(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 428 040**

(51) Int. CI.:

G06F 21/56 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **09.06.2006 E 06744158 (4)**

(97) Fecha y número de publicación de la concesión europea: **04.09.2013 EP 1891571**

(54) Título: **Resistencia a la propagación de código y datos no deseados**

(30) Prioridad:

09.06.2005 GB 0511749

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
05.11.2013

(73) Titular/es:

GLASSWALL (IP) LIMITED (100.0%)
Kensal House, 77 Springfield Road
Chelmsford, Essex CM2 6JG, GB

(72) Inventor/es:

SCALES, NICHOLAS JOHN

(74) Agente/Representante:

UNGRÍA LÓPEZ, Javier

DESCRIPCIÓN

Resistencia a la propagación de código y datos no deseados

Esta invención se refiere a sistemas informáticos y métodos de funcionamiento de tales sistemas para resistir la propagación de código y datos no deseados.

En la década pasada, los sistemas informáticos han sido cada vez más objeto de ataques mediante código no deseado. Los ejemplos más extremos (hasta ahora) de código no deseado son los virus informáticos. Un virus informático, al igual que su homónimo biológico, es capaz de infectar una máquina y a continuación, a partir de ahí, infectar a otros, apropiándose de los recursos del sistema de correo electrónico para enviar mensajes de correo electrónico que contienen el virus desde un ordenador a muchos otros, usando la libreta de direcciones de cada ordenador al que llega.

El ancho de banda desaprovechado resultante es una molestia para los usuarios. Además, muchos virus realizan alguna acción no deseada en cada ordenador al que llegan, lo que puede incluir, por ejemplo, la eliminación de archivos.

Los virus suelen llegar como código ejecutable, en un archivo adjunto por separado, pero también pueden ocultarse en partes de un correo electrónico, de modo que pueden activarse sin necesidad de que un usuario separe y ejecute el código explícitamente. Muchas aplicaciones, tales como procesadores de texto, hojas de cálculo y bases de datos, incluyen potentes lenguajes de secuencias de instrucciones, que permiten a un archivo que parece ser un documento incluir una secuencia de instrucciones capaz de realizar ciertas operaciones. Los creadores de virus han hecho uso de tales lenguajes de secuencias de instrucciones para escribir los virus de secuencia de instruciones, de manera que los adjuntos de un correo electrónico que incluyen archivos que parecen ser documentos pueden albergar un virus oculto.

Los virus no son la única forma de código no deseado. Es común que los programas "gratis" se distribuyan con "programas espías" ocultos, que pueden, por ejemplo, instalarse de manera encubierta en el ordenador de un usuario y, posteriormente, pueden informar de los sitios web visitados u otras transacciones a un ordenador remoto. Algunos programas espías, provocan la visualización de publicidad no deseada. Algunos programas espías tratarán de provocar que un módem marque repetidamente un número de tasa alta, en el que el creador del programa espía recibe ingresos de un operador de telecomunicaciones. Otros tipos de código perjudicial incluyen programas informáticos dañinos, gusanos y puertas de atrás.

Mientras que los virus se propagan por sí mismos de un ordenador a otro, otras formas de código no deseado se distribuyen por correo electrónico basura, por distribución oculta en el disco, o, cada vez más, por la descarga de un sitio web visitado sin darse cuenta. Todos estos tipos de código no deseado tienen en común el hecho de que su existencia o su propósito real es ocultarse de los propietarios y usuarios de los ordenadores a los que van dirigidos. Mientras que algunos tipos son relativamente inofensivos, otros tienen la capacidad de destruir datos empresariales valiosos y, por lo tanto, se ha desarrollado una industria para el suministro de software antivirus.

El software antivirus, como se conoce en la actualidad, consiste en un programa que se ejecuta en el ordenador que debe protegerse. Tales programas funcionan normalmente en modo monitor, en el que los archivos que deben accederse se comprueban en busca de virus cada vez que se accede al archivo, y en un modo de exploración en el que se exploran todos los archivos en una localización determinada (por ejemplo, una unidad de disco). Los proveedores de programas antivirus monitorizan los brotes de virus y, cuando se detecta un nuevo virus, las compañías de los programas de antivirus analizan el virus y extraen los datos que se pueden usar para detectar el virus. A continuación, estos datos se pondrán a disposición de los ordenadores que ejecutan el programa antivirus específico interesado; normalmente, proporcionándole en el sitio web de la compañía del programa de antivirus para su descarga.

Los virus se detectan de diversas maneras diferentes. Puede almacenarse una cadena de código característico que forma parte del virus y explorarse los archivos entrantes en busca de la presencia de esa cadena, que por lo tanto actúa como una "firma" o "huella digital" para el virus. Como alternativa, los virus pueden detectarse por su comportamiento previsto; el código fuente o los archivos de secuencia de instrucciones pueden tratarse para detectar operaciones predeterminadas que son características de un virus.

Desafortunadamente, los virus, al igual que sus homólogos biológicos, pueden "mutarse" fácilmente, pequeños cambios en el código, equivalentes a la sustitución de letras mayúsculas y minúsculas, pueden cambiar la firma del virus. Por lo tanto, los archivos de datos para detectar virus, mediante cualquier método, están llegando a ser extremadamente grandes, y el tiempo empleado por los programas antivirus está aumentando proporcionalmente a medida que está creciendo el número de firmas o reglas que deben comprobarse. Mientras que esto puede ser aceptable en el modo de exploración de virus, en el modo monitor está añadiendo una latencia cada vez mayor al tiempo necesario para acceder a los archivos. Además, como las descargas se hacen más grandes y se requieren con más frecuencia, el riesgo de que un usuario fracase al descargar las actualizaciones necesarias, y por lo tanto,

se quede sin protección contra el virus más reciente (y por lo tanto el más peligroso), es alta.

Por lo tanto, la presente invención tiene un enfoque totalmente diferente a la protección frente a código no deseado.

El documento US2005/081057 desvela un método para prevenir la explotación de un mensaje de correo electrónico y el sistema del mismo.

Sumario

Se proporciona un método implementado por ordenador de resistencia a la propagación de código y datos no deseados en una electrónica como se establece en la reivindicación 1.

De acuerdo con un aspecto de la presente invención, se proporciona un método de recepción de un archivo electrónico que contiene datos de contenido en un formato de datos predeterminado, comprendiendo el método las etapas de: recibir el archivo electrónico, determinar el formato de datos, tratar los datos de contenido, a fin de determinar si se ajustan al formato de datos predeterminado, y si los datos de contenido se ajustan al formato de datos predeterminado, regenerar los datos tratados para crear un archivo electrónico regenerado en el formato de datos.

También se proporcionan los sistemas informáticos correspondientes, los programas y los medios para realizar tales programas.

Una realización de la invención funciona para analizar cada archivo recibido y, a continuación, reconstituir a partir de él un archivo de sustitución. Debido a que el archivo original no se almacena el mismo directamente, o se accede, en el ordenador que debe protegerse, no es, el mismo, capaz de hacer daño a ese ordenador. Puede, por ejemplo, almacenarse en una forma de bit invertida u otra forma en la que no se puede ejecutar. Por otro lado, se genera el archivo de sustitución usando un generador de rutina que puede generar solo código "limpio" y los datos. Por tanto, es incapaz de generar código no deseado coincidente con cualquier código en un archivo recibido.

Parte de la presente invención puede basarse en una nueva aplicación de algunas verdades conocidas desde hace tiempo acerca de los archivos de ordenador. La gran mayoría de los archivos que se importan en un ordenador hoy en día están en formatos de archivo normalizados. Los programas propietarios crean sus propios formatos de archivo (y los datos destinados a usarse por los programas deben ajustarse a esos formatos) pero hay suficiente demanda para el intercambio de datos entre los distintos programas propietarios que, en primer lugar, un programa propietario se suministra a menudo con filtros de importación para leer los datos escritos por otros, y, en segundo lugar, existen varios formatos que no están asociados con ningún programa propietario. Ejemplos de tales formatos genéricos son el texto ASCII, el formato de texto enriquecido (RTF), el lenguaje de marcado de hipertexto (HTML) y el lenguaje de marcado extensible (XML).

Por lo tanto, los datos en los archivos deben ajustarse con precisión a las normas rígidas si deben leerse por cualquier programa de aplicación, y los formatos usados por los diferentes archivos son ampliamente conocidos. Los presentes inventores se han dado cuenta de que, aunque los formatos usados por los archivos permiten una amplia variación, la gran mayoría de los archivos contienen datos que cumplen algunas limitaciones pragmáticas relativamente reducidas. Por ejemplo, la mayoría de los sistemas operativos y las aplicaciones aceptan títulos de ficheros de gran longitud, pero la mayoría de los usuarios, la mayor parte del tiempo, usan nombres de archivo cortos y fácilmente reconocibles.

En consecuencia, el análisis realizado por una realización de la presente invención puede comprender detectar si los datos que de otro modo se ajustan a la especificación del tipo de archivo supuesto violan los límites pragmáticos. Estas restricciones del "mundo real" permiten a la presente invención detectar los archivos aceptables “normales”. Cualquier contenido de archivo que no corresponda a los límites pragmáticos de este tipo no se pasa al programa generador y por lo tanto, no alcanza el ordenador de los usuarios en una forma ejecutable.

Por lo tanto, se verá que una realización de la presente invención funciona de una manera fundamentalmente diferente a los programas antivirus conocidos. Los programas antivirus conocidos tienen por objeto detectar virus, y pasan todo lo que no se detecta como un virus. Por lo tanto, siempre fallan en proteger al usuario del peligro más grande; a saber, el de los virus desconocidos. Cada nuevo virus que se acomete ya debe haber infectado a un número de ordenadores antes de que llegue a la atención de las compañías de antivirus.

Además, aun cuando se instale el software antivirus, y se posea un conjunto de actualizaciones de los datos detectados, los virus se almacenarán normalmente en el disco duro u otro medio del ordenador protegido antes de que puedan detectarse por el software antivirus. Si, por alguna razón, el software antivirus falla en su ejecución, el virus está en su lugar y puede activarse.

