(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 751 359**

(51) Int. CI.:

G06Q 30/02 (2012.01)
H04L 9/32 (2006.01)
H04W 12/02 (2009.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **28.11.2013 PCT/EP2013/003601**

(87) Fecha y número de publicación internacional: **05.06.2014 WO14082749**

(96) Fecha de presentación y número de la solicitud europea: **28.11.2013 E 13798583 (4)**

(97) Fecha y número de publicación de la concesión europea: **17.07.2019 EP 2926308**

(54) Título: **Método de anonimización mediante transmisión de un conjunto de datos entre diferentes entidades**

(30) Prioridad:

28.11.2012 WO PCT/EP2012/004920
24.10.2013 EP 13005086

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
31.03.2020

(73) Titular/es:

TELEFÓNICA GERMANY GMBH & CO. OHG (100.0%)
Georg-Brauchle-Ring 50
80992 München, DE

(72) Inventor/es:

UKENA, JONATHAN y
SCHÖPF, PHILIPP

(74) Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

DESCRIPCIÓN

Método de anonimización mediante transmisión de un conjunto de datos entre diferentes entidades

La presente invención se refiere a un método para anonimización mediante la transmisión de un conjunto de datos de eventos de al menos una entidad proveedora de datos que suministra el conjunto de datos de eventos a al menos una entidad de agregación de datos que agrega el conjunto de datos de eventos en el que el conjunto de datos incluye al menos un identificador que se identifica en al menos un usuario de la entidad proveedora de datos.

Los sistemas de comunicación permiten la comunicación entre dos o más entidades. Una entidad en el contexto de este documento debe definirse como un área específica que es organizacional separada de otras áreas debido a especificaciones lógicas, físicas o legales. Además de mover datos de carga útil entre estas entidades, los sistemas de comunicación deben generar, recopilar y procesar datos de gestión como direcciones, ubicaciones, descripciones de servicios, etc. Por ejemplo, para un servidor web que se comunica con un ordenador de cliente, el servidor web necesita procesar la dirección IP de un cliente, las URL solicitadas, información de encabezado HTTP y datos de sesión. En un sistema de comunicación móvil, datos adicionales tales como información de ubicación, tipo de servicios o identificadores que identifican el dispositivo móvil (IMEI), las tarjetas SIM (IMSI) se procesan. Además, cada relación de comunicación crea datos adicionales a los que se hace referencia a continuación como un conjunto de datos.

Además, dichos sistemas/redes, en particular, sistemas de comunicación móvil, podrían recopilar continuamente datos adicionales nombrados como datos de eventos de ubicación durante la operación regular del sistema/red. Cada conjunto de datos de eventos está relacionado con un evento específico de un suscriptor individual. Los eventos pueden ser activados por un suscriptor/usuario, la red o por un dispositivo que no tiene importancia para su posterior procesamiento. El conjunto de datos incluye varios atributos que describen diferentes propiedades del evento desencadenado. Estos conjuntos de datos de eventos están asociados con un identificador personal que permite la asignación del conjunto de datos de eventos a un suscriptor individual del sistema de comunicación.

Además, los operadores de sistemas de comunicación registran datos relacionados con el cliente, tal como los datos de contacto y la información del contrato. La recopilación de estos datos es necesaria para fines de facturación o para estar disponible para las autoridades. A continuación, dichos datos se definen como datos de relación con el cliente (CRM). Los datos de CRM pueden agregarse para formar datos de clase de cliente.

Debido a la retención de esta información, tales sistemas llamados entidades de suministro de datos, en particular, sistemas de comunicación móvil, ofrecen la posibilidad de proporcionar información sobre los hábitos de los suscriptores, por ejemplo, con respecto a sus datos de ubicación durante un intervalo de tiempo definido. Estos datos pueden usarse para crear perfiles de ubicación para sitios geográficos o para derivar patrones dinámicos de movimiento de multitudes. En este contexto, la información podría ser útil para una amplia gama de aplicaciones en el área de servicios de tráfico, servicios de ciudad inteligente, servicios de optimización de infraestructuras, servicios de información de minoristas, servicios de seguridad y muchos más. Por lo tanto, es deseable proporcionar la información generada en forma adecuada a las partes llamadas entidades de agregación de datos que se benefician de aplicaciones como las mencionadas anteriormente. Tales partes podrían incluir consejos locales, empresas de transporte público e infraestructura como proveedores de transporte público o proveedores de electricidad, minoristas, grandes organizadores de eventos u organismos de seguridad pública y muchos más usos y usuarios aún desconocidos.

Sin embargo, es obligatorio proporcionar esta información de forma anónima para proteger la privacidad de cada individuo, en particular, cada suscriptor del sistema de comunicación móvil. En consecuencia, el proveedor del sistema de comunicación móvil (entidad proveedora de datos) que proporciona esta información solo debe proporcionar información extraída de datos anónimos y agregados sin vender información personal divulgada. La divulgación de cualquier información personal está estrictamente prohibida en ciertas jurisdicciones, pero también puede ser no deseada en otras, el seguimiento e identificación de individuos debe evitarse en cualquier circunstancia.

El documento US 2009/0048977 A1 divulga una infraestructura publicitaria que proporciona anuncios individuales de usuario a un usuario de una red de comunicación móvil. Por lo tanto, un terminal de acceso inalámbrico puede obtener un perfil de mensaje de orientación. Para fines de anonimización, un identificador del usuario está encriptado por una función hash. El intercambio de datos entre el terminal del suscriptor y la infraestructura publicitaria se cifra mediante un proceso de cifrado simétrico. Métodos y sistemas similares se describen en los documentos US 2009/0327488 A1 y US 2011/0264527 A1.

Un proceso de cifrado iterativo se divulga. En la publicación de David L. Chaum *"untraceable electronic mail return addresses, and digital pseudonyms".*

Es el objeto de la invención proporcionar un método para la anonimización de los datos recopilados o utilizados dentro de un sistema o red arbitrario y que cada uno está relacionado con un suscriptor individual del sistema.

El objeto mencionado anteriormente se resuelve mediante un método según la combinación de características de la reivindicación 1. Realizaciones preferidas son la materia de las reivindicaciones dependientes 2 a 16.

De acuerdo con la reivindicación 1 de la invención, se propone un método para anonimizar uno o más identificadores de un conjunto de datos. Esto se logra realizando una técnica de ofuscación muy específica combinada con la transmisión de dicho conjunto de datos desde al menos una entidad proveedora de datos que suministra el conjunto de datos a al menos una entidad de agregación de datos que agrega el conjunto de datos. El conjunto de datos incluye al menos un identificador que identifica al menos un suscriptor de un "sistema o red", en particular, una red de comunicación móvil. El conjunto de datos ya existe o se genera en la entidad proveedora de datos. La entidad de agregación de datos está interesada en la información contenida en dicho conjunto de datos para cualquier propósito de aplicación, tal como análisis de datos para crear perfiles de ubicación para sitios geográficos o la derivación de patrones dinámicos de movimiento de multitudes.

Para mantener la privacidad del al menos un suscriptor de la entidad proveedora de datos, se realizan las siguientes etapas del método para ofuscar la información personal sobre el suscriptor:

a. realizar un cifrado básico no reversible del al menos un identificador utilizando un mecanismo de cifrado con una determinada vida útil,
b. realizar iterativamente un cifrado adicional de dicho identificador cifrado básico para n veces, lo que da como resultado un identificador cifrado adicional n veces con n igual o mayor que 1,
en el que el cifrado adicional comprende las etapas de agregar un componente aleatorio a dicho identificador cifrado y cifrar la salida usando cifrado asimétrico con una clave pública,
en el que cada iteración usa un componente aleatorio con una vida útil más corta que la vida útil de la iteración anterior y una clave pública diferente a la de la iteración anterior, en el que la vida útil de la primera iteración es más corta que la vida útil determinada de acuerdo con la etapa b,
c. transmitir el conjunto de datos de evento caracterizado por el identificador cifrado adicional n veces por lo menos a una entidad de agregación de datos.
d. deshacer al menos una iteración de los cifrados adicionales en la primera entidad de agregación de datos de recepción que da como resultado un identificador cifrado adicional n-1 veces en el que el primer agregador de datos deriva al menos del índice estadístico de una serie de conjuntos de datos de eventos relacionados con el mismo identificador cifrado adicional n-1 veces y recopilado dentro de la vida útil del conjunto de componentes aleatorios durante el cifrado adicional n-1,
e. el primer agregador de datos reenvía dichos índices estadísticos derivados, junto con su identificador cifrado adicional n-1 veces adicionales a al menos otra entidad de agregación de datos en la que la recepción de al menos otra entidad de agregación de datos deshace al menos una iteración del identificador de cifrado adicional n-1 que resulta en el identificador cifrado adicional n-2 veces o el identificador cifrado básico no reversible y almacena los índices estadísticos recibidos relacionados con el mismo identificador cifrado adicional n-2 veces o el identificador cifrado básico no reversible dentro de la vida útil del conjunto de componentes aleatorios durante la iteración de deshacer del identificador de cifrado adicional n-2 o la cierta vida útil utilizada durante el cifrado básico no reversible.

En la etapa a, se realiza un primer cifrado del al menos un identificador utilizando un mecanismo de cifrado con una vida útil específica, es decir, el mecanismo produce la misma salida dentro de dicho intervalo de tiempo (vida útil). En particular, dicho cifrado se realiza mediante una función de hash con una semilla que se crea en un entorno seguro y nunca es conocida por ninguna de las entidades involucradas. Dicho entorno seguro, por ejemplo, podría establecerse mediante el uso de hardware criptográfico adecuado o un componente de software debidamente asegurado. Dicho componente también podría distribuir la semilla secreta a diferentes entidades. Esto permitiría, por ejemplo, que incluso diferentes compañías generen una representación cifrada idéntica para el mismo identificador. La semilla se puede cambiar después de un cierto intervalo de tiempo (vida útil), preferiblemente de forma regular. Esta primera etapa se denomina anonimización de base y conduce a un identificador personal llamado anonimato de base.

En una segunda etapa (etapa b), un componente aleatorio, por ejemplo, una serie de caracteres aleatorios que definen una cadena de caracteres, que comprende preferiblemente caracteres alfanuméricos, en particular, una serie de dígitos y/o letras y/o cualquier otro tipo de símbolos, se agrega al identificador ofuscado cifrado único. De nuevo, el componente aleatorio puede cambiarse después de un cierto intervalo de tiempo, preferiblemente de forma regular. La vida útil del componente aleatorio es menor que la vida útil en la etapa b, en particular, según la semilla utilizada en la base de anonimización.

Después de agregar el componente aleatorio, el resultado se cifra mediante un mecanismo de cifrado asimétrico. La entidad proveedora de datos utiliza la clave pública de la entidad de agregación de datos para el cifrado. El resultado de la etapa b) es un identificador personal encriptado adicional anónimo y 1 vez (n = 1).

La etapa b se llama cifrado adicional. El cifrado adicional se puede realizar en varias iteraciones (n veces) que conducen a identificadores ofuscados múltiples intercalados (identificador personal cifrado adicional n veces y anónimo de base). El componente aleatorio de una iteración se cambia después de un período de tiempo más corto que el de la iteración anterior. Debido a estas iteraciones, el resultado de una determinada iteración solo es accesible si las iteraciones ejecutadas después ya se han invertido.

En una realización preferida particular, la última iteración usa una vida útil que es válida solo para un evento único, es decir, el componente aleatorio cambia después de cada evento (conjunto de datos). La última iteración de cifrado adicional siempre tiene que usar un componente aleatorio que cambia constantemente, para que el resultado de esta última iteración de cifrado no tenga estabilidad en absoluto. El número de encriptaciones adicionales define así el número máximo de diferentes tipos de identificadores anónimos (con posibles vidas diferentes por tipo) que podrían usarse dentro de una cadena de varias entidades de agregación de datos que procesan sucesivamente los datos. Debido a estas iteraciones, el resultado de una determinada iteración solo es accesible si las iteraciones ejecutadas después ya se han invertido.

En la siguiente etapa c, el conjunto de datos que incluye el identificador personal cifrado adicional n veces y anónimo de base se transmite al menos a una entidad de agregación de datos.

En la última etapa d, la al menos una entidad de agregación de datos obtiene el conjunto de datos y deshace el último cifrado adicional (n veces). Por lo tanto, la clave privada apropiada del par de claves asimétricas mencionado anteriormente se utiliza para revelar la combinación de un identificador personal cifrado adicional n-1 veces y anónimo de base junto con el componente aleatorio. Basado en el conocimiento de cómo ambos valores se combinaron, el componente aleatorio puede ser eliminado. Esto termina en el identificador encriptado adicional n-1 veces y anónimo de base. La vida útil de dicho identificador se estableció en el cifrado adicional n-1. Dentro de esta vida útil, el agregador de datos puede vincular conjuntos de datos inicialmente relacionados con un mismo individuo y realizar cálculos basados en los conjuntos de datos combinados.

Todo el método se denomina Proceso de anonimización multinivel (MAP) que se puede ejecutar en instancias paralelas. La configuración de una instancia MAP define la anonimización base (duración de la semilla secreta), así como n iteraciones de encriptaciones adicionales (para cada una: duración del componente aleatorio, forma de combinar componente aleatorio con identificador cifrado, clave pública para usar para el cifrado). El resultado de tal instancia es una anidación de n diferentes tipos de identificadores, cada uno potencialmente con una vida útil específica para una serie de entidades de agregación de datos que procesan datos posteriormente en una ruta de transmisión de datos. De ese modo, la vida útil según la etapa a, en particular, la vida útil de la semilla secreta define la vida útil máxima para cualquier tipo de identificadores dentro de las diferentes entidades de agregación de datos. Cada iteración de cifrado adicional solo puede definir una vida útil más corta que la que se definió en la iteración anterior. Preferentemente, la última iteración de cifrado adicional siempre usa un componente aleatorio que cambia constantemente, de modo que el resultado del MAP en la al menos una entidad proveedora de datos no tiene estabilidad en absoluto.

Más allá de dicha definición para operar posteriormente entidades de agregación de datos, también sería posible definir diferentes tipos de identificadores anónimos para entidades de agregación de datos que operan en rutas de transmisión de datos paralelas. Esto se haría definiendo diferentes configuraciones MAP que se ejecutan en iteraciones paralelas MAP.

La invención describe un proceso de anonimización de niveles múltiples, implementado preferiblemente como parte de un sistema de comunicación en la al menos una entidad de suministro de datos y un sistema informático genérico en la al menos una entidad de agregación de datos. Al realizar las etapas antes mencionadas, al menos un identificador personal incluido en el conjunto de datos se anonimiza para evitar cualquier inferencia en el individuo o usuario respectivo fuera de la entidad proveedora de datos. Preferentemente, tan pronto como se inicia la ejecución del método, la entidad proveedora de datos no puede leer y/o modificar y/o descifrar el identificador personal. La anonimización de base y n iteraciones de encriptaciones adicionales se ejecutan en un sistema cerrado en al menos una entidad proveedora de datos. Por lo tanto, dicha entidad no tendrá acceso a ningún resultado intermedio.

