Esquema de Cifrado de Datos con Capacidades de Búsqueda

Melissa Brigitthe Hinojosa-Cabello, Miguel Morales-Sandoval
{mhinojosa,mmorales}@tamps.cinvestav.mx
CINVESTAV Tamaulipas
Ciudad Victoria, Tamaulipas, México

RESUMEN

En la actualidad, el cómputo en la nube facilita el acceso a una gran variedad de recursos bajo demanda. Uno de los servicios de tecnologías de la información con mayor demanda es el almacenamiento en la nube. Su auge se debe principalmente a que cada vez se genera una mayor cantidad de datos, los cuales requieren almacenarse de tal forma que se puedan compartir fácilmente con otros usuarios. Asimismo, se espera que dicha información se encuentre disponible en todo momento y desde cualquier dispositivo para cualquier consumidor que tenga autorizado acceder a los datos. Si bien el uso de servicios de almacenamiento en la nube trae consigo grandes ventajas, también existen importantes retos, tales como la confidencialidad de los datos y control de acceso hacia éstos. Aunque la confidencialidad se puede alcanzar mediante el cifrado, dicho proceso dificulta la búsqueda de información. Con esto, una de las grandes ventajas provistas por el almacenamiento en la nube se ve mermada. Además, al cifrar los datos es necesario que el propietario imponga y maneje las restricciones de acceso a su información. Para resolver el problema de las búsquedas sobre datos cifrados ha surgido la técnica criptográfica Searchable Encryption. Ésta cuenta con tres enfoques de solución, sin embargo, no todos son completamente viables para ser implementados en entornos reales de almacenamiento y compartición de datos en la nube. Es por ello que en este trabajo de investigación se propone definir y construir un esquema de cifrado de datos, bajo el enfoque ABSE, que permita preservar las capacidades de búsqueda y recuperación de información. Dicho enfoque se fundamenta en el cifrado basado en atributos (ABE), por lo que es capaz de garantizar control de acceso de grano fino y, al mismo tiempo, permite compartir información con múltiples usuarios.

PALABRAS CLAVE

Criptografía, cifrado basado en atributos, almacenamiento en la nube, compartición de datos, recuperación de información, Searchable Encryption

1. INTRODUCCIÓN

El almacenamiento en la nube es un modelo de servicio en el cual se almacenan, gestionan y respaldan datos de forma remota en servidores externos administrados por un proveedor de servicios. Dado que generalmente el acceso a dichos datos se realiza a través de Internet, es posible administrar la información en cualquier momento y prácticamente a través de cualquier dispositivo digital [1]. El modelo simple del almacenamiento en la nube contempla tres actores, tal como se muestra en la Figura 1:

1. El **propietario de los datos**, que externaliza una colección de documentos hacia un servidor de almacenamiento en la nube.
2. El **proveedor del servicio de almacenamiento**, que es responsable de almacenar los datos de los propietarios y hacerlos disponibles a los usuarios autorizados.
3. El **consumidor de los datos**, el cual realiza operaciones de consulta sobre los datos del propietario, usando para ello un servicio de búsqueda del proveedor.

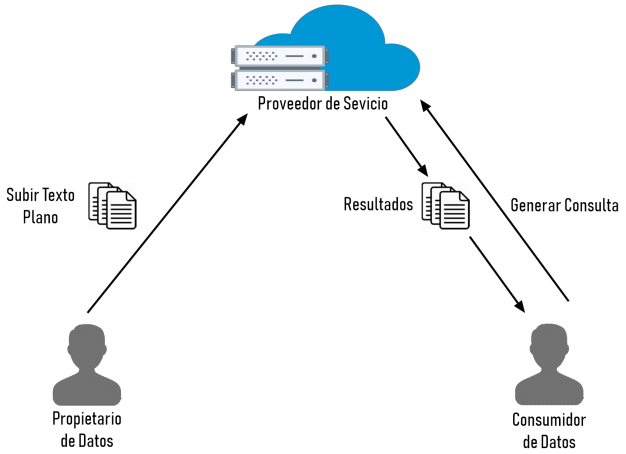

Figura 1: Modelo de operación del almacenamiento en la nube.

Los constantes avances tecnológicos y el aumento en la cantidad de datos que se generan día con día han hecho que los servicios de almacenamiento en la nube sean de vital importancia para usuarios que requieren almacenar o respaldar su información. Con mayor frecuencia, tanto personas como organizaciones eligen resguardar su información en servidores externos a través de servicios de almacenamiento en la nube.

