(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 946 277**

(51) Int. CI.:

G06F 21/60 (2013.01)
G06F 21/57 (2013.01)
G06F 21/62 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **02.08.2018 PCT/AU2018/050808**

(87) Fecha y número de publicación internacional: **07.02.2019 WO19023756**

(96) Fecha de presentación y número de la solicitud europea: **02.08.2018 E 18840744 (9)**

(97) Fecha y número de publicación de la concesión europea: **22.03.2023 EP 3662402**

(54) Título: **Un sistema, método, programa informático y señal de datos para identificar un software capaz de capturar información de identificación personal**

(30) Prioridad:

02.08.2017 AU 2017903065

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.07.2023

(73) Titular/es:

SOURCE OF TRUE PTY LTD (100.0%)
9 Prospect Street
Waverley, NSW 2024, AU

(72) Inventor/es:

JOWETT, ROBIN

(74) Agente/Representante:

ELZABURU, S.L.P

DESCRIPCIÓN

Un sistema, método, programa informático y señal de datos para identificar un software capaz de capturar información de identificación personal

Campo técnico de la Invención

La presente invención se refiere a un sistema según el preámbulo de la reivindicación 5, un método según el preámbulo de la reivindicación 1, un programa informático y una señal de datos para identificar un software capaz de capturar información de identificación personal.

Las realizaciones de la invención encuentran un uso particular, pero no exclusivo, en la identificación de código ejecutable en sitios web que capturan y almacenan información que puede ser utilizada para identificar personalmente a un individuo y/o usuario del sitio web.

En una realización específica de la invención, el software determina si la información de identificación personal introducida en un primer sitio web por un usuario se comunica a un sitio web o base de datos de terceros sin el consentimiento informado del usuario.

Antecedentes

Existe una necesidad creciente de proteger la Información de Identificación Personal (PII). PII es cualquier dato que se puede utilizar para identificar potencialmente a un individuo específico. En términos generales, cualquier información que ayude en la identificación de un individuo específico (en otras palabras, la información que se puede utilizar para desanonimizar datos que de otro modo se considerarían anónimos) se considera PII. Los ejemplos obvios de información PII incluyen parte o la totalidad del nombre de un individuo, su número de carnet de conducir, su número de pasaporte, su número de teléfono personal, etc.

Sin embargo, otra información que no se puede considerar PII de inmediato, pero que se puede utilizar (junto con otra información) para identificar a un individuo, también puede ser sensible y también puede ser necesario protegerla. Tal información puede incluir coordenadas GPS de la ubicación de un individuo, números de Identidad de Equipo de Estación Móvil Internacional (IEMI) que identifican el teléfono móvil de un individuo, información sobre afiliaciones religiosas, políticas o de organizaciones privadas e información médica, por nombrar algunos ejemplos.

La recopilación y/o el almacenamiento inapropiados de PII e información que se puede utilizar para identificar a los individuos se ha convertido en un problema de privacidad creciente, particularmente a medida que las empresas y los departamentos gubernamentales trasladan sus servicios en línea y requieren que los individuos divulguen la PII a través de redes y a través de intermediarios y proveedores de terceros, para que el individuo interactúe con la empresa o el departamento gubernamental.

Código mal escrito, falta de comprensión sobre qué información debería guardarse en las bases de datos o en la memoria caché (o en cookies y otros almacenes de información permanentes o semipermanentes asociados con transacciones basadas en la web), el uso de intermediarios que no son transparentes, y el código malicioso que está diseñado para capturar deliberadamente PII sensible, todos contribuyen a exponer a los individuos a la captura inapropiada y al posible uso inapropiado e ilegal de su PII.

Los gobiernos, en los últimos años, han implementado leyes y regulaciones en un intento de proporcionar garantías jurídicas y normas para proteger a los individuos. Sin embargo, dado que los sitios web, otros servicios en línea y las aplicaciones suelen ignorar, ignoran, están diseñados para eludir u operar a través de fronteras jurisdiccionales, leyes y regulaciones solo pueden proporcionar una protección limitada a los individuos.

En el artículo OLEKSII STAROV Y COL, "Extended Tracking Powers" (“Potencias de Rastreo Extendidas”), WORLD WIDE WEB, COMITÉ DIRECTIVO DE CONFERENCIAS INTERNACIONALES WORLD WIDE WEB, REPÚBLICA Y CANTÓN DE GINEBRA SUIZA, (20170403), doi:10.1145/3038912.3052596, ISBN 978-1-4503-4913-0 , páginas 1481 - 1490, XP058327240 [I] se describe que las extensiones tienen acceso a contenido y funcionalidad que no está disponible para las páginas web, y se proporciona un estudio a gran escala de la fuga de privacidad que permiten las extensiones.

El documento US2016/142423 A1 describe un método para detectar malware (“programa maligno”) en un flujo de red a al menos un ordenador principal. El método puede incluir la inicialización de un perfil de navegador correspondiente a un primer sitio web que tiene una primera fuente de sitio web y una primera pluralidad de características de contenido, por lo que se graba una primera pluralidad de características de contenido y se filtra el flujo de red en busca de una segunda característica de contenido dentro de una segunda pluralidad de características de contenido asociadas con un segundo sitio web. El método determina si la segunda característica de contenido coincide con una primera característica de contenido.

