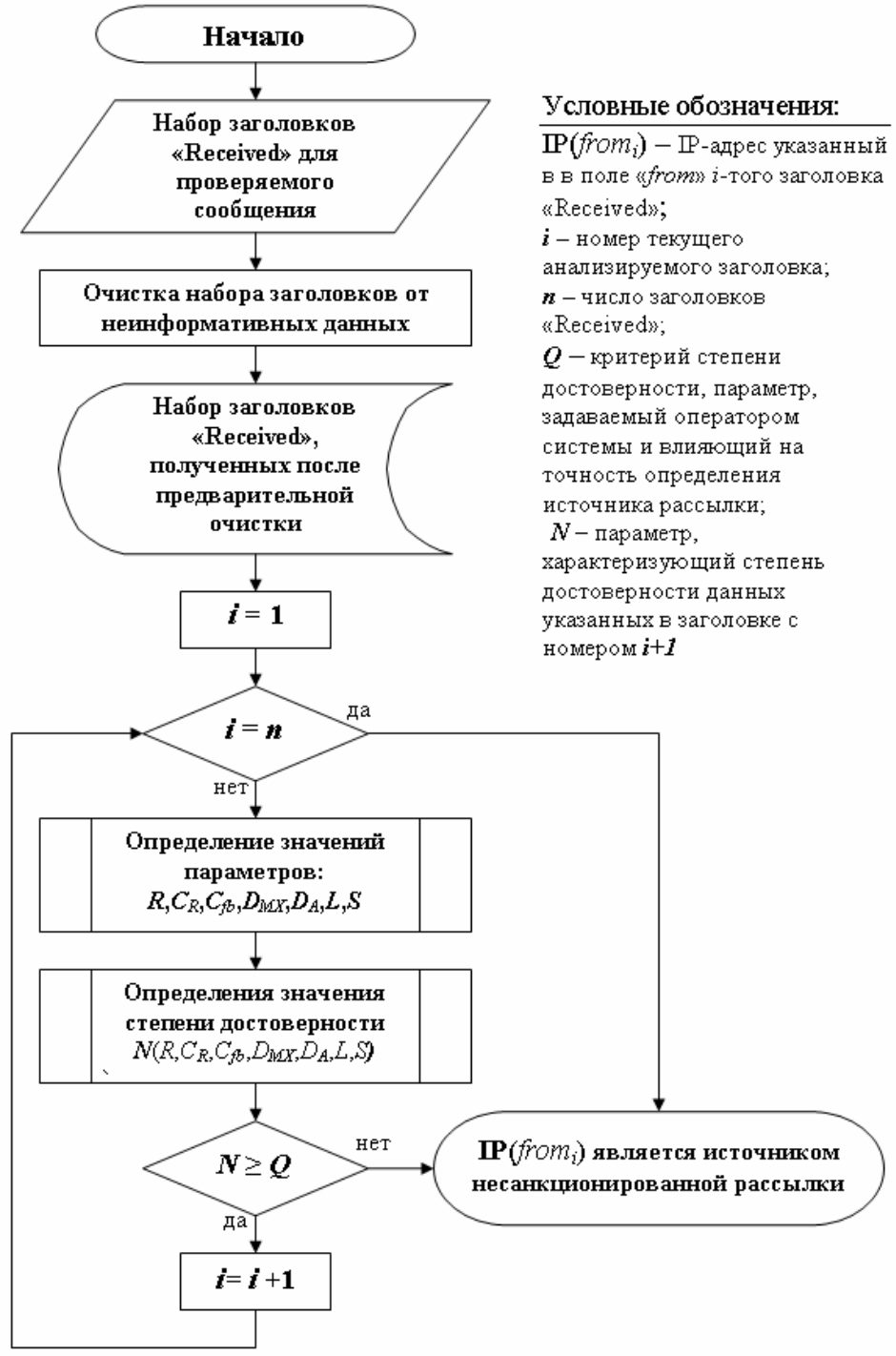

Рисунок 3. Алгоритм определения источника несанкционированной рассылки.

