РОССИЙСКАЯ ФЕДЕРАЦИЯ

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(19) **RU** (11) **2 586 853**(13) **C2**

(51) МПК
H04L 12/58 (2006.01)
G06Q 10/10 (2012.01)

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2013108772/08, 22.08.2011**

(24) Дата начала отсчета срока действия патента:
22.08.2011

Приоритет(ы):
(30) Конвенционный приоритет:
31.08.2010 US 12/872,691

(43) Дата публикации заявки: **10.09.2014** Бюл. № **25**

(45) Опубликовано: **10.06.2016** Бюл. № **16**

(56) Список документов, цитированных в отчете о поиске: **US 2005/0081059 A1, 14.04.2005. US 2006/0168055 A1, 27.07.2006. US 2010/0195503 A1, 05.08.2010. US 2004/0167964 A1, 26.08.2004. RU 2381551 C2, 10.02.2010.**

(85) Дата начала рассмотрения заявки PCT на национальной фазе: **27.02.2013**

(86) Заявка PCT:
US 2011/048664 (22.08.2011)

(87) Публикация заявки PCT:
WO 2012/030576 (08.03.2012)

Адрес для переписки:
129090, Москва, ул. Б. Спасская, 25, строение 3, ООО "Юридическая фирма Городисский и Партнеры"

(72) Автор(ы):
АНДЕРСЕН Ханс Кристиан (US),
ПАНАСЮК Анатолий (US),
РЕМАНИ Венката Соманадха Сарма (US),
КАС Барт (US)

(73) Патентообладатель(и):
МАЙКРОСОФТ ТЕКНОЛОДЖИ ЛАЙСЕНСИНГ, ЭлЭлСи (US)

(54) АДАПТИВНЫЙ ВЫБОР ПРАВИЛ СКАНИРОВАНИЯ ЭЛЕКТРОННЫХ СООБЩЕНИЙ

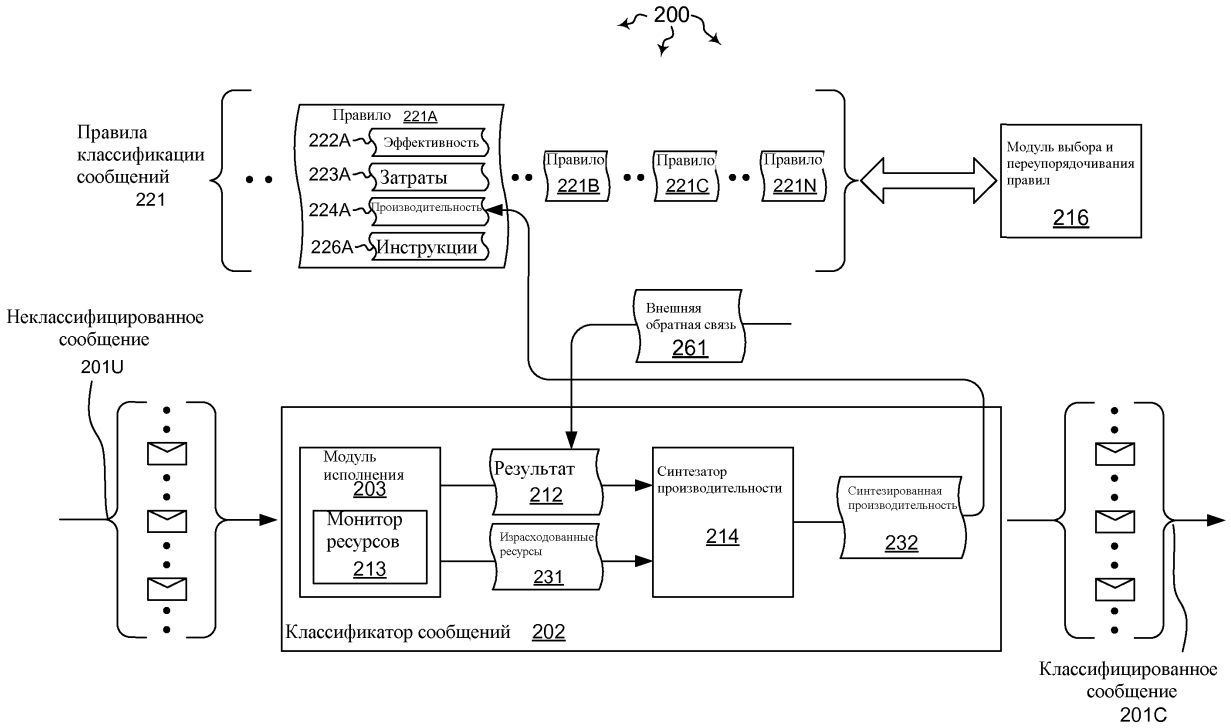
(57) Реферат:

Изобретение относится к способам и системе адаптивного выбора правил классификации электронных сообщений. Технический результат заключается в повышении эффективности выбора правил классификации электронных сообщений. Система содержит процессор, системную память, компьютерный накопитель, хранящий правила обнаружения SPAM и исполняемые инструкции, которыми реализуются классификатор сообщений и модуль выбора и переупорядочивания правил, при этом классификатор сообщений принимает сообщение электронной почты, применяет правило обнаружения SPAM, для которого вычисляет вероятность того, что сообщение электронной почты представляет собой SPAM, измеряет количество ресурсов, израсходованных на применение правила обнаружения SPAM и синтезирует показатель производительности из вычисленных результатов и измеренных затрат ресурсов для правила обнаружения SPAM, определяющий производительность классификации электронного сообщения как SPAM, при этом модуль выбора и переупорядочивания правил сравнивает синтезированный показатель производительности с существующими показателями

RU 2586853 C2

RU 2586853 C2

производительности для правил обнаружения SPAM, и выбирает новое подмножество правил обнаружения SPAM для использования в классификации последующих принятых сообщений электронной почты на основе результатов сравнения. 3 н. и 17 з.п. ф-лы, 2 табл., 5 ил.



Фиг. 2

RU 2586853 C2

RU 2586853 C2

RUSSIAN FEDERATION

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(19) **RU** (11) **2 586 853** (13) **C2**

(51) Int. Cl.
H04L 12/58 (2006.01)
G06Q 10/10 (2012.01)

(12) ABSTRACT OF INVENTION

(21)(22) Application: **2013108772/08, 22.08.2011**

(24) Effective date for property rights:
22.08.2011

Priority:
(30) Convention priority:
31.08.2010 US 12/872,691

(43) Application published: **10.09.2014** Bull. № **25**

(45) Date of publication: **10.06.2016** Bull. № **16**

(85) Commencement of national phase: **27.02.2013**

(86) PCT application:
US 2011/048664 (22.08.2011)

(87) PCT publication:
WO 2012/030576 (08.03.2012)

Mail address:
129090, Moskva, ul. B. Spasskaja, 25, stroenie 3, OOO "JUridicheskaja firma Gorodisskij i Partnery"

(72) Inventor(s):
ANDERSEN KHans Kristian (US),
PANASYUK Anatolij (US),
REMANI Venkata Somanadkha Sarma (US),
KAS Bart (US)

(73) Proprietor(s):
MAJKROSOFT TEKNOLODZHI LAJSENSING, ElElSi (US)

(54) ADAPTIVELY SELECTING ELECTRONIC MESSAGE SCANNING RULES

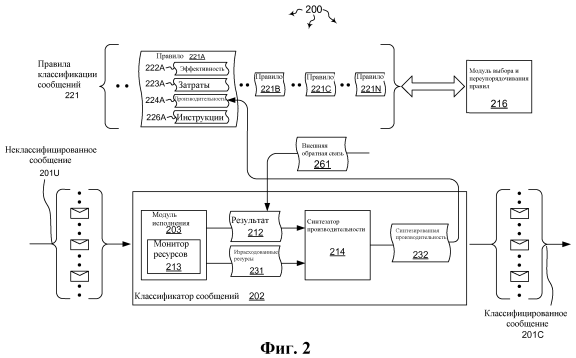
(57) Abstract:

FIELD: information technology.

SUBSTANCE: invention relates to methods and system for adaptively selecting electronic message scanning rules. System includes a processor, a system memory, a computer storage device that stores rules for scanning SPAM and executable instructions implementing a classifier messages and selection module and re-ordering of rules, when message classifier receives e-mail messages, applies SPAM scanning rules for which probability of receiving SPAM messages from e-mail is calculated, and measures amount of resources spent on application of SPAM scanning rules and synthetizes performance index from calculated results and measured resource overheads for SPAM scanning rules which determines performance of electronic message classification as SPAM, when module for selecting and re-ordering rules compares synthesised performance index with existing performance indexes for SPAM scanning rules, and selects new subset of SPAM scanning rules for using classification of next e-mail messages based on comparison results.

EFFECT: improving efficiency of electronic message classification rules.

20 cl, 2 tbl, 5 dwg



Фиг. 2

RU 2586853 C2
RU 2586853 C2

ПРЕДШЕСТВУЮЩИЙ УРОВЕНЬ ТЕХНИКИ

[0001] Компьютерные системы и связанные с ними технологии влияют на многие аспекты жизни общества. Действительно, способность компьютерных систем обрабатывать информацию изменила то, как мы живем и работаем. Компьютерные системы в настоящее время обычно выполняют множество задач (например, обработка текста, планирование, учет и т.п.), которые до появления компьютерных систем выполнялись вручную. В последнее время компьютерные системы были соединены друг с другом и другими электронными устройствами для формирования проводных и беспроводных компьютерных сетей, по которым компьютерные системы и другие электронные устройства могут передавать электронные данные. Соответственно, производительность множества вычислительных задач распределяется между несколькими различными компьютерными системами и/или несколькими различными вычислительными средами.

[0002] Во многих вычислительных средах электронные сообщения, такие как, например, сообщения электронной почты, используются для законного обмена информацией между пользователями компьютерных систем. Однако такие вычислительные среды также предоставляют пользователям невостребованные и/или нежелательные сообщения, часто именуемые как SPAM. Множество различных технологий было разработано для поиска и блокирования SPAM.

[0003] Технологии поиска SPAM должны, как правило, согласовать набор показателей, включающий в себя: эффективность, точность, производительность и время ожидания. Эффективность относится к тому, в какой мере SPAM может быть идентифицирован и остановлен. Точность к тому, в какой мере допустимые сообщения ошибочно идентифицируются как SPAM (например, доля ошибочных срабатываний). Производительность относится к потреблению ресурсов, связанному с идентификацией сообщения как SPAM или допустимого. Время ожидания относится к тому, на какое время каждое отдельное сообщение задерживается при передаче в результате сканирования.

[0004] Согласование между этими показателями является относительно сложной задачей, так как улучшение в одной области, как правило, означает деградацию в одной или нескольких других областях. Например, более агрессивное анти-SPAM обнаружение (повышенная эффективность) может вести к более высокому уровню ошибочных срабатываний (сниженная точность) и/или более высокой загрузке ЦПУ благодаря более сложным алгоритмам обработки (повышенное потребление ресурсов).

[0005] Кроме того, некоторая комбинация этих показателей часто отображается в Service Level Agreements ("SLA") - соглашениях об уровне обслуживания, которые поддерживаются поставщиком услуг. Например, поставщик анти-SPAM услуг может согласиться поддерживать эффективность не ниже, чем X, точность не ниже, чем Y, и т.д. Нарушение условий SLA, например обладание эффективностью меньше, чем X, в течение некоторого количества времени, для поставщика анти-SPAM услуг может быть поводом для некоторого денежного возмещения заказчику.

[0006] Однако в то же время службы анти-SPAM, как правило, сталкиваются с высокой изменчивостью нагрузки системы. Например, на протяжении любого конкретного дня, по выходным и сезонно, объем SPAM и/или объем допустимых электронных сообщений может изменяться. К сожалению, это может привести к тому, что поставщики услуг используют избыточное обеспечение. Например, общепринятое конструктивное решение заключается в том, чтобы построить службу поиска с достаточной мощностью, чтобы гарантировать SLA во время пиковой нагрузки, которая

может быть в три-пять раз выше, чем средняя нагрузка.