El enfoque inventivo da como resultado una separación estricta (técnica) de los identificadores individuales relacionados en la al menos una entidad proveedora y no los identificadores individuales relacionados en la al menos una entidad de agregación de datos. Esta separación es irreversible en una sola entidad, ya sea en la entidad proveedora de datos o en la entidad de agregación de datos. Por lo tanto, una conclusión para un individuo basada en el identificador utilizado en la al menos una entidad de agregación de datos siempre necesitaría la cooperación de dicha entidad con la al menos una entidad proveedora.

En base a esta separación técnica, el esfuerzo de descubrir la identidad de un usuario basada en el identificador utilizado en al menos una entidad de agregación se correlaciona directamente con el nivel de separación organizacional entre dicha entidad de agregación y la al menos una entidad de suministro de datos. En otras

palabras, cuanto mayor sea el nivel de separación organizacional, mayor será el esfuerzo para descubrir la identidad de un individuo en función del identificador utilizado en al menos una entidad de agregación. Por ejemplo, se garantiza un alto nivel de separación organizativa si el al menos un proveedor de datos y la al menos una entidad de agregación de datos se implementan en diferentes premisas legales, como sería el caso en empresas independientes. Además, puede concebirse que la al menos una entidad proveedora de datos y/o cualquier socio fiable (que se definirá más adelante) y/o la al menos una entidad que agregue datos estén representados por dos procesos independientes ejecutados en el mismo sistema y/o en sistemas físicamente separados, en particular, ubicados en diferentes lugares de instalación.

Dentro de la definición de la ley alemana de protección de datos, los datos se definen como anónimos en el mismo momento en que el esfuerzo por revelar la identidad de un individuo se considera desproporcionado con respecto a los costes, tiempo y mano de obra. Sin un esfuerzo tan desproporcionado, los datos se consideran seudonimizados únicamente (§ 3 Abs. (6) BDSG). Por lo tanto, la posibilidad de influir en el esfuerzo es fundamental para alcanzar el anonimato con respecto a la ley alemana de protección de datos. La invención permite esto mediante la introducción de un método que realiza una separación técnica (estricta) de identificadores relacionados individuales y no individuales que es irreversible dentro de una sola entidad. En base a esto, el esfuerzo real de desanonización es directamente ajustable por el nivel elegido de separación organizacional.

Según una realización preferida, el método permite a la al menos una entidad que suministra datos establecer el horizonte temporal en el que la al menos una entidad de agregación de datos obtiene estabilidad para sus identificadores anónimos. Por lo tanto, la al menos una entidad proveedora de datos siempre tiene el control de cuánto tiempo una o más entidades de agregación de datos pueden vincular conjuntos de datos ocurridos inicialmente para un mismo usuario. De nuevo, esto es muy importante por razones de protección de datos, especialmente cuando la entidad de agregación de datos trabaja con datos dinámicos relacionados con el comportamiento de los usuarios (por ejemplo, datos de ubicación). En tal caso, la duración del intervalo de tiempo en el que los datos podrían combinarse (en base a un identificador anónimo estable) influye directamente en la posibilidad de que al menos una entidad de agregación de datos descubra patrones específicos. Tal patrón podría permitir una conclusión indirecta para el individuo basada en la combinación con información adicional.

En el ejemplo de los datos de ubicación, este podría ser un patrón de movimiento único. La comparación de dicho patrón con información adicional de terceros (por ejemplo, verificación basada en la ubicación en los datos proporcionados por el usuario dentro de las redes sociales) podría ofrecer una posibilidad teórica para revelar indirectamente la identidad de los usuarios. Por supuesto, esta amenaza aumenta con el intervalo de tiempo en el que los eventos de datos dinámicos podrían asignarse a un mismo identificador anónimo. Esto se debe a que, por un lado, el patrón se vuelve más único, por otro lado, aumenta la probabilidad de encontrar información adicional adecuada.

En un aspecto preferido de la invención, la al menos una entidad de suministro de datos es una red de comunicación móvil. Si es así, al menos un identificador que identifica a un suscriptor de la entidad proveedora de datos puede ser un identificador que identifica a un suscriptor y/o un dispositivo móvil en un sistema de comunicación móvil, tal como el IMEI, IMSI o identificadores similares proporcionados por el estándar técnico de comunicación móvil (por ejemplo, dirección MAC) o definidos por un software que utiliza dicha infraestructura técnica (por ejemplo, credenciales de cuenta en aplicaciones móviles o sistemas operativos).

En otro aspecto preferido de la invención, el conjunto de datos incluye al menos un atributo de datos asociado con el al menos un identificador/usuario e incluye al menos uno del grupo de tipo de evento y/o marca de tiempo de un evento y/o ubicación de un evento. En detalle, en un sistema de comunicación móvil, un conjunto de datos puede comprender un IMSI como identificador y otros atributos de datos que caracterizan un evento actualmente activado por el usuario o dispositivo móvil respectivo, por ejemplo, una solicitud de servicio (voz, datos, etc.) o un proceso de búsqueda de posición ejecutado o cualquier otro posible evento activo o pasivo que pueda activarse en una red de comunicación móvil.

Por ejemplo, al menos un atributo del conjunto de datos incluye información de eventos del usuario que caracteriza el tipo de evento capturado actualmente y/o la marca de tiempo de un evento y/o la ubicación actual del usuario en dicha marca de tiempo. En detalle, en un sistema de comunicación móvil, el conjunto de datos puede comprender el IMSI como un identificador y el tipo de servicio, por ejemplo, SMS, voz de conmutador de circuito, datos de conmutación de paquetes, y/o la marca de tiempo y/o la posición exacta o estimada del usuario que exige el tipo de servicio respectivo. La posición, por lo tanto, también podría derivarse indirectamente por la posición de las torres de telefonía celular, etc.

En una realización preferida, una instancia del MAP está configurada para usar una semilla secreta con una larga vida útil en la anonimización de base, así como dos encriptaciones adicionales mediante las cuales la primera encriptación adicional usa un componente aleatorio con una corta vida útil y una clave pública proporcionada como una parte de un par de claves asimétricas por una entidad de agregación de datos (en lo sucesivo denominado "Agregador de datos"). El segundo cifrado adicional utiliza un componente aleatorio que cambia constantemente

para cada llamada de función junto con la clave pública de otra entidad de agregación de datos (en lo sucesivo, "Socio de confianza"). Las dos entidades de agregación de datos que operan posteriormente en la ruta de transmisión de datos por la cual el Socio de confianza es la primera y el Agregador de datos la segunda entidad.

La combinación del conjunto de datos y el identificador personal encriptado adicional 2 veces se transmite al Agregador de datos a través del Socio de confianza. El Socio de confianza deshace el segundo cifrado adicional y, opcionalmente, procesa y/o transforma la información incluida en los datos antes de reenviarlos al Agregador de datos. La descodificación del identificador personal cifrado adicional 2 veces en el Socio de confianza y el identificador personal cifrado adicional 1 vez en el lado del Agregador de datos se realiza descifrando en función de la clave privada respectiva del mecanismo de cifrado asimétrico aplicado. Por lo tanto, es necesario que la última iteración de cifrado adicional se base en la clave pública del Socio de confianza. La iteración anterior debe basarse en la clave pública del Agregador de datos que recibe el identificador del Socio de confianza.

Sin embargo, no es necesario que el Socio de confianza o el Agregador de datos estén informados sobre la cadena de caracteres utilizada concreta como componente aleatorio respectivo. Es suficiente que cada entidad sepa de qué manera la cadena de caracteres se ha combinado con la cadena del identificador cifrado de la iteración anterior. El Agregador de datos llegará al identificador encriptado de la iteración anterior o al identificador personal anónimo base de larga duración de acuerdo con la etapa a de la reivindicación 1 borrando el número conocido de caracteres en las posiciones conocidas en la cadena resultante del descifrado asimétrico.

En una realización preferida, las vidas de la vida útil larga de acuerdo con el cifrado básico y/o las vidas cortas de acuerdo con el cifrado adicional se pueden variar, en particular, dependiendo de los atributos de los conjuntos de datos de eventos. Por ejemplo, la larga vida útil se define como válida por un año. La primera iteración del cifrado adicional puede usar una corta vida útil de aproximadamente 24 horas. Es obvio que otros intervalos de tiempo o combinaciones de diferentes tiempos de vida son concebibles. El método selecciona preferiblemente el tiempo de vida apropiado para la anonimización básica dependiendo del identificador y/o el atributo del conjunto de datos a transmitir. La vida útil define un intervalo de tiempo dentro del identificador cifrado que la entidad de destino (socio de confianza y/o agregador de datos) considera válida.

Como se mencionó anteriomente, la última iteración del cifrado adicional se descifra en al menos un Socio de confianza utilizando su respectiva clave privada y eliminando el componente aleatorio que da como resultado un identificador cifrado con una vida útil corta. Los conjuntos de datos transmitidos al Socio de confianza solo son aplicables dentro de la corta vida útil de su identificador encriptado de corta vida asociado. Esta es una medida de precaución para evitar la derivación de patrones de un cierto número de conjuntos de datos almacenados que crean un historial de eventos detallado.

Por lo tanto, se prefiere especialmente las tendencias extrapoladas de una colección de conjuntos de datos durante un intervalo de tiempo definido. Estas tendencias se denominan índices estadísticos derivados de una serie de conjuntos de datos de eventos y/o atributos de datos relacionados con el mismo identificador cifrado que se ha descifrado en el Socio de confianza y se ha recopilado durante la vida útil de dicho identificador cifrado.

También es posible que al menos un índice estadístico se refiera a un índice de actividad que caracterice la frecuencia y/o probabilidad de un determinado identificador cifrado asociado con un determinado evento, por ejemplo, llamadas salientes/SMS de un suscriptor asociado con dicho identificador cifrado.

El atributo de probabilidad o frecuencia de los índices estadísticos representa una probabilidad relativa y/o frecuencia absoluta y/o una clase de probabilidad/frecuencia particular de uno o algunos de los atributos recopilados dentro de los conjuntos de datos. La probabilidad estadística puede definir la ocurrencia de un evento definido en una ubicación definida. El atributo de frecuencia puede incluir un valor numérico que representa el número de eventos producidos en una ubicación definida. El valor de referencia caracteriza el resultado del índice estadístico, es decir, el atributo con la mayor probabilidad/frecuencia. Finalmente, los índices estadísticos incluyen un atributo que caracteriza la probabilidad de error del valor de referencia.

El Agregador de datos recopila las probabilidades y/o frecuencias recibidas relacionadas con dicho identificador cifrado anónimo base dentro del largo período de vida útil. En ese caso, los conjuntos de datos recopilados dentro de la vida útil corta de acuerdo con el Socio de confianza y transformados en índices estadísticos se pueden almacenar en el Agregador de datos como índices estadísticos a largo plazo que sobrevivirán a la vida útil larga seleccionada del evento respectivo. En particular, el cálculo de índices estadísticos a largo plazo que permiten declaraciones de probabilidad sobre movilidad y actividad basadas en datos de eventos de ubicación es posible sin guardar ningún historial de eventos detallado. Los índices estadísticos pueden mantenerse y usarse dentro de un intervalo de tiempo largo, por ejemplo, más de un año. Los datos del evento de ubicación se pueden descartar después de un breve intervalo de tiempo, por ejemplo, después de 24 h, para minimizar el riesgo de que los patrones de movimiento se puedan derivar de los datos de eventos de ubicación almacenados. Al usar índices en su lugar, no se puede extraer ninguna conclusión sobre la identidad real de un usuario generador de datos de eventos de ubicación.

Preferentemente, puede ser importante reconocer y filtrar las irregularidades para generar declaraciones estadísticas precisas. El filtrado puede realizarse integrando valores de comparación históricos durante un período de tiempo más largo. Los enfoques sugeridos permiten derivar declaraciones de probabilidad sobre movilidad y actividad basadas en datos anónimos reales sin guardar ningún historial de eventos detallado que pueda usarse para derivar patrones que permitan sacar conclusiones sobre la identidad de un individuo.

Otra realización de la invención introduce una técnica de cómo los diferentes tipos de información que se han calculado en base a diferentes tipos de identificadores anónimos pueden combinarse de forma anónima con respecto a los identificadores utilizados.

Para algunas aplicaciones, puede ser necesario combinar diferentes tipos de información generados en diferentes entidades de agregación de datos o dentro de al menos una entidad de agregación mediante el uso de diferentes tipos de identificadores anónimos para cada tipo de información. Tal combinación podría ser requerida dentro de una de dichas entidades de agregación de datos, así como dentro de al menos otra (nueva) entidad de agregación de datos. Al realizar tales combinaciones, es muy importante mantener el anonimato de los datos, dentro de cada una de las entidades involucradas. El método de traducción considera este criterio crítico con respecto a los identificadores utilizados, manteniendo separados los diferentes tipos de identificadores anónimos utilizados internamente en cualquier momento. Preferiblemente, esto se archiva mediante una coincidencia indirecta a través de un componente temporal único, como una cadena de caracteres o un número alfanumérico, por ejemplo.

En el caso de que una misma entidad de agregación de datos posea uno o más tipos de información y quiera combinarlos, el proceso de coincidencia debe realizarse en una entidad de agregación de datos separada.

Todas las entidades de agregación de datos involucradas podrían clasificarse en uno de los siguientes papeles:

- "integrador" que combina diferentes tipos de información basada en al menos una tabla interna utiliza un identificador personal anónimo de base específico
- "extractor" que posee al menos un tipo de información que debe combinarse con al menos otro tipo de información en una entidad integradora
- "híbrido" que posee al menos un tipo de información y desea combinarlo con al menos otro tipo de información propia y/o quiere enriquecer al menos un tipo de información propia con al menos otro tipo de información que obtendrá al menos un extractor

Según esta definición, se realizan las siguientes etapas:

a) Construir una "gran tabla de coincidencia" que contenga una tupla de datos para cada identificador personal de al menos una entidad proveedora de datos. Cada tupla de datos consiste en:

- un identificador personal encriptado adicional 1 vez y anónimo de base para cada tipo diferente de información que debe extraerse de al menos un extractor y/o entidad de agregación de datos híbrida
- un identificador personal encriptado adicional 1 vez y anónimo de base para cada tabla interna dentro de al menos un integrador y/o entidad de agregación de datos híbrida
- un componente temporal único

b) Basado en la gran tabla de correspondencias: Generar una "pequeña tabla de coincidencia" para cada tipo de identificador personal encriptado adicional 1 vez y anónimo de base que se construyó en la etapa a). Por lo tanto, cada tabla pequeña incluye una tupla de datos que contiene la base anónima y un identificador personal encriptado adicional 1 vez junto con el componente temporal único

c) Distribución de al menos una tabla pequeña adecuada a todas las entidades de agregación de datos involucradas. En el caso de una entidad de agregación de datos híbrida, la pequeña tabla de coincidencia que debe usarse para la integración se distribuye a otra, entidad de agregación de datos separada.

d) Espera de la entidad de agregación de datos separada, cada entidad de agregación de datos obtiene al menos una tabla pequeña adecuada y deshace el primer cifrado adicional y, por lo tanto, revela los identificadores personales anónimos de base.

e) Dependiendo de su clasificación, las diferentes entidades de agregación de datos realizan las siguientes acciones:

- Integrador: Guarda la pequeña tabla de coincidencia interna para usarla para la integración posterior. Es decir, guarda cada combinación de identificador personal anónimo base y componente temporal único asociado. Espera los conjuntos de datos que se enviarían con un componente temporal único como identificador. Al recibir dichos conjuntos de datos, busca el identificador personal anónimo interno de la base en la pequeña tabla de coincidencia y procesa los conjuntos de datos basados en este identificador.
- Extractor: Guarda la pequeña tabla de coincidencia interna y la usa para extracción y reenvío. Es decir, busca la información necesaria basada en la base interna del identificador personal anónimo. Reemplaza el

identificador personal anónimo interno de la base interna con el componente temporal único asociado dentro de los conjuntos de datos y reenvía dichos conjuntos de datos a al menos una entidad de agregación de datos clasificada como integrador o híbrido.