La solicitud publicada de Estados Unidos US 2003/0145213 desvela un sistema donde se detecta una secuencia de instrucciones o código malicioso en un archivo. A continuación, el archivo se reconstruye en una plantilla y el código

malicioso se elimina de la plantilla para proporcionar una versión limpia del archivo.

A modo de contraste completo, la presente invención no tiene por objeto detectar virus, o incluso rechazar el comportamiento general similar a un virus. En su lugar, puede rechazar todos los archivos entrantes por completo, y sustituir en su lugar, cuando sea posible, los archivos generados que no pueden contener código y datos no deseados. Por lo tanto, puede evitarse que el código y los datos no deseados alcancen el disco duro del ordenador para protegerse en forma ejecutable, y no pueden propagarse desde un ordenador a otro.

En este punto, puede mencionarse que la solicitud publicada de Estados Unidos 2003/229810 desvela una propuesta de un "cortafuegos óptico" para la protección contra los virus. Por razones que se harán evidentes en breve, no se cree que este sistema se haya puesto en práctica (o que podría ponerse en práctica). Describe un sistema en el que un ordenador cortafuegos recibe un archivo tal como un archivo de imagen, y muestra la imagen en la pantalla del ordenador cortafuegos. Una matriz sensora óptica explora la imagen y a continuación la imagen explorada se suministra al destinatario previsto. Cualquier virus que se ocultase en la imagen no se muestra, y por consiguiente, no se transmite en la imagen explorada. En una variante, puede usarse un mapa de bits de la pantalla en lugar de una visualización de la pantalla real.

Por diversas razones, el cortafuegos "acoplador óptico" proporcionado en la solicitud de patente de Estados Unidos mencionada anteriormente no podía proporcionar una protección eficaz y segura contra los virus.

Por ejemplo, la reproducción usando el software de reconocimiento óptico de caracteres (OCR) puede proporcionar información inexacta. Además, la reproducción de imágenes usando la técnica de vídeo puede proporcionar imágenes de calidad inferior a la prevista. También, el ordenador que recibe el archivo de entrada puede llegar a infectarse si el archivo de entrada contiene un virus.

Por otro lado, analizando y, a continuación regenerando los archivos en lugar de ejecutarlos, mostrarlos, y explorarlos ópticamente, una realización de la presente invención es capaz de proporcionar archivos de sustitución que en la gran mayoría de los casos emulan estrechamente al archivo original (si está libre de código no deseado) con el fin de hacer la sustitución transparente.

Los formatos de archivo varían en su complejidad. En un extremo, los archivos de texto tienen un formato simple. Los archivos que pueden contener secuencias de comandos o secuencias de instrucciones (como archivos de procesadores de texto u hojas de cálculo) son de complejidad intermedia, mientras que los archivos que contienen código solo pueden analizarse plenamente mediante un analizador de código. Mientras que tal análisis de código es, a la larga, posible de acuerdo con la presente invención, las realizaciones de la invención pueden funcionar convenientemente para eliminar todas las secuencias de instrucciones y de comandos de los archivos de documentos, y no pasar cualquiera de los archivos que constan únicamente de programas, código, secuencias de instrucciones o comandos.

Será evidente inmediatamente que habrá ocasiones frecuentes en que los usuarios pueden desear recibir tales archivos. Por consiguiente, en una realización preferida, la presente invención puede funcionar a lo largo de una parte del filtro que está dispuesto para filtrar los archivos mediante una fuente, de manera que siempre pasan los archivos (o archivos de un tipo determinado) de ciertas fuentes, y rechazan tales archivos de otras.

Por lo tanto, mientras que una realización de la presente invención puede bloquear que los usuarios reciban código en los archivos de todas las fuentes, el filtro en paralelo permite solo tales archivos que proceden de fuentes conocidas. Por lo tanto, los usuarios pueden recibir, por ejemplo, archivos de los administradores de sistemas o sitios web certificados que serían rechazadas por la invención. Identificando solo esas fuentes de las cuales un usuario desea recibir el código, la presente invención puede bloquear el código no deseado.

Debido a que la presente invención puede funcionar detectando la conformidad con las normas de archivo, y el comportamiento de usuario normal, en lugar de detectando virus, las actualizaciones frecuentes son innecesarias; tales actualizaciones solo son necesarias en el momento donde grandes cambios para una ganancia convencional de aceptación generalizada, o cuando el comportamiento del usuario ha cambiado básicamente, ambas son procesos lentos, en comparación con la velocidad frenética con la que se deben distribuir las actualizaciones de antivirus. Del mismo modo, ya que el número de pruebas que se deben realizar se mantiene más o menos estable en el tiempo, no hay aumento en el tiempo en la latencia para los programas de partida.

Estos y otros aspectos, realizaciones y ventajas de la invención se discutirán en la siguiente descripción y en las reivindicaciones. Se describirán ahora las realizaciones de la invención, a modo de ejemplo solamente, con referencia a los dibujos adjuntos en los que:

La figura 1A muestra un diagrama de bloques de un sistema de archivo electrónico de acuerdo con una realización de la presente invención;
La figura 1B muestra un sistema informático adaptado para su uso en las realizaciones de la presente invención;

La figura 1C muestra un diagrama de flujo de un proceso de acuerdo con una realización de la presente invención;
La figura 2 muestra un diagrama de bloques de un sistema de correo electrónico de acuerdo con una segunda realización de la presente invención;
La figura 3 muestra un diagrama de flujo de un proceso de acuerdo con una segunda realización de la presente invención;
La figura 4 muestra un esquema de ejemplo de las diferentes partes que forman un correo electrónico;
La figura 5 muestra un diagrama de flujo de un proceso de acuerdo con una tercera realización de la presente invención.

Primera realización

Se indica un esquema del sistema básico que muestra una primera realización de la presente invención en la figura 1A. Un archivo 101 electrónico se crea en una fuente y se transmite a través de un medio 103 de transmisión. El medio 103 de transmisión puede ser cualquier medio adecuado para transmitir archivos electrónicos, incluyendo sistemas cableados y sistemas inalámbricos. El archivo 101 electrónico pasa a través del medio 103 de transmisión de la manera normal hasta que alcanza su destino. En esta realización, se instala una aplicación 105 AV (antivirus) en un sistema de destino. La aplicación 105 AV funciona de tal manera que no se permite a los datos dentro del archivo electrónico entrante entrar en el sistema 107 operativo de destino hasta que se han analizado contra un formato permitido predefinido, y, si se determina que los datos son admisibles, se regeneran. Por lo tanto, la aplicación 105 AV determina si se debe permitir al archivo 101 electrónico pasar a través del sistema 107 operativo.

La figura 1B muestra un sistema informático adaptado para implementar las realizaciones de la presente invención. El ordenador 109 recibe el archivo 101 electrónico de entrada en una interfaz 111 de entrada. La interfaz 111 de entrada está conectada a un microprocesador 113, que está dispuesto para realizar diversos procesos en el archivo recibido. El microprocesador 113 incluye un analizador 115. El microprocesador 113 está conectado además a un dispositivo 117 de memoria, a una unidad 119 de disco y a un número de interfaces (121, 123) para habilitar las conexiones que deben hacerse a los dispositivos de salida tales como una pantalla 125 y un teclado 127.

De manera que no se permite a los archivos ejecutables entrantes ejecutarse de forma automática cuando entran en la aplicación AV, el sistema se encarga de que los datos que componen los archivos electrónicos entrantes se almacenen en la memoria en cualquier formato codificado adecuado.

En esta realización, el método codificado invierte el orden de los bits en un byte. Es decir, los bits de 0 al 7 se reciben en orden, pero se almacenan en una manera inversa de bit de tal manera que el bit 0 se cambia con el bit 7, el bit 1 se cambia con el bit 6, el bit 2 se cambia con el bit 5 y el bit 3 se cambia con el bit 4. Por lo tanto, como un ejemplo, un byte que comprende 1 0 1 1 0 0 0 0 se almacenaría en el orden siguiente: 0 0 0 0 1 1 0 1. De esta manera, cualquier código ejecutable no es capaz de ejecutarse de forma automática y por lo tanto, cualquiera de los archivos electrónicos infectados no son capaces de infectar a la aplicación AV o al sistema operativo de destino.

Como alternativa a la aplicación AV que se localiza en el destino previsto del archivo, la aplicación AV puede localizarse también en la fuente, en alguna parte en el medio de la transmisión o en otro lugar, siempre que sea capaz de analizar el archivo electrónico en un punto a lo largo de su trayectoria de transmisión.

La figura 1C muestra un diagrama de flujo de las etapas básicas realizadas por la aplicación 105 AV en esta primera realización con el fin de determinar si el archivo 101 electrónico se le permite pasar a través del sistema 107 operativo de destino. En la etapa S109, el archivo 101 electrónico se introduce en la aplicación 105 AV usando cualquier medio adecuado. El medio de entrada puede variar dependiendo del tipo de archivo electrónico que se recibe y el medio sobre el que se está transmitiendo. En esta realización, se acepta el archivo 101 electrónico en la aplicación AV.

En la etapa S111, el análisis se realiza mediante un dispositivo de análisis de conformidad para determinar si el archivo 101 electrónico se ajusta a un formato predeterminado. La aplicación AV se diseña solo para permitir pasar a los archivos electrónicos que se ajustan a uno de una pluralidad de formatos conocidos almacenados, admisibles, predefinidos. Generalmente, un archivo se compone de datos de contenido codificados y dispuestos de acuerdo con una especificación de tipo de archivo que comprende un conjunto específico de reglas, teniendo asociado cada tipo de archivo (texto, HTML, XML, hoja de cálculo, y así sucesivamente) un conjunto de reglas. Los tipos de archivo comunes se indican algunas veces mediante el sufijo del nombre de archivo (por ejemplo, .pdf, .txt, .doc), y también o como alternativa por los primeros bytes de datos en el archivo. Muchos tipos de archivos incluyen un encabezamiento que indica alguna cosa acerca de la estructura del archivo, seguido de los datos de contenido (por ejemplo, texto, números, datos de audio o de imagen).