El documento US2017161520 describe un sistema y un método para determinar la información de identidad comprometida.

Teniendo estos problemas en mente es como se ha desarrollado la presente invención.

Compendio de la Invención

Este objeto se resuelve mediante un método que tiene las características de la reivindicación 1 y un sistema que tiene las características de la reivindicación 5.

En un primer aspecto, se proporciona un método para identificar software capaz de capturar información de identificación personal, que comprende las etapas de conectarse a un servidor remoto a través de un dispositivo local, enviar al menos una solicitud de datos y recuperar al menos un paquete de datos en respuesta a la solicitud, determinar si los datos solicitados recibidos se originaron en un servidor de terceros distinto del servidor remoto y, de ser así, determinar si los datos recibidos contienen evidencia de la presencia de información de identificación personal.

Según la invención, la etapa de determinar si los datos recibidos contienen información de identificación personal incluye la etapa de determinar si los datos recibidos incluyen datos que tienen un patrón de firma específico.

En una realización, el método comprende la etapa adicional de determinar si los datos tienen el patrón de firma específico comparando los datos con una base de datos de patrones de firma.

En una realización, el método incluye la etapa adicional de enviar datos con el patrón de firma específico a un módulo de identificación dispuesto para clasificar los datos en datos que contienen información de identificación personal o en datos que no contienen información de identificación personal.

En una realización, los datos recibidos se envían a un usuario para permitir que el usuario clasifique los datos, en donde que la clasificación elegida por el usuario se utiliza para proporcionar entrada al módulo de identificación.

En un segundo aspecto, se proporciona un sistema para identificar software capaz de capturar información de identificación personal, que comprende un módulo dispuesto para conectarse a un servidor remoto a través de un dispositivo local, estando dispuesto además el módulo para enviar al menos una solicitud de datos y recuperar al menos un paquete de datos en respuesta a la solicitud, y un módulo de identificación dispuesto para recibir al menos un paquete de datos y determinar si los datos recibidos se originaron en un servidor de terceros distinto del servidor remoto y, de ser así, determinar si los datos recibidos contienen evidencia de la presencia de información de identificación personal.

En un tercer aspecto, se proporciona un programa informático, que incluye al menos una instrucción y está dispuesto para ser ejecutable en un sistema informático, en donde, tras la ejecución, el sistema informático realiza el método según el primer aspecto de la invención.

En un cuarto aspecto, se proporciona una señal de datos, que incluye al menos una instrucción codificada y está dispuesta para ser recibida y ejecutada en un sistema informático, en donde, tras la recepción y ejecución, el sistema informático realiza el método según el primer aspecto de la invención.

Breve descripción de los dibujos

Otras características de la presente invención se describen con más detalle en la siguiente descripción de varias realizaciones no limitativas de la misma. Esta descripción se incluye únicamente con el propósito de ejemplificar la presente invención. No debería entenderse como una restricción al amplio compendio, exposición o descripción de la invención tal como se establece anteriormente. La descripción se hará con referencia a los dibujos adjuntos en los que:

La FIGURA 1 es un sistema informático ejemplar que es capaz de hacer funcionar un dispositivo, sistema, método y/o programa informático según una realización de la presente invención; y
la FIGURA 2 es un diagrama de flujo que ilustra un método según una realización de la invención.

Descripción detallada de las realizaciones preferidas

La presente invención se refiere en general a un sistema, método, programa informático y señal de datos para identificar un software capaz de capturar información de identificación personal. En particular, las realizaciones de la invención proporcionan un "complemento" de sitio web que es capaz de interactuar con un navegador web, aunque se entenderá que otras realizaciones pueden encontrar uso como aplicaciones de software independientes, o aplicaciones, que pueden configurarse para operar en cualquier sistema informático adecuado, incluyendo dispositivos móviles informáticos y de telecomunicaciones.

Más detalladamente, un aspecto de las realizaciones descritas en la presente memoria proporciona un método para identificar un software capaz de capturar información de identificación personal. El método comprende las etapas de, conectarse a un servidor remoto a través de un dispositivo local, enviar al menos una solicitud principal de datos y recuperar al menos un paquete de datos en respuesta a al menos una solicitud principal y determinar si hay solicitudes secundarias. generadas a partir de al menos una respuesta principal que se van a transmitir a un servidor de terceros

distinto del servidor remoto. Si es así, el método determina además si los datos enviados al servidor de terceros contienen evidencia de la presencia de información de identificación personal. Tal método puede implementarse como un programa informático, puede integrarse en un dispositivo de hardware (por ejemplo, un dispositivo portátil dispuesto para conectarse físicamente a un sistema informático) o puede codificarse en una señal de datos.

