INSTITUTO POLITÉCNICO NACIONAL

Escuela Superior de Cómputo

ESCOM

Trabajo Terminal

"Protocolo criptográfico para el almacenamiento sin duplicados en la nube, resistente a ataques por fuerza bruta."

2016-B045

Presentan

Eder Jonathan Aguirre Cruz

Diana Leslie González Olivier

Jhonatan Saulés Cortés

Directora

Dra. Sandra Díaz Santiago

Mayo 2017

Índice

Índice de Figuras

Índice de Tablas

Capítulo 1

Introducción

1.1. Contexto

En el nuevo ambiente de las tecnologías de la información se encuentran los usuarios de estas tecnologías y las organizaciones, cualquier movimiento de almacenamiento masivo puede ser realizado mediante modelos basados en el cómputo nube, es decir, el almacenamiento en la nube. Asimismo, al manejar un gran volumen de información, los usuarios buscan la posibilidad de que al almacenar esos datos puedan ser accedidos a ellos de manera fácil [16].

El cómputo nube es un término general utilizado para nombrar así a la provisión de servicios de almacenamiento a través de Internet que ha sido utilizado para facilitar el cambio de los modelos de negocios, agilizar procesos y reducir los costos de operación. Uno de los mayores beneficios que ofrece este servicio es la virtualización de los centros de datos, que pueden operar de manera automatizada, sin necesidad de la presencia de una persona física y ser pueden ser gestionados en cualquier momento y tiempo.De acuerdo con un estudio realizado por la consultora Market Research Media, el cloud computing generará $270,000 millones de dólares en 2020, por lo que empresas como Google, Amazon, IBM, Oracle y Apple han adoptado este sistema como parte del servicio brindado a sus consumidores, por ejemplo Google Drive o iCloud, a través de los cuales, con sólo estar conectados a Internet, los usuarios tienen la posibilidad de utilizarlos [3].

Básicamente el almacenamiento en la nube se caracteriza por 5 puntos esenciales que son:

- **Autoservicio on-demand o pago por evento**
- **Acceso ubicuo a la red (uso de los servicios cuando sea y donde sea)**
- **Fondo común de recursos**
- **Rápida elasticidad**
- **Servicio medido** [10].

1.2. Problemática

Hoy en día el manejo de información en la sociedad juega un papel importante en el desarrollo de las actividades que la conforman. Millones de personas en el mundo tienen la

facilidad de acceder a un dispositivo electrónico que les permite manipular esta información o almacenarla para posteriormente darle un uso específico. La información que circula en dispositivos electrónicos es mayor a la memoria disponible que ofrecen estos, a medida que el volumen de información aumenta, también lo hace la demanda para los servicios de almacenamiento en línea [2]. Un gran incremento en el uso de estos servicios implica tener más infraestructura y personal para que los sistemas de almacenamiento tengan más capacidad y puedan cubrir la demanda que se presenta en el mercado. Si bien el almacenamiento logró dar buenos resultados al cliente en sus primeras etapas, ahora la preocupación por el incremento de infraestructura para seguir dando esos resultados se ha incrementado considerablemente [1].
El cambio en las estrategias de negocio y la explosión de datos digitales se ha lanzado enormes demandas de alto volumen y almacenamiento de datos eficiente. Debido a los limitados recursos financieros y altos gastos de almacenamiento de datos electrónicos, los usuarios prefieren almacenar sus datos en los entornos de nube, el almacenamiento en la nube permite a sus usuarios transferir sus datos y aplicaciones en la web para que puedan operar esos programas sin ninguna infraestructura física necesaria. Hay recursos limitados de almacenamiento y de red en el sistema de nube. La totalidad de los servicios en la nube que se han ofrecido hasta ahora, permite a los usuarios detener los problemas mediante el uso de los dos aspectos importantes de la fiabilidad y elasticidad [1].

Una de las principales razones del incremento en el tamaño en la estructura de almacenamiento de servicios en linea es la duplicación de archivos por varios y diferentes usuarios, existen muchas copias en la nube de un mismo archivo que se encuentra presente en diferentes cuentas de usuarios. Por ejemplo $\boldsymbol{n}$cantidad de usuarios pueden subir la misma canción a la nube, por lo tanto esta se encuentra almacenada en las $\boldsymbol{n}$ cantidad de cuentas que tiene registro la nube, esta misma canción que se encuentra almacenada está cubriendo un espacio en la memoria del servicio, si se tuviera una sóla copia almacenada de esta canción se ahorraría mucho espacio en la nube que podría utilizarse para el aalmacenamiento de un archivo diferente . Según un estudio [7] realizado por HP se estima que hay 1 Exabyte de datos almacenados en la nube, además de 2012 a 2017, las cargas de trabajo de los centros de datos crecerán 2.3 veces, mientras que en la nube aumentarán 3.7 veces, lo cual implica que el Exabyte que se estima se podría llegar a triplicar y las empresas que proporcionan estos servicios disminuyen su oferta en el mercado.

Sumado al agraviante problema de crecimiento en la estructura de almacenamiento en línea, existe un elemento que pocas organizaciones y usuarios contemplan para el almacenamiento de su información, y es la seguridad. Las plataformas que ahora ofrecen el servicio de almacenamiento en la nube, no contemplan la protección e integridad de la información, este servicio está sujeto a los ataques de adversarios que están interesados en el robo, manipulación o alteración de la información importante para los usuarios. Dicho servicio carece de algún esquema de seguridad y por tanto no se puede garantizar la permanencia de los usuarios en el consumo del servicio en la plataforma de almacenamiento. Dejar sin resolver la seguridad de la infofrmación, puede generar negativas consecuenicias hacia cualquier usuario de la plataforma de almacencamiento, cualquier individuo que tenga acceso a los archivos almacenados en ésta plataforma, podrá visualizar el contenido de estos, comprometiendo la fidelidad e integridad de la información para poder usarla en situaciones que pudieran llegar a perjudicar al usuario.

1.3. Solución propuesta

La eliminación de duplicación de datos es una experiencia progresiva que puede disminuir drásticamente la cantidad de información de respaldo almacenada eliminando todos los datos redundantes como se ilustra en la figura 1.1. Al evitar la duplicación de datos explota el consumo de almacenamiento mientras que permite a las tecnologías de información recuperar más datos de respaldo de líneas cercanas durante más tiempo. Esto recupera enormemente la capacidad del disco de copia de seguridad establecida, alterando la forma en que los datos están protegidos. En general, la eliminación de duplicados compara la información nueva con la información actual de los trabajos anteriores de copia de seguridad o archivado y elimina las redundancias en la nube reduciendo la asignación de almacenamiento dentro de esta, puede reducir las necesidades de almacenamiento en hasta un *80 %* para archivos y copias de seguridad que los usuarios resguardan en la nube. Las ventajas de no tener duplicados en la nube incluyen una mayor capacidad de almacenamiento y ahorro presupuestario, al igual que la minimización del ancho de banda para menos costosa y más rápida la repetición de la información fuera de la reserva simplificando y mejorando la gestión del almacenamiento de datos [?].

