(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 701 874**

(51) Int. CI.:

H04L 12/54 (2013.01)
H04L 9/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **25.10.2007 PCT/CA2007/001909**

(87) Fecha y número de publicación internacional: **08.05.2008 WO08052317**

(96) Fecha de presentación y número de la solicitud europea: **25.10.2007 E 07816059 (5)**

(97) Fecha y número de publicación de la concesión europea: **12.09.2018 EP 2080324**

(54) Título: **Método basado en reputación y sistema para determinar una probabilidad de que un mensaje sea no deseado**

(30) Prioridad:

31.10.2006 US 554746

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
26.02.2019

(73) Titular/es:

WATCHGUARD TECHNOLOGIES, INC. (100.0%)
505 Fifth Avenue South, Suite 500
Seattle, WA 98014 , US

(72) Inventor/es:

GABE, CHRISTOPHER JOHN

(74) Agente/Representante:

ELZABURU, S.L.P

DESCRIPCIÓN

Método basado en reputación y sistema para determinar una probabilidad de que un mensaje sea no deseado

Campo de la invención

La presente invención se refiere a un sistema y a un método para usar una reputación, derivada para un creador de mensajes, para determinar una probabilidad de que un mensaje sea no deseado. Más específicamente, la presente invención se refiere a un método y a un sistema de producción de una métrica de reputación para creadores de mensajes, usando al menos una tupla de características del mensaje para identificar al creador del mensaje, cuya métrica se puede usar para determinar la probabilidad de que un mensaje sea no deseado.

Antecedentes de la invención

El correo electrónico no deseado, comúnmente conocido como SPAM, se define generalmente como correo electrónico no solicitado en masa, típicamente con propósitos comerciales. El SPAM es un problema significativo para los administradores y usuarios de correo electrónico. En el mejor de los casos, el SPAM utiliza recursos en los sistemas de correo electrónico, requiere tiempo del titular de la cuenta de correo electrónico para revisar y borrar y generalmente es frustrante y molesto. En el peor de los casos, el SPAM puede incluir software malicioso y puede dañar el software, los sistemas y/o los datos almacenados.

Las comunicaciones de voz basadas en el Protocolo de Inicio de Sesión (SIP) también están sometidas a mensajes no deseados y tales mensajes no deseados también se conocen en la presente memoria como SPAM. Aunque aún no es común, se espera que el SPAM relacionado con la voz llegue a ser un problema común a medida que más usuarios migren del servicio telefónico ordinario (POTS) a comunicaciones de voz basadas en SIP. Por ejemplo, es posible enviar mensajes comerciales no solicitados a cada buzón de correo de voz en una organización, utilizando los recursos del sistema y desperdiciando el tiempo de los usuarios para revisar y/o borrar los mensajes de SPAM.

Se ha acometido mucho trabajo en los últimos años para combatir el problema creciente del SPAM. Uno de los métodos usados hasta la fecha para reducir el SPAM de correo electrónico no deseado es el uso del filtrado bayesiano en donde se examina el contenido de los correos electrónicos recibidos en busca de contenido específico para tomar una decisión estadística en cuanto a si el correo electrónico constituye SPAM. Un mensaje que se considera que es SPAM se puede marcar como tal y/o dirigir a una carpeta de almacenamiento seleccionada o borrar del sistema. Aunque tales filtros reconocen muchos mensajes de SPAM, los creadores de los mensajes de SPAM están constantemente cambiando sus mensajes, a menudo con éxito, en intentos de engañar a los filtros.

La solicitud de patente publicada de EE.UU. pendiente de tramitación 2007/0209067 de Fogel, presentada el 21 de febrero de 2006 y titulada “System and Method For Providing Security For SIP-Based Communications” describe un aparato de seguridad y algunos métodos que pueden ser útiles para reducir la aparición de SPAM de voz y los contenidos de esta solicitud se incorporan en la presente memoria por referencia.

Otro método comúnmente empleado hasta la fecha es el uso de listas negras que identifican las direcciones IP desde las cuales se han recibido previamente mensajes considerados que son no deseados y las cuales consideran todos los mensajes posteriores desde esas direcciones IP como que son mensajes no deseados. Aunque las listas negras pueden ser eficaces, sufren de ser de granularidad muy gruesa ya que no distinguen entre mensajes enviados por un usuario de buena fe en una dirección IP y SPAM enviados por creadores de SPAM desde la misma dirección IP.

En su lugar, una vez que la dirección IP se ha identificado e incluido en una lista negra como que es una dirección IP usada para crear SPAM, los mensajes de los usuarios de buena fe ya no se aceptarán más en los sistemas que han incluido en una lista negra la dirección IP. Como muchos Proveedores de Servicios de Internet (ISP) alojan múltiples correos electrónicos y/o dominios SIP en una única dirección IP, esta lista negra de dominios puede afectar a un gran número de usuarios de buena fe.

Más recientemente, se han empleado técnicas basadas en la reputación para ayudar a identificar mensajes no deseados. Tales técnicas basadas en la reputación comprenden sistemas de bases de datos que mantienen estadísticas de una dirección IP y estas estadísticas se compilan a partir de la salida de otros sistemas anti-SPAM, tales como el filtro bayesiano o los sistemas SIP mencionados anteriormente. Las estadísticas indican la frecuencia con la que se transmite SPAM desde la dirección IP y pueden incluir otra información tal como si la dirección IP de envío es una dirección estática o dinámica.

Las técnicas basadas en la reputación se basan en un análisis de la actividad pasada de una dirección IP para proporcionar una indicación de una probabilidad de que un nuevo mensaje enviado desde esa dirección IP sea SPAM.

Cuando se recibe un mensaje en un servidor de correo electrónico o intermediario SIP, la reputación de la dirección IP de origen se comprueba en la base de datos y la “reputación” (es decir, las estadísticas compiladas) para esa dirección IP se puede usar como una de las entradas a un proceso anti-SPAM.

Otra técnica basada en la reputación para correos electrónicos se describe en el documento, “Sender Reputation in a Large Webmail Service”, de Bradley Taylor, presentado en CEAS 2006 - Tercera conferencia sobre correo electrónico y antispam, 27-28 de julio de 2006, Mountain View, CA. Esta técnica crea una reputación para cada dominio (que se autentica a través de otros medios) desde el cual se recibe un mensaje de correo electrónico y usa la reputación creada como entrada a un proceso de detección de SPAM.