2. PLANTEAMIENTO DEL PROBLEMA

Una de las principales preocupaciones de los propietarios de datos en el contexto del almacenamiento en la nube es la confidencialidad de su información, ya que el proveedor del servicio, al tener pleno acceso a sus datos, pudiera usarlos, por ejemplo, para aprender de ellos sin su consentimiento. Si bien los proveedores del servicio pueden ser honestos y no divulgar la información de los usuarios, también es cierto que pueden ser curiosos y que a partir de los datos de los que disponen pueden aprender o derivar información. En este contexto, podría decirse que el principal adversario en el servicio de almacenamiento en la nube es el proveedor mismo [2, 3].

De acuerdo con [4] los proveedores de servicio de almacenamiento pueden categorizarse como *terceros de confianza* (TTP), *honestos pero curiosos* (HBC) o *semi-honestos pero curiosos* (SHBC). En el primer caso, los usuarios del servicio confían totalmente en el proveedor, respaldado esencialmente por los acuerdos de nivel de servicio (SLAs) y el prestigio de su marca. En el modelo HBC, el proveedor se considera honesto porque almacena los archivos de sus clientes y realiza las funciones esperadas de acuerdo al protocolo establecido en los SLAs. No obstante, trata de inferir tanta información como le sea posible a partir de las acciones e interacciones de los usuarios. En el modelo SHBC, un proveedor semi-honesto pero curioso se comporta como en el modelo HBC, pero, adicionalmente, puede ejecutar la totalidad de las operaciones de búsqueda o solamente un fragmento de éstas y devolver al usuario todos o una porción de los resultados que ha encontrado de forma honesta.

Una alternativa de solución para el propietario es cifrar su información previo a enviarla al proveedor de almacenamiento en la nube y resguardar sus claves criptográficas de descifrado de tal forma que él sea el único que pueda acceder a sus datos en claro. Sin embargo, dicha alternativa supone dos inconvenientes:

- Las capacidades de búsqueda y recuperación de datos del proveedor no se pueden aprovechar si los datos se encuentran cifrados. El propietario tendría que descargar toda su información, descifrarla y, posteriormente, aplicar un algoritmo de búsqueda y recuperación de forma local.
- La compartición de información con consumidores de datos se vuelve una tarea compleja, ya que sería responsabilidad del propietario implementar un mecanismo para distribuir de forma eficiente las claves criptográficas para poder descifrar los datos.

Searchable Encryption (SE) es una técnica criptográfica cuyo objetivo es mantener la privacidad de los datos (confidencialidad), al mismo tiempo que brinda a los usuarios funciones de recuperación de información. Mediante SE es posible para el propietario almacenar datos cifrados en servidores externos poco confiables y garantizar a aquellos usuarios autorizados la posibilidad de realizar operaciones de búsqueda. De esta forma se aprovechan al máximo las ventajas de la computación en la nube y se evitan sobrecargas de procesamiento y comunicación del lado del usuario.

Mediante SE se protege no solo la información de los usuarios de servicios de almacenamiento, sino también las consultas enviadas al servidor, los resultados de éstas, así como los datos y/o archivos recuperados derivado de dichas consultas. Asimismo, SE permite garantizar la seguridad de documentos y palabras clave, patrones de búsqueda y patrones de acceso de los usuarios. De esta forma, se minimizan los riesgos de filtrado de información dado que el servidor externo no puede aprender a partir de las consultas de los usuarios ni de los documentos bajo su reguardo [5, 6]. Aunque el concepto de SE es simple, la construcción de esquemas eficientes de este tipo representa una tarea compleja y se mantiene hasta ahora como un problema abierto.

Si bien en la literatura existen algunos esquemas SE propuestos, también es cierto que estos esquemas no han sido completamente estudiados ni evaluados para ser usados en la práctica en escenarios reales. De los enfoques existentes de SE, ABSE[1] es el más adecuado para escenarios de almacenamiento en la nube debido a que:

1. Permite implementar mecanismos de control de acceso criptográfico, con lo cual se puede resolver el problema de distribución de información a los consumidores.
2. Permite el cifrado de los datos, con lo cual se garantiza la confidencialidad de los datos del propietario.
3. Provee mecanismos de recuperación de información en colecciones cifradas, con lo cual es posible que el servidor realice tareas de búsqueda sin comprometer la confidencialidad de los datos ni la eficiencia del lado de productores y consumidores.

