TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

UNIVERSIDAD DE ALMERIA

ESCUELA SUPERIOR DE INGENIERÍA

"Desarrollo de una Técnica de Seudonimización de Datos Personales basada en criptografía"

Curso 2020/2021

Alumno/a:

Mustapha Chakra

Director/es:

José Andrés, Moreno Ruiz

AGRADECIMIENTOS

En primer lugar, quiero dar las gracias a mi tutor *José Andrés Moreno Ruiz* por haberme ayudado en la elaboración de mi proyecto y el haberme dado libertad para desarrollarlo y llevarlo a cabo con mi enfoque y a mi ritmo.

Seguidamente, quiero agradecer a todos los profesores que me han impartido clase en la carrera, como ha dicho Ever Garrison: “Un profesor es una brújula que activa los imanes de la curiosidad, el conocimiento y la sabiduría en los alumnos”.

Como no, debo de agradecer a mis padres *Abdelkader Chakra* y *Kaida Choukri* que día a día y a lo largo de toda mi vida, hayan creído en mi persona y en mi capacidad más que yo mismo. A mis amigos por aburrirles con mis discursillos informáticos, y muy especialmente a mi esposa *Ilham Ouadou* por su paciencia y ayuda, y a mi hijo *Tayssir* por darle sentido a mi vida, así como la motivación que me hace falta para luchar día a día.

Por último, quiero agradecer de corazón la ayuda recibida por mis compañeros de clase en más de una ocasión y decir que sin ellos no hubiese sido ni la mitad de ameno.

TABLA DE CONTENIDO

INTRODUCCIÓN:

La capacidad de generar y recopilar información sobre las personas en nuestra sociedad actual de la información desafía las regulaciones de protección de datos, que hasta ahora parecían proporcionar instrumentos apropiados para supervisar la legalidad de los flujos de datos.

Nuestra privacidad está ahora bajo constante amenaza. El progreso tecnológico continuo compromete la seguridad de los datos almacenados, que es probable que sufran ataques externos e incluso pueden ser utilizados para fines ilegítimos por aquellos que los poseen. Cada día, se recogen y almacenan una cantidad ingente de información personal como resultado de haber compartido dicha información, independientemente de que se tenga o no conocimiento de ello.

Por estas razones, hasta ahora se han discutido temas de "anonimato" y "técnicas de anonimización", considerando especialmente las preocupaciones derivadas de los fenómenos de análisis y predicción del comportamiento de los usuarios en línea, empezando por el "Big data" y la expansión de bases de datos tradicionales en las organizaciones públicas y privadas.

El Reglamento de la Unión Europea (2016/679) del Parlamento Europeo y el Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de dichos datos[1], **que ha entrado en vigor desde el 25 mayo del año 2018**, introduce en el artículo 4, entre sus definiciones, un nuevo concepto jurídico:

"*Seudonimización: el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyen a una persona física identificada o identificable*"

El Reglamento Europeo no define procedimientos contemplados en la legislación española como "disociar" o "anonimizar", sino que introduce un nuevo concepto jurídico, que en la actualidad no se reconoce en la legislación española vigente: *seudonimizar*[2].

Se entiende que el proceso de *seudonimizar* consiste en separar la información original, de tal modo que sin volverla a unir o asociarla, no sería posible identificar a las personas físicas. El ejemplo que encajaría en dicho proceso sería el de aplicar un código a una muestra biológica que se envía a analizar a un laboratorio. De este modo, en el laboratorio obtendrían una información "seudonomizada", y sin disponer de información adicional no podrían conocer de qué persona física se trata. Por lo tanto, efectuando un proceso inverso o reversible a la *seudonimización* se podría volver a obtener el dato de la persona física identificada.

Este trabajo presenta una arquitectura de seguridad para la privacidad de datos, que es estrictamente controlada por el propietario de los datos, es decir, es él quien decide a quién se le da acceso a sus datos, lo que elimina la confianza requerida en los administradores, especialmente los de bases de datos. Puesto que confiar en una única estrategia de seguridad tiene sus inconvenientes, integramos técnicas de seudonimización y cifrado para superar sus defectos individuales y crear un protocolo que utiliza pseudónimos como mecanismo de control de acceso y protege las claves criptográficas secretas por un modelo de seguridad basado en capas.

[1] Se deroga por la Directiva 95/46/CE, Reglamento general de protección de datos.
[2] No aparece recogido en el diccionario de la Real Academia Española, pero sí en el Diccionario panhispánico del español jurídico.

OBJETIVOS

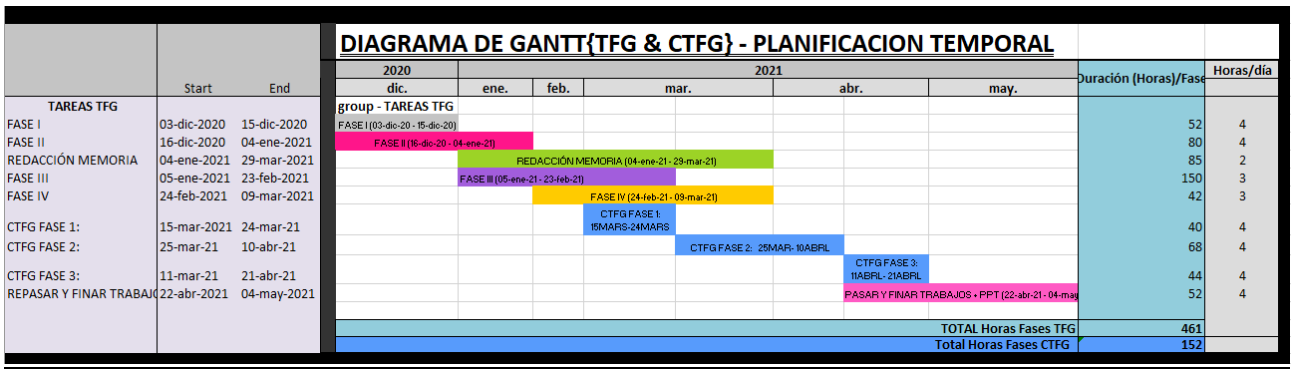
Este trabajo presenta un desarrollo de un sistema de seguridad para la privacidad de los datos personales, la cual está estrictamente controlada por el propietario de los datos. A continuación, se exponen los objetivos principales de nuestro trabajo:

- Integración de las técnicas de seudonimización y cifrado de datos, con el fin de establecer una metodología para un modelo de seguridad basado en capas.

- Aplicación de los algoritmos criptográficos para asegurar la privacidad, integridad y autenticidad de la información.

- Desarrollo de una técnica específica de seudonimización, en la que la persona es el legítimo dueño de sus datos, y puede acceder a ellos en tiempo real, y solo puede dar consentimiento a una segunda persona para acceder a sus datos o a parte de ellos.

PLANIFICACIÓN TEMPORAL:



FASE I: ESTUDIAR LOS ENFOQUES EXISTENTES PARA TRATAR LA PRIVACIDAD Y CONFIDENCIALIDAD DE LOS DATOS PERSONALES

1. Introducción:

En el mundo en que vivimos, la privacidad de los datos personales se ha convertido en una realidad abstracta, ya que, y en términos legales, la línea divisoria entre lo que es personal y lo que es información pública no está totalmente definida.

Además, hay otros aspectos de la cultura de seguridad de la información que, a menudo, se pasan por alto, y que se deberían cuestionar: ¿Se permite el uso de su información y datos para futuras ofertas comerciales?

Cuando se producen estas situaciones, nos enfrentamos a la solicitud de consentimiento expreso, y después de eso también tenemos que hacer frente a un consentimiento tácito al procesamiento y visualización de los datos personales.
La existencia de un consentimiento tácito es la puerta abierta a la violación de los derechos de los ciudadanos, especialmente con respecto a la confidencialidad de la información y la privacidad de los ciudadanos. La mayoría de los sistemas de información con los que tratamos a diario se nos presenta como una herramienta de valor añadido para nuestra comodidad, productividad y reducción de costes.

De hecho, la cultura de la seguridad y el comportamiento preventivo de los ciudadanos respecto de la información que muestran, es el garante de sus derechos, teniendo en cuenta que vivimos en una sociedad cada vez más competitiva, donde las personas no son conscientes ni tienen el conocimiento suficiente sobre el funcionamiento de la privacidad en la red. Y, por lo tanto, deben existir unos mecanismos y sistemas que sean eficaces y nos garanticen un nivel alto de protección, para proteger la privacidad de las personas con relación a la confidencialidad de los datos personales.
Estos mecanismos son los que llamamos “enfoques”, es lo que vamos a ver en este apartado. Estudiaremos los distintos enfoques existentes, basados en el análisis de los conceptos, que nos llevan a avanzar hacia al núcleo de este estudio.

2. Conceptos y Análisis:

2.1. Tratamiento de Los datos personales:

Según el Reglamento del Parlamento Europeo con relación a la protección de las personas físicas en lo que respecta al tratamiento de datos personales, destacamos los siguientes conceptos y definiciones:

2.1.1. DATOS PERSONALES:

Según el Reglamento del Parlamento Europeo[3], los datos personales hacen referencia a “toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona”.

Los datos personales son datos que se refieren a individuos/personas, y los podemos dividir en dos tipos principales: datos identificables (directamente o indirectamente) y datos anonimizados. De tal forma que cada tipo tiene sus propias reglas de procesión.

Los datos también pueden ser identificativos (por ejemplo, nombre y número de identificación personal, etc.) o descriptivos (por ejemplo, información de salud, información financiera, etc.)

2.1.1.1. Principales tipos de datos personales:

[3] Reglamento del Parlamento Europeo, recogido en el decreto 2016/679

Los datos directamente identificables son datos que tienen un nombre (o número) de identificación personal adjunto a ellos, mientras que los datos des-identificados son aquellos que permiten la identificación indirecta utilizando un número de serie u otra información como identificador.

Los datos indirectamente identificables son los que nos permiten reconocer datos anonimizados de otra manera, a través de una combinación de los datos descriptivos proporcionados en el conjunto de datos. La identificación indirecta puede darse de forma aleatoria o sistemática. Si el conjunto de datos permite volver a identificar sistemáticamente todas las partes (o las esenciales) de los datos, entonces, en este caso deben tratarse como datos personales identificables.

Los datos indirectamente identificables corresponden a datos anónimos, excepto si el conjunto de datos contiene una clave que señala de nuevo a un identificador personal, o si el conjunto de datos es tan amplio que los datos en su conjunto permiten la identificación del individuo.

A pesar de esta distinción, los datos anónimos no son datos personales, en el sentido estricto, porque no debería ser posible rastrear los datos de los individuos.

2.1.2. Tratamientos automatizados y Elaboración de perfiles:

2.1.2.1. TRATAMIENTO:

Según el Reglamento de la Unión Europea[4], el concepto de tratamiento hace referencia *a "cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción[5]."*

2.1.2.2. ELABORACIÓN DE PERFILES:

Según el Reglamento de la Unión Europea[6], la elaboración de perfiles hace referencia a *"toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física".*

2.1.2.3. La regulación de las decisiones individuales automatizadas

El Reglamento General de Protección de Datos (GDPR) de 2016, extiende la protección contra las decisiones adoptadas únicamente sobre la base de un tratamiento automatizado para abarcar no sólo el perfil de los datos, sino también cualquier otra forma de tratamiento automatizado.

Todos los principios de protección de datos se aplican a dicho tratamiento, pero quizás lo más importante son los requisitos del primer principio, que estipula que el tratamiento de los datos personales debe ser lícito,

[4] Reglamento de la Unión Europea, en el decreto 2016/679

[5] El concepto "tratamiento" también fue abordado por M. Schulz (M. SCHULZ, 2016)

[6] Reglamento de la Unión Europea, en el decreto 2016/679

justo y transparente. Aunque esto puede parecer sencillo, la aplicación práctica al aprendizaje automático de cada elemento de este principio es probable que sea todo un desafío.

El apartado 1 del artículo 22 del GDPR concede a los interesados el derecho a no estar sujetos a la toma de decisiones, incluida la elaboración de perfiles, basándose únicamente en la toma de decisiones automatizada, que produzca efectos jurídicos que les afecten de manera similar.

Los datos personales utilizados para las decisiones automatizadas, incluida la elaboración de perfiles, sólo deben recopilarse para fines específicos, explícitos y legítimos, y no se permite el procesamiento posterior que no sea utilizado con tales fines.

Por lo que respecta a la transparencia, los artículos 13[7] y 14, del GDPR obligan a los responsables del tratamiento a informar a los interesados (en el momento de la recopilación) sobre «la existencia de una toma de decisiones automatizada», Información significativa sobre la lógica implicada, así como la importancia y las consecuencias previstas de dicho tratamiento [...] ».

Por último, si bien se ha prestado considerable atención a los peligros de la ausencia de la transparencia en los procesos de toma de decisiones algorítmicas, no debe olvidarse que la toma de decisiones en los seres humanos suele estar influida por prejuicios, tanto conscientes como inconscientes e incluso por el metabolismo. De hecho, si bien puede ser extremadamente difícil asegurar una transparencia completa en los procesos de toma de decisiones automatizados, incluso los responsables humanos bien intencionados son susceptibles a prejuicios de los que ni siquiera son conscientes.

Privacidad y Confidencialidad:

En la guía estándar para los principios de confidencialidad, privacidad, acceso y seguridad de datos para la información de salud, la *Sociedad Estadounidense de Pruebas y Materiales sobre Informática Sanitaria, privacidad, confidencialidad y acceso* define los siguientes principios:

2.1.3. Privacidad:

Ilustración 1. Principios y fundamentos de la privacidad (fuente: deloitte.com)

[7] Apartado 2, letra f), y apartado 2, letra g)

El concepto de privacidad según A. Buckovich (1999) hace referencia a *"el derecho de los individuos a ser dejados solos y protegidos contra la invasión física o psicológica o el mal uso de sus bienes. Incluye la libertad de intrusión o de observación en los asuntos privados, el derecho a mantener el control sobre determinada información personal y la libertad de actuar sin interferencia externa"*.

2.1.4. Confidencialidad:

Según A. Buckovich (1999) la confidencialidad hace referencia a la *"condición otorgada a los datos o información que indica que es sensible por alguna razón, y por lo tanto debe ser protegida contra el robo, la divulgación, el uso indebido, o ambos, y debe ser divulgada solamente a personas u organizaciones autorizadas con un necesita saber."*

2.2. Protección de privacidad:

Contrariamente a lo que solemos pensar, la **información personal** está muy relacionada con nuestra privacidad, ya que pueden determinar, por ejemplo, nuestro estilo de vida y nuestros comportamientos de compra (lugar de residencia, el enfoque de compras, ocio), nuestra intimidad (los productos que son más aficionados), o en nosotros mismos (discusiones en los foros, la pertenencia a un sindicato o a un partido político). Esta información circula en un mundo sin fronteras en nuestra actualidad, que podría ser perjudicial en un momento dado, ya que el derecho al olvido no es obvio.

Además, si no tenemos cuidado, será fácil dañar irreversiblemente nuestro espacio íntimo y nuestros derechos fundamentales.

3. Enfoque para estabilizar la privacidad y la confidencialidad de los datos personales:

3.1. Vista general:

Internet y el acceso a la Web continúan creciendo a ritmos agigantados en todo el mundo. Las personas están utilizando la Web para la comunicación, como un recurso de información, así como una forma de comprar productos y servicios. Al mismo tiempo, los avances tecnológicos han hecho posible que las organizaciones rastreen los movimientos de los consumidores en la Web. Los propietarios de sitios web pueden recopilar, almacenar, transferir y analizar grandes cantidades de datos de las personas que visitan sus sitios web. Dado que el Internet trasciende las fronteras geográficas, las personas que navegan o compran productos a través de la Web tienen poca idea de dónde están alojados los sitios web que visitan

La cuestión de la privacidad de los datos personales es a menudo el principal temor que enfrentan las personas que utilizan Internet. En este apartado examinamos los diferentes enfoques de privacidad y la confidencialidad de los datos personales existentes.

La preocupación sobre la privacidad de las personas y la confidencialidad de sus datos obliga a los responsables a la renovación y actualización de las estructuras legislativas que tratan el tema de la protección de datos.

Antes de hablar sobre el enfoque legislativo, hay que hablar sobre los sistemas organizativos que tienen la especialidad y la responsabilidad para producir soluciones estratégicas y estándares legislativos y organizativos que garantizan la protección de las personas a nivel de su privacidad y la confidencialidad de sus datos.

Por último, viene el enfoque técnico que tiene un papel importante, el de preparar una plataforma legislativa y organizativa, donde se desarrollan y aplican las distintas metodologías tecnológicas para proteger los datos de las personas frente el uso no autorizado, o en el caso de robo o violación de datos (Steinke, 2002).

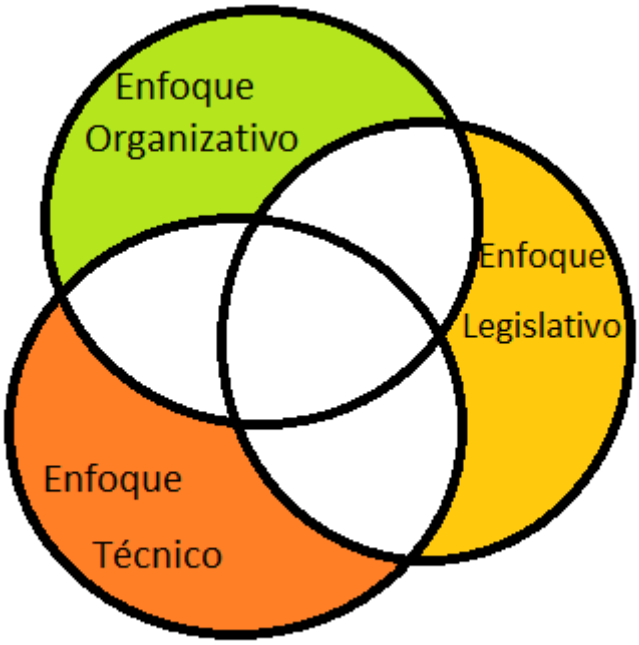


Ilustración 2. Los principales Enfoque para estabilizar la privacidad y la confidencialidad

3.2. Enfoque Organizativo

3.2.1. Directiva de Protección de Datos EU:

Para estandarizar la protección de la privacidad de los datos, en 1995 la Unión Europea promulgó la Directiva de Protección de Datos de la Unión Europea (DPP, 1995) que entró en vigor en 1998. Algunos de los requisitos de la directiva son:

- Una organización debe informar a las personas acerca de los propósitos para los cuales recopila y usa información sobre ellos, cómo comunicarse con la organización y los tipos de terceros a los cuales divulga la información.
- Una organización debe ofrecer a los individuos la oportunidad de optar por si su información puede ser utilizada para un propósito aparte de aquel para el cual fue originalmente recopilada. Para información sensible como condiciones médicas, origen racial o étnico, opiniones políticas, etc., los consumidores deben tener una opción específica de elección antes de que la información sea revelada a un tercero.

- Cada organización que gestione datos personales debe tomar medidas razonables para garantizar su seguridad e integridad.

- Las personas deben tener acceso a su información personal y ser capaces de corregirla. También deben existir mecanismos para garantizar el cumplimiento de la Directiva, el recurso a las personas afectadas por el incumplimiento y las consecuencias para la organización cuando no se respete la Directiva.

- Se prohíbe a las corporaciones y los gobiernos utilizar virtualmente cualquier registro personal para cualquier propósito que no sea el original, sin permiso explícito. (Esta es probablemente una de las principales razones por las que las empresas europeas no hacen tanto marketing de bases de datos o dirigidas a individuos basándose en sus perfiles demográficos como las empresas estadounidenses.)

- La directiva también requiere la creación de agencias gubernamentales de protección de datos, el registro de bases de datos con esas agencias y, a veces, incluso la aprobación previa antes de que ciertos datos puedan ser procesados.

- Los datos personales de los ciudadanos de la UE sólo podrán transferirse a países que no sean del bloque de 15 naciones que adopten estas normas o se considere que proporcionan una "protección adecuada" para los datos.

3.2.2. Agencia Española de Protección de Datos (AEPD):

La AEPD fue creada por el Real Decreto 428/1993, de 26 de marzo, modificado por la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal. Esta enmienda aplicaba la Directiva 95/46 / CE. La agencia fue creada en el contexto de la Constitución española de 1978, artículo 18.4, en la que se señala que "la ley restringirá el uso de la informática para proteger el honor y la privacidad de personas y familias de los ciudadanos españoles, así como el ejercicio pleno de sus derechos ", tal como fue elaborado por la Ley Orgánica 5/1992.

3.2.2.1. Actividades de AEPD:

La AEPD es una autoridad de derecho público que goza de "absoluta independencia de la Administración Pública". Y tiene como responsabilidad (AEPD, 2021):

- Información sobre sus actividades y el derecho a la protección de datos personales (incluyendo 450 entrevistas y 850 "impactos" en los medios).
- Asistencia directa en respuesta a consultas ciudadanas (47.741 en 2007).
- Procedimientos para proteger los derechos de las personas de acceder, rectificar, cancelar y objetar. Los más comunes son procesos de cancelación (62%) y acceso (32%).
- Registro de sistemas de archivo (1.017.266 entradas totales).
- Procedimientos de inspección y sanción (399 procedimientos de sanción resueltos con multas por importe de 19,6 millones de euros).
- Promoción del Real Decreto 1720/2007.
- Cooperación con agencias internacionales y con las comunidades autónomas de Cataluña , Euskadi y Madrid.

- Evaluación de los riesgos emergentes, incluidos los datos personales en Internet, la generalización de los sistemas de videovigilancia, la supervisión de los empleadores por la videovigilancia, la biometría y el uso de Internet, y la intensificación de los flujos internacionales de datos.

En respecto a este último punto, la AEPD abogó por: (AEPD, 2021)

- Desarrollar procedimientos que permitan la protección del derecho de autor de una manera compatible con el derecho fundamental a la protección de datos.
- Regulación de la publicación anónima de sentencias dictadas por los Tribunales de Justicia
- Regulación de los sistemas internos de denuncia de irregularidades a disposición de los trabajadores dentro de las empresas, delineando las actividades en las que puede ser necesario establecer estos sistemas y garantizando la confidencialidad de los que informan y los derechos de los denunciados.
- Elaboración de planes específicos de política pública para la protección de menores en Internet.
- El aumento de precaución a fin de evitar que el intercambio no deseado de los datos personales sensibles en Internet a través de la red P2P.
- Fomento de la autorregulación entre los medios de comunicación para garantizar la privacidad y la protección de los datos personales, fomentando un mayor respeto por el uso en relación con las disposiciones de protección de datos.
- Acciones de orientación ciudadana sobre el uso de garantías de confidencialidad para los destinatarios de correos electrónicos.
- Plan de Fomento de “Buenas Prácticas” en materia de garantía de la privacidad en los Boletines y Diarios Oficiales, mediante la adopción de medidas que, sin afectar su finalidad, limitarán la recopilación de información personal por parte de los buscadores de Internet.
- Estrategia Local destinada a conformar la instalación de cámaras de control de tráfico a las disposiciones sobre protección de datos personales.

3.3. Enfoque Legislativo

3.3.1. La Protección de datos en la UE

3.3.1.1. Reforma de las Normas de protección de datos de la UE en enero de 2012:

La Comisión Europea presentó su reforma de la protección de datos de la UE en enero de 2012 para hacer que Europa encaje en la era digital. Más del 90% de los europeos dicen que quieren los mismos derechos de protección de datos en toda la UE - e independientemente de dónde se procesen sus datos.
El Reglamento es un paso esencial para reforzar los derechos fundamentales de los ciudadanos en la era digital y facilitar los negocios mediante la simplificación de las normas para las empresas en el mercado único digital. Una sola ley también eliminará la actual fragmentación y las costosas cargas administrativas, lo que supondrá un ahorro para las empresas de unos 2.300 millones de euros al año. La Directiva relativa al sector de la policía y la justicia penal protege el derecho fundamental de los ciudadanos a la protección de datos cuando los datos personales son utilizados por las autoridades penales. En particular, velará por que los datos personales de las víctimas, testigos y sospechosos de delitos estén debidamente protegidos y facilitará la cooperación transfronteriza en la lucha contra la delincuencia y el terrorismo (DGJC-CE, 2016).

1.1.1.1. El mercado único digital-EU (DSM)

El propósito original de la Unión Europea era fomentar el comercio entre los Estados miembros eliminando barreras y fomentando la libre circulación de bienes, servicios y personas. Un área donde se percibe que las barreras permanecen son los bienes y servicios digitales. Ahora están en la mira de la Presidencia de Juncker, y un pilar central de su plataforma al tomar el poder en noviembre de 2014 fue introducir un "Mercado Único Digital".
El 6 de mayo de 2015, se anunció la estrategia de mercado único digital (DSM) de la Comisión Europea (Comisión). El ámbito de aplicación de la estrategia DSM es ambicioso, abarcando una amplia gama de ámbitos como el comercio electrónico transfronterizo, la prevención del Geobloqueo injustificado, la reforma de los marcos jurídicos de derechos de autor, audiovisuales y comunicaciones, la competencia y el IVA. La estrategia estableció también un calendario estricto para lograr los cambios deseados (TaylorWessing, 2015).

1.1.1.2. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo

El Reglamento General de Protección de Datos (GDPR) (Reglamento 2016/679) es un reglamento por el que el Parlamento Europeo, el Consejo de la Unión Europea y la Comisión Europea tienen la intención de reforzar y unificar la protección de datos para todos los individuos dentro de la Unión Europea (UE). También se ocupa de la exportación de datos personales fuera de la UE.

El GDPR tiene como objetivo principal dar control a los ciudadanos y residentes sobre sus datos personales y simplificar el entorno regulador de los negocios internacionales mediante la unificación de la regulación dentro de la UE. (CEU, June 2015)
Cuando el GDPR surta efecto, sustituirá a la Directiva de protección de datos (oficialmente Directiva 95/46 / CE) [2] de 1995. El Reglamento fue adoptado el 27 de abril de 2016. Se convierte en ejecutivo a partir del 25

de mayo de 2018 tras una transición de dos años y, a diferencia de una directiva, no obliga a los gobiernos nacionales a aprobar ninguna legislación autorizada, por lo que es directamente vinculante y aplicable (Blackmer, 2016).

❖ Alcance:

La aplicación del Reglamento está dirigida al responsable del tratamiento de datos (organización que recopila datos de los residentes de la UE) o el procesador (organización que procesa datos en nombre del controlador de datos, por ejemplo, los proveedores de servicios en la nube) o el propietario de datos (persona). Además, el Reglamento también se aplica a las organizaciones basadas fuera de la Unión Europea si recogen o procesan datos personales de residentes de la UE.

Según la Comisión Europea, "*datos personales son cualquier información relativa a un individuo, ya se trate de su vida privada, profesional o pública. Puede ser cualquier cosa de un nombre, una dirección de casa, una foto, una dirección de correo electrónico, un banco de información sobre los sitios web de redes sociales, información médica o la dirección IP de un ordenador*" (Commission, 2012). El reglamento no se aplica al tratamiento de datos personales para actividades de seguridad nacional o aplicación de la ley; sin embargo, el paquete de reforma de la protección de datos incluye una directiva de protección de datos para el sector de la policía y la justicia penal, que establece normas sólidas sobre el intercambio de datos personales a nivel nacional, europeo e internacional.

❖ Las principales modificaciones relativas a los datos personales son las siguientes

- **Ampliación de la definición de datos personales y la provisión de nuevas definiciones.** "El número de identificación, los datos de localización están expresamente incluidos en el significado de los datos personales. Además, se proporcionan nuevas definiciones para nociones tales como: "violación de datos personales", "datos genéticos", "datos biométricos", "datos relativos a la salud", "grupo de empresas", "establecimiento principal", "niño" y otros.

- **Normas más estrictas para obtener el consentimiento del interesado.** El Reglamento prevé expresamente que el consentimiento debe ser "explícito" y que la aceptación tiene que ser " en una declaración o mediante una acción positiva que no deja lugar a equivoque". La inactividad o el silencio no pueden representar el consentimiento. La declaración de consentimiento tiene que ser distinta de las declaraciones relativas a otros aspectos (por ejemplo, la aceptación de términos y condiciones generales). El consentimiento puede ser retirado por el interesado. En lo que respecta al consentimiento del niño, "cuando el menor tenga menos de 16 años, dicho trámite sólo será lícito si y en la medida en que el titular de la responsabilidad parental sobre el niño autoriza o da su consentimiento". Los Estados miembros pueden rebajar la edad límite, a un mínimo de 13 años.

- **Ampliación significativa del área geográfica de aplicación de la normativa de datos personales de la Unión Europea.** El Reglamento se aplica al tratamiento de datos personales "en el contexto de las actividades de un establecimiento de un responsable del tratamiento o de un transformador en la Unión", independientemente de si el tratamiento de datos se realiza realmente dentro o fuera de la Unión Europea. Además, el Reglamento se aplica a "el tratamiento de datos personales de personas físicas que residen en la Unión por un responsable no establecido en la Unión, cuando las actividades de tratamiento estén relacionadas con: i) la oferta de bienes o servicios a dichas personas; ii) al control de su comportamiento en la medida en que su comportamiento se produzca dentro de la Unión", por un controlador que no esté establecido en la Unión Europea.

- **Provisión de nuevos derechos para los interesados**. Los titulares de los datos tienen nuevos derechos en virtud del Reglamento, incluido el derecho al olvido (la solicitud de borrar los datos personales) o la rectificación, el derecho a la portabilidad de los datos (es decir, el derecho a transmitir los datos a otro responsable) derecho a oponerse al tratamiento de datos personales, incluso mediante perfiles. Los controladores tienen nuevas obligaciones correlacionadas con los derechos de los interesados.

- **Perfiles.** El Reglamento estipulaba una nueva restricción para los controladores en lo que respecta a la adopción de decisiones basadas en el procesamiento automatizado, si dicho procesamiento "produce efectos jurídicos relacionados con esta persona o los afecta significativamente". No obstante, dicho tratamiento se autoriza con arreglo al consentimiento expreso del interesado, si es necesario para celebrar un contrato o si está expresamente autorizado por una legislación de la Unión o de un Estado miembro.

