(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 986 176**

(51) Int. Cl.:

G06F 21/62 (2013.01)
G06F 11/34 (2006.01)
H04L 9/40 (2012.01)
H04L 67/02 (2012.01)
H04L 67/50 (2012.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **12.04.2022 E 22167889 (9)**

(97) Fecha y número de publicación de la concesión europea: **03.04.2024 EP 4261724**

(54) Título: **Anonimizador de sesiones**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
08.11.2024

(73) Titular/es:

RED BULL GMBH (100.0%)
Am Brunnen 1
5330 Fuschl am See, AT

(72) Inventor/es:

PLÖTZENEDER, MAXIMILIAN y
SCHMIRL, WOLFGANG

(74) Agente/Representante:

CURELL SUÑOL, S.L.P.

DESCRIPCIÓN

Anonimizador de sesiones

Campo técnico

La presente invención se refiere a un método implementado por ordenador para proporcionar información identificable personalmente anonimizada así como a un medio de almacenamiento no transitorio, legible por ordenador, correspondiente y a un sistema informático correspondiente.

Antecedentes

Mantener seguros los datos es un desafío tecnológico que afecta a un amplio espectro de tecnologías convencionales y nuevas en la era de la digitalización y los macrodatos. En particular, el creciente grado de interconexiones e intercambio de datos, tales como mediante Internet, genera una alta demanda para mantener segura la información identificable personalmente (PII).

Esto se aplica especialmente, aunque no de forma exclusiva, a herramientas de analítica web que están configuradas para realizar un seguimiento de, analizar y comunicar informes sobre el tráfico de sitios web. En el mundo actual, las herramientas de analítica web permiten realizar un seguimiento de usuarios individuales, almacenar datos de usuario en una base de datos y dejar que las organizaciones analicen registros de usuario individuales o combinados para obtener hallazgos, tales como hallazgos para optimizar la experiencia del usuario. Según la tecnología convencional ilustrada en la figura 1, un usuario de un dispositivo de cliente 1 lleva a cabo acciones en un sitio web y en cuanto el usuario da consentimiento, una herramienta de analítica de terceros 1000 alojada en un servidor en comunicación con el dispositivo de cliente 1 comienza a realizar un seguimiento de las acciones del usuario. En particular, los datos sobre los que se realiza el seguimiento, incluida la PII, tal como direcciones de IP, se envían directamente desde el sitio web en el dispositivo de cliente 1 al servidor 1000 que aloja la herramienta de analítica de terceros. El usuario puede apreciar la experiencia de usuario optimizada, pero también desea que su PII sea segura.

Los intentos convencionales de mejorar la seguridad de los datos de la PII en el contexto de herramientas de analítica web podan u ofuscan datos sobre los que se realiza el seguimiento de manera insuficiente o permiten de algún otro modo extraer conclusiones tanto sobre el usuario como sobre el dispositivo de cliente que se ha usado para visitar el sitio web.

"Web Tracking: Mechanisms, Implications, and Defenses" ["Seguimiento web: mecanismos, implicaciones y defensas"], ACTAS DEL IEEE, vol. 105, n.º 8, 1 de Agosto de 2017 (01-08-2017), de TOMASZ BUJLOW ET AL, describe que los servidores *proxy* son entidades intermediarias en una comunicación entre dos partes. Algunos de los servidores *proxy* ofrecen servicios de anonimización, que se fundamentan, entre otras cosas, en ocultar la dirección de IP del usuario o/y eliminar *cookies* de solicitudes HTTP.

Las solicitudes de patente US2010145960 A1 y US2021026990 A1 proporcionan enseñanzas adicionales referentes al campo técnico de la invención.

Sumario de la invención

Los inventores han reconocido que los intentos convencionales son insuficientes cuando se trata de mantener segura la información identificable personalmente (PII). La información identificable personalmente identifica de forma directa o indirecta un dispositivo de cliente y/o un usuario del dispositivo de cliente, de manera que es necesario un grado mayor de seguridad de los datos para la PII. La PII puede incluir, por ejemplo, un identificador de sesión asociado a un usuario de un dispositivo de cliente, una parte de una dirección de IP del dispositivo de cliente, un agente de usuario de un navegador utilizado por un usuario en el dispositivo de cliente, una dirección de contacto asociada a un usuario del dispositivo de cliente, un nombre asociado a un usuario del dispositivo de cliente, una versión del navegador y/u otros identificadores.

En particular, los intentos existentes seguirían permitiendo que editores o terceros estableciesen una conexión entre el cliente respectivo (por ejemplo, el navegador Google Chrome) y la herramienta de analítica web (por ejemplo, Google Analytics) al acceder a ambos simultáneamente. Por ejemplo, esto se puede lograr a través de un seguimiento en tiempo real, que muestra el sitio web que visita un usuario en el navegador del cliente, así como la totalidad del resto de datos correspondientes transmitidos a la herramienta de analítica web. De esta manera, se sigue pudiendo acceder a la PII de forma no deseada.

De este modo, un objetivo de la presente invención es reducir la probabilidad de acceso no autorizado a información identificable personalmente.

Este objetivo se alcanza con el método que define la reivindicación independiente 1, con el por lo menos un medio

de almacenamiento no transitorio legible por ordenador que define la reivindicación independiente 20 y con el sistema informático que define la reivindicación independiente 21. Algunas de las formas de realización preferidas se describen en las reivindicaciones dependientes, en la descripción y en las figuras.

Según un aspecto general de la presente invención, se proporciona un método para proporcionar información identificable personalmente anonimizada, en el que el método se implementa mediante uno o más primeros ordenadores, comprendiendo el método: obtener una primera solicitud, incluyendo la primera solicitud primeros datos que indican acciones que ocurrieron en un dispositivo de cliente y segundos datos que están asociados a los primeros datos y que se basan en información identificable personalmente asociada al dispositivo de cliente, en donde los segundos datos incluidos en la primera solicitud obtenida son datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente asociada con el dispositivo de cliente o son la información identificable personalmente asociada con el dispositivo de cliente; poner en cuarentena por lo menos los segundos datos, en donde la puesta en cuarentena por lo menos incluye almacenar por lo menos los segundos datos en un módulo de almacenamiento de datos; recuperar datos del módulo de almacenamiento de datos, en donde los datos recuperados se basan en los segundos datos almacenados; generar una tercera solicitud para su transmisión a un servidor externo con respecto a dicho uno o más ordenadores, en donde la tercera solicitud incluye los primeros datos y datos sintéticos asociados a los primeros datos, en donde los datos sintéticos de la tercera solicitud se basan en los datos recuperados, y en donde los datos sintéticos de la tercera solicitud son los datos sintéticos de la primera solicitud o se sintetizaron sobre la base de la información identificable personalmente incluida en la primera solicitud. Los datos sintéticos incluyen una versión ofuscada, reducida y/o a la que se ha aplicado una función *hash*, de la información identificable personalmente. Los datos recuperados se recuperan del módulo de almacenamiento de datos, o la tercera solicitud es transmisible al servidor externo, solo en cuanto o solo después de que expire una asociación de por lo menos una parte de la información identificable personalmente al dispositivo de cliente o su usuario. El método incluye, además, transmitir la tercera solicitud generada al servidor externo.

Los datos recuperados del módulo de almacenamiento de datos pueden ser los segundos datos (almacenados) (tales como la PII o los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente antes de su almacenamiento en calidad de segundos datos en el módulo de almacenamiento de datos), o pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente de los segundos datos almacenados. Los datos sintéticos incluidos en la tercera solicitud pueden ser los datos recuperados del módulo de almacenamiento de datos o pueden ser datos sintéticos que se sintetizaron sobre la base de información identificable personalmente de los datos recuperados.

En particular, aspectos de la presente invención ponen en cuarentena datos, que están asociados a datos que indican acciones que ocurrieron en un dispositivo de cliente (datos sobre los cuales se realiza un seguimiento) y que se basan en PII, antes de poner estos datos a disposición para su transmisión a un servidor externo, tal como un Servidor que aloja un *software* o herramienta de analítica web. Por consiguiente, se reduce significativamente la probabilidad de establecer una conexión entre un cliente (por ejemplo, un navegador) y un servidor externo (por ejemplo, una herramienta de analítica web) al acceder a ambos simultáneamente. En particular, se dificulta significativamente el seguimiento en tiempo real, que muestra el sitio web que está visitando un usuario en el dispositivo de cliente, así como la totalidad del resto de datos correspondientes transmitidos al servidor externo (por ejemplo, *software* de analítica web). Por consiguiente, aspectos de la presente invención reducen la probabilidad de accesos no autorizados a información identificable personalmente.

La cuarentena puede finalizar después de la expiración de una cantidad de tiempo predeterminada. Según aspectos preferidos, los datos recuperados se recuperan del módulo de almacenamiento de datos, tal como para (preparar) la transmisión de la tercera solicitud al servidor externo, o la tercera solicitud es transmisible al servidor externo, solo después de que haya expirado una cantidad de tiempo predeterminada desde un tiempo en el que ocurrieron parte de las acciones (por ejemplo, las primeras) en el dispositivo de cliente. Al involucrar la cuarentena para los datos durante una cantidad de tiempo predeterminada, la cantidad de tiempo se puede determinar de tal manera que la probabilidad antes descrita de acceso no autorizado se reduzca adicionalmente o incluso se minimice a cero. En aspectos particularmente ventajosos de la presente invención, la cantidad de tiempo predeterminada está en concordancia con la expiración, o supresión permanente en el dispositivo de cliente, de una asociación de por lo menos una parte de la información identificable personalmente al dispositivo de cliente. Es decir, la cantidad de tiempo se puede determinar de tal manera que los datos que pasaron por la cuarentena se transmitan únicamente al servidor externo (por ejemplo, un *software* de analítica web), en cuanto o después de que haya expirado o se haya suprimido permanentemente en el dispositivo de cliente una o cualquier asociación entre la PII y el usuario o el cliente, tal como un identificador (por ejemplo, *Cookie*). Es decir, al producirse la recepción de los datos sintéticos, que se sintetizaron sobre la base de la PII, el servidor externo (por ejemplo, el *software* de analítica web) no puede asociar el usuario o el dispositivo de cliente a los datos recibidos, ni siquiera si el servidor externo pudiera aplicar ingeniería inversa a la PII a partir de los datos sintéticos recibidos. Por consiguiente, la información identificable personalmente incluso puede mantenerse absolutamente segura hasta un nivel incluso que cumpla con el Reglamento General de Protección de Datos (GDPR). En particular, aspectos de la invención también permiten enviar datos a cualquier servidor externo que aloje cualquier *software* de analítica, con independencia de la ubicación del servidor externo.

Para las ventajas antes descritas de los aspectos de la presente invención, no es esencial si los datos sintéticos se sintetizaron sobre la base de la PII por parte de un proveedor tercero externo o mediante aspectos de la presente invención. Además, no es esencial para la presente invención si los datos sintéticos se sintetizaron basándose en la PII mediante aspectos de la presente invención antes o después del almacenamiento de los datos en el módulo de almacenamiento de datos. En otras palabras, no es esencial si los datos almacenados en el módulo de almacenamiento de datos ya son datos sintéticos que se sintetizaron basándose en la PII o si los datos almacenados en el módulo de almacenamiento de datos todavía incluyen PII que se convertirá en datos sintetizados después de la recuperación de los datos del módulo de almacenamiento de datos y antes de la transmisión al servidor externo. Todas estas materializaciones de la síntesis de datos son compatibles con o quedan cubiertas por la presente invención.

Según un aspecto preferido de la presente invención, la síntesis de datos tiene lugar antes del almacenamiento de los datos en el módulo de almacenamiento de datos. En particular, los segundos datos incluidos en la primera solicitud obtenida y/o los datos recuperados del módulo de almacenamiento de datos pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente asociada con el dispositivo de cliente. En ese caso, el método también puede incluir: mediante un segundo o segundos ordenadores, recibir una segunda solicitud del dispositivo de cliente, en donde la segunda solicitud incluye los primeros datos y la información identificable personalmente, y generar la primera solicitud sobre la base de la segunda solicitud, en donde la generación de la primera solicitud incluye sintetizar los datos sintéticos incluyendo convertir por lo menos parte de la información identificable personalmente incluida en la segunda solicitud en los datos sintéticos e incluir en la primera solicitud los datos sintéticos en lugar de la información identificable personalmente. Dicho uno o más segundos ordenadores pueden encontrarse entre dicho uno o más primeros ordenadores, o el segundo o segundos ordenadores pueden ser externos pero estar acoplados comunicativamente a dicho uno o más primeros ordenadores, tal como un servidor sintetizador que esté ubicado aguas arriba del módulo de almacenamiento de datos. Hacer que la síntesis de datos tenga lugar antes del almacenamiento de los datos en el módulo de almacenamiento de datos es particularmente ventajoso en términos de seguridad de los datos, ya que los datos que se almacenan en el módulo de almacenamiento de datos y se procesan posteriormente antes de la transmisión al servidor externo ya se han liberado de parte o la totalidad de la PII, de tal manera que obtener acceso a los datos no es suficiente de por sí para establecer una relación de los datos con un cliente o usuario en particular. Es decir, la anonimización de la PII se lleva a cabo antes y se reduce el riesgo de filtración de la PII, tal como a través de un ataque sobre el módulo de almacenamiento de datos. Asimismo, se puede obtener una vista previa de los datos sintéticos antes de almacenar los datos en el módulo de almacenamiento de datos, lo cual permite reducir errores cuando el sistema se somete a prueba o cuando se procesan los datos antes de la transmisión al servidor externo.

Según otro aspecto preferido de la presente invención, la síntesis de datos tiene lugar después del almacenamiento de los datos en el módulo de almacenamiento de datos. En particular, los segundos datos incluidos en la primera solicitud obtenida y/o los datos recuperados del módulo de almacenamiento de datos son la información identificable personalmente asociada con el dispositivo de cliente. La generación de la tercera solicitud también puede incluir entonces: sintetizar los datos sintéticos, incluyendo convertir por lo menos parte de los segundos datos o datos recuperados en los datos sintéticos e incluir en la tercera solicitud los datos sintéticos en lugar de los segundos datos o datos recuperados. Hacer que la síntesis de datos tenga lugar después del almacenamiento de los datos en el módulo de almacenamiento de datos tiene otras ventajas. Por ejemplo, los recursos informáticos requeridos para la síntesis de datos pueden reservarse hasta que los datos se recuperen realmente del módulo de almacenamiento de datos para su transmisión al servidor externo. Asimismo, se pueden liberar de carga componentes de procesado de datos (por ejemplo, un receptor) aguas arriba del módulo de almacenamiento de datos, que tendrán entonces mejores tiempos de reacción y proporcionarán una mejor experiencia de usuario en la parte de presentación del sistema. Además, normalmente el tamaño de los datos sintéticos es mayor que el de los datos originales, de manera que se puede reducir la cantidad de memoria requerida en el módulo de almacenamiento de datos.

Según un aspecto de la presente invención, la puesta en cuarentena puede incluir, además, la recuperación de los datos recuperados o segundos datos del módulo de almacenamiento de datos (por ejemplo, los datos almacenados se pueden recuperar durante la cuarentena, tal como para la síntesis de los datos sintéticos) y/ o la síntesis de los datos sintéticos. Los segundos datos pueden recuperarse del módulo de almacenamiento de datos para sintetizar los datos sintéticos sobre la base de los segundos datos recuperados. A continuación, o bien los datos sintéticos se pueden almacenar en el módulo de almacenamiento de datos antes de que los datos sintéticos se recuperen del módulo de almacenamiento de datos (por ejemplo, en forma de datos recuperados que se basan en los segundos datos almacenados) y reenviarse al sistema con vistas a su inclusión en la tercera solicitud para su transmisión al servidor externo, o bien los datos sintéticos no se almacenan en el módulo de almacenamiento de datos, sino que se prefiere reenviarlos directamente al sistema con vistas a su inclusión en la tercera solicitud para su transmisión al servidor externo.

En cualquier caso, la tercera solicitud no es transmisible al servidor externo durante la puesta en cuarentena hasta que finalice la puesta en cuarentena. Por ejemplo, los datos que se basan en los segundos datos almacenados y

que van a recuperarse del módulo de almacenamiento de datos, tal como a efectos de (preparar) la transmisión de la tercera solicitud al servidor externo, pueden no ser recuperables durante la puesta en cuarentena hasta que finalice la puesta en cuarentena. Es decir, también es posible que los datos que se basan en los segundos datos almacenados y que van a recuperarse del módulo de almacenamiento de datos únicamente se pongan a disposición para su recuperación del módulo de almacenamiento de datos, tal como a efectos de (preparar) la transmisión de la tercera solicitud lo cual puede ser después de que se sinteticen los datos sintéticos, una vez o después de que ya haya finalizado la puesta en cuarentena, es decir, que ya haya terminado la cuarentena. Alternativamente, los datos que se basan en los segundos datos almacenados y que van a recuperarse del módulo de almacenamiento de datos, tal como a efectos de (preparar) la transmisión de la tercera solicitud al servidor externo, pueden ser recuperables durante la puesta en cuarentena y antes de que finalice la puesta en cuarentena, pero la tercera solicitud solo se pone a disposición para su transmisión al servidor externo una vez que finalice la puesta en cuarentena, es decir, que termine la cuarentena. En otras palabras, siempre que la tercera solicitud no se envíe al servidor externo que aloja la analítica de terceros antes de que finalice la puesta en cuarentena, se pueden recuperar datos del módulo de almacenamiento de datos, tal como a efectos (únicamente) de la síntesis de los datos sintéticos, antes de que finalice la cuarentena. En por lo menos algunos de estos casos, la recuperación de los segundos datos o datos recuperados del módulo de almacenamiento de datos y/o la síntesis de los datos sintéticos forman parte de la puesta en cuarentena.