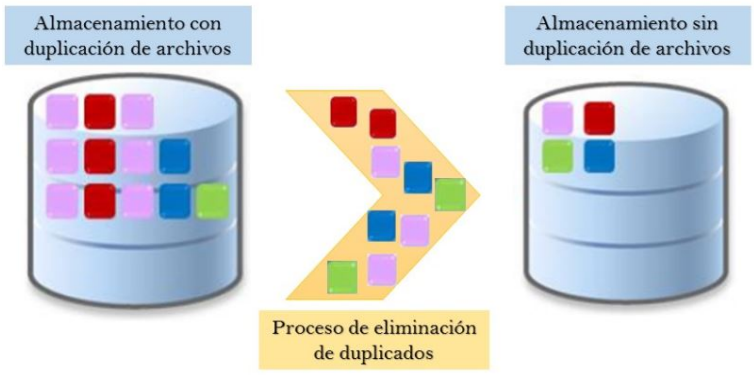

Figura 1.1: Solución Propuesta

El objetivo es almacenar más datos en menos espacio mediante la segmentación de los archivos en pequeños trozos de tamaño variable*(32 a 128 KB)*, la identificación de fragmentos duplicados, manteniendo una sola copia de cada trozo. Las copias repetidas del trozo se sustituyen por una referencia a la única copia. Los trozos se comprimen y luego son organizados en contenedores especiales de archivos en la carpeta Información del volumen del sistema. Para garantizar la privacidad de los datos obtenidos después del proceso de eliminación de duplicación, es posible utilizar algoritmos criptográficos [11].

Una posible solución para la protección a los datos y eliminar duplicaciones de estos, es echar mano de la criptografía. Ciencia que se encarga del estudio de técnicas para transformar la información a una forma que no pueda entenderse a simple vista; sin embargo, el objetivo de la Criptografía no es sólo mantener los datos secretos, sino también protegerlos contra modificación o manipulación y comprobar la fuente de los mismos [8].

Está ciencia que matiene la información segura se encuentra dividida en dos grandes tipos: **Criptográfia simétrica** y la **Criptografía asimétrica**.

- *La criptografía simétrica* o también llamada criptografía de llave privada, basa su seguridad en una sóla llave que se comparte entre dos usuarios que quieren compartir información, dicha llave es utilizada para cifrar un archivo al ser enviado al otro usuario y este utilizará la misma llave para descifrarlo cuando lo reciba.

- *La criptografía asimétrica* o criptografía de llave pública involucra el uso de un par de llaves para cada usuario que desea comunicarse, estas llaves llamadas pública y privada. Para que un usuario envíe un archivo a otro usuario necesita cifrar el archivo con la llave pública de ese usuario al que se desea enviar, y cuando lo reciba ese usuario lo deberá descifrar con su llave privada o secreta. De esta manera se evita el compartir llaves para cifrar y descifrar como sucede en la criptografía simétrica y reduce los riesgos de un ataque de adversarios.

Puesto que ambas cuestiones, la eliminación de duplicados y la privacidad de la información, son importantes, se ha comenzado a proponer mecanismos que solucionen ambos problemas de manera conjunta, que son: Dupless [2], ABS: the apportioned backup system. [5], Flud Backup [17], SIGOPS Oper. Syst. [6], TahoeFS [21].

1.4. Justificación

En la actualidad millones de personas usan los servicios de almacenamiento que ofrece la nube, ya sean gratuitos o privados, este número de personas ha ido en un incremento exponencial lo cual hace que el espacio de almacenamiento disminuya, entonces ¿Cómo podría mitigar el problema de almacenamiento y tener privacidad de los datos al mismo tiempo?

Usando la criptografía clásica para poder cifrar un archivo se utiliza una clave privada la cuál es distinta para cada usuario, cada vez que se cifra un archivo el resultado de este es diferente para cada intento. Por tanto no se puede evitar la duplicación de archivos utilizando este mecanismo de la criptografía y se deben implementar soluciones más robustas.

Una solución para tener privacidad y evitar duplicación la proporcionó John R. Douceur, la cual dice que teniendo a M que será el contenido de un archivo de aquí en adelante denominado el mensaje, el cliente primero calcula una clave $K \leftarrow H(M)$ mediante la aplicación de una función de hash criptográfica H al mensaje y luego calcula el texto cifrado $C \leftarrow E(K,M)$ a través de un esquema de cifrado simétrico determinista. El derivado del mensaje K se almacena por separado cifrándolo con una llave por cliente. Un segundo cliente B cifra el mismo archivo M que producirá el mismo C, evitando la duplicación. [14]

En el artículo publicado por Mihir Bellare, Sriram Keelveedhi, Thomas Ristenpart, nombrado "DupLESS: Server-Aided Encryption for Deduplicated Storage" [2], se observó que uno de los principales problemas al que nos enfrentamos es que el esquema de cifrado solo es seguro cuando el espacio de mensajes es demasiado grande, por lo tanto agentes externos pueden provocar agravios a la integridad de la información de los usuarios.

Si bien esta solución se ocupa de la duplicación de archivos deja muy vulnerable el aspecto de la privacidad, ya que ante un espacio de mensajes pequeño las amenazas del adversario son demasiadas. Si se tuvieran como ejemplo 1000 mensajes, para el adversario sería muy fácil intentar encontrar la clave, probando las 1000 claves posibles generadas con la función hash, hasta descifrar el archivo, por lo tanto se comprueba que un espacio de 1000 mensajes sigue siendo pequeño.

Es por ello que este trabajo terminal tiene como principal meta atacar esta problemática de privacidad, proponiendo una arquitectura del sistema que a través de un servidor de llaves se generaran llaves de acuerdo al contenido del archivo, para con esta se pueda cifrar y luego almacenar en la nube donde se eludirá la duplicación de archivos. Dicha arquitectura se explica con detalle en el siguiente apartado.

1.5. Objetivos

1.5.1. Objetivo General

Desarrollar un protocolo criptográfico para evitar la duplicación de archivos almacenados en la nube, garantizando la privacidad de los usuarios contra adversarios cuando el espacio de mensajes es pequeño, utilizando algoritmos criptográficos para su implementación.