- **Obligación de designar a un oficial de protección de datos.** Si (i) el procesamiento es realizado por una empresa que emplea 250 o más personas; o ii) las actividades principales del responsable del tratamiento o del procesador consisten en operaciones de tratamiento que, por su naturaleza, alcance y / o finalidad, obliguen a un control periódico y sistemático de los interesados, el responsable del tratamiento está obligado a designar un (empleado o proveedor de servicios). El oficial de protección de datos es responsable ante la alta dirección del controlador.

- **Obligación de notificar a la autoridad supervisora en caso de incumplimiento de la seguridad de los datos personales.** Los controladores tienen que notificar a la autoridad supervisora dentro de las 72 horas siguientes a tomar conocimiento de una violación. Cuando el incumplimiento de los datos personales pueda afectar desfavorablemente a la protección de los datos personales o la privacidad de la persona afectada, el responsable del tratamiento también deberá comunicar la infracción de datos personales a la persona afectada sin demora injustificada.

- **Obligaciones, evaluación de impacto, protección implícita**. Los controladores deben demostrar que respetan el Reglamento, incluso mediante la aplicación de políticas transparentes y fácilmente accesibles sobre el tratamiento de los datos personales y el ejercicio de los derechos del interesado. El Reglamento establece que los controladores y procesadores deben llevar a cabo una evaluación del impacto de las operaciones de tratamiento con datos personales si el tratamiento presenta un riesgo específico para los derechos y libertades del interesado. Además, en virtud del Reglamento, los responsables del tratamiento y los transformadores deben aplicar medidas técnicas y organizativas de tal forma que el tratamiento cumpla el nivel de seguridad de acuerdo con los riesgos que entrañe dicho tratamiento y con los datos personales que deben protegerse.

- **Sanciones significativas.** Las multas de hasta 20 millones de euros, es decir, un 4% del volumen de negocios mundial anual, pueden aplicarse a los controladores y operadores por incumplimiento del Reglamento. La sanción se impondrá caso por caso, teniendo en cuenta determinados criterios, tales como: la naturaleza, la gravedad y la duración de la infracción, el carácter intencional o negligente de la infracción, el grado de responsabilidad de la persona física o jurídica, así como de los incumplimientos anteriores de los mismos, las medidas y procedimientos técnicos y organizativos aplicados y el grado de cooperación con la autoridad de supervisión para remediar la infracción (Blackmer, 2016).

3.4. Enfoque Técnico:

Vista general:

En este apartado nos concentramos sobre las vistas técnicas posibles para avanzar en nuestro estudio sobre los enfoques para estabilización de la confidencialidad y la privacidad de los datos personales, teniendo en cuenta las recomendaciones que vienen en los distintos artículos Del GDPR (Reglamento General de Protección de datos de EU) y lo tomamos como referencia principal.
La seguridad de los datos desempeña un papel prominente en el nuevo Reglamento General de Protección de Datos (GDPR), que refleja su relación simbiótica con los modernos regímenes generales de privacidad.

En comparación con la Directiva 95/46/CE, el GDPR impone obligaciones más estrictas a los procesadores y controladores de datos con respecto a la seguridad de los datos, ofreciendo al mismo tiempo más orientación sobre las normas de seguridad adecuadas.

De conformidad con el artículo 32, al igual que el artículo 17 de la Directiva, los controladores y procesadores están obligados a "aplicar las medidas técnicas y organizativas apropiadas" teniendo en cuenta "el estado de la técnica y los costos de ejecución" y "la naturaleza, alcance, contexto", no obstante, el GDPR ofrece sugerencias específicas sobre qué tipos de medidas de seguridad podrían considerarse "apropiadas para el riesgo, "Incluyendo:

- La seudonimización y cifrado de datos personales.
- La capacidad de garantizar la confidencialidad, la integridad, la disponibilidad de los sistemas y servicios de procesamiento.
- La capacidad de restaurar la disponibilidad y el acceso a los datos personales de manera oportuna en caso de un incidente físico o técnico.
- Un proceso para evaluar regularmente la efectividad de las medidas técnicas y organizativas para garantizar la seguridad del procesamiento.

3.4.1. Desafíos de la gestión de la seguridad de la información:

Debido a la naturaleza multidimensional de la gestión de la seguridad de la información, los desafíos que enfrentan las organizaciones son tan complejos y conflictivos como sus objetivos para la gestión de la seguridad de la información. Requiere que la dirección de una organización se ocupe de la seguridad de la información de una manera completa y global. El principal desafío es alcanzar los objetivos generales de la seguridad de la información, que incluyen:

- Salvaguardar información confidencial, crítica y propietaria del acceso no autorizado, divulgación o modificación;
- Proteger los sistemas de información y apoyar los recursos informáticos de la pérdida, daño y destrucción;
- Proporcionar a la dirección de la organización una garantía razonable en cuanto a la integridad, confidencialidad y disponibilidad de la información y los activos de la información (las herramientas o utilidades para el desarrollo y soporte de los sistemas de información);

- Reconociendo y adoptando todas las regulaciones y leyes legales concernientes a la confidencialidad, disponibilidad e integridad de la información crítica (Hong K.S, 2005)

3.4.2. Gestión de riesgo para sistemas de información:

Un proceso de detección, análisis y reducción de los riesgos de TI, que pueden afectar o ya ha impactado al sistema de información de la empresa. La principal tarea de gestión de riesgos está basada en el conjunto de medidas (controles), que permiten reducir el nivel de riesgo a un valor asequible.

Ilustración 3. Gestión de Riesgos

Las necesidades de aplicar la gestión de riesgo vienen de que las tecnologías de la información aumentan enormemente las posibilidades de negocio. Pero es obvio que nuevas posibilidades inevitablemente se relacionan con nuevas amenazas, que pueden conducir a las pérdidas financieras y de otro tipo en la empresa.

3.4.2.1. La Evaluación De Riesgo:

La evaluación de riesgos es el primer proceso en la metodología de gestión de riesgos. Las organizaciones utilizan la evaluación del riesgo para determinar el alcance de la amenaza potencial y el riesgo asociado con un sistema de TI a lo largo de su SDLC. El resultado de este proceso ayuda a identificar los controles apropiados para reducir o eliminar el riesgo durante el proceso de mitigación del riesgo, como se analiza en la Sección (2.4.2.2).

El riesgo es una función de la probabilidad de que una fuente dada de amenaza ejerza una vulnerabilidad potencial particular y el impacto resultante de ese evento adverso en la organización. Para determinar la probabilidad de un evento adverso futuro, las amenazas a un sistema de TI deben analizarse junto con las vulnerabilidades potenciales y los controles existentes para el sistema de TI.

El impacto se refiere a la magnitud del daño que podría ser causado por el ejercicio de una amenaza de una vulnerabilidad. El nivel de impacto se rige por los impactos potenciales y, a su vez, produce un valor relativo para los activos y recursos de TI afectados (por ejemplo, la criticidad y la sensibilidad de los componentes y datos del sistema de TI). La metodología de evaluación de riesgos abarca nueve etapas principales:

3.4.2.1.1. Caracterización del sistema:

Al evaluar los riesgos para un sistema de TI, el primer paso es definir el alcance del esfuerzo. En este paso, se identifican los límites del sistema de TI, junto con los recursos y la información que constituyen el sistema. La caracterización de un sistema de TI establece el alcance del esfuerzo de evaluación de riesgos, delimita los límites operativos de autorización (o acreditación) y proporciona información (por ejemplo, hardware, software, conectividad del sistema y personal responsable de división o soporte) esencial para definir el riesgo.

3.4.2.1.2. Identificación de la vulnerabilidad:

El objetivo de este paso es identificar las potenciales fuentes de amenazas y compilar una declaración de amenaza que enumere posibles fuentes de amenazas que sean aplicables al sistema informático que se está evaluando. Una fuente de amenaza se define como cualquier circunstancia o evento con el potencial de causar daño a un sistema de TI. Las amenazas comunes pueden ser naturales, humanos o ambientales.

Al evaluar las fuentes de amenazas, es importante considerar todas las posibles amenazas-fuentes que podrían causar daño a un sistema de TI y su entorno de procesamiento.
Tomamos como ejemplo, una declaración de amenaza para un sistema de TI ubicado en un desierto puede no incluir "inundación natural" debido a la baja probabilidad de que ocurra tal evento, las amenazas ambientales como el estallido de una cañería que podría inundar rápidamente una sala de computadoras y causar daños a los recursos de TI de una organización.

Los seres humanos pueden ser fuentes de amenaza a través de actos intencionales, como ataques deliberados de personas maliciosas o empleados descontentos, o actos no intencionados, como negligencia y errores. Un ataque deliberado puede ser un intento malintencionado de obtener acceso no autorizado a un sistema de TI (por ejemplo, a través de la adivinación de contraseñas) para comprometer la integridad, la disponibilidad o la confidencialidad del sistema y de los datos, o ser benigno, pero no obstante intencionado, intentan eludir la seguridad del sistema.

3.4.2.1.3. Identificación de la vulnerabilidad

Una vulnerabilidad es un fallo o debilidad en los procedimientos de seguridad del sistema, diseño, implementación, o controles internos que podrían ser ejercidos y dan lugar a un fallo de seguridad o una violación de la política de seguridad del sistema.

El análisis de la amenaza a un sistema de TI debe incluir un análisis de las vulnerabilidades asociadas con el entorno del sistema. El objetivo de este paso es desarrollar una lista de vulnerabilidades del sistema (fallos o debilidades) que podrían ser explotadas por las posibles fuentes de amenazas.

- **Métodos para identificar las vulnerabilidades de Sistemas de TI:**

Los métodos recomendados para identificar las vulnerabilidades del sistema son el uso de fuentes de vulnerabilidad, el rendimiento de las pruebas de seguridad del sistema y el desarrollo de una lista de

verificación de requisitos de seguridad. Cabe señalar que los tipos de vulnerabilidades que existirán y la metodología necesaria para determinar si las vulnerabilidades están presentes generalmente variarán según la naturaleza del sistema informático y la fase en la que se encuentre, en el SDLC (El ciclo de vida de desarrollo de sistemas):

- Si el sistema de TI todavía no se ha diseñado, la búsqueda de vulnerabilidades debería centrarse en las políticas de seguridad de la organización, los procedimientos de seguridad planificados y las definiciones de requisitos del sistema y los análisis de productos de seguridad de los vendedores o desarrolladores (por ejemplo, documentos técnicos).

- Si el sistema de TI está siendo implementado, la identificación de vulnerabilidades debe ampliarse para incluir información más específica, como las características de seguridad planificadas descritas en la documentación de diseño de seguridad y los resultados de la prueba y evaluación de certificación del sistema.

- Si el sistema de TI está operativo, el proceso de identificación de vulnerabilidades debe incluir unos análisis de las características de seguridad del sistema de TI y los controles de seguridad, técnicos y de procedimiento, utilizados para proteger el sistema.

❖ **Prueba de seguridad del sistema:**

Los métodos proactivos, que emplean las pruebas del sistema, se pueden usar para identificar las vulnerabilidades del sistema de manera eficiente, dependiendo de la criticidad del sistema de TI y los recursos disponibles (por ejemplo, fondos asignados, tecnología disponible, personas con experiencia para realizar la prueba). Los métodos de prueba incluyen:

- Herramienta automatizada de escaneo de vulnerabilidades:

  La herramienta de escaneo de vulnerabilidades automatizadas se utiliza para escanear un grupo de hosts o una red para servicios vulnerables conocidos (por ejemplo, el sistema permite el Protocolo anónimo de transferencia de archivos [FTP], la transmisión de correos electrónicos). Sin embargo, debe tenerse en cuenta que algunas de las vulnerabilidades potenciales identificadas por la herramienta de exploración automatizada pueden no representar vulnerabilidades reales en el contexto del entorno del sistema. Por ejemplo, algunas de estas herramientas de análisis califican vulnerabilidades potenciales sin considerar el entorno y los requisitos del sitio. Algunas de las "vulnerabilidades" marcadas por el software de escaneo automatizado pueden no ser vulnerables para un sitio en particular, pero pueden configurarse de esa manera porque su entorno lo requiere. Por lo tanto, este método de prueba puede producir falsos positivos.

- Prueba de seguridad y evaluación (ST & E):

  Es otra técnica que se puede usar para identificar las vulnerabilidades del sistema de TI durante el proceso de evaluación de riesgos. Incluye el desarrollo y la ejecución de un plan de prueba (por ejemplo, secuencia de comandos de prueba, procedimientos de prueba y resultados esperados de las pruebas). El propósito de las pruebas de seguridad del sistema es probar la efectividad de los controles de seguridad de un sistema de TI, ya que se han aplicado en un entorno operativo. El objetivo es garantizar que los controles aplicados cumplan con las especificaciones de seguridad aprobadas para el software y el hardware e implementar la política de seguridad de la organización o cumplir con los estándares de la industria.

➢ Pruebas de penetración:

Las pruebas de penetración se pueden utilizar para complementar la revisión de los controles de seguridad y garantizar que se protejan las diferentes facetas del sistema de TI. Las pruebas de penetración, cuando se emplean en el proceso de evaluación de riesgos, se pueden usar para evaluar la capacidad de un sistema informático para resistir intentos intencionales de eludir la seguridad del sistema. Su objetivo es probar el sistema de TI desde el punto de vista de una fuente de amenaza e identificar posibles fallos en los esquemas de protección del sistema de TI.

3.4.2.1.4. Análisis de control

El objetivo de este paso es analizar los controles que la organización ha implementado o está planeando implementar para minimizar o eliminar la probabilidad de que una amenaza ejerza una vulnerabilidad del sistema.

Para obtener una calificación general de probabilidad que indique, si se pueda ejercer una vulnerabilidad potencial dentro de la construcción del entorno de amenaza, debe considerarse la implementación de controles actuales o planificados.

➢ **Métodos de control:**

Los controles de seguridad abarcan el uso de métodos técnicos y no técnicos. Los controles técnicos son salvaguardas que se incorporan al hardware, software o firmware de la computadora (por ejemplo, mecanismos de control de acceso, mecanismos de identificación y autenticación, métodos de encriptación, software de detección de intrusiones). Los controles no técnicos son los controles operacionales y de gestión, como las políticas de seguridad; procedimientos operacionales; y personal, seguridad física y ambiental.

➢ **Técnica de análisis de control:**

El desarrollo de una lista de verificación de los requisitos de seguridad o el uso de una lista de verificación disponible será útil para analizar los controles de manera eficiente y sistemática. La lista de verificación de los requisitos de seguridad se puede usar para validar el incumplimiento de seguridad y el cumplimiento. Por lo tanto, es esencial actualizar dichas listas de verificación para reflejar los cambios en el entorno de control de una organización (por ejemplo, cambios en políticas de seguridad, métodos y requisitos) para garantizar la validez de la lista de verificación.

3.4.2.1.5. Determinación de la probabilidad:

Para obtener una calificación general de probabilidad que indique la probabilidad de que se pueda ejercer una vulnerabilidad potencial dentro de la construcción del entorno de amenaza asociado, se deben considerar los siguientes factores de gobierno:

- Motivación y capacidad de la fuente de la amenaza.
- Naturaleza de la vulnerabilidad.
- Existencia y efectividad de los controles actuales.

La probabilidad de que una vulnerabilidad potencial pueda ser ejercida por una determinada fuente de amenaza puede describirse como alta, media o baja. La Tabla 1 describe estos tres niveles de probabilidad.

Nivel de probabilidad	Definición de probabilidad
Alta	Una fuente de amenaza muy fuerte, frente unos controles para detectar la vulnerabilidad que son ineficaces.
Media	La fuente de amenaza esta fuerte y es capaz, pero existen controles que pueden impiden el ejercicio exitoso de la vulnerabilidad.
Baja	La fuente de amenaza carece de motivación o capacidad, frente la existencia de unos controles para prevenir o, al menos, impedir de manera significativa la vulnerabilidad de ser ejercida.

Tabla 1. Niveles de probabilidad

3.4.2.2. Sistema de Información de Gestión de Riesgos (RMIS: Risk Management Information System)

Un Sistema de Información de Gestión de Riesgos (RMIS) es un sistema informático integrado de información que se utiliza para ayudar a los responsables de la toma de decisiones a evaluar los riesgos y hacer un seguimiento de toda la información relevante. Esta información incluye la exposición al riesgo, las medidas de protección de riesgos y la gestión del riesgo. Ejemplos de información almacenada incluyen medidas de control de pérdidas, valores de propiedad, registros de reclamaciones anteriores y pólizas de seguro relevantes.

Además de servir como una forma de almacenar información, el sistema también debe ser capaz de proporcionar informes relevantes para los tomadores de decisiones.
Los sistemas actuales deben ser flexibles para permitir el cambio, y también deben ser accesibles desde diferentes ubicaciones o incluso tipos de dispositivos. Tradicionalmente, estos sistemas de gestión de riesgos se centraban en la información de víctimas, y esto incluiría información sobre edificios, vehículos y responsabilidad, pero los sistemas actuales podrían comenzar a enfocarse en otras áreas de riesgos, como la exposición en línea (Rhodes, 2015).

3.4.3. Nivel de Madurez:

*Ilustración 4. Niveles de Madurez (*Freitas, 2010)

El concepto de modelos de madurez se está aplicando cada vez más en el campo de los sistemas de información como un enfoque para el desarrollo organizacional o como medio de evaluación organizacional. Cualquier marco sistemático para llevar a cabo la evaluación comparativa y la mejora del rendimiento puede ser considerado como un modelo y si tiene procesos de mejora continua se puede considerar un modelo de madurez. La madurez implica un sistema completo. Generalmente, en la literatura constitutiva, la madurez implica un sistema perfeccionado o explícitamente definido, administrado, medido y controlado. También es un progreso en la demostración de una capacidad específica o en la realización de un objetivo desde una fase inicial hasta una etapa final deseada (F.Saleh, 2011).

FASE II: SEUDONIMIZACIÓN Y CRIPTOGRAFÍA

1. Introducción:

Apartar del mes de mayo de 2018, al menos en Europa, un punto de inflexión para la privacidad de los datos y la protección de la información personal. En este apartado, nos centraremos presentar las distintas técnicas para mejorar la seguridad y la protección de datos en general y los datos personales en especial que están almacenados en los sistemas de información de un organismo, describiendo los principales métodos de desidentificación de datos personales, como la seudonimización, la anonimización y la encriptación.

Las organizaciones suelen ejecutar un sistema de TI de producción en el que se encuentran los datos personales, a saber, los correos electrónicos de los clientes, nombres, direcciones, cuentas bancarias e incluso información de tarjetas de crédito. Del mismo modo, la base de datos puede contener información similar sobre empleados, proveedores y socios.

Los datos confidenciales pueden escaparse de estos sistemas, pero es una práctica buena y común restringir el acceso al sistema de TI, y es difícil obtener acceso no autorizado a las bases de datos de producción relevantes. Sin embargo, los sistemas de soporte y las aplicaciones a menudo están vinculados a estos

sistemas de producción, por lo que pueden exportar datos de las bases de datos de producción, incluidos los datos personales. Normalmente, estos sistemas de soporte no tienen reglas estrictas, y los datos personales confidenciales a menudo se pueden encontrar en sistemas de almacenamiento no seguros, como en una unidad de disco portátil común. Desde allí, la fuga de datos es muy fácil y las organizaciones pierden el control de estos datos. Un ejemplo común es en la exportación de datos de producción al departamento de marketing o incluso a una agencia externa.

➢ Los datos que no tenemos no pueden tener fugas:

El principio básico es que solo debemos recopilar, procesar y almacenar datos personales que realmente se necesitan. La organización debe completar el inventario de datos personales, identificar los campos de datos que contienen información personal confidencial y explorar sus motivos para recopilar y almacenar esta información. Si el riesgo de violación de datos es mayor que el valor agregado, entonces es apropiado dejar de recopilar y almacenar esta información. Un ejemplo podría ser una lista de correo para una campaña de boletín informativo. Dicha lista puede representar un alto riesgo para la organización en comparación con su valor agregado. Además, durante el proceso de inventario de datos personales, puede descubrir que el propósito original de la recopilación de datos ya no es aplicable, pero los datos aún se están recopilando. Esto es, por supuesto, un riesgo completamente innecesario y debe eliminarse sin demora.

➢ Des-identificar cualquier otra cosa:

Cuando se trata de información personal que necesitamos recopilar, procesar o almacenar por razones comerciales, el acceso debe limitarse únicamente a los sistemas de TI y al personal necesarios. Este enfoque debe ser automatizado y monitoreado. Además, se debe crear un registro de auditoría que almacene registros de todas las veces que se acceda a los datos. Estas medidas sirven para demostrar que la protección de los datos personales confidenciales es una preocupación seria para la organización y que no se toma a la ligera. Este enfoque también es una acción preventiva eficiente contra un posible ataque interno. Dichas medidas también serán útiles si se produce una fuga de datos y la organización necesita pruebas en su investigación.

Todas las exportaciones de datos de esta base de datos central restringida deben pasar por un proceso de desidentificación. El objetivo de la identificación es ajustar los datos personales de forma tal que ya no sea posible identificar a la persona que los origina.

Todos los campos de datos confidenciales se des-identifican para eliminar información privada y retener el valor de la información de los datos para su posterior análisis o investigación. Es necesario garantizar la protección contra la divulgación estadística, que es un método que permite la identificación de la persona mediante la investigación avanzada de datos (por ejemplo, determinar la identidad de una persona desde una ubicación geográfica conocida de esa persona en algún momento).

2. La Criptografía (encriptación o cifrado de los datos):

Introducción:

El cifrado de datos traduce los datos a otra forma o código, de modo que solo las personas con acceso a una clave secreta (formalmente llamada clave de descifrado) o una contraseña puedan leerlo. Los datos cifrados se conocen comúnmente como *texto cifrado*, mientras que los datos no cifrados se llaman *texto claro*. Actualmente, la encriptación es uno de los métodos de seguridad de datos más populares y efectivos utilizados por las organizaciones. Existen dos tipos principales de encriptación de datos: **encriptación asimétrica**, también conocida como encriptación de clave pública y **encriptación simétrica**.

❖ La función principal del cifrado de datos

El objetivo del cifrado de datos es proteger la confidencialidad de los datos digitales, ya que se almacenan en sistemas informáticos y se transmiten a través de Internet u otras redes informáticas. El estándar de cifrado de datos desactualizado (DES) ha sido reemplazado por modernos algoritmos de cifrado que juegan un papel fundamental en la seguridad de los sistemas de TI y las comunicaciones.

❖ Proceso y tipos de encriptación:

El proceso de cifrado de datos consiste en ciertos pasos. Los datos pasan a través de una fórmula matemática llamada algoritmo, que los convierte en datos encriptados llamados texto cifrado. Estos algoritmos crean una clave y luego encapsulan el mensaje con esta clave.

Hay dos tipos de encriptaciones: **asimétrica** y **simétrica**.

A. Criptografía asimétrica (criptografía de clave pública):

La criptografía asimétrica, también conocida como “criptografía de clave pública'', usa claves públicas y privadas para cifrar y descifrar datos. Las claves son simplemente números de una gran magnitud que se han emparejado pero que no son idénticos (asimétricos). Una clave, del par, se puede compartir con todos y se le llama clave pública. La otra clave del par se mantiene en secreto, a la que se le llama clave privada. Cualquiera de las claves se puede usar para encriptar un mensaje y la clave opuesta a la utilizada al encriptar el mensaje, se utiliza para el descifrado.

Muchos protocolos como SSH, OpenPGP, S/MIME y SSL/TLS dependen de la criptografía asimétrica para el cifrado y las funciones de firma digital. También se usan en programas de software, como navegadores, que necesitan establecer una conexión segura a través de una red insegura como Internet o necesitan validar una firma digital.

La intensidad de la encriptación está directamente relacionada con el tamaño de la clave y la duplicación de la longitud de la clave ofrece un aumento exponencial de la fuerza, aunque sí perjudica el rendimiento. A medida que aumenta la potencia informática y se descubren algoritmos de factorización más eficientes, también aumenta la capacidad de factorizar números cada vez mayores.

Para que el cifrado asimétrico entregue confidencialidad, integridad, autenticidad y no inseguridad, los usuarios y sistemas deben tener la certeza de que la clave pública es auténtica, que pertenece a la persona o entidad reclamada y que no ha sido manipulada o reemplazada por otra clave pública de un tercero malicioso. No hay una solución perfecta para este problema de autenticación de clave pública.

Una infraestructura de clave pública (PKI), donde las autoridades de certificación de confianza certifican la propiedad de los pares de claves y certificados, es el enfoque más común, pero los productos de cifrado se basan en la muy buena privacidad. El modelo (PGP) (incluido OpenPGP) se basa en un modelo de autenticación descentralizado denominado red de confianza, que se basa en el respaldo individual del vínculo entre el usuario y la clave pública.

Whitfield Diffie y Martin Hellman, investigadores de la Universidad de Stanford, primero propusieron públicamente la encriptación asimétrica en su artículo "Nuevas instrucciones en criptografía" (1977). El concepto había sido propuesto de forma independiente y encubierta por James Ellis varios años antes, mientras trabajaba para la Sede de Comunicaciones del Gobierno (GCHQ), la organización de inteligencia y

seguridad británica. El algoritmo asimétrico como se describe en el documento de Diffie-Hellman utiliza números elevados a potencias específicas para producir claves de descifrado.

a. Un esquema de cifrado de clave pública tiene seis partes principales:

- **Texto sin formato:** este es el mensaje de texto al que se aplica un algoritmo.
- **Algoritmo de cifrado:** realiza operaciones matemáticas para realizar sustituciones y transformaciones en el texto sin formato.
- **Claves públicas y privadas:** es un par de claves donde una se usa para el cifrado y la otra para el descifrado.
- **Texto cifrado:** este es el mensaje cifrado o codificado producido al aplicar el algoritmo al mensaje de texto claro mediante la clave.

b. El proceso de encriptación:

El proceso de encriptación de datos asimétricos tiene los siguientes pasos:

- El proceso de encriptación comienza convirtiendo el texto en un código pre-hash. Este código se genera usando una fórmula matemática.
- Este código pre-hash es encriptado por el software usando la clave privada del remitente.
- La clave privada se generaría utilizando el algoritmo utilizado por el software.
- El código pre-hash cifrado y el mensaje se vuelven a cifrar usando la clave privada del remitente.
- El siguiente paso es que el remitente del mensaje recupere la clave pública de la persona a la que está destinada esta información.
- El remitente encripta la clave secreta con la clave pública del destinatario, por lo que sólo el destinatario puede descifrarla con su clave privada, concluyendo así el proceso de cifrado.

c. El proceso de desencriptación:

El proceso de descifrado de datos asimétricos tiene los siguientes pasos:

- El destinatario usa su clave privada para descifrar la clave secreta.
- El destinatario usa su clave privada junto con la clave secreta para descifrar el código cifrado, previamente hash y el mensaje cifrado.
- El destinatario recupera la clave pública del remitente. Esta clave pública se utiliza para descifrar el código pre-hash y para verificar la identidad del remitente.
- El destinatario genera un código posterior al hash del mensaje. Si el código pos-hash es igual al código hash previo, entonces esto verifica que el mensaje no haya sido cambiado en ruta.

d. Los algoritmos asimétricos:

Los algoritmos asimétricos usan dos claves interdependientes, una para cifrar los datos y la otra para descifrarla. Esta interdependencia proporciona una serie de características diferentes, las más importantes probablemente sean las firmas digitales que se utilizan entre otras cosas para garantizar que un mensaje fue creado por una entidad particular o para autenticar sistemas o usuarios remotos.

1. *Intercambio de claves Diffie-Hellman:*

El intercambio de claves Diffie-Hellman (DHKE) es un método criptográfico para intercambiar de forma segura claves criptográficas (protocolo de acuerdo de claves) a través de un canal público (inseguro) de forma que la comunicación que se escucha por encima no revela las claves. Las claves intercambiadas se utilizan posteriormente para la comunicación cifrada (por ejemplo, utilizando un cifrado simétrico como AES).

DHKE fue uno de los primeros protocolos de clave pública, que permite a dos partes intercambiar datos de forma segura, de modo que, si alguien husmea la comunicación entre las partes, la información intercambiada puede ser revelada.

El método Diffie-Hellman (DH) es un esquema anónimo de acuerdo de claves: permite que dos partes que no se conocen previamente establezcan conjuntamente una clave secreta compartida a través de un canal inseguro.

El método DHKE es resistente a los ataques de "sniffing" (interceptación de datos), pero es vulnerable a los ataques "man-in-the-middle" (el atacante retransmite en secreto y posiblemente altera la comunicación entre dos partes).

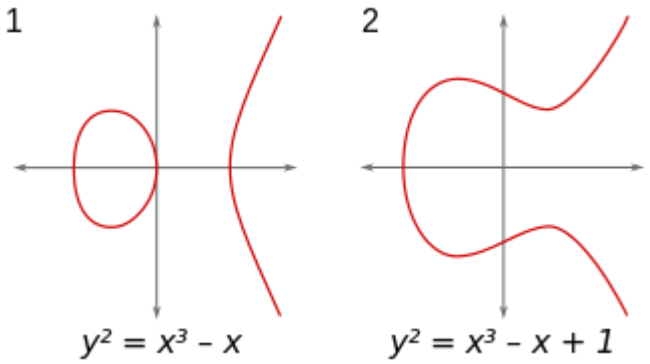
El protocolo de intercambio de claves Diffie-Hellman puede implementarse utilizando logaritmos discretos (el algoritmo clásico DHKE) o utilizando criptografía de curva elíptica (el algoritmo ECDH).