Con independencia de si la síntesis de datos tiene lugar antes o después del almacenamiento de los segundos datos en el módulo de almacenamiento de datos, la expiración de la cantidad de tiempo predeterminada, tal como el momento en el que expira o se suprime permanentemente en el dispositivo de cliente la asociación de por lo menos una parte de la información identificable personalmente al dispositivo de cliente (por ejemplo, el identificador, tal como una *Cookie*), se puede coordinar preferentemente con la recuperación de los datos del módulo de almacenamiento de datos a efectos de la transmisión de la tercera solicitud al servidor externo o con la transmisión de la tercera solicitud al servidor externo. Es decir, los datos únicamente se pueden recuperar del módulo de almacenamiento de datos a efectos de la transmisión de la tercera solicitud al servidor externo en cuanto se produzca o después de la expiración de la cantidad de tiempo predeterminada o la tercera solicitud solo se transmite al servidor externo en cuanto se produzca o después de la expiración de la cantidad de tiempo predeterminada. Antes de este vencimiento o expiración, con medios apropiados (por ejemplo, bloqueos) se impiden las operaciones correspondientes (recuperación de datos del módulo de almacenamiento de datos o transmisión de la tercera solicitud al servidor externo). La expiración de la cantidad de tiempo predeterminada puede coincidir con la expiración o la supresión permanente, en el dispositivo de cliente, de la asociación de por lo menos una parte de la PII al dispositivo de cliente. La expiración de la cantidad de tiempo predeterminada puede coincidir con la expiración o la supresión permanente, en el dispositivo de cliente, de cualquier o un identificador, tal como una *Cookie*, asociado a por lo menos una parte de la información identificable personalmente.

Aspectos de la presente invención proporcionan una o más de las siguientes ventajas. Los proveedores de plataformas digitales pueden realizar seguimientos de, modificar y almacenar datos, que son generados por consumidores y/o leídos de sus dispositivos, de una manera que es improbable que se pueda hacer referencia personal a un individuo o dispositivo, logrando así una anonimización mejorada de la PII. Algunos aspectos de la presente invención hacen uso de la posibilidad de aislar datos (cuarentena), preferentemente hasta que ya no se pueda deducir el dispositivo y/o la persona que los está utilizando. En otras palabras, no se transmiten datos al servidor externo hasta que se haya eliminado cualquier tipo de asociación o identificador, lo cual mantiene segura la PII. Adicionalmente, aspectos de la presente invención especifican medios particularmente ventajosos para sintetizar datos sintéticos sobre la base de la PII para lograr un grado potenciado adicional de anonimización. De hecho, aspectos de la presente invención posibilitan una anonimización total de la PII. Aspectos de la presente invención proporcionan medidas de seguridad de datos para garantizar que solo se reciba, procese y almacene información identificable personalmente anonimizada en servidores externos, lo cual resulta especialmente ventajoso cuando se usan herramientas de analítica en dichos servidores externos.

A partir de la siguiente descripción se pondrán de manifiesto otras ventajas y el anterior sumario no pretende limitar el alcance de la invención.

Breve descripción de los dibujos

Se describirán diversos aspectos y formas de realización en referencia a las siguientes figuras. Cabe observar que las figuras no se han dibujado necesariamente a escala. Los elementos que aparecen en múltiples figuras se indican con el mismo número de referencia en todas las figuras en las que aparecen.

La figura 1 ilustra un sistema convencional que involucra una interacción entre un cliente y una analítica de terceros según la técnica anterior.

La figura 2 ilustra un sistema según una primera forma de realización de la presente invención en el que la síntesis de datos tiene lugar antes de que se almacenen los datos en el módulo de almacenamiento de datos.

La figura 3 ilustra un eje en el tiempo que es aplicable igualmente a la primera forma de realización y a la

segunda forma de realización.

La figura 4 ilustra un sistema según la primera forma de realización de la presente invención en el que los datos se recuperan del módulo de almacenamiento de datos una vez que se ha producido la expiración del tiempo T1 o T2 de la figura 3 o después de dicha expiración.

La figura 5 ilustra un sistema de acuerdo con una segunda forma de realización de la presente invención en el que la síntesis de datos tiene lugar después de que se almacenen los datos en el módulo de almacenamiento de datos.

La figura 6 ilustra un sistema según una versión de la segunda forma de realización de la presente invención en el que los datos se recuperan del módulo de almacenamiento de datos una vez que se ha producido la expiración del tiempo T1 o T2 de la figura 3 o después de dicha expiración.

La figura 7 ilustra un sistema según otra versión de la segunda forma de realización de la presente invención en el que los datos se recuperan del módulo de almacenamiento de datos antes de la expiración del tiempo T1 o T2 de la figura 3, pero en la que los datos sintéticos sintetizados por el sintetizador se transmiten a la analítica de terceros una vez que se ha producido la expiración del tiempo T1 o T2 de la figura 3 o después de dicha expiración.

La figura 8 ilustra más detalles sobre el sistema según la primera y segunda formas de realización de la presente invención en el que la síntesis de datos se ilustra como opcional y puede tener lugar en uno o más cualesquiera de los diversos componentes del sistema antes o después de que los datos se almacenen en el módulo de almacenamiento de datos.

La figura 9 ilustra un objeto de datos según la invención que mapea diferentes valores correspondientes a un elemento de información identificable personalmente con un mismo valor sintetizado.

Descripción detallada

La figura 2 ilustra un sistema según una primera forma de realización de la presente invención en el que tiene lugar una síntesis de datos antes de que los datos se almacenen en el módulo de almacenamiento de datos.

Un método llevado a cabo por el sistema de la figura 2 se refiere a la provisión de información identificable personalmente anonimizada, en el que el método se implementa mediante uno o más primeros ordenadores del sistema (servidor 100 de la figura 8), comprendiendo el método: obtener una primera solicitud 110, incluyendo la primera solicitud 110 primeros datos que indican acciones que ocurrieron en un dispositivo de cliente 1 y segundos datos que están asociados a los primeros datos y que se basan en información identificable personalmente asociada con el dispositivo de cliente 1; poner en cuarentena por lo menos los segundos datos, en donde la puesta en cuarentena por lo menos incluye almacenar por lo menos los segundos datos en un módulo de almacenamiento de datos 102; recuperar datos del módulo de almacenamiento de datos 102, en donde los datos recuperados se basan en los segundos datos almacenados; y generar una tercera solicitud 130 para su transmisión a un servidor 1000 externo a dicho uno o más ordenadores, en donde la tercera solicitud 130 incluye los primeros datos y datos sintéticos asociados a los primeros datos, en donde los datos sintéticos se basan en los datos recuperados y se sintetizaron sobre la base de la información identificable personalmente. Los segundos datos pueden ser o incluir la información identificable personalmente o pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente. Los datos recuperados del módulo de almacenamiento de datos pueden ser los segundos datos (almacenados) (tales como la PII o los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente antes de su almacenamiento en calidad de segundos datos en el módulo de almacenamiento de datos), o pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente de los segundos datos almacenados. Los datos sintéticos incluidos en la tercera solicitud pueden ser los datos recuperados del módulo de almacenamiento de datos o pueden ser datos sintéticos que se sintetizaron sobre la base de información identificable personalmente de los datos recuperados. Los datos sintéticos pueden ser una versión anonimizada, reducida y/u ofuscada de la información identificable personalmente.

El servidor externo 1000 se puede configurar para alojar un *software* de herramienta de analítica de terceros para analizar acciones que ocurrieron en el dispositivo de cliente 1.

Según la primera forma de realización, los segundos datos incluidos en la primera solicitud 110 obtenida y/o los datos recuperados del módulo de almacenamiento de datos 102 son los datos sintéticos que fueron sintetizados por el sintetizador 10 sobre la base de la información identificable personalmente asociada con el dispositivo de cliente 1.

La síntesis de datos sintéticos la puede llevar a cabo un sintetizador 10. El sintetizador 10 puede alojarse en un servidor de terceros externo al sistema subyacente a la presente invención. Alternativamente, el sintetizador 10

puede alojarse en uno o más ordenadores del sistema según la invención.

El método llevado a cabo por el sistema puede incluir opcionalmente, además, las operaciones asociadas a la síntesis de datos según es llevada a cabo por el sintetizador 10. El método llevado a cabo por el sistema según la primera forma de realización puede comprender además, entonces: por parte de uno o más segundos ordenadores del sistema, recibir una segunda solicitud 20, 120 del dispositivo de cliente 1, en donde la segunda solicitud 20, 120 incluye los primeros datos y la información identificable personalmente, y generar la primera solicitud 110 sobre la base de la segunda solicitud 20, 120, en donde la generación de la primera solicitud 110 incluye sintetizar los datos sintéticos incluyendo convertir en los datos sintéticos por lo menos parte de la información identificable personalmente incluida en la segunda solicitud 20, 120 e incluir en la primera solicitud 110 los datos sintéticos en lugar de la información identificable personalmente.

La segunda solicitud 20, 120 puede ser una primera solicitud de seguimiento generada en el dispositivo de cliente 1 para el seguimiento de las acciones de un usuario en el dispositivo de cliente 1. Dicho uno o más segundos ordenadores se encuentran entre el primer o primeros ordenadores, o en donde dicho uno o más segundos ordenadores del sistema son externos con respecto a dicho uno o más primeros ordenadores pero están acoplados comunicativamente a los mismos.

La figura 3 ilustra un eje en el tiempo que es aplicable igualmente a la primera forma de realización y a la segunda forma de realización descrita más adelante.

Los datos que se transmiten en tiempo real podrían infringir la seguridad de los datos y/o reglamentos legales (GDPR), en la medida en que una comprobación simultánea tanto en el dispositivo de cliente como en la base de datos de la herramienta de analítica podría identificar potencialmente a un usuario y/o dispositivo. Para impedir dicha identificación, aspectos de la presente invención ofrecen el siguiente planteamiento aplicable a cada uno de los aspectos y formas de realización descritos en la presente.

Eventos y datos que se han generado en el cliente 1 se envían desde el cliente 1, tal como desde el navegador o aplicaciones del cliente, al sistema de la presente invención donde los datos se almacenarán temporalmente en un módulo de almacenamiento de datos 102 para preferentemente solo ser reenviados al servidor de analítica 1000 una vez que se pueda garantizar que en el dispositivo 1 del cliente se ha eliminado permanentemente cualquier identificador.

Es decir, el sistema de la presente invención almacenará preferentemente todos los datos hasta que se hayan eliminado los identificadores (por ejemplo, *cookies*), y solo entonces los reenvía al servidor de analítica 1000 (por ejemplo, Google Analytics).

La expiración de la sesión del lado del cliente en T1 en la figura 3 significa que el dispositivo de cliente, en T1, dejará de producir incidencias (datos de acciones) asociadas a un ID de sesión específico que son transmitidas por el cliente. La expiración de la sesión del lado del servidor en T2 en la figura 3 significa que el sistema de esta invención deja de aceptar incidencias nuevas en T2 y considera que las incidencias están listas para su expedición/transmisión al servidor externo 1000. T3 puede ser un retardo aceptado máximo de datos dentro de la analítica de terceros antes de que se pierdan los datos y/o T3 puede venir dado por un momento en el que el servidor externo (*software* de herramienta de analítica) deja de aceptar solicitudes o incidencias nuevas atribuidas a un ID de sesión específico. T1 a T3 son lapsos de tiempo respectivos después de T0.

En T0, la sesión puede iniciarse en el dispositivo de cliente 1. En otras palabras, entre el tiempo T0 y T1, se producen incidencias (datos de acciones) en el dispositivo de cliente 1 por acciones llevadas a cabo por el usuario en el dispositivo de cliente (por ejemplo, clics del usuario en iconos de una página web). En T1 como muy tarde expirarán o se eliminarán la *cookie* de ID de sesión, u otros identificadores o asociaciones entre el cliente/usuario y la PII. Después de T1, no se generan incidencias nuevas con el mismo ID de sesión.

En o después de T1, tal como a más tardar en T2, es decir, una cierta cantidad de tiempo de seguridad tras T1, el sistema de la invención puede finalizar la cuarentena de los datos recogidos. Puede que los datos almacenados no sean recuperables a partir del módulo de almacenamiento de datos 102 (por ejemplo, bloqueo de la recuperación de datos) a efectos de la transmisión de la tercera solicitud al servidor externo durante la cuarentena hasta que los datos almacenados se hayan puesto explícitamente a disposición (por ejemplo, levantamiento del bloqueo) para la recuperación a partir del módulo de almacenamiento de datos 102, tal como para una síntesis seguida por una transmisión de la tercera solicitud 130. Alternativamente, los datos pueden ser recuperables en el módulo de almacenamiento de datos durante la cuarentena pero puede que no sean transmisibles al servidor externo 1000 durante la cuarentena hasta que la tercera solicitud se ponga a disposición para su transmisión al servidor externo. Por ejemplo, es posible que los datos se pongan a disposición para su recuperación a partir del módulo de almacenamiento de datos a efectos de la transmisión de la tercera solicitud, lo cual puede ser después de que se sinteticen los datos sintéticos, solo cuando la puesta en cuarentena ya haya finalizado en T1 o T2. Alternativamente, los datos que se basan en los segundos datos almacenados y que se van a recuperar del módulo de almacenamiento de datos, tal como a efectos de preparar la transmisión de la tercera solicitud al servidor

externo, pueden ser recuperables durante la puesta en cuarentena, y antes de que finalice la puesta en cuarentena en T1 o T2, pero la tercera solicitud solo se pone a disposición para su transmisión hacia el servidor externo una vez que ha finalizado la puesta en cuarentena en T1 o T2. En otras palabras, siempre que la tercera solicitud 130 no se envíe al servidor externo 1000 que aloja la analítica de terceros antes de que termine la puesta en cuarentena en T1 o T2, se pueden recuperar datos del módulo de almacenamiento de datos 102, tal como para la síntesis de los datos sintéticos, antes de que termine la cuarentena en T1 o T2. En ese caso, la recuperación de los segundos datos a partir del módulo de almacenamiento de datos 102 y/o la síntesis de los datos sintéticos pueden formar parte de la puesta en cuarentena.

Por ejemplo, los datos recuperados se recuperan del módulo de almacenamiento de datos 102, o la tercera solicitud 130 es transmisible al servidor externo 1000, únicamente después de que haya expirado una cantidad de tiempo predeterminada T1 o T2 desde un tiempo T0 en el que ocurrieron parte de las acciones (por ejemplo, las primeras) en el dispositivo de cliente 1. Los primeros datos (datos de acciones) están asociados a un sello de tiempo y en donde dicho uno o más primeros ordenadores tienen acceso a un reloj para determinar la expiración de la cantidad de tiempo predeterminada (T1 o T2).

Por ejemplo, los segundos datos se recuperan del módulo de almacenamiento de datos 102 para la transmisión de la tercera solicitud al servidor externo, o la tercera solicitud 130 es transmisible al servidor externo 1000, solamente en T1 o T2 cuando expire o se haya suprimido permanentemente en el dispositivo de cliente 1 una asociación de por lo menos una parte de la información identificable personalmente al dispositivo de cliente 1. La asociación de la parte de la información identificable personalmente al dispositivo de cliente 1 expiró o se suprimió permanentemente en el dispositivo de cliente 1 haciendo que un programa informático, tal como una aplicación, cambiase en el dispositivo de cliente 1 un valor de la parte de la información identificable personalmente a un valor nuevo.

Los segundos datos se pueden almacenar en el módulo de almacenamiento de datos 102 en asociación con un sello de tiempo que está asociado a las acciones que ocurrieron en el dispositivo de cliente 1, en donde el tiempo en el que la asociación o el identificador expiró o se suprimió permanentemente en el dispositivo de cliente 1 se determina sobre la base del sello de tiempo. El sello de tiempo puede indicar un tiempo T0 en el que se inició la asociación de la por lo menos una parte de la información identificable personalmente al dispositivo de cliente 1, tal como el identificador, y en el que el primer o primeros ordenadores tienen acceso a un reloj tal que el tiempo (T1 o T2) en el cual la asociación o el identificador expiró o se suprimió permanentemente en el dispositivo de cliente 1 se determina además sobre la base de una lectura del reloj.

Por ejemplo, los segundos datos se recuperan del módulo de almacenamiento de datos 102, o la tercera solicitud 130 es transmisible al servidor externo 1000, solamente en T1 o T2 cuando expiró o se suprimió permanentemente en el dispositivo de cliente 1 cualquier o un identificador, tal como una *Cookie*, asociado a por lo menos una parte de la información identificable personalmente.

Como se describe de forma adicional posteriormente en el contexto de la figura 8, el retardo de T2 con respecto a T1 puede venir dado por la latencia de componentes del sistema y/o representa una memoria intermedia temporal de la que se prefiere que garantice que el cliente haya limpiado el identificador (por ejemplo, el navegador del cliente ha limpiado la *cookie*: la *Cookie* expiró o se suprimió en el dispositivo de cliente 1) antes de que tenga lugar cualquier comunicación con el servidor externo 1000.

A todas las incidencias, que se atribuyen a una sesión específica, se les asigna un sello de tiempo de expiración de la sesión (momento en el que expiró o se suprimió el identificador, tal como una *cookie*). Esto se puede implementar de manera que se produzca en T2 o antes, es decir, un poco más tarde que T1, después de que se suprima o expire la *cookie*. En el momento de la supresión o expiración, al sistema se le notificará que la sesión ha terminado y a continuación el mismo tomará esta notificación y posibilitará la transmisión de las incidencias asociadas a la sesión hacia el servidor externo 1000.