Выходными данными для предложенного алгоритма будет IP-адрес узла, о котором можно однозначно сказать, что он *участвовал* в передаче исходного почтового сообщения.

Для успешной работы алгоритма необходимо выполнение следующих условий:

- заголовок Received, добавленный последним (на принимаемой стороне), является истинным;
- вероятность подделки заголовков почтовыми системами (доверенными узлами) считается незначительной;
- все почтовые MTA добавляют запись «Received» в набор заголовков;
- формат заголовков должен удовлетворять требованиям протокола (RFC 2821 и RFC 2822);
- на узле, являющемся распространителем вредоносных программ или спама, не должно быть работающих почтовых служб или их эмуляторов, прослушивающих TCP порты 25,110,143.

В четвёртой главе «Автоматизированная система выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты» рассматриваются вопросы, связанные с разработкой концептуальной модели, функциональной структурой и ограничениями, реализацией модели на базе алгоритмического обеспечения. Также приведено описание разработанной автоматизированной системы, её основные характеристики, проведено тестирование системы на тестовой выборке почтовых сообщений и даны методические рекомендации по улучшению её работы.

Модель автоматизированной системы выявления сетевых узлов представим состоящей из следующих четырёх функциональных блоков:

- подсистема сбора данных о нарушениях безопасности;
- подсистема определения начальной точки генерации почтового сообщения (а также сбора дополнительной информации о нём);
- подсистема выбора и реализации ответных действий (по отношению к начальной точке следования пакетов);
- подсистема аудита и подготовки отчётов.

Упрощением модели системы выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты, является то обстоятельство, что найденный узел может являться промежуточным.

Основные ограничения модели являются следствием ограничений алгоритма определения источника несанкционированной рассылки.

Модель автоматизированной системы реализована в виде программного решения для операционных систем семейства Windows (2000/XP/2003).

Схема функциональной структуры разрабатываемой концептуальной модели приведена на рисунке 4.