[0007] На практике результат проектирования под пиковую нагрузку относительно ресурсов бывает (потенциально существенно) недозагружен значительную часть времени. Сканирование, как правило, включает в себя фиксированное количество этапов и/или использование фиксированного количества правил сканирования с ограниченным, если таковой имеется, фактором доступных ресурсов. Таким образом, в непиковое время используется фиксированное количество этапов и/или правил для сканирования сообщения, даже если доступны ресурсы для дополнительного сканирования. Таким образом, проектирование под пиковую нагрузку представляет собой нежелательную форму перспективы затрат по продаже продукции, но тем не менее необходимую для целей SLA.

[0008] Дополнительные осложнения могут возникнуть, когда поддерживаются всевозможные различные уровни обслуживания, такие как, например, постоянные заказчики, премиальные заказчики, низкобюджетные заказчики и т.п., каждые из которых, как правило, с различными метриками, определенными в их SLA. Зачастую предложения премиумных услуг поступают с SLA, которые гарантируют наивысший уровень обслуживания (например, повышенная точность, меньшее время ожидания и т.д.), нуждающийся в больших вычислительных/процессорных со стороны поставщика услугах.

[0009] Одно конструктивное решение для обработки различных уровней обслуживания представляет собой для нас одну общую анти-SPAM услугу для всех уровней обслуживания. Каждый уровень обслуживания ограничен количеством правил и/или этапов обработки, через которые проходит сообщение, на основании требований SLA. Например, электронная почта премиальных заказчиков может проходить через десять этапов обработки, тогда как электронная почта базового заказчика может проходить только лишь через пять этапов обработки. Затраты на обслуживание базовых заказчиков снижены за счет более низкого качества сканирования (например, сниженной эффективности), даже когда ресурсы для дополнительного сканирования могут быть доступны. В дополнение к более низкому качеству сканирования, базовые заказчики также более подвержены целевым атакам за счет использования слабых мест в уровне защиты, предоставленном базовым заказчикам (предсказуемость системы).

[0010] Другое общеизвестное решение состоит в том, чтобы установить две отдельные системы, одну для премиальных заказчиков, а другую для базовых заказчиков. Каждая система спроектирована, чтобы сбалансировать качество услуг и затраты на услуги в соответствии с типом заказчика. К сожалению, этот тип системы нуждается в двойной инфраструктуре, что ведет к более высоким общим затратам, так же как и основная проблема использования избыточного обеспечения в целях соответствия SLA во время пиковой нагрузки.

КРАТКОЕ ИЗЛОЖЕНИЕ СУЩНОСТИ ИЗОБРЕТЕНИЯ

[0011] Настоящее изобретение распространяется на способы, системы и компьютерные программные продукты для адаптивного выбора правил сканирования электронных сообщений. В некоторых вариантах осуществления выбираются адаптивно правила, используемые для классификации электронных сообщений. Принимаются одно или несколько электронных сообщений. Для каждого одного или из нескольких электронных сообщений каждое правило классификации сообщений из предварительно выбранного подмножества правил классификации электронных сообщений применяется к данному электронному сообщению. Предварительно выбранное подмножество правил классификации электронных сообщений представляет собой подмножество

множества доступных правил классификации электронных сообщений.

[0012] Для каждого правила классификации электронных сообщений из предварительно выбранного подмножества правил классификации электронных сообщений вычисляется результат, который показывает вероятность того, что электронное сообщение обладает заданными характеристиками сообщения. Измеряются затраты ресурсов, показывающие количества ресурсов, потребляемых для применения правила классификации электронных сообщений к электронному сообщению. Сохраняются вычисленный результат и измеренные затраты ресурсов, связанные с применением каждого правила классификации электронной почты к каждому электронному сообщению.

[0013] Показатель производительности для каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации сообщений синтезируется из сохраненных вычисленных результатов и измеренных затрат ресурсов для данного правила классификации сообщений. Синтезированные показатели производительности сравниваются с существующими показателями производительности для правил классификации электронных сообщений, входящих во множество доступных правил классификации электронных сообщений. Выделяется новое подмножество правил классификации электронных сообщений из множества доступных правил классификации электронных сообщений на основании, по меньшей мере частично, результатов сравнения синтезированного показателя производительности с существующим показателем производительности. Новое подмножество правил классификации электронных сообщений предназначено для использования в классификации последующих принятых электронных сообщений. Соответственно, правила классификации сообщений могут вводиться и выводиться из использования для адаптации к изменяющимся образцам содержимого сообщения.

[0014] Данное краткое изложение сущности изобретения предоставляется для знакомства с подборкой концепций в упрощенной форме, которые дополнительно описаны ниже в подробном описании. Данное краткое изложение сущности изобретения не предназначено ни для определения ключевых признаков или существенных признаков заявленного изобретения, ни для использования в качестве помощи в определении объема заявленного изобретения.

[0015] Дополнительные признаки и преимущества изобретения будут установлены далее в описании, которое следует ниже, и частично будут очевидны из описания или могут быть изучены при осуществлении изобретения. Признаки и преимущества изобретения могут быть реализованы и получены посредством инструментов и комбинаций, подробно указанных в прилагаемой формуле изобретения. Эти и другие признаки настоящего изобретения станут очевидными в наиболее полной мере из нижеследующих описания и формулы изобретения или могут быть изучены при осуществлении изобретения, как изложено ниже.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0016] Для того, чтобы описать, каким образом вышеизложенные и другие преимущества и признаки изобретения могут быть получены, более подробное описание изобретения, кратко описанного выше, будет представлено со ссылками на конкретные варианты осуществления, которые показаны на прилагаемых чертежах. Понимая, что эти чертежи изображают только типичные варианты осуществления изобретения и поэтому не должны рассматриваться как ограничивающие его объем, изобретение будет описано и объяснено с дополнительными спецификой и детализацией посредством прилагаемых чертежей.

[0017] На фигуре 1 показан пример компьютерной архитектуры, которая обеспечивает адаптивную классификацию электронных сообщений.

[0018] На фигуре 2 показан пример компьютерной архитектуры, которая обеспечивает адаптивный выбор правил, использующихся для классификации электронных сообщений.

[0019] На фигуре 3 показан пример логической блок-схемы способа адаптивной классификации электронных сообщений.

[0020] На фигуре 4 показан пример логической блок-схемы способа адаптивного выбора правил, используемых для классификации электронных сообщений.

[0021] На фигуре 5 показан еще один пример компьютерной архитектуры, которая обеспечивает адаптивное сканирование электронных сообщений и адаптивный выбор правил, используемых для классификации электронных сообщений.

ПОДРОБНОЕ ОПИСАНИЕ

[0022] Настоящее изобретение распространяется на способы, системы и компьютерные программные продукты для адаптивного выбора правил сканирования электронных сообщений. В некоторых вариантах осуществления правила, используемые для классификации электронных сообщений, выбираются адаптивно. Принимаются одно или несколько электронных сообщений. Для каждого из одного или нескольких электронных сообщений каждое правило классификации сообщений из предварительно выделенного подмножества правил классификации электронных сообщений применяется к электронному сообщению. Предварительно выбранное подмножество правил классификации электронных сообщений представляет собой подмножество множества доступных правил классификации электронных сообщений.

[0023] Для каждого правила классификации электронных сообщений из предварительно выбранного подмножества правил классификации электронных сообщений вычисляется результат, который показывает вероятность того, что электронное сообщение обладает заданными характеристиками сообщения. Измеряются затраты ресурсов, показывающие количество ресурсов, потребляемых для применения правила классификации электронных сообщений к электронному сообщению. Сохраняются вычисленный результат и измеренные затраты ресурсов, связанные с применением каждого правила классификации электронной почты к каждому электронному сообщению.

[0024] Показатель производительности для каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации сообщений синтезируется из сохраненных вычисленных результатов и измеренных затрат ресурсов для данного правила классификации сообщений. Синтезированные показатели производительности сравниваются с существующими показателями производительности для правил классификации электронных сообщений, входящих во множество доступных правил классификации электронных сообщений. Выделяется новое подмножество правил классификации электронных сообщений из множества доступных правил классификации электронных сообщений на основании, по меньшей мере частично, результатов сравнения синтезированного показателя производительности с существующим показателем производительности. Новое подмножество правил классификации электронных сообщений предназначено для использования в классификации последующих принятых электронных сообщений. Соответственно, правила классификации сообщений могут вводиться и выводиться из состава услуги для адаптации к изменяющимся образцам содержания сообщения.

[0025] В других вариантах осуществления адаптивно классифицируются электронные сообщения. Электронное сообщение, отправленное от отправителя к получателю,

принимается в заданное время. Уровень обслуживания, применимый к принятому электронному сообщению, определяется на основе одного или более из отправителя и получателя.

[0026] Уровень обслуживания устанавливает по меньшей мере значение минимальной эффективности и ряд значений максимальных затрат на сканирование электронных сообщений. Значение минимальной эффективности отражает минимальную совокупную общую эффективность, при которой комбинация правил классификации сообщений должна удовлетворять уровню обслуживания. Каждое значение максимальных затрат из ряда значений максимальных затрат соответствует различным обозначенным временным диапазонам и отражает общее количество ресурсов, которые могут быть использованы для применения правил классификации сообщений к электронному сообщению. Максимальное значение затрат из ряда максимальных значений затрат выбирается для использования, когда сканирование принятого электронного сообщения основано на заданном времени в течение обозначенного временного диапазона для выбранного максимального значения затрат.

[0027] Одно или несколько правил классификации сообщений применяются к принятому электронному сообщению. Каждое правило классификации сообщений имеет измеренную эффективность, подсчитанные затраты ресурсов и производительность, вычисленную на основе измеренной эффективности с учетом подсчитанных затрат ресурсов. Измеренная эффективность отражает вероятность надлежащей идентификации электронного сообщения как имеющего заданные характеристики сообщения. Одно или несколько правил классификации сообщений применяются в порядке производительности до тех пор, пока не будет достигнута минимальная совокупная общая эффективность, определенная в уровне обслуживания.

[0028] Каждое правило классификации сообщений применяется к электронному сообщению, чтобы сгенерировать результат, показывающий вероятность того, что данное электронное сообщение обладает заданными характеристиками сообщения. Подсчитанные затраты ресурсов для примененного правила классификации сообщений добавляется к совокупному количеству израсходованных ресурсов. Совокупное количество израсходованных ресурсов вычисляется посредством суммирования подсчитанных затрат ресурсов ранее примененных правил классификации сообщений в одном или нескольких правилах классификации сообщений.

[0029] Определяют, меньше ли совокупное количество израсходованных ресурсов, чем выбранное максимальное значение затрат. Дополнительные правила классификации сообщений применяются к электронным сообщениям на основании данного определения. Когда количество израсходованных ресурсов меньше, чем выбранное максимальное значение затрат, больше правил электронных сообщений применяется к принятому электронному сообщению, в результате чего эффективность превышает заданную в уровне обслуживания. Когда количество израсходованных ресурсов по меньшей мере равно выбранному максимальному значению затрат, правила электронных сообщений применяются к еще одному другому электронному сообщению.

[0030] Варианты осуществления настоящего изобретения могут включать в себя или задействовать компьютер специального назначения или общего назначения, включающий в себя компьютерное аппаратное обеспечение, такое как, например, один или несколько процессоров и системная память, как рассмотрено более подробно ниже. Варианты осуществления в рамках объема настоящего изобретения также включают в себя физические и прочие считываемые компьютером носители информации для доставки или хранения исполняемых компьютером инструкций и/или структур данных.

Такие считываемые компьютером носители информации могут быть любыми доступными носителями информации, к которым может осуществлять доступ компьютерная система общего назначения или специального назначения. Считываемые компьютером носители информации, которые хранят исполняемые компьютером инструкции, представляют собой физические накопители. Считываемые компьютером носители информации, которые несут в себе исполняемые компьютером инструкции, представляют собой среды передачи данных. Таким образом, в качестве примера и без ограничения, варианты осуществления изобретения могут включать в себя по меньшей мере два отдельных различных вида считываемых компьютером носителей информации: компьютерные накопители (устройства) и среды передачи данных.

[0031] Компьютерные накопители (устройства) включают в себя RAM (ОЗУ), ROM (ПЗУ), EEPROM (ЭСППЗУ), CD-ROM или другой накопитель на оптических дисках, накопитель на магнитных дисках или другие магнитные накопительные устройства или любые другие средства, которые могут быть использованы для хранения необходимого программного кода в форме исполняемых компьютером инструкций или структур данных и которые могут быть доступны для компьютера общего назначения или специального назначения.