Según aspectos preferidos de la presente invención, el método llevado a cabo por el sistema de la presente invención puede incluir, además, transmitir la tercera solicitud 130 generada al servidor externo 1000, preferentemente en donde la tercera solicitud 130 no incluye la información identificable personalmente. La tercera solicitud 130 puede transmitirse al servidor externo 1000 solo después de que haya expirado la cantidad de tiempo predeterminada T1 o T2. Por ejemplo, la tercera solicitud 130 puede transmitirse al servidor externo 1000 solamente en cuanto o solamente después de que expire o se suprima permanentemente en el dispositivo de cliente 1 la asociación de por lo menos una parte de la información identificable personalmente al dispositivo de cliente 1.

No obstante, la tercera solicitud 130 se transmite preferentemente al servidor externo 1000 antes de la expiración del límite de tiempo T3 para aceptar solicitudes según son suministradas por el *software* de analítica de terceros alojado en el servidor externo 1000.

Según aspectos preferidos de esta invención, la expiración de la sesión del lado del servidor en T2 es posterior a

la expiración de la sesión del lado del cliente en T1 para garantizar de manera fiable que el servidor externo 1000 no reciba ningún dato del que se haya realizado un seguimiento cuando la sesión en el cliente todavía está activa y para lograr así una anonimización total. Esta restricción garantiza que no se produzca, en ningún instante de tiempo, un solapamiento del cliente y la herramienta analítica en el conocimiento de los datos (por ejemplo, un ID de sesión almacenado en el cliente así como en la base de datos de analítica) que potencialmente podría posibilitar una identificación en tiempo real.

La figura 4 ilustra una vista más detallada del sistema según la primera forma de realización de la presente invención de la figura 2. Según la figura 4, los datos recuperados que se basan en los segundos datos almacenados se recuperan del módulo de almacenamiento de datos 102 una vez que se ha producido o después del vencimiento o expiración del tiempo T1 o T2 de la figura 3 (y preferentemente se transmiten con la tercera solicitud 130 a la analítica de terceros 1000 antes del T3 de la figura 3).

Como ilustra la figura 4, la síntesis de datos sintéticos la lleva a cabo el sintetizador 10 antes de que los datos se almacenen en el módulo de almacenamiento de datos 102, como también muestra la figura 2. En particular, se recibe una segunda solicitud 20, 120 (solicitud de seguimiento 20, 120) desde el dispositivo de cliente 1, en donde la segunda solicitud 20, 120 incluye los primeros datos que indican acciones que ocurrieron en un dispositivo de cliente 1. En el ejemplo meramente ilustrativo mostrado en la figura 4, los primeros datos indican que se han visto los vídeos AB, CD y EF. La segunda solicitud 20, 120 incluye, además, información identificable personalmente. En el ejemplo ilustrativo mostrado en la figura 4, la PII incluye la información de que estos vídeos se han visto durante la sesión 1234 en la dirección de IP 112.12.2.2. Cada una de estas visualizaciones de vídeo es una “incidencia” para la misma sesión.

En el sintetizador 10 ubicado aguas arriba del módulo de almacenamiento de datos 102, la PII se convierte en datos sintéticos. En otras palabras, el sintetizador 10 sintetiza datos sintéticos para eliminar la PII. En el ejemplo ilustrativo mostrado en la figura 4, la PII “sesión 1234” se sustituye por “sesión abcd”, donde el valor “1234” se sustituye por “abcd”.

De este modo, la conversión en los datos sintéticos puede incluir: sustituir un primer valor de un primer elemento de la información identificable personalmente por un valor sintetizado que es diferente del primer valor; e incluir los datos sintéticos con el valor sintetizado en la primera solicitud 110.

La dirección de IP, que es también PII, se puede omitir o eliminar, sustituir parcialmente por valores nuevos o convertir en una región geográfica, tal como un país. Por ejemplo, uno o más valores de un subconjunto de la dirección de IP, tal como la mitad de la dirección de IP, se pueden sustituir por uno o más valores predeterminados (por ejemplo, cero), de manera que, a partir de los datos sintéticos sintetizados para la dirección de IP, se pueda deducir solamente una región geográfica, tal como un país, y no una ubicación más detallada del cliente, tal como una ciudad. Según otro ejemplo, el sistema según aspectos de la invención puede almacenar asociaciones entre intervalos de direcciones de IP y regiones geográficas correspondientes, tales como países, y una dirección de IP particular se puede convertir en una región geográfica correspondiente sobre la base de las asociaciones almacenadas. Por ejemplo, el primer intervalo correspondiente a la dirección de IP puede estar asociado a una primera región geográfica y un segundo intervalo correspondiente a la dirección de IP, que es diferente del primer intervalo, puede estar asociado a una segunda región geográfica que es diferente de la primera región geográfica.

De este modo, la sustitución del primer valor por el valor sintetizado puede incluir: eliminar el primer valor, de tal manera que la primera solicitud 110 no incluya y no esté asociada con el primer valor eliminado para el primer elemento de información identificable personalmente.

En el ejemplo ilustrativo de la figura 4, la dirección de IP se convierte en el país “Austria” en el que se originó la segunda solicitud 20, 120 o en el que está ubicado el dispositivo de cliente 1. La dirección de IP puede estar contenida en un encabezamiento asociado a la segunda solicitud 20, 120 recibida del cliente 1. El valor sintetizado “Austria” puede ser un atributo de la primera solicitud 110 o puede estar contenido en un encabezamiento asociado a la primera solicitud 110. El valor sintetizado, en este caso “Austria”, indicativo de la región geográfica asociada con el dispositivo de cliente 1 es generado por dicho uno o más ordenadores que alojan el módulo de almacenamiento de datos 102 o por otro u otros ordenadores que están en comunicación con el módulo de almacenamiento de datos 102. Por ejemplo, los ordenadores que alojan el módulo de almacenamiento de datos 102 pueden recibir la dirección de IP y pueden convertir la dirección de IP en una región geográfica, tal como el país (“Austria”).

La segunda solicitud 20, 120 también puede comprender otra información, tal como una versión del navegador utilizada por el usuario cuando se llevan a cabo las acciones, y esta información de un navegador también se puede ofuscar o modificar en los datos sintéticos, como se explica adicionalmente más adelante en el contexto de la figura 9.

A continuación, el sintetizador 10 incluye los datos sintéticos en calidad de segundos datos en la primera solicitud 110 y transmite la primera solicitud 110 al módulo de almacenamiento de datos 102. El sintetizador 10 está

conectado comunicativamente con el dispositivo de cliente 1 y con el módulo de almacenamiento de datos 102. El sintetizador 10 puede estar alojado en el mismo ordenador o servidor o en uno diferente en comparación con el módulo de almacenamiento de datos 102. En general, la síntesis de datos llevada a cabo por el sintetizador 10 es opcional para esta invención y también puede ser proporcionada por un servicio de terceros alojado en un servidor externo al servidor informático que aloja el módulo de almacenamiento de datos 102. Alternativamente, la síntesis de datos puede formar parte de la invención, y la presente invención especifica maneras particularmente ventajosas que contribuyen a potenciar adicionalmente la seguridad de los datos de la PII a través de mejoras de la anonimización de la PII.

Al producirse la recepción u obtención de la primera solicitud 110, el módulo de almacenamiento de datos 102 almacena por lo menos los segundos datos, que en este caso son los datos sintéticos que fueron sintetizados por el sintetizador 10, en correspondencia con una cantidad de tiempo predeterminada. Durante esta cantidad de tiempo predeterminada, los datos almacenados están en cuarentena.

En cuanto o después de que T1 o T2 venza o expire, la cuarentena termina y los datos almacenados se pueden recuperar del módulo de almacenamiento de datos 102. En T1 o T2, puede considerarse que la sesión ha llegado a su fin, ya que ha expirado una asociación entre la PII y el usuario o el dispositivo de cliente 1, tal como la expiración o supresión permanente de la *cookie* “1234” para la sesión 1234 en el dispositivo de cliente 1.

Como ilustra el ejemplo de la figura 4, en o después del vencimiento o expiración de T1 o T2, es decir, en o después del final de la cuarentena, el módulo de almacenamiento de datos 102 transmite los datos almacenados y recuperados a un expedidor 107 que está configurado para generar la tercera solicitud 130 para su transmisión al servidor externo 1000 que aloja la analítica de terceros. En esta tercera solicitud 130, no hay presencia de ninguna PII, sino que, en su lugar, los datos sintéticos se incluyen en la tercera solicitud 130, que se recuperó del módulo de almacenamiento de datos 102 y que fue sintetizada por el sintetizador 10 sobre la base de la PII de la segunda solicitud 20, 120 antes de su inclusión en la primera solicitud 110. Además, la tercera solicitud 130 incluye los primeros datos (por ejemplo, vídeo AB visto) que indican las acciones que ocurrieron en el dispositivo de cliente 1.

Es decir, cada una de la primera solicitud 110 y la tercera solicitud 130 es una solicitud de seguimiento adicional generada con versiones ofuscadas o anonimizadas de la información identificable personalmente de la segunda solicitud 20, 120 para evaluar las acciones del usuario en el dispositivo de cliente 1. En la medida en la que el servidor externo 1000 puede recibir la tercera solicitud 130 solo después del vencimiento o expiración de T1 o T2, no es posible un acceso simultáneo a la tercera solicitud 130 y a la sesión 1234 en el dispositivo de cliente 1 por parte de la analítica de terceros. Cuando la tercera solicitud 130 llega al servidor externo 1000 que aloja la analítica de terceros, la sesión 1234 ya ha expirado, por ejemplo debido a que la *cookie* 1234 ya haya expirado o ya se haya suprimido permanentemente en el dispositivo de cliente 1. Es decir, la cuarentena reduce la probabilidad o incluso impide por completo un acceso simultáneo a la tercera solicitud 130 y la sesión en el dispositivo de cliente 1, lo cual potencia los datos o incluso garantiza la seguridad de los datos para la PII y también conduce al cumplimiento del GDPR. Los medios preferidos adicionales para proporcionar una síntesis de datos mejorada en el sintetizador 10 potencian adicionalmente la seguridad de los datos de la PII al contribuir a una anonimización total.

La figura 5 ilustra un sistema de acuerdo con una segunda forma de realización de la presente invención en el que la síntesis de datos tiene lugar en el sintetizador 10 después de que los segundos datos se almacenen en el módulo de almacenamiento de datos 102.

También en esta forma de realización, un método llevado a cabo por el sistema se refiere a la provisión de información identificable personalmente anonimizada, en donde el método se implementa mediante uno o más primeros ordenadores del sistema (por ejemplo, el servidor 100 de la figura 8), comprendiendo el método: obtener una primera solicitud 110, incluyendo la primera solicitud 110 primeros datos que indican acciones que ocurrieron en un dispositivo de cliente 1 y segundos datos que están asociados a los primeros datos y que se basan en información identificable personalmente asociada con el dispositivo de cliente 1; poner en cuarentena por lo menos los segundos datos, en donde la puesta en cuarentena por lo menos incluye almacenar por lo menos los segundos datos en un módulo de almacenamiento de datos 102; recuperar datos del módulo de almacenamiento de datos 102, en donde los datos recuperados se basan en los segundos datos almacenados; y generar una tercera solicitud 130 para su transmisión a un servidor 1000 externo a dicho uno o más primeros ordenadores, en donde la tercera solicitud 130 incluye los primeros datos y datos sintéticos asociados a los primeros datos, en donde los datos sintéticos se basan en los datos recuperados y se sintetizaron sobre la base de la información identificable personalmente. Los segundos datos pueden ser o incluir la información identificable personalmente o pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente. Los datos recuperados del módulo de almacenamiento de datos pueden ser los segundos datos (almacenados) (tales como la PII o los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente antes de su almacenamiento en calidad de segundos datos en el módulo de almacenamiento de datos), o pueden ser los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente de los segundos datos almacenados. Los datos sintéticos incluidos en la tercera solicitud pueden ser los datos recuperados del módulo de almacenamiento de datos o pueden ser datos sintéticos que se sintetizaron sobre la base de información

identificable personalmente de los datos recuperados. Los datos sintéticos pueden ser una versión anonimizada, reducida y/u ofuscada de la información identificable personalmente.

El servidor externo 1000 se puede configurar para alojar el *software* de herramienta de analítica de terceros con el fin de analizar acciones que ocurrieron en el dispositivo de cliente 1.

De acuerdo con la segunda forma de realización, los segundos datos incluidos en la primera solicitud 110 obtenida y/o los datos recuperados del módulo de almacenamiento de datos 102 son la información identificable personalmente asociada con el dispositivo de cliente 1.

El método llevado a cabo por el sistema puede incluir opcionalmente, además, las operaciones asociadas a la síntesis de datos según es llevada a cabo por el sintetizador 10. En esta forma de realización, el sintetizador 10 está ubicado aguas abajo del módulo de almacenamiento de datos 102, pero antes de que los datos se transmitan al servidor externo 1000 que aloja la analítica de terceros. Los datos sintéticos sintetizados por el sintetizador 10 se pueden almacenar en el módulo de almacenamiento de datos 102 antes de que los datos sintéticos se recuperen del módulo de almacenamiento de datos (en calidad de datos recuperados que se basan en los segundos datos) y se reenvíen con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000, o los datos sintéticos pueden reenviarse directamente con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000, como se muestra, por ejemplo, en las figuras 6 y 7. El sintetizador 10 puede estar alojado en uno o más ordenadores del sistema según la invención.

La generación de la tercera solicitud 130 como operación del método llevado a cabo por el sistema de acuerdo con la segunda forma de realización puede comprender: sintetizar los datos sintéticos incluyendo convertir en los datos sintéticos por lo menos parte de los segundos datos o datos recuperados e incluir en la tercera solicitud 130 los datos sintéticos en lugar de los segundos datos o datos recuperados.

La figura 6 ilustra una vista más detallada del sistema según una versión de la segunda forma de realización de la presente invención ilustrada por la figura 5. Según la versión de la figura 6, los datos recuperados que se basan en los segundos datos almacenados se recuperan del módulo de almacenamiento de datos 102 una vez que se ha producido o después del vencimiento o expiración del tiempo T1 o T2 de la figura 3 (y la tercera solicitud se transmite preferentemente a la analítica de terceros 1000 antes que el T3 de la figura 3). Es decir, la cuarentena termina en T1 o T2 después de lo cual los segundos datos se recuperan del módulo de almacenamiento de datos 102 en calidad de datos recuperados que se basan en los segundos datos y los segundos datos recuperados se reenvían en el sistema hacia el sintetizador 10.

Como ilustra la figura 6, la síntesis de datos sintéticos la lleva a cabo el sintetizador 10 después de que los segundos datos se almacenen en el módulo de almacenamiento de datos 102 y después de que los segundos datos se recuperen del módulo de almacenamiento de datos 102 (en calidad de datos recuperados que se basan en los segundos datos), como muestra la figura 5. La ubicación exacta del sintetizador 10 después del módulo de almacenamiento de datos 102 y antes de la transmisión al servidor de analítica de terceros 1000 no es esencial en varias ubicaciones posibles como se describe adicionalmente en el contexto de la figura 8. En particular, el sintetizador 10 puede estar ubicados aguas arriba con respecto al expedidor 107 o puede estar integrado en el expedidor 107.

En particular, se recibe una primera solicitud 110 (solicitud de seguimiento) desde el dispositivo de cliente 1, en donde la primera solicitud 110 incluye los primeros datos que indican acciones que ocurrieron en un dispositivo de cliente 1. En el ejemplo meramente ilustrativo mostrado en la figura 6, los primeros datos indican que se han visto los vídeos AB, CD y EF. La primera solicitud 110 incluye, además, información identificable personalmente. En el ejemplo ilustrativo mostrado en la figura 6, la PII incluye la información de que estos vídeos se han visto durante la sesión 1234 en la dirección de IP 112.12.2.2. Cada una de estas visualizaciones de vídeo es una “incidencia” para la misma sesión.

En el sintetizador 10 ubicado aguas abajo del módulo de almacenamiento de datos 102, la PII se convierte en datos sintéticos. En otras palabras, el sintetizador 10 sintetiza datos sintéticos para eliminar la PII. En el ejemplo ilustrativo mostrado en la figura 6, la PII “sesión 1234” se sustituye por “sesión abcd”, donde el valor “1234” se sustituye por “abcd”.

De este modo, la conversión en datos sintéticos puede incluir: sustituir un primer valor de un primer elemento de la información identificable personalmente por un valor sintetizado que es diferente del primer valor; e incluir en la tercera solicitud 130 los datos sintéticos con el valor sintetizado.

La dirección de IP, que también es PII, se puede omitir o eliminar, sustituir parcialmente por valores nuevos o se puede convertir en una región geográfica, tal como un país. Por ejemplo, uno o más valores de un subconjunto de la dirección de IP, tal como la mitad de la dirección de IP, pueden sustituirse por uno o más valores predeterminados (por ejemplo, cero), de manera que, a partir de los datos sintéticos sintetizados para la dirección de IP, se puede deducir solamente una región geográfica, tal como un país, y no una ubicación más detallada del cliente, tal como

una ciudad. Según otro ejemplo, el sistema según aspectos de la invención puede almacenar asociaciones entre intervalos de direcciones de IP y regiones geográficas correspondientes, tales como países, y una dirección de IP particular se puede convertir en una región geográfica correspondiente sobre la base de las asociaciones almacenadas. Por ejemplo, el primer intervalo correspondiente a la dirección de IP puede estar asociado a una primera región geográfica y un segundo intervalo correspondiente a la dirección de IP, que es diferente del primer intervalo, puede estar asociado a una segunda región geográfica que es diferente de la primera región geográfica.

