

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 755 780**

51 Int. Cl.:

G06F 21/56 (2013.01)

G06N 20/00 (2009.01)

H04L 29/06 (2006.01)

H04W 12/12 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.09.2012 E 12184590 (3)**

97 Fecha y número de publicación de la concesión europea: **21.08.2019 EP 2610776**

54 Título: **Análisis estático y de comportamiento automatizado mediante la utilización de un espacio aislado instrumentado y clasificación de aprendizaje automático para seguridad móvil**

30 Prioridad:

16.09.2011 US 201161535804 P

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

23.04.2020

73 Titular/es:

VERACODE, INC. (100.0%)

65 Network Drive

Burlington, MA 01803, US

72 Inventor/es:

TITONIS, THEODORA H.;

MANOHAR-ALERS, NELSON R. y

WYSOPAL, CHRISTOPHER J.

74 Agente/Representante:

RIZZO , Sergio

ES 2 755 780 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Análisis estático y de comportamiento automatizado mediante la utilización de un espacio aislado instrumentado y clasificación de aprendizaje automático para seguridad móvil

CAMPO

- 5 [0001] El presente sistema y método está relacionado en general con la seguridad para dispositivos móviles y, más en concreto, con un análisis de aplicaciones automático que utiliza un espacio aislado instrumentado y una clasificación de aprendizaje automático para evaluar la seguridad de una aplicación móvil.

ANTECEDENTES

- 10 [0002] El *malware*, o *software malicioso*, es *software* diseñado para infiltrarse o dañar un sistema informático. Algunos ejemplos de *malware* incluyen virus informáticos, gusanos, troyanos, *spyware*, *adware* fraudulento, *scareware*, *crimeware*, y *rootkits*. Entre las formas de ataque se incluyen intentos de leer, alterar o destruir datos o poner en riesgo el sistema operativo del ordenador para tomar el control de la máquina. La motivación principal para el desarrollo y uso del *malware* es el beneficio económico.

- 15 [0003] A fin de conseguir el mayor impacto, el *malware* se crea normalmente para fijar como objetivo los dispositivos y sistemas operativos que tienen la mayor cuota de mercado. Debido al aumento del número de dispositivos móviles en todo el mundo, se ha incrementado drásticamente el número de variantes de *malware* que tienen como objetivo estos dispositivos. Los dispositivos móviles de consumo y de empresa están expuestos a un número sin precedentes de amenazas para la seguridad, que incluyen un aumento de un 400 % del *malware* para Android desde junio de 2010 a enero de 2011 (*Malicious Mobile Threats Report 2010/2011*, Juniper Networks Global Threat Center).
- 20

- [0004] Además de los vectores de ataque estándar que suponen una amenaza para la seguridad informática tradicional, los dispositivos móviles también son vulnerables a una amplia variedad de nuevas vulnerabilidades de seguridad que interceptan el micrófono, la cámara y el GPS. Si el *malware* tiene acceso a la raíz en un dispositivo móvil, es probable que tenga acceso al correo electrónico, a las credenciales bancarias, a los contactos, e incluso a la ubicación física del usuario.
- 25

- [0005] El *software antimalware* actual para dispositivos móviles se basa en una arquitectura utilizada de manera tradicional por los ordenadores personales. Este método utiliza firmas generadas a partir de análisis heurísticos rudimentarios para identificar y defenderse contra los ataques. Los dispositivos móviles no pueden soportar el proceso intensivo para memoria y CPU de consultar decenas de millones de firmas de *malware*. Los sistemas *antimalware* basados en firmas son prácticamente ineficaces a la hora de detectar variantes de día cero o desconocidas anteriormente. El *malware* no puede ser detectado a menos que ya se hayan obtenido muestras, que un profesional con experiencia haya realizado un análisis manual detallado, que se hayan generado firmas, y que se hayan distribuido actualizaciones a los usuarios. Este proceso puede llevar desde unas horas hasta varios días, quedándose algunas vulnerabilidades sin parchear durante años.
- 30

- [0006] El documento US 2011047594 (A1) está dirigido a un sistema y método para proporcionar asesoramiento sobre aplicaciones en dispositivos de comunicación móviles como teléfonos inteligentes, *netbooks* y tabletas. Un servidor reúne información sobre aplicaciones móviles, analiza las aplicaciones y produce una evaluación que puede asesorar a los usuarios sobre una variedad de factores, que incluyen seguridad, privacidad, impacto en la batería, impacto en el rendimiento, y uso de la red.
- 35

- [0007] En el documento WO 2008/103286 (A2), la evaluación de seguridad y la comprobación de la vulnerabilidad de aplicaciones de *software* se lleva a cabo, al menos en parte, en función de los metadatos de la aplicación a fin de determinar un nivel de seguridad apropiado y un plan de comprobación asociado que incluye múltiples tipos de análisis.
- 40

- [0008] El documento WO 2009/097610 (A1) da a conocer un sistema de detección de intrusos que recoge elementos obtenidos al ejecutar una aplicación en una máquina virtual, y analiza los elementos obtenidos mediante algoritmos de aprendizaje automático.
- 45

- [0009] Debido al volumen y al aumento de la sofisticación del *malware*, los analistas deben priorizarse en función de la prevalencia de la infección, la velocidad a la que se extiende, el impacto en la seguridad, y el esfuerzo requerido para eliminarlo. Los analistas de *malware* están capacitados para seguir una técnica en tres pasos, que incluye un análisis de superficie, un análisis del tiempo de ejecución y un análisis estático. Este proceso comienza con el análisis más directo y que utilice menos recursos y progresa hasta técnicas que requieren el mayor tiempo y esfuerzo. El análisis estático es la técnica más eficaz para determinar lo que hace realmente el *malware*, pero este nivel de análisis está reservado normalmente para el *malware* de mayor prioridad porque es demasiado costoso en lo que se refiere a esfuerzo y recursos.
- 50

- [0010] El uso de técnicas de ofuscación como *packers* binarios, cifrado, y código automodificable por parte de los programadores de *malware* hace que los análisis estáticos sean aparentemente imposibles. Cuando se lleva a cabo un análisis estático, el analista de *malware* se basa en su experiencia individual. Con base en este
- 55

conocimiento, categorizan las muestras en familias de manera que las nuevas variantes se puedan comparar con *malware* que hayan visto antes. Hay una escasez de analistas de *malware* con esta habilidad. Incluso en el Equipo de Respuesta para Emergencias Informáticas de EE.UU. (US-CERT, por sus siglas en inglés), un líder de confianza en ciberseguridad, hay pocas personas capaces de realizar este nivel de trabajo (*Building a Malware Analysis Capability*, CERT, 12/07/2011 Gennari *et al.*).

[0011] Se siguen encontrando aplicaciones maliciosas en páginas web de terceros y tiendas de aplicaciones. Muchas webs de terceros alojan aplicaciones sin las debidas diligencias. Las tiendas de aplicaciones de Google, Apple y Amazon emplean un proceso de evaluación principalmente manual que es ineficaz e ineficiente. Se sigue encontrando un número alarmante de aplicaciones maliciosas en el Marketplace de Google y en la Appstore de Amazon. Las aplicaciones maliciosas pueden haberse convertido en un problema persistente para Google, que ha tenido que limpiar su tienda varias veces. Debido a alertas mandadas por un tercero, han sacado más de 50 aplicaciones en marzo de 2011, 36 en mayo y 10 aplicaciones maliciosas más en junio.

[0012] El proceso manual de evaluación de aplicaciones está alienando a los desarrolladores legítimos, que se están frustrando por todo el tiempo que tardan en que aprueben su aplicación y la publiquen en las tiendas de aplicaciones. De manera adicional, los desarrolladores son incapaces de llevar a cabo pruebas de seguridad, rendimiento, estabilidad y regresión adecuadas para asegurar la calidad antes de presentar su aplicación para su distribución porque existen pocos entornos de prueba para sistemas operativos móviles.

[0013] Las compañías móviles están en una batalla aparentemente interminable contra el *malware* para los recursos de red, los ingresos de los operarios y la confianza de los abonados. Se les cobra por proteger la seguridad de los consumidores a la vez que defienden sus activos de red principal del *malware* que consume banda ancha. Las compañías se enfrentan a pérdidas de ingresos atribuidas al *malware*, incluidas aquellas que mandan mensajes SMS premium o indeseados, se utilizan para ataques de denegación de servicio, o para dañar los dispositivos móviles de sus clientes, lo que daría como resultado una rescisión de abonados.

[0014] Los consumidores ignoran en gran medida las aplicaciones móviles maliciosas o anómalas, o aplicaciones, que están instaladas en sus dispositivos móviles. A fin de protegerse, se les aconseja que investiguen al editor de una aplicación, que comprueben los permisos de la aplicación, y que no instalen aplicaciones de tiendas de aplicaciones o páginas web de terceros. La mayoría de consumidores concederán permisos a las aplicaciones sin consideración, y no se tomarán su tiempo para investigar la fuente.

[0015] Las empresas y las agencias del gobierno suelen permitir a los empleados que utilicen sus propios dispositivos móviles para el trabajo, aumentando el riesgo de que entre *malware* en la red de empresa. Se están publicando actualmente programas de *software* de empresa reservados normalmente para PC como aplicaciones que permiten el acceso a información privada y financiera desde dispositivos móviles de empresa y personales. Los departamentos de IT se encuentran en una situación de clara desventaja debido que la proliferación de dispositivos móviles en la empresa desafía la estrategia de seguridad predominante de endurecer el perímetro y controlar el acceso a la red interna.

RESUMEN

[0016] Según un modo de realización, el presente sistema soporta *middleware* reutilizable, al que se le hace referencia como un servicio en la nube, y extremos extensibles, a los que se les hace referencia como la aplicación de suscriptor y el espacio aislado. Los extremos del presente sistema están designados como componentes de complementos previstos para dirigirse al sistema operativo del dominio de la plataforma y no están limitados a dispositivos móviles. Desde el punto de vista del servicio en la nube, la aplicación de suscriptor y el espacio aislado, que incluyen un servidor AV, los extremos son proveedores de código binario y un vector de características numéricas correspondientes para dicho código binario, respectivamente. El servicio en la nube incluye un servidor web, un controlador, un distribuidor, una base de datos, una consola, y componentes de agrupación en clústeres y visualización.

[0017] El presente sistema proporciona un medio automatizado para identificar aplicaciones maliciosas. Miles de programas de *malware*, un número que siempre está en aumento, se encuentran en libre circulación y se dirigen desde miles de puntos de origen hasta millones de usuarios. Un analista de *malware* es notificado de la aplicación cuestionable cuando se quejan suficientes usuarios, o de manera alternativa, si un analista de *malware* está explorando las tiendas de aplicaciones a mano o mediante alguna automatización primitiva. Si el analista de *malware* disecciona la aplicación de manera adecuada, podría encontrar la firma de código estático, las llamadas al sistema, o incluso el comportamiento de red, que provocó que los usuarios se quejaran. Este proceso manual actualmente no es escalable.

[0018] El presente sistema reduce la canalización sin filtro de aplicaciones de *malware* en libre circulación a un goteo de fuentes, que se reduce de manera adicional mediante la visualización y los trazados de conectividad. Actualmente, las aplicaciones anómalas se identifican antes en el proceso de distribución, a diferencia de esperar que los usuarios se quejen tras una amplia distribución cuando el daño ya está hecho.

[0019] El presente sistema proporciona la habilidad de analizar, identificar, comparar y archivar posible *malware* de una manera rápida, eficiente y en volumen. El proceso de automatización de un extremo a otro permite que

los analistas de *malware*, los proveedores de tiendas de aplicaciones, los desarrolladores de aplicaciones, las compañías de telefonía, los consumidores y las empresas identifiquen acciones maliciosas y clasifiquen rápidamente un comportamiento de amenaza de manera sistemática. Este proceso automatizado mitiga las exigencias innecesarias de recursos valiosos. El espacio aislado instrumentado proporciona un mecanismo transversal de Interfaz Gráfica de Usuario (GUI, por sus siglas en inglés) inteligente que simula cómo un usuario puede interactuar con la aplicación. El sistema sustituye lo que hasta el momento había sido un proceso manual que requería un número de diversas aplicaciones.

[0020] Los registros de salida del análisis de comportamiento proporcionan a un analista información detallada de las acciones del *malware*, que incluye, pero sin carácter limitativo, un resumen del análisis, resultados de análisis antivirus de terceros, registros completos de simulaciones en espacios aislados, capturas de pantalla, resúmenes y detalles de la cobertura transversal de la GUI, resúmenes y detalles de la actividad de la red, resúmenes y detalles del alcance de la IP de la red observado durante la simulación en un espacio aislado, resúmenes y análisis detallados anotados para registros de alto nivel como el gestor de actividades y los registros de eventos, resúmenes y detalles de la ejecución transversal de la interfaz de usuario, resúmenes y análisis anotados detallados de los registros de llamadas del sistema operativo de bajo nivel, análisis resumidos y anotados a lo largo de una línea de tiempo integrada a través de dichos registros, resúmenes y detalles del análisis de la integridad del sistema de archivos, resúmenes y detalles de los objetos de archivo transferidos por la red identificados que incluyen los resultados del análisis antivirus, resúmenes y detalles de la actividad del navegador, cronologías de comportamiento y perfiles estadísticos extraídos de las llamadas al sistema operativo, llamadas a la biblioteca a nivel de aplicación, así como las operaciones del sistema de archivos, perfiles de la memoria y/o la CPU, resúmenes y detalles de las alertas de detección de intrusiones, resúmenes y detalles de la carga de tráfico de red impuesta por los servidores de anuncios, y resúmenes y detalles del alcance de la red a los sitios maliciosos de la aplicación durante su ejecución.

[0021] Según un modo de realización, el análisis estático se automatiza mediante un proceso de decompilación de la aplicación y de extracción de una forma rudimentaria del código fuente original. La funcionalidad básica del *software* se registra a la vez que se mantiene escéptica a los aspectos concretos del código subyacente. Estas funciones básicas del *software* incluyen, pero sin carácter limitativo, métodos finales públicos, llamadas de API de base, invocaciones de método directas, constantes de cadenas e invocaciones de la API de la interfaz para HTTP, SMS, URL, redes, GPS y telefonía. También es un aspecto de la presente invención que la invención proporcione medios para una evaluación de riesgos inferencial de las capacidades del binario de la aplicación a la vez que se mantiene escéptica al flujo de control y de datos del binario. Al implementar una directiva a nivel de capacidades, pueden identificarse los binarios de la aplicación peligrosa y evitar que entren en la red de empresa y/o que se instalen en los dispositivos móviles de los empleados. Un análisis estático avanzado, que incluye la creación de un flujo de control y un gráfico de flujo de datos completos, puede llevarse a cabo de manera opcional para determinar con más detalle el comportamiento del binario de la aplicación. Un gráfico de flujo de datos completo puede determinar si los comportamientos peligrosos, como la filtración de datos, realmente ocurren solo con el análisis estático. Un gráfico de flujo de datos completo puede determinar si realmente se filtran datos sensibles del dispositivo. Un análisis estático rudimentario sin un flujo de datos completo puede ser capaz de determinar que se accede a la información personal y que la aplicación transfiere datos del dispositivo a través de una red, pero no puede determinar que la información personal son los datos que se transfieren hacia fuera del dispositivo. Un análisis estático con un flujo de datos completo puede determinar si son los datos sensibles los que se están transmitiendo hacia fuera del dispositivo utilizando técnicas de comunicación inseguras.

[0022] En el presente documento, se hace referencia a la recopilación combinada de los análisis estáticos, así como de comportamiento mencionados anteriormente, como el conjunto de programas de análisis.

[0023] En la actualidad, se requiere un esfuerzo humano significativo para identificar amenazas, extraer características de las amenazas, y codificar las características en *software* para detectar las amenazas. Según un modo de realización del presente sistema, este proceso laborioso se automatiza mediante el uso del aprendizaje automático y técnicas de minería de datos. Estas técnicas pueden reemplazar el equivalente de cientos de miles de horas de análisis detallados de especialistas. El proceso consiste en un conjunto de algoritmos, programas informáticos que llevan a cabo tareas basadas en los datos de entrada que aprenden con el tiempo según se introducen más datos, o muestras de entrenamiento, en el sistema. Al final de este periodo de aprendizaje, que es en realidad muy poco tiempo, el resultado es un modelo informático que es equivalente, y la mayoría de las veces mejor, que un humano entrenado para llevar a cabo la tarea de identificar *malware* en un dispositivo móvil.

[0024] A efectos de clasificación, las aplicaciones, las que tienen *malware* conocido y programas benignos, se ejecutan en el espacio aislado que genera los informes de análisis estáticos y de comportamiento. Después se extraen las características de estos informes. Un ejemplo de característica incluye, pero sin carácter limitativo, el programa que intenta acceder a una URL o dirección IP, cuántos cambios está haciendo a los archivos de inicialización del sistema operativo, etc. Estas acciones se convierten en un conjunto de datos que se introduce en un método de clasificación como, por ejemplo, pero sin carácter limitativo, una regresión logística o una máquina vectorial de apoyo.

[0025] El presente sistema entrena los sistemas y métodos de clasificación para reconocer *malware* del mismo modo que un cliente de correo electrónico puede reconocer el correo basura sin que un humano lo revise realmente.

5 **[0026]** Con el sistema actual, se evita que una aplicación maliciosa llegue a los consumidores al engancharse a la red de distribución de aplicaciones, agilizando la cola del análisis de aplicaciones, y mediante el etiquetado automático de aplicaciones anómalas al principio del proceso de distribución. Google, Amazon, Apple y otras tiendas de aplicaciones que utilizan el sistema actual pueden asegurar a sus clientes que las aplicaciones de sus tiendas de aplicaciones no representan ningún peligro. Además de encontrar *malware* furtivo y de día cero, los proveedores de tiendas de aplicaciones también pueden emplear un proceso de revisión para encontrar
10 aplicaciones que se bloqueen u otras acciones indeseables debido a la mera incompetencia de los desarrolladores.

[0027] Una forma eficaz para que los desarrolladores estén por delante de las vulnerabilidades de seguridad es crear aplicaciones de forma segura, desde cero. El sistema actual proporciona un medio automatizado para realizar pruebas de vulnerabilidad a lo largo del ciclo de vida del desarrollo de aplicaciones. Ofrece una solución
15 completa y que ahorra tiempo para desarrolladores de aplicaciones, gestores de construcción, equipos de control de calidad (QA, por sus siglas en inglés), probadores de intrusión y auditores de seguridad. Los desarrolladores pueden introducir de manera involuntaria *malware* o comportamientos de riesgo no deseados en sus aplicaciones cuando utilizan bibliotecas binarias de código creado por terceros. A menudo se utiliza código binario compartido para implementar funcionalidades como el soporte multiplataforma o la funcionalidad
20 requerida de una red publicitaria. Este código binario compartido puede analizarse durante el proceso de desarrollo utilizando análisis estáticos y de comportamiento para que la aplicación no herede *malware* o comportamientos de riesgo no deseados del código compartido.

[0028] Con el sistema actual, las compañías de telefonía pueden detectar y eliminar *malware* en su red antes de que se propague. Pueden ofrecer a sus abonados una experiencia móvil segura y receptiva al eliminar el
25 *malware* que consume ancho de banda de su red. Las compañías de telefonía tienen un conocimiento global de la situación y un archivo correspondiente de *malware* detectado, que incluye el punto de origen y el canal de distribución basado en el tiempo. Además, las compañías de telefonía móvil pueden extender el sistema actual para ofrecer nuevas vías para la generación de ingresos mediante controles corporativos y ofertas de conexiones directas a medida.

30 **[0029]** El sistema actual ofrece una protección proactiva y completa contra el *malware* sin depender de actualizaciones de firmas. La presente invención también proporciona medios para evaluar el rendimiento de la huella de un binario de aplicación (como por ejemplo, pero sin limitarse a la evaluación de riesgos, el alcance de la red, el rendimiento de la CPU, los requisitos de memoria y el uso de ancho de banda) en el dispositivo móvil. La presente invención también proporciona medios para marcar y disuadir uso de tales binarios de aplicaciones
35 marcados en dispositivos móviles de consumo.

[0030] La presente invención proporciona medios para detectar una fuga de información sensible escondida durante la ejecución del binario de la aplicación dentro de un espacio aislado. Al consumidor se le asegura que su ubicación, contactos, búsquedas en la web, música, fotos, correos electrónicos, mensajes de texto, llamadas de teléfono y credenciales de inicio de sesión están seguras de aplicaciones maliciosas en su dispositivo móvil.

40 **[0031]** La presente invención proporciona medios para hacer uso de análisis previos e identificar un binario de aplicación con características de comportamiento y/o estáticas anómalas dentro de una faceta concreta del conjunto de programas de análisis. Con este fin, para un binario de aplicación determinado, la invención compara un perfil derivado de una o más facetas del conjunto de programas de análisis con un perfil general/basado en agregados (por ejemplo, a lo largo de algún subconjunto de binarios de aplicación seleccionados en función de
45 algunos criterios como el intervalo de tiempo o la clasificación del *malware*) a fin de identificar anomalías dentro de dicha faceta del paquete de aplicaciones. Por ejemplo, esto permite identificar binarios de aplicación que consumen ancho de banda que no se consideran necesariamente *malware*.

[0032] Un lector experto en la materia puede apreciar que un pequeño número de binarios de aplicación pueden no estar lo suficientemente entrenados por el autómatas de la GUI (por ejemplo, debido a avisos de contraseña/nombre de usuario, errores y/o requisitos de configuración del entorno). La presente invención
50 proporciona medios para marcar y poner en la cola de forma autónoma estos binarios de aplicación para una posterior interacción con la interfaz de usuario asistida por humanos. La presente invención especifica el uso de un entorno de espacio aislado compatible pero sin conexión que conduce a la generación de un conjunto de programas de análisis compatible. Según el aspecto de la invención mencionado anteriormente descrito en el párrafo [0028], los binarios de aplicación que tengan un rendimiento menor de manera anómala con respecto al
55 perfil general/basado en agregados para la faceta transversal de la GUI del conjunto de programas de análisis se marcan para una inspección manual por parte de un operario.

[0033] El conocimiento de la situación de la seguridad móvil para la empresa se consigue mediante la percepción del *malware* dentro de un dispositivo o red, la comprensión del tipo de *malware*, y la proyección del estado del
60 *malware* en el futuro. El presente sistema proporciona seguridad de extremo y puede ayudar a los

5 departamentos de IT a garantizar el cumplimiento de las políticas de seguridad al asegurar que solo los dispositivos que se ajusten a estas y con aplicaciones no maliciosas puedan acceder a redes empresariales y servidores de correo electrónico. Mediante la implementación de una directiva a nivel de aplicación, pueden identificarse y eliminarse amenazas en los dispositivos móviles de los empleados tan pronto como aparezcan antes de que lleguen a la empresa.

[0034] Estas y otras características, aspectos y ventajas del sistema actual se entenderán mejor haciendo referencia a las siguientes descripciones y reivindicaciones. Este resumen es una introducción de conceptos. No pretende identificar las características clave ni debería utilizarse para limitar el alcance de la reivindicación.

BREVE DESCRIPCIÓN DE LAS FIGURAS

10 **[0035]** Los dibujos adjuntos, que se incluyen como parte de la presente memoria, ilustran el modo de realización preferido en la actualidad y, junto con la descripción detallada proporcionada anteriormente y la descripción detallada del modo de realización preferido proporcionada a continuación, sirven para explicar y enseñar los principios del presente sistema.

15 La figura 1 ilustra un dispositivo móvil de ejemplo conectándose a un servicio en la nube para su uso con el presente sistema, según un modo de realización.

La figura 1A ilustra un proveedor de una tienda de aplicaciones móviles que se conecta con el servicio en la nube según un modo de realización del presente sistema.

La figura 1B ilustra un ejemplo de web de compañías móviles de aplicaciones móviles conectándose a un servicio en la nube según un modo de realización del presente sistema.

20 La figura 1C ilustra un ejemplo de interfaz de aplicación de suscriptor para su utilización con el presente sistema, según un modo de realización.

La figura 1D ilustra un ejemplo de proceso de envío de una aplicación de suscriptor para su utilización con el presente sistema, según un modo de realización.

25 La figura 2 ilustra un ejemplo de componentes del servicio en la nube para su utilización con el presente sistema, según un modo de realización.

La figura 2A ilustra un ejemplo de interfaz de subida de una aplicación para su utilización con el presente sistema, según un modo de realización.

La figura 2B ilustra un ejemplo de componentes de complementos del servicio en la nube para su utilización con el presente sistema, según un modo de realización.

30 La figura 3 ilustra un ejemplo de distribuidor para su utilización con el presente sistema, según un modo de realización.

La figura 3A ilustra un ejemplo de aplicación de suscriptor que envía una solicitud de espacio aislado a un servicio en la nube para su utilización con el presente sistema, según un modo de realización.

35 La figura 3B ilustra un ejemplo de aplicación de suscriptor que solicita el progreso de un proceso de un servicio en la nube para su utilización con el presente sistema, según un modo de realización.

La figura 3C ilustra un ejemplo de un servicio en la nube que envía una notificación de finalización a un proceso de aplicación de suscriptor para su utilización con el presente sistema, según un modo de realización.

40 La figura 3D ilustra un ejemplo de interfaz para recibir una notificación de compleción para su utilización con el presente sistema, según un modo de realización.

La figura 3E ilustra un ejemplo de interfaz para un dispositivo móvil que muestra un enlace a un informe de análisis para su uso con el presente sistema, según un modo de realización.

La figura 4A ilustra un ejemplo de inicio de un espacio aislado para su utilización con el presente sistema, según un modo de realización.

45 La figura 4B ilustra un ejemplo de ciclo de vida de una aplicación dentro de un servicio en la nube según un modo de realización del presente sistema.

La figura 4C ilustra un ejemplo de *shell* o intérprete de comandos instrumentado en relación con un sistema operativo principal e invitado según un modo de realización del presente sistema.

50 La figura 5 ilustra un ejemplo de proceso para un servidor web administra una solicitud de espacio aislado según un modo de realización del presente sistema.

La figura 6 ilustra un ejemplo de proceso para un servidor AV (antivirus) según un modo de realización del presente sistema.

- La figura 7 ilustra un ejemplo de proceso para un controlador para su utilización con el presente sistema, según un modo de realización.
- La figura 7A ilustra un ejemplo de proceso para un distribuidor para su utilización con el presente sistema, según un modo de realización.
- 5 La figura 7B ilustra un ejemplo de proceso para una interfaz transversal de usuario autónomo durante la ejecución de un binario de aplicación en un espacio aislado, según un modo de realización.
- La figura 8A ilustra una tabla de base de datos de solicitudes de ejemplo para su utilización con el presente sistema, según un modo de realización.
- 10 La figura 8B ilustra una tabla de base de datos de aplicaciones de ejemplo para su utilización con el presente sistema, según un modo de realización.
- La figura 8C ilustra una tabla de base de datos de un mapa de controlador de ejemplo para su utilización con el presente sistema, según un modo de realización.
- La figura 8D ilustra una tabla de base de datos de registros de ejemplo para su utilización con el presente sistema, según un modo de realización.
- 15 La figura 8E ilustra una tabla de base de datos de análisis AV de ejemplo para su utilización con el presente sistema, según un modo de realización.
- La figura 8F ilustra una tabla de base de datos de progreso de ejemplo para su utilización con el presente sistema, según un modo de realización.
- 20 La figura 9 ilustra un ejemplo de interfaz de informe de análisis exhaustivo para su utilización con el presente sistema, según un modo de realización.
- La figura 10 ilustra un ejemplo de interfaz de informe de usuario final para su utilización con el presente sistema, según un modo de realización.
- La figura 11A ilustra un ejemplo de ejecución correcta de un proceso de solicitud de espacio aislado para su utilización con el presente sistema, según un modo de realización.
- 25 La figura 11B ilustra un ejemplo de generación de características para un identificador de solicitudes según un modo de realización del presente sistema.
- La figura 11C ilustra un ejemplo de proceso para evaluar la validez de los registros, según un modo de realización del presente sistema.
- 30 La figura 11D ilustra un ejemplo de análisis de un archivo de registro según un modo de realización del presente sistema.
- La figura 12 ilustra un ejemplo de generación de características numéricas a partir de un archivo de registro de red según un modo de realización del presente sistema.
- La figura 12A ilustra un ejemplo de análisis estático automatizado según un modo de realización del presente sistema.
- 35 La figura 13 ilustra secciones de ejemplo de un vector de características según un modo de realización del presente sistema.
- La figura 14 ilustra un ejemplo de una agrupación en clústeres de un conjunto finito de vectores de características en un conjunto finito de clústeres según un modo de realización del presente sistema.
- 40 La figura 15 ilustra un ejemplo de agrupación de clústeres en línea de nuevos vectores de características frente a clústeres predefinidos según un modo de realización del presente sistema;
- La figura 16A ilustra un ejemplo de etiquetado de un clúster para representar binarios de aplicación con características similares según un modo de realización del presente sistema.
- La figura 16B ilustra, con fines comparativos, un ejemplo de representación alternativa (mediante una agrupación en clústeres jerárquica) de un clúster en dos dimensiones.
- 45 La figura 16C ilustra un ejemplo de matriz de distancia correspondiente entre pares de vectores de características y cómo proporciona la base para generar clústeres según un modo de realización del presente sistema.
- La figura 17 ilustra un ejemplo de una tabla de base de datos de vectores de características según un modo de realización del presente sistema.
- 50 La figura 18 ilustra un ejemplo de computación de una asignación de clúster inicial para un conjunto de vectores de características según un modo de realización del presente sistema.

La figura 18A ilustra una tabla de clústeres de ejemplo.

La figura 19 ilustra un ejemplo de asignación de un nuevo vector de características en un conjunto calculado previamente de clústeres según un modo de realización del presente sistema.

La figura 20 ilustra una consola de ejemplo según un modo de realización del presente sistema.

5 La figura 21A ilustra un ejemplo de vista de estadísticas básicas de la página de análisis de la base de datos de la consola del sistema según un modo de realización del presente sistema.

La figura 21B ilustra análisis de agrupaciones en clústeres y redes presentados en una vista de ejemplo de la página de análisis de la base de datos según un modo de realización del presente sistema.

10 La figura 21C ilustra análisis de una evaluación de riesgo y de un objeto de descarga en una vista de ejemplo de la página de análisis de la base de datos según un modo de realización del presente sistema.

La figura 22 ilustra un ejemplo de una interfaz de vista del sistema de la consola del sistema según un modo de realización del presente sistema.

La figura 23 ilustra un ejemplo de interfaz de vista de controlador de la consola del sistema según un modo de realización del presente sistema.

15 La figura 24 ilustra un ejemplo de una interfaz de vista de un servidor web de la consola del sistema según un modo de realización del presente sistema.

La figura 24A ilustra un ejemplo de vista de una vista de un servidor antivirus de la consola del sistema según un modo de realización del presente sistema.

20 La figura 25 ilustra un ejemplo de vista de GeolIP (geolocalización de dirección de protocolo de internet (IP)) según un modo de realización del presente sistema.

La figura 26 ilustra un ejemplo de interfaz de análisis estático de visualización de red según un modo de realización del presente sistema.

La figura 27 ilustra un ejemplo de interfaz de análisis interactivo de visualización de red según un modo de realización de la presente invención.

25 La figura 28 ilustra un ejemplo de vista de distribuidor de la consola del sistema según un modo de realización del presente sistema.

La figura 29 ilustra un ejemplo de interfaz de vista de una Tarjeta de informe del identificador de solicitud según un modo de realización del presente sistema.

30 La figura 30 ilustra un ejemplo de interfaz de vista de una tarjeta rápida según un modo de realización del presente sistema.

La figura 31 ilustra un ejemplo de regla de evaluación de riesgo y sus elementos constitutivos.

35 **[0036]** Cabe destacar que las figuras no están necesariamente dibujadas a escala, y que elementos con estructuras o funciones similares se representan en general con numerales de referencia similares con fines ilustrativos a lo largo de las figuras. También cabe destacar que las figuras solo pretenden facilitar la descripción de los varios modos de realización descritos en el presente documento. Las figuras no describen necesariamente todos los aspectos de lo expuesto en el presente documento, y no limitan el alcance de las reivindicaciones.

DESCRIPCIÓN DETALLADA

40 **[0037]** Los sistemas de la técnica anterior proporcionan dos métodos para examinar la presencia de *malware* en un dispositivo móvil. En la primera opción, puede instalarse un detector AV en el dispositivo móvil para llevar a cabo un análisis estático de aplicaciones que también estén guardadas en el dispositivo. No obstante, mientras que los análisis AV normalmente hacen un uso intensivo de recursos, los dispositivos móviles, en cambio, no lo hacen. Por este motivo, algunos proveedores de análisis AV pueden elegir proporcionar tablas de búsqueda simples que buscan asignar un atributo único de una aplicación (p. ej., MD5, SHA1) a un análisis de infección calculado previamente, donde la presencia de dicho análisis en dicha tabla de búsqueda indica una aplicación de *malware* conocida e identificada previamente. Sin embargo, la detección de una aplicación infectada depende de dos factores, la versión de la tabla de búsqueda, y la población de un análisis de infección para dicha aplicación en la tabla de búsqueda.

45 **[0038]** En una segunda opción, el dispositivo móvil (mediante un análisis AV o mediante una página web) puede enviar una o más de sus aplicaciones a análisis estáticos, remotos, basados en la web. No obstante, los análisis estáticos son insensibles a nuevo *malware* que aún no ha sido analizado a mano y lo que es más importante, pueden emerger aspectos importantes de un comportamiento anómalo y malicioso de una aplicación durante la aplicación y esta información no se está aprovechando.

[0039] Principalmente, en cualquier caso, el conocimiento agregado de millones de análisis de este tipo no se está analizando para descubrir o investigar patrones anómalos que pueden emerger del análisis de análisis agregados.

5 **[0040]** El presente sistema incluye un sistema informático en red que permite a los suscriptores móviles, y a otros, enviar aplicaciones móviles para que se analicen a fin de determinar comportamientos anómalos y maliciosos, utilizando datos adquiridos durante la ejecución de la aplicación en un entorno controlado y altamente instrumentado para el que el análisis se basa en las ejecuciones, así como en datos agregados comparativos adquiridos a través de muchas ejecuciones de uno o más suscriptores.

10 **[0041]** La figura 1 ilustra un dispositivo móvil (1) de ejemplo que contiene la aplicación de suscriptor (10) que permite la conexión (15) con nuestro servicio en la nube (20) a través de Internet (30) para enviar (40) alguna aplicación corresidente (p. ej., 5, 6) para su análisis.

15 **[0042]** Para los dispositivos móviles (p. ej., 1), puede encontrarse una aplicación en uno de dos estados: empaquetado (es decir, en un formato agregado listo para la distribución a dispositivos móviles) y sin empaquetar (es decir, instalada en un dispositivo móvil (1)). La aplicación de suscriptor (10) puede enviar cualquier aplicación instalada (por ejemplo, 5) que se encuentre en el dispositivo. Asimismo, el sistema proporciona dos medios alternativos para el envío de aplicaciones empaquetadas al servicio en la nube (20). En primer lugar, una página web de carga proporciona medios para enviar una aplicación desde cualquier tipo de ordenador (por ejemplo, un ordenador personal) que tenga acceso a la aplicación empaquetada (prevista para dispositivos móviles). En segundo lugar, la API de un servicio en la nube (véase la FIG. 1A), permite el envío por
20 lotes en una o más aplicaciones empaquetadas (almacenadas en cualquier tipo de ordenador) al servicio en la nube (20).

25 **[0043]** La figura 1A ilustra un ejemplo de tiendas de aplicaciones para móviles (como el Android Market (50) o el Amazon Market (55)) haciendo uso de dicha API del servicio en la nube (60) para conectarse a nuestro servicio en la nube (20) por Internet (30) a fin de enviar por lotes (p. ej., 71 (se muestra), 72, 73, etc.) múltiples aplicaciones (p. ej., 6, 7, 8) para un análisis de *malware*. En la presente invención, cada tienda de aplicaciones está asociada con una clave de tienda de aplicaciones única (parte del campo *datSolic* que se encuentra en cada envío). Una clave de tienda de aplicaciones representa una secuencia de caracteres únicos predefinidos para asignarlos en el servicio en la nube a una fuente de envío (por ejemplo, el Amazon Market, una compañía móvil, una empresa, el público en general) y asociarlos con una cuenta del servicio en la nube. Cada envío realizado se asocia con una clave de la tienda de aplicaciones. El servicio en la nube rastrea la asociación de cada envío a su clave concreta de la tienda de aplicaciones. Después, las fuentes de envío pueden utilizar su correspondiente clave de tienda de aplicaciones para recuperar el estado de uno (o más) envíos realizados con la misma clave de tienda de aplicaciones para cualquier intervalo de tiempo arbitrario. Una fuente de envío cuenta además con informes analíticos basados en agregados que proporcionan datos resumidos extraídos de subconjuntos de
35 informes asociados con claves de la tienda de aplicaciones. Para mayor flexibilidad, una fuente de envío puede asociarse a varias claves de tienda de aplicaciones. Por ejemplo, una tienda de aplicaciones puede tener campañas de envío trimestrales, cada una de ellas realizada con una clave de tienda de aplicaciones diferente.

40 **[0044]** Tal como se muestra también en la FIG. 1B, el sistema actual puede ser utilizado por una entidad (como las compañías móviles (82) que son compatibles con dispositivos móviles (por ejemplo, 81) o una empresa que soporta usuarios con dispositivos móviles) para complementar un *proxy* de cortafuegos a nivel de aplicación (83) al proporcionar un filtrado de aplicaciones de *malware* a través del servicio en la nube (85), analizando las aplicaciones que se descarguen, independientemente de su origen o la genealogía de la tienda de aplicaciones o en el sitio web de descarga (84). Por ejemplo, una descarga de una aplicación (86a) se envía a través de la conexión de la compañía móvil (por ejemplo, la conexión GSM de un dispositivo móvil) y es reconocida (87a) como una descarga por un cortafuegos a nivel de aplicación (83), que sustituye la respuesta normal (88a, 88b) por una secuencia aumentada que comprende los pasos de recuperar la aplicación (88a, 88b) y, a continuación, comprobar (89a) el servicio en la nube (85) para determinar el estado de la aplicación. El servicio en la nube (85) responde con el estado conocido de la aplicación, en este caso, se muestra que está infectada (89b) El cortafuegos a nivel de aplicación (83) retransmite (87b) esta información a las compañías móviles (82) que pueden decidir detener la descarga o, de manera alternativa, publicar una notificación (86b) de que se está
45 descargando una aplicación infectada y permitir que el Usuario decida qué medida tomar. En algunos modos de realización, un dispositivo de administración de dispositivos móviles (MDM, por sus siglas en inglés) puede actuar como remitente. El MDM puede reconocer que el servicio en la nube no tiene información sobre una aplicación concreta y extraer la aplicación o determinar los metadatos de la aplicación y enviarlos al servicio en la
50 nube.

55 **[0045]** En algunos casos, las propias aplicaciones pueden no estar disponibles o puede que no se permita enviarlas al servicio en la nube para su análisis. En tales casos, los metadatos de la aplicación que describen las aplicaciones como el *hash*, el nombre, el desarrollador, la versión, la fecha de creación, y el tamaño pueden enviarse al servicio en la nube como un *proxy* para la aplicación. Luego el servicio en la nube puede adquirir la
60 aplicación de una tienda de aplicaciones para su análisis. Este enfoque no requiere que un cliente envíe su copia de la aplicación, que puede estar prohibido debido a las políticas de privacidad u otros acuerdos legales.

5 [0046] La ejecución de una solicitud puede llevar un periodo de tiempo variable (por ejemplo, de un minuto a dos docenas). Por este motivo, el presente sistema proporciona medios para cualquiera de los métodos del remitente mencionados anteriormente para, tras el envío, recibir (por correo electrónico) un informe posterior, que informa de que el análisis se ha completado. La figura 1C muestra una captura de pantalla de la recopilación de la dirección de correo electrónico del usuario por parte de la aplicación de suscriptor (10). Un panel de ajustes (2) permite al usuario introducir un correo electrónico (3). Tal como se ha mencionado, una vez que los resultados del análisis están disponibles, el servicio en la nube enviará un correo electrónico al usuario que contiene un enlace a un informe que detalla el análisis.

10 [0047] Para mejorar el tiempo de respuesta del servicio en la nube en este modo de realización, estos análisis en espacios aislados para las aplicaciones pueden almacenarse en la memoria caché en términos de un identificador único derivado de una aplicación (como por ejemplo, pero sin carácter limitativo, MD5 y/o SHA1), permitiendo por tanto una respuesta al análisis casi instantánea a todas menos a la primera solicitud para el servicio en la nube de una aplicación determinada.

15 [0048] La figura 1D ilustra un ejemplo de una aplicación de suscriptor (10) con una huella reducida. Un subproceso (31) en la aplicación de suscriptor (10) permite que el usuario seleccione (32) una aplicación instalada (por ejemplo, 5 en la FIG. 1). Luego, rellena (33) el mensaje para enviarlo al servicio en la nube, y después envía (34) un mensaje de solicitud de espacio aislado (es decir, enviar la aplicación (11)) al servicio en la nube. Entre otras cosas, el mensaje de solicitud de espacio aislado remite una copia de la aplicación corresidente del dispositivo móvil (1) al servicio en la nube (20). Un diseño encadenado permite que se envíen otras aplicaciones; esto es, sin tener que esperar a que se complete una solicitud de espacio aislado pendiente.

20 [0049] Un subproceso (12) permite que la aplicación de suscriptor consulte (37) al servicio en la nube (mediante el mensaje de solicitud de progreso (1)) las actualizaciones del estado para cualquier solicitud de espacio aislado pendiente. En respuesta a cualquiera de estas consultas, el servicio en la nube manda un mensaje de actualización del progreso (43) a la aplicación de suscriptor, que después almacena esta(s) actualización(es) en la base de datos (130). Para reducir el drenaje de la batería del dispositivo móvil, este proceso de consulta sigue un retardo variable que se deteriora cuanto más antiguo se haga el envío a menos que se renueve (por ejemplo, 1, 2, 4, 8, 1, 2, ...) debido a una nueva revisión explícita (por el usuario) del estado del envío pendiente. Finalmente, al recibir una actualización del progreso (43) que indica que se ha completado (42) una solicitud de espacio aislado pendiente, la aplicación de suscriptor (10) muestra un icono de notificación (véase la FIG. 3D) en la barra de notificación del dispositivo móvil.

25 [0050] El servicio en la nube (20) representa un servicio de planificación de carga escalable que empareja la solicitud de espacio aislado con la capacidad disponible en función de algunos criterios como por ejemplo, pero sin carácter limitativo, la capacidad de cálculo disponible, el tipo de recursos disponible, el estado de la conectividad, la tasa de fallo, la identidad de la aplicación y/o del dispositivo de la solicitud de espacio aislado, etc.

30 [0051] La figura 2 ilustra un ejemplo de servicio en la nube (20) que consiste en un servidor web (100) visible de forma externa, que tiene acceso a una red de área local interna (110), en la que el controlador (120), la base de datos (130), la consola (160) y uno o más nodo(s) distribuidor(es) (140, 150) están interconectados.

35 [0052] El servidor web (100) proporciona soporte para una serie de API web que permiten que los usuarios interactúen de forma remota con el sistema, un controlador (120) se utiliza para asignar solicitudes a un distribuidor (por ejemplo, 140), proporcionando soporte para uno o más espacios aislados. Todos los componentes inician y finalizan operaciones con respecto a una base de datos (130) Esta base de datos (130) proporciona una memoria persistente a lo largo de los componentes del servicio en la nube. Los componentes del servicio en la nube pueden estar distribuidos en distintos nodos o redes. Un planificador (145) pone en la cola y ralentiza la velocidad de la solicitudes entrantes presentadas al servidor web en una carga máxima al controlador (120). El distribuidor (p. ej., 140) invoca un componente de conjunto de programas de análisis (155) tras completar una simulación en espacio aislado a fin de aplicar una serie de análisis forenses (como por ejemplo, pero sin carácter limitativo, una clasificación de aprendizaje automático, análisis estáticos, análisis basados en agregados, análisis de redes, análisis de cronogramas, análisis de evaluación de *malware*, y generación de informes) a los registros de ejecución producidos por un espacio aislado dentro de un distribuidor. Los registros de ejecución se almacenan en el distribuidor dentro de la base de datos y se recuperan por el conjunto de programas de análisis para un análisis *post mortem*. Los análisis llevados a cabo por el conjunto de programas de análisis no necesitan estar situados en el mismo servidor, siempre y cuando estén conectados con la base de datos (130).

40 [0053] Una consola (125) está disponible mediante la web y proporciona soporte para consultas del progreso en tiempo real, del estado, y de los resultados seleccionados de conformidad con algunos criterios como por ejemplo, pero sin carácter limitativo, una clave de tienda de aplicaciones común, el intervalo de tiempo, o ambos. La consola (125) comprende varias vistas de proyección diferentes que representan abstractos estadísticos de las solicitudes seleccionadas. Algunas de estas vistas de proyección son páginas del servidor web (126) que proporcionan actualizaciones en tiempo real de solicitudes entrantes e indican velocidades de cola y estados de finalización entre otras cosas, una página de controlador (127) que proporciona actualizaciones en tiempo real de

solicitudes programadas e indica los resultados de *malware* así como los análisis de la calidad de la ejecución para estos entre otras cosas, una página de distribuidor (128) que proporciona actualizaciones en tiempo real de la utilización del espacio aislado (p. ej., 141, 142) a lo largo de uno o más distribuidores (p. ej., 140), y una página de base de datos (129) que proporciona actualizaciones en tiempo real de análisis basados en agregados para la evaluación de amenazas de *malware* para un subconjunto de solicitudes determinado.

[0054] Un emulador (p.ej., 330, 331) representa una imagen de un dispositivo móvil específico (p.ej., un dispositivo Android genérico, un dispositivo iOS genérico) que pueden personalizarse para reunir mediciones con alta ocultación. Un espacio aislado (p.ej., 320) representa una *shell* de flujo de control y de datos envuelta alrededor de un emulador de este tipo diseñado de este modo a fin de entrenar un emulador en una secuencia de etapas resiliente, invariable y predefinida, como por ejemplo, pero sin carácter limitativo:

- (a) crear imágenes limpias del dispositivo emulador (p.ej., 331),
- (b) colocar datos como una selección aleatorizada u ordenada previamente de correos electrónicos, datos de identidad de usuarios, datos de identidad de dispositivos, credenciales de inicio de sesión, números de tarjetas de crédito, números de teléfono, historial de coordenadas GPS, direcciones de internet, historial de navegación por internet, entradas en una agenda de contactos y direcciones, mensajes de sistema de mensajes cortos (SMS), registro de llamadas, grabaciones de audio y video, muestras de archivos de texto y datos, etc., en el dispositivo emulador,
- (c) iniciar una recogida de medidas,
- (d) instalar el binario de aplicación en el dispositivo emulador,
- (e) iniciar el binario de aplicación
- (f) limitar y monitorizar de manera interactiva la interfaz de usuario del progreso de la ejecución de la aplicación iniciada,
- (g) finalizar y/o reiniciar la aplicación,
- (h) finalizar la recogida de medidas,
- (i) recuperar los registros y datos obtenidos del dispositivo emulador, y finalmente,
- (j) propagar la liberación (es decir, la planificar la disponibilidad) del dispositivo emulador de nuevo a su distribuidor asociado (p.ej., 140).

Por estas razones, posteriormente, la *shell* de espacio aislado también puede denominarse *shell* instrumentada.

[0055] Sobre todo, en un entorno de servicio en la nube autónomo y de implementación, la *shell* de espacio aislado también debe añadir resiliencia significativa, monitorización, documentación, recuperación y medidas de limitación en el estado del dispositivo emulado, el progreso, y la disponibilidad en cada y durante cada una de estas etapas. En un modo de realización, la página del controlador (127) proporciona una monitorización a tiempo real del progreso de una solicitud en el servicio en la nube con una granularidad suficiente para monitorizar los pasos anteriores (a-j). La página del distribuidor (128) proporciona una detección en tiempo real de la disponibilidad y la utilización de espacios aislados.

[0056] Los distribuidores (p.ej., 140, 150) proporcionan acceso a entornos de computación virtualizados que ejecutan solicitudes de espacios aislados. El controlador (120) proporciona la planificación y la asignación de la solicitud de espacio aislado en los recursos informáticos disponibles. El servidor web (100) proporciona interfaces para consumidores y/o empresas para recibir solicitudes de espacio aislado y monitorizar el progreso de las mismas. Una base de datos (130) proporciona acceso a tablas de datos compartidos sobre el servicio en la nube, sus solicitudes del espacio aislado y las conclusiones resultantes.

[0057] La figura 2A ilustra un ejemplo de interfaz de carga de aplicación para su utilización con el presente sistema, según un modo de realización. La subida se consigue mediante el acceso a la dirección web del sistema (161), que proporciona un formulario alojado en la web que comprende:

- una casilla de selección de solicitud (162) en la que se selecciona uno de entre varios análisis,
- un control de selección de archivos (163) que permite al usuario especificar la aplicación móvil para subirla,
- un campo de correo electrónico (164), especificado por el usuario y utilizado por el sistema para enviar al usuario un enlace al informe al que puede accederse a través de la web, y
- un botón de enviar (165), que inicia la subida del binario de la aplicación móvil en el servicio en nube.

[0058] La figura 2B ilustra un ejemplo de servicio en la nube (20) que se conecta (por medio de una red de área local (110) o internet (30)) a componentes de complementos (170, 172, 180, 182). Los componentes de complementos comprendían, pero sin carácter limitativo, uno o más servidores AV, que se conectan mediante

una API de antivirus y uno o más espacios aislados, que implementan la especificaciones del *software* de la API del espacio aislado.

[0059] La API del antivirus (175) permite que el servicio en la nube (20) se conecte a servidores de AV potencialmente diferentes. La API del antivirus (175) permite:

- 5
- el envío de una solicitud de análisis AV con respecto a una solicitud de espacio aislado concreta,
 - la extracción de la aplicación de la base de datos (130), y
 - ingresar los datos del resultado del análisis antivirus en la base de datos (130).

10 **[0060]** Nótese que un servidor AV ejecuta solicitudes de análisis de manera asíncrona con respecto a la ejecución de una solicitud de espacio aislado dentro del servicio en la nube. La notificación del resultado del escaneo y el envío dan como resultado actualizaciones asíncronas a la base de datos (130). Pueden utilizarse múltiples servidores AV para distribuir la carga de análisis así como para conseguir una verificación cruzada de resultados de análisis AV de distintos proveedores.

15 **[0061]** La API de espacio aislado (185) permite al servicio en la nube (20) conectarse a espacios aislados (180, 182) basados en *hardware* (182) o en *software* (180). La API de espacio aislado (185) asegura el cumplimiento de las especificaciones del *software* proporcionadas posteriormente en la FIG. 7B. La API del espacio aislado puede permitir que el servicio en la nube interactúe con una entidad altamente instrumentada (es decir, basada en *hardware* o en *software*) que emule un dispositivo móvil del tipo especificado y un sistema operativo como por ejemplo, pero sin carácter limitativo, matrices de simuladores de *software*, dispositivos de *hardware* «con *rooting* o *jailbreak*», o una combinación de estos. El *rooting* o *rooteo* es una técnica de escalación de privilegios y/o un proceso que permite a los usuarios de dispositivos móviles que ejecuten el sistema operativo invitado alcanzar un control privilegiado (conocido como «acceso a raíz») en un subsistema de Android que permite que las aplicaciones del dispositivo *rooteado*, si son capaces, superen las limitaciones que las compañías y los fabricantes de *hardware* ponen en los dispositivos, dando como resultado la habilidad de alterar o reemplazar configuraciones y aplicaciones del sistema, ejecutar binarios de aplicación especializados que requieren permisos a nivel de administrador, o llevar a cabo otras operaciones que de lo contrario serían inaccesibles para un usuario de Android normal. El *rooteo* es análogo a los dispositivos con *jailbreak* que ejecutan el sistema operativo iOS de Apple.

20 **[0062]** La figura 3 ilustra un ejemplo de nodo distribuidor (140) único, que consiste en un servidor con subprocesos (300) que se ejecuta sobre un sistema operativo principal (310). El servidor con subprocesos (300) proporciona acceso a un conjunto finito de espacios aislados (320, 321, 322, 323, etc.), que podría estar creado sobre emuladores basados en *hardware* (330) y/o basados en *software* (p. ej., 331, 332, 333, etc.) de un dispositivo móvil (1). Cada uno de estos emuladores se ejecuta sobre instancias separadas (y posiblemente diferentes) de sistemas operativos invitados (p. ej., 341, 342, 343, etc.). Cada sistema operativo invitado de este tipo (como por ejemplo, pero sin carácter limitativo, AndroidOS o iOS) se virtualiza de manera independiente sobre un sistema operativo principal (310) común (como por ejemplo, pero sin carácter limitativo, Ubuntu o Windows) que se ejecuta en un nodo distribuidor (p. ej., 140). En el modo de realización preferido, cada nodo distribuidor puede asociarse con un número máximo de virtualizaciones concurrentes definidas en función de algunos criterios como el número de CPU disponible, el ancho de banda de red y/o la memoria.