[0032] "Сеть" определена как одна или несколько линий передачи данных, которые позволяют перемещать электронные данные между компьютерными системами и/или модулями и/или другими электронными устройствами. Когда информация передается или предоставляется через сеть или другие коммуникационные соединения (как проводные, так и беспроводные или сочетание проводных и беспроводных) на компьютер, компьютер соответственно рассматривает соединение как среду передачи данных. Среда передачи данных может включать в себя сеть и/или линии передачи данных, которые могут быть использованы для передачи необходимых средств программного кода в форме исполняемых компьютером инструкций или структур данных и которые могут быть доступны для компьютера общего назначения или специального назначения. Комбинации вышеперечисленного также должны быть включены в область считываемых компьютером носителей информации.

[0033] Дополнительно, при досягаемости различных компьютерных системных компонентов, средства программного кода в форме исполняемых компьютером инструкций или структур данных могут быть переданы автоматически от сред передачи данных к компьютерным накопителям (устройствам) (или наоборот). Например, исполняемые компьютером инструкции или структуры данных, принятые по сети или линии передачи данных, могут быть буферизированы в RAM сетевого интерфейсного модуля (например, "NIC") и затем со временем перемещены в RAM компьютерной системы и/или менее энергозависимый компьютерный накопитель в компьютерной системе. Таким образом, следует понимать, что компьютерные накопители (устройства) могут быть включены в состав компонентов компьютерной системы, которые также (или даже в первую очередь) используют среды передачи данных.

[0034] Исполняемые компьютером инструкции включают в себя, например, инструкции и данные, которые во время исполнения в процессоре побуждают компьютер общего назначения, компьютер специального назначения или обрабатывающее устройство специального назначения к исполнению некоторой функции или группы функций. Исполняемые компьютером инструкции могут представлять собой, например, двоичные файлы, инструкции промежуточного формата, такого как язык ассемблера, или даже исходный код. Несмотря на то что изобретение описано на языке, характерном для структурных признаков и/или методологических действий, следует понимать, что

объем изобретения, определяемый в прилагаемой формуле изобретения, не обязательно ограничивается описанными функциями или действиями, описанными выше. Скорее, описанные признаки и действия раскрыты как пример формы реализации формулы изобретения.

[0035] Специалисты в данной области техники оценят, что изобретение может осуществляться в сетевом компьютерном окружении со многими типами конфигураций компьютерных систем, в том числе персональными компьютерами, настольными компьютерами, портативными компьютерами, обработчиками сообщений, портативными устройствами, многопроцессорными системами, микропроцессорной или программируемой бытовой электронной аппаратурой, сетевыми персональными компьютерами, миникомпьютерами, большими электронными вычислительными машинами, мобильными телефонами, карманными персональными компьютерами, пейджерами, маршрутизаторами, коммутационными устройствами и т.п. Изобретение также может осуществляться в окружениях распределенных систем, где локальная и удаленная компьютерные системы, которые соединены (либо по проводным линиям передачи данных, либо беспроводным линиям передачи данных, либо с помощью комбинации проводных и беспроводных линий передачи данных) по сети, обе выполняют задачи. В окружении распределенной системы программные модули могут быть расположены как на локальных, так и на удаленных запоминающих устройствах.

[0036] В основном, варианты осуществления изобретения относятся к динамически (и потенциально непредсказуемо) меняющейся глубине/полноте классификации электронных сообщений для защиты от нежелательного содержания сообщения (например, SPAM, вирусов, цифровой утечки и т.д.). Поддерживается минимальная эффективность и, когда имеющиеся ресурсы позволяют, она может быть превышена для предоставления повышенной защиты. Оптимальное подмножество доступных правил классификации сообщений может быть выбрано на основе каждого сообщения. Выбор правил базируется на доступных системных ресурсах, минимальной желаемой эффективности (например, определенной в соглашении об уровне обслуживания ("SLA") и характеристиках правил. Обратная связь может быть использована для оптимизации подмножеств правил классификации.

[0037] Таким образом, в рамках описания изобретения и последующей формулы изобретения "классификация сообщений" включает в себя классификацию электронных сообщений (например, сообщений электронной почты, сообщений службы коротких сообщений ("SMS"), файлов и т.п.) по различным "классам" на основе характеристик сообщения (или файла), таких как, например, содержимое, размер сообщения, прикрепленные файлы, доменные имена категории бизнес-потребитель, регион происхождения, отправитель, получатель, время, дата и т.д.

[0038] В некоторых вариантах осуществления электронное сообщение классифицируется, чтобы определить уровень обслуживания (например, согласно SLA), соответствующий электронному сообщению. Уровень обслуживания определяет дальнейшее применение правил классификации сообщений к электронному сообщению. Уровень обслуживания может задавать, каких типов и сколько других правил классификации сообщений должно быть применено к электронному сообщению. Например, правила классификации, которые особенно эффективны для классификации сообщений в одной стране, могут быть менее эффективны для классификации сообщений в другой стране, и наоборот.

[0039] В некоторых вариантах осуществления дополнительная классификация относится к определению, представляет ли собой электронное сообщение нежелательное

и/или невостребованное электронное сообщение (например, SPAM), либо электронное сообщение содержит вредоносную программу или, напротив, зараженную и/или опасную (например, вирусы, программу-шпион, троянский конь и т.п.), либо происходит утечка конфиденциальной информации в электронном сообщении и т.д. Например, система предотвращения утечек информации ("DLP") может использовать правила для определения того, включает ли в себя электронное сообщение конфиденциальную информацию.

[0040] На фигуре 1 показан пример компьютерной архитектуры 100, которая обеспечивает адаптивную классификацию электронного сообщения. Ссылаясь на фигуру 1, компьютерная архитектура 100 включает в себя классификатор 102 сообщений, определитель 107 уровня обслуживания, таймер 108, корректировочные коэффициенты 118, правила 121 классификации сообщений и соглашения 131 об уровне обслуживания. Каждый из проиллюстрированных компонентов соединен с другим по сети (или представляет собой ее часть), такой как, например, локальная сеть ("LAN"), глобальная сеть ("WAN") и даже Интернет. Соответственно, каждый из проиллюстрированных компонентов, как и любые другие соединенные компьютерные системы и их компоненты, могут производить данные, относящиеся к сообщениям, и обмениваться данными, относящимися к сообщениям (например, датаграммами протокола Интернет ("IP") и других протоколов высокого уровня, которые используют датаграммы IP, таких как протокол управления передачей данных ("TCP"), протокол передачи гипертекста ("HTTP"), упрощенный протокол пересылки электронной почты ("SMTP") и т.д.), по сети.

[0041] Правило 121 содержит множество правил классификации сообщений, таких как, например, правила с 121A по 121N, которые могут быть использованы для классификации электронных сообщений. Каждое правило может показывать эффективность, затраты, производительность и может включать в себя инструкции. Эффективность показывает, насколько правдоподобно правило идентифицирует сообщение, точно как, тем или иным образом, нежелательное на основании используемого типа сканирования. Например, эффективность правила для обнаружения SPAM может показывать, насколько правдоподобно правило выявляет SPAM без ложных срабатываний. Затраты показывают (например, предполагаемое) количество системных ресурсов, которые расходуются, когда исполняемый модуль выполняет инструкции правила. Производительность показывает, насколько эффективно правило, на основе эффективности с точки зрения потребления ресурсов. В некоторых вариантах осуществления производительность представляет собой частное от деления эффективности на затраты. Инструкции исполняются для формирования результата, относящегося к классификации электронного сообщения (например, определения того, представляет ли собой электронное сообщение SPAM, содержит ли оно вредоносную программу, содержит ли оно конфиденциальную информацию и т.п.).

[0042] В большинстве случаев классификатор 102 сообщений сконфигурирован классифицировать электронные сообщения на основании характеристик электронного сообщения. Как проиллюстрировано, классификатор 102 сообщений включает в себя модуль 103 исполнения, монитор 104 затрат и монитор 106 эффективности. Модуль 103 исполнения сконфигурирован исполнять инструкции (например, скрипты или другой исполняемый код), содержащиеся в полученном правиле. Инструкции производят индивидуальный результат, который может быть использован как точка данных для классификации электронного сообщения. Например, индивидуальный результат может показать, представляет ли собой электронное сообщение нежелательное и/или

невостребованное электронное сообщение (например, SPAM), заражено ли оно или опасно, содержит ли конфиденциальную информацию и т.д. Модуль 103 исполнения может накапливать индивидуальные результаты исполнения целого ряда различных правил. Классификатор 102 сообщений может использовать накопленные индивидуальные результаты для классификации сообщений.

[0043] Монитор 104 затрат сконфигурирован отслеживать текущие затраты ресурсов, связанные со сканированием электронного сообщения. Когда правила исполняются, монитор 104 затрат сохраняет общие затраты ресурсов для всех правил в отношении электронного сообщения. В некоторых вариантах осуществления, когда каждое правило исполняется, затраты на правило добавляются к затратам ресурсов всех прежде исполненных правил.

[0044] Монитор 106 эффективности сконфигурирован отслеживать текущую эффективность сканирования электронного сообщения. Когда правила исполняются, монитор 106 эффективности сохраняет общую эффективность для всех правил, исполняемых в отношении электронного сообщения. В некоторых вариантах осуществления, когда каждое правило исполняется, эффективность правила добавляется к эффективности всех прежде исполненных правил.

[0045] Соглашение 129 об уровне обслуживания содержит множество SLA, в том числе SLA 131. Каждое SLA включает в себя минимальную эффективность и одну или несколько затрат. Каждые затраты применимы к заданному диапазону даты/времени. Минимальная эффективность отражает накопленную эффективность (т.е. сумму эффективностей множества правил классификации), которая достигается в процессе сканирования сообщения (даже если потребление ресурсов превышено). Таблица 1 представляет собой пример эффективности в пересчете на SLA в зависимости от типа заказчика.

Таблица 1

Тип заказчика	Минимальная эффективность
Базовый заказчик	75
Премиальный заказчик	100

[0046] Таблица 1 показывает, что минимальная эффективность (т.е. накопленная эффективность, полученная в результате применения множества правил классификации) равна 75 для базовых заказчиков и 100 для премиальных заказчиков. Другие факторы также могут быть учтены при назначении минимальной эффективности в SLA.

[0047] Одна или каждая из нескольких затрат учитывает временной диапазон и максимальное потребление. Каждая пара временного диапазона/максимального потребления отображает, какие максимальные затраты ресурсов для применения правил рассматриваются для сообщения, когда сообщение принято в рамках временного диапазона. Пары временного диапазона/максимального потребления могут меняться или быть одинаковыми для разных уровней обслуживания. В некоторых вариантах осуществления пары временного диапазона/максимального потребления назначены в общедоступной таблице так, что пары временного диапазона/максимального потребления одинаковы для многих SLA. В других вариантах осуществления пары временного диапазона/максимального потребления могут быть назначены на основании SLA, например, путем включения в SLA. Таблица 2 представляет собой пример пары временного диапазона/максимального потребления.

Таблица 2	
Время суток	Максимальные затраты
Часы пик	50
Обычное время	75
Время непиковой нагрузки	100

[0048] Таблица 2 показывает, что максимальные затраты ресурсов на применение правил классификации в час пик составляют 50, в обычное время составляют 75, а во время непиковой нагрузки равны 100. Могут также быть учтены и другие факторы.

[0049] Максимальные затраты могут меняться со временем. Если на сервер классификации сообщений устанавливается дополнительное аппаратное обеспечение и таким образом достигается большая вычислительная мощность, возможность максимальных затрат может возрасти. С другой стороны, если к обслуживанию добавляются дополнительные заказчики или нагрузка внезапно увеличивается, возможность максимальных затрат может уменьшиться.

[0050] В некоторых вариантах осуществления минимальная эффективность рассматривается с более значительным отношением к максимальным затратам. В таких вариантах осуществления ресурсы с превышением максимальных затрат могут быть израсходованы, чтобы гарантировать, что минимальная эффективность достигнута. Если минимальная эффективность достигнута с использованием меньших ресурсов, чем максимальные затраты, то могут быть применены дополнительные правила классификации для повышения эффективности до тех пор, пока максимальные затраты не будет достигнуты или превышены.

