(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 761 603**

(51) Int. CI.:

G06F 21/56 (2013.01)
G06F 21/53 (2013.01)
H04L 29/06 (2006.01)
H04L 29/08 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **04.07.2017 PCT/EP2017/066565**

(87) Fecha y número de publicación internacional: **11.01.2018 WO18007350**

(96) Fecha de presentación y número de la solicitud europea: **04.07.2017 E 17734745 (7)**

(97) Fecha y número de publicación de la concesión europea: **25.09.2019 EP 3479281**

(54) Título: **Procedimiento y sistema informático para determinar una puntuación de amenaza**

(30) Prioridad:

04.07.2016 EP 16177783

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
20.05.2020

(73) Titular/es:

CYAN SECURITY GROUP GMBH (100.0%)
Krotenthallergasse 8
1080 Wien, AT

(72) Inventor/es:

ARNOTH, PETER y
CSERNA, MARKUS

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

DESCRIPCIÓN

Procedimiento y sistema informático para determinar una puntuación de amenaza

La presente invención se refiere a la seguridad de TI y, más particularmente, a un procedimiento, a un producto de programa de ordenador y a un sistema de ordenador para determinar una puntuación de amenaza de un documento electrónico.

En relación con los documentos electrónicos la seguridad informática se refiere principalmente a la detección y/o la prevención de comportamientos maliciosos causados por documentos electrónicos. Los documentos electrónicos pueden ser páginas web, secuencias de comandos o archivos de documentos. Se puede acceder a ellos mediante un navegador web u otro visor de documentos. El comportamiento malicioso se refiere a acciones que pueden dañar el sistema informático que accede al documento electrónico, en particular acciones que pueden dañar los datos almacenados en ese sistema informático, y acciones que comprometen la seguridad o la privacidad de los datos, en particular por transmisión ilegítima o uso indebido de datos accesibles en ese sistema informático

En el pasado, la detección de comportamiento malicioso causado por los documentos electrónicos se ha logrado mediante la operación de programas especializados, como por ejemplo motores antivirus o escáneres de virus, en el sistema informático para acceder a los documentos electrónicos. Dichos programas especializados generalmente monitorean el comportamiento de otros programas que se ejecutan en el sistema informático para actividades de programas que se clasifican como comportamiento malicioso (por ejemplo, acceder y modificar el sistema operativo del sistema informático, instalar procesos en segundo plano o interceptar controladores de dispositivos de interfaz humana). El monitoreo se realiza mientras el usuario está operando el sistema informático e interactuando con el documento electrónico. Una vez que se detecta cualquier evento malicioso, el programa especializado toma medidas inmediatas para bloquear cualquier otra acción desencadenada por el documento electrónico y eliminar el documento electrónico del sistema informático. Es decir, una vez que un documento electrónico se clasifica como "malicioso", cualquier análisis posterior del documento electrónico o su comportamiento se interrumpe (o un usuario del sistema informático lo desactiva manualmente). Este enfoque tiene varias desventajas: primero, el programa especializado debe estar presente y ejecutarse localmente en el dispositivo del usuario, consumiendo así memoria y recursos informáticos en este dispositivo; segundo, al detectar un comportamiento malicioso solo cuando ocurre en el dispositivo del usuario, el programa especializado puede proporcionar protección contra los efectos del documento electrónico pero no puede evitar el acceso al documento electrónico, lo que desperdicia el ancho de banda para transferir el documento electrónico al dispositivo del usuario; tercero, el programa especializado se limita a la detección de efectos locales y no puede proteger contra los efectos fuera del dispositivo (remotos) activados por el documento electrónico (por ejemplo, modificación de archivos almacenados de forma remota).

Para superar algunas de las desventajas, se han desplegado medidas de seguridad remotas en forma de filtros de acceso. Dichos filtros de acceso pueden implementarse con servidores proxy a través de los cuales el dispositivo del usuario accede a documentos electrónicos. Los filtros de acceso analizan el contenido de cualquier documento electrónico accedido antes de transmitirlo al dispositivo del usuario. El análisis se realiza mediante la búsqueda de ciertos patrones predefinidos en el contenido, que previamente se ha encontrado que indican un comportamiento malicioso del documento electrónico. No se realiza ninguna representación o ejecución real del documento electrónico. En consecuencia, las estrategias para evitar la detección por dichos filtros de acceso han evolucionado. Por ejemplo, partes de documentos electrónicos pueden ofuscarse y/o cifrarse hasta que no coincidan con ningún patrón predefinido aplicado por el filtro de acceso durante su análisis de contenido estático. Por lo tanto, los filtros de acceso actuales no pueden reemplazar el uso de programas especializados locales para una mayor seguridad del dispositivo del usuario.

Como medida de seguridad adicional modernos navegadores web utilizan un entorno controlado, es decir, un entorno muy controlado y limitado, para la carga y representación de páginas web. Las limitaciones impuestas por el entorno controlado en el documento dentro restringen la mayoría de los efectos del comportamiento malicioso al entorno controlado en sí. Esto se logra separando el contexto en el que se carga el documento y se ejecutan complementos o secuencias de comandos desde una capa de presentación que tiene acceso directo a partes del dispositivo anfitrión (por ejemplo, el dispositivo del usuario que ejecuta el navegador web). El entorno controlado de un navegador web logra la seguridad al imponer ciertas restricciones de comportamiento. Generalmente no monitoriza el comportamiento que cumple con dichas restricciones. En consecuencia, no está en condiciones de combinar información sobre diferentes comportamientos permisibles pero sospechosos para proporcionar cualquier evaluación o categorización de un documento. Fuera del campo de la seguridad de TI, se sabe analizar el comportamiento dinámico de los documentos electrónicos en entornos controlados para garantizar la calidad. En estas aplicaciones, el documento electrónico bajo prueba se carga y se procesa en un entorno controlado de documentos. Luego se realiza una serie de pruebas, en las que para cada prueba se observa si una determinada interacción predefinida con el documento electrónico en el entorno controlado de documentos conduce a un comportamiento esperado predefinido. El resultado de dicho procedimiento de prueba es un "aprobado" cuando cada una de las pruebas requeridas realizadas ha conducido al comportamiento esperado respectivo o un "fallo" en caso contrario. Un ejemplo de este tipo de caja de documentos es PhantomJS (ver http://phantomjs.org/). PhantomJS proporciona un entorno controlado de documentos adaptado para el uso con páginas web; específicamente implementa un entorno controlado de navegador web para pruebas y monitoreo de sitios web sin

cabeza.

El documento US 2011/0289582 A1 muestra un aparato y un sistema para anotar y sitios web de clasificación y un procedimiento de operación. El sistema utiliza un emulador de navegador para cargar sitios web desde una o más identificaciones de recursos uniformes. La puntuación y la calificación se basan en el comportamiento de un sitio web caracterizado por las API, funciones y bibliotecas de programas invocadas. Con respecto al comportamiento dinámico del sitio web, el procedimiento se limita al examen de JavaScript inicial y JavaScript solicitado por un JavaScript predecesor. Incluye el examen de las respuestas recibidas para peticiones de clientes diferidos. En general, el procedimiento se limita al examen del comportamiento autoactivo del sitio web. Por lo tanto, un sitio web malicioso puede suplantar la puntuación y evitar una calificación negativa, por ejemplo, al desencadenar un comportamiento malicioso solo con la interacción del usuario.

