# Custodia de larga duración de documentos electrónicos

**ÁLVARO DE LA ESCALERA ARMIÑO**
Fábrica Nacional de Moneda y Timbre

**RESUMEN**
Determinados documentos electrónicos (en concreto los que incorporan firmas electrónicas), incorporan elementos que caducan con e transcurso del tiempo y puede que hagan imposible la verificación de los mismos. Se exponen los mecanismos criptográficos y de procedimiento necesarios para garantizar que esto no ocurra.

**PALABRAS CLAVE**
Custodia documental, hash, cifrado, firma electrónica, certificado de clave publica, infraestructura de clave publica, fechado digital.

**ABSTRAC**
Certain electronic records (in particular those that they incorporate electronic signature), incorporate elements that expire with the course of the time and can that makes impossible the verification of such. The cryptographic mechanisms and of necessary procedure to guarantee that this does not happen are exposed.

**KEYWORDS**
Documentary safekeeping, hash, coding, electronic signature, certificate of key publish, key infrastructure publishes, dated.

## LOS DOCUMENTOS ELECTRÓNICOS Y EL CONCEPTO DE "ORIGINAL"

Cuando nos referimos a documentos impresos o manuscritos, estamos acostumbrados a distinguir entre el documento "original" y las posibles "copias" que se pueda hacer de dicho documento. En el caso de documentos electrónicos, si estos mediante algún procedimiento de impresión son trasladados a papel, se puede seguir hablando de documentos originales siempre y cuando se agreguen determinados elementos (como por ejemplo, sellos o firmas). Pero lamentablemente de un documento que permanece en formato electrónico, siempre que no se le dote de ciertos añadidos, no se puede asegurar su procedencia, ni la fecha de su creación, y mucho menos que no haya sido alterado en alguna forma. Por lo tanto de la misma manera que a un documento impreso se le añaden determinados elementos, veremos que para que pueda existir un "original electrónico" (de alguna manera), es necesario añadir nuevos elementos a nuestros documentos.

## CONCEPTOS CRIPTOGRÁFICOS

Antes de comenzar a hablar de esos elementos que se añaden a los documentos electrónicos, es necesario hablar brevemente de dos herramientas criptográficas que son empleadas tanto en el llamado "sellado electrónico" como en la firma electrónica. Estas herramientas son el *hash* o resumen y el *cifrado*.

## LAS FUNCIONES HASH

El *hash* o resumen criptográfico consiste en aplicar una determinada transformación a una información almacenada como números de manera tal que dicha transformación cumpla con tres requisitos fundamentales.

1. Una función hash debe de proporcionar una transformación "de una sola vía", esto quiere decir que teniendo el resultado del hash, sea inviable el poder obtener de nuevo la información original.
2. Una función hash debe de ser resistente a colisiones. Dados dos documentos, la probabilidad de

que produzcan el mismo hash debería ser despreciable.

3. El tamaño del hash es habitualmente mucho menor que el de la información de partida.

Aunque sean dos los requisitos fundamentales, se podría añadir un tercero y es que pequeñas modificaciones en el documento fuente produzcan hashes muy diferentes.

Veamos un mal ejemplo de función hash (mal ejemplo porque no cumple la condición 2 anterior) pero que sirve para ilustrar el funcionamiento de una función hash.

Tomemos el texto "Esto es un texto". Si este texto está codificado en ASCII vemos que se codifica así "69 115 116 111 32 101 115 32 117 110 32 116 101 120 116 111". Apliquemos una sencilla operación a esta codificación (sumemos todos los números de la misma), y vemos que el resultado de este mal hash es 1514.

Si un hash debe "identificar" un documento, el ejemplo de la suma no sirve como hash, dado que sin ir mas lejos el texto "Páguese 1000000000000001" se codifica en ASCII como "80 97 103 117 101 115 101 32 32 49 48 48 48 48 48 48 48 48 48 48 48 48 48 48 49" y casualmente la suma de esos números es 1514, el mismo resultado que el anterior. Como veremos mas adelante, la firma electrónica de documentos se basa entre otras cosas en el cifrado de un hash.... realmente este hash es muy pero que muy malo.