2. *Algoritmo asimétrico RSA (Rivest, Shamir y Adleman):*

RSA (Rivest-Shamir-Adleman), el algoritmo asimétrico más utilizado, está integrado en el protocolo SSL/TLS que se utiliza para proporcionar seguridad de comunicaciones a través de una red informática. RSA deriva su seguridad de la dificultad computacional de factorizar números enteros grandes que son el producto de dos números primos grandes. Multiplicar dos números primos grandes es fácil, pero la dificultad de determinar los números originales del factorizado total forma la base de la seguridad de la criptografía de clave pública. El tiempo que lleva factorizar el producto de dos primos suficientemente grandes se considera que está más allá de las capacidades de la mayoría de los atacantes. Las claves RSA suelen tener una longitud de 1024 o 2048 bits, pero los expertos creen que las claves de 1024 bits podrían romperse en un futuro próximo, razón por la cual el gobierno y la industria se están moviendo a una longitud de clave mínima de 2048 bits.

3. *Criptografía de curva elíptica (ECC):*

Criptografía de curva elíptica (ECC) está ganando adeptos en muchos expertos en seguridad como una alternativa a RSA para implementar criptografía de clave pública. ECC es una técnica de cifrado de clave pública basada en la teoría de curva elíptica que puede crear claves criptográficas más rápidas, más pequeñas y eficientes.



Representación gráfica en un sistema de coordenadas cartesianas de curvas elípticas sobre{R} (contributors, 2018).

ECC genera claves a través de las propiedades de la ecuación de curva elíptica. Para romper ECC, se debe calcular un logaritmo discreto de curva elíptica, y resulta que este es un problema significativamente más difícil que la factorización. Como resultado, los tamaños de clave ECC pueden ser significativamente más pequeños que los requeridos por RSA, pero ofrecen una seguridad equivalente con menor potencia de cálculo y uso de recursos de batería, lo que lo hace más adecuado para aplicaciones móviles que RSA

4. *Algoritmos Hash:*

Una función hash criptográfica tiene una función algo diferente a otros algoritmos criptográficos. Se usa para devolver un valor basado en un dato, un archivo o mensaje, por ejemplo. Cualquier cambio accidental o intencionado de los datos cambiará este valor hash. Un buen algoritmo hash debería hacer imposible crear una entrada inicial que produzca un valor hash específico o permitir que la entrada original se calcule a partir del valor hash. MD5 y SHA-1 fueron algoritmos hash ampliamente utilizados, pero ahora se consideran débiles y están siendo reemplazados por SHA-224, SHA-256, SHA-384 o SHA-512, a veces denominados colectivamente SHA-2.

Microsoft, Google y Mozilla han anunciado planes para eliminar el soporte SHA-1 de sus productos de navegador. Aunque todavía no se han informado ataques sobre las variantes de SHA-2, son algorítmicamente similares a SHA-1 y por lo tanto un nuevo estándar hash, SHA-3, se seleccionará de manera similar a AES en los próximos años. Como puede ver, el panorama de la criptografía está en constante cambio y para estar al tanto de los últimos desarrollos, siga las noticias y recomendaciones de organismos de normalización como el Instituto Nacional de Estándares y Tecnología.

- Firmas digitales como una aplicación de los algoritmos Hash:

Las firmas digitales se basan en criptografía asimétrica y pueden proporcionar garantías de la evidencia de origen, identidad y estado de un documento electrónico, transacción o mensaje, así como también el reconocimiento del consentimiento informado por parte del firmante. Para crear una firma digital, el software de firma (como un programa de correo electrónico) crea un hash unidireccional de los datos electrónicos que se deben firmar. La clave privada del usuario se usa para encriptar el hash, devolviendo un valor que es exclusivo de los datos hash. El hash cifrado, junto con otra información como el algoritmo hash, forma la firma digital. Cualquier cambio en los datos, incluso en un solo bit, da como resultado un valor de hash diferente. Este atributo permite a otros validar la integridad de los datos mediante el uso de la clave pública del firmante para descifrar el hash. Si el hash descifrado coincide con un segundo hash calculado de los mismos datos, prueba que los datos no han cambiado desde que se firmó. Si los dos hashes no coinciden, los datos se han manipulado de algún modo (lo que indica un fallo de integridad) o la firma se ha creado con

una clave privada que no corresponde a la clave pública presentada por el firmante (lo que indica un fallo de autenticación).

B. Criptografía simétrica:

Los algoritmos de clave simétrica son algoritmos para la criptografía que utilizan las mismas claves criptográficas tanto para el cifrado del texto simple como para el descifrado del texto cifrado. Las claves, en la práctica, representan un secreto compartido entre dos o más partes que se puede utilizar para mantener un enlace de información privada. Este requisito de que ambas partes tengan acceso a la clave secreta es uno de los principales inconvenientes del cifrado de clave simétrica, en comparación con el cifrado de clave pública (también conocido como cifrado de clave asimétrica).

Al usar esta forma de encriptación, es esencial que el emisor y el receptor tengan una forma de intercambiar claves secretas de manera segura. Si alguien conoce la clave secreta y puede descifrar el algoritmo, las comunicaciones serán inseguras. También existe la necesidad de un fuerte algoritmo de encriptación. Lo que esto significa es que, si alguien tuviera un texto cifrado y un mensaje de texto claro correspondiente, no podría determinar el algoritmo de encriptación. Hay dos métodos para atacar el cifrado convencional: fuerza bruta y criptoanálisis. La fuerza bruta es como suena; usando un método (computadora) para encontrar todas las combinaciones posibles y eventualmente determinar el mensaje de texto claro. El criptoanálisis es una forma de ataque que ataca las características del algoritmo para deducir un texto claro específico o la clave utilizada.

a. Un esquema de cifrado simétrica tiene cinco partes principales:

- **Texto sin formato:** este es el mensaje de texto al que se aplica un algoritmo.
- **Algoritmo de cifrado:** realiza operaciones matemáticas para realizar sustituciones y transformaciones en el texto sin formato.
- **Clave secreta:** esta es la entrada para el algoritmo, ya que la clave dicta el resultado encriptado.
- **Texto cifrado:** este es el mensaje cifrado o codificado producido al aplicar el algoritmo al mensaje de texto claro utilizando la clave secreta.
- **Algoritmo de descifrado:** este es el algoritmo de cifrado inverso. Utiliza el texto cifrado y la clave secreta para derivar el mensaje de texto claro.

b. Tipos de los algoritmos de clave simétrica:

Los algoritmos simétricos se pueden dividir en dos tipos: cifrado de flujo y cifrado de bloque. Los cifrados de flujo encriptan un solo bit de texto plano a la vez, mientras que los cifrados de bloque toman una cantidad de bits (típicamente 64 bits en cifrados modernos) y los encriptan como una sola unidad.

1. ***Cifrados de Flujo (Stream cipher):***

Un cifrado de flujo es un algoritmo de encriptación que encripta 1 bit o byte de texto plano a la vez. Utiliza una secuencia infinita de bits pseudoaleatorios como la clave. Para que una implementación de cifrado de flujo permanezca segura, su generador pseudoaleatorio debe ser impredecible y la clave nunca debe reutilizarse. Los cifrados de flujo están diseñados para aproximarse a un cifrado idealizado, conocido como el Pad Único.

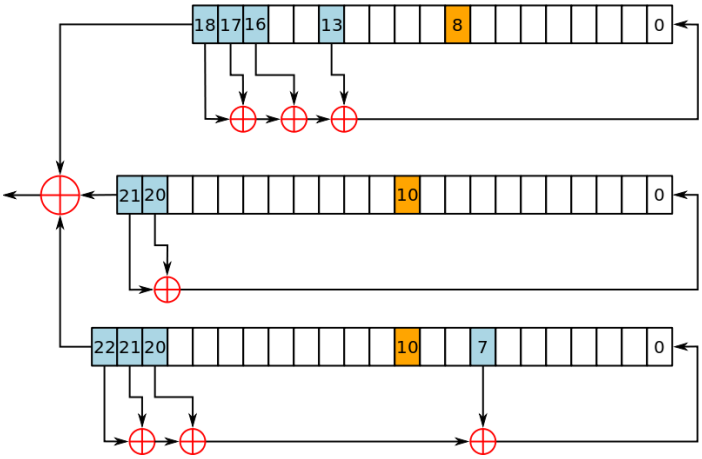


Ilustración 5. El funcionamiento del generador de flujo de claves en A5 / 1, un cifrado de flujo basado en LFSR utilizado para cifrar las conversaciones de los teléfonos móviles. (Wikipedia, 2021)

El One-Time Pad, que se supone que emplea una clave puramente aleatoria, puede alcanzar el "secretismo perfecto". Es decir, se supone que es completamente inmune a los ataques de fuerza bruta. Su problema es que, para crear dicho cifrado, su clave debe ser tan larga o incluso más larga que el texto sin formato. En otras palabras, si tiene un archivo de video de 500 MegaByte que le gustaría encriptar, necesitaría una clave de al menos 4 Gigabits de longitud.

Claramente, mientras que la información de alto secreto o asuntos de seguridad nacional pueden justificar su uso, tal cifrado sería poco práctico para el uso público cotidiano. La clave de un cifrado de flujo ya no es tan larga como el mensaje original. Por lo tanto, ya no puede garantizar el "secreto perfecto". Sin embargo, aún puede alcanzar un alto nivel de seguridad.

Entre los algoritmos populares del cifrado de flujo enumeramos los siguientes:

- RC4 - RC4, que significa Rivest Cipher 4, es la cifra de flujo más utilizada, en particular en software. También se lo conoce como ARCFOUR o ARC4. RC4 se ha utilizado en varios protocolos, como WEP y WPA (ambos protocolos de seguridad para redes inalámbricas), así como en TLS. Desafortunadamente, estudios recientes revelaron vulnerabilidades en RC4, lo que provocó que Mozilla y Microsoft recomendaran que se inhabilitara siempre que fuera posible. De hecho, RFC 7465 prohíbe el uso de RC4 en todas las versiones de TLS.

  Estos hallazgos recientes seguramente permitirán que otras cifras de flujo (por ejemplo, SALSA, Sosemanuk, Panama y muchas otras, que ya existen pero que nunca alcanzaron la misma popularidad que RC4) emergieran y posiblemente tomaran su lugar.

2. ***Cifrados de Bloque (Block cipher):***

Un cifrado de bloques es un algoritmo de encriptación que cifra un tamaño fijo de n bits de datos a la vez, conocido como bloque. Los tamaños habituales de cada bloque son 64 bits, 128 bits y 256 bits. Por ejemplo, un cifrado de bloques de 64 bits incluirá 64 bits de texto sin formato y lo encriptará en 64 bits de texto cifrado. En los casos en que los trozos de texto plano son más cortos que el tamaño del bloque, se invocan los esquemas de relleno. La mayoría de las cifras simétricas utilizadas hoy en día son en realidad del cifrados de bloque. DES, Triple DES, AES, IDEA y Blowfish son algunos de los algoritmos de encriptación utilizados comúnmente que se incluyen en este grupo y que vamos a enumerar en los siguientes puntos:

- **DES**, que significa *Data Encryption Standard*, solía ser el sistema de cifrado de bloques más popular del mundo y se utilizaba en varias industrias. Todavía es popular hoy en día, pero solo porque generalmente se incluye en discusiones históricas de algoritmos de encriptación. El algoritmo DES se convirtió en un estándar en los EE. UU. En 1977. Sin embargo, ya se ha demostrado que es vulnerable a los ataques de fuerza bruta y otros métodos criptoanalíticos. DES es un cifrado de 64 bits que funciona con una clave de 64 bits. En realidad, 8 de los 64 bits en la clave son bits de paridad, por lo que el tamaño de la clave es técnicamente de 56 bits de largo.

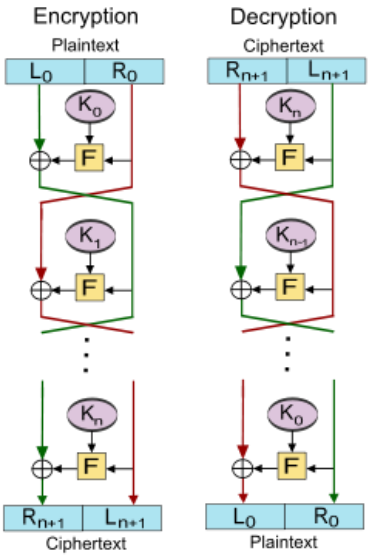


Ilustración 6. Muchos sistemas de cifrado por bloques, como DES y Blowfish, utilizan estructuras conocidas como cifras de Feistel. (Wikimedia Foundation, 2017)

- **3DES**: Como su nombre lo indica, 3DES es un cifrado basado en DES. Es prácticamente DES que se ejecuta tres veces. Cada operación DES puede usar una clave diferente, con cada clave de 56 bits de largo. Al igual que DES, 3DES tiene un tamaño de bloque de 64 bits. Aunque 3DES es mucho más potente que DES, también es mucho más lento (aproximadamente 3 veces más lento). Debido a que muchas organizaciones descubrieron que 3DES es demasiado lento para muchas aplicaciones, nunca se convirtió en el último sucesor de DES. Esa distinción está reservada para el siguiente cifrado de nuestra lista: AES.

- **AES**: un estándar del gobierno federal de EE. UU. Desde 2002, AES o Advanced Encryption Standard es posiblemente el cifrado de bloques más utilizado en el mundo. Tiene un tamaño de bloque de 128 bits y admite tres tamaños de clave posibles: 128, 192 y 256 bits. Cuanto más largo es el tamaño de la clave, más fuerte es el cifrado. Sin embargo, las claves más largas también resultan en procesos de encriptación más largos.

- **Blowfish**: este es otro cifrado de bloque popular (aunque no tan utilizado como AES). Tiene un tamaño de bloque de 64 bits y admite una clave de longitud variable que puede oscilar entre 32 y 448 bits. Una de las cosas que hace que Blowfish sea tan atractivo es que no tiene patente y es de uso libre.

- **Twofish** - Este cifrado está relacionado con Blowfish, pero no es tan popular (todavía). Es un cifrado de bloque de 128 bits que admite tamaños de clave de hasta 256 bits de longitud (John Carl, 2017).

3. Seudonimización:

Introducción:

El uso no autorizado o el uso indebido de nuestra Información de Identificación Personal (PII), como el nombre, número de seguro social, fecha de nacimiento, apellido de soltera de la madre, lugar de nacimiento, etc., puede ocasionar el robo de identidad y otros delitos relacionados con la suplantación. por no hablar de vergüenza, inconveniencia y gasto. Para aquellas organizaciones que recopilan PII, pero no las protegen, existen serias ramificaciones legales y financieras, razón por la cual más organizaciones se centran en la mitigación del riesgo de los datos.

Hospitales, agencias gubernamentales, corporaciones, instituciones financieras y otros que mantienen registros de clientes y pacientes que contienen PII deben cumplir con las leyes de privacidad de datos.

Las empresas que manejan datos actualmente se enfrentan a limitaciones impuestas por las leyes de protección de datos. El Reglamento General de Protección de Datos (GDPR) de la UE entrará en vigor en mayo de 2018 e introducirá regulaciones más firmes e impondrá sanciones más severas por no cumplir con estas leyes. Las regulaciones de protección de datos son necesarias para proteger la seguridad de los datos y la privacidad de las personas.

Según el REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO EN EL ARTÍCULO 4(DEFINICIONES – EL PUNTO 5), la *Seudonimización es "el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable;"*

Con los datos seudónimizados, dependiendo de la técnica utilizada, se puede revertir el desprendimiento de ciertos campos / identificadores del registro de datos personales. No confunda la seudonimización con encriptación, una técnica de protección de datos que también es recomendada por el GDPR, pero es algo completamente diferente. También hay una clara diferencia entre la seudonimización y la anonimización. Citando el Artículo 29 Opinión del Grupo de Trabajo 05/2014 sobre Técnicas de Anonimización: *"Los datos seudónimizados no pueden equipararse a información anonimizada ya que continúan permitiendo que un sujeto de datos individual sea individualizable y vinculable a través de diferentes conjuntos de datos".*

A lo largo de este apartado aclaramos un poco más la técnica de seudonimización y sus diferencias al comparando con otras técnicas de protección de datos.

Diferencias entre la seudonimización y la anonimización:

Está claro que en el mundo de la seguridad de los datos existe una clara distinción entre dos términos, a saber: la seudonimización y la anonimización. Estas técnicas de datos son distintas en un aspecto principal. Para los datos de seudonimización, el sujeto sustituye su identidad de forma que debe haber un caso de información adicional para reconocer al sujeto de datos, mientras que, para la anonimización, como su nombre lo indica, destruye cualquier forma de reconocer los datos del sujeto. Es bastante pertinente comprender la distinción entre estos dos términos, ya que ambas categorías de datos se clasifican en categorías muy diferentes en la regulación con la invención de GDRP.

Conclusión:

Así podemos decir que la seudonimización es un método de identificación de datos a través de la ***sustitución con un valor reversible*** *y constante, donde la anonimización es la* ***destrucción de los datos identificables****.*

A. Técnicas de anonimización y seudonimización:

a. Técnicas de anonimización:

El paso de procesamiento de la anonimización de los datos personales es el último segundo legal que estos datos entran dentro del alcance de las leyes de protección de datos de la UE como datos personales. La opinión del WP29 (Grupo de Trabajo del Artículo 29) considera varias técnicas de anonimización:

1) Noise addition:

Esto significa que se agrega una imprecisión a los datos originales. Por ejemplo, un médico puede medir su peso correctamente, pero después de agregar ruido, muestra un ancho de banda de peso de +/- 10 Kg.

2) Sustitución:

Los valores de información de los datos originales se reemplazan con otros parámetros. Por ejemplo, en lugar de indicar la altura del paciente con 175 cm, este valor se sustituye por la palabra "azul". Si la altura del paciente es de 180 cm, se registra como "amarillo". La sustitución a menudo se combina con la adición de ruido.

3) Agregación:

Para no ser señalado, una persona se agrupa con varias otras personas que comparten algunos o todos los datos personales, es decir, su lugar de residencia y edad. Por ejemplo, un conjunto de datos no captura a los habitantes de San Francisco con ciertas características, sino a los habitantes del norte de California.

K- anónimo es una forma de agregación. El proceso impide la re-identificación eliminando parte de la información, pero dejando intactos los datos para su uso futuro. Si se libera el conjunto de datos depurado y la información para cada persona no se puede distinguir de al menos k-1 personas, se considera k-anónimo. Un método de anonimato k es la supresión de datos. Puede suprimir datos reemplazando un valor con un marcador de posición. Por ejemplo, en lugar de "29 años", el valor es "X". Otro método es generalizar los datos. En lugar de "29 años", la entrada es "entre 25 y 35".

4) Privacidad diferencial:

Esto entra en juego cuando una empresa le da acceso a un tercero a un conjunto de datos anónimos. Una copia de los datos originales permanece con la compañía, y el destinatario externo solo recibe un conjunto de datos anónimos. Se aplican técnicas adicionales, como la adición de ruido, antes de la transferencia del conjunto de datos. La privacidad diferencial se aplica cuando un tercero autorizado solicita datos.

b. Técnicas de seudonimización:

Las técnicas de seudonimización son diferentes de las técnicas de anonimización. Con la anonimización, se borran los datos en busca de cualquier información que pueda servir como identificador de un sujeto de datos. La seudonimización no elimina toda la información de identificación de los datos, sino que simplemente reduce la capacidad de enlace de un conjunto de datos con la identidad original de un individuo (por ejemplo, a través de un esquema de cifrado). La opinión WP29 (Grupo de Trabajo del Artículo 29) proporciona los siguientes ejemplos seleccionados de técnicas de seudonimización:

1) La Encriptación con clave secreta:

Cifrado con clave secreta: en este caso, el titular de la clave puede volver a identificar trivialmente a cada sujeto de datos mediante el descifrado del conjunto de datos porque los datos personales aún están contenidos en el conjunto de datos, aunque en forma cifrada. Suponiendo que se aplicó un esquema de encriptación de última generación, el descifrado solo puede ser posible con el conocimiento de la clave.

2) Las funciones hash:

Una función hash es una función matemática que convierte un valor numérico de entrada en otro valor numérico comprimido. La entrada a la función hash es de longitud arbitraria, pero la salida siempre es de longitud fija.

Los valores devueltos por una función hash se denominan mensaje resumen o simplemente valores hash. La siguiente imagen ilustra la función hash:

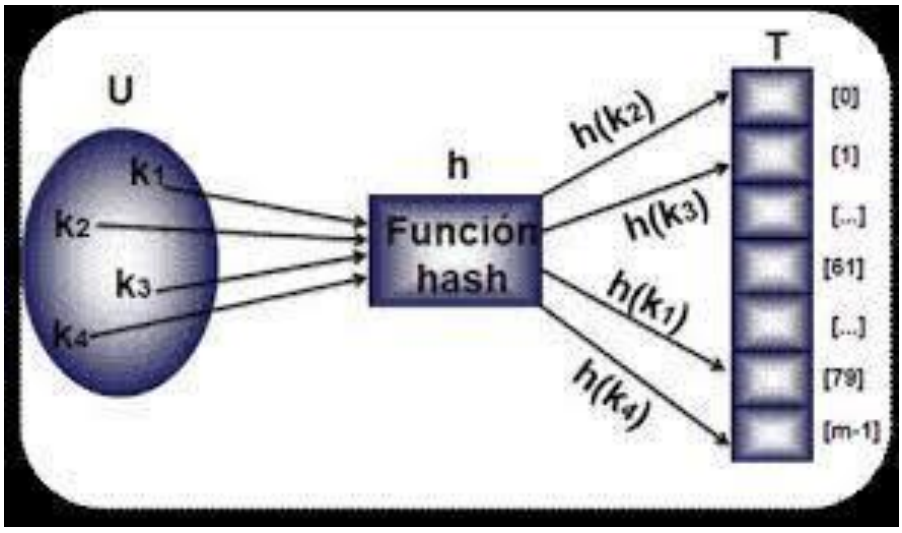


Ilustración 7. Representación gráfica de una Tabla Hash. (Guerra, 2012)

Las funciones Hash son una herramienta popular porque se pueden calcular rápidamente. Se usan para asignar datos de cualquier tamaño a códigos de un tamaño fijo. Por ejemplo, los nombres de Cédric Burton, Sára Gabriella Hoffman y John M. Smith se pueden codificar en "01", "02" y "03". Sin importar el nombre, el valor del hash siempre tendrá dos dígitos.

✔ Características de las funciones hash:

- *Una salida de longitud fija (valor hash):*
    - La función hash oculta datos de longitud arbitraria a una longitud fija. Este proceso a menudo se conoce como **hash de los datos**.
    - En general, el hash es mucho más pequeño que los datos de entrada, por lo que las funciones hash a veces se llaman **funciones de compresión**.
    - Como un hash es una representación más pequeña de datos más grandes, también se lo conoce como **resumen**.
    - La función hash con salida de n bits se conoce como **función hash de n bits**. Las funciones hash populares generan valores entre 160 y 512 bits.
- *La eficiencia de la operación:*
    - En general, para cualquier función hash h con entrada x, el cálculo de h(x) es una operación rápida.
    - Las funciones hash computacionalmente son mucho más rápidas que una encriptación simétrica.

✔ Propiedades de las funciones hash:

Para ser una herramienta criptográfica efectiva, se desea que una función Hash posea las siguientes propiedades:

- Resistencia Pre-imagen:
  Esta propiedad significa que debe ser computacionalmente difícil invertir una función hash. En otras palabras, si una función Hash h produce un valor hash z, entonces debería ser un proceso difícil encontrar cualquier valor de entrada x que que coincida con el hash en z. Así que esta propiedad protege contra un atacante que solo tiene un valor hash e intenta encontrar la entrada.

- 2º Resistencia pre-imagen:
  Con una entrada y su hash, debería ser difícil encontrar una entrada diferente con el mismo hash. Y de otra manera, si una función hash h para una entrada x produce un valor hash h(x), entonces debería ser difícil encontrar cualquier otro valor de entrada tal que h(y)=h(x). Esta propiedad de función hash protege contra una atacante que tiene un valor de entrada y su hash, y quiere reemplazar un valor diferente como valor válido en lugar de valor de entrada original.

- Resistencia a colusión:
  Con esta propiedad sería difícil encontrar dos entradas diferentes de cualquier longitud que den como resultado el mismo hash. Esta propiedad se conoce como función hash libre de colisión.
  En otra palabra, para una función hash h, es difícil encontrar dos entradas diferentes 'x' y 'y' tales que h(x)=h(h(y).
  Como la función hash comprime la función con una longitud de hash fija, es imposible que una función hash tenga colisiones. Esta propiedad de colisión libre solo hay que confirmar que estas colisiones deberían ser difíciles de encontrar.
  Esta propiedad hace que sea muy difícil para una atacante encontrar dos valores de entrada con el mismo hash. Además, si una función Hash es resistente a colisiones, es la segunda resistente a la imagen previa.

✔ **Funciones hash populares:**
Veamos brevemente algunas funciones hash populares:

- **Message Digest (MD):**
  - La familia MD comprende funciones hash MD2, MD4, MD5 y MD6. Fue adoptado como el estándar de Internet RDC 1321. Es una función hash de 128 bits.
  - Los resúmenes de MD5 se han utilizado ampliamente en el mundo del software para proporcionar seguridad sobre la integridad del archivo transferido.

- **Función Secure Hash (SHA):**
  La familia de SHA se compone de cuatro algoritmos SHA; SHA-0, SHA-1, SHA-2 y SHA-3. Aunque de la misma familia, hay diferencias estructurales.

  - La versión original es SHA-0, una función hash de 160 bits, fue publicada por el Instituto Nacional de Estándares y Tecnología (NIST) en 1993. Tenía pocas debilidades y no se fué muy popular. Más tarde, en 1995, SHA-1 fue diseñado para corregir presuntas debilidades de SHA-0.

  - SHA-1 es el más utilizado de las funciones hash SHA existentes. Se emplea en varias aplicaciones y protocolos ampliamente utilizados, incluida la seguridad de la capa de conexión segura (SSL).

  - En 2005, se obtuvo un método para descubrir colisiones para SHA-1 dentro de un marco de tiempo práctico, lo que hace que la empleabilidad a largo plazo de SHA-1 sea dudosa.

  - La familia SHA-2 tiene otras cuatro variantes SHA, SHA-224, SHA-256, SHA-384 y SHA-512, dependiendo de la cantidad de bits en su valor hash. Aún no se han reportado ataques exitosos en la función hash SHA-2.

  - En octubre de 2012, el NIST eligió el algoritmo Keccak como el nuevo estándar SHA-3. Keccak ofrece muchos beneficios, como rendimiento eficiente y buena resistencia a los ataques.

- **RIPEMD:**

El RIPEND es un acrónimo de RACE "Integrity Primitives Evaluation Message Digest". Este conjunto de funciones hash fue diseñado por una comunidad de investigación abierta y generalmente conocido como una familia de funciones hash europeas.

  - El conjunto incluye RIPEND, RIPEMD-128 y RIPEMD-160. También existen versiones de 256 y 320 bits de este algoritmo.
  - RIPEMD original (128 bit) se basa en los principios de diseño utilizados en MD4 y se encontró que proporciona una seguridad cuestionable. La versión RIPEMD de 128 bits se presentó como un reemplazo de solución rápida para superar vulnerabilidades en el RIPEMD original.
  - RIPEMD-160 es una versión mejorada y la versión más utilizada en la familia. Las versiones de 256 y 320 bits reducen la posibilidad de una colisión accidental, pero no tienen niveles de seguridad más altos en comparación con RIPEMD-128 y RIPEMD-160, respectivamente.

- **Whirlpool:**

Esta es una función hash de 512 bits.

  - Se deriva de la versión modificada de "Advanced Encryption Standard" (AES). Uno de los diseñadores fue Vincent Rijmen, creador de AES.
  - Se han lanzado tres versiones de Whirlpool; a saber, WHIRLPOOL-0, WHIRLPOOL-T y WHIRLPOOL.

3) *Función hash con clave almacenada:*

Corresponde a una función hash particular que utiliza una clave secreta como una entrada adicional (esto difiere de una función hash salada ya que la sal comúnmente no es secreta). Un controlador de datos puede reproducir la función en el atributo usando la clave secreta, pero es mucho más difícil para un atacante reproducir la función sin conocer la clave, ya que el número de posibilidades a probar es lo suficientemente grande como para ser poco práctico.

4) *Encriptación determinista o función hash con la eliminación de clave*

Esta solución permite disminuir el riesgo de vinculación entre los datos personales en el conjunto de datos y los relacionados con el mismo individuo en otro conjunto de datos donde se usa un seudónimo diferente. Teniendo en cuenta un algoritmo de vanguardia, será computacionalmente difícil para un atacante descifrar o reproducir la función, ya que implicaría probar todas las claves posibles, dado que la clave no está disponible.

5) *Tokenización:*

La tokenización es un proceso mediante el cual ciertos componentes de datos son sustituidos por un equivalente no sensible. Ese equivalente se llama token. El token no tiene valor explotable, pero sirve como identificador. Es una referencia que se remonta a los datos originales.

El mapeo de datos originales a un token usa métodos que hacen que los tokens no sean reversibles en ausencia del sistema de tokenización, por ejemplo, utilizando tokens creados a

partir de números aleatorios (staff, 2013). El sistema de tokenización debe estar seguro y validado utilizando las mejores prácticas de seguridad aplicable a la protección de datos confidenciales, almacenamiento seguro, auditoría, autenticación y autorización. El sistema de tokenización proporciona a las aplicaciones de procesamiento de datos la autoridad y las interfaces para solicitar tokens o destokenizar a los datos confidenciales.

✔ **El proceso de tokenización:**

• La aplicación pasa los datos necesarios para ser tokenizados junto con la información de autenticación al sistema de tokenización;
• El sistema de tokenización comprueba la validez de la información de autenticación.
Si la autenticación falla, el proceso se detiene y la información se envía al sistema de recopilación de eventos. Esto permitirá a los administradores identificar problemas y administrar adecuadamente el sistema. Si la autenticación es correcta, el sistema pasa al siguiente paso;

• El sistema de tokenización genera, basado en algoritmos criptográficos unidireccionales, el token para los datos pasados y ambos se almacenan en la base de datos altamente segura;
• El nuevo token se pasa a la aplicación para un uso posterior. El punto crítico de este sistema y el objetivo más atractivo para los piratas informáticos es la bóveda de datos donde se almacenan los datos confidenciales reales. La bóveda debe estar protegida con sólidas capacidades de encriptación y un sistema mejorado de administración de claves que asegurará que los datos confidenciales solo sean accedidos por personas y aplicaciones autorizadas.