- Híbrido: Guarda la pequeña tabla de coincidencia y lleva la información análoga al extractor según el identificador personal anónimo interno de la base. Que reordena la tabla de información transportada basada en el componente temporal único. Reenvía esta tabla reordenada a la entidad de agregación de datos separada que posee la pequeña tabla de coincidencia que podría usarse para la integración. Dicha entidad neutral de agregación de datos reemplaza todos los componentes temporales únicos por la base anónima asociada y los identificadores personales cifrados adicionales. En base a estos identificadores, la tabla se reordena nuevamente y se envía de nuevo a la entidad de agregación de datos híbrida. Aquí se deshace el primer cifrado adicional. En función del identificador personal anonimizado de la base interna resultante, la entidad de agregación de datos híbrida puede procesar los datos.

Dependiendo de los requisitos específicos (por ejemplo, restricciones de por vida del identificador personal anónimo de base dentro de la entidad integradora o de agregación de datos híbrida), podría ser necesario repetir las etapas a-e mencionadas anteriormente de forma regular. El método de integración introducido mediante el uso de una entidad de agregación de datos separada (neutral) también podría usarse opcionalmente para integrar información de diferentes entidades de agregación de datos clasificadas como extractoras en lugar de híbridas. Esto podría ser útil especialmente en los casos en que la combinación de información diferente necesita ser verificada nuevamente con respecto a los riesgos de desanonización potencialmente dados a través del propio tipo de información. En estos casos, la entidad de agregación de datos separada también podría realizar las transformaciones necesarias para garantizar el anonimato antes de enviar la información combinada a la al menos una entidad de agregación de datos clasificada como integrador.

La generación de los diferentes tipos de identificador personal encriptado adicional 1 vez y anónimo de base dentro de la etapa a se realiza ejecutando instancias MAP paralelas. La configuración de cada instancia incluye una anonimización base y un cifrado adicional. Para las configuraciones que conciernen a entidades de agregación de datos clasificadas como extractor o híbrido, cada anonimización de base usa la misma semilla secreta que también se usa para la anonimización de los conjuntos de datos que ya están almacenados en dichas entidades. Para las entidades de agregación de datos clasificadas como integradoras, esto no es necesario. Es suficiente tener una semilla secreta con cierta vida útil. El cifrado adicional no depende de la clasificación de la entidad de agregación de datos. Para cada configuración, solo es importante elegir la clave privada adecuada de la entidad de agregación de datos respectiva y cambiar el componente aleatorio para cada nueva tabla de coincidencia grande (en cada llamada funcional).

A modo de ejemplo para el proceso ALIP para derivar índices estadísticos, los índices estadísticos derivados a largo plazo deben integrarse a una determinada base de datos en la entidad de agregación de datos donde ya almacena otro tipo de información con un tipo diferente de identificador personal anónimo de base del que es utilizado para la derivación de los índices estadísticos a largo plazo. El agregador de datos en este caso podría clasificarse como híbrido. Dicha base de datos puede administrar los datos recopilados en base al identificador cifrado adicional 1 vez y anónimo de base utilizado para los conjuntos de datos ya almacenados en el agregador de datos. En una realización preferida, el proceso de mapeo se puede basar en una tabla de traducción que incluye tuplas, cada una de las cuales consta de una base anónima anonimizada de larga vida útil y un identificador cifrado adicional de 1 vez utilizado para los cálculos a largo plazo de acuerdo con la etapa a, un identificador encriptado adicional 1 vez y anónimo de base, en particular, con un identificador de corta duración y un número único. La tabla puede generarse en paralelo al proceso general MAP multinivel.

La invención está relacionada además con un sistema de comunicación para realizar el método según la invención o según una realización preferida de la invención. Es obvio que el sistema de comunicación se caracteriza por las propiedades y ventajas según el método de la invención. Por lo tanto, una descripción repetitiva se considera innecesaria.

Las ventajas y propiedades adicionales de la presente invención deberían describirse sobre la base de tres realizaciones preferidas mostradas en las figuras. Las figuras muestran

Figura 1a: una descripción esquemática sobre el proceso básico de anonimización multinivel (MAP),

Figura 1b: una descripción esquemática de la implementación de iteración múltiple MAP de acuerdo con la invención,

Figura 2a, 2b: dos versiones de una descripción arquitectónica sobre un sistema para la implementación MAP

Figura 3a, 3b: dos versiones de una descripción arquitectónica para el proceso de indexación a largo plazo anónimo según la invención,

Figura 4: una descripción arquitectónica sobre un sistema para el proceso de indexación anónimo a largo plazo de acuerdo con una realización preferida de la presente invención,

Figura 5: una realización de ejemplo para calcular índices estadísticos de acuerdo con la invención y

Figura 6a, 6b: dos versiones de otra realización preferida de la presente invención.

La figura 1a ilustra la idea fundamental de la presente invención. La idea básica de la presente invención se refiere a un procedimiento de anonimización de datos para permitir el uso de datos de ubicación masiva para aplicaciones de big data con pleno respeto de los estándares europeos de protección de datos. Los datos de ubicación masiva serán recopilados por proveedores de redes de comunicaciones móviles o inalámbricas, así como proveedores que recopilan información que se basa en otras tecnologías de ubicación como GPS, Galileo, GLONASS, Compass, redes de sensores, etc., que también pueden poseer información personal detallada y verificada sobre sus usuarios. Sin embargo, aplicaciones sin información personal concreta también son posibles. Un ejemplo es la aplicación de direcciones MAC dentro de una red WIFI.

Además, los proveedores de redes móviles pueden extraer datos de eventos de ubicación de sus usuarios. Toda la información se combina en conjuntos de datos anónimos que pueden ser de interés para diferentes aplicaciones ofrecidas por empresas de terceros.

Por ejemplo, los proveedores de redes móviles pueden vender o proporcionar los datos anónimos y agregados a los consejos locales, empresas de transporte público, empresas de infraestructuras como proveedores de transporte público o proveedores de electricidad, minoristas, los principales organizadores de eventos o seguridad pública que utilizan dicha información para mejorar sus procesos de toma de decisiones.

Los conjuntos de datos proporcionados también pueden analizarse para determinar cuántas personas visitan un área por tiempo, género y edad. Las organizaciones podrán analizar los movimientos de las multitudes en cualquier lugar por hora, día, semana o mes, y hacer comparaciones comparables por área, así como también comprender los patrones de captación.

Una aplicación particular de los datos podría ser la implementación de ciudades inteligentes. El análisis de datos antes mencionado podría usarse para analizar el volumen de tráfico en ciertos distritos de la ciudad. Por lo tanto, el ayuntamiento puede optimizar la ingeniería vial sobre la base de dicha información. Por supuesto, dicha información es útil para cada planificación de construcción teniendo en cuenta la cantidad de usuarios/visitantes potenciales.

Sin embargo, es obligatorio cuidar la privacidad de cada usuario y la información personal del usuario. Por lo tanto, el objetivo de la presente invención es definir un proceso que permita la anonimización real en lugar de solo la seudoanonimización. Al dividir el proceso en varias etapas del proceso que se ejecutaron dentro de diferentes premisas legales de entidades independientes, se evita la posibilidad de generar una tabla de asignación entre identificadores anónimos y no anónimos.

Como se puede ver en la figura 1a, un proveedor de datos como primera entidad y al que se hace referencia como DS está conectado comunicativamente a través de una red pública o privada virtual a un agregador de datos como una segunda entidad y referenciado como DA. La entidad proveedora de datos DS puede ser cualquier proveedor de datos de movimiento y/o personales. Esto incluye, por ejemplo, proveedores de redes móviles o inalámbricas, así como proveedores que recopilan información que se basa en otras tecnologías de ubicación como GPS, Galileo, GLONASS, Compass, redes de sensores, La siguiente argumentación se basará en el caso ejemplar de un sistema de red móvil que proporciona los conjuntos de datos antes mencionados que contienen datos personales, así como datos de eventos de ubicación sobre sus usuarios.

Como se describe, DS y DA están físicamente separados y asignados a sistemas independientes. Por lo general, DS y DA cumplen diferentes tareas que pueden asignarse a diferentes usuarios que tienen diferentes perfiles de autoridad en un sistema común o que se realizan dentro de diferentes zonas de seguridad de un sistema común.

Las realizaciones ejemplares de acuerdo con las figuras se basan en un sistema de red móvil como un proveedor de datos DS que proporciona los conjuntos de datos antes mencionados que contienen datos personales, así como datos de eventos de ubicación sobre sus suscriptores. Cada suscriptor individual de la red global de DS se identifica mediante un PID de identificador personal que podría ser un identificador conocido como IMSI de un suscriptor. Para tener una verdadera anonimización de acuerdo con las normas europeas de protección de datos, es necesario, entre otras cosas, separar el PID inicial y su parte contraria, el O-PID (identificador personal ofuscado). En este contexto, el esfuerzo de reunir estos dos identificadores tiene que ser irrazonablemente alto en comparación con el rendimiento que se podría obtener con tal acción.

Este requisito se cumple si la separación se realiza físicamente dentro de las premisas de dos entidades legalmente independientes mediante las cuales una entidad solo conoce el PID y la otra solo el O-PID. Sin embargo, la separación de DS y DA también se puede realizar mediante una de las posibilidades alternativas propuestas anteriormente. En cualquier caso, es necesario encriptar y transmitir el O-PID a una entidad denominada como el agregador de datos DA. Ese identificador personal se combina con un conjunto de datos con atributos de datos adicionales que describen un determinado evento de ubicación. Por ejemplo, estos atributos de datos de eventos caracterizan una acción de un suscriptor en un lugar determinado. Los posibles atributos son el tipo de evento, la ubicación del evento y la marca de tiempo. En este ejemplo, el cifrado solo se realiza para el identificador personal, pero también se puede hacer para otros datos.

La ofuscación de los datos sensibles debe realizarse mediante un proceso de anonimización de múltiples niveles (MAP) realizado en el DS para proteger la privacidad del usuario. En la figura 1a se ofrece una descripción general de las etapas necesarias.

En una primera etapa 1, se realiza una anonimización de base aplicando un algoritmo de hash con clave no reversible para el PID, donde la semilla/clave (clave DS) solo es conocida por el proveedor de datos DS. Dicho algoritmo de hash debe ser una función hash criptográfica fuerte. Diferentes claves DS pueden estar disponibles en el lado DS con diferentes vidas como ST/LT (corto tiempo/largo tiempo), por ejemplo. La salida de la primera etapa del método es un único PID ofuscado al que se hace referencia como O-PID. La vida útil de dicho O-PID depende del intervalo en que se cambia la clave DS. Es decir, si la clave DS es, por ejemplo, constante durante 24 horas, el DA obtendrá un identificador ofuscado estático durante exactamente ese período de tiempo. El tipo de clave DS utilizada para ofuscar el PID depende del conjunto de datos/atributos de datos que se transmiten al DA o socio de confianza TP en combinación con el PID ofuscado. Por ejemplo, se utiliza una clave a corto plazo (clave ST) para ofuscar el PID que se envía en combinación con los datos de la clase del cliente en el que se utiliza una clave LT para el proceso MAP cuando se ofusca el PID para transmitir conjuntos de datos de eventos de ubicación.

En una segunda etapa 2, un componente aleatorio RC o cadena, por ejemplo, preferiblemente se agrega un número aleatorio de varios dígitos al O-PID de salida del procedimiento de anonimización de base de acuerdo con la primera etapa 1. Se observa que el componente aleatorio podría insertarse en cualquier posición del O-PID en el que el DA debe conocer la posición. Se observa además que cualquier otra cadena de caracteres generada aleatoriamente y cualquier otro procedimiento de combinación de las dos cadenas podría ser apropiado. La longitud del intervalo del componente aleatorio utilizado también podría ser variable, pero tiene que ser conocido por el DA. La salida de la segunda etapa se marca como O-PID + RC.

En la última etapa 3, se ejecuta un cifrado de segundo nivel sobre la base de un mecanismo de cifrado asimétrico que utiliza la clave pública DA-Pub-Clave de la segunda entidad DA. El cifrado asimétrico se aplica al resultado de la etapa 2 O-PID + RC resuiting en un resultado que está marcado como OO-PID. En consecuencia, el PID está doblemente ofuscado para proteger la privacidad del usuario.

La vida útil del identificador encriptado doble OO-PID solo depende del intervalo en el que se cambia el componente aleatorio utilizado en la etapa 2. Esto significa que el OO-PID es constante siempre que el RC sea constante, lo cual es importante para los cálculos realizados en el OO-PID por una determinada entidad (por ejemplo, un Socio de confianza que construye índices estadísticos como se describirá más adelante). Por el contrario, el valor real del componente aleatorio no es necesario para la decodificación del OO-PID en el DA.

Las etapas 1 a 3 se implementan en una unidad atómica de trabajo. Es imposible que el proveedor de datos DS lea o escriba cualquier información generada entre las etapas individuales. La combinación de las etapas 2, 3 se denomina encriptación adicional "AE".

En el lado del agregador de datos, el descifrado DA se ejecuta en el cifrado adicional de acuerdo con la etapa 3 utilizando su clave privada DA-Priv-Clave para descifrar el identificador cifrado recibido OO-PID. El resultado O-PID + RC se procesará aún más borrando el número conocido de dígitos al final de la cadena que representa el componente aleatorio. El resultado resultante es el O-PID. La duración de este identificador cifrado único O-PID en el lado del agregador de datos DA se define por la longitud del intervalo de la clave DS generada. Si ha transcurrido el intervalo de la clave DS, se generará una nueva clave DS y, por lo tanto, un nuevo O-PID en el DS.

El PID original solo es visible en el lado del proveedor de datos DS ya que el lado del agregador de datos DA solo conoce el identificador cifrado único O-PID. Por lo tanto, es imposible construir un catálogo (una tabla que asigna cada PID no anónimo a su parte contraria anónima, el O-PID) dentro de las instalaciones de una sola parte.

El resultado del proceso de anonimización de niveles múltiples (MAP) explicado anteriormente es que el proveedor de datos DS no puede encontrar el PID ofuscado. Lo mismo se aplica al agregador de datos DA que no puede encontrar el PID original sobre la base del PID ofuscado suministrado.

La figura 1b muestra una versión extendida MAP de acuerdo con la figura 1a llamada implementación de iteración múltiple MAP. La presente invención se refiere a dicha implementación MAP de iteración múltiple. Dicha versión extendida MAP se utiliza principalmente para construir índices estadísticos a largo plazo de conjuntos de datos recopilados.

