Cifrado CP-ABE para el ecosistema Apache Hadoop

Aitor Osa, Iñaki Garitano, Ignacio Arenaza-Nuño, Urko Zurutuza, Mikel Iturbe
Departamento de Electrónica e Informática
Escuela Politécnica Superior
Mondragon Unibertsitatea
Goiru 2, E-20500 Arrasate-Mondragón
Email: aitor.osal@alumni.mondragon.edu, {igaritano,iarenaza,uzurutuza,miturbe}@mondragon.edu

Resumen- **La socialización de herramientas y soluciones Big Data junto con la aparición de dispositivos Internet of Things ha aumentado tanto el volumen como el interés hacia el valor de los datos. Bien sea por el posible valor o conocimiento que se puede obtener a partir del análisis como el carácter sensible de los datos, en muchos casos su protección resulta indispensable, especialmente cuando se hace uso de infraestructura de terceros. Así, es necesario ofrecer a los usuarios mecanismos que protejan la seguridad y la privacidad de sus datos. A lo largo de este trabajo se exponen algunos de los retos en la implementación del algoritmo de cifrado Ciphertext-Policy Attribute-Based Encryption (CP-ABE) sobre el ecosistema Apache Hadoop. Dicho algoritmo ofrece una protección de grado fino, pudiendo llegar a definir múltiples atributos sobre los cuales se establecerá un esquema de cifrado adecuado para cada caso.**

Index Terms- **CP-ABE, Ciberseguridad, Apache Hadoop, Apache Ranger, Transparent Data Encryption.**

Tipo de contribución: *Investigación en desarrollo*

I. Introducción

Tanto la oferta como la demanda de infraestructura de nube pública han aumentado considerablemente a lo largo de los últimos años. Son muchos los actores, como las pequeñas y medianas empresas, incluso particulares, que prefieren hacer uso de nube pública cuando de ingesta, almacenamiento y análisis de grandes volúmenes de datos se trata. Así, evitan o reducen la inversión que supone montar la infraestructura necesaria *in-house*. Sin embargo, el uso de infraestructura pública presenta una serie de riesgos asociados, ya que compartir el espacio físico donde residen los datos puede dar lugar a accesos no autorizados tanto de agentes externos a la infraestructura como a personal interno, suponiendo un riesgo de seguridad y privacidad considerable.

La protección de los datos mediante el control de acceso es una labor correspondiente al propietario de los datos, quien es el responsable de protegerlos [1]. Dicho control debe restringir el acceso solo a aquellos usuarios autorizados [2], realizando así un control acceso granular.

Sin embargo, el control de acceso granular no es el único mecanismo que permite proteger los datos [3]. En este sentido, el cifrado de los datos almacenados en infraestructura pública aumenta su nivel de protección, dificultando su compresión aun disponiendo de acceso físico.

II. Trabajos Relacionados

Apache Hadoop es una herramienta de almacenamiento y análisis de grandes volúmenes de datos usado ampliamente en la actualidad, pero presenta un pobre mecanismo de seguridad [4]. Según Zhou *et al.* [3] Apache Hadoop no garantiza la confidencialidad de datos de manera robusta, debido a que los métodos de cifrado *Identity-based Encryption* y *Public Key Infraestructure* ofrecen al proveedor de infraestructura acceso a todos los datos, resultando en mecanismos no recomendables para este tipo de herramientas. Aunque Apache Hadoop ofrezca la posibilidad de cifrar los datos mediante el algoritmo de cifrado *Advanced Encryption Standard* (AES), Zhou *et al.* [4] proponen la implementación de CP-ABE como mecanismo de cifrado robusto.

Desde que el mecanismo de cifrado y control de acceso robusto CP-ABE fue presentado por Bethencourt *et al*. [2], se ha empleado en diversos campos, desde el cifrado de comunicaciones en redes móviles de escenarios de recuperación de desastres [5] hasta el cifrado de datos en dispositivos *Internet of Things* (IoT) [7], pasando por el cifrado de datos en la nube [8].

El trabajo presentado por Roy *et al.* [5] hace uso del método de cifrado de CP-ABE en ámbitos de falta de conectividad continuada. En situaciones en las que la conectividad no es continua resulta importante que los datos permanezcan cifrados de forma indefinida. Así, debido a que CP-ABE proporciona acceso de control de grado fino, resulta un mecanismo adecuado para el uso en sistemas donde los retrasos de red sean frecuentes e incluso cuando se pierda el acceso durante varios días.

Así mismo, el trabajo presentado por Jo *et al.* [7] propone el uso de CP-ABE sobre dispositivos IoT para aplicaciones móviles con restricciones de consumo energético. Aunque distintos trabajos como el presentado por Guo *et al.* [6], declaran que CP-ABE presenta una desventaja frente a otros métodos de cifrado como *Identity-Based Encription* o *Multi-Identity Single-Key Encryption* debido a la longitud dinámica de las claves de descifrado, las soluciones presentadas por Jo *et al.* [7] y Guo *et al.* [6] proponen nuevos métodos de generación de claves de descifrado de longitud constantes lo cual facilitaría el uso de CP-ABE sobre dispositivos de poca capacidad de procesamiento.

III. Acercamiento al Problema

A lo largo del presente trabajo, el cual todavía está en curso, se está realizando la implementación del método de

cifrado CP-ABE realizada por Bethencourt *et al.* [2], sobre Apache Hadoop[1] y Apache Ranger[2], el cual es un framework que sirve para monitorizar y gestionar la seguridad integral de los datos y los procesos del ecosistema Apache Hadoop.