De este modo, la sustitución del primer valor por el valor sintetizado puede incluir: eliminar el primer valor, de tal manera que la tercera solicitud 130 no incluya y no esté asociada con el primer valor eliminado correspondiente al primer elemento de información identificable personalmente.

En el ejemplo ilustrativo de la figura 6, la dirección de IP se convierte en el país “Austria” en el cual se originó la primera solicitud 110 o en el cual está ubicado el dispositivo de cliente 1. La dirección de IP puede estar contenida en un encabezamiento asociado a la primera solicitud 110 recibida del cliente 1. El valor sintetizado “Austria” puede ser un atributo de la tercera solicitud 130 o puede estar contenido en un encabezamiento asociado a la tercera solicitud 130. El valor sintetizado, en este caso “Austria”, indicativo de la región geográfica asociada con el dispositivo de cliente 1 también puede ser generado por dicho uno o más ordenadores que alojan el módulo de almacenamiento de datos 102 o por otro u otros ordenadores que estén en comunicación con el módulo de almacenamiento de datos 102. Por ejemplo, los ordenadores que alojan el módulo de almacenamiento de datos 102 pueden recibir la dirección de IP y pueden convertir la dirección de IP en una región geográfica, tal como el país (“Austria”).

La primera solicitud 110 también puede comprender otra información, tal como una versión del navegador utilizada por el usuario cuando se llevan a cabo las acciones, y esta información de un navegador también se puede ofuscar o modificar en los datos sintéticos, como se explica adicionalmente más adelante en el contexto de la figura 9.

Al producirse la recepción u obtención de la primera solicitud 110, el módulo de almacenamiento de datos 102 almacena por lo menos los segundos datos, que en este caso son la PII, en correspondencia con una cantidad de tiempo predeterminada. Durante esta cantidad de tiempo predeterminada, los datos almacenados están en cuarentena.

El sintetizador 10 está conectado comunicativamente con el módulo de almacenamiento de datos 102. El sintetizador 10 puede estar alojado en el mismo ordenador o servidor o en uno diferente en comparación con el módulo de almacenamiento de datos 102. En general, la síntesis de datos especifica formas particularmente ventajosas que contribuyen a potenciar adicionalmente la seguridad de los datos de la PII a través de mejoras de la anonimización de la PII.

En cuanto o después de que T1 o T2 venza o expire, la cuarentena termina y los datos almacenados pueden recuperarse del módulo de almacenamiento de datos 102 y transmitirse al sintetizador 10. En T1 o T2, puede considerarse que la sesión ha llegado a su fin, debido a que ha expirado una asociación entre la PII y el usuario o el dispositivo de cliente 1, tal como la expiración o supresión permanente de la *cookie* “1234” para la sesión 1234 en el dispositivo de cliente 1.

Como ilustra el ejemplo de la figura 6, en o después del vencimiento o expiración de T1 o T2, es decir, en o después del final de la cuarentena, el sintetizador 10 recibe los segundos datos recuperados del módulo de almacenamiento de datos 102 y sintetiza los datos sintéticos sobre la base de los segundos datos (PII) recuperados del módulo de almacenamiento de datos 102. El sintetizador 10 o el expedidor 107 incluye los datos sintéticos en la tercera solicitud 130. A continuación, el expedidor 107 transmite la tercera solicitud 130 al servidor externo 1000, preferentemente antes del vencimiento de T3. T3 puede ser un retardo aceptado máximo de datos dentro del sistema de analítica antes de que se pierdan los datos.

En esta tercera solicitud 130, no hay presencia de ninguna PII, sino que, en su lugar, los datos sintéticos se incluyen en la tercera solicitud 130, que fue sintetizada por el sintetizador 10 sobre la base de la PII de la primera solicitud 110 antes de su inclusión en la tercera solicitud 130. Además, la tercera solicitud 130 incluye los primeros datos (por ejemplo, vídeo visto AB) que indican las acciones que ocurrieron en el dispositivo de cliente 1.

Es decir, la tercera solicitud 130 es una solicitud de seguimiento adicional generada con versiones ofuscadas o anonimizadas de la información identificable personalmente de la primera solicitud 110 para evaluar las acciones del usuario en el dispositivo de cliente 1. En la medida en la que el servidor externo 1000 puede recibir la tercera solicitud 130 solo después del vencimiento o expiración de T1 o T2, la no es posible un acceso simultáneo a la tercera solicitud y a la sesión 1234 en el dispositivo de cliente 1 por parte de analítica de terceros. Cuando la tercera solicitud 130 llega al servidor externo 1000 que aloja la analítica de terceros, la sesión 1234 ya ha expirado, por ejemplo debido a que la *cookie* 1234 ya ha expirado o ya se ha suprimido permanentemente en el dispositivo de cliente 1. Es decir, la cuarentena reduce la probabilidad o incluso impide por completo un acceso simultáneo a la tercera solicitud y a la sesión en el dispositivo de cliente, lo cual potencia los datos o incluso garantiza la seguridad de los datos para la PII y también conduce al cumplimiento del GDPR. Los medios preferidos adicionales

para proporcionar una síntesis de datos mejorada en el sintetizador 10 potencian adicionalmente la seguridad de los datos de la PII al contribuir a una anonimización total.

La figura 7 ilustra una vista detallada del sistema según otra versión de la segunda forma de realización de la presente invención que se ilustra en la figura 5. En la figura 7, los segundos datos se recuperan del módulo de almacenamiento de datos antes del final de la cuarentena, es decir, antes del vencimiento o expiración del tiempo T1 o T2 de la figura 3. Durante la cuarentena, el sintetizador 10 sintetiza los datos sintéticos sobre la base de los segundos datos almacenados y a continuación los mismos bien se pueden almacenar en el módulo de almacenamiento de datos 102 antes de que los datos sintéticos se recuperen del módulo de almacenamiento de datos y se reenvían con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000, o bien se pueden reenviar directamente (es decir, sin almacenamiento en el módulo de almacenamiento de datos) con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000. En cualquier caso, los datos sintéticos del sintetizador 10 pueden reenviarse hacia el expedidor 107 durante la cuarentena, pero se transmiten con la tercera solicitud 130 al servidor de analítica de terceros 1000 solamente después del final de la cuarentena como se muestra en la figura 7, es decir, en o después del vencimiento o expiración del tiempo T1 o T2 de la figura 3 (y preferentemente antes que el T3 de la figura 3). Alternativamente, según una variante no mostrada en la figura 7, los datos sintéticos pueden ser sintetizados por el sintetizador 10 sobre la base de los segundos datos almacenados y a continuación se almacenan en el módulo de almacenamiento de datos 102 durante la cuarentena, pero se pueden recuperar del módulo de almacenamiento de datos 102, en calidad de datos recuperados que se basan en los segundos datos, y reenviar en el sistema hacia el expedidor 107 con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000 solamente después del final de la cuarentena, es decir, en cuanto se haya producido o después del vencimiento o expiración de T1 o T2.

Como ilustra la figura 7, la síntesis de datos sintéticos la lleva a cabo el sintetizador 10 después de que los segundos datos se almacenen en el módulo de almacenamiento de datos 102 pero antes del final de la cuarentena en T1 o T2. Es decir, la puesta en cuarentena incluye la recuperación de los segundos datos almacenados a partir del módulo de almacenamiento de datos 102 y la síntesis de los datos sintéticos sobre la base de los segundos datos recuperados. La ubicación exacta del sintetizador 10 aguas abajo del módulo de almacenamiento de datos 102 y aguas arriba del expedidor no es esencial y son posibles varias ubicaciones según se ha mencionado anteriormente y según se describe de manera adicional en el contexto de la figura 8. En particular, el sintetizador 10 puede estar ubicado aguas arriba con respecto al expedidor 107 o puede estar integrado en el expedidor 107.

En particular, se recibe u obtiene una primera solicitud 110 (solicitud de seguimiento) del dispositivo de cliente 1, en donde la primera solicitud 110 incluye los primeros datos que indican acciones que ocurrieron en un dispositivo de cliente 1. En el ejemplo meramente ilustrativo mostrado en la figura 7, los primeros datos indican que se han visto los vídeos AB, CD y EF. La primera solicitud 110 incluye, además, información identificable personalmente. En el ejemplo ilustrativo mostrado en la figura 7, la PII incluye la información de que estos vídeos se han visto durante la sesión 1234 en la dirección de IP 112.12.2.2. Cada una de estas visualizaciones de vídeo es una “incidencia” para la misma sesión.

Al producirse la recepción u obtención de la primera solicitud 110, el módulo de almacenamiento de datos 102 almacena por lo menos los segundos datos, que en este caso son la PII. Durante el almacenamiento y durante la cuarentena, en este caso los datos almacenados pueden recuperarse, tal como mediante el sintetizador 10 para la síntesis de datos.

En el sintetizador 10 ubicado aguas abajo del módulo de almacenamiento de datos 102, la PII se convierte en datos sintéticos. En otras palabras, el sintetizador 10 sintetiza datos sintéticos para eliminar la PII. En el ejemplo ilustrativo mostrado en la figura 7, la PII “sesión 1234” se sustituye por “sesión abcd”, donde el valor “1234” se sustituye por “abcd”.

De este modo, la conversión en los datos sintéticos puede incluir: sustituir un primer valor de un primer elemento de la información identificable personalmente por un valor sintetizado que es diferente del primer valor; e incluir en la tercera solicitud 130 los datos sintéticos con el valor sintetizado.

La dirección de IP, que también es PII, se puede omitir o eliminar, sustituir parcialmente por valores nuevos o se puede convertir en una región geográfica, tal como un país. Por ejemplo, uno o más valores de un subconjunto de la dirección de IP, tal como la mitad de la dirección de IP, pueden sustituirse por uno o más valores predeterminados (por ejemplo, cero), de manera que, a partir de los datos sintéticos sintetizados para la dirección de IP, se pueda deducir solo una región geográfica, tal como un país, y no una ubicación más detallada del cliente, tal como una ciudad. Según otro ejemplo, el sistema según aspectos de la invención puede almacenar asociaciones entre intervalos de direcciones de IP y regiones geográficas correspondientes, tales como países, y una dirección de IP particular se puede convertir en una región geográfica correspondiente sobre la base de las asociaciones almacenadas. Por ejemplo, el primer intervalo correspondiente a la dirección de IP puede estar asociado a una primera región geográfica y un segundo intervalo correspondiente a la dirección de IP, que es diferente del primer intervalo, puede estar asociado a una segunda región geográfica que es diferente de la primera región geográfica.

De este modo, la sustitución del primer valor por el valor sintetizado puede incluir: eliminar el primer valor, de tal manera que la tercera solicitud 130 no incluya y no esté asociada con el primer valor eliminado en correspondencia con el primer elemento de información identificable personalmente.

En el ejemplo ilustrativo de la figura 7, la dirección de IP se convierte en el país "Austria" en el cual se originó la primera solicitud 110 o en el cual está ubicado el dispositivo de cliente 1. La dirección de IP puede estar contenida en un encabezamiento asociado a la primera solicitud 110 recibida del cliente 1. El valor sintetizado "Austria" puede ser un atributo de la tercera solicitud 130 o puede estar contenido en un encabezamiento asociado a la tercera solicitud 130. El valor sintetizado, en este caso "Austria", indicativo de la región geográfica asociada con el dispositivo de cliente 1 también puede ser generado por dicho otro uno o más ordenadores que alojan el módulo de almacenamiento de datos 102, el sintetizador 10 y/o el expedidor 107, o por otro u otros ordenadores que estén en comunicación con estos componentes del sistema. Por ejemplo, los ordenadores que alojan el sistema pueden recibir la dirección de IP y pueden convertir la dirección de IP en una región geográfica, tal como el país ("Austria").

La primera solicitud 110 también puede comprender otra información, tal como una versión del navegador utilizada por el usuario cuando se llevan a cabo las acciones, y esta información de un navegador también se puede ofuscar o modificar en los datos sintéticos, como se explica adicionalmente más adelante en el contexto de la figura 9.

El sintetizador 10 está conectado comunicativamente con el módulo de almacenamiento de datos 102. El sintetizador 10 puede estar alojado en el mismo ordenador o servidor o en uno diferente en comparación con el módulo de almacenamiento de datos 102. En general, la síntesis de datos especifica formas particularmente ventajosas que contribuyen a potenciar adicionalmente la seguridad de los datos de la PII a través de mejoras de la anonimización de la PII.

Antes de que T1 o T2 venza o expire, el sintetizador 10 recupera los segundos datos almacenados en el módulo de almacenamiento de datos 102 y lleva a cabo la ofuscación o anonimización de la PII generando los datos sintéticos sobre la base de la PII.

El sintetizador 10 o el expedidor 107 incluye los datos sintéticos en la tercera solicitud 130. En cuanto o después de que T1 o T2 venza o expire, la cuarentena termina y los datos sintéticos pueden transmitirse con la tercera solicitud 130 al servidor externo 1000 por parte del expedidor 107, preferentemente antes de que T3 venza. Por ejemplo, el sintetizador 10 puede transmitir los datos sintéticos al expedidor 107 en cuanto o después de que T1 o T2 venza o expire. Alternativamente, el sintetizador 10 está integrado en el expedidor 107, iniciando dicho expedidor 107 la transmisión de la tercera solicitud 130 al servidor de analítica de terceros 1000 únicamente en cuanto o después de que venza o expire T1 o T2.

Al vencer o expirar T1 o T2, puede considerarse que la sesión ha llegado a su fin, ya que ha expirado una asociación entre la PII y el usuario o el dispositivo de cliente 1, tal como la expiración o supresión permanente de la *cookie* "1234" para la sesión 1234 en el dispositivo de cliente 1.

En la tercera solicitud 130 transmitida al servidor externo 1000, no hay presencia de ninguna PII, sino que, en su lugar, los datos sintéticos se incluyen en la tercera solicitud 130, que fue sintetizada por el sintetizador 10 sobre la base de la PII recuperada del módulo de almacenamiento de datos 102 e incluida en la primera solicitud 110. Además, la tercera solicitud 130 incluye los primeros datos (por ejemplo, vídeo visto AB) que indican las acciones que ocurrieron en el dispositivo de cliente 1.

Es decir, la tercera solicitud 130 es una solicitud de seguimiento adicional generada con versiones ofuscadas o anonimizadas de la información identificable personalmente de la primera solicitud 110 para evaluar las acciones del usuario en el dispositivo de cliente 1. En la medida en la que el servidor externo 1000 también en este caso de la figura 7 puede recibir la tercera solicitud 130 únicamente después del vencimiento o expiración de T1 o T2, no es posible un acceso simultáneo a la tercera solicitud y a la sesión 1234 en el dispositivo de cliente 1 por parte de la analítica de terceros. Cuando la tercera solicitud 130 llega al servidor externo 1000 que aloja la analítica de terceros, la sesión 1234 ya ha expirado, por ejemplo debido a que la *cookie* 1234 ya ha expirado o ya se ha suprimido permanentemente en el dispositivo de cliente 1. Es decir, la cuarentena reduce la probabilidad o incluso impide por completo un acceso simultáneo a la tercera solicitud y a la sesión en el dispositivo de cliente, lo cual potencia los datos o incluso garantiza la seguridad de los datos para la PII y también conduce al cumplimiento del GDPR. Los medios preferidos adicionales para proporcionar una síntesis de datos mejorada en el sintetizador 10 potencian adicionalmente la seguridad de los datos de la PII al contribuir a una anonimización total.

La figura 8 ilustra más detalles sobre el sistema según la primera y segunda formas de realización de la presente invención que ilustran las figuras 2 a 7. El sistema de la invención incluye el servidor de cuarentena 100 como ejemplo de dicho uno o más primeros ordenadores. El servidor de cuarentena 100 puede alojar el módulo de almacenamiento de datos 102 y el expedidor 107 a los que se ha hecho referencia por medio de uno cualquiera de los aspectos y formas de realización descritos anteriormente. El expedidor 107 está configurado para transmitir datos al servidor externo 1000 que aloja el *software* de analítica de terceros. El servidor 1000 no es un

subcomponente del sistema según la invención. El sintetizador 10 puede alojarse en un servidor operado por un servicio de terceros externo al sistema de esta invención. Alternativamente, el servidor de cuarentena 100 puede alojar además como componente opcional el sintetizador 10. En cualquier caso, el servidor de cuarentena 100 puede alojar además, como componentes opcionales, un receptor 101, un abonado 105 y una cola de espera de eventos 106.

El sistema puede incluir el sintetizador 10 opcional al que se ha hecho referencia en uno cualquiera de los aspectos y formas de realización descritos anteriormente. En la figura 8, la síntesis de datos se ilustra como opcional para la invención y puede tener lugar en el sintetizador 10 que puede ubicarse en uno o más cualesquiera de los diversos componentes del sistema mostrado en la figura 8 antes o después de que se almacenen los segundos datos en el módulo de almacenamiento de datos 102. En particular, el sintetizador 10 puede alojarse en un servidor sintetizador 10 que forma parte del sistema según aspectos de esta invención pero que es independiente del servidor de cuarentena 100 que está en comunicación con el servidor sintetizador 10.