Sin entrar en mas profundidades sobre los hashes, simplemente decir que si que existen funciones que cumplen perfectamente los requisitos anteriores, como pueden ser SHA-1[1], SHA2[2], RIPEMD[3] ....

Valga como ejemplo el resultado de aplicar la función SHA1 a los ejemplos anteriores:

| Texto | SHA1 |
|---|---|
| Esto es un texto | 9a7b00dd5bc100c4e7b600a9c28800aab6ae00 |
| Esto es un Texto | a21700e4cb840029099a3dae9bc300063d8726 |
| Paguese 1000000000000001 | 5d0800c4488e00fa0c19004a09580e72d5980d |

Véase como el simple hecho de cambiar una mayúscula por una minúscula consigue que el hash sea completamente diferente, por no decir del texto que en el ejemplo anterior producía el mismo hash.

Para el lector que desee comprobar todo esto por si mismo, en http://www.paulschou.com/tools/xlate/ está disponible una herramienta en línea que permite la codificación y decodificación de textos, así como la obtención de hashes de los mismos.

## EL CIFRADO

Cifrar en el sentido criptográfico, consiste en realizar una operación sobre una información que llamaremos A, obteniendo una información B, de tal manera que no se pueda conocer A menos que se conozcan bien ciertas características del proceso, bien la "clave" de cifrado.

Como vemos, entre otras cosas, el cifrado se diferencia del hash es que este último es *reversible*, bajo determinadas circunstancias se puede obtener la información de partida a partir de la información cifrada.

Desde tiempos antiguos se utiliza el cifrado, en general para tratar de asuntos de estado o de guerra, como se dice que hacía César[4]. El cifrado utilizado por César era el denominado "cifrado por desplazamiento", según este cifrado, se fija un desplazamiento y se sustituye la letra del mensaje por la correspondiente del alfabeto desplazado. Si tenemos un desplazamiento de 3:

Entonces la frase "Alea jacta est" se cifraría como "Dñhd mdfwd hvw", se puede practicar de manera sencilla con este tipo de cifrado en

http://www.sccs.swarthmore.edu/users/03/julieg/h

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | Ñ | O | P | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | E | F | G | H | I | J | K | L | M | Ñ | N | O | P | R | S | T | U | V | W | X | Y | Z | A | B | C |

[1] http://en.wikipedia.org/wiki/SHA-1

[2] http://en.wikipedia.org/wiki/SHA-1

[3] http://en.wikipedia.org/wiki/RIPEMD

[4] *Si tenía que decir algo confidencial, lo escribía usando el cifrado, esto es, cambiando el orden de las letras del alfabeto, para que ni una palabra pudiera entenderse. Si alguien quiere decodificarlo, y entender su significado, debe sustituir la cuarta letra del alfabeto, es decir, la D por la A, y así con las demás.* — Suetonio, Vida de Julio César

w14cipher.html (lamentablemente, perdemos nuestra querida ñ).

Está claro que incluso para esa época, el cifrado de sustitución no era seguro y mejoró, apareciendo los cifrados por sustitución, de los cuales el cifrado por desplazamiento no es mas que un subconjunto. En los cifrados por sustitución se hace que el alfabeto por el que se sustituye cada palabra no se obtenga mediante un desplazamiento del alfabeto normal, sino que se procura que sea algo más aleatorio. Se podría decir que es entonces cuando nace lo que se conoce como "clave". Como curiosidad decir que loa cifrados por sustitución han continuado siendo usados hasta nuestros días (las novelas de espías de la segunda guerra mundial están llenas de ejemplos de mensajes cifrados en los que la "clave" era una determinada edición de un determinado libro, y el mensaje se descifraba sustituyendo los números del mensaje por las palabras del libro referenciadas por los números). Como curiosidad matemática, decir que en la actualidad el único algoritmo de cifrado considerado (y demostrado) matemáticamente seguro es un algoritmo de sustitución en el que el alfabeto que se utiliza es indefinidamente grande y construido aleatoriamente con un generador de números aleatorios adecuado.