[0051] Определитель 107 уровня обслуживания сконфигурирован определять уровень обслуживания, соответствующий принятому электронному сообщению. На основе характеристик сообщения и времени/даты определитель 107 уровня обслуживания может определить соответствующее SLA из соглашений 131 об уровне обслуживания. Таймер 108 может сохранять дату и время дня и отсылать эту информацию анализатору 107 уровня обслуживания, когда электронное сообщение принято. Определитель уровня обслуживания может отослать минимальную эффективность и максимальные затраты на сообщение классификатору 102 сообщений. Для каждого правила классификации сообщений классификатор 102 сообщений может сравнивать накопленную эффективность с максимальной эффективностью и суммарные затраты с максимальными затратами, чтобы определить, сколько и каких правил классификации применить к принятому сообщению.

[0052] Корректировочные коэффициенты 118 определяют некоторый процентный показатель того, что дополнительные правила классификации должны применяться к электронному сообщению, даже если минимальная эффективность уже достигнута, а максимальные затраты уже достигнуты или превышены. Корректировочные коэффициенты 118 позволяют правилам классификации, которые иначе были бы пропущены (например, в связи с их производительностью), выполняться время от времени. В некоторых вариантах осуществления корректировочные коэффициенты 118 показывают процентный показатель того, что каждое правило из правил 121 классификации сообщений должно применяться к электронному сообщению.

[0053] На фигуре 3 показана логическая блок-схема последовательности операций примера способа 300 адаптивной классификации электронного сообщения. Способ 300 будет описан в отношении компонентов и данных компьютерной архитектуры 100.

[0054] Способ 300 включает в себя действие приема электронного сообщения в указанное время, когда электронное сообщение посылается от отправителя к получателю (действие 301). Например, классификатор 102 сообщений может принять сообщение 101U во время 114 (как показывает таймер 108). Сообщение 101U может включать в себя характеристики 111 сообщения, в том числе адрес отправителя и адрес получателя.

[0055] Способ 300 включает в себя действие идентификации уровня обслуживания, применимого к принятому электронному сообщению, на основе одного или более из отправителя и получателя, уровня обслуживания, причем уровень обслуживания задает по меньшей мере значение минимальной эффективности и набор максимальных значений затрат, причем значение минимальной эффективности отражает минимальную накопленную общую эффективность, для которой комбинация правил классификации сообщений должна удовлетворять уровню обслуживания, причем каждое максимальное значение затрат из набора максимальных значений затрат соответствует отличающемуся от других обозначенному временному диапазону, каждое максимальное значение затрат отражает общее количество ресурсов, которое может быть использовано, чтобы применить правила классификации сообщений к электронному сообщению (действие 302). Например, определитель 107 уровня обслуживания может принимать характеристики 111 сообщения и время 114. На основании характеристик 111 сообщения (например, адресов отправителя и/или получателя) определитель 107 уровня обслуживания может идентифицировать SLA 131 как применимое для классификации сообщения 101U.

[0056] Как изображено, SLA 131 задает минимальную эффективность 132 и затраты 133. Затраты 133 включают в себя пары временного диапазона/максимальных затрат, в том числе временной диапазон 134A/максимальные затраты 136A, временной диапазон 134B/максимальные затраты 136B, временной диапазон 134C/максимальные затраты 136C и т.д.

[0057] Способ 300 включает в себя действие выбора максимального значения затрат из набора максимальных значений затрат, которое будет использовано во время сканирования принятого электронного сообщения на основании заданного времени в рамках временного диапазона, назначенного для выбранного максимального значения затрат (действие 303). Например, определитель 107 уровня обслуживания может определить, что время 114 находится в пределах временного диапазона 134A. В ответ на это определитель 107 уровня обслуживания может выбрать максимальные затраты 136A, которые будут использованы при сканировании неклассифицированного сообщения 101U.

[0058] Определитель 107 уровня обслуживания может переслать минимальную эффективность 132 и максимальные затраты 136A в классификатор 102 сообщений. Классификатор 102 сообщений может использовать минимальную эффективность 132 и максимальные затраты 136A, чтобы определить, когда остановить применение правил классификации сообщений к неклассифицированному сообщению 101U.

[0059] Способ 300 включает в себя действие применения одного или нескольких правил классификации сообщений к принятому электронному сообщению, причем каждое правило классификации сообщений обладает измеренной эффективностью, подсчитанными затратами ресурсов и вычисленной производительностью, основанной на измеренной эффективности с учетом подсчитанных затрат ресурсов, причем измеренная эффективность отражает вероятность того, что электронные сообщения будут должным образом классифицированы как обладающие заданными

характеристиками сообщений, при этом одно или несколько правил классификации сообщений применяются в порядке производительности, пока не будет достигнута минимальная накопленная общая эффективность, определенная в уровне обслуживания (действие 304). Например, классификатор сообщений может применять правила из набора правил 121 в порядке производительности, пока не будет достигнута минимальная эффективность 132 (т.е. 60).

[0060] На изображенных правилах возможно, что производительность 124A (т.е. 4) наиболее высока для правил из правил 121. Таким образом, правило 121A представляет собой первое правило, применяемое к неклассифицированному сообщению 101U. В результате применения правила 121A накопленная эффективность 162, равная 8, сравнивается с эффективностью 122A. Классификатор 102 сообщений определяет, что 8 меньше, чем 60, так что дополнительные правила классификации подлежат применению для достижения минимальной эффективности 132A.

[0061] Может быть, что производительность 124B (т.е., 3) является следующей наивысшей для правил из числа правил 121. Таким образом, правило 121B представляет собой следующее правило, применяемое к неклассифицированному сообщению 101U. В результате применения правила 121B накопленная эффективность 162 равна 11, что эквивалентно сумме эффективности 122A и эффективности 122B. Классификатор 102 сообщений определяет, что 11 меньше, чем 60, поэтому дополнительные правила классификации будут применяться для достижения минимальной эффективности 132A.

[0062] Может быть, что производительность 124C (т.е. 2,8) является наивысшей для правил из числа правил 121. Таким образом, правило 121C представляет собой следующее правило, применяемое к неклассифицированному сообщению 101U. В результате применения правила 121C накопленная эффективность 162 равна 65, что эквивалентно сумме эффективности 122A, эффективности 122B и эффективности 122C. Классификатор 102 сообщений определяет, что 65 больше, чем 60, поэтому дополнительные правила классификации не требуются, чтобы удовлетворить SLA 131.

[0063] Для каждого из одного или нескольких применяемых правил классификации сообщений способ 300 включает в себя действие применения правила классификации сообщений к электронному сообщению, чтобы выработать результат, показывающий вероятность того, что электронное сообщение обладает заданными характеристиками сообщения (действие 305). Например, модуль 103 исполнения может исполнить инструкции 126A в отношении неклассифицированного сообщения 101U, чтобы выработать результат 112. Результат 112 показывает вероятность того, что неклассифицированное сообщение 101U представляет собой нежелательное и/или невостребованное электронное сообщение, зараженное или опасное сообщение, содержит конфиденциальную информацию и т.д. Модуль 103 исполнения может сохранять результаты 112 в накопленных результатах 113. Инструкции 126B и 126C могут быть аналогичным образом исполнены в отношении неклассифицированного сообщения 101U, чтобы выработать результаты. Эти результаты могут также быть сохранены в накопленных результатах 113.

[0064] Каждого из для одного или нескольких применяемых правил классификации сообщений способ 300 включает в себя действие добавления подсчитанных затрат ресурсов для примененного правила классификации сообщений к совокупному количеству израсходованных ресурсов, причем совокупное количество израсходованных ресурсов вычисляется путем суммирования подсчитанных затрат ресурсов ранее примененных правил классификации сообщений из числа одного или нескольких правил классификации сообщений (действие 306). Например, в результате применения правил

121A, 121B и 121C накопленные затраты 161 равны 21, что эквивалентно сумме затрат 123A, затрат 123B и затрат 123C.

[0065] Способ 300 включает в себя действие определения того, меньше ли совокупное количество израсходованных ресурсов, чем выбранное максимальное значение затрат (действие 307). Например, монитор 104 затрат может определить, меньше ли накопленные затраты 161, чем максимальные затраты 136A. Способ 300 включает в себя действие применения дополнительных правил классификации сообщений к электронным сообщениям на основании данного определения (действие 308). Например, классификатор 102 сообщений может применить дополнительные правила классификации сообщений к электронным сообщениям на основании того, меньше ли накопленные затраты 161, чем максимальные затраты 136A, или нет.

[0066] Как проиллюстрировано в компьютерной архитектуре 100, в результате достижения минимальной эффективности 132A накопленные затраты 161 (т.е. 21) меньше, чем максимальные затраты 136A (т.е. 25). Таким образом, дополнительные правила классификации могут быть применены к неклассифицированному сообщению 101U, чтобы повысить эффективность классификации неклассифицированного сообщения 101U.

[0067] Например, может быть так, что производительность 124D (1,75) является наибольшей для правил из числа правил 121. Таким образом, правило 121D представляет собой следующее правило, применяемое к неклассифицированному сообщению 101U. Таким образом, модуль 103 исполнения может исполнять инструкции 126D в отношении неклассифицированного сообщения 101U, чтобы выработать результат и сохранить результаты в накопленных результатах 113. В результате применения правила 121D накопленные затраты 161 становятся равными 29, что эквивалентно сумме затрат 123A, затрат 123B, затрат 123C и затрат 123D. (Эффективность 122D, по существу, игнорируется после того, как минимальная эффективность 132A уже была достигнута.) После того как накопленные затраты 161 (т.е. 29) превысили максимальные затраты 136A (т.е. 25), никакие дополнительные правила не применяются к неклассифицированному сообщению 101U.

[0068] С другой стороны, если в результате достижения минимальной эффективности 132A после применения правила 121C накопленные затраты 161 сравнялись или превысили максимальные затраты 136A (т.е. 25), никакие дополнительные правила не применяются к неклассифицированному сообщению 101U. Тем не менее, минимальная эффективность 132A продолжает достигаться.

[0069] Когда минимальная эффективность 132A достигнута и максимальные затраты 136A достигнуты или превышены, классификатор 102 сообщений может обратиться к корректировочным коэффициентам 118. Классификатор 102 сообщений может использовать корректировочные коэффициенты 118 для определения того, должны ли дополнительные правила классификации быть применены к неклассифицированному сообщению 101U. Если да, то классификатор 102 сообщений применяет одно или несколько (или все оставшиеся) правила из числа правил 121 классификации сообщений, такие как, например, правило 121E, к неклассифицированному сообщению 101U. Использование корректировочных коэффициентов 118 позволяет оценить и соответственно изменить функциональные характеристики (например, эффективность и затраты) других неиспользуемых или ограниченно используемых правил классификации сообщений. На основании изменений частота использования таких правил классификации может вырасти. Например, оценка функциональных характеристик старого правила может показать, что старое правило стало сейчас более

эффективным в связи с изменением шаблонов SPAM.

[0070] Когда никакие дополнительные правила не применяются к неклассифицированному сообщению 101U, классификатор 102 сообщений может использовать накопленные результаты 113 для классификации неклассифицированного сообщения 101U. Например, исходя из накопленных результатов 113 классификатор 102 сообщений может классифицировать неклассифицированное сообщение 101U как допустимое сообщение или как нежелательное и/или невостребованное сообщение (например, SPAM), как включающее или не включающее в себя вредоносную программу, как включающее или не включающее в себя конфиденциальную информацию и т.п. Классификатор 102 сообщений может выдавать классифицированное сообщение 101C, чтобы показать классификацию.

[0071] В результате выдачи классифицированного сообщения 101C классификатор 102 сообщений может перейти к классификации следующего электронного сообщения.

[0072] На фигуре 2 показан пример компьютерной архитектуры 200, которая обеспечивает адаптивный выбор правил, используемых для классификации электронных сообщений. Ссылаясь на фигуру 2, компьютерная архитектура 200 включает в себя классификатор 202 сообщений, правила 221 классификации сообщений и модуль 216 выбора и переупорядочивания. Каждый из проиллюстрированных компонентов соединен друг с другом по сети (или представляет собой ее часть), такой как, например, локальная сеть ("LAN"), распределенная сеть ("WAN") и даже Интернет. Соответственно, каждый из проиллюстрированных компонентов, как и любые другие соединенные компьютерные системы и их компоненты, могут производить данные, относящиеся к сообщениям, и обмениваться данными, относящимися к сообщениям (например, датаграммами протокола интернет ("IP") и других протоколов высокого уровня, которые используют датаграммы IP, таких как протокол управления передачей данных ("TCP"), протокол передачи гипертекста ("HTTP"), упрощенный протокол пересылки электронной почты ("SMTP")), по сети.