Aunque las técnicas basadas en la reputación pueden ser una mejora sobre las listas negras, sufren de algunos de los mismos problemas y, en particular, sufren de una falta de granularidad lo que puede dar como resultado que todos los mensajes de una dirección IP o todos los mensajes de un dominio sean identificados como SPAM porque se ha enviado SPAM previamente desde esa dirección IP o dominio. Como se ha mencionado anteriormente, esto puede dar como resultado que un gran número de usuarios de buena fe se vean afectados negativamente como resultado de las actividades de unos pocos creadores de SPAM.

Se desea tener un sistema y un método basados en la reputación para determinar una probabilidad de que un mensaje sea no deseado, lo que permite una granularidad más fina en el seguimiento de las reputaciones.

El documento US 2005/0204012 describe un sistema y un método para utilizar características tolerantes a fallos de un protocolo de entrega de mensajes para evitar la aceptación de un mensaje de un remitente desconocido por un servidor de mensajes de un destinatario, a menos que o bien el destinatario o bien el administrador del servidor den un permiso explícito. Cuando un remitente se pone en contacto con un servidor de mensajes de recepción controlado por el sistema descrito y solicita la entrega de un mensaje, el servidor de recepción solicita que el sistema determine si el mensaje se debería aceptar para su entrega. El sistema crea una o más tuplas de identidad usando la información del sobre del mensaje proporcionada por el servidor de mensajes, y usa éstas para consultar las bases de datos para determinar si se debería rechazar o aceptar el mensaje del remitente.

El documento US 2006/0168024A1 describe el uso de las reputaciones del remitente para la prevención de SPAM. Las estadísticas y las heurísticas en tiempo real se construyen, almacenan, analizan y usan para formular un nivel de reputación del remitente para su uso en la evaluación y control de una conexión de un remitente dado con un agente de transferencia de mensajes o un destinatario de correo electrónico.

Es el objeto de la presente invención permitir una determinación mejorada de una probabilidad en cuanto a si un mensaje recibido es un mensaje no deseado.

El objeto se resuelve por la materia objeto de las reivindicaciones independientes.

Las realizaciones preferidas de la presente invención se definen por las reivindicaciones dependientes.

Compendio de la invención

Es un objeto de la presente invención proporcionar un método y un sistema novedosos basados en la reputación para determinar la probabilidad de que un mensaje sea no deseado, el cual obvia o mitiga al menos una desventaja de la técnica anterior.

Según un primer aspecto de la presente invención, se proporciona un método de determinación de una probabilidad de que un mensaje recibido sea un mensaje no deseado, que comprende los pasos de: (i) recibir un mensaje en un sistema de mensajería; (ii) reenviar a un motor de reputación un conjunto de identificadores preseleccionados con relación al origen del mensaje, al menos uno de los identificadores que es en forma de una tupla, una mitad de la cual representa datos que no se pueden falsificar por el creador del mensaje recibido; (iii) comprobar las bases de datos en el motor de reputación para determinar las métricas de reputación determinadas previamente para los identificadores reenviados y devolver cualquier métrica de reputación determinada previamente al sistema de mensajería; (iv) hacer una primera determinación en el sistema de mensajería de una probabilidad en cuanto a si el mensaje recibido es no deseado usando un primer conjunto de criterios, incluyendo las métricas de reputación devueltas; y (v) marcar el mensaje como que es o bien deseado o bien no deseado según la primera determinación.

Preferiblemente, la mitad de la tupla que no se puede falsificar es la dirección IP del creador del mensaje. También preferiblemente, el método comprende además los pasos de: (vi) hacer una segunda determinación en el sistema de mensajería en cuanto a si el mensaje recibido es no deseado sin usar ninguna métrica de reputación devuelta; y (vii) reenviar la segunda determinación al motor de reputación para actualizar las bases de datos y las métricas de reputación respectivas para incluir la segunda determinación.

Según otro aspecto de la presente invención, se proporciona un entorno de mensajería que emplea un servicio de reputación al determinar una probabilidad en cuanto a si los mensajes recibidos son no deseados, que comprende: una pluralidad de servidores de mensajes interconectados por una red de comunicaciones, al menos uno de la pluralidad de servidores de mensajes que incluye una función anti-SPAM para determinar una probabilidad en cuanto a si los mensajes recibidos son no deseados; una pluralidad de clientes de mensajes conectados a los respectivos de la pluralidad de servidores de mensajes y operables para recibir mensajes desde los mismos; y un motor de reputación operable para comunicarse con el al menos un servidor de mensajes, el motor de reputación que mantiene un conjunto de bases de datos asociando una métrica de reputación con cada uno de un conjunto de

identificadores preseleccionados con relación a los orígenes de los mensajes, al menos uno de los identificadores que es en forma de una tupla, una mitad de la cual no se puede falsificar por el creador del mensaje recibido, la función anti-SPAM que opera para reenviar el conjunto de identificadores preseleccionados al motor de reputación que devuelve las métricas de reputación almacenadas en sus bases de datos para cualquiera de los identificadores y la función anti-SPAM que usa las métricas de reputación devueltas para hacer una primera determinación de la probabilidad en cuanto a si un mensaje recibido es no deseado.

La presente invención proporciona un sistema y un método para proporcionar un servicio de reputación para su uso en entornos de mensajería que emplea estadísticas compiladas, que representan si los mensajes de SPAM se han recibido previamente desde el creador del mensaje, o desde creadores relacionados, en un proceso de toma de decisiones para mensajes recién recibidos. Los sistemas de mensajes que reciben un mensaje reenvían un conjunto de identificadores con relación a los orígenes del mensaje, tales como la dirección IP de origen del mensaje, una tupla identificadora del dominio y la dirección IP desde la cual se recibió supuestamente el mensaje y una tupla de identificadora del usuario y la dirección IP desde la cual el mensaje fue supuestamente recibido a un motor de reputación. El motor de reputación mantiene las bases de datos para cada identificador y cada una de estas bases de datos incluye al menos una métrica de reputación asociada derivada de mensajes recibidos considerados previamente y las determinaciones hechas por los sistemas de correo electrónico en cuanto a una probabilidad de que sean SPAM. El motor de reputación devuelve las métricas de reputación asociadas, en su caso, para los identificadores al sistema de mensajes, que entonces puede hacer una determinación, con las métricas devueltas, de una probabilidad en cuanto a si el mensaje es SPAM.