Sin embargo, aunque de forma teórica ABSE cumple con los requerimientos de seguridad y garantiza las capacidades de búsqueda del proveedor del servicio de almacenamiento en la nube, en la literatura no se encuentran construcciones de ABSE que hayan sido evaluadas en escenarios reales. Por un lado, los requerimientos funcionales básicos en un esquema SE no son suficientes para una aplicación eficiente en el entorno del almacenamiento en la nube. Por otro lado, un requerimiento no funcional muy importante de un esquema ABSE es el nivel de seguridad que es capaz de ofrecer.

De acuerdo con diversos estándares internacionales [3, 7], como el reporte NIST SP800-57 y el proyecto ECRYPT-CSA D5.4, actualmente los niveles de seguridad esperados son al menos de 128-bits. No obstante, en la literatura no existe una construcción de un esquema ABSE que permita implementarse en niveles de seguridad mayores a 80-bits, el cual representa un nivel de seguridad obsoleto. Realizar la construcción de un esquema ABSE para un nivel de seguridad de 128-bits o más no es trivial. Ello implica cambiar la definición de los algoritmos que conforman al esquema ABSE para poder operar con emparejamientos bilineales asimétricos, los cuales son objetos matemáticos basados en criptografía de curva elíptica. Otro requerimiento no funcional relevante para un esquema ABSE es la eficiencia, el cual es crítico para que el esquema realmente pueda utilizarse en aplicaciones reales. Sin embargo, los esquemas basados en emparejamientos bilineales y en curvas elípticas son tradicionalmente costosos computacionalmente.

3. TRABAJO RELACIONADO

En la literatura existen diversos trabajos relacionados con la definición e implementación de esquemas de búsqueda sobre colecciones de datos cifrados. Como se mencionó anteriormente, Searchable

[1] Attribute-Based Searchable Encryption

Encryption cuenta con diversos enfoques, cuya implementación en escenarios reales conlleva ciertas ventajas y desventajas. No obstante, no todos ellos son realmente eficientes o apropiados para poner en práctica, particularmente en ambientes tan complejos como el cómputo en la nube. En este contexto, la implementación del enfoque ABSE resulta más conveniente dado que éste garantiza un control de acceso de grano fino a partir de la definición de políticas en función de un conjunto de atributos que caracterizan a las entidades.

Una de las primeras propuestas centradas en el enfoque basado en atributos de SE es presentada en [8] por Wenhai Sun et al. En ésta se plantea un esquema escalable de búsquedas de una o varias palabras clave sobre datos cifrados en un contexto multi-propietario y multi-usuario. La autorización de búsqueda se basa en el uso de CP-ABE[2] con el objetivo de proveer un control de acceso de grano fino a nivel de archivos. No obstante, para cada archivo se genera su propio índice seguro con base en una estructura de acceso de compuertas AND, lo cual implica el control, del lado del servidor, de tantas estructuras de acceso como archivos tenga el propietario.

Por otro lado, en [9] se propone un esquema en el cual los propietarios de datos se aseguran de la confidencialidad de su información mediante la definición de una política de control de acceso. Así, únicamente aquellos usuarios que cumplan con los atributos definidos en dicha política podrán realizar búsquedas y, posteriormente, acceder al contenido legible de los archivos del propietario. En este enfoque, se garantiza además la privacidad del usuario al evitar que el propietario conozca qué palabras clave buscan los usuarios sobre los datos que mantiene externalizados. Sin embargo, el alcance de esta propuesta se limita al uso de datos estáticos y requiere que las credenciales de acceso de propietarios y usuarios de datos sean enviadas por una autoridad de confianza a través de canales de comunicación privados.

En la propuesta de solución presentada en [10] se describe la implementacion de SE multi-propietario y multi-usuario basado en CP-ABE. En éste, se utiliza un índice seguro y un esquema simétrico para cifrar los archivos de los propietarios de datos. De esta forma, solamente cuando los atributos de un usuario satisfacen la política de acceso, éste podrá acceder a los datos cifrados del propietario de los datos. Así, el proveedor de servicio solamente tiene la facultad de ejecutar operaciones de búsqueda sobre la colección de datos cifrados y no puede obtener o derivar información a partir de la información que almacena. Si bien se utiliza un enfoque simétrico para cifrar los archivos con el fin de asegurar la eficiencia del sistema, no se especifica cómo se soluciona el problema de compartición de llaves asociado al uso de cifrado simétrico.

El trabajo realizado en [11] combina los esquemas SSE[3] y ABE, por lo que se considera de tipo cifrado híbrido. Por una parte, los propietarios de datos garantizan la confidencialidad de su información al cifrarla utilizando SSE. Por otro lado, mediante CP-ABE se garantiza un control de acceso de grano fino. De esta forma, los propietarios de datos pueden compartir sus archivos con múltiples usuarios de forma segura y eficiente. Sin embargo, esta propuesta no ha sido probada en un entorno multi-cloud, donde diferentes usuarios u organizaciones pueden estar utilizando plataformas en la nube completamente distintas.