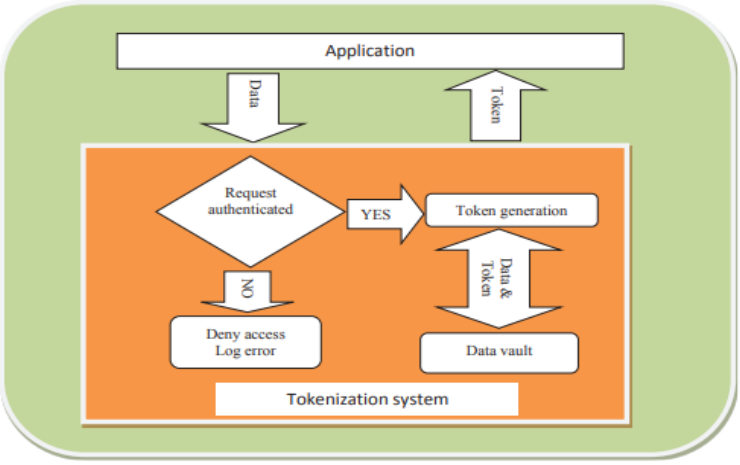


Ilustración 8. El proceso de tokenización

El desarrollo de una nueva generación de sistemas de tokenización sin vault o almacenamiento (Jonathan Care, 2015) intenta eliminar estas limitaciones que no se basarán en una base de datos de almacenamiento central y eliminarán las limitaciones correspondientes.

✔ **Diferencia entre la encriptación y la tokenización:**

La tokenización y la encriptación "clásica" protegen eficazmente los datos si se implementan correctamente, y una solución de seguridad ideal utilizará ambos. Si bien son similares en

ciertos aspectos, la tokenización y el cifrado clásico difieren en algunos aspectos clave. Ambos son métodos de seguridad de datos criptográficos y básicamente tienen la misma función, sin embargo, lo hacen con diferentes procesos y tienen diferentes efectos sobre los datos que están protegiendo.
La tokenización es un enfoque no matemático que reemplaza los datos sensibles con sustitutos no sensibles sin alterar el tipo o la longitud de los datos. Esta es una distinción importante del cifrado porque los cambios en el tipo y longitud de los datos pueden hacer que la información sea ilegible en sistemas intermedios, como las bases de datos. Los datos tokenizados son seguros, pero aun así pueden ser procesados por sistemas heredados, lo que hace que la tokenización sea más flexible que el cifrado clásico.

Otra diferencia es que los tokens requieren significativamente menos recursos computacionales para procesar. Con tokenización, los datos específicos se mantienen total o parcialmente visibles para su procesamiento y análisis, mientras que la información confidencial se mantiene oculta. Esto permite que los datos tokenizados se procesen más rápidamente y reduce la tensión en los recursos del sistema. Esto puede ser una ventaja clave en sistemas que dependen de un alto rendimiento.

FASE III: DISEÑO Y DESARROLLO DEL SISTEMA DE SEUDONIMIZACIÓN

Introducción:

Se trata de un sistema de seudonimización de los datos personales o sensibles que difiere de los enfoques existentes en que tiene una capacidad de integrar de forma segura el uso primario y secundario de los datos personales, el sistema proporciona la técnica de la seudonimización (o la anonimizarían rastreable) para garantizar la confidencialidad de los registros de datos personales situados en la base de datos del sistema, de tal forma que los registros de datos serán desacoplados de la información que identifican al propietario de los datos correspondientes, permitiendo el uso secundario (uso externo, Usuarios del sistema de información de un organismo...) sin pasos de anonimización adicionales. Al contrario de los usos secundarios en los que no es necesario la identificación de los participantes individuales como propietarios de datos.

1. Conceptos técnicos:

1.1. Autentificación:

La autenticación es un elemento absolutamente esencial de un modelo de seguridad típico. Es el proceso de confirmar la identificación de un usuario (o en algunos casos, una máquina) que está intentando iniciar sesión o acceder a recursos. Existen varios mecanismos de autenticación diferentes, pero todos cumplen este mismo propósito.

Hay una serie de componentes involucrados en la consecución de estos objetivos. Una forma es asignar permisos de acceso a los recursos que especifican qué usuarios pueden o no pueden acceder a esos recursos y en qué circunstancias. Sin embargo, los permisos de acceso sólo funcionan si hay capacidad de verificar la identidad del usuario que intenta acceder a los recursos.

En este apartado, examinaremos el papel desempeñado por la autenticación en los sistemas en general, tipos populares de autenticación, cómo funciona la autenticación y los métodos de autenticación más comúnmente utilizados.

Métodos de autentificación:

Hay varios medios físicos mediante los cuales puede proporcionar sus credenciales de autenticación al sistema. El más común, pero no el más seguro, es la autenticación de contraseña.
El ambiente empresarial competitivo de hoy en día exige opciones que ofrecen más protección cuando los recursos de red incluyen datos altamente confidenciales. Las tarjetas inteligentes y los tipos de autenticación biométrica proporcionan esta protección adicional.

"Toda organización tiene, o debería tener, una política de seguridad relativa a la protección de las aplicaciones, de los datos o también de los sistemas del SI. Esta política de seguridad puede definir niveles mínimos de autenticación en función de la criticidad del recurso utilizado" (Evidian, 2015).

Podemos ver unos de los métodos más utilizados:

1.1.1. El identificador y la contraseña:

Los sistemas de identificación de usuario y contraseña están entre las formas más antiguas de autenticación digital. Estos tipos de autenticación, que simplemente piden a un usuario que ingrese su ID y contraseña para obtener acceso al sistema, son fáciles de implementar y usar, pero también implican enormes riesgos de seguridad.

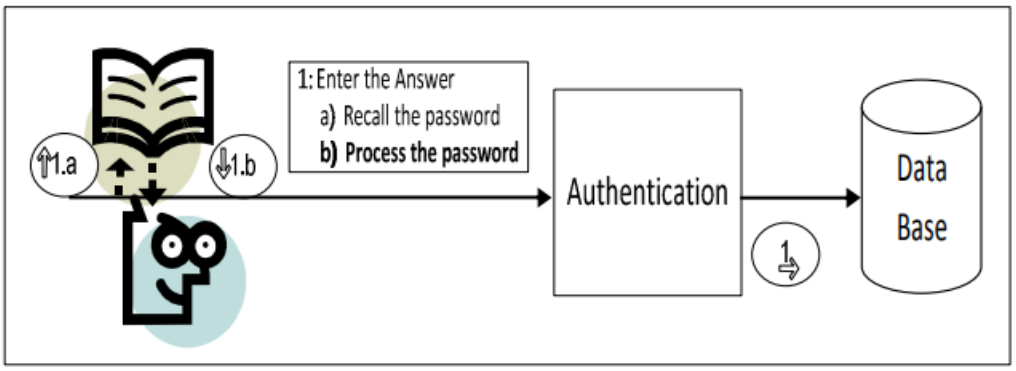


Ilustración 9. Factor of authentication i.e. Something you process (Shakir Ullah Shah, 2009)

Para preservar la seguridad de la red, las contraseñas deben ser "fuertes", es decir, deben contener una combinación de caracteres alfanuméricos y símbolos, no deben ser palabras que se encuentran en un diccionario y deben ser relativamente largas (Ocho caracteres o más). En resumen, no deben ser fácilmente adivinados.

La autenticación de contraseña es vulnerable a una contraseña "cracker" que usa un ataque de fuerza bruta (probando todas las combinaciones posibles) o que usa un protocolo "sniffer" para capturar paquetes si las contraseñas no se cifran cuando se envían a través de la red (techtarget, 2008).

1.1.2. Tarjetas inteligentes:

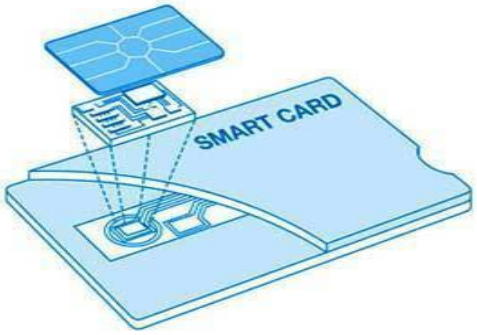


Ilustración 10: diseño sencillo de una tarjeta inteligente (redoxtechnologies, 2020)

Las tarjetas inteligentes son dispositivos de tamaño de tarjeta de crédito que contienen un chip de computadora pequeño que se utiliza para almacenar claves públicas y privadas y otra información personal utilizada para identificar a una persona y autenticarlo en el sistema. Iniciar sesión en la red con una tarjeta inteligente requiere que inserte físicamente la tarjeta en (o deslice a través de ella) un lector y, a continuación, introduzca un número de identificación personal (PIN) de la misma manera que utiliza una tarjeta ATM (*Automatic Teller Machine* - cualquier tarjeta de pago emitida por una institución financiera que permite a un cliente acceder a un cajero automático (ATM) para realizar transacciones tales como depósitos, retiros de dinero en efectivo, obtener información de cuenta, etc.).

Las tarjetas inteligentes utilizan la autenticación basada en criptografía asimétrica y proporcionan una seguridad más fuerte que una contraseña, ya que, para obtener acceso, el usuario debe estar en posesión física de la tarjeta y debe conocer el PIN.

❖ **Las tarjetas inteligentes actuales están en dos formatos:**

- **Las tarjetas de contacto,** que requieren un lector para facilitar la conexión bidireccional. La tarjeta debe insertarse en un dispositivo que toque los puntos de contacto de la tarjeta, lo que facilita la comunicación con el chip de la tarjeta. Las tarjetas de contacto vienen en modelos de 3 voltios y 5 voltios, al igual que las CPUs de escritorio actuales. Los lectores de tarjetas de contacto se construyen comúnmente en edificios y activos pertenecientes a empresas o vendedores, teléfonos celulares, dispositivos portátiles, dispositivos independientes que se conectan al puerto serial o USB (Universal Serial Bus), ranuras para tarjetas para laptop y teclados.

- **Las tarjetas sin contacto** son las que utilizan acopladores de proximidad para obtener información desde y hacia el chip de la tarjeta. Una antena se enrolla alrededor de la circunferencia de la tarjeta y se activa cuando la tarjeta se irradia a una distancia específica del acoplador. La configuración de la antena de la tarjeta y el acoplador facilitan estados conectados de un par de centímetros a un par de pies. La transmisión bidireccional está codificada y puede ser cifrada usando una combinación de los algoritmos de chip codificados, de un proveedor de tarjeta: números de sesión generados al azar, el certificado del titular de la tarjeta, la clave secreta o el número de identificación personal (PIN). La sofisticación de la conexión puede facilitar conexiones separadas y discretas con múltiples tarjetas si se encuentran dentro del alcance del acoplador, ya que las tarjetas sin contacto no requieren contacto físico con un lector.

❖ **El estándar ISO 7816:**

Las normas internacionales rigen las características físicas de las tarjetas inteligentes. La ISO 7816[8] y las normas posteriores cubren los parámetros de fabricación, las características físicas y eléctricas, la ubicación de los puntos de contacto, los protocolos de comunicación, el almacenamiento de datos y más. Por ejemplo, el tamaño de una tarjeta está cubierto por la Organización Internacional de Normalización. La disposición y el formato de datos, sin embargo, pueden variar de vendedor a vendedor.

Además de los estándares físicos y de manufactura, existe un número creciente de estándares para aplicaciones de proveedores específicos. Los vendedores de tarjetas de crédito, vendedores de teléfonos celulares, bancos de Estados Unidos y europeos, agencias de crédito y agencias de débito son ejemplos de organizaciones que están adaptando aplicaciones y procedimientos de tarjetas inteligentes orientadas exclusivamente a los servicios que ofrecen ya las empresas con las que hacen negocios.

Los dos mayores proveedores de sistemas operativos para tarjetas inteligentes son MAOSCO (un consorcio de la industria) y Microsoft (Clercq, 2017).

❖ **Tarjetas inteligentes basados en PKI (Infraestructura de Clave Pública):**

Una infraestructura de clave pública (PKI) admite la distribución e identificación de claves públicas de cifrado, lo que permite a los usuarios y las computadoras intercambiar datos de forma segura a través de redes como Internet y verificar la identidad de la otra parte.

Sin PKI, la información sensible aún puede ser encriptada (asegurando la confidencialidad) e intercambiada, pero no habría seguridad de la identidad (autenticación) de la otra parte. Cualquier forma de datos confidenciales intercambiados a través de Internet depende del PKI para la seguridad. Una PKI típica consiste en hardware, software, políticas y estándares para administrar la creación, administración, distribución y revocación de claves y certificados digitales. Los certificados digitales están en el corazón de PKI, ya que afirman la identidad del sujeto del certificado y vinculan esa identidad a la clave pública contenida en el certificado.

En general una PKI está formada por tres entidades distintas:

- ✓ **La autoridad de registro (SA).** Esta entidad se encarga de las operaciones administrativas, tales como la verificación de la identidad del usuario o el seguimiento de las solicitudes.
- ✓ **La autoridad de certificación (CA).** Esta entidad se encarga de las tareas de creación de certificados o firma de las listas de revocación.
- ✓ **La autoridad de depósito (AD).** Esta entidad se encarga de la conservación de seguridad de los certificados.

Al almacenar firmas electrónicas y certificados en una tarjeta, se pueden obtener todos los beneficios de la tecnología PKI. Las tarjetas inteligentes proporcionan autenticación en cualquier lugar, no sólo la autenticación en el escritorio particular donde se descarga originalmente el certificado. Además de su portabilidad, los certificados basados en tarjetas son mucho más seguros que los almacenados en discos duros. Debido a estos factores, las tarjetas inteligentes son posiblemente los medios más eficientes y seguros de almacenar credenciales electrónicas.

[8]ISO 7816 es un estándar internacional relacionado con las tarjetas de identificación electrónicas, en especial las tarjetas inteligentes, gestionado conjuntamente por la Organización Internacional de Normalización (ISO) y Comisión Electrotécnica Internacional (IEC). Se trata de una extensión de la ISO 7810 (Wikimedia Foundation, 2017).

❖ ***Sistema de administración de las tarjetas inteligentes:***

El CMS[9] gestiona la emisión, autenticación, retiro, cancelación, etc. de las tarjetas inteligentes y la gestión de su ciclo de vida completo. Las funciones principales en CMS son la administración de tarjetas, la administración del titular de la tarjeta, la solicitud, la administración de usuarios y la administración de roles. El sistema AMS administra la billetera electrónica, el control de acceso y otras aplicaciones añadidas después de la emisión. Ambos sistemas (AMS y CMS) comparten una base de datos centralizada.

1.1.3. Los tokens de contraseña de una sola vez (OTP)

Son otra forma de autenticación que requiere dos factores, estos tokens están programados para generar y mostrar nuevas contraseñas a ciertos intervalos. Para acceder a un sistema, un usuario debe ingresar su ID de usuario y contraseña, que es el primer factor de autenticación y luego proporcionar el PIN que aparece en el token, que es el segundo factor de autenticación.

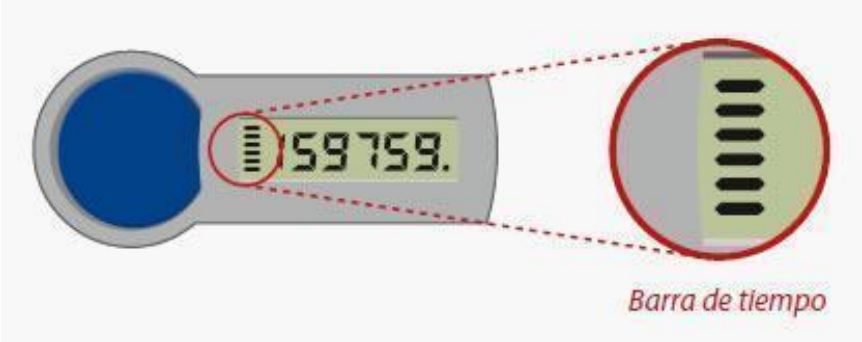


Ilustración 11. Ejemplo de un token de contraseña (Nación, 2017)

El PIN proporcionado por el token está cambiando constantemente - aproximadamente cada 30-60 segundos, dependiendo de cómo está programado - y eso hace que sea extremadamente difícil para un hacker usar ese PIN para efectuar un acceso malicioso. Incluso si el atacante robase con éxito el PIN, antes de que entrara al sistema, ya habría cambiado.

1.1.4. Autenticación biométrica:

La biometría es un método de autenticación que utiliza huellas dactilares o exploraciones faciales e iris o reconocimiento de voz para identificar usuarios. Un dispositivo de escaneo biométrico toma los datos biométricos de un usuario, como un patrón de iris o un escaneo de huellas dactilares, y lo convierte en información digital que un ordenador puede interpretar y verificar. Dado que es más difícil para un pirata informático malintencionado acceder a los datos biométricos de una persona, y es poco probable que un usuario pierda o mal uso de sus datos biométricos, esta forma de tecnología proporciona un mayor nivel de seguridad que otros métodos de identificación.

[9] *CMS: Card Management System*

Ilustración 12 La autenticación biométrica (Group, 2016)

La biometría se puede utilizar tanto para el acceso físico a edificios corporativos como para el acceso interno a computadoras y sistemas empresariales. La biometría se utiliza más a menudo como una forma de autenticación en un sistema más amplio de autenticación de dos factores o *Multi-factores*, ya que la mayoría de las implementaciones biométricas también requieren que los empleados ingresen identificadores de usuario y contraseñas. Falta la referencia bibliográfica

1.1.5. La autenticación Multi-factores

Un factor de autenticación es un elemento que se sabe (código secreto), que se posee (apoyo físico) o que es (biométrica).
Hablamos de la autenticación multifactorial, cuando se utiliza más de un factor para una misma operación de autenticación.

Ejemplos de sistema de autenticación a 1 factor:	• **Identificador + contraseña (elemento que se sabe),** • **Definición sin contacto (elemento que se posee),** • **Biométrica o identificador + biométrica (elemento que es).**
Ejemplos de sistema de autenticación a 2 factores:	• Tarjeta inteligente + código PIN (elementos **que se posee Y que se sabe**), • Tarjeta inteligente + biométrica (elemento **que se posee Y que es**), • Biométrica + contraseña (elemento **que es Y que se sabe**).
Ejemplo de sistema de autenticación a 3 factores:	• Tarjeta inteligente + cifra PIN + biométrica (elementos que se posee Y que se sabe Y que es).

La multiplicación del número de factores de autenticación aumenta el nivel de seguridad general, pero plantea los siguientes problemas:

- El ciclo de vida de cada factor debe administrarse: inicialización de las contraseñas y códigos PIN, distribución de las tarjetas inteligentes, etc.
- La ergonomía de utilización puede volverse demasiado pesada para los usuarios.
- Se añaden los costes de los periféricos (tarjetas inteligentes, lectores, sensores biométricos). Además, la carga del servicio de ayuda al usuario va a aumentar para administrar el conjunto de estos métodos (desbloqueo de las contraseñas y códigos PIN, distribución de las tarjetas, formación de los usuarios a la biométrica, etc).
-

1.2. Autorización y Control de Acceso:

Para tener una autorización exitosa y esquemas de control de acceso, se necesitan dos cosas: buena autenticación y buenas políticas.

Los mecanismos de control de acceso son un elemento de diseño necesario y crucial para la seguridad de cualquier aplicación. En general, una aplicación web debe proteger los datos de front-end y back-end y del sistema mediante la implementación de restricciones de control de acceso sobre qué pueden hacer los usuarios, a qué recursos tienen acceso y qué funciones pueden realizar en los datos. Idealmente, un esquema de control de acceso debería proteger contra la visualización, modificación o copia no autorizada de datos. Además, los mecanismos de control de acceso también pueden ayudar a limitar la ejecución de código malicioso o acciones no autorizadas a través de un atacante que explote las dependencias de la infraestructura (servidor DNS, servidor ACE, etc.).

La autorización y el Control de acceso son términos que a menudo se intercambian por error. La autorización es el acto de verificar si un usuario tiene el permiso adecuado para acceder a un archivo en particular o realizar una acción en particular, suponiendo que el usuario se haya autenticado correctamente.
La autorización está muy centrada en credenciales y depende de reglas específicas y listas de control de acceso preestablecidas por los administradores de la aplicación web o los propietarios de los datos. Las verificaciones de autorización típicas implican consultar la pertenencia a un grupo de usuarios en particular, la posesión de una autorización particular o buscar a ese usuario en la lista de control de acceso aprobada de un recurso.

Modelos de control del acceso:

El control de acceso es básicamente identificar a una persona que realiza un trabajo específico, autenticarlo mirando su identificación, y luego darle a esa persona solo la llave de la puerta o la computadora a la que necesita acceder. En el mundo de la seguridad de la información, uno consideraría esto como otorgar un permiso individual para ingresar a una red a través de un nombre de usuario y contraseña, lo que les permite acceder a archivos, computadoras u otro hardware o software que la persona requiera y garantizar que tener el nivel correcto de permiso (es decir, solo lectura) para hacer su trabajo. Entonces, ¿cómo se puede otorgar el nivel correcto de permiso a un individuo para que pueda realizar sus tareas?, aquí es donde cobran importancia los modelos de control de acceso.

Los modelos de control de acceso tienen cuatro características: control de acceso obligatorio (MAC), control de acceso basado en roles (RBAC), control de acceso discrecional (DAC) y control de acceso basado en reglas (RBAC o RB-RBAC). Veamos cada uno de ellos y lo que implican (techotopia, 2016).

1.2.1. El modelo de Control de acceso obligatorio (MAC)

El modelo de Control de acceso obligatorio (MAC) proporciona sólo la gestión del propietario y del custodio de los controles de acceso. Esto significa que el usuario final no tiene control sobre las configuraciones que proporcionan privilegios a nadie. Actualmente, hay dos modelos de seguridad asociados con MAC: Biba y Bell-LaPadula. El modelo Biba se centra en la integridad de la información, mientras que el modelo Bell-LaPadula se centra en la confidencialidad de la información.

- **El modelo Biba:** es una configuración en la que un usuario con poco margen de seguridad puede leer información de nivel superior (llamada "read up") y un usuario con una autorización de alto nivel puede escribir para niveles más bajos de autorización (llamado "write down"). El modelo Biba se utiliza típicamente en negocios donde los empleados de niveles inferiores pueden leer información de mayor nivel y los ejecutivos pueden escribir para informar a los empleados de nivel inferior.

- **El modelo Bell-LaPadula:**

Por otro lado, es una configuración donde un usuario en un nivel superior (es decir, Top Secret) solo puede escribir en ese nivel y no más bajo (llamado "write up"), pero también puede leer en niveles más bajos (llamado "read down"). El modelo Bell-LaPadula se desarrolló para fines gubernamentales y / o militares, donde si uno no tiene el nivel de autorización correcto y no necesita conocer cierta información, no tienen ningún negocio con la información (techotopia, 2016).

1.2.2. El modelo de Control de acceso basado en roles (RBAC):

El modelo RBAC (The Role Based Access Control) proporciona control de acceso según la posición que ocupa una persona en una organización. Por lo tanto, en lugar de asignar permisos de John como administrador de seguridad, la posición del administrador de seguridad ya tiene permisos asignados. En esencia, John solo necesitaría acceder al perfil del administrador de seguridad. RBAC hace la vida más fácil para el administrador del sistema de la organización. El gran problema con este modelo de control de acceso es que, si John requiere acceso a otros archivos, tiene que haber otra forma de hacerlo, ya que los roles solo están asociados con la posición; de lo contrario, los gerentes de seguridad de otras organizaciones podrían tener acceso a archivos para los que no están autorizados (techotopia, 2016).

1.2.3. El modelo de control de acceso discrecional (DAC):

El DAC (Discretionary Access Control) es el modelo menos restrictivo comparado con el modelo MAC, que es más restrictivo. El DAC permite un control completo e individual sobre cualquier objeto que posea junto con los programas asociados con esos objetos. Esto le da al DAC dos debilidades principales: en primer lugar, le da al usuario final el control total para establecer las configuraciones de nivel de seguridad para otros usuarios, lo que podría dar como resultado que los usuarios tengan privilegios más altos de lo que deberían; en segundo lugar, y lo que es peor, los permisos que tiene el usuario final se heredan en otros programas que ejecuta. Esto significa que el usuario final puede ejecutar malware sin saberlo y el malware podría aprovechar los privilegios potencialmente de alto nivel que posee el usuario final (techotopia, 2016).

1.2.4. El Control de acceso basado en reglas (RBAC o RB-RBAC):

El cuarto y último modelo de control de acceso es el Control de acceso basado en reglas, también conocido con el acrónimo RBAC o RB-RBAC, que asignará de forma dinámica roles a los usuarios según los criterios definidos por el custodio o el administrador del sistema. Por ejemplo, si a alguien solo se le permite el acceso a archivos durante ciertas horas del día, el Control de acceso basado en reglas sería la herramienta de elección. Las "reglas" adicionales del control de acceso basado en reglas que requieren implementación pueden necesitar ser "programadas" en la red por el custodio o el administrador del sistema en forma de código en lugar de "marcar la casilla (techotopia, 2016).

2. Discusión sobre los Algoritmos de Encriptación Utilizados:

2.1. Vista general:

En este apartado vamos a aplicar una selección entre los distintos algoritmos que hemos visto en los apartados anteriores (Fase III-i) cogiendo la mejor opción para cada aplicación de nuestro sistema.

Tendremos lugar dos dimensiones para nuestra selección, de una parte, miraremos el método de encriptación del algoritmo si es simétrico o asimétrico. Y de otra parte el medio o el lugar donde vamos a utilizar el algoritmo, en nuestro caso tenemos las tres capas de seguridad del sistema, donde en cada capa se aplicará un algoritmo de cifrado adecuado.

2.2. Algoritmos por tipo de encriptación:

2.2.1. Seleccionar entre los algoritmos asimétrico:

Los protocolos criptográficos modernos utilizan cada vez más algoritmos asimétricos como RSA y ECC debido a su flexibilidad y capacidad mejorada para gestionar las claves.
RSA, que se desarrolló a finales de los 70, se ha convertido en un algoritmo de elección para seguridad en internet. Donde la criptografía de curva elíptica (ECC), que fue la primera propuesta en la década de los 80 se está utilizando cada vez más por varias razones.
Hay importantes diferencias entre los dos lo que justifica una cuidadosa comparación.

El nivel de seguridad en los sistemas se está convirtiendo en una preocupación principal como cabría esperar. La mayoría de los expertos en criptografía recomiendan que los sistemas actuales ofrezcan al menos 128 bits de seguridad, pero ¿qué significa eso realmente? Tenga en cuenta que esto no es lo mismo que la longitud de la clave, como muchos pueden pensar. La seguridad proviene de la combinación del algoritmo específico y su longitud de clave. Por ejemplo, generalmente se piensa que se pueden lograr 128 bits de seguridad con claves AES de 128 bits, claves de curva elíptica de 256 bits y claves RSA de 3072 bits. Si se ignoran los problemas de implementación, estos algoritmos con las longitudes de clave especificadas generalmente tendrán el mismo nivel de seguridad (BlueKrypt, 2015). Las implementaciones típicas de RSA actualmente emplean claves de 1024 o 2048 bits, pero ambas son menos seguras que AES-128.

Bits de seguridad	Algoritmo de cifrado simétrico	Tamaño mínimo (bits) de las claves públicas	
		RSA	ECC

80	Skipjack	1024	160
112	3DES	2048	224
128	AES-128	3072	256
192	AES-192	7680	384
256	AES-256	15360	512

Tabla 3. Comparación de seguridad para varias combinaciones de tamaño de clave-algoritmo (Agency, 2009)

Las longitudes de las claves generalmente aumentan con el tiempo a medida que aumenta el cómputo disponible para los atacantes, que es una manifestación de la tasa de armas criptográficas. Algunos expertos ahora sugieren que se emplee AES-256 para el cifrado de datos en lugar del protocolo AES-128 aceptado previamente. Si utiliza curvas elípticas para la gestión de claves (es decir, la clave de sesión de cifrado / descifrado) de una sesión AES-256, entonces se necesitará una clave de sesión de curva elíptica de 512 bits, como se muestra en la Tabla anterior (tabla-3). Para lograr el mismo nivel de seguridad con el cifrado RSA, se requieren claves de 15.360 bits, lo que es computacionalmente inviable en los sistemas integrados de hoy. Este marcado contraste entre la viabilidad de ECC sobre RSA para sistemas integrados indica que ECC es el algoritmo del futuro para sistemas integrados.

Dicho esto, la seguridad del algoritmo no importa si un atacante puede obtener las claves a través de otros métodos. Este punto no se puede enfatizar lo suficiente. La seguridad comienza y termina con lo bien que están protegidas las claves. Además del almacenamiento deficiente de claves, las implementaciones de algoritmos débiles o defectuosos, la generación incorrecta de números aleatorios y los ataques agresivos en los sistemas de punto final también pueden degradar la seguridad.