A. Importancia de CP-ABE

El método de cifrado CP-ABE ofrece la posibilidad de cifrar los datos en base a un conjunto de atributos, los cuales pueden estar asociados a entidades de distinto carácter, como por ejemplo las personas. Así, cada entidad puede disponer de un conjunto de atributos distinto, dando lugar a un acceso a datos en base al conjunto de atributos disponible.

Apache Hadoop permite el almacenamiento de grandes volúmenes de datos de forma distribuida y fiable. Para ello, replica los datos tanto de forma local como a través de la red de comunicaciones entre los distintos dispositivos que forman el clúster de almacenamiento. En este sentido, la implementación de CP-ABE como mecanismo de cifrado ofrece un cifrado granular de datos tanto para los datos almacenados de forma estática, como para los datos en tránsito entre los distintos dispositivos del clúster.

B. Implementación sobre Apache Hadoop y Apache Ranger

Los objetivos del presente trabajo son principalmente (1) la implementación del conjunto de herramientas de cifrado CP-ABE realizada por Bethencourt *et al.* [2] sobre Apache Hadoop y (2) la integración de la gestión de CP-ABE, la gestión de atributos y usuarios, en Apache Ranger.

La implementación original de CP-ABE está realizada en el lenguaje de programación C. Sin embargo, tanto Apache Hadoop como Apache Ranger están escritos en el lenguaje de programación Java. Debido a ello y con el objetivo de evitar cualquier error de implementación asociada a la reescritura del código de CP-ABE en el lenguaje de programación Java, se ha hecho uso de la tecnología *Java Native Interface* (JNI). Así, se han creado distintos métodos en Java, los cuales hacen de interfaz entre el código en Java de Apache Hadoop y Apache Ranger y las funciones originales en C, evitando de esta manera alterar el código original de CP-ABE.

A partir de la versión 2.6, Apache Hadoop ofrece la característica *transparent data encryption*, la cual hace uso del algoritmo criptográfico AES. Debido a que el algoritmo AES es un algoritmo de cifrado de clave simétrico, la misma clave es usada tanto para el cifrado como el descifrado de los datos. Además, la longitud de dicha clave solo puede variar entre 128, 192 o 256 bits.

CP-ABE por su parte utiliza claves distintas tanto para el cifrado como el descifrado de los datos. Además, dependiendo del número de atributos asociados, la longitud de cada clave de descifrado es variable.

Apache Hadoop está diseñado para realizar la gestión y uso de las claves de cifrado de AES, claves simétricas de longitud fija. Así, su diseño no comprende el uso de claves distintas para el cifrado y descifrado de los datos, resultando ser un reto para la implementación de CP-ABE. Así mismo, el uso de claves de longitud variable puede inducir en la necesidad de realizar un cambio en el diseño de Apache Hadoop, lo cual puede resultar en un trabajo de dimensiones considerables.

A su vez, las versiones más recientes de Apache Ranger no están diseñadas para soportar la gestión de atributos y atributos asociados a usuarios. Esta es una característica imprescindible si, una vez implementado CP-ABE sobre Apache Hadoop, se pretende facilitar una gestión centralizada de la seguridad de los datos.

IV. CONCLUSIONES

La seguridad junto a la privacidad de los datos sigue siendo uno de los mayores problemas del Big Data actual [4]. Por ello, en el presente trabajo se ha presentado una aproximación para extender Apache Hadoop y Apache Ranger, con el objetivo de ofrecer a los usuarios la posibilidad de cifrar los datos junto con un control de acceso de grado fino.

AGRADECIMIENTOS

Este trabajo ha sido desarrollado por el grupo de Sistemas Inteligentes para Sistemas Industriales, financiado por el Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco.

REFERENCIAS

[1] Priyanka Rajput and Pankaj Kawadkar. Highly Secure Method based on Ciphertext Policy Attribute based Encryption in Hadoop System. International Journal of Computer Applications, Foundation of Computer Science, 2014, 103

[2] John Bethencourt, Amit Sahai and Brent Waters. Ciphertext-Policy Attribute-Based Encryption. IEEE Symposium on Security and Privacy, 2007, 321-334

[3] Zhou, H. and Wen, Q. A new solution of data security accessing for Hadoop based on CP-ABE. IEEE 5th International Conference on Software Engineering and Service Science, 2014, 525-528

[4] RezaeiJam, M.; Khanli, L. M.; Akbari, M. K. & Javan, M. S. A survey on security of Hadoop. 4th International Conference on Computer and Knowledge Engineering (ICCKE), 2014, 716-721

[5] Roy, S and Chuah, M. Secure data retrieval based on ciphertext policy attribute-based encryption (CP-ABE) system for the DTNs. Citeseer, 2009

[6] Guo, F.; Mu, Y.; Susilo, W.; Wong, D. S. & Varadharajan, V. CP-ABE with constant-size keys for lightweight devices. IEEE transactions on information forensics and security, IEEE, 2014, 9, 763-771

[7] Minho Jo, Vanga Odelu, Ashok Kumar Das, Muhammad Khurram Khan and Kim-Kwang Raymond Choo. Expressive CP-ABE Scheme for Mobile Devices in IoT satisfying Constant-size Keys and Ciphertexts. IEEE Access, 2017, PP, 1-1

[8] Guojun Wang, Qin Liu and Jie Wu. Hierarchical Attribute-based Encryption for Fine-grained Access Control in Cloud Storage Services. Proceedings of the 17th ACM Conference on Computer and Communications Security, ACM, 2010, 735-737

[1] https://hadoop.apache.org/

[2] https://ranger.apache.org/