Alternativamente, el sintetizador 10 puede alojarse en el servidor de cuarentena 100 según aspectos de esta invención, tal como en el receptor 100 aguas arriba del módulo de almacenamiento de datos 102 o en el abonado 105 o el expedidor 107 aguas abajo del módulo de almacenamiento de datos 102 o en cualquier otro componente alojado en el servidor de cuarentena 100. Tener el sintetizador en el abonado o expedidor permite liberar de carga componentes de procesado de datos (por ejemplo, un receptor) aguas arriba del módulo de almacenamiento de datos, que tendrán entonces tiempos de reacción mejorados y proporcionarán una mejor experiencia de usuario en la parte de presentación del sistema. Además, normalmente el tamaño de los datos sintéticos es mayor que el de los datos originales, de manera que se puede reducir la cantidad de memoria requerida en el módulo de almacenamiento de datos. Tener el sintetizador en el expedidor en lugar del abonado permite mantener una mayor eficiencia para el abonado de manera que este último reciba de manera eficiente los datos asociados a las sesiones que han expirado. Al tener el sintetizador en el receptor o más aguas arriba del receptor, la anonimización de la PII se lleva a cabo lo antes posible y se reduce el riesgo de fugas de la PII, tal como a través de un ataque al módulo de almacenamiento de datos. Asimismo, se puede obtener una vista previa de los datos sintéticos antes de almacenar los datos en el módulo de almacenamiento de datos, lo cual permite reducir errores cuando se somete a pruebas el sistema o cuando se procesan los datos antes de la transmisión al servidor externo.

Un usuario interactúa con un sitio web 2 (por ejemplo, en un navegador o aplicación) en el dispositivo de cliente 1 y se realiza un seguimiento de las acciones del usuario, tales como todas las acciones pertenecientes a una sesión común del navegador, aplicación o sitio web, para su evaluación mediante un *software* de analítica alojado en un servidor externo 1000. Desde el dispositivo de cliente 1 se puede transmitir una segunda solicitud 20 o 120 al servidor de cuarentena 100. Esta segunda solicitud 20, 120 puede incluir primeros datos que identifican las acciones y puede incluir, además, la información identificable personalmente, PII, que puede identificar de manera directa o indirecta al usuario o al dispositivo de cliente 1. La PII puede incluir, por ejemplo, un identificador de sesión asociado con el usuario del dispositivo de cliente 1, una parte de una dirección de IP del dispositivo de cliente 1, un agente de usuario de un navegador utilizado por el usuario en el dispositivo de cliente 1, una dirección de contacto asociada con el usuario del dispositivo de cliente 1, y/o un nombre asociado con el usuario del dispositivo de cliente 1.

La segunda solicitud 20 puede pasar por uno o más módulos de transformación 3, 4 opcionales (por ejemplo un equilibrador de carga de Akamai) que están configurados para transformar un elemento de información de la PII en una forma ofuscada o anonimizada de este elemento de información. Los módulos de transformación 3, 4 pueden estar alojados en el servidor de cuarentena 100 o pueden estar alojados en otro u otros ordenadores.

Por ejemplo, el módulo de transformación 3, 4 se puede configurar para convertir una dirección de IP asociada con el dispositivo de cliente 1 en una región geográfica, tal como un país, que está asociada con el dispositivo de cliente 1 al tiempo que eliminado u omitiendo de la solicitud la dirección de IP. Por ejemplo, la dirección de IP puede estar contenida en un encabezamiento asociado a la solicitud enviada por el cliente 1 al módulo de transformación 3, 4. El valor sintetizado que indica la región geográfica, según lo sintetiza el módulo de transformación 3, 4 a partir de la dirección de IP, puede ser un atributo de la segunda solicitud 20 o puede estar contenido en un encabezamiento asociado a la segunda solicitud 20.

Este encabezamiento puede indicar la región geográfica, tal como el país, en la cual se ha generado la solicitud del usuario o que se ha introducido en la red o sistema de la figura 8. En lugar de enviar la dirección de IP real del cliente 1 o usuario al *software* de analítica en el servidor 1000 y basarse en ella para transfigurar la dirección de IP en la ubicación del usuario, el sistema enviará explícitamente la región geográfica del usuario, tal como un país, al servidor 1000. La dirección de IP real del usuario se puede descartar.

Por ejemplo, la solicitud del usuario con la dirección de IP 188.105.236.52 entrará en el módulo de transformación (por ejemplo, red Akamai en Alemania). A continuación, el módulo de transformación 3, 4 convertirá la dirección de IP en la región geográfica correspondiente, tal como el país correspondiente. La información del país será reenviada por el módulo de transformación 3, 4 hacia el servidor de cuarentena 100 y finalmente al *software* de analítica alojado en el servidor 1000. Este cambio dentro de la cadena de suministro de datos hace que aumente

la seguridad por no transmitir ninguna parte de la dirección de IP al sistema de analítica (por ejemplo, Google Analytics), lo cual hace que resulte más duro para cualquier parte identificar a un usuario sobre la base de la dirección de IP. La solicitud de los usuarios se puede enviar al siguiente punto de entrada.

Como alternativa a los módulos de transformación 3, 4, el sintetizador 10 puede sintetizar datos sintéticos para la dirección de IP. Por ejemplo, la dirección de IP, que también es PII, se puede omitir o eliminar, sustituir parcialmente por valores nuevos, o se puede convertir en una región geográfica, tal como un país, por parte del sintetizador 10.

Por ejemplo, uno o más valores de un subconjunto de la dirección de IP, tal como la mitad de la dirección de IP, pueden sustituirse por uno o más valores predeterminados (por ejemplo, cero), de manera que, a partir de los datos sintéticos sintetizados para la dirección de IP, se puede deducir solo una región geográfica, tal como un país, y no una ubicación más detallada del cliente, tal como una ciudad. El subconjunto de la dirección de IP se puede sustituir por valores predeterminados y el subconjunto restante de la dirección de IP, cuyo valor o valores no se sustituyen por un valor sintetizado, es indicativo de una región geográfica, tal como un país, asociada a una ubicación del dispositivo de cliente 1. Por ejemplo, la dirección de IP 188.105.236.52 será sintetizada por el sintetizador 10 obteniendo 188.105.0.0 antes de su reenvío. Esto crea efectivamente grupos de IP que pueden incluir hasta 65025 dispositivos y, por lo tanto, no permiten ninguna forma de identificación personal. Esta anonimización, así como el esfuerzo de reducción, que se realiza mediante aspectos de la presente invención, ofrece anonimización al grado máximo, al tiempo que se sigue pudiendo informar a nivel de región (por ejemplo, el país).

Según otro ejemplo, el propio sintetizador 10 o el servidor de cuarentena 100 puede almacenar asociaciones entre intervalos de direcciones de IP y regiones geográficas correspondientes, tales como países, y el sintetizador 10 puede convertir una dirección de IP particular en una región geográfica correspondiente sobre la base de las asociaciones almacenadas. Por ejemplo, el primer intervalo correspondiente a la dirección de IP puede estar asociado a una primera región geográfica y un segundo intervalo correspondiente a la dirección de IP, que es diferente del primer intervalo, puede estar asociado a una segunda región geográfica que es diferente de la primera región geográfica.

En el servidor de cuarentena 100, se implementa un método para proporcionar información identificable personalmente anonimizada, comprendiendo el método: obtener la primera solicitud 110 en el receptor 101 o el módulo de almacenamiento de datos 102, incluyendo la primera solicitud 110 primeros datos que indican acciones que ocurrieron en el dispositivo de cliente 1 y segundos datos que están asociados a los primeros datos y que se basan en información identificable personalmente asociada con el dispositivo de cliente 1; poner en cuarentena por lo menos los segundos datos, en donde la puesta en cuarentena por lo menos incluye almacenar por lo menos los segundos datos 103 en el módulo de almacenamiento de datos 102; recuperar, tal como por parte del abonado 105 o el sintetizador 10, datos, tales como los segundos datos, a partir del módulo de almacenamiento de datos 102, en donde los datos recuperados se basan en los segundos datos almacenados; y generar, por parte del expedidor 107 o el sintetizador 10, una tercera solicitud 130 para su transmisión por el expedidor 107 al servidor 1000 externo al servidor de cuarentena 100, en donde la tercera solicitud 130 incluye los primeros datos y datos sintéticos asociados a los primeros datos, en donde los datos sintéticos se basan en los datos recuperados y fueron sintetizados por el sintetizador 10 sobre la base de la información identificable personalmente.

En caso de que el sintetizador 10 se sitúe aguas arriba del módulo de almacenamiento de datos 102 pero dentro del servidor de cuarentena 100, tal como en el receptor 101, 20 y 120 pueden hacer referencia a la misma segunda solicitud recibida del cliente 1. En caso de que el sintetizador 10 se sitúe aguas arriba del módulo de almacenamiento de datos 102 pero fuera del servidor de cuarentena 100, tal como en el servidor sintetizador 10 independiente, 120 y 110 pueden hacer referencia a la misma primera solicitud obtenida por el servidor de cuarentena 100.

Después de que el receptor 101 en el servidor 100 reciba la segunda solicitud 20, 120, un sintetizador 10 ubicado en el receptor 101 puede sintetizar los segundos datos sobre la base de la PII en caso de que la síntesis de datos no se haya producido ya en el servidor sintetizador 10 externo al servidor de cuarentena 100. No obstante, puede que no se produzca ninguna síntesis de datos sintéticos previa al almacenamiento de los segundos datos en el módulo de almacenamiento de datos 102, debido a que dicha síntesis puede tener lugar aguas abajo del módulo de almacenamiento de datos 102, tal como en el abonado 105 o el expedidor 107. En caso de que la síntesis de datos sintéticos tenga lugar en el sintetizador 10 aguas abajo del módulo de almacenamiento de datos 102, tal como en el abonado 105, los datos sintéticos sintetizados por el sintetizador 10 pueden no almacenarse en el módulo de almacenamiento de datos 102 sino que se prefiere que se puedan reenviar directamente (por ejemplo, mediante la cola de espera de eventos 106) al expedidor 107 con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000 por el expedidor 107. Alternativamente (no mostrado en la figura 8), los datos sintéticos sintetizados por el sintetizador 10, tal como por el abonado 105, se pueden almacenar en el módulo de almacenamiento de datos 102 antes de que los datos sintéticos se recuperen del módulo de almacenamiento de datos 102 (en calidad de datos recuperados que se basan en los segundos datos) y reenviar (por ejemplo, mediante la cola de espera de eventos 106) al expedidor 107 con vistas a su inclusión en la tercera solicitud 130 para ser transmitidos al servidor externo 1000 por el expedidor 107.

En cualquier caso, los segundos datos se ponen en cuarentena, incluyendo el almacenamiento de los segundos datos 103 en el módulo de almacenamiento de datos 102. Poner en cuarentena por lo menos los segundos datos puede incluir almacenar los segundos datos asociados a una clave, SN, en el módulo de almacenamiento de datos 102, preferentemente en donde la clave está asociada a la información identificable personalmente asociada a los segundos datos. La puesta en cuarentena de por lo menos los segundos datos puede incluir, además, poner en cuarentena los primeros datos (datos de acciones). Poner en cuarentena los primeros y segundos datos puede incluir almacenar los segundos datos y los primeros datos, ambos asociados a la clave, en el módulo de almacenamiento de datos 102.

Por ejemplo, los segundos datos 103 se pueden almacenar con datos almacenados 102 de tal manera que la clave, que puede estar asociada a un identificador de sesión que identifica la sesión en el dispositivo de cliente 1, vincula todos los elementos de datos de los primeros y/o segundos datos que se están almacenando en el módulo de almacenamiento de datos 102. Por ejemplo, todas las “incidencias” asociadas al mismo ID de sesión se pueden almacenar en asociación mutua y/o juntas, como se ilustra en la figura 8.

En la figura 8, S1 representa el ID de sesión respectivo. El valor de S1 puede ser el propio ID de sesión respectivo, o una clave asociada con el ID de sesión, dependiendo de si la síntesis de los datos sintéticos ya ha tenido lugar cuando se almacenan los segundos datos 103. En caso de síntesis de datos sintéticos por parte del sintetizador 10 antes del almacenamiento de los segundos datos (datos sintéticos) en el módulo de almacenamiento de datos 102, S1 será una Clave sintética S1 (por ejemplo, 04d0fde5cc3160ea220cf4535b3239a8e36d475213d2f77301553eca84203122). En caso de síntesis de datos sintéticos por el sintetizador 10 después del almacenamiento de los segundos datos (PII) en el módulo de almacenamiento de datos 102, S1 será el valor original (por ejemplo 1682390.1648740198429), tal como de la *Cookie* u otro identificador.

Las incidencias respectivas representan valores tales como H1=Visitas de una página en el sitio web, o plataforma digital, H2= Inicio de Vídeo de los Vídeos “abc”, y H3=fin de Vídeo “abc”. Junto con las incidencias se pueden determinar valores tales como el “ID de Sesión” o “Dirección de IP”, y los mismos se asociarán a la sesión correspondiente.

El receptor 101 puede coger todas las solicitudes de incidencia que se reciben y puede gestionar la autenticación y la comunicación con el módulo de almacenamiento de datos 102. El receptor 100 se puede configurar para: Crear objetos de sesión nuevos para incidencias, utilizando los IDs de sesión como clave; y agregar los datos de incidencias al objeto de sesión. El receptor 101 puede ser un servicio ligero que se puede replicar múltiples veces y cada receptor puede tener la capacidad de actuar independientemente de los otros receptores. Esto permitiría un escalado horizontal rápido con una carga mayor.

El módulo de almacenamiento de datos 102, que puede ser una base de datos, permite un almacenamiento temporal eficiente. El módulo de almacenamiento de datos o la base de datos puede proporcionar almacenamiento de datos en memoria.

El módulo de almacenamiento de datos 102 se puede configurar para admitir persistencia de datos. No obstante, la persistencia de datos requiere la escritura de datos (por lo menos periódicamente) en disco, pero solo se requiere en caso de fallo de la base de datos. El módulo de almacenamiento de datos 102 puede ofrecer la posibilidad de ejecución en un escenario de alta disponibilidad a través de un patrón de agrupamiento de *proxies*, en el que se aprovisionan por lo menos uno, dos o tres nodos maestros, cada uno de ellos con un nodo de reserva dedicado. Este nodo de reserva actúa como respaldo en caso de que el(los) nodo(s) maestro(s) se averíe(n). Si se averían tanto el maestro como su respaldo vinculado, hay disponibles por lo menos una o dos particiones más en el agrupamiento. Esta redundancia que puede ser hasta séxtuple ofrecerá garantías de una disponibilidad muy alta. En general, el módulo de almacenamiento de datos 102 puede proporcionar un buen compromiso entre disponibilidad, consistencia de datos y latencia.

El módulo de almacenamiento de datos 102 puede proporcionar la posibilidad de crear notificaciones sobre espacios de claves, lo cual hace uso de una característica de publicación y suscripción (Pub/Sus) 104 para difundir todas las claves que han superado su tiempo de vida (TTL) en un canal dedicado. En esta invención, estas claves se pueden representar con los IDs de sesión. Por lo tanto, a cada aplicación que se sitúe a la escucha de este canal (uno o más Abonados 105) se le informaría de todos los IDs de sesión que han expirado o se suprimieron y, por lo tanto, la misma puede gestionar un procesado adicional.

El abonado 105 se puede configurar para recibir una notificación de que la asociación de la información identificable personalmente al dispositivo de cliente 1, tal como el identificador asociado a la información identificable personalmente, ha expirado o se ha suprimido permanentemente en el dispositivo de cliente 1. El abonado 105 también se puede configurar para publicar en una cola de espera de eventos 106 información sobre la asociación o identificador que ha expirado o se ha suprimido permanentemente.

El abonado 105 puede ser un servicio que se sitúa a la escucha del canal de notificación que difunde IDs de sesión que han expirado para procesarlos. Puesto que todos los abonados 105 pueden recibir deliberadamente todos los mensajes, puede tener una funcionalidad limitada y reducida y puede presentar un caudal teórico mayor que la frecuencia de publicación del módulo de almacenamiento de datos 102. La emisión de incidencias hacia el servidor de analítica 1000 utilizando el protocolo HTTP directamente desde el abonado 105 podría conducir a tiempos prolongados de los ciclos (incluso cuando se realiza de forma asíncrona) y podría introducir fases en las que se publiquen más notificaciones de expiración de sesión de las que se pueden procesar.

Por lo tanto, el abonado 105 puede recibir todos los mensajes desde el módulo de almacenamiento de datos 102 pero únicamente publica la información de sesiones que han expirado en la cola de espera de eventos 106. La notificación a través de la funcionalidad "pub/sus" 104 de las notificaciones sobre espacios de claves se puede basar en el paradigma de "disparar y olvidar". Esto significa que se dispara una notificación, incluso si no hay disponible en ese momento ningún abonado, lo cual conduciría a una situación en la que se almacenan datos de sesiones en el módulo de almacenamiento de datos 102 sin que ningún proceso aguas abajo tenga nunca conocimiento de su expiración. Esto podría acumular objetos residuales que nunca se limpian o reenvían al servidor de analítica 1000. Por lo tanto, se prefiere desplegar un mínimo de dos o tres abonados 105 para garantizar la tolerancia a fallos y evitar objetos residuales. Puesto que todos los abonados procesarán los mismos mensajes, se puede implementar deduplicación de datos aguas abajo de los abonados 105.