En otras palabras, un aspecto amplio de las realizaciones descritas en la presente memoria proporciona un método para identificar un software capaz de capturar información de identificación personal donde la captura inapropiada de información de identificación personal puede presentar riesgos financieros y de seguridad para los usuarios del sistema y/u otros miembros del público.

También se proporciona un sistema para identificar un software capaz de capturar información de identificación personal, que comprende un módulo dispuesto para conectarse a un servidor remoto a través de un dispositivo local, estando dispuesto además el módulo para enviar al menos una solicitud de datos y recuperar al menos un paquete de datos en respuesta a la solicitud, y un módulo de identificación dispuesto para recibir al menos un paquete de datos recuperado y determinar si los datos recibidos se originaron en un servidor de terceros distinto del servidor remoto y, de ser así, determinar si los datos recibidos recibido contienen evidencia de la presencia de información de identificación personal.

Una realización del método está codificada en un sistema informático, tal como el sistema informático mostrado en la FIGURA 1.

En la FIGURA 1 se ha mostrado un diagrama esquemático de un sistema informático, que en esta realización es un servidor 100 adecuado para utilizar con una realización de la presente invención. El servidor 100 se puede utilizar para ejecutar aplicaciones y/o servicios del sistema, tales como servicios de comercio electrónico, servicios bancarios o de seguros, servicios gubernamentales o cualquier otro servicio en el que se requiera que un usuario revele información de identificación personal para interactuar con el servidor.

Con referencia a la FIGURA 1, el servidor 100 puede comprender componentes adecuados necesarios para recibir, almacenar y ejecutar instrucciones informáticas apropiadas. Los componentes pueden incluir un procesador 102, una memoria 104 de solo lectura (ROM), una memoria 106 de acceso aleatorio (RAM), dispositivos de entrada/salida tales como unidades 108 de disco, dispositivos 110 de entrada remotos o conectados (tales como un dispositivo informático móvil, un teléfono inteligente o un ordenador personal de 'escritorio'), y uno o más enlaces 114 de comunicaciones.

El servidor 100 incluye instrucciones que pueden instalarse en la ROM 104, la RAM 106 o las unidades 112 de disco y pueden ser ejecutadas por el procesador 102. Puede proporcionarse una pluralidad de enlaces 114 de comunicaciones que pueden conectarse de diferentes maneras a uno o más dispositivos informáticos 110 tales como servidores, ordenadores personales, terminales, dispositivos informáticos portátiles o inalámbricos, o dispositivos de comunicación móvil tales como un teléfono móvil (celular). Al menos uno de una pluralidad de enlaces 114 de comunicaciones puede estar conectado a una red informática externa a través de una red de telecomunicaciones.

En una realización particular, el dispositivo puede incluir una base 116 de datos que puede residir en el dispositivo 112 de almacenamiento. Se entenderá que la base de datos puede residir en cualquier dispositivo de almacenamiento adecuado, que puede incluir unidades de estado sólido, unidades de disco duro, unidades ópticas o unidades de cinta magnética. La base 116 de datos puede residir en un solo dispositivo de almacenamiento físico o puede distribuirse entre múltiples dispositivos de almacenamiento.

El servidor 100 incluye un sistema operativo adecuado 118 que también puede residir en un dispositivo de almacenamiento o en la ROM del servidor 100. El sistema operativo está dispuesto para interactuar con la base de datos y con uno o más programas informáticos para hacer que el servidor realice las etapas, funciones y/o procedimientos requeridos.

En términos generales, las realizaciones de la invención se refieren a un método, sistema y programa informático (o una señal de datos) dispuestos para interactuar con el servidor a través de uno o más dispositivos remotos que están conectados al servidor a través de la red de comunicaciones. Los dispositivos remotos incluyen un software de "navegador" (es decir, un software capaz de reproducir Lenguaje de Marcas de Hipertexto (HTML) y tecnologías de navegador web asociadas, incluyendo programas de lenguaje JavaScript, Adobe Flash, Perl y otros métodos de entrada y salida de datos), y una realización de la invención adopta la forma de una aplicación "complementaria" (que de aquí en adelante se denominará aplicación de PII) que interactúa con el software del navegador. Sin embargo, se entenderá que la aplicación de PII también puede adoptar la forma de una aplicación independiente, y que la aplicación de PII también puede adoptar la forma de una aplicación basada en servidor, como se describirá con más detalle más adelante.

Con referencia ahora a la FIGURA 2, se ha mostrado un diagrama 200 de flujo para un método de operación de una aplicación de PII complementaria según una realización de la invención.

Se instancia una sesión de navegador de prueba y se dirige al navegador, ya sea manualmente o a través de una

secuencia de comandos automatizada instanciada y operada por la aplicación de PII, para acceder a una página web o serie de páginas web en el sitio web que se va a probar.

Todos los datos transmitidos desde la sesión de navegador de prueba al sitio web y recibidos por el navegador de prueba desde el sitio web se organizan para pasar a través de un servidor (1) de recopilación de datos http. El servidor de recopilación de datos http está configurado para registrar todas las instancias de datos enviados y recibidos por la sesión de navegador de prueba.