Al contrario de la implementación básica MAP según la figura 1a, la extensión inventiva MAP realiza el cifrado adicional (AE) de forma iterativa con al menos dos iteraciones. Sin embargo, es posible un número arbitrario n de iteraciones. El número de iteraciones depende del número de entidades de agregación de datos/TP de socios de confianza que están disponibles para descifrar las iteraciones respectivas del cifrado adicional. Cada clave pública de cada iteración está dedicada a una entidad de agregación de datos definida/TP de socio de confianza que comprende la clave privada asignada.

El intervalo de constancia de la semilla utilizada (clave DS) en el DS es igual a la vida útil del O-PID y, en consecuencia, determina la vida útil máxima de todos los identificadores ofuscados n veces en la parte de transmisión en el medio. La vida útil concreta de cada identificador ofuscado n veces en los diversos TP en la ruta de transmisión es igual a la vida útil de cada componente aleatorio utilizado en la iteración AE adecuada.

La semilla utilizada (clave DS) de la anonimización básica tiene la vida útil más larga, llamado como LT de por vida, por ejemplo, un año. La vida útil del componente aleatorio RC de las iteraciones se denomina ST de vida corta. La vida útil de los componentes aleatorios disminuye con un número creciente de iteraciones. Por ejemplo, el componente aleatorio de la primera iteración de AE tiene una vida útil corta de 24 horas y el componente aleatorio de la última iteración de AE se cambia para cada conjunto de datos para evitar la creación de tablas de arco iris en DS.

En la figura 2a se ofrece una descripción general de la arquitectura de un sistema para obtener una explicación básica de la implementación del proceso MAP básico. Describe una solución técnica para el anonimato de diferentes conjuntos de datos entregados por un único proveedor de datos DS.

La anonimización y la transmisión de estos conjuntos de datos a un único agregador de datos DA se procesan mediante procesos completamente separados que se ejecutan en el proveedor de datos DS. Los diferentes conjuntos de datos se pueden combinar en función de sus identificadores iguales en el agregador de datos DA. Esta realización es apropiada, si el proveedor de datos DS está sujeto a restricciones legales o de otro tipo con respecto a la combinación de conjuntos de datos específicos en forma no anónima. Con respecto a los estándares europeos de protección de datos, esto se aplica a la combinación de datos de eventos de ubicación con datos personales de clientes, por ejemplo.

Por lo tanto, todo el proceso se subdivide en dos procesos independientes de anonimización de niveles múltiples (MAP), donde los PID de los identificadores personales (como elementos unificadores entre los conjuntos de datos) se anonimizan y transmiten por separado al agregador de datos DA, junto con sus respectivos conjuntos de datos. De ese modo, el primer proceso MAP es responsable de transmitir los llamados atributos de evento de ubicación, incluido el tipo de evento, una marca de tiempo y la ubicación del usuario.

El segundo proceso es responsable de transmitir los atributos que clasifican a los usuarios/identificadores en diferentes grupos de clases de usuarios, por ejemplo, género o grupos de edad.

Como se puede ver en la figura 2a, ambos procesos ejecutan el cifrado de primer nivel sobre la base de una clave DS idéntica. La vida útil de esta clave se ha establecido a modo de ejemplo en 24 horas. Para distinguir este tipo de clave DS de otras claves requeridas para otras aplicaciones, además de tener un nombre general para una mayor argumentación, esta clave DS se define como clave a corto plazo o clave ST (en referencia a su vida útil relativamente corta). En una segunda etapa, se agregan números aleatorios individuales RN a los resultados de las primeras etapas. Los números aleatorios se cambian para cada procedimiento de encriptación de cualquier conjunto de datos nuevo. Por lo tanto, los números aleatorios RN diferirán dentro del primer MAP para cada evento de ubicación individual y entre los propios dos procesos de MA.

En la tercera etapa, se ejecuta un cifrado de segundo nivel en el O-PID + RN mediante el uso de DA-Pub-Clave A. Esta clave se genera a partir del agregador de datos DA como parte de un par de claves asimétricas A. El DA proporciona el DA-Pub-Clave A al DS con el fin de realizar el cifrado de segundo nivel. Más tarde, esto podría descifrarse mediante el uso de la DA-Priv-Clave A, que solo DA conoce. En este contexto, la "A" mayúscula se entiende como un contador. Al tener más de un par de claves asimétricas, se pueden distinguir por la mayúscula (par de claves A, B, C,...). En la realización dada, ambos MAP que realizan el cifrado de segundo nivel emplean la misma clave pública DA-Pub-Clave A. Los resultados de las terceras etapas son PID de doble cifrado. Dado que estos PID dobles encriptados se basan en O-PID generados mediante el uso de una clave ST a corto plazo como se define en la etapa uno, el identificador doble encriptado OO-PID debe llamarse ST-OO-PID. Se transmiten diferentes ST-OO-PID en combinación con sus respectivos atributos a través de diferentes rutas de transmisión por cada MAP.

Debido al hecho de que el número aleatorio es diferente dentro de cada procedimiento de cifrado de segundo nivel, los ST-OO-PID cifrados resultantes son únicos. Es decir, un ST-OO-PID específico siempre pertenece a un conjunto de datos específico, por lo tanto, la vida útil de un ST-OO-PID se limita a un solo evento (que en este contexto incluye la generación de datos de eventos de ubicación, así como datos de clase de cliente). En consecuencia, no es posible combinar ninguna clase de clientes (género, grupo de edad) y datos del evento (tipo de evento, marca de tiempo, ubicación) ni varios conjuntos de datos de eventos de ubicación dentro de las instalaciones de DS del proveedor de datos o en la ruta de transmisión al agregador de datos DA.

La combinación de datos mencionada anteriormente solo puede ser realizada por el agregador de datos DA. Por lo tanto, el agregador de datos primero descifra el ST-OO-PID con la clave privada respectiva (DA-Priv-Clave A) correspondiente a la DA-Pub-Clave A proporcionada al proveedor de datos DS. La cadena de salida incluye el O-PID además del número aleatorio RN. Como DA sabe el número de dígitos bloqueados para el RN al final de la cadena, el agregador de datos simplemente elimina el número de dígitos para obtener el O-PID. Basado en este elemento unificador, el DA puede combinar los datos durante un período de tiempo correspondiente a la vida útil de la clave ST utilizada por el DS para generar el O-PID. Por lo tanto, el DA podría combinar varios eventos de ubicación junto con datos de clase de cliente para un O-PID estático durante un período de tiempo de 24 horas.

Una versión ligeramente diferente de la realización descrita anteriormente se muestra en la figura 2b. Este modelo de proceso difiere del que se muestra en la figura 2b en el hecho de que los dos procesos de MA son realizados por dos agregadores de datos diferentes DS1 y DS2. En el caso de un operador de red móvil, esto podría ser necesario, por ejemplo, si partes de la infraestructura (como la red móvil, por ejemplo) se subcontratan a otras empresas.

Para generar O-PID iguales, ambos agregadores de datos deben usar la misma clave DS y la misma técnica con respecto a la adición de números aleatorios a este resultado. El agregador de datos podría proporcionar diferentes claves públicas de diferentes conjuntos de claves. Para que sea simple, en este ejemplo, ambos DS funcionan con las mismas claves públicas, a saber, DA-Pub-Clave A.

En la figura 3a se representa una realización que describe el proceso de indexación anónima a largo plazo (ALIP). Esta realización debería permitir el cálculo de índices estadísticos a largo plazo que permitan declaraciones de probabilidad sobre movilidad y actividad basadas en datos de eventos de ubicación, sin guardar ningún historial de eventos detallado. El desafío general se plantea por el hecho de que la precisión de tales declaraciones estadísticas depende directamente de la cantidad de datos disponibles para la derivación de las declaraciones. Si, por ejemplo, un operador de red móvil desea calcular el código postal donde probablemente vive un identificador personal anónimo (O-PID), una probabilidad respectiva es estadísticamente derivable contando el número de eventos asignados a este O-PID en diferentes ubicaciones entre las 7 pm y las 7 am en días hábiles. En este contexto, es muy importante reconocer y filtrar las irregularidades (vacaciones, viajes de negocios, etc.) para generar declaraciones precisas. El filtrado puede realizarse integrando valores de comparación históricos durante un período de tiempo más largo. Este argumento podría extenderse a muchas otras aplicaciones de movilidad y actividad (por ejemplo, código postal de trabajo, comportamiento de uso promedio de varios servicios,...). Por lo general, la precisión de las declaraciones estadísticas calculadas depende de la cantidad de datos de eventos de ubicación recopilados dentro de un intervalo de tiempo dado. Cuantos más datos se usen para el cálculo, mejor será la precisión. Además, la precisión también aumentará si se extiende el intervalo de tiempo definido para la recopilación de datos.

Sin embargo, un número creciente de datos de eventos de ubicación recopilados y almacenados aumenta proporcionalmente la probabilidad de una identificación exitosa de patrones de movimiento únicos a partir de los datos almacenados. Estos patrones permiten llegar a una conclusión sobre la identidad real de una persona o al revés. Según las normas europeas de protección de datos, es obligatorio minimizar el riesgo de identificación de patrones. A la luz de lo anterior, el proceso ALIP de construir índices a largo plazo sin guardar el historial detallado de eventos es otro componente central de esta invención.

Las siguientes variaciones de procesos de indexación a largo plazo anónimos (ALIP, mostrados en la figura 3a y la figura 3b) describen soluciones técnicas para tal problema basadas en la técnica general del proceso de anonimización multinivel (MAP) como se presentó anteriormente. La idea básica prevé dividir diferentes partes del proceso, así como la extensión de los datos visibles entre las diferentes entidades participantes. Además de los proveedores de datos conocidos y los agregadores de datos, se presenta un TP de socio de confianza como una nueva instancia entre DS y DA. El socio de confianza TP crea índices estadísticos basados en conjuntos de datos de eventos de ubicación reales durante un período a corto plazo y reenvía solo estos índices (pero ningún dato de eventos de ubicación reales) a un DA. El DA respectivo asocia estos índices con valores de comparación históricos durante un período a largo plazo.

La siguiente explicación describe una posible variación del proceso visualizada en la figura 3a. En esta variación, un único DA es responsable del cálculo de los índices a largo plazo, así como de la agregación de datos. Dado que la agregación de datos se basa en un identificador personal ofuscado a corto plazo (ST-O-PID) y los índices a largo plazo se identifican con un identificador personal ofuscado a largo plazo (LT-O-PID), el DA no puede combinar

ambos conjuntos de datos directamente dentro de sus instalaciones. Una combinación siempre requiere una traducción a través del socio de confianza. No obstante, en algunos casos, puede desearse distribuir el cálculo de índices a largo plazo y las tareas de agregación de datos en dos agregadores de datos independientes (DA 1 y DA 2). Dicha versión modificada respectiva de la realización de acuerdo con la figura 3a se visualiza en la figura 3b y funciona de forma análoga a la versión descrita anteriormente y a continuación.

En el proveedor de datos, los datos del evento de ubicación del lado DS consisten en un identificador personal PID y el tipo de evento de atributos de datos, marca de tiempo, la ubicación se aplica a dos tipos diferentes de procesos de anonimización basados en la lógica MAP como se describió anteriormente. El primer MAP funciona con la misma DS-Clave a corto plazo que también se usa para cifrar datos de eventos de ubicación como se describe en la primera realización de esta invención. Por lo tanto, la clave también cambia a modo de ejemplo cada 24 horas (ST-Clave). El ST-O-PID saliente se agrega mediante un número aleatorio cambiante RN que cambia cada vez que este primer MAP se realiza con un número conocido de dígitos y luego se encripta por segunda vez mediante el uso de la clave pública proporcionada por el DA (DA-Pub- Clave A). El primer MAP se activa una vez cada vez que el segundo MAP genera un nuevo identificador personal doblemente ofuscado a largo plazo LT-OO-PID (véase más adelante). El resultado del primer proceso es un único ST-OO-PID que el DA puede descifrar para obtener un ST-O-PID como un identificador único para la combinación con otras fuentes de datos en un momento posterior. La vida útil de este ST-O-PID en el DA será de 24 horas en el ejemplo dado.

El segundo MAP opera con una DS-Clave a largo plazo (clave LT). De acuerdo con la realización ilustrada de la figura 3a, la clave LT se cambia una vez al año. El cifrado de primer nivel del PID con la clave LT resulta en un LT-O-PID. En la segunda etapa, se agrega un número aleatorio (con un número conocido de dígitos) al LT-O-PID. En el ejemplo dado, el RN cambia en los mismos períodos de tiempo que la clave ST. Por lo tanto, la RN es constante durante 24 horas. Dentro del cifrado de segundo nivel, la combinación de LT-O-PID + RN se ofusca al usar otra clave pública del DA (DA-Pub-Clave B). Por lo tanto, el LT-OO-PID resultante es constante para todos los eventos transmitidos al socio de confianza TP dentro de las 24 horas. Después de que el agregador de datos decodifica el LT-OO-PID, un DA LT-O-PID constante durante un año está disponible en el DA.

Como se mencionó anteriormente, el primer MAP para generar el ST-OO-PID se activa solo uno cada la vez, cuando el segundo MAP genera un nuevo LT-OO-PID (en el ejemplo descrito en el presente documento, esto es cada 24 horas cuando cambia el RN del segundo MAP). En este momento, el ST-OO-PID (sin más atributos) y el respectivo LT-OO-PID (fuera de un conjunto completo de datos de ubicación que incluye todos los atributos) se transmiten al socio de confianza TP. Al mismo tiempo, el LT-OO-PID (sin más atributos) también se reenvía al DA. En este caso, se utiliza para devolver los valores de los índices a largo plazo, ya que se han calculado hasta el final de la última vida útil de LT-OO-PID (esto se describe en detalle más adelante).

El TP guarda la asignación entre ST-OO-PID y LT-OO-PID en una tabla de traducción ST/LT. Luego, el socio de confianza crea índices estadísticos basados en los conjuntos de datos de eventos de ubicación que entrega el DS a través del segundo MAP dentro de la vida útil del LT-OO-PID (aquí: 24 horas). Como estos índices se calculan dentro de un intervalo a corto plazo, se denominan índices a corto plazo. En las figuras 3a y 3b muestra un único índice (ST-Índice 1) como representante de algunos índices potenciales que podrían calcularse en este punto. El término índice estadístico en esta descripción debe incluir valores de frecuencia simples (por ejemplo, número de envíos de SMS en el corto plazo), así como valores de probabilidad (por ejemplo, el 80 % de los eventos entre las 7 p.m. y las 7 a.m. ocurrieron en una región geográfica que tiene el código postal 80639). Los eventos de ubicación originales se descartan después de haber sido procesados para calcular varios índices estadísticos. Antes del final de la vida útil de LT-OO-PID, un nuevo conjunto de datos con el LT-OO-PID, así como todos los índices ST construidos para este identificador dentro del período de corto plazo en el TP, se envían al agregador de datos DA.