Un procedimiento similar para la puntuación y la clasificación se describe en el documento US 8.990.945 B1, con limitaciones similares.

El documento US 2013/0014020 A1 da a conocer un sistema y procedimiento para la gestión de una base de datos de reputación de los sitios web y que indica una reputación de un sitio web a un cliente que accede dicho sitio web.

El documento WO 2010/002816 A1 muestra un sistema y un procedimiento para la categorización de sitios web y para usar dicha categorización para operar un filtro de acceso.

El documento US 9.104.867 B1 muestra un procedimiento para analizar malware, que se ejecuta en una máquina virtual después de un control estático. La máquina virtual puede ser cualquier tipo de máquina virtual, como, por ejemplo, emulación de hardware, virtualización completa, paravirtualización y máquinas virtuales de virtualización a nivel de sistema operativo. Por lo tanto, este procedimiento se aplica a la virtualización a nivel de sistema operativo o inferior.

Un objeto de la presente invención es proporcionar una clasificación de seguridad de documentos electrónicos, que pueden ser utilizados para operar un filtro de acceso remoto desde un dispositivo de usuario y que es más difícil de falsificar que los esquemas conocidos para clasificación de seguridad.

Con el fin de alcanzar el objetivo mencionado anteriormente, la presente invención proporciona un procedimiento como se define en el primer momento, que comprende las etapas de:

- cargar y representar el documento electrónico en un entorno controlado de documentos;
- consultar una lista de todos los elementos de navegación disponibles en el documento electrónico desde el entorno controlado de documentos;
- controlar el entorno controlado de documentos para simular la interacción del usuario con el documento electrónico basado en la lista consultada,
- mientras carga y representa el documento electrónico y controla el entorno controlado del documento para simular la interacción del usuario con el documento electrónico, monitoriza el entorno controlado del documento en busca de eventos activados por el documento electrónico y que pertenecen a una de al menos dos clases de eventos predefinidas;
- registrar cada evento observado junto con una clase de evento respectiva a la que pertenece;
- determinar una puntuación de amenaza del documento electrónico basado en ponderaciones numéricas predefinidas asociadas con cada una de las clases de eventos predefinidas a las que pertenecen los eventos registrados.

Con el fin de alcanzar el objetivo mencionado anteriormente, la presente invención proporciona un producto de programa de ordenador como se define en el primer momento, que comprende partes de programa, que cuando se carga en un ordenador están diseñados para llevar a cabo las etapas del procedimiento del presente procedimiento.

Con el fin de alcanzar el objetivo mencionado anteriormente, la presente invención proporciona un sistema de ordenador como se define en el primer momento, que comprende:

- un módulo de entorno controlado de documentos para cargar y representar documentos electrónicos,
- un módulo de simulación de interacción conectado al módulo de entorno controlado del documento y configurado para consultar en el entorno controlado del documento una lista de todos los elementos de navegación disponibles en el documento electrónico y para controlar el módulo de entorno controlado del documento para simular la interacción del usuario con el documento electrónico basado en la lista consultada,
- un módulo de monitoreo conectado al entorno controlado de documentos y configurado para monitorear el módulo de entorno controlado de documentos para eventos que pertenecen a una de al menos dos clases de eventos predefinidas y para registrar cada evento observado junto con una clase de evento respectiva a la que pertenece, y
- un módulo de puntuación conectado al módulo de monitorización y configurado para determinar una puntuación de amenaza basada en ponderaciones numéricas predefinidas asociadas con cada una de las clases

de eventos predefinidas a las que pertenecen los eventos registrados por el módulo de monitorización.

En particular, la presente invención se puede aplicar a las páginas web, es decir, documentos electrónicos accesibles a través de la World Wide Web usando un navegador web para la visualización de dichos documentos en un dispositivo de usuario para su visualización y/o interactuar con el documento, por lo general identificados por un localizador uniforme de recursos (URL). Básicamente, la presente invención proporciona una clasificación de seguridad conductual de documentos electrónicos basada en una serie de comportamientos "sospechosos" predefinidos diferentes, cada uno asociado con una determinada calificación de seguridad. En este contexto, "comportamiento" se refiere al comportamiento en tiempo de ejecución y significa que los documentos electrónicos se analizan simulando la situación en un dispositivo de usuario y cargando y haciendo que el documento sea similar a un visor de documentos utilizado por el usuario final. Para este propósito, se utiliza un entorno controlado de documentos, que proporciona un entorno realista pero seguro para cargar, procesar e interactuar con el documento electrónico. Como se mencionó anteriormente, dichos entornos limitados de documentos se han utilizado previamente para garantizar la calidad de los documentos electrónicos, es decir, para encontrar errores y defectos en el documento electrónico. Por ejemplo, cuando se prueba una página web, el entorno controlado web correspondiente implementa un motor de navegador de Internet completo con interfaces de programación hacia/desde otras máquinas o secuencias de comandos.

En el entorno controlado de documentos se procesan documentos electrónicos como si se carga y mostrada en un visor de documentos real, pero en vez cargado y rendido sin una interfaz humano o salida gráfica sólo con el propósito de controlar el comportamiento y, opcionalmente, la automatización de las interacciones con el documento. Todas las interacciones del usuario con el documento electrónico se implementan a través de interfaces de programación (API) que se pueden controlar desde fuera del entorno controlado. Las interacciones del usuario como hacer clic en los botones o seguir los enlaces en una página web (simulando la interacción del usuario con un clic del ratón) o la ejecución de formularios web se pueden controlar mediante procedimientos de automatización a través de secuencias de comandos predefinidas. También es posible simular interacciones "pasivas", como pasar el puntero del ratón sobre un contenido específico o parte del documento, que pueden provocar que ciertas reacciones en forma de secuencias de comandos se ejecuten dentro del entorno controlado. Estas reacciones son nuevamente monitoreadas, registradas y procesadas cuando se determina una puntuación de amenaza del documento electrónico.

El puntuación de amenaza es una clasificación numérica de los riesgos de seguridad asociados con el acceso al documento electrónico correspondiente en un dispositivo de usuario. Como tal, un filtro de acceso remoto consciente de la puntuación de amenaza de un documento electrónico puede denegar el acceso de un dispositivo de usuario a dicho documento en función de un rango aceptable determinado de puntuaciones de amenaza, si la puntuación del documento en cuestión está fuera de dicho rango aceptable. El rango aceptable se puede ajustar para cada filtro de acceso y/o para cada dispositivo de usuario, según un nivel de amenaza aceptable o un riesgo de seguridad aceptable. Alternativamente, se puede realizar una categorización general de cada documento electrónico comparando la puntuación de amenaza determinado con un umbral predefinido (marca de agua alta) y colocando todos los documentos que tienen una puntuación de amenaza que excede dicho umbral en una lista negra que permite filtros de acceso u otros dispositivos de seguridad para evitar el acceso a estos documentos.

El presente procedimiento comprende la etapa de, mientras que el control del entorno controlado de documentos, el control del entorno controlado de documentos a la interacción del usuario simular con el documento electrónico. Al simular las interacciones del usuario, el comportamiento sospechoso desencadenado solo por las interacciones del usuario puede observarse y tenerse en cuenta al determinar la puntuación de amenaza. Por lo tanto, el análisis de seguridad del documento electrónico logra una cobertura más completa de los casos de uso del documento electrónico, lo que hace que en general sea más difícil falsificar el análisis; en particular, un sitio web malicioso no puede evitar la detección de comportamientos maliciosos activando dicho comportamiento malicioso solo ante la interacción del usuario.