Breve descripción de los dibujos

Las realizaciones preferidas de la presente invención se describirán ahora, a modo de ejemplo solamente, con referencia a las Figuras adjuntas, en donde:

la Figura 1 muestra una representación esquemática de un entorno de mensajería de correo electrónico que emplea un servicio de reputación según la presente invención;

la Figura 2 es un diagrama de flujo de una parte del método del servicio de reputación de la Figura 1; y

la Figura 3 es un diagrama de flujo de otra parte del método del servicio de reputación de la Figura 1.

Descripción detallada de la invención

Un entorno de mensajería de correo electrónico que incorpora un servicio de reputación según la presente invención, se indica de manera general en 20 en la Figura 1. Aunque la realización ilustrada es una realización de mensajería de correo electrónico, la presente invención también es aplicable a otros entornos de mensajería, tales como voz sobre IP (VoIP) basada en SIP, etc.

Por ejemplo, como es sabido por los expertos en la técnica, además de la dirección IP de origen, los mensajes SIP (es decir, INVITE, etc.) usados para configurar una comunicación de voz incluyen campos para un identificador de usuario (un nombre o número de teléfono, etc.) y un dominio. De este modo, aunque la siguiente discusión se relaciona con un entorno de correo electrónico según la presente invención, será evidente para los expertos en la técnica que el mismo método y sistema generales se pueden emplear también para comunicación de voz basada en SIP.

El entorno de mensajería 20 incluye al menos un cliente de correo electrónico 24 que se conecta a un sistema de correo electrónico 28. El sistema de correo electrónico 28 incluye al menos un servidor de correo electrónico 32, que proporciona servicios de correo electrónico entrante y saliente, y un aparato de seguridad de correo electrónico 36, como el cortafuegos de correo electrónico MXtreme™ vendido por el cesionario de la presente invención. El aparato de seguridad de correo electrónico 36 proporciona servicios anti-SPAM, como se describe además a continuación, y puede proporcionar otros servicios de seguridad. El entorno 20 también incluye una pluralidad de clientes de correo electrónico 40 que están conectados a servidores de correo electrónico 44 que proporcionan servicios de correo electrónico entrante y saliente.

En el caso de un entorno de mensajería basado en SIP, el aparato de seguridad 36 puede ser un producto de seguridad SIP, tal como el cortafuegos SIP SIPassure™ vendido por el cesionario de la presente invención.

Cada uno del sistema de correo electrónico 28 y los servidores de correo electrónico 44 están interconectados por una red 48, tal como Internet, y cada dispositivo conectado a la red 48 se identifica en la misma mediante una dirección única. En el caso ilustrado de Internet, a cada dispositivo se le asigna una dirección IP única (Protocolo de Internet) que comprende un conjunto de cuatro valores que oscilan entre 0 y 255 (por ejemplo, 75.127.34.65).

Como es sabido por los expertos en la técnica, cada servidor de correo electrónico 44 y cada sistema de correo electrónico 28 pueden alojar uno o más dominios (es decir, mail.com, example.co.uk, house.org, etc.) para los clientes de correo electrónico 44 que sirve. De este modo, dos o más dominios pueden enviar o recibir mensajes en la misma dirección única para el sistema de correo electrónico 28 o el servidor de correo electrónico 44 que los aloja.

Como se emplea en la presente memoria, el término “dominio” se pretende que comprenda cualquier indicador adecuado para la dirección no-IP y la parte no específica del usuario del creador de un mensaje bajo consideración. Como se ha indicado anteriormente, los dominios comprenderán típicamente la parte de una dirección de correo electrónico o un número de teléfono SIP, etc., a la derecha del símbolo “@” (por ejemplo, “example.com”, no obstante, en algunos países, también se añade un código de país al dominio (es decir, guys.co.uk), donde co.uk es el ccTLD (código de país Dominio de Nivel Superior) y en algunos casos también se puede incluir un indicador de dominio adicional (es decir, el “mail” en mail.zap.co.uk). Con propósitos de coherencia, se prefiere que los identificadores de los dominios usados en la presente invención comprendan el TLD o ccTLD, y el primer identificador a la izquierda del TLD o ccTLD (es decir, guys.co.uk o zap.co.uk).

Además, un dominio se puede alojar en dos o más servidores de correo electrónico 44 o sistemas de correo electrónico 28 a los que se asignan diferentes direcciones únicas. De hecho, esto es bastante común, especialmente si un dominio es particularmente grande (es decir, mail.google.com o mail.yahoo.com). De este modo, un correo electrónico enviado desde tal dominio puede originarse desde una cualquiera de dos o más direcciones únicas.

Un motor de reputación 52, según la presente invención, también está conectado a la red 48, y se le asigna una dirección única dentro de la misma. El motor de reputación 52 se puede comunicar con sistemas de correo electrónico 28 autorizados, a través de la red 48, como se describe a continuación.

La Figura 2 muestra un diagrama de flujo de un método según una realización de entorno de correo electrónico de la presente invención. El método comienza en el paso 100 en donde un sistema de correo electrónico recibe un mensaje. Como se emplea en esta memoria, el término mensaje se pretende que comprenda el mensaje completo, incluyendo las cabeceras, la información del sobre (estructura MIME, etc.), los campos de datos SIP, la marca de tiempo del recibo, el texto del mensaje (en su caso), etc.

En el paso 104, se crea un conjunto de identificadores para el mensaje y estos identificadores se relacionan generalmente con aspectos del origen del mensaje. En esta realización de la invención, se crean tres identificadores únicos para el mensaje. Específicamente, se crea un identificador para la dirección IP de origen desde la cual se recibió el mensaje, se crea un identificador para la tupla del dominio y de la dirección IP desde la cual se envío el mensaje (por ejemplo, example.com) y se crea un identificador para la tupla del usuario y de la dirección IP desde la que se envió el mensaje (por ejemplo, johnsmith@example.com).

Como es difícil para un creador de mensajes no deseados enmascarar o falsificar la dirección IP de origen, se prefiere incluir la dirección IP de origen como una mitad de las tuplas de usuario y de dominio. Incluyendo el usuario o doma como la otra mitad de las tuplas identificadoras, las métricas de reputación se pueden aplicar con una granularidad más fina que en la técnica anterior.