1.5.2. Objetivos Específicos

- Evitar la duplicación de archivos que sean almacenados por los usuarios de la nube
- Proteger ante los adversarios la información de los usuarios de la nube
- Establecer un esquema de autenticación de usuarios
- Reducir la pérdida y filtración de información de los usuarios de la nube

Capítulo 2

Preliminares.

El contenido de este capítulo abordará temas estrechamente relacionados con la criptografía, la seguridad de la información y las implicaciones que ésta podria traer si esta se encuentra corrompida por algun adversario. También, este capítulo contiene información acerca de los 2 tipos de criptografía que existen mencionando los diferentes esquemas de cifrado y los modos de operación que son utilizados por algunos de estos. De igual forma se describe con detalle los servicios que ofrece el cómputo nube, haciendo énfasis en un servicio en particular que es el de almacenamiento que se utilizará para la implementación de este protocolo criptogrráfico.

2.1. Definiciones.

2.2. Definiciones.

- **Criptografía.** La Criptografía es la ciencia que se encarga del estudio de técnicas matemáticas relacionadas con aspectos de seguridad para transformar la información a una forma que no pueda entenderse a simple vista; sin embargo, el objetivo de la Criptografía no es sólo mantener los datos secretos, sino también protegerlos contra modificación y comprobar la fuente de los mismos [9].

- **Criptoanálisis.** Es la ciencia que se ocupa del análisis de un texto cifrado para obtener la información original sin conocimiento de la clave secreta, esto es, de forma ilícita rompiendo así los procedimientos de cifrado establecidos por la Criptografía, por lo que se dice que Criptoanálisis y Criptografía son ciencias complementarias pero contrarias. El criptoanálisis es el arte de descifrar comunicaciones cifradas sin conocer las llaves [22].

2.2.1. Servicios criptográficos.

Los servicios criptográficos son aquellos que garantizan en un sistema de información la adquisición, almacenamiento, procesamiento y transmisión de la información y para lograrlo se valen de uno o más objetivos fundamentales.

- **Confidencialidad.** Es un servicio utilizado para mantener el contenido de la información de todos, excepto los autorizados a tenerla. El secreto es un término sinónimo

de confidencialidad y privacidad. Hay numerosos enfoques para proporcionar confidencialidad, que van desde la protección física a los algoritmos matemáticos que hacen que los datos sean ininteligibles.

- **Autenticación.** Es un servicio relacionado con la identificación. Esta función se aplica tanto a las entidades como a la propia información. Dos partes que participan en una comunicación deben identificarse entre sí. La información entregada a través de un canal debe ser autenticada en cuanto al origen, fecha de origen, contenido de los datos, tiempo enviado, etc. Por estas razones este aspecto de la criptografía suele subdividirse en dos clases principales: autenticación de entidad y autenticación de origen de datos. La autenticación de origen de datos proporciona implícitamente la integridad de los datos (si se modifica un mensaje, la fuente ha cambiado).

- **Integridad.** Es un servicio que se ocupa de la alteración no autorizada de los datos. Para asegurar la integridad de los datos, se debe tener la capacidad de detectar la manipulación de datos por parte de algún adversario. La manipulación de datos incluye cosas tales como inserción, supresión y sustitución.

- **No repudio.** Es un servicio que impide a una entidad negar compromisos o acciones anteriores. Cuando surgen disputas debido a que una entidad niega que se tomaron ciertas acciones, es necesario un medio para resolver la situación. Por ejemplo, una entidad puede autorizar la compra de una propiedad por otra entidad y posteriormente denegar que se concedió dicha autorización. Se necesita un procedimiento que involucre a un tercero de confianza para resolver la disputa [9].

2.3. Ataques a servicios criptográficos.

Un ataque es una violación a la seguridad de la información realizada por intrusos que tienen acceso físico al sistema sin ningún tipo de restricción, su objetivo es robar la información o hacer que ésta pierda valor relativo, o que disminuyan las posibilidades de su supervivencia a largo plazo.

- **Ataque sólo con texto cifrado.** Este caso es cuando el criptoanalista sólo conoce el criptograma y el algoritmo con que fue generado; con esta información pretende obtener el texto en claro.

- **Ataque con texto original conocido.** En esta situación el criptoanalista conoce mensajes en claro seleccionados por él mismo y sus correspondientes criptogramas, así como el algoritmo con que éstos fueron generados; aquí el objetivo es conocer la clave secreta y poder descriptar libremente cualquier texto.

- **Ataque con texto cifrado escogido.** El criptoanalista conoce el algoritmo de cifrado, así como un criptograma seleccionado por él mismo y su correspondiente texto en claro, su objetivo es obtener el mensaje en claro de todo criptograma que intercepte.

- **Ataque con texto escogido.** En este caso el criptoanalista además de conocer el algoritmo de cifrado y el criptograma que quiere descriptar, también conoce el criptograma de un texto en claro que él elija y el mensaje en claro de un criptograma también elegido por él [23].

2.4. Criptografía Simétrica.

Los esquemas criptográficos simétricos también se conocen como esquemas o algoritmos de clave simétrica, clave secreta y de clave única. Consideremos un esquema de cifrado que consiste en los conjuntos de transformaciones de cifrado y descifrado *Ee:* $e \in \mathcal{K}$ y *Dd:* $d \in \mathcal{K}$, respectivamente, donde K es el espacio clave. El esquema de cifrado se dice que es de clave simétrica si para cada par asociado de cifrado/descifrado de claves *(e, d)*, es computacionalmente "fácil" para determinar d conociendo sólo e, y determinar e a partir de d. Desde $e = d$ en los esquemas de cifrado de clave simétrica más prácticos, la clave simétrica término se convierte apropiado [9].
Cuando exisen dos usuarios, que quieren comunicarse para compartir información a través de un canal inseguro que puede ser Internet, teléfonos móviles o comunicación LAN inalámbrica, etc, se presenta un problema, ya que existe algún adversario que tiene acceso a ese canal de comunicación, este tipo de escucha no autorizada se llama espionaje. En esta situación, la criptografía simétrica ofrece una solución: el usuario cifra su mensaje x usando un algoritmo simétrico, dando el texto cifrado y. El usuario destinatario recibe el texto cifrado y descifra el mensaje, si se tiene un algoritmo de cifrado fuerte, el texto cifrado se verá como bits aleatorios al adversario y no contendrá ninguna información que le sea útil [12].

Los algoritmos criptográficos simétricos tienen dos versiones: cifrador en bloque y cifrador de flujo. El beneficio del uso de un algoritmo simétrico radica en el procesamiento rápido para cifrar y descifrar un alto volúmen de datos. El cifrado simétrico es una táctica eficaz de almacenamiento de información sensible en una base de datos, un registro o archivo [18].