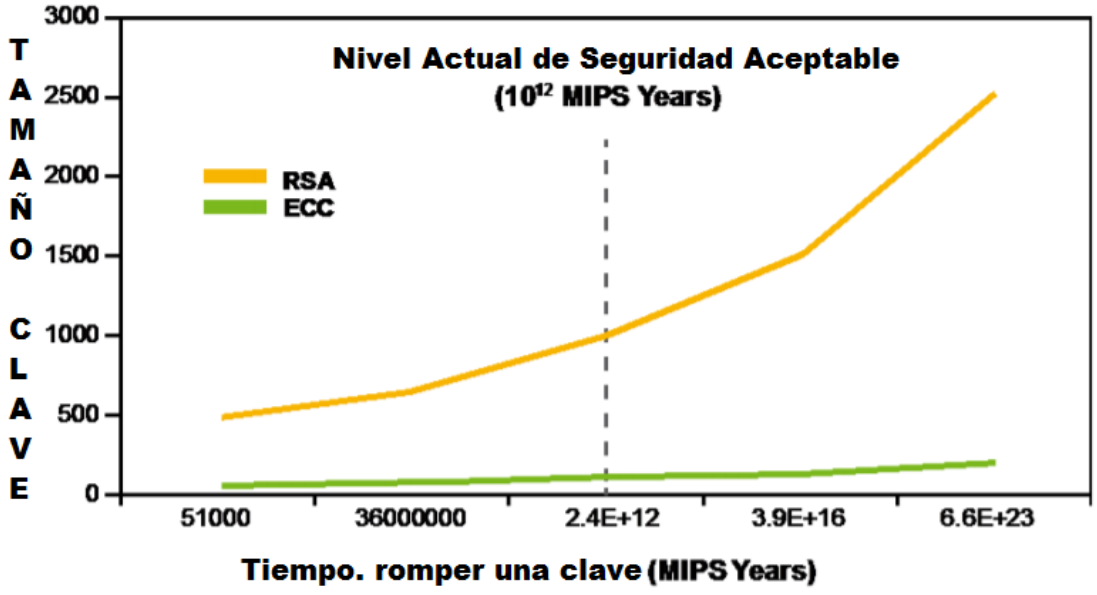


Figura 12. Rendimiento de RSA y ECC (Fuente: RSA) (M. Alimohammadi, 2014)

Esta gráfica presenta otra forma de ver el rendimiento de RSA y ECC. Compara qué longitudes de clave de cada algoritmo proporcionará un nivel de seguridad medido en el tiempo en MIPS-años para romper la seguridad.
Está claro que ECC es más eficiente.

Ansiedad de rendimiento:

Cuando se trata de rendimiento a niveles de seguridad de 128 bits, se informa que RSA es 10 veces más lento que ECC para operaciones de clave privada como la generación de firmas o la administración de claves.
La disparidad de rendimiento se expande dramáticamente a niveles de seguridad de 256 bits, donde el RSA es de 50 a 100 veces más lento. La generación de claves de RSA también es muy lenta en comparación con la generación de claves ECC, ya que la RSA es de 100 a 1000 veces más lenta. Sin embargo, esto puede o no ser una consideración importante en los sistemas que generan claves con poca frecuencia. Sí importa para ciertos protocolos o políticas que requieren una generación de claves más frecuente.
La validación de firmas de claves públicas generalmente es más rápida con RSA en comparación con ECC, lo que puede proporcionar un beneficio.

Ancho de banda:

Cuando se trata del ancho de banda de la red, la preocupación número uno se relaciona con el algoritmo simétrico utilizado para el cifrado de mensajes y la codificación de autenticación de mensajes (MAC) para la verificación de integridad (esto no está relacionado con la elección de RSA frente a ECC). Los sistemas integrados más pequeños pueden iniciar sesiones con más frecuencia, o la autenticación asimétrica puede ser un porcentaje mayor del tráfico general y el tamaño de las claves y firmas puede marcar la diferencia. En el nivel de seguridad de 128 bits, las claves públicas y las firmas son seis veces más grandes para RSA que para ECC. Las claves privadas son 12 veces más grandes para RSA en comparación con ECC en el nivel de seguridad de 128 bits. El tamaño de la clave generalmente no tiene impacto en el rendimiento, pero el tamaño importa cuando se trata del costo de almacenamiento seguro de las claves.

Conclusiones:

Debido a los problemas de seguridad, la mayoría de los nuevos protocolos criptográficos se están alejando de RSA a curvas elípticas. Esa transición se está produciendo aún más rápido en el espacio integrado donde los beneficios de costo / rendimiento de ECC rápidamente se vuelven significativos.

2.2.2. Seleccionar entre algoritmos simétrico:

En la (fase III, apartado i) hemos hablado sobre los distintos algoritmos simétricos/Asimétricos que se utilizan en las distintas aplicaciones, en esta sesión hacemos una comparativa basando sobre unas métricas determinadas. Esta comparativa nos va a ayudar a seleccionar el algoritmo que vamos a utilizar en las distintas encriptaciones simétricas para las capas de seguridad de nuestro sistema de seguridad.

Los algoritmos que entran en esta comparativa son los siguientes:

DES, 3DES, AES, RSA y Blowfish con la idea de prevenir los ataques de adivinanzas.

Métricas de evaluación:

i- Tiempo de cifrado: el tiempo necesario para convertir texto sin formato en texto cifrado es el tiempo de cifrado.

ii- Tiempo de descifrado: el tiempo para recuperar texto sin formato del texto cifrado se denomina tiempo de descifrado.

iii-Memoria utilizada: las diferentes técnicas de encriptación requieren diferentes tamaños de memoria para su implementación. Este requisito de memoria depende de la cantidad de operaciones a realizar por el algoritmo, el tamaño de clave utilizado, los vectores de inicialización utilizados y el tipo de operaciones.

iv-Efecto de avalancha: en la criptografía, una propiedad llamada difusión refleja la fuerza criptográfica de un algoritmo. Si hay un pequeño cambio en una entrada, la salida cambia significativamente. Esto también se llama efecto de avalancha.

v- Entropy: es la aleatoriedad recopilada por una aplicación para su uso en criptografía que requiere datos aleatorios. La falta de entropía puede tener un impacto negativo en el rendimiento y la seguridad.

vi-Número de bits necesarios para la codificación óptima: el número de bits necesarios para codificar un carácter cifrado debe ser menor. Desde entonces, el bit cifrado se transmite a través de una red después de la codificación.

Resultados y discusiones:

Los resultados se analizan en función de la implementación realizada en Priyadarshini Patil, (2016) y Polimon J, (2008).

i- La Figura 13 muestra que el algoritmo blowfish registra el tiempo de cifrado más rápido, y el algoritmo RSA registra el tiempo de cifrado más lento. En función del tiempo de encriptación, seleccionaremos la técnica del blowfish para una evaluación adicional.

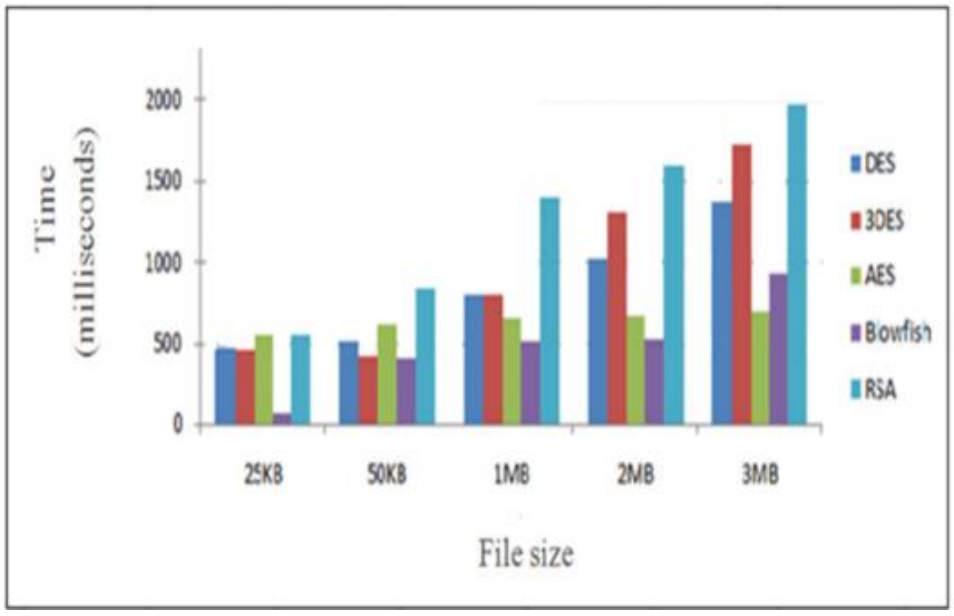


Figura 13. Tiempo de cifrado frente a tamaño de archivo para DES, 3DES, AES, Blowfish y RSA

ii- La Figura 14 muestra que el tiempo de descifrado para todos los algoritmos es más rápido que el tiempo de cifrado. Además, el algoritmo blowfish registra el tiempo de descifrado más rápido y el algoritmo RSA registra el tiempo de descifrado más lento. Sobre la base de la función de tiempo de descifrado, seleccionaremos la técnica del blowfish para ser considerada en el siguiente nivel de evaluación.

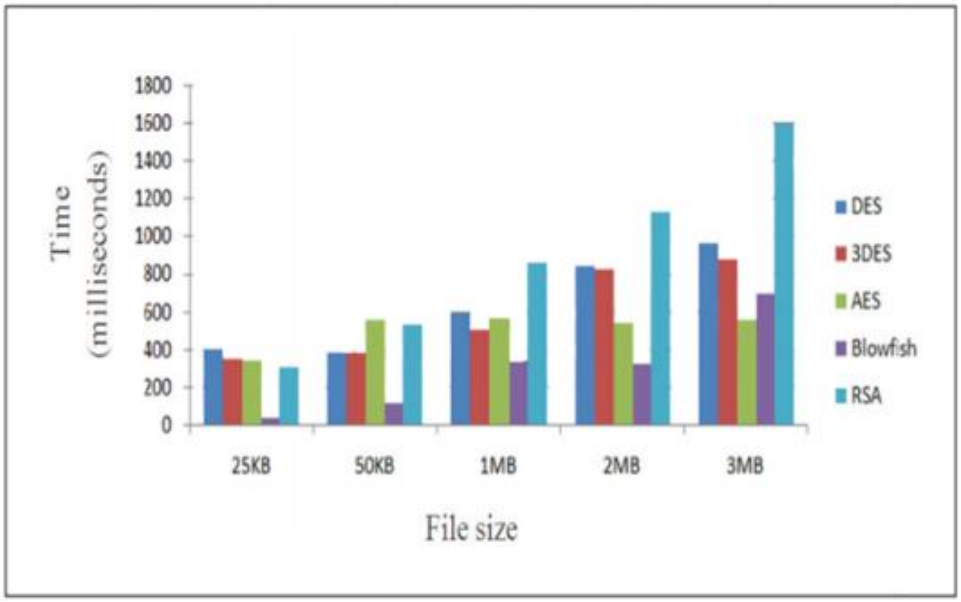


Figura 14. Tiempo de descifrado frente a tamaño de archivo para DES, 3DES, AES, Blowfish y RSA

iii- A continuación, en la tabla 4, se presenta la memoria utilizada para las operaciones unitarias para todas las técnicas criptográficas que estudiamos. Blowfish consumió menos almacenamiento de memoria que otros tipos, mientras que RSA usa la memoria más alta.

Algoritmo	**Memoria utilizada (KB)**
DES	18,2
3DES	20,7
AES	14,7
Blowfish	9,38
RSA	31,5

Tabla 4. Memoria utilizada para las operaciones unitarias

iv- La Figura 15 muestra que AES manifiesta el mayor efecto de avalancha, mientras que RSA manifiesta el menor efecto de avalancha. Esto ha devuelto la atención a AES para su posterior análisis y mejoras.

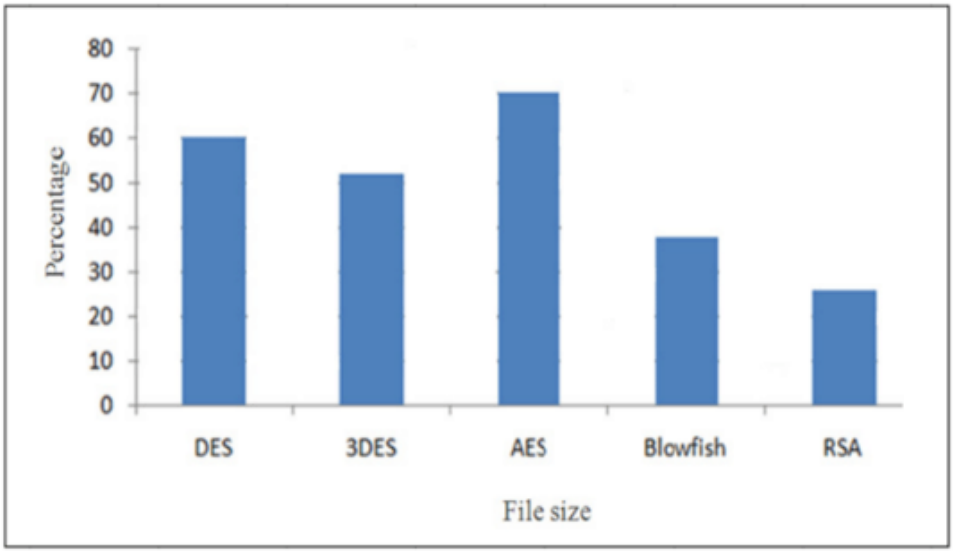


Figura 15. Efecto de avalancha frente a tamaño de archivo para DES, 3DES, AES, Blowfish y RSA

v- Como la prueba de entropía y el experimento final. La Tabla 5 muestra que el pez globo registra la entropía promedio más alta por byte de cifrado. Eso debería resaltar los logros del algoritmo Blowfish para considerar un nuevo aspecto de seguridad.

Algoritmo	**Promedio de entropía por byte de cifrado**
DES	2,9477
3DES	*2,9477*
AES	*3,84024*
Blowfish	*3,93891*
RSA	*3,0958*

Tabla 5. Valores medios de entropía.

vi- La Tabla 6 presenta AES exige que el mayor número de bits se codifican de manera óptima, mientras que DES exige que el menor número de bits se codifican de manera óptima.

Algoritmo	**Número promedio de bits demandados para codificar de manera óptima un byte de datos cifrados**
DES	*27*
3DES	*40*
AES	*256*
Blowfish	*128*
RSA	*44*

Tabla 6. Longitud de codificación óptima

Conclusión:

Cada uno de los algoritmos criptográficos tiene puntos débiles y puntos fuertes. Seleccionamos el algoritmo criptográfico en función de las demandas de la aplicación que se utilizará. A partir de los resultados del experimento y la comparación, el algoritmo de ***pez globo (o Blowfish)*** es la elección perfecta en caso de tiempo y memoria según los criterios de ataques de adivinación y las funciones requeridas, ya que registra el tiempo más corto entre todos los algoritmos. Además, consume el mínimo de memoria de almacenamiento. Si la confidencialidad y la integridad son factores importantes, se puede seleccionar el algoritmo AES. Si la demanda de la aplicación es el ancho de banda de la red, el DES es la mejor opción.

2.3. Algoritmos por capas de seguridad:

2.3.1. La Criptografía en la capa de autentificación:

Esta capa de seguridad requiere una rapidez y eficacia en sus aplicaciones criptográficas, y como sabemos que en el proceso de autentificación de los usuarios hablamos de multifactores para acceder uno de ellos es el uso de la tarjeta inteligente lo que significa la aplicación de la criptografía asimétrica (claves públicos internos/externos).

Y si cogemos como base las conclusiones que hemos sacado en la sesión 3.2.1 sobre la comparativa de los dos algoritmos asimétricos, podemos ver que la criptografía de curva elíptica (**ECC**) cumple los requisitos para forzar el proceso de autentificación de nuestro sistema.

Así que cogemos el ECC como la primera opción de nuestra selección para los algoritmos asimétricos.

2.3.2. La Criptografía en la capa de autorización & seudonimización:

En estas capas se necesita acceso a las carpetas y ficheros solo para los usuarios autorizados mediante el par de las claves asimétricas internos, así quedamos con la misma selecciona de la sesión anterior (3.2.1), usando la criptografía de curva elíptica (**ECC**) para la implementar la seguridad con claves asimétricas.

Y para encriptar los ficheros de los autorizados se necesita una implementación de la criptografía simétrica para generar los claves simétricos internos del usuario o las claves de apertura de la sesión de los usuarios autorizados.

Si nos basamos en las conclusiones de la sesión (3.2.2), donde hemos elegido el algoritmo simétrico *pez globo (o **Blowfish**)* como la opción preferida. los resultados de esta comparativa nos ayudan mucho a seleccionar el algoritmo simétrico **Blowfish** como la mejor opción que nos puede garantizar un nivel alto de seguridad para proteger los ficheros seudónimizados

2. Funcionamiento y Metodología del Sistema de Seudonimización:

Introducción:

En este capítulo se describe la funcionalidad y la metodología del sistema de seudonimización (Heurix, y otros, 2012).

El sistema está compuesto de un modelo de seguridad basado en capas, concretamente 3, cada una se encarga de: autenticación, autorización y una capa para los datos seudonimizados.

En los siguientes apartados, se describe en profundidad el diseño y desarrollo, así como las propiedades de cada capa. También, se desarrollan distintos roles y tipos de usuarios como pueden ser los propietarios

de los datos, los usuarios autorizados y los afiliados, que son los que forman parte de los actores del sistema y estudiaremos la relación existente entre ellos en el sistema.

Además, se va a tratar la arquitectura que va a tener el sistema. Como caso concreto de arquitectura, se propone el uso de una tarjeta inteligente como token de seguridad y construir un módulo de seguridad del Hardware HSM para gestionar las distintas operaciones de seguridad, de manera conjunta entre el servidor de seudonimización donde se aplican los algoritmos criptográficos, la tarjeta inteligente del usuario y un servidor almacenamiento en el cual se crea un centro para guardar las claves secretas de los usuarios y las bases de datos para guardar los datos seudónimizados (Heurix, y otros, 2012).

También se desarrolla un modelo de datos para el sistema, basándose en el diseño propuesto en el artículo Heurix, y otros(2012). En este punto se detalla el modelo relacional entre los elementos que forman parte de la estructura de este modelo. Y, por último, se tratan también las metodologías aplicadas para la búsqueda de registros, como por ejemplo, comprobar si el usuario tiene acceso y permisos suficientes, según políticas de seguridad aplicadas.

2.1. Modelo de seguridad basado en capas:

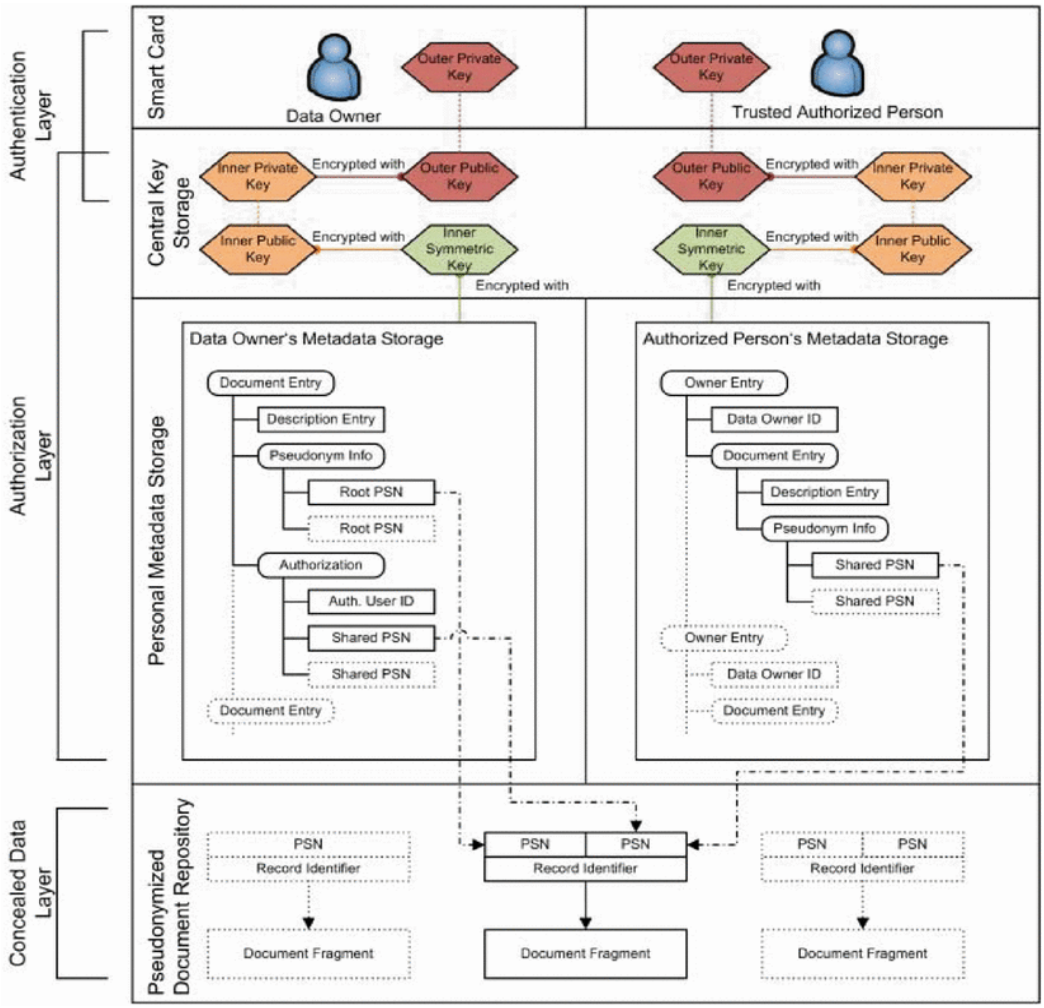


Ilustración 1. Modelo de seguridad basado en capas (Hawaii International Conference on System Sciences, 2012)

2.1.1. Capa de Autentificación de usuario:

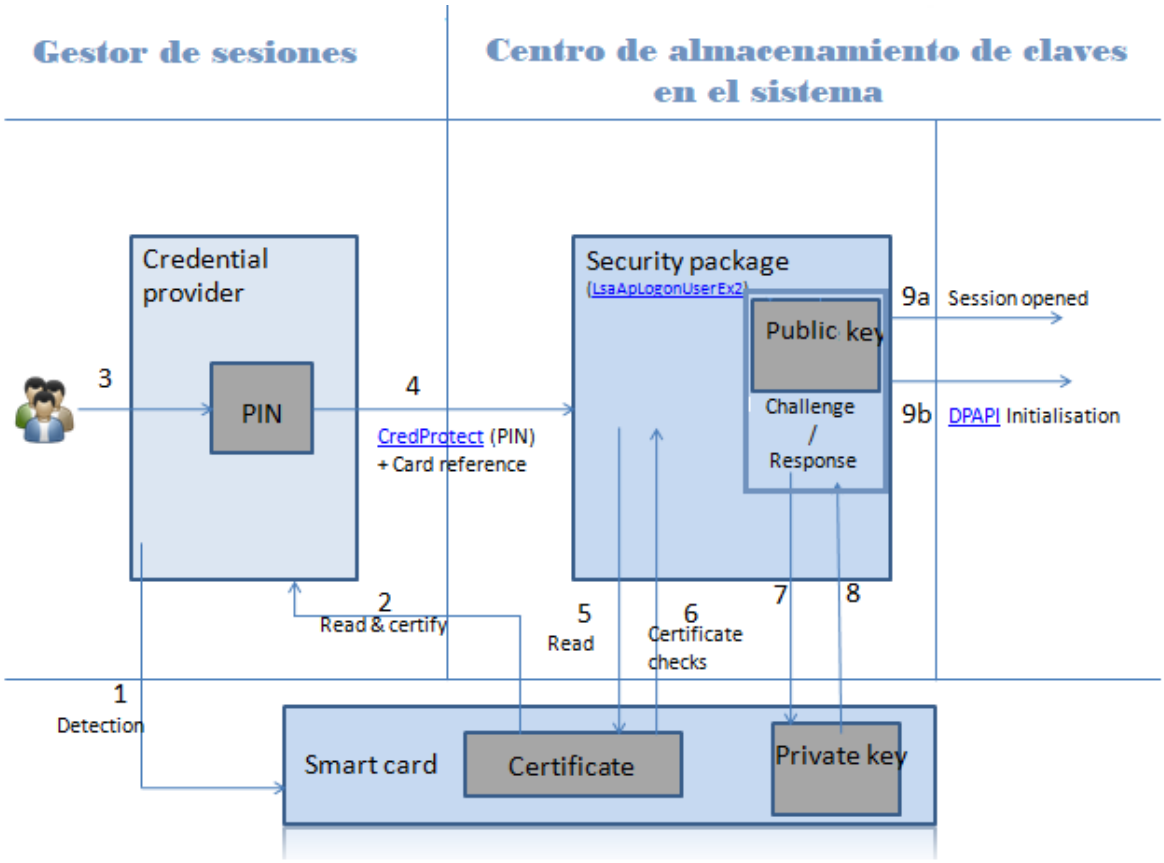


Figura 2. Capa de Autentificación - Proceso de autenticación interactivo (mysmartlogon, 2013)

Como política de seguridad hemos utilizado la criptografía asimétrica integrando un método de autenticación de dos factores (Tarjeta inteligente con el PIN de verificación –ver Figura 2) de esta forma garantizamos un nivel alto de la confidencialidad de los metadatos personales en el sistema de información.

Las claves externas públicas y privadas forman la capa externa, la *capa de autenticación*, la cual es responsable de identificar inequívocamente al usuario correspondiente. Junto con el identificador del usuario, la clave privada externa representa las credenciales de autenticación, que se almacenan junto con la clave pública del servidor en la tarjeta inteligente del usuario. En combinación con el PIN correcto, la tarjeta inteligente proporciona autenticación de dos factores, en la que el procedimiento de autenticación involucra tanto el par de contactos externos del servidor del usuario como el del servidor de claves del sistema.

Como podemos ver en la Figura 2, hay una serie de pasos a conseguir para un proceso de autenticación interactivo:

1) **Detección de la tarjeta inteligente del usuario:**
Hemos hablado anteriormente sobre las tarjetas inteligentes y el rol que juegan para una autenticación más segura, como vemos en la parte de debajo de la Figura 2, la tarjeta con su certificado instalada y su clave privada, después de que el usuario va a meter su tarjeta en lector de tarjetas en el sistema de autenticación (la parte del gestor de sesiones), este último tiene configurado un proveedor de credenciales, en este caso está configurado para verificar el segundo factor de autenticación que es el PIN.
Los proveedores de credenciales son el mecanismo principal para la autenticación de usuarios. Actualmente, son el único método para que los usuarios prueben su identidad, lo que es necesario para el inicio de sesión y otros escenarios de autenticación del sistema.
Los sistemas operativos proporcionan una variedad de proveedores de credenciales como parte del mismo sistema, como la contraseña y el PIN.

2) **Leer certificado de la tarjeta:**
Después de detectar la tarjeta y autentificarla con el PIN del usuario, viene el paso de leer el contenido de la tarjeta e identificar su certificado que está instalado en la misma. Los certificados y sus claves identifican al usuario de la tarjeta inteligente para obtener información sobre los certificados.

3) **El usuario autentifica con los dos factores (tarjeta + PIN):**
Cuando los usuarios caminan hacia una estación de trabajo configurada adecuadamente e insertan sus tarjetas inteligentes en un lector de tarjetas inteligentes conectado, el sistema inicia un proceso de inicio de sesión similar. Sin embargo, en lugar de ingresar un nombre de usuario y contraseña, los usuarios ingresan su PIN, que desbloquea la tarjeta inteligente. Este proceso es un ejemplo de autenticación de dos factores: el primer factor es algo que tienes (es decir, la tarjeta inteligente), y el segundo factor es algo que sabes (es decir, el PIN) [*ver sesiones anteriores sobre factores de autenticación*].

4) **Continuación del paso anterior:**
En un entorno de dominio, la estación de trabajo (gestor de sesiones) envía el certificado en la tarjeta inteligente a un Centro de distribución de claves en nuestro caso sería el centro de almacenamiento de las claves en el sistema.

5) **Verificación del certificado (Leer certificado)**
Se explica junto con el siguiente paso.

6) **Verificación del certificado (extraer certificado):**

El centro de almacenamiento de claves de sistema (CACS) verifica la validación del certificado para la tarjeta insertada para el correspondiste usuario, crea una clave de sesión de inicio de sesión, encripta la clave de sesión con la clave pública en el certificado y envía la clave de sesión cifrada al CACS, este último transfiere la clave de sesión de inicio de sesión cifrada a la tarjeta inteligente para su descifrado.

7) 8) **Intercambio de las claves externas asimétricas:**

La tarjeta inteligente realiza junto con el centro de almacenamiento de claves de sistema (CACS) todas las funciones criptográficas que involucran el certificado y su clave privada. [*este proceso lo explicaremos posteriormente*]

9) Inicio de sesión:

Y por tanto una sesión abierta con la generación de una clave de sesión para inicializar una nueva sesión del usuario y acceder a sus carpetas destinos en el sistema de información, según las autorizaciones que tiene.

2.1.2. Capa de autorización:

Antes de que el sistema autorice al usuario el acceso a la capa más interna donde están situados los identificadores de registros de datos, esta capa de autorización juega el rol de un puente y no deja pasar directamente a los registros de datos hasta que se desencripta la clave simétrica interna que no se puede descifrar si no estás autorizado.