La ejecución de la solicitud HTTP real para cada incidencia con respecto al servidor de analítica 1000 puede resultar lenta debido al tiempo de respuesta de los Servidores seleccionados como objetivo. Por lo tanto, aspectos de la presente invención proponen el uso de la cola de espera de eventos 106 que actúa como memoria intermedia para todas las incidencias de analítica, que se consideran como incluidas en una lista blanca y se pueden reenviar al servidor 1000. La cola de espera de eventos 106 puede ayudar a acoplar con holgura los servicios que filtran todas las incidencias (notificaciones sobre espacios de claves y abonado 105) y el servicio que ejecutará la solicitud HTTP real (Expedidor 107). Posibilita el uso de un patrón de "pull", en el que el expedidor 107 acepta mensajes de la cola de espera 106 siempre que el Expedidor 107 esté listo, lo cual actúa como memoria intermedia en picos de tráfico.

La publicación de mensajes en la cola de espera 106 se puede lograr de manera más eficiente usando compresión, tratamiento por lotes y/u protocolos de comunicación optimizados. Si expiran más sesiones de las que un abonado 105 individual puede publicar en la cola de espera 106, esto podría conducir a un atasco del tráfico en la memoria del abonado lo cual puede derivar en un desbordamiento, en caso de que este estado esté en curso durante un periodo prolongado. El abonado 105 quedaría entonces indisponible temporalmente y no se expedirían todas las sesiones sin procesar de su memoria. Para superar esta implicación, en aspectos de la presente invención se optimiza la publicación en la cola de espera de tareas a través de un tratamiento por lotes y compresión así como una comunicación asíncrona, para optimizar capacidades del caudal.

El expedidor 107 se puede configurar para recolectar con solicitud previa [*pull*] la información sobre la asociación o identificador que ha expirado o se ha eliminado permanentemente de la cola de espera 106, y para recuperar, del módulo de almacenamiento de datos 102 y sobre la base de la información recolectada, los segundos datos, y preferentemente también los primeros datos, para la generación de la tercera solicitud 130.

El expedidor puede ser responsable de recolectar con solicitud previa, a partir de la cola de espera 106, IDs de sesión que han expirado y recuperar todas las incidencias atribuidas al ID de sesión a partir de la base de datos. Después de esto, el expedidor 107 reenvía las incidencias de HTTP al servidor de analítica 1000 (por ejemplo, Google Analytics). De manera similar al Receptor 101, el expedidor 107 debería ser preferentemente un servicio ligero que pueda replicarse múltiples veces y tener la capacidad de actuar independientemente de las otras replicaciones del expedidor, lo cual a su vez permite un escalado horizontal rápido con una carga creciente.

En otras palabras, puede que el abonado 105 no recupere datos del módulo de almacenamiento de datos 102, sino que únicamente se le informe de IDs de sesión que han expirado, los cuales posteriormente se transfieren a la cola de espera de eventos 106. No obstante, el expedidor 107 lee datos del módulo de almacenamiento de datos 102 y reenvía los datos recuperados al servidor externo 1000 que aloja la analítica. Las claves SN expiran cuando se supera un umbral de tiempo específico, tal como T1 o T2. El sistema, según ilustra la figura 8, sabe que la sesión ha terminado y que la clave correspondiente, que refleja el ID de sesión, ha expirado. El mensaje con la sesión que ha expirado se difunde a través del canal de publicación y suscripción 104. El abonado 105 está a la escucha de este canal 104 y recibe mensajes que incluyen el ID de sesión y el ID de sesión en la cola de espera de eventos 106. El expedidor 107 toma un ID de sesión aleatorio de la cola de espera de eventos 106, recupera todas las incidencias asociadas al ID de sesión a partir del módulo de almacenamiento de datos 102 y envía los datos recuperados al servidor externo 1000 que aloja la analítica solamente en cuanto se produzca el vencimiento o vencimiento de T1 o T2 o después del mismo, pero preferentemente antes del vencimiento o expiración de T3.

El sistema que ilustra la figura 8 se puede hacer funcionar en concordancia con cada una de las figuras 2 a 7. T2 se fija entre T1 y T3 preferentemente de tal manera que a) la diferencia de tiempo entre T2 y T1 tenga en cuenta el retardo máximo de las operaciones que tienen lugar antes del almacenamiento de los segundos datos en el

módulo de almacenamiento de datos 102, tales como las operaciones llevadas a cabo por el servidor sintetizador 10 externo al servidor de cuarentena 100 o por el receptor 101, y b) la diferencia de tiempo entre T2 y T3 tenga en cuenta un retardo máximo de las operaciones que tienen lugar en el servidor de cuarentena 100 después del módulo de almacenamiento de datos 102, tales como las operaciones llevadas a cabo por el abonado 105, la cola de espera de eventos 106 y/o el expedidor 107. Esto se puede realizar para evitar una pérdida de datos. Por ejemplo: T1 = 3.5 horas después de T0, T2 = 3.75 horas después de T0, y T3 = 4.0 horas después de T0.

Con respecto al retardo de T2 con respecto a T1 y a), se pueden insertar incidencias nuevas en el módulo de almacenamiento de datos 102 después de que haya expirado la sesión del lado del servidor T2. Esto podría ser el resultado de una interrupción temporal del funcionamiento del receptor 101 o problemas de comunicación con el receptor 101. Como consecuencia, se puede crear en el módulo de almacenamiento de datos 102 un objeto de sesión nuevo que contenga las incidencias retardadas. Dichas incidencias se pueden entregar después de la expiración de T3 y pueden ser rechazadas por el *software* de analítica alojado en el servidor externo 1000. La probabilidad de un escenario tan improbable se puede reducir adicionalmente aumentando la separación temporal entre T1 (por ejemplo, adelantando T1) y T2 (por ejemplo, posponiendo T2). En relación con el retardo de T3 con respecto a T2 y b, puede producirse una disrupción temporal del abonado, de la cola de espera de eventos o del expedidor, de manera que puede que se entreguen incidencias después de la expiración de T3 y las mismas pueden ser rechazadas por el *software* de analítica alojado en el servidor externo 1000, dando como resultado pérdida de datos. La probabilidad de este escenario, que es improbable, se puede reducir adicionalmente aumentando la separación temporal entre T2 (por ejemplo, adelantando T2) y T3 (por ejemplo, posponiendo T3).

En general, cualquier latencia o interrupción del funcionamiento también se puede reducir desplegando múltiples replicaciones de cada uno de los servicios 101-107 alojados en el servidor de cuarentena 100 para lograr redundancia.

La figura 9 ilustra un objeto de datos según aspectos de la invención, que mapea diferentes valores correspondientes a un elemento de información identificable personalmente a un mismo valor sintetizado.

Como se ha mencionado, la información identificable personalmente puede incluir, por ejemplo, un identificador de sesión asociado a un usuario del dispositivo de cliente 1, una parte de una dirección de IP del dispositivo de cliente 1, un agente de usuario de un navegador utilizado por un usuario en el dispositivo de cliente 1, una dirección de contacto asociada a un usuario del dispositivo de cliente 1 y/o un nombre asociado a un usuario del dispositivo de cliente 1. Es posible otra PII y la misma queda cubierta por la presente invención. Aspectos de la presente invención pueden sintetizar todos estos diferentes elementos de información identificable personalmente antes de transmitirlos al servidor externo que sirve de alojamiento para el *software* de analítica.

En cada una de las anteriores figuras 1 a 8, se han descrito las operaciones opcionales para ofuscar, reducir y/o anonimizar información identificable personalmente, tal como cualquiera de los elementos de PII mencionados en la presente (por ejemplo, un agente de usuario, una dirección de IP, un ID de sesión), por medio de la síntesis de datos sintéticos sobre la base de la información identificable personalmente. También se puede lograr anonimización eliminando parte de los valores antes de que se genere la tercera solicitud.

En este contexto, se han descrito aspectos de la invención tales que la información identificable personalmente se puede convertir en los datos sintéticos. La conversión en los datos sintéticos puede incluir: sustituir un primer valor de un primer elemento de la información identificable personalmente por un valor sintetizado que sea diferente del primer valor; e incluir en la primera solicitud 110 o la tercera solicitud 130 los datos sintéticos con el valor sintetizado.

A continuación, la sustitución del primer valor por el valor sintetizado puede incluir: eliminar el primer valor, de tal manera que la primera solicitud 110 o tercera solicitud 130 correspondiente no incluya y no esté asociada con el primer valor eliminado en correspondencia con el primer elemento de información identificable personalmente. El valor sintetizado se puede obtener a partir de un objeto de datos 900 que mapee diferentes valores para el primer elemento de información identificable personalmente con el valor sintetizado. La figura 9 ilustra un objeto de datos 900 ejemplificativo.

En particular, aspectos de la presente invención pueden sintetizar el agente de usuario, la dirección de IP y el ID de sesión antes de transmitirlos al servidor externo que aloja el *software* de analítica. La manera de sintetizar la dirección de IP se describió más arriba, tal como en el contexto de la figura 8 y el módulo de transformación 3, 4 o en el contexto de la sustitución de una parte de la dirección de IP por uno o más valores predeterminados.

Un agente de usuario del navegador del usuario puede ser una descripción de la configuración del *software* con la que el usuario visita un sitio web. El agente de usuario puede presentarse como parte de los encabezamientos de las solicitudes HTTP que inciden en el servidor 100 (en esta etapa se puede suprimir la totalidad del resto de encabezamientos de la solicitud HTTP). Por ejemplo, un agente de usuario típico tiene este aspecto:

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/97.0.4692.71

Safari/537.36

Este agente de usuario específico contiene la información de que un usuario utiliza, como sistema operativo, “Windows 10” y, como navegador, “Google Chrome”. Las versiones de los sistemas operativos y del navegador pueden ser muy detalladas y variar en un espacio de hasta diez dígitos (por ejemplo, 97.0.4692.71). Esto conduce a una cardinalidad elevada en la que solo un puñado de usuarios podría tener la versión 97.0.4692.71, lo cual podría dar como resultado un atributo que sirve para identificar a personas. Según aspectos de la presente invención, el método incluye identificar el navegador del usuario y su sistema operativo. Por tanto, no se requiere la granularidad completa que el agente de usuario proporciona de forma nativa y, por lo tanto, desde la perspectiva del análisis, la misma se puede modificar. No obstante, puede que algún *software* de analítica, tal como Google Analytics, no acepte fijar directamente el navegador/sistema operativo. Por lo tanto, se puede enviar un agente de usuario sintético completo para hacer que se informe con precisión sobre el navegador/sistema operativo.

Típicamente, para usuarios de configuraciones de *software* comunes, tales como “Google Chrome” + “Windows”, puede haber docenas de usuarios con el mismo agente de usuario. No obstante, para configuraciones de *software* menos comunes, tales como “Firefox” + “Linux”, no se puede garantizar un agente de usuario que no sea exclusivo. Por lo tanto, aspectos de la presente invención protegerán al usuario contra su identificación a través de su agente de usuario potencialmente exclusivo, intercambiando su agente de usuario real por un agente de usuario genérico y sintético que siga brindando las mismas propiedades genéricas, tales como el nombre del navegador y el sistema operativo.

Esto se puede lograr mediante aspectos de la invención que detectan el navegador y el sistema operativo a partir del agente de usuario en el sistema según la figura 8. En cuanto estas propiedades sean conocidas, las mismas se sustituirán por un agente de usuario universal que proporcione las mismas propiedades. Por ejemplo, el resumen genérico de la figura 9 muestra cómo agentes de usuario diferentes que tienen las mismas propiedades de navegador/sistema operativo se sustituirán por una versión genérica del agente de usuario con las mismas propiedades.

Para crear la lista de agentes de usuario sintéticos, aspectos de la invención analizan la prevalencia histórica de agentes de usuario originales y solamente hicieron uso de aquellas combinaciones (navegador + sistema operativo) que representaban por lo menos 250 entradas de usuario diferentes por mes. Todas las combinaciones que estén representadas por menos de 250 entradas de usuario por mes pueden descartarse y sustituirse por un agente de usuario vacío. Esta lista se revisará con una frecuencia anual para garantizar que se dispone de una metodología representativa y precisa.

Como se ha descrito, la conversión en los datos sintéticos puede incluir: sustituir un primer valor de un primer elemento de la información identificable personalmente (por ejemplo, un ID de sesión) por un valor sintetizado que es diferente del primer valor; e incluir en la primera solicitud 110 o la tercera solicitud 130 los datos sintéticos con el valor sintetizado. El valor sintetizado se puede obtener mediante una operación que se aplica al primer valor, en donde la operación incluye una función *hash* unidireccional. Las operaciones pueden incluir, además, complementar el primer valor con otro valor antes de que se aplique la función *hash* unidireccional al primer valor complementado. El otro valor puede ser un valor arbitrario, preferentemente en donde el otro valor incluye texto arbitrario.

Por ejemplo, un ID de sesión es un identificador que se puede utilizar para agrupar múltiples incidencias en un flujo continuo de interacciones, que comunica las acciones que ocurrieron en una plataforma (por ejemplo, en primer lugar se cargó la página, a continuación se dio inicio a un vídeo, todo ello dentro de la misma visita). El propio ID se puede almacenar en una *cookie*, directamente en el navegador del dispositivo. En configuraciones convencionales de herramientas de analítica, el ID de sesión se usa también para distinguir múltiples sesiones del mismo usuario cuando se visita la plataforma varias veces. No obstante, estos esfuerzos de distinción también podrían, potencialmente, identificar un dispositivo a través de la *cookie*, la cual persiste hasta que se suprima o expire. Esto podría dar como resultado un riesgo potencial para la seguridad de los datos y una infracción de los reglamentos legales.

Según aspectos de la presente invención, se fija un ID de sesión exclusivo para cada visita mediante una *cookie* propia que expirará después de 30 minutos de inactividad o después de 3.5 horas de actividad. El identificador, tal como la *cookie*, puede expirar a las 3.5 horas, pero son posibles otras expiraciones, tales como a las 10 horas. En otras palabras, la *cookie* se puede suprimir del navegador de manera automática después de que haya terminado la sesión. Para proteger el propio ID contra la denominada sincronización de *cookies*, en la que diferentes herramientas pueden sincronizar IDs de *cookies* (propias) y fusionar IDs entre sitios web, el valor del ID de sesión en el cliente se sustituirá por un “*hash* con sal”. Una función *hash* es una función unidireccional que siempre traducirá un contenido A en un contenido B. No obstante, puesto que es una función unidireccional, es imposible deducir el contenido A en caso de que llegases a conocer el contenido B. No obstante, si se conoce la función *hash*, podría ser que un atacante pudiese conjeturar continuamente sobre el ID de Sesión y validarlo con respecto al valor *hash* conocido. Aunque el ataque denominado de “fuerza bruta” para conjeturar sobre los IDs no es práctico debido al tamaño elevado de los potenciales IDs, los atacantes que tuvieran conocimiento de todos los IDs de

sesión durante una cierta ventana de tiempo dispondrían de un conjunto significativamente reducido del que elegir. Considerando el hecho de que solo existe un puñado de funciones *hash* seguras, preferentemente el valor de la *cookie* del ID de sesión no solamente se puede someter a una función *hash*, sino que también se puede complementar con un fragmento de texto arbitrario antes del proceso *hash* real. Esto hace que resulte prácticamente imposible conjeturar sobre el ID de sesión sometido a una función *hash*, incluso si se conociera el valor de la *cookie* del ID de sesión además de la función *hash*, lo cual crea un efecto de síntesis.

En conclusión, aspectos descritos de la presente invención permiten despersonalizar datos de usuarios y/o dispositivos de clientes antes de que se envíen a sistemas de analítica tales como Google Analytics. Se consigue que la identificación de un usuario/ o dispositivo sea cada vez más improbable o que ni siquiera sea posible. Aspectos de la presente invención proporcionan metodologías para anonimizar y sintetizar datos en un entorno escalable, que es independiente del propio sistema de analítica. Además, se describieron medios preferidos para mejorar la disponibilidad del sistema, la latencia del sistema y la escalabilidad.

Los aspectos y formas de realización antes descritos de la tecnología descrita en la presente se pueden implementar según cualquiera de entre numerosas maneras. Por ejemplo, las formas de realización se pueden implementar usando *hardware*, *software* o una combinación de los mismos. Cuando se implementan en *software*, el código de *software* se puede ejecutar en cualquier procesador o colección de procesadores adecuado, ya sea proporcionado en un único ordenador o distribuido entre múltiples ordenadores. Dichos procesadores se pueden implementar en forma de circuitos integrados, con uno o más procesadores en un componente de circuito integrado, incluidos componentes de circuito integrado disponibles comercialmente y conocidos en la técnica con denominaciones tales como chips de CPU, chips de GPU, microprocesador, microcontrolador o coprocesador. Alternativamente, un procesador se puede implementar en circuitería personalizada, tal como un ASIC, o circuitería semipersonalizada que resulte de la configuración de un dispositivo lógico programable. Como otra alternativa adicional más, un procesador puede ser una parte de un circuito o dispositivo semiconductor mayor, ya sea disponible comercialmente, semipersonalizado o personalizado. Como ejemplo específico, algunos microprocesadores disponibles comercialmente tienen múltiples núcleos de tal manera que uno o un subconjunto de esos núcleos puede constituir un procesador. No obstante, un procesador se puede implementar usando circuitería en cualquier formato adecuado.

Además, debe apreciarse que un “ordenador”, tal como el cliente 1, se puede materializar en cualquiera de entre una serie de formas, tales como un ordenador montado en bastidor, un ordenador de sobremesa, un ordenador portátil o un ordenador de tipo tableta. Adicionalmente, un ordenador puede estar integrado en un dispositivo que no se considere de manera general un ordenador pero con capacidades de procesado adecuadas, incluido un Asistente Digital Personal (PDA), un teléfono inteligente o cualquier otro dispositivo electrónico portátil o fijo adecuado.