30 **[0063]** La figura 3A ilustra un ejemplo de aplicación de suscriptor (205, 10) que envía (245) una solicitud de espacio aislado (200) que comprende datos como por ejemplo, pero sin carácter limitativo, datos de identidad del dispositivo (210), datos de identidad de usuario (220), datos de identidad de aplicación (230), y datos de identidad de red (240), al servicio en la nube (20). Por ejemplo, los datos de identidad del dispositivo comprenden datos como, pero sin carácter limitativo, el fabricante, la marca, la dirección MAC, y/o el número de serie del dispositivo móvil. Los datos de identidad del usuario comprenden datos como, pero sin carácter limitativo, la cuenta de usuario, la dirección de correo electrónico, el número de empleado de la empresa, la clave de identificación de la tienda de aplicaciones, etc. Los datos de identidad de la aplicación comprenden datos como, pero sin carácter limitativo, el MD5/SHA1 de la aplicación, el nombre del paquete, el nombre del archivo, o el código binario. Finalmente, los datos de identidad de la red comprenden datos como, pero sin carácter limitativo, la IP asignada del usuario, la subred, las coordenadas GPS. Los campos anteriores pueden proporcionarse de manera opcional con la excepción del nombre de archivo de la aplicación, el MD5, el código binario, y el correo electrónico del usuario.

40 **[0064]** En un modo de realización, antes de considerar cualquier solicitud de espacio aislado (200), el servicio en la nube (20) determina primero si la clave de la tienda de aplicaciones suministrada (271) de una solicitud de espacio aislado (200) es válida. Esto permite que el servicio en la nube ponga en la cola según la prioridad las solicitudes de espacio aislado en función de su clave de tienda de aplicaciones suministrada, así como una denegación de servicio para claves de tienda de aplicaciones inválidas. Por ejemplo, una clave de tienda de aplicaciones puede ser inválida o estar asociada a una cuenta caducada o sin fondos, mientras que se puede dar una prioridad menor a una solicitud de espacio aislado asociada a una clave de tienda de aplicaciones pública/compartida que a solicitudes de espacio aislado que paguen fuentes de envío.

5 **[0065]** La figura 3A también muestra que un mensaje de solicitud de espacio aislado (245) es gestionado por un subproceso (255) en el servicio en la nube (20) que luego remite (270) una respuesta a la aplicación de suscriptor (205) que transmite un identificador de solicitud (250) asignado a dicha solicitud de espacio aislado (245). Luego, el servicio en la nube pone en la cola (256) la solicitud de una evaluación de espacio aislado de la solicitud de espacio aislado para un procesamiento posterior. El identificador de solicitud (250) está hecho para ser único, así como válido a través de todos los componentes del servicio en la nube.

10 **[0066]** La figura 3B ilustra un ejemplo de interacción posterior entre el servicio en la nube (20) y la aplicación de suscriptor (10). Un subproceso (280) del progreso de las solicitudes (282) de la aplicación de suscriptor (10) actualiza la información al enviar un mensaje de solicitud de progreso (284) al servicio en la nube (20) utilizando el identificador de solicitud (250) asignado. El servicio en la nube (20) consulta la base de datos (130), que recupera el último registro de actualización de progreso (26) para dicho identificador de solicitud (250). Luego se reenvía (286, 287) el registro a la aplicación de suscriptor (10), que después actualiza el estado de la solicitud de espacio aislado correspondiente (tal como se describe en la FIG. ID)

15 **[0067]** La figura 3C ilustra un ejemplo de servicio en la nube (20) que envía una notificación de finalización (260) a la aplicación de suscriptor (10) de que se ha completado la solicitud de espacio aislado (200). Al mismo tiempo, el servicio en la nube (20) manda un enlace de informe (225) al correo electrónico especificado previamente en los datos de identidad del usuario (220). Luego, después de que la aplicación de suscriptor (10) reciba el mensaje de notificación de finalización (260), se actualiza el estado de la solicitud de espacio aislado correspondiente (tal como se describe en la FIG. 1D). En un modo de realización, los enlaces de los informes están protegidos del acceso por minería de datos por parte de rastreadores web y motores de búsqueda mediante claves de paso y el acceso a HTTP seguras.

25 **[0068]** La presente invención proporciona medios para precalcular (es decir, antes de que acceda un usuario) los informes de análisis así como para generar un informe de análisis a petición (es decir, en función del acceso en tiempo real por un usuario). Para gestionar de manera eficiente la recuperación y la actualización de los informes de análisis, la presente invención prevé el uso del almacenamiento en caché, el control de versiones, la revalidación y la generación a petición de informes de análisis. Por ejemplo, es deseable que los informes de análisis generados previamente incorporen nuevas mejoras, formatos y/o actualizaciones a su conjunto de programas de análisis subyacente y/o datos subyacentes hechas tras su generación. Por ejemplo, sería deseable propagar actualizaciones relacionadas con facetas del conjunto de programas de análisis como por ejemplo, pero sin carácter limitativo, actualizaciones debido a un nuevo resultado del antivirus, actualizaciones debido a análisis mejorados de aprendizaje automático, actualizaciones debidas a evaluaciones y directivas de riesgo mejoradas/a medida, actualizaciones debidas a las listas negras de nuevas webs maliciosas, actualizaciones debidas a análisis basados en agregados del alcance de la red IP, actualizaciones debidas a cambios en las directivas relacionados con la restricción del tráfico de red, actualizaciones debidas a cambios en las directivas relacionadas con la detección de la intrusión y la privacidad, actualizaciones debidas a mejoras en los metadatos del binario de aplicación, etc.) que pueden surgir desde la última vez que se generó un informe.

30 **[0069]** La figura 3D ilustra un ejemplo de vista del dispositivo móvil (1) con la aplicación de suscriptor (10) tras recibir la notificación de finalización (260 en la FIG. 3C) y mostrar un icono de notificación (270) en la vista del historial de notificaciones de sistema (273) en la pantalla del dispositivo. Al hacer clic en una notificación, se muestra la pantalla de resultados del análisis (FIG. 3E). Se muestra una notificación (p. ej., 270, 272, 274) para cada envío. La pantalla muestra el historial (265) de aplicaciones escaneadas, que muestra si están infectadas (271) o no (p. ej., 272, 274). También pueden mostrarse las fechas para envío. Asimismo, al hacer clic en una entrada individual se abre una nueva pantalla (véase la FIG. 3E) con información detallada sobre los resultados de dicho análisis y que contiene un enlace al informe en línea.

45 **[0070]** La figura 3E ilustra un ejemplo de dispositivo móvil (1) con la aplicación de suscriptor (10) después de que se haya prestado atención al icono de notificación (270), lo que resulta en la visualización posterior de detalles sobre el resultado del análisis (275) y que proporciona un enlace (276) al informe que documenta el análisis así como una visualización clara (277) del resumen de los resultados (es decir, infectado en este caso).

50 **[0071]** La figura 4A ilustra un ejemplo de entorno aislado (320) en un nodo distribuidor (p. Ej., 140). Un servidor con subprocesos (300) utiliza datos de identidad del dispositivo (210) (p. ej., la versión del sistema operativo, el tipo de dispositivo) para adquirir una imagen de emulador (p. ej., 321) para el dispositivo móvil (1). Después, el servidor con subprocesos (300) lanza un entorno aislado (320) que comprende una *shell* altamente instrumentada (400) alrededor de un emulador (p. ej., 331) de un dispositivo móvil (p. ej., 1) que interactúa tanto con el sistema operativo principal (310) como con el sistema operativo invitado seleccionado (341). El sistema operativo invitado proporciona un entorno de virtualización que permite la monitorización y el control de la ejecución de la aplicación en el emulador. Para permitir esta interacción, hay un puente principal-invitado (360) desde el sistema operativo principal (310) hasta el sistema operativo invitado seleccionado (p. ej., 341) que permite que los comandos se reenvíen desde el sistema operativo principal (310) hasta el sistema operativo invitado (310), así como que los datos fluyan en ambas direcciones.

60 **[0072]** La figura 4B ilustra un ejemplo de ciclo de vida de una aplicación (40) dentro del servicio en la nube (20). Después de subirla, un identificador de solicitud (402) es asignado a una solicitud de espacio aislado y su

aplicación se almacena (401) en la base de datos del servicio en la nube (130), indexada por su identificador de solicitud asignado (250). Si se determina que un análisis para dicha aplicación no está disponible previamente, se programará ejecución de la solicitud de espacio aislado. De lo contrario, se proporciona y se devuelve un análisis. Almacenar los análisis en la memoria caché es una característica deseable para disminuir la latencia de la respuesta tal como se indica en la FIG. 1B. La determinación de si un análisis almacenado en caché existe puede basarse en criterios como por ejemplo, pero sin carácter limitativo, si se ha analizado previamente la misma aplicación para este consumidor, para otro consumidor, o para cualquier consumidor en función de datos como, pero sin carácter limitativo, los datos de identidad de usuario. Por ejemplo, una búsqueda de análisis previa puede estar limitada a los consumidores solo del país Z, o de la empresa X, o de las compañías móviles Y. Según un modo de realización, el presente sistema permite un modo de reemplazo que permitiría que una aplicación analizada correctamente se vuelva a enviar para un análisis sin depender de análisis pasados. Esta evaluación del análisis de *malware* normalmente requiere una revalidación de análisis pasados y es deseable dicha característica. Si la solicitud de espacio aislado se ha de ejecutar, se asigna primero (402) a un espacio aislado (p. ej., 321) que se encuentra en algún nodo distribuidor (p. ej., 140). Después, un subproceso (300) en dicho distribuidor recupera (403) la aplicación (p. ej., 40) de la base de datos (130) y lo copia (404) en el sistema de archivos del sistema operativo principal (310). Después, se selecciona un sistema operativo invitado (p. ej., 341), se virtualiza y se inicializa (405) con una imagen de un dispositivo móvil (p. ej., 1) y después, la aplicación (p. ej., 40) es instalada (406), por el sistema operativo principal (310) mediante el puente principal-invitado (360, en la Fig. 4A) correspondiente, en el sistema operativo invitado seleccionado (p. ej., 341). Finalmente, después, la aplicación (p. ej., 40) se analiza (407), y la imagen del sistema operativo huésped seleccionado (p. ej., 341) se borra.

[0073] En un modo de realización previsto, la determinación de si utilizar un informe de análisis almacenado en caché puede someterse a una validación adicional utilizando un modelo de actualización de dependencia contra los varios elementos constituyentes del conjunto de programas de análisis de una manera análoga al uso de archivos Make en un sistema de archivos. De manera específica, la validez de un informe de análisis almacenado en caché depende de la novedad de su conjunto de programas de análisis asociado. Por su parte, la validez de cada uno de sus análisis depende de elementos relacionados con el control (como por ejemplo, pero sin carácter limitativo, analizadores y algoritmos), datos (como por ejemplo, pero sin carácter limitativo, archivos de registro y tablas), y/o la presentación (como por ejemplo, pero sin carácter limitativo, formatos XML/HTML). En el modo de realización previsto, una regla de dependencia simple valida el uso de un informe de análisis almacenado en caché si los elementos de control, datos y presentación constituyentes mencionados anteriormente para todos los análisis dentro de un conjunto de programas de análisis no se han modificado desde la marca de tiempo del informe de análisis generado almacenado en caché.

[0074] La figura 4C ilustra un ejemplo de *shell* instrumentada (400) en relación con el sistema operativo principal (310), el sistema operativo invitado seleccionado (p. ej., 341) y su correspondiente puente principal-invitado (360). Ilustra que los datos de comportamiento de red (410) se capturan fuera del sistema operativo invitado (341), las métricas de rendimiento (420), los eventos (430) a nivel de emulador (p. ej., 331), y los eventos (440) de API a nivel de sistema operativo invitado (p. ej., 341) se capturan en el sistema operativo invitado (341). La salida de esta instrumentación se almacena en archivos de registro (p. ej., 425, 435, 445, etc.) y se almacena después en el sistema operativo principal (310). Ahí, el distribuidor almacenará posteriormente estos archivos de registro en la base de datos (130) indexada por el identificador de solicitud único (250) correspondiente asociado con la solicitud de espacio aislado (200) que se acaba de analizar. En la presente invención, un emulador (p. ej., 311) está dotado de un acceso bidireccional a Internet y a la red. En un modo de realización, el acceso de red desde/hacia un emulador (p. ej., 331) es asignado por el sistema operativo invitado (p. ej., 341) subyacente a las interfaces de red del sistema operativo principal (310) auxiliar que luego proporciona acceso a las redes.

[0075] Se sabe que un binario de aplicación podría utilizar un tráfico HTTP seguro para ocultarse o por razones maliciosas. En un modo de realización previsto, todo el tráfico HTTP seguro desde/hacia un emulador (331) concreto es interceptado, inspeccionado, registrado y retransmitido por un *proxy* HTTPS interceptor (como BURP [<http://www.portswigger.net/burp/proxy.html>]) situado en el sistema operativo principal (310). El registro de transacciones resultante permitiría un análisis de contenido forense de transacciones HTTP seguras como por ejemplo, pero sin carácter limitativo, alertas de detección de intrusos, identificación de objetos transferidos por HTTP como por ejemplo, pero sin carácter limitativo, parámetros, fugas y archivos, y detección inferencial de la presencia de tráfico de red comprimido. Por estos medios, la presente invención proporciona medios para detectar la presencia de una carga de *malware* en múltiples etapas. Por ejemplo, un binario de *malware* en dos etapas ofrece su carga maliciosa por medio de la distribución de un binario de aplicación de primera etapa relativamente no malicioso que una vez ejecutado, simplemente descarga un binario de aplicación malicioso de segunda etapa que luego se instala y se inicia en un dispositivo móvil.

[0076] La presente invención proporciona medios para identificar, reconstruir y analizar objetos descargados así como subidos que se encuentran en las secuencias de red de captura. La presente invención identifica, reconstruye y analiza de manera autónoma la presencia de *malware* en objetos transferidos por la red, como por ejemplo, pero sin carácter limitativo, imágenes PNG/JPG/GIF, documentos de texto/HTML, archivos PDF, y objetos *flash*. En un modo de realización previsto, todos los objetos transferidos por la red identificables podrán

ser sometidos además (en función de criterios como el tipo de archivo) a análisis especializados como por ejemplo, pero sin carácter limitativo, análisis de vulnerabilidad de seguridad de JavaScript, validación de HTML/CSS, vulnerabilidades de seguridad de Adobe PDF/Flash, evaluación de riesgo de binarios de aplicación de segunda etapa, etc. De este modo, la aplicación proporciona medios para detectar la descarga de un objeto de red malicioso de segunda etapa intencionado o no (por ejemplo, una vulnerabilidad de seguridad de *flash*) de un binario de aplicación de primera etapa aparentemente no malicioso.

[0077] La presente invención también proporciona medios para reducir la velocidad de envío conjunta Y de múltiples fuentes de envío. En el modo de realización preferido, el servidor web (100) pone en la cola una versión persistente de la solicitud de espacio aislado (200) e inmediatamente devuelve el identificador de solicitud única asignado (250) de vuelta a la fuente de envío origen (p. ej., una tienda de aplicaciones) mediante el mensaje de respuesta de espacio aislado (515). Un planificador/servidor de puesta en cola (145) planifica después la solicitud de espacio aislado puesta en cola al enviar el mensaje de solicitud de espacio aislado (510) al controlador (120) pero en alguna velocidad X fijada en función de algunos criterios (como por ejemplo, pero sin carácter limitativo, la velocidad de finalización por hora y la capacidad del emulador disponible y/o modificada). De este modo, la invención proporciona medios de control de regulación para aplicar una reducción de escala de una velocidad X de envío arbitraria hasta una velocidad de procesamiento de Y. Estos medios permiten que una tienda de aplicaciones sea capaz de enviar en lotes miles de binarios de aplicación para su análisis sin espera. En un modo de realización previsto, el planificador/servidor de puesta en cola (145) también proporciona una planificación de prioridad de solicitudes de espacio aislado en función de criterios como por ejemplo, pero sin carácter limitativo, las claves de tienda de aplicaciones, el GeoIP de inicio de sesión/correo electrónico, etc., de solicitudes de espacio aislado en la cola.

[0078] La presente invención también proporciona acceso a la interfaz web a componentes de análisis individuales seleccionados del conjunto de programas de análisis como por ejemplo, pero sin carácter limitativo, evaluaciones de riesgo inferenciales mediante análisis estáticos de marcas rojas, análisis de red resumidos y con detalles, análisis de resultados resumidos en una página, y clasificación de *malware* para el binario de aplicación determinado. En el modo de realización preferido, si el binario de aplicación no se valida para un informe de análisis almacenado en caché, todas estas solicitudes de análisis especializados también dan como resultado la prórroga de una solicitud subyacente en una solicitud de espacio aislado normal sujeta al conjunto de programas de análisis exhaustivo. En el presente documento, a estas solicitudes se les hace referencia como solicitudes de prórroga.

[0079] La presente invención está diseñada para la escalabilidad de solicitudes de espacio aislado concurrentes. El controlador mantiene una relación de uno a varios con múltiples nodos distribuidores. En esencia, el controlador proporciona un servicio de asignación entre un identificador de solicitud único (250) para un espacio aislado adecuado seleccionado de un número arbitrario de nodos distribuidores (p. ej., 145) en función de algunos criterios de adecuación individuales (p. ej., el tipo de dispositivo) y/o agregados (p. ej., el equilibrio de carga) y donde los nodos distribuidores pueden no estar bien colocados en la red. Por su parte, un nodo distribuidor puede tener uno o más espacios aislados por nodo y el número de espacios aislados por distribuidor puede establecerse para ser fijo o variable en función de algunos criterios como por ejemplo, pero sin carácter limitativo, un número de núcleos de CPU y/o la memoria disponible presentes dentro del nodo distribuidor.

[0080] Como mínimo, un distribuidor solo necesita el conocimiento de dicha asignación (es decir, un identificador de solicitud único para un espacio aislado disponible), a fin de recuperar todos los datos de la solicitud de espacio aislado asociados de la base de datos compartida por la red (130). Tras completar una solicitud de espacio aislado, un distribuidor solo necesita almacenar todos los registros de ejecución (p. ej., 425, 410) extraídos durante la simulación de la solicitud de espacio aislado en la base de datos compartida en la red (130). Los nodos distribuidores pueden desplegarse a gran escala en parques de servidores accesibles por la red (como los servicios en la nube de Amazon) para abordar las cuestiones de escalabilidad mientras que los componentes restantes del servicio en la nube pueden ejecutarse en cualquier otra parte. Para implementaciones de parques de servidores a gran escala puede ser necesario reducir la carga de red impuesta en el componente de base de datos (130) compartida por la red. Un lector experto en la materia apreciará que podría utilizarse una base de datos local/almacenada en caché después para reducir esta carga de red. En concreto, cada partición o subconjunto de distribuidores remotos (es decir, en parques de servidores) se asociarían a una base de datos local (al parque de servidores) y luego a cada base de datos local se le asignaría una planificación de actualizaciones por lotes para transmitir por lotes de manera eficiente registros de ejecución completados en la base de datos compartida por la red. Por eficiencia, una base de datos local puede truncarse de manera periódica de tales registros transferidos con éxito.

[0081] El modo de realización preferido se basa en el controlador para emitir una notificación asíncrona (de una asignación) a un nodo distribuidor. Además, en un modo de realización previsto, un distribuidor puede sondear la base de datos compartida por la red para dicha asignación en su lugar. No obstante, el modo de realización preferido es más susceptible de un despliegue a gran escala por evitar este esfuerzo de sondeo. Asimismo, el nodo controlador del modo de realización preferido proporciona medios para permitir un punto de toma de decisiones capaz de gestionar, repartir, activar y desactivar instancias dinámicas de nodos distribuidores para atender la demanda de los servicios.

[0082] La figura 5 ilustra un proceso de ejemplo para un servidor web (100) que tramita una solicitud de espacio aislado (200). El servidor web (100) asigna un identificador de solicitud único (250), inicia un nuevo subproceso (p. ej., 501), y registra la solicitud de espacio aislado (200) en la base de datos (130) indexada por dicho identificador de solicitud único (250). El servidor web (100) luego envía (506) una solicitud de análisis (505) que desencadena un análisis de antivirus asíncrono (es decir, análisis estático) desde un servidor AV (p. ej., 570). Cabe destacar que la solicitud de análisis (505) envía el identificador de solicitud (250) al servidor AV (p. ej., 170). El subproceso (501) procede después a enviar (507) una solicitud de reparto de espacio aislado (510) al controlador (120) que solicita una asignación (adecuada) de esta solicitud de espacio aislado (p. ej., 200) en un espacio aislado en algún nodo distribuidor (p. ej., 140), basándose esto en criterios como los recursos de distribuidor disponibles y/o los datos de identidad del dispositivo (210). El subproceso (501) espera (512) que finalice (p. ej., 515) la solicitud de espacio aislado (200) y registra (514) los resultados en la base de datos (130). En algún momento, los análisis AV asíncronos de la aplicación notificarán (517) su finalización y posteriormente actualizarán (516) la base de datos (130) con los resultados de análisis correspondientes.

[0083] La solicitud de espacio aislado puede terminarse debido a un tipo de finalización con éxito (515), almacenándose en caché (520) o un tiempo de expiración (530) del temporizador que controla la duración máxima permitida para la ejecución de cualquier solicitud de espacio aislado.

[0084] La figura 6 ilustra un servidor AV (170) de ejemplo. Primero, inicia un subproceso (610) para gestionar cada nueva solicitud de escaneo (505). El subproceso (610) utiliza el identificador de solicitud único (250) enviado con la solicitud de análisis (505) y recupera (615) la aplicación (20) de la base de datos (130). A continuación, calcula el MD5/SHA1 (616) para la aplicación (20) y consulta (620) la base de datos (130) para determinar si los resultados del análisis AV (630) para la misma aplicación (20) ya han sido registrados por alguna solicitud anterior de espacio aislado (es decir, distinta a 200). Si es así, almacena (621) los resultados en la base de datos (130), esta vez indexados por el identificador de solicitud (250) determinado y termina (622). De lo contrario, escanea (650) la aplicación y espera (670) a que la finalización del análisis desencadene (660) el almacenamiento de los resultados del análisis AV en la base de datos (130), indexada por el identificador de solicitud (250) y termina (680).

[0085] La figura 7 ilustra un servidor de controlador (120) de ejemplo. Tras recibir una solicitud de espacio aislado (510), se genera un nuevo subproceso (705) para encontrar un espacio aislado disponible (y/o adecuado) (p. ej., 320) para completar dicha solicitud de espacio aislado. El presente sistema dispone que dicho reparto de espacios aislados disponibles a una solicitud de espacio aislado pendiente se basa en criterios como por ejemplo, pero sin carácter limitativo, la disponibilidad, la carga del servidor, el tipo de dispositivo. El presente sistema dispone que el emulador elegido para ejecutar una solicitud se selecciona de entre los dispositivos con mayor cuota de mercado establecidos en los respectivos lanzamientos de SO de dispositivos soportados.

[0086] El presente sistema prevé que el espacio aislado se elija con criterios como, pero sin carácter limitativo, la identificación de dispositivo (210) del dispositivo móvil utilizado para enviar la solicitud o el remitente de la identidad de dispositivo de la API.

[0087] El controlador (120) mantiene una estructura de asignación de controlador que rastrea el reparto de espacios aislados a solicitudes de espacio aislado. La estructura de asignación de controlador también se utiliza para determinar qué espacios aislados están disponibles en ese momento y dónde están. La asignación del controlador es una estructura de datos compartidos que se actualiza a lo largo de todos los subprocesos de solicitud de espacio aislado y por tanto, el controlador aplica la integridad de acceso de esta estructura de datos compartidos por dichos subprocesos concurrentes mediante el uso de un bloqueo compartido.

[0088] El subproceso (705) intenta planificar (713) su solicitud de espacio aislado en un espacio aislado y si tiene éxito (714), después se bloquea (740) en el nuevo espacio aislado y luego se registra (741) la solicitud en la estructura de asignación del controlador en la base de datos (130), indexada por el identificador de solicitud (510, 250) para dicha solicitud de espacio aislado (p. ej., 200) y devoluciones (742).

[0089] Si no hay ningún espacio aislado disponible (720), el subproceso (705) hace que la solicitud de reparto de espacio aislado (510) se quede latente (730) durante un tiempo determinado (715) y luego pasa a volver a intentar la planificación (713). Si es necesario, este proceso se repite hasta que o bien la solicitud de espacio aislado se planifique o bien en circunstancias extraordinarias, hasta que se haya hecho un número de intentos máximo.

[0090] La figura 7A ilustra un distribuidor de ejemplo. Tras recibir un mensaje de solicitud de distribuidor (743) del controlador, el distribuidor actualiza (745) el estado de progreso (748) de esta solicitud de espacio aislado. A continuación, inicia un temporizador (746) para controlar el tiempo máximo asignado a la ejecución de la solicitud de espacio aislado. A continuación, inicia (747) una configuración de virtualización que ejecutará la solicitud de espacio aislado en un espacio aislado especificado y espera (749, 752) un evento de finalización de la solicitud (751) que indica que se ha completado con éxito la ejecución en el espacio aislado asignado. De manera alternativa, es posible que se detenga la ejecución del espacio aislado (750). A continuación, almacena (743) los archivos de registro estáticos y de comportamiento obtenidos de la ejecución de la solicitud de espacio aislado en la tabla de registros (754) de la base de datos. Además, calcula (757) y almacena (758) una métrica

relacionada con la calidad o validez de estos archivos de registro para evaluar la idoneidad de uso asociada con la ejecución de la solicitud de espacio aislado. Luego, aplica análisis de procesamiento posterior como por ejemplo, pero sin carácter limitativo, un análisis de red (769), generación de vectores de características, etc. y luego rellena (760) la base de datos con los resultados. Por último, actualiza el estado de progreso de esta solicitud de espacio aislado e informa al controlador de la finalización de la solicitud de espacio aislado mediante el mensaje (763) de finalización de la solicitud.

[0091] La figura 7B ilustra una API espacio aislado de ejemplo. Al recibir una solicitud de espacio aislado (759), se inicia el espacio aislado (760). En primer lugar, realiza etapas de inicialización como la recuperación de un emulador adecuado para el dispositivo móvil y el inicio del dispositivo emulado. Luego, la aplicación se recupera (762) de la base de datos (130) y luego se instala en el dispositivo (763). Luego se instala (764) y se inicia (765) la instrumentación. Luego, se inicia la aplicación (770) y se ejecuta su interfaz de usuario (771). Después de completar la ejecución de la interfaz de usuario de la aplicación, se finaliza la aplicación (772), se detiene la instrumentación (773) y se recogen los archivos de registro resultantes (778).

[0092] La especificación de *software* mencionada anteriormente para la API del espacio aislado puede satisfacerse mediante diferentes componentes de complementos, como por ejemplo, pero sin carácter limitativo, una emulación de *software* virtualizada de un dispositivo móvil, un dispositivo móvil físico, una combinación de los mismos.

[0093] De acuerdo con un modo de realización, dicha ejecución (771) de la interfaz de usuario (UI) de una aplicación está compuesta de etapas tales como, sin limitarse a:

- identificar el conjunto de elementos de la interfaz de usuario presentes en una ventana de actividad de la interfaz de usuario,
- seleccionar, según algunos criterios, un elemento de interfaz de usuario de dicho conjunto,
- interactuar con dicho elemento de interfaz de usuario,
- descubrir si dicha interacción con un elemento de interfaz de usuario da lugar a un cambio en la ventana de actividad de la interfaz de usuario,
- actualizar una matriz de visitas de actividad con transiciones *descubiertas* entre una ventana de actividad de interfaz de usuario a la misma o una ventana de actividad de interfaz de usuario diferente a través de dicha interacción con un elemento de interfaz de usuario,
- iterar todos los elementos de la interfaz de usuario de una ventana de actividad de la interfaz de usuario, y/o
- iterar toda la ventana de actividad de la interfaz de usuario descubierta.

[0094] Los resúmenes de la matriz de visitas de la actividad observaron transiciones de interfaz de usuario de la aplicación determinada mediante tuplas del tipo:

[DesdeVentanaDeActividad, Interacción(Elemento) → HastaVentanaDeActividad], que se corresponde con el estado actual (es decir, DesdeVentanaDeActividad), transición (es decir, debido a la Interacción(Elemento)), un estado siguiente (es decir, HastaVentanaDeActividad), respectivamente de la máquina de estados finitos descubierta al analizar la interfaz de usuario del binario de aplicación.

[0095] También es una característica de la presente invención que los binarios de aplicación que no se han comportado adecuadamente en recorridos de GUI autónomos se marquen para su posterior envío a una cola de operador manual. Estos binarios de aplicación serán ejecutados después por un humano según los pasos del procedimiento definidos en un entorno de espacio aislado controlado de manera manual. La presente invención prescribe que dicha indicación de interacciones de interfaz de usuario fallidas sean detectadas por medio de un análisis detallado basado en agregados del perfil establecido de la característica de recorrido de interfaz de usuario (UI, por sus siglas en inglés) resultante. El perfil de recorrido de UI contiene características como por ejemplo, pero sin carácter limitativo, el número total de actividades encontradas, el número total de elementos de UI descubiertos, el número total de elementos de UI con los que se interactúa, el número total de reinicios de recorrido requeridos, etc.

[0096] La presente invención proporciona medios para evaluar el rendimiento histórico al examinar con detalle los archivos de registro de instrumentación producidos por todas las simulaciones de espacio aislado dentro de un intervalo de tiempo arbitrario en un distribuidor. La presente invención planifica de manera autónoma un análisis de la evaluación del distribuidor que examina la presencia de un conjunto conocido de marcadores de progreso instrumentales y válidos de los registros de instrumentación del espacio aislado. De este modo, el sistema es capaz de producir evaluaciones de si en un intervalo de tiempo arbitrario, las simulaciones de espacio aislado parecían completarse con éxito o de lo contrario, cuántas no han podido llevar a cabo ni detallar qué secuencias de fallo se observaron y en qué etapa se observaron los fallos. De este modo, cuando un conjunto de

binarios de aplicación conocidos se envía de manera periódica, la presente invención proporciona medios para comprobar la regresión autónoma.

5 [0097] La presente invención mantiene una variedad de tablas de resultados de análisis almacenadas en caché utilizadas para agilizar el tiempo de respuesta del servicio en la nube a las solicitudes. Para agilizar la respuesta, es una característica del presente sistema que todas estas tablas se indexen tanto por el identificador de solicitud único como por los MD5 asociados para el binario de aplicación subyacente. Una lista de ejemplo de estas tablas comprende lo siguiente pero sin carácter limitativo:

- una tabla de conexiones de red que detalla todas las transacciones de red,
- 10 • una tabla de marcas rojas que detalla todos los resultados con marcas rojas identificados de análisis estáticos inferenciales,
- una tabla de objetos transferidos por la red que detalla todos los objetos transferidos por la red identificados,
- varias tablas basadas en eventos que detallan todos los eventos a nivel de sistema operativo invitado (p. ej., AndroidOS) registrados durante la ejecución,
- 15 • una tabla de detección de intrusiones que detalla todas las alertas observadas basadas en la red para filtraciones, páginas maliciosas, *malware*, etc.,
- una tabla de capturas de pantalla que detalla todas las capturas de pantalla extraídas durante la ejecución,
- 20 • una tabla de evaluación de *malware* que detalla los detalles de evaluación y confianza sobre un binario de aplicación,
- una tabla de conclusiones que detalla las conclusiones expuestas sobre un binario de aplicación,
- una tabla de validación de registros que detalla información sobre la validez de registros extraídos,
- una tabla de recorridos de interfaz de usuario que detalla la información sobre las conclusiones del recorrido y la cobertura observada durante la ejecución, y
- 25 • una tabla de integridad de archivo que detalla la información sobre los cambios en el sistema de archivo observados durante la ejecución.

[0098] La presente invención mantiene una variedad de tablas de apoyo utilizadas para apoyar las operaciones del servicio en la nube. Una lista de ejemplo de estas tablas comprende lo siguiente pero sin carácter limitativo:

- 30 • una tabla de solicitudes en cola que detalla una versión persistente de las solicitudes de espacio aislado entrantes y que se utiliza para (1) ralentizar la velocidad de llegada entrante al servidor web de las solicitudes de espacio aislado a una velocidad de QoS fija en el controlador y (2) permitir reinicios del servicio en nube con la pérdida de las solicitudes de espacio aislado pendientes,
- una tabla de tareas sin conexión que detalla todos los análisis que han de realizarse de forma forense después de la extracción satisfactoria de los registros de ejecución de una solicitud de espacio aislado, y que se utiliza para (1) desencadenar el análisis especificado contra los registros de ejecución asociados con el identificador de solicitud único especificado,
- 35 • una tabla de solicitudes asignadas que documenta los detalles asociados a qué solicitudes del entorno de pruebas se almacenaron en caché o no en lo que respecta a una asignación entre identificadores de solicitud únicos,
- 40 • una tabla de solicitudes de prórroga que documenta todas las solicitudes de análisis especializadas (por ejemplo, evaluación de riesgos de marcas rojas, análisis de redes, clasificación de *malware*) colocadas en el servicio en nube y utilizadas para (1) iniciar y documentar una solicitud de prórroga de espacio aislado para cada una de ellas,
- una tabla de claves de paso del informe que asigna una clave de paso a cada identificador de solicitud único y que se utiliza para (1) limitar y validar el acceso a la red para el informe de análisis obtenido con un identificador único de solicitud,
- 45 • una tabla de tiendas de aplicaciones que detalla todas las claves de tienda de aplicaciones para cada tienda de aplicaciones y sus estadísticas de uso actuales, y
- 50 • una tabla de solicitudes de tienda de aplicaciones que detalla la clave de tienda de aplicaciones suministrada en cada solicitud de espacio aislado y que se utiliza para (1) realizar proyecciones analíticas de consola sobre datos con respecto a tiendas de aplicaciones.

5 **[0099]** La presente invención proporciona medios para mantener el total del crédito y el uso por tienda de aplicaciones. En un modo de realización, una vez que se recibe una solicitud de servicio en la nube (por ejemplo, una solicitud de espacio aislado), el crédito total para la tienda de aplicaciones correspondiente asociado a la clave de tienda de aplicaciones suministrada se actualiza para reflejar el coste en créditos de la solicitud. La invención también proporciona medios para apoyar de forma gratuita los envíos al servicio en nube; una clave pública de tienda de aplicaciones se rellena de forma predeterminada y se asocia a los envíos que no se pagan. En un modo de realización previsto, el total del crédito asociado se repone periódicamente en función de algunos criterios establecidos, como por ejemplo, pero sin carácter limitativo, el número máximo de envíos por hora, día y/o tipo de API.

10 **[0100]** La presente invención mantiene una variedad de reglas y tablas de apoyo utilizadas para apoyar las operaciones por parte del conjunto de programas de análisis. Una lista de ejemplo de estas tablas (y usos) comprende lo siguiente pero sin carácter limitativo:

- 15 • una tabla de servidores de anuncios que enumera páginas de servidores de anuncios conocidos comercialmente y que se utiliza para (1) detallar qué contenido y/o transacciones de tráfico de red llegaron a los servidores de anuncios conocidos,
- una tabla de páginas maliciosas que enumera páginas maliciosas y se utiliza para (1) detallar qué contenido y/o transacciones de tráfico de red alcanzaron las páginas maliciosas conocidas por nombre,
- una tabla de IP maliciosas que enumera las direcciones de Internet maliciosas y se utiliza para (1) detallar qué contenido y/o transacciones de tráfico de red llegaron a sitios maliciosos conocidos por dirección IP,
- 20 • una tabla de aplicaciones por defecto que enumera el conjunto de binarios de aplicación que se encuentran instalados por defecto en un dispositivo y que se utilizan para (1) evitar el reprocesamiento de dichas aplicaciones por defecto,
- una tabla de metadatos que enumera todos los metadatos proporcionados por la tienda de aplicaciones para los binarios de aplicación y que se utiliza para (1) ampliar los informes de análisis con dichos metadatos relevantes para dichos binarios de aplicación,
- 25 • unas reglas de intrusión que enumeran reglas de detección de intrusos y se utiliza para (1) configurar la aplicación forense de análisis de contenido de detección de intrusos sobre los registros de tráfico de la red que analizan contenido, como por ejemplo, pero sin carácter limitativo, datos colocados previamente, correos electrónicos, contraseñas, cuentas, nombres de usuario, *tokens*, mensajes SMS, identificadores, números de teléfono, direcciones y páginas web y direcciones maliciosas y firmas de *malware*, y
- 30 • unas reglas de marcas rojas que enumeran las marcas rojas de evaluación de riesgos, utilizadas para (1) configurar la aplicación de análisis estático de evaluación de riesgos inferenciales sobre el binario de la aplicación, utilizado para inferir riesgos a través de (pero sin limitarse a) llamadas API, *tokens*, adyacencia de llamadas API y/o *tokens*, y el nivel de ofuscación.

35 **[0101]** En un modo de realización previsto, se permitirá que una tienda de aplicaciones suministre metadatos personalizados para rellenar la tabla de metadatos mencionada anteriormente. Al esta tabla con MD5 así como la clave de la tienda de aplicaciones, ahora será posible mejorar los informes de análisis generados por esta tienda de aplicaciones con datos, como, pero sin carácter limitativo: (1) la marca de la tienda de aplicaciones, (2) los metadatos de la tienda de aplicaciones seleccionados para el binario de aplicación (p. ej., número de descargas, estructura de costes, información del desarrollador), y (3) retroalimentación seleccionada, información, reseñas y recomendaciones del contenido de la tienda de aplicaciones.

40 **[0102]** La figura 10 ilustra un informe de usuario final (1000) de ejemplo previsto para el consumidor y puesto a disposición de un consumidor mediante un localizador uniforme de recursos (URL, por sus siglas en inglés) enviado por correo electrónico por parte del servicio en la nube al correo electrónico asociado con la solicitud de espacio aislado correspondiente. El informe consiste en un área de encabezado (1050) común y una serie de secciones de informe (p. ej., 1010, 1020, 1030, 1040) que contienen los resultados del análisis para cada análisis forense especializado (p. ej., análisis estático, análisis de red, análisis del sistema de archivos, etc.). Se proporciona un acceso a cada una de estas secciones mediante una barra de navegación de contenidos (1005) que enumera un enlace con el nombre de cada uno de los análisis implementados. Los contenidos de la sección del informe seleccionada actualmente se presenta en el área (1006) debajo de la barra de navegación de contenidos. A modo de ejemplo, también se muestran ejemplos de la apariencia del contenido de estas secciones: análisis estático (1010), análisis transversal de la interfaz de usuario (1020), análisis de red (1030) y análisis del sistema de archivos (1040). La invención proporciona medios para personalizar el contenido del informe seleccionando los enlaces de acceso que se mostrarán en la barra de navegación de contenidos (1005) de un informe en función de algunos criterios como la fuente de envío y/o la clave de la tienda de aplicaciones. Un área de encabezado (1050) es común a todas las secciones del informe y se utiliza para enfatizar los atributos básicos y las conclusiones resumidas sobre el análisis del binario de la aplicación. El encabezado tiene

dos secciones, una sección de estadísticas basada en texto (1050) y una sección gráfica de clasificación de *malware* (1060).

- 5 • La sección de estadísticas del encabezado (1050) comprende atributos determinados como por ejemplo, pero sin carácter limitativo: un número de consulta único, la fecha en que se realizó el análisis, el nombre de archivo suministrado para el binario de la aplicación, el nombre del paquete identificado del binario de la aplicación, el dispositivo emulado de destino utilizado para evaluar el binario de la aplicación, el sistema operativo invitado de destino utilizado para evaluar el binario de la aplicación, el MD5 asociado con el binario de la aplicación, así como los atributos derivados tales como la etiqueta de *malware* asociada por el análisis del binario de la aplicación, la etiqueta de aprendizaje automático asociada al binario de la aplicación, y la validez del registro y/o la calidad asociadas con los registros extraídos después de la ejecución del binario de la aplicación. En un modo de realización previsto, esta sección mostrará una indicación de calidad relativa del análisis del recorrido de la interfaz de usuario autónoma.
- 10
- 15 • La sección de clasificación gráfica (1060) contiene una indicación visual de la confianza del análisis de que el *malware* es sospechoso de ser malicioso. La presente invención utiliza una escala numérica (como por ejemplo, pero sin carácter limitativo, que evalúa de 0 a 10) para evaluar el aumento en confianza. Por ejemplo, una evaluación de *malware* de 0 indica que el análisis no percibió motivos de preocupación, mientras que una evaluación de *malware* de 10 indica que el binario de aplicación exhibe un comportamiento malicioso conocido o derivado de uno conocido y por tanto las evaluaciones de *malware* que hay en medio indican una acumulación de motivos de preocupación. En un modo de realización, este indicador visual se mejora con información adicional que pretende subrayar la naturaleza de la exposición al riesgo asociada con el binario de aplicación (p. ej., exposición a la red, filtraciones de identidad/privacidad, etc.).
- 20

[0103] El informe consiste en varias secciones distintas como por ejemplo, pero sin carácter limitativo:

- 25 • una sección de conceptos básicos (1000), que comprende resultados principales esenciales y resumidos sobre el análisis del binario de aplicación, metadatos opcionales (públicos) sobre el binario de aplicación, una captura de pantalla inicial al iniciar el binario de aplicación, y un resumen y detalles de la cronología integrada de eventos obtenida a partir de la correlación basada en el tiempo de los registros del sistema operativo invitado (como por ejemplo, pero sin carácter limitativo, el registro de eventos, el registro de actividades, el registro de conexiones de red, el registro de detecciones de intrusos, etc.);
- 30 • una sección de marcas rojas (1010), que comprende un perfil de evaluación del riesgo comparativo para el binario de aplicación que documenta el riesgo por área de interés del usuario final (a lo que se hace referencia como una categoría de marcas rojas), resúmenes y detalles de activaciones de reglas y categorías de marcas rojas, resúmenes y detalles gráficos del contexto que rodea la activación de cada regla de marcas rojas, y la identificación de un binario de aplicación que exhibe una evaluación del riesgo similar.
- 35 • una sección de autómata de interfaz de usuario (1020), que comprende estadísticas comparativas sobre las ventanas de descubrimiento, los elementos de UI, los estímulos aplicados a estos, y los recorridos resultantes descubiertos por el autómata de UI de conformidad con hasta tres técnicas de recorrido diferentes;
- 40 • una sección de red (1030), que comprende análisis de red y perfiles comparativos basados en agregados para el tráfico de paquetes, resolución de nombres de dominio (DNS, por sus siglas en inglés) para direcciones de internet y análisis por GeoIP del acceso a la red, análisis de detección de intrusos, análisis de archivos transferidos por red, análisis de conexión TCP, análisis de conexión UDP, y análisis de las transacciones HTTP;
- 45 • una sección de sistema de archivos (1040), que comprende un perfil comparativo basado en agregados y resultados del análisis de la integridad del sistema de archivos que documenta archivos sin cambios, añadidos, renombrados, eliminados y modificados durante la ejecución del binario de aplicación.
- 50 • una sección de antivirus (1050), que comprende resultados del análisis AV y resultados del análisis de aprendizaje automático sobre características estáticas y de comportamiento; entre las secciones mostradas en la Fig. 10. Otras secciones que no se muestran en la Fig. 10 son las siguientes, pero sin carácter limitativo:
- 55 • visualización de redes sociales para la conectividad de red con un etiquetado de servidores de anuncios, de páginas maliciosas, países, subredes y binarios de aplicación infectados.
- información de análisis estático básica como, pero sin carácter limitativo, permisos declarados, archivos, métodos, constantes, cadenas, intenciones declaradas, servicios, archivo de instrumentación del espacio aislado;

- análisis del rendimiento basado en los subprocesos, memoria y CPU y perfil comparativo basado en agregados, y
- resumen estadístico para llamadas del sistema dentro del sistema operativo principal y perfil comparativo basado en agregados asociado.

5 **[0104]** El modo de realización proporciona soporte para un análisis detallado basado en agregados para la mayoría de componentes (por ejemplo, características de red, características de integridad del sistema de archivos, características de llamadas de API del sistema, características de rendimiento, características del recorrido de la interfaz de usuario, etc.) del conjunto de programas de análisis. Para llevar a cabo un análisis detallado basado en agregados para un conjunto de características concreto de un binario de aplicación

10 determinado, el sistema compone un perfil compuesto de dos o más características de interés en dicho conjunto de características. El sistema almacena cada uno de estos perfiles calculados, dando como resultado una recogida acumulativa de dichos perfiles. Cuando el análisis de detección de similitudes y/o anomalías se ha de generar para un análisis determinado de un binario de aplicación, primero se normaliza el perfil asociado (por ejemplo, con respecto a las estadísticas asociadas con la recogida actual de dichos perfiles). Luego, los análisis de similitudes y/o anomalías se aplican entre el perfil normalizado y un conjunto de perfiles de una base de datos

15 de perfiles asociados (p. ej., perfiles de red, perfiles de recorrido de interfaz de usuario, etc.). En un modo de realización, este conjunto de perfiles puede comprender el conjunto completo de perfiles de la base de datos o un subconjunto seleccionado por algunos criterios de muestreo, como por ejemplo, sin carácter limitativo, los asociados a la misma clave de tienda de aplicaciones. En la presente invención, la detección de anomalías se basa en una comparación de puntuaciones z actuales para el perfil frente a las puntuaciones z para los miembros del conjunto de perfiles seleccionados de la base de datos. Un lector experto en la materia puede apreciar que pueden aplicarse fácilmente técnicas de clasificación más avanzadas a este modelo. En el sistema actual, la detección de similitud se basa en (pero no se limita a) el agrupamiento difuso mediante vectores de distancia euclidianos de puntuaciones z actuales para el perfil frente a las puntuaciones z para los miembros del

20 conjunto de perfiles seleccionados de la base de datos. Un lector experto en la materia puede apreciar que pueden aplicarse fácilmente técnicas de similitud más refinadas a este modelo. Este aspecto de la presente invención permite identificar y magnificar la presencia de una anomalía aislada en el comportamiento de un binario de aplicación determinado que de lo contrario podría ser suavizado durante la agregación de múltiples conjuntos de características. Por ejemplo, a través de este análisis detallado basado en agregados es posible identificar si un binario de aplicación impone una carga de red inusual, una carga de red inusual basada en servidores de anuncios, un alcance de GeolIP inusual, un comportamiento de interfaz de usuario inusual, un rendimiento de CPU inusual en comparación con decenas de miles o más de otros binarios de aplicación. Este aspecto de la presente invención atrae intereses comerciales que no se limitan a las preocupaciones de análisis de *malware*, sino que extiende la conveniencia de la invención actual a la supervisión, vigilancia y modificación de las capacidades de la web por parte de los administradores.

35

[0105] Es una característica del modo de realización que se realice un análisis sobre uno o más archivo(s) de registro. En la presente invención, el análisis comprende elementos como, pero sin carácter limitativo:

- un párrafo orientado al usuario final de su objetivo y la naturaleza de los resultados importantes (p. ej., «Los datos del sistema minan tu binario de aplicación frente a otros miles para evaluar mejor el riesgo asociado con la instalación y ejecución de la aplicación. La siguiente gráfica indica cómo se compara el perfil de riesgo para tu binario de aplicación frente al de otros miles»).
 - un sumario y/o resumen estadístico de características seleccionadas extraídas de todos los contenidos del archivo(s) de registro específico(s). Por ejemplo, el análisis de red genera un resumen que comprende elementos como, pero sin carácter limitativo: una cantidad de tráfico de red consumido por servidores de anuncios, un número de páginas web maliciosas visitadas, la presencia y el número de filtraciones de red detectadas de datos infiltrados, la presencia y el número de firmas de *malware* detectadas (por ejemplo, referencias a páginas de comando y control, comandos, firmas, etc.), la presencia y el número de datos de identidad/privacidad transmitidos a servidores que no son de anuncios, la distribución geográfica del alcance de la red en cuanto a países y direcciones de Internet, número y estado de la infección para tipos identificados de objetos transferidos por red, etc.
 - un análisis basado en agregados como, pero sin carácter limitativo, una selección de un perfil para características de archivo de registro seleccionadas y evaluar este perfil frente al conjunto de otros perfiles recogidos de manera similar.
 - una presentación orientada al usuario final de los contenidos del archivo(s) de registro específico(s) que comprenden, pero sin carácter limitativo, la abstracción de contenido como la agrupación de eventos relacionados (p. ej., representación abreviada de la actividad observada, el servicio y/o las secuencias de transición de eventos), selección y/o filtrado de contenido en función de algunos criterios como la importancia y/o la prioridad, anotación y/o resaltado del contenido en función de criterios como la relevancia para la evaluación de directivas maliciosas, sospechosas, arriesgadas, y/o que ponen en
- 40
- 45
- 50
- 55
- 60

- una documentación de manera autónoma de los resultados de alto interés del usuario final en función de, pero sin limitarse a, la malicia, la peligrosidad, la desconfianza, la anomalía de los resultados derivados de cualquiera de los elementos de análisis anteriores.

5 Además, cada análisis está asociado con un extractor de características, destinado a identificar y detallar las características numéricas seleccionadas de cualquiera de los elementos de análisis anteriores para su posterior uso en la aplicación de los métodos de clasificación de aprendizaje automático.

10 **[0106]** Por ejemplo, el análisis de marcas rojas proporciona un resumen y detalles para reglas de evaluación de riesgo activadas así como categorías de evaluación de riesgo según se obtienen durante el análisis estático del binario de aplicación, evaluación basada en agregados del perfil de evaluación del riesgo resultante frente a una selección de otros perfiles para identificar anomalías en la exposición al riesgo, contexto del documento gráfico y/o de texto anotado y destacado que rodea la activación inferencial de cada regla de evaluación de riesgo. El análisis también genera una serie de resultados como, pero sin carácter limitativo, anomalías clasificadas basadas en agregados encontradas durante la clasificación del perfil de evaluación de riesgo para el binario de aplicación y la presencia de riesgos muy peligrosos como el *rooteo* del dispositivo y las infecciones de *malware*. En la Fig. 10 se muestra una vista parcial de los resultados (1010) de este análisis.

15 **[0107]** En concreto, tal como se muestra en la FIG. 7B, el administrador de la GUI (771, 780) implementa recorridos similares a los de profundidad (ver 785, 794) recursivos para permitir el descubrimiento autónomo de la interfaz de usuario subyacente de una aplicación arbitraria (por ejemplo, 40). En su núcleo interno, consiste en dos etapas básicas de consulta del estado actual de la GUI (791) y luego elegir un método de interacción apropiado (790) para la pantalla de GUI determinada. De esta manera, se realiza un recorrido autónomo y sin supervisión de la mayoría (si no de todas) las ventanas de la GUI (793) y sus elementos de UI constitutivos (787). Los bucles infinitos se evitan mediante el uso de una estructura de datos visitada (792b). La recursión finaliza cuando no se pueden descubrir más elementos nuevos de ventanas y de la interfaz de usuario.

20 **[0108]** De acuerdo con un modo de realización, la interacción con un elemento de interfaz de usuario se basa en métodos, heurística y/o procedimientos en función del tipo de objeto subyacente (por ejemplo, botón, área de texto, panel, tabulador, lienzo, vista web, casilla de verificación, etiqueta).

25 **[0109]** De acuerdo con un modo de realización, los mecanismos de sincronización se utilizan para lograr dos objetivos. En primer lugar, si el recorrido de la interfaz gráfica de usuario lleva más tiempo del máximo asignado para que una solicitud se ejecute dentro de un espacio aislado, la solicitud de espacio aislado concluye sin problemas. En segundo lugar, si el recorrido de la interfaz gráfica de usuario requiere una fracción de la cantidad máxima de tiempo asignada para que se ejecute una solicitud de espacio aislado, el espacio aislado reinicia los recorridos de la interfaz gráfica de usuario, pero esta vez, primero con recorridos aleatorios y luego si se aprovecha el tiempo, con la incorporación de métodos de interacción de la interfaz de usuario más complejos basados en la heurística sobre objetos y ventanas de la interfaz de usuario. Este mecanismo permite que el espacio aislado interactúe con las aplicaciones que no siguen las plantillas de interfaz de usuario y los elementos de interacción tradicionales; como por ejemplo, en el caso de las aplicaciones de juegos basadas en la visualización de elementos gráficos vectoriales mapeados en un lienzo.

30 **[0110]** Según un modo de realización, dicho ejercicio de la interfaz de usuario de una aplicación se realiza de forma autónoma, es decir, *sin* intervención humana.

35 **[0111]** El sistema actual prevé medios para identificar de forma autónoma a través de algunos criterios, tales como, pero no limitados a, la validez de los archivos de registro resultantes, si una solicitud de espacio aislado debe considerarse como una excepción de espacio aislado y esta también debe ser enviada a un espacio aislado que permita la interacción humana con el fin de mejorar la calidad o la validez de los archivos de registro resultantes.

40 **[0112]** Las figuras 8A a 8F ilustran ejemplos de las tablas mantenidas por la base de datos (130) a fin de respaldar el modo de realización tal como se ha descrito hasta ahora en las figuras 1 a 7B. Ilustra seis tablas, la tabla de solicitudes (800), la tabla de aplicaciones (825), la tabla de asignación de controlador (850), la tabla de registros (875), la tabla de análisis AV (890) y la tabla de progreso (845).

45 **[0113]** La figura 8A ilustra un ejemplo de tabla de solicitudes (800) que almacena datos de identidad previos a la solicitud (que comprenden datos de identidad del dispositivo (210, 805), los datos de identidad del usuario (220, 810), los datos de identidad de la aplicación (230, 815), y los datos de identidad de red (240, 820)) tal como los proporciona una solicitud de espacio aislado (200) en el servicio en la nube (20), con la excepción de la aplicación (40). Cada registro es indexado (801) por el identificador de solicitud (250) correspondiente y se introduce la marca de tiempo (802) de su registro. Por ejemplo, se muestra que el identificador de solicitud 102 consiste en el envío de un binario de *TankHero.apk* desde un dispositivo *Droid2* situado en la IP 65.30.0.1, e indica que los resultados del análisis deberían mandarse por correo electrónico a *X@ABC.NET*.