[0073] Правило 221 содержит множество правил классификации сообщений, таких как, например, правила с 221A по 221N, которые могут быть использованы для классификации электронных сообщений. Аналогично правилам 121, каждое правило из числа правил 221 может показывать эффективность, затраты, производительность и может включать в себя инструкции.

[0074] В большинстве случаев классификатор 202 сообщений сконфигурирован классифицировать электронные сообщения на основании характеристик электронного сообщения. Например, классификатор сообщений может принять неклассифицированные сообщения 201U как входную информацию и выработать классифицированные сообщения 201C как выходную информацию. Каждое сообщение из классифицированных сообщений 201C может быть классифицировано, например, чтобы показать, представляет ли собой сообщение SPAM, содержит ли оно вредоносную программу, содержит ли оно конфиденциальную информацию и т.д.

[0075] Как проиллюстрировано, классификатор 202 сообщений включает в себя модуль 203 исполнения, дополнительно включая в себя монитор 213 ресурсов и синтезатор 214 производительности. Модуль 203 исполнения сконфигурирован исполнять инструкции (например, скрипты или другой исполняемый код), содержащиеся в полученном правиле. Инструкции производят индивидуальный результат (потенциально с задействованием внешней обратной связи с пользователями), который может быть использован как точка данных для классификации электронного сообщения. Монитор 213 ресурсов может отслеживать (например, по существу, в режиме реального

времени) количество различных израсходованных ресурсов (например, системной памяти, процессорного времени, пропускной способности сети и т.д.) во время исполнения правила.

[0076] Синтезатор 214 производительности может принимать результат и показатель израсходованных ресурсов и синтезировать и обновлять производительность для примененного правила. Результаты и израсходованные ресурсы для примененного правила могут также быть использованы для обновления эффективности и/или затрат по данному правилу для согласованности с синтезированной производительностью.

[0077] Таким образом, затраты и эффективность каждого правила классификации могут быть измеряемыми значениями, измеренными в определенный момент времени (например, при применении), и могут меняться со временем. По мере развития шаблонов и содержимого SPAM правило классификации может становиться более или менее эффективным. Если конкретная ставшая частью истории SPAM-кампания переживает возрождение в большом количестве, старое правило может внезапно стать более эффективным. Дополнительно, по мере того, как программное обеспечение усовершенствуется и оптимизируется, затраты на правило могут уменьшаться.

[0078] Модуль 216 выбора и переупорядочивания может выбирать правила из числа правил классификации сообщений 221 для применения к электронному сообщению (например, на основании производительности). Модуль 216 выбора и переупорядочивания также может сортировать правила классификации сообщений 221 (например, на основании производительности).

[0079] На фигуре 4 показана логическая блок-схема примерного способа 400 адаптивного выбора правил, использующихся для классификации электронных сообщений. Способ 400 будет описан в отношении компонентов и данных компьютерной архитектуры 200.

[0080] Способ 400 включает в себя действие приема одного или нескольких электронных сообщений (действие 401). Например, классификатор 202 сообщений может принимать неклассифицированные сообщения 201U.

[0081] Для каждого из одного или нескольких электронных сообщений способ 400 включает в себя действие применения каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации электронных сообщений к электронному сообщению, причем предварительно выбранное подмножество правил классификации электронных сообщений является подмножеством множества правил классификации электронных сообщений (действие 402). Например, классификатор 202 сообщений может применить правила 221A-221C к каждому сообщению из неклассифицированных сообщений 201U (например, на основании минимальной эффективности и максимальных затрат в SLA, а также, возможно, корректировочных коэффициентов).

[0082] Для каждого правила классификации электронных сообщений из предварительно выбранного подмножества правил классификации электронных сообщений способ 400 включает в себя действие, в котором правило электронного сообщения вычисляет результат, показывающий вероятность того, что электронное сообщение обладает заданными характеристиками сообщения (действие 403). Например, модуль 203 исполнения может исполнить инструкции 226A в отношении неклассифицированного сообщения из 201U, чтобы выработать результат 212. Результат 212 может показывать вероятность того, что сообщение из 201U представляет собой нежелательное электронное сообщение, представляет собой зараженное или опасное электронное сообщение, содержит конфиденциальную информацию и т.п. (например,

на основании назначенного получателя сообщения). Результаты для правил 221B и 221C также могут быть вычислены.

[0083] Внешняя обратная связь (например, от пользователя) может быть использована в вычисленном результате. Например, внешняя обратная связь 261 может быть использована в результате 212. Внешняя обратная связь может повышать или понижать вычисленную эффективность на основании пользовательского восприятия эффективности. Когда это целесообразно, внешняя обратная связь может также быть использована в вычисляемых результатах для правил 221B и 221C.

[0084] В некоторых вариантах осуществления электронные сообщения, содержащие непойманные SPAM, вредоносные программы или конфиденциальную информацию (ложноотрицательные результаты), так же как и допустимые сообщения, классифицированные как содержащие SPAM, вредоносные программы или конфиденциальную информацию (ошибочные срабатывания), подписываются на дополнительный анализ. Данный тип обратной связи может также быть использован для настройки баллов эффективности.

[0085] Для каждого правила классификации электронных сообщений из предварительно выбранного подмножества правил классификации электронных сообщений способ 400 включает в себя действие по измерению затрат ресурсов, которое показывает количество ресурсов, потребленных, чтобы применить правило классификации электронных сообщений к электронному сообщению (действие 404). Например, монитор 213 ресурсов может измерить затраты ресурсов, показывающие количество израсходованных ресурсов 231, затраченных при исполнении инструкций 226A в отношении сообщения из 201U. Затраты потребления ресурсов для правил 221B и 221C могут также быть измерены.

[0086] Способ 400 включает в себя действие по сохранению вычисленного результата и подсчитанных затрат ресурсов, связанных с применением каждого правила классификации электронной почты к каждому электронному сообщению (действие 405). Например, классификатор 202 сообщений может сохранить результат 212 и израсходованные ресурсы 231 вместе с затратами ресурсов на исполнение правила 221A в отношении других сообщений из неклассифицированных сообщений 201U. Результаты и затраты ресурсов на исполнение правил 221B и 221C в отношении сообщений из неклассифицированных сообщений 201U могут также быть сохранены.

[0087] Для каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации сообщений способ 400 включает в себя действие синтезирования показателя производительности из сохраненных вычисленных результатов и подсчитанных затрат ресурсов для правила классификации сообщений (действие 406). Таким образом, для каждого из правил 221A, 221B и 221C синтезатор 214 производительности может синтезировать показатель производительности из сохраненных вычисленных результатов и подсчитанных затрат ресурсов. Например, для правила 221A синтезатор 214 производительности может синтезировать синтезированную производительность 232 из результата 212 и израсходованных ресурсов 231, а также из вычисленных результатов и подсчитанных затрат ресурсов от применения правила 221A к другим сообщениям из неклассифицированных сообщений 201U. Производительности могут также быть синтезированы для правил 221B и 221C.

[0088] Затем классификатор 202 сообщений может заменить 224A синтезированной производительностью 232. Эффективность 222A и затраты 223A также могут быть обновлены по мере необходимости для согласования с синтезированной

производительностью 232. Производительности, эффективности и затраты для правил 221B и 221C также могут быть обновлены по мере необходимости.

[0089] Способ 400 включает в себя действие сравнения синтезированных показателей производительности с существующими показателями производительности для правил классификации электронных сообщений, содержащихся во множестве правил классификации электронных сообщений (действие 407). Например, синтезированная производительность 232 может быть сравнена с производительностями, заключенными в других правилах классификации сообщений 221. Синтезированные производительности для правил 221B и 221C также могут быть сравнены с производительностями, заключенными в других правилах классификации сообщений 221.

[0090] Способ 400 включает в себя действие выбора нового подмножества правил классификации электронных сообщений из состава множества правил классификации электронных сообщений для использования в классификации принимаемых в дальнейшем электронных сообщений на основании по меньшей мере части результатов сравнения синтезированных показателей производительности с существующими показателями производительности (действие 408). Например, на основании синтезированных производительностей правила 221A, 221B и 221C могут стать более или менее эффективными относительно друг друга, так же как и относительно других правил классификации сообщений 221. Таким образом, одно или несколько из правил 221A, 221B и 221C могут быть отброшены, когда выбирается новое подмножество правил (например, на основании SLA) для классификации электронных сообщений.

[0091] В некоторых вариантах осуществления затраты и эффективность для правил классификации сообщений непрерывно пересчитываются (например, в течение дня), по мере того как предпринимаются оперативные измерения в отношении относительной эффективности каждого правила по классификации сообщений (например, задержания SPAM, вредоносных программ, конфиденциальной информации и т.д.) и актуальных наблюдаемых затратах на выполнение правил. Для наиболее эффективных правил поступает больше данных об эффективности данных правил и затратах, так как они выполняются в отношении большего количества сообщений. Для менее эффективных правил корректировочный коэффициент (или случайная возможность), такой как, например, 1%, приводит к тому, что собирается по меньшей мере минимальное количество обновлений информации о затратах и эффективности. Пересчитываются затраты и эффективность и, следовательно, баллы производительности, которые используются, чтобы упорядочить правила. Последующие электронные сообщения классифицируются с использованием правил классификации сообщений, которые выбраны на основе обновленных баллов.

[0092] По мере того как дописываются новые правила (например, чтобы задержать новые типы SPAM или вредоносных программ), размер свода правил растет. Вновь введенные правила могут быть введены с баллом эффективности 0 и баллом затрат 1, что приводит к баллу производительности 0 и ставит правило в самый конец списка. Со временем, по мере того как правило применяется к сообщениям в соответствии с корректировочными коэффициентами, достаточное количество реальных данных может быть в итоге накоплено для вычисления более реалистичных значений затрат и эффективности нового правила, и, таким образом, назначено больше баллов производительности. Поскольку баллы производительности пересчитаны, новое правило автоматически перемещается на свое оптимальное место в списке.

[0093] Со временем свод правил может вырасти слишком большим для практически возможного выполнения на основании текущих корректировочных коэффициентов

(например, 1%). Таким образом, должен быть добавлен еще один уровень, в котором правила с баллом производительности меньше, чем, скажем, 0,1, выполняются на основе пониженных корректировочных коэффициентов, таких как, например, 0,1%. Результаты таких правил с пониженной производительностью могут даже не использоваться для классификации сообщений, а вместо этого использоваться только для выработки обновленной информации о затратах и эффективности.

[0094] На фигуре 5 показан пример компьютерной архитектуры 500, которая обеспечивает адаптивное сканирование электронных сообщений и адаптивного выбора правил, которые используются для классификации электронных сообщения.

[0095] Неклассифицированную почту 501U принимают. Информация об отправителе/получателе отсылается к клиентскому классу 531. Клиентский класс 531 устанавливает минимальную эффективность для использования при классификации неклассифицированного сообщения 501U. На стадии 541 минимальная эффективность для клиентского класса 531 достигается за счет исполнения правил 511A, 511B и 511C. На стадии 542 одно или несколько дополнительных правил, в том числе правило 511D, запускаются рационально, насколько позволяет доступность 504 ресурсов. На стадии 543 решение не исполнять одно или несколько других правил вплоть до правила 511N корректируется на основании случайной возможности 518, и эти одно или несколько других правил вплоть до 511N исполняются. Выходные данные представляют собой классифицированную почту 501C (например, как SPAM или допустимую) на основании результатов правил с 511A по 511N.

[0096] Данные о работоспособности собираются на этапе выполнения 503 каждого правила для правил с 511A по 511N. Обновленные баллы затрат записываются обратно в правила с 511A по 511N. Результат на выходе каждого правила 512 определяется для правил с 511A по 511N как положительный (например, это SPAM) или отрицательный (например, это допустимо). Внешняя обратная связь 561 включается в рассмотрение, чтобы установить ошибочные срабатывания и ложные отрицания в результатах на выходе. Обновленные баллы эффективности записываются обратно в правила с 511A по 511N. Показатели производительности пересчитываются и правила переупорядочиваются на основании пересчитанных показателей производительности.