El DA descifra el LT-OO-PID utilizando la clave privada apropiada (DA-Priv-Clave B). El próximo identificador a largo plazo LT-O-PID permite al DA combinar los nuevos índices a corto plazo recibidos del TP con valores históricos para los mismos índices. Por lo tanto, el DA primero guarda los nuevos valores en su base de datos (historial de índices a corto plazo) y luego calcula nuevos índices a largo plazo (por ejemplo, LT-Índice 1) basándose en todos los valores (nuevos e históricos) en la base de datos. Estos índices a largo plazo se combinan en un nuevo conjunto de datos y se guardan con el LT-O-PID como un identificador hasta el comienzo de un nuevo período de LT-OO-PID.

Al comienzo del nuevo período, se envía un nuevo LT-OO-PID desde el DS al DA como se describió anteriormente. Este LT-OO-PID se descifra mediante el uso de la DA-Priv-Clave B para encontrar el LT-O-PID apropiado. Posteriormente, los índices a largo plazo que se han almacenado para el LT-O-PID en la base de datos del historial de índices a corto plazo hasta ese momento se combinan con el LT-OO-PID recién recibido. Si no hay valores históricos disponibles en la base de datos, se aplica un valor ficticio "n.a." a todos los atributos del conjunto de datos recién generado antes de retransmitir dicho conjunto de datos al TP.

A medida que el TP obtiene este nuevo conjunto de datos con LT-OO-PID y los índices a largo plazo calculados por el DA, busca el ST-OO-PID apropiado que guardó para el LT-OO-PID en la tabla de traducción ST/LT anteriormente. Después de cambiar el LT-OO-PID con el ST-OO-PID, se reenvía el conjunto de datos al DA nuevamente. Para

asegurar, que el DA solo podría usar la tabla de traducción dentro de un período de vida LT-OO-PID, la asignación ST/LT utilizada se puede eliminar después.

Nuevamente en el DA, el ST-OO-PID se descifra con la DA-Priv-Clave A. El ST-O-PID saliente como un identificador único finalmente permite la combinación de los índices a largo plazo con otras fuentes de datos como el evento de ubicación y datos de clase de cliente. Al transmitir los índices a largo plazo que se han calculado hasta el final del último período de vida de LT-OO-PID al comienzo del nuevo período de vida de LT-OO-PID, el ALIP garantiza la disponibilidad de índices a largo plazo (o valores ficticios) en cada momento de la vida útil de ST-O-PID.

La solución técnica dada brinda la posibilidad de calcular índices estadísticos anónimos a largo plazo sin guardar datos de eventos de ubicación durante un período de tiempo más largo. Como ya se mencionó anteriormente, la segunda versión de esta solución (figura 3b) funciona de la misma manera, pero difunde el cálculo de índices a largo plazo y la combinación de varias fuentes de datos a dos agregadores de datos diferentes.

La figura 4 muestra una posible realización de la presente invención. Describe el proceso ALIP antes mencionado de una manera ligeramente diferente.

La figura 4 muestra una solución técnica para el anonimato de diferentes conjuntos de datos entregados por un único proveedor de datos DS. La anonimización y la transmisión de estos conjuntos de datos a un único agregador de datos DA se procesan mediante procesos completamente separados que se ejecutan en el proveedor de datos DS. Los diferentes tipos de conjuntos de datos se pueden combinar sobre la base de los identificadores iguales O-PID en el agregador de datos DA.

Todo el proceso se subdivide en dos procesos independientes de anonimización de niveles múltiples (MAP) 10, 20 donde los identificadores personales PID (como elementos únicos entre los conjuntos de datos) se anonimizan por separado y se transmiten al agregador de datos junto con sus respectivos conjuntos de datos. De ese modo, el primer proceso MAP 10 es responsable de transmitir los llamados datos de clase de cliente que incluyen atributos que clasifican a los suscriptores en diferentes grupos de clases de suscriptores, por ejemplo, género o grupos de edad. El identificador personal es anonimizado por el MAP básico de acuerdo con la figura 1a utilizando una semilla (clave DS) con una vida útil corta. El ST-OO-PID resultante se transmite junto con el conjunto de datos CCD al agregador de datos DA. El cifrado adicional se invierte en DA en el bloque 11, el ST-O-PID resultante se almacena con el conjunto de datos asociado en el bloque 12.

El segundo proceso MAP 20 es responsable de transmitir los llamados conjuntos de datos de eventos de ubicación con atributos que incluyen el tipo de evento, una marca de tiempo cuando ocurrió el evento y la ubicación del suscriptor que define la ubicación donde ocurrió el evento. El conjunto de datos de ubicación incluye obligatoriamente al menos una marca de tiempo, otros atributos como tipo de evento y ubicación son opcionales. Similar a la realización según las figuras 3a, 3b, el PID para cada conjunto de datos de eventos de ubicación también se anonimiza dos veces por dos operaciones MAP en el proceso MAP 20.

La primera operación MAP del proceso MAP 20 es exactamente la misma de acuerdo con las figuras 3a, 3b. Por lo tanto, la clave también cambia a modo de ejemplo cada 24 horas (ST-Clave). El ST-O-PID saliente se agrega mediante un componente aleatorio cambiante RC que cambia cada vez que este primer MAP se realiza con un número conocido de dígitos y luego se encripta por segunda vez mediante el uso de la clave pública proporcionada por el DA (DA-Pub- Clave A). El primer MAP se activa una vez cada vez que la segunda operación MAP genera un nuevo identificador personal ofuscado a largo plazo LT-OOO-PID (véase más adelante). El resultado del primer proceso es un único ST-OO-PID que el DA puede descifrar para obtener un ST-O-PID como un identificador único para la combinación con otras fuentes de datos en un momento posterior. La vida útil de este ST-O-PID en el DA será de 24 horas en el ejemplo dado. Dicho ST-OO-PID también se envía a través del TP al DA. La etapa intermedia divulgada ("Filtrado de datos de eventos de ubicación") en el TP no es relevante para la comprensión de la presente invención.

La segunda operación MAP difiere ligeramente de la de las figuras 3a, 3b. Básicamente, la segunda operación MAP se refiere a la implementación representada en la figura 2. La anonimización básica del PID con la clave LT (por ejemplo, un año) da como resultado un LT-O-PID. En la segunda etapa, se ejecutan dos iteraciones del cifrado adicional. El componente aleatorio de la primera iteración cambia en los mismos períodos de tiempo que la clave ST, es decir, por cada 24 horas. El componente aleatorio de la segunda y última iteración cambia con cada conjunto de datos de eventos de ubicación.

Por lo tanto, el LT-OOO-PID resultante se transmite al TP y la última iteración se invierte en el TP utilizando la clave privada del TP. Por lo tanto, el LT-OO-PID resultante es constante para todos los eventos transmitidos al socio de confianza TP dentro de las 24 horas.

En el bloque 60, el socio de confianza TP crea índices estadísticos basados en los conjuntos de datos de eventos de ubicación que entrega el DS a través de la segunda operación MAP dentro de la vida útil del LT-OO-PID (la vida útil de RC es de 24 horas). Como estos índices se calculan dentro de un intervalo a corto plazo, se denominan índices a corto plazo. Los índices estadísticos incluyen varios atributos, incluyendo un tipo de índice, una referencia, un valor y un alfa. El tipo de índice de atributo puede caracterizar un cierto tipo de evento, por ejemplo, SMS. El valor del atributo incluye valores de frecuencia simples (por ejemplo, número de envíos de SMS durante el período a corto plazo), así como valores de probabilidad (por ejemplo, el 80 % de los eventos entre las 7 p.m. y las 7 a.m. ocurrieron en una región geográfica que tiene el código postal 80639). Los eventos de ubicación originales se descartan después de haber sido procesados para calcular varios índices estadísticos. Antes del final de la vida útil de LT-OO-PID, un nuevo conjunto de datos con el LT-OO-PID, así como todos los índices ST construidos para este identificador dentro del período de corto plazo en el TP, se envían al agregador de datos DA.

El DA descifra el LT-OO-PID utilizando la clave privada apropiada (DA-Priv-Clave B). El próximo identificador a largo plazo LT-O-PID permite al DA combinar los nuevos índices a corto plazo recibidos del TP con valores históricos para el mismo LT-O-PID. Por lo tanto, en la etapa 70, el DA primero guarda los nuevos valores en su base de datos (historial de índices a corto plazo) y luego calcula nuevos índices a largo plazo basándose en todos los valores (nuevos e históricos) en la base de datos. Estos índices a largo plazo se combinan en un nuevo conjunto de datos y se guardan con el LT-O-PID como un identificador hasta el comienzo de un nuevo período de LT-OO-PID.

Puede ser conveniente almacenar estos índices a largo plazo en una base de datos común para todos los conjuntos de datos recibidos en el DA. Por lo tanto, un algoritmo de mapeo similar al descrito por la realización de acuerdo con la figura 3a, 3b es necesario. Sin embargo, el enfoque para generar y usar la tabla de traducción difiere ligeramente del de las figuras 3a, 3b.

De acuerdo con la realización de la figura 4, se genera una tabla de traducción en el DS cuando se ejecutan ambas operaciones MAP del proceso MAP 20. DS comprende un componente generador de tabla de traducción 30 que genera diferentes identificadores anónimos mediante el uso de diferentes instancias MAP con semilla adecuada para la anonimización de la base.

En la realización representada de la figura 4, el componente 30 genera una tabla que incluye tuplas de ST-OO-PID (con semilla constante de 24 h para anonimización de base) y LT-OO-PID (con semilla constante de un año para anonimización de base). En otras realizaciones, son posibles identificadores anónimos más diferentes (X-OO-PID, Y-OO-PID).

Además, el componente 30 también genera un "número" aleatorio único para cada tupla de identificadores anónimos. El resultado es una tabla con una columna para todas las representaciones de cada identificador anónimo y una columna adicional que contiene el número aleatorio único para cada tupla.
[ST-OO-PID | LT-OO-PID | ... | Número aleatorio único].

La tabla generada se reenvía a un componente de distribución de la tabla de traducción que también se implementa en el DS (no se muestra en la figura 4). Dicho componente de distribución de la tabla de traducción define cuál de los identificadores anónimos debe usarse como ID de integración (en la realización de la figura 4 se selecciona el ST-OO-PID). Posteriormente, el componente de distribución genera una tabla de dos columnas para cada tipo de identificador anónimo que contiene las representaciones del identificador anónimo respectivo y los números aleatorios únicos apropiados.

Las tablas resultantes pueden verse así:

[ST-OO-PID | Número aleatorio único] (Tabla 1)
[LT-OO-PID | Número aleatorio único] (Tabla 2)

En otras realizaciones, son concebibles más tablas para diferentes componentes de extracción: [X-OO-PID | Número aleatorio único] etc.

La tabla 1 marcada como "tabla de integración" se reenvía al componente de integración 40 ubicado en el TP. Cada "tabla de extracción" (tabla 2) se reenvía al componente de extracción 80 apropiado ubicado en el DA.

El componente de extracción 80 recibe la "tabla de extracción" apropiada (tabla 2) e invierte la primera iteración del cifrado adicional en todos los identificadores doblemente ofuscados almacenados en la tabla, es decir, todos los LT-OO-PID se descifran en LT-O-PID) (bloque 81).

En una siguiente etapa, los respectivos LT-O-PID almacenados con sus índices a largo plazo en la base de datos del componente 70 se reemplazan por números aleatorios únicos apropiados. La "tabla de atributos" resultante con identificadores aleatorios únicos se reenvía al componente de integración 40.

El componente de integración 40 recibe la "Tabla de integración" del componente de distribución 30. Además, la "tabla de atributos" resultante se recibe del componente de extracción 80. El componente de integración 40 combina todos los atributos de las diferentes "Tablas de atributos" por el número aleatorio único dado de cada "Tabla de atributos" y reemplaza los números aleatorios únicos por el identificador cifrado de destino apropiado, que se refiere al ST-OO-PID. La "tabla de atributos" recientemente combinada que incluye los índices a largo plazo y el ST-OO-PID se envía de vuelta al DA para almacenar la información recopilada en una base de datos común 100.

La figura 5 muestra un posible enfoque para construir índices a largo plazo que se implementa en el TP. La figura muestra diferentes eventos de ubicación activados por un solo suscriptor en diferentes ubicaciones A, B, C durante un cierto intervalo de tiempo. Diferentes eventos pueden ser una llamada telefónica o SMS saliente/entrante, El suscriptor se caracteriza por un cierto IMSI que permanece constante durante el intervalo de tiempo investigado.

La columna 200 muestra la implementación de iteración múltiple MAP para ALIP. Dado que el componente aleatorio utilizado para la última iteración de los cambios de cifrado adicionales para cada evento, cada LT-OOO-IMSI resultante difiere entre sí.

En la siguiente etapa 300, se calculan índices a corto plazo en el TP. Por lo tanto, la última iteración es revertida por la clave privada del TP. Como se puede ver en la Figura 5, los LT-OO-IMSI resultantes son constantes dentro de un cierto intervalo de tiempo (intervalo a corto plazo) que se refiere a la vida útil del componente aleatorio utilizado para la primera iteración del cifrado adicional. El TP ahora calcula un índice de actividad para una determinada ubicación. Los eventos ABA ocurrieron durante el primer intervalo a corto plazo 100. Dado que se han activado dos eventos en la ubicación A y solo se ha activado un evento B en la ubicación B, el valor de referencia dentro de dicho intervalo de tiempo 100 es A con un valor de probabilidad del 66 %. Se calcula un valor de referencia B diferente con una probabilidad del 60 % para el intervalo de tiempo 101 en el que el valor de referencia A para el intervalo de tiempo 102 se ha determinado con una probabilidad del 87 %.

Los índices calculados a corto plazo se transmiten después de cada intervalo de tiempo 100, 101, 102 al DA. En el DA, el LT-OO-IMSI se descifra, lo que da como resultado un único LT-O-IMSI ofuscado que permanece constante durante el intervalo de tiempo LT, que es de un año. Los índices a corto plazo que se reciben dentro del intervalo de tiempo LT pueden asociarse con un LT-O-IMSI común y los nuevos índices a largo plazo pueden calcularse sobre la base de los índices a corto plazo recopilados. Según el ejemplo mostrado en la figura 5, los eventos arbitrarios son activados por un solo suscriptor en la ubicación A dentro de un año con una probabilidad del 76,5 %.

Otra realización de la presente invención se muestra en las figuras 6a y 6b. Esta realización proporciona una solución para una aplicación de terceros para proporcionar atributos de datos adicionales (Atributo 1...Atributo n) identificados por un identificador personal secundario (SID), que también es conocido por el DS (variante 1 según la figura 6a), o por el PID original del DS (variante 2; figura 6b). En este contexto, la tercera parte actúa como una entidad de entrega de datos similar al DS. Una diferencia importante entre el DS y la entidad de tercera parte es que la tercera parte no realiza el MAP por sí misma.

Según una primera variante (figura 6a), un socio de confianza TP compara el conjunto de datos recibido de la tercera parte con un identificador personal doblemente encriptado a corto plazo (ST-OO-PID) generado (a través de un MAP) y proporcionado por el DS. Tal escenario podría ser razonable si la tercera parte no está dispuesta o no puede proporcionar sus datos directamente al DS. La siguiente argumentación describe este proceso basado en un identificador secundario (SID). También es posible realizar el proceso de correspondencia del socio de confianza TP utilizando el identificador personal PID. Por lo tanto, el PID simplemente necesita ser encriptado de la misma manera descrita para el SID. Es decir, debe encriptarse una segunda vez en paralelo al MAP en un proceso de un solo nivel.