50 **[0114]** La figura 8B ilustra una tabla de aplicaciones de ejemplo (825) que contiene una entrada para cada solicitud de espacio aislado (200) que almacena la aplicación (40) que ha de analizarse y su MD5/SHA1 (830) calculado junto con su nombre de archivo (835) y un nombre de paquete (840). La entrada está indexada (826)

por el identificador de solicitud (250) correspondiente. Por ejemplo, muestra que el identificador de solicitud 104 consiste en el envío de un binario de *cellfire.apk*, cuyo nombre formal de paquete de Java es *com.cellfire.android*, con algunos MD5 determinados correspondientes al código binario determinado que empieza con *0x0234...*

5 **[0115]** La figura 8C ilustra un ejemplo de tabla de asignación de controlador (850) que contiene una entrada para cada solicitud de espacio aislado (200) que almacena la asignación de dicha solicitud en un espacio aislado (p. ej., 321, 855) en algún nodo distribuidor (p. ej., 170, 860). La entrada está indexada (851) por el identificador de solicitud (250) correspondiente y se introduce la marca de tiempo (852) del registro. Por ejemplo, muestra que el identificador de solicitud 105 se asignó al nodo distribuidor llamado *dirac.tti* el 8 de agosto utilizando un dispositivo *Droid1* con la versión 2.2 del sistema operativo Android.

10 **[0116]** La figura 8D ilustra un ejemplo de tabla de registros (875) que contiene una entrada para cada archivo de registro (p. ej., 410, 420, 430, 440, etc.) producida por un espacio aislado (p. ej., 321) durante el análisis de la solicitud de espacio aislado (p. ej., 200). Cada entrada está indexada (876) por el identificador de solicitud (250) correspondiente y el tipo de registro (880) asociado con el archivo de registro (p. ej., 410). La marca de tiempo (877) del registro también se introduce para cada entrada. Por ejemplo, se muestra que para el identificador de solicitud 104, dos archivos de registro (llamados del sistema de nivel bajo y de red) se han registrado el 8 de agosto a las 12:15.

15 **[0117]** La figura 8E ilustra una tabla de análisis AV (890) de ejemplo que registra el resultado (896) de un análisis AV para una aplicación determinada (p. ej., 40), que resulta en una o más tuplas indexadas por el identificador de solicitud (250) y el nombre del análisis AV (895). Además, se proporciona un campo de descripción detallada (897) y la marca de tiempo (898) para el registro. Por ejemplo, muestra que para el identificador de solicitud 125, el análisis AV de marca A completó su análisis el 9 de agosto e informó de que la infección afirma ser *Falsa* y en consecuencia describió el análisis como *No se encontró nada*. Por otra parte, para el identificador de solicitud 124, dos marcas distintas de análisis AV A y B han completado sus análisis afirmando ambas que la aplicación correspondiente está infectada (896) pero mostrando que no están de acuerdo en la descripción de la infección (897).

20 **[0118]** La figura 8F ilustra un ejemplo de tabla de progresos (845) que contiene una entrada para cada ETAPA (846) (p. ej., REQ START, REQ END, FALLO, etc.) de cada componente (847) (p. ej., servidor web, controlador, distribuidor, espacio aislado, servidor AV) alcanzado en el procesamiento de un identificador de solicitud determinado (p. ej., 250) durante el análisis de una solicitud de espacio aislado (p. ej., 200). La entrada está indexada por el identificador de solicitud (848) correspondiente y se introduce una marca de tiempo de registro (848) durante el registro.

25 **[0119]** La figura 9 ilustra un ejemplo de informe de desarrollador (900) generado después de que se analicen los archivos de registro para un espacio aislado. El informe es accesible mediante el identificador de solicitud (250), que está etiquetado en el informe (915). El informe consiste en varias secciones distintas como por ejemplo, pero sin carácter limitativo: resumen/metadatos de la aplicación (920), resultados del análisis AV (925), demografía de la GeolIP (930), archivo de registro automatizado de GUI (935), análisis de red (940), visualización de red (945), archivo de registro de la instrumentación de espacio aislado (950), actividades de Android (955), eventos de aplicación de Android (960), cronología a nivel de sistema (965), perfil (970), rendimiento de la CPU (975), análisis estático/DEX (980), cambios en el sistema de archivos/MD5 (985), contenido del caché de la web (990), y análisis de agrupaciones en clústeres (995).

30 **[0120]** Según un modo de realización, puede generarse un informe escalonado que incluye, pero sin carácter limitativo, un informe de usuario final simplificado que es generado por una proyección del informe exhaustivo del desarrollador. En concreto, el informe del usuario final es generado a condición de que se pueda acceder a dicho informe desde un dispositivo móvil y de conformidad con las restricciones colocadas por dicho tipo de dispositivos en el tamaño de visualización y/o del archivo.

35 **[0121]** La figura 10 ilustra un informe de usuario final (1000) de ejemplo previsto para el consumidor y puesto a disposición de un consumidor mediante un localizador uniforme de recursos (URL, por sus siglas en inglés) enviado por correo electrónico por parte del servicio en la nube al correo electrónico asociado con la solicitud de espacio aislado correspondiente

40 **[0122]** La figura 11A ilustra un proceso de ejemplo para una ejecución posterior satisfactoria de una solicitud de espacio aislado (200). Muestra que la serie de archivos de registro (p. ej., 410, 420, 430, 440) generados por el espacio aislado (y almacenados en la base de datos (130)) se analizan después para generar características numéricas correspondientes (p. ej., 1115, 1125, 1135, 1145). Después, estas características (p. ej., 1115, 1125, 1135, 1145) se combinan para producir el vector de características (1200) para el identificador de solicitud especificado.

45 **[0123]** La figura 11B ilustra un ejemplo de generación de características para un identificador de solicitud arbitrario (1150, 250). Primero, se recupera el conjunto de tipos de registros (1155) asociado con el identificador de respuesta, luego se calcula la validez del conjunto de archivos de registro (1160), y se inicializa el vector de características. Después, se hace una iteración (1162) frente a cada archivo de registro, donde cada archivo de registro se analiza para extraer características numéricas de este (1167) y luego se actualiza el vector de

características (1168). Este proceso se itera para cada uno de los archivos de registro. Finalmente, cuando se han procesado todos los archivos de registro (1163), se almacena el vector de características actualizado (1169) en la base de datos (130).

5 **[0124]** Según un modo de realización, el presente sistema proporciona medios para evaluar la validez de los resultados de la ejecución de una solicitud de espacio aislado por medio de algunos criterios de evaluación como, pero sin carácter limitativo, el tamaño de los archivos de registro, los contenidos de los archivos de registro resultantes (p. ej., un archivo TCPDUMP válido), la presencia o la ausencia de palabras clave (p. ej., RECORRIDO GUI COMPLETADO).

10 **[0125]** La figura 11C ilustra un proceso de ejemplo para evaluar la validez de los registros producidos por la ejecución de una solicitud de espacio aislado. Por defecto, se asigna un identificador de solicitud determinado (250) a una validez de registro por defecto de L0 (1165), no obstante, si la ejecución de la solicitud de espacio aislado dio como resultado registros con datos válidos y un recorrido satisfactorio de la interfaz GUI, se asigna una validez de registro de L1(1167). Finalmente, si la ejecución dio como resultado un archivo de registro que también ha monitorizado con éxito la actividad de red, se asigna una validez de registro de L2. Estas clasificaciones pueden redefinirse y/o extenderse.

15 **[0126]** La figura 11D ilustra un ejemplo de análisis (1170) de un archivo de registro. Se leen los datos del archivo de registro (1172) y se obtiene un resumen estadístico (1174) del archivo de registro. Las estadísticas de resumen resultante se combinan después (si es necesario) y se transforman (1176) en características numéricas. Por ejemplo, para un archivo de registro de red, se calculan aquí las características numéricas como, pero sin limitarse a, el rendimiento medio entre conexiones salientes, el número de conexiones TCP, UDP, HTTP iniciadas, el número total de bytes enviados y/o recibidos. Finalmente, se devuelven las características numéricas (1178).

20 **[0127]** Según un modo de realización, el resumen estadístico y las características numéricas resultantes también pueden derivarse de la aplicación de técnicas conocidas de procesamiento del lenguaje natural (NLP, por sus siglas en inglés) que incluyen, pero sin carácter limitativo:

- análisis de la frecuencia de términos (TF, por sus siglas en inglés) de conformidad con algunos criterios como la importancia alta (TDIDF),
- proximidad temporal/espacial aproximada en los documentos (PROXIMIDAD),
- secuencias de *tokens* en un documento (NGRAMAS),
- 30 • minería de datos de expresiones habituales (REGEX),
- recuentos de tipos de elementos en un documento (RECUENTOS),
- sumas de valores de elementos en un documento (SUMAS),
- diferencias (DIFS) entre versiones de documentos similares,
- aserciones booleanas (BOOLEANOS) en combinaciones de los anteriores.

35 **[0128]** La lista de ejemplo de aplicaciones de estos métodos de extracción de características consiste, pero sin carácter limitativo: un BOOLEANO puede utilizarse para afirmar si una aplicación tiene metadatos de mercado asociados, puede utilizarse un RECUENTO para identificar el número de páginas maliciosas alcanzadas, una SUMA para totalizar la carga de red del servidor de anuncios impuesta, una PROXIMIDAD para inferir si el binario de aplicación tiene capacidades de *rooteo*, pueden utilizarse NGRAMAS para construir una firma de comportamiento de bajo nivel sobre secuencias de llamadas de sistema del sistema operativo principal, puede utilizarse REGEX para identificar una dirección de protocolo de internet (IP) y el TF puede utilizarse para construir una firma de comportamiento a nivel de aplicación sobre la distribución de llamadas a distintas API de sistema operativo invitado.

40 **[0129]** Es una característica del presente modo de realización que el análisis estadístico también pueda aplicarse al código de bytes Dalvik del sistema operativo de Android del binario de aplicación. Esto tiene el beneficio de que el código de bytes Dalvik puede extraerse fácilmente del binario de aplicación. Es también un aspecto de la invención actual que los análisis estáticos mencionados anteriormente realizados por el sistema no necesiten conciencia de control y datos del código específicos del binario de la aplicación.

45 **[0130]** A estas etapas se les denomina etapas de extracción de características. En la presente invención, un análisis compendia estas etapas de extracción de características y genera un componente único (1250, 1260, etc.) de un vector de características. No obstante, un análisis puede correlacionar múltiples resultados de extracción de características descritos anteriormente de distintos archivos de registro (p. ej., registro de tráfico de red, registro de detección de intrusiones, registro de transacciones HTTPS) en un componente único (como un componente de red) de un vector de características asociado con la ejecución del espacio aislado del binario de aplicación.

[0131] En la presente invención, estos análisis se realizan de forma forense (es decir, tras la ejecución del binario de aplicación en un espacio aislado en un nodo distribuidor). El planificador pone en cola, limita, reparte e inicia un análisis tras la finalización de una simulación de espacio aislado en función de algunos criterios como, pero sin carácter limitativo, la precedencia aplicable y/o la concurrencia entre análisis. Tal como se ha señalado, cada análisis toma uno o más archivos de registro y produce un informe de análisis y un componente de vector de características. Esto posibilita que la presente invención permita la asignación de distintos servidores a distintos análisis, posiblemente configurados con recursos especializados y/o *software* adaptado a la naturaleza de los análisis a fin de que se lleven a cabo dentro de este. Por ejemplo, en el presente modo de realización, los análisis que consumen muchos recursos informáticos como aquellos basados en métodos de clasificación de aprendizaje automático se asignan a servidores internos con gran rendimiento.

[0132] La figura 13 ilustra secciones de ejemplo que constituyen secciones del vector de características (1200) producidas a efectos del análisis de aprendizaje automático. El vector de características está indexado por el identificador de solicitud (1205, 250). Las secciones constituyentes son las siguientes, pero sin carácter limitativo:

- características de resumen de red (1210), como el número de páginas visitadas, el número de bytes enviados, el número de bytes recibidos, el número de conexiones TCP, el número de conexiones UDP, el número de IP distintas visitadas, el número de transacciones HTTP, la carga de tráfico de los anunciantes, etc.;
- características de GeolP (1215), como el número de países alcanzados, número de conexiones por país, número de subredes alcanzadas, etc.;
- características de alerta (de detección de intrusos) de red (1220), como el número de sitios maliciosos alcanzados, el número de firmas de *malware* de red observadas, el número de filtraciones de datos de privacidad, identidad y/o archivos infiltrados, número de transacciones a servidores de anuncios, número de objetos de red descargados;
- características de comportamiento de bajo nivel (es decir, del sistema operativo principal) (1230), como los recuentos totales, la duración media, y la duración total para todos los tipos conocidos de llamadas de sistema operativo (p. ej., escrituras en memoria, escrituras en disco, selección de archivos, esperas de red, etc.);
- características de comportamiento de alto nivel (es decir, del sistema operativo invitado y/o del dispositivo de emulación) (1240), como el número total de tipos diferentes de llamadas de API de AndroidOS observados, el número total de NGRAMAS de tamaño 2 observadas para las llamadas de API de AndroidOS, etc.;
- características de integridad/cambios del sistema de archivos (1250), como el número de archivos modificados, eliminados, añadidos, duplicados, etc.;
- características de rendimiento del sistema operativo invitado, como una desviación media y estándar para varios indicadores de rendimiento (p. ej., la CPU, la memoria, el número de subprocesos) observados durante la ejecución del binario de aplicación;
- características de análisis estáticos (1260), como el tamaño del binario de aplicación, el número de constantes de cadenas, y los resultados de un perfil de evaluación de riesgo inferencial para el binario de aplicación (p. ej., el número de bloques que contienen referencias asociadas con el acceso a *socket*, número de bloques que contienen referencias asociadas con capacidades criptográficas, número de bloques que contienen referencias asociadas con las capacidades de *rooteo* del dispositivo, número de bloques que contienen referencias asociadas con las capacidades de recuperación de identidad, número de bloques que contienen referencias asociadas con el acceso al sistema de archivos, número de bloques que contienen referencias asociadas con el envío de mensajes SMS, número de bloques que contienen referencias asociadas con la interceptación de llamadas telefónicas de voz, número de bloques que contienen referencias asociadas con el soporte para el acceso a FTP, número de bloques que contienen referencias asociadas con la recuperación de coordenadas GPS, número de bloques que indican niveles inusuales de ofuscación del código, etc.);
- características de metadatos de aplicación (1222), como el número de descargas, la puntuación media, etc.; y/o
- la métrica de validez (1270, 1167) calculada (1167) para los archivos de registro correspondientes asociados con dicho identificador de solicitud (1205).

[0133] La presente invención también proporciona medios para la extracción de características sobre características como, pero sin carácter limitativo, la evaluación del alcance y la calidad de los recorridos de GUI, el alcance y la densidad del tráfico de servidores de anuncios en relación con el tráfico de red general, la presencia de infecciones en archivos descargados, interacciones con *scripts* maliciosos del lado del servidor,

carga del tráfico de red situada en países distintos a EE. UU. vs. carga de tráfico situada en EE. UU., indicaciones de anomalías de clasificación en la clasificación de perfiles basados en agregados frente al conjunto de perfiles generales para cualquier conjunto de características concreto (por ejemplo, anomalías en el perfil de evaluación de riesgo, el perfil de rendimiento del sistema, el perfil de integridad del archivo, el perfil de análisis de red, etc.), etc.

[0134] Por ejemplo, puede utilizarse un BOOLEANO para afirmar si una aplicación tiene metadatos de mercado asociados, pueden utilizarse NGRAMAS para construir una firma de comportamiento de bajo nivel, puede utilizarse un REGEX para identificar direcciones de protocolo de internet (IP), y puede utilizarse un TF para construir una firma de comportamiento a nivel de aplicación.

[0135] La figura 12 ilustra un ejemplo de generación de características numéricas al analizar el archivo de registro de red (410, 1110) para extraer características (1115) como, pero sin carácter limitativo, estadísticas resumidas de red (1116) y estadísticas de red detalladas (1117). Las estadísticas de red resumidas se recogen a lo largo de todas las conexiones (p. ej., conexiones TCP (1111, 1112), y conexiones UDP (1113, 1114). Por ejemplo, algunos ejemplos de características consideradas son el número total de conexiones TCP, UDP (1117f, 1117e), el número de IP distintas con las que se interactúa (1117h), el número de transacciones TCP completadas (1117g), el total de bytes enviados y recibidos (1117a, 1117b), el total de paquetes enviados y recibidos (1117c, 1117d), y las medidas derivadas como la velocidad de transferencia media, el retardo medio en las transferencias de ida y vuelta, el número de subredes con las que se interactúa, el número de alertas de IDS/*Snort* generadas, etc. A este respecto, se recogen estadísticas de red detalladas para cada conexión (1116) y algunos ejemplos son números de paquetes enviados y recibidos, el número de bytes enviados y recibidos, el rendimiento en cada dirección, el retardo en cada dirección.

[0136] Por ejemplo, se aplica un análisis conceptualmente similar a los resultados de la aplicación de herramientas de análisis estáticos sobre el código binario de la aplicación enviada (40), tal como se ilustra en la Fig. 12A.

[0137] La figura 12A ilustra un ejemplo de análisis estático automatizado (1300) que es aplicado por el servicio en la nube (20) sobre cualquier aplicación determinada (40). En primer lugar, en (1305), un identificador de solicitud determinado (p. ej., 250) se utiliza para recuperar la aplicación correspondiente (40) de la base de datos (130). Después, utilizando un decompilador, se realiza ingeniería inversa a la aplicación en el código fuente de Java (1310). Después, se aplica una etapa opcional (1315) de embellecimiento de este código para desofuscar el código fuente. A continuación, en las etapas (1320, 1325, 1330) se aplican operaciones de minería de datos relevantes sobre este código base con respecto a las definiciones de función (1320), las constantes de cadenas (1325) y las invocaciones de API (1330). Después, en las etapas (1335, 1340, 1345), los registros producidos por estas etapas de minería de datos se analizan para producir resúmenes estadísticos (similar a la Fig. 12) para cada uno. El presente sistema también realiza análisis del manifiesto de la aplicación (1355) así como con respecto a las características relacionadas con el tamaño (1360) del código fuente base. Finalmente, el conjunto de características calculadas en el análisis estático se adjuntan (1635) y se devuelven (1370). El vector de características resultante (1370, 1200) consiste en una matriz de características numéricas, organizadas normalmente en secciones con respecto al archivo de registro/componente que produjo dichas características, con cada característica numérica expresada como un valor real o un valor integral.

[0138] Según un modo de realización, se llevan a cabo etapas similares para extraer características numéricas similares de otros registros, como archivos de registro de la memoria/CPU (420, 1120, 1125), archivos de registro de eventos API a nivel de dispositivo/emulador (430, 1130, 1135), y un archivo de registro de eventos API a nivel de sistema operativo invitado (440, 1140, 1145).

[0139] En el presente sistema, el objetivo de los métodos de clasificación de aprendizaje automático (a los que se les denomina en el presente documento «Clasificadores») es decidir si un binario de aplicación de muestra se incluye o no en un conjunto de miembros especificado. Un experto en la materia aprecia que una vez que los conjuntos de datos y los conjuntos de entrenamiento se acumulan, se implementarán otras técnicas de clasificación supervisadas más susceptibles para conjuntos de datos más grandes en el servicio en la nube. En concreto, se prevé el uso de máquinas de soporte vectorial y árboles de decisión en función de los vectores de características mencionados anteriormente y/o subconjuntos de sus componentes.

[0140] La presente invención proporciona medios para enfoques de métodos de clasificación múltiple:

- Clasificación por analogía a los binarios de *malware*, que comprenden (pero sin carácter limitativo) las etapas de:

(1) comparar el vector de características de un binario de aplicación de muestra con los vectores de características de un conjunto de binarios de *malware* conocidos;

(2) determinar si existe un alineamiento estrecho adecuado (es decir, una coincidencia) de conformidad con algunos criterios establecidos como la distancia euclidiana mínima entre vectores de característica correspondientes; y

- (3) después, si existe dicha coincidencia, recuperar y crear informes sobre las propiedades de la coincidencia más próxima (como, pero sin carácter limitativo, la distancia euclidiana y las probabilidades asociadas, la identidad del binario de *malware* coincidente, y la naturaleza de su infección, y el número de otras coincidencias próximas en similitud) y en caso contrario, crear informes sobre la falta de dicha coincidencia.
- 5
- Clasificación por analogía a los binarios de aplicaciones no infectadas, que comprenden (pero sin carácter limitativo) las etapas de:
 - (1) seleccionar un subconjunto representativo de binarios de aplicación en función de algunos criterios (como, pero sin carácter limitativo, un conjunto sin filtrar y/o sin reducir, la asociación a una clave de tienda de aplicaciones, la popularidad por total de descargas (p. ej., las aplicaciones más descargadas), el alcance de la funcionalidad (p. ej., mp3, SMS, teléfono, navegación, aplicación de juegos, etc.), y/o propiedades estadísticas (p. ej., agrupar centroides en clústeres y/o el centro de masas);
 - (2) comparar el vector de características de un binario de aplicación de muestra con los vectores de característica de dicho subconjunto de binarios de aplicación;
 - (3) determinar si un alineamiento estrecho adecuado (es decir, una coincidencia) existe de conformidad con algunos criterios establecidos como la distancia euclidiana mínima entre vectores de características correspondientes; y
 - (4) después, si existe dicha coincidencia, recuperar y crear informes sobre las propiedades de la coincidencia más próxima (como, pero sin carácter limitativo, la distancia euclidiana y las probabilidades asociadas, la identidad del binario de aplicación coincidente, y las propiedades o atributos del binario de aplicación como, pero sin carácter limitativo, la evaluación del riesgo, el informe del análisis en caché, etc.) y en caso contrario, crear informes sobre la falta de dicha coincidencia.
- 10
- 15
- 20
- Clasificación por analogía a binarios de aplicación arbitrarios (ya sean de *malware*, sin *malware*, o una combinación de ambos), que comprenden (pero sin carácter limitativo) las siguientes etapas 1, 2, 3 y 4 descritas anteriormente.
- 25
- [0141]** La presente invención también proporciona los ámbitos de dos métodos de clasificación:
- Clasificador de ámbito generalizado, que consiste en un método de clasificación destinado a abordar una amplia población de binarios de aplicación (ya sean *malware*, sin defectos de *malware* conocidos, o sean una combinación de ambos); y
 - Clasificador de ámbito especializado, que consiste en un método de clasificación destinado a abordar una población reducida de binarios de aplicación (ya sean *malware*, sin defectos de *malware* conocidos, o sean una combinación de ambos).
- 30
- [0142]** Según un modo de realización, el presente sistema genera un vector de análisis que proporciona un *proxy* representativo para el análisis y ejecución de una solicitud de espacio aislado. Cabe destacar que estos indicadores numéricos representan características extraídas del archivo de registro correspondiente y para propósitos de aprendizaje automático, estas características se seleccionan y se dice que son características resumidas representativas de datos subyacentes (p. ej., archivos de registro).
- 35
- [0143]** En este sentido, un extractor de características está definido por tanto específicamente para cada componente (p. ej., 1250, 1260, etc.) de un vector de características (1200) y las características numéricas resultantes para todos estos componentes se ensamblan en un vector de características único (1200) que se dice que es representativo de la aplicación correspondiente (40).
- 40
- [0144]** La figura 13 ilustra un ejemplo de secciones constituyentes del vector de características (1200) producidos a efectos del análisis de aprendizaje automático. El vector de características está indexado por el identificador de solicitud (1203, 250). Las secciones constituyentes son las siguientes, pero sin carácter limitativo:
- 45
- características resumidas de red (1210),
 - características de GeolP (1215),
 - características de alerta (intrusión) de red (1220),
 - características de comportamiento de bajo nivel (es decir, sistema operativo invitado) (1230),
 - características de comportamiento de alto nivel (es decir, emulación del dispositivo) (1240),
 - características de integridad/cambios en el sistema de archivos (1250),
- 50

- características de rendimiento (p. ej., CPU, memoria, número de subprocesos) (1255),
- características de análisis estático (1260),
- características de metadatos de la aplicación (1222), y/o
- la métrica de validez (1270, 1167) calculada (1167) para los archivos de registro correspondientes asociados con dicho identificador de solicitud (1205).

[0145] Las técnicas de aprendizaje automático requieren una etapa de entrenamiento con casos de entrenamiento etiquetados de los cuales extraer su respuesta aprendida. Por este motivo, también se proporcionan etiquetas de infección (binarios) antivirus (1280) para ser utilizados para análisis de clasificación automática de entrenamiento mediante técnicas de aprendizaje automático como, pero sin carácter limitativo, máquinas de soporte vectorial (SVM), árboles de decisión, redes bayesianas y agrupaciones en clústeres.

[0146] Un experto en la materia apreciará que debido a la emergencia reciente de aplicaciones móviles, al pequeño número de análisis de *malware* diseccionados por humanos disponible, y a los enfoques de novedad tomados por el *malware* móvil (p. ej., la interceptación de la privacidad mediante SMS, teléfono, GPS, etc.), el número de casos infectados con los que entrenar es significativamente pequeño aunque engañoso, y que un sistema que extraiga de manera robusta, sistemática y anónima vectores de características de aplicaciones es una herramienta extremadamente valiosa.

[0147] Según un modo de realización, el presente sistema permite un análisis de *malware* para dispositivos móviles al proporcionar una plataforma de servicio en la nube que aprende del comportamiento agregado de aplicaciones analizadas. El presente sistema utiliza técnicas de agrupamiento de aprendizaje automático aplicadas sobre vectores de características para ayudar a clasificar aplicaciones, en función de sus vectores de características correspondientes (obtenidos mediante una combinación de análisis estáticos y de comportamiento), en clústeres de comportamiento de aplicaciones bien definidos. Las características se extraen tanto del comportamiento de ejecución observado de una aplicación, como de los análisis estáticos de la aplicación, y estas características se utilizan para representar o redirigir mediante un *proxy* una aplicación concreta (p. ej., 40) mediante su vector de características correspondiente (p. ej., 2000).

[0148] En el presente sistema, el objetivo de los métodos de clasificación de aprendizaje automático (a los que se les denomina en el presente documento «Clasificadores») es decidir si un binario de aplicación de muestra se incluye o no en un conjunto de miembros especificado. Un experto en la materia aprecia que una vez que los conjuntos de datos y los conjuntos de entrenamiento se acumulan, se implementarán otras técnicas de clasificación supervisadas más susceptibles para conjuntos de datos más grandes en el servicio en la nube. En concreto, se prevé el uso de máquinas de soporte vectorial y árboles de decisión en función de los vectores de características mencionados anteriormente y/o subconjuntos de sus componentes.

[0149] La presente invención proporciona medios para enfoques de métodos de clasificación múltiple:

- Clasificación por analogía a los binarios de *malware*, que comprenden (pero sin carácter limitativo) las etapas de:
 - (1) comparar el vector de características de un binario de aplicación de muestra con los vectores de característica de un conjunto de binarios de *malware* conocidos;
 - (2) determinar si existe un alineamiento estrecho adecuado (es decir, una coincidencia) de conformidad con algunos criterios establecidos como la distancia euclidiana mínima entre vectores de característica correspondientes; y
 - (3) después, si existe dicha coincidencia, recuperar y crear informes sobre las propiedades de la coincidencia más próxima (como, pero sin carácter limitativo, la distancia euclidiana y las probabilidades asociadas, la identidad del binario de *malware* coincidente, y la naturaleza de su infección, y el número de otras coincidencias próximas en similitud) y en caso contrario, crear informes sobre la falta de dicha coincidencia.
- Clasificación por analogía a los binarios de aplicaciones no infectadas, que comprenden (pero sin carácter limitativo) las etapas de:
 - (1) seleccionar un subconjunto representativo de binarios de aplicación en función de algunos criterios (como, pero sin carácter limitativo, un conjunto sin filtrar y/o sin reducir, la asociación a una clave de tienda de aplicaciones, la popularidad por total de descargas (p. ej., las aplicaciones más descargadas), el alcance de la funcionalidad (p. ej., mp3, SMS, teléfono, navegación, aplicación de juegos, etc.), y/o propiedades estadísticas (p. ej., agrupar centroides en clústeres y/o el centro de masas);
 - (2) comparar el vector de características de un binario de aplicación de muestra con los vectores de característica de dicho subconjunto de binarios de aplicación;

- (3) determinar si un alineamiento estrecho adecuado (es decir, una coincidencia) existe de conformidad con algunos criterios establecidos como la distancia euclidiana mínima entre vectores de características correspondientes; y
- 5 • (4) después, si existe dicha coincidencia, recuperar y crear informes sobre las propiedades de la coincidencia más próxima (como, pero sin carácter limitativo, la distancia euclidiana y las probabilidades asociadas, la identidad del binario de aplicación coincidente, y las propiedades o atributos del binario de aplicación como, pero sin carácter limitativo, la evaluación del riesgo, el informe del análisis en caché, etc.) y en caso contrario, crear informes sobre la falta de dicha coincidencia.
- 10 • Clasificación por analogía a binarios de aplicación arbitrarios (ya sean de *malware*, sin *malware*, o una combinación de ambos), que comprenden (pero sin carácter limitativo) las siguientes etapas 1, 2, 3 y 4 descritas anteriormente.

[0150] La presente invención también proporciona los ámbitos de dos métodos de clasificación:

- 15 • Clasificador de ámbito generalizado, que consiste en un método de clasificación destinado a abordar una amplia población de binarios de aplicación (ya sean *malware*, sin defectos de *malware* conocidos, o sean una combinación de ambos); y
- Clasificador de ámbito especializado, que consiste en un método de clasificación destinado a abordar una población reducida de binarios de aplicación (ya sean *malware*, sin defectos de *malware* conocidos, o sean una combinación de ambos).

20 **[0151]** La presente invención proporciona medios para la aplicación de métodos de clasificación de múltiples niveles, con clasificadores de ámbito especializado rápidos que son sucesivos hasta que se identifica una coincidencia para el binario de aplicación de muestra y en caso contrario, seguido de la aplicación de uno o más clasificadores de ámbito generalizado ligado a la computación.

25 **[0152]** Un experto en la materia aprecia que dicho diseño de métodos de clasificación de múltiples niveles es apto para un cálculo eficiente en un entorno de servidores de clústeres.

30 **[0153]** Un experto en la materia aprecia que dicho diseño de métodos de clasificación de múltiples niveles es apto para la aplicación de diferentes métodos de clasificación (por ejemplo, máquinas de vectores de apoyo, agrupación, árboles de decisión, etc.) –posiblemente de forma simultánea–, al mismo binario de aplicación de muestra. De este modo, la presente invención permite distintos métodos de clasificación, ya sea con datos de entrenamiento iguales o distintos, respecto al mismo binario de aplicación. En el presente modo de realización, la presencia de una coincidencia se selecciona de los resultados en función de algunos criterios (como la presencia de cualquier coincidencia, un consenso entre uno o más métodos de clasificación, etc.).

35 **[0154]** Un experto en la materia apreciará que dicho diseño de métodos de clasificación de múltiples niveles es apto para un perfeccionamiento progresivo de la determinación del ámbito de la clasificación mediante el desarrollo e incorporación de clasificadores de ámbito generalizados y/o especializados. Por ejemplo, la presente invención permite un desarrollo e incorporación sencillos de clasificadores de ámbito especializado para binarios de *malware* descubiertos relativamente recientes que

- (1) no se hayan identificado como tales,
- (2) posean propiedades de agrupación en clústeres similares pero
- 40 (3) no obstante indiquen anomalías de clasificador con respecto a binarios de aplicación que no son *malware* al entrenar un nuevo clasificador de ámbito limitado con dichas muestras.

45 **[0155]** Un experto en la materia apreciará que es posible que el *malware* sea lo suficientemente inteligente para enmascarar un comportamiento similar a otras aplicaciones y por esta razón, el presente sistema se basa en aspectos de características exhaustivos y extensivos (p. ej., análisis estático (1260), la huella del rendimiento (1255), la firma del sistema operativo (1230), para formular el vector de características para una aplicación de tal manera que aumente las formas y la profundidad de los que observamos simultáneamente en una aplicación. El motivo es que una aplicación de *malware* que imita una aplicación sin *malware* si se mira, debe invocar aun así determinados recursos primitivos y especiales para determinar que está siendo monitorizada y que dichos eventos serían capturados por algunos de los aspectos de características anteriores. Por ejemplo, las primitivas de virtualización ahora pueden estar presentes, los picos de CPU pueden ser notables, y puede estar presente un perfil de sistema de sistema operativo distinto que corresponde a dicha comprobación del proceso.

55 **[0156]** Un experto en la materia aprecia que una vez que los conjuntos de datos y los conjuntos de entrenamiento se acumulan, se implementarán otras técnicas de clasificación sin supervisión en el servicio en la nube. En concreto, se prevé el uso de máquinas de soporte vectorial y árboles de decisión en función de los vectores de características mencionados anteriormente y/o subconjuntos de sus componentes.

[0157] La figura 14 ilustra un ejemplo de agrupación de un conjunto finito de vectores de características (1401, 1402, 1403, 1404, 1405, 1406, etc.) en un conjunto finito de clústeres. Esto se consigue mediante algoritmos de agrupación en clústeres (p. ej., agrupaciones jerárquicas, k-medias) que dan como resultado la identificación de varios clústeres (1420, 1430, 1440, 1450) mostrados como elipses que contienen uno o más vectores de características (1402; 1403, 1405, 1406; 1401, 1404; 1407). También muestra los centroides correspondientes (1421, 1431, 1441, 1451) de estos clústeres (1420, 1430, 1440, 1450) y que un centroide representa el centro del elipsoide que abarca el clúster. También muestra que es posible que un vector de características (p. ej., 1401) esté cerca de más de un elipsoide (1440, 1450). Para esta aplicación, se permite que cada vector de características sea un miembro de exactamente un clúster. Finalmente, ilustra la identificación del vector de características más representativo de un clúster dado que el vector de características en el conjunto de vectores de características de un clúster que tiene la distancia mínima al centroide de su clúster (p. ej., 1405, 1404, 1407); en el caso de que exista la misma distancia (1402, 1406), se elige el vector de características con el identificador de solicitud más bajo (p. ej., 1402). Los resultados de esta etapa se utilizan para buscar y producir una asignación de clústeres que consiste en el número de clústeres utilizados, los centroides de estos clústeres, y el conjunto de miembros de un clúster en cuanto a los vectores de características.

[0158] El presente sistema implementa el paso ligado a la computación descrito anteriormente lejos de los principales flujos de trabajo de ejecuciones de espacio aislado paralelas. Por ejemplo, el presente sistema permite que esta computación se lleve a cabo en alguna otra parte del sistema siempre que se proporcione acceso a la base de datos compartida por la red (130). El modo de realización preferido del presente sistema implementa esta etapa ligada a la computación lejos de los principales flujos de trabajo de ejecuciones de espacio aislado paralelas. Lleva a cabo esta etapa en un nodo adecuado para tareas ligadas a la computación.

[0159] La figura 15 ilustra un ejemplo de conceptualización de vista de tiempo del proceso de agrupación en clústeres en línea de nuevo un flujo de nuevo vectores de características (1501, 1502, 1503, 1504, 1505, 1506) frente a un conjunto de clústeres predefinidos (1420, 1430, 1440) y sus asignaciones correspondientes (1420, 1430, 1420, 1440, 1430) y en caso contrario, la detección de anomalías (1506). El vector de características anómalo (1506) exhibe una distancia a los centroides (1520, 1530, 1540) de los clústeres predefinidos correspondientes (1420, 1430, 1440) que no exhiben importancia estadística para los miembros de los mismos. El vector de características resultante se actualiza como una anomalía en la base de datos (130) (tal como se describe en la Fig. 17 y la Fig. 18), indexado de nuevo por el identificador de solicitud correspondiente (250).

[0160] La figura 16A ilustra un clúster pequeño de ejemplo de vectores de características en un espacio bidimensional (mediante agrupamientos de K-medias). La figura 16A muestra cinco vectores de características f1 (, f2, f3, f4, f5 (1601, 1602, 1603, 1604, 1605, respectivamente) y cómo cuatro de ellos pueden visualizarse en un clúster (1610) cuyo centroide (1600) puede visualizarse como el centro de la elipse del clúster (1610). Muestra que el vector de características f5 (1605) no se sitúa lo suficientemente cerca de la elipse para afirmar de manera confidente que forme parte del clúster (1610).

[0161] La figura 16B ilustra un ejemplo de representación alternativa (mediante una agrupación en clústeres jerárquica) de un clúster en dos dimensiones. El mismo clúster (1610b) se mostró con los mismos vectores de características (f1, f2, f3, f4) en la Fig. 16A como el clúster al que se señala con (1610).

[0162] La figura 16C ilustra un ejemplo de matriz de distancia correspondiente (1620) y cómo proporciona la base para generar clústeres. La matriz de distancia (1620) está compuesta por la medida de la distancia calculada (p. ej., distancia euclidiana) entre dos pares cualquiera de vectores de características del conjunto de datos. En esta ilustración, el conjunto consiste en cinco vectores de características (f1, f2, f3, f5, f5) y la disposición de los clústeres resultante (1630, 1631) indica que cuatro de ellos tienen cortas distancias por pares entre sí (f1, f2, f3, f4) y que no puede decirse que el vector de características restante f5 esté cerca de cualquier otra cosa y en este caso, se considera un clúster por sí mismo. En consecuencia, se crean dos clústeres c0 (1630) y c1 (1631).

[0163] Según un modo de realización, un clúster se etiqueta de conformidad con los criterios relacionados con los vectores de características constituyentes del clúster. Por ejemplo, un clúster se etiqueta para expresar una representación a un conjunto de aplicaciones con características de análisis estático y de comportamiento suficientemente similares. Para cada clúster (p. ej., 1610), el centroide (p. ej., 1600) es calculado por la aplicación de un algoritmo de agrupación en clústeres en el conjunto de todos los vectores de características. Por tanto, para cada clúster (p. ej., 1610), se genera un vector de distancia (1630) al calcular la distancia euclidiana (o similar) de cada elemento del clúster (f1, f2, f3, f4) (1601, 1602, 1603, 1604, respectivamente) hasta su centroide (1600). El elemento más cercano (1604) (es decir, el vector de características (1604) que produce la menor distancia euclidiana, d_{04}) al centroide (1600) se elige que sea el vector de características representativo para dicho clúster (1610), y utilizando el identificador de solicitud (250), el nombre de archivo 1640, 835) de la aplicación correspondiente (p. ej., 40) se recupera de la base de datos (130). De este modo, cada clúster (p. ej., 1610) es nombrado según su aplicación (más representativa) y por tanto se dice que múltiples aplicaciones en el mismo clúster son similares a esta en cuanto al comportamiento, la aplicación más representativa del clúster (1610). Por ejemplo, puede generarse un clúster por aplicaciones cuyos desarrolladores compartieron el código como interfaces de servidor/API, siguieron plantillas/patrones de diseño como elementos de interfaz de GUI, y/o

introdujeron el uso de mecanismos periféricos similares como servidores de retransmisión de anuncios del mismo nivel. Finalmente, si más de una aplicación exhibe la distancia mínima al centroide, entonces se elige la aplicación con el identificador de solicitud más bajo.

5 **[0164]** La figura 17 ilustra una tabla de ejemplo utilizada para almacenar vectores de características (1200) en la base de datos (130). Una tabla de vectores de características (1700) se utiliza para almacenar cada vector de características generado (1200), indexado de nuevo por su identificador de solicitud correspondiente (250). Además, se proporciona una etiqueta de clúster (1710), y una marca de anomalía de clúster (1720).

10 **[0165]** La figura 18 ilustra un cálculo de ejemplo de una asignación de clúster inicial para un conjunto de vectores de característica (1805); aquí los vectores de características duplicados se eliminan y se toman medidas para la limpieza de datos. Los vectores de características se escalan (1810) a intervalos numéricos estables; más tarde, se almacenará la transformación de la escala. Se calcula la matriz de distancia por pares entre todos los pares de vectores de características (1815), utilizando algunas medidas de distancia multidimensionales como la distancia euclidiana. A continuación, se determina un número objetivo de clústeres (1820). Este número se elige de acuerdo con varios criterios, como el grado de agrupación en clústeres óptimo o se determina a través de alguna heurística con respecto al tamaño del conjunto de datos. Luego, se utiliza una técnica de agrupación en clústeres, como la agrupación en clústeres jerárquica (1830), para encontrar la pertenencia de los vectores de características (1805) a los clústeres. A continuación, utilizando dichos resultados de agrupación en clústeres, se recuperan los centroides de los clústeres (1835) y también se recupera la pertenencia de los vectores de características a los clústeres (1840). La pertenencia de los vectores de características a clústeres obtenida en (1840) se registra en la tabla de vectores de características (1700) de la base de datos (130), indexada de nuevo por el identificador de solicitud (250) de cada vector de características. Finalmente, los clústeres, sus miembros y el factor de escala actual se almacenan (1850) en la base de datos (130).

25 **[0166]** La figura 18A ilustra una tabla de clústeres de ejemplo (1860) que se utiliza para almacenar cada clúster indexado por nombre de clúster (es decir, su aplicación más representativa) (1861), su tamaño actual (1862), la marca de tiempo de su creación (1863), un vector de pertenencia en serie (1864), su centroide actual (1865), y la fecha de la última modificación (1867),

30 **[0167]** La figura 19 ilustra un ejemplo de asignación en línea de un nuevo vector de características en un conjunto precalculado de clústeres (representado en cuanto a sus centroides) tal como se describe en la Fig. 18. En primer lugar, teniendo en cuenta el identificador de solicitud proporcionado, se recupera el vector de características (1900) de la base de datos (130). El vector de características se escala (1905) posteriormente utilizando la transformación de escala utilizada actualmente ((1810). Luego, se recuperan los centroides (1910) y se calcula la distancia por pares del vector de características frente a cada centroide actual (1915). Se identifica el centroide con la distancia mínima al nuevo vector de características (1920). Luego, se recupera el conjunto de vectores de características asociado actualmente con este clúster (1925, 1930) y se calcula la matriz de distancia para todas las distancias por pares entre este nuevo vector de características y los miembros del clúster (1935). Luego, utilizando esta matriz de distancia, se aplica una prueba aproximada o imprecisa de pertenencia (1940, 1945). Si este nuevo vector de características tiene una fuerte pertenencia al conjunto existente, el nuevo vector de características se asigna al mejor centroide elegido y se actualiza la base de datos. De lo contrario, el vector de características es etiquetado como una anomalía de agrupamiento (1960). La tabla de vectores de características (1700) de la base de datos (130) se actualiza después con esta asignación (1970), indexada de nuevo por el correspondiente identificador de solicitud (250). Esto puede ocurrir incluso cuando el nuevo vector de características hace que las estadísticas de pertenencia al clúster sean significativamente diferentes; entonces, se rechaza la hipótesis de que el nuevo vector de características puede asignarse a este centroide.

45 **[0168]** Un experto en la materia aprecia que la prueba de la pertenencia es de naturaleza difusa y aproximada y con el tiempo, los vectores de características entrantes serán asignados a los centroides de una manera que podría degenerar las asignaciones de agrupaciones en clústeres. Por este motivo, el presente sistema asocia un desencadenante para un evento de reagrupamiento en clústeres, que esencialmente invoca las etapas de la Fig. 18 con un conjunto de los vectores de características que se han de elegir por criterios como todos, un subconjunto de los más activos, los anteriores más un conjunto de los centroides más activos, una muestra ponderada de la pertenencia a clústeres, una muestra aleatoria de las pertenencias a clústeres, todos (o un subconjunto de) los datos disponibles.

50 **[0169]** Recalcular una nueva asignación de clústeres (1830) que contabiliza vectores de características (1910) que no están presentes en el conjunto original de vectores de características (1810) se consigue de la siguiente manera. Según se añaden nuevos vectores de características, el conjunto de clústeres predefinidos (1420, 1430, 1440) puede requerir que se actualice para representar cambios en los miembros entre vectores de características previos, la aparición de nuevos clústeres, el cálculo de nuevos centroides de clústeres y nombres de clústeres, y la asignación de nuevos vectores de características a clústeres predefinidos. El conjunto de clústeres predefinidos (1840) y sus centroides (1845) se recupera (2020) para después recuperar todos los vectores de características constituyentes (2030) de la tabla de vectores de características (1700) de la base de datos (130) así como la pertenencia de sus vectores de características constituyentes 1810). Los nuevos

vectores de características (2010) se combinan con los vectores de características constituyentes (2030) para generar el conjunto actual de vectores de características (2040). El conjunto actual de vectores de características se escala (2050) y se almacena la transformación de la escala (2060). Después, se invocan las etapas del diagrama de flujo de la Fig. 18, dando como resultado la generación de un nuevo subconjunto de clústeres (2070), sus centroides (2075), los nombres de los clústeres (2080), y la asignación de miembros (2085) del conjunto actual de vectores de características (2040) a los nuevos clústeres (2070).

[0170] Según un modo de realización, la generación de un nuevo subconjunto de clústeres predefinidos se basa en criterios como el número de anomalías observadas hasta ahora, el número de muestras de *malware* conocido que no es tenido en cuenta por el presente conjunto de clústeres predefinidos, una métrica de validación basada en la importancia estadística de la pertenencia de vectores de características a clústeres, y/o la última vez desde que se llevó a cabo el último evento de agrupamiento para generar clústeres predefinidos.

[0171] Según un modo de realización, un evento de reagrupación en clústeres puede volver a asignar un vector de características asignado previamente a un clúster distinto a un nuevo clúster en función de la disponibilidad de análisis posteriores. Un informe de usuario final es generado de forma dinámica, a petición del usuario, que contiene la mayoría de resultados actualizados sobre cualquier identificador de solicitud.

[0172] La presente invención proporciona medios automatizados para identificar posibles candidatos de día cero (o de lo contrario, falsos positivos) que representan formas análogas en el comportamiento de familias de *malware* conocidas por medio de una similitud suficiente en función de algunos criterios como la distancia euclidiana entre el vector de características de un binario de aplicación a un clúster de una familia de *malware* conocida.

[0173] La figura 20 ilustra un modelo de ejemplo del componente de consola (2100) para su utilización con el presente sistema, según un modo de realización. Los usuarios se registran en la consola para interactuar mediante una de varias vistas preconstruidas que realizan proyecciones analíticas sobre análisis almacenados en la base de datos (130). La consola se compone de las siguientes vistas: una vista de controlador (2130), una vista de servidor web (2120), una vista de base de datos (2160), una vista de servidor AV (2140), y una vista de distribuidor (2150). Los usuarios seleccionan una vista al hacer clic en un icono que corresponde a cada vista descrita anteriormente en una barra lateral de selección de vista de sistema (2110). Las vistas funcionan en selecciones de análisis en función de criterios como la clave de tienda de aplicaciones, la fuente del envío, el intervalo de tiempo, etc. La consola comprende un servidor web (300) que proporciona un acceso de usuario seguro en Internet, recupera entradas y parámetros (como la clave de tienda de aplicaciones, intervalo de tiempo, estado de la infección, identificador de solicitud, etc.) de dicho usuario, y ejecuta los *scripts* (310) correspondientes asociados con dichas vistas de análisis preconstruidas de los contenidos de la base de datos en función de dichas entradas y parámetros. Los *scripts* recuperan resultados de análisis, detalles y resúmenes de la base de datos y los preparan para la presentación en la vista de consola correspondiente.

[0174] La vista de base de datos proporciona análisis exhaustivos derivados de análisis basados en agregados aplicados sobre varias facetas de los resultados del conjunto de programas de análisis para dichos identificadores de solicitud en áreas como, pero sin carácter limitativo, fuentes de envío, análisis de marcas rojas, detección de intrusos, alcance de la red, tablas de fusión para la fuente AV así como el rendimiento del clasificador de aprendizaje automático etc. Las Figs. 21A, 21B y 21C ilustran capturas de pantalla complementarias de una vista de base de datos de ejemplo. La vista de base de datos proporciona análisis basados en agregados para un subconjunto arbitrario de análisis seleccionados en función de algunos criterios como, por ejemplo, pero sin carácter limitativo, la pertenencia a la clave de tienda de aplicaciones, el intervalo de tiempo, la expresión regular aplicada sobre el nombre de un paquete o de una aplicación, y/o las combinaciones de los anteriores. El sistema actual permite que un usuario de la consola recupere los análisis sólo para los envíos asociados a las claves de tienda de aplicaciones del usuario. El sistema actual también proporciona medios para seleccionar un conjunto de programas de análisis sobre el que realizar la vista de base de datos. Por ejemplo, un control de intervalo de tiempo (2155) permite especificar un intervalo de tiempo (1/1/2011:11:20AM a 1/10/2011:1030AM) utilizando vistas de calendario (2156). De manera similar, un campo de texto permite recuperar los análisis que coincidan con la restricción basada en texto especificada en uno de los varios campos de análisis seleccionados previamente (por ejemplo, dirección de Internet, nombre del paquete). Ambas restricciones se pueden aplicar a los resultados de los análisis de subconjuntos, pero primero se aplica la restricción del intervalo de tiempo.

[0175] La figura 21A ilustra una vista parcial de ejemplo de la página de base de datos que contiene los contenidos de análisis de las secciones relacionadas con

- el cuadro de selección de intervalo de tiempo (2155);
- la casilla de restricción adicional (2156) que se ha de imponer a las entradas seleccionadas, como las expresiones regulares, sobre la presencia de un nombre de archivo coincidente, una dirección de protocolo de Internet o una regla de bandera roja activada (evaluación de riesgos);

- la sección de estadísticas básicas (2157) que contiene un resumen que comprende el número de análisis, el número de binarios de aplicación diferentes enviados, el número de binarios de aplicación infectados de acuerdo con un oráculo de banco de pruebas/de referencia una fuente AV externa, una fuente AV interna y el clasificador de agrupación en clúster de aprendizaje automático del sistema, el número de fuentes de envío diferentes, el número de días que abarcan estos envíos, el número total de direcciones IP salientes únicas, el número total de direcciones IP entrantes únicas y estadísticas sobre el tiempo de finalización medio y estadísticas sobre el tiempo de finalización medio y la carga de análisis almacenada en caché vs. sin almacenar en caché;
- resumen y detalle de las fuentes de envío (2158), como el número de binarios de aplicación enviados y el intervalo de tiempo para las fuentes de envío observadas representadas en términos de direcciones de Internet, y
- resumen y detalle del alcance de la red entrante y saliente (2159) (como el volumen de paquetes y el paquete medio por dirección de Internet) tal como se observa en todos los análisis encontrados dentro de un intervalo de tiempo especificado.

15 **[0176]** La figura 21B ilustra una vista parcial de ejemplo de la página de la base de datos que contiene los contenidos de análisis de las secciones relacionadas con

- los detalles comparativos de las estadísticas de infección (2160) (entre el análisis AV interno, el análisis AV externo y el clasificador de agrupación de aprendizaje automático),
- las tablas de confusión comparativas (2161) para documentar verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos en términos de aplicaciones y vectores de características para estos (análisis AV interno, análisis AV externo y clasificador de agrupación de aprendizaje automático) cuando cada uno de ellos se compara con el mismo oráculo banco de pruebas/de referencia, y
- estadísticas de red y detalles del tipo de archivo para objetos descargados/cargados (2162) tal y como se observa en todos los análisis encontrados dentro de un intervalo de tiempo especificado.

20 **[0177]** La figura 21C ilustra una vista parcial de ejemplo de la página de la base de datos que contiene contenidos de análisis de las secciones relacionadas con

- resúmenes y detalles (2170) de las alertas de detección de intrusos, y
- resúmenes y detalles de las normas de evaluación del riesgo de marcas rojas activadas (2171), que comprenden comparaciones resumidas y detalladas de las normas de evaluación del riesgo activadas entre los análisis para binarios de aplicación infectados frente a análisis de binarios de aplicación sin infectar, tal como se observa en todos los análisis encontrados dentro de un intervalo de tiempo específico.

25 **[0178]** La figura 22 ilustra un ejemplo de barra lateral de la vista del sistema (2110) que proporciona el medio principal de navegación entre las vistas de la consola construidas previamente en el sistema, según un modo de realización del presente sistema. Al hacer clic en un indicador de componentes (por ejemplo, 2240) se abre la vista de componentes correspondiente (por ejemplo, 2140) de la consola. Ilustra los indicadores que muestran medidas de rendimiento a corto plazo relacionadas con cada una de las vistas enumeradas en la FIG. 20. El indicador de servidor web (2220) muestra una medida relacionada con la velocidad de solicitudes enviadas por hora. El indicador de controlador (2230) muestra una medida relacionada con el número de solicitudes completadas por hora. El indicador del servidor AV (2240) muestra una medida relacionada con el número de análisis completados por hora. El indicador de distribuidor (2250) muestra una medida relacionada con el número de espacios aislados actualmente en uso.