Todos los datos transmitidos a dominios de terceros (es decir, dominios no principales que son diferentes del dominio del sitio web) se filtran y se hacen pasar a un módulo (2) de descompresión y descodificación.

El módulo de descompresión y descodificación descomprime el paquete http y descodifica los datos que residen en los campos de cabecera, cookies, URL y cuerpos de solicitud.

El conjunto de datos descodificados se hace pasar luego a un módulo (3) de Detección de Señales de PII. El módulo utiliza una base de datos de patrones de señales para escanear “señales” que indiquen la presencia de datos de PII dentro del conjunto de datos. Un patrón de señal simplificado para detectar la presencia de una dirección de correo electrónico en un flujo de datos podría representarse como la expresión regular:

```
/\A([\w+\-].?)+@[a-z\d\-]+(\.[a-z]+)*\.[a-z]+\z/i
```

La salida se denomina "vector de señal" de PII. Un vector de señal de PII es una Máquina de Vectores de Soporte Multidimensional (SVM) - Sistema de Aprendizaje Automático donde cada dimensión refleja la presencia o ausencia de datos de PII específicos. Un vector de señal simple podría tener las siguientes dimensiones: género, dirección, ciudad, país, código postal y edad, por nombrar un ejemplo simple.

Sin embargo, en muchos casos de la vida real, el vector de señal puede tener muchas más dimensiones, tales como coordenadas de GPS, números de identificación, información relativa a las finanzas de un individuo, su historial de compras, etc., dependiendo del caso de uso específico de la realización. Por ejemplo, al completar una solicitud de préstamo en el sitio web de una institución financiera, el vector puede incluir dimensiones que están destinadas a capturar posibles "fugas" de PII, tales como activos y pasivos actuales, instituciones financieras actuales utilizadas por el individuo, incluso la cantidad de crédito que se busca. Se entenderá que un experto en la técnica entenderá los tipos de dimensiones requeridas para cualquier caso de uso particular, y tales variaciones están dentro del alcance de un experto en la técnica.

El vector de señal se hacer pasar al SVM, (4) que clasifica el vector para determinar si contiene datos de PII. Se entenderá que en el contexto de la presente memoria descriptiva, la SVM es un módulo de software que implementa un modelo de aprendizaje supervisado con uno o más algoritmos de aprendizaje asociados que analiza datos utilizados para clasificación y análisis de regresión. Se puede encontrar un manual básico sobre SVM en, por ejemplo, https://en.wikipedia.org/wiki/Supportvectormachine, que describe parte de la teoría subyacente que sustenta el funcionamiento de diferentes tipos de SVM conocidas.

Dado un conjunto de ejemplos de entrenamiento, cada uno marcado como perteneciente a una u otra de dos categorías, la SVM inicialmente utiliza un algoritmo de entrenamiento que crea un modelo de datos de PII y no PII. Luego, el algoritmo utiliza el modelo para comparar el vector con el modelo y, sobre la base de la comparación, asigna el vector a la categoría de datos de PII o a la categoría de datos de no PII, creando una clasificación lineal binaria no probabilística.

En otras palabras, en términos matemáticos (o estadísticos), el algoritmo SVM crea un modelo que es una representación de cada vector como un "punto en el espacio", mapeado para que los ejemplos de las categorías de datos de PII y de no PII se dividan por un espacio libre lo más amplio posible.

Luego, los nuevos vectores se mapean en el espacio existente y se predice que pertenecen a la categoría de PII o de no PII en base al lado del espacio en el que se encuentren. De esta manera, a medida que la aplicación de PII encuentra más ejemplos de datos de PII y de no PII, el modelo se refina con cada ejemplo, creando así un proceso de aprendizaje donde, con el tiempo, el conjunto de datos se vuelve más grande y, por lo tanto, estadísticamente más preciso.

En algunas realizaciones, particularmente donde hay un gran número de vectores, el módulo SVM puede residir en un sistema informático diferente al de la aplicación de PII. Por ejemplo, la aplicación de PII puede ser un complemento para un navegador web, pero puede comunicarse con un servidor remoto que contiene el módulo o la aplicación SVM. Esto puede ser necesario cuando se utiliza un algoritmo SVM complejo que requiere un poder de cómputo sustancial para categorizar correctamente cada vector.

Por supuesto, se entenderá que la SVM es solo un ejemplo de un algoritmo de aprendizaje y clasificación que se puede utilizar para categorizar datos de PII y de no PII. Otros ejemplos de algoritmos apropiados pueden incluir (sin limitación) algoritmos de árboles de decisión, algoritmos de redes neuronales, algoritmos de aprendizaje profundo, algoritmos de lógica inductiva, árboles de decisión (p. ej., bosque aleatorio), cuantificación de vectores de aprendizaje y algoritmos de aprendizaje basados en reglas.

Si no se encuentran datos de PII, el proceso termina y la falta de datos de PII se comunica al usuario.