Además de la generación del ST-OO-PID, el DS también define una nueva clave para el cifrado del identificador secundario SID (Clave SI). Por un lado, el DS utiliza esta clave para cifrar todos los SID en su base de datos con un cifrado simple (de un nivel) para obtener el O-SID. Por otro lado, la clave se proporciona a la tercera parte para habilitar el mismo proceso en sus conjuntos de datos. El resultado en el DS es una tabla de asignación que vincula todos los O-SID con el ST-OO-PID apropiado. La tabla se transmite al socio de confianza TP.

La tercera parte también aplica el cifrado de un nivel al identificador SID incluido en sus conjuntos de datos con atributos adicionales. Los conjuntos de datos resultantes, incluido el O-SID como identificador, se transmiten al TP.

El TP ahora realiza una búsqueda en la base de datos en el DB de coincidencia de identificador para encontrar el ST-OO-PID apropiado para el O-SID de cada conjunto de datos recibido de la tercera parte. Después de reemplazar el identificador, el nuevo conjunto de datos, incluido el ST-OO-PID y los atributos adicionales de la tercera parte, se reenvía al agregador de datos DA.

El DA descifra el ST-OO-PID (de acuerdo con la lógica MAP descrita anteriormente) y obtiene el ST-O-PID. En función de este identificador único, el DA puede realizar la combinación con otras fuentes de datos, como datos de

eventos de ubicación, datos de clase de clientes o índices a largo plazo.

En la segunda variante (figura 6b) no se desea ocultar la información de la tercera parte del DS. En este caso, la tercera parte podría simplemente transferir sus datos al DS, donde se cifra a través de un MAP de la misma manera que cualquier fuente de datos interna del DS. Después de descifrar el ST-OO-PID en el DA, es posible una combinación de datos de la manera común.

REIVINDICACIONES

1. Un método para anonimización mediante la transmisión de un conjunto de datos de eventos desde al menos una entidad proveedora de datos que suministra el conjunto de datos de eventos a varias entidades de agregación de datos que agregan el conjunto de datos de eventos, en el que el conjunto de datos incluye al menos un identificador que identifica al menos un usuario de la entidad de suministro de datos y al menos un atributo asociado con al menos un identificador/usuario e incluye al menos uno del grupo de tipo de evento y/o la marca de tiempo de un evento y/o la ubicación de un evento, comprendiendo el método las etapas de:

a. realizar un cifrado básico no reversible del al menos un identificador utilizando un mecanismo de cifrado con una determinada vida útil,
b. realizar de forma iterativa un cifrado adicional de dicho identificador cifrado básico n veces, con n igual o mayor que 2 dando como resultado un identificador cifrado adicional,
en el que el cifrado adicional comprende las etapas de agregar un componente aleatorio a dicho identificador cifrado y cifrar la salida usando cifrado asimétrico con una clave pública,
en el que cada iteración utiliza un componente aleatorio con una vida útil más corta que la vida útil de la vida útil determinada según la etapa a y que la vida útil de la iteración anterior y una clave pública diferente a la de la iteración anterior,
c. transmitir el conjunto de datos de evento **caracterizado por** el identificador cifrado adicional n veces a una primera entidad de agregación de datos, y
d. deshacer al menos una iteración de los cifrados adicionales en la primera entidad de agregación de datos de recepción que da como resultado un identificador cifrado adicional n-1 veces en el que el primer agregador de datos deriva al menos del índice estadístico de una serie de conjuntos de datos de eventos relacionados con el mismo identificador cifrado adicional n-1 veces y recopilado dentro de la vida útil del conjunto de componentes aleatorios durante el cifrado adicional n-1,
e. el primer agregador de datos reenvía dichos índices estadísticos derivados, junto con su identificador cifrado adicional n-1 veces adicionales a al menos otra entidad de agregación de datos en la que la recepción de al menos otra entidad de agregación de datos deshace al menos una iteración del identificador cifrado adicional n-1 veces que resulta en el identificador cifrado adicional n-2 veces o el identificador cifrado básico no reversible y almacena los índices estadísticos recibidos relacionados con el mismo identificador cifrado adicional n-2 veces o el identificador cifrado básico no reversible dentro de la vida útil del conjunto de componentes aleatorios durante la iteración de deshacer del n-2 identificador cifrado adicional o la cierta vida útil utilizada durante el cifrado básico no reversible.

2. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la al menos una entidad de suministro de datos y la al menos una entidad de agregación de datos están representadas por dos procesos independientes ejecutados en sistemas físicamente separados, en particular, ubicados en diferentes lugares de instalación.

3. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que las vidas de la vida útil determinada de la etapa a y/o las vidas de los componentes aleatorios pueden variar, en particular, dependiendo de los atributos de los conjuntos de datos de eventos.

4. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la vida útil del componente aleatorio de la última iteración cambia para cada conjunto de datos.

5. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el cifrado básico se basa en un algoritmo de hashing que usa una semilla con una larga vida útil.

6. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el al menos un índice estadístico derivado incluye al menos un atributo que incluye al menos uno del grupo de tipo de índice y/o referencia y/o valor de probabilidad y/o frecuencia y/o probabilidad de error.

7. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que al menos un índice estadístico se refiere a un índice de movilidad que caracteriza la frecuencia y/o probabilidad de un identificador encriptado adicional n veces y anónimo asociado con una determinada ubicación.

8. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que al menos un índice estadístico se refiere a un índice de actividad que caracteriza la frecuencia y/o probabilidad de una cierta base anónima y un identificador cifrado adicional n veces asociado con un determinado evento.

9. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que se genera una gran tabla de coincidencia que contiene al menos una tupla de datos para cada identificador personal de la al menos una entidad proveedora de datos en la que cada tupla de datos consiste preferiblemente en un identificador personal encriptado

adicional 1 vez y anónimo de base para cada tipo diferente de información que debe extraerse de al menos un extractor y/o entidad de agregación de datos híbrida y/o un identificador personal encriptado adicional 1 vez y anónimo para cada tabla interna dentro de al menos un integrador y/o entidad de agregación de datos híbrida y/o al menos un componente temporal único.

10. El método de acuerdo con la reivindicación 9, en el que se genera al menos una tabla de coincidencia pequeña para cada tipo identificador personal de encriptado adicional 1 vez y anónino de base incluido en la tabla de coincidencia grande en la que cada tabla pequeña preferiblemente incluye una tupla de datos que contiene la base anonimizada e identificador personal encriptado adicional 1 vez junto con el componente temporal único.

11. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que al menos una entidad de suministro de datos es una red de comunicación móvil.

12. Un sistema de comunicación para realizar el método de acuerdo con una cualquiera de las reivindicaciones anteriores.

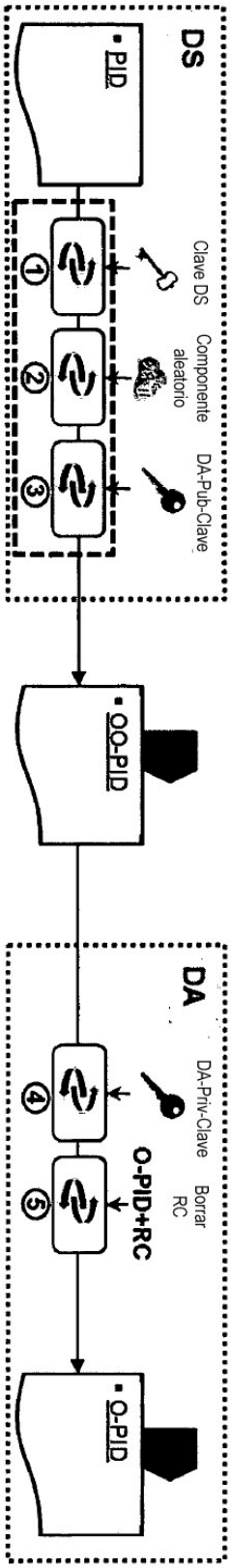

DS
PID
Clave DS
Componente aleatorio
DA-Pub-Clave
1
2
3
OO-PID
DA
DA-Priv-Clave
Borrar RC
O-PID+RC
4
5
O-PID

Fig. 1a

Fig. 1b

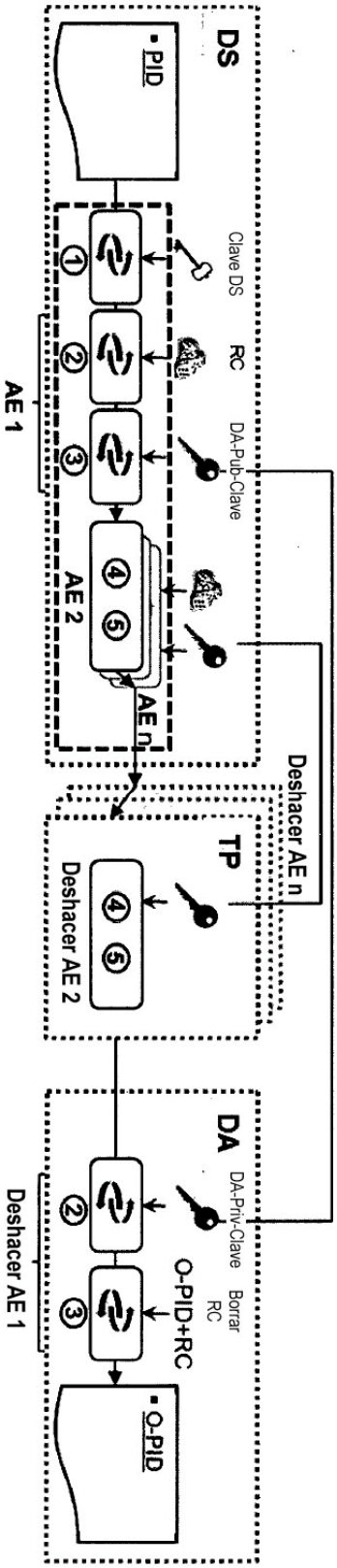

DS
PID
Clave DS
RC
DA-Pub-Clave
AE 1
AE 2
AE n
Deshacer AE n
TP
Deshacer AE 2
DA
DA-Priv-Clave
Borrar RC
O-PID+RC
Deshacer AE 1
O-PID

Fig. 2a

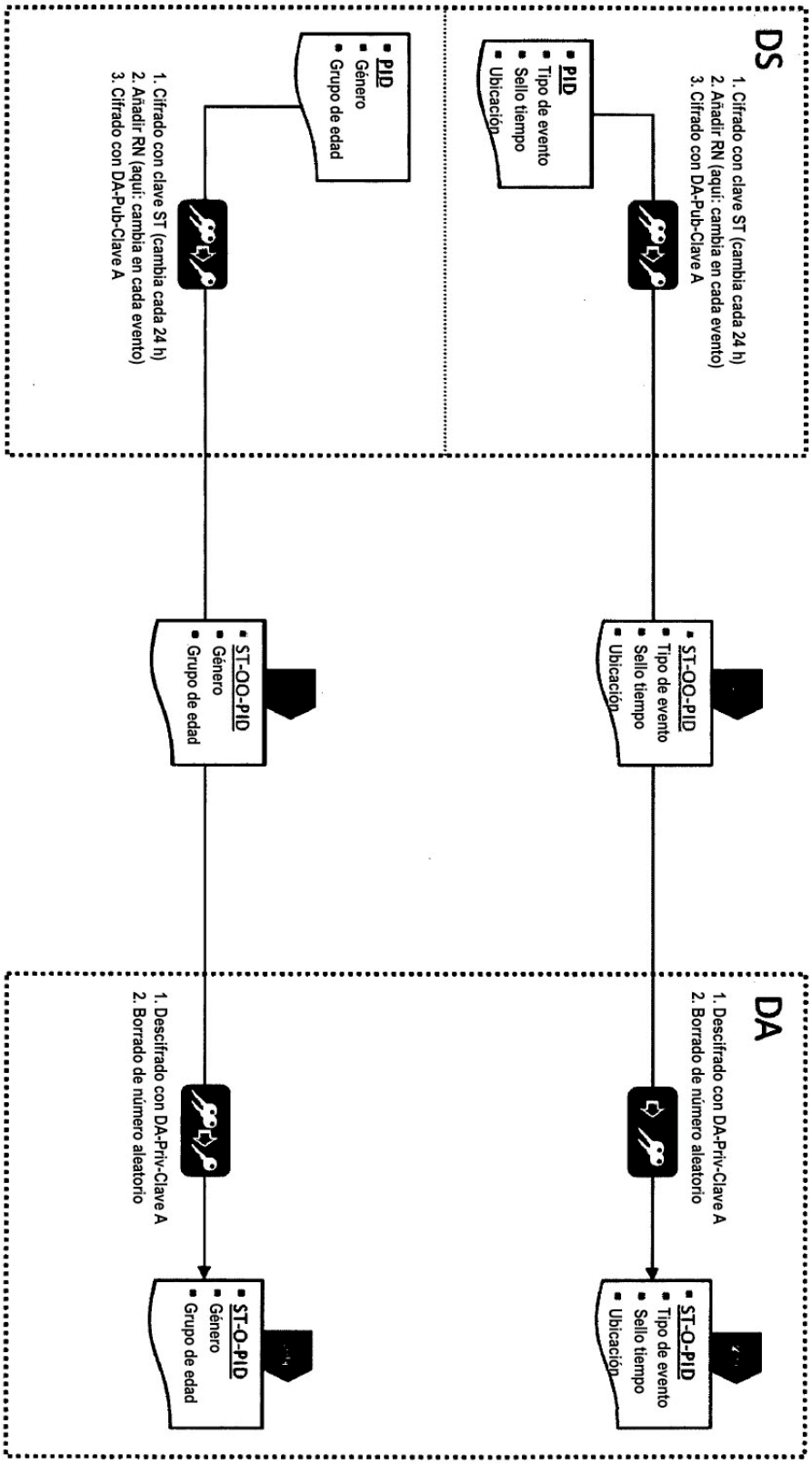

DS
1. Cifrado con clave ST (cambia cada 24 h)
2. Añadir RN (aquí: cambia en cada evento)
3. Cifrado con DA-Pub-Clave A
PID
Tipo de evento
Sello tiempo
Ubicación
PID
Género
Grupo de edad
1. Cifrado con clave ST (cambia cada 24 h)
2. Añadir RN (aquí: cambia en cada evento)
3. Cifrado con DA-Pub-Clave A
ST-OO-PID
Tipo de evento
Sello tiempo
Ubicación
ST-OO-PID
Género
Grupo de edad
DA
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
Tipo de evento
Sello tiempo
Ubicación
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
Género
Grupo de edad