Así como la criptografía tiene grandes ventajas para la solución en la comunicación de dos agentes a través de un canal inseguro, también cuenta con ciertas desventajas que son:

- La seguridad depende de un secreto compartido entre el emisor y el receptor.
- La administración de las claves no es escalable.
- La distribución manual de llaves es costosa, ocupa mucho tiempo y es propensa a errores.
- La distribución de claves debe hacerse a través de algún medio seguro como centros de distribución de llaves, implementación de algoritmos, etc [ ].

El esquema de cifrado simétrico se puede representar a través de la siguiente figura 2.1.

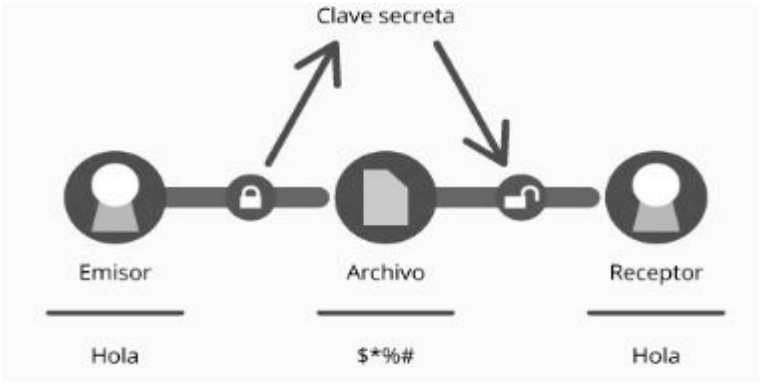

Figura 2.1: Diagrama de Criptografía Simétrica.

La sintaxis de un esquema de cifrado simétrico, esta dada por la siguiente definición.

Definición 2.1 *Un esquema de cifrado simétrico está conformado por una tripleta de algoritmos* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$*, definidos como se describe a continuación:*

- *El algoritmo generador de claves* Gen *selecciona una llave* K *al azar del conjunto de llaves* $\mathcal{K}$*, esto se denotará como* $K \xleftarrow{\$} \mathcal{K}$*. Esta clave* K *será usada por los algoritmos* Enc *y* Dec*, esta clave la compartirán emisor y receptor.*
- *El algoritmo de cifrado* Enc*, toma como entrada un texto en claro* $M \in \mathcal{M}$ *y una clave* K *generada por* Gen *y regresa un texto cifrado* $C \in \mathcal{C}$*. Usualmente esto se denota como* $C \leftarrow \mathsf{Enc}_K(M)$*.*
- *El algoritmo de descifrado* Dec*, toma como entrada un texto cifrado* C *y una llave* K *y regresa* M*. Esta operación se denota por* $M \leftarrow \mathsf{Dec}_K(C)$*. Para que cualquier algoritmo de cifrado simétrico funcione correctamente, se debe garantizar que para todas las llaves posibles en* $\mathcal{K}$ *y todos los posibles mensajes* $\mathcal{M}$*,*

$$\mathsf{Dec}_K(\mathsf{Enc}_K(M)) = M.$$

2.5. Criptografía asimétrica.

La criptografía asimétrica es por definición aquella que utiliza dos claves diferentes para cada usuario, una para cifrar que se le llama clave pública y otra para descifrar que es la clave privada. Los algoritmos asimétricos son diferentes a los simétricos en un sentido muy importante [18]. Cuando se genera una clave simétrica, simplemente se escoge un número aleatorio de la longitud apropiada. Al generar claves asimétricas el proceso es más complejo. [18].

Características de la Criptografía asimétrica:

- Se basa en operaciones matemáticas complejas.
- Se ejecuta de 100 a 1000 veces más lento que los algoritmos simétricos.

[18]

Los beneficios de la criptografía asimétrica son la solución a los problemas de la criptografía simétrica, pues las claves públicas pueden ser distribuidas con toda tranquilidad, no valen de nada sin las claves privadas. El cifrado asimétrico se emplea muy frecuentemente para pasar con seguridad una clave privada, que posteriormente, será la que se utilice para cifrar y/o descifrar otra información. El cifrado asimétrico puede ser representado como aparece en la figura **??**.

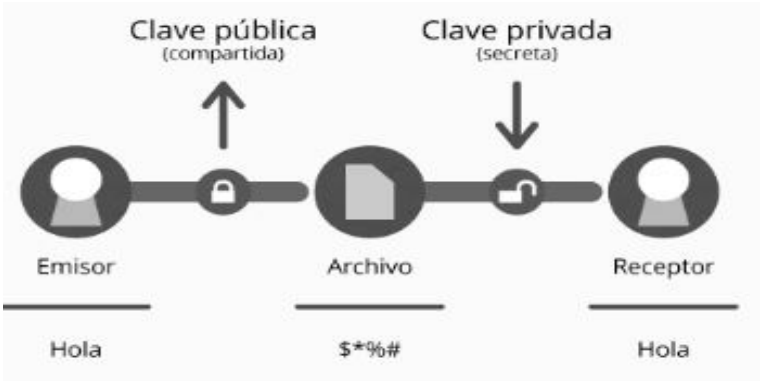

Figura 2.2: Diagrama de Criptografía Asimétrica.

2.6. Cifrado por bloques.

Los algoritmos de cifrado por bloques toman bloques de tamaño fijo del texto en claro y producen un bloque de tamaño fijo de texto cifrado, generalmente del mismo tamaño que la entrada. El tamaño del bloque debe ser lo suficientemente grande como para evitar ataques de texto cifrado. La asignación de bloques de entrada a bloques de salida debe ser uno a uno para hacer el proceso reversible y parecer aleatoria.
Para la asignación de bloques los algoritmos de cifrado simétrico realizan sustituciones y permutaciones en el texto en claro hasta obtener el texto cifrado.
La sustitución es el reemplazo de un valor de entrada por otro de los posibles valores de salida, en general, si usamos un tamaño de bloque k, el bloque de entrada puede ser sustituido por cualquiera de los bloques posibles. La permutación es un tipo especial de sustitución en el que los bits de un bloque de entrada son reordenados para producir el bloque cifrado, de este modo se preservan las estadísticas del bloque de entrada (el número de unos y ceros).

Los algoritmos de cifrado por bloques iterativos funcionan aplicando en sucesivas rondas una transformación a un bloque de texto en claro. La misma función es aplicada a los datos usando una subclave obtenida de la clave secreta proporcionada por el usuario. El número de rondas en un algoritmo de cifrado por bloques iterativo depende del nivel de seguridad deseado.