Asimismo, un ordenador puede tener uno o más dispositivos de entrada y de salida. Estos dispositivos se pueden utilizar, entre otras cosas, para presentar una interfaz de usuario. Ejemplos de dispositivos de salida que pueden usarse para proporcionar una interfaz de usuario incluyen impresoras o pantallas de visualización para la presentación visual de salidas y altavoces u otros dispositivos generadores de sonido para presentaciones audibles de salidas. Los ejemplos de dispositivos de entrada que se pueden utilizar para una interfaz de usuario incluyen teclados y dispositivos señaladores, tales como ratones, paneles táctiles y tabletas digitalizadoras. Como ejemplo adicional, un ordenador puede recibir información de entrada a través de reconocimiento de voz o en otro formato audible.

Dichos ordenadores se pueden interconectar mediante una o más redes de cualquier forma adecuada, incluyendo en forma de red de área local o red de área extensa, tal como una red empresarial o Internet 121. Dichas redes se pueden basar en cualquier tecnología adecuada y pueden funcionar según cualquier protocolo adecuado y pueden incluir redes inalámbricas, redes por cable o redes de fibra óptica.

Asimismo, los diversos métodos o procesos esbozados en la presente se pueden codificar en forma de *software* que sea ejecutable en uno o más procesadores que utilicen uno cualquiera de una variedad de sistemas operativos o plataformas. Adicionalmente, dicho *software* se puede escribir utilizando cualquiera de una serie de lenguajes de programación y/o herramientas de programación o guionizado de instrucciones adecuados, y también se puede compilar en forma de código de lenguaje de máquina ejecutable o código intermedio que se ejecute en un entorno de trabajo o máquina virtual.

A este respecto, la presente invención se puede materializar en forma de un medio de almacenamiento legible por ordenador (o múltiples medios legibles por ordenador) (por ejemplo, una memoria de ordenador, uno o más discos flexibles, discos compactos (CD), discos ópticos, discos de vídeo digitales (DVD), cintas magnéticas, memorias *flash*, configuraciones de circuitos en Matrices de Puertas Programables in Situ u otros dispositivos semiconductores, u otro medio de almacenamiento informático tangible) codificado con uno o más programas que, cuando se ejecutan en uno o más ordenadores u otros procesadores, llevan a cabo métodos que implementan las diversas formas de realización de la invención analizadas anteriormente. Como se pone de manifiesto a partir de los ejemplos anteriores, un medio de almacenamiento legible por ordenador puede guardar información durante

un tiempo suficiente para proporcionar instrucciones ejecutables por ordenador en un formato no transitorio. Dicho medio o medios de almacenamiento legibles por ordenador pueden ser transportables, de tal manera que el programa o programas almacenados en los mismos se puedan cargar en uno o más ordenadores diferentes u otros procesadores para implementar diversos aspectos de la presente invención según se ha analizado anteriormente. Tal como se usa en la presente, el término "medio de almacenamiento legible por ordenador" abarca únicamente un medio no transitorio legible por ordenador que se puede considerar que es una manufactura (es decir, un artículo de manufactura) o una máquina. De manera alternativa o adicional, la invención se puede materializar en forma de un medio legible por ordenador que no sea un medio de almacenamiento legible por ordenador, tal como una señal que se propaga.

Los términos "algoritmo", "servicio", "código de programa", "programa informático" o "*software*" se utilizan en la presente en un sentido genérico para referirse a cualquier tipo de código informático o conjunto de instrucciones ejecutables por ordenador que se pueden utilizar para programar un ordenador u otro procesador con el fin de implementar diversos aspectos de la presente invención según se ha analizado más arriba. El módulo de almacenamiento de datos, el receptor, el abonado, el expedidor, el módulo de transformación y el sintetizador se pueden implementar con *software* informático y/o *hardware* informático. Adicionalmente, debería apreciarse que, según un aspecto de esta forma de realización, no es necesario que uno o más de los programas informáticos que, cuando se ejecutan, llevan a cabo métodos de la presente invención, residan en un único ordenador o procesador, sino que pueden estar distribuidos de forma modular entre una serie de ordenadores o procesadores diferentes para implementar diversos aspectos de la presente invención.

Las instrucciones ejecutables por ordenador pueden presentarse en muchos formatos, tales como módulos de programa, ejecutados por uno o más ordenadores u otros dispositivos. En general, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etcétera, que llevan a cabo tareas particulares o implementan tipos de datos abstractos particulares. Típicamente, la funcionalidad de los módulos de programa se puede combinar o distribuir según se desee en diversas formas de realización.

Asimismo, el registro de datos se puede almacenar en medios legibles por ordenador en cualquier formato adecuado. Para simplificar la ejemplificación, se puede mostrar que los registros de datos tienen entradas de registro de datos. No obstante, se puede utilizar cualquier mecanismo adecuado para establecer ubicaciones dentro del registro de datos con el fin de almacenar datos, tales como valores de parámetros.

Diversos aspectos de la presente invención se pueden utilizar de manera individual, combinados o en una variedad de disposiciones no analizadas específicamente en las formas de realización antes descritas y, por lo tanto, no se limitan en su aplicación a los detalles y disposición de componentes expuestos en la descripción anterior o ilustrados en los dibujos. Por ejemplo, aspectos descritos en una forma de realización se pueden combinar de cualquier manera con aspectos descritos en otras formas de realización.

Asimismo, la invención se puede materializar en forma de un método, del cual se ha proporcionado un ejemplo. Las acciones llevadas a cabo como parte del método se pueden ordenar de cualquier forma adecuada. Por consiguiente, se pueden construir formas de realización en las que se llevan a cabo acciones en un orden diferente al ilustrado, lo cual puede incluir llevar a cabo algunas acciones simultáneamente, incluso aunque se muestren como acciones secuenciales en formas de realización ilustrativas.

Además, algunas acciones se describen como efectuadas por un "usuario" o "jugador". Debería apreciarse que no es necesario que un "usuario" o "jugador" sea un único individuo, y que, en algunas formas de realización, acciones atribuibles a un "usuario" o "jugador" pueden ser llevadas a cabo por un equipo de individuos y/o un individuo en combinación con herramientas asistidas por ordenador u otros mecanismos. Debería apreciarse que no es necesario que un "usuario" o "jugador" sea un individuo y podría ser una máquina, tal como en forma de bot.

El uso de términos ordinales tales como "primer", "segundo", "tercero", etcétera, en las reivindicaciones para modificar un elemento de una reivindicación no implica por sí mismo ninguna prioridad, precedencia u orden de un elemento de una reivindicación con respecto a otro o el orden temporal en el que se llevan a cabo acciones de un método, sino que se utilizan meramente como indicativos para distinguir un elemento de una reivindicación que tiene una cierta denominación con respecto a otro elemento que tiene la misma denominación (excepto por el uso del término ordinal) con el fin de distinguir los elementos de las reivindicaciones. Asimismo, la fraseología y terminología utilizadas en la presente tienen fines descriptivos y no deberían considerarse como limitativas. El uso de "incluir", "comprender" o "tener", "contener", "involucrar" y variaciones de los mismos en la presente está destinado a abarcar los elementos enumerados tras ellos y sus equivalentes, así como elementos adicionales.

Además, aunque se indican ventajas de la presente invención, debería apreciarse que no toda forma de realización de la invención descrita en la presente incluirá cada una de las ventajas descritas. Puede que algunos aspectos y formas de realización no implementen ninguna de las características descritas como ventajosas en la presente y, en algunos casos, se pueden implementar una o más de las características descritas para lograr formas de realización adicionales. Por consiguiente, esta descripción y estos dibujos se aportan únicamente a título de ejemplo.

Tras haber descrito varios aspectos de por lo menos dos formas de realización de esta invención, debe apreciarse que a aquellos versados en la materia se les ocurrirán fácilmente diversos cambios, modificaciones y mejoras. Dichos cambios, modificaciones y mejoras están destinados a formar parte de esta divulgación, y están destinados a situarse dentro del alcance de la invención según definen las reivindicaciones adjuntas.

REIVINDICACIONES

1. Método para proporcionar información identificable personalmente anonimizada, en el que el método se implementa mediante uno o más primeros ordenadores, comprendiendo el método:

obtener una primera solicitud (110), incluyendo la primera solicitud (110) unos primeros datos que indican acciones que ocurrieron en un dispositivo de cliente (1) y unos segundos datos que están asociados a los primeros datos y que se basan en información identificable personalmente asociada con el dispositivo de cliente (1), en el que los segundos datos incluidos en la primera solicitud (110) obtenida son datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente asociada con el dispositivo de cliente (1) o es la información identificable personalmente asociada con el dispositivo de cliente (1);

poner en cuarentena por lo menos los segundos datos, incluyendo la puesta en cuarentena por lo menos almacenar por lo menos los segundos datos en un módulo de almacenamiento de datos (102);

recuperar datos del módulo de almacenamiento de datos (102), basándose los datos recuperados en los segundos datos almacenados;

generar una tercera solicitud (130) para su transmisión a un servidor (1000) externo a dicho uno o más primeros ordenadores, incluyendo la tercera solicitud (130) los primeros datos y datos sintéticos asociados a los primeros datos, basándose los datos sintéticos de la tercera solicitud (130) en los datos recuperados, y siendo los datos sintéticos de la tercera solicitud (130) los datos sintéticos de la primera solicitud (110) o siendo sintetizados sobre la base de la información identificable personalmente incluida en la primera solicitud (110),

en el que los datos sintéticos incluyen una versión ofuscada, reducida y/o a la que se ha aplicado una función *hash*, de la información identificable personalmente, y

en el que los datos recuperados se recuperan del módulo de almacenamiento de datos (102), o la tercera solicitud (130) es transmisible al servidor externo (1000), solamente en cuanto o solamente después de que expire una asociación de por lo menos una parte de la información identificable personalmente con el dispositivo de cliente (1) o su usuario; y

transmitir la tercera solicitud (130) generada al servidor externo (1000).

2. Método según la reivindicación 1, en el que los datos recuperados se recuperan del módulo de almacenamiento de datos (102), o la tercera solicitud (130) es transmisible al servidor externo (1000), solamente después de que haya expirado una cantidad de tiempo predeterminada (T1, T2) desde un tiempo (T0) en el que ocurrieron algunas de las acciones en el dispositivo de cliente (1), estando la cantidad de tiempo predeterminada en concordancia con la expiración de la asociación de la parte de la información identificable personalmente al dispositivo de cliente (1) o su usuario.

3. Método según la reivindicación 1 o 2, en el que la asociación de por lo menos una parte de la información identificable personalmente con el dispositivo de cliente (1) o su usuario ha expirado por expiración o supresión permanente, en el dispositivo de cliente (1), de un identificador, tal como una *Cookie*, que está asociado a la parte de la información identificable personalmente.

4. Método según una cualquiera de las reivindicaciones anteriores, en el que los datos recuperados a partir del módulo de almacenamiento de datos (102) son los datos sintéticos que se sintetizaron sobre la base de la información identificable personalmente asociada con el dispositivo de cliente (1).

5. Método según la reivindicación 4, que comprende asimismo:

por parte de uno o más segundos ordenadores,

recibir una segunda solicitud (20, 120) del dispositivo de cliente (1), incluyendo la segunda solicitud (20, 120) los primeros datos y la información identificable personalmente, y

generar la primera solicitud (110) sobre la base de la segunda solicitud (20, 120), incluyendo la generación de la primera solicitud (110):

sintetizar los datos sintéticos, incluyendo convertir por lo menos alguna información identificable personalmente incluida en la segunda solicitud (20, 120) en los datos sintéticos; e

incluir en la primera solicitud (110) los datos sintéticos en lugar de la información identificable personalmente.

6. Método según una cualquiera de las reivindicaciones 1 a 3, en el que los datos recuperados a partir del módulo

de almacenamiento de datos (102) son la información identificable personalmente asociada con el dispositivo de cliente (1).

7. Método según la reivindicación 6, en el que la generación de la tercera solicitud (130) incluye:

sintetizar los datos sintéticos, incluyendo convertir por lo menos algunos los segundos datos o datos recuperados en los datos sintéticos; e

incluir en la tercera solicitud (130) los datos sintéticos en lugar de los segundos datos o datos recuperados.

8. Método según una cualquiera de las reivindicaciones anteriores, en el que el servidor externo (1000) está configurado para alojar un *software* de herramienta de analítica de terceros para analizar acciones que ocurrieron en el dispositivo de cliente (1).

9. Método según una cualquiera de las reivindicaciones anteriores, en el que la tercera solicitud (130) no incluye la información identificable personalmente.

10. Método según la reivindicación 9, en el que la tercera solicitud (130) se transmite al servidor externo (1000) solamente después de que expire la cantidad de tiempo predeterminada (T1, T2) desde que ocurrieron algunas de las acciones en el dispositivo de cliente (1).

11. Método según la reivindicación 9 o 10, en el que la tercera solicitud (130) se transmite al servidor externo (1000) solamente en cuanto o solamente después de que expire la asociación de la parte de la información identificable personalmente al dispositivo de cliente (1) o su usuario.

12. Método según una cualquiera de las reivindicaciones 1 a 11, en el que la conversión en los datos sintéticos incluye:

sustituir un primer valor de un primer elemento de la información identificable personalmente por un valor sintetizado que es diferente del primer valor; e

incluir los datos sintéticos con el valor sintetizado en la primera solicitud (110) o la tercera solicitud (130).

13. Método según la reivindicación 12, en el que la sustitución del primer valor por el valor sintetizado incluye:

eliminar el primer valor, de tal manera que la primera solicitud (110) o tercera solicitud (130) correspondiente no incluya y no esté asociada con la primera solicitud eliminada en correspondencia con el primer elemento de información identificable personalmente.

14. Método según la reivindicación 12 o 13, en el que el valor sintetizado se obtiene a partir de un objeto de datos (900) que mapea con el valor sintetizado valores diferentes para el primer elemento de información identificable personalmente.

15. Método según una cualquiera de las reivindicaciones 12 a 14, en el que el primer valor está contenido en un encabezamiento asociado a la primera solicitud (110) o la segunda solicitud (20, 120), y/o en el que el valor sintetizado es un atributo de o está contenido en un encabezamiento asociado a la primera, segunda o tercera solicitud (20, 110, 120, 130).

16. Método según una cualquiera de las reivindicaciones 12 a 15, en el que el valor sintetizado se obtiene mediante una operación que se aplica al primer valor, incluyendo la operación una función *hash* unidireccional.

17. Método según una cualquiera de las reivindicaciones anteriores, en el que los datos recuperados no han sido recuperables a partir del módulo de almacenamiento de datos (102) durante la cuarentena hasta que los datos recuperados se pusieron a disposición para su recuperación a partir del módulo de almacenamiento de datos (102), o en el que la tercera solicitud no es transmisible al servidor externo (1000) durante la cuarentena hasta que la tercera solicitud se ponga a disposición para su transmisión hacia el servidor externo (1000).

18. Método según una cualquiera de las reivindicaciones 1 a 17, en el que la tercera solicitud (130) se transmite al servidor externo (1000) antes de la expiración de un límite de tiempo (T3) para aceptar solicitudes según son proporcionadas por el *software* de analítica de terceros alojado en el servidor externo (1000).

19. Método según una cualquiera de las reivindicaciones anteriores, en el que la puesta en cuarentena incluye asimismo recuperar los segundos datos o datos recuperados y/o la síntesis de los datos sintéticos.

20. Por lo menos un medio de almacenamiento no transitorio legible por ordenador que almacena instrucciones ejecutables por ordenador que, cuando son ejecutadas por uno o más ordenadores, consiguen que el ordenador

u ordenadores lleven a cabo el método según una cualquiera de las reivindicaciones anteriores.

21. Sistema informático, que comprende:

por lo menos un procesador de *hardware* informático; y

por lo menos un medio de almacenamiento no transitorio legible por ordenador que almacena instrucciones ejecutables por procesador que, cuando son ejecutadas por dicho por lo menos un procesador de *hardware* informático, consiguen que dicho por lo menos un procesador de *hardware* informático lleve a cabo el método según una cualquiera de las reivindicaciones 1 a 19.