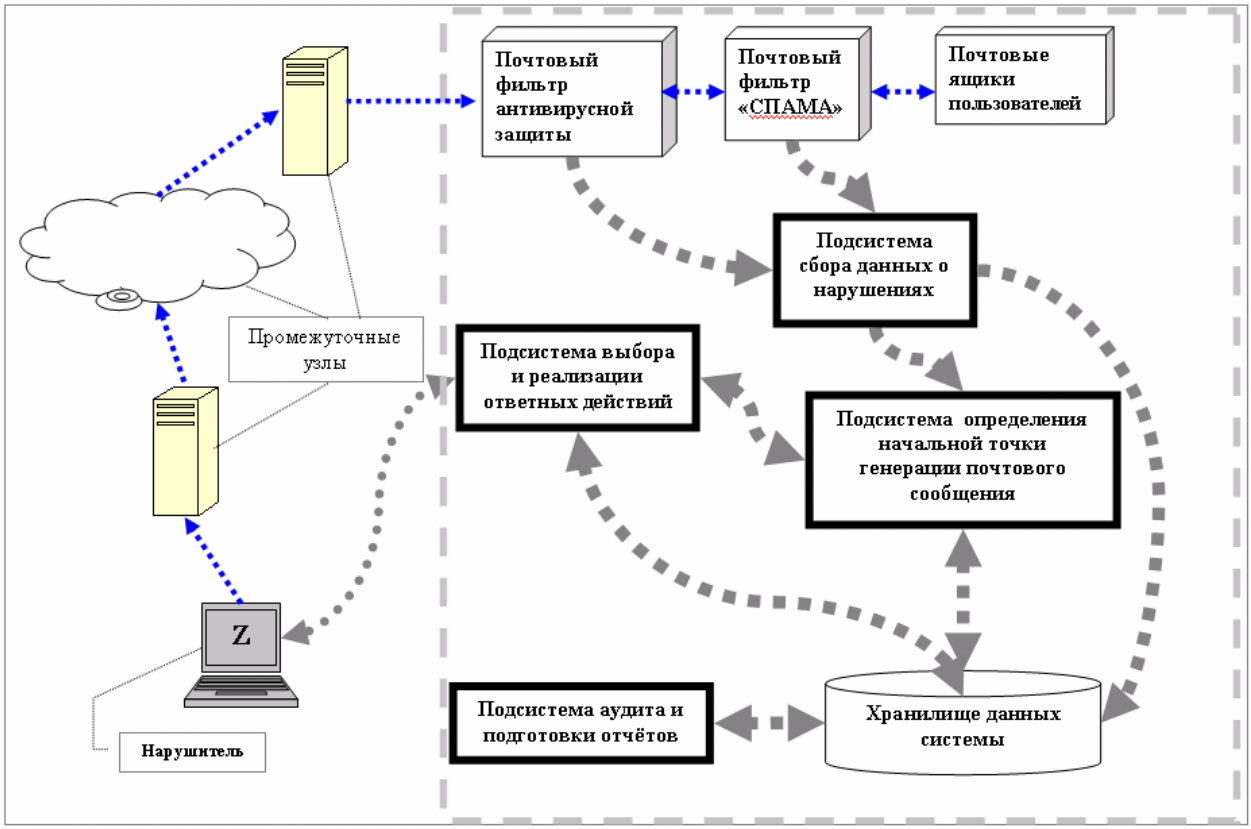

Рис. 4. Функциональная структура концептуальной модели.

В процессе выявления источника рассылки используется разработанный алгоритм определения фиктивных заголовков, в котором основным регулируемым параметром является критерий степени достоверности ***Q.*** С целью определения оптимального значения данного критерия была рассмотрена выборка заголовков «Received». 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, попавшие в тестовую выборку, были проверены на наличие фиктивных заголовков «вручную». Проверка выборки, состоящей из **247** сообщений, показала:

- в 65% (162) сообщений присутствовали, а в 33% (81) сообщений отсутствовали фиктивные заголовки;
- в 2% (4) сообщений определить присутствие (отсутствие) фиктивных заголовков не удалось.

Оптимальное значение критерия ***Q*** определялось, исходя из следующих положений:

- количество заголовков, неправильно классифицированных как фиктивные, но являющихся истинными, должно быть минимальным;
- должны отсутствовать заголовки, неправильно классифицированные как истинные, но являющиеся фиктивными.

Итоговые данные, собранные на основании анализа тестовой выборки, приведены в таблице 2.

Таблица 2.

Q	**Удовлетворяют условию $N \geq Q$**	**Ложно отброшенные / в “( )” – не отнесены ни в какую категорию**	**Ложно принятые / в “( )” – не отнесены ни в какую категорию**
0	247	0 (0)	162 (4)
1	116	4 (0)	35 (4)
2	101	4 (1)	21 (3)
3	83	5 (2)	5 (2)
4	65	18 (2)	0 (2)
5	65	18 (2)	0 (2)
6	56	25 (4)	0 (0)
7	36	45 (4)	0 (0)
8	36	45 (4)	0 (0)
9	32	49 (4)	0 (0)
10	32	49 (4)	0 (0)
11	27	54 (4)	0 (0)
12	25	56 (4)	0 (0)
13	23	58	0 (0)
14	19	62	0 (0)
15	18	63	0 (0)
16	13	68	0 (0)
17	3	78	0 (0)
18	2	79	0 (0)
19	2	79	0 (0)
20	2	79	0 (0)

Рассмотрим диаграмму, приведённую на рис.3, которая построена на основании таблицы 2:

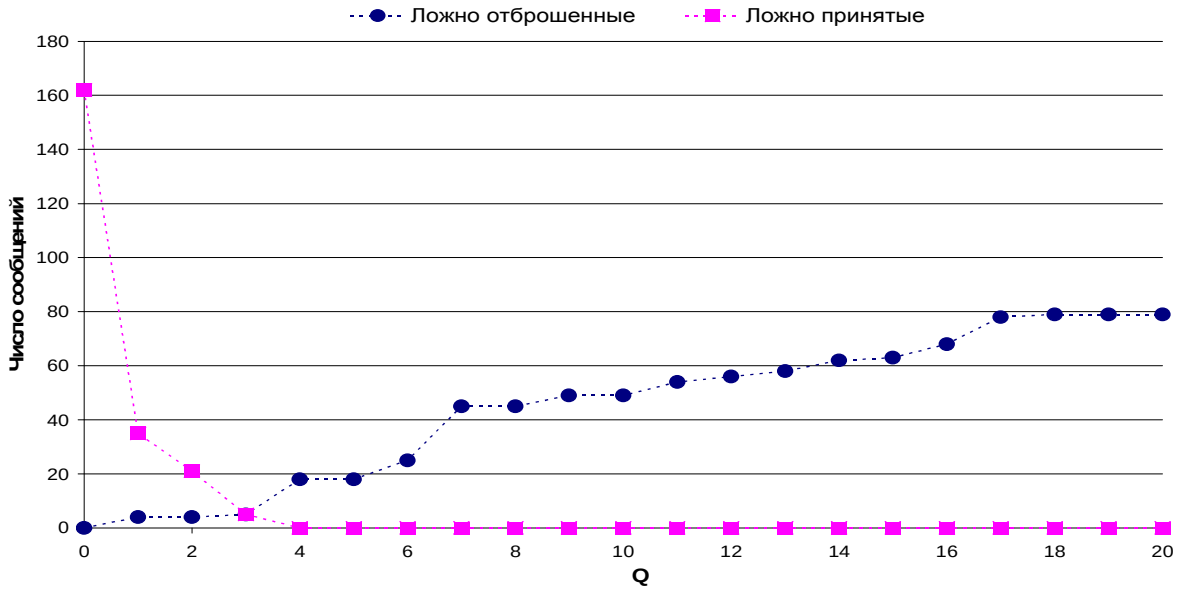

Рис. 3. Зависимость числа неправильно классифицированных заголовков от значения критерия степени достоверности ***Q***.

Из этой диаграммы видно, что для данной тестовой выборки число заголовков, неправильно классифицированных как истинные, стремится к 0 при значении критерия $\boldsymbol{Q} \geq 4$.

Проверка работы автоматизированной системы на тестовой выборке показала её эффективность в плане правильности определения источника несанкционированной рассылки и времени реакции. Число заголовков, неправильно классифицированных как истинные, но являющихся фиктивными, стремится к 0 при значении критерия $\boldsymbol{Q} \geq 4$. При выборе значения критерия $\boldsymbol{Q}$ равным 4 число заголовков, неправильно классифицированных как фиктивные, но являющихся истинными, не превышает 8%, а заголовки, неправильно классифицированные как истинные, отсутствуют.

Дальнейшее улучшение работы системы предлагается проводить по следующим основным направлениям:

- добавление дополнительных модулей;
- увеличение производительности системы;
- обеспечение безопасности информационных потоков внутри системы при разделении её на несколько программно-аппаратных модулей.

Первое направление – это добавление дополнительных модулей, которые реализуют новые механизмы сбора информации, а также выбора и реализации ответных действий по отношению к скомпрометированному сетевому узлу.

Задачу увеличения производительности системы, т.е. числа сообщений, обрабатываемых системой в единицу времени, можно решать различными способами. В первую очередь предполагается провести оптимизацию исходного кода модулей автоматизированной системы. Вместе с тем, желателен перевод исходного кода с интерпретируемого языка программирования (в настоящие время – PERL) на компилируемый (например C/C++). Это позволит значительно увеличить производительность системы, даже не рассматривая вопросы оптимизации исходного кода.

Для увеличения производительности целесообразно также разнести функциональные модули на независимые аппаратные платформы, каждая из которых будет настроена на выполнение задач конкретного модуля.

При разделении системы на несколько программно-аппаратных модулей станет актуальной задача обеспечения безопасности информационных потоков внутри системы. Для решения задачи защиты информационных потоков системы можно предложить использование технологий VPN на основе криптографической защиты передаваемых данных.