[0097] Настоящее изобретение может быть осуществлено в других специфических формах, не отступая от его сущности или существенных характеристик. Описанные варианты должны быть рассмотрены во всех отношениях только как иллюстративные и не ограничивающие. Объем изобретения, таким образом, определяется прилагаемой формулой изобретения, а не вышеизложенным описанием. Все изменения, которые подходят по смыслу и широте охвата эквивалентов формулы изобретения, должны быть включены в ее объем.

Формула изобретения

1. Способ, реализуемый в компьютерной системе, включающей в себя один или несколько процессоров и системную память, причем компьютерная система включает в себя множество правил классификации электронных сообщений, при этом способ предназначен для адаптивного выбора правил, использующихся для классификации электронных сообщений, причем способ включает в себя:

этап приема, на котором принимают одно или несколько электронных сообщений;

для каждого из этих одного или нескольких электронных сообщений этап вычисления результата, на котором вычисляют результат, показывающий вероятность того, что электронное сообщение обладает заданными характеристиками сообщения, посредством

применения каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации электронных сообщений;

этап измерения затрат ресурсов, на котором измеряют затраты ресурсов, показывающие количество ресурсов, израсходованных на применение каждого правила классификации электронных сообщений к каждому из упомянутых одного или нескольких электронных сообщений;

для каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации сообщений:

этап, на котором синтезируют показатель производительности из измеренных результатов и подсчитанных затрат ресурсов для правила классификации сообщений, при этом показатель производительности представляет собой показатель, определяющий производительность классификации электронного сообщения;

этап, на котором сравнивают синтезированные показатели производительности с существующими показателями производительности для правил классификации электронных сообщений, входящих в состав упомянутого множества правил классификации электронных сообщений; и

этап выбора нового подмножества правил классификации электронных сообщений, на котором выбирают новое подмножество правил классификации электронных сообщений из состава упомянутого множества правил классификации электронных сообщений для использования в классификации последующих принятых электронных сообщений на основе, по меньшей мере частично, результатов сравнения синтезированных показателей производительности с существующими показателями производительности.

2. Способ по п. 1, в котором этап выбора нового подмножества правил классификации электронных сообщений включает в себя этап, на котором выбирают новое подмножество правил классификации электронных сообщений в соответствии с соглашением об уровне обслуживания (SLA).

3. Способ по п. 1, в котором этап выбора нового подмножества правил классификации электронных сообщений включает в себя этап, на котором переупорядочивают упомянутое множество правил классификации электронных сообщений на основе баллов производительности.

4. Способ по п. 1, в котором при приеме одного или нескольких электронных сообщений принимают одно или несколько сообщений электронной почты.

5. Способ по п. 1, в котором при приеме одного или нескольких электронных сообщений принимают одно или несколько сообщений службы коротких сообщений (SMS).

6. Способ по п. 1, в котором при приеме одного или нескольких электронных сообщений принимают один или несколько файлов.

7. Способ по п. 1, в котором упомянутое множество правил классификации электронных сообщений используется для классификации электронных сообщений как нежелательные (SPAM) или как допустимые.

8. Способ по п. 1, в котором упомянутое множество правил классификации электронных сообщений используется для классификации электронных сообщений как содержащие вредоносные программы или не содержащие вредоносные программы.

9. Способ по п. 1, в котором упомянутое множество правил классификации электронных сообщений используется для классификации электронных сообщений как содержащие конфиденциальную цифровую информацию или не содержащие конфиденциальную цифровую информацию.

10. Способ, реализуемый в компьютерной системе, включающей в себя один или несколько процессоров и системную память, причем компьютерная система включает в себя множество правил классификации электронных сообщений, при этом способ предназначен для адаптивного выбора правил, использующихся для классификации электронных сообщений, причем способ включает в себя:

этап приема, на котором принимают одно или несколько электронных сообщений;

для каждого из этих одного или нескольких электронных сообщений:

этап, на котором применяют каждое правило классификации сообщений из предварительно выбранного подмножества правил классификации электронных сообщений к электронному сообщению, причем предварительно выбранное подмножество правил классификации электронных сообщений представляет собой подмножество упомянутого множества правил классификации электронных сообщений;

для каждого правила классификации электронных сообщений из предварительно выбранного подмножества правил классификации электронных сообщений:

этап, на котором вычисляют результат для правила классификации электронных сообщений, показывающий вероятность того, что электронное сообщение обладает заданными характеристиками сообщения;

этап, на котором измеряют затраты ресурсов, показывающие количество ресурсов, израсходованных на применение правила классификации электронных сообщений к электронному сообщению;

этап, на котором сохраняют вычисленный результат и измеренные затраты ресурсов, связанные с применением каждого правила классификации электронной почты к каждому электронному сообщению;

для каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации сообщений:

этап, на котором синтезируют показатель производительности из сохраненных вычисленных результатов и измеренных затрат ресурсов для правила классификации сообщений, при этом показатель производительности представляет собой показатель, определяющий производительность классификации электронного сообщения;

этап, на котором сравнивают синтезированные показатели производительности с существующими показателями производительности для правил классификации электронных сообщений, входящих в состав упомянутого множества правил классификации электронных сообщений; и

этап, на котором выбирают новое подмножество правил классификации электронных сообщений из состава упомянутого множества правил классификации электронных сообщений для использования в классификации последующих принятых электронных сообщений на основе, по меньшей мере частично, результатов сравнения синтезированных показателей производительности с существующими показателями производительности.

11. Способ по п. 10, дополнительно включающий в себя этап, на котором, до применения каждого правила классификации сообщений из предварительно выбранного подмножества правил классификации электронных сообщений, выбирают выбираемое подмножество правил классификации электронных сообщений на основе вычисленных баллов производительности.

12. Способ по п. 10, дополнительно включающий в себя:

этап, на котором получают внешнюю обратную связь относительно применения по меньшей мере одного правила классификации сообщений к электронному сообщению; и

этап, на котором включают внешнюю обратную связь в вычисленный результат применения этого по меньшей мере одного правила классификации сообщений к электронному сообщению.

13. Способ по п. 12, в котором полученная внешняя обратная связь показывает, что вычисленный результат применения упомянутого по меньшей мере одного правила классификации сообщений к электронному сообщению представляет собой одно из ложноотрицательного результата и ошибочного срабатывания.

14. Способ по п. 12, дополнительно включающий в себя этап, на котором обновляют баллы эффективности для упомянутого по меньшей мере одного правила классификации сообщений на основе вычисленного результата, включающего в себя внешнюю обратную связь.

15. Способ по п. 10, в котором этап выбора нового подмножества правил классификации электронных сообщений содержит этап, на котором выбирают новое подмножество правил классификации электронных сообщений в соответствии с соглашением об уровне обслуживания (SLA).

16. Способ по п. 10, в котором этап выбора нового подмножества правил классификации электронных сообщений содержит этап, на котором переупорядочивают упомянутое множество правил классификации электронных сообщений на основе баллов производительности.

17. Система для адаптивного выбора правил обнаружения нежелательных сообщений (SPAM), содержащая:

один или несколько процессоров;

системную память;

один или несколько компьютерных накопителей, на которых хранится множество правил обнаружения SPAM и на которых хранятся исполняемые инструкции, которыми реализуются классификатор сообщений и модуль выбора и переупорядочивания правил,

при этом классификатор сообщений сконфигурирован:

принимать одно или несколько сообщений электронной почты;

для каждого из этих одного или нескольких сообщений электронной почты применять каждое правило обнаружения SPAM из предварительно выбранного подмножества правил обнаружения SPAM к сообщению электронной почты, причем предварительно выбранное подмножество правил обнаружения SPAM представляет собой подмножество упомянутого множества правил обнаружения SPAM; и

для каждого правила обнаружения SPAM из предварительно выбранного подмножества правил обнаружения SPAM:

вычислять результат, показывающий вероятность того, что сообщение электронной почты представляет собой SPAM;

измерять затраты ресурсов, показывающие количество ресурсов, израсходованных на применение правила обнаружения SPAM, к каждому из упомянутых одного или нескольких сообщений электронной почты; и

синтезировать показатель производительности из вычисленных результатов и измеренных затрат ресурсов для правила обнаружения SPAM, при этом показатель производительности представляет собой показатель, определяющий производительность классификации электронного сообщения как SPAM на основе частного от деления вычисленных результатов и измеренных затрат ресурсов; и

при этом модуль выбора и переупорядочивания правил сконфигурирован:

сравнивать синтезированные показатели производительности с существующими показателями производительности для правил обнаружения SPAM, входящих в состав

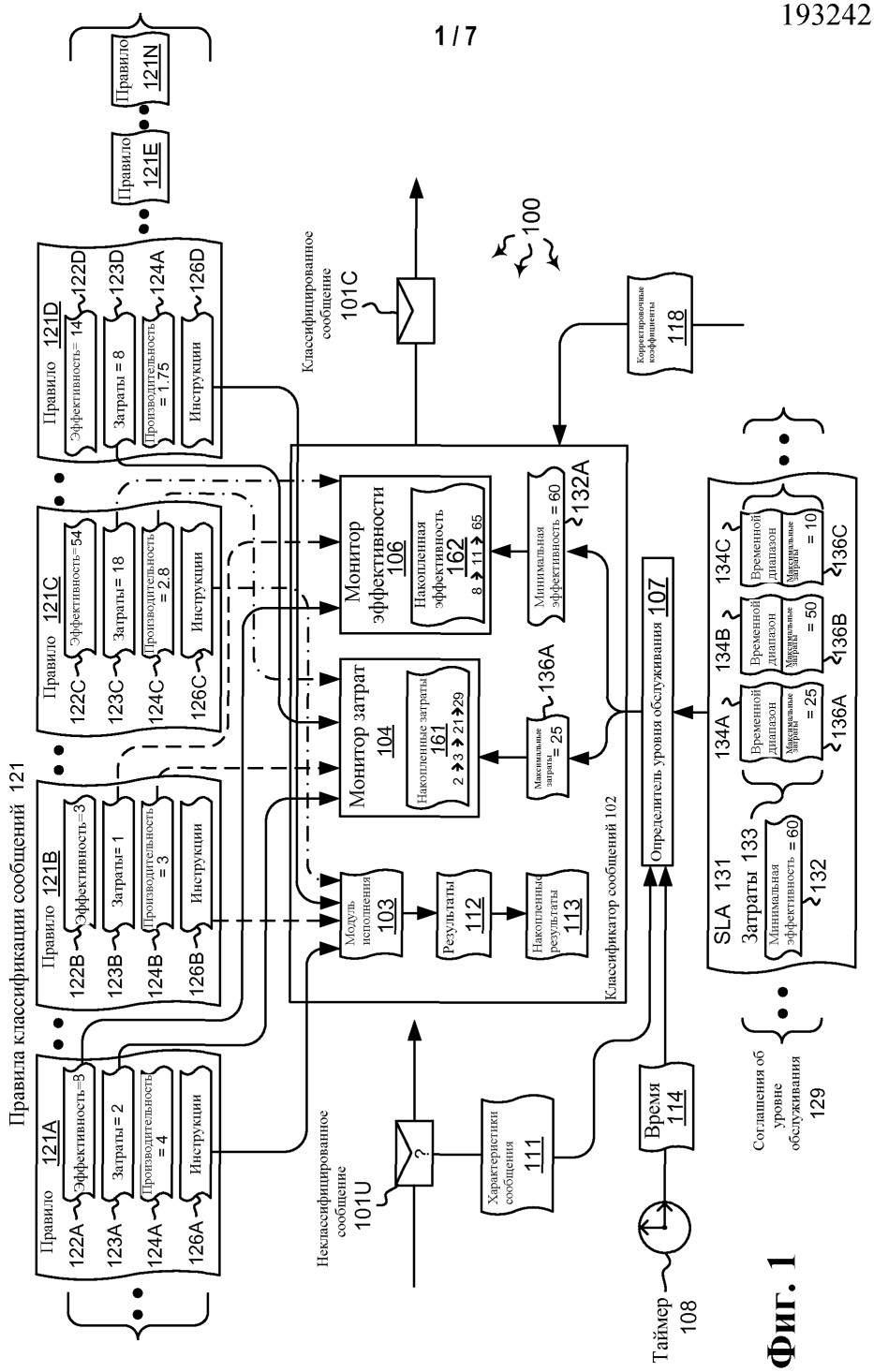
упомянутого множества правил обнаружения SPAM; и

выбирать новое подмножество правил обнаружения SPAM для использования в классификации последующих принятых сообщений электронной почты на основе, по меньшей мере частично, результатов сравнения синтезированных показателей производительности с существующими показателями производительности.