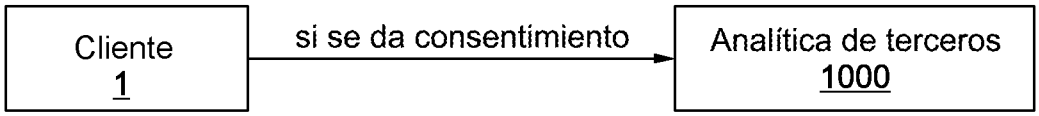

Fig. 1 (técnica anterior)

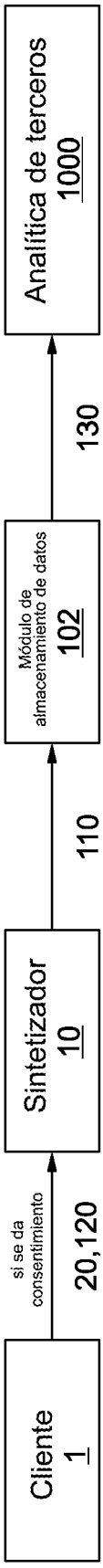
Cliente
1
si se da consentimiento
20,120
Sintetizador
10
110
Módulo de almacenamiento de datos
102
130
Analítica de terceros
1000

Fig. 2

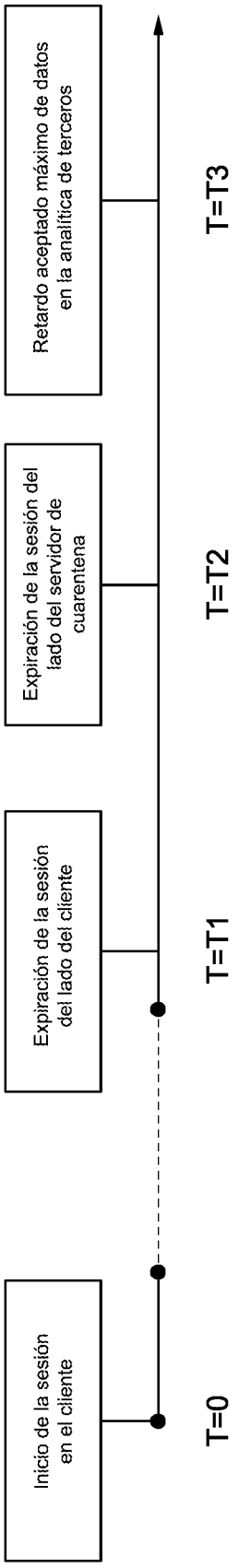
Inicio de la sesión en el cliente
T=0
Expiración de la sesión del lado del cliente
T=T1
Expiración de la sesión del lado del servidor de cuarentena
T=T2
Retardo aceptado máximo de datos en la analítica de terceros
T=T3

Fig. 3

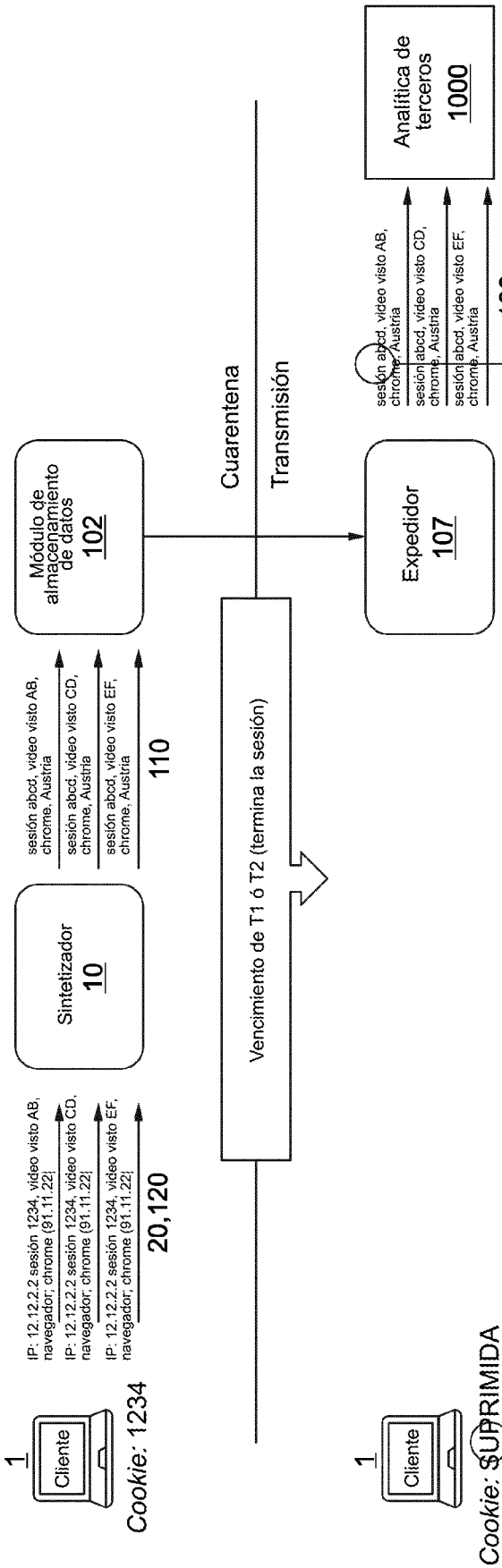
1
Cliente
Cookie: 1234
IP: 12.12.2.2 sesión 1234, video visto AB, navegador; chrome (91.11.22)
IP: 12.12.2.2 sesión 1234, video visto CD, navegador; chrome (91.11.22)
IP: 12.12.2.2 sesión 1234, video visto EF, navegador; chrome (91.11.22)
20,120
Sintetizador
10
sesión abcd, video visto AB, chrome, Austria
sesión abcd, video visto CD, chrome, Austria
sesión abcd, video visto EF, chrome, Austria
110
Módulo de almacenamiento de datos
102
Cuarentena
Transmisión
Vencimiento de T1 ó T2 (termina la sesión)
Expedidor
107
sesión abcd, video visto AB, chrome, Austria
sesión abcd, video visto CD, chrome, Austria
sesión abcd, video visto EF, chrome, Austria
Analítica de terceros
1000
130
1
Cliente
Cookie: $UPRIMIDA

Fig. 4

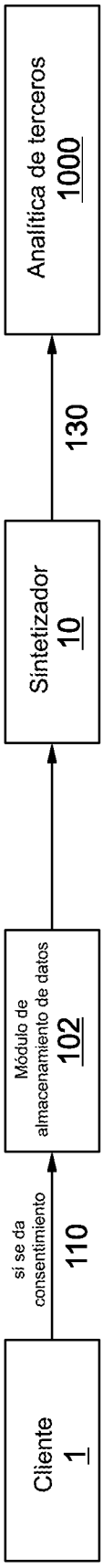
Cliente
1
sí se da consentimiento
110
Módulo de almacenamiento de datos
102
Sintetizador
10
130
Analítica de terceros
1000

Fig. 5

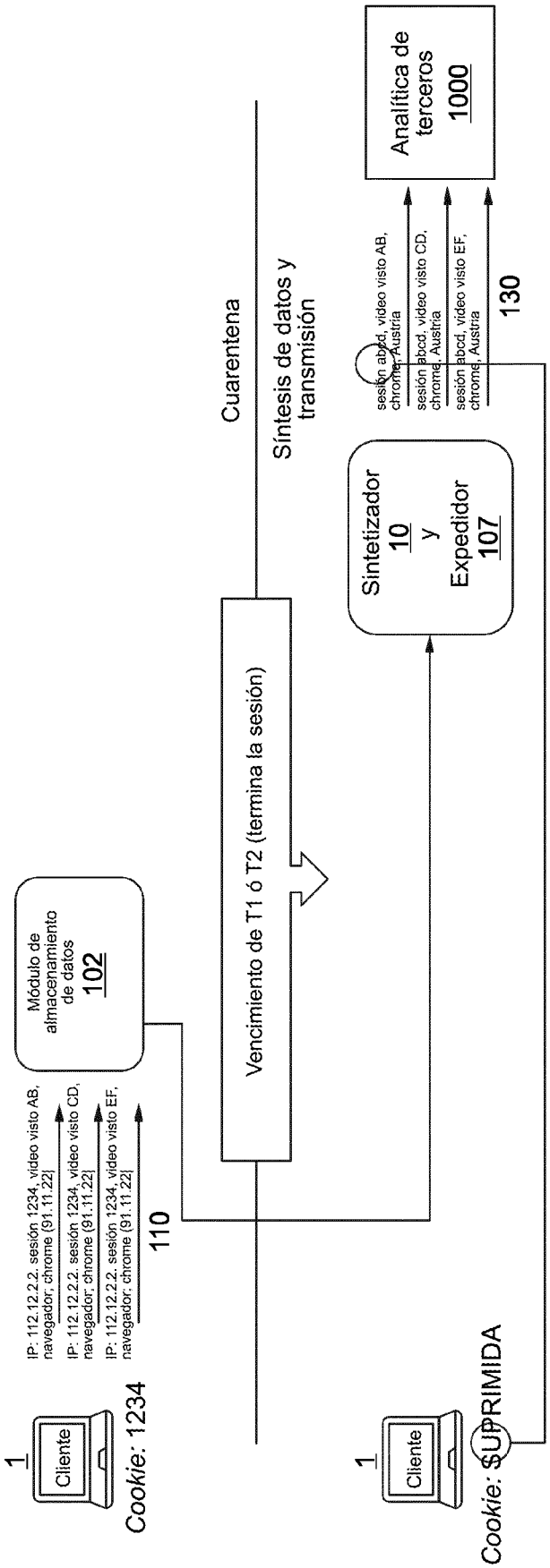
1
Cliente
Cookie: 1234
IP: 112.12.2.2, sesión 1234, video visto AB, navegador: chrome (91.11.22)
IP: 112.12.2.2, sesión 1234, video visto CD, navegador: chrome (91.11.22)
IP: 112.12.2.2, sesión 1234, video visto EF, navegador: chrome (91.11.22)
110
Módulo de almacenamiento de datos 102
Vencimiento de T1 ó T2 (termina la sesión)
Cuarentena
Síntesis de datos y transmisión
Sintetizador 10 y Expedidor 107
sesión abcd, video visto AB, chrome, Austria
sesión abcd, video visto CD, chrome, Austria
sesión abcd, video visto EF, chrome, Austria
Analítica de terceros 1000
130
1
Cliente
Cookie: $UPRIMIDA

Fig. 6

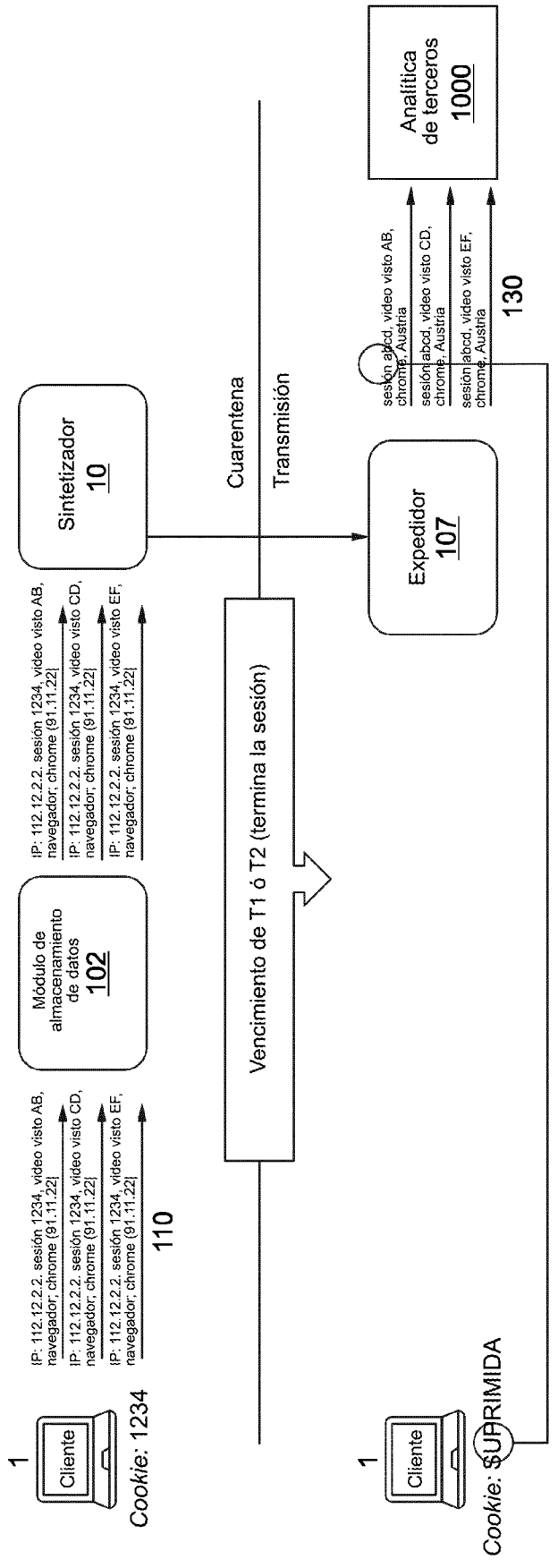
1
Cliente
Cookie: 1234
110
IP: 112.12.2.2, sesión 1234, video visto AB, navegador; chrome (91.11.22)
IP: 112.12.2.2, sesión 1234, video visto CD, navegador; chrome (91.11.22)
IP: 112.12.2.2, sesión 1234, video visto EF, navegador; chrome (91.11.22)
Módulo de almacenamiento de datos
102
Vencimiento de T1 ó T2 (termina la sesión)
Sintetizador
10
Cuarentena
Transmisión
Expedidor
107
sesión abcd, video visto AB, chrome, Austria
sesión abcd, video visto CD, chrome, Austria
sesión abcd, video visto EF, chrome, Austria
130
Analítica de terceros
1000
1
Cliente
Cookie: $URRIMIDA

Fig. 7

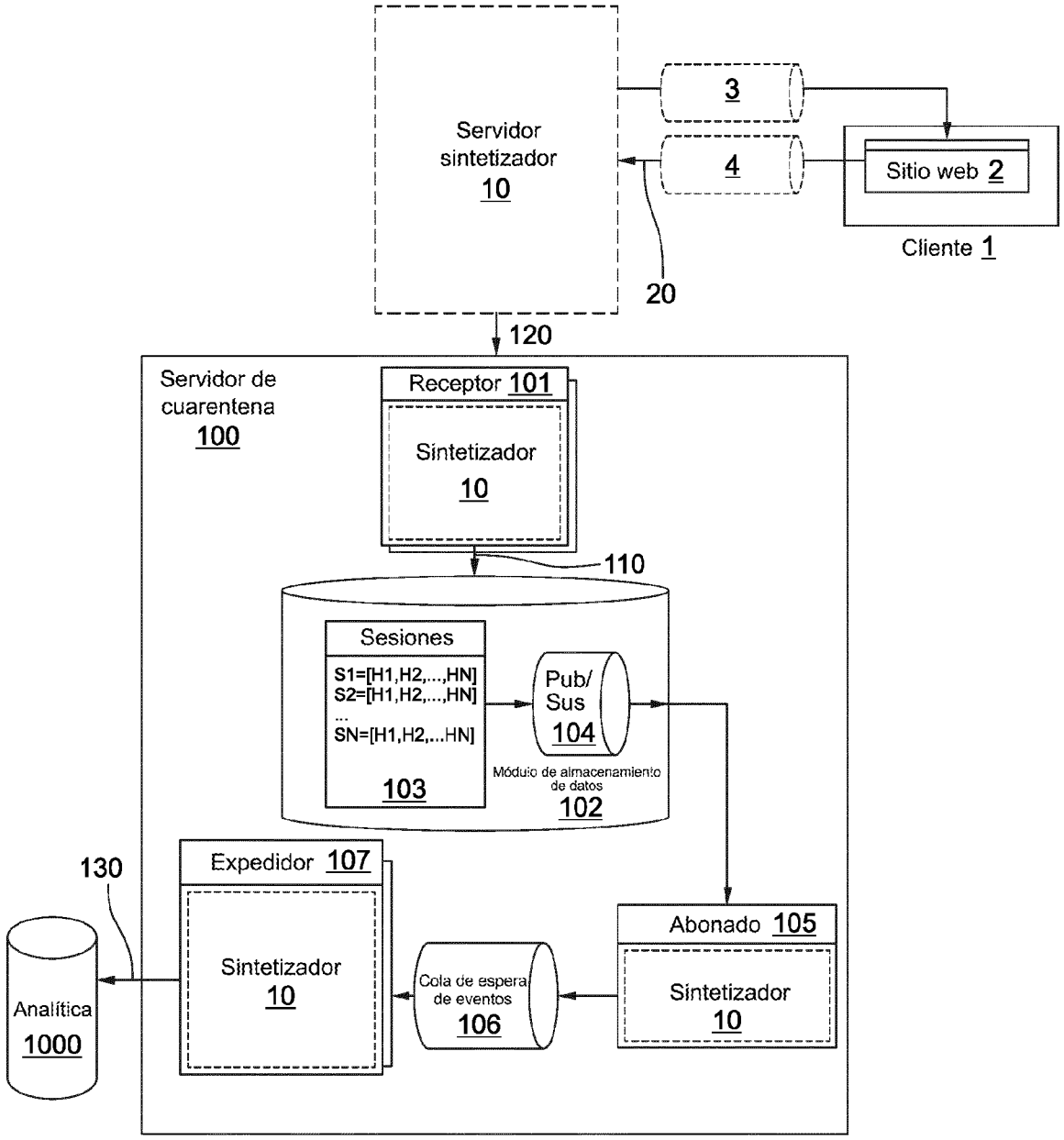
Servidor sintetizador 10
3
4
Sitio web 2
Cliente 1
20
120
Servidor de cuarentena 100
Receptor 101
Sintetizador 10
110
Sesiones
S1=[H1,H2,...,HN]
S2=[H1,H2,...,HN]
...
SN=[H1,H2,...HN]
103
Pub/ Sus 104
Módulo de almacenamiento de datos 102
130
Expedidor 107
Sintetizador 10
Analítica 1000
Cola de espera de eventos 106
Abonado 105
Sintetizador 10

Fig. 8

Navegador/ Sistema operativo	Agente de usuario original	Agente de usuario sintético
Safari /iOS	Mozilla/5.0 (iPhone; CPU iPhone OS 15_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.1 Mobile/15E148 Safari/604.1	Mozilla/5.0 (iPhone; CPU iPhone OS 15_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.1 Mobile/15E148 Safari/604.1
Safari /iOS	Mozilla/5.0 (iPhone; CPU iPhone OS 15_2_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.2 Mobile/15E148 Safari/604.1	Mozilla/5.0 (iPhone; CPU iPhone OS 15_1 like Mac OS X) AppleWebkit/605.1.15 (KHTML, like Gecko) Version/15.1 Mobile/15E148 Safari/604.1
Safari /iOS	iPhone XR/iOS-14.8.1/Version: 5.9.1/Build: 3	Mozilla/5.0 (iPhone; CPU iPhone OS 15_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.1 Mobile/15E148 Safari/604.1

Fig. 9