Específicamente movimientos del ratón una interacción de usuario simulado puede comprender simulados, clics de ratón simulados y/o entradas del teclado. Los movimientos y clics del ratón pueden realizarse de acuerdo con algunos o todos los elementos de entrada (enlaces, botones, etc.) reconocidos al representar el documento, por ejemplo, moviendo el ratón sobre uno de dichos elementos y simulando un clic del ratón. Las entradas del teclado se pueden realizar con respecto a algunos o todos los elementos de formulario reconocidos al presentar el documento, por ejemplo, asignando foco a uno de dichos elementos de formulario y simulando una secuencia de pulsaciones de teclas.

Con respecto al comportamiento monitorizado preferiblemente uno o más de las al menos dos clases de eventos predefinidos son peticiones de recursos de red del entorno controlado de documentos. Al monitorear el comportamiento del entorno controlado de documentos en busca de peticiones de recursos de red sospechosas, los efectos remotos provocados por los documentos electrónicos y reconocidos como comportamiento sospechoso pueden tenerse en cuenta al determinar la puntuación de amenaza.

Específicamente los datos y/o el contenido de cualquier petición de recursos de red del entorno controlado de documentos meta pueden analizarse para uno o más de las siguientes clases de petición de recursos de red, que

pueden ser parte de las al menos dos clases de eventos predefinidos: peticiones de recursos de ubicaciones diferentes de un origen del documento electrónico (por ejemplo, sitios web externos sospechosos, mientras que estos sitios web ya son una fuente conocida de código malicioso (detección previa) o de contenido dudoso (categorización de contenido web); peticiones de recursos de ubicaciones en países diferentes al origen del documento electrónico; peticiones de documentos para los cuales una puntuación de amenaza fuera de un rango aceptable predefinido se ha determinado anteriormente peticiones de recursos de ubicaciones que coinciden con un patrón definido en una lista negra de ubicaciones, en particular la carga de secuencias de comandos sospechosas de dominios extranjeros (indicativos de "secuencias de comandos entre sitios"); solicita la transferencia de testigos de datos (por ejemplo, cookies, identificadores o claves de acceso) a ubicaciones que coinciden con un patrón definido en la lista negra de ubicaciones, en particular, mientras que la ubicación/dominio de destino puede ser conocido por servir contenido dudoso (esta práctica también se conoce como "robo de cookie"); y/o recursos que coinciden con un patrón predefinido de contenido malicioso. Un recurso puede ser otro documento electrónico (por ejemplo, una secuencia de comandos o un archivo de imagen) o un punto final de la API que desencadena efectos adicionales cuando se accede a él.

Además, con respecto el comportamiento monitorizado, preferiblemente uno o más de las al menos dos clases de eventos predefinidos son invocaciones función de secuencia de comandos dentro del entorno controlado de documentos. Las invocaciones de funciones de secuencia de comandos generalmente proporcionan un enlace confiable a las actividades desencadenadas por el documento electrónico en general y, por lo tanto, también a un comportamiento sospechoso.

En concreto, ha resultado instructivo analizar funciones de secuencia de comandos invocadas durante la ejecución del secuencia de comandos para una o más de las siguientes clases de invocación de función, que son parte de las al menos dos clases de eventos predefinidos: invocaciones de funciones que manipulan el documento representado sin interacción del usuario; invocaciones de funciones que desencadenan una descarga sin interacción del usuario (que es una forma común de distribuir malware a los usuarios que visitan un sitio web llamado "conducir por descarga"); y/o invocaciones de funciones enumeradas en una lista negra de funciones (por ejemplo, que comprende la función de JavaScript "eval"). Las funciones que manipulan el documento representación sin interacción del usuario a menudo indican un intento de desviar las interacciones del usuario y engañar al usuario para que realice una acción que no sea su intención principal ("superposiciones" o "cebos de clic"; por ejemplo, al plantar un control invisible sobre un legítimo control para provocar la activación del control invisible cuando el usuario realmente tiene la intención de operar el control legítimo, como un enlace o botón). Las funciones que desencadenan una descarga sin interacción del usuario son una forma común de distribuir malware a los usuarios que visitan un sitio web ("conducir por descarga"). Las listas negras de funciones se pueden usar para restringir en general el comportamiento aceptable en términos de funciones de secuencia de comandos, por ejemplo, penalizando las funciones que acceden a los recursos locales de los dispositivos del usuario y/o las funciones utilizadas para la ofuscación (aunque las partes ofuscadas se ejecutarán de todos modos dentro del entorno controlado del documento y se contabilizarán en la evaluación general de seguridad del documento electrónico).

Además, una o más de las al menos dos clases de eventos predefinidos pueden ser cambios del documento. La monitorización de este tipo de evento puede complementar o reemplazar la monitorización de eventos de invocación de funciones de secuencia de comandos, ya que no se refiere a las funciones como tales, sino a los efectos de su ejecución en el documento electrónico dentro del entorno controlado de documentos. De este modo, los tipos específicos de cambios en el documento pueden interpretarse como comportamiento sospechoso y contabilizarse en la puntuación de amenaza.

En particular, cambios del documento pueden ser analizados para una o más de las siguientes clases de cambio de documentos, que son parte de las al menos dos clases de eventos predefinidos: cambia la introducción de elementos ocultos o invisibles en el documento; y/o cambios que introducen elementos que se refieren a recursos de ubicaciones diferentes de un origen del documento electrónico. Dichos recursos externos pueden ser otros documentos electrónicos, por ejemplo, secuencias de comandos adicionales, a los que no se ha hecho referencia en el documento electrónico original y, por lo tanto, habrían evitado el análisis de contenido estático. Los elementos ocultos o invisibles generalmente indican un intento de desviar las interacciones del usuario para desencadenar efectos no deseados como se describió anteriormente. Este enfoque se puede utilizar para confirmar ciertas acciones basadas en las credenciales proporcionadas por el usuario en un momento anterior (por ejemplo, confirmar una autorización de seguridad o autorización de transferencia o enviar mensajes a nombre del usuario).

Los eventos registrados pueden ser almacenados junto con las respectivas clases de eventos en una base de datos de eventos. Esto permite recopilar un patrón o perfil de comportamiento del documento electrónico en cuestión. Según el perfil de comportamiento almacenado, se puede realizar una reevaluación posterior de la puntuación de amenaza, por ejemplo, cuando los recursos a los que hace referencia un documento determinado se clasifican posteriormente como maliciosos, lo que afecta la puntuación de amenaza de todos los documentos de referencia. Además, dicha información registrada permite reproducir la puntuación de amenaza determinada y analizar y reajustar las ponderaciones que conducen a una determinada puntuación de amenaza.

Para un acceso rápido y fácil por medio de dispositivos de seguridad tales como filtros de acceso, el puntuación de amenaza determinada del documento electrónico se puede almacenar en una base de datos de filtro. A diferencia de

una lista negra, que almacena solo una categorización de amenazas identificando solo documentos maliciosos, el almacenamiento de la puntuación de amenaza permite un control más detallado sobre un rango aceptable y para ajustes o personalizaciones posteriores con respecto al rango aceptable para diferentes aplicaciones.