Los datos de contenido pueden incluir parámetros (por ejemplo, etiquetas para indicar que los datos de contenido se deben presentar en negrita). Las reglas que componen la especificación del tipo de archivo pueden especificar los valores o el intervalo de lo que tales parámetros pueden asumir. También pueden especificar, por ejemplo, los valores permitidos o el intervalo de valores que los datos contenidos pueden asumir.

Un programa de aplicación capaz de abrir un archivo de un tipo específico incluye un analizador para aplicar las reglas que componen la especificación del tipo de archivo a un archivo, para extraer los datos de contenido para la presentación o el procesamiento. Por ejemplo, una aplicación de procesamiento de texto puede ser capaz de abrir archivos en su formato de archivo propietario (por ejemplo, Microsoft Word™), los formatos de archivos propietarios de otras aplicaciones de procesamiento de textos, y los formatos de archivos genéricos tales como el formato de texto enriquecido (RTF), ASCII y HTML. Un programa de aplicación capaz de almacenar datos de contenido como un archivo de un tipo específico incluye un generador para aplicar a los datos de contenido las reglas que componen la especificación del tipo de archivo para generar un archivo en el formato requerido.

En la presente realización se almacena, para cada tipo de archivo, el formato predeterminado. El formato predeterminado, generalmente, incluye las reglas que componen la especificación del archivo. Sin embargo, los formatos predeterminados solo incluyen las reglas relativas a las partes usadas frecuentemente del formato. Además, los formatos predeterminados incluyen reglas adicionales que limitan los valores y/o los intervalos que pueden tomar el contenido y los parámetros, de manera que solo incluyen comúnmente y frecuentemente los valores e intervalos usados. Por lo tanto, solo pueden analizarse esas partes de un archivo de un tipo dado que consisten exclusivamente de datos y parámetros que ocurren frecuentemente o comúnmente de acuerdo con el correspondiente formato predeterminado almacenado de esta realización.

Ejemplos de componentes de tipos de datos a los que no se les permite pasar a través del sistema (debido a que los formatos predeterminados no incluyen las reglas relativas a los mismos, ya que se usan con poca frecuencia) son secuencias de instrucciones complejas en archivos de tratamiento de texto y tramas-I en páginas HTML. Ejemplos de valores de datos de uso poco frecuente que no se les permite pasar a través del sistema (debido a que los formatos predeterminados se limitan a los valores que los excluyen) son caracteres de control en un archivo ASCII que no sean los caracteres de uso común TAB, CR/LF y LF.

El dispositivo de análisis de conformidad determina si el archivo electrónico está en el formato que dice que está, y que todos los parámetros se ajustan al formato predeterminado asociado con ese tipo de archivo electrónico específico. Si el archivo electrónico no se ajusta a alguno de los formatos predeterminados no se regenera, y entonces se bloquea de manera eficaz, y preferentemente se elimina en la etapa S113. Sin embargo, si el archivo electrónico se ajusta al formato predeterminado, se extraen de él los datos de contenido (y se almacenan temporalmente en una estructura de datos), y se regenera (a partir de la estructura temporal de datos) mediante el dispositivo de análisis de conformidad en el formato predeterminado asociado con ese tipo de archivo electrónico, para componer un archivo de sustitución, como se muestra en la etapa S115.

A continuación, se reenvía el archivo electrónico regenerado, por ejemplo, al sistema operativo en la etapa S117 con el fin de que se procese de una manera normal. Todos los datos de contenido que pueden extraerse a partir del archivo usando las reglas que componen el formato predeterminado se extraen y se regeneran, y por lo tanto, las partes que no se pueden extraer, no pueden regenerarse.

De esta manera, debido a la comprobación de conformidad y a la regeneración del archivo, los virus son incapaces de entrar e infectar el sistema operativo; de hecho, nada pero los datos de contenido en un formato que ocurre comúnmente se extraen y, por consiguiente se regeneran.

En las situaciones donde un mensaje electrónico puede romperse en sub-partes, algunas sub-partes del mensaje electrónico pueden ajustarse a un formato predeterminado, mientras que otras sub-partes no pueden ajustarse. En situaciones como ésta, la aplicación AV determina si el número total de sub-partes que se ajustan cumple una prueba de sustancialidad (por ejemplo, si la mayoría, o las partes más importantes, se ajustan), y si es así, regenera las sub-partes del mensaje electrónico que se ajustan.

Las sub-partes del mensaje que no se ajustan no se regenerarán. En su lugar, la aplicación AV inserta un texto de advertencia pertinente en el mensaje electrónico informando al destinatario de que a parte del mensaje no se le permitió pasar. Como opción, este texto de advertencia puede indicar las razones para no permitir que la sub-parte pase.

Además, también puede bloquearse una parte dentro de una sub-parte del archivo electrónico, es decir, no regenerada y preferentemente eliminada, si no se ajusta al formato predeterminado permitido para esa parte. Esto es, por ejemplo, si una cadena de caracteres en un archivo electrónico ASCII incluye un carácter de control (por ejemplo, el carácter “BEL”), esta cadena de caracteres puede sustituirse con una advertencia de texto insertado por la aplicación AV informando al destinatario previsto que la cadena se ha quedado fuera de esta parte del archivo electrónico regenerado porque la parte no se ajusta al formato predeterminado. El dispositivo de análisis de conformidad no busca específicamente los caracteres de control que no están permitidos (por ejemplo, el carácter “BEL”), pero en cambio pasa solo aquellos caracteres de control que están permitidos, como se define mediante el formato permitido predeterminado.

Como alternativa, el carácter de control que no se ajusta podría sustituirse por un espacio o eliminarse completamente. Las diferentes opciones elegidas dependen de, por ejemplo, el entorno en el que se está ejecutando

la aplicación AV y cómo de importante es para al menos el mínimo de información de conformidad que debe permitirse pasar a través de la aplicación AV a su destino.

Ahora se describirá una alternativa adicional a la primera realización. Tras la determinación por la aplicación AV de que un archivo electrónico o una sub-parte del mismo, no se ajusta y por lo tanto no se le permite pasar a través del sistema operativo de destino, el archivo electrónico original se pasa a una aplicación de filtro de amenazas que determina si hay alguna amenaza asociada con el archivo electrónico, o la sub-parte del mismo.

La determinación se realiza en base a lo que el sistema espera recibir desde ciertas fuentes. El sistema realiza esta determinación revisando una lista de tipos de datos contra una lista predeterminada de las fuentes almacenadas en la memoria, para ver si el tipo de datos se acepta desde esa fuente; en otras palabras, los correos electrónicos se filtran por la fuente. Por lo tanto, si se reciben los archivos que contienen datos que no se ajustan desde la misma fuente, donde los datos que no se ajustan se conocen no como una amenaza, se permite que los datos que no se ajustan originales pasen al sistema operativo. De esta manera, el sistema que comprende la aplicación AV y la aplicación de filtro de amenazas permite dinámicamente que la mayoría de los archivos electrónicos seguros pasen a su destino previsto.

Segunda realización

A continuación, se describe la segunda realización, los archivos electrónicos son correos electrónicos transmitidos a través de Internet, desde un emisor a un proveedor de servicios de Internet (ISP). El ISP reenvía los correos electrónicos a un servidor de cliente de correo electrónico, con lo que a la recepción, el servidor de cliente de correo electrónico reenvía el correo electrónico al buzón de entrada del destinatario.

La figura 2 muestra un esquema de un sistema de correo electrónico de acuerdo con esta realización en la que se incorpora la aplicación AV de la presente invención. Se reenvía un correo electrónico mediante un remitente desde una localización 201 fuente. El correo electrónico se reenvía a través de Internet 203 a un proveedor 205 de servicios de Internet (ISP), determinado por el nombre de dominio incorporado dentro del correo electrónico. El servidor 207 de cliente de correo electrónico del receptor se conecta al ISP 205 a través de conexiones abiertas directas. La primera conexión es una conexión 209 de salida del protocolo para la transferencia simple de correo electrónico (SMTP) para reenviar el correo electrónico saliente desde el servidor 207 de cliente de correo electrónico al ISP 205. Una segunda conexión es una conexión 211 entrante POP (Protocolo de Oficina Postal), que recupera el correo electrónico desde el ISP 205.

La aplicación 105 AV está situada en el ISP 205. La aplicación 105 AV reside en los puertos de entrada/salida conectados al servidor 207 de cliente de correo electrónico del destinatario, con el fin de analizar todos los correos electrónicos entrantes y salientes que se envían y se reciben por el servidor 207 de cliente de correo electrónico.

En esta realización, la aplicación 105 AV es una parte de código de ordenador, que se implementa usando técnicas de programación de ordenador conocidas. Todos los correos electrónicos que se envían al servidor 207 de cliente de correo electrónico deben pasar a través de la aplicación 105 AV antes de que los correos electrónicos sean capaces de entrar en el servidor 207 de cliente de correo electrónico. De forma similar, todos los correos electrónicos reenviados por el servidor de cliente de correo electrónico al ISP 205 deben pasar a través de la aplicación 105 AV antes de que entren en el ISP 205.

La aplicación 105 AV analiza el mensaje de correo electrónico entrante tratando los datos a medida que entran en la aplicación. Como en la primera realización, los datos se almacenan en un modo codificado con el fin de detener la ejecución de cualquiera de los archivos ejecutables. La aplicación 105 AV determina si las partes separadas del correo electrónico entrante se ajustan a un formato admisible predeterminado, y, si la parte se ajusta, regenera cada parte del mensaje de correo electrónico. Por lo tanto, cualquier virus dentro de cualquier correo electrónico no se le permite pasar para infectar el sistema del destinatario, ni pasar desde el sistema del destinatario al ISP.

Se usa un dispositivo de análisis de conformidad en esta realización para analizar los tipos de datos específicos para ver si se ajustan a un formato predefinido para ese tipo de datos (como se discutió en la primera realización) y extraer esos datos de contenido que se ajustan. A continuación, el dispositivo de análisis de conformidad regenera los datos usando el formato permitido predefinido para ese tipo de datos. Cada tipo de datos se analiza y se regenera mediante su propio dispositivo de análisis de conformidad específico.