En la actualidad, los sistemas de cifrado se pueden dividir principalmente en dos grupos, cifrado de clave simétrica y cifrado de clave asimétrica. Como su nombre indica, en el primer grupo la clave utilizada tanto para el cifrado como para el descifrado de la información es la misma, y por el contrario, en la criptografía de clave asimétrica, se utilizan distintas claves para el cifrado y el descifrado.

Aunque los algoritmos de clave simétrica son muy importantes, baste decir que se emplean generalmente cuando se requiere tratar con un elevado volumen de datos (son en general mas eficientes que sus primos de clave asimétrica) y cuando se tiene resuelto (o se resuelve por otros medios) el problema de cómo se intercambia entre los participantes esta clave simétrica.

Precisamente para solucionar el problema de cómo conseguir cifrar datos obviando el problema de la distribución de claves (dado que cualquiera que intercepte un mensaje y su correspondiente clave de un sistema de cifrado simétrico automáticamente es capaz de descifrarlo), es como surgen los sistemas de clave asimétrica.

En los sistemas de clave asimétrica se generan al mismo tiempo dos claves, una denominada clave pública que se distribuye libremente y una denominada clave privada que permanece debidamente custodiada. Estas claves, tienen la particularidad de que, cuando se realiza una operación de cifrado con una de ellas, el descifrado se realiza con la otra y viceversa. De esta manera, además del cifrado de por si suficiente beneficio, se obtiene lo siguiente:

Se resuelve el problema de la distribución de las claves, dado que la clave pública se puede distribuir sin problemas.

Se puede tener la certeza del origen de la información. Cualquier información que se pueda descifrar con una determinada clave pública *solamente ha podido ser cifrada con su correspondiente clave privada*. Esto como veremos es uno de los pilares de la firma electrónica.

Veamos como puede funcionar la firma electrónica, y como anteriormente con el hash, para hacerlo comprensible.... emplearemos un ejemplo de un cifrado no demasiado bueno.

Supongamos que tenemos la operación matemática multiplicación. Como sabemos, para cualquier número $z$ (excepto para el 0) se puede encontrar otro número $i$ tal que:

$z \times i = 1$

O lo que es lo mismo, $i$ es el *inverso* de $z$

En un mal sistema criptográfico, a alguien se le podría ocurrir seleccionar un número al azar, calcular su inverso para la operación multiplicación, y llamar al número seleccionado *clave privada* ($K_1$), y al inverso *clave pública* ($K_2$).

De esta manera supongamos que queremos que alguien nos envíe un mensaje cifrado, para ello, primero generamos nuestras claves privada y pública (por ejemplo, 5 y 1/5), y enviamos nuestra clave pública (1/3) a quien (o quienes) queremos que nos envíen mensajes cifrados. Una vez nuestro interlocutores tienen nuestra clave pública, nos quieren mandar cifrado el mensaje "25". Para ello hacen la operación:

($K_2$) x mensaje = cifrado

O sea:

25 x 1/5 = 5

Y transmiten el número 5 que es el mensaje cifrado.

Para descifrarlo, nosotros hacemos la operación

($K_1$) x cifrado = mensaje

5 x 5 = 25

Y ya tenemos el mensaje descifrado.

Es obvio que el sistema de cifrado anterior es rematadamente malo, dado que es inmediato obtener a partir de la clave privada la clave pública, pero imaginemos que en determinadas condiciones no sea una tarea trivial el

obtener ese inverso, esto es lo que se hace en RSA[5]. En RSA se generan aleatoriamente dos números primos muy grandes $p$ y $q$, se multiplican estos dos números y se obtiene un número $m$. A la pareja $(p,m)$ se le llama clave privada, y a la pareja $(q,m)$ clave pública. Pues bien, como para números tan grandes la factorización de $m$(o lo que en la escuela se llama descomposición en factores primos) no es una tarea trivial, no es posible obtener $p$ a partir de $q$ y viceversa, y basándose en obtener potencias de números *módulo m* ( y en los inversos de esa operación) es posible el cifrar y descifrar los mensaje.