La sustitución es el reemplazo de un bloque de n bits por otro bloque de n bits en un espacio de 2^k [4]. Los cifradores por bloques mas usados son AES (Advanced Encryption Standard, por sus siglas en inglés) y DES (Data Encryption Standard, por sus siglas en inglés). [19]

Los cifradores por bloques pueden ser representados como se ve en la figura **??**.

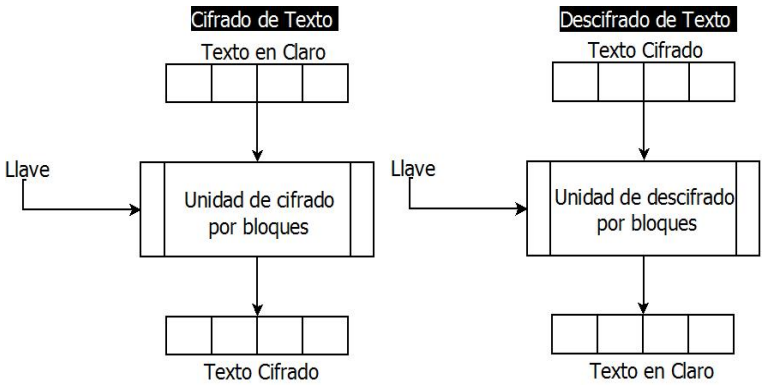

Figura 2.3: Diagrama de Cifradores por Bloques

2.7. RSA

El algoritmo de clave pública RSA fué creado en 1978 por Rivest, Shamir y Adlman, y es el sistema criptográfico asimétrico más conocido y usado. Estos señores se basaron en el artículo de Diffie-Hellman sobre sistemas de llave pública, crearon su algoritmo y fundaron la empresa RSA Data Security Inc., que es actualmente una de las más prestigiosas en el entorno de la protección de datos. El sistema RSA se basa en el hecho matemático de la dificultad de factorizar números muy grandes. Para factorizar un número el sistema más lógico consiste en empezar a dividir sucesivamente éste entre 2, entre 3, entre 4,..., y así sucesivamente, buscando que el resultado de la división sea exacto, es decir, de resto 0, con lo que ya tendremos un divisor del número.

Ahora bien, si el número considerado es un número primo (el que sólo es divisible por 1 y por él mismo), tendremos que para factorizarlo habría que empezar por 1, 2, 3,..... hasta llegar a él mismo, ya que por ser primo ninguno de los números anteriores es divisor suyo. Y si el número primo es lo suficientemente grande, el proceso de factorización es complicado y lleva mucho tiempo.

Basado en la exponenciación modular de exponente y módulo fijos, el sistema RSA crea sus claves de la siguiente forma:

- Se buscan dos números primos lo suficientemente grandes: p y q (de entre 100 y 300 dígitos).
- Se obtienen los números n = p * q y Ø = (p-1) * (q-1).
- Se busca un número e tal que no tenga múltiplos comunes con Ø.
- Se calcula d = e-1 mod Ø, con mod = resto de la división de números enteros. Y ya con estos números obtenidos, n es la clave pública y d es la clave privada. Los números p, q y Ø se destruyen. También se hace público el número e, necesario para alimentar el algoritmo.

El cálculo de estas claves se realiza en secreto en la máquina en la que se va a guardar la clave privada, y una vez generada ésta conviene protegerla mediante un algoritmo criptográfico simétrico.

En cuanto a las longitudes de claves, el sistema RSA permite longitudes variables, siendo aconsejable actualmente el uso de claves de no menos de 1024 bits (se han roto claves de

hasta 512 bits, aunque se necesitaron más de 5 meses y casi 300 computadoras trabajando juntas para hacerlo).

RSA basa su seguridad es ser una función computacionalmente segura, ya que si bien realizar la exponenciación modular es fácil, su operación inversa, la extracción de raices de módulo Ø no es factible a menos que se conozca la factorización de e, clave privada del sistema.

RSA es el más conocido y usado de los sistemas de clave pública, y también el más rápido de ellos. Presenta todas las ventajas de los sistemas asimétricos, incluyendo la firma digital, aunque resulta más útil a la hora de implementar la confidencialidad el uso de sistemas simétricos, por ser más rápidos. Se suele usar también en los sistemas mixtos para encriptar y enviar la clave simétrica que se usará posteriormente en la comunicación cifrada. [15]

2.8. Firmas a ciegas.

Las firmas a ciegas son un tipo especial de firmas digitales en las que se firma algo que no se conoce. Para hacer firmas a ciegas se utilizan factores de opacidad, para ocultar el mensaje original que se requiere que esté firmado, y así la autoridad no pueda conocer lo que está firmando. Por lo tanto, el propósito de una firma a ciegas es evitar que el firmante B conozca el mensaje que firma; y así posteriormente, sea incapaz de asociar el mensaje que firmó con el remitente A. Entonces, las firmas a ciegas tienen aplicación en varias situaciones. A continuación se mencionan dos de ellas:

- Cuando se utiliza dinero electrónico. En este caso, m representa un valor monetario que A (el cliente) tiene derecho a gastar. Y así, cuando m y s(m) se presentan a B (el banco) para efectuar el pago, B es incapaz de identificar al cliente que originalmente le dio ese dinero electr´onico a firmar, pues le fue enviado de manera oculta. Lo anterior permite que la identidad de A permanezca anónima, y sus movimientos financieros no puedan ser monitoreados.

- En las elecciones electrónicas también pueden utilizarse las firmas a ciegas, ya que se requiere que B (una autoridad electoral) no conozca la identidad de A (el votante) debido a que el voto debe efectuarse de manera anónima. Sin embargo, es necesario que A demuestre que su voto m es válido. Lo cual se logra cuando A presenta ante B la firma s(m). Y se sabe de antemano que B no puede asociar s(m) a A, debido a que el votante previamente le envió a B su voto m pero de forma oculta para que se lo firmara. [13]

2.9. Funciones Hash.

A continuación se describirán las características de las *funciones hash*, también conocidas como *funciones de resumen*. Las funciones hash basan su definición en funciones de un solo sentido (*one-way functions*, en inglés). Una función de un solo sentido es aquella que para un valor x, es muy fácil calcular $f(x)$, pero es muy difícil hallar $f^{-1}(x)$. Es complicado en general, hallar funciones de este tipo y probar que lo son.