No obstante, se contempla que otros identificadores, o bien además de o bien en lugar de, estos tres identificadores se pueden emplear si se desea, pero se recomienda que se emplee al menos un identificador, en forma de una tupla donde al menos una mitad de la tupla no pueda ser falsificada o enmascarada fácilmente. Por ejemplo, se puede emplear una tupla identificadora basada en la dirección IP de origen y otros datos en el mensaje (estructura MIME, etc.). Como otro ejemplo que puede ser más aplicable en, pero no limitado a, comunicaciones de voz basadas en SIP, se puede emplear un identificador que comprenda una tupla del usuario o dominio y la hora en que se recibió el mensaje (incremento de hora o media hora), en la medida que la hora en que el mensaje fue recibido no se puede falsificar ni imitar por el creador. Esto puede ser útil, en la medida que un creador de mensajes no deseados puede crear o iniciar tales mensajes fuera de las horas laborales normales o a otras horas particulares.

Es preferible que, por razones de privacidad, los identificadores que puedan tener implicaciones de privacidad, tales como el identificador para la tupla del dominio y de la dirección IP y el identificador para la tupla del usuario y de la dirección IP, se creen a través de una función unidireccional que evita que una tercera parte sea capaz de analizar el identificador para recuperar la información específica del usuario (es decir, nombre de usuario y/o dominio).

Por consiguiente, en una presente realización de la invención, se usa una función de comprobación aleatoria SHA1 para crear identificadores de valor de comprobación aleatoria para los identificadores de cada una de las tuplas para el dominio de origen y para el usuario. Además, esto puede proporcionar una ventaja en que la longitud de los identificadores llega a ser coherente. No obstante, la presente invención no se limita al uso de funciones de comprobación aleatoria, SHA 1 o de otro modo, y se puede emplear cualquier función unidireccional adecuada, como se les ocurrirá a los expertos en la técnica.

De este modo, preferiblemente: el identificador de la dirección IP es la dirección IP, o una representación adecuada (texto ASCII, hexadecimal, etc.) de la dirección IP; el identificador de la tupla del dominio y de la dirección IP es una representación de comprobación aleatoria del dominio y de la dirección IP; y el identificador de la tupla del usuario y de la dirección IP es una representación de comprobación aleatoria del usuario y de la dirección IP.

También se contempla que, en muchos casos, se preferirá que el identificador de la dirección IP solamente identifique una parte de la dirección IP, tal como los tres primeros octetos de la dirección IP, en la medida que muchos ordenadores centrales grandes tendrán servidores de correo electrónico a los que se les asignan direcciones IP consecutivas (es decir, 75.127.34.64, 75.127.34.65, 75.127.34.66, etc.). En tal caso, una parte de la

dirección IP, tal como los tres primeros octetos (es decir, 75.127.34) puede constituir una identificación suficiente del origen de los mensajes de estos ordenadores centrales. En tal caso, los identificadores adecuados tratados anteriormente solamente incluirán la parte seleccionada de la dirección IP.

Aunque el uso de una reputación asociada con una dirección IP se conoce a partir de la técnica anterior, la presente invención (a diferencia de la técnica anterior) emplea uno o más identificadores de grano más fino con relación al creador/origen de los mensajes en combinación con la dirección IP u otro atributo que no es fácil de falsificar. En particular, en una realización preferida actualmente de la invención, el conjunto de identificadores incluye una tupla del dominio y de la dirección IP (es decir, example.com y 75.127.34.65) y una tupla del usuario y de la dirección IP (es decir, jsmith@example.com y 75.127.34.65).

Preferiblemente, se almacenará una métrica de reputación adecuada para la tupla de grano más fino, en este ejemplo que comprende el usuario y la dirección IP, como se describe a continuación. Si no se almacena tal métrica de reputación para una tupla particular del usuario y de la dirección IP, se considerará la siguiente tupla de grano más fino, que en este ejemplo es la tupla para el dominio y la dirección IP. La reputación del identificador de la dirección IP solamente necesita ser considerada si no está disponible otra métrica de reputación, de grano más fino.

Volviendo ahora a la Figura 2, en el paso 108, los identificadores creados se envían al motor de reputación 52 a través de la red 48.

Con referencia ahora a la Figura 3, en el paso 112, el motor de reputación 52 recibe los identificadores creados para el mensaje recibido desde el aparato de seguridad de correo electrónico 36. El motor de reputación 52 contiene una base de datos para cada una de las categorías de identificadores (dirección IP, tupla del dominio y de la dirección IP, tupla del usuario y de la dirección IP, etc.) enviados desde el dispositivo de seguridad de correo electrónico 36.

El motor de reputación 52 busca cada base de datos con el identificador recibido respectivo. Si ya existe una entrada en la base de datos respectiva, entonces el motor de reputación 52 recupera la métrica de reputación almacenada en la base de datos respectiva para ese identificador.

En una presente realización, la métrica de reputación incluye preferiblemente al menos un par de recuentos, un recuento que representa el número total de mensajes recibidos en cualquier sistema de correo electrónico 28 en el entorno 20 que coopera con el motor de reputación 52 y el segundo recuento que representa el número de mensajes recibidos en cualquier sistema de correo electrónico 28 en el entorno 20 que coopera con el motor de reputación 52 que se han identificado como que son mensajes de SPAM. No obstante, como será evidente para los expertos en técnica, la métrica de reputación puede ser cualquier métrica adecuada o un conjunto de métricas tales como un porcentaje o una puntuación numérica producida según una fórmula ponderada adecuadamente, etc. y también puede incluir recuentos de mensajes previos encontrados que contienen virus, recuentos de mensajes mal formados recibidos previamente, recuentos de ataques de recolección de directorios reconocidos, etc.

En el paso 116, las métricas de reputación recuperadas de la base de datos para cada identificador se devuelven al sistema de correo electrónico 28. Las métricas reales devueltas pueden ser una métrica combinada derivada de los datos almacenados en las bases de datos o pueden ser los datos reales almacenados, etc. En el mejor de los casos, el motor de reputación 52 tendrá una métrica de reputación almacenada para cada identificador (es decir, dirección IP; tupla del dominio y de la dirección IP; y tupla del usuario y de la dirección IP) asociado con el mensaje recibido y estas métricas de reputación se pueden usar por sistema de correo electrónico 28 como se describe a continuación.

No obstante, se contempla que, en muchos casos, el motor de reputación 52, por ejemplo, no tendrá una métrica de reputación almacenada para el identificador que representa a una tupla particular del usuario y de la dirección IP. En tal caso, el motor de reputación 52 empleará las métricas que tiene, esto es, las métricas para la dirección IP y la tupla del dominio y de la dirección IP.