30 **[0179]** La figura 23 ilustra una vista de controlador (2300) de ejemplo según un modo de realización del presente sistema. Proporciona una vista en tiempo real del estado actual (2325) de los análisis seleccionados procesados por el servicio en nube. Para cualquier solicitud de espacio aislado (200, por ejemplo, 2335), muestra los datos de solicitud introducidos (*datSolic*) como el identificador de solicitud, el nombre de la aplicación, la fecha de planificación y el estado de la solicitud dentro del flujo de trabajo del servicio en nube, expresados en términos de un componente y una etapa dentro de dicho componente. Un panel de intervalos de tiempo (2315), expresado en relación con la fecha/hora de inicio y fecha/hora de finalización, controla el intervalo de tiempo de las consultas sobre los análisis seleccionados de los contenidos de la base de datos (130). Los valores del intervalo de tiempo se expresan en relación con una fecha/hora de inicio, y una fecha/hora de finalización se comparte en todas las vistas de la consola y controla el intervalo de tiempo de todas las consultas subyacentes desde dentro de la consola hasta la base de datos (130). Un panel de búsqueda (2320) permite restringir de manera adicional la selección de análisis basados en criterios como la coincidencia de nombres de paquetes, direcciones de Internet y reglas de evaluación de riesgos activadas dentro de un análisis. Un panel de validez de registro (2330) identifica la distribución de la calidad de los archivos de registro extraídos para un análisis y se

calcula de nuevo para el intervalo de tiempo seleccionado. Como en todas las vistas de la consola, la vista del sistema (2310, 2110) muestra el estado actual de los otros componentes y proporciona acceso a las otras vistas de componentes a través del indicador de componentes correspondiente. Un panel de asignación de solicitudes (2325) identifica todos los análisis seleccionados que se encuentran en el intervalo especificado y/o sujetos a una restricción específica. Las filas individuales (por ejemplo, 2335) de esta vista (2325) permiten al usuario ver la vista correspondiente de la tarjeta de informe del identificador de solicitud (véase la FIG. 29), permitiendo el acceso a una tarjeta rápida resumida (véase la FIG. 30) y a informes detallados (informe del usuario final, véase la FIG. 10) de los análisis orientados al usuario final para el identificador de solicitud determinado en dicha fila (por ejemplo, 2335). Un resumen para un panel de asignación de solicitudes (2340) proporciona una vista estadística resumida del rendimiento del servicio en la nube (20). El promedio del promedio por hora estimado para varios intervalos de tiempo dentro del intervalo de tiempo seleccionado se muestra para el número de solicitudes de espacio aislado (por ejemplo, 200). Del mismo modo, un gráfico muestra el recuento total observado en cada una de las diversas etapas de progreso del flujo de trabajo de evaluación del espacio aislado. Algunos ejemplos de dichas etapas del flujo de trabajo de evaluación del espacio aislado en el servicio en la nube son: «recibido en el servidor web», «programado en el controlador», «recibido en el distribuidor», «iniciado en el espacio aislado», «completado en el espacio aislado», «evaluado por el conjunto de programas de análisis». Estas estadísticas se muestran para tres intervalos, actualmente 1/3, 2/3, y 3/3 del tiempo seleccionado (2315).

[0180] La figura 24 ilustra una vista de servidor web (2120) de ejemplo según un modo de realización del presente sistema. La vista del servidor web incluye tres paneles: el encabezado de control (2401), la tabla de solicitudes (2402) y los gráficos de resumen (2403). La tabla de solicitudes (2420) proporciona una vista en tiempo real del estado actual de las solicitudes de envío seleccionadas (por ejemplo, 2425) enviadas al servicio en nube, que se ha determinado que se realizaron dentro del intervalo de tiempo especificado por el usuario (2405) y/o la condición de búsqueda (2420). La tabla de solicitudes (2420) muestra las filas de cada envío, ya esté pendiente o se haya completado, que se ajuste a los criterios seleccionados. Cada fila muestra datos de identificación de la solicitud de espacio aislado, como el identificador de la solicitud, el nombre de la aplicación, la fecha de envío y el estado de finalización y los resultados de la solicitud (por ejemplo, completados, algunos registros incompletos, pero con una alta calificación de *malware*, como en (2425)) para todas las solicitudes que cumplen los criterios especificados. Como también se encuentra en la vista de controlador (2300), las filas individuales (por ejemplo, 2335) de la tabla de solicitudes (2325) permiten el acceso a la vista de tarjeta de informe del identificador de solicitud correspondiente (ver FIG. 29). Dicho informe proporciona información de cuentas del flujo de trabajo, así como acceso a través de enlaces a la tarjeta rápida (ver FIG. 30) y al informe del usuario final (ver FIG. 10) para el identificador de solicitud determinado.

[0181] Tanto la vista del servidor web (2300) como la vista del controlador (2400) proporcionan también una vista de análisis agregado (2310, 2410) que permite realizar análisis de visualización de red sobre una selección arbitraria de filas de la tabla de solicitud correspondiente (2325, 2420). La selección se realiza al seleccionar filas dentro de la tabla de solicitudes correspondiente. Se utilizan técnicas de interfaz de usuario web estándar (específicamente, clic-mayus-subrayar) de selección de tablas discontinua para permitir que el usuario especifique cualquier conjunto arbitrario de identificadores de solicitud de la tabla de solicitudes correspondiente (2402). La consola proporciona apoyo para un análisis de visualización de red manual (2408, ver la Fig. 26) y/o para un análisis de asignación de GeolP (2409, ver la Fig. 25) de los análisis seleccionados.

[0182] La figura 24A ilustra una vista de ejemplo de la vista del servidor AV (2475), según un modo de realización del presente sistema. Comprende una tabla de solicitudes (2480) construida de forma similar a la descrita anteriormente, pero con un estado de seguimiento de las solicitudes de análisis individuales para análisis antivirus de complementos, si los hubiera, puestos a disposición del sistema. Cada fila proporciona datos de identificación de la solicitud y el estado y/o resultados del análisis antivirus. Un ejemplo detallado de datos de identificación de la solicitud incluye el nombre de la aplicación, la firma MD5, el tipo de análisis, las horas de inicio y finalización, el estado de la infección y el tipo de infección. La vista del servidor AV (2475) también proporciona una línea de tiempo de la carga de solicitudes presentada a los análisis antivirus de complementos internos que permite identificar deficiencias en el procesamiento de solicitudes por parte de dichos análisis, así como proporcionar medios para modificar la capacidad computacional de dichos análisis.

[0183] La figura 25 ilustra un análisis de distribución de GeolP de ejemplo (2600) (ilustrado para algunas selecciones de análisis de ejemplo), según un modo de realización. El análisis de distribución de GeolP proporciona una distribución geográfica de la actividad de Internet (ver FIG. 12) observada durante la ejecución de los binarios de aplicación correspondientes asociados a los análisis seleccionados. Este análisis proporciona medios para proporcionar desgloses detallados para el tráfico de red (en términos de bytes y/o paquetes transferidos) por conexión, espacio aislado, subred y/o país (2609). El resumen y los detalles de estos desglosamientos se utilizan para especificar los nodos (2610) y los bordes (2620) entre dichos nodos para su uso posterior en la generación de la correspondiente visualización del gráfico de red del tráfico agregado de red seleccionado.

[0184] La distribución de GeolP se obtiene consultando bases de datos de GeolP para una distribución de una IP/DNS determinada en unas coordenadas geográficas y trazando la coordenada resultante en el mapa. Según

un modo de realización, la distribución de GeoIP no se limita al nivel de detalle en función del país, sino también al estado, la ciudad y la calle, en función del nivel de detalle de las bases de datos geográficas disponibles. La distribución de GeoIP también permite identificar otras etiquetas de mapeo GeoIP con etiquetas de infección asociadas al identificador de solicitud correspondiente. En un modo de realización, la distribución de GeoIP también se puede presentar opcionalmente en una vista de línea de tiempo, permitiendo la visualización de actualizaciones de la distribución de GeoIP en función del tiempo en el mapa actual, donde dichas actualizaciones se muestran con respecto al orden temporal de los identificadores de solicitud seleccionados.

[0185] La figura 26 ilustra un análisis de visualización de red de ejemplo (2500) que muestra una visualización de un gráfico de red social del tráfico de red agregado seleccionado, derivado del análisis de red (véase la Fig. 12) observado durante la ejecución de los identificadores de solicitud seleccionados. El análisis de redes sociales se utiliza para ilustrar la relación entre el tráfico TCP/UDP entre conexiones que desglosan IP, subredes comunes entre estas IP, y países para estas.

[0186] El sistema actual, según un modo de realización, permite que la visualización de red (2408) identifique otros datos nodales de etiquetas con etiquetas de infección extraídas para el identificador de solicitud correspondiente. En un modo de realización, la visualización de red también puede presentarse opcionalmente en una vista de línea de tiempo, lo que permite mostrar las actualizaciones de la visualización de red (2408) en función del tiempo a la visualización actual, en la que dichas actualizaciones se muestran con respecto al orden cronológico de los identificadores de solicitud seleccionados. De acuerdo con un modo de realización, la visualización de red (2408) puede utilizarse para monitorizar la evolución y propagación de infecciones y anomalías de clústeres.

[0187] La figura 27 ilustra una interfaz de análisis de visualización de red interactiva de ejemplo (2550) según un modo de realización de la presente invención, y adecuada para la interacción del usuario por un usuario de la consola. Los elementos de la interfaz se pueden mover y seleccionar haciendo clic en cada uno de ellos. Los países (por ejemplo, EE.UU. (2560), China (2570) asociados con direcciones de Internet que se encuentran en los análisis seleccionados para ser alcanzados por espacios aislados (por ejemplo, 2565) son identificados y escalados mediante una medida proporcional a su carga de tráfico agregado. Se identifican también las direcciones de protocolo de Internet (IP) relevantes (por ejemplo, 2575). Cada nodo seleccionado por el usuario en el gráfico se desglosa en el panel lateral (2555). Cada una de estas listas (por ejemplo, 2556) incluye el nombre del nodo, la función y los enlaces a diversas herramientas de descubrimiento del origen, tales como, pero sin limitarse a, la resolución inversa de nombres de dominio y las entradas de registro en Internet. En un modo de realización previsto, también se proporciona un enlace a las fichas rápidas relevantes de análisis asociadas a dicho nodo. En el sistema actual, los nodos de solicitud (representados mediante el identificador de solicitud) que se ha determinado que están infectados se diferencian visualmente por el color. Del mismo modo, las direcciones IP maliciosas también se diferencian por el color cuando están presentes. El sistema actual ayuda a la identificación de las direcciones de Internet que presentan medidas de alta centralidad. El usuario de la consola puede recuperar de forma interactiva el estado de infección de los espacios aislados de comunicación para estos, así como los registros de origen de red asociados a las direcciones de Internet correspondientes, a fin de ayudar a evaluar las tendencias de la red asociadas a los análisis infectados seleccionados con arreglo a algunos criterios, como los análisis infectados conocidos, los análisis de día cero o los análisis de falsos positivos.

[0188] La figura 28 ilustra una interfaz de vista de distribuidor (2150) de ejemplo para su utilización con el presente sistema, según un modo de realización. La vista de distribuidor permite monitorizar (3010) en tiempo real la asignación de solicitudes de espacio aislado que se ejecutan (por ejemplo, 3020) dentro de un nodo distribuidor. También permite la identificación de ejecuciones de espacio aislado que exceden los límites de tiempo, como por ejemplo en el caso de un emulador interbloqueado y/o que no se haya utilizado recientemente (3030). Se coloca uno o más límites de tiempo (en el peor de los casos) sobre la ejecución de cualquier espacio aislado y que al excederse cualquiera de estos límites de tiempo, el servicio en la nube termine de manera forzada la ejecución del espacio aislado.

[0189] La figura 29 ilustra una vista de ejemplo de la tarjeta de informe del identificador de solicitud (3100) para su uso con el sistema actual de acuerdo con un modo de realización. La vista 3100 resume los datos clave internos extraídos de la base de datos sobre el progreso de una solicitud concreta dentro del flujo de trabajo del servicio en la nube y los componentes de *malware* que comprenden, pero sin carácter limitativo, al menos uno de entre:

- una sección de mensajes emergentes que proporciona la imagen de marca del informe y la imagen del indicador visual de calificación de malware (3105);
- una sección de estadísticas básicas del informe (3110) que proporciona parámetros de identificación del envío;
- una sección de asignación de caché (3120) que proporciona información sobre el uso y la identidad de los resultados de un análisis en caché;

- una sección de acceso a los informes de análisis (3130) que proporciona enlaces a los diversos informes de usuario final y de desarrollador proporcionados por el sistema;
- una sección de distribución geográfica (3140) y una sección detallada y resumida del acceso a la red por país (3150);
- 5 • una sección de asignación de distribuidores (3160) que proporciona un resumen de la asignación de la solicitud del espacio aislado a un espacio aislado dentro de un distribuidor;
- una sección de progreso del flujo de trabajo (3335) que proporciona un resumen del progreso de la solicitud de espacio aislado dentro de los componentes; y
- 10 • una sección de desglose del registro (3340) que proporciona un resumen de los registros extraídos para la solicitud de espacio aislado.

Es una característica del presente sistema que una recarga de la vista de tarjeta de identificador de solicitud recupere los valores más actuales para cada una de estas secciones, permitiendo al operador monitorizar el progreso gradual de una solicitud de espacio aislado a través de esta vista del sistema.

15 **[0190]** También proporciona acceso al informe del espacio aislado (véase la FIG. 9) y muestra una distribución de GeoIP (3105) de la actividad de red (véase la FIG. 12) observada durante la ejecución del identificador de solicitud (3110) determinado. Se muestran los campos de datos seleccionados para la tabla de solicitudes, la tabla de aplicaciones, la tabla de registros, la tabla de progreso y la tabla de análisis AV correspondientes.

20 **[0191]** La figura 30 ilustra un informe de ejemplo de tarjeta rápida para su uso con el sistema actual, de acuerdo con una personificación. El informe de tarjeta rápida proporciona un resumen destacado de los resultados clave sobre la evaluación de riesgos de una aplicación móvil. El informe de tarjeta rápida incluye, sin carácter limitativo, al menos uno de entre:

- una sección de mensajes emergentes que proporciona la imagen de marca del informe (3300) y la imagen del indicador visual de calificación de malware (3305);
- 25 • una sección de estadísticas básicas del informe (3310) que proporciona parámetros de identificación del envío, tales como la fuente del envío, la fecha de presentación, etc., así como que proporciona un enlace (3315) al informe de análisis exhaustivo (véase FIG. 10) para el envío;
- una sección de evaluación de riesgos (3320) que proporciona un resumen y un detalle de las reglas de evaluación de riesgos activadas, desglosando para cada una de dichas reglas, datos de evaluación tales como, pero no limitados a, al menos uno de las categorías de riesgo, descripción de riesgos, clasificación de riesgos, puntuación de riesgos e intensidad de riesgos;
- 30 • una sección de alcance de red (3325) que proporciona un resumen y detalles de las conexiones de red, desglosando para cada uno datos de evaluación tales como, pero sin limitarse a, la dirección de Internet, el país y la carga de tráfico, independientemente de que se sepa o no que dicha dirección de Internet es maliciosa;
- 35 • una sección de detección de intromisiones en la red (3330) que proporciona un resumen y detalles de las alertas de intromisión, desglosando para cada uno estos datos de evaluación, tales como, pero sin carácter limitativo, al menos uno entre la prioridad de la alerta, clasificación de la alerta, descripción de la alerta, un recuento y direcciones de Internet asociadas a la alerta;
- una sección (3335) de resultados de análisis antivirus estático que proporciona un resumen y detalles de los resultados de los análisis antivirus, desglosando para cada uno estos datos de evaluación, tales como, pero sin carácter limitativo, al menos uno de entre el estado de la infección, el tipo de infección, el nombre del análisis, la versión del análisis; y
- 40 • una sección de resultados de clústeres (3340) que proporciona un resumen y detalles de los resultados del método de clasificación de aprendizaje automático, desglosando para cada uno dichos datos de evaluación, tales como, pero sin carácter limitativo, al menos uno de entre el método de clasificación, el resultado de clasificación, la alineación/clase de la clasificación, la probabilidad, la confianza o el recuento por consenso.

45 **[0192]** La presente invención proporciona medios para calcular sin intervención del usuario una clasificación de *malware* para un binario de aplicación que comprende contribuciones numéricas de al menos uno o más de, pero sin limitarse a:

- una puntuación de consenso de clúster y/o un nivel que indique confianza en la asignación de clúster;
- la probabilidad para un resultado de clasificación asociado a un método de clasificación de aprendizaje automático;

- el perfil de evaluación de riesgos asociado al análisis estático del binario de aplicación;
- el perfil de detección de intrusiones en la red obtenido a partir del análisis de red del binario de aplicación en el entorno del espacio aislado instrumentado;
- 5 • un perfil de medidas relacionado con la evaluación de la ofuscación del código dentro del código del binario de aplicación;
- una lista negra de binarios de aplicaciones infectados conocidos;
- una lista blanca de binarios de aplicación con falsos positivos conocidos;
- una descarga de objetos infectados transferidos por la red; y
- un acceso a sitios y/o direcciones de Internet maliciosas conocidas.

10 **[0193]** La presente invención proporciona medios automatizados para identificar posibles candidatos a *malware* de día cero (y, en caso contrario, falsos positivos) a partir de análisis de clasificación (y sus correspondientes aplicaciones móviles) basados en la presencia de una calificación alta de *malware* que no ampliamente aceptada y/o que se sabe que se considera *malware*.

15 **[0194]** La presente invención apoya la realización de inferencias de evaluación de riesgos autónomas sobre un código binario de aplicación móvil (conocido en el presente documento como binario de aplicación) a través del análisis estático sobre dicho binario de aplicación mediante la aplicación de una o más reglas de evaluación de riesgos y la generación de una evaluación de riesgos, como, pero sin limitarse a, una puntuación numérica, un perfil estadístico, una advertencia de texto, y/o una advertencia gráfica.

20 **[0195]** La figura 31 proporciona una ilustración de ejemplo de la especificación de una regla de evaluación de riesgos (3400). Cada una de estas reglas de evaluación de riesgos comprende la especificación de al menos uno o más de:

- uno o más términos de búsqueda independientes (por ejemplo, 3405, 3410, 3415) utilizados para activar la regla en función de los criterios especificados en la misma;
- 25 • una categoría de evaluación del riesgo (3341) utilizada para evaluar el nivel de exposición al riesgo en zonas de interés para el usuario final o para el vector de ataque;
- una explicación y/o descripción del riesgo (3342) utilizada para proporcionar una breve explicación adecuada para su visualización en dispositivos móviles y/o en informes en línea;
- una fuente de referencia de riesgo (3430) utilizada para proporcionar una referencia que documenta de manera adicional el riesgo asociado;
- 30 • una puntuación de riesgo (3450) utilizada para actualizar la puntuación de riesgo total asociada con el binario de aplicación; y
- una generación de contexto de riesgo (3460) utilizada para determinar si el contexto que activó dicha regla debe incluirse en el informe de análisis y, en tal caso, el formato de dicho contexto.

35 En el modo de realización preferido, la selección del conjunto de reglas de evaluación de riesgos de interés a evaluar y sus puntuaciones de riesgo asociadas pueden especificarse de forma independiente en función de criterios establecidos por, pero sin carácter limitativo, las preferencias individuales de los usuarios finales y/o los administradores de empresas.

[0196] La presente invención proporciona medios para evaluar una o más de dichas reglas de evaluación de riesgos frente a dicho binario de aplicación que comprenden:

- 40 • generar una lista exhaustiva de divisiones del modelo de documento (en adelante denominados bloques básicos) del código de bytes de un binario de aplicación al dividir el código de bytes por medio de algunos criterios tales como, pero sin limitarse a, la localización espacial aproximada, funciones, métodos, constantes, cadenas, manifiestos, permisos, archivos adjuntos, encabezados, etc.;
- 45 • recuperar una regla de evaluación de riesgos que comprenda un número variable N de términos de búsqueda independientes;
- para cada uno de estos términos de búsqueda en una regla, realizando una búsqueda en la lista de divisiones para determinar el conjunto de bloques básicos de código en los que se afirma el término de búsqueda;
- recuperar los resultados de hasta N resultados de búsqueda (independientes) si es necesario;
- 50 • recuperar la intersección de los conjuntos resultantes;

- reclamar la activación de una regla de evaluación de riesgos basada en una intersección resultante que no es nula;
 - acumular y reunir la puntuación de riesgo general por categoría de riesgo en función de la contribución a la puntuación de cada regla de evaluación de riesgos activada;
- 5
- agrupar las normas de evaluación de riesgos en categorías de riesgo; y
 - generar un vector de evaluación de riesgo acumulado basado en categorías para dicho binario de aplicación.

[0197] La presente invención proporciona una evaluación de riesgos autónoma en la que las actualizaciones de dichas reglas de evaluación de riesgos son aplicadas por el sistema en todas las evaluaciones posteriores de conjuntos de programas de análisis. Además, es un aspecto de la presente invención que los análisis previos puedan reevaluarse con respecto a las reglas de evaluación de riesgos actualizadas y/o a la acumulación de análisis. En el sistema, el análisis de evaluación de riesgos se realiza dentro del servicio en la nube, acumulando los resultados de análisis de la evaluación de los análisis entrantes de una o más fuentes de envío. La presente invención proporciona medios para agregar y comparar de manera autónoma dicho vector de evaluación de riesgo acumulado basado en categorías para un binario de aplicación determinado con vectores previamente calculados de algún conjunto de binarios de aplicación que identifican anomalías y similitudes en vectores de riesgo acumulado basados en categorías y que generan resultados que deben ser incluidos en el informe de análisis de dicho binario de aplicación. En un modo de realización, la comparación de vectores de evaluación de riesgos acumulados basados en categorías se realiza utilizando métodos de comparación como, pero sin carácter limitativo, técnicas de semejanzas como la agrupación en clústeres, y técnicas de detección de anomalías como puntuaciones z, (es decir) la clasificación basada en variables normalizada estándar.

[0198] En un modo de realización, la selección de vectores de evaluación de riesgos que se han de utilizar en la comparación basada en agregados descrita anteriormente puede estar basada y limitada de manera adicional por criterios, como por ejemplo, que estén derivados del usuario final, la clave de tienda de aplicaciones, un conjunto de vectores de evaluación de riesgos representativos de infecciones por malware, o un perfil de directiva de riesgo derivada de la empresa o derivada de la compañía móvil.

[0199] Un lector experto en la materia apreciaría que dicho análisis de puntuación de la evaluación de riesgos también podría realizarse dentro de un dispositivo móvil mediante técnicas rudimentarias de decompilación de Java y basándose en una base de datos de vectores de evaluación de riesgos acumulados basados en categorías previamente calculados.

[0200] La presente invención proporciona medios automatizados para identificar posibles candidatos de día cero (o de otro modo, falsos positivos) que representen formas polimórficas de reglas de evaluación de riesgos conocidas por medios tales como la evaluación anómala de vectores de evaluación de riesgos acumulados y/o la similitud de éstos con la de instancias de *malware* conocidas.

[0201] En la descripción anterior, con fines explicativos, se establece una nomenclatura específica para proporcionar una comprensión completa de la presente revelación. No obstante, resultará evidente para un experto en la materia que todos estos detalles específicos no son necesarios para poner en práctica lo expuesto en la presente revelación.

[0202] Algunas partes de las descripciones detalladas del presente documento se presentan en términos de algoritmos y representaciones simbólicas de operaciones en bits de datos dentro de una memoria de ordenador. Estas descripciones y representaciones algorítmicas son los medios utilizados por los expertos en el procesamiento de datos para transmitir de la manera más eficaz el contenido de su trabajo a otros expertos en la materia. Un algoritmo está en el presente documento, y en general, concebido para ser una secuencia coherente de etapas que conducen a un resultado deseado. Las etapas son aquellas que requieren manipulaciones físicas de cantidades físicas. Normalmente, aunque no necesariamente, estas cantidades adoptan la forma de señales eléctricas o magnéticas que pueden almacenarse, transferirse, combinarse, compararse y manipularse de otro modo. Se ha demostrado que en ocasiones es conveniente, principalmente por razones de uso común, referirse a estas señales como bits, valores, elementos, símbolos, caracteres, términos, números o similares.

[0203] Sin embargo, debe tenerse en cuenta que todos estos y otros términos similares deben asociarse con las cantidades físicas apropiadas y que son simplemente etiquetas convenientes que se aplican a estas cantidades. A menos que se indique específicamente lo contrario, como se desprende del siguiente análisis, se aprecia que, a lo largo de la descripción, los análisis en los que se utilizan términos como «procesar», «calcular», «cálculo», «determinar», «visualizar» o similares, se refieren a la acción y los procesos de un sistema informático, o de un dispositivo informático electrónico similar, que manipula y transforma datos representados como cantidades físicas (electrónicas) dentro de los registros y las memorias del sistema informático en otros datos representados de forma similar como cantidades físicas dentro de las memorias o los registros o registros del sistema informático, o bien en otros dispositivos de transmisión, visualización o almacenamiento de información de este tipo.

5 **[0204]** La presente exposición también se refiere a un aparato para realizar las operaciones del presente documento. Este aparato puede estar construido especialmente para los fines requeridos, o puede comprender un ordenador de uso general activado o reconfigurado de forma selectiva por un programa informático almacenado en el ordenador. Dicho programa informático puede almacenarse en un medio de almacenamiento legible por ordenador, como por ejemplo, pero sin carácter limitativo, cualquier tipo de disco, incluidos disquetes, discos ópticos, CD-ROM y discos ópticos magnéticos, memorias de sólo lectura (ROM), memorias de acceso aleatorio (RAM), EPROM, EEPROM, tarjetas magnéticas u ópticas, o cualquier tipo de medio adecuado para almacenar instrucciones electrónicas, y cada uno de ellos acoplado a un bus de sistema informático.

10 **[0205]** Los algoritmos presentados en el presente documento no están relacionados de manera inherente con ningún ordenador u otro aparato en particular. Varios sistemas de propósito general, servidores informáticos, u ordenadores personales pueden utilizarse con programas de acuerdo con lo expuesto en este documento, o puede ser conveniente construir un aparato más especializado para realizar los pasos requeridos del método. La estructura requerida para una variedad de estos sistemas se derivará de la descripción anterior. Se apreciará que pueden utilizarse varios lenguajes de programación para implementar lo descrito en la exposición tal como se describe en el presente documento.

15 **[0206]** Además, las diversas características de los ejemplos representativos y de las afirmaciones dependientes pueden combinarse de maneras que no se enumeran de manera específica y explícita, a fin de proporcionar nuevos modos de realización útiles de lo descrito en el presente documento. También se señala expresamente que todos los intervalos de valores o indicaciones de grupos de entidades dan a conocer todos los posibles valores intermedios o entidades intermedias a los efectos de la exposición original, así como a los efectos de restringir la materia objeto reivindicada.

20 **[0207]** Se entiende que los modos de realización descritos en el presente documento tienen el propósito de aclarar y no se debe considerar que limitan el objeto de la divulgación. El alcance de la invención está definido por las reivindicaciones.

25

REIVINDICACIONES

1. Método para evaluar la calidad de aplicaciones móviles, estando realizado el método mediante un entorno de ordenadores en red que comprende un servicio basado en la nube para dispositivos móviles, comprendiendo el método las etapas de método:
 - 5 realizar una evaluación del riesgo de análisis estático de código binario asociado a una aplicación móvil que está siendo enviada por una fuente de envío;
 - 10 examinar un comportamiento de ejecución de la aplicación móvil dentro de un entorno de espacio aislado instrumentado agregando los resultados del análisis del comportamiento de ejecución y del análisis estático;
 - 15 generar, en función de los resultados del análisis agregados del comportamiento de ejecución y del análisis estático, un vector de características que comprende: (i) una característica de resumen de red, (ii) una característica de comportamiento basada en un sistema operativo, y (iii) una característica de análisis estático, donde el vector de característica es un vector de análisis compuesto por uno o más conjuntos de características derivados del análisis de datos relativos a la aplicación seleccionados del grupo que comprende características de ejecución de la aplicación y el análisis de características estáticas de la aplicación;
 - 20 acumular resultados de análisis de un conjunto seleccionado de vectores de características generados previamente;
 - 25 aplicar una o más técnicas de aprendizaje automático en los resultados de análisis acumulados;
 - realizar una reevaluación de agrupaciones en clústeres periódica del conjunto de vectores de características resultando por tanto en un conjunto de clústeres de clasificación;
 - comparar evaluaciones estadísticas agregadas de los resultados del análisis para la aplicación móvil con evaluaciones estadísticas agregadas asociadas a un conjunto de análisis previos utilizando las mismas o distintas aplicaciones móviles; y
 - realizar una clasificación predictiva de la aplicación móvil utilizando el vector de características y las técnicas de aprendizaje automático en los resultados de análisis acumulados, obteniendo estadísticas predictoras que describen características de calidad y vulnerabilidad de aplicaciones móviles.
2. Método según la reivindicación 1, comprendiendo además, por el servicio en la nube, la generación de un informe de análisis que comprende al menos uno de: una evaluación de riesgos que identifica características de comportamiento sospechoso de la aplicación móvil; una calificación de confianza de *malware* que indica la confianza de la evaluación de riesgos; una calificación de riesgos de *malware* que indica la peligrosidad de los riesgos asociados; y una etiqueta de *malware* que indica detalles sobre la naturaleza de los riesgos asociados a la aplicación móvil.
3. Método según la reivindicación 1, comprendiendo además, por el servicio en la nube, el análisis de archivos de registro generados durante la ejecución y el análisis de la aplicación móvil.
4. Método según la reivindicación 2, comprendiendo además, por el servicio en la nube, el análisis de una o más fuentes seleccionadas del grupo de transacciones del protocolo de Internet, archivos transferidos por la red, el alcance de la red, la carga de red desde/hacia páginas de difusión de anuncios, conexiones de red a sitios maliciosos, el recorrido de interfaz de usuario, el alcance geográfico de Internet, alertas de intrusión en la red, el histograma de API del sistema operativo, el histograma de API a nivel de aplicación, el perfil de utilización de recursos, y cambios en el sistema de archivos.
5. Método según la reivindicación 1, comprendiendo además, por el servicio en la nube, proporcionar técnicas de aprendizaje automático de clasificación seleccionadas del grupo que comprende árboles de decisión y máquinas de vector de soporte para generar predictores sobre la aplicación móvil en función de los datos de comportamiento agregados de múltiples aplicaciones.
6. Sistema para evaluar la calidad de aplicaciones móviles, comprendiendo el sistema:
 - al menos una memoria que almacena instrucciones ejecutables por ordenador; y
 - al menos una unidad de procesamiento para ejecutar las instrucciones almacenadas en la memoria, donde la ejecución de las instrucciones resulta en una o más aplicaciones comprendiendo en conjunto:
 - 50 un componente de servidor web para recibir solicitudes para analizar una pluralidad de aplicaciones móviles;
 - un componente de controlador para planificar las solicitudes recibidas;
 - un módulo de distribuidor para proporcionar un espacio aislado adecuado para ejecutar la solicitud en función de uno o más criterios asociados con la solicitud y ejecutar la solicitud en el espacio aislado; y
 - 55 un componente de base de datos para almacenar los datos resultantes de la ejecución de la solicitud;
 - donde el componente de servidor web, el componente de controlador, el módulo de distribuidor y el componente de base de datos se proporcionan mediante un servicio basado en la nube;
 - donde el componente de base de datos está configurado para acumular resultados de análisis de una o más fuentes de envío;

donde el módulo de distribuidor está configurado para: realizar una evaluación del riesgo de análisis estático de código binario asociado con una aplicación móvil que está siendo enviada por una fuente de envío;

5 examinar el comportamiento de ejecución de la aplicación móvil dentro de un entorno de espacio aislado instrumentado;

agregar resultados de análisis del comportamiento de ejecución y el análisis estático; agregar resultados de análisis para generar uno o más vectores de características, comprendiendo cada vector de características: (i) una característica de resumen de red, (ii) una característica de comportamiento basada en un sistema operativo, y (iii) una característica de análisis estático, donde el vector de característica es un vector de análisis compuesto por uno o más conjuntos de características derivados del análisis de datos relativos a la aplicación seleccionados del grupo que comprende características de ejecución de la aplicación y el análisis de características estáticas de la aplicación;

15 acumular resultados de análisis de un conjunto seleccionado de vectores de característica generados previamente;

aplicar una o más técnicas de aprendizaje automático en los resultados de análisis acumulados; realizar una reevaluación de agrupaciones en clústeres periódica del conjunto de vectores de características resultando por tanto en un conjunto de clústeres de clasificación; y

20 comparar evaluaciones estadísticas agregadas de los resultados del análisis para la aplicación móvil con evaluaciones estadísticas agregadas asociadas a un conjunto de análisis previos utilizando las mismas o distintas aplicaciones móviles;

y

25 donde el componente de base de datos está configurado además para llevar a cabo una clasificación predictiva de la aplicación móvil utilizando el uno o más vectores de características y las técnicas de aprendizaje automático en los resultados de análisis acumulados, obteniendo estadísticas predictoras que describen características de calidad y vulnerabilidad de aplicaciones móviles.

7. Sistema según la reivindicación 6, donde la ejecución de la solicitud comprende realizar inferencias de evaluación de riesgo autónomas en el código binario de la aplicación móvil mediante un análisis estático del código binario de la aplicación al aplicar y evaluar una o más reglas de evaluación de riesgos durante la ejecución de la solicitud.