Así, dentro del presente procedimiento la puntuación amenaza determinado puede preferiblemente ser comparado con uno o más rangos predefinidos de las puntuaciones de amenaza, en el que cada rango está asociado con una categoría de amenaza. El resultado de la comparación determina una categoría de amenaza del documento electrónico; por ejemplo, dice "inseguro" de los documentos "seguros". Por ejemplo, cualquier documento electrónico que tenga una puntuación de amenaza por debajo de una marca de agua baja predefinida puede clasificarse como inofensivo, mientras que los documentos que tienen una puntuación de amenaza entre dicha marca de agua baja y una marca de agua alta predefinida pueden clasificarse como cuestionables y los documentos que tienen una puntuación de amenaza por encima de dicha marca de agua alta pueden clasificarse como maliciosos. Como consecuencia, se puede denegar el acceso a documentos electrónicos maliciosos y los documentos electrónicos cuestionables pueden someterse a una revisión adicional, por ejemplo, por personal de seguridad especializado.

En una aplicación preferida de la presente invención el acceso a un documento electrónico se filtra mediante la determinación de una categoría de amenaza de un documento electrónico a ser visitada y negar el acceso cuando el documento electrónico pertenece a una categoría de seguridad predefinida (que sería la categoría "inseguro" en el ejemplo anterior).

Con referencia ahora a los dibujos, en los que las figuras son para ilustrar la presente invención y no para limitarla,

La figura 1 muestra esquemáticamente una arquitectura de un sistema para usar la presente invención con una pluralidad de entornos controlados de documentos que funcionan en paralelo; y
La figura 2 muestra esquemáticamente la estructura de un sistema informático y un procedimiento para determinar una puntuación de amenaza de un documento electrónico de acuerdo con la presente invención.

El sistema 1 se muestra en la figura 1 para el uso de la presente invención está adaptado para poblar una base 2 de datos de filtro sobre la base de una lista 3 de entrada de los documentos electrónicos, por ejemplo, páginas web identificadas por direcciones URL. Los proveedores de servicios de Internet (ISP), en particular los proveedores de banda ancha móvil pueden acceder a la base 2 de datos de filtro para operar filtros de acceso (no mostrados) para proteger a sus clientes del acceso a páginas web maliciosas. En base a la lista 3 de entrada, una pluralidad de rastreadores 4 web (también llamados "arañas") navegan permanentemente por la web e imitan el comportamiento de los usuarios humanos mientras analizan las páginas web visitadas. Utilizan las entradas de la lista 3 de entrada como puntos de partida y navegan, cargan y procesan automáticamente el contenido de todas las páginas web visitadas. Cada rastreador 4 web usa un entorno controlado de documentos para páginas web (en resumen, un entorno controlado web) para cargar y analizar una página web visitada en busca de comportamiento malicioso sin mostrar realmente ningún resultado gráfico de la página web en una pantalla.

Puede haber un número fijo o dinámico de rastreadores 4 web utilizados en paralelo, dependiendo de los recursos disponibles de un ordenador o plataforma de alojamiento. Cada rastreador 4 web puede implementarse en máquinas virtuales separadas. Alternativamente, el alojamiento puede ser realizado por un proveedor de Función como Servicio (FaaS), en el que cada uno de los rastreadores 4 web corresponde a una llamada de función. Por lo tanto, el sistema informático es un entorno adaptativo, que permite un análisis de una gran cantidad de documentos electrónicos y páginas web de manera oportuna (es decir, seguimiento y análisis de nuevos documentos tan pronto como aparecen) sin interacción humana. Dicho esto, será ventajoso distribuir rastreadores 4 web geográficamente y/o cambiar con frecuencia los rangos de IP empleados para evitar la detección y evasión de documentos y programas maliciosos. También podría ser ventajoso limitar artificialmente el ancho de banda de una conexión y retrasar el comportamiento de carga de los rastreadores 4 web para que el sistema permanezca sin ser detectado.

Las direcciones URL en la lista 3 de entrada se envían a los rastreadores 4 web por un despachador 5. El despachador 5 monitoriza un conjunto de rastreadores 4 web (cada uno de los cuales comprende una instancia de entorno controlado) y espera hasta que uno de los rastreadores 4 web esté en estado inactivo. Si no hay disponible un rastreador 4 web inactivo, se puede crear una nueva instancia, dependiendo de los recursos disponibles y la potencia informática. Una vez que un rastreador web inactivo está disponible, el despachador 5 reinicia el rastreador web inactivo y envía una entrada de la lista 3 de entrada al rastreador web inactivo. Tan pronto como finaliza el análisis de la página web correspondiente, el rastreador 4 web envía los resultados de su análisis (URL accedida, eventos observados) a un recopilador 6, que recopila esos resultados y actúa como un módulo de puntuación del sistema 1.

En base a los resultados recibidos, el recolector 6 determina una puntuación de amenaza dependiendo de una pluralidad de ponderaciones numéricas predefinidas obtenidas de una base 20 de datos de ponderaciones. La base 20 de datos de ponderaciones comprende una ponderación numérica para cada una de las clases de eventos predefinidas a las que pertenecen los eventos registrados por el rastreador 4 web. Para determinar la puntuación de amenaza, el recolector 6 resume los eventos observados ponderados con sus respectivas ponderaciones numéricas. La suma resultante de ponderaciones numéricas es la puntuación de amenaza de la página web analizada. Opcionalmente, las correlaciones entre los eventos observados también pueden tenerse en cuenta al

determinar la puntuación de amenaza; por ejemplo, cuando se observa una combinación de ciertos eventos (correlacionados), la puntuación de amenaza se puede aumentar además de la suma de las ponderaciones numéricas de los eventos individuales debido a una "coincidencia de combinación". Por ejemplo, una observación combinada de los eventos "manipulación del documento procesado", "transferencia de datos a una ubicación en la lista negra" y "activación de descarga de fondo" puede ser penalizada de esta manera.

El colector 6 escribe la puntuación de amenaza y la URL de la página web analizada a la base 2 de datos de filtro. La base 2 de datos de filtro puede contener preferiblemente las puntuaciones de amenaza determinadas, así como todos los comportamientos maliciosos detectados (eventos) que han sido detectados y registrados por los rastreadores 4 web en relación con el documento electrónico procesado. Los documentos electrónicos que tienen una puntuación de amenaza por encima de un cierto umbral (marca de agua alta) se marcan como "malos" o "inseguros" y, opcionalmente, se colocan en una lista negra separada. Los documentos con una puntuación de amenaza de cero (marca de agua baja) se marcan como "buenos" o "seguros" y, opcionalmente, se colocan en una lista blanca separada. Los documentos con una puntuación de amenaza entre la marca de agua alta y la marca de agua baja se marcan como "revisión de necesidades" y, opcionalmente, se colocan en una lista separada para análisis humano/manual posterior, después de lo cual se pueden clasificar manualmente como "seguros" o "inseguros".