В заключении приведены результаты и выводы, полученные автором в ходе выполнения работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

В процессе выполнения работы были получены нижеследующие основные результаты.

1. Показано, что в силу специфики протокола передачи электронных почтовых сообщений использование систем типа IDS/IPS, а также системы антивирусной фильтрации почтового трафика и анти-спамовских фильтров для определения источников несанкционированной рассылки является неэффективным.
2. Предложен, проанализирован и реализован новый подход к обеспечению информационной безопасности компьютерных систем, основанный на своевременном определении очагов активности вредоносных программ, проявляемой в виде рассылки электронных почтовых сообщений, содержащих спам или вирусы, с целью последующего воздействия на них.
3. Показано, что невозможно, в общем случае, определение источника почтовой рассылки в рамках протокола SMTP.
4. Впервые разработана математическая модель выявления фальсифицированных (фиктивных) служебных заголовков почтового сообщения. Отличительной особенностью модели является введение понятия степени достоверности заголовка и условий, указывающих на присутствие фиктивных заголовков, формализованных математически.
5. Впервые предложен метод анализа заголовка электронного почтового сообщения с целью локализации сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама, с учётом возможного присутствия в нём фальсифицированных данных.
6. Впервые разработан алгоритм определения фиктивных данных в служебных заголовках электронного почтового сообщения, основанный на анализе значений заданных параметров.
7. Предложен алгоритм определения параметра «показатель корреляции заголовков», входящего в алгоритм определения фиктивных данных в служебных заголовках.
8. Впервые разработана концептуальная модель и функциональная структура программного обеспечения автоматизированной системы выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты.
9. На основе комплексного подхода с учетом всей доступной информации, связанной с решением данной задачи, реализована автоматизированная система выявления сетевых узлов, участвующих в распространении вредоносного программного обеспечения и спама в системах электронной почты. Автоматизированная система, реализующая предложенные методы анализа заголовков, позволила существенно снизить время, необходимое на разбор заголовков с целью

определения источников рассылки вредоносного программного обеспечения.

10. Определено оптимальное значение критерия степени достоверности заголовка на примере тестовой выборки. При выборе значения критерия Q равным 4 число заголовков, неправильно классифицированных как фиктивные, но являющихся истинными, не превышает 8%, а заголовки, неправильно классифицированные как истинные, отсутствуют.

Основные результаты диссертационной работы изложены в 6 печатных трудах:

1. Рудик К.П. Определение заражённых вредоносными программами сетевых узлов путём анализа заголовков распространяемых ими электронных почтовых сообщений, содержащих инфицированные вложения. // Сборник научных трудов конференции «XIV ОБЩЕРОССИЙСКАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ. Методы и технические средства обеспечения безопасности информации» С-Перербург СПбГПУ 2005 С. 99.
2. Рудик К.П. Построение активной сетевой защиты // Сборник научных трудов конференции «XI Всероссийская научно-техническая конференция. Проблемы информационной безопасности в системе высшей школы», Москва, 2004г. С. 130-131.
3. Рудик К.П. Способы сбора сетевой информации о нарушителе.// «Безопасность информационных технологий» М. МИФИ 2004. №2. С. 76-79.
4. Рудик К.П. Выявление источников рассылки вредоносного программного обеспечения при использовании протокола SMTP// Сборник научных трудов конференции «XIII Всероссийская научно-техническая конференция. Проблемы информационной безопасности в системе высшей школы», Москва, 2006г. С. 98-99
5. Рудик К.П. Выявление сетевых узлов, участвующих в распространении вредоносных программ и спама в системах электронной почты .// «Безопасность информационных технологий» М. МИФИ 2008. №1. С. 35-46.
6. Рудик К.П. «Исследование способов выявления сетевых узлов, участвующих в несанкционированной рассылке сообщений электронной почты»// Сборник научных трудов конференции XVII Общероссийская научно-техническая конференция. Методы и технические средства обеспечения безопасности информации» С-Петербург СПбГПУ 2008г.