## LA FIRMA ELECTRÓNICA

Dado que como hemos visto anteriormente, si se realiza una operación de cifrado asimétrico sobre un documento, se puede asumir que para descifrar un mensaje cifrado con una clave pública es imprescindible la clave privada o viceversa, como consecuencia de esto se puede imaginar que si alguien cifra algo con su clave privada, todo el mundo podría (conociendo la clave pública) comprobar que ese documento efectivamente ha sido cifrado con dicha clave, y asumir que quien lo cifró está en posesión de la misma. Nos encontramos entonces ante lo que podría ser perfectamente un mecanismo de firma electrónica, pero lamentablemente nos faltan aún dos cosas para conseguir que dicha firma sea operativa, en primer lugar, limitaciones de los algoritmos de cifrado asimétricos (son lentos, y p.e. en el caso de RSA no se pueden cifrar documentos si el resultado de la operación es mayor que el módulo $m$), y en segundo, todavía queda por resolver de que manera puede alguien confiar en que el poseedor de una clave privada es quien dice ser.

En el primer caso, esto se resuelve aplicando al documento antes del cifrado asimétrico una función hash. La función hash "reduce" el tamaño del documento (por eso también se les llama funciones resumen), y está ligada estrechamente como hemos visto antes con el documento original. Así, los algoritmos de firma electrónica consisten en realizar un hash a un documento y posteriormente realizar una operación de cifrado asimétrico con una clave privada. El algoritmo de firma en el que se utiliza como hash SHA1 y como cifrado RSA se conoce como sha1WithRSASignature. El proceso de verificación de la firma es sencillo, se descifra la firma con la clave pública (con lo cual se obtiene el hash), se obtiene el hash del documento, y si coinciden el hash descifrado con el hash obtenido, la firma es correcta, en cualquier otro caso bien el documento ha sido manipulado, bien el emisor del documento no está en posesión de la correspondiente clave privada.

Respecto al segundo caso, existen dos soluciones, una basada en la denominada "red de confianza" y otra basada en las "terceras partes de confianza".

## LA RED DE CONFIANZA

Imaginemos dos personas, A y B, se intercambian sus claves de una manera segura (ya porque se las pasan en soporte magnético o por cualquier otro medio). Si la persona A confía en la B (y viceversa), si alguna otra persona tiene contacto con la A pero no con la B podría pedir a la A por ejemplo que le firmase su clave pública, añadiendo información sobre su identidad, y que enviase este documento firmado a la persona B. E esta manera a B le ha llegado ya de una manera confiable la clave pública de C ligada a su identidad. A este documento electrónico en el que se firma una clave pública ligada a una identidad se le conoce como *certificado electrónico, certificado digital* o *certificado de clave pública* (también por certificados X.509 en referencia a la norma de ITU-T que especifica el formato de los certificados de clave pública mas utilizados, pero este nombre es menos genérico). Lamentablemente, este modelo tiene algunas lagunas a efectos legales (por el propio concepto de red, la responsabilidad está muy distribuida) y aunque documentos firmados de esta manera puedan ser presentados ante un tribunal como prueba en caso de litigio, la legislación española no lo reconoce como *firma electrónica reconocida*[6], y no le da el mismo valor que a la firma manuscrita.

## TERCERAS PARTES DE CONFIANZA

Si existiese una organización "de confianza", se podría modificar el modelo de la red de confianza de tal manera que toda la distribución de las claves públicas y emisión de los certificados digitales recayese en esa organización. De esta manera, aunque dos usuarios de esta infraestructura no tuviesen relación alguna entre ellos, como confían en esta "tercera parte de confianza", pueden confiar en que la clave pública de un sujeto realmente pertenece a él, y que este sujeto esta en posesión de su correspondiente clave privada dado que todo esto viene firmado por el organismo emisor de estos "certificados digitales".

A estas terceras partes de confianza la legislación espa-

[5] http://en.wikipedia.org/wiki/RSA
[6] ley 59/2003 del 19 de Diciembre

ñola las denomina "prestadores de servicios de certificación", y en la literatura son conocidas como "terceras partes de confianza" o "autoridades de certificación". A toda la red en general se le conoce como "infraestructura de clave pública" (o también pos sus siglas en inglés, PKI de *Public Key Infraestructure)*. A la firma electrónica que se realiza utilizando *certificados reconocidos*, emitidos por un prestador de servicios de certificación *reconocido* por la legislación española, la legislación le da la misma validez que a la firma manuscrita.