Como se indica en la figura 2 cada uno de los rastreadores 4 web es en sí mismo un sistema informático que comprende un módulo 7 de entorno controlado de documentos y un motor 13 de inteligencia de amenazas. El módulo 7 de documento entorno controlado se implementa con un motor de navegador estándar sin una interfaz gráfica de usuario para acelerar la ejecución de la carga de páginas web. Tiene una API para permitir el control y la monitorización del motor del navegador (por ejemplo, PhantomJS, consulte http://phantomjs.org/). Para eliminar la posibilidad de que el entorno controlado sea detectado por código malicioso y, como resultado, no se realice el comportamiento sospechoso (evasión de entorno controlado), se pueden usar los mismos motores de navegador que se usan en los navegadores de escritorio (Chrome, Firefox, Safari, Internet Explorer) dentro del entorno controlado de documentos como motores de entorno controlado. El motor de entorno controlado, por ejemplo, podría estar utilizando el motor de representación del navegador WebKit con un motor de JavaScript integrado. Las API permiten conectar el motor 13 de inteligencia de amenazas o más específicamente un módulo 8 de monitoreo y un módulo 21 de simulación de interacción del motor 13 de inteligencia de amenazas. El motor 13 de inteligencia de amenazas está configurado para controlar el módulo 7 de entorno controlado de documentos para realizar el acceso automático y la navegación del documento electrónico analizado, todo mientras se monitorean las amenazas relevantes de seguridad que emanan de dicho documento. Se puede implementar como un proceso separado que interactúa con el módulo 7 de entorno controlado de documentos a través de su API y controla y monitoriza estrictamente el motor del navegador dentro del módulo 7 de entorno controlado de documentos.

El rastreador 4 web espera el envío de una URL por el despachador 5. Tan pronto como se envía una URL, el módulo 7 de entorno controlado del documento comienza con el análisis de la página web proporcionada. Dado que el motor del navegador no necesita una representación gráfica real, la potencia de cómputo requerida para cargar la página web se reduce considerablemente en comparación con un navegador convencional y una gran cantidad de entornos controlados web se pueden operar en paralelo en una sola máquina física.

El módulo 7 de entorno controlado de documentos está configurado para cargar (etapa 12) y procesar un documento 10 electrónico (una página web) basado en una URL 11 de entrada enviada al módulo 7 de entorno controlado de documentos. La URL 11 de entrada es un punto de entrada a un sitio web (por ejemplo, http://www.cnn.com/) y actúa como un punto de partida para el análisis. El proceso de carga y representación de la página web correspondiente es similar a ingresar una URL en la barra de navegación de un navegador web cliente estándar, excepto por la ausencia de un dispositivo de salida gráfico. El motor de entorno controlado procesa el documento 10 electrónico y se procesa como si fuera necesaria una salida gráfica. En este proceso, todos los recursos externos se solicitan a los servidores web de destino a los que se refiere el documento 10 electrónico (por ejemplo, el código de la página web) y las secuencias de comandos ejecutados en el motor 15 de secuencias de comandos. El motor 15 de secuencia de comandos está configurado para ejecutar secuencias de comandos que están incrustadas en el documento electrónico o a las que se refiere el documento electrónico. Dichas secuencias de comandos pueden reaccionar activamente ante las acciones de un usuario o realizar acciones en segundo plano cuando se accede al documento electrónico (por ejemplo, páginas web que utilizan AJAX, API web o manipulación DOM).

El módulo 7 de entorno controlado de documentos comprende un motor 14 de interacción para realizar interacciones activas del usuario con el documento 10 electrónico. Las acciones de usuario activas realizadas pueden ser controladas por el motor 13 de inteligencia de amenazas. Se pueden usar para simular las acciones que se realizan cuando un ser humano lee y navega por el documento electrónico. Dichas acciones pueden ser cualquiera de los movimientos del ratón y clics y/o entradas de teclado.

El módulo 8 de monitorización está conectado al módulo 7 de entorno controlado de documentos. Específicamente, se adjunta a una pluralidad de ganchos pasivos (como administradores de eventos o devoluciones de llamada) proporcionados por la API del módulo 7 de entorno controlado de documentos. La selección de ganchos pasivos mediante los cuales el motor 13 de inteligencia de amenazas está conectado al módulo 7 de entorno controlado de documentos corresponde al menos a dos clases de eventos predefinidas monitoreadas por comportamiento

sospechoso. Por simplicidad en la representación ejemplar mostrada en la figura 2, el módulo 8 de monitorización comprende solo tres puntos 16, 17, 18 de entrada para ganchos pasivos. En particular, el módulo 8 de monitorización comprende un primer punto 16 de entrada para un enlace de E/S de red, un segundo punto 17 de entrada para un enlace de ejecución de secuencia de comandos y un tercer punto 18 de entrada para un enlace de cambio de documento. El módulo 8 de monitorización está configurado para monitorear el módulo 7 de entorno controlado de documentos para eventos que pertenecen a una de al menos dos clases de eventos predefinidas. Además, está configurado para registrar cada evento observado junto con una clase de evento respectiva a la que pertenece en una base 19 de datos de eventos como se explicará con más detalle a continuación. La base 19 de datos de eventos recopila todos los eventos recibidos mientras el documento 10 electrónico se procesa en el entorno controlado de documentos.

El primer punto 16 de entrada es llamada cuando el documento 10 electrónico procesado por el módulo 7 de entorno controlado de documentos lleva a cabo una red de E/S (es decir, haciendo el tráfico de red entrante o saliente). Por ejemplo, recibe eventos del módulo 7 de entorno controlado de documentos cada vez que se solicita un recurso de red externo desde una página web. Dicha petición de recursos puede ser típicamente la carga de una imagen, de partes de documentos de soporte (por ejemplo, código HTML adicional), de hojas de estilo o de código de secuencia de comandos adicional (por ejemplo, archivos JavaScript externos). Además, las secuencias de comandos ejecutados dentro del entorno controlado de documentos pueden iniciar la comunicación de red con un servidor de destino para intercambiar información (por ejemplo, AJAX o API web). Los metadatos de tales comunicaciones de red (es decir, el anfitrión y el puerto de destino), así como los contenidos (es decir, tipo de recurso, contenido de recurso) pueden ser analizados por el motor 13 de inteligencia de amenazas con el fin de encontrar comunicación para (a) objetivos diferentes al objetivo inicial, es decir, la ubicación del sitio web, (b) objetivos que se encuentran en un país diferente al objetivo inicial, (c) objetivos que ya son conocidos por recursos maliciosos, o (d) contenido de recursos que contener patrones de comportamiento malicioso. Todos los eventos maliciosos detectados después de analizar el desencadenante del primer punto 16 de entrada se registran en la base 19 de datos de eventos.

El segundo punto 17 de entrada es llamado cuando el documento 10 electrónico procesado por el módulo 7 de entorno controlado de documentos requiere la ejecución de una secuencia de comandos (por ejemplo, cuando se llama el motor JavaScript). Recibe eventos del módulo 7 de entorno controlado de documentos cada vez que se ejecuta una función de secuencia de comandos en el entorno controlado de documentos. El código de secuencia de comandos sospechoso se puede detectar cuando (a) ejecuta funciones de secuencia de comandos que manipulan el documento electrónico (por ejemplo, el árbol DOM de una página web) sin interacción del usuario (dicho comportamiento puede indicar superposiciones o cebo de clic), (b) desencadena una descarga sin interacción del usuario (dicho comportamiento puede indicar una unidad de descarga), o (c) ejecuta funciones utilizadas para ocultar el código del análisis de código estático o la coincidencia de patrones (por ejemplo, "eval()" en JavaScript). Todos los eventos observados que cumplen uno de esos criterios después de analizar el desencadenante del segundo punto 16 de entrada se registran en la base 19 de datos de eventos.