Sin embargo, si el algoritmo, compuesto por el módulo de detección de señales de PII y el sistema de clasificación de aprendizaje automático, determina la presencia de datos de PII (5), entonces el sistema puede informar de este hallazgo directamente al usuario o puede, en ciertas realizaciones, enviar los datos al usuario o a un tercero, para que el usuario o el tercero puedan realizar una revisión. La revisión da como resultado que se proporcionen comentarios a la SVM, de modo que la SVM pueda "aprender" de la entrada proporcionada por el usuario y/u otra parte. Se entenderá que en el contexto de la realización descrita en la presente memoria, los términos "otra parte" y "otra parte" pueden referirse a una persona o pueden referirse a otra aplicación de software y/o sistema informático.

Los datos de entrenamiento (vectores de soporte) (6) también se pueden extraer del módulo de detección de señales con el fin de entrenar otras instancias de la aplicación de PII. Es decir, los datos de entrenamiento pueden cargarse en una base de datos central (no mostrada en las Figuras) que luego puede utilizar los datos de entrenamiento para proporcionar actualizaciones a otras instancias del software de PII. Un ejemplo de un conjunto de datos de entrenamiento simple se ha mostrado a continuación en la Tabla 1. El conjunto de datos se proporciona como un ejemplo simplificado para un experto en la técnica, y se entenderá que en una realización de la vida real, tal conjunto de datos sería más complejo. El presente ejemplo se proporciona únicamente en beneficio de la brevedad y la facilidad de comprensión, y no debe tomarse ninguna glosa del ejemplo y la tabla a continuación para limitar o variar de otro modo el significado simple de la invención reivindicada en la presente memoria, como lo entendería un experto en la técnica.

Tabla 1: Conjunto de Datos de Entrenamiento Simple

Vector	Datos de Entrenamiento	Pii detectado
10000	https://mydomain.com/unsubscribe?e=robin@email.com	Verdadero
00010	cd2=56%20young%20street	Falso
00000	c22=3146	Falso
00100	c23=25y	Falso
00011	v1=1%20Dover%20Road,Australia	Verdadero

Se entenderá que las realizaciones descritas anteriormente se han descrito como un complemento a un navegador web, pero que la descripción contempla otras formas equivalentes de implementación, tales como una aplicación independiente, una aplicación para un dispositivo móvil, un módulo que se incorpora en otro tipo de software (tal como un software antivirus), o cualquier otra forma adecuada en que se pueda implementar el concepto inventivo y la invención reivindicada.

Ventajas

Una de las ventajas de las realizaciones y la invención más amplia descrita en la presente memoria es que la invención elimina la responsabilidad de los consumidores (es decir, usuarios de redes informáticas, sitios web comerciales y gubernamentales, aplicaciones de comercio electrónico, etc.) de asumir la responsabilidad total de determinar si un sitio web, portal o aplicación cumple con los requisitos de privacidad y retención de datos. Siempre que los datos recibidos por el navegador se filtren a través de la aplicación de PII, la posibilidad de que la PII se publique accidentalmente o se recopile maliciosamente se reduce considerablemente.

Además, la aplicación de PII según una realización de la invención no requiere que el propietario o administrador de un sitio web permita que un usuario acceda a cualquier parte del sistema informático. Dicho de otro modo, la aplicación de PII puede proporcionar al usuario basándose únicamente en la información disponible públicamente proporcionada por el sitio web al navegador del usuario. No requiere ningún conocimiento interno del sitio web. Esto elimina la necesidad de que haya permisos otorgados por el propietario o administrador del sitio web y tampoco es necesario integrar ningún aspecto de la realización en un sitio web de destino.

Dado que los sitios web, los portales y las aplicaciones actualmente no tienen restricciones en la forma en que recopilan y procesan datos (desde una perspectiva técnica más que legal), se deduce que un usuario, cuando se le presenta un sitio web, tiene poco conocimiento o protección contra la captura o divulgación de PII, lo que a su vez

puede causar problemas de privacidad, seguridad, responsabilidad, técnicos y éticos. Por lo tanto, las realizaciones descritas en la presente memoria proporcionan una solución técnica mediante la cual los usuarios pueden realizar su propia evaluación de la idoneidad o legalidad de interactuar con un sitio web, portal o aplicación antes de decidir introducir su propia PII. En otras palabras, las realizaciones descritas en la presente memoria proporcionan una solución técnica a un problema que, en el pasado, se ha tratado por medios legales (es decir, no técnicos), que es una solución poco elegante que no hace nada para mejorar el problema de captura de PII inapropiada, sino que sólo puede buscar remediar el problema después del hecho.

Además, los desarrolladores y/u operadores de sitios web, portales y/o aplicaciones que recopilan PII pueden utilizar la aplicación de PII para probar su sitio web, portal y/o aplicación para asegurarse de que cumplen con todas las obligaciones legales y éticas para con los usuarios de su sitio web. La aplicación de PII proporciona a los operadores la capacidad de monitorizar, controlar y/o administrar la recopilación de PII. Como tal, la aplicación funciona como una protección para los usuarios finales, pero también de manera importante como una herramienta de control para los administradores y desarrolladores de sitios web.