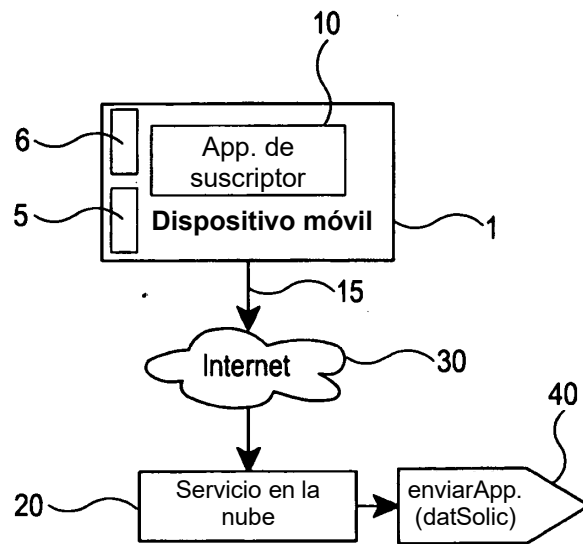


Figura 1

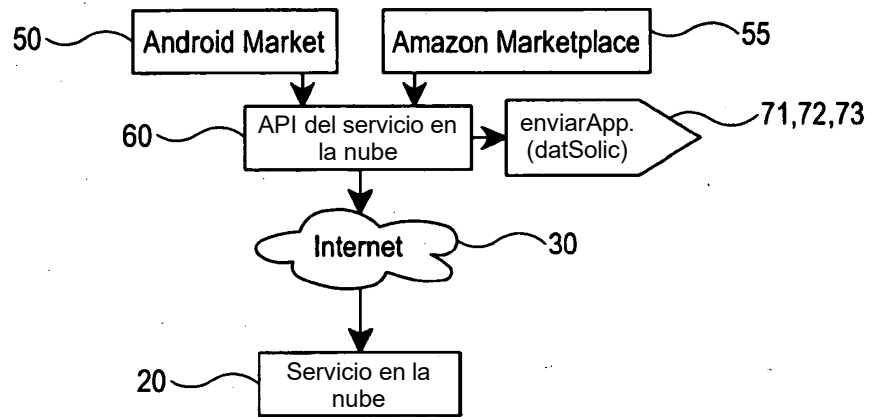


Figura 1A

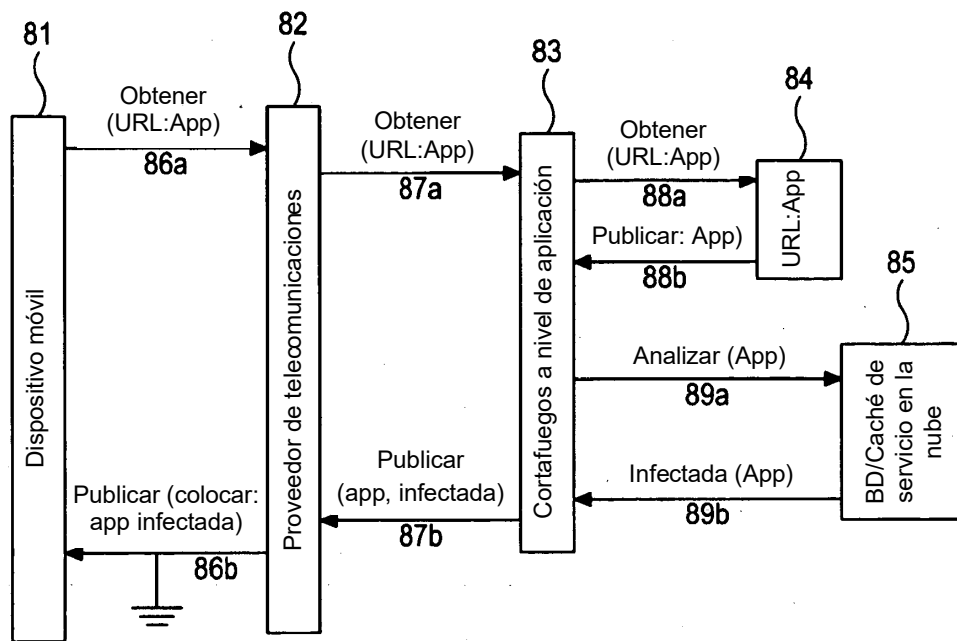


Figura 1B

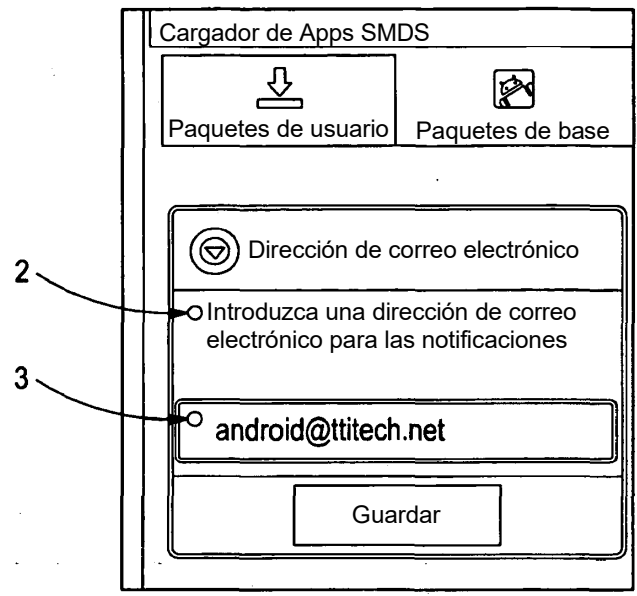


Figura 1C

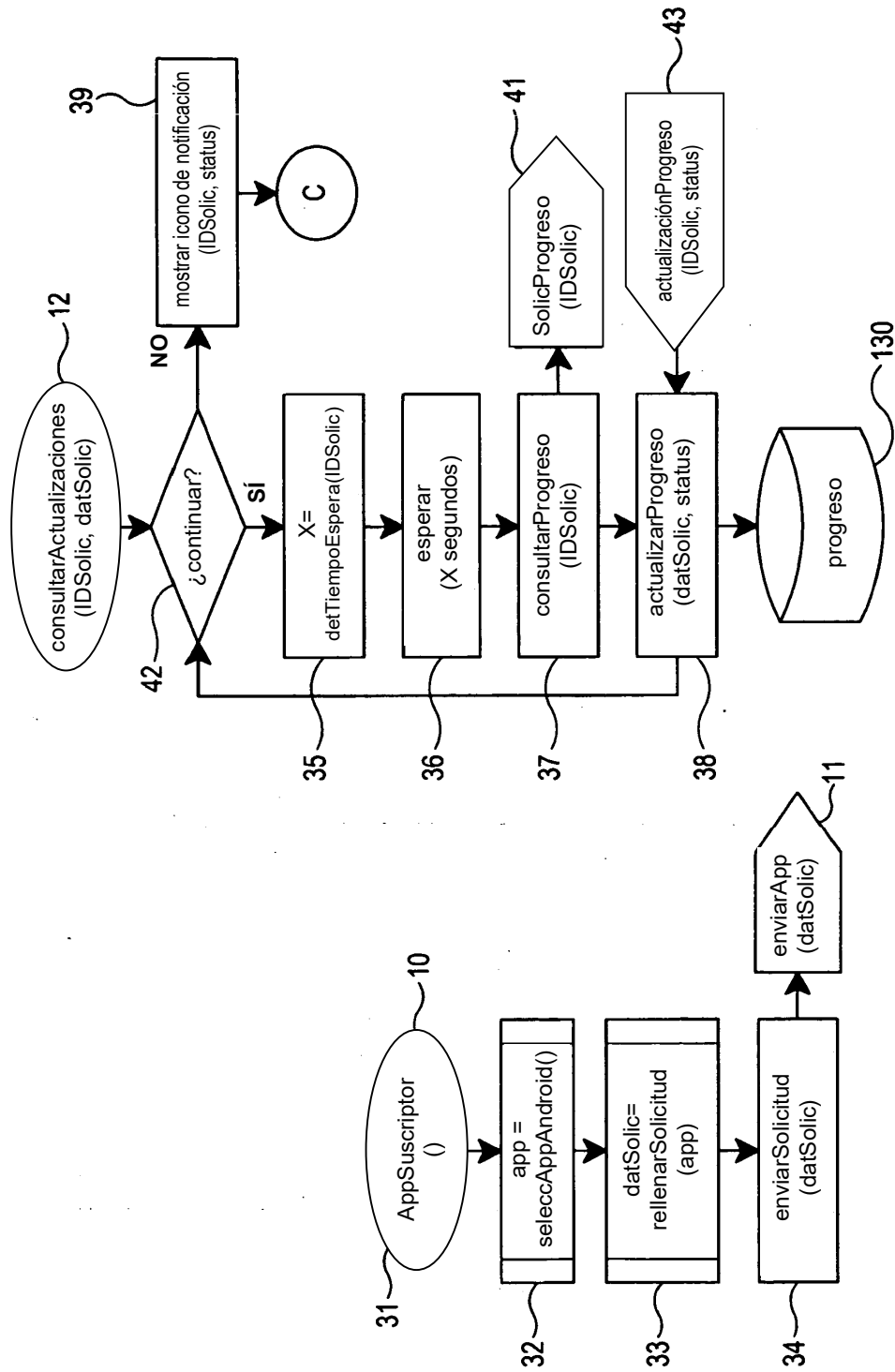


Figura 1D

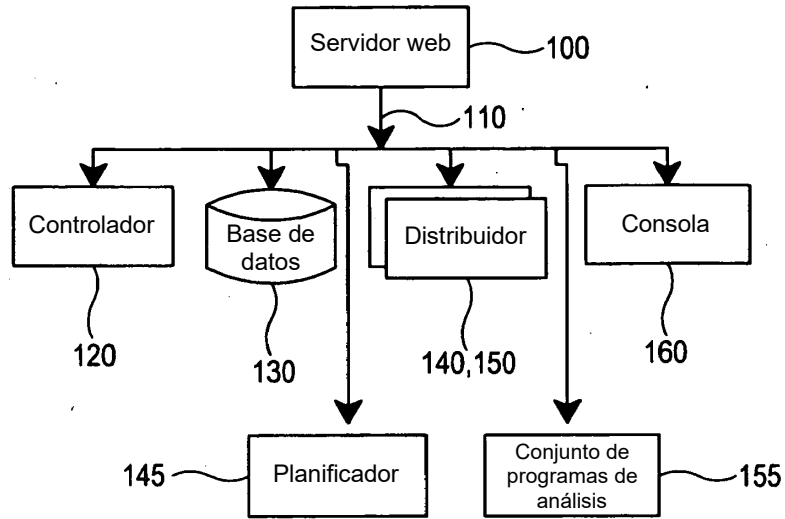


Figura 2

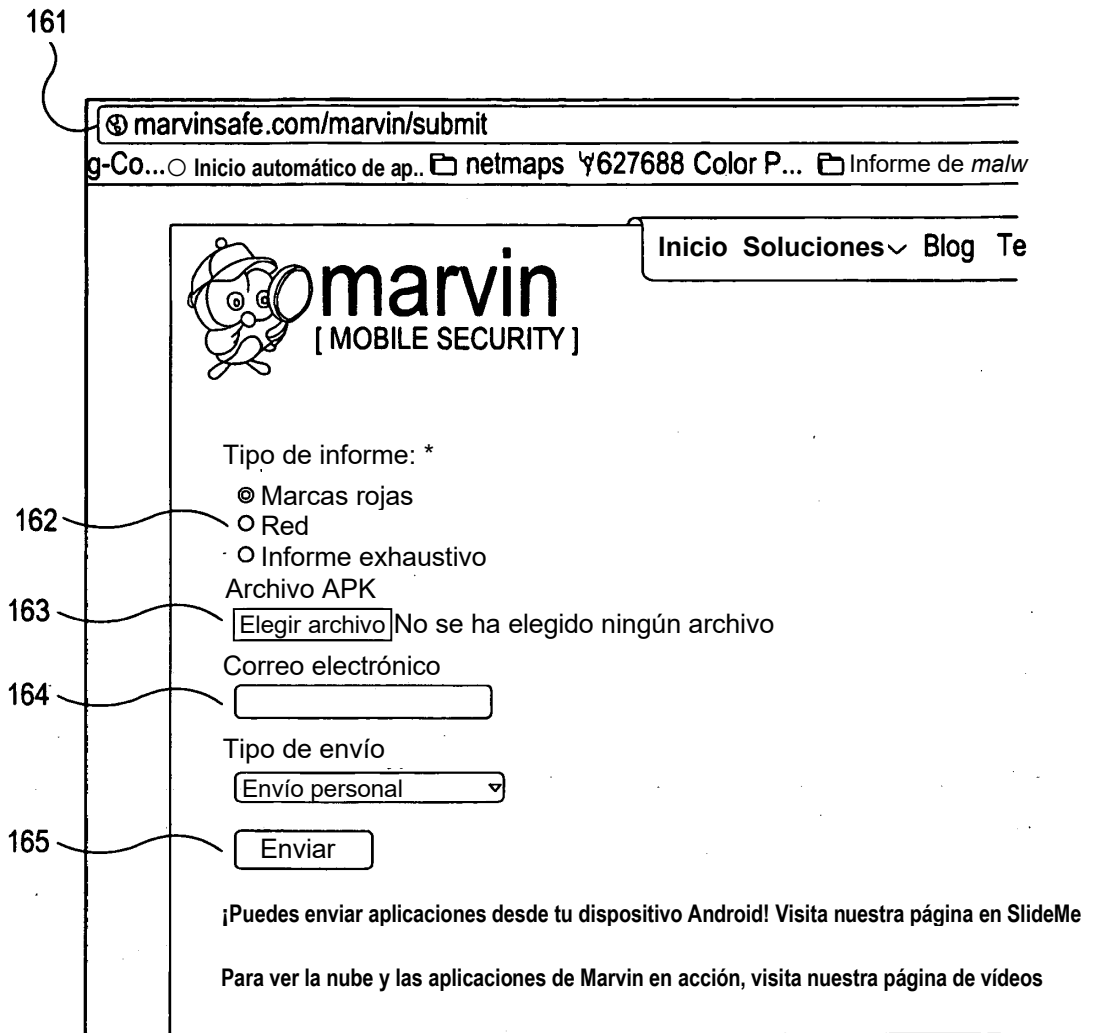


Figura 2A

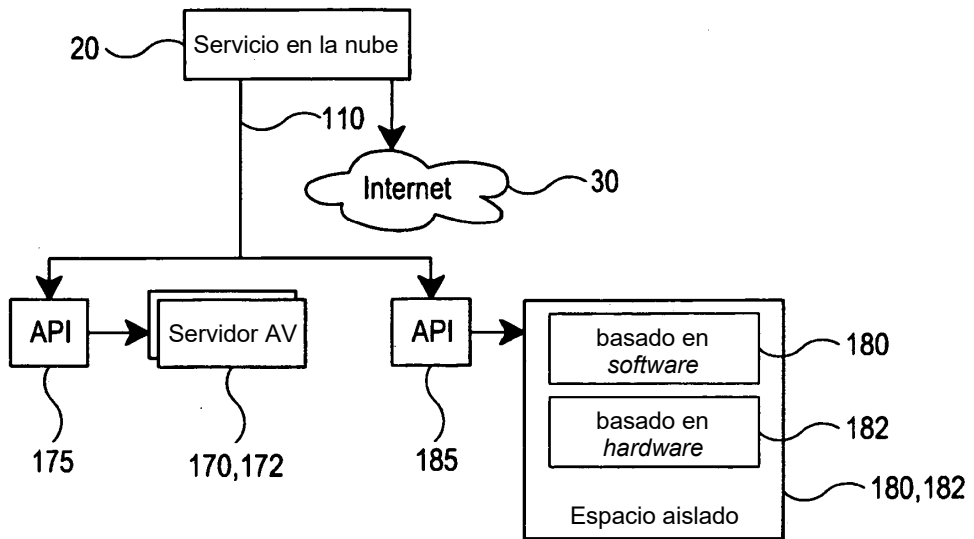


Figura 2B

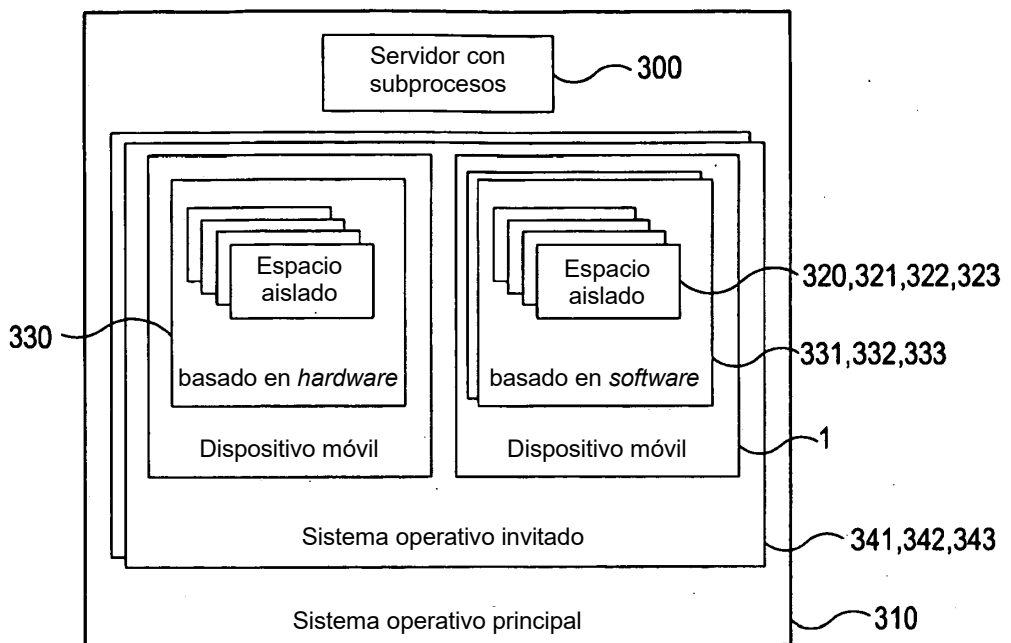


Figura 3

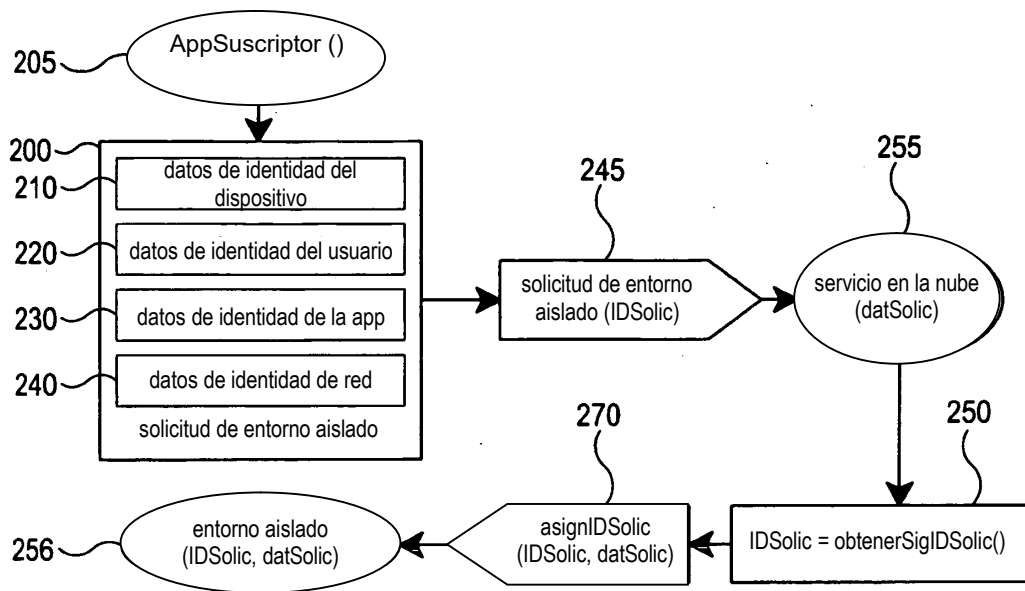


Figura 3A

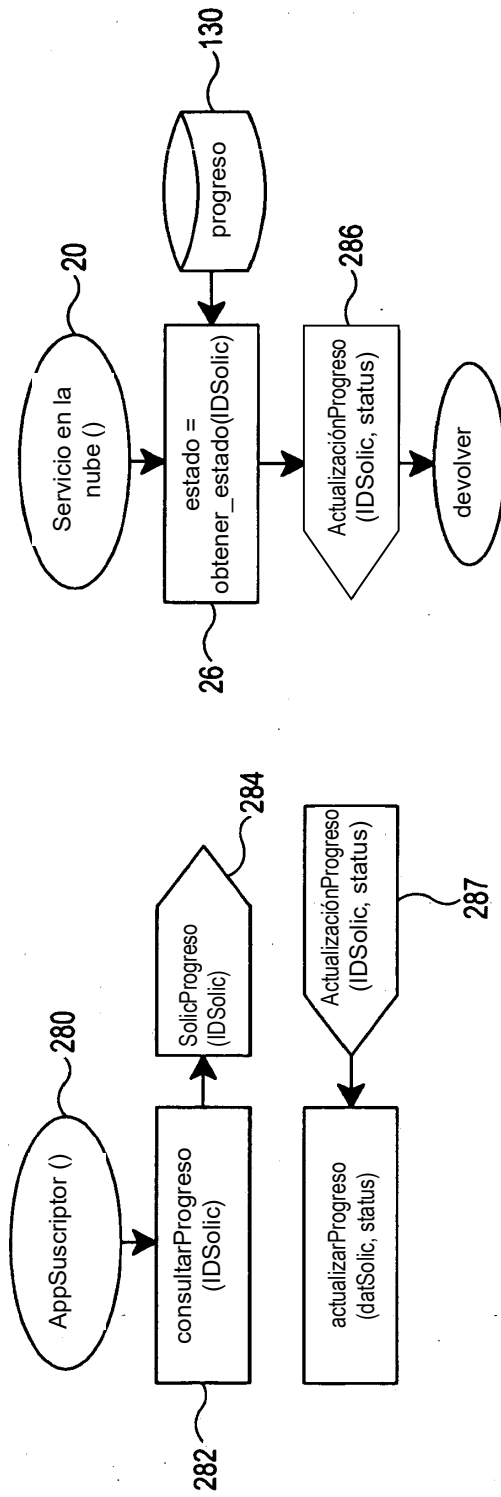


Figura 3B

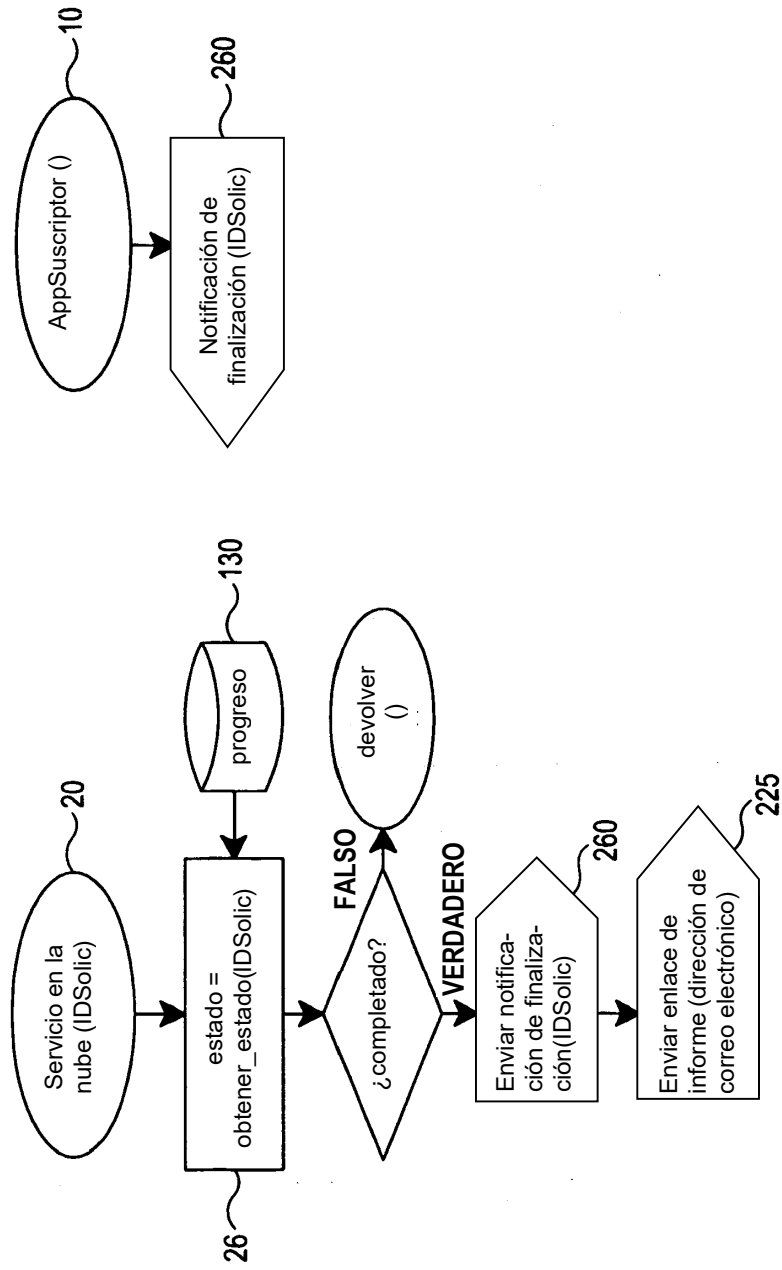


Figura 3C

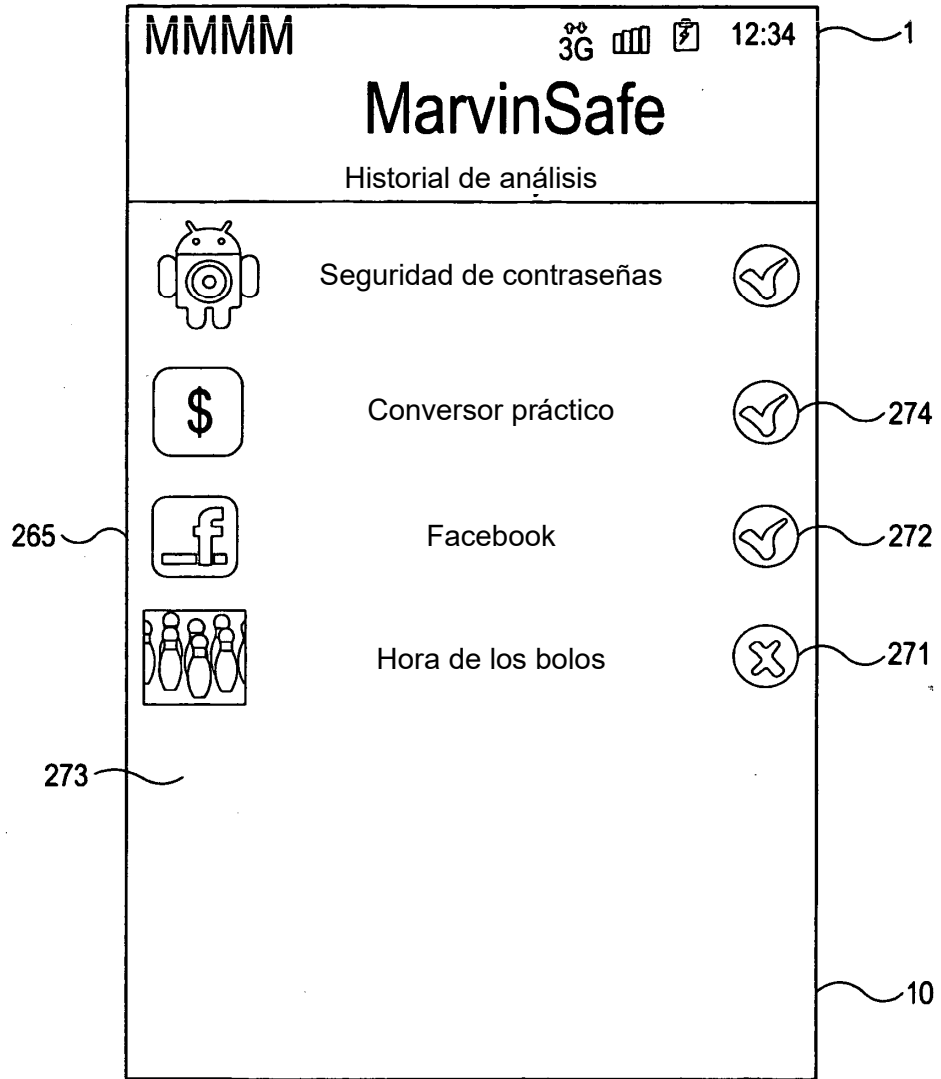


Figura 3D

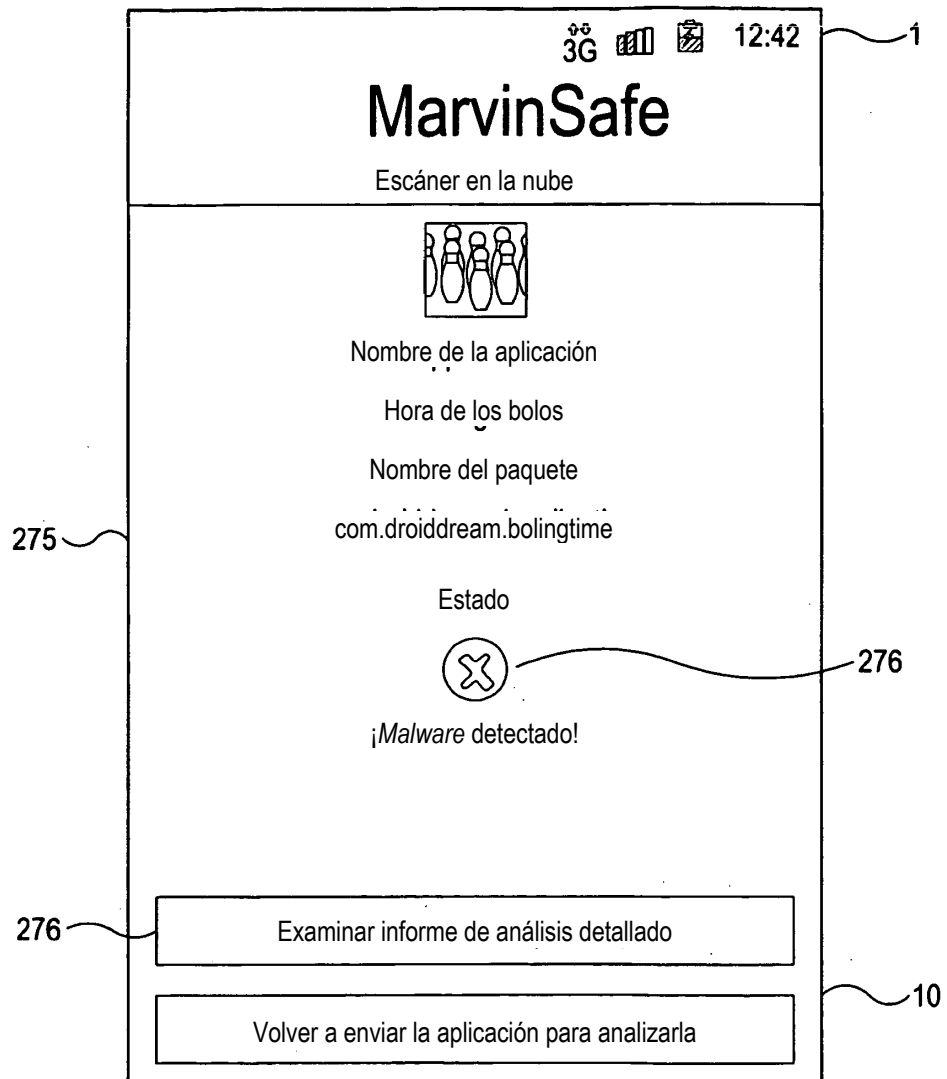


Figura 3E

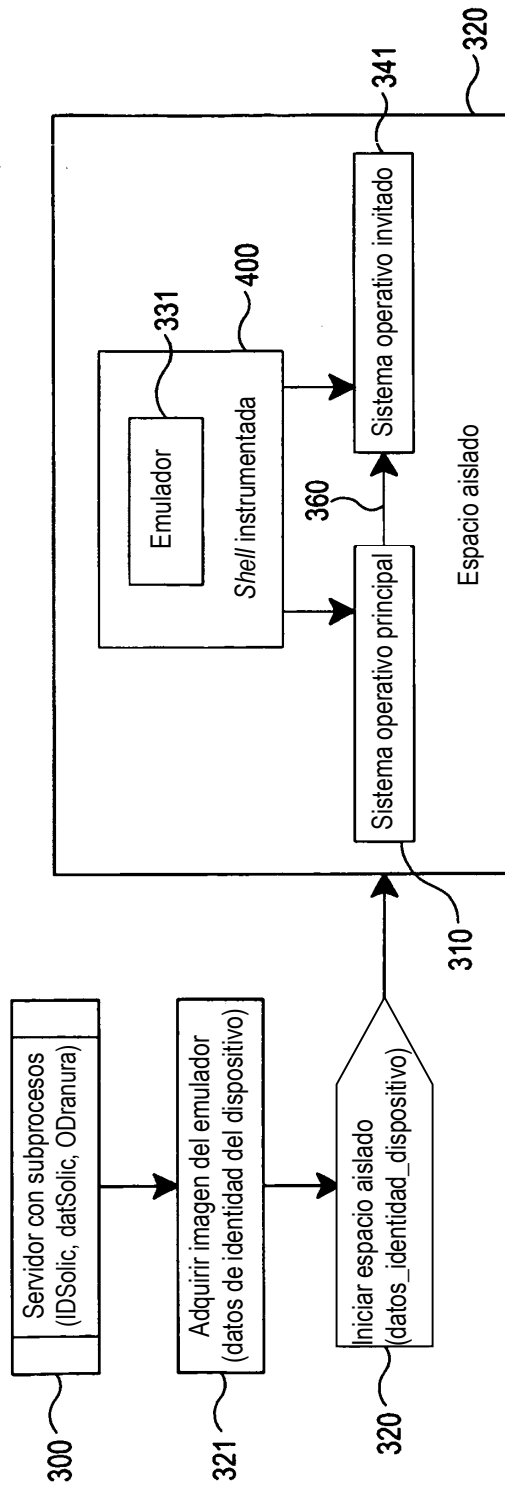


Figura 4A

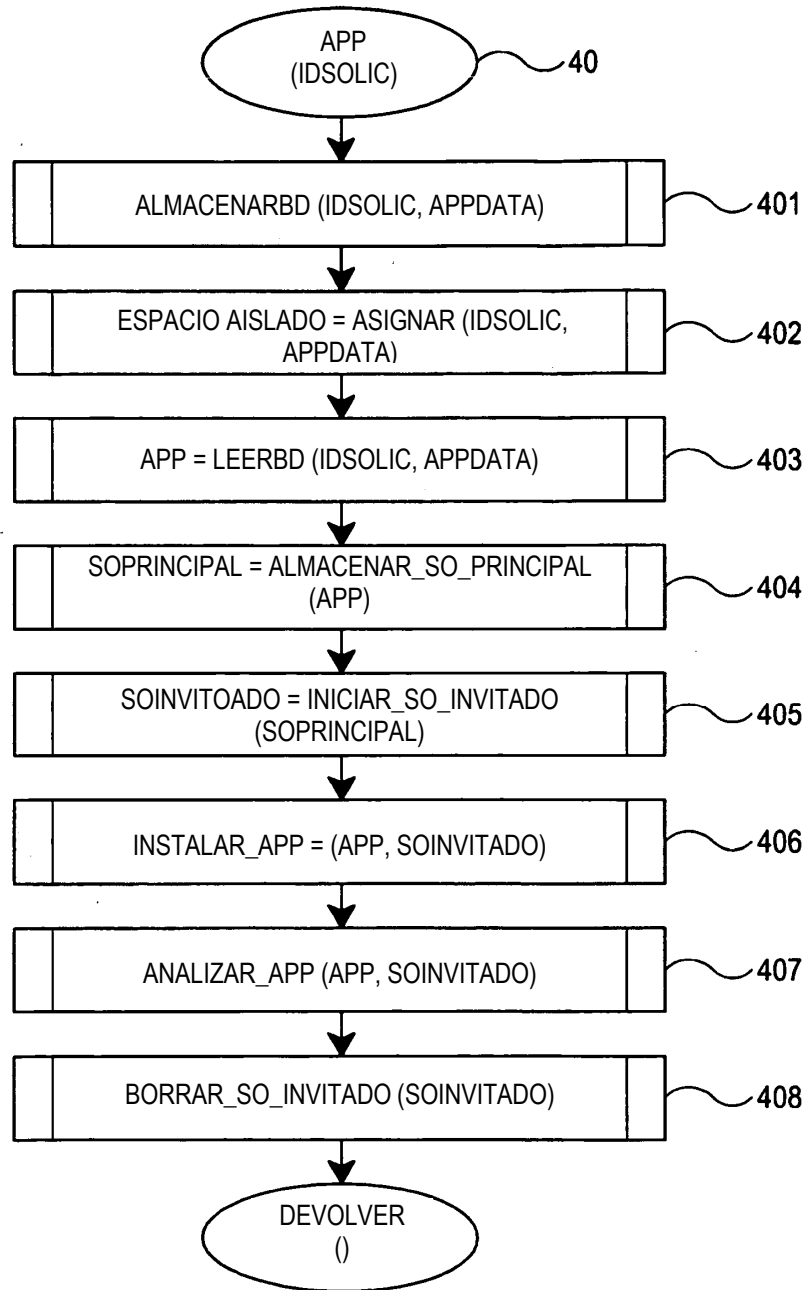


Figura 4B

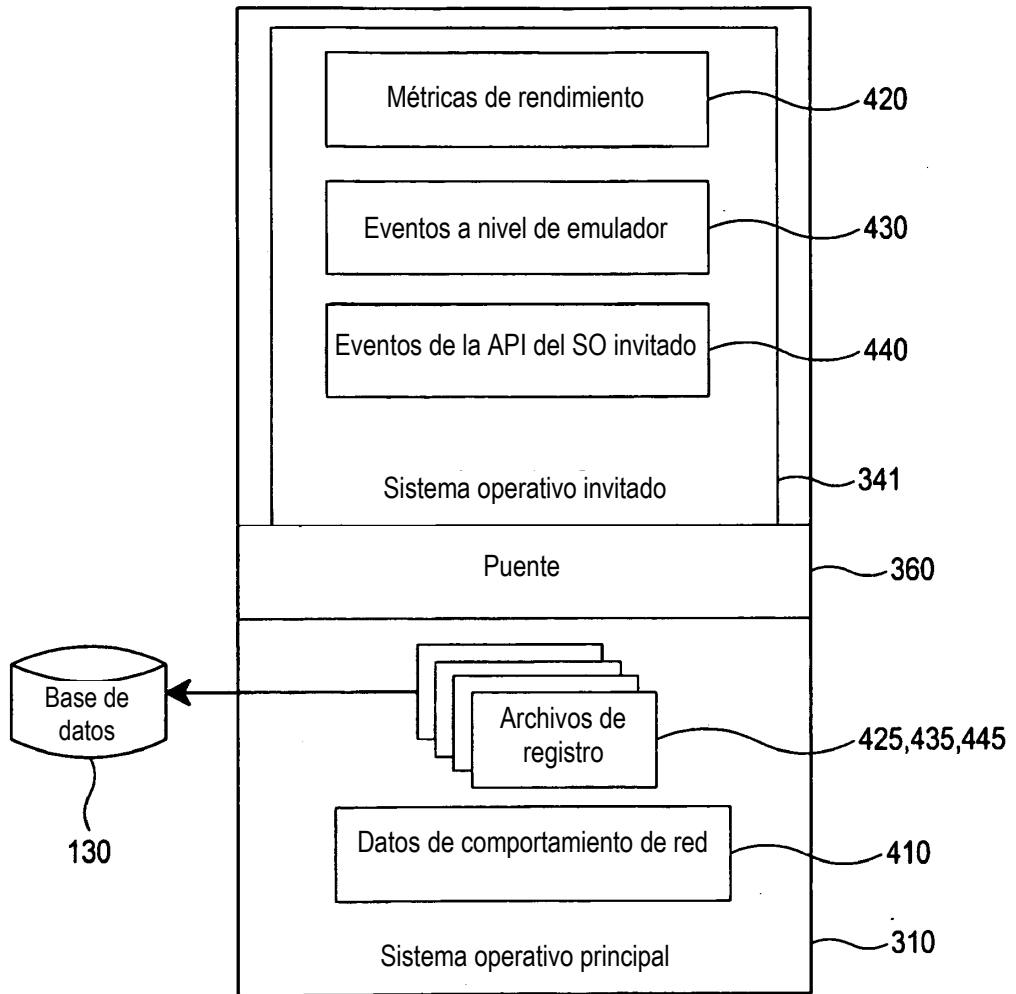


Figura 4C

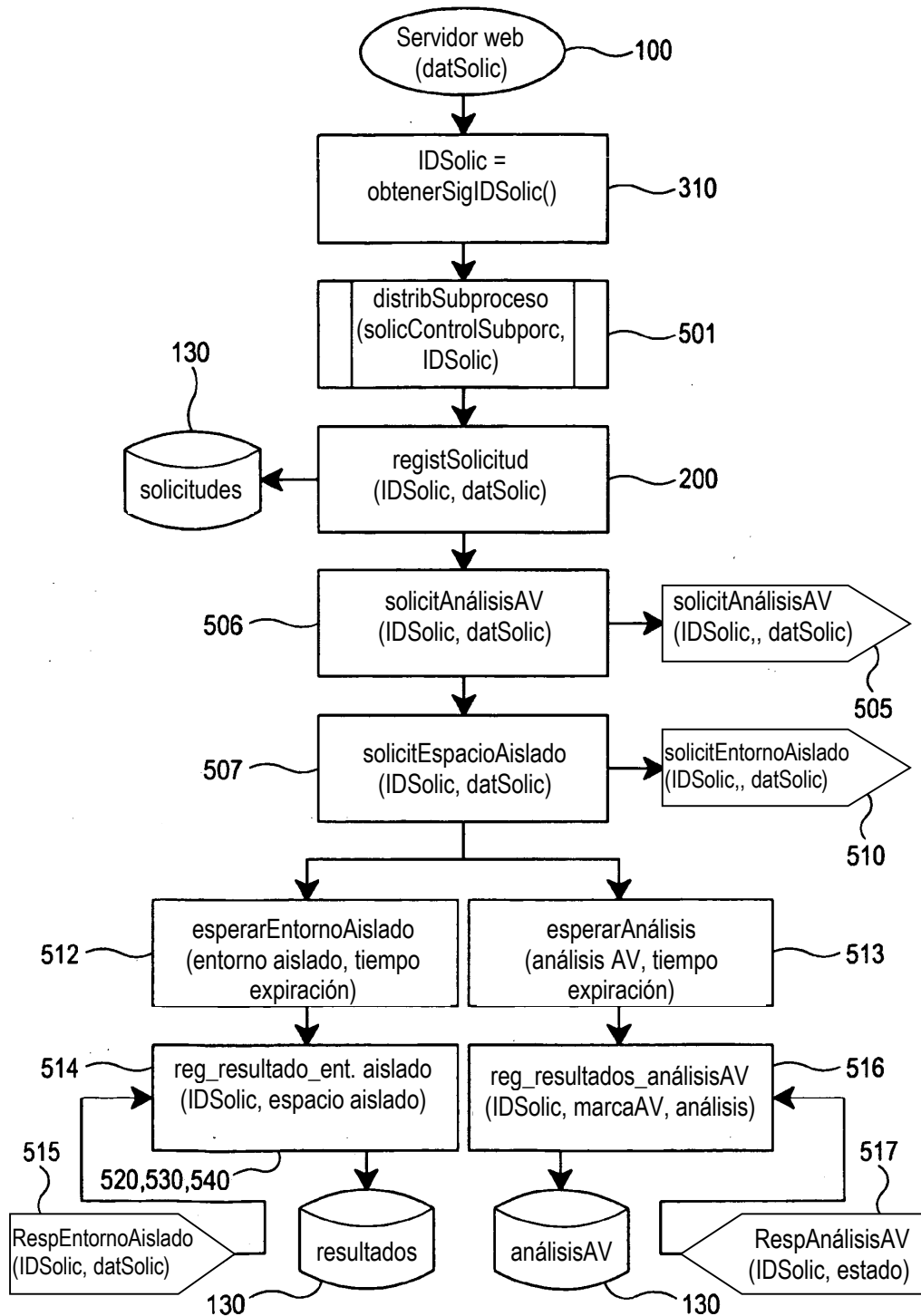


Figura 5

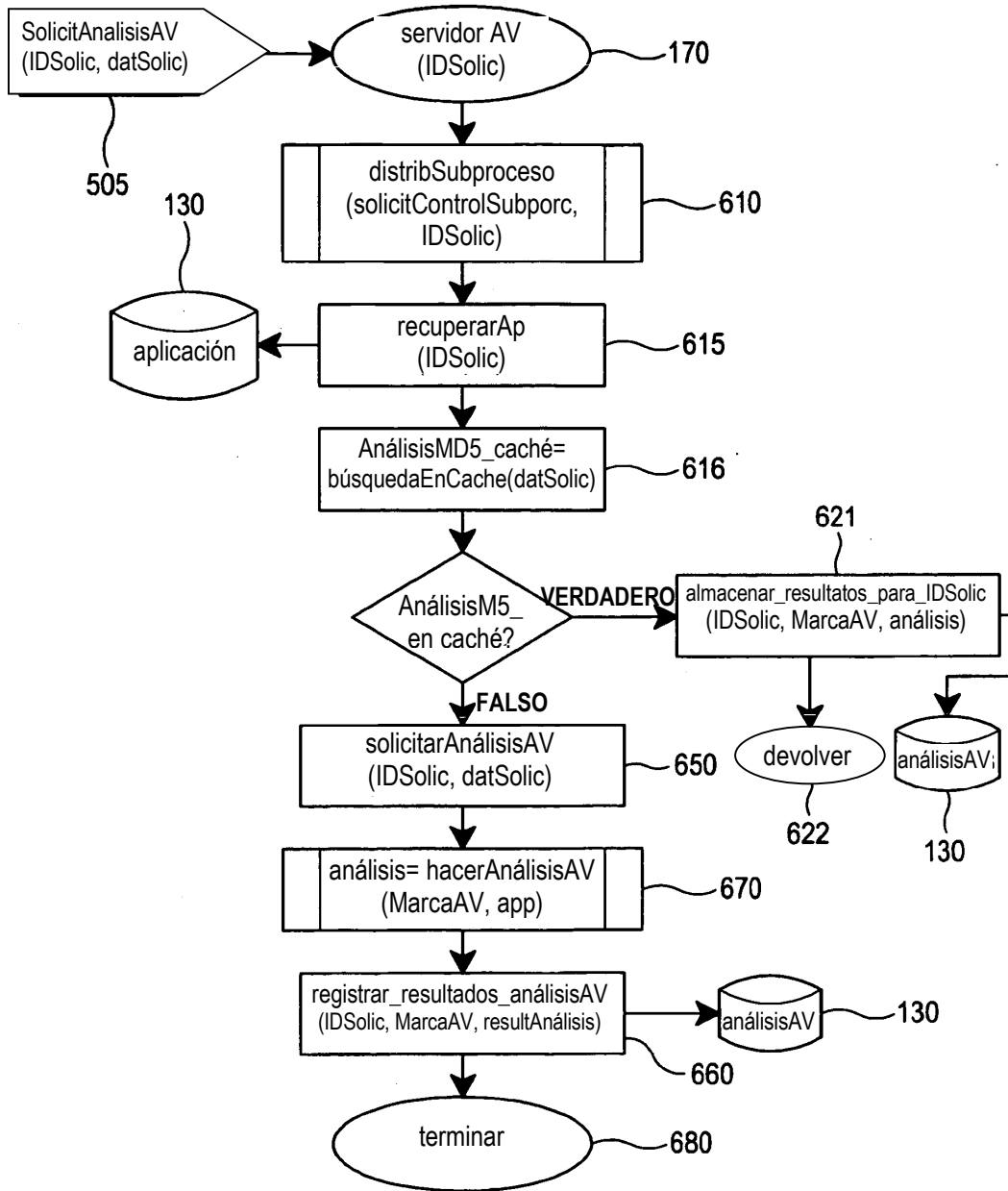


Figura 6

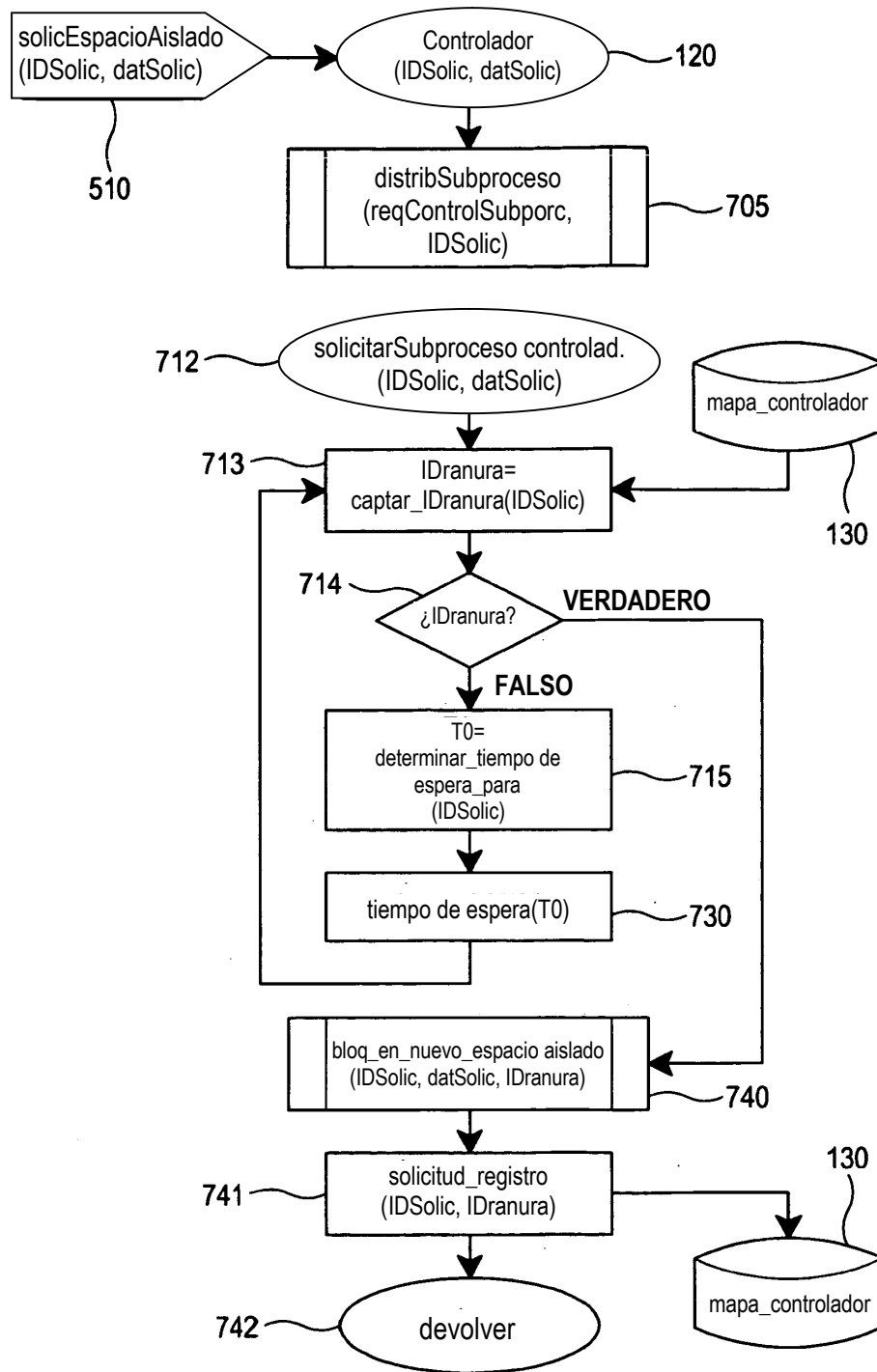


Figura 7

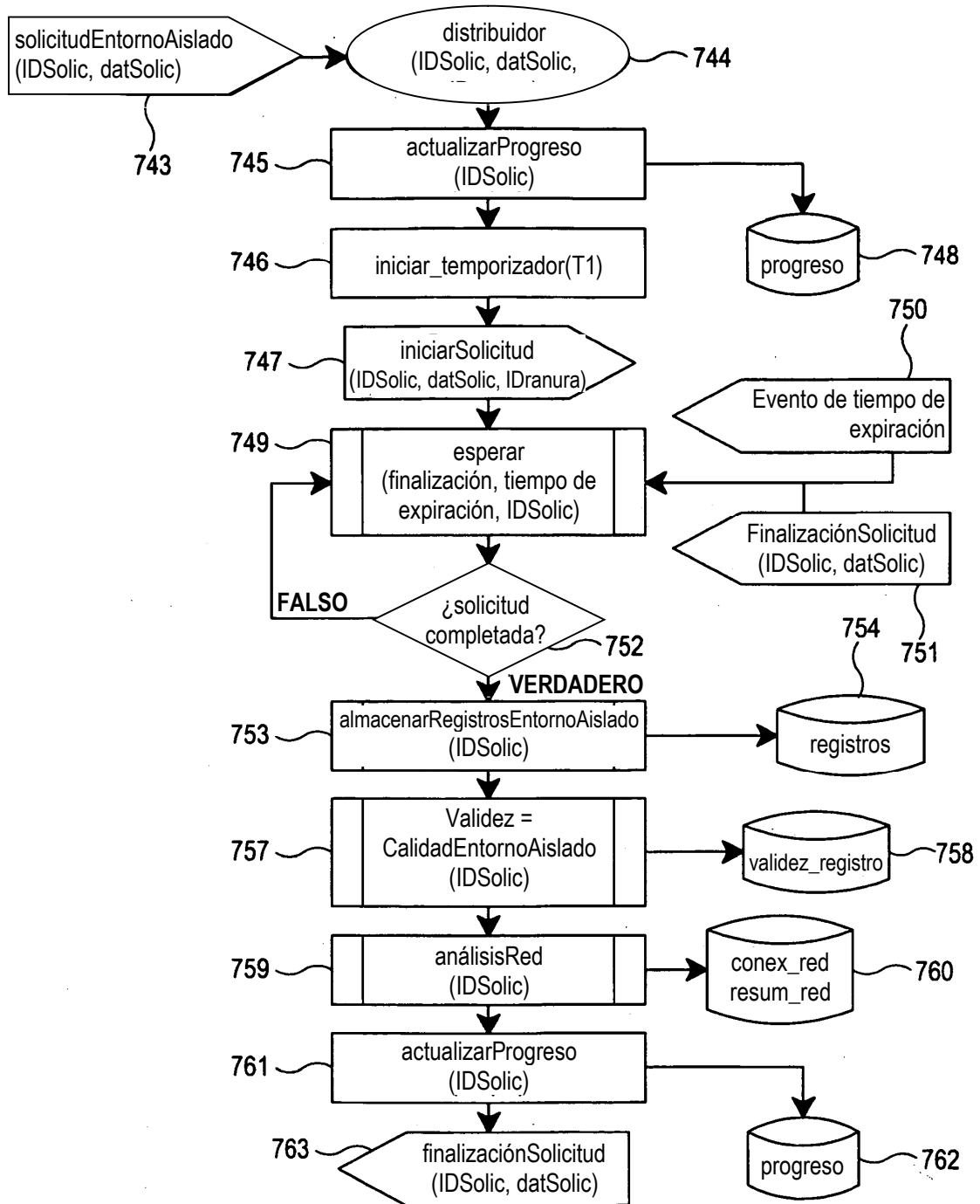


Figura 7A

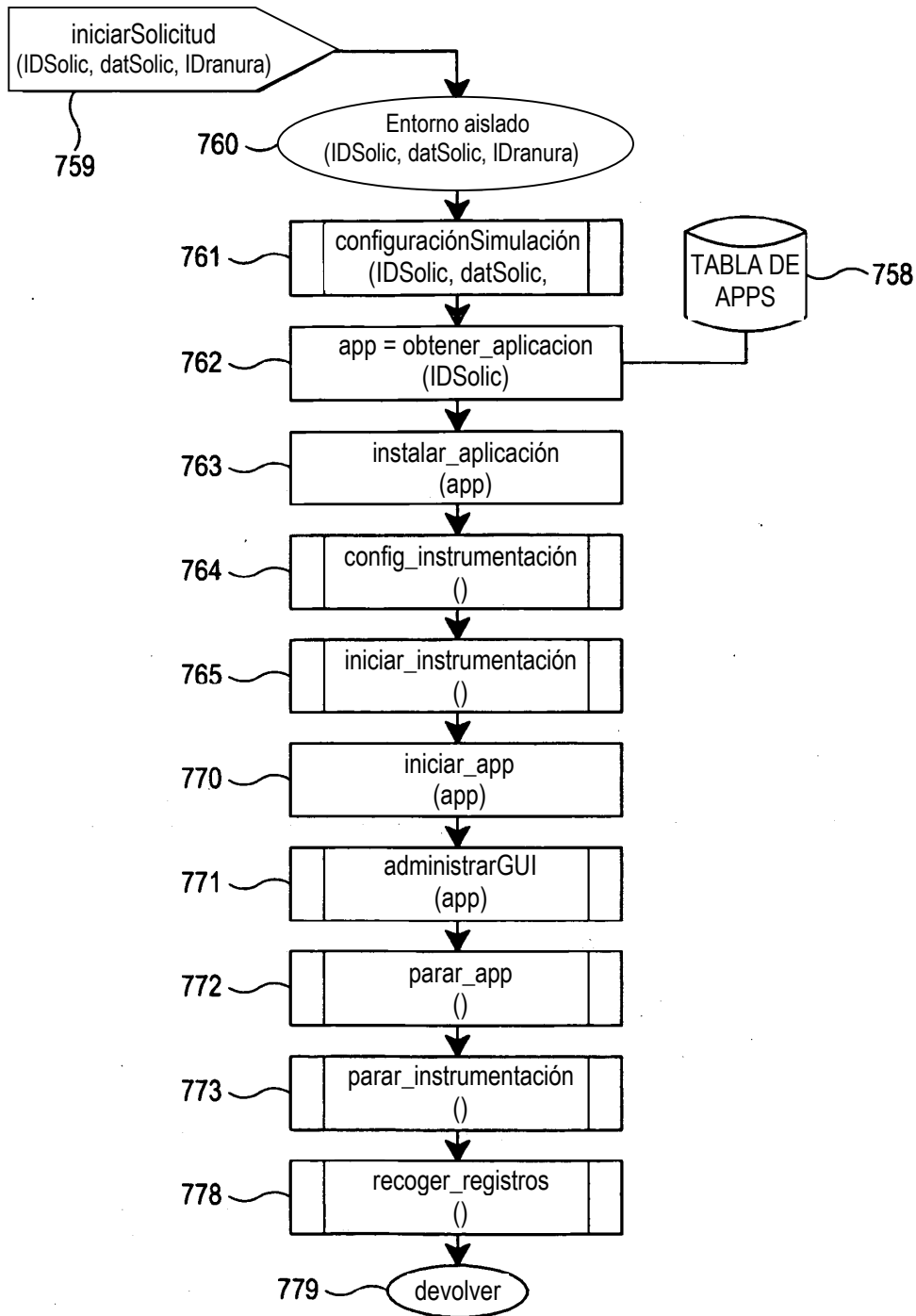


Figura 7B

Continúa en la Hoja 23/94

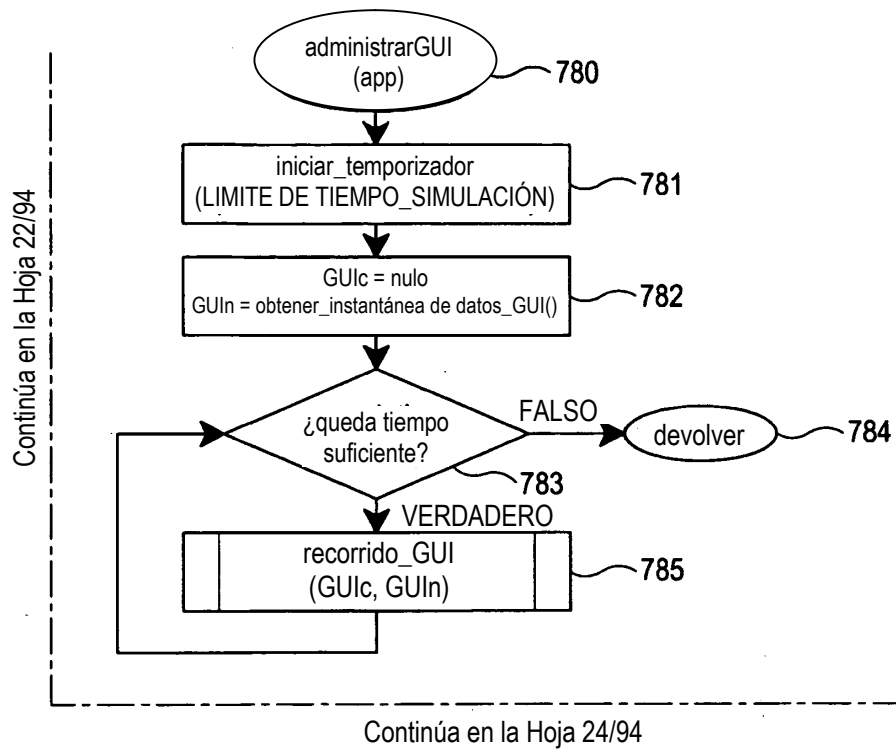


Figura 7B (continuación)

Continúa en la Hoja 23/94

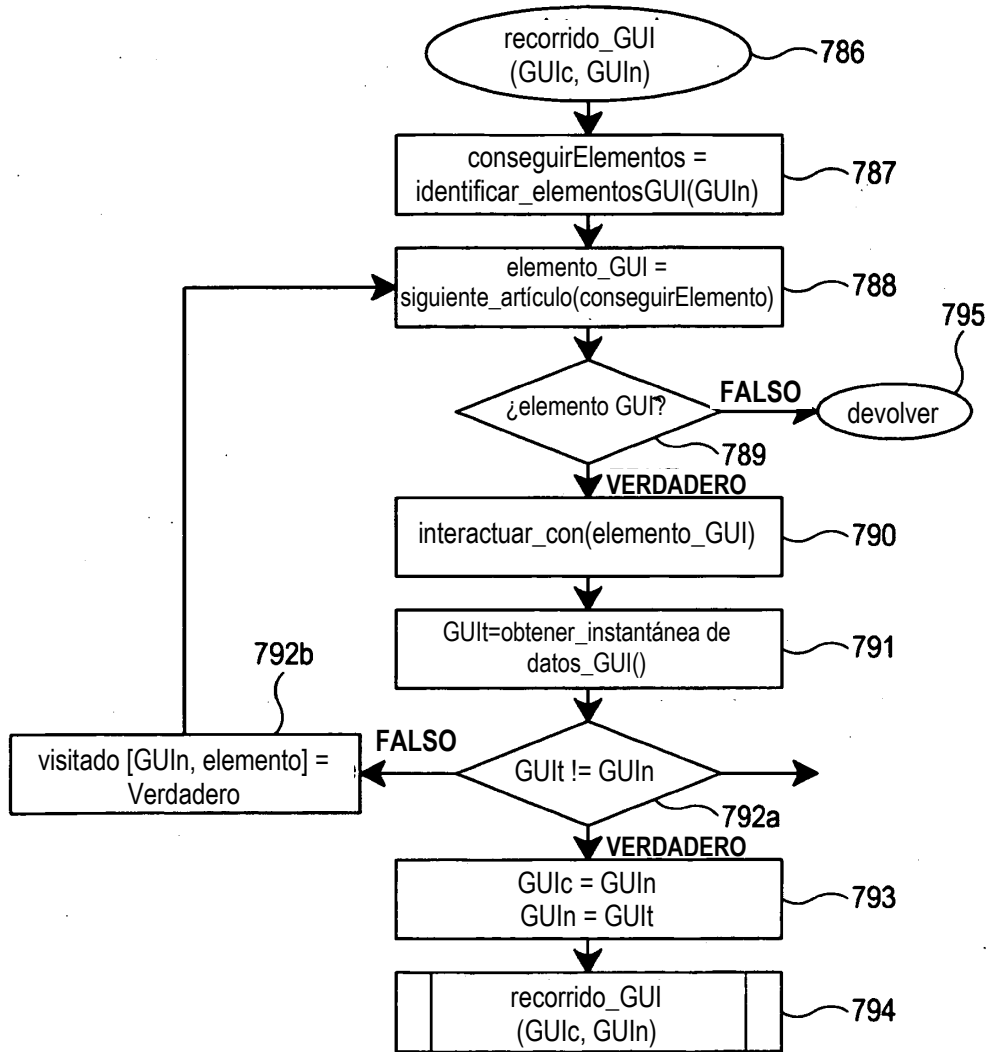


Figura 7B (continuación)

250	210,805	220,810	230,815	240,820	
801	ID_SOLIC	ID_DISPOSITIVO	ID_USUARIO	ID_APLICACIÓN	ID_RED
101	DROID1, SN1, ...	N@TTI.NET, ACC1, ...	YOUTUBE.APK, MD5, ...	172.30.0.1, GEOIP, ...	
102	DROID2, SN2, ...	X@ABC.NET, ACC2, ...	TANKHERO.APK, MD5, ...	65.30.0.1, GEOIP, ...	
103	DROID1, SN1, ...	N@TTI.NET, ACC1, ...	GOOGLEMAP.APK, MD5, ...	172.30.0.1, GEOIP, ...	

Figura 8A

250	830	835	840	40
IDSOLIC	MDS	NOMBRE ARCHIVO	NOMBRE PKG	DATOS DE APP
104	efedc2479dc34...	cellfire.apk	com.cellfire.android	0x0234ac200a...
105	75e843da4b6...	antibody2.apk	creafire.com.antibody2.k.lite	0x1ffa102e32....
106	20468ea02281d...	todolist.apk	com.android.it.todolist	0xff12aaa342a...

Figura 8B

250	170,860	321,855	852	851
ID SOLIC	DISTRIBUIDOR	ENTORNO AISLADO	MARCA DE TIEMPO	SO_DISPOSITIVO
104	DIRAC.TTITECH	5554	Lun 8 Ago 12:05:34	DROID1/SDK2.2
105	DIRAC.TTITECH	5556	Lun 8 Ago 12:05:55	DROID1/SDK2.2
106	HAHN.TTITECH	5560	Lun 8 Ago 12:07:01	DROID2/SDK2.3

Figura 8C

250	880	410	877
876	TIPO DE REISTRO	DATOS DE REGISTRO	MARCA DE TIEMPO
104	SEGUIMIENTO DE RED	0x34ffa234a244a...	Lun 8 Ago 12:15:34
104	LLAMADAS DE SISTEMA	escribirmem(), iocfl(1), leermem()...	Lun 8 Ago 12:15:45
105	SIMULACIÓN GUI	ClickVentana(), enviarClave(), CerrarVentana()	Lun 8 Ago 12:17:21

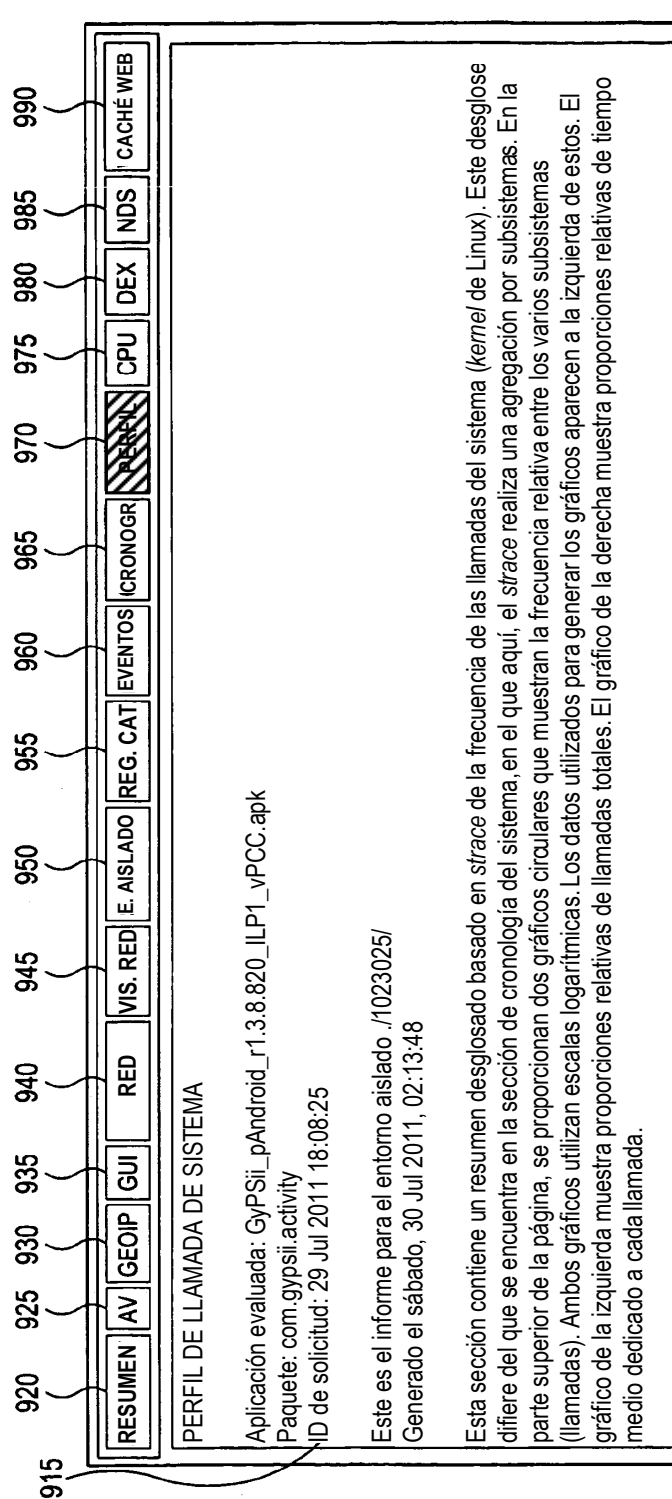
Figura 8D

250	895	896	897	898
IDSOLIC	ANÁLISIS	INFECTADO	DESCRIPCIÓN	MARCA DE TIEMPO
124	A	Verdadero	Troyano, envío de SMS clandestinos...	Mar 9 Ago 13:02:9
124	B	Verdadero	Marcador clandestino	Mar 9 Ago 13:02:9
155	A	Falso	No se encontró nada	Mar 9 Ago 14:02:5

Figura 8E

848	847	846	848
IDSOLIC	COMPONENTE	FASE	MARCA DE TIEMPO
104	CONTROLADOR	SOLIC_RECIBIDA	Lun 8 Ago 12:15:34
104	ESPACIO AISLADO	INSTAL_APP_CORRECTA	Lun 8 Ago 12:15:45
105	ESPACIO AISLADO	INSTAL_APP_CORRECTA	Lun 8 Ago 12:17:21

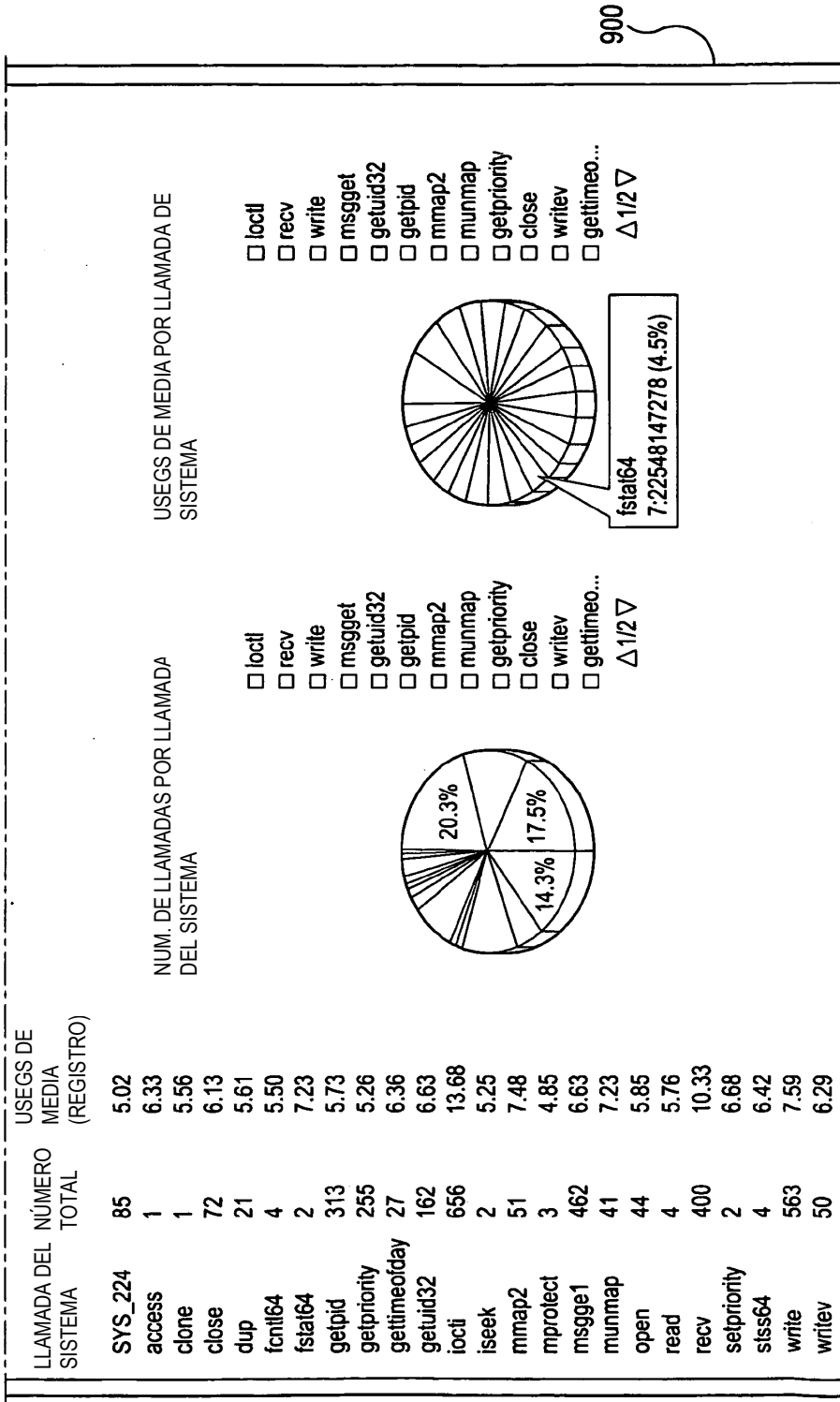
Figura 8F



Continúa en la Hoja 32/94

Figura 9


Continúa en la Hoja 31/94

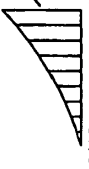


900

Figura 9 (continuación)

1000

 **marvin**
[SEGURIDAD MÓVIL] 1050

Consulta	#3019498	Evaluación de malware	1060
Fecha	2012-08-03 05:18 AM		10/10 [INFECTADO]
Nombre de archivo	00c154b42fd483196d303618582420b89cedbf46.apk		
Paquete	com.zhangling.danti275		
Dispositivo de análisis	Motorola Droid		
SO de Android	2.2		
MD5	f914b6b8f545f6cd78ec2e06b9796998		
Etiqueta de malware	[INFECTADO]		
Etiqueta de clúster	0 L2-NF 3019498		
Validez de registro	L2: all logs collected		

1005

Conc. básicos	Red	Antivirus	Marcas rojas	Automata de IU	Androguard	Sist. Archivos	Desarrollador	Copyright
---------------	-----	-----------	--------------	----------------	------------	----------------	---------------	-----------

1006

Resumen	Cap.	Pantalla	Esc.	Tiempo
Descripción básica de la aplicación, p. ej., su firma MD5, el nombre del paquete, etc.				
NOM. ARCHIVO / /3019498/ /00c154b42fd483196d303618582420b89cedbf46.apk				
PAQUETE com.zhangling.danti275				
MD5 f914b6b8f545f6cd78ec2e06b9796998				
ES VÁLIDO Sí				
CÓDIGO DE LA VERSIÓN 19				

Continúa en la hoja 34-94

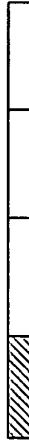
Figura 10

Continúa en la Hoja 33/94

NOMBRE DE LA VERSIÓN 3.0.1	
SDK MIN	Ninguno
SDK OBETIVO	Ninguno
Detección de intrusiones (filtraciones)	Se detectaron filtraciones de dabs del dispositivo
Detección de intrusiones (malware)	Se detectaron firmas de <i>malware</i>
Evaluación de <i>malware</i>	El análisis obtuvo una clasificación de <i>malware</i> inusualmente alta
Perfil de riesgo	*La aplicación com.zhanging.danti275 (MD5: f914b6b8f545f6cd78ec2e96b9796998 se comparó con miles de otras aplicaciones con respecto a marcas rojas conocidas de análisis estáticos de Marvin (a fecha del 2 de julio de 2012). La aplicación manifiesta unos niveles de exposición al riesgo ANORMALMENTE ALTO en las siguientes categorías de marca roja: APP INFECTADA, TUS SMS, TUS ARCHIVOS La aplicación manifiesta unos niveles de exposición al riesgo ANORMALMENTE ALTOS en categorías muy críticas. La aplicación se registra como INFECTADA . Tenga cuidado.
Transacción TGP	Se reiniciaron algunas conexiones TCP

Continúa en la Hoja 35/94

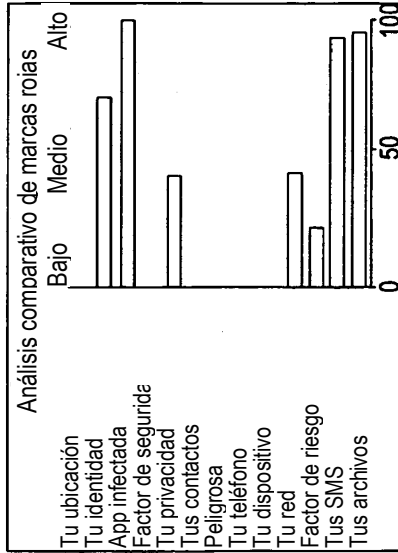
Figura 10 (continuación)



1010

Marvin examina los datos de tu aplicación frente a otras miles para evaluar mejor el riesgo asociado con la instalación y ejecución de la aplicación. El siguiente gráfico indica cómo se compara el perfil de riesgo de tu aplicación con el de otras miles aplicaciones.