Por su parte, en [12] se propone un esquema ABE jerárquico en el cual una colección de documentos que comparten una misma estructura de acceso puede cifrarse en conjunto, en lugar de cifrarse individualmente. Dicha estructura es construida basándose en el modelo TF-IDF[4] y en los atributos asignados a los archivos para organizar los vectores de documentos que permitirán la recuperación de información. No obstante, el algoritmo de búsqueda profunda diseñado para lograr la recuperación de datos emplea una estrategia voraz a partir de la cual no es posible asegurar la convergencia a una solución adecuada. Asimismo, la eficiencia de esta propuesta solamente fue analizada de forma teórica y evaluada mediante simulaciones, por lo que, de forma práctica, no se garantiza su efectividad.

4. PROPUESTA DE INVESTIGACIÓN

En la literatura no exite un esquema ABSE que se haya construido teniendo en cuenta la eficiencia y evaluación en escenarios de almacenamiento en la nube. De acuerdo al estado actual de la literatura sobre esquemas ABSE se deriva que:

1. Los requerimientos funcionales de ABSE en escenarios de almacenamiento no se han identificado explícitamente.
2. No se cuenta con esquemas ABSE con construcciones que permitan implementar niveles de seguridad mayores a 128-bits.
3. No se han identificado construcciones ABSE que se hayan evaluado experimentalmente en escenarios de almacenamiento y compartición de datos en la nube en donde se haya demostrado su eficiencia.

Dado lo anterior, en este trabajo de investigación se propone diseñar e implementar un esquema de cifrado de datos, bajo el enfoque ABSE, que permita preservar las capacidades de búsqueda y recuperación de información. Es decir, que opere bajo el modelo de ataque semi-honesto pero curioso en escenarios reales de almacenamiento y compartición de datos en la nube, de tal forma que cumpla con los requerimientos funcionales mínimos en dichos entornos y, a la vez, cubra aspectos no funcionales, como nivel de seguridad y eficiencia.

Para ello, es necesario identificar los requerimientos esenciales de un esquema ABSE para su uso en entornos de almacenamiento y compartición de datos en la nube, así como el algoritmo de indexamiento de texto cifrado que resulta más eficiente para su implementación en un esquema de este tipo. Lo anterior permitirá definir los módulos principales y la arquitectura del esquema, considerando aspectos funcionales, criptográficos, de recuperación de información, entre otros. De esta forma se realizará la construcción del esquema ABSE que permita su despliegue en aplicaciones reales proveyendo niveles de seguridad iguales o mayores a 128-bits. Asimismo, será posible determinar la factibilidad del esquema ABSE propuesto, mediante su evaluación experimental bajo escenarios similares a los del entorno de almacenamiento y compartición de datos en la nube.

[2] Ciphertext-Policy Attribute-Based Encryption
[3] Searchable Symmetric Encryption
[4] Frecuencia de Término - Frecuencia Inversa de Documento

En la Figura 2 se muestra el modelo de operación del esquema propuesto. Como se puede observar, un esquema bajo el enfoque ABSE considera los mismos actores que el modelo básico de operación de SE y el del almacenamiento en la nube. No obstante, en ABSE existe un actor adicional a los tres principales contemplados en SE. Dicho actor es denominado autoridad de confianza y posee una pareja de claves, las cuales se encuentran correlacionadas mediante criptografía basada en emparejamientos bilineales. Además, la autoridad de confianza se encarga de administrar los atributos de los usuarios del sistema y proveerles sus claves secretas de usuario, así como de proporcionar a los propietarios de los datos la clave pública a partir de la cual podrán cifrar tanto el índice seguro como la colección de documentos a externalizar.

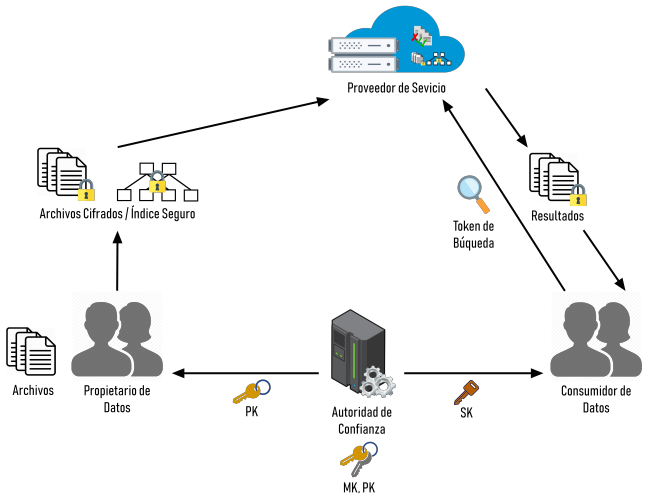

Figura 2: Modelo de operación del esquema de cifrado de datos con capacidades de búsqueda.