Como corolario, los operadores que utilizan la aplicación de PII cumplirían con sus obligaciones éticas como proveedores de servicios y, por lo tanto, serían más atractivos para los consumidores y/o usuarios de su sitio web, portal y/o aplicación. Como tal, el uso de la aplicación de PII proporciona a los operadores la oportunidad de demostrar su "buena ciudadanía corporativa" y, así, aumentar su base de seguidores y/o clientes. Dado que la protección de la PII se vuelve no solo legalmente necesaria sino un problema más visible entre los usuarios del sitio web, el uso de la aplicación de PII también puede mejorar las preocupaciones de responsabilidad legal y/o reducir las primas de seguro para los operadores, lo que también proporciona ventajas financieras y de mercado además de las ventajas técnicas.

Descargos de responsabilidad

A lo largo de esta memoria descriptiva, a menos que el contexto requiera lo contrario, la palabra "comprende" o variaciones tales como "comprende" o "que comprende", se entenderá que implica la inclusión de un número entero indicado o grupo de números enteros, pero no la exclusión explícita de cualquier otro número entero o grupo de números enteros.

Los expertos en la técnica apreciarán que las realizaciones descritas en la presente memoria son susceptibles de variaciones y modificaciones obvias distintas de las descritas específicamente, y se pretende que las reivindicaciones más amplias cubran todas tales variaciones y modificaciones. Los expertos en la técnica también comprenderán que el concepto inventivo que sustenta las reivindicaciones más amplias puede incluir cualquier número de etapas, características y conceptos mencionados o indicados en la memoria descriptiva, ya sea de forma individual o colectiva, y cualquier combinación de cualquiera de las dos o más de las etapas o características pueden constituir una invención.

Cuando las definiciones de los términos seleccionados utilizados en la presente memoria se encuentran dentro de la descripción detallada de la invención, se pretende que tales definiciones se apliquen a la invención reivindicada. Sin embargo, si no se definen explícitamente, todos los términos científicos y técnicos utilizados en la presente memoria tienen el mismo significado que comúnmente entienden los expertos en la técnica a la que pertenece la invención.

Aunque no es necesario, las realizaciones descritas con referencia al método, programa informático, señal de datos y aspectos del sistema pueden implementarse a través de una interfaz de programación de aplicaciones (API), un kit de desarrollo de aplicaciones (ADK) o como una serie de bibliotecas de programas para su uso por parte de un desarrollador, para la creación de aplicaciones de software que se utilizarán en una o más plataformas o dispositivos informáticos, tales como un sistema operativo de terminal o de ordenador personal o un dispositivo informático portátil, un sistema operativo para teléfonos inteligentes o tabletas o dentro de una estructura de servidor más grande, tal como una "granja de datos" o dentro de un sistema de procesamiento de transacciones más grande.

En general, dado que los módulos de programa incluyen rutinas, programas, objetos, componentes y archivos de datos que realizan o ayudan en la realización de funciones, se entenderá que la funcionalidad del método, el programa informático y la señal de datos definidos en la presente memoria pueden distribuirse a lo largo de un número de rutinas, programas, objetos o componentes para lograr la misma funcionalidad que la realización y la invención más amplia reivindicada en la presente memoria. Tales variaciones y modificaciones están contempladas por el inventor y están dentro del alcance de los expertos en la técnica.

También se apreciará que cuando los métodos y sistemas de la presente invención y/o realizaciones se implementan mediante sistemas informáticos o se implementan parcialmente mediante sistemas informáticos, entonces se puede utilizar cualquier arquitectura de sistema informático apropiada sin apartarse del concepto inventivo. Esto incluye ordenadores independientes, ordenadores en red y dispositivos informáticos dedicados que no utilizan "software" como se entiende coloquialmente (tales como matrices de puertas programables en campo).

Cuando los términos "ordenador", "sistema informático" y "dispositivo informático" se utilizan en la memoria descriptiva, estos términos pretenden cubrir cualquier disposición apropiada de hardware informático para implementar el concepto inventivo y/o las realizaciones descritas en la presente memoria.

Cuando los términos "complemento", "aplicación de PII", "aplicación de software" y "aplicación" se utilizan en la memoria descriptiva cuando se hace referencia a una realización de la invención, estos términos pretenden cubrir cualquier software apropiado que sea capaz de realizar la funciones y/o lograr los resultados como se describe ampliamente en la presente memoria.

Cuando se haga referencia a normas, métodos y/o sistemas de comunicación, se entenderá que los dispositivos, servidores, etc., que constituyen la realización o interactúan con la realización pueden transmitir y recibir datos a través de cualquier mecanismo de hardware y protocolo de software adecuados, incluyendo los protocolos de comunicaciones alámbricos e inalámbricos, tales como, pero no limitados a, los protocolos de telecomunicaciones 2G, 3G y 4G, Wi-Fi, Bluetooth, otras frecuencias de radio, ópticas, acústicas, magnéticas, GPS/GPRS o cualquier otra forma o método de comunicación que puede estar disponible de vez en cuando.