El tercer punto 18 de entrada es llamada cuando se cambia el documento 10 electrónico dictada por el módulo 7 de entorno controlado de documentos (por ejemplo, cuando se añade un elemento de un modelo de árbol de objetos de documento (DOM) de una página web, modificado o reemplazado). Recibe eventos del módulo 7 de entorno controlado de documentos cada vez que cambia la apariencia visible del documento electrónico (por ejemplo, según lo definido por el árbol DOM). Después del tercer punto 18 de entrada, se puede detectar un comportamiento malicioso como (a) introducción de elementos ocultos/invisibles (HTML) en el documento electrónico, o (b) introducción de nuevos elementos (HTML) que se refieren a recursos externos (por ejemplo, imágenes o secuencias de comandos de dominios extranjeros). Dado que los cambios en la apariencia visual son comúnmente utilizados por las páginas web dinámicas legítimas (AJAX), el comportamiento malicioso es más difícil de detectar en dichas páginas web. En cualquier caso, todos los eventos observados que cumplen uno de los criterios definidos se registran en la base 19 de datos de eventos.

También el módulo 21 de simulación interacción está conectado al módulo 7 de entorno controlado de documentos. Específicamente, está unido por los desencadenantes 24, 25, 26 de acción a una pluralidad de puntos de entrada (como funciones o rutinas) proporcionados por la API del módulo 7 de entorno controlado de documentos. La selección de puntos de entrada a los que están conectados los desencadenantes 24, 25, 26 de acción del motor 13 de inteligencia de amenazas corresponde a las interacciones del usuario simuladas por el módulo 21 de simulación de interacción.

El primer desencadenante 24 de acción simula la navegación acciones sobre el documento 10 electrónico. Dichas acciones de navegación incluyen seguir activamente un enlace a través de un objeto de navegación dentro del documento electrónico (por ejemplo, la página web). El motor 13 de inteligencia de amenazas, mediante la API del módulo 7 de entorno controlado de documentos, simula tales acciones de navegación, por ejemplo, simulando un clic del ratón en un elemento de navegación del documento 10 electrónico. Las acciones de navegación simuladas se realizan revisando una lista de todos los elementos de navegación disponibles en el documento 10 electrónico. Dicha lista se consulta a través de la API del módulo 7 de entorno controlado de documentos, que ha procesado el documento y, por lo tanto, conoce todos los elementos contenidos, incluidos los elementos de navegación.

El segundo desencadenante 25 de acción simula los movimientos del ratón sobre el documento 10 electrónico

representado. Sirve para simular que un ser humano mueve un ratón cuando visualiza el documento 10 electrónico, por ejemplo, sobre imágenes y otro contenido. Dichos movimientos del ratón pueden ser utilizados por documentos 10 electrónicos para activar la ejecución de secuencias de comandos que de otro modo permanecerían inactivos. Por ejemplo, tales desencadenantes se usan a menudo para cargar anuncios o reproducir sonido o música. La simulación de los movimientos del ratón se logra a través de puntos de entrada dedicados de la API del módulo 7 de entorno controlado de documentos.

El tercer desencadenante 26 de acción simula la interacción con los campos de formulario en el documento 10 electrónico. La disponibilidad de campos de formulario en el documento 10 electrónico (por ejemplo, formularios web como un formulario de inicio de sesión o un formulario de contacto) se detecta por la presencia de elementos de entrada dedicados, que pueden consultarse a través de la API del módulo 7 de entorno controlado de documentos. Si se detecta la presencia de al menos un campo de formulario, el motor 13 de inteligencia de amenazas simula el llenado de datos en el campo de formulario por medio de la API del módulo 7 de entorno controlado de documentos. Por ejemplo, se pueden usar datos de entrada de listas ejemplares preparadas o se pueden generar datos aleatorios para completar los campos del formulario. Dicha entrada puede ser utilizada por el documento 10 electrónico para activar funciones de secuencia de comandos. Estos se usan comúnmente para validar datos de entrada, pero pueden modificarse para causar un comportamiento malicioso que, por lo tanto, puede ser detectado por el motor 13 de inteligencia de amenazas.

Cuando el módulo 7 de entorno controlado de documentos ha terminado de procesar el documento electrónico y todos los eventos relevantes se han registrado en la base 19 de datos de eventos como se describió anteriormente, un informe 23 de la base 19 de datos de eventos que comprende todos los eventos registrados se transmite al recopilador 6 como explicado en relación con la figura 1 para determinar una puntuación de amenaza. Las ponderaciones numéricas utilizados por el recolector 6 para determinar la puntuación de amenaza pueden ser los siguientes para las clases de eventos mencionadas anteriormente (en la práctica, el número de clases de eventos será mayor y las ponderaciones se ajustarán manualmente para lograr la experiencia del consumidor deseada):

Clase de evento	Ponderación
peticiones de red de recursos externos	1
petición de red a un país diferente	1
petición de red de documento "inseguro"	5
petición de red a una ubicación en la lista negra	4
transferencia de datos a una ubicación en la lista negra	5
petición de patrón "malicioso» de coincidencia de recursos	4
manipulación del documento representado	3
descarga de fondo desencadenante	3
invocación de la función de secuencia de comandos en la lista negra	3
introducción de elementos ocultos o invisibles	4
introducción de referencia a recursos externos	2

Por ejemplo, cuando una página web se ha realizado una petición de red a un país diferente y provocó una descarga de fondo, su puntuación de amenaza será 1 + 3 = 4. Cuando una página web no ha mostrado ningún comportamiento sospechoso, su puntuación de amenaza será cero.

El motor 13 de inteligencia de amenazas comprende además un módulo 22 que está conectado a un gancho pasivo del módulo 7 de entorno controlado de documentos y puede ser desencadenado en los cambios de estado del entorno controlado de documentos monitorización del estado (por ejemplo, entorno controlado de carga, entorno controlado listo, URL recibido, comenzar a cargar, terminar de cargar, comenzar a representar, etc.).

Tras la inicialización del motor 13 de inteligencia de amenazas, las API del módulo 7 de entorno controlado de

documentos están conectados al motor 13 de inteligencia amenazas como se describe anteriormente. En particular, los ganchos pasivos están conectados a los puntos 16, 17, 18 de entrada del motor 13 de inteligencia de amenazas, de modo que ciertos eventos en el motor de entorno controlado de documentos se informan al motor 13 de inteligencia de amenazas y pueden usarse para analizar el comportamiento del documento 10 electrónico.

En lo que sigue, el proceso de determinación de una puntuación de amenaza del documento 10 electrónico se describirá con referencia al sistema 1 tal como se muestra en la figura 1 y un rastreador 4 web del sistema 1 como se muestra en la figura 2.

En primer lugar, un documento 10 electrónico se carga y representa mediante el módulo 7 de entorno controlado de documentos funcionamiento del entorno controlado de documentos. Mientras se carga y presenta el documento 10 electrónico, el módulo 8 de monitorización monitoriza el entorno controlado de documentos para detectar eventos activados por el documento 10 electrónico y que pertenecen a una de al menos dos clases de eventos predefinidas. Por ejemplo, cuando el entorno controlado de documentos es un entorno controlado web, la actividad del motor del navegador empleado se controla estrictamente con la ayuda de interfaces en el motor del navegador para detectar cualquier comportamiento sospechoso. Cada evento observado durante el monitoreo se registra en la base 19 de datos de eventos junto con una clase de evento respectiva a la que pertenece. Las clases de eventos monitorizadas generalmente incluyen diferentes tipos de peticiones de recursos de red del entorno controlado de documentos, invocaciones de funciones de secuencia de comandos dentro del entorno controlado de documentos y cambios de documentos del documento 10 electrónico representado.