De manera similar, se contempla que, en algunos casos, el motor de reputación 52 no tendrá una métrica de reputación para cualquiera de los identificadores que representan una tupla particular del usuario y de la dirección IP o una tupla del dominio y de la dirección IP. En tal caso, el motor de reputación 52 devolverá la métrica de reputación para la dirección IP. También es posible que el motor de reputación 52 no tenga una métrica de reputación almacenada para cualquiera de los tres identificadores, en cuyo caso se devuelve una métrica de reputación NULA al sistema de correo electrónico 28. No obstante, en el mejor de los casos, el sistema de correo electrónico 28 se dota con métricas de reputación para el mensaje recibido para cada una de la dirección IP, la tupla del dominio y de la dirección IP y la tupla del usuario y de la dirección IP.

En el paso 120, el sistema de correo electrónico 28 recibe las métricas de reputación del motor de reputación 52 y el aparato de seguridad 36 hace una determinación en cuanto a si el mensaje recibido es SPAM. Esta determinación se puede hacer de cualquier manera adecuada, como se les ocurrirá a los expertos en la técnica, y en una presente realización de la invención se logra con un proceso de Análisis de Testigo Estadístico Bayesiano que se ejecuta en el aparato de seguridad de correo electrónico 36.

El método real de uso de las métricas de reputación cuando se determina una probabilidad de que el mensaje recibido sea SPAM no está particularmente limitado y una variedad de alternativas será evidente para los expertos

en la técnica, algunas de las cuales son triviales. Por ejemplo, si se devuelven métricas de reputación para cada una de la dirección IP, la tupla del dominio y de la dirección IP y la tupla del usuario y de la dirección IP, y si esas métricas reflejan todas una alta probabilidad de que el mensaje recibido no sea SPAM, entonces hay una alta probabilidad de que el mensaje recibido no sea SPAM y la probabilidad determinada de que el mensaje recibido sea SPAM reflejará esto (es decir, es poco probable que el mensaje sea SPAM).

En un caso más interesante, si se recibe un mensaje y si la métrica de reputación para la dirección IP indica una reputación relativamente escasa (es decir, se han recibido previamente grandes cantidades de SPAM desde esta dirección IP) pero la métrica de reputación para la tupla del dominio y de la dirección IP indica una reputación relativamente buena (es decir, se ha recibido previamente muy poco SPAM desde este dominio en esta dirección IP), la probabilidad determinada de que el mensaje sea SPAM indicará que es probable que el mensaje no sea SPAM. Este tipo de análisis se puede usar para diferenciar entre múltiples dominios alojados en la misma dirección IP donde se usan uno o más dominios para originar SPAM mientras que los otros dominios en la misma dirección IP se usan por usuarios legítimos.

De manera similar, si la métrica de reputación para la tupla del usuario y de la dirección IP es muy favorable (es decir, se ha recibido previamente muy poco, en su caso, SPAM desde este usuario en esta dirección IP) mientras las métricas de reputación para la dirección IP y la tupla para el dominio y la dirección IP son relativamente malas (es decir, se han recibido previamente cantidades altas de mensajes de SPAM) la probabilidad determinada de que el mensaje sea SPAM puede indicar que es probable que el mensaje no sea SPAM. Este tipo de análisis se puede usar para diferenciar entre usuarios buenos y malos alojados en el mismo dominio.

El proceso de determinación de una probabilidad de que un mensaje recibido sea no deseado puede emplear las métricas de reputación devueltas desde el motor de reputación 52 en una amplia variedad de maneras, como se les ocurrirá a los expertos en la técnica. Como será evidente, una variedad de interpretaciones adecuadas se pueden realizar a partir de las métricas de reputación. Específicamente, saber que no se ha observado que una tupla de usuario particular o tupla de dominio envíe mensajes no deseados antes, pero ha estado enviando mensajes deseados puede proporcionar un nivel de confianza razonablemente alto de que se desea un mensaje recién recibido.

En el paso 124, el mensaje recibido se procesa por el sistema de correo electrónico 28, según la determinación en cuanto a si el mensaje recibido es SPAM hecho en el paso 120, según las políticas establecidas para funciones anti-SPAM en el sistema de correo electrónico 28.

En el paso 128, la determinación en cuanto a si el mensaje recibido es SPAM se vuelve a calcular sin usar las métricas de reputación devueltas desde el motor de reputación 52. En el paso 132, esta determinación “libre de reputación” se envía entonces, a través de la red 48, al motor de reputación 52. En una presente realización de la invención, esta determinación de probabilidad de que el mensaje es SPAM es una determinación binaria (por ejemplo, SPAM o NO SPAM), pero también se contempla que otras determinaciones, tales como los valores que representan una probabilidad de que el mensaje sea SPAM, se pueden emplear si se desea.

Se contempla además que las métricas de reputación del motor de reputación 52 se puedan modificar por una variedad de otros procesos, incluyendo realimentación proactiva del destinatario del mensaje. Las técnicas de realimentación del destinatario, tales como proporcionar un control de interfaz de usuario en los clientes de correo electrónico 24 con el que el usuario puede indicar que un mensaje recibido particular se ha identificado incorrectamente como no deseado, o viceversa, son bien conocidas y se contempla que tales técnicas de realimentación también se puedan incluir dentro de la presente invención, como será evidente para los expertos en la técnica.

En el paso 136, el motor de reputación 52 recibe los identificadores de mensaje y la determinación libre de reputación en cuanto a si el mensaje es SPAM y en el paso 140, el método se completa a medida que el motor de reputación 52 actualiza sus métricas de reputación almacenadas para reflejar la determinación de probabilidad de SPAM recibida desde el sistema de correo electrónico 28 en el paso 136.

Si en el paso 116 el motor de reputación 52 no tenía una métrica almacenada para uno o más de los identificadores recibidos, se crean registros adecuados en las bases de datos en el motor de reputación 52 para esos identificadores y esos registros se actualizan para reflejar la determinación de probabilidad de SPAM libre de reputación recibida desde el sistema de correo electrónico 28 en el paso 136.

Este método de dos iteraciones de determinación de una probabilidad de que el mensaje recibido sea SPAM (con métricas de reputación y sin métricas de reputación) se prefiere actualmente para reducir la posibilidad de que se induzca un comportamiento inestable en el motor de reputación 52, o bien intencionalmente por los creadores de SPAM o bien involuntariamente. No obstante, se contempla que se pueden emplear otros mecanismos, tales como realimentación o mecanismos de retardo, o bien además de o bien en lugar de, el método de dos iteraciones, como se les ocurrirá a los expertos en la técnica.

