ISSN: 2249-7196
IJMRR/April 2022/ Volume 12/Issue 2/ 01-12
CHOWDOJU UTTHEJ/ International Journal of Management Research & Review

An Efficient and Fine-grained Big Data Access Scheme with Privacy-Preserving Policy

CHOWDOJU UTTHEJ[1], GUDIMETLA SREECAR SARMAA[2], MEDIDHA VINEEL KUMAR,[3] BHUKYA SUMAN[4]
SUPERVISOR,Y LAXMI PRASANNA
Associate Professor
ANURAG ENGINEERING COLLEGE
AUTONOMOUS
(Affiliated to JNTU-Hyderabad,Aprroved by AICTE-New Delhi)
ANANTHAGIRI (V) (M), SURYAPETA (D), TELANGANA-508206

Abstract: *How to control the access of the huge amount of big data becomes a very challenging issue, especially when big data are stored in the cloud. Ciphertext-Policy Attribute based Encryption (CP-ABE) is a promising encryption technique that enables end-users to encrypt their data under the access policies defined over some attributes of data consumers and only allows data consumers whose attributes satisfy the access policies to decrypt the data. In CP-ABE, the access policy is attached to the ciphertext in plaintext form, which may also leak some private information about end-users. Existing methods only partially hide the attribute values in the access policies, while the attribute names are still unprotected. In this paper, we propose an efficient and fine-grained big data access control scheme with privacy-preserving policy. Specifically, we hide the whole attribute (rather than only its values) in the access policies. To assist data decryption, we also design a novel Attribute Bloom Filter to evaluate whether an attribute is in the access policy and locate the exact position in the access policy if it is in the access policy. Security analysis and performance evaluation show that our scheme can preserve the privacy from any LSSS access policy without employing much overhead.*

Keywords: *Ciphertext-Policy Attribute based Encryption, Cloud computing, fine grained access control*.

I. INTRODUCTION

In the era of big data, a huge amount of data can be generated quickly from various sources (e.g., smart phones, sensors, machines, social networks, etc.). Towards these big data, conventional computer systems are not competent to store and process these data. Due to the flexible and elastic computing resources, cloud computing is a natural fit for storing and

processing big data. With cloud computing, end-users store their data into the cloud, and rely on the cloud server to