Cada evento de petición de recurso de red desencadenado por el entorno controlado de documentos se analiza con respecto a sus metadatos y/o contenido. Según los metadatos y/o el contenido, los eventos de petición de recursos de red que coinciden con una de las siguientes clases se registran como eventos sospechosos para su consideración en la puntuación de amenaza: peticiones de recursos de ubicaciones diferentes de un origen del documento electrónico, peticiones de recursos de ubicaciones en países diferentes al origen del documento electrónico, peticiones de documentos para los cuales una puntuación de amenaza fuera de un rango aceptable predefinido se ha determinado anteriormente (y se almacena en la base 2 de datos de filtro), peticiones de recursos de ubicaciones que coinciden con un patrón definido en una lista negra de ubicaciones, solicita transferir testigos de datos a ubicaciones que coinciden con un patrón definido en la lista negra de ubicaciones y recursos que coinciden con un patrón predefinido de contenido malicioso.

Cada evento de invocación de función de secuencia de comandos desencadenado por el entorno controlado de documentos se analiza con respecto a la función de secuencia de comandos respectiva. Los eventos en los que la función de secuencia de comandos invocada coincide con una de las siguientes clases se registran como eventos sospechosos para su consideración en la puntuación de amenaza: invocaciones de funciones que manipulan el documento representación sin interacción del usuario, invocaciones de funciones que desencadenan una descarga sin interacción del usuario e invocaciones de funciones enumeradas en una lista negra de funciones.

Finalmente, cada evento de cambio de documento desencadenado por el entorno controlado de documentos se analiza con respecto al tipo respectivo de cambio realizado en el documento electrónico. Específicamente, los cambios que introducen elementos ocultos o invisibles en el documento y/o los cambios que introducen elementos que se refieren a recursos de ubicaciones diferentes de un origen del documento electrónico se registran como eventos sospechosos para su consideración en la puntuación de amenaza.

Mientras que el control del entorno controlado de documentos para las clases de eventos descritos anteriormente, el módulo 7 de entorno controlado de documentos es controlado por el módulo 21 de simulación interacción del motor 13 de nivel de amenaza para simular las interacciones del usuario con el documento 10 electrónico. Las interacciones del usuario simuladas por el módulo 21 de simulación de interacción comprenden movimientos simulados del ratón, clics simulados del ratón y/o entradas del teclado. Todos los eventos que ocurren durante el proceso se registran y almacenan junto con las respectivas clases de eventos en la base 19 de datos de eventos.

Cuando finaliza el análisis, que incluye la carga, la representación y la simulación de la interacción del usuario, el colector 6 determina una puntuación de amenaza del documento electrónico basándose en ponderaciones numéricas predefinidas almacenadas en la base 20 de datos de ponderaciones que está asociada con cada una de las clases de eventos predefinidas a las que pertenecen los eventos grabados.

La puntuación amenaza determinada del documento electrónico se almacena en una base 2 de datos de filtro. Además, cada puntuación de amenaza determinado se compara con un primer rango de puntuaciones de amenaza asociados con una categoría de amenaza "buena" o "segura" (por ejemplo, una marca de agua baja en la puntuación de amenaza cero), con un segundo rango de puntuaciones de amenaza asociados con una categoría de amenaza de "revisión de necesidades" (entre marca de agua baja y marca de agua alta) y con un tercer rango de puntuaciones de amenaza asociados con una categoría de amenaza "mala» o "insegura» (por encima de la marca de agua alta). El resultado de la comparación, es decir, la categoría de amenaza correspondiente también se almacena en la base 2 de datos de filtro junto con la puntuación de amenaza. Según las ponderaciones numéricas ejemplares mencionadas anteriormente, la marca de agua baja puede ser 4 y la marca de agua alta puede ser 6.

Dentro del ámbito de la presente invención de acuerdo con un procedimiento para filtrar el acceso a documentos electrónicos una categoría de amenaza de un documento electrónico a ser visitada puede ser determinada mediante la consulta de la base 2 de datos de filtro y negar el acceso cuando el documento electrónico pertenece a la categoría de amenaza "insegura".

El procedimiento realizado por los rastreadores 4 web en el ejemplo anterior puede preferiblemente ser definido en un producto de programa de ordenador para determinar una puntuación de amenaza de un documento electrónico, las partes de programa que comprenden producto de programa de ordenador, que cuando se cargan en un ordenador están diseñados para realizar las etapas del procedimiento descritos anteriormente. Este producto de programa informático puede distribuirse posteriormente, por ejemplo, a proveedores de Plataforma como servicio (PaaS) o Función como servicio (FaaS), donde el producto de programa informático se ejecuta para lograr el objetivo de la presente invención.

REIVINDICACIONES

1. Procedimiento para determinar una puntuación de amenaza de un documento electrónico (10), comprendiendo el procedimiento las etapas de:

- cargar y representar el documento (10) electrónico en un entorno controlado de documentos;
- consultar una lista de todos los elementos de navegación disponibles en el documento (10) electrónico desde el entorno controlado de documentos;
- controlar el entorno controlado del documento para simular la interacción del usuario con el documento (10) electrónico basado en la lista consultada;
- al cargar y representar el documento (10) electrónico y al controlar el entorno controlado del documento para simular la interacción del usuario con el documento electrónico, monitorear el entorno controlado del documento en busca de eventos activados por el documento (10) electrónico y pertenecer a uno de al menos dos eventos predefinidos clases
- registrar cada evento observado junto con una clase de evento respectiva a la que pertenece;
- determinar una puntuación de amenaza del documento (10) electrónico basado en ponderaciones numéricas predefinidas asociadas con cada una de las clases de eventos predefinidas a las que pertenecen los eventos registrados.

2. Procedimiento según la reivindicación 1, **caracterizado porque** una interacción simulada del usuario comprende movimientos simulados del ratón, clics del ratón y/o entradas del teclado simulados.

3. Procedimiento según una de las reivindicaciones anteriores, **caracterizado porque** una o más de las al menos dos clases de eventos predefinidas son peticiones de recursos de red del entorno controlado de documentos.

4. Procedimiento según la reivindicación 3, **caracterizado porque** los metadatos y/o el contenido de cualquier petición de recursos de red del entorno controlado de documentos se analizan para una o más de las siguientes clases de petición de recursos de red, que forman parte de al menos dos clases de eventos predefinidas:

- peticiones de recursos desde lugares distintos del origen del documento electrónico;
- peticiones de recursos desde ubicaciones en países diferentes al origen del documento electrónico;
- peticiones de documentos para los cuales una puntuación de amenaza fuera de un rango aceptable predefinido se ha determinado anteriormente;
- peticiones de recursos de ubicaciones que coinciden con un patrón definido en una lista negra de ubicaciones;
- peticiones de transferir testigos de datos a ubicaciones que coinciden con un patrón definido en la lista negra de ubicaciones; y/o
- recursos que coinciden con un patrón predefinido de contenido malicioso.

5. Procedimiento según una de las reivindicaciones anteriores, **caracterizado porque** una o más de las al menos dos clases de eventos predefinidas son invocaciones de funciones de secuencia de comandos dentro del entorno controlado del documento.

6. Procedimiento según la reivindicación 5, **caracterizado porque** las funciones de secuencia de comandos invocadas durante la ejecución de secuencia de comandos se analizan para una o más de las siguientes clases de invocación de funciones, que forman parte de al menos dos clases de eventos predefinidas:

- invocaciones de funciones que manipulan el documento representado sin interacción del usuario;
- invocaciones de funciones que desencadenan una descarga sin interacción del usuario; y/o
- invocaciones de funciones enumeradas en una lista negra de funciones.