Fig. 2b

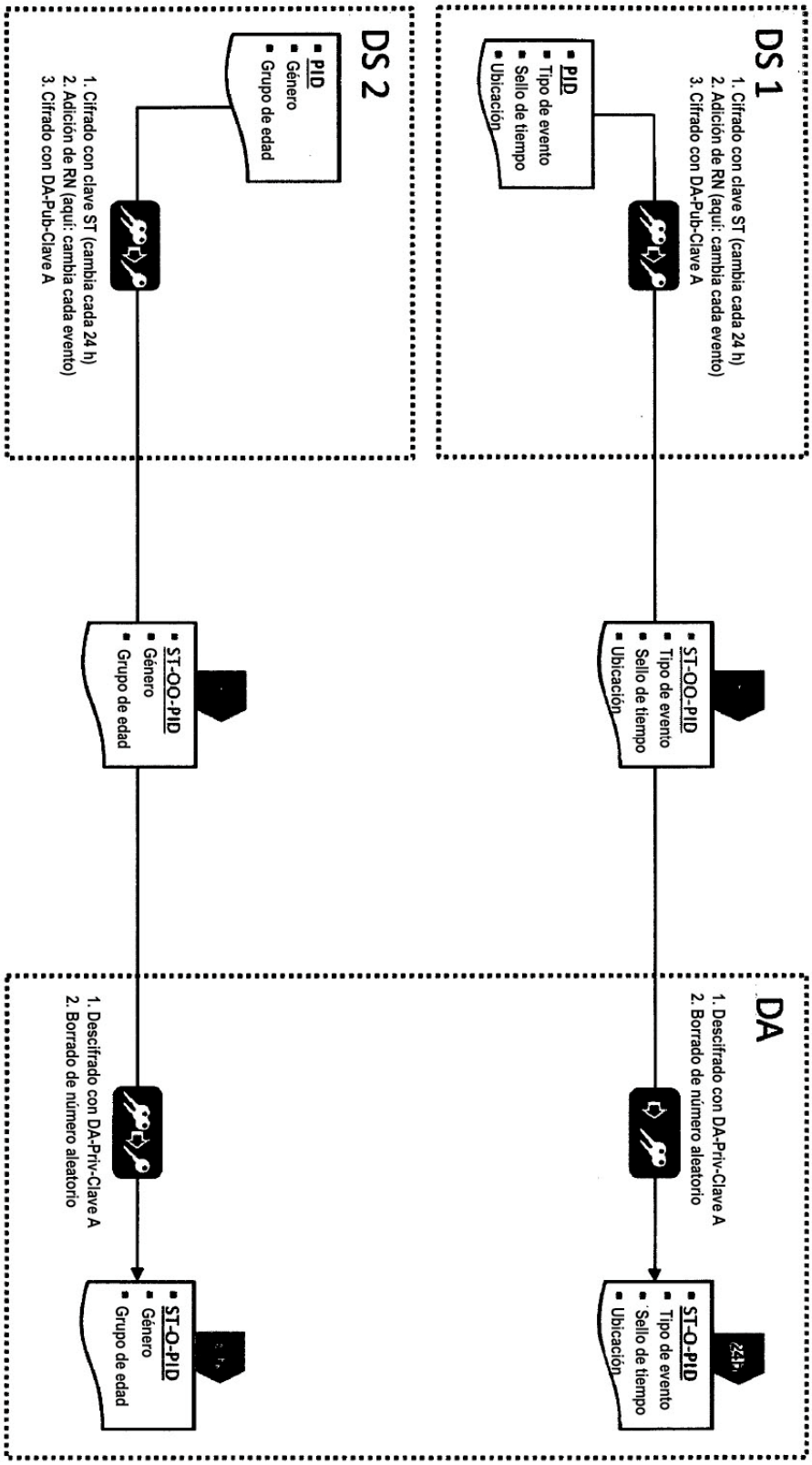

DS 1
1. Cifrado con clave ST (cambia cada 24 h)
2. Adición de RN (aquí: cambia cada evento)
3. Cifrado con DA-Pub-Clave A
PID
Tipo de evento
Sello de tiempo
Ubicación
DS 2
PID
Género
Grupo de edad
1. Cifrado con clave ST (cambia cada 24 h)
2. Adición de RN (aquí: cambia cada evento)
3. Cifrado con DA-Pub-Clave A
ST-OO-PID
Tipo de evento
Sello de tiempo
Ubicación
ST-OO-PID
Género
Grupo de edad
DA
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
Tipo de evento
Sello de tiempo
Ubicación
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
Género
Grupo de edad

Fig. 3a

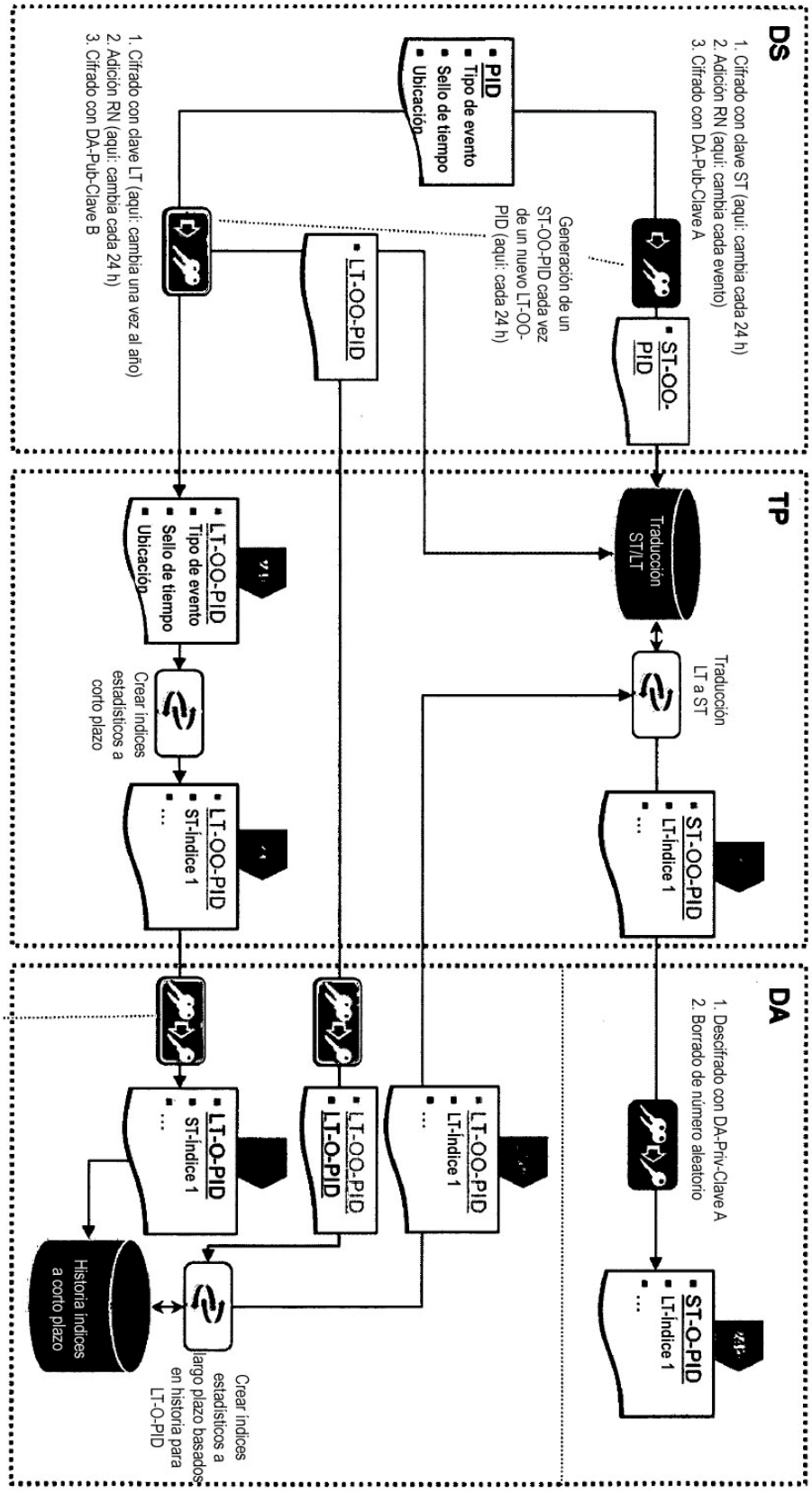

DS
1. Cifrado con clave ST (aquí: cambia cada 24 h)
2. Adición RN (aquí: cambia cada evento)
3. Cifrado con DA-Pub-Clave A
PID
Tipo de evento
Sello de tiempo
Ubicación
ST-OO-PID
Generación de un ST-OO-PID cada vez de un nuevo LT-OO-PID (aquí: cada 24 h)
LT-OO-PID
1. Cifrado con clave LT (aquí: cambia una vez al año)
2. Adición RN (aquí: cambia cada 24 h)
3. Cifrado con DA-Pub-Clave B
TP
Traducción ST/LT
Traducción LT a ST
ST-OO-PID
LT-Índice 1
LT-OO-PID
Tipo de evento
Sello de tiempo
Ubicación
Crear índices estadísticos a corto plazo
LT-OO-PID
ST-Índice 1
DA
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
LT-Índice 1
LT-OO-PID
LT-Índice 1
LT-OO-PID
LT-O-PID
LT-O-PID
ST-Índice 1
Historia índices a corto plazo
Crear índices estadísticos a largo plazo basados en historia para LT-O-PID
1. Descifrado con DA-Priv-Clave B
2. Borrado de número aleatorio

Fig. 3b

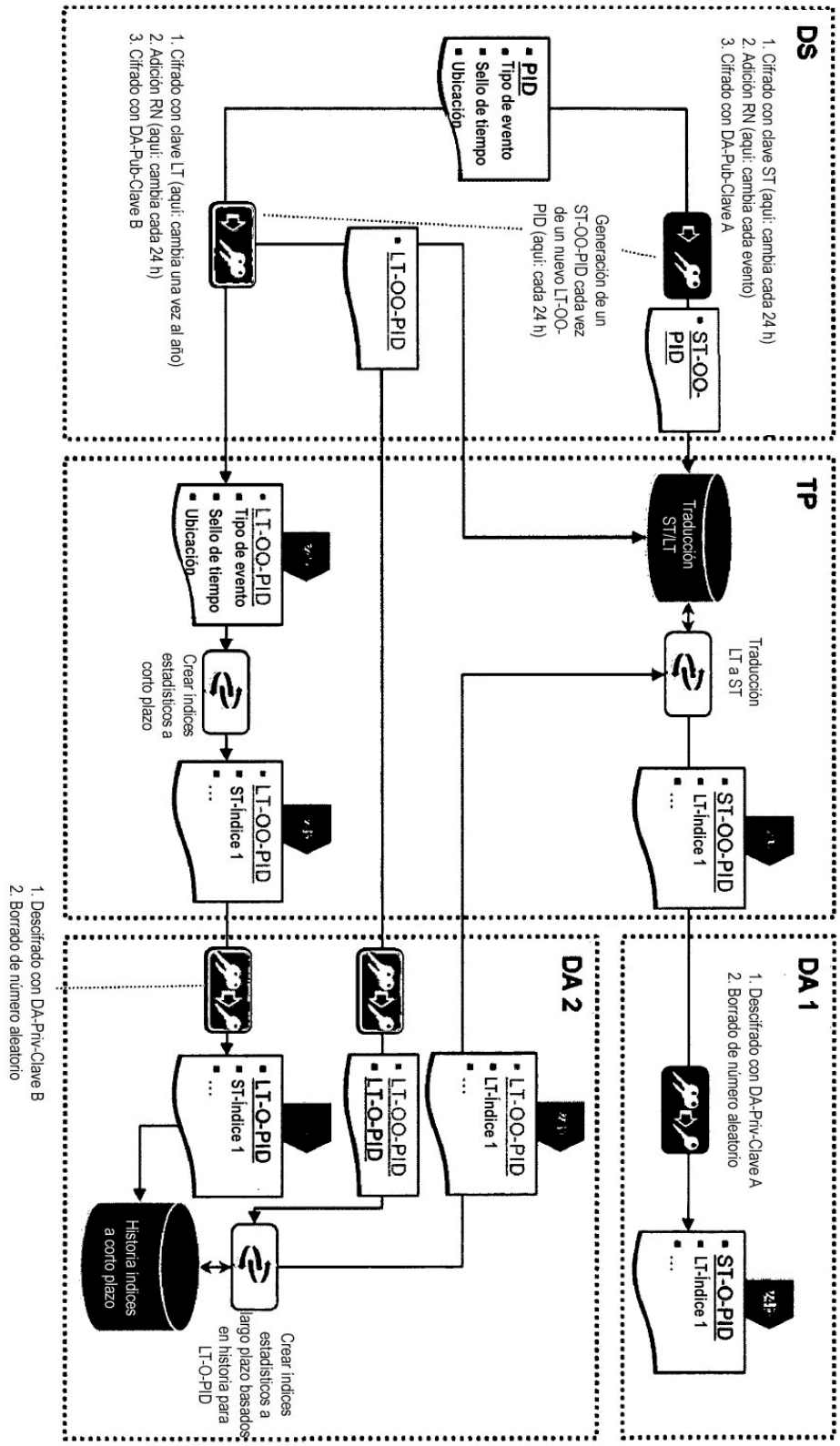

DS
1. Cifrado con clave ST (aquí: cambia cada 24 h)
2. Adición RN (aquí: cambia cada evento)
3. Cifrado con DA-Pub-Clave A
PID
Tipo de evento
Sello de tiempo
Ubicación
ST-OO-PID
Generación de un ST-OO-PID cada vez de un nuevo LT-OO-PID (aquí: cada 24 h)
LT-OO-PID
1. Cifrado con clave LT (aquí: cambia una vez al año)
2. Adición RN (aquí: cambia cada 24 h)
3. Cifrado con DA-Pub-Clave B
TP
Traducción ST/LT
Traducción LT a ST
ST-OO-PID
LT-Índice 1
LT-OO-PID
Tipo de evento
Sello de tiempo
Ubicación
Crear índices estadísticos a corto plazo
LT-OO-PID
ST-Índice 1
DA 1
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
LT-Índice 1
DA 2
LT-OO-PID
LT-Índice 1
LT-OO-PID
LT-O-PID
LT-OO-PID
ST-Índice 1
LT-O-PID
Historia índices a corto plazo
Crear índices estadísticos a largo plazo basados en historia para LT-O-PID
1. Descifrado con DA-Priv-Clave B
2. Borrado de número aleatorio