Perfil de riesgo:



*La aplicación com.zhangling.danti275 (MD5: f914b6b8f545f6cd78ec2e9669796998 se comparó con miles de otras aplicaciones con respecto a marcas rojas conocidas de análisis estáticos de Marvin (a fecha del 2 de julio de 2012). La aplicación manifiesta unos niveles de exposición al riesgo ANORMALMENTE ALTO en las siguientes categorías de marca roja: APP INFECTADA, TUS SMS, TUS ARCHIVOS. La aplicación manifiesta unos niveles de exposición al riesgo ANORMALMENTE ALTOS en categorías muy críticas. La aplicación se registra como **INFECTADA**. Tenga cuidado.

Figura 10 (continuación)

Continúa en la Hoja 35/94

1020

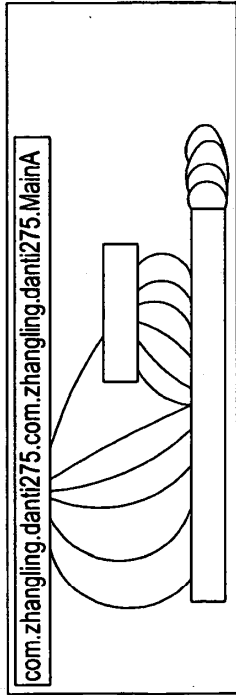
Conc. básicos	Red	Antivirus	Marcas rojas	Autómata de IU	Androguard	Sist. Archivos	Desarrollador	Copyright
<p>Ventana emergente</p> <p>Esta sección describe el recorrido del autómata de IU sobre la aplicación y documenta las actividades de IU, ventanas y transiciones que se descubrieron durante la interacción con la aplicación. El autómata de IU intenta hasta tres tipos distintos de recorrido en una aplicación, siempre que quede suficiente tiempo para la ejecución. Los resultados de cada recorrido están contenidos en un encabezado de subsección. «Ejecución de GUI #».</p> <p>Esta sección permite a los usuarios entender cómo se ejecutó la IU de la aplicación dentro del entorno aislado así como a entender los defectos y/o problemas de usabilidad con la IU de la aplicación determinada.</p> <p>Ejecución de GUI 1</p> <p>-----</p> <p>Nombre del paquete: com.zhangling.danti275 Actividad de raíz [app]: com.zhangling.danti275.mainA Ventana original: com.zhangling.danti275.com.zhangling.danti275.mainA Ventanas totales encontradas: 4</p> <p>-----</p> <p>Modo: interacción de componente Comienzo del análisis: 06-03 06:26:33 Finalización del análisis: 06-03 06:29:11</p> <p>-----</p>								

Continúa en la Hoja 37/94

Figura 10 (continuación)

Continúa en la Hoja 36/94

Número de reinicios: 0



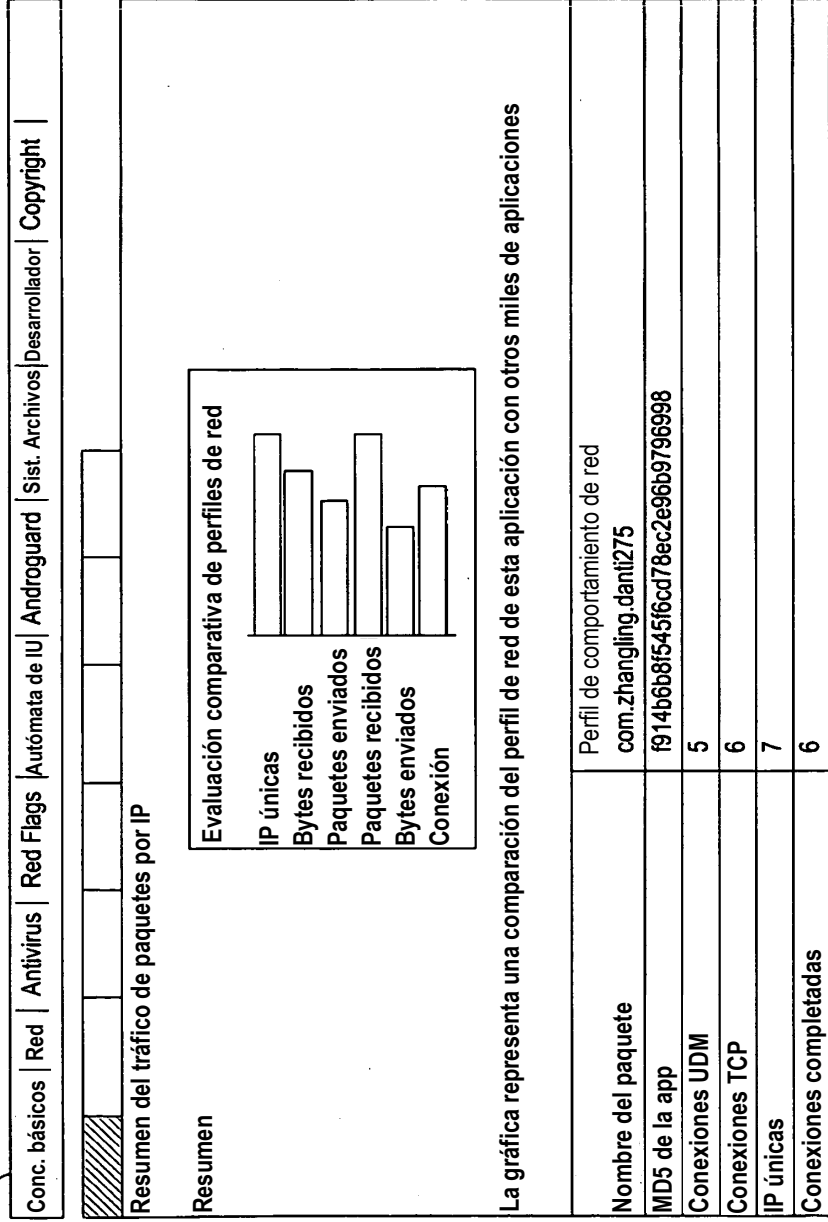
ID Borde	Nombre de clase	Etiquetas de borde			Fecha y hora
		ID	Texto		
A1	android.widget.Button	id/button3		06-03 05:28:55	
B1	android.widget.TextView	id/message		06-03 05:28:39	
C1	android.widget.ImageView	id/icon	Ninguno	06-03 05:26:44	
D1	android.widget.Button	id/button2		06-03 05:29:03	
E1	android.widget.ImageView	id/titleDivider	Ninguno	06-03 05:28:31	
F1	android.widget.Button	id/button1		06-03 05:28:47	

Continúa en la Hoja 38/94

Figura 10 (continuación)

Continúa en la Hoja 37/94

1030



Continúa en la Hoja 39/94

Figura 10 (continuación)

Continúa en la Hoja 38/94

	Transmitidos	Recibidos
Bytes (kB)	5	45
Paquetes	130	26

Conexiones salientes

Volumen de tráfico saliente por IP

*Representa la cantidad de datos enviados en cada IP en bytes.

Conexiones salientes		
Puerto origen	IP destino	Tipo
	225.5.155.18	TCP
		6

Continúa en la Hoja 40/94

Figura 10 (continuación)

1040

Conc. básico	Red	Antivirus	Marcas rojasAutomata de IU	Androguard	Sist. Archivos	Desarrollador	Copyright
--------------	-----	-----------	----------------------------	------------	----------------	---------------	-----------

Resumen de recuento de archivos

***PERFIL DE CAMBIO DE SISTEMA DE ARCHIVOS**

MEDIA INUSUAL ANORMAL

Categoría	Frecuencia
0	0
1	2
2	2
3	1
4	1
5	2
6	2
7	2
8	2
9	2
10	2

*La aplicación com.zhangling.danti275 (MD5: f914b6b8f545f6cd78ecd78ec2e96b9796998) se comparó con otras aplicaciones con respecto a la intensidad de los cambios del sistema de archivos. Las estadísticas sobre la modificación del sistema inducida por la ejecución de esta aplicación se compararon con las de otras aplicaciones. Marvin encontró lo siguiente: No se observó ninguna desviación significativa.

Creados	14
Borrados	5
Modificados	26

Figura 10 (continuación)

Continúa en la Hoja 42/94

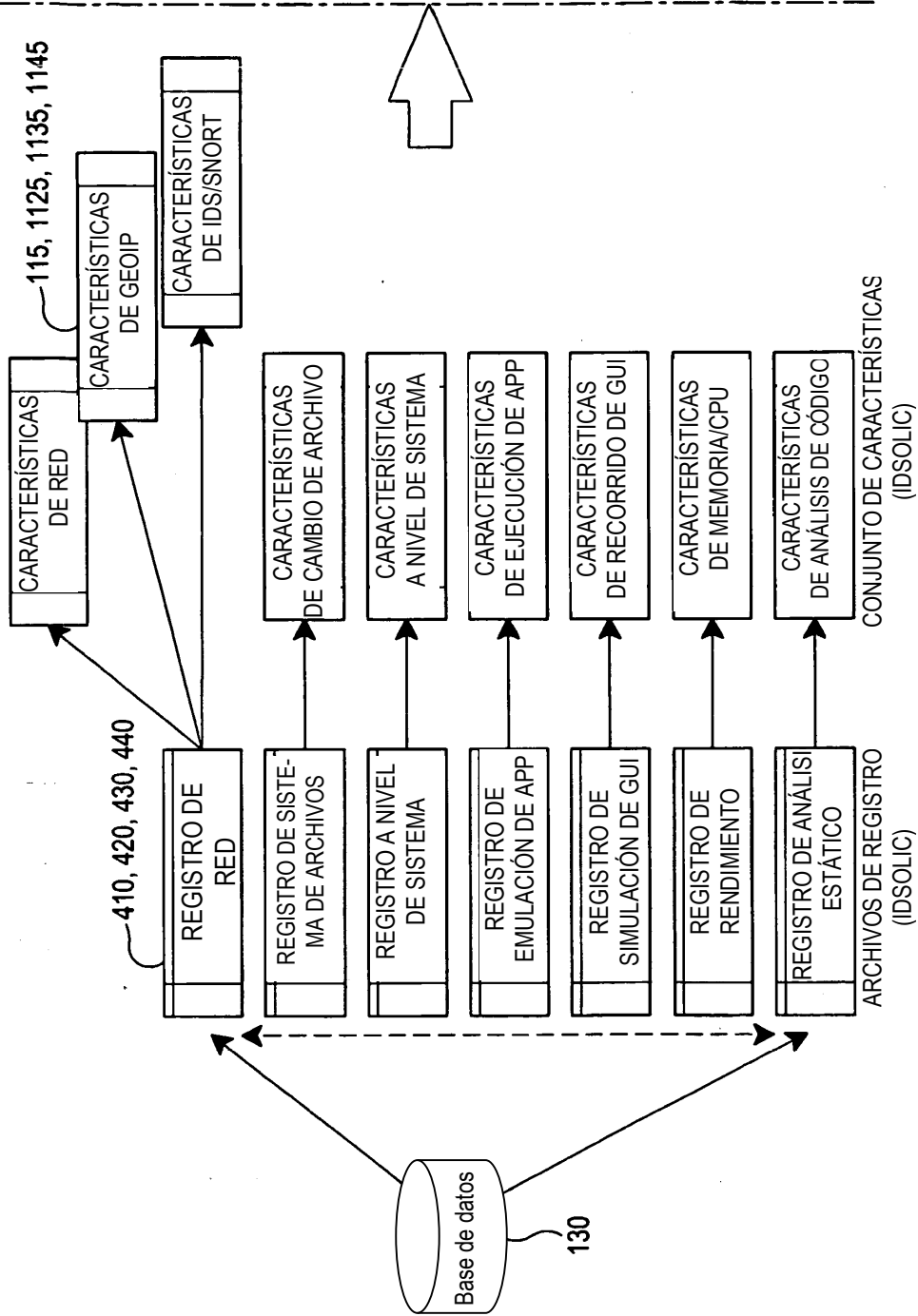


Figura 11A

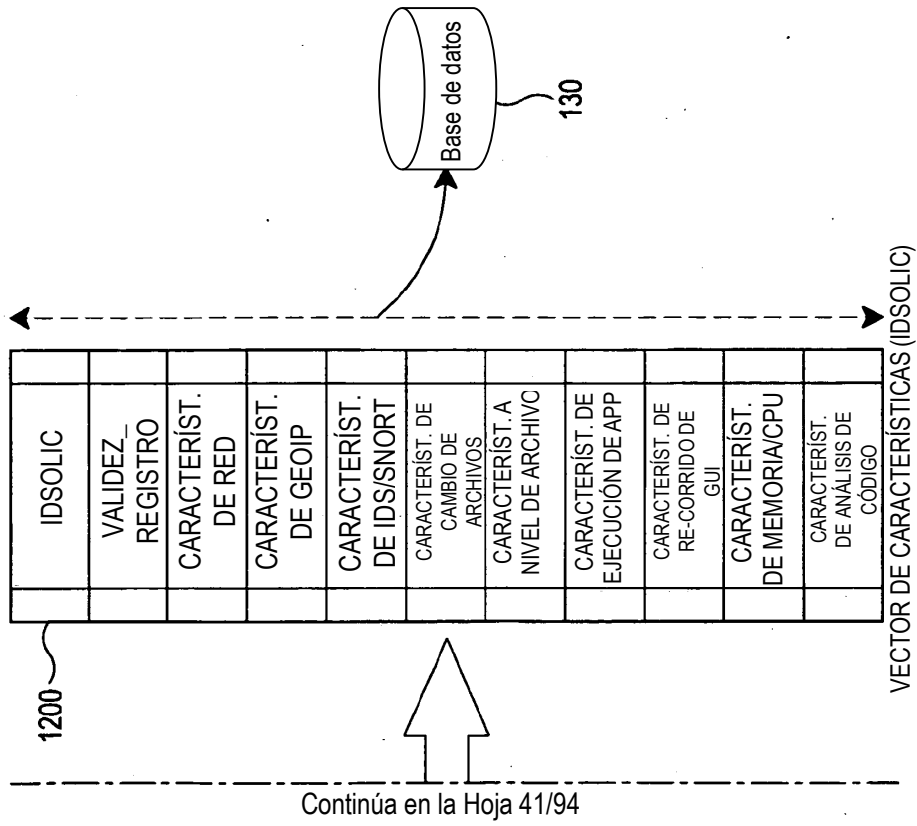


Figura 11A (continuación)

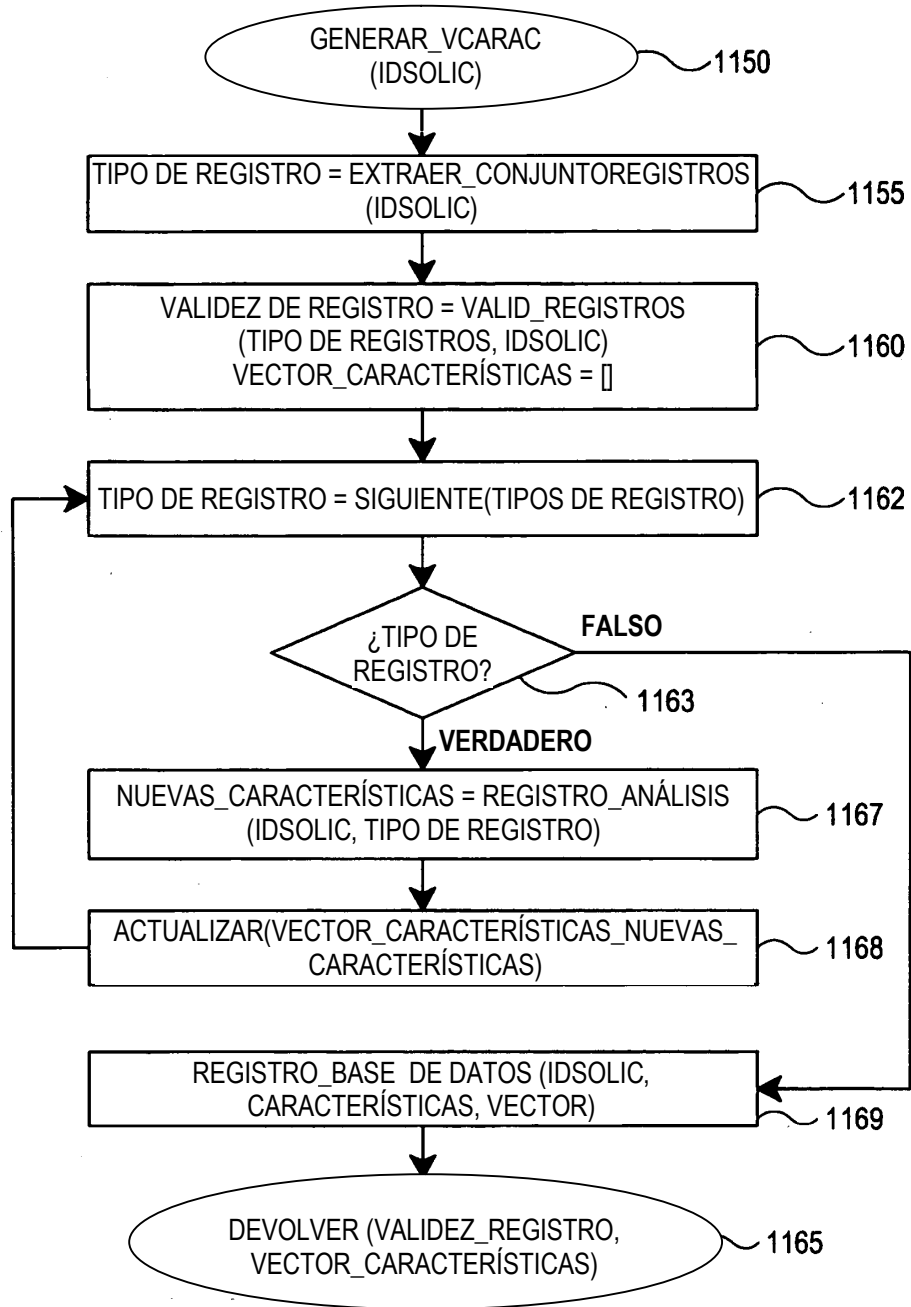


Figura 11B

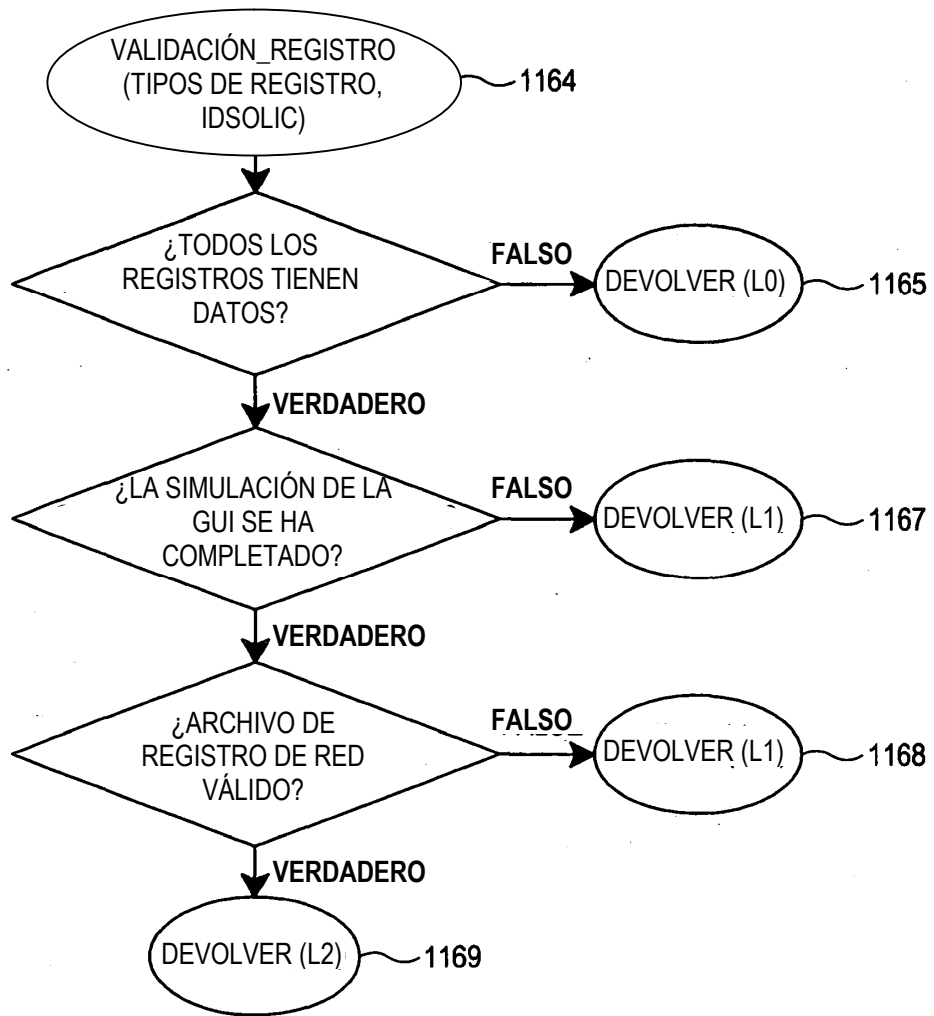


Figura 11C

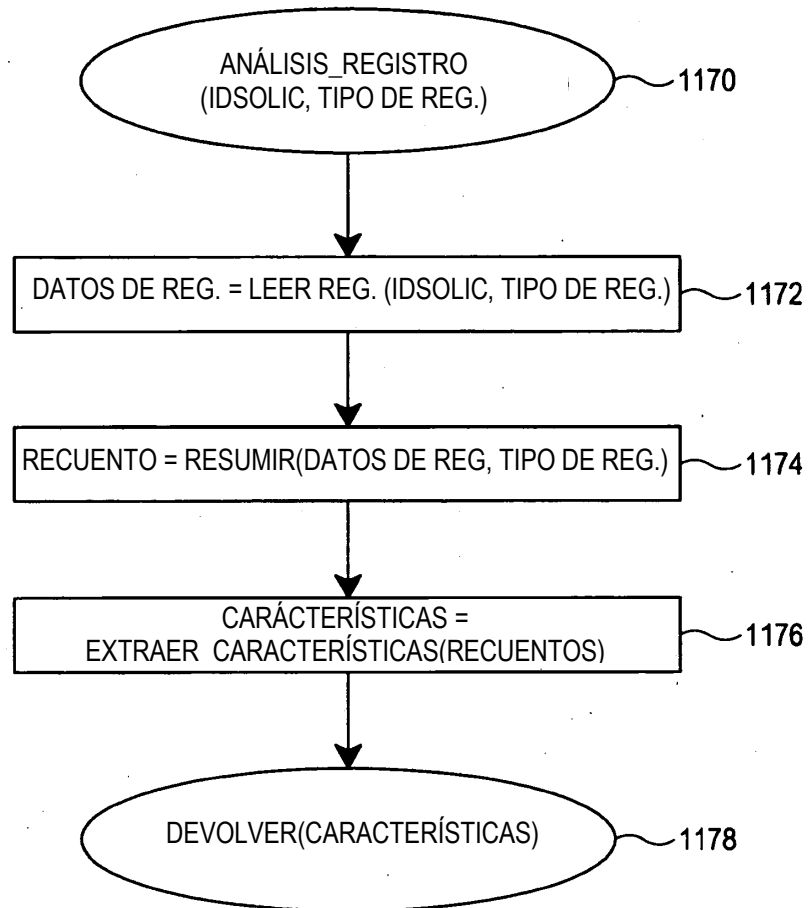


Figura 11D

Continúa en la Hoja 47/94

<p>Conexión TCP 1: Serv. principal a: 10.0.2.15.43599 Serv. principal b: 216.156.194.121:80 conex. complet.: no (SYNs: 0) (FINs: 2) primer paquete: Vie 29 jul 19:53:07.099224.2011 último paquete: Vie 29 jul 19:56:55. tiempo transcurr: 0:03:48.257037 paquetes totales: 28</p>		<p>Conexión TCP 6: Serv. principal q: 10.0.2.15.60939 Serv. principal r: 65.121.209.90.80 conex. complet.: Si primer paquete: Vie 29 jul 20:00:47.609438.2011 último paquete: Vie 29 jul 20:02:43.083234.2011 tiempo transcurr: 0:01:55.473795 paquetes totales: 13</p>	
<p>a -> b:</p>	<p>b -> a:</p>	<p>q -> r:</p>	<p>r -> q:</p>
<p>paquetes totales: 14</p>	<p>paquetes totales: 7</p>	<p>paquetes totales: 6</p>	<p>paquetes totales: 6</p>
<p>paq. recon. enviad.: 14</p>	<p>paq. recon. enviad.: 6</p>	<p>paq. recon. enviad.: 6</p>	<p>paq. recon. enviad.: 6</p>
<p>bytes únicos env.: 1899</p>	<p>bytes únicos env.: 475</p>	<p>bytes únicos env.: 286</p>	<p>bytes únicos env.: 286</p>
<p>paq. datos reales: 8</p>	<p>paq. datos reales: 2</p>	<p>paq. datos reales: 1</p>	<p>paq. datos reales: 1</p>
<p>bytes de dat. real.: 1899</p>	<p>bytes de dat. real.: 475</p>	<p>bytes de dat. real.: 286</p>	<p>bytes de dat. real.: 286</p>
<p>paq. desordenad.: 0</p>	<p>paq. desordenad.: 0</p>	<p>paq. desordenad.: 0</p>	<p>paq. desordenad.: 0</p>
<p>paq. datos insert.: 8</p>	<p>paq. datos insert.: 2</p>	<p>paq. datos insert.: 1</p>	<p>paq. datos insert.: 1</p>
<p>paq. SYN/FIN env.: 0/1</p>	<p>paq. SYN/FIN env.: 1/1</p>	<p>paq. SYN/FIN env.: 1/1</p>	<p>paq. SYN/FIN env.: 1/1</p>
<p>n.ºs solicitados: 0 bytes</p>	<p>n.ºs solicitados: 1460 bytes</p>	<p>n.ºs solicitados: 1460 bytes</p>	<p>n.ºs solicitados: 1460 bytes</p>
<p>tiempo inactiv.: 60202.4 ms</p>	<p>tiempo inactiv.: 115341.7 ms</p>	<p>tiempo inactiv.: 115342.1 ms</p>	<p>tiempo inactiv.: 115342.1 ms</p>
<p>capac. proceso: 8 Bps</p>	<p>capac. proceso: 4 Bps</p>	<p>capac. proceso: 2 Bps</p>	<p>capac. proceso: 2 Bps</p>
<p>muestras RTT: 9</p>	<p>muestras RTT: 4</p>	<p>muestras RTT: 3</p>	<p>muestras RTT: 3</p>
<p>media RTT: 0.2 ms</p>	<p>media RTT: 6.6 ms</p>	<p>media RTT: 0.4 ms</p>	<p>media RTT: 0.4 ms</p>
<p>devest RTT: 0.1 ms</p>	<p>devest RTT: 13.0 ms</p>	<p>devest RTT: 0.2 ms</p>	<p>devest RTT: 0.2 ms</p>
<p>1111</p>	<p>1112</p>		

Continúa en la Hoja 48/94

Figura 12

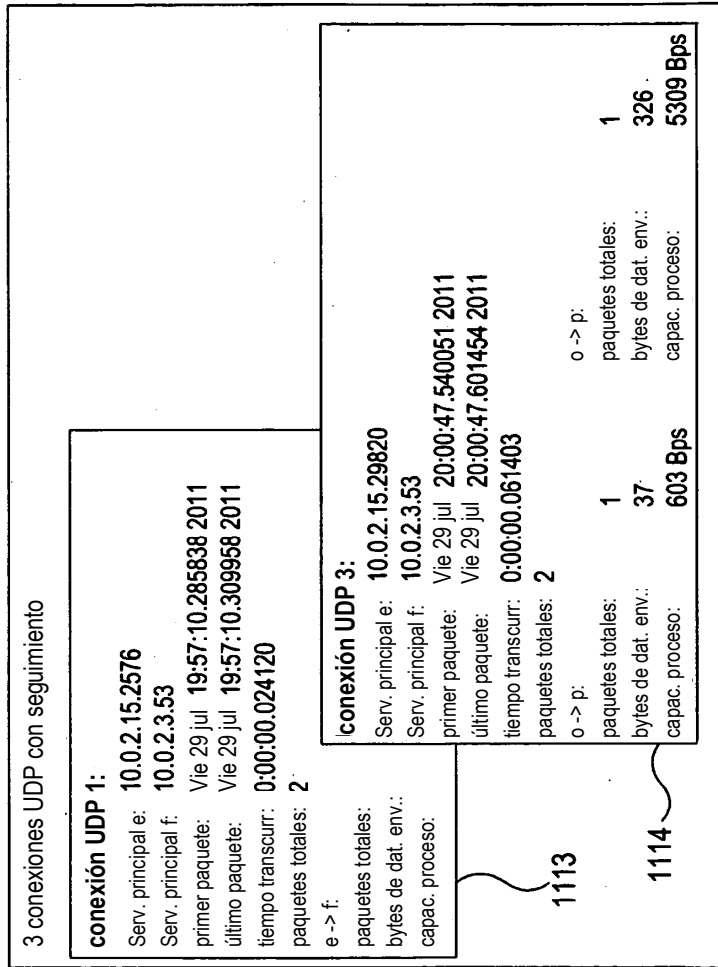


Figura 12 (continuación)

Continúa en la Hoja 46/94

Continúa en la Hoja 49/94

Continúa en la Hoja 46/94

1116												
IDSolic	numc	tiempo c	tipo c	de ip	de puerto	a ip	a puerto	bytes enviados	bytes recibid.	paq. enviados	paq. recibid.	rtin. enviad.
1023175	1	0:00:00.024120	UDP	10.0.2.15	12576	10.0.2.3	53	37	326	1	1	0.0
1023175	6	0:01:55.473795	TCP	10.0.2.15	60939	65.121.209.90	80	475	286	7	6	6.6
1023175	5	0:00:25.835322	TCP	10.0.2.15	45933	70.32.132.54	80	168	2824	7	6	33.3
1023175	4	0:02:25.916593	TCP	10.0.2.15	39432	70.32.132.54	80	168	2824	7	6	34.1
1023175	1	0:03:48.257037	TCP	10.0.2.15	43590	216.156.194.121	80	1899	1144	14	14	0.2

1117												
IDSolic	fecha_red	infectado	infección	nbyte env.	nbyte recib.	npaq env.	npaq recib.	nconex udp	nconex tcp	num_compl	IP unic.	
1023175	0:09:35.984010	0		4722	7405	102	6	3	6	4	6	
1023174	0:03:13.392808	0		5241	108921	252	20	10	13	5	6	
1023171	0:00:36.798329	0		3820	17997	115	14	7	11	5	8	
1023169	0:18:11.698288	0		66142	11415456	17741	10	5	25	6	7	
1023167	0:06:40.518065	0		7048	26555	114	4	2	3	2	4	
1023166	0:06:53.225792	0		3099	2432	52	6	3	7	0	6	
1023161	0:07:42.957626	0		72387	326792	1310	30	15	73	65	17	
1023159	0:12:22.401086	0		1107	16091	88	4	2	8	5	3	
1023158	0:09:13.616603	0		123001	3393506	5127	60	30	89	84	20	
1023153	0:00:00.000000	0		0	0	0	0	0	0	0	0	

Figura 12 (continuación)

rttin_recib.	thrup_env.	thrup_recib
8.8	1534	13516
8.4	4	2
8.4	7	78
9.1	1	14
8.3	8	5

Continúa en la Hoja 48/94

Figura 12 (continuación)

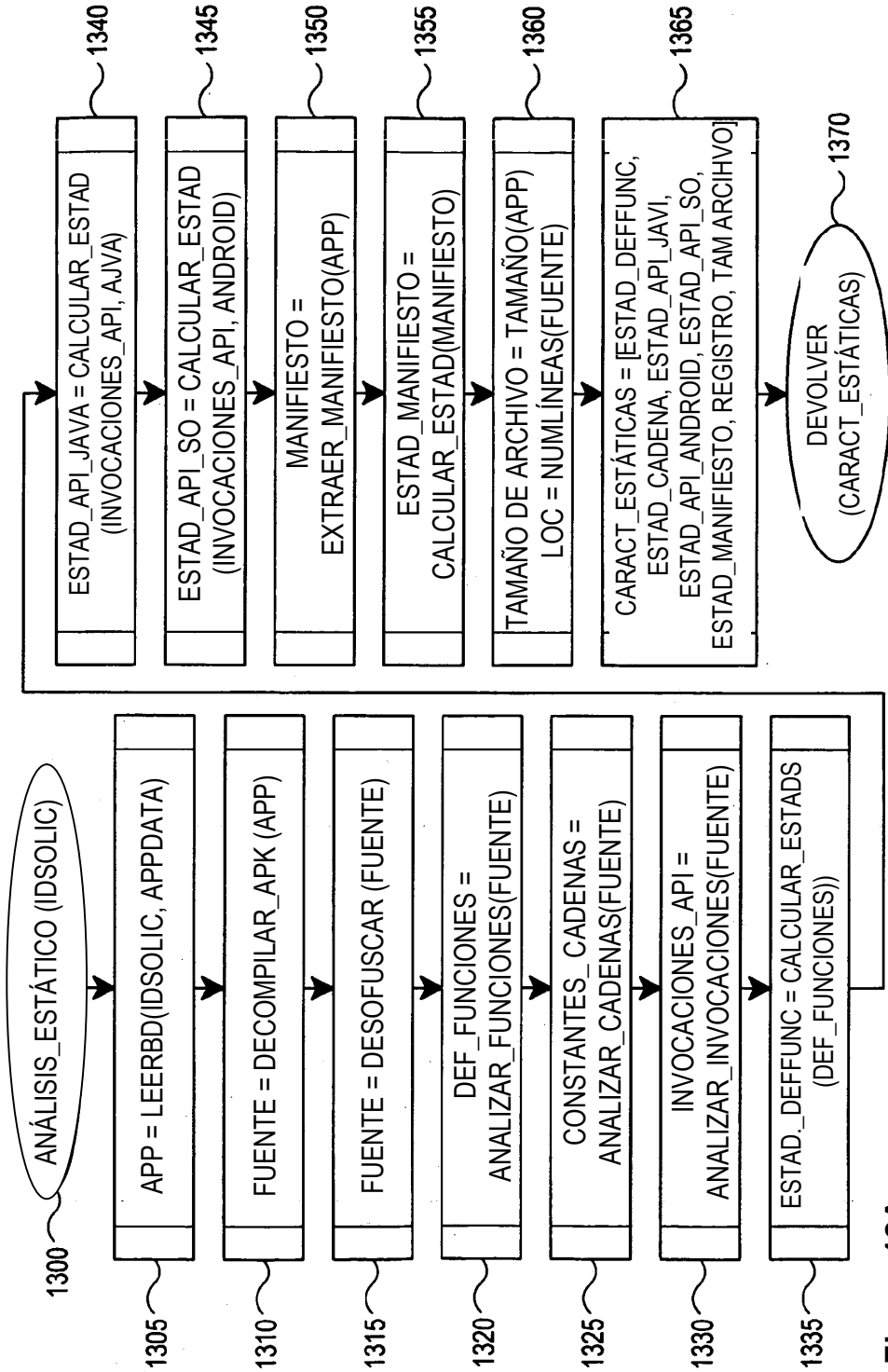


Figura 12A

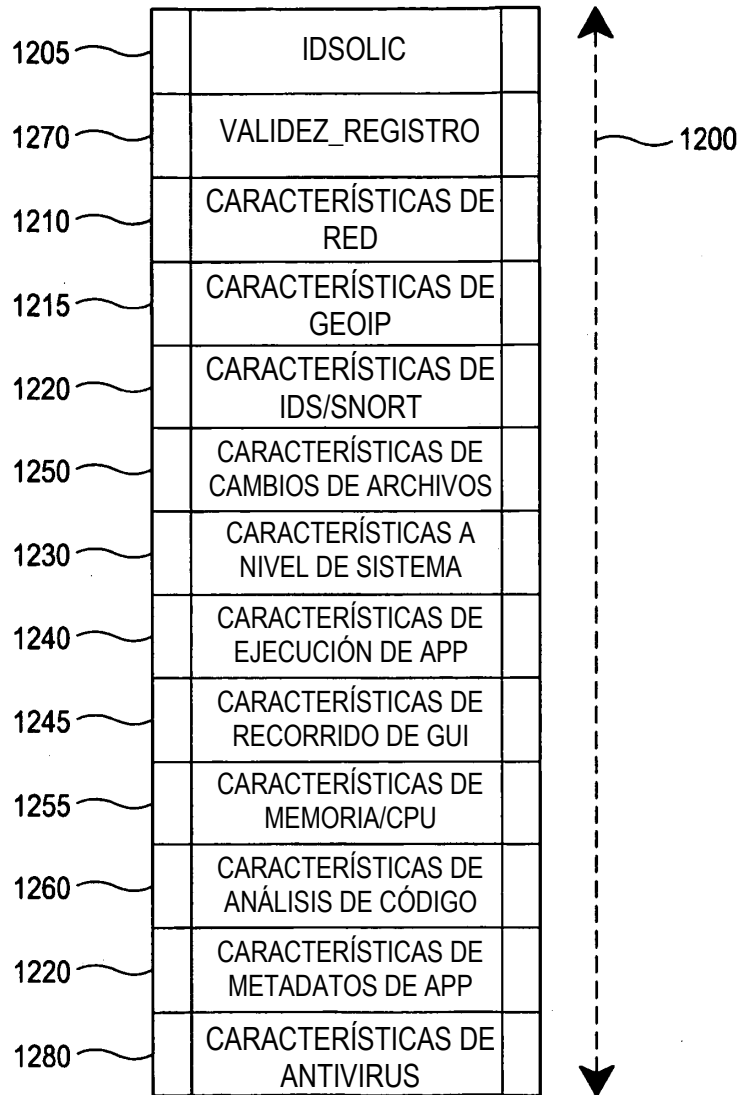


Figura 13

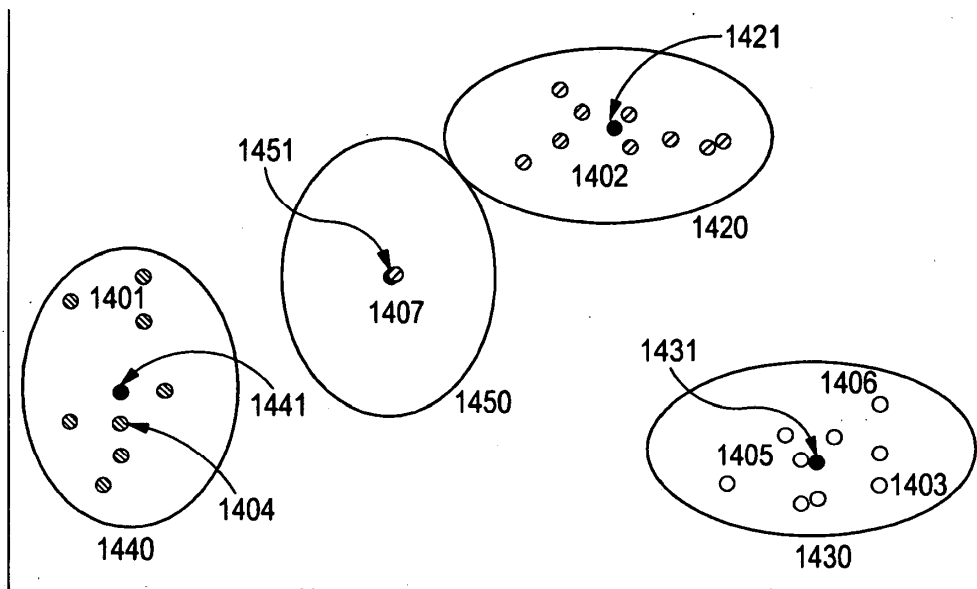


Figura 14

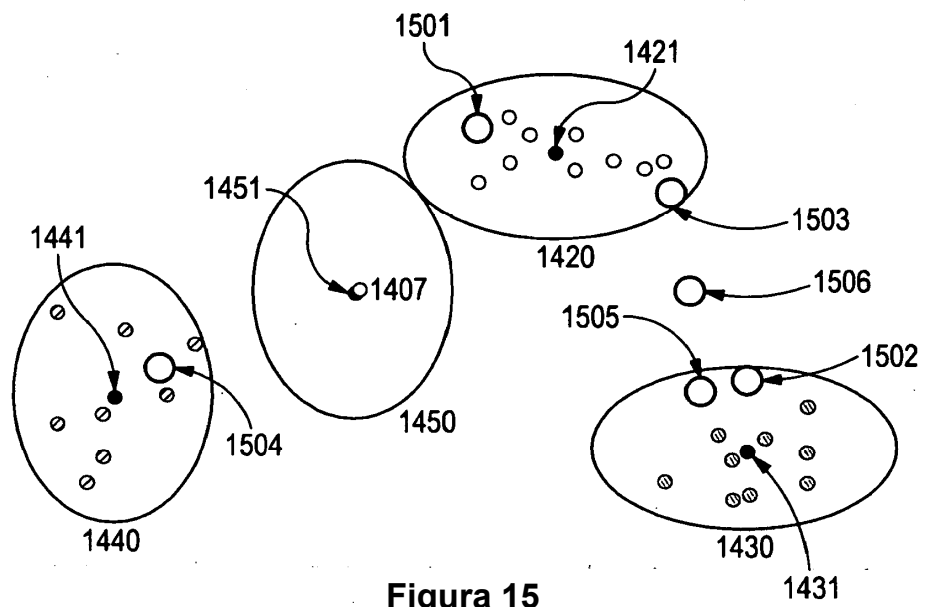


Figura 15

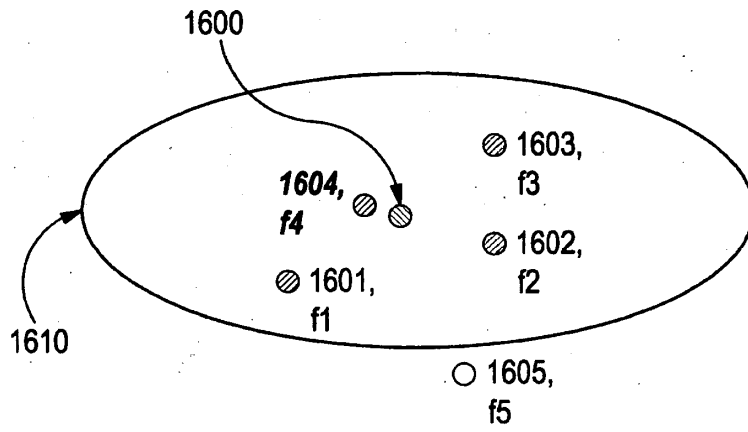


Figura 16A

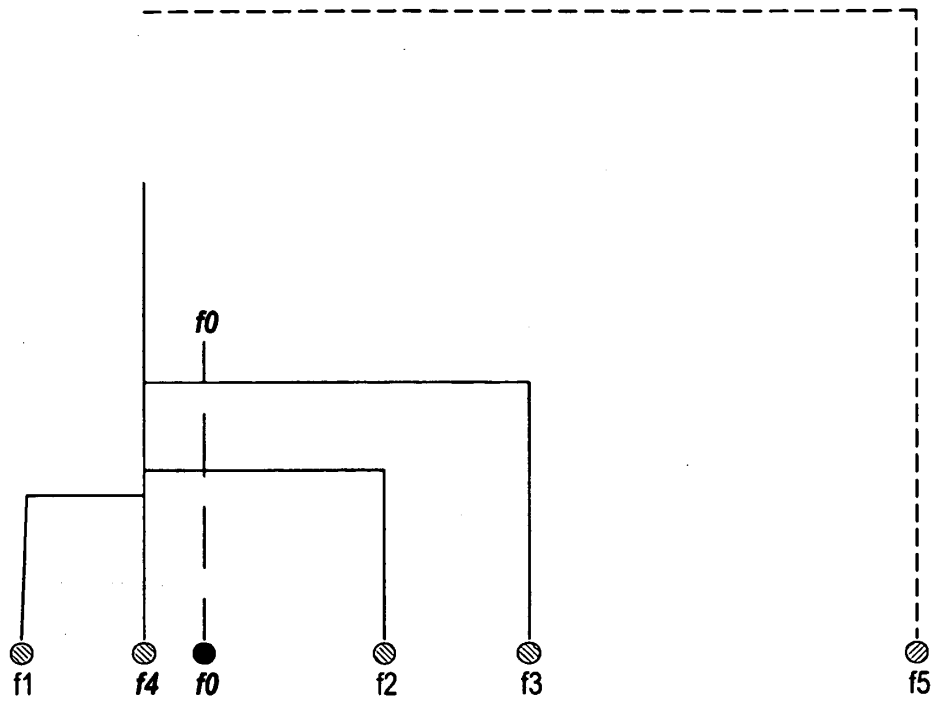


Figura 16B

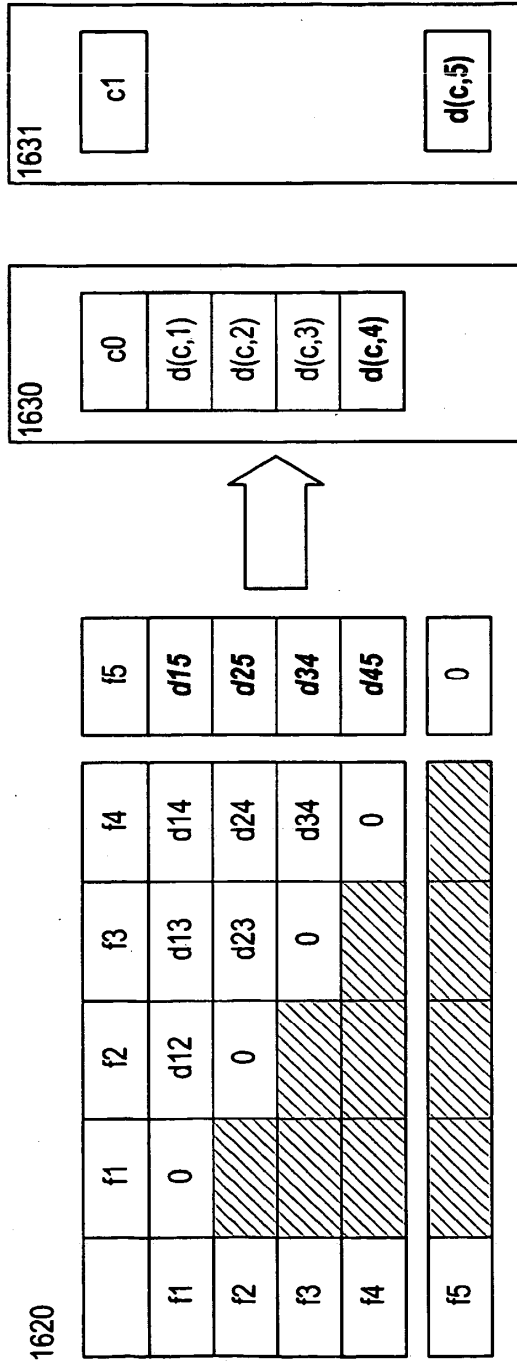


Figura 16C

IDSolic	DATOS_VCARACT	ETIQUETA_CLÚSTER	CENTROIDE	ANOMALÍA	MARCA DE TIEMPO
1204	REP_APP1	NO	NO	Lun 8 ago 12:15:34
1205	REP_APP3	SÍ	NO	Lun 8 ago 12:15:45
1206	NINGUNO	NA	SÍ	Lun 8 ago 12:17:21