Cada dispositivo de análisis de conformidad ejecuta un conjunto de reglas específicas sobre los datos dependiendo del tipo de datos recibidos. Las reglas se definen mediante la especificación oficial predefinida para el tipo de archivo, y los ejemplos que ocurren habitualmente en el mundo real (y por lo tanto seguro) de los tipos de datos conocidos. En general, las reglas permiten solo un subconjunto de archivos que se ajusten a la especificación del tipo de archivo, pero pueden relajar ciertas reglas de la especificación oficial cuando éstas se incumplen habitualmente. Por ejemplo, las direcciones de correo electrónico no deberían contener espacios, pero algunas aplicaciones de correo electrónico más populares incumplen esta regla, por lo que los correos electrónicos que violan la especificación en este respecto son comunes, y por lo tanto el formato predeterminado para analizar

correos electrónicos de acuerdo con esta realización acepta direcciones de correo electrónico que contienen un espacio, y por lo tanto los análisis de la realización y los extractos de tales direcciones de correo electrónico.

También, el dispositivo de análisis de conformidad puede comprobar un parámetro determinado dentro de un archivo de datos. Por ejemplo, si el encabezado indica que el archivo es un archivo RTF (formato de texto enriquecido), a continuación, se leen los primeros bytes de datos para determinar si esto es correcto.

La figura 3 muestra un diagrama de flujo de cómo funciona un sistema que incorpora una aplicación AV de acuerdo con esta realización. Como puede verse en la figura 3, en la etapa S301, el correo electrónico se recibe en el ISP a través de la conexión entrante SMTP.

En la etapa S303, un dispositivo de análisis de conformidad del protocolo realiza un proceso para leer el formato básico del correo electrónico entrante, y regenera el correo electrónico de manera que se ajusta a los protocolos de correo electrónico básicos. Un lector de correo electrónico que no se ajusta lee el correo electrónico. A continuación, los datos leídos se pasan a un escritor de correo electrónico que los ajusta a los protocolos de correo electrónico básicos. De esta manera, las no conformidades comunes se convierten en un correo electrónico que se ajusta. Por ejemplo, si la dirección de correo electrónico del destinatario está mal formada, el escritor del correo electrónico la reescribe de forma que se ajuste.

Un ejemplo adicional es cuando se recibe un mensaje de correo electrónico sin un encabezado "De:". En este caso, el mensaje de correo electrónico se encapsula en un mensaje de correo electrónico completamente nuevo con un encabezado "De:".

También se realizan otros parámetros en el correo electrónico para ajustarle. Por ejemplo, la longitud de la línea, que se usen los códigos de caracteres ASCII correctos, que se use la codificación de base 64 correcta cuando sea apropiado, la información de encabezado intacta (“Para:”, “Asunto:”, etc.), un espacio entre el encabezado y el cuerpo del correo electrónico, y así sucesivamente.

Si el correo electrónico está tan mal formado parte de él no se puede volver a reescribir, entonces se determina si todavía existe un correo electrónico razonable si se ha perdido la parte que no se ajusta. Si se determina que el proceso todavía resulta en un correo electrónico razonable, se reescribe el correo electrónico con la parte que falta que no se ajusta. Se puede insertar un texto de advertencia en su lugar.

Además, el dispositivo de análisis de conformidad del protocolo puede rechazar la totalidad del correo electrónico. Por ejemplo, si el dispositivo de análisis de conformidad del protocolo detecta que la codificación de base 64 que no se ajusta se está usando en gran parte de los datos dentro del correo electrónico, el correo electrónico se rechaza completamente en la etapa S305.

Si el dispositivo de análisis de conformidad del protocolo determina que el correo electrónico se ajusta a los protocolos de correo electrónico, se regenera mediante el dispositivo de análisis de conformidad del protocolo y se pasa a la siguiente etapa en el proceso.

Todos los correos electrónicos deberían ajustarse a la norma RFC actual para el correo electrónico (es decir, el RFC 822 y sus sucesores). Esta norma define cómo se forma el correo electrónico. Después de que el correo electrónico pase a través del dispositivo de análisis de conformidad del protocolo, el dispositivo de análisis de conformidad del RFC 822 comprueba para ver si el correo electrónico se ajusta a la norma RFC 822. El dispositivo de análisis de conformidad del RFC 822 realiza esta comprobación de conformidad, rompiendo primero el correo electrónico en sus partes componentes separadas encontrando los límites dentro del correo electrónico (como se discute más adelante), y a continuación, tratando cada componente del correo electrónico para ver si se ajusta al RFC 822.

Se entenderá que se necesitarían actualizaciones cuando se actualice la norma RFC para asegurarse de que el dispositivo de análisis de conformidad del RFC 822 es capaz de comprobar la conformidad de todos los tipos de datos conocidos.

Como es bien sabido, un correo electrónico se compone de un número de partes separadas, como se muestra, por ejemplo, en la figura 4. El correo electrónico empieza con un encabezado 401 RFC 822, que define una serie de campos, tales como "De:", "Para:" y "Asunto:", etc. El siguiente es el encabezado 403 MIME, que define una serie de campos para su uso en el protocolo de extensión, tal como el "Tipo de contenido:” que define el texto usado para indicar el límite entre las diferentes partes del correo electrónico.

Después de los encabezados (401 y 403), se indica el primer límite 405. La siguiente parte del correo electrónico empieza con un encabezado 407 MIME adicional, que define el formato usado en esta parte. En este ejemplo, esta parte comprende el asunto del texto que se mostrará en un formato de texto. Por lo tanto, sigue el bloque 409 de texto. Al final del bloque 409 de texto está un límite 411 adicional.

Un encabezado 413 MIME adicional indica el formato que estará en la siguiente parte del correo electrónico. En este ejemplo, la siguiente parte del correo electrónico es un bloque 415 mezcla de texto y HTML formateado. Un límite 417 adicional indica el final de esta parte del correo electrónico.

En la última parte del correo electrónico, el encabezado 419 MIME final indica el tipo de datos de un archivo adjunto al correo electrónico, que en este caso es un archivo zip. El archivo 421 ZIP está codificado en base 64 y se añade al correo electrónico. A continuación, un límite 423 final indica el final del correo electrónico.

En la etapa S307 en la figura 3, el dispositivo de análisis de conformidad del RFC 822 trata los caracteres ASCII que forman el correo electrónico usando un analizador. A continuación, el dispositivo de análisis de conformidad del RFC 822 es capaz de detectar los límites en el correo electrónico y comprobar para ver si ciertos parámetros se ajustan a un formato predeterminado aceptable conocido. Por ejemplo, el dispositivo de análisis de conformidad del RFC 822 comprueba la longitud de la línea para ver si se ajusta a la norma RFC 822 y así solo se regeneran las longitudes de línea de 2.000 o menores.

Se pueden realizar otras comprobaciones para ver si los datos tratados en el correo electrónico se ajustan a la norma RFC 822. Por ejemplo, se comprueba si los caracteres dentro del correo electrónico se conocen como caracteres ASCII aceptables tal como se define en la norma, si la información en el encabezado es como se define en la norma y si la longitud del encabezado se ajusta a la definición de la norma. Estas comprobaciones listadas son ejemplos simplemente de un gran grupo de comprobaciones diferentes que realiza el dispositivo de análisis de conformidad del RFC 822 (el resto de las cuales serán evidente para un experto en la materia), y como tal, esta invención no se limita a las enumeradas anteriormente.

Así como el análisis de los datos tratados para ver si se ajustan a la norma RFC 822 básica, el dispositivo de análisis de conformidad del RFC 822 también comprueba para ver si ciertos parámetros se ajustan a ejemplos del mundo real de los correos electrónicos de la norma RFC 822. Es decir, la especificación de ciertos parámetros puede dejarse abierta para que los definan los usuarios, mientras que, en el mundo real, se usarán solo los valores razonables. Por ejemplo, un correo electrónico, por lo general, solo comprende un número mínimo de partes. Por lo tanto, si se recibe un correo electrónico que incluye 1.000 límites, esto no sería un ejemplo del mundo real de los correos electrónicos de la norma RFC 822, y por esto debería bloquearse, es decir, no se regeneraría y preferentemente se eliminaría, mediante el dispositivo de análisis de conformidad del RFC 822.

Para cada parte de componente del correo electrónico que contiene datos que necesitan una comprobación de conformidad adicional, la parte de componente se reenvía a la etapa S309, en esta realización, en paralelo a un dispositivo de análisis de conformidad distinto dependiendo del tipo de datos de la parte correspondiente. Es decir, si la parte de correo electrónico que se analiza se define como texto, los caracteres ASCII que componen el texto se reenvían a un dispositivo de análisis de conformidad de texto. Si la parte de correo electrónico que se analiza se define como un archivo TIFF, los caracteres que componen el archivo TIFF se envían a un dispositivo de análisis de conformidad TIFF.

En la etapa S309, cada uno de los dispositivos de análisis de conformidad analiza los datos que le han reenviado para ver si se ajustan a su formato supuesto. Si los datos se ajustan se regeneran mediante el dispositivo de análisis de conformidad. Si existe alguna no conformidad dentro de los datos, los datos bien se excluyen, o, si es posible, se regeneran mediante el dispositivo de análisis de conformidad al que se ajustan. Un ejemplo de la regeneración de los datos al que se ajustan es el de añadir corchetes anidados en un archivo RTF donde sea que se estén perdiendo.

Si un correo electrónico comprende una anidación de diferentes tipos de datos, los dispositivos de análisis de conformidad se llaman de forma recursiva, de manera que varios dispositivos específicos se ejecutan en secuencia y cada uno se pone en espera en cada punto en que se descubre un tipo adicional de datos. De esta manera, un correo electrónico con un archivo zip, que incluye un documento de procesamiento de texto, que incluye un archivo de imagen JPEG podría ejecutarse a través de la secuencia de los diferentes dispositivos de análisis de conformidad (zip, procesamiento de textos, JPEG) con el fin de desplegarse a través del anidamiento de archivos y analizar cada archivo en secuencia. Al final del análisis, el archivo se vuelve a montar usando las partes regeneradas que se ajustan.