## EL PROBLEMA DEL TIEMPO Y EL SELLADO DIGITAL

Por mucho que el sistema operativo diga que un determinado archivo (o documento electrónico) ha sido creado o modificado en una fecha determinada, esta fecha no es fiable dado que bastaría cambiar la fecha (o la hora) del ordenador donde se edita ese documento para que apareciese la fecha y la hora que mas pudieran interesar.

Para determinados documentos electrónicos (como por ejemplo un contrato) en el que intervienen dos partes, si en el documento queda constancia de la fecha y la hora, y el documento en su totalidad es firmado por ambas partes, el problema de la fecha de validez del documento quedaría solucionado pero ¿cómo se resuelve este problema en el mundo electrónico en otros casos?. Por ejemplo, cuando se necesita certificar que un documento existía en un momento determinado en el tiempo pero a priori no se conoce a quien se va a presentar ese documento y esa fecha como prueba.

De la misma manera que para vincular identidades con claves surge la figura de la *Autoridad de certificación*, para fechar documentos electrónicos surge otra tercera parte de confianza denominada *Autoridad de fechado digital* (también conocida por sus siglas en inglés de TSA o *Time Stamping Authority*).

Una autoridad de fechado digital como se ha dicho antes, es una tercera parte de confianza que en lugar de emitir *certificados de clave pública*, emite *certificados de tiempo* (o *Time Stamp Token* en inglés). Veamos que mecanismos se utilizan para conseguir el fechado digital de un documento.

En primer lugar, dado que este servicio se utiliza a través de una red de comunicaciones, no conviene que se aplique un fechado digital a documentos arbitrariamente grandes, en segundo lugar, tampoco interesa que ni la tercera parte de confianza (ni cualquier otra parte de menos confianza) pueda conocer el documento, por lo tanto como hemos visto anteriormente, acude al rescate la función hash, realizamos un hash del documento.

Una vez realizado el hash, es necesario que el servidor de fechado digital reconozca que lo que se envía es una solicitud de fechado digital, para ello, el cliente de fechado construye una petición con un formado determinado por la RFC 3161[7] formando el denominado *Time Stamp Request*.

El solicitante del fechado envía su solicitud a la autoridad de fechado, si se cumplen unas determinadas condiciones (básicamente que la petición de fechado esté correctamente construida), la autoridad de fechado añade la fecha y hora obtenida de una fuente de confianza (según la legislación española, la hora oficial española es la proporcionada por el Real Observatorio de la Armada de San Fernando, en Cádiz), toma el hash del documento, lo firma electrónicamente, y lo devuelve al solicitante en un *Time Stamp Token* cuyo formato se especifica también en la RFC 3167.

De esta manera, mediante el fechad digital, aunque no se asegure estrictamente la fecha de creación de un documento, si que se asegura la existencia de dicho documento a partir de un instante determinado de tiempo, y, como ocurre con la firma electrónica, que dicho documento no ha sido manipulado.

## LA CUSTODIA DE DOCUMENTOS

En determinadas circunstancias, podría ser conveniente que documentos electrónicos sean almacenados de una manera segura y confidencial durante largos (o no tan largos) periodos de tiempo. Analicemos los servicios que debería proporcionar un sistema de custodia de documentos.

Almacenamiento seguro de documentos.
Recuperación de documentos.
Cifrado de documentos (opcional).
Fechado de documentos (opcional).
Almacenamiento de datos de verificación de documentos (opcional).
Renovación de documentos (opcional).

## ALMACENAMIENTO SEGURO DE DOCUMENTOS

El servicio de custodia deberá ser capaz de almacenar los documentos de manera segura y durante largos

[7] "Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP) " http://www.ietf.org/rfc/rfc3161.txt

periodos de tiempo. Para ello, deberá contar con los adecuados medios para efectuar copias de respaldo programadas y periódicas. Asimismo, y dependiendo de la importancia de los documentos, podría llegar a ser conveniente el contar con un centro de respaldo en una ubicación geográfica distinta.

## RECUPERACIÓN DE DOCUMENTOS

De poco o nada sirve almacenar documentos electrónicos si estos no pueden ser recuperados cuando sea necesario de una manera efectiva.

## CIFRADO

Para documentos confidenciales, se precisa que estos se almacenen cifrados. Para ello, el servicio de custodia debería proporcionar mecanismos para asegurar que los documentos puedan ser almacenados cifrados de manera que:

El que envió el documento a custodia sea capaz de descifrar el documento.

Cualquier otro no autorizado no sea capaz de descifrar el documento.

En caso de extravió de la clave de cifrado, el sistema sea capaz, bajo las debidas medidas de seguridad, de descifrar el documento.

## FECHADO

El sistema deberá ser capaz de fechar electrónicamente los documentos, de esta manera se da a los documentos una fecha de entrada al sistema fiable y, se tienen constancia de la existencia de un determinado documento a partir de un determinado instante de tiempo.

El sistema será capaz de almacenar este fechado digital de manera conjunta al documento.

## ALMACENAMIENTO DE DATOS DE VERIFICACIÓN

Si los documentos a almacenar van firmados electrónicamente, y estas firmas han de ser verificadas en un futuro, es conveniente que en el instante del almacenamiento se verifique la validez de la firma y de todos los componentes asociados a la misma. Asimismo, se deberá almacenar todos estos datos de verificación conjuntamente con el documento, y proceder al fechado conjunto de los mismos.

Veamos en que consisten estos datos de verificación. Anteriormente hemos visto que un componente fundamental de la firma electrónica es el denominado *certificado de clave pública*, el cual liga una identidad a un material criptográfico. Pues bien, si por cualquier circunstancia ese certificado se invalida antes de su periodo de expiración (por extravío de las claves, por que el usuario piensa que alguien puede conocerlas, etc...) se hace necesario que una infraestructura de clave pública proporcione mecanismos para la *revocación* de ese certificado. Esos mecanismos se implementan de dos maneras, bien mediante la publicación de las llamadas *listas de revocación* (en inglés CRL), bien proporcionando un servicio en línea de validación del estado de los certificados (OCSP, *Online Certificate Status Protocol).*

Pues bien, el sistema de custodia debería ser capaz de obtener esos datos, incorporarlos al documento, y fecharlos de manera conjunta para poder demostrar a posteriori que en el momento del archivo esas firmas electrónicas eran perfectamente válidas.

## RENOVACIÓN DE DOCUMENTOS

Es posible que, bien por avances en la criptografía, bien por avances en la capacidad de proceso de los ordenadores, los mecanismos utilizados anteriormente tanto para el fechado digital como para la firma electrónica se vean comprometidos. Para ello, se dota a los certificados utilizados de un determinado periodo de validez no excesivamente grande (del orden de tres años) en los que se asume que no se van a ver comprometidos los mecanismos criptográficos utilizados. Pero nos podríamos preguntar ¿qué ocurre si es necesario verificar la validez de firmas electrónicas con mucha posterioridad a su realización?. Claramente, y en el caso general, es muy probable que los mismos actores que intervinieron en la firma del documento no estén disponibles para volver a firmarlo (o que se nieguen). Por lo tanto, es necesario que de una manera sencilla se pueda prolongar la validez de estros documentos.

La manera sencilla de prolongar esta validez es mediante el fechado digital. Si sabemos que un documento es válido en un instante de tiempo determinado, se puede fechar el documento en su totalidad en ese instante, y, con posterioridad, se puede proceder a su fechado periódicamente (incluyendo los posibles fechados anteriores). De esta manera, como los posteriores fechados se supone están realizados utilizando certificados que no están caducados, algoritmos criptográficos robustos, y una longitud de claves adecuada, se puede prolongar la validez en el tiempo de esos documentos de manera indefinida.

**Referencias:**


# David Kahn, Codebreakers — The Story of Secret Writing, 1967. ISBN 0684831309.

# F. L. Bauer, Decrypted Secrets, 2ª edición, 2000, Springer. ISBN 3540668713.