250

1710

1720

Figura 17

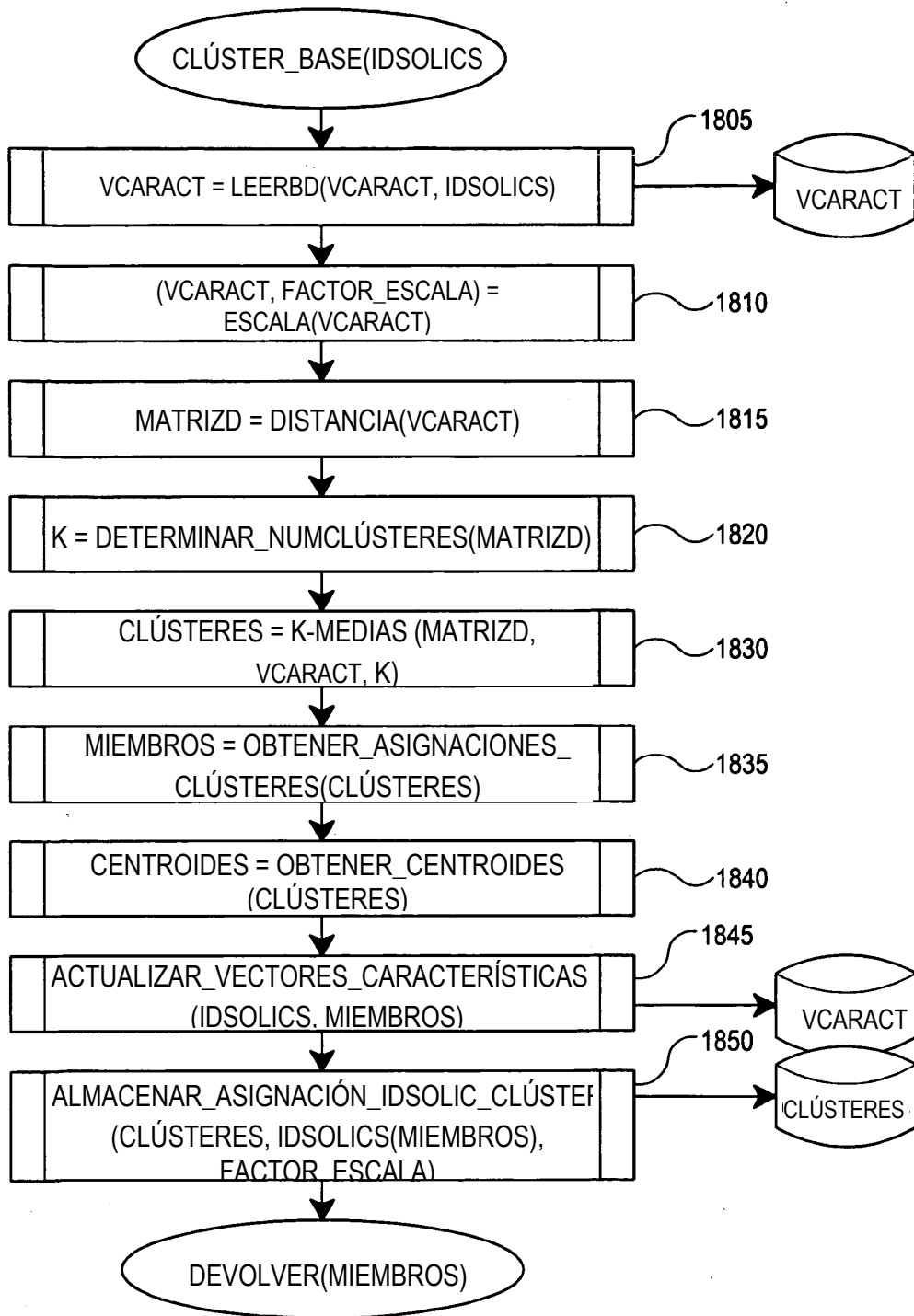
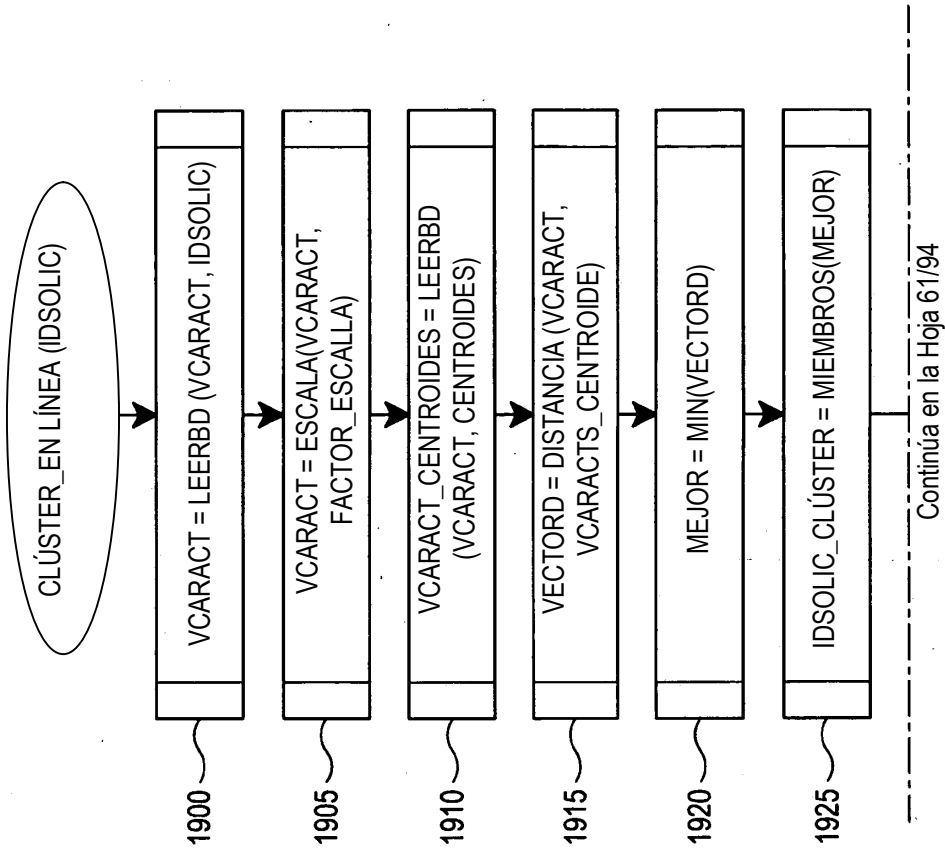


Figura 18

CLÚSTERES

ETIQ. CLÚSTER	TAMAÑO	CREADO EL	MIEMBRO	CENTROIDE	ÚLTIM. MODIFIC
WINTERGAMES. APK		Dom 7 Ago 01:00:12	NO	NO	Lun 8 Ago 12:15:34
YOUTUBE.APK	Dom 7 Ago 01:00:12	SÍ	NO	Lun 8 Ago 12:15:45
TANKHERE. APK	Dom 7 Ago 01:00:34	N	SÍ	Dom 7 Ago 01:00:34

Figura 18A



Continúa en la Hoja 61/94

Figura 19

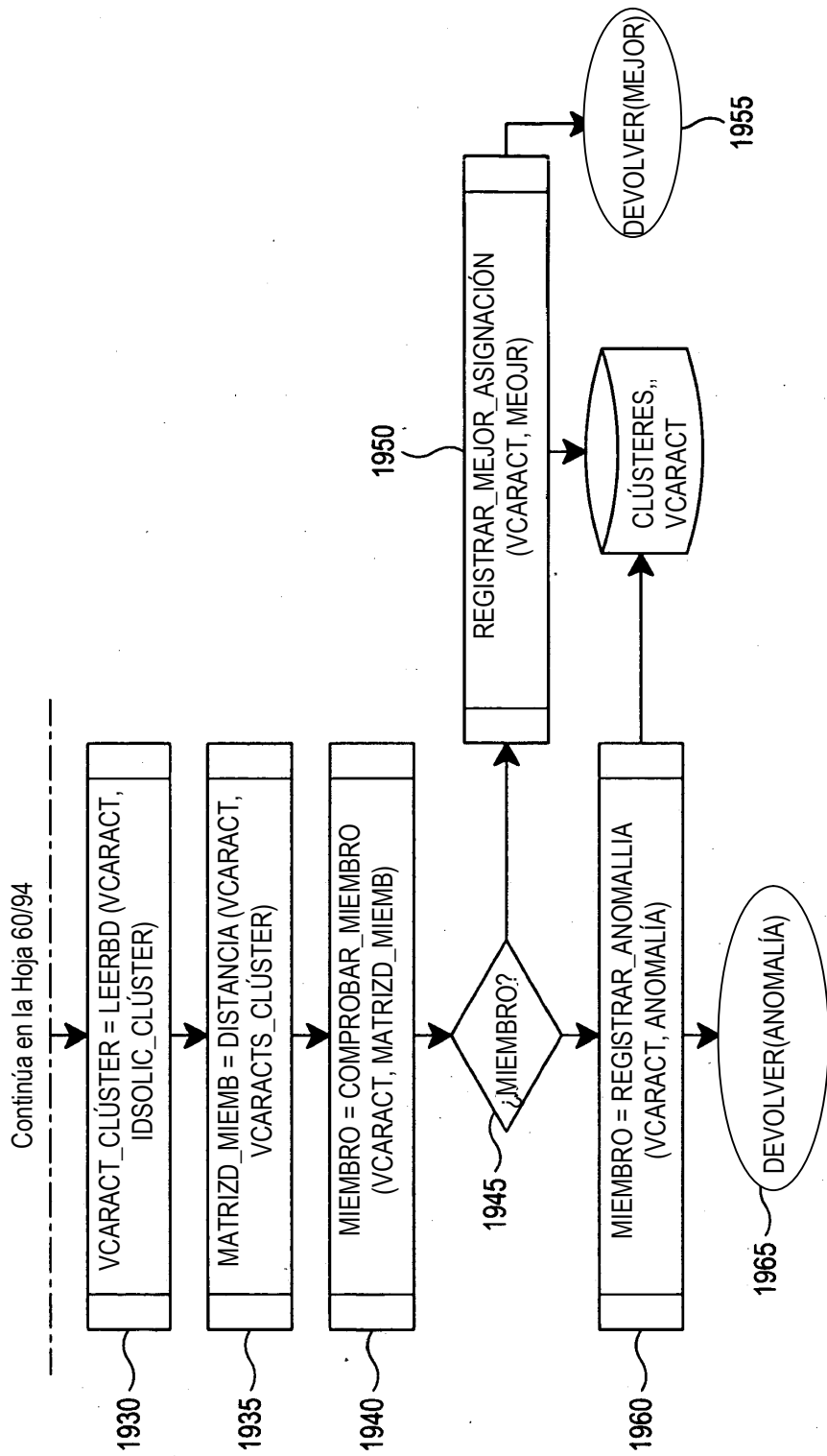


Figura 19 (continuación)

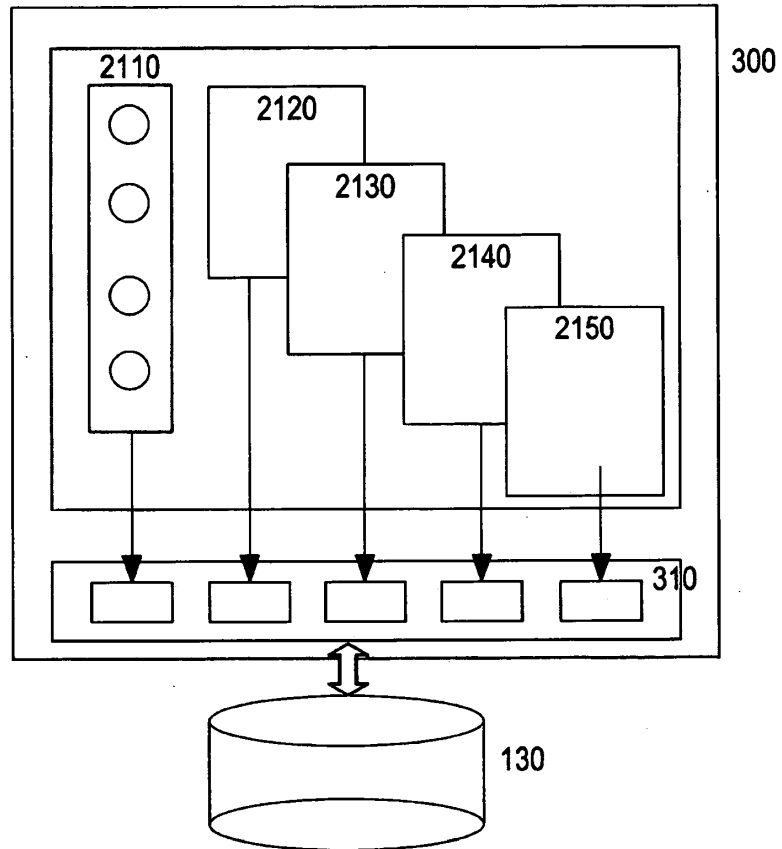
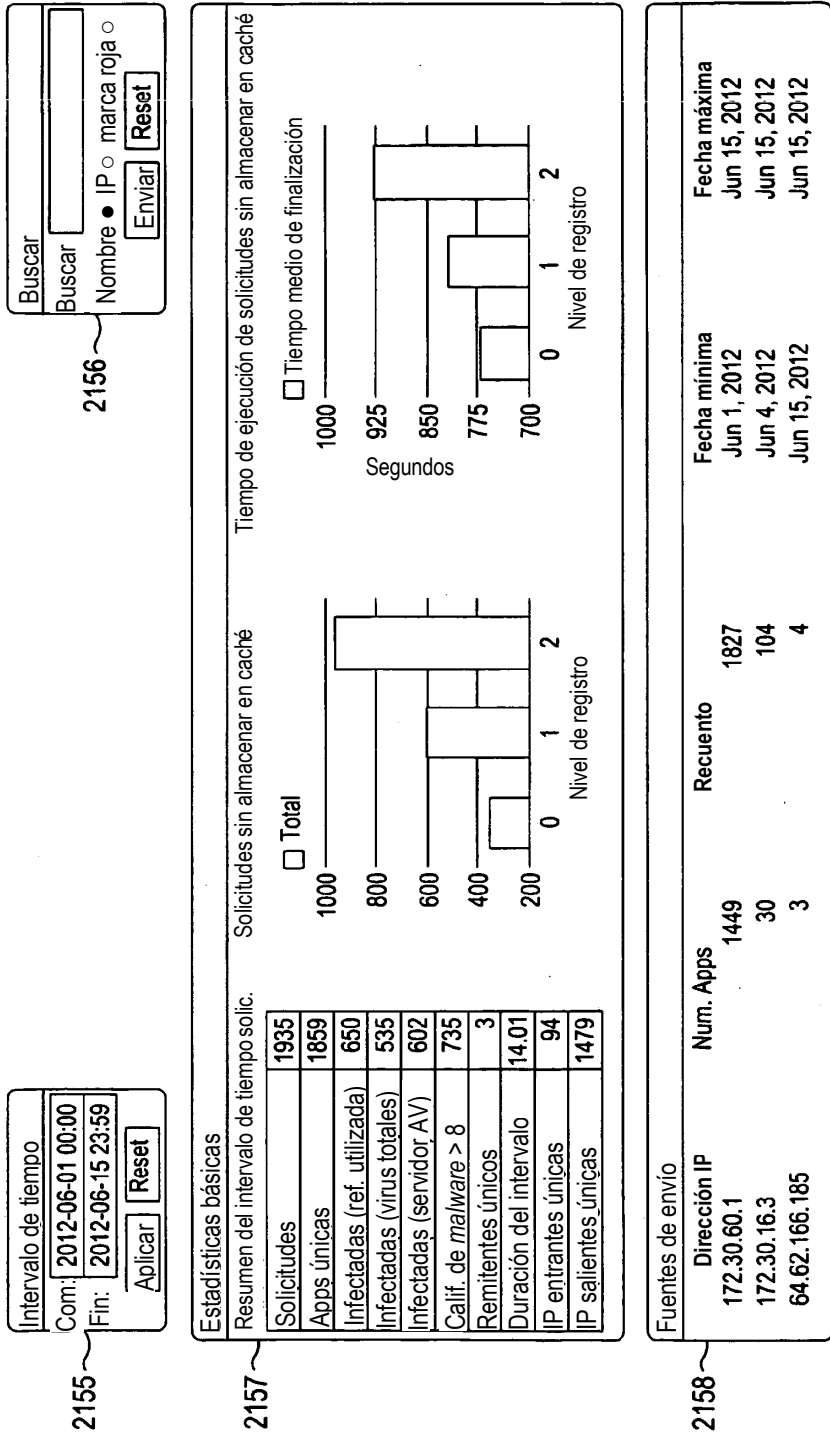


Figura 21



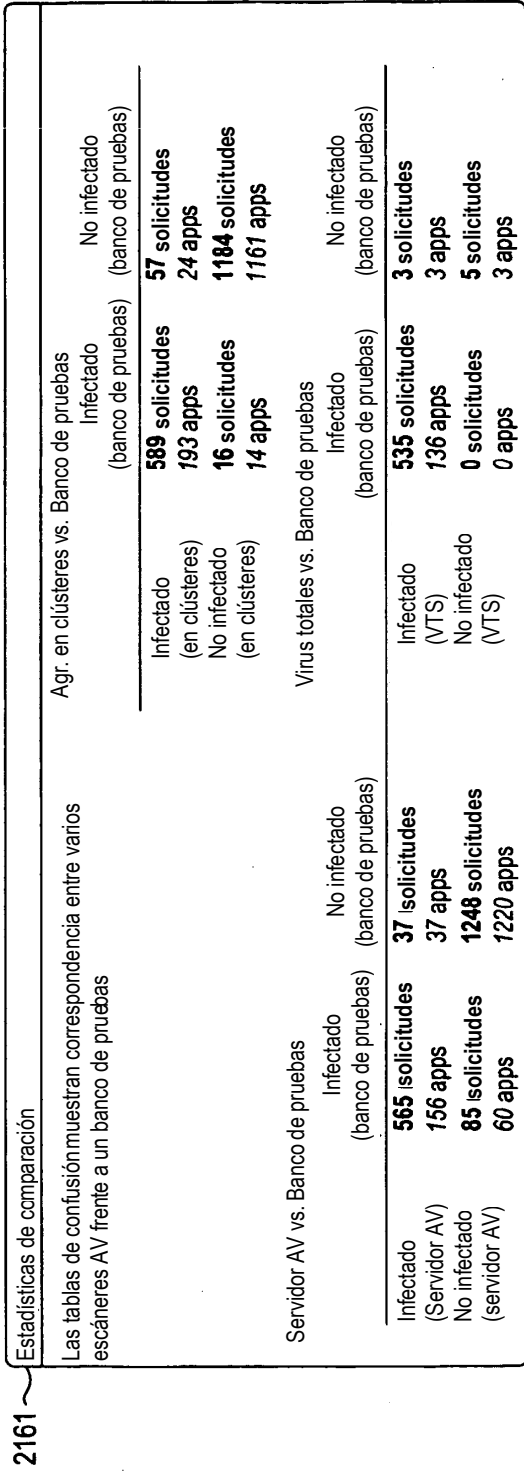
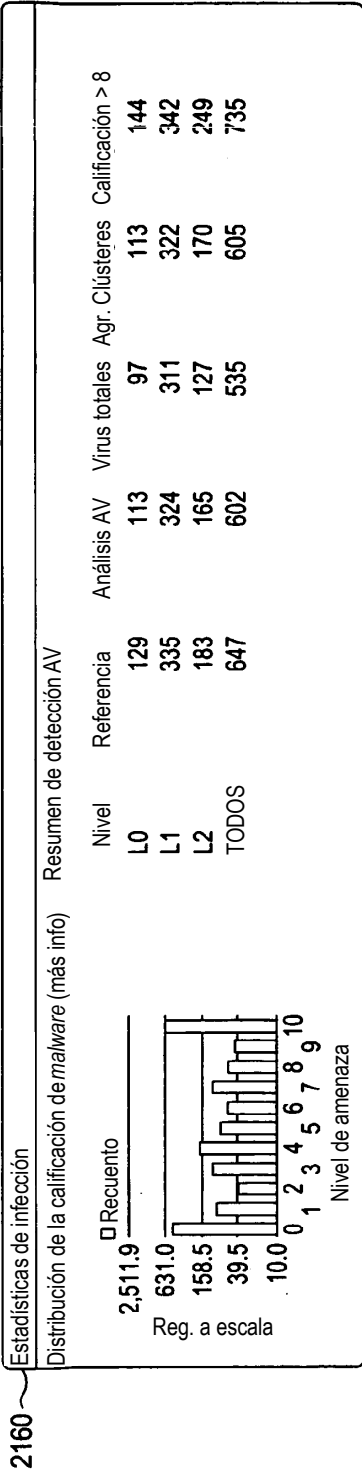
Continúa en la Hoja 64/94

Figura 21A

Continúa en la Hoja 63/94

Conexión de red de espacio aislado		IP que llegan al espacio aislado					
Espacios aislados de IP solicitados		IP		IP que llegan al espacio aislado		IP que llegan al espacio aislado	
Destino	Recuento	Volumen del paq.	Paq. medios	Fuente de la IP	Recuento	Volumen del paq.	Paq. medios
202.44.8.40	27	47014	1741	10.0.2.15	37447	748158	19
173.45.82.141	1919	13881	7	58.83.143.22	12	486	40
74.125.127.82	28	9952	355	112.125.53.4	6	200	33
173.192.187.130	6	8826	1471	58.63.244.77	54	135	2
122.11.61.10	538	8520	15	123.103.103.131	1	37	37
10.0.2.3	7243	7257	1	70.32.132.54	20	33	1
123.196.120.182	1359	6944	5	67.222.106.169	21	30	1
58.63.244.76	250	6868	27	123.196.120.182	6	24	4
58.63.244.65	211	6767	32	122.11.61.106	6	20	3
58.83.143.22	215	6620	30	58.63.244.76	9	18	2

Figura 21A (continuación)



Continúa en la Hoja 66/94

Figura 21B

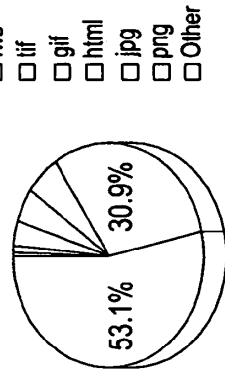
Continúa en la Hoja 65/94

2162

Estadísticas de archivos de red

Tipos de archivo descargados (más info)

Tipo	kB totales	Recuento	kB medios
avi	0	0	0
pdf	0	0	0
bmp	1.6	4	0.4
mpg	1437.88	15	95.86
fws	1795.5	15	119.7
tif	7928.73	302	26.25
gif	11582.41	1427	8.12
html	13191.25	2363	5.58
jpg	69782.69	2149	32.47
png	119934.55	3920	30.59



Tipos de archivo subidos (más info)

Tipo	kB totales	Recuento	kB medios
jt-g	0	0	0
gif	0	0	0
bmp	0	0	0
html	0	0	0
avi	0	0	0
pdf	0	0	0
tif	0	0	0
png	0	0	0
mpg	0.52	6	0.09
tws	10.3	17	0.61

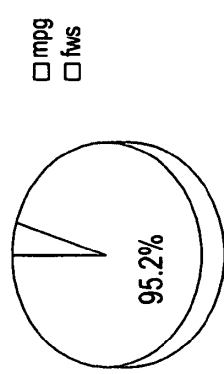
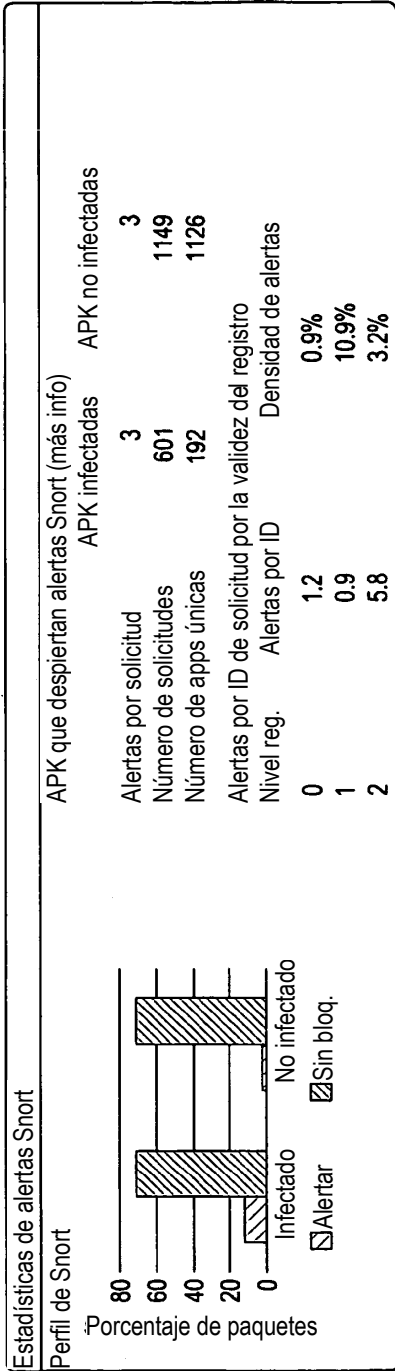


Figura 21B (continuación)



2170

Estado de marcas rojas

Comparación de marcas rojas (más info)

Estad. infección	N.º solicitud	N.º apps	Activaciones totales	Tipos de marcas rojas totales	Activaciones/App totales
No infectadas	1192	1167	155619	145	133.00
Infectadas	643	210	53423	123	254.00

Activaciones de marcas rojas predominantes

Tipo de marca roja	Marca roja	Número
TU RED	Firma para descargar recursos URL	
TU RED	Firma para navegación web insegura	
TU RED	Firma para subir recursos URL mediante un HTTP Post	
TU IDENTIDAD	Firma para recuperar información sobre tu tipo de dispositivo	
TU RED	Firma para buscar consultas de recursos URL mediante HTTP Get	
TU IDENTIDAD	Firma para recuperar tu identificador gsm imei único del teléfono	
TU PRIVACIDAD	Firma mejorada para monitorizar tu estado de red	

2171

Continúa en la Hoja 68/94

Figura 21C

Continúa en la Hoja 67/94

Tipo de marca roja	Marca roja activada	Estado de infección asociado	Activaciones medias	N.º solicitudes
TU PRIVACIDAD	Firma mejorada para monitorizar tu estado de red	1092		
TUS ARCHIVOS	Firma para ser curioso sobre el contenido del directorio de la tarjeta sd	1046		
TUS ARCHIVOS	Firma para acceder a tu tarjeta sd	910		
FACTOR DE RIESGO	Firma para recuperar información sensible sobre tu red	851		
TU IDENTIDAD	Firma para recuperar tu id gsm imsi de suscriptor	840		
FACTOR DE RIESGO	Firma para depender de un temporizador y/o una estructura de retardo	809		
TU UBICACIÓN	Firma para comprobar tu última geolocalización	760		
TU PRIVACIDAD	Firma para una distribución de anuncios convencional	693		
Activaciones de marcas rojas predominantes por estado de infección				
TU RED	Firma para obtener recursos ftp	Infectado	231.1	56
FACTOR DE RIESGO	Firma mejorada para mandar correos por programación	Infectado	157.8	4
FACTOR DE RIESGO	Firma mejorada para mandar correos por programación	No infectado	101.5	11
TU RED	Firma para descargar recursos url pero mediante librería apache	No infectado	72.4	550
TU RED	Firma para obtener recursos ftp	No infectado	55.2	94
TU TELÉFONO	Firma para monitorizar tus estadísticas de tráfico de red	No infectado	45.7	9
TU PRIVACIDAD	Firma para leer cookies HTTP	No infectado	37.2	33
FACTOR DE RIESGO	Firma para exponer un objeto en serie compartido en la red	No infectado	34.2	11
TU PRIVACIDAD	Firma para una distribución de anuncios convencional	No infectado	31.5	558
PELIGROSO	Firma para iniciar procesos de java nativos y enviar datos	No infectado	25.0	2

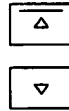


Figura 21C (continuación)

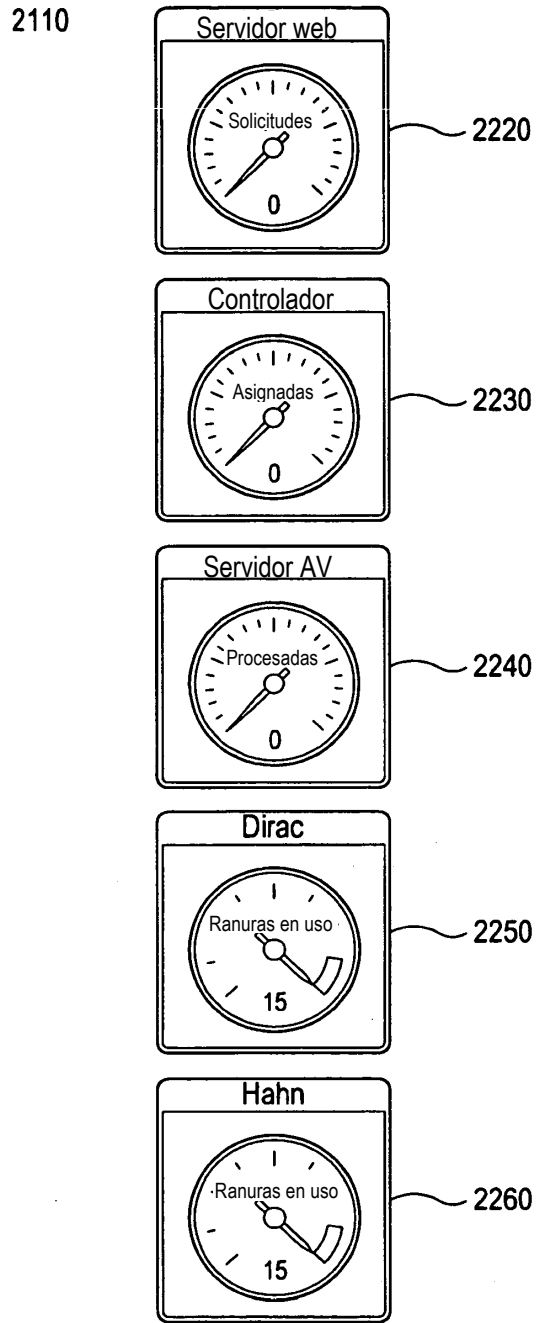


Figura 22

Continúa en la Hoja 71/94

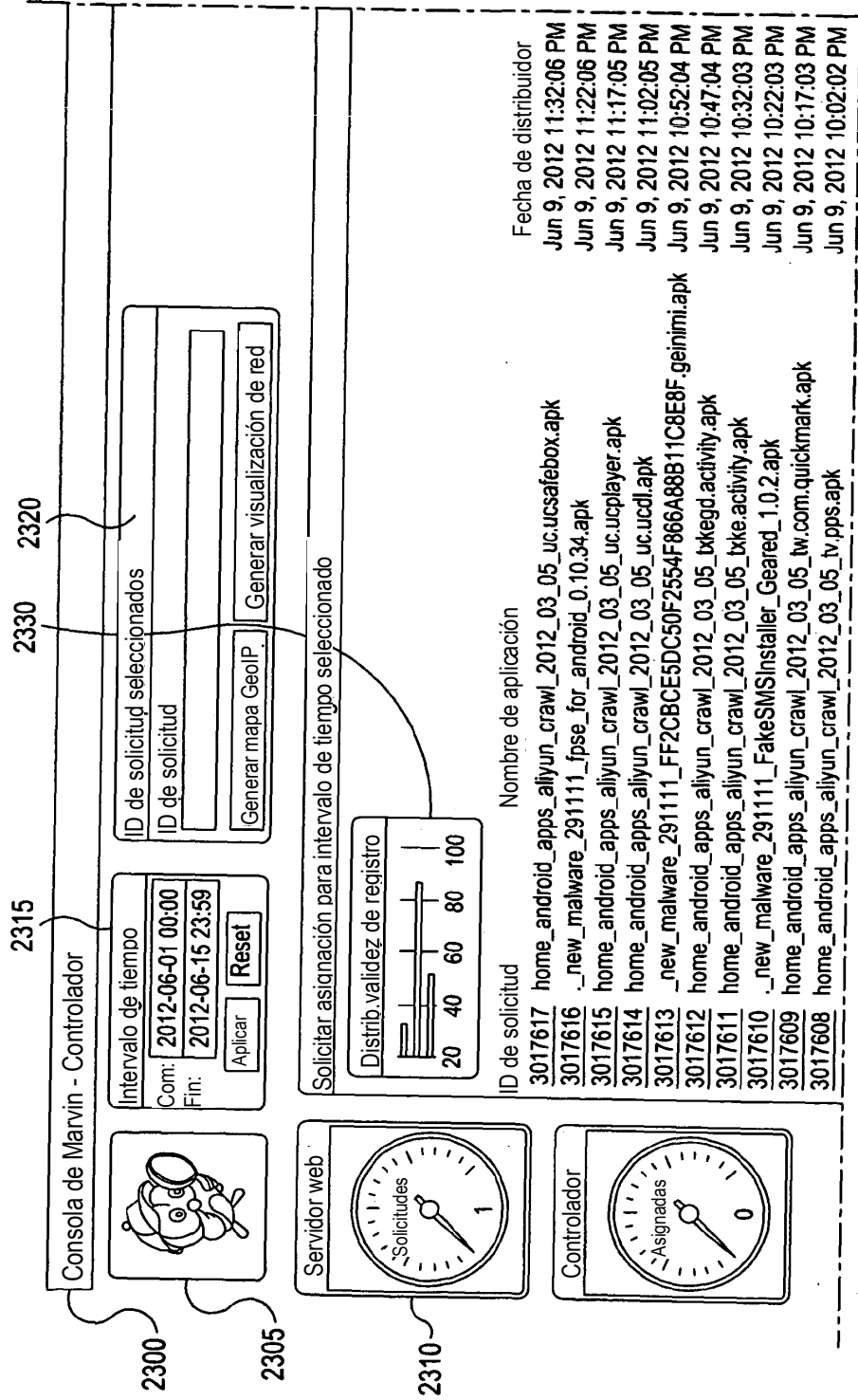


Figura 23

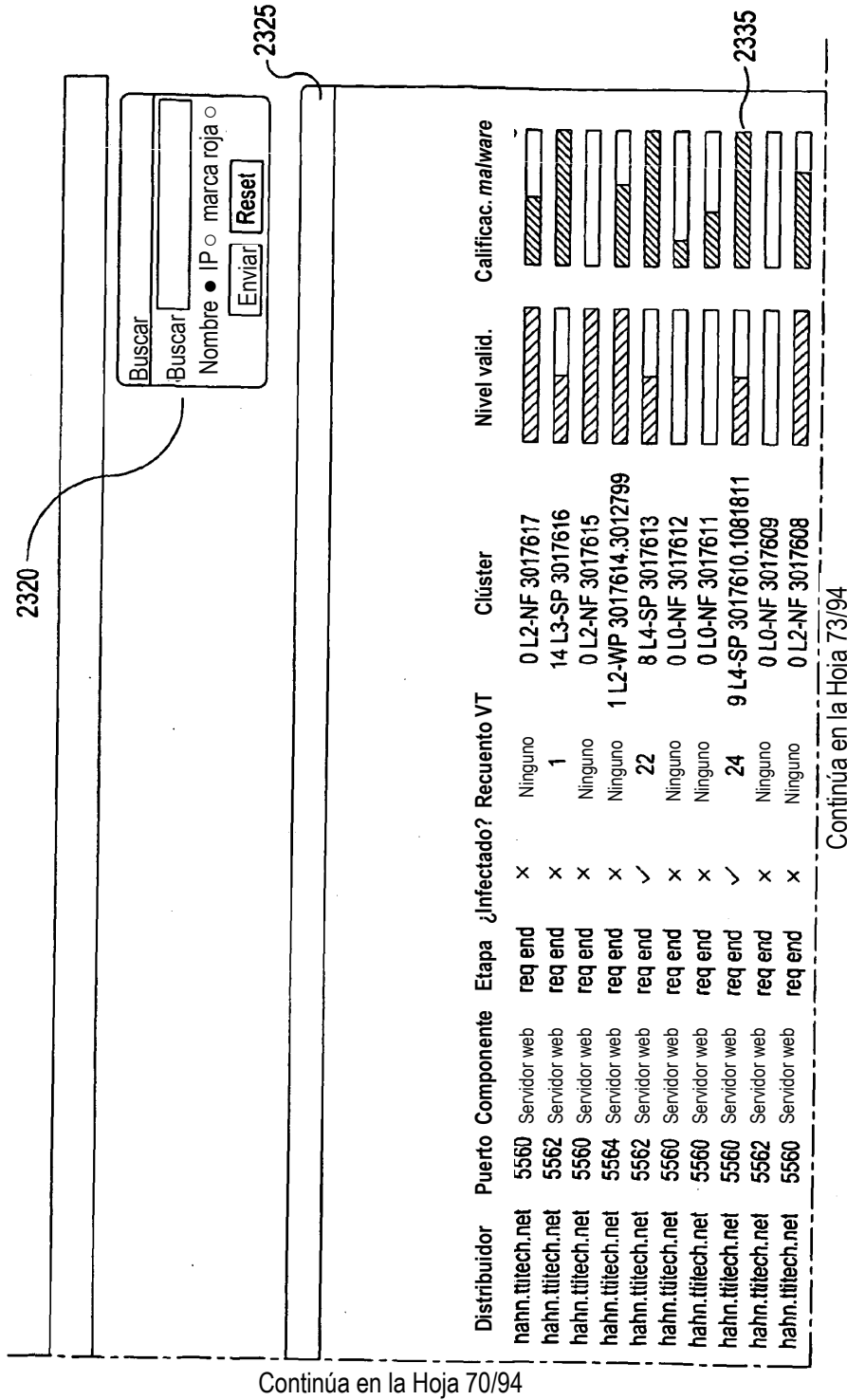


Figura 23 (continuación)

Continúa en la Hoja 70/94

Continúa en la Hoja 73/94

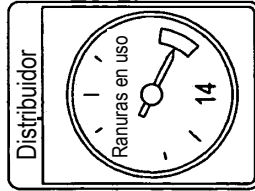
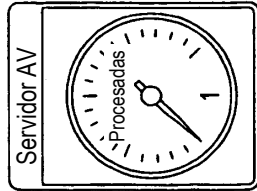
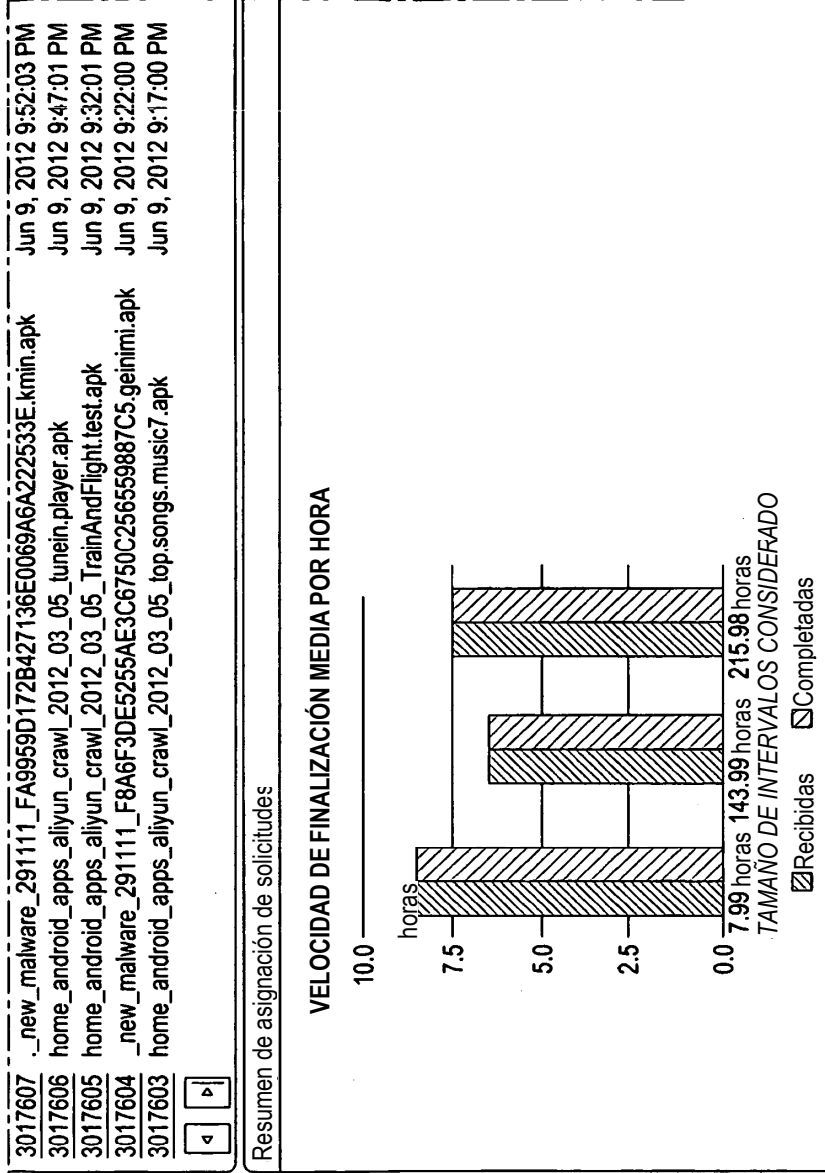


Figura 23 (continuación)

Continúa en la Hoja 71/94

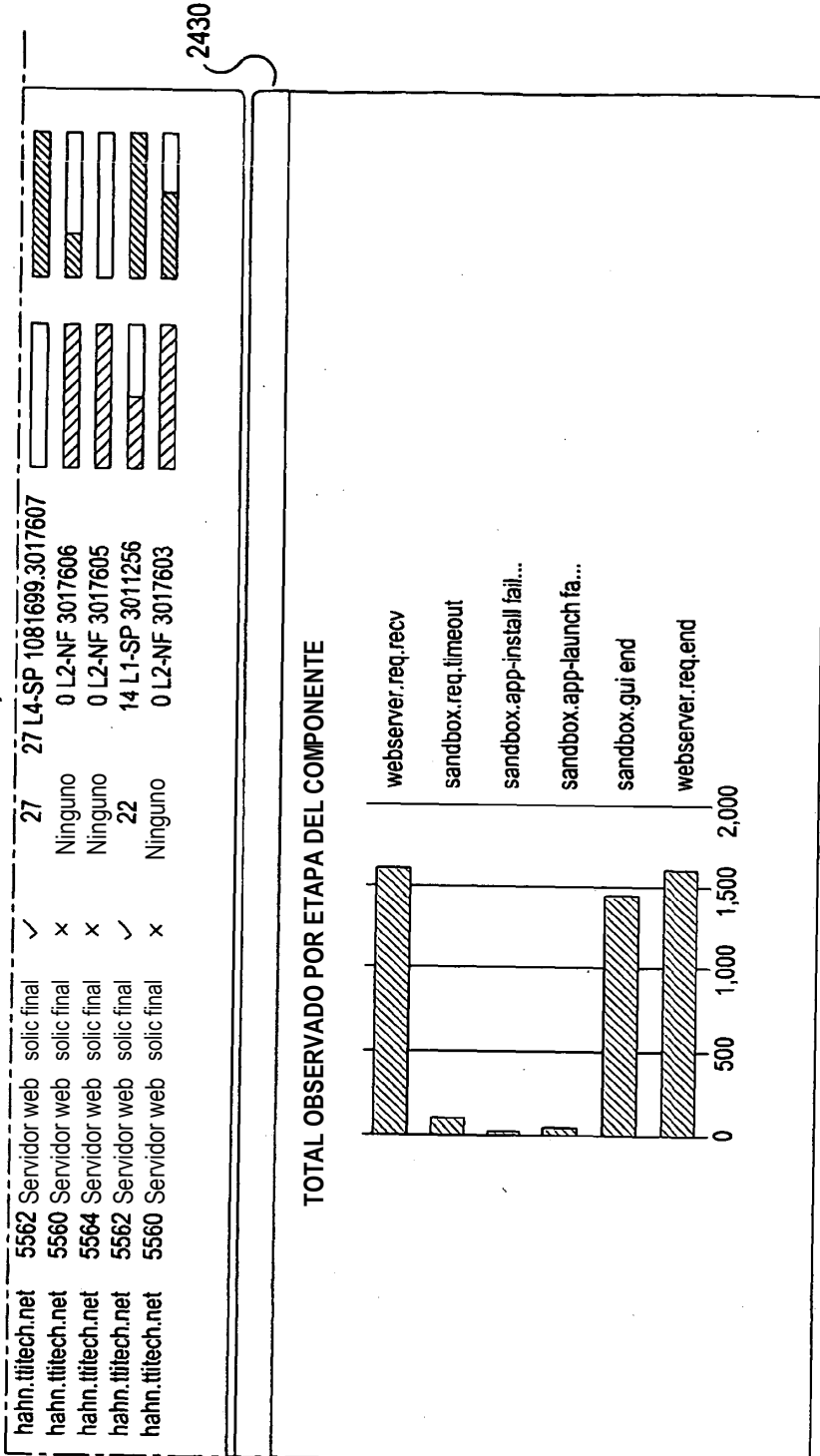
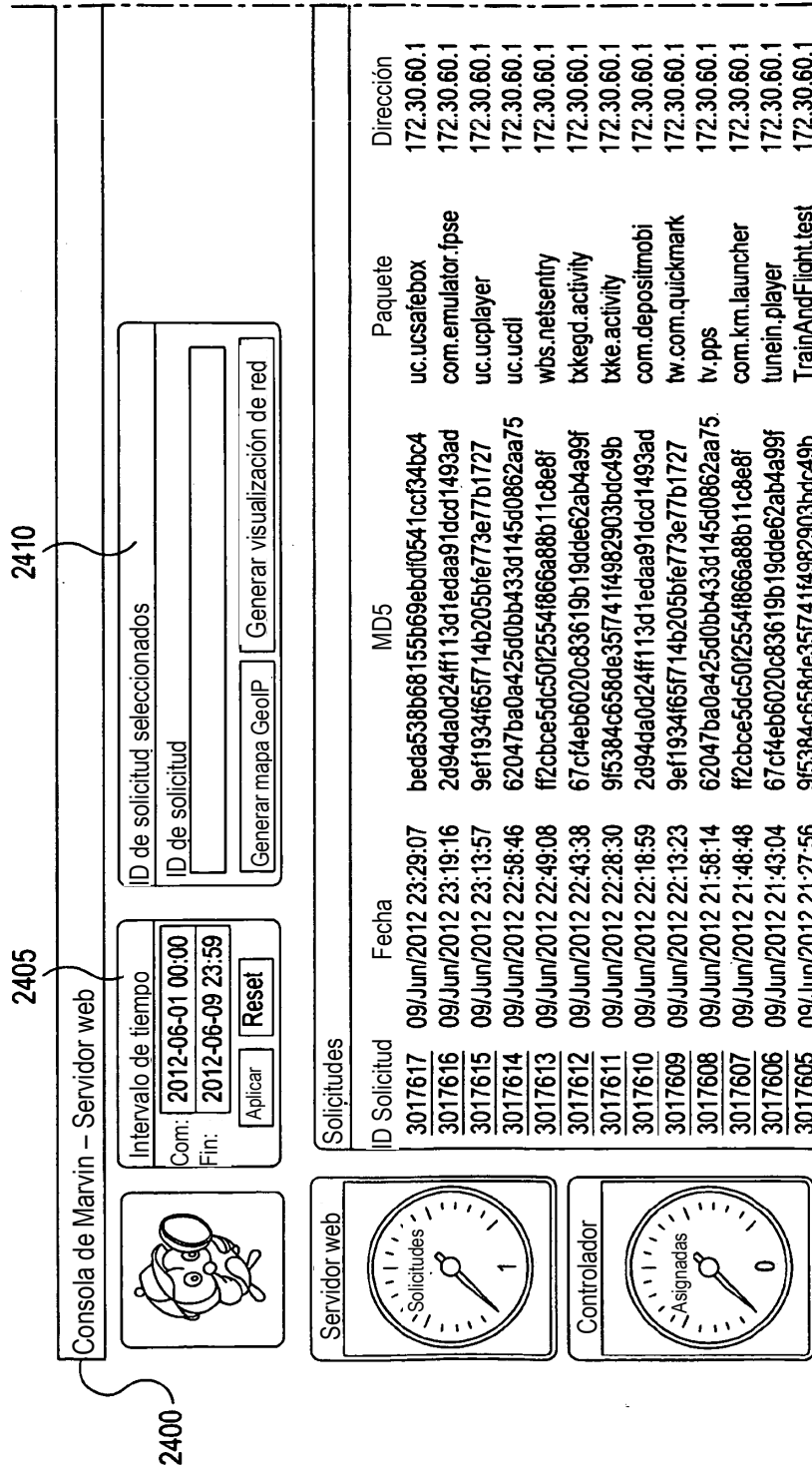


Figura 23 (continuación)

Continúa en la Hoja 72/94

Continúa en la Hoja 76/94



Continúa en la Hoja 75/94

Figura 24

Continúa en la Hoja 77/94

Continúa en la Hoja 74/94

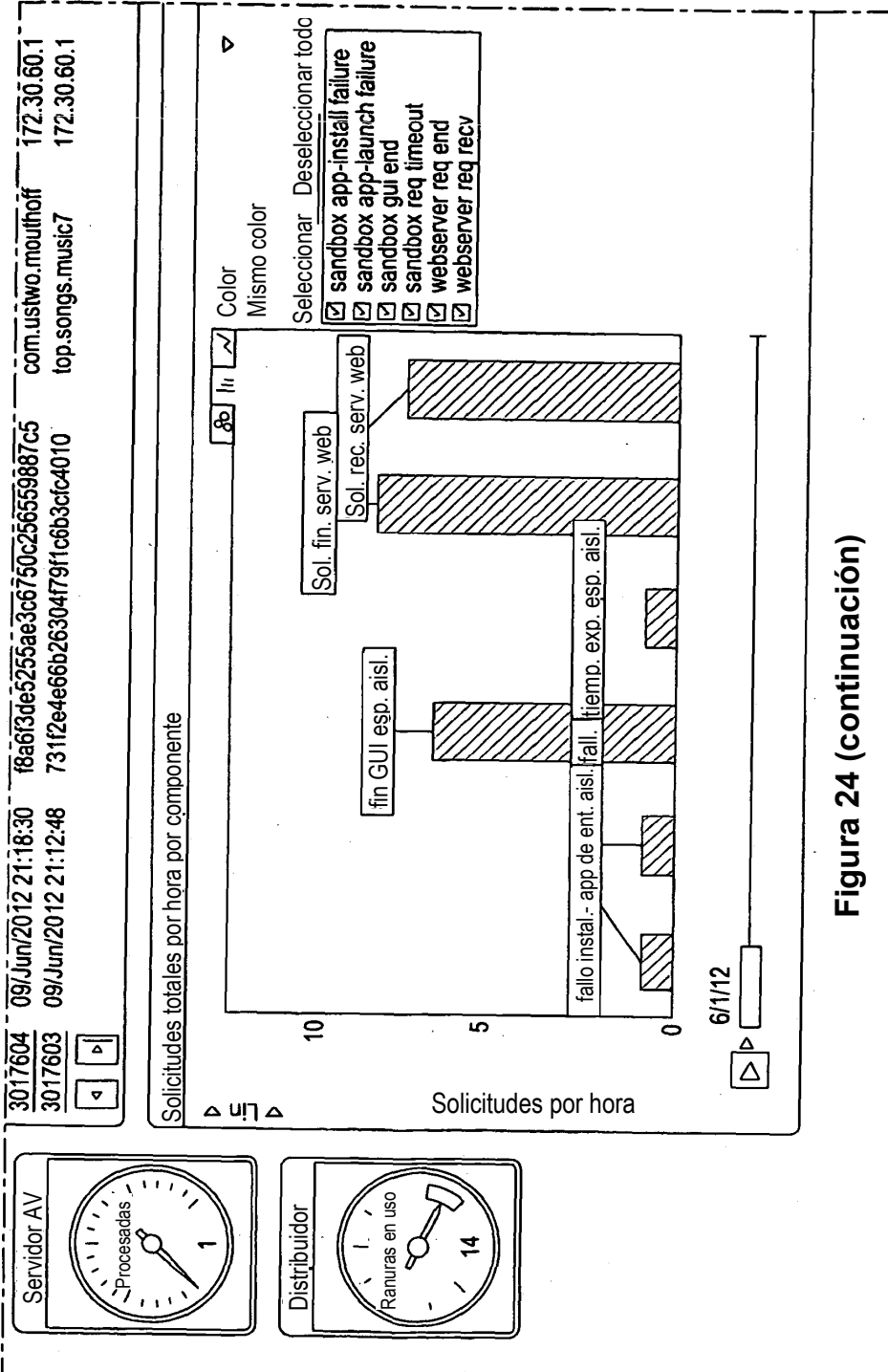


Figura 24 (continuación)

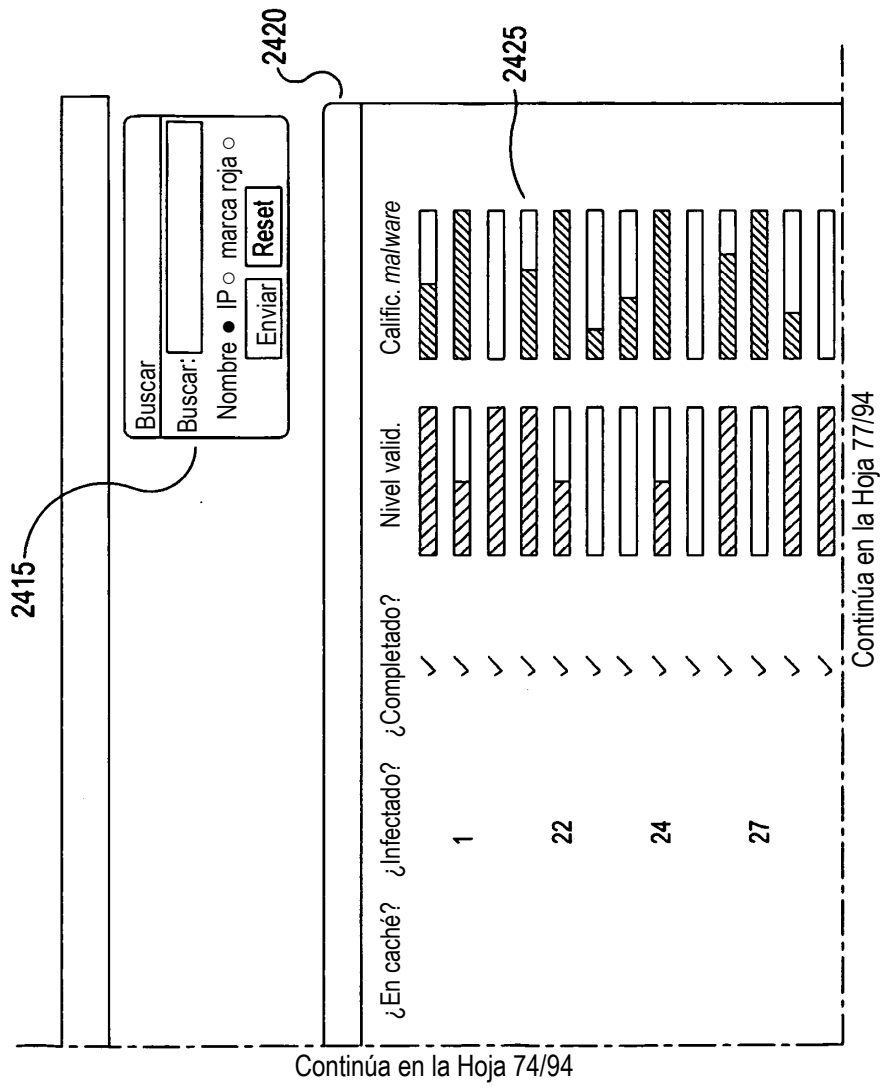


Figura 24 (continuación)