Tras una determinación en la etapa S311 en la que se han regenerado suficientes partes del correo electrónico para formar adecuadamente un correo electrónico coherente, comprensible y útil, se vuelven a montar los datos usando el dispositivo de análisis de conformidad del RFC 822 con las partes regeneradas, como se muestra en la etapa S313. Esto asegura que el correo electrónico regenerado se reenvía en el formato correcto.

A continuación, la aplicación AV reenvía el correo electrónico regenerado al destinatario previsto usando el protocolo SMTP, como se indica en la etapa S315.

Sin embargo, si la aplicación AV determina en la etapa S311 que suficientes partes del correo electrónico no se han regenerado para formar un correo electrónico útil, se rechaza el correo electrónico en la etapa S317. Durante la

etapa S317, el texto de advertencia se envía al destinatario previsto del correo electrónico informándole de que un correo electrónico destinado a él se ha rechazado por el sistema. El texto de advertencia puede incluir detalles de por qué se ha eliminado el mensaje e información adicional destinada a ayudar al destinatario a identificar al remitente, o el motivo de por qué se rechazó el correo electrónico.

A continuación, se describen en detalle algunos ejemplos de dispositivos de análisis de conformidad para su uso en esta realización, que podrían usarse durante la etapa S309. Como se muestra en la etapa S309, la parte de componente de un correo electrónico que pretende ser texto, tanto sobre la base de la información en el encabezado RFC 822, el encabezado MIME o la extensión del archivo, se pasa a un dispositivo de análisis de conformidad de texto. El dispositivo de análisis de conformidad de texto trata los datos de texto para determinar si se ajustan a su formato admisible predeterminado como se describe a continuación.

Como hay un número de diferentes tipos de archivo de texto, tales como, por ejemplo, la variable separada por coma (CSV) y el formato de texto enriquecido (RTF), el dispositivo de análisis de conformidad de texto debe primero diferenciar qué tipo de archivo de texto pretenden ser los datos tratados. Todos los archivos adjuntos al correo electrónico tendrán una extensión de archivo asociada con él que indica cuál es el tipo de archivo que debería ser. El dispositivo de análisis de conformidad de texto analiza la extensión del archivo tratado dentro del encabezado MIME para determinar si el archivo de texto es un archivo ASCII puro. Si es así, solo es necesario usar un dispositivo de análisis conformidad ASCII, como se describe a continuación.

Sin embargo, si el dispositivo de análisis de conformidad de texto, tras el análisis, determina que el archivo de texto es un tipo de archivo que no sea un ASCII puro, por ejemplo un archivo CSV, entonces también se llamará a un dispositivo de análisis de conformidad CSV para analizar y regenerar los datos CSV. Sin embargo, en primer lugar el dispositivo de análisis de conformidad ASCII analiza los caracteres ASCII que componen el archivo de texto dentro del correo electrónico para ver si la cadena de texto se ajusta al formato predeterminado ASCII, y, si hay conformidad, regenera el archivo ASCII.

El dispositivo de análisis de conformidad ASCII trata los datos para asegurarse de que el archivo se ajusta al formato predefinido ASCII mínimo. Por ejemplo, el dispositivo de análisis de conformidad ASCII solo permite que los caracteres ASCII 32 a 127, y cuatro caracteres de control, "salto de línea" (LF= 10), "retorno de carro" (CR= 13), “tabulador” (TAB= 9) y “tabulador vertical" (VT= 11) se regeneren y pasen a través del sistema.

Otros caracteres de control, tales como el carácter de campana (BEL= 7), no están en el formato permitido predeterminado para un archivo ASCII, tal como se define mediante la aplicación AV. Por lo tanto, el dispositivo de análisis de conformidad ASCII no regenera el carácter “BEL” en el bloque de códigos ASCII que se tratan, pero rechazará ese carácter ASCII.

Otros ejemplos de análisis del dispositivo de análisis de conformidad ASCII realizados son:

- ¿Es la longitud de línea natural menor que 1.024 caracteres?
- ¿Son las longitudes de palabra menores que 25 caracteres?
- ¿Está el porcentaje de espacios para caracteres por debajo de un límite predefinido?

Si en cualquier momento el dispositivo de análisis de conformidad ASCII no es capaz de regenerar los datos de esa parte del código ASCII debido a que no se ajusta al formato predeterminado de base, el dispositivo de análisis de conformidad ASCII comprueba los datos para ver si se ajustan a algún otro tipo de código ASCII. Por ejemplo, el código fuente, BinHex, Base 64. Si los datos se ajustan a otro tipo de código ASCII, los datos se reenvían al dispositivo de análisis de conformidad pertinente para ese tipo de ASCII, lo que sería, para los ejemplos mostrados anteriormente, un dispositivo de análisis de conformidad de código fuente, un dispositivo de análisis de conformidad BinHex o un dispositivo de análisis de conformidad Base 64. Se entenderá que un archivo de código ASCII Base 64 puede incluir también otros tipos de archivos dentro de los datos codificados. A continuación, estos otros tipos de archivos se remitirían también al dispositivo de análisis de conformidad del tipo de archivo pertinente, y así sucesivamente.

Los dispositivos de análisis de conformidad para los tipos adicionales de código ASCII tendrían además limitaciones de conformidad de los datos dentro de esta parte del correo electrónico. Por ejemplo, el archivo podría comprobarse para ver si es un código estructurado adecuadamente, tiene longitudes de línea correctas, y así sucesivamente. Una vez que cada dispositivo de análisis de conformidad ha determinado que se ajustan los datos de contenido y de parámetro, y en consecuencia se han extraído, los datos de contenido extraídos se regeneran en el formato predefinido admisible usando el dispositivo de análisis de conformidad.

Una vez que el dispositivo de análisis de conformidad ASCII ha terminado su tarea, los datos ASCII regenerados se reenvían al dispositivo de análisis de conformidad de texto pertinente del que los datos pretenden ser. En esta realización, el archivo de texto es un archivo CSV, y por lo tanto los datos se reenvían al dispositivo de análisis de conformidad CSV.

Los ejemplos de las comprobaciones realizadas por el dispositivo de análisis de conformidad CSV son los siguientes. El dispositivo de análisis de conformidad CSV trata los datos ASCII para asegurar que no hay párrafos de texto largos, como párrafos que no son parte del formato predefinido para los archivos CSV. Cualquier dato que no puede tratarse porque no se ajusta se rechaza mediante el dispositivo de análisis de conformidad CSV. El dispositivo de análisis de conformidad CSV comprueba también, por ejemplo, para ver si el número de delimitadores se ajusta al número predeterminado normal de delimitadores en un archivo CSV. Cuando el dispositivo de análisis de conformidad CSV determina que los datos se ajustan, los datos se regeneran en el mismo formato.

De este manera, solo las partes de un archivo de texto que se ajusten a un formato predeterminado se les permite pasar a la siguiente fase de la aplicación AV. Solamente las partes que se ajusten del archivo de texto se regeneran con las otras partes de tipos de datos regeneradas antes de montarse de nuevo y reenviarse a su destino. Por lo tanto, cualquiera de las partes del correo electrónico que contenga un virus no se ajusta y entonces debería bloquearse, es decir, no se regenera y preferentemente se elimina. Cualquiera de las partes que no se ajuste no se le permite pasar a través de la aplicación AV e infectar el sistema operativo.

Un ejemplo adicional de dispositivo de análisis de conformidad es un dispositivo de análisis de conformidad TIFF (Formato de Fichero de Imagen Etiquetado), usado para analizar y regenerar archivos TIFF.

Un archivo TIFF tiene un formato estructurado con un conjunto de directorios y etiquetas dispuestas en un formato predefinido. No es posible determinar si los propios datos de la imagen representan una imagen significativa. Sin embargo, el dispositivo de análisis de conformidad TIFF trata y analiza los datos de la imagen para asegurar que cae dentro de los límites predefinidos.

La información del encabezado en el archivo TIFF se trata y se analiza para ver si la información correcta está completa e intacta. Por ejemplo, el dispositivo de análisis de conformidad TIFF comprueba para ver si la información de encabezado incluye los campos de resolución, tamaño y profundidad que están dentro de los límites razonables para una imagen TIFF. Además, el dispositivo de análisis de conformidad TIFF determina si el número de bandas indicadas en el encabezado coincide con los datos de la imagen.

Los archivos TIFF se comprimen normalmente, por lo general, usando técnicas de compresión LZW (Lempel-Ziv-Welch). Cada banda TIFF se descomprime mediante el dispositivo de análisis de conformidad para ver si la longitud de la banda está dentro de límites predefinidos razonables. Por ejemplo, si la longitud de la banda no es igual o menor que un límite máximo del tamaño de la imagen (por ejemplo, mayor que un tamaño de papel A0 normalizado), la banda se rechaza. Tan pronto como el dispositivo de análisis de conformidad TIFF rechaza una banda, se rechaza todo el archivo TIFF.

El dispositivo de análisis de conformidad TIFF realiza también un análisis de las etiquetas (es decir, datos de parámetros) dentro del archivo TIFF. Las etiquetas se comprueban contra un formato permitido predefinido para ver si, por ejemplo, las etiquetas están en el orden especificado (de acuerdo con el directorio de información de etiqueta en el encabezado) y las etiquetas se interrelacionan entre sí de la manera correcta.

Cuando el dispositivo de análisis de conformidad TIFF determina que los datos se ajustan al formato admisible predefinido, los datos se regeneran para crear un archivo TIFF regenerado que tiene el nombre del archivo original (donde el nombre de archivo se ajusta al formato predeterminado). El archivo TIFF regenerado se reenvía al servidor de correo electrónico para volverse a montar en un mensaje de correo electrónico.

También es posible tener otros tipos de imágenes dentro del propio archivo TIFF. Por ejemplo, las imágenes JPEG pueden encapsularse dentro del archivo TIFF. Si se detecta un tipo de imagen diferente mediante el dispositivo de análisis de conformidad TIFF, éste reenvía los datos asociados con esa imagen a un dispositivo de análisis de conformidad adicional, en este ejemplo, un dispositivo de análisis de conformidad JPEG. A continuación, el dispositivo de análisis de conformidad JPEG, trata y analiza los datos para ver si se ajustan a un formato JPEG esperado, y si es así, regenera los datos en el formato JPEG. A continuación, los datos regenerados se vuelven a montar en el archivo TIFF regenerado, que se usa, a continuación, para volver a montar un correo electrónico regenerado. A continuación, este correo electrónico se pasa al servidor de correo electrónico.