Fig. 4

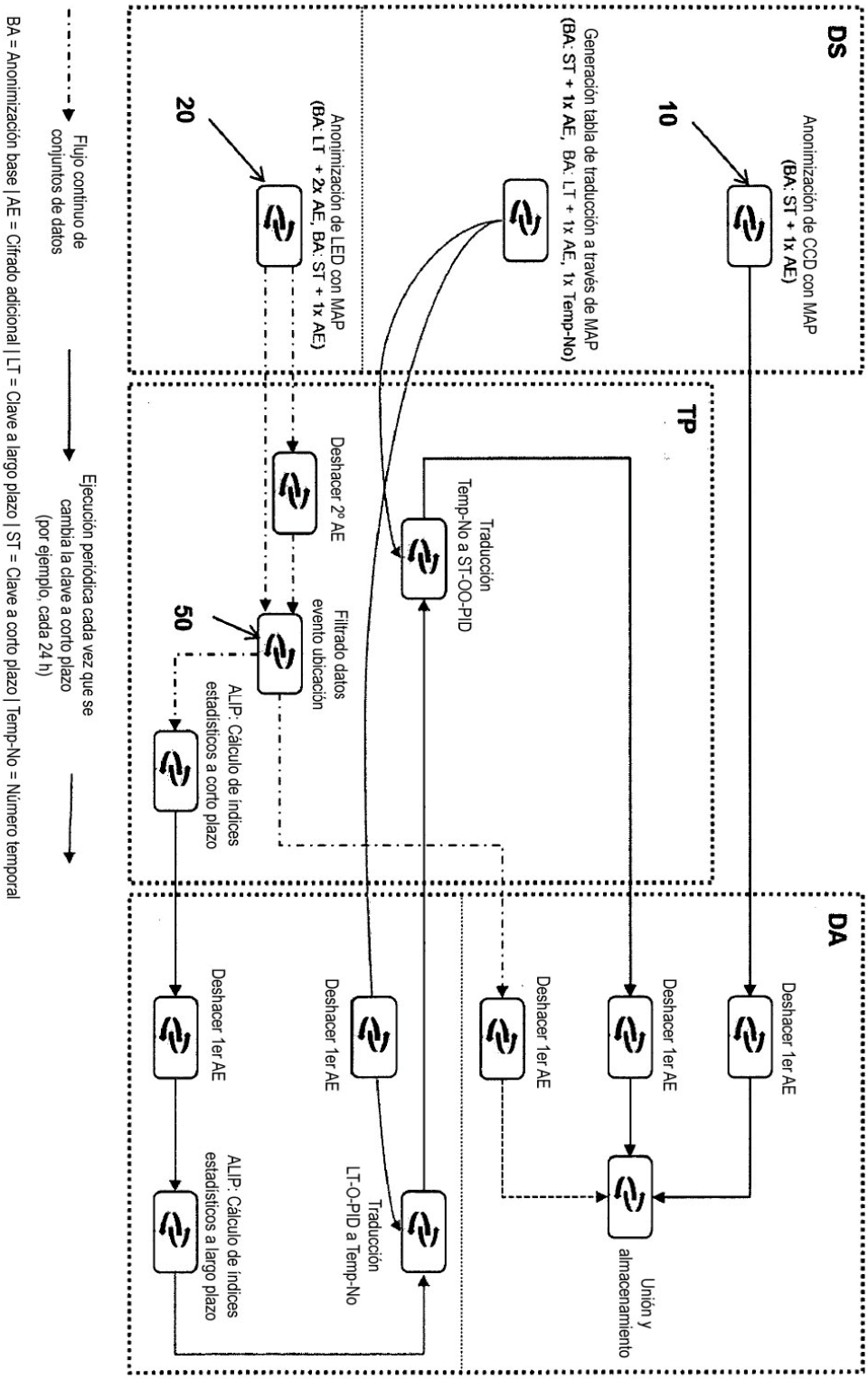

DS
Anonimización de CCD con MAP (BA: ST + 1x AE)
10
Generación tabla de traducción a través de MAP (BA: ST + 1x AE, BA: LT + 1x AE, 1x Temp-No)
Anonimización de LED con MAP (BA: LT + 2x AE, BA: ST + 1x AE)
20
TP
Deshacer 2º AE
Filtrado datos evento ubicación
50
Traducción Temp-No a ST-OO-PID
ALIP: Cálculo de índices estadísticos a corto plazo
DA
Deshacer 1er AE
Deshacer 1er AE
Deshacer 1er AE
Unión y almacenamiento
Deshacer 1er AE
Traducción LT-O-PID a Temp-No
ALIP: Cálculo de índices estadísticos a largo plazo
Flujo contínuo de conjuntos de datos
Ejecución periódica cada vez que se cambia la clave a corto plazo (por ejemplo, cada 24 h)
BA = Anonimización base | AE = Cifrado adicional | LT = Clave a largo plazo | ST = Clave a corto plazo | Temp-No = Número temporal

Fig. 5

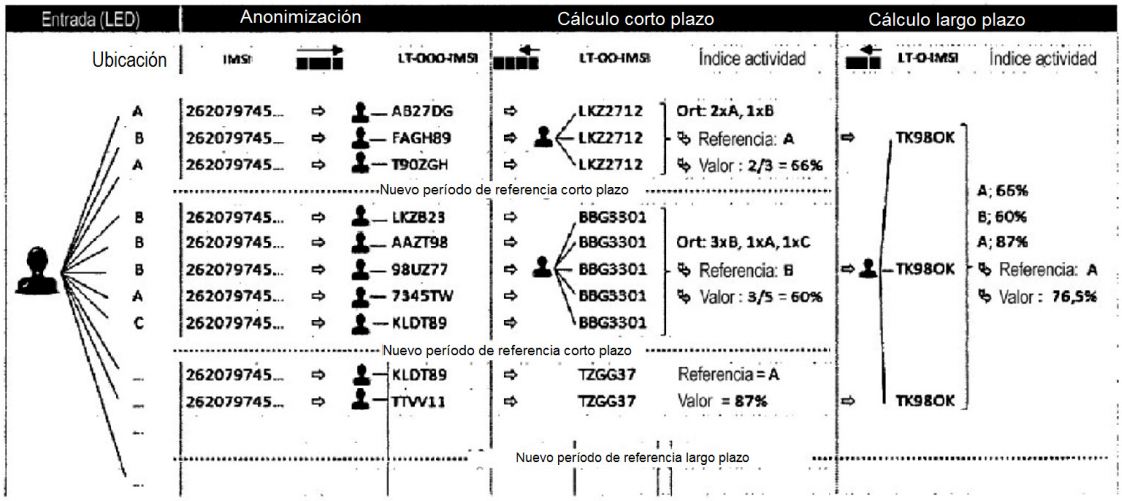

Entrada (LED)
Anonimización
Cálculo corto plazo
Cálculo largo plazo
Ubicación
IMSI
LT-OOO-IMSI
LT-OO-IMSI
Índice actividad
LT-O-IMSI
Índice actividad
A 262079745... AB27DG LKZ2712 Ort: 2xA, 1xB
B 262079745... FAGH89 LKZ2712 Referencia: A TK98OK
A 262079745... T90ZGH LKZ2712 Valor : 2/3 = 66%
Nuevo período de referencia corto plazo
A; 66%
B; 60%
A; 87%
B 262079745... LKZB23 BBG3301
B 262079745... AAZT98 BBG3301 Ort: 3xB, 1xA, 1xC
B 262079745... 98UZ77 BBG3301 Referencia: B TK98OK Referencia: A
A 262079745... 7345TW BBG3301 Valor : 3/5 = 60% Valor : 76,5%
C 262079745... KLDT89 BBG3301
Nuevo período de referencia corto plazo
262079745... KLDT89 TZGG37 Referencia = A
262079745... TTVV11 TZGG37 Valor = 87% TK98OK
Nuevo período de referencia largo plazo

Fig. 6a

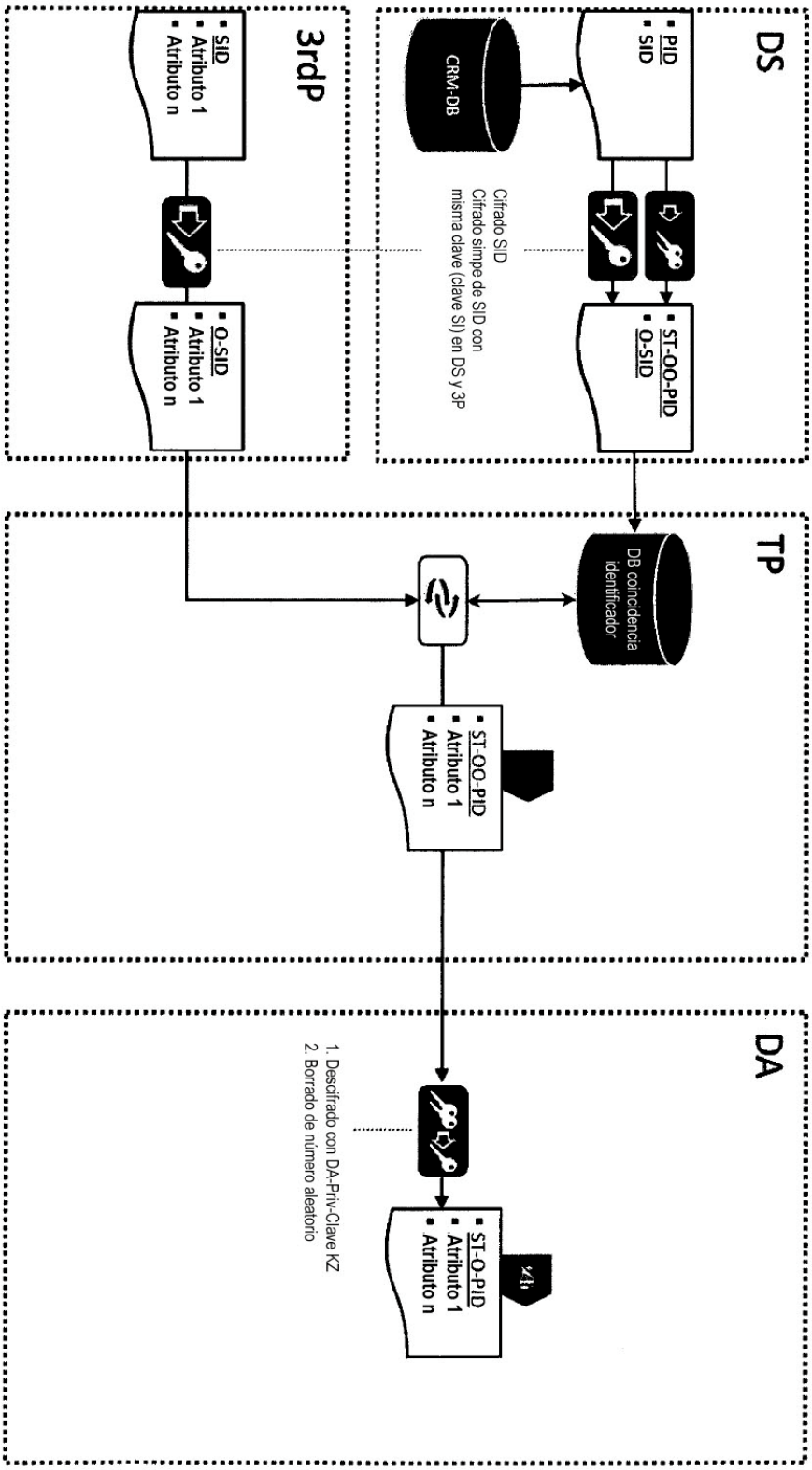
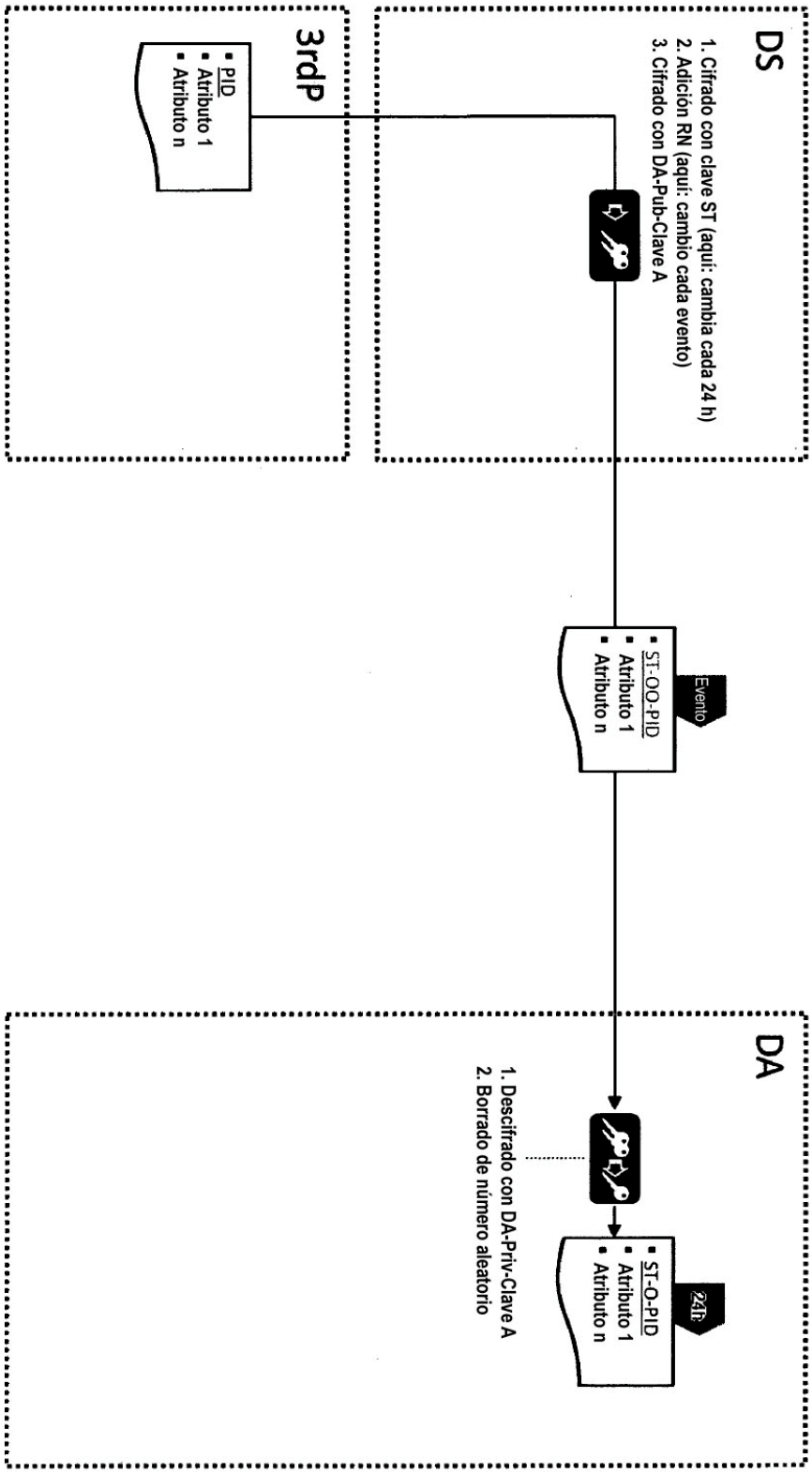

DS
PID
SID
CRM-DB
Cifrado SID
Cifrado simpe de SID con misma clave (clave SI) en DS y 3P
ST-OO-PID
O-SID
3rdP
SID
Atributo 1
Atributo n
O-SID
Atributo 1
Atributo n
TP
DB coincidencia identificador
ST-OO-PID
Atributo 1
Atributo n
DA
1. Descifrado con DA-Priv-Clave KZ
2. Borrado de número aleatorio
ST-O-PID
Atributo 1
Atributo n

Fig. 6b


DS
1. Cifrado con clave ST (aquí: cambia cada 24 h)
2. Adición RN (aquí: cambio cada evento)
3. Cifrado con DA-Pub-Clave A
3rdP
PID
Atributo 1
Atributo n
ST-OO-PID
Atributo 1
Atributo n
Evento
DA
1. Descifrado con DA-Priv-Clave A
2. Borrado de número aleatorio
ST-O-PID
Atributo 1
Atributo n
24h