Además de las métricas de reputación de la dirección IP, la tupla del dominio y de la dirección IP y la tupla del usuario y de la dirección IP tratadas anteriormente, se contempla además que la presente invención también puede

devolver una indicación de una probabilidad de que un mensaje recibido sea de un dominio falsificado. Como es bien conocido por los expertos en la técnica, es un asunto relativamente fácil para el creador de un mensaje de SPAM representar el mensaje como proveniente de un dominio distinto del dominio desde el cual se envía realmente y esto se conoce comúnmente como “suplantación de identidad”. Aunque los sistemas tales como “Claves de Dominio” y “SPF” se han desarrollado para más difícil la suplantación de identidad, tales sistemas requieren que la participación activa/los pasos se acometan por el titular del dominio y que muchos titulares de dominios no tomen tales pasos, reduciendo por ello la efectividad de estos sistemas.

Para detectar la suplantación de identidad de dominios, o bien en lugar de usar “Claves de Dominio” o SPF o bien además de, el motor de reputación 52 también puede mantener una base de datos anti-suplantación de identidad de registros que relacionan cada dominio con cada dirección IP desde la cual se han recibido previamente mensajes desde ese dominio. En tal caso, el aparato de seguridad 36 también enviará un identificador de dominio al motor de reputación 52. El motor de reputación 52 usará este identificador de dominio para situar el registro adecuado en la base de datos anti-suplantación de identidad y comparará el identificador de dirección IP enviado, como se ha tratado anteriormente, con los identificadores de dirección IP almacenados en el registro para el dominio identificado. En el paso 116, el motor de reputación 52 también puede devolver entonces una métrica falsa que comprende una indicación en cuanto a si se han recibido previamente mensajes no de SPAM desde el dominio de la dirección IP identificada. Esta métrica falsa se puede establecer cuando el dominio no se ha asociado previamente con la dirección IP identificada y se ha borrado cuando se han asociado previamente el dominio y la dirección IP.

En el paso 120, el sistema de correo electrónico 28 puede usar la métrica falsa, además de las métricas de reputación devueltas, al recalcular una probabilidad de que el mensaje recibido sea SPAM y en el paso 132 el motor de reputación 52 también puede actualizar la base de datos anti-suplantación de identidad, si se requiere.

Como será evidente ahora, la presente invención se refiere a un método y a un sistema para proporcionar un servicio de reputación para su uso en entornos de mensajería de correo electrónico. Las estadísticas, que representan si los mensajes de SPAM se han recibido previamente de direcciones IP, dominios y/o usuarios respectivos, se incorporan en un proceso de toma de decisiones para los mensajes recibidos.

Los sistemas de mensajes que reciben un mensaje reenvían un identificador de la dirección IP de origen del mensaje, un identificador del dominio desde el cual se recibió supuestamente el mensaje y un identificador del usuario desde el cual se recibió supuestamente el mensaje a un motor de reputación.

El motor de reputación mantiene las bases de datos para cada uno de: el identificador de la dirección IP de origen; el identificador la tupla del dominio y de la dirección IP; y el identificador de la tupla del usuario y de la dirección IP. Cada una de estas bases de datos incluye una métrica de reputación asociada derivada de los mensajes recibidos considerados previamente y las determinaciones hechas por los sistemas de mensajes en cuanto a una probabilidad de que sean SPAM.

El motor de reputación devuelve las métricas de reputación asociadas, en su caso, para el identificador de la dirección IP, el identificador de la tupla del dominio y de la dirección IP y el identificador de la tupla del usuario y de la dirección IP al sistema de correo electrónico, que puede calcular entonces una determinación con las métricas devueltas en cuanto a si el mensaje es SPAM. El mensaje se maneja, según la determinación calculada y el mensaje se maneja entonces según una política definida.

Una vez que el mensaje ha sido manejado según la política, el cálculo en cuanto a si el mensaje es SPAM se vuelve a realizar, sin la consideración de las métricas de reputación devueltas desde el motor de reputación para obtener una determinación “libre de reputación”, y esta determinación libre de reputación se reenvía al motor de reputación para usar para actualizar, posiblemente con otra información suministrada desde el destinatario del mensaje u otros métodos, sus bases de datos adecuadamente.

El motor de reputación también puede devolver una métrica falsa al sistema de mensajes si el mensaje se ha originado en una dirección IP desde la cual el motor de reputación no ha visto previamente mensajes originados por el dominio identificado.

Las realizaciones de la invención descritas anteriormente se pretende que sean ejemplos de la presente invención y se pueden efectuar alteraciones y modificaciones a las mismas por los expertos en la técnica, sin apartarse del alcance de la invención que se define únicamente por las reivindicaciones adjuntas a la misma.

REIVINDICACIONES

1. Un método de determinación de una probabilidad de que un mensaje recibido sea un mensaje no deseado, que comprende los pasos de:

(i) recibir un mensaje en un sistema de mensajería;

(ii) reenviar a un motor de reputación un conjunto de identificadores preseleccionados con relación al origen del mensaje, el conjunto de identificadores preseleccionados que incluye una dirección IP desde la cual se originó el mensaje recibido, una tupla de un dominio en el que se originó supuestamente recibido y la dirección IP desde la cual se originó el mensaje recibido, y una tupla de un usuario que supuestamente originó el mensaje y la dirección IP desde la cual se originó el mensaje recibido;

(iii) comprobar las bases de datos en el motor de reputación para determinar métricas de reputación determinadas previamente para los identificadores reenviados y devolver cualquier métrica de reputación determinada previamente al sistema de mensajería, cada métrica de reputación que es indicativa de una cantidad de mensajes no deseados recibidos previamente a través del identificador respectivo;

(iv) hacer una primera determinación en el sistema de mensajería de una probabilidad en cuanto a si el mensaje recibido es no deseado usando un primer conjunto de criterios, incluyendo las métricas de reputación devueltas; y

(v) marcar el mensaje como que es o bien deseado o bien no deseado según la primera determinación.

2. El método según la reivindicación 1 que comprende además los pasos de:

(vi) hacer una segunda determinación en el sistema de mensajería de una probabilidad en cuanto a si el mensaje recibido es no deseado sin usar ninguna métrica de reputación devuelta; y

(vii) reenviar la segunda determinación al motor de reputación para actualizar las bases de datos y las métricas de reputación respectivas para incluir la segunda determinación.