Una opción adicional disponible en esta realización es la aplicación AV para insertar un texto de advertencia en lugar de las partes que no se ajustan del correo electrónico. Es decir, si un dispositivo de análisis de conformidad trata los datos de la parte que no se ajusta y determina que una porción de la parte no se ajusta al formato admisible predeterminado, tras la regeneración del correo electrónico, el dispositivo de análisis de conformidad inserta el texto de advertencia en lugar de la parte que no se ajusta, informando al destinatario previsto del correo electrónico que se ha rechazado una porción del correo electrónico por la aplicación AV. Como alternativa, si un dispositivo de análisis de conformidad rechaza toda una parte de un correo electrónico debido a que no se ajusta, la aplicación AV inserta el texto de advertencia dentro del correo electrónico informando al destinatario previsto que se ha bloqueado una parte del correo electrónico, es decir, no regenerada y preferentemente eliminada, por la aplicación AV.

Tercera realización

En referencia a la figura 5, se describirá ahora una tercera realización de la presente invención.

Esta tercera realización incorpora todas las características de la segunda realización, incluyendo cualquiera de las opciones discutidas en relación a la segunda realización.

La figura 5 muestra un diagrama de flujo de un proceso de acuerdo con esta tercera realización.

Esta realización se refiere a la situación después de que la aplicación AV ha bloqueado una porción, una parte o la totalidad de un correo electrónico (denominada "parte que no se ajusta" en esta realización). En la etapa S501, la aplicación AV realiza la determinación de si la parte no se ajusta y, si es así, se debe bloquear. Si se bloquea por la aplicación AV, la parte que no se ajusta se reenvía a una aplicación de filtro de amenazas para determinar si la parte que no se ajusta es una amenaza, como se muestra en la etapa S503.

La aplicación de filtro de amenazas determina si la parte que no se ajusta es considerada una amenaza real en base a las preferencias del usuario del sistema. El sistema ha almacenado en su memoria una lista de tipos de archivo y fuentes asociadas a estos tipos de archivos que no se consideran una amenaza. Por lo tanto, el sistema puede determinar, en base al remitente del archivo y al tipo de archivo, si se debe permitir pasar al archivo.

Si la determinación en la etapa S503 determina que el tipo de archivo no es uno de los indicados como que es admisible por la fuente asociada, se bloquea en la etapa S505.

Si el tipo de archivo se considera que debe ser admisible, la parte que no se ajusta evita la aplicación AV en la etapa S507. La aplicación AV regenera el resto del archivo recibido en la etapa S509, y vuelve a montar las partes que se ajustan regeneradas y las partes que no se ajustan evitadas del archivo en la etapa S511.

Por ejemplo, si un sistema bancario recibe de un remitente conocido un gran número de correos electrónicos incluyendo hojas de cálculo que incorporan secuencias de instrucciones complicadas, estos pueden estar fuera del formato predeterminado permitido para una secuencia de instrucciones dentro de un archivo adjunto de hoja de cálculo, y por lo tanto, el dispositivo de análisis de conformidad de la secuencia de instrucciones bloquearía esta parte del correo electrónico.

Sin embargo, como el sistema bancario es capaz de determinar quién está enviando los correos electrónicos, y el remitente está registrado como un socio de confianza del sistema bancario en una base de datos para estos tipos de archivos, no se considera que la hoja de cálculo en el correo electrónico sea una amenaza. Por lo tanto, el usuario del sistema puede configurar la aplicación de filtro de amenazas para permitir que estas partes de secuencia de instrucciones que no se ajustan eviten la aplicación AV y puedan volver a montarse en el correo electrónico con las partes regeneradas del correo electrónico.

Como alternativa, la aplicación de filtro de amenazas puede funcionar en un modo mediante el que se determine si se debería permitir a un archivo regenerado recibido desde la aplicación AV que continúe a través del sistema de destino. Si la aplicación AV recibe un archivo que incluye partes que no se ajustan, que en sí mismas no son suficientemente no ajustadas para que la aplicación AV rechace todo el archivo en el acto, pero da como resultado un archivo que se ajusta regenerado que es diferente básicamente del archivo original, el archivo regenerado se reenvía a la aplicación de filtro de amenazas. Por ejemplo, el tamaño del archivo original puede ser considerablemente más grande que un tamaño de archivo que se ajusta regenerado debido a un gran número de palabras individuales reescritas dentro de una secuencia de instrucciones que no se regenera mediante la aplicación AV.

La aplicación de filtro de amenazas realiza una determinación de si se está enviando el tipo de archivo desde una fuente autorizada para ese tipo de archivo, y si es así, permitirá que el tipo de archivo pase a través del sistema.

Realizaciones Adicionales

Se entenderá que las realizaciones de la presente invención se describen en el presente documento solo a modo de ejemplo, y que pueden hacerse diversos cambios y modificaciones sin apartarse del alcance de la invención.

Se entenderá que la presente invención puede implementarse en cualquier sistema donde los archivos electrónicos se muevan desde un origen a un destino. El método de envío de los archivos electrónicos para los fines de la presente invención no se limita a ningún método específico. Es decir, por ejemplo, los archivos electrónicos pueden transferirse de un componente a otro componente del hardware de un sistema informático. Como alternativa, por ejemplo, los archivos electrónicos pueden transferirse a través de una interfaz aérea desde una estación base a un dispositivo de teléfono móvil. También, por ejemplo, los archivos electrónicos pueden transmitirse a través de una red de área local (LAN), una red de área amplia (WAN) o a través de Internet.

Además, se entenderá que, como una opción adicional para cualquier realización descrita anteriormente, puede proporcionarse una función de anulación para que los usuarios de forma manual anulen cualquiera de las determinaciones realizadas por, o la aplicación AV o la aplicación de filtro de amenazas cuando se recibe el archivo electrónico. Es decir, cuando un dispositivo de análisis de conformidad dentro de la aplicación AV bloquea una porción, una parte o la totalidad del correo electrónico, debido a su no conformidad, todavía se da al usuario una opción de permitir que la no conformidad se regenere y se vuelva a montar en el correo electrónico.

Un ejemplo de la realización de esta opción es suministrar al destinatario previsto un texto de advertencia que le pregunta si debe permitirse que la no conformidad del correo electrónico analizado pase a través del sistema como si se ajustase al formato permitido predefinido. Una respuesta a esta advertencia proporciona al dispositivo de análisis de conformidad una instrucción para regenerar, si es posible, y volver a montar el correo electrónico. O, como alternativa, se permite al mensaje original evitar tanto a la aplicación AV como a la aplicación de filtro de amenazas y pasar a través del sistema sin regeneración.

Además, se entenderá que la aplicación AV como se describe en la segunda realización puede localizarse en otro lugar que en el servidor de correo electrónico del ISP. Por ejemplo, la aplicación AV puede localizarse en e instalarse en el servidor de cliente de correo electrónico del destinatario. De esta manera, cualquiera de los correos electrónicos reenviados por el servidor de cliente de correo electrónico al buzón de entrada del destinatario en una unidad de disco duro son los correos electrónicos regenerados como se ha descrito anteriormente.

Además, se entenderá que la aplicación AV puede cablearse en un dispositivo semiconductor, tal como, pero no limitado a silicio, arseniuro de galio (GaAs), fosfuro de indio (InP). Es decir, la aplicación AV tiene una tarea cuantificable, que no requiere la necesidad de actualizaciones en el proceso de definir un formato que se ajuste predefinido. Las instrucciones necesarias para realizar la tarea de la aplicación AV, incluyendo el tratamiento, el análisis, la regeneración y la vuelta a montar pueden realizarse en cualquier dispositivo semiconductor adecuado. Además, las instrucciones necesarias para implementar la aplicación AV pueden almacenarse en un dispositivo de memoria semipermanente o permanente. Entonces, el dispositivo de memoria sería capaz de funcionar para ejecutar la aplicación AV en asociación con un procesador conectado. Entonces, en estos casos, puede proporcionarse la invención independientemente del ordenador que debe protegerse, como un dispositivo independiente (por ejemplo, en una tarjeta tal como una tarjeta de módem, una tarjeta de adaptador de red o el controlador de la unidad de disco) incluyendo el hardware del procesador y la memoria independientes a los del ordenador que debe protegerse. Esto tiene la ventaja de aislar el archivo electrónico entrante por completo del sistema de archivos y otros recursos del ordenador que deben protegerse, y su almacenamiento en una localización que normalmente no se puede escribir ni actualizar, con el fin de evitar los ataques de "puerta de atrás" en la propia aplicación AV; en otras palabras, un nivel de seguridad física. El dispositivo semiconductor puede consistir de un procesador y un dispositivo de memoria donde el procesador ejecuta la aplicación AV desde el dispositivo de memoria y almacena los archivos entrantes en el dispositivo de memoria para aislarlos.

Además, se entenderá que el dispositivo semiconductor descrito anteriormente puede proporcionarse como una parte de cualquier tarjeta de red adecuada usando métodos convencionales. De esta manera, la tarjeta de red puede utilizarse en una red de comunicaciones como un medio para asegurar que la red está protegida de código y datos no deseados regenerando los archivos electrónicos recibidos usando los métodos descritos.

Además, se entenderá que los archivos electrónicos tal como se describe en la primera realización pueden recibirse mediante un dispositivo informático, donde los archivos electrónicos se almacenan en un dispositivo de memoria extraíble. Por ejemplo, los archivos electrónicos pueden almacenarse en un dispositivo de disco USB, una tarjeta inteligente, un dispositivo de memoria digital segura (SD), un dispositivo de memoria de tarjeta multimedia (MMC), una tarjeta compact flash (CF) de tipo 1 o 2, una tarjeta de medios inteligente (SM), una tarjeta XD, un disquete, una unidad ZIP, un disco duro portátil o cualquier otro dispositivo de memoria adecuado que pueda conectarse, directamente o a través de un medio inalámbrico, a un dispositivo informático.