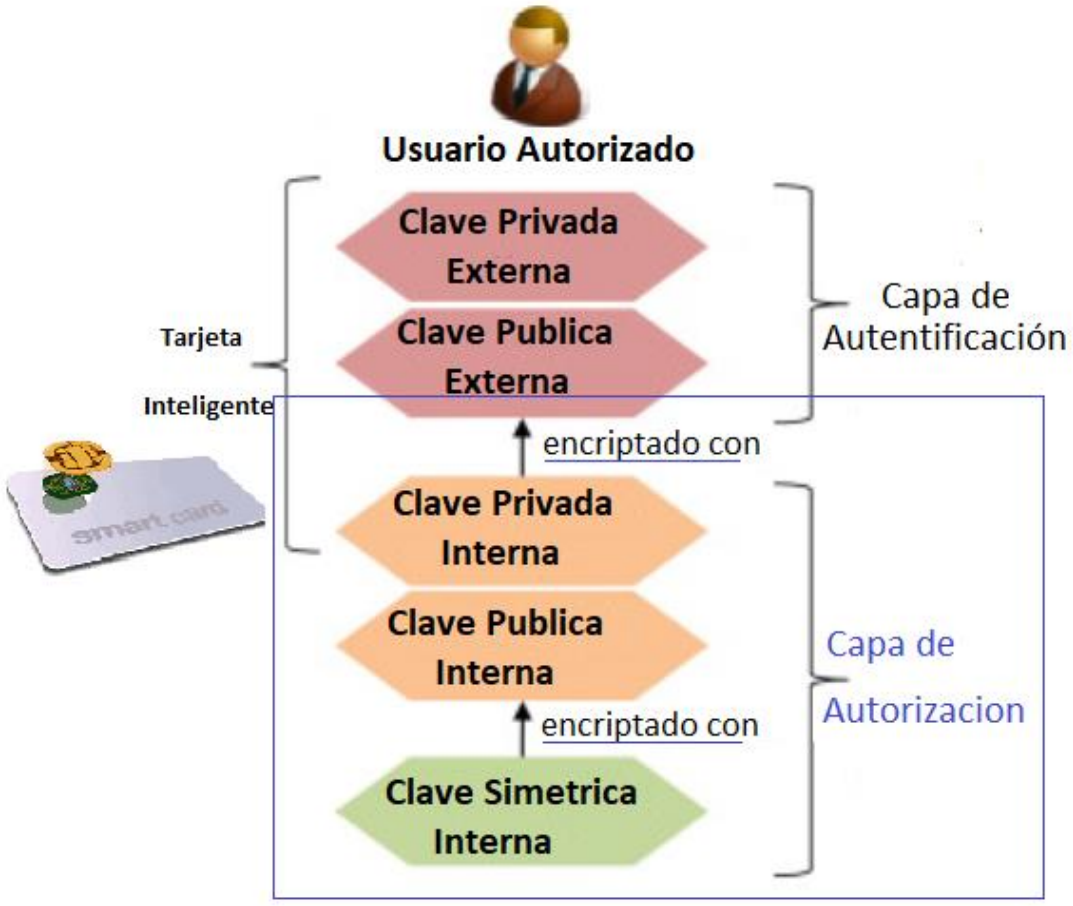
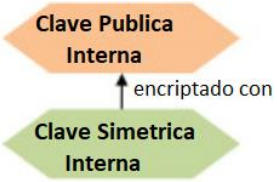


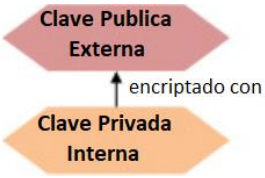
Figura 3. Capa de Autorización de un usuario.

La capa de autorización consiste en un par de claves asimétricas internas del usuario (las cajetas con color naranja – ver Figura 3) y la clave simétrica interna (las cajetas con color verde – ver Figura 3). Mientras que las claves privadas externas del usuario están creadas en la tarjeta

inteligente (las cajetas con color rosa – ver ***Figura 3***) del usuario, que está emitido especialmente para el usuario.



Las otras claves están almacenadas en la base de datos de seudonimización donde las claves secretas están cifradas, esto significa que la clave simétrica interna está cifrada con la clave pública interna.



Mientras que la clave privada interna está encriptada con la clave pública externa.
Por lo tanto, la clave privada externa es necesaria para descifrar la clave simétrica interna, que finalmente descifra los seudónimos (registros de identificación de los datos) y, por tanto, proporciona el acceso a la capa más interna, la capa de datos seudonimizados. (ver ***Figura 3***)

2.1.3. Capa de datos seudonimizados:

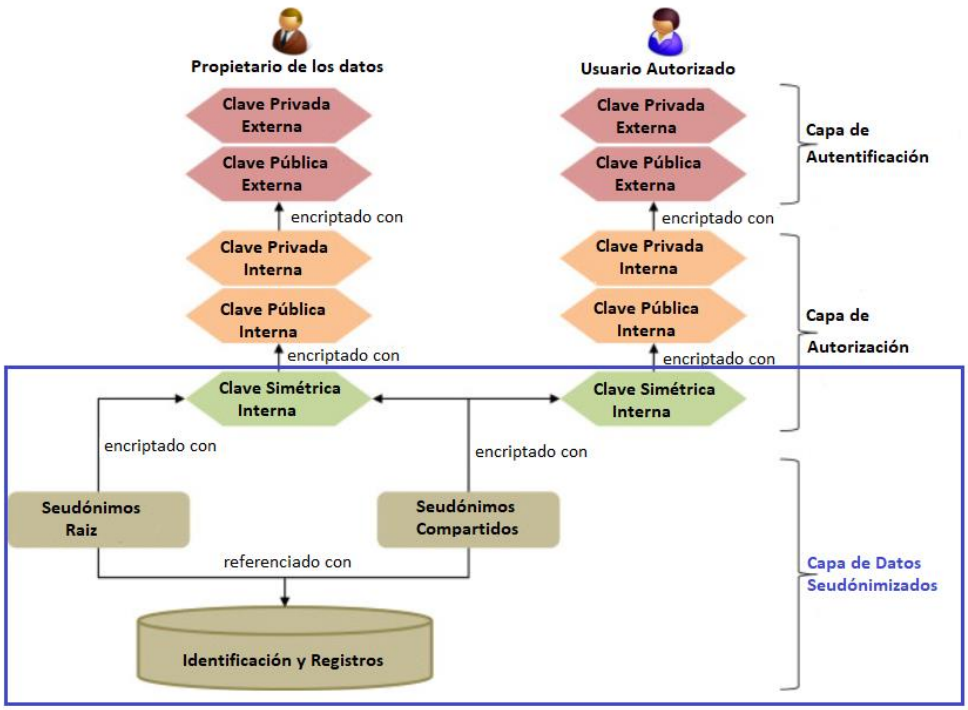


Figura 4 - Capa de datos Seudonimizados de un usuario.

Es la capa más interna del modelo de seguridad donde son situados los Seudónimos que son registros de identificación de los registros de datos personales, estos seudónimos o identificadores están cifrados con la clave simétrica interna (color verde- ver ***Figura 4***) asociada a cada usuario.

Al descifrar los seudónimos con la clave simétrica interna, el usuario finalmente puede volver a vincular el registro de datos con su registro de identificación correspondiente.
Los seudónimos están divididos por dos tipos: seudónimos raíz y compartidos, que los vamos a explicar en las siguientes sub-secciones:

2.1.3.1. Seudónimo raíz:

Los seudónimos raíz están disponibles únicamente para el propietario de los datos y representan los principales identificadores de acceso a los registros de datos. Cada vez que se almacenan nuevos registros de datos, se asigna un nuevo seudónimo raíz a cada registro individual (parte derecha inferior de la ***Figura 4***).
Como podemos ver en la ***Figura 4***, cada seudónimo raíz está encriptado con la clave simétrica interna del propietario de datos, referenciados con los registros del mismo propietario de datos, que es el único que tiene acceso, pero con una excepción; si no existe un usuario afiliado que tenga asignado una autorización especial del propietario de datos, para acceder a los seudónimos raíz y por tanto puede ver los registros[10].

2.1.3.2. Seudónimo Compartido

Los seudónimos compartidos representan las autorizaciones de acceso para usuarios de confianza. Así decimos que los seudónimos compartidos son creados por el propietario de los datos y compartidos solo con el usuario de confianza en particular, y para cada autorización de acceso individual se crea un nuevo seudónimo compartido, por lo que este tipo de seudónimos y cómo podemos ver en la ***Figura 4,*** están encriptados junto con las dos claves simétricas internas del correspondiente propietario de datos y del usuario autorizado.

2.2. Funcionamiento del Sistema

2.2.1. Roles de los usuarios del sistema

El sistema puede soportar tres roles principales diferentes: el propietario de los datos, el afiliado y el autorizado. El propietario de los datos que es quien tiene el control sobre sus datos personales, situados en el sistema de información, y es el único que puede crear las autorizaciones del acceso a sus datos en los registros específicos del sistema de información (para los usuarios autorizados), así como la concesión de acceso completo equivalentes al acceso seudónimo *Root* (para usuarios afiliados). Las autorizaciones en este contexto no se refieren a las autorizaciones de acceso en el sentido tradicional, sino proporcionar al autorizado la capacidad de volver a vincular los identificadores (seudónimos) con los registros correspondientes.

2.2.1.1. Relación entre roles

Si un usuario autorizado tiene que asociar el vínculo entre un registro de datos y su identificación correspondiente, esto se realiza mediante la creación de un nuevo conjunto de seudónimos compartidos, que están cifrados mediante las claves simétricas internas del autorizado junto con la del propietario de datos para que ambos pueden descifrar esta relación de autorización.

[10] Detallado el usuario de la sesión 4.4

En contraste con los usuarios autorizados, un usuario afiliado tiene una autorización total y especial del propietario de los datos y tiene el acceso a la clave privada interna del propietario de datos, y por tanto, puede descifrar la clave simétrica interna del propietario de datos.
Por lo tanto, el afiliado es capaz de descifrar los vínculos entre todos los pseudónimos raíz y compartidos relacionados con el propietario de los datos.

2.2.2. Arquitectura General del Sistema

Para una protección óptima de las claves secretas, proponemos utilizar tarjetas inteligentes como tokens de seguridad de propiedad del usuario para la autenticación[11]. La combinación de tarjeta inteligente y PIN de usuario proporciona autenticación de dos factores y, por lo tanto, es significativamente más segura que usar una combinación simple de nombre de usuario / contraseña. En este contexto, la tarjeta inteligente es una tarjeta de microcontrolador seguro y resistente a manipulaciones (contacto) con un área segura de almacenamiento de clave y motores criptográficos basados en hardware para algoritmos comunes, como RSA o AES (más detalles sobre los algoritmos criptográficos en tercera fase de esta memoria (Heurix, y otros, 2012).

Como las tarjetas inteligentes pueden perderse o dañarse, empleamos un mecanismo de respaldo basado en el intercambio secreto, en el siguiente ejemplo podemos explicar de manera general, cómo funciona esta técnica de compartir o intercambiar secretos:

Suponemos que un conjunto de datos D se divide en n piezas de forma que D sea fácilmente reconstruible a partir de cualquier k piezas, pero incluso el conocimiento completo de las piezas k-1 no revela absolutamente ninguna información sobre D. Esta técnica permite la construcción de robustos esquemas de administración de claves para sistemas criptográficos que pueden funcionar de manera segura y confiable incluso cuando las desgracias destruyen la mitad de las piezas y las brechas de seguridad exponen todas menos una de las piezas restantes.
Además de las tarjetas inteligentes, los otros componentes de la arquitectura general son la seudonimización & el módulo de lógica de consulta y el proveedor de almacenamiento12 (A.Shamir, 1979).

[11] Detallado en el apartado 2.1.1
[12] Consultar Figura 5

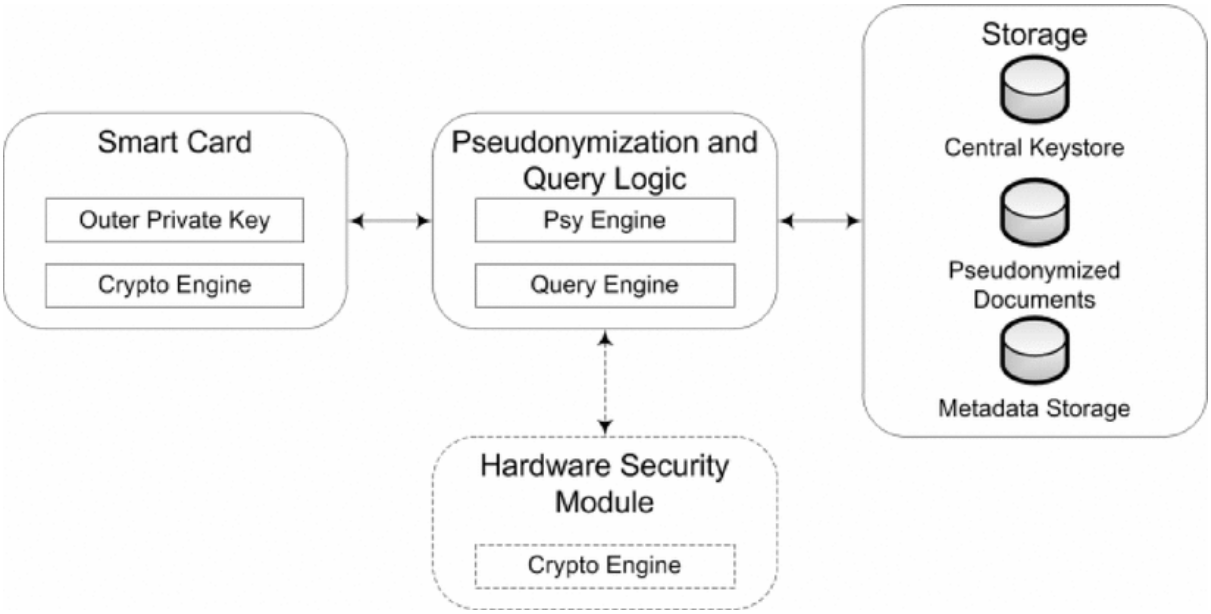


Figura 5. Arquitectura Técnica General.

El proveedor del almacenamiento incluye el almacén de claves central, los documentos seudonimizados y el almacenamiento personal de metadatos.

El almacén de claves central contiene todas las claves de usuario (aparte de la clave privada externa) donde las claves secretas se almacenan cifradas según la arquitectura de la capa de seguridad[13].

El almacenamiento de documentos seudonimizados contiene los fragmentos del documento y los seudónimos de texto claro asociados, mientras que el almacenamiento de metadatos persiste en los metadatos del documento que se pueden buscar en registros organizados en bases de datos cifradas y específicas del usuario (virtuales). Debido al estado encriptado del almacén de claves y el contenido de metadatos y la estructura de datos seudonimizados de los documentos (de carga), el proveedor de almacenamiento no necesita confiar plenamente en la confidencialidad.

En cuanto al módulo lógico, que es responsable de la seudonimización del documento y la consulta y la vinculación de fragmentos de documentos, distinguimos dos escenarios:

- **El escenario local:** el módulo lógico está situado en la máquina del cliente del usuario. La tarjeta inteligente puede actuar como módulo criptográfico primario sin depender del servidor (es decir, la estación de trabajo) de modo que todas las operaciones de criptografía se ejecuten dentro del entorno seguro de la tarjeta. Esto también incluye que todas las claves secretas están disponibles en texto plano solo dentro de la tarjeta inteligente.

- **El escenario central:** el módulo lógico se ubica en el servidor (es decir, el proveedor de almacenamiento) y proporciona el servicio de seudonimización.
  En este caso, la tarjeta inteligente solo actúa como token de autenticación que descifra la clave privada interna después de ser recuperada del almacén de claves central. La clave privada interna actúa como token de descifrado y se transfiere al lado del servidor donde se descifra la clave simétrica interna, recuperada del almacén de claves. La clave simétrica

[13] consultar la sesión 2.1

interna se utiliza para las operaciones criptográficas reales, por lo tanto, siempre permanece del lado del servidor.
En este caso, proponemos utilizar un módulo de seguridad de hardware (HSM) para mantener las claves protegidas durante el uso. Al igual que las tarjetas inteligentes, los HSM son dispositivos criptográficos a prueba de manipulaciones con soporte dedicado para algoritmos comunes, pero a un nivel de rendimiento considerablemente más alto.

Para la autenticación mutua se emplea un procedimiento de autenticación de desafío / respuesta que involucra el par de llaves externas del usuario y el par de llaves asimétricas del proveedor de almacenamiento. En el escenario central, la clave pública del proveedor de almacenamiento (o del servidor) también protege la clave privada interna del usuario durante la transferencia cliente / servidor. También en el escenario central con un HSM, la seguridad puede mejorarse aún más mediante el cifrado de las asignaciones de seudónimo / registro de identificador (ver Figura 1) con una clave lógica que solo conoce el HSM. En este caso, se deben cumplir las siguientes condiciones para poder recuperar un documento en particular:
(i) Usuario autenticado que proporciona su clave simétrica interna secreta,
(ii) el HSM con su clave lógica, y
(iii) el acceso al almacenamiento de metadatos, así como a los fragmentos de documentos seudonimizados.

2.2.3. Módulo de seguridad del Hardware (HSM)

Un módulo de seguridad de hardware es un dispositivo de seguridad basado en hardware que genera, almacena y protege claves criptográficas. Proporciona la base para una autoridad de certificación de campus seguro de alto nivel. Los módulos de certificación también están disponibles en software, pero un dispositivo de hardware proporciona un mayor nivel de seguridad.
Los beneficios de seguridad y rendimiento ofrecidos por un dispositivo de seguridad de hardware proporcionan un componente crítico en la administración y almacenamiento de claves privadas dentro de una infraestructura de seguridad. Los HSM también proveen la infraestructura necesaria para los sectores financiero, gubernamental y de salud para ajustarse a las normas industriales y regulatorias.
La seguridad de una autoridad de certificado depende de las herramientas adecuadas y de los procesos o políticas que utilizan dichas herramientas. Estas son algunas de las protecciones específicas que se pueden lograr al combinar un módulo de seguridad de hardware con prácticas y procesos institucionales efectivos.
El contenido del módulo de hardware se puede copiar a otros dispositivos de hardware. Si un conjunto de hardware se destruye, un conjunto de copias de seguridad permanece en un dispositivo de hardware duplicado o en un token o conjunto de tarjetas de repuesto dependiendo de su modelo HSM. La protección del contenido se logra con la combinación de las protecciones de hardware y buenas prácticas operativas. (CREN, 2001)

2.2.3.1. Personalizar el uso del HSM en el desarrollo

El uso de HSM y tarjetas inteligentes, como módulos criptográficos, funciona de la siguiente manera: Las funciones criptográficas de la tarjeta inteligente sólo se requieren para los pasos de autenticación del usuario y los cifrados y, los descifrados que implican la clave de sesión utilizada para proteger el canal de comunicación entre el usuario, su estación de trabajo y el servidor de seudónimos.

El HSM maneja las principales tareas criptográficas (ver **Figura 5**), incluyendo la parte del servidor de la autenticación del usuario, el cifrado/descifrado de los seudónimos, el descifrado de la clave simétrica interna del usuario con su clave privada interna y el cifrado de las acciones de respaldo.

Estas acciones se crean para distribuir de forma segura las partes de la clave privada interna del usuario a los titulares de acciones dedicados (operadores, administradores de sistema,..) para que se puede reconstruirla en caso de que la tarjeta inteligente esté perdida o dañada almacenando las claves del usuario (con la excepción de las claves privadas externas) en la base de datos y cargarlas cuando sea necesario en lugar de mantenerlas en el HSM en todo momento hace que la sustitución del HSM sea más fácil en caso de fallo o robo.

La clave privada externa del servidor todavía necesita copias de seguridad, por ejemplo, compartición de secretos de umbral similar a la copia de seguridad de las claves privadas internas del usuario.

2.3. Modelo de datos de nuestro sistema:

2.3.1. Descripción del modelo de datos:

Los modelos de datos se refieren a las interrelaciones lógicas y el flujo de datos entre diferentes elementos de datos involucrados en el mundo de la información. También documentan la forma en que se almacenan y recuperan los datos.

Y como podemos ver en la figura abajo (ver Figura 6) que representa el modelo de datos de nuestro sistema:

Al medio de la figura más arriba, están colocados los registros de identificación, relacionados por los dos lados con los seudónimos de Identificación correspondientes (Identificador Raíz/compartido: estaban explicados también en la sesión 3.1.3) de la forma siguiente:

(Relación a): Los registros de identificación están relacionados con los Identificadores raíces mediante una cordialidad de (**1: n**), donde cada Identificador raíz, representa la cabecera que contiene los datos de identificación de un seudónimo raíz, para un propietario de datos.
A su mismo cada Identificador raíz está relacionado con su correspondiente seudónimo raíz (**Relación b)** con una cordialidad (**1: 1**) mediante un enlace encriptado con la clave simétrica interna del propietario de datos. Y luego cada seudónimo raíz hace referencia (**Relación c)** a su correspondiente registro de datos del mismo propietario.

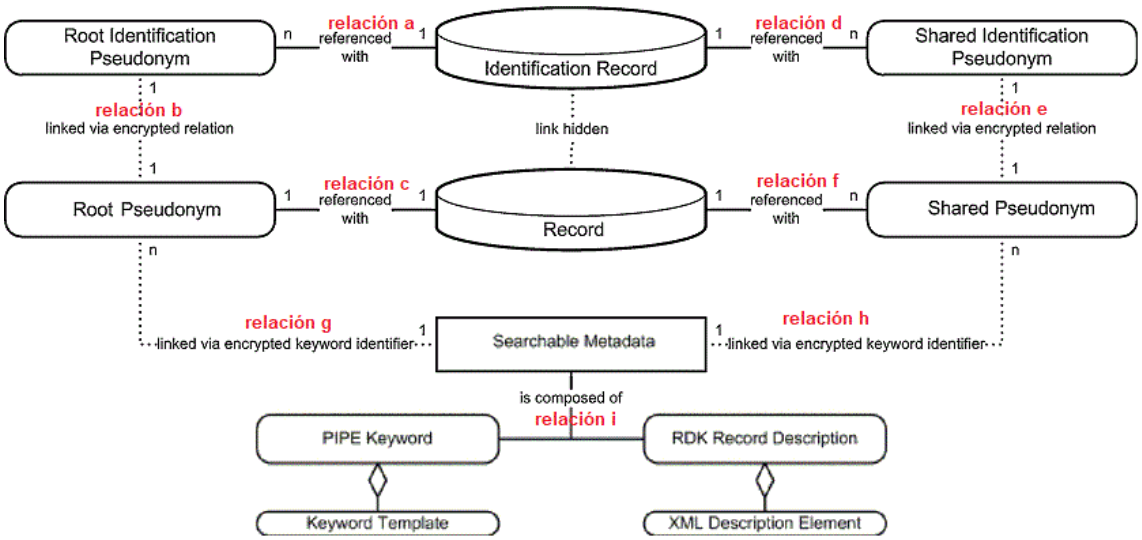


Figura 6. Módulo de datos (Hawaii International Conference on System Sciences, 2012)

(relación d) Por la derecha, el registro de identificación está relacionado con **n** Identificadores compartidos. Donde cada Identificador compartido está relacionado **(relación e)** con su correspondiente seudónimo compartido, mediante un enlace encriptado con la clave simétrica interna de ambos usuarios el propietario de los datos y del usuario autorizado, así ambos usuarios puedan descifrarlo utilizando sus correspondientes claves.

A su vez el seudónimo compartido hace referencia **(relación f)** a su correspondiente registro de datos, que en este caso el usuario autorizado puede acceder (ver Figura 6).
El vínculo entre un Identificador (seudónimo de identificación) y el registro de los datos está ocultado y viene representado por el enlace que está entre el Identificador y su correspondiente seudónimo.

(Relación g) y (Relación h): Los seudónimos (raíz/compartidos) están relacionados con el servidor de los metadatos mediante un enlace encriptado con la clave simétrica interna del usuario.
Tanto el propietario de datos como el usuario autorizado pueden buscar sus registros mediante las palabras claves que están guardadas como plantillas en el servidor de metadatos, donde cada seudónimo tiene asignado una palabra clave.
La misma palabra clave se asigna a los seudónimos referenciados al mismo registro, como podemos ver en la figura arriba **(Relación g) y (Relación h)** (ver Figura 6).

El servidor de los metadatos está compuesto por dos partes **Relación i)**, la primera que hemos dicho anteriormente donde se guardan las palabras claves que están asignada a los seudónimos, la segunda parte donde se guardan los archivos con formato XML, que contienen una descripción de los registros de los usuarios, estos archivos están encriptados con la clave simétrica interna corresponde a cada usuario, ya que pueden contener datos identificativos del usuario.

2.3.2. Metodología de consultas:

Para consultar a unos seudónimos particulares (y por tanto a registros), a cada seudónimo se le asigna una palabra clave específica (la misma palabra clave se asigna a los seudónimos raíz que

se hace referencia con el mismo registro de datos), que también puede utilizarla para los dimas seudónimos o registros, si es aplicable (ver Figura 6) (A.Shamir, 1979).

Y para evitar la sobrecarga de procesamiento de los datos cifrados, se propone almacenar las palabras clave en textos claros para que sean utilizables o compartidas por todos los usuarios, mientras que el identificador de palabras clave identifica el vínculo entre los seudónimos y las palabras clave.

De nuevo, el enlace (identificador de palabra clave o el mapeo de seudónimo) se cifra con las claves simétricas internas del propietario y / o del usuario autorizado. Ya que se pueden incluir información confidencial o de identificaciones personales, las palabras clave están altamente estructuradas y construidas a partir de plantillas de palabras clave predefinidas (tipo de documento (imagen, texto, etc.), mediante el uso de los archivos XML cifrados.

Estructura XML:

Suponemos que en los servidores de los metadatos se almacenan los archivos en formato XML, para el buen funcionamiento de consultas de datos en los registros, hay posibilidad de desarrollar un módulo de descripción de registros basado en XML (o RDK: Record Description Kit) (M. Schrefl, 2007) que permite el procesamiento de consultas y actualizaciones sobre documentos XML encriptados almacenados en servidores confiables, explotando la semántica estructural de registros XML para devolver ciertas partes del documento sin descifrar todo el registro, y utiliza los siguientes mecanismos:

A. El Etiquetado de esquema:

Un documento XML se define mediante una definición de esquema como un esquema XML o DTD. A cada elemento, atributo y nodo de texto del documento XML se le asigna una etiqueta única que codifica la semántica estructural de los elementos (información de ruta), lo que da como resultado un esquema de etiquetado de esquema. Este esquema de etiquetado permite consultar partes específicas del documento XML (Fragmentos) y por lo tanto nos da una mejora a la eficiencia de las consultas porque ciertas partes de consulta pueden procesarse sin tener acceso a la base de datos.

B. Estructura del índice:

Para garantizar la rapidez de las consultas ejecutadas frecuentemente, las estructuras de índice se crean en una sintaxis similar a XPath. Estas estructuras admiten consultas típicas tales como las consultas de rangos o consultas de información estructural.

- Desarrollamos la estructura de índice de los Seudónimos con sus correspondientes registros usando XPATH:

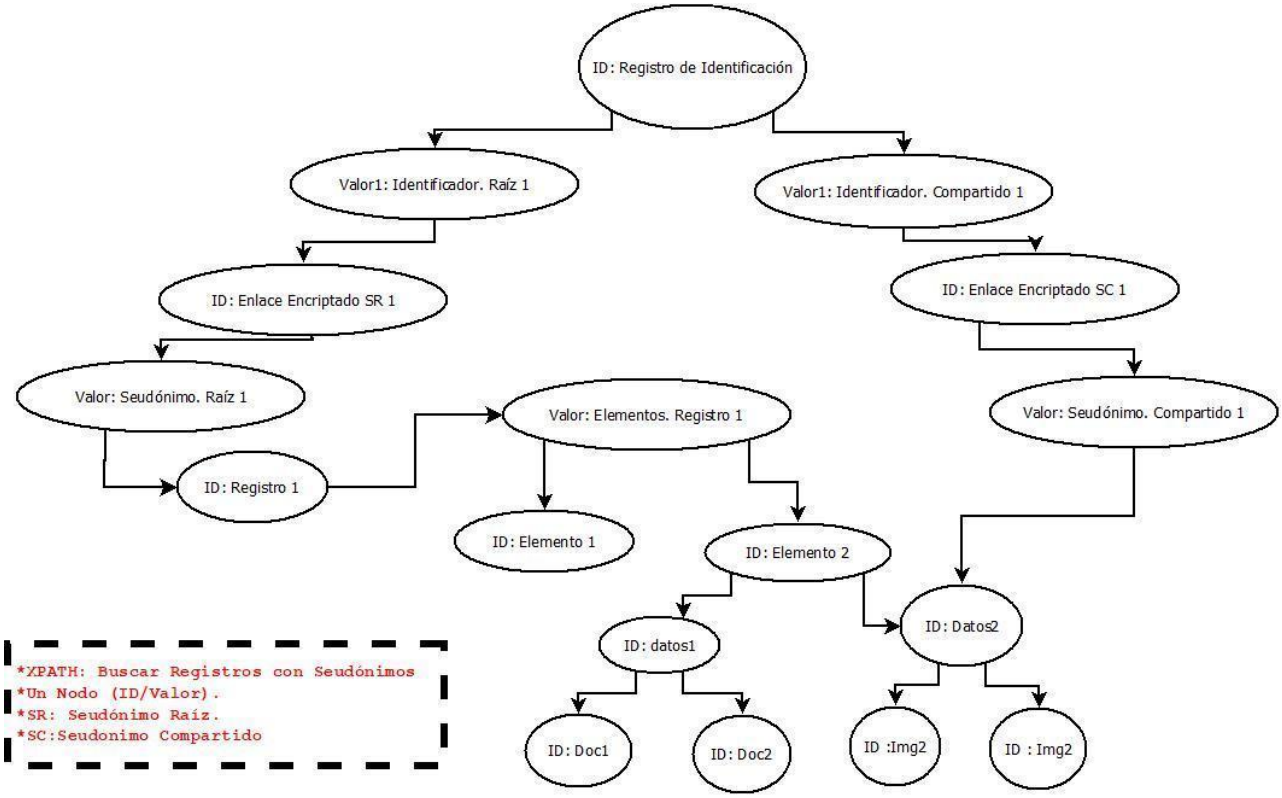


Figura 7. Búsqueda de Registros usando XPATH en XML

Suponemos que un propietario de datos quiere acceder a su registro de datos (Nodo: Registro1), que contiene dos elementos (Nodos: Elemento1 y Elemento2) cada elemento contiene datos (Nodos: datos1 y datos2), en este caso el Nodo de (datos1) contiene dos documentos (Nodos: Doc1 y Doc2) y el Nodo (datos2) contiene dos Imágenes (Nodos: Img1 y Img2).
El Mismo propietario para este caso, ha autorizado a un usuario (Usuario Autorizado) a acceder a la carpeta que está en el nodo (datos2) del segundo elemento (elememto2) de su registro de datos. (ver-Figura 7).