Definición 2.2 *Una función hash, es una función de un solo sentido cuya entrada m es un mensaje de longitud arbitraria y la salida es una cadena binaria de longitud fija. Al resumen o*

hash de un mensaje m, se le denotará como $h(m)$. Una función hash debe tener las siguientes propiedades:

- *Para cualquier mensaje m, debe ser posible calcular $h(m)$ eficientemente.*
- *Dado $h(m)$, debe ser computacionalmente difícil, hallar un mensaje m', tal que $h(m) = h(m')$.*
- *Debe ser computacionalmente difícil, hallar dos mensajes m y m' tales que $h(m) = h(m')$.*

Entre las funciones hash que se usan para criptografía están: MD2, MD4, MD5, donde MD significa *Message Digest*, y el algoritmo estándar al momento de escribir éstas notas es el *Secure Hash Algorithm* por sus siglas en inglés SHA. La MD5 fue diseñada por Ron Rivest, toma como entrada un mensaje de longitud arbitraria y proporciona como salida una cadena binaria de 128 bits. El mensaje de entrada se procesa por bloques de 512 bits. La SHA 256 fue diseñada por en NIST (National Institute of Standards and Technology) y se estableció como estándar en 1993. Recibe como entrada un mensaje con longitud menor a 2^{64} bits y como salida se obtiene una cadena binaria de 160 bits. Al igual que el MD5, se procesa en bloques de 512 bits [20].

2.10. Cómputo Nube.

El cómputo nube definido así por el NIST, es un modelo para permitir un acceso a la red ubicuo, es decir, que se encuentra presente en todas partes al mismo tiempo y conveniente a un conjunto de recursos informáticos configurables (por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios) que se puede aprovisionar y liberar rápidamente con un esfuerzo mínimo de gestión o una interacción entre el proveedor de servicios. Este modelo de cómputo nube se compone de 5 características esenciales, 3 modelos de servicio y 4 modelos de despliegue.

Características:

- **Auto-servicio bajo demanda.**
  Un consumidor puede proporcionar unilateralmente capacidades del tiempo del servidor y el almacenamiento en red, según se necesite automáticamente sin interacción con cada proveedor de servicios.
- **Amplio acceso a la red.**
  Las capacidades están disponibles a través de la red y se accede a través de mecanismos que promueven el uso por plataformas de clienteheterogéneas finas o gruesas (por ejemplo, teléfonos móviles, tablets, computadoras portátiles y estaciones de trabajo)
- **Agrupación de recursos.**
  Los recursos informáticos del proveedor se agrupan para servir a múltiples consumidores utilizando un modelo de múlti- usuario, con diferentes recursos físicos y virtuales asignados dinámicamente y reasignados de acuerdo con la demanda del consumidor. Hay una sensación de independencia de ubicación en que el cliente generalmente no

tiene control o conocimiento sobre la ubicación exacta de los recursos proporcionados, pero puede especificar la ubicación en un nivel superior de abstracción (por ejemplo, país, estado o centro de datos). Ejemplos de recursos incluyen almacenamiento, procesamiento, memoria y ancho de banda de la red.

- **Elasticidad rápida.**
  Las capacidades pueden ser suministradas elásticamente y liberadas, en algunos casos de forma automática, para escalar rápidamente hacia fuera y hacia adentro proporcional a la demanda. Para el consumidor, las capacidades disponibles para la provisión a menudo parecen ser ilimitadas y pueden ser apropiadas en cualquier cantidad en cualquier momento.

- **Servicio medido.**
  Los sistemas de cómputo nube controlan y optimizan automáticamente el uso de recursos aprovechando una capacidad de medición en algún nivel de abstracción apropiado al tipo de servicio (por ejemplo, almacenamiento, procesamiento, ancho de banda y cuentas de usuario activas). El uso de recursos puede ser monitoreado, controlado y reportado, proporcionando transparencia tanto para el proveedor como para el consumidor del servicio utilizado.

Modelos de servicio.

- **Software como Servicio (SaaS).**
  La capacidad proporcionada al consumidor es utilizar las aplicaciones del proveedor que se ejecutan en una infraestructura en la nube. Las aplicaciones son accesibles desde varios dispositivos cliente a través de una interfaz de cliente ligero, como un navegador web (por ejemplo, correo electrónico basado en web) o una interfaz de programa. El consumidor no gestiona ni controla la infraestructura oculta de la nube, incluyendo la red, los servidores, los sistemas operativos, el almacenamiento o incluso las capacidades de las aplicaciones individuales, con la posible excepción de las limitadas configuraciones específicas de la configuración de la aplicación.

- **Plataforma como Servicio (PaaS).**
  La capacidad proporcionada al consumidor es desplegar en la infraestructura de la nube aplicaciones creadas por el consumidor, utilizando lenguajes de programación, bibliotecas, servicios y herramientas soportadas por el proveedor. El consumidor no gestiona ni controla la infraestructura oculta de la nube, incluyendo la red, los servidores , sistemas operativos o de almacenamiento, pero tiene control sobre las aplicaciones desplegadas y, posiblemente, configuración de configuración para el entorno de alojamiento de aplicaciones.

- **Infraestructura como Servicio (IaaS).**
  La capacidad proporcionada al consumidor es proveer procesamiento, almacenamiento, redes y otros recursos de computación fundamentales donde el consumidor es capaz de desplegar y ejecutar software arbitrario, que puede incluir sistemas operativos y aplicaciones. El consumidor no gestiona ni controla la infraestructura subyacente de la nube, sino que tiene control sobre los sistemas operativos, el almacenamiento y las aplicaciones implementadas; Y posiblemente un control limitado de componentes de red selectos (por ejemplo, firewalls de host).

Modelos de despliegue.

- **Nube privada.**
  La infraestructura de la nube está preparada para el uso exclusivo de una sola organización que comprende varios consumidores (por ejemplo, unidades de negocio). Puede ser propiedad, administrado y operado por el órgano.

- **Nube de la comunidad.**
  La infraestructura de la nube está preparada para uso exclusivo por una comunidad específica de consumidores de organizaciones que tienen preocupaciones compartidas (por ejemplo, misión, requisitos de seguridad, política y consideraciones de cumplimiento). Puede ser propiedad, administrado y operado por una o más de las organizaciones de la comunidad, un tercero, o una combinación de ellos, y puede existir dentro o fuera de las instalaciones.

- **Nube pública.**
  La infraestructura de la nube está preparada para el uso abierto por el público en general. Puede ser propiedad, administrado y operado por una organización comercial, académica u gubernamental, o alguna combinación de ellos. Existe en las instalaciones del proveedor de la nube.

- **Nube híbrida.**
  La infraestructura de la nube es una composición de dos o más infraestructuras de nube distintas (privadas, comunitarias o públicas) que siguen siendo entidades únicas, pero están unidas por una tecnología estandarizada o propietaria que permite la portabilidad de datos y aplicaciones (por ejemplo, burbujas de nube para equilibrar la carga entre Nubes).

Problemas en Cómputo Nube.