Además, se entenderá que un sistema operativo como se describe en esta aplicación puede ser cualquier sistema que usa archivos. Por ejemplo, un sistema embebido, un encaminador, una tarjeta de red o similar.

Además, se entenderá que pueden utilizarse otros métodos de codificación para asegurar que cualquiera de los archivos ejecutables recibidos no puede ejecutarse de forma automática. Por ejemplo, el método de codificación almacena cada par de bytes de entrada usando un método de intercambio de bytes. En este ejemplo, si se están recibiendo 6 bytes, A B C D E F, por la aplicación AV con el byte A que se recibe el primero y el byte F que se recibe el último, se almacenan en la memoria en el orden siguiente: B A D C F E. El primer byte (A) se almacena en una segunda posición de memoria, y el segundo byte (B) se almacena en una primera posición de memoria. Esta inversión se produce en las posiciones de memoria posteriores para cada par de bytes recibidos. De esta manera, cualquier código ejecutable no es capaz de ejecutarse de forma automática y por lo tanto cualquiera de los archivos electrónicos infectados no es capaz de infectar a la aplicación AV o al sistema operativo de destino.

Para evitar cualquier duda, se solicita por la presente la protección para todas y cada una de las nuevas realizaciones descritas anteriormente, por separado y en combinaciones.

Habiendo descrito los diversos aspectos y realizaciones de la invención y las modificaciones de los mismos, los expertos en la materia apreciarán que la invención puede modificarse en disposición y detalle sin apartarse de los principios de la misma. Reivindicamos todas las realizaciones, variaciones y modificaciones que entran dentro del alcance de las reivindicaciones siguientes.

REIVINDICACIONES

1. Un método implementado por ordenador de resistencia a la propagación de código y datos no deseados en un archivo electrónico, comprendiendo el método:

recibir un archivo electrónico entrante donde el archivo electrónico entrante es un correo electrónico que tiene una pluralidad de partes desde un remitente, conteniendo cada parte de dicho archivo datos de contenido en un tipo de archivo de datos predeterminado, teniendo cada tipo de archivo de datos un conjunto de reglas asociado; incluyendo dichas reglas las reglas que componen la especificación del tipo de archivo y las reglas adicionales que limitan los valores y/o los intervalos que contienen y los parámetros que pueden asumir
determinar un tipo de archivo de datos predeterminado supuesto de cada parte
tratar los datos de contenido de cada parte de acuerdo con las reglas asociadas con el tipo de archivo de datos predeterminado supuesto;
determinar si los datos de contenido de cada parte se ajustan a las reglas asociadas con el tipo de archivo de datos predeterminado supuesto;
regenerar las partes que se ajustan de los datos de contenido tratados, tras una determinación positiva a partir de los medios de determinación, para crear un archivo electrónico regenerado de sustitución en el tipo de archivo de datos predeterminado supuesto, conteniendo dicho archivo electrónico regenerado de sustitución los datos de contenido regenerados;
bloquear las partes de los datos contenidos tratados que no se ajustan a las reglas asociadas con el tipo de archivo de datos predeterminado supuesto con el fin de bloquear su inclusión en el archivo electrónico regenerado de sustitución;
almacenar una lista de tipos de archivo y las fuentes asociadas con dichos tipos de archivo que no se consideran una amenaza; reenviar las partes que no se ajustan a un filtro de amenaza;
determinar mediante el filtro de amenaza para cada parte que no se ajusta si se debe permitir que esa parte que no se ajusta pase sobre la base de la lista almacenada y el remitente del archivo y el tipo de archivo de datos; y
permitir que una parte que no se ajusta evite el bloqueo e incluir la parte que no se ajusta que lo ha evitado en el archivo electrónico regenerado de sustitución que determinó que es admisible.

2. Un método de acuerdo con la reivindicación 1, que comprende además almacenar el archivo electrónico entrante en un formato codificado en la memoria.

3. Un método de acuerdo con la reivindicación 2, donde cada byte de datos se almacena en un orden invertido de bit.

4. Un método de acuerdo con la reivindicación 2, donde los datos se almacenan de manera que cada par de bytes de datos recibidos se coloca en un orden de memoria invertido.

5. Un medio legible por ordenador que comprende un programa de ordenador adaptado para realizar el método de una cualquiera de las reivindicaciones 1 a 4.

6. Un dispositivo semiconductor que comprende un medio de memoria que incluye instrucciones para realizar el método de una cualquiera de las reivindicaciones 1 a 4.

7. Un dispositivo semiconductor de acuerdo con la reivindicación 6, donde el dispositivo semiconductor es un dispositivo de memoria semipermanente o permanente.