REIVINDICACIONES

1. Un método para identificar un software capaz de capturar información de identificación personal, que comprende las etapas de:

conectarse a un servidor remoto a través de un dispositivo local,
conectarse al servidor remoto a través del dispositivo local se hace para enviar al menos una solicitud de datos y recuperar al menos un paquete de datos en respuesta a la solicitud; y
el método comprende además las etapas de:

determinar si los datos solicitados recibidos se originaron en un servidor de terceros distinto del servidor remoto y, de ser así:

determinar si los datos recibidos contienen evidencia de la presencia de información de identificación personal al determinar si los datos recibidos incluyen datos que tienen un patrón de firma específico, permitiendo de este modo determinar si la información de identificación personal introducida en el servidor remoto se comunicó al servidor de terceros sin el consentimiento informado del usuario.

2. Un método según la reivindicación 1, que comprende la etapa adicional de determinar si los datos tienen el patrón de firma específico comparando los datos con una base de datos de patrones de datos de firma.

3. Un método según la reivindicación 2, que comprende la etapa adicional de enviar datos con el patrón de firma específico a un módulo de identificación dispuesto para clasificar los datos en datos que contienen información de identificación personal o en datos que no contienen información de identificación personal.

4. Un método según cualquiera de las reivindicaciones 1 a 3, en donde los datos recibidos se envían a un usuario para permitir que el usuario clasifique además los datos, en donde la clasificación elegida por el usuario se utiliza para proporcionar entrada al módulo de identificación.

5. Un sistema para identificar un software capaz de capturar información de identificación personal, comprendiendo el sistema:

un módulo dispuesto para conectarse a un servidor remoto a través de un dispositivo local; y
un módulo de identificación;
el módulo está además dispuesto para enviar al menos una solicitud de datos y recuperar al menos un paquete de datos en respuesta a la solicitud; y
el módulo de identificación está dispuesto para:

recibir el al menos un paquete de datos solicitado, y
determinar si los datos recibidos se originaron en un servidor de terceros distinto del servidor remoto y, de ser así, determinar si los datos recibidos contienen evidencia de la presencia de información de identificación personal determinando si los datos recibidos incluyen datos que tienen un patrón de firma específico, permitiendo de este modo determinar si la información de identificación personal introducida en el servidor remoto se comunicó al servidor de terceros sin el consentimiento informado del usuario.

6. Un sistema según la reivindicación 5, en donde el módulo de identificación incluye además una rutina que determina si los datos tienen el patrón de firma específico comparando los datos con una base de datos de patrones de datos de firma.

7. Un sistema según la reivindicación 5 o la reivindicación 6, en donde el módulo de identificación envía datos con el patrón de firma específico a un módulo de identificación dispuesto para clasificar los datos en datos que contienen información de identificación personal o en datos que no contienen información de identificación personal.

8. Un sistema según cualquiera de las reivindicaciones 5 a 7, en donde el módulo de identificación envía datos recibidos a un usuario para permitir que el usuario clasifique los datos, en donde la clasificación elegida por el usuario se utiliza como entrada al módulo de identificación.

9. Un programa informático, que incluye al menos una instrucción y está dispuesto para ser ejecutable en un sistema informático, en donde, tras la ejecución, el sistema informático realiza el método de al menos una de las reivindicaciones 1 a 4.

10. Una señal de datos, que incluye al menos una instrucción codificada y dispuesta para ser recibida y ejecutada en un sistema informático, en donde, tras la recepción y ejecución, el sistema informático realiza el método de al menos una de las reivindicaciones 1 a 4.