De forma que el usuario autorizado sólo puede acceder al sub-registro en el nodo (datos2).
A continuación, se muestra el código que hemos desarrollado con la sintaxis de XPATH:

```
Registro_de_Identificación id="Registro_ID" >
       <Identificador_Raíz_1 id= "enlace_Encreptado_SR1">
       <EncryptedData xmlns='http://enlace_Encreptado_SR1#'
Type='http://Encripted_Link'>
        <CipherData>
          <CipherValue>A23B45C56</CipherValue>
        </CipherData>
     </EncryptedData>

              <Seudónimo_Raíz_1 id= "registro 1" contenido="elementos_SR1">
                     <Elemento1 id="elemento1">
                            //Vacio;
                     </Elemento1>
                     <Elemento2 id="elemento2" contenido="datos">
                            <Datos1 id="datos1" contenido="documentos">
                                   <Doc1 id="doc1">
                                          contenido= doc1.docx;
                                   </Doc1>
                                   <Doc2 id="doc2">
                                          contenido= doc2.docx;
                                   </Doc2>
                            </Datos1>
                            <Datos2 id="datos2" contenido="imagenes">
                                   <Img1 id="img1">
                                          contenido= Img1.docx;
                                   </Img1>
                                   <Img2 id="img2">
                                          contenido= Img2.docx;
                                   </Img2>
                            </Datos2>
                     </Elemento2>
              </Seudónimo_Raíz_1>
       </Identificador_Raíz_1>

       <Identificador_Compartido_1 id= "enlace_Encreptado_SC1">
              <EncryptedData
xmlns='http://enlace_Encreptado_SC1#'Type='http://Encripted_Link'>
        <CipherData>
          <CipherValue>A23B45C56</CipherValue>
        </CipherData>
     </EncryptedData>
              <Seudónimo_Comparido_1 id= "Datos2" contenido="Datos2">
                            <Datos2 id="datos2" contenido="imagenes">
                                   <Img1 id="img1">
                                          contenido= Img1.docx;
                                   </Img1>
                                   <Img2 id="img2">
                                          contenido= Img2.docx;
                                   </Img2>
                            </Datos2>
              </Seudónimo_Comparido_1>
       </Identificador_Compartido_1>
                     </Registro_de_Identificación>
```

Figura 8. estructura del índice mediante la sintaxis XPATH

C. Estructura de almacenamiento de XML:

Los documentos XML se almacenan fragmentados en pares (clave, valor), con etiquetas como claves hash criptográficas y el elemento XML se corresponde como un valor cifrado.
La fragmentación depende del tamaño de las partes o trozos de información utilizadas dentro del documento. El cifrado y el descifrado se puede llevar a cabo con algoritmos criptográficos simétricos estándar (por ejemplo, AES). Esto garantiza que la estructura y el contenido del documento XML se oculte a los no autorizados. Los datos almacenados en este formulario incluyen estructuras de índice, metadatos, como información de esquema, así como los documentos XML reales.

Si fijamos en al ejemplo desarrollado en el punto anterior, podemos observar lo que hemos explicado en el párrafo anteriormente (ver-Figura 11):

```
<Identificador_Compartido_1 id= "enlace_Encreptado_SC1">
    <EncryptedData xmlns='http://enlace_Encreptado_SC1#'Type='http://Encripted_Link'>
      <CipherData>
        <CipherValue>A23B45C56</CipherValue>
      </CipherData>
    </EncryptedData>
    <Seudónimo_Comparido_1 id= "Datos2" contenido="Datos2">
          <Datos2 id="datos2" contenido="imagenes">
              <Img1 id="img1">
                  contenido= Img1.docx;
              </Img1>
              <Img2 id="img2">
                  contenido= Img2.docx;
              </Img2>
          </Datos2>
    </Seudónimo_Comparido_1>
</Identificador_Compartido_1>
```

Figura 11: Ejemplo de fragmentos - XML cifrado

De esta forma, para que un usuario pueda acceder al enlace del seudónimo compartido tienes que ser autorizado, así puede descifrar el enlace para poder ver el contenido de registro (**datos2**) con que está referenciado el seudónimo. Podemos ver en este ejemplo también que el registro (**datos2**) está dividido en dos fragmentos **[**clave: **Datos2;** valor: **Imágenes (**Img1, Img2**)].**

3. Diagramas de Flujo y de Secuencias UML:

3.1. Autentificación del Usuario:

3.1.1. Diagrama de flujo para la Autentificación del Usuario:

La autenticación de usuario (ver **Figura 8**) implica la autenticación mutua del usuario que utiliza la tarjeta inteligente con el servidor de autenticación, usando su par de claves externas y el par de *nonces o tokens* del servidor (números seleccionados aleatoriamente por el servidor y se usan una vez, y tienen funcionamientos del PIN) como desafíos de usuario / servidor.

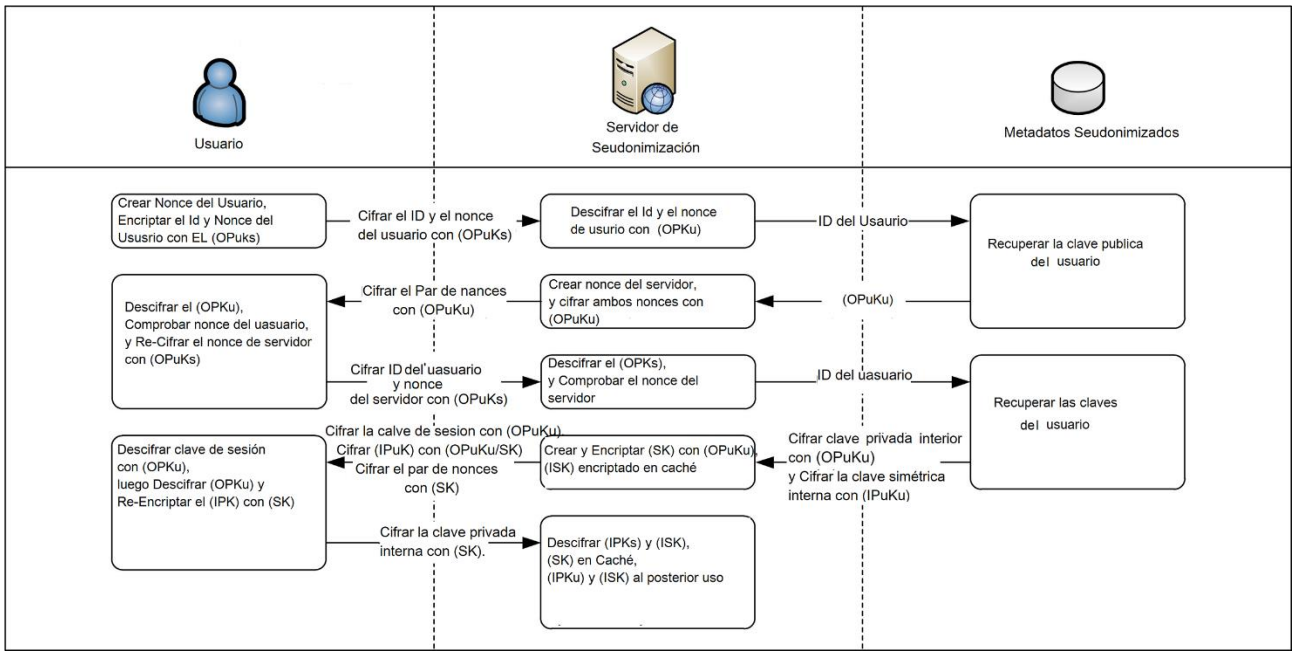


Figura 8. Diagrama de flujo de *Autenticación del Usuario*

Una vez que se confirman ambas identificaciones, la clave privada interna del usuario se recupera de la base de datos de seudonimización y se transfiere a la tarjeta inteligente del usuario para ser descifrada con la clave privada externa del usuario.

Con la clave privada interna descifrada, la clave simétrica interna del usuario se puede descifrar dentro del HSM en el servidor de seudonimización y se almacenan en caché para unas operaciones adicionales junto con la clave privada interna del usuario. Además, se genera una clave de sesión (SK) en el HSM y de forma segura.

3.1.2. Diagrama de secuencias UML para la Autentificación del Usuario:

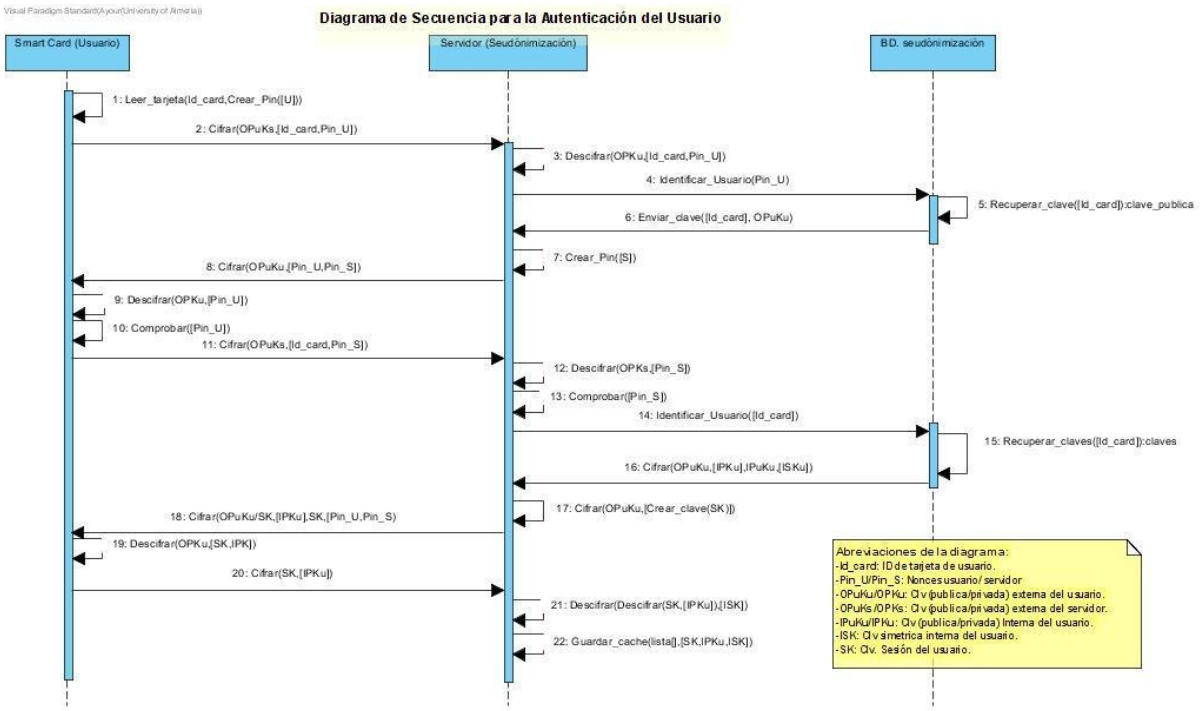


Diagrama 1: Diagrama de secuencias UML para la Autentificación del Usuario

3.1.2.1. Descripción de los métodos:

- **1: Leer_tarjeta(id_card , crear_Pin(u)):**

Para acceder al sistema se requiere una autenticación con la tarjeta inteligente, este proceso de verificación de la validación de tarjeta empieza cuando el usuario mete su tarjeta en el lector correspondiente a la misma. El método tiene dos parámetros: el número de tarjeta y el pin del usuario, que tecla el usuario por teclado. Al final de este proceso, el sistema toma una decisión con respecto a la tarjeta, dependiendo de si es válida o no.

Cuando se detecta que la tarjeta no está registrada en el sistema de verificación o el pin que teclea el usuario es incorrecto, se genera un mensaje de error en la pantalla para informar al usuario.

Al contrario, la validación de la tarjeta insertada significa el paso al siguiente método del proceso de autentificación del usuario.

- **2: cifrar(OPuKs, [id_card,Pin_U]):**

Mediante este método se realiza el cifrado del número de la tarjeta del usuario (**id_card**) y el código tecleado por el usuario (**Pin_U**), mediante la clave pública externa del servidor de seudonimización.

Con la idea de recuperar las claves públicas externas correspondientes a la tarjeta de autentificación del mismo usuario en la base de datos.

- **3: Descifrar (OPuKs, [id_card,Pin_U]):**

A nivel del servidor de seudonimización se descifran los cardinales de la tarjeta (**id_card,Pin_U**), con la misma clave pública externa del servidor.

- **4: Identificar_uasurio(Pin_U):**

Es un método que sirve para consultar las claves públicas de la tarjeta correspondiente al usuario en la base de datos de la seudonimización.

- **5: Recuperar_clave(id_card):**

A nivel de la base de datos, empieza el proceso de la recuperación o la búsqueda de la clave pública externa del usuario, pasando como parámetro el número de la tarjeta como un identificador.

- **6: Enviar_clave(id_card):OPuKu:**

Cuando se recupera la clave pública del usuario en la base de datos, se envía al servidor de seudonimización, mediante este método, pasando el número de la tarjeta como parámetro.

- **7: Crear_Pin([s]):**

El servidor en este momento genera un código pin, como nonce o token del servidor.

- **8: Cifrar(OPuKu,[Pin_U,Pin_S]):**

El servidor cifra el par de nonces [Pin_U,Pin_S] mediante la clave pública externa del usuario.

- **9: Descifrar (OPKu, [Pin_U]):**

En el sistema cliente del usuario se descifra el nonce del usuario, mediante la clave privada externa del usuario (OPKu).

- **10: comprobar(Pin_U):**

El método de comprobación de nonce del usuario (Pin_U) tecleado por el mismo en el sistema de cliente.

- **11: Cifrar(OPuKs,[id_card, Pin_S]):**

Posteriormente, se cifran el número de la tarjeta id_card y el nonce de servidor Pin_S mediante la clave pública externa del servidor.

- **12: Descifrar (OPks, [Pin_S]):**

A nivel del servidor de seudonimización se descifra el nonce del servidor mediante su clave privada externa.

- **13: comprobar (Pin_S):**

Después descifrar el Pin_S, pasa al proceso de la comprobación del mismo, mediante el sistema del servidor.

- 14: **identeficar_usuario (id_card):**

Consultar las claves del usuario en la base de datos de seudonimización.

- 15: **recuperar_claves (id_card):**

Al recibir una petición de consultar las claves del usuario con su id_card, este método encarga de recuperar la clave privada interna (IPKu), la clave simétrica interna (ISKu) y la clave de sesión (SK)en la base de datos.

- **16: cifrar (OPuKu, [IPKu], IPuKu,[ISKu])**

Después de recuperar la IPKu y la ISKu del usuario se cifran en este método mediante la OPuKu y la IPuKu respectivamente.

- **17: cifrar (OPuKu, [SK]):**

Luego se cifra la clave de sesión con la clave pública externa del usuario y se entregan al nivel del servidor de seudonimización.

- **18: cifrar ({OPuKu,Sk},[IPKu],{SK},[pin_U,pin_S]:**

Con la clave pública externa y clave de sesión, a nivel del servidor, se cifra la clave privada interna del usuario, en otro lado solo con la clave de sesión se cifran los nonces del usuario/servidor.

- **19: Descifrar (OPKu, [SK,IPKu]):**

El sistema de cliente del usuario descifrar la clave de su sesión y la clave privada interna para usarlas.

- **20: cifrar (SK, [IPKu]):**

Posteriormente se cifra la clave privada interna del usuario con su clave de sesión.

- **21: descifrar (descifrar (SK, [IPKu]), [ISKu]):**

Después de descifrar la clave privada interna del usuario con la clave de la sesión, se descifra la clave simétrica interna del usuario.

- **22: Guardar_cache (lista[], SK, IPKu,ISKu):**

Se guardan los claves SK, IPKu y ISKu en la cache para el futuro uso.

3.2. Recuperación de Registros de datos

3.2.1. Diagrama de flujo para recuperar Registros:

Para recuperar un registro de datos en particular (ver **figura 7**), en primero el usuario debe consultar los seudónimos cifrados particulares, creando una palabra clave utilizando las plantillas de palabras clave, recuperando el identificador de palabra clave correspondiente y consultando el identificador cifrado para

encontrar los seudónimos cifrados coincidentes, Es decir, las asignaciones cifradas de seudónimo asociadas con el identificador de palabra clave cifrado.

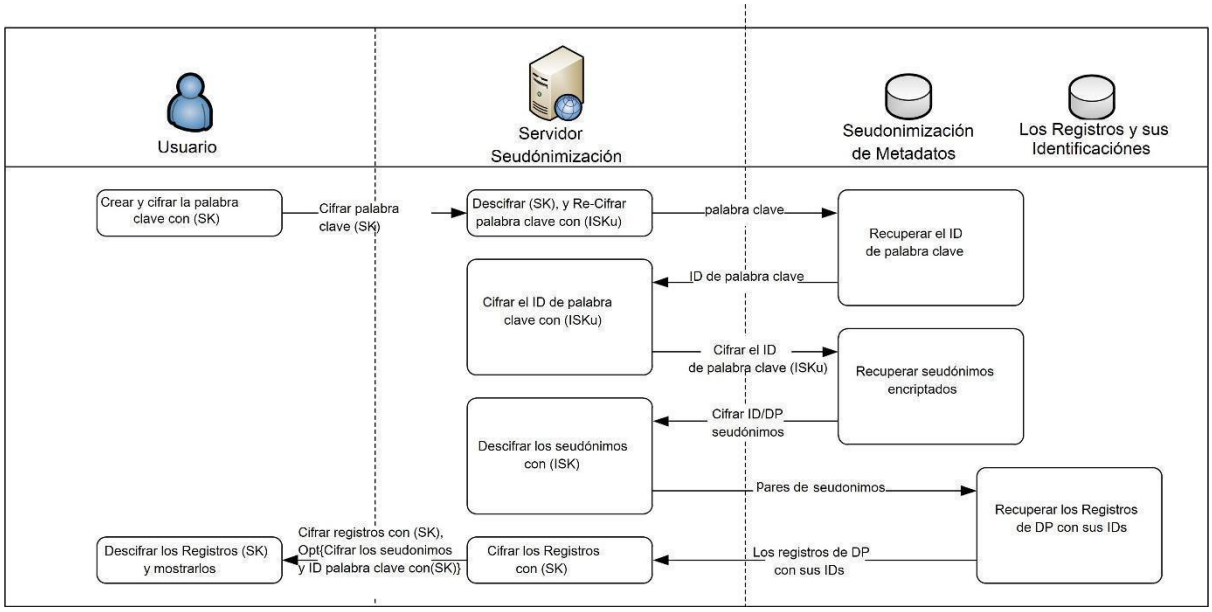


Figura 7. Diagrama de flujo de *Recuperación de Registros*

Los pares de seudónimo se descifran con la clave simétrica interna del usuario y los seudónimos de texto sin formato se utilizan para recuperar la identificación de los registros y sus registros correspondientes, que se transfieren y se muestran al usuario.
Opcionalmente, los seudónimos y el identificador de palabras clave también se transfieren al usuario.

3.2.2. Diagrama de secuencias UML para recuperar Registros:

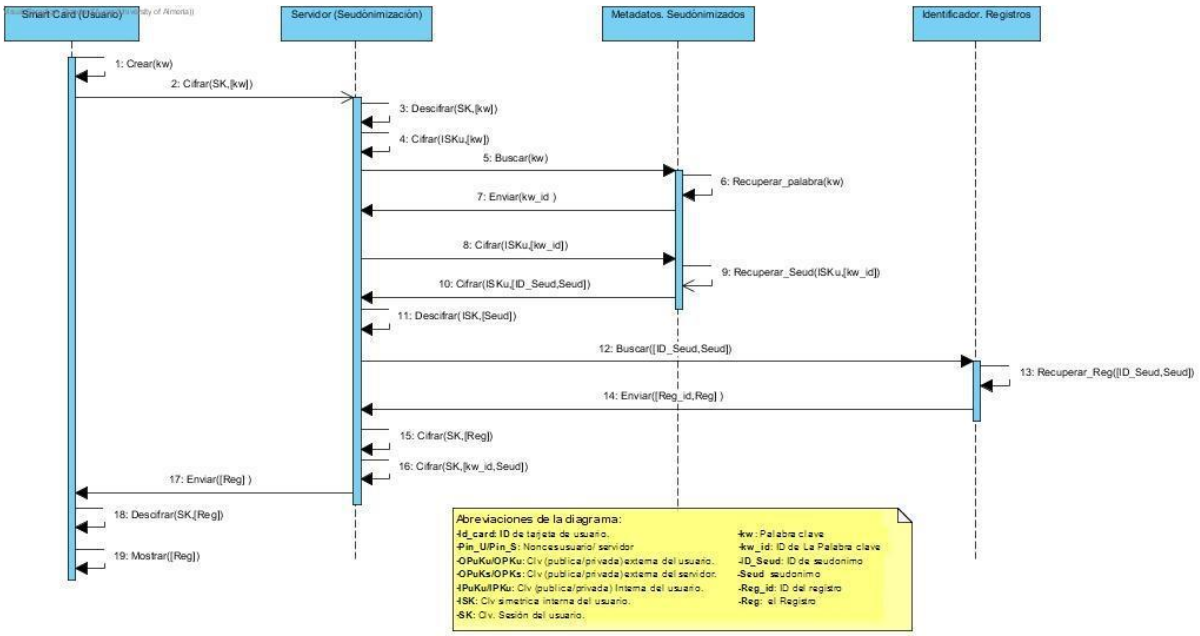


Diagrama 2: Diagrama de secuencias UML para recuperar Registros:

3.2.2.1. Descripción de los métodos:

En los puntos siguientes describimos los distintos métodos que forman parte del diagrama de secuencia arriba para la recuperación de los registros de datos en la base de datos de seudonimización de Metadatos:

- **La recuperación del identificador de la palabra clave:**

- **1: crear (kw):**

Crear una palabra clave utilizando plantilla de palabras claves recomendada por el sistema, el método tiene como parámetro de entrada una palabra de clave (kw) teclado por el usuario.

- **2: cifrar (SK,[kw]):**

Después de crear una palabra de clave, se cifra mediante la clave de sesión (SK) del usuario

- **3: descifrar (SK,[kw]):**

En el servidor de seudonimización, se descifra la palabra clave que ha creado el usuario para usarla.

- **4: cifrar (ISKu,[kw]):**

Antes de consultar y buscar el identificador correspondiente a esta palabra clave en la base de datos de seudonimización del metadato, se cifra la palabra clave con la clave simétrica interna del usuario (ISKu).

- **5: buscar (kw):**

Un método de búsqueda de identificador de palabra clave en BD de Metadatos, y tiene como parámetro de entrada (kw).

- **6: recuperar_palabra (kw):**

En la BD de Metadatos se recupera el identificador de la palabra clave, pasando kw como parámetro de entrada.

- **7: enviar (id_kw):**

Posteriormente, se envía el identificador encontrado correspondiente a la palabra clave, al servidor de seudonimización.

❖ **La recuperación del par de seudónimos correspondiente a id de la palabra clave:**

- **8: cifrar (ISKu,[id_kw]):**

Ahora viene la etapa de recuperar el par de seudónimo correspondiente al identificador de palabra clave (id_kw), en este método se cifra el (id_kw) mediante la clave simétrica interna del usuario.

- **9: recuperar_seud (ISKu,[id_kw]):**

Después de descifrar el id_kw con la ISKu a nivel de BD de metadatos, se empieza a recuperar el par de seudónimo (id seudónimo y el seudónimo de texto) correspondiente al identificador de la palabra clave (id_kw).

- **10: cifrar (ISKu,[id_seud,seud]):**

Mediante este método, se cifra el par de seudónimos (id seudónimo y el seudónimo de texto) con la clave simétrica interna del usuario, y se envía al servidor de seudonimización.

- **11: descifrar (ISKu,[id_seud,seud]):**

En el servidor se descifra el par de seudónimos con la clave ISKu, para usarlo luego en la recuperación del registro correspondiente.

❖ **La recuperación de los registros:**

- **12: buscar (id_seud,seud):**

Mediante el par de seudónimos, se busca el registro o registros relacionado con este par de seudónimos en la base de datos de identificadores y registros.

- **{13: recuperar_reg (id_seud,seud)} y {14: enviar(id_reg, reg)}:**

El método {13} recupera los registros correspondientes al par de seudónimos a nivel de la base de datos de registros, y posteriormente con el método {14} se envían los registros recuperados al servidor de Seudonimización.

- **15: cifrar (SK,[reg]) y 16: cifrar (SK, [id_kw, [id_seud,seud]])**

A nivel del servidor de Seudonimización, el primer método cifra los registros recuperados mediante la clave SK, y el segundo método, cifra los seudónimos y el identificador de palabra clave con la misma clave SK.

- **17: enviar (reg):**

Este método se encarga de enviar los registros recuperados al sistema cliente del usuario solicitante.

- **18: descifrar (ISK, [reg]):**

En el sistema del usuario se descifran los registros recuperados mediante la clave de sesión.

- **19: mustrar (reg)**:

Y finalmente se muestran los registros solicitados y se visualizan al usuario solicitante.

3.3. Autorización del Usuario:

3.3.1. Diagrama de flujo para crear una autorización a un usuario:

Para que un usuario Autorizado tenga el acceso a enlazar el vínculo entre el registro de identificación del propietario de datos y un registro de sus datos personales particular (ver **Figura 8)**, se crea un nuevo par de seudónimo compartido como relación de autorización.

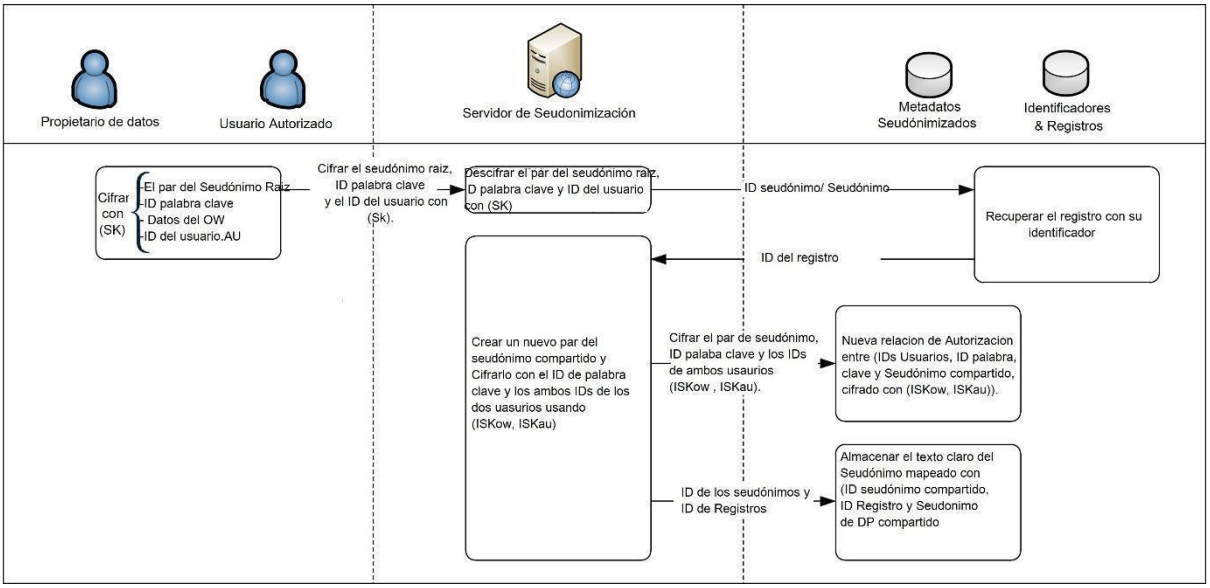


Figura 8. Diagrama de flujo de Autorización del usuario

El propietario de datos primero tiene que recuperar el par de seudónimo raíz y el identificador de palabra clave correspondiente al registro de datos a compartir. Además, tanto el propietario como el usuario autorizado tienen que ser autenticados en la misma estación de trabajo, de manera que ambos identificadores del usuario estén conocidos en interface de trabajo cliente, mientras que ambas claves simétricas internas se almacenan en la caché del HSM del servidor de seudonimización.

El par de seudónimo raíz se transfiere al servidor de seudonimización junto con los identificadores de usuario y el identificador de palabra clave y los identificadores de registro correspondientes se recuperan utilizando las asignaciones de texto claro de registro / seudónimo.

A continuación, el servidor selecciona aleatoriamente un nuevo par de seudónimos compartidos, que primero se cifra con las claves simétricas internas de ambos usuarios (junto con los IDs de los usuarios y el identificador de palabra clave) y luego los almacena en la base de datos como relación de autorización. Finalmente, los seudónimos de texto claro se hacen referencia a continuación con los identificadores de registros recuperados para crear dos nuevas asignaciones de registro / seudónimo.

3.3.2. Diagrama de secuencias UML para crear una autorización a un usuario:

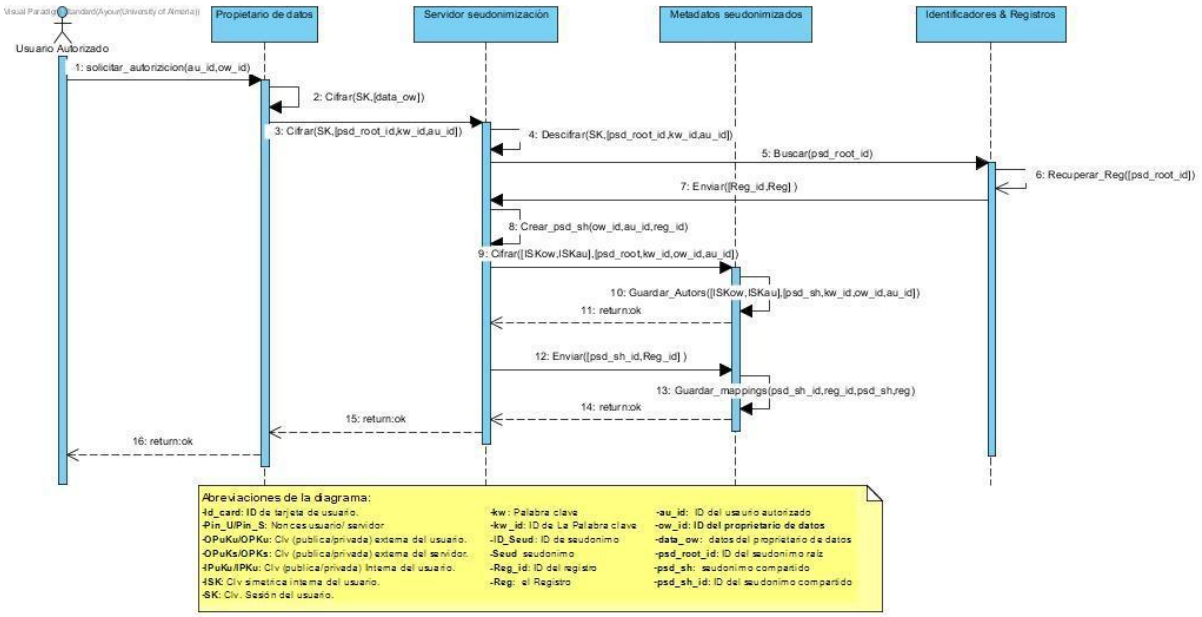


Diagrama 3: Diagrama de secuencias UML para crear una autorización a un usuario:

3.3.2.1. Descripción de los métodos:

- **1: solicitar_autorizacion (au_id,ow_id):**

El usuario autorizado manda una solicitud al propietario de datos mediante su sistema cliente, el método tiene como parámetro los identificadores de los dos usuarios, *au_id* como identificador del usuario autorizado, y *ow_id* como identificador del propietario de datos.