El desarrollo del proyecto de investigación se divide en 3 fases, a partir de las cuales progresivamente se obtienen avances en la construcción del esquema ABSE propuesto y en el cumplimiento de los objetivos planteados.

- **Fase 1 –** Definición de los requerimientos esenciales de un esquema ABSE
- **Fase 2 –** Construcción del esquema propuesto
- **Fase 3 –** Diseño de pruebas y evaluación del rendimiento del esquema

5. PRINCIPALES RESULTADOS Y CONTRIBUCIONES

Los principales resultados esperados del desarrollo de este trabajo de investigación se enuncian a continuación.

Contribución tecnológica:

- Un nuevo esquema de seguridad que opere en el modelo *semi-honesto pero curioso* utilizando emparejamientos bilineales asimétricos.

Contribución científica y académica:

- Evaluación experimental del esquema que demuestre la viabilidad de su uso en entornos reales.
- Prototipo experimental que permita evaluar nuevas técnicas criptográficas y de recuperación de información en el entorno del almacenamiento en la nube.

6. CONCLUSIONES

Si bien en el estado del arte se pueden encontrar diversos trabajos relativos a la implementación de esquemas ABSE, dichas propuestas aún cuentan con importantes áreas de oportunidad. Por ejemplo, en todos los casos los niveles de seguridad provistos son inferiores al estándar actual de 128-bit. Asimismo, la mayoría no han sido implementadas en ambientes reales y complejos, por lo que su eficiencia y viabilidad solamente han sido evaluadas de forma teórica. Es por ello que resulta de vital importancia contar con un esquema que permita garantizar los servicios de seguridad requeridos en entornos de almacenamiento y compartición de datos en la nube y que, además, preserve las funcionalidades de búsqueda y demuestre su viabilidad de uso en escenarios reales con base en los resultados de la evaluación de un prototipo experimental.

7. AGRADECIMIENTOS

Este proyecto de investigación se realiza en el marco del proyecto 281565 del Fondo de Investigación para la Educación SEP-CONACYT, bajo la dirección del Dr. Miguel Morales Sandoval.

REFERENCIAS

[1] Cisco Networking Academy, "Connecting networks," tech. rep., Cisco Systems, Inc., 2016.
[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Proceedings of Financial Cryptography: Workshop on Real-Life Cryptographic Protocols and Standardization 2010*, vol. 6054, pp. 136–149, Microsoft Corporation, Springer, 01 2010.
[3] M. Morales-Sandoval, J. L. González-Compeán, A. Díaz-Pérez, and V. J. Sosa-Sosa, "A pairing-based cryptographic approach for data security in the cloud," *International Journal of Information Security*, vol. 17, pp. 441–461, 08 2018.
[4] J. Wang and Z. A. Kissel, *Introduction to Network Security*. Wiley Publishing, Inc., 2nd. ed., 2015.
[5] K. Chamili, J. Nordin, W. Ismail, and A. Radman, "Searchable encryption: A review," *Security and its Applications*, pp. 79–88, 12 2017.
[6] Y. Wang, J. Wang, and X. Chen, "Secure searchable encryption: A survey," *Communications and Information Networks*, vol. Vol. 1, pp. 52–65, 12 2016.
[7] D. Giry, "Nist recommendation for key management (2016)."
[8] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, vol. 108, pp. 226–234, IEEE, 05 2014.
[9] Q. Zheng, S. Xu, and G. Ateniese, "VABKS: Verifiable attribute-based keyword search over outsourced encrypted data," *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 522–530, 2014.
[10] W. Guo, X. Dong, Z. Cao, and J. Shen, "Efficient attribute-based searchable encryption on cloud storage," vol. 1087, IOP Publishing, 2018.
[11] A. Michalas, "The lord of the shares: Combining attribute-based encryption and searchable encryption for flexible data sharing," 2018.
[12] N. Wang, J. Fu, B. K. Bhargava, and J. Zeng, "Efficient retrieval over documents encrypted by attributes in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 13, 2018.