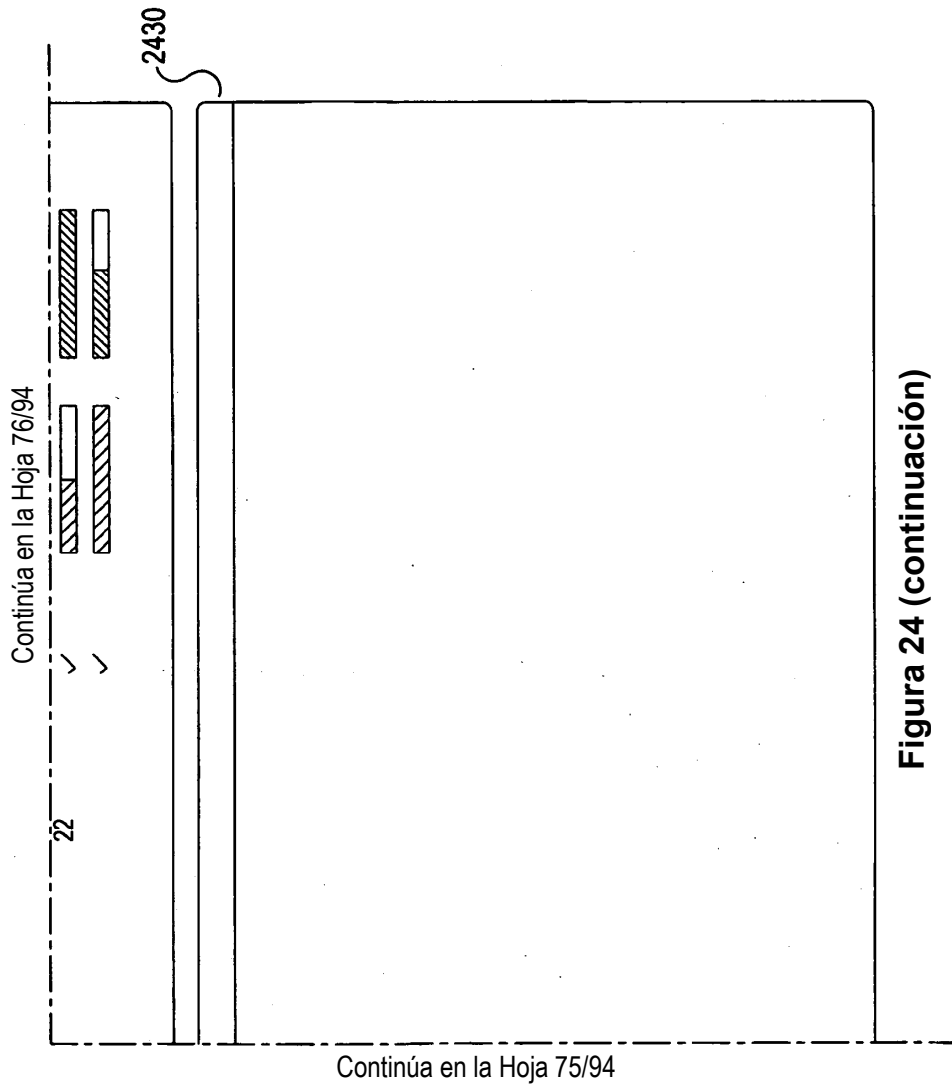
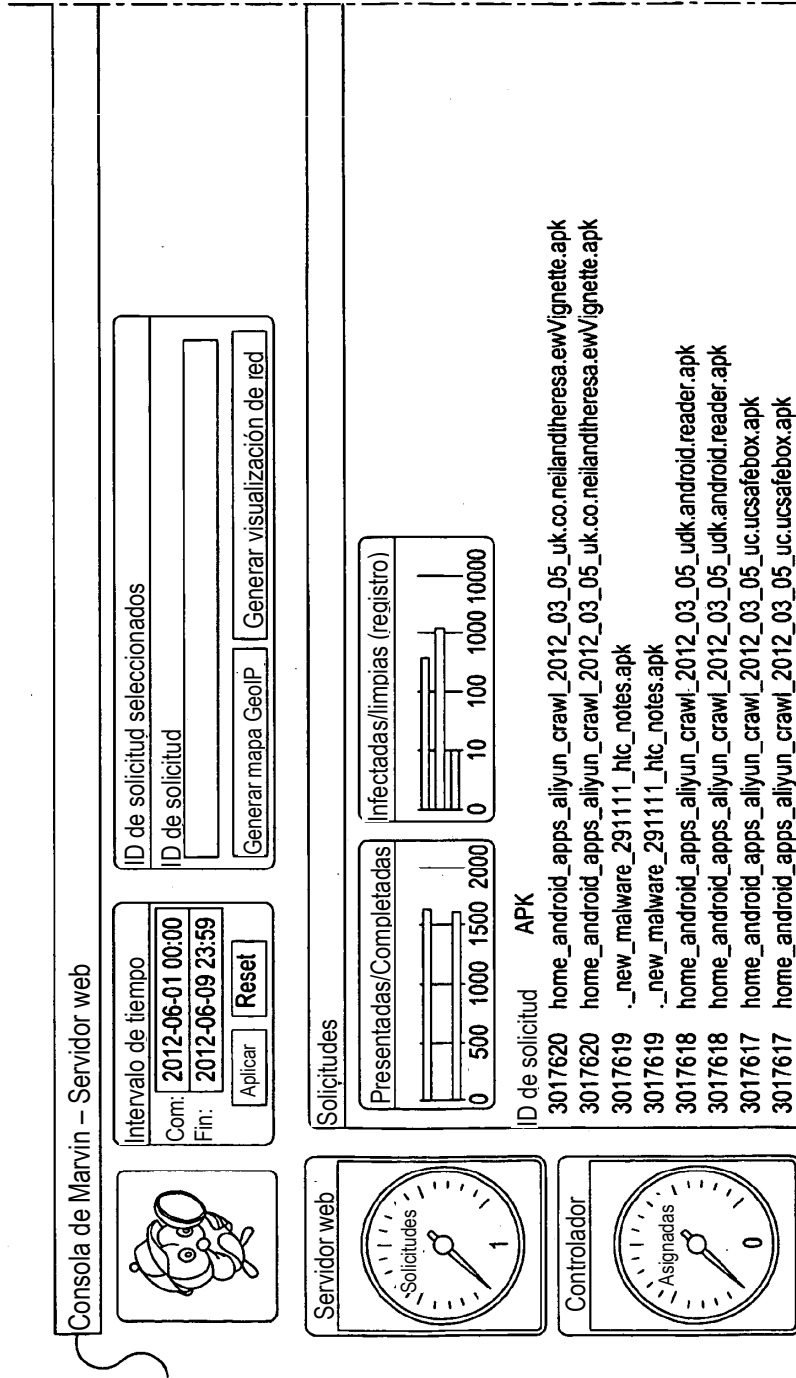


Figura 24 (continuación)

Continúa en la Hoja 79/94



Continúa en la Hoja 80/94

Figura 24A

2475

2480

Buscar

Nombre IP marca roja

Enviar

MD5	AV	Fecha presentación	Fecha inicio	Fecha fin	¿Infectada? ¿Infectada con?
5ccfa14be5e5108841022d459e6794a1	DrWeb	2012-06-09 23:59	2012-06-09 23:59	2012-06-09 23:59	x No se encontró nada.
5ccfa14be5e5108841022d459e6794a1	ClamAV	2012-06-09 23:59	2012-06-09 23:59	2012-06-09 23:59	x No se encontró nada.
95c6ca341aa4ac9ce54252883a55da22	DrWeb	2012-06-09 23:49	2012-06-09 23:49	2012-06-09 23:49	✓ Malware detectado.
95c6ca341aa4ac9ce54252883a55da22	ClamAV	2012-06-09 23:49	2012-06-09 23:49	2012-06-09 23:49	x No se encontró nada.
3abbac36c09310237c7c3ae7440cf28	DrWeb	2012-06-09 23:44	2012-06-09 23:44	2012-06-09 23:44	x No se encontró nada.
3abbac36c09310237c7c3ae7440cf28	ClamAV	2012-06-09 23:44	2012-06-09 23:44	2012-06-09 23:44	x No se encontró nada.
bedaa538b68155b69ebd0541ccf34bc4	DrWeb	2012-06-09 23:29	2012-06-09 23:29	2012-06-09 23:29	x No se encontró nada.
bedaa538b68155b69ebd0541ccf34bc4	ClamAV	2012-06-09 23:29	2012-06-09 23:29	2012-06-09 23:29	x No se encontró nada.

Continúa en la Hoja 81/94

Figura 24A (continuación)

Continúa en la Hoja 78/94

Continúa en la Hoja 81/94

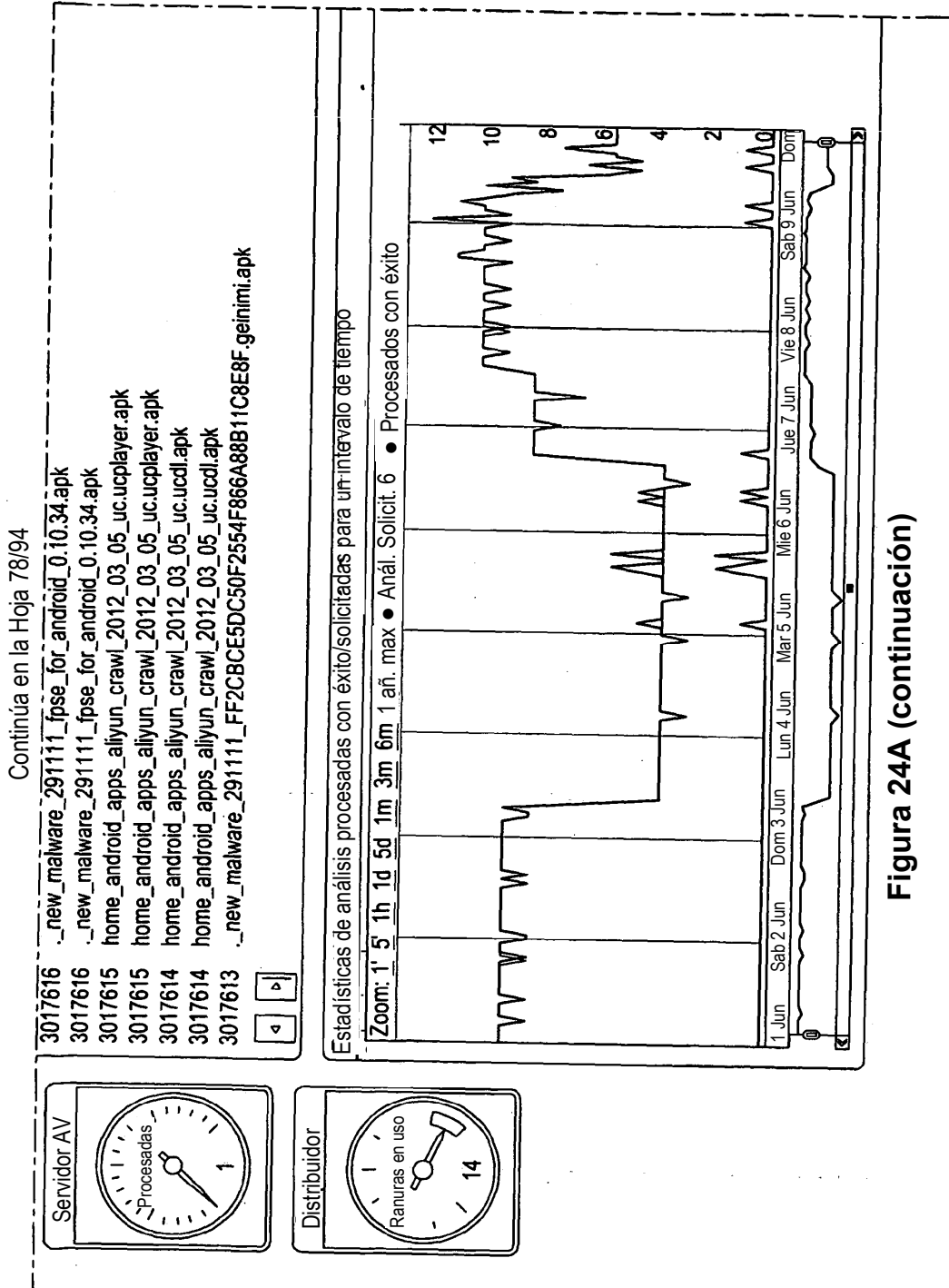


Figura 24A (continuación)

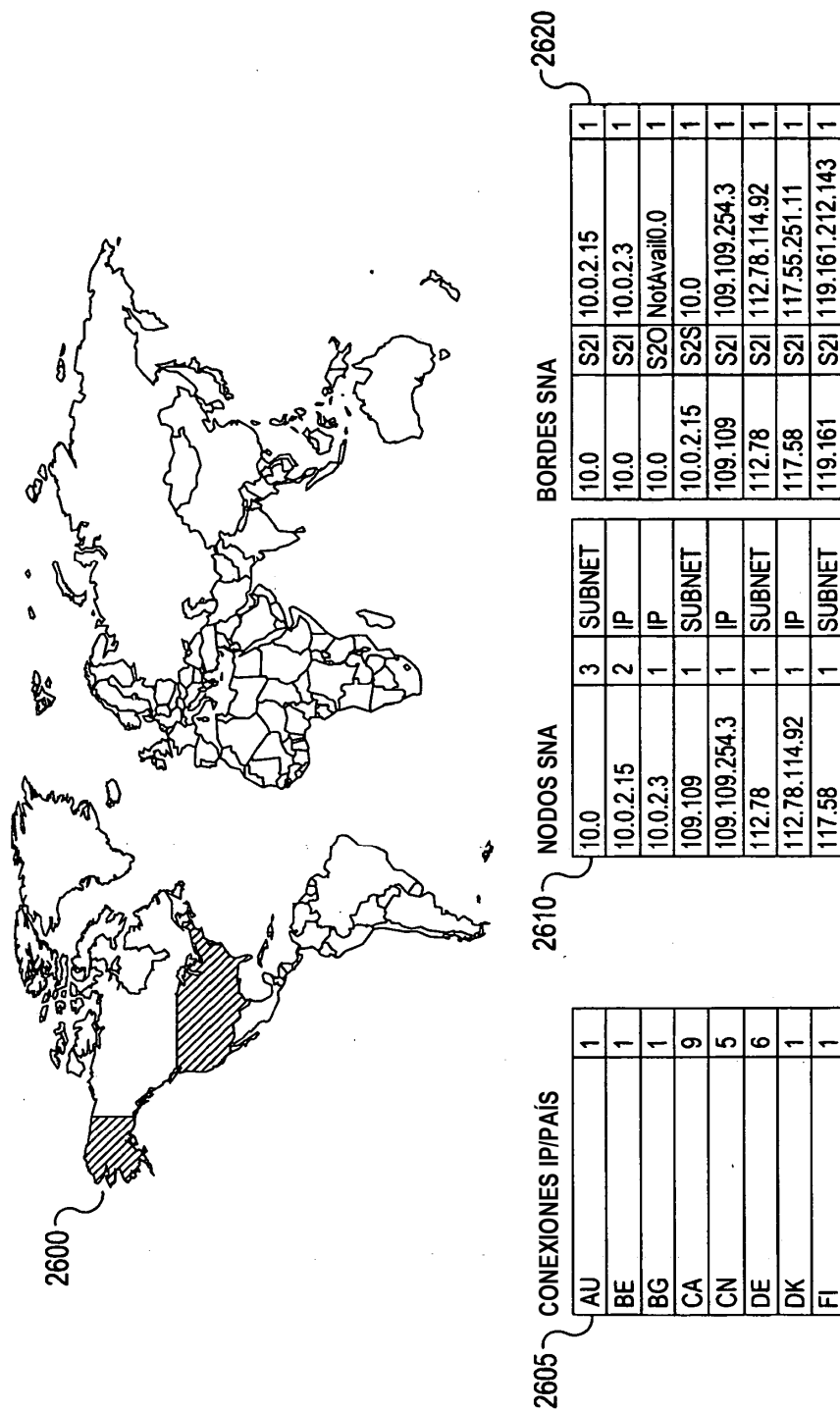
Continúa en la Hoja 79/94

2d94da0d24ff113d1edaa91dcd1493ad	DrWeb	2012-06-09 23:19	2012-06-09 23:19	2012-06-09 23:19	x	No se encontró nada.
2d94da0d24ff113d1edaa91dcd1493ad	ClamAV	2012-06-09 23:19	2012-06-09 23:19	2012-06-09 23:19	x	No se encontró nada.
9ef1934f65714b205bfe773e77b1727	DrWeb	2012-06-09 23:14	2012-06-09 23:14	2012-06-09 23:14	x	No se encontró nada.
9ef1934f65714b205bfe773e77b1727	ClamAV	2012-06-09 23:14	2012-06-09 23:14	2012-06-09 23:14	x	No se encontró nada.
62047ba0a425d0bb433d145d0862aa75	DrWeb	2012-06-09 22:58	2012-06-09 22:58	2012-06-09 22:58	x	No se encontró nada.
62047ba0a425d0bb433d145d0862aa75	ClamAV	2012-06-09 22:58	2012-06-09 22:58	2012-06-09 22:58	x	No se encontró nada.
ff2cbce5dc5012554f866a88b11c8e8f	DrWeb	2012-06-09 22:49	2012-06-09 22:49	2012-06-09 22:49	✓	Malware detectado.

2485

Continúa en la Hoja 80/94

Figura 24A (continuación)



Continúa en la Hoja 83/94

Figura 25

Continúa en la Hoja 82/94

FR	4	117.58.251.11	1	IP	119.161	S21	119.161.213.99	1
GB	9	119.161	2	SUBRED	119.42	S21	119.42.227.250	1
IE	3	119.161.212.143	1	IP	12.129	S21	12.129.242.20	1
IL	1	119.161.213.99	1	IP	128.30	S21	128.30.52.37	1
JP	5	119.42	1	SUBRED	129.33	S21	129.33.27.97	1
NL	2	119.42.227.250	1	IP	146.101	S21	146.101.146.121	1
NoDisp10.0	3	12.129	1	SUBRED	157.166	S21	157.166.173.170	1
NoDisp209.54	2	12.129.242.20	1	IP	157.166	S21	157.166.224.212	1
NoDisp239.255	1	128.30	1	SUBRED	157.166	S21	157.166.224.25	1
RU	1	128.30.52.37	1	IP	157.166	S21	157.166.255.19	1
SE	1	129.33	1	SUBRED	157.166	S21	157.166.255.222	1
SG	1	129.33.27.97	1	IP	159.53	S21	159.53.60.148	1
UA	1	146.101	1	SUBRED	159.53	S21	159.53.84.11	1
US	544	146.101.146.121	1	IP	161.221	S21	161.221.85.120	1
		157.166	5	SUBRED	165.193	S21	165.193.245.178	1
		157.166.173.170	1	IP	165.193	S21	165.193.245.41	1

2605

2610

2650

Figura 25 (continuación)

VISUALIZACIÓN DE REDES IP COLECTIVAS ALCANZADAS POR REQUIDS [1029542, 1029570]

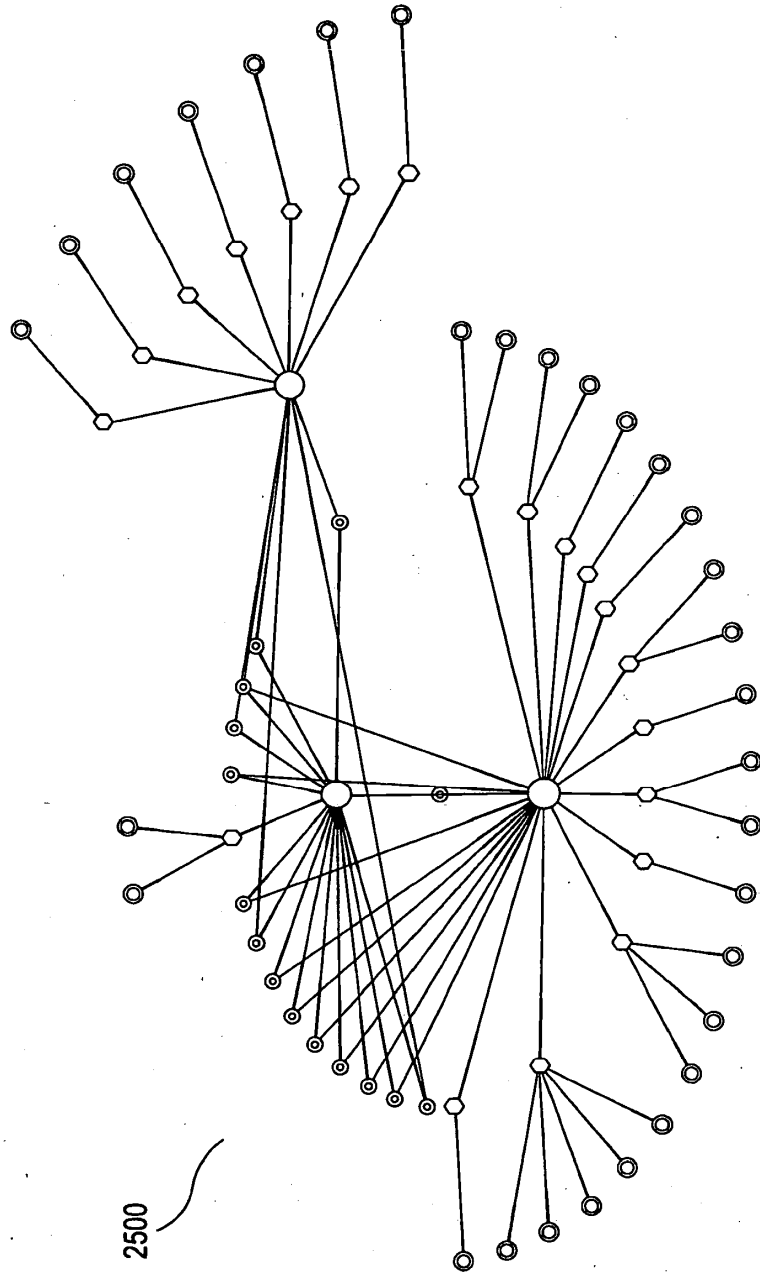
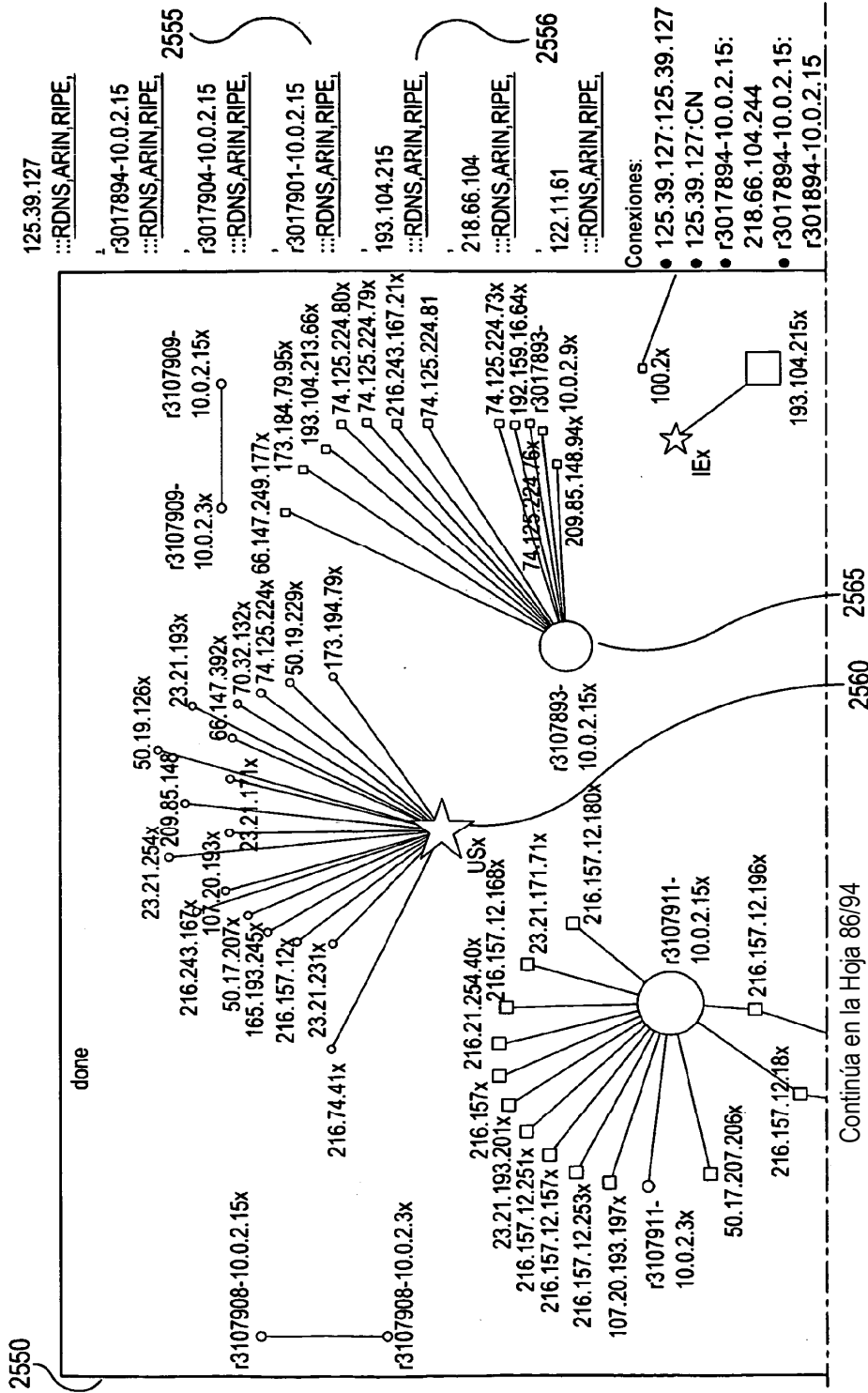


Figura 26



- 125.39.127
 - ...RDNS,ARIN,RIPE,
 - r3017894-10.0.2.15
 - ...RDNS,ARIN,RIPE,
 - 2555
 - r3017904-10.0.2.15
 - ...RDNS,ARIN,RIPE,
 - r3017901-10.0.2.15
 - ...RDNS,ARIN,RIPE,
 - 193.104.215
 - ...RDNS,ARIN,RIPE,
 - 218.66.104
 - ...RDNS,ARIN,RIPE,
 - 2556
 - 122.11.61
 - ...RDNS,ARIN,RIPE,
- Conexiones:
- 125.39.127:125.39.127
 - 125.39.127:CN
 - r3017894-10.0.2.15:
 - 218.66.104.244
 - r3017894-10.0.2.15:
 - r301894-10.0.2.15

Figura 27

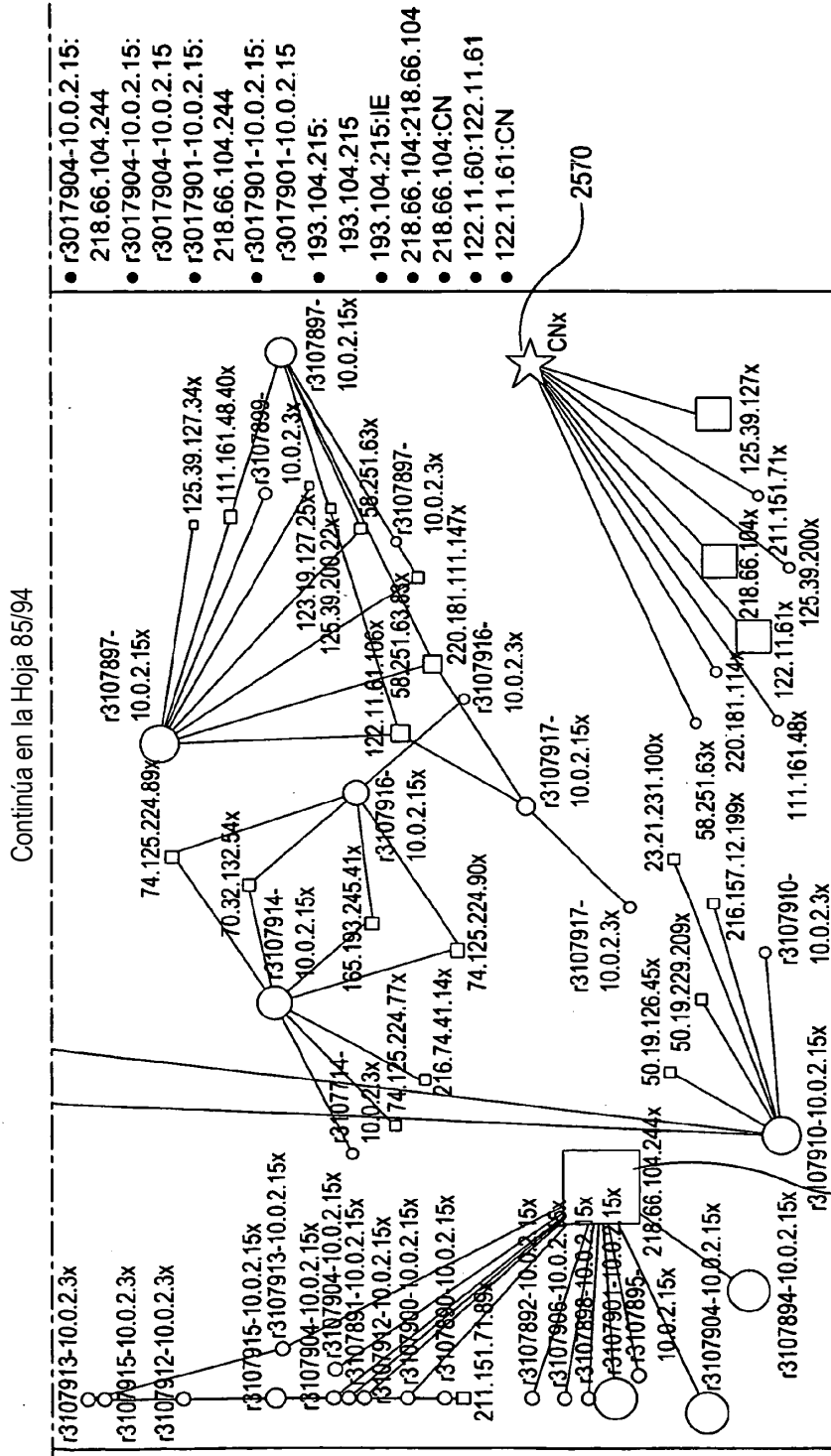
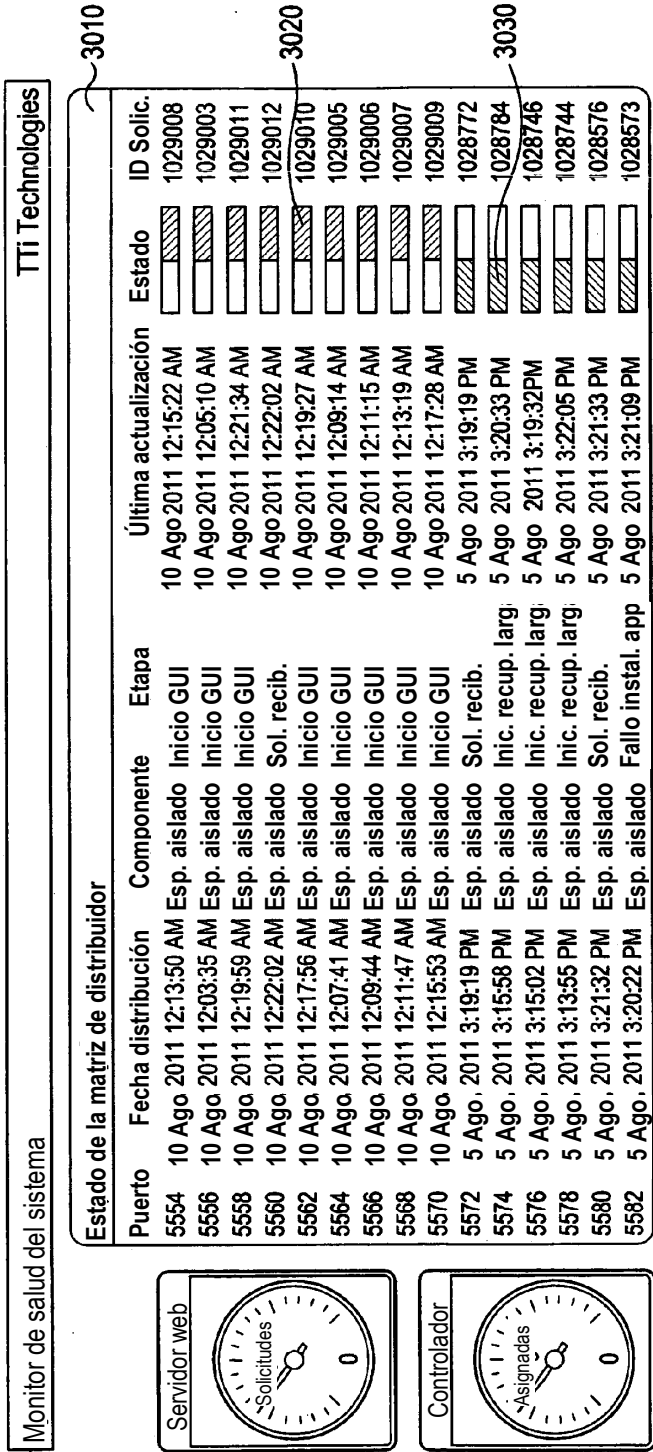


Figura 27 (continuación)

3000



Continúa en la Hoja 88/94

Figura 30

Continúa en la Hoja 87/94

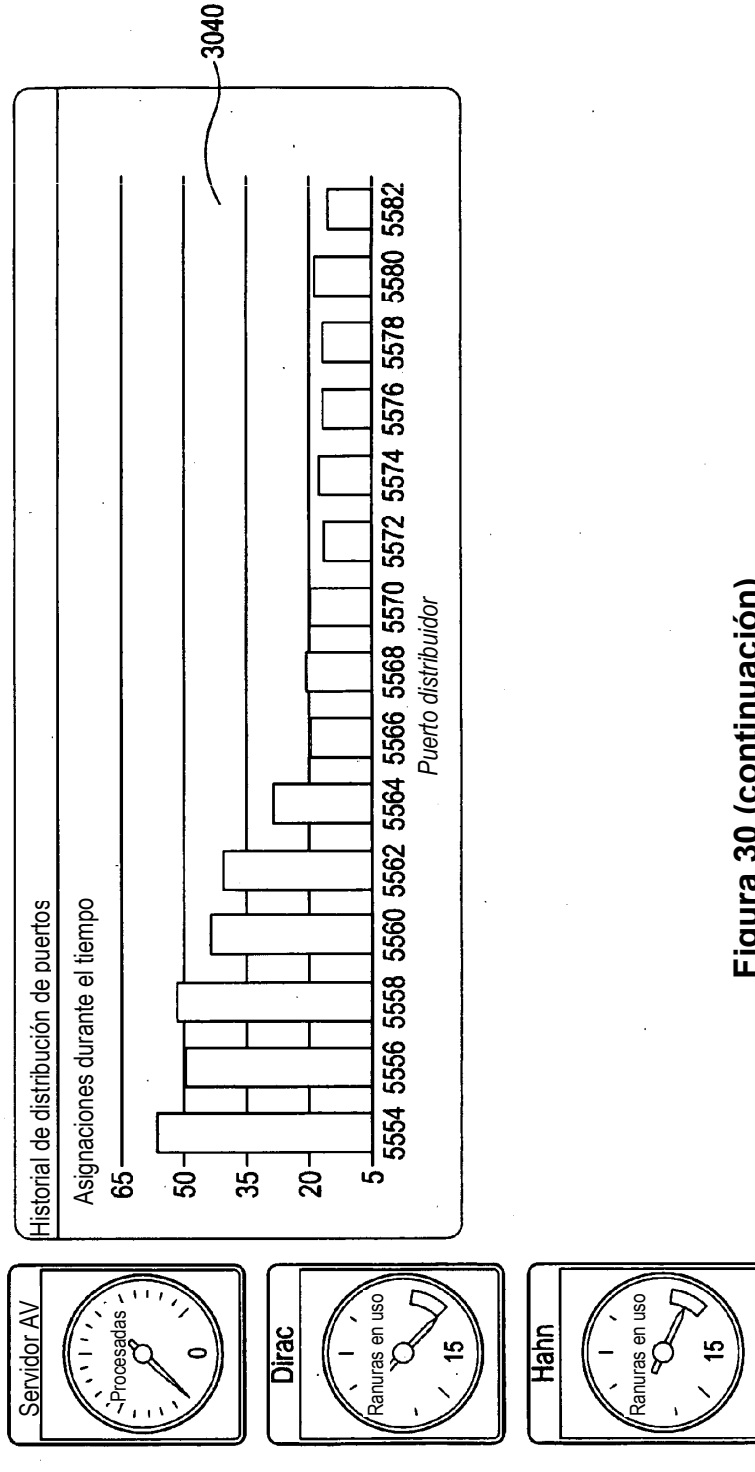
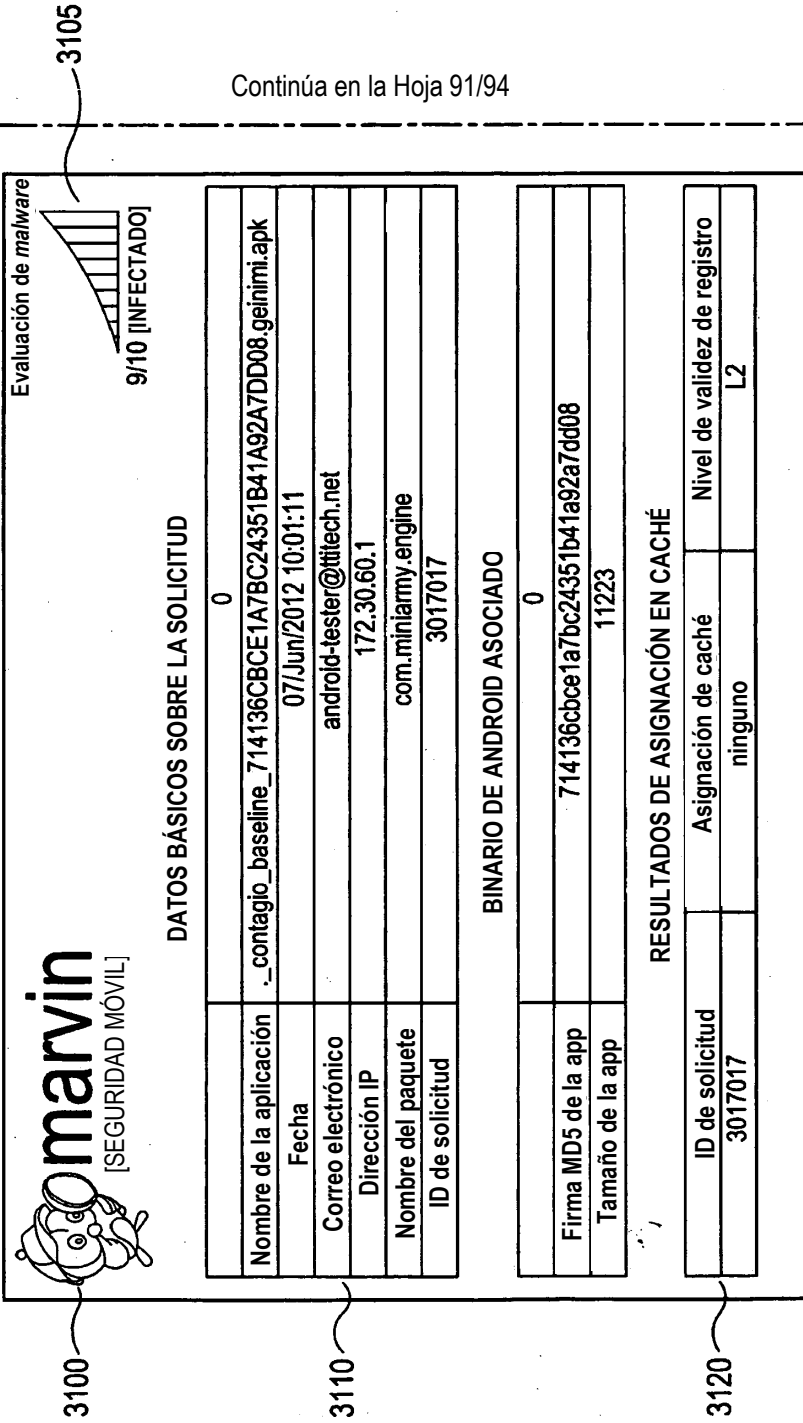


Figura 30 (continuación)



Continúa en la Hoja 90/94

Figura 31

Continúa en la Hoja 89/94

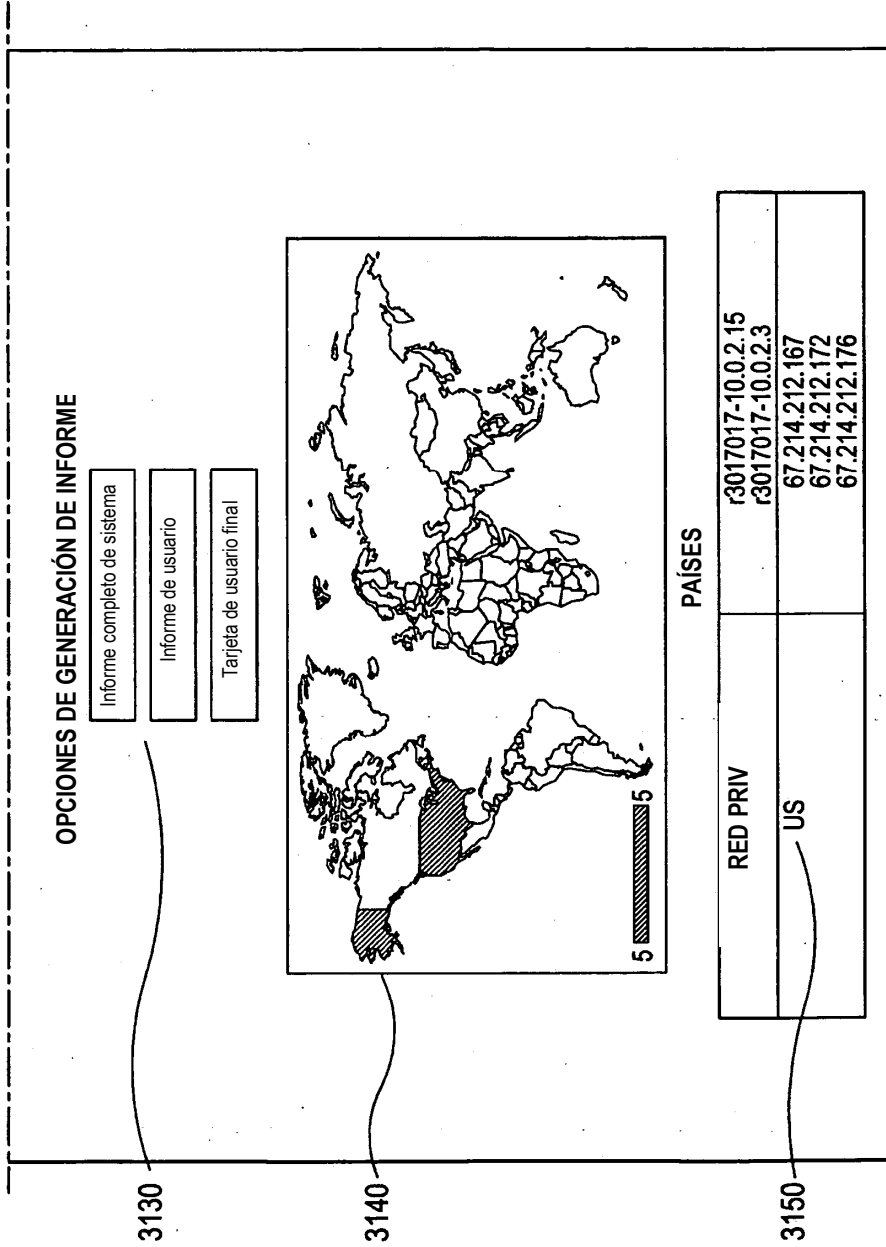
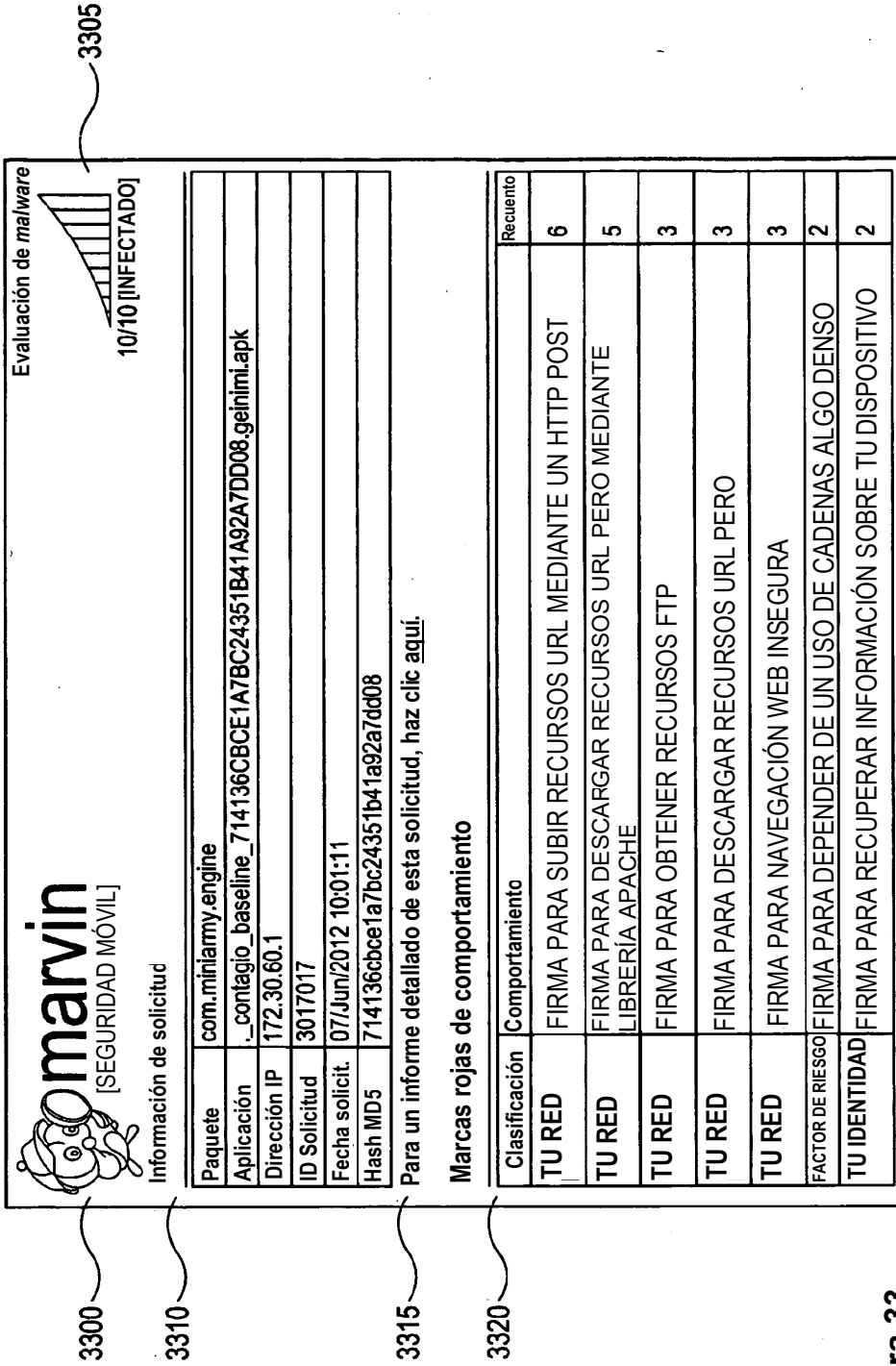


Figura 31 (continuación)

3160	SOLICITAR ASIGNACIÓN				
		0			
	Fecha distribución	1339077794			
	Nodo distribuidor	hahn.ttitech.net			
	Puerto distribuidor	5564			
3170	PROGRESO DE LA SOLICITUD A TRAVÉS DE LOS COMPONENTES DE MARVIN				
	Componente	Etapa del componente	ID solicitud	Marca de tiempo	
	0	servidor web	"solic recib"	"3017017"	1339077676.0401821
	1	esp. aisl. AV	"solic recib"	"3017017"	1339077683.95
	2	esp. aisl. AV	"solic final"	"3017017"	1339077684.17
	3	controlador	"solic recib"	"3017017"	1339077794.5644851
	4	distribuidor	"solic recib"	"3017017"	1339077794.6224661
	5	esp. aislado	"solic recib"	"3017017"	1339077798.2598491
	6	esp. aislado	"app instal. con éxito"	"3017017"	1339078073.085664
	7	esp. aislado	"app instal. con éxito"	"3017017"	1339078078.2627261
	8	esp. aislado	"inicio de GUI"	"3017017"	1339078118.139673
	9	esp. aislado	"final. de GUI"	"3017017"	1339078822.7045419
	10	esp. aislado	"inicio recup. registr."	"3017017"	1339078843.566833
	11	esp. aislado	"fin recup. registr."	"3017017"	1339078916.3190839
	12	esp. aislado	"inicio recup. registr."	"3017017"	1339078932.126332
	13	distribuidor	"solic final"	"3017017"	1339078955.71737
	14	controlador	"solic final"	"3017017"	1339078961.1264441
	15	servidor web	"solic final"	"3017017"	1339078961.1928339
3180	REGISTRO GENERADO APA ESTA SOLICITUD				
		Tamaño de registro (kB)	Tipo de registro		
	0	22	db		
	1	122	events		
	2	18	gui		
	3	95	logcat		
	4	26	md5sums		
	5	96	output		
	6	120	png		
	7	21659	scalls		
	8	1	ssum		
	9	88	tcpdump		
	10	114	top		

Continúa en la Hoja 89/94

Figura 31 (continuación)



Continúa en la Hoja 93/94

Figura 33

Continúa en la Hoja 92/94

3325	<p>Conexiones a Internet por país</p> <table border="1"> <thead> <tr> <th data-bbox="560 1424 592 1588">País</th> <th data-bbox="560 990 592 1424">GS conexión</th> </tr> </thead> <tbody> <tr> <td data-bbox="596 1424 655 1588">RED PRIV</td> <td data-bbox="596 990 655 1424">r3017017-10.0.2.15 r3017017-10.0.2.3</td> </tr> <tr> <td data-bbox="660 1424 751 1588">US</td> <td data-bbox="660 990 751 1424">67.214.212.167 67.214.212.172 67.214.212.176</td> </tr> </tbody> </table>	País	GS conexión	RED PRIV	r3017017-10.0.2.15 r3017017-10.0.2.3	US	67.214.212.167 67.214.212.172 67.214.212.176
País	GS conexión						
RED PRIV	r3017017-10.0.2.15 r3017017-10.0.2.3						
US	67.214.212.167 67.214.212.172 67.214.212.176						
3330	<p>Alerta de Snort</p> <p>No hay alertas para la solicitud</p>						
3335	<p>Resultados del escáner antivirus</p> <p>2 de 2 informes de infección</p> <table border="1"> <tr> <td data-bbox="983 1424 1042 1588">Análisis</td> <td data-bbox="983 990 1042 1424">Etiqueta</td> </tr> </table>	Análisis	Etiqueta				
Análisis	Etiqueta						
3340	<p>Resultados de agrupamiento en clústeres</p> <table border="1"> <tr> <td data-bbox="1107 1424 1166 1588">Etiqueta clúster</td> <td data-bbox="1107 990 1166 1424">"15 L2-SP_3017848.3017274.3009145.3017849"</td> </tr> </table>	Etiqueta clúster	"15 L2-SP_3017848.3017274.3009145.3017849"				
Etiqueta clúster	"15 L2-SP_3017848.3017274.3009145.3017849"						

Figura 33 (continuación)

3400
3420 REGLA QUÉ = (r 'flurry/android', 'onEvent/logEvent', 'FlurryAgent'),
3430 URL='MARVIN: http://www.marvinsafe.com,
http://support.flurry.com/sdkdocs/android/classcom_1_1flurry_1_androi
3440 DESCRIPCIÓN = [TU PRIVACIDAD] – Firma para activar 3ª parte
detallada?',
3450 RIESGO=10
3460 PNG=modo(Verdadero,Ninguno) EOR 3441 3442

Figura 34