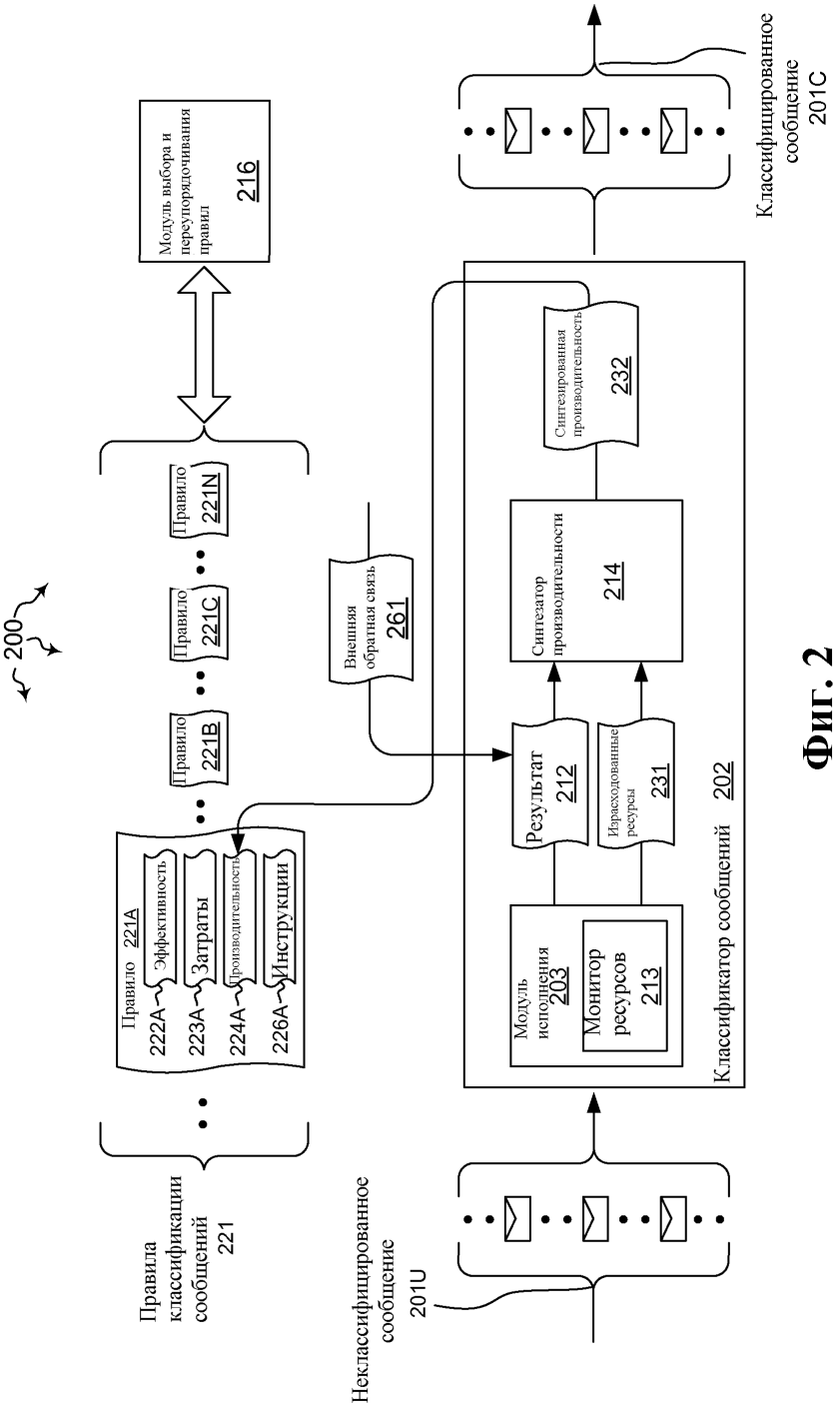
18. Система по п. 17, в которой модуль выбора и переупорядочивания правил, будучи сконфигурированным выбирать новое подмножество правил обнаружения SPAM, сконфигурирован выбирать новое подмножество правил обнаружения SPAM в соответствии с соглашением об уровне обслуживания (SLA).

19. Система по п. 17, в которой модуль выбора и переупорядочивания правил, будучи сконфигурированным выбирать новое подмножество правил обнаружения SPAM, сконфигурирован переупорядочивать упомянутое множество правил обнаружения SPAM на основе баллов производительности.

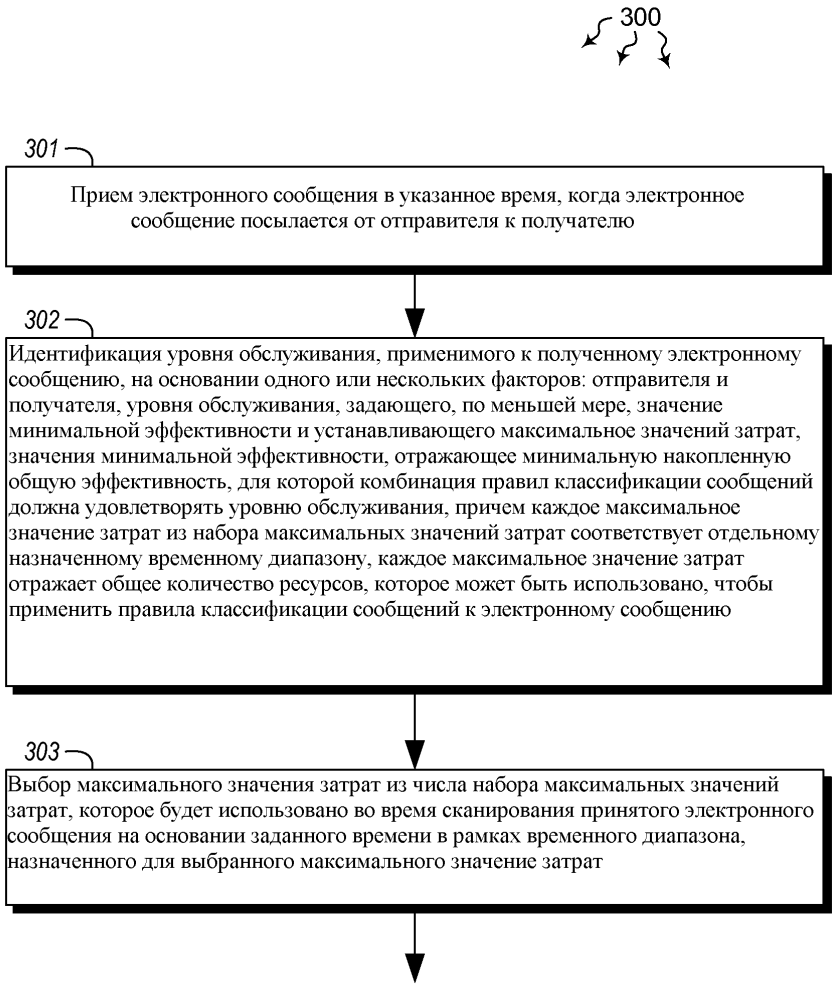
20. Система по п. 17, в которой классификатор сообщений, будучи сконфигурированным принимать одно или несколько сообщений электронной почты, сконфигурирован принимать сообщения электронной почты из Интернет.


Правила классификации сообщений 121
Правило 121A
Эффективность=3
Затраты = 2
Производительность = 4
Инструкции
122A
123A
124A
126A
Правило 121B
Эффективность=3
Затраты= 1
Производительность = 3
Инструкции
122B
123B
124B
126B
Правило 121C
Эффективность=54
Затраты= 18
Производительность = 2.8
Инструкции
122C
123C
124C
126C
Правило 121D
Эффективность= 14
Затраты = 8
Производительность = 1.75
Инструкции
122D
123D
124A
126D
Правило 121E
Правило 121N
Неклассифицированное сообщение 101U
Классифицированное сообщение 101C
100
Классификатор сообщений 102
Монитор затрат 104
Накопленные затраты 161
2 → 3 → 21 → 29
Максимальные затраты = 25
136A
Монитор эффективности 106
Накопленная эффективность 162
8 → 11 → 65
Минимальная эффективность = 60
132A
Модуль исполнения 103
Результаты 112
Накопленные результаты 113
Характеристики сообщения 111
Определитель уровня обслуживания 107
Корректировочные коэффициенты 118
Время 114
Таймер 108
SLA 131
Затраты 133
Минимальная эффективность = 60
132
134A
Временной диапазон
Максимальные затраты = 25
136A
134B
Временной диапазон
Максимальные затраты = 50
136B
134C
Временной диапазон
Максимальные затраты = 10
136C
Соглашения об уровне обслуживания 129

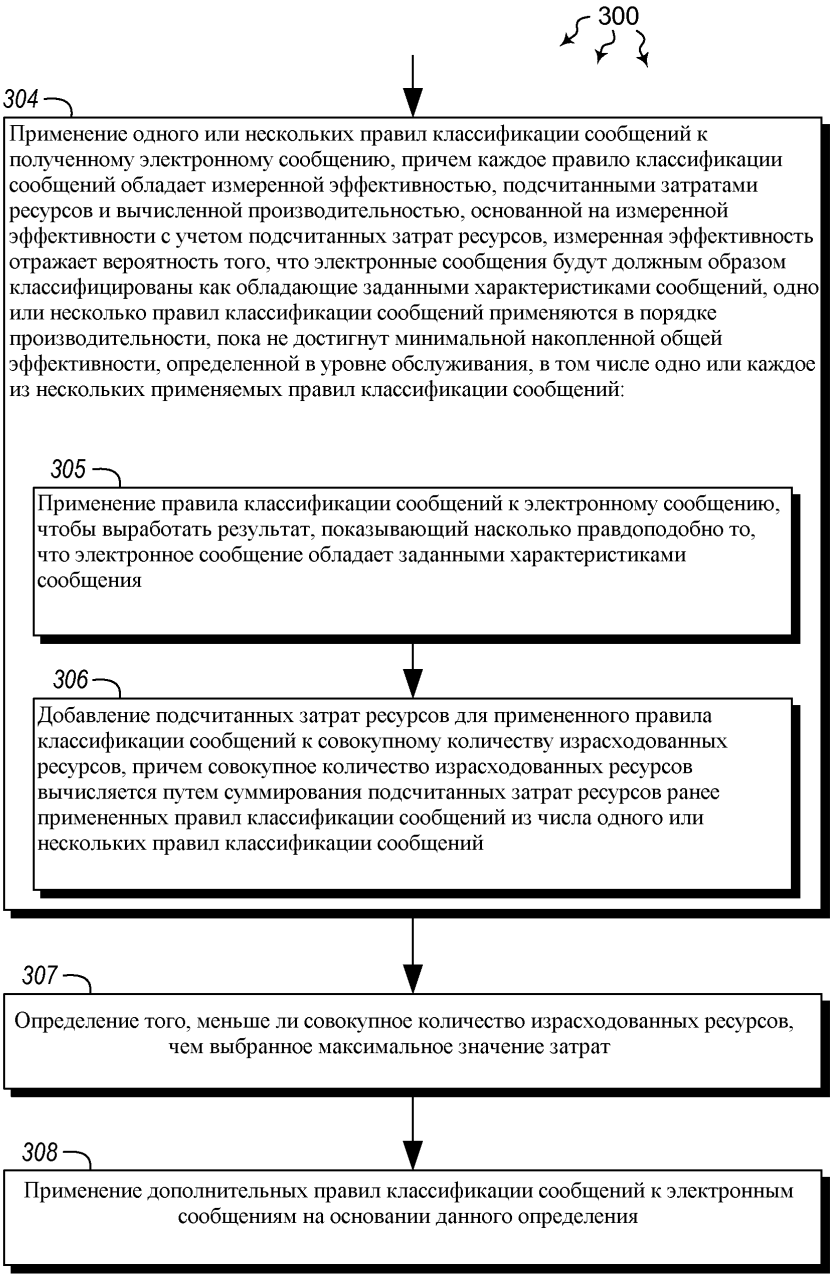
Фиг. 1


200
Правила классификации сообщений 221
Правило 221A
222A Эффективность
223A Затраты
224A Производительность
226A Инструкции
Правило 221B
Правило 221C
Правило 221N
Модуль выбора и переупорядочивания правил 216
Внешняя обратная связь 261
Неклассифицированное сообщение 201U
Классификатор сообщений 202
Модуль исполнения 203
Монитор ресурсов 213
Результат 212
Израсходованные ресурсы 231
Синтезатор производительности 214
Синтезированная производительность 232
Классифицированное сообщение 201C