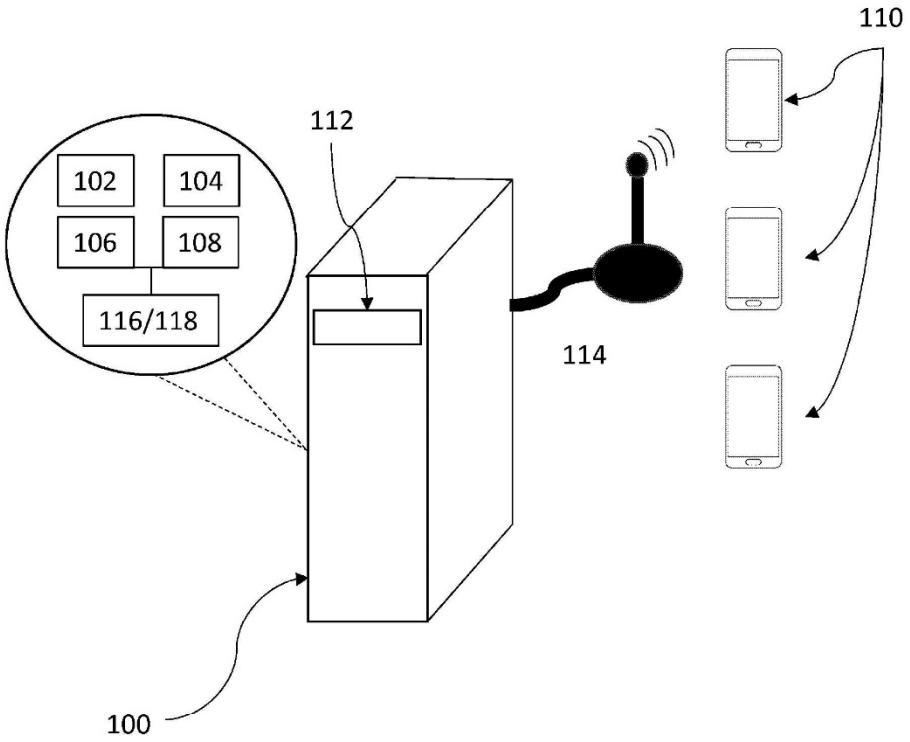

FIG. 1

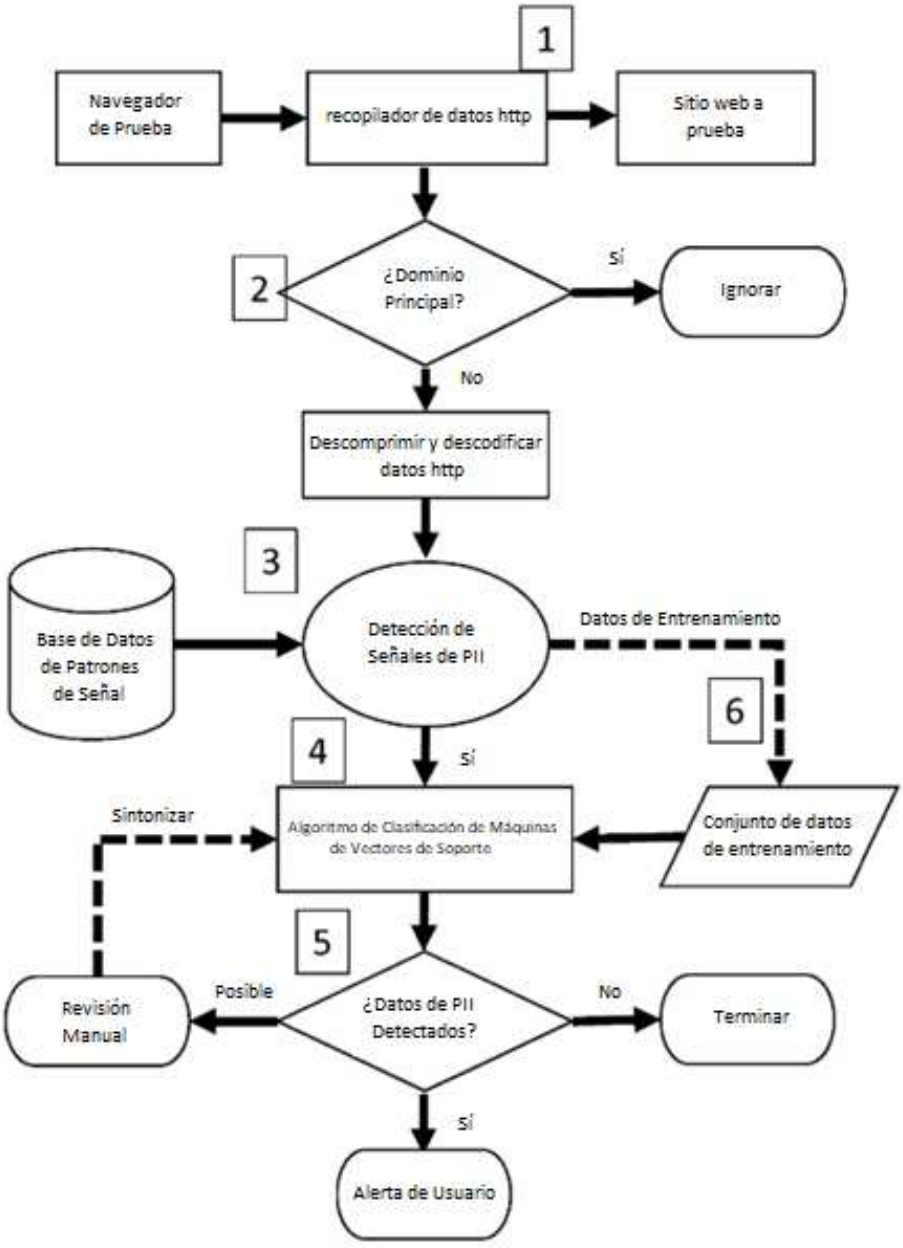
1
Navegador de Prueba
recopilador de datos http
Sitio web a prueba
2
¿Dominio Principal?
Sí
Ignorar
No
Descomprimir y descodificar datos http
3
Base de Datos de Patrones de Señal
Detección de Señales de PII
Datos de Entrenamiento
6
4
Sí
Sintonizar
Algoritmo de Clasificación de Máquinas de Vectores de Soporte
Conjunto de datos de entrenamiento
5
Posible
¿Datos de PII Detectados?
No
Revisión Manual
Terminar
Sí
Alerta de Usuario

FIG. 2