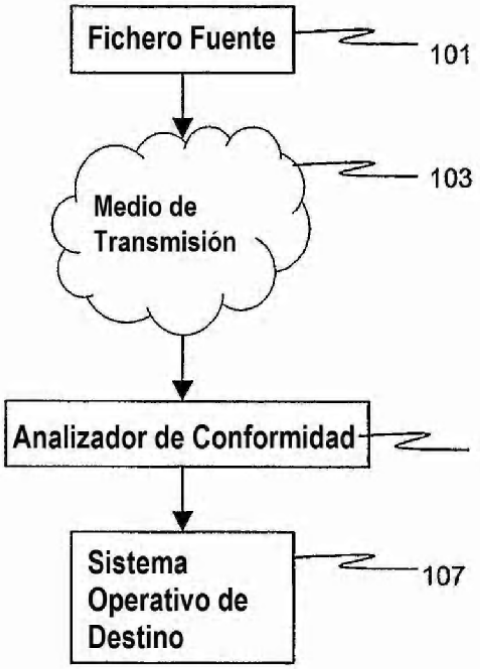

FIG. 1A

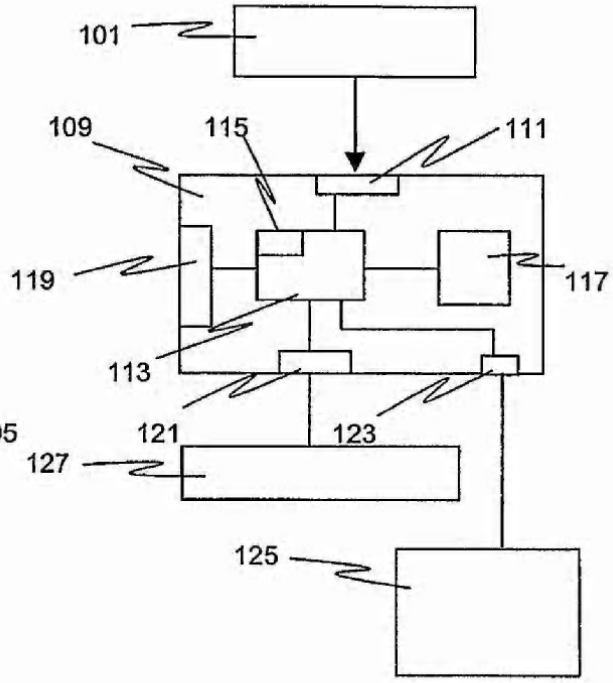

FIG. 1B

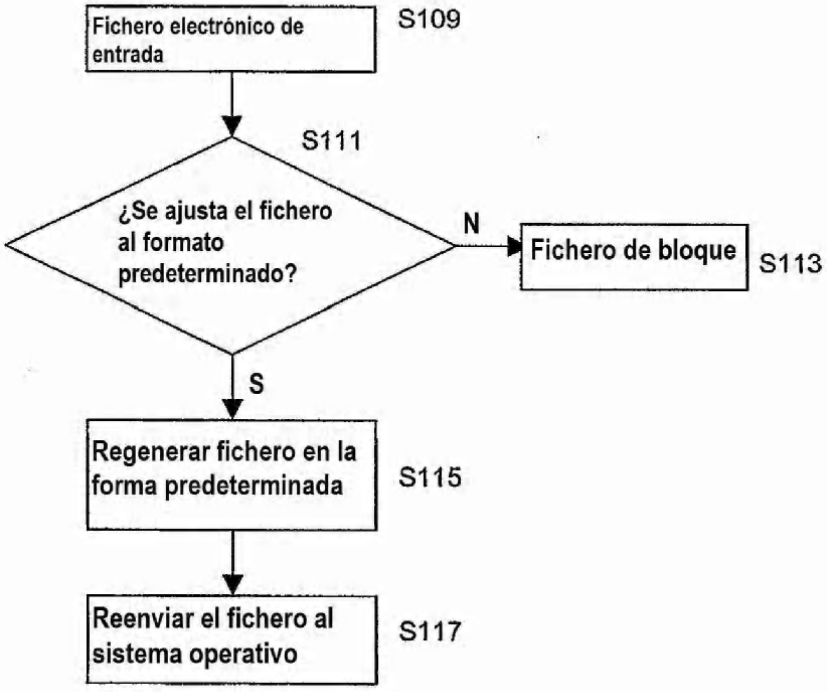

FIG. 1C

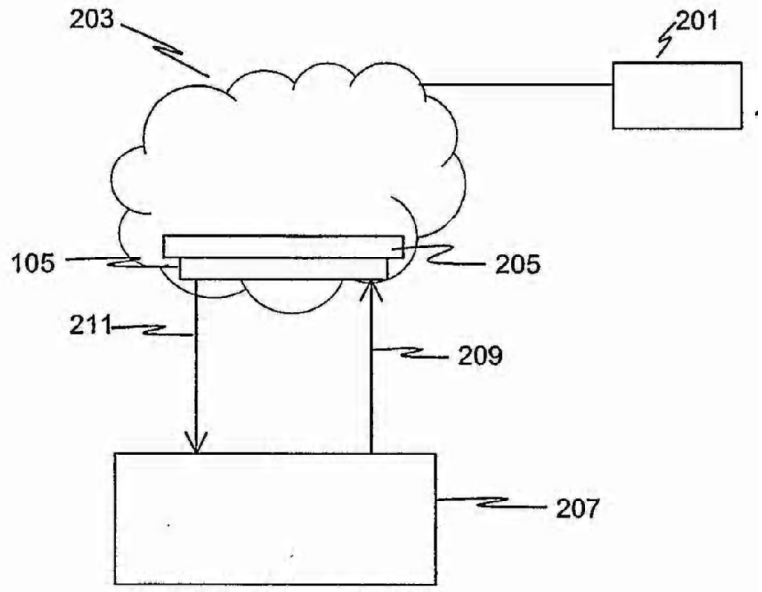

FIG. 2

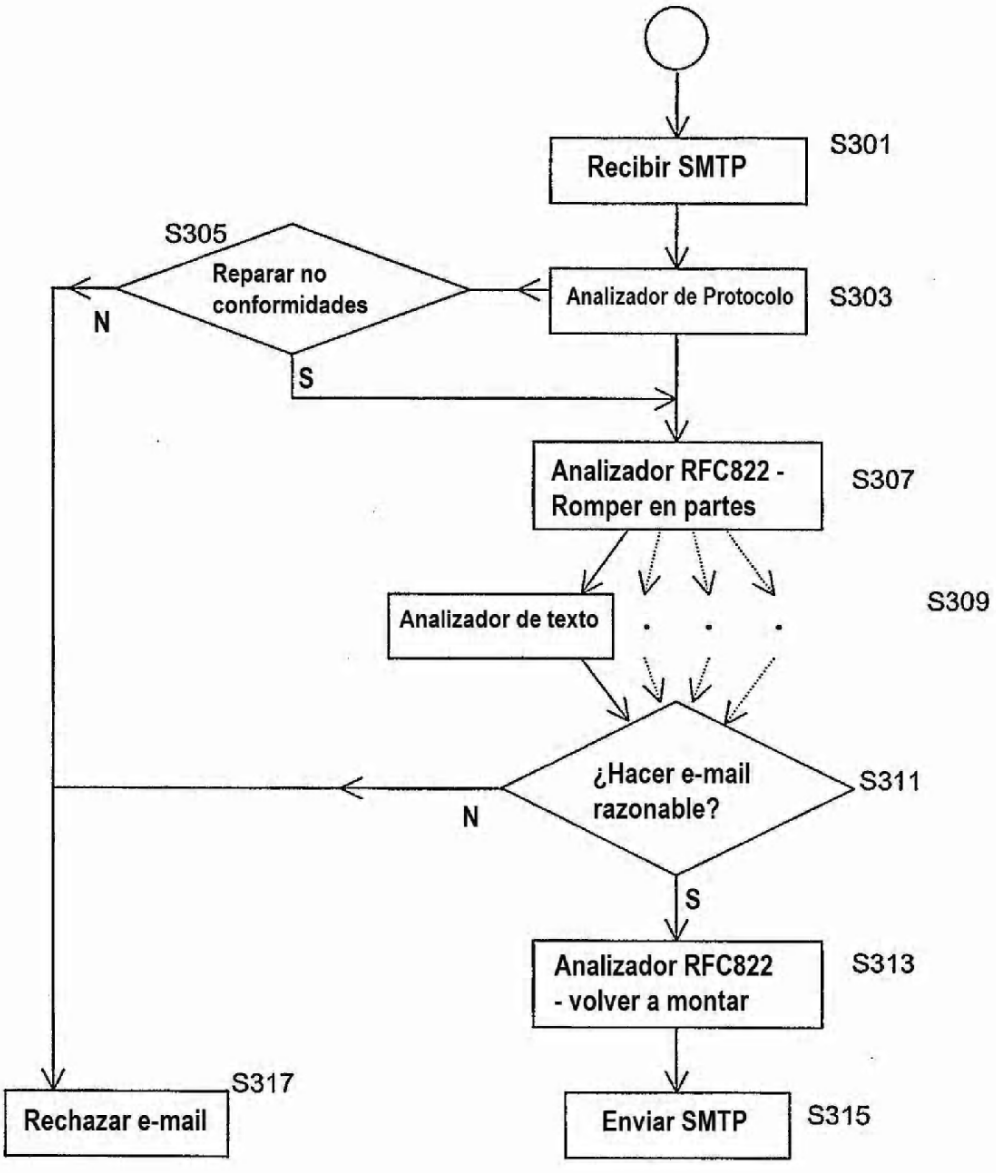
Recibir SMTP
S301
Analizador de Protocolo
S303
S305
Reparar no conformidades
N
S
Analizador RFC822 - Romper en partes
S307
Analizador de texto
S309
¿Hacer e-mail razonable?
S311
N
S
Analizador RFC822 - volver a montar
S313
Enviar SMTP
S315
Rechazar e-mail
S317

FIG. 3

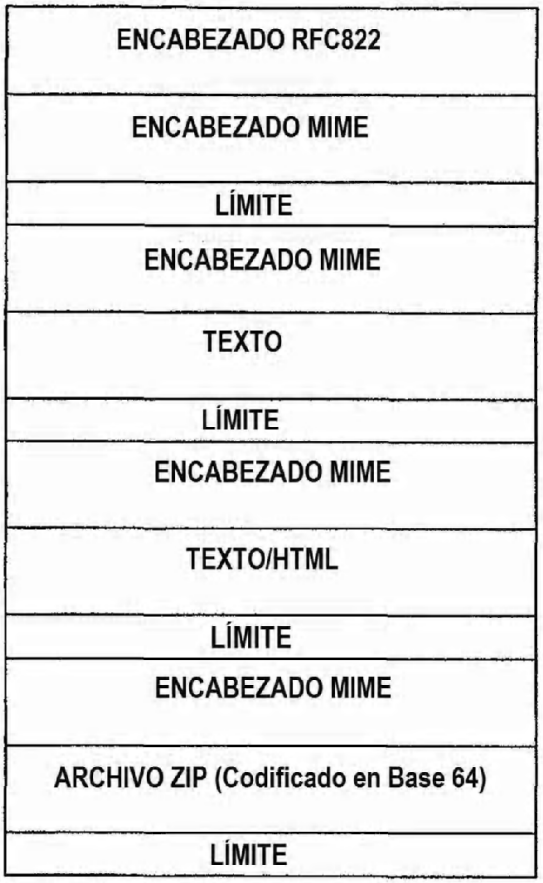
ENCABEZADO RFC822
401
ENCABEZADO MIME
403
LÍMITE
405
ENCABEZADO MIME
407
TEXTO
409
LÍMITE
411
ENCABEZADO MIME
413
TEXTO/HTML
415
LÍMITE
417
ENCABEZADO MIME
419
ARCHIVO ZIP (Codificado en Base 64)
421
LÍMITE
423

FIG. 4

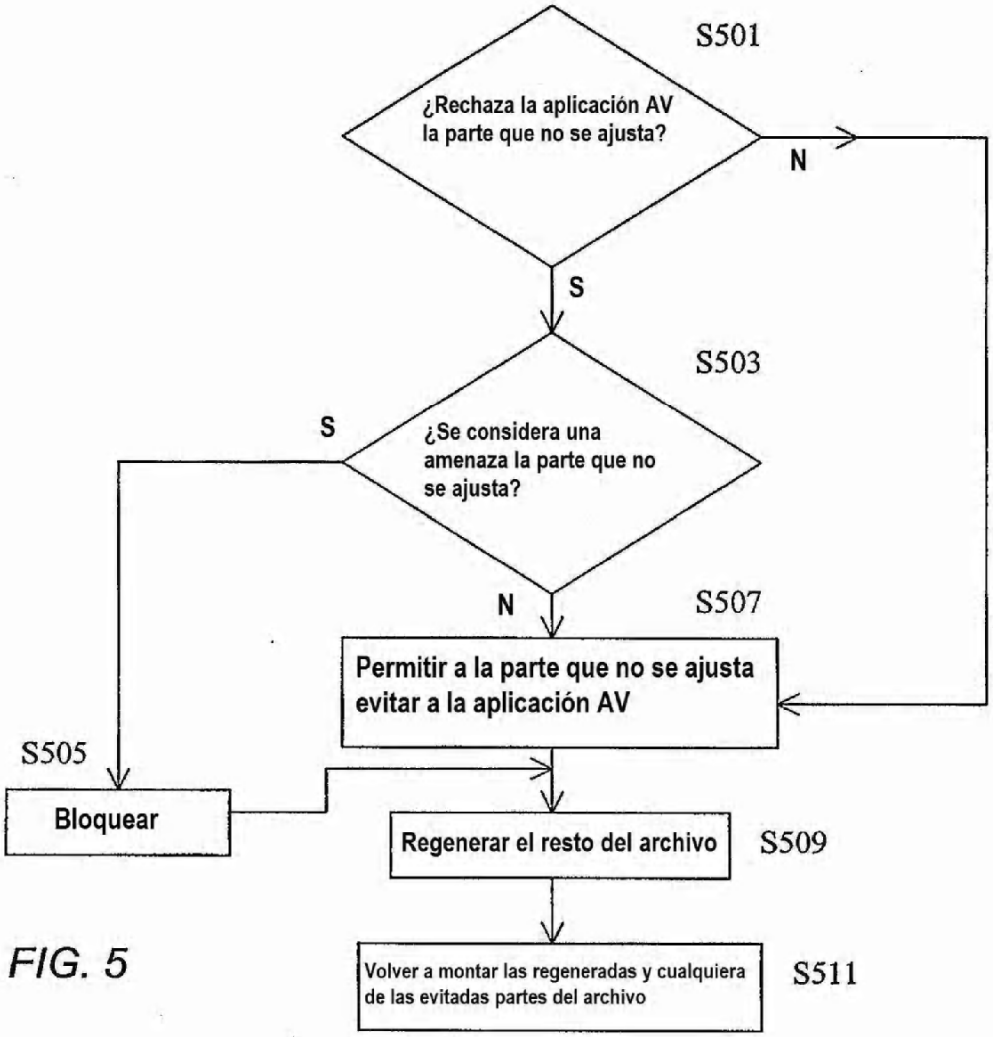
S501
¿Rechaza la aplicación AV la parte que no se ajusta?
N
S
S503
S
¿Se considera una amenaza la parte que no se ajusta?
N
S507
Permitir a la parte que no se ajusta evitar a la aplicación AV
S505
Bloquear
Regenerar el resto del archivo
S509
Volver a montar las regeneradas y cualquiera de las evitadas partes del archivo
S511

FIG. 5