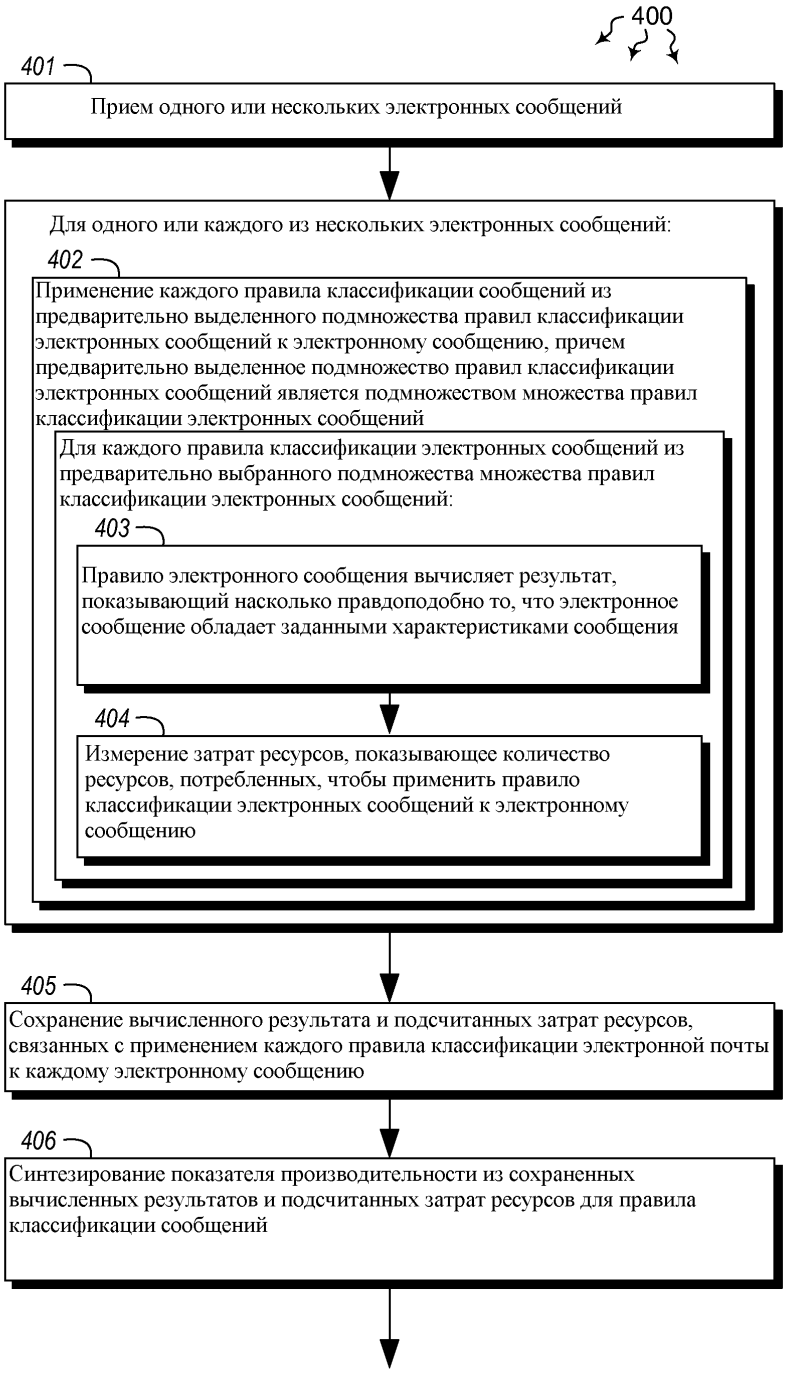
Фиг. 2


300
301
Прием электронного сообщения в указанное время, когда электронное сообщение посылается от отправителя к получателю
302
Идентификация уровня обслуживания, применимого к полученному электронному сообщению, на основании одного или нескольких факторов: отправителя и получателя, уровня обслуживания, задающего, по меньшей мере, значение минимальной эффективности и устанавливающего максимальное значений затрат, значения минимальной эффективности, отражающее минимальную накопленную общую эффективность, для которой комбинация правил классификации сообщений должна удовлетворять уровню обслуживания, причем каждое максимальное значение затрат из набора максимальных значений затрат соответствует отдельному назначенному временному диапазону, каждое максимальное значение затрат отражает общее количество ресурсов, которое может быть использовано, чтобы применить правила классификации сообщений к электронному сообщению
303
Выбор максимального значения затрат из числа набора максимальных значений затрат, которое будет использовано во время сканирования принятого электронного сообщения на основании заданного времени в рамках временного диапазона, назначенного для выбранного максимального значение затрат

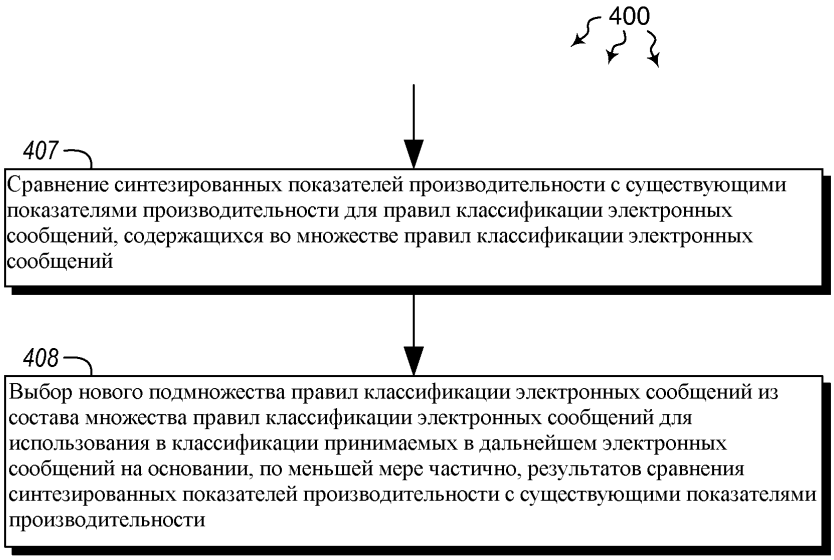
Фиг. 3



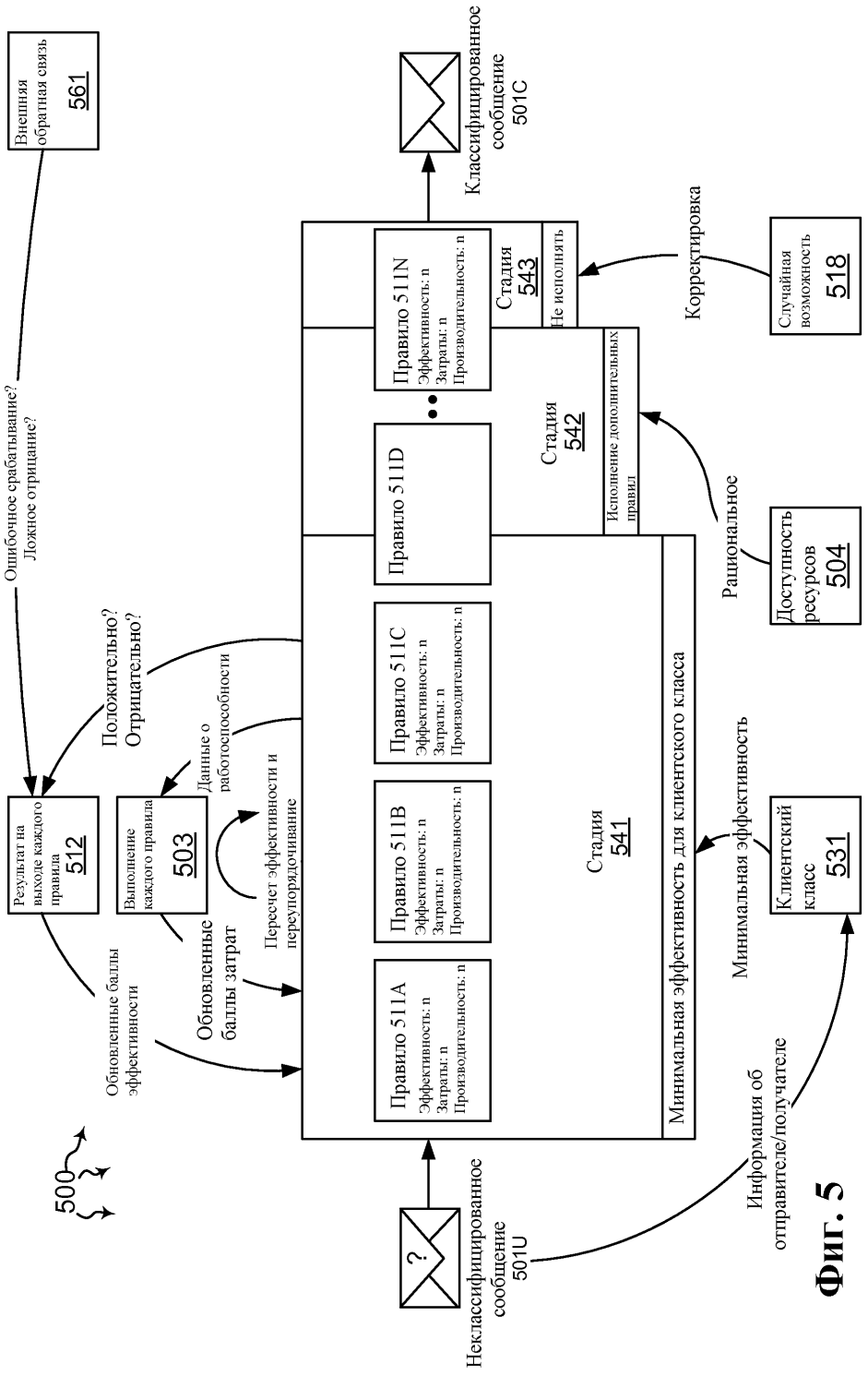
Фиг. 3
(продолжение)


400
401
Прием одного или нескольких электронных сообщений
Для одного или каждого из нескольких электронных сообщений:
402
Применение каждого правила классификации сообщений из предварительно выделенного подмножества правил классификации электронных сообщений к электронному сообщению, причем предварительно выделенное подмножество правил классификации электронных сообщений является подмножеством множества правил классификации электронных сообщений
Для каждого правила классификации электронных сообщений из предварительно выбранного подмножества множества правил классификации электронных сообщений:
403
Правило электронного сообщения вычисляет результат, показывающий насколько правдоподобно то, что электронное сообщение обладает заданными характеристиками сообщения
404
Измерение затрат ресурсов, показывающее количество ресурсов, потребленных, чтобы применить правило классификации электронных сообщений к электронному сообщению
405
Сохранение вычисленного результата и подсчитанных затрат ресурсов, связанных с применением каждого правила классификации электронной почты к каждому электронному сообщению
406
Синтезирование показателя производительности из сохраненных вычисленных результатов и подсчитанных затрат ресурсов для правила классификации сообщений

Фиг. 4



Фиг. 4
(продолжение)


500
Внешняя обратная связь 561
Ошибочное срабатывание?
Ложное отрицание?
Положительно?
Отрицательно?
Результат на выходе каждого правила 512
Выполнение каждого правила 503
Данные о работоспособности
Пересчет эффективности и переупорядочивание
Обновленные баллы эффективности
Обновленные баллы затрат
Правило 511A
Эффективность: n
Затраты: n
Производительность: n
Правило 511B
Эффективность: n
Затраты: n
Производительность: n
Правило 511C
Эффективность: n
Затраты: n
Производительность: n
Правило 511D
Правило 511N
Эффективность: n
Затраты: n
Производительность: n
Стадия 541
Минимальная эффективность для клиентского класса
Стадия 542
Исполнение дополнительных правил
Стадия 543
Не исполнять
Некласифицированное сообщение 501U
Классифицированное сообщение 501C
Информация об отправителе/получателе
Минимальная эффективность
Клиентский класс 531
Рациональное
Доступность ресурсов 504
Корректировка
Случайная возможность 518

Фиг. 5