- **2: cifrar(sk,[data_ow]):**

Este método se eliminará

- **3: cifrar (SK, [psd_root_id, kw_id,ow_id,au_id]):**

Antes de mandar la petición al servidor de seudonimización, para recuperar los registros vinculados con el seudónimo raíz, este método se encarga de cifrar los identificadores de seudónimo raíz, la palabra clave y los dos usuarios respectivamente mediante la clave de sesión *SK*.

- **4: descifrar (SK, [psd_root_id,kw_id,au_id]):**

A nivel del servidor se descifran los identificadores del seudónimo raíz, palabra clave y el usuario autorizado respectivamente, para el siguiente uso.

- **5: buscar (psd_root_id):**

El servidor consulta los registros vinculados con el seudónimo raíz, en la base de datos de los registros e identificadores, el método lo pasamos el identificador del seudónimo como parámetro.

- **6: recuperar_reg (psd_root_id):**

En la Base de datos de registros e identificadores, se recuperan los registros vinculados con el seudónimo que pasamos como parámetro.

- **7: enviar (reg_id, reg):**

Después de localizar los registros vinculados con el seudónimo, se envían los registros con sus correspondientes identificadores al servidor de seudonimización.

- **8: crear_psd_sh (ow_id,au_id, reg_id):**

Después de recuperar los registros solicitados, viene el proceso de formar una relación de autorización entre el solicitante de registros y el propietario de los mismos registros, mediante la creación de un seudónimo compartido como una autorización a acceder a los registros solicitados, pasando como parámetros los dos identificadores de usuarios y el identificador del registro solicitado.

- **9: cifrar (ISKow,ISKau, [psd_root,kw_id,ow_id,au_id]):**

Posteriormente, se cifran los siguientes elementos: seudónimo raíz, identificador de palabra clave y los dos identificadores del usuario autorizado y propietario de datos respectivamente con las dos claves simétricas internas de ambos usuarios.

- **10: guardar_autors (ISKow,ISKau, [psd_sh,kw_id,ow_id,au_id]):**

En la base de datos de Metadatos de Seudonimización, se guarda la autorización creada anteriormente, cifrando el seudónimo compartido, el identificador de palabra clave, el ID de propietario de datos y el ID del usuario autorizado con la clave simétrica interna de ambos usuarios.

- **11: return: ok:**

Se manda una confirmación al servidor de que se ha guardado la relación de autorización con éxito.

- **12: enviar (psd_sh_id, reg_id):**

El servidor de Seudonimización envía una petición mediante el identificador de seudónimo compartido y el identificador de del registro recuperado, para crear una referencia entre seudónimos de texto claro con los correspondientes registros recuperados.

- **13: guardar_mapings (psd_sh_id, reg_id,psd_sh,reg):**

Finalmente, se guarda el mapping de la asignación de registro/seudónimos, pasando los siguientes parámetros, el seudónimo compartido y el registro recuperado con sus correspondientes identificadores.

- **14: return: ok:**

Se manda una confirmación al servidor de que se ha guardado con éxito la relación de asignación de registro/seudónimos.

- **15: {return: ok: y 16: return: ok}:**

Los usuarios reciben una confirmación de que, se ha permitido al usuario autorizado, a acceder a los registros solicitados.

3.4. Afiliación del Usuario:

3.4.1. Diagrama de flujo para la Afiliación del Usuario:

Al igual que con las autorizaciones, una afiliación de usuario (ver **Figura 9** Figura 8) requiere que tanto el propietario de datos y el Personal de confianza como usuario afiliado sean autenticados en la misma estación de trabajo.

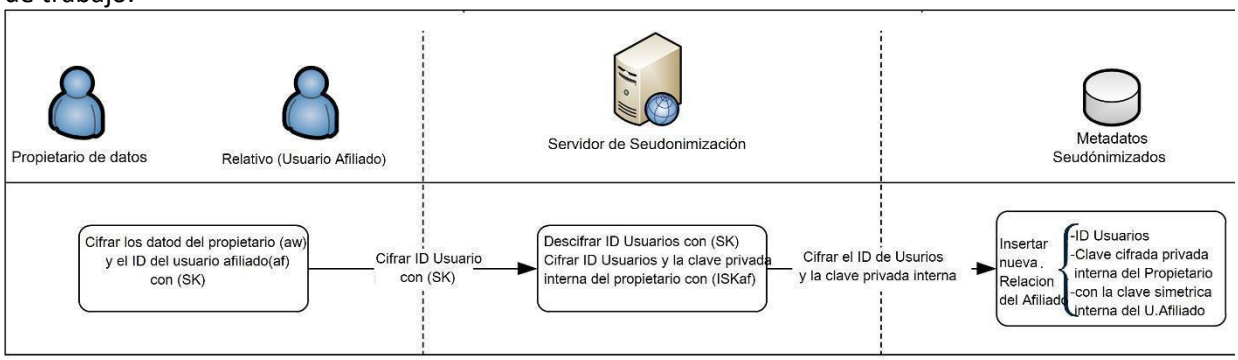


Figura 9. Diagrama de flujo de Afiliación del usuario

A continuación, ambos identificadores de usuario se transfieren al servidor de seudonimización donde se cifran con las claves simétricas internas de ambos usuarios. Finalmente, todos los elementos se guardan en el almacenamiento de metadatos de seudonimización como relación de afiliación.

3.4.2. Diagrama de secuencia UML para la Afiliación del Usuario:

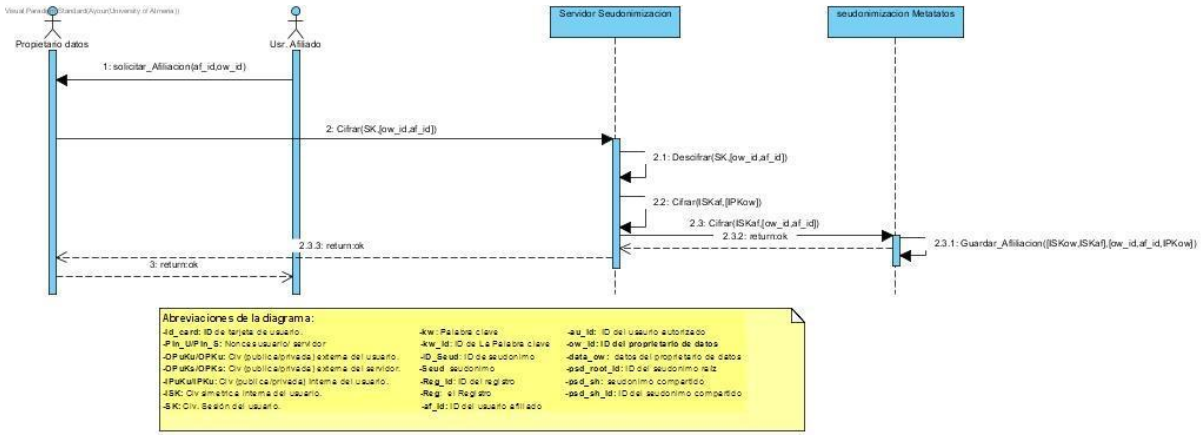


Diagrama 4: Diagrama de secuencia UML para la Afiliación del Usuario:

5.4.2.1. Descripción de los métodos

- **1: solicitar_afiliacion(af_id,ow_id):**

Para estabilizar una relación de afiliación de un usuario de confianza con el propietario de usuario, este último manda una petición al propietario de datos, mediante su sistema cliente. Este método tiene dos parámetros de entrada, el *ID* de usuario afiliación y el *ID* de usuario del propietario de datos.

- **2: cifrar (SK, [ow_id,af_id]):**

Antes de transferir la petición de afiliación al servidor de seudonimización, se cifran los dos identificadores de ambos usuarios mediante la clave de sesión.

- **2.1: descifrar (SK, [ow_id,af_id]):**

A nivel del servidor de seudonimización, se descifran los identificadores de ambos usuarios con la clave de sesión para el siguiente uso.

- **2.2: cifrar (SKaf, [IPKow]):**

Un paso muy importante para el proceso de afiliación es la encriptación de la clave privada interna del propietario de datos mediante la clave semítica interna del usuario afiliado a nivel del servidor. Esto significa que se está estabilizando una relación de confianza.

- **2.3. Cifrar (ISKaf, [ow_id,af_id]):**

Antes de guardar esta relación en BD Metadatos, se protegen los identificadores de ambos usuarios con la clave semítica interna del usuario afiliado, y después se mandan a la base de datos de Metadatos.

- **2.3.1. Guardad_afiliacion ({ISKow,ISKaf}, [ow_id,af_id,IPKow])**

Finalmente, se guarda la relación de afiliación, cifrando los dos identificadores de ambos usuarios y la clave privada interna de propietario de datos mediante las claves semíticas de ambos usuarios (el afiliado y el propietario de datos).

- **2.3.2. Return ok:**

Devolver y enviar el "ok" al servidor de seudonimización después de guardar una relación de afiliación en la base de datos de Metadatos.

- **2.3.3. Return ok:**

El mismo funcionamiento del método anterior, el servidor envía el "**ok**" al propietario de datos, de que a esta creado una relación del mismo con un usuario de confianza (afiliado).

- **3. Return ok:**

El propietario de datos manda un "**ok**" al usuario afiliado.

3.5. Almacenamiento de datos

3.5.1. Diagrama de flujo para el Almacenamiento de datos

Desde el punto de vista del propietario de datos, el almacenamiento de datos personales ***(Figura 11)*** requiere que se obtenga un par de seudónimos raíz, como referencia al identificador del registro. Además, el mismo propietario de datos crea una nueva palabra clave e introduce el nuevo registro de datos en la estación de trabajo.

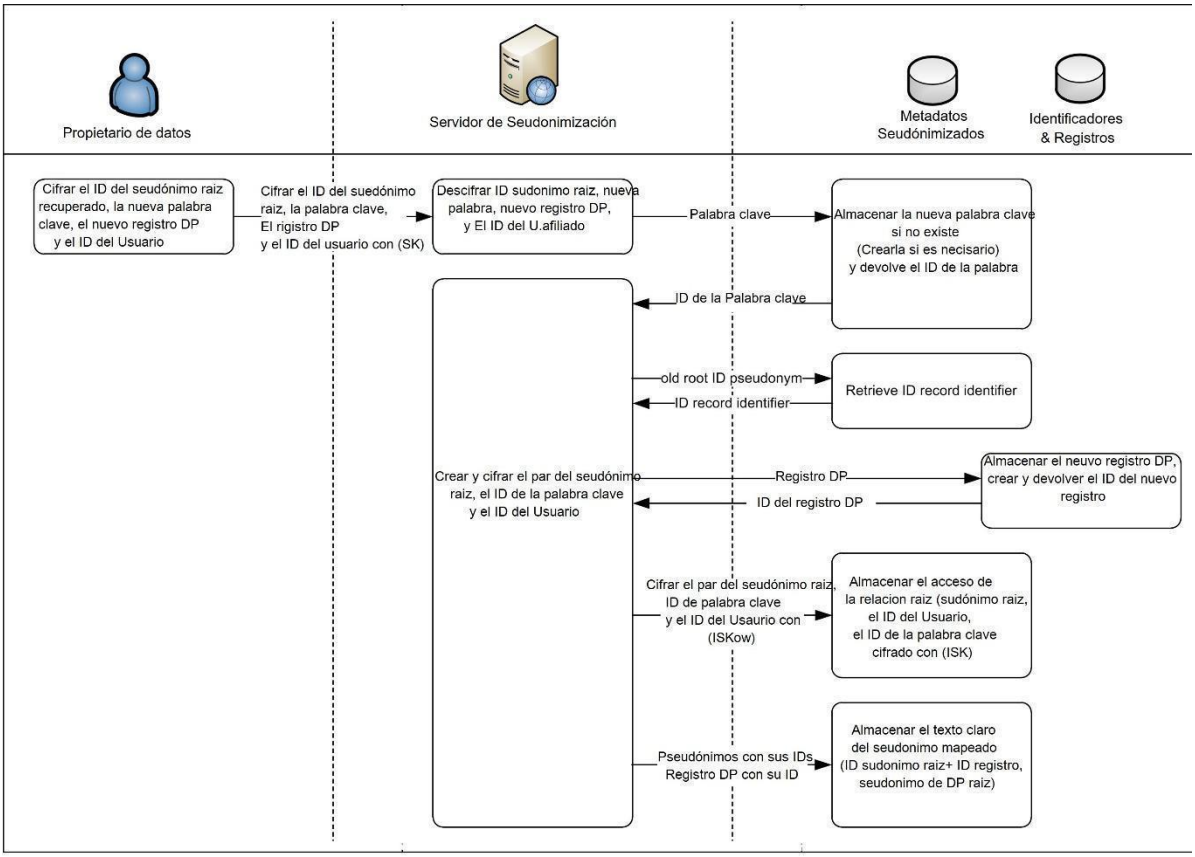


Figura 11. Diagrama de flujo de Almacenamiento de datos personales

A continuación, el seudónimo, la nueva palabra clave, el nuevo registro de datos y el identificador de usuario se transfieren al servidor de seudonimización, donde se almacena la palabra clave (y su identificador se determina por el motor de base de datos) y también el identificador de registro recuperado. El nuevo registro se almacena en la base de datos de registros y su identificador de registro devuelto al servidor. A continuación, el servidor crea un nuevo par de seudónimo raíz y lo almacena encriptado con el identificador de palabra clave y el identificador de usuario como acceso *Root*.

3.5.2. Diagrama de secuencia UML para el Almacenamiento de datos

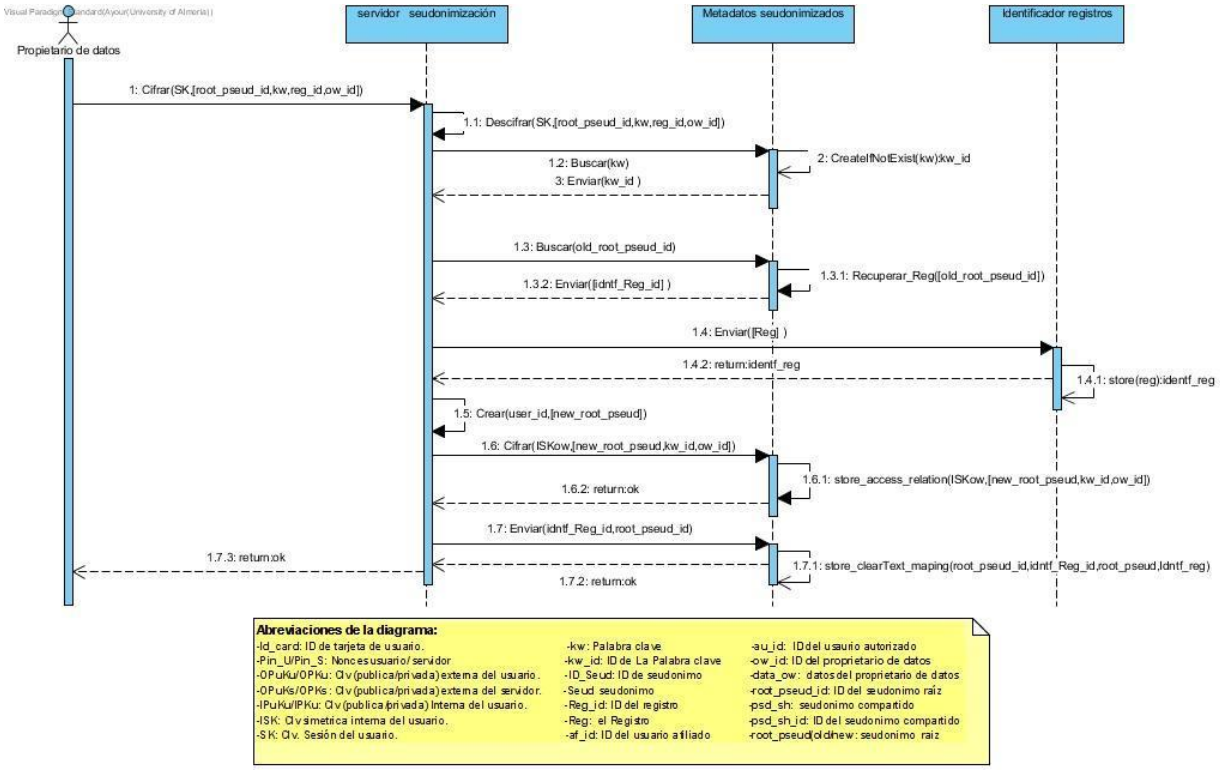


Diagrama 5: Diagrama de secuencia UML para el Almacenamiento de datos

3.5.2.1. Descripción de los métodos:

❖ La creación del identificador de una nueva palabra clave en BD:

➢ 1. Cifrar (SK, [root_pseud_id,kw,reg_id,ow_id]):

Antes de crear la palabra clave que refiere al nuevo registro a almacenar, este método encarga de cifrar el identificador de seudónimo raíz, el texto de palabra clave a almacenar, el identificador del registro actual del propietario de datos y el identificador del mismo usuario (propietario de datos) mediante la clave de sesión de este último, con la idea de protegerlos antes de mandarlos al servidor de seudonimización.

➢ 1.1. Descifrar (SK, [root_pseud_id,kw,reg_id,ow_id]):

A nivel del servidor de seudonimización, se descifran los elementos encriptados en el método anterior, mediante la clave de sesión.

➢ 1.2. Buscar (kw):

Es un método de consultar la existencia y solicitar la registración de esta nueva palabra clave si no existe, asignamos como parámetro de entrada el texto de la palabra clave.

➢ 2. CreateNotExiste (wk): kw_id

Este método devuelve el identificador de la palabra clave consultada por el servidor si está en la base de datos, y se almacena en el caso de la inexistencia en la *DB*.

➢ 3. Enviar (kw_id):

Se envía el identificador de la nueva palabra clave al servidor de la seudonimización.

❖ La recuperación del esquema de los registros y los seudónimos raíces antiguos del usuario:

➢ 1.3. Buscar (old_root_pseud_id):

Para conocer el esquema actual de los registros del propietario de datos se manda una petición de consulta a la base de datos, asignando como parámetro de entrada, el identificador del seudónimo raíz actual.

➢ 1.3.1. Recuperar_Reg (old_root_pseud_id):

Se recupera el registro actual del usuario, mediante su correspondiente identificador de seudónimo.

➢ 1.3.2. Enviar (reg_id):

La base de datos de identificadores y registros envía el *ID* de registro consultado.

❖ Almacenamiento del nuevo registro, y la creación de su correspondiente par de seudónimos

➢ **Los métodos**: {1.4: enviar(reg)}, {1.4.1: store(reg):reg_id} y {1.4.2: return():reg_id}

El servidor manda una petición para almacenar el registro en la base de datos (1.4), después de almacenarlo (1.4.1), la base de datos genera un identificador de manera automática, y se envía al servidor de seudonimización.

➢ 1.5: Crear (ow_id, [new_root_pseud]):

Posteriormente el servidor se encarga de crear un nuevo seudónimo correspondiente al registro almacenado y se asignará al propietario de datos mediante su identificador.

➢ 1.6: Cifrar (ISKow, [new_root_pseud,kw_id,ow_id]):

Para proteger esta creación se encriptan los identificadores del nuevo seudónimo, de la palabra clave y del usuario, mediante la clave simétrica interna del propietario de datos, luego se envían a la base de datos.

❖ Almacenamiento de la nueva relación (usuario, palabra clave y el registro) en la base de datos, y el mapping de texto claro del registro y su correspondiente seudónimo:

➢ {1.6.1: store_access_relation (ISKow,[new_root_pseud,kw_id,ow_id])} y {1.6.2:return():ok}:
A nivel de la base de datos se almacena y se crea una relación con los elementos encriptados mediante el método anterior (1.6). Se envía un mensaje de respuesta al servidor.

➢ 1.7: enviar (reg_id,root_seud_id]):
El servidor manda una petición para almacenar un mapping del texto claro del nuevo registro.

➢ 1.7.1: store_clearText_maping (root_pseud_id,reg_id,root_pseud,reg):
Finalmente, se almacena un mapping de texto claro con los identificadores del nuevo registro y su correspondiente seudónimo raíz y sus ambos textos claros.

➢ {1.7.2: return (): ok} y {1.7.3: return(): ok}:
Son métodos de mensajes de respuestas para informar el estado de operación de almacenamiento del nuevo registro al servidor y el propietario de datos respectivamente.

FASE IV: CONCLUSIONES Y VÍAS FUTURAS:

1. Conclusiones:

- RGPD requiere que se tomen todas las medidas técnicas y organizativas apropiadas para proteger los datos personales, y la seudonimización puede ser un método apropiado de elección si se desea mantener la utilidad de datos.

- A partir de todo lo que se ha tratado en este trabajo, se puede decir que las técnicas de seudonimización representan actualmente una de las mejores soluciones de seguridad, que se puede aplicar frente cualquier tratamiento de datos, ya sea autorizado o no autorizado.

- En comparación con la anonimización, la seudonimización es una opción mucho más sofisticada, ya que te deja la clave para "desbloquear" los datos. De esta manera, los datos no se consideran directamente identificativos, y tampoco se anonimiza, por lo que no pierden su valor original.

- Hace que los datos sean identificables si es necesario, pero inaccesibles para los usuarios no autorizados y permite a los procesadores y controladores de datos reducir el riesgo de una posible violación de datos y salvaguardar los datos personales.

- La estructura de capas de seguridad para los sistemas dedicados a la gestión de los datos sensibles y con carácter personal, facilita la gestión de vigilar la información aplicando el protocolo y nivel de seguridad adecuado a cada partición y capa en el sistema.

- La criptografía de datos, como técnica de seudonimización, da una ventaja a nivel de seguridad a los sistemas de información, cuando se aplica de manera correcta a cara a la tipología criptográfica o el algoritmo utilizado sobre el elemento (fichero, ruta, campo…) que se quiere proteger.

2. Vías Futuras:

- A nivel general, uno de los puntos claves, sería la presentación de estas técnicas y que se desarrollen más proyectos que estudian y trabajan en este tema, para lanzar la tecnología y tener más cobertura para adquirir una protección adecuada a los sistemas de información, como se ha recomendado en RGPD.

- Prototipo CTFG. Aplicar mejoras de seguridad sobre un sistema de información que gestiona datos sensibles que tienen carácter personal, utilizando Microsoft Azure Como plataforma de trabajo.

- En el futuro próximo, elaborar proyectos de investigación/Máster que pueden dar un salto y un avance en la utilización de las técnicas de seudonimización a nivel empresarial.

BIBLIOGRAFÍA

A Comprehensive Evaluation of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish. **Priyadarshini Patil, Prashant Narayankar, Narayan D.G., Meena S.M. 2016.** 2016, Procedia Computer Science - Elsevier B.V., Vol. 78, págs. 617 – 624.

A Goal for Privacy, Confidentiality, and Security of Health Information. **Suzy A. Buckovich. 1999.** 1999, ournal of the American Medical Informatics Association, págs. 122–133.

AEPD. 2021. Agencia Española de Protección de Datos. [En línea] 2021. [Citado el: 30 de 05 de 2021.] www.aepd.es/es.

Automated design of a lightweight block cipher with Genetic Programming. **Polimon J, Hernández-Castro JC, Estévez-Tapiador JM, Ribagorda A. 2008.** 1, 2008, International Journal of Knowledge-Based and Intelligent Engineering Systems , Vol. 12, págs. 3-14.

Blackmer, W. Scott. 2016. nuevo Reglamento General de Protección de Datos de la UE. *infolawgroup.com.* [En línea] 5 de Mayo de 2016. [Citado el: 2021 de 03 de 8.] http://www.infolawgroup.com.

Blackmer, W.S. 2016. "GDPR: Getting Ready for the New EU General Data Protection Regulation". [En línea] 05 de 2016. [Citado el: 05 de 03 de 2021.] https://www.infolawgroup.com/insights/2016/05/articles/gdpr/gdpr-getting-ready-for-the-new-eu-general-data-protection-regulation.

Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses. **Commission, European. 2012.** 2012, press-release Database.

contributors, EcuRed. 2018. Curva elíptica. [En línea] 11 de 12 de 2018. [Citado el: 15 de 03 de 2021.] https://www.ecured.cu/index.php?title=Curva_el%C3%ADptica&oldid=3268735.

Data privacy approaches from US and EU perspectives. **Steinke, Gerhard. 2002.** 2002, Elsevier.

Datos personales y protección de la privacidad en la era de las nuevas tecnologías. **Girard-Oppici, Carole. 2015.** 2015, Net-iris, págs. 2-4.

F.Saleh, Dr.Malik. 2011. researchgate. [En línea] Enero de 2011. www.researchgate.net.

Freitas, Vidalina De. 2010. Propuesta de metodología de gestión de seguridad de las TIC's para el sector universitario venezolano. [En línea] 2010. http://www.revistaespacios.com/a10v31n01/10310152.html.

Guerra, Yulaine Arias. 2012. [En línea] Universidad de las Ciencias Informáticas, 11 de 2012. https://www.monografias.com/trabajos94/documentacion-biblioteca-estructuras-datos-avanzadas/documentacion-biblioteca-estructuras-datos-avanzadas2.shtml.

Heurix, Johannes, Karlinger, Michael y Neubauer, Thomas. 2012. Pseudonymization with Metadata Encryption for Privacy-Preserving Searchable Documents. *ieeexplore.ieee.org.* [En línea] 4-7 de Jan de 2012. ieeexplore.ieee.org/document/6149189.

Hong K.S, Chi Y.P. , Chao L.R. , Tang J.H. 2005. Una teoría del sistema integrado de gestión de la seguridad de la información. [En línea] 12 de 2005. [Citado el: 25 de 03 de 2021.] https://doi.org/10.1108/09685220310500153.

How to Share a Secret. **A.Shamir. 1979.** 1979, Communications of the ACM, 22(11), págs. 612-613.

Introduction to the EU Digital Single Market (DSM). **TaylorWessing. 2015.** 2015, united-kingdom.taylorwessing.com.

J. Heurix, M. Karlinger and T. Neubauer. 2012. Pseudonymization with Metadata Encryption for Privacy-Preserving Searchable Documents. *Hawaii International Conference on System Sciences, Maui, HI, 2012, pp. 3011-3020.* [En línea] 02 de February de 2012. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6149189&isnumber=6148595.

John Carl, Villanueva. 2017. jscape.com. *jscape.com.* [En línea] 2017. http://www.jscape.com/blog/stream-cipher-vs-block-cipher.

Jonathan Care, R.K. 2015. Market Guide for Merchant/Acquirer Tokenization of Payment. [En línea] 2015. http://www.gartner.com/document/3177018?ref=solrAll&refval=173965545&qid=1adf8c9d58697bdfcaeba59872b301df..

Los 7 métodos de Autentificación más utilizados. **Evidian. 2015.** 2015, www.evidian.com, pág. 6.

M. SCHULZ, J.A. HENNIS-PLASSCHAERT. 2016. *REGLAMENTO (UE) 2016/679.* Bruselas : EL PARLAMENTO EUROPEO Y DEL CONSEJO EU , 2016.

New Factor of Authentication: Something You Process. **Shakir Ullah Shah, Fazl-e-Hadi, Abid Ali Minhas. 2009.** 2009, IEEE Xplore.

—. **Shakir Ullah Shah, Fazl-e-Hadi, Abid Ali Minhas. 2009.** Islamabad, Pakistan : s.n., 2009. 2009 International Conference on Future Computer and Communication. pág. 103.

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). **CEU, Council of the European Union. June 2015.** June 2015, http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf, págs. 2-3.

Protección de datos personales. **DGJC-CE. 2016.** 2016, http://ec.europa.eu.

redoxtechnologies. 2020. [En línea] 2020. [Citado el: 12 de 03 de 2021.] http://www.redoxtechnologies.com/.

Rhodes, Angus. 2015. What is an RMIS? [En línea] Marzo de 2015. [Citado el: 03 de 03 de 2021.] http://blog.ventivtech.com.

Shamir, Adi. 1979. cs.jhu.edu. *cs.jhu.edu.* [En línea] 11 de 1979. [Citado el: 26 de 02 de 2021.] https://cs.jhu.edu/~sdoshi/crypto/papers/shamirturing.pdf.

Wikimedia Foundation, Inc. 2017. wikipedia.org. *wikipedia.org.* [En línea] 21 de noviembre de 2017. https://en.wikipedia.org/wiki/Feistel_cipher#/media/File:Feistel_cipher_diagram_en.svg.

Wikipedia, colaboradores de. 2021. Cifrador de flujo. [En línea] 03 de 2021. [Citado el: 18 de 04 de 2021.] https://es.wikipedia.org/w/index.php?title=Cifrador_de_flujo&oldid=136161667.

UNIVERSIDAD DE ALMERÍA

Resumen/Abstract

The goal of this Project relays on the analysis and development of pseudonymization techniques based in cryptography, to protect sensible data. During the project, many topics are covered like data protection analysis takin into account European law, also from technical approach to cover the nowadays challenges of information security. The development of the thesis has been with the main focus on the enterprise world, to better apply and understanding the main concepts related to information security, like the mentioned “pseudonymization” to give a better data protection and to accomplish the EU recommendations.

GRADO EN INGENIERÍA INFORMÁTICA, 2020/2021