3. El método de la reivindicación 1 donde el paso (iii) también comprende reenviar al sistema de mensajería una métrica falsa que representa si el motor de reputación ha recibido previamente identificadores que indican que uno o más mensajes recibidos previamente del dominio en el que se ha recibido el mensaje recibido supuestamente originado desde la dirección IP en la que se originó el mensaje recibido y donde la primera determinación también emplea la métrica falsa.

4. El método de la reivindicación 1 en donde el sistema de mensajería es un sistema de mensajería de correo electrónico.

5. El método de la reivindicación 1 en donde el sistema de mensajería es un sistema de comunicación de voz basado en SIP.

6. El método de la reivindicación 1 en donde los identificadores para el dominio en el que se originó supuestamente el mensaje recibido y el usuario que supuestamente originó el mensaje se crean a partir de la información del dominio y del usuario mediante una función unidireccional.

7. Un entorno de mensajería que emplea un servicio de reputación al determinar una probabilidad en cuanto a si los mensajes recibidos son no deseados, que comprende:

una pluralidad de servidores de mensajes interconectados por una red de comunicaciones, al menos uno de la pluralidad de servidores de mensajes que incluye una función anti-SPAM para determinar una probabilidad en cuanto a si los mensajes recibidos son no deseados;

una pluralidad de clientes de mensajes conectados a los respectivos de la pluralidad de servidores de mensajes y operables para recibir mensajes desde los mismos;

y un motor de reputación operable para comunicarse con al menos un servidor de mensajes, el motor de reputación que mantiene un conjunto de bases de datos asociando una métrica de reputación con cada uno de un conjunto de identificadores preseleccionados con relación a los orígenes de los mensajes, el conjunto de identificadores preseleccionados que incluye una dirección IP desde la cual se originó el mensaje recibido, una tupla de un dominio en el cual se originó supuestamente el mensaje recibido y la dirección IP desde la cual se originó el mensaje recibido, y una tupla de un usuario que originó supuestamente el mensaje y la dirección IP a partir del cual se originó el mensaje recibido, cada métrica de reputación que es indicativa de una cantidad de mensajes no deseados recibidos previamente a través del identificador respectivo, la función anti-SPAM que opera para reenviar el conjunto de identificadores preseleccionados al motor de reputación que devuelve las métricas de reputación almacenadas en sus bases de datos para cualquiera de los identificadores y la función anti-SPAM que usa las métricas de reputación devueltas para hacer una primera determinación de una probabilidad en cuanto a si un mensaje recibido es no deseado.

8. Un entorno de mensajería según la reivindicación 7 en donde la función anti-SPAM también hace una segunda determinación de una probabilidad en cuanto a si un mensaje recibido es no deseado, la segunda determinación que se hace independiente de las métricas de reputación devueltas, y reenviar esa segunda determinación al motor de reputación para actualizar adecuadamente las métricas de reputación almacenadas en el motor de reputación para reflejar la segunda determinación.

9. Un entorno de mensajería según la reivindicación 7 en donde al menos algunos de los identificadores se crean a partir de la información del mensaje mediante una función unidireccional.

10. Un entorno de mensajería según la reivindicación 7 en donde el sistema de mensajería es un sistema de correo electrónico.