7. Procedimiento según una de las reivindicaciones anteriores, **caracterizado porque** una o más de las al menos dos clases de eventos predefinidas son cambios de documentos.

8. Procedimiento según la reivindicación 7, **caracterizado porque** los cambios en el documento se analizan para una o más de las siguientes clases de cambio de documento, que forman parte de al menos dos clases de eventos predefinidas:

- cambios que introducen elementos ocultos o invisibles en el documento; y/o
- cambios que introducen elementos que se refieren a recursos de ubicaciones diferentes de un origen del documento electrónico.

9. Procedimiento según una de las reivindicaciones anteriores, **caracterizado por** almacenar los eventos grabados junto con las clases de eventos respectivas en una base (19) de datos de eventos.

10. Procedimiento según una de las reivindicaciones anteriores, **caracterizado por** almacenar la puntuación de amenaza determinado del documento electrónico en una base (2) de datos de filtro.

11. Procedimiento según una de las reivindicaciones anteriores, **caracterizado por** comparar la puntuación de amenaza determinada con uno o más rangos predefinidos de puntuaciones de amenaza, en el que cada rango está

asociado con una categoría de amenaza.

12. Procedimiento para filtrar acceso a un documento electrónico determinando una categoría de amenaza de un documento electrónico al que se accederá de acuerdo con la reivindicación 11 y denegando el acceso cuando el documento electrónico pertenece a una categoría de amenaza predefinida.

13. Producto de programa de ordenador para determinar una puntuación de amenaza de un documento electrónico, comprendiendo el producto de programa de ordenador partes del programa, que cuando se carga en un ordenador está diseñado para realizar las etapas de un procedimiento de acuerdo con una de las reivindicaciones 1 a 12.

14. Sistema (1) informático para determinar una puntuación de amenaza de un documento (10) electrónico, comprendiendo el sistema (1) informático:

- un módulo (7) de entorno controlado de documentos para cargar y representar un documento (10) electrónico,
- un módulo (21) de simulación de interacción conectado al módulo (7) de entorno controlado de documentos y configurado para consultar en el entorno controlado de documentos una lista de todos los elementos de navegación disponibles en el documento (10) electrónico y para controlar el módulo (7) de entorno controlado de documentos simular la interacción del usuario con el documento (10) electrónico basado en la lista consultada,
- un módulo (8) de monitoreo conectado al módulo (7) de entorno controlado de documentos y configurado para monitorear el módulo (7) de entorno controlado de documentos para eventos que pertenecen a una de al menos dos clases de eventos predefinidas y para registrar cada evento observado junto con un evento respectivo clase a la que pertenece, y
- un módulo de puntuación conectado al módulo (8) de monitorización y configurado para determinar una puntuación de amenaza basada en ponderaciones numéricas predefinidas asociadas con cada una de las clases de eventos predefinidas a las que pertenecen los eventos registrados por el módulo (8) de monitorización.

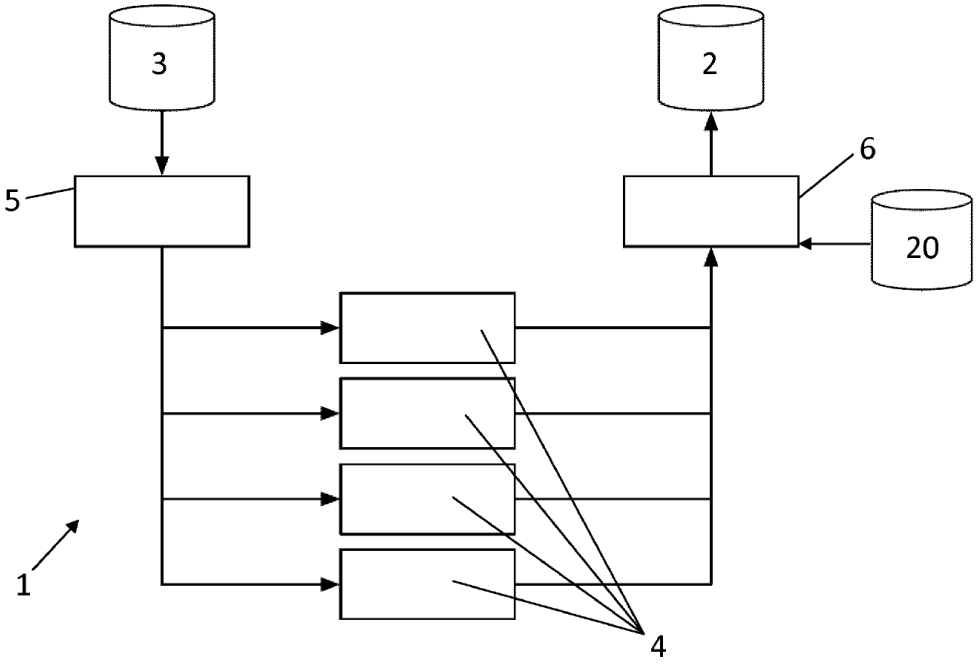

Fig. 1

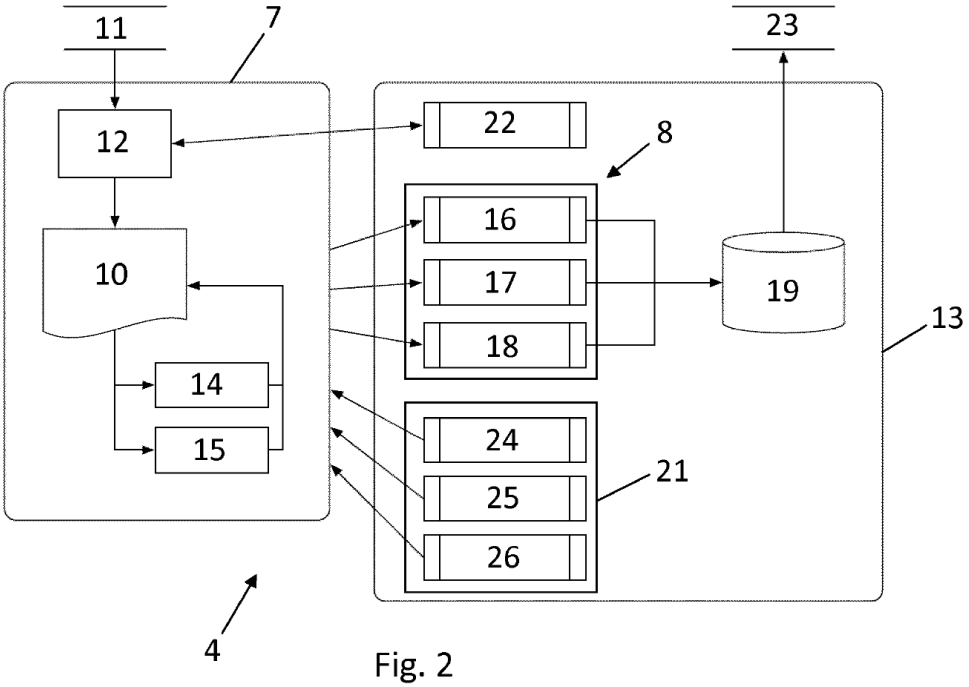

Fig. 2