- **Interfaces y API poco seguros.**
  Generalmente los proveedores de servicios en la nube ofrecen una serie de interfaces y API (del inglés, Application Programming Interface) para controlar e interactuar con los recursos. De este modo, toda la organización, el control, la provisión y la monitorización de los servicios cloud se realiza a través de estos API o interfaces. Dado que todo (autenticación, acceso, cifrado de datos, etc.) se realiza a través de estas herramientas, se hace necesario que los interfaces estén diseñados de forma segura, evitando así los problemas de seguridad, tanto los que son intencionados como los que se producen de forma accidental.

- **Pérdida o fuga de información.**
  Existen muchas formas en las que los datos se pueden ver comprometidos. Por ejemplo, el borrado o modificación de datos sin tener una copia de seguridad de los originales, supone una pérdida de datos. En la nube, aumenta el riesgo de que los datos se vean comprometidos ya que el número de interacciones entre ellos se multiplica debido a la propia arquitectura de la misma. Esto deriva en pérdida de imagen de la compañía, daños económicos y, si se trata de fugas, problemas legales, infracciones de normas, etc.

- **Secuestro de sesión o servicio.**
  En un entorno en la nube, si un atacante obtiene las credenciales de un usuario del

entorno puede acceder a actividades y transacciones, manipular datos, devolver información falsificada o redirigir a los clientes a sitios maliciosos.

Capítulo 3

Análisis y Diseño

3.1. Business Process Model and Notation (BPMN)

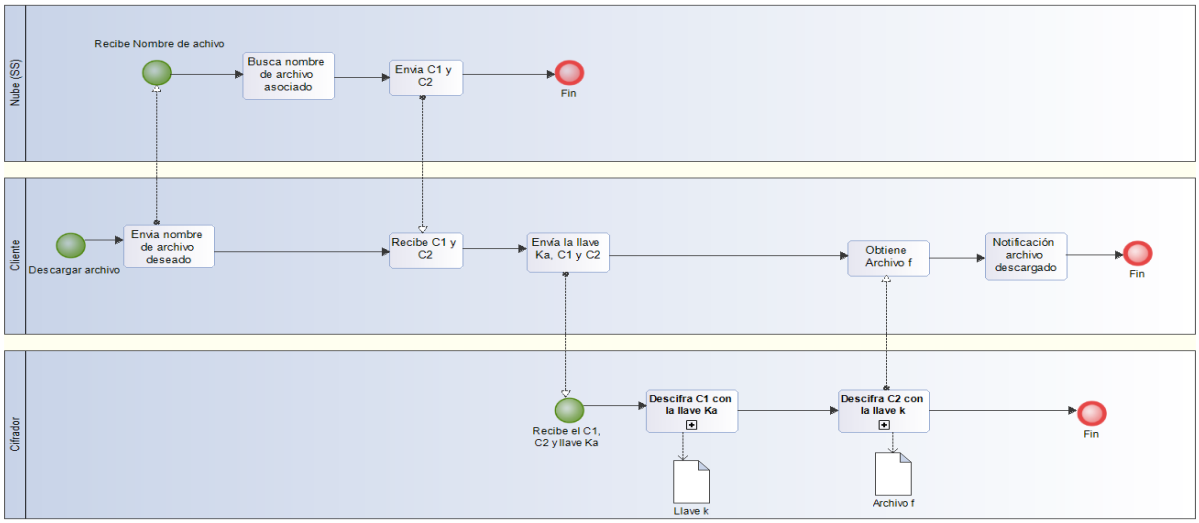

Figura 3.1: BPMN Subir archivo.

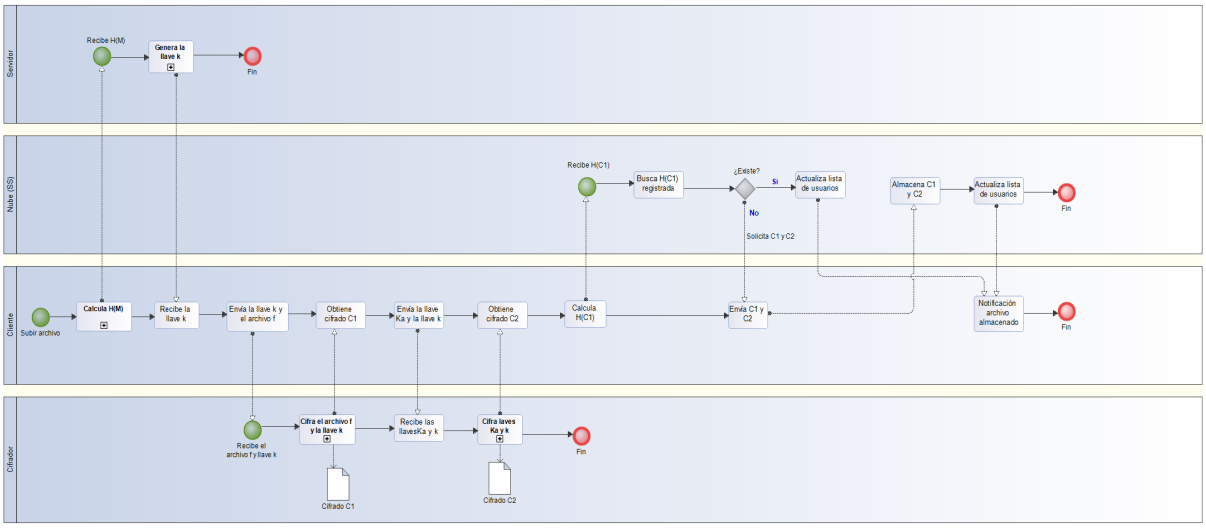

Figura 3.2: BPMN Descargar archivo.

3.2. Requerimientos Funcionales.

Servidor de Llaves	
ID	**Descripción**
RF – SLL1	El sistema permitirá la generación de llaves de usuario a través de una clave secreta (Kg) propia del servidor de llaves.
RF – SLL2	El sistema permitirá la firma a ciegas (y) de cualquier archivo que se desee almacenar.