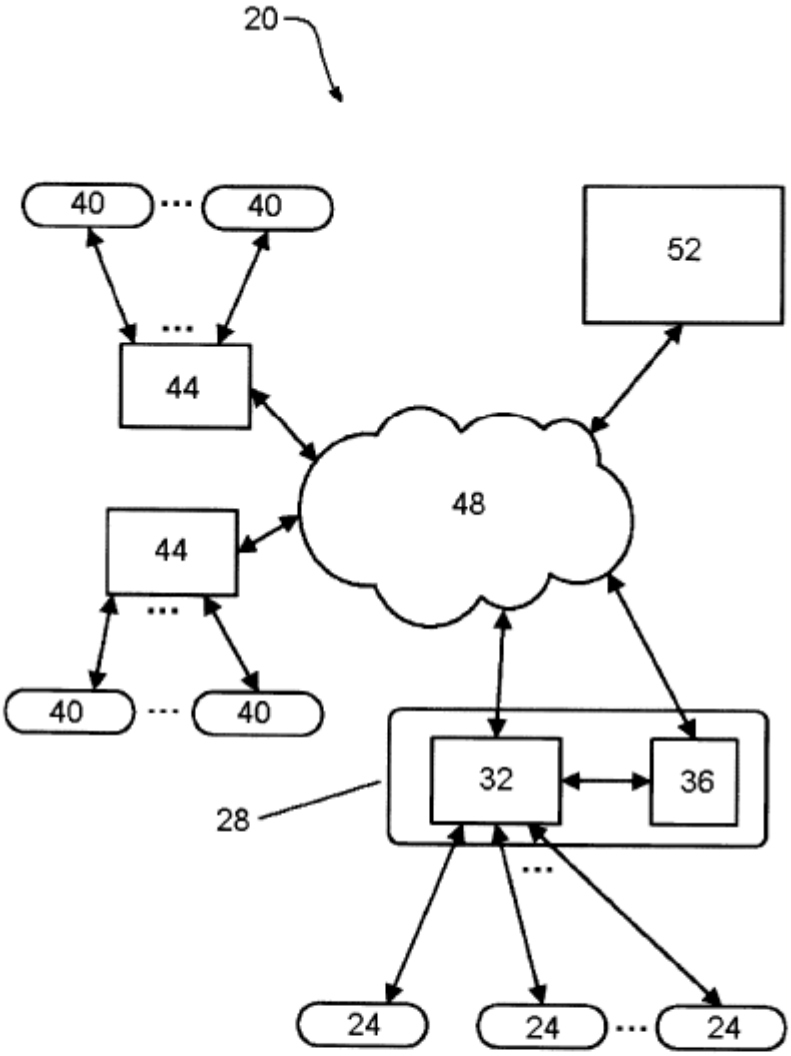

Fig. 1

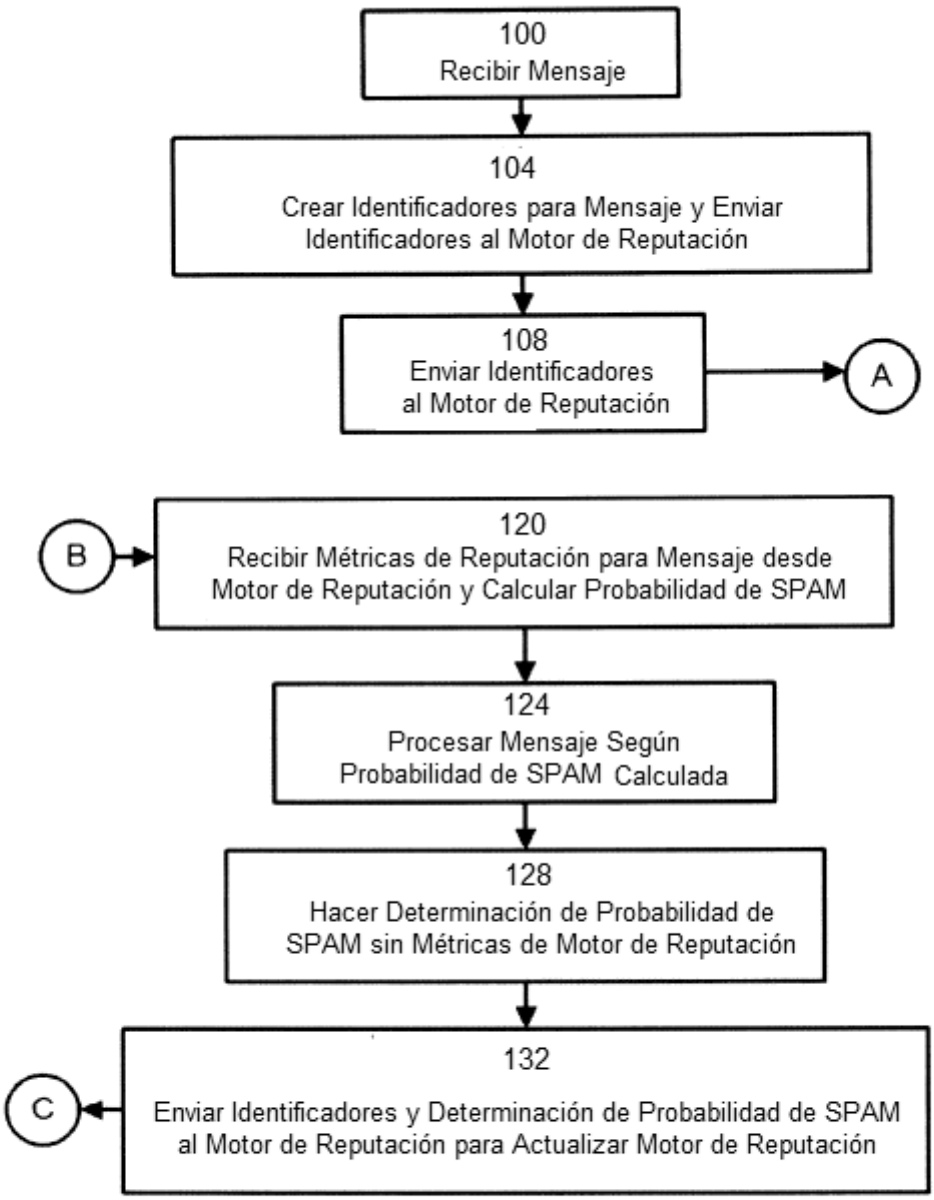
100
Recibir Mensaje
104
Crear Identificadores para Mensaje y Enviar Identificadores al Motor de Reputación
108
Enviar Identificadores al Motor de Reputación
A
B
120
Recibir Métricas de Reputación para Mensaje desde Motor de Reputación y Calcular Probabilidad de SPAM
124
Procesar Mensaje Según Probabilidad de SPAM Calculada
128
Hacer Determinación de Probabilidad de SPAM sin Métricas de Motor de Reputación
132
Enviar Identificadores y Determinación de Probabilidad de SPAM al Motor de Reputación para Actualizar Motor de Reputación
C

Fig. 2

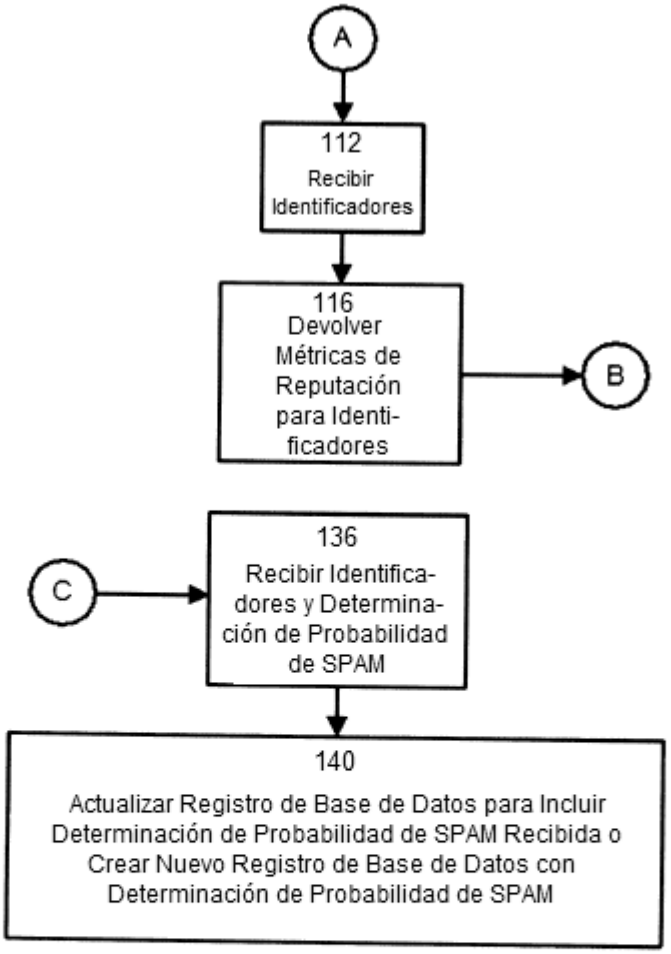
A
112
Recibir Identificadores
116
Devolver Métricas de Reputación para Identificadores
B
C
136
Recibir Identificadores y Determinación de Probabilidad de SPAM
140
Actualizar Registro de Base de Datos para Incluir Determinación de Probabilidad de SPAM Recibida o Crear Nuevo Registro de Base de Datos con Determinación de Probabilidad de SPAM

Fig. 3