Tabla 3.1: Requerimientos funcionales del servidor de llaves

Cliente	
ID	**Descripción**
RF – CL1	El sistema permitirá al usuario gestionar archivos: Subir, Descargar, Eliminar
RF – CL2	El sistema permitirá al usuario subir un archivo (F) cifrado al servicio de almacenamiento
RF – CL3	El sistema permitirá al usuario descargar un archivo (F) descifrado elejido de su lista de archivos en el servicio de almacenamiento
RF – CL4	El sistema permitirá al usuario eliminar un archivo (F) cuando el usuario elige alguno de su lista de archivos cargados en el servicio de almacenamiento
RF – CL5	El sistema generará la llave (K) correspondiente a la firma (Y) que envió el servidor de llaves
RF – CL6	El sistema cifrará el archivo (F) que el usuario a solicitado
RF – CL7	El sistema descifrará el archivo (F) que el usuario a solicitado

Tabla 3.2: Requerimientos funcionales del cliente

Servicio de almacenamiento (Nube)	
ID	**Descripción**
RF – SA1	El sistema permitirá al servicio de almacenamiento gestionar archivos: Almacenar, Descargar y Eliminar.
RF – SA2	El sistema permitirá al servicio de almacenamiento guardar cualquier cifrado que el usuario solicite cargar.
RF – SA3	El sistema permitirá al servicio de almacenamiento descargar cualquier cifrado que el usuario tenga en su lista de archivos.
RF – SA4	El sistema permitirá al servicio de almacenamiento eliminar cualquier cifrado que el usuario tenga en su lista de archivos.

Tabla 3.3: Requerimientos funcionales del Servicio de almacenamiento (Nube)

3.3. Requerimientos No Funcionales.

Requerimientos No Funcionales		
RNF1	Eficiencia	▪ El servidor de llaves tendrá la capacidad de realizar 1000 peticiones de gestión de almacenamiento de archivos por segundo. ▪ El sistema podrá funcionar de forma correcta con usuarios conectados de manera concurrente. ▪ Los archivos que sean gestionados dentro del servidor de almacenamiento, deben ser actualizados en la base datos y la visualización de cada cliente de manera casi inmediata.
RNF2	Fiabilidad	▪ La pérdida de consultas en el servidor de llaves es menor a 3 veces el máximo de consultas realizadas. ▪ Los archivos almacenados en el servidor de almacenamiento deben ser recuperados por el usuario al instante en que este lo solicite. ▪ El tiempo de latencia que existe entre el servidor de llaves y el cleinte será de máximo 118ms.
Sigue en la página siguiente.		

ID	Atributo	Descripción
RNF3	Seguridad	▪ El sistema almacenará los datos de los usuarios y sus contraseñas en una base de datos MySQL, dichos datos serán modificados mínimo 2 veces al año. ▪ Se autenticarán los clientes antes de comenzar el proceso de generación de llaves de archivo. ▪ El servidor de llaves firmará claves para un sólo mensaje a la vez sin saber el contenido de éste. ▪ El inicio de sesión de usuarios estará protegido en un canal seguro utilizando almoritmos criptográficos. ▪ Las funciones hash de archivos a almacenar utilizarán la función criptográfica SHA-(256)
RNF4	Mantenibilidad	▪ Cuaquier nuevo requerimiento funcional o no funcional tendrá que ser analizado y diseñado para poder cuantificar las implicaciones que este tendrá sobre el funcionamiento del sistema. ▪ El sistema contará con un plan de pruebas que facilitará la identificación de posibles fallas existentes en el funcionamiento de este.

Tabla 3.4: Requerimientos no funcionales del sistema

Referencias

[1] R. Bellare, Keelveedhi. *Message-locked encryption and secure deduplication.*, volume 7881. EUROCRYPT, 2013.

[2] R. Bellare, Keelveedhi. Dupless: Server-aided encryption for deduplicated storage., 2013:429.

[3] F. Ceballos. Cloud computing, detonador de competitividad. *Forbes*, 2013.

[4] D. Chakraborty and F. Rodríguez-Henríquez. Block cipher modes of operation from a hardware implementation perspective. In Ç. K. Koç, editor, *Cryptographic Engineering*, pages 321–363. Springer, 2009.

[5] T. C. y. P. A. Cooley J. Abs: the apportioned backup system. MIT Laboratory for Computer, 2004.

[6] M. C. y. N. B. Cox L. *SIGOPS Oper. Syst.* Pastiche: making backup cheap and easy, 2002.

[7] P. HP. Estadísticas que todos deberían conocer sobre cloud computing, 2016. `http://www.popa.hn/index.php/es/soluciones/96-otras-noticias/152-20-estadisticas-que-todos-los-cios-deberian-conocer-sobre-cloud-computing`.

[8] L. B. Jaquelina. Fundamentos de criptografía. Universidad Nacional Autónoma de México, 2012. `http://redyseguridad.fi-p.unam.mx/proyectos/criptografia/criptografia/index.php/1-panorama-general/11-concepto-decriptografia`.

[9] A. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[10] Microsoft. "cómputo en la nube": nuevo detonador para la competitividad de méxico. *Instituto Mexicano pala la Competitividad*, 2012.

[11] Microsoft. Data deduplication overview. Biblioteca TechNet, 2015. `https://technet.microsoft.com/enus/library/hh831602(v=ws.11).aspx`.

[12] C. Paar and J. Pelzl. *Understanding Cryptography - A Textbook for Students and Practitioners*. Springer, 2010.

[13] G. Z. C. Patricia. Diseño y desarrollo de un sistema para elecciones electrónicas seguras (seles). Master's thesis, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, 2005.

[14] D. J. R., A. A., B. W. J., S. D., and T. M. *Reclaiming space from duplicate files in a serverless distributed file system.* ICDCS, 2002.

[15] s/a. Rsa. Herramientas WEB para la enseñanza de Protocolos de Comunicación. `http://neo.lcc.uma.es/evirtual/cdd/tutorial/presentacion/rsa.html`.

[16] S/A. Almacenamiento en la nube, ventajas y retos. *D-Link Building Networks for people*, 2011.

[17] s/a. Flud backup, 2011. `http://flud.org/wiki/Flud_Backup`.

[18] s/a. Cifrado simetrico. Guía de Gnu Privacy Guard, 2015. `https://www.gnupg.org/gph/es/manual/c190.html#AEN201`.

[19] T. O. Sergio. Introducción a la criptología. *InfoCentreUV*, 2003.

[20] W. Stallings. *Cryptography and Network Security: Principles and Practice.* Pearson Education, 5a edition, 2002.

[21] H. D. y. W. N. Wilcox-O´Hearn Z. *Tahoe: The least-authority.* In Proceedings of the 4th ACM, 2008.

[22] E. A. M. y M.C. Ma. Jaquelina López Barrientos. Fundamentos de criptografía. Universidad Nacional Autónoma de México, 2012. `http://redyseguridad.fi-p.unam.mx/proyectos/criptografia/criptografia/index.php/1-panorama-general/11-concepto-de-criptografia`.

[23] E. A. M. y M.C. Ma. Jaquelina López Barrientos. Fundamentos de seguridad informática. Universidad Nacional Autónoma de México, 2012. `http://redyseguridad.fi-p.unam.mx/proyectos/criptografia/criptografia/index.php/1-panorama-general/14-ataques/142-ataques-a-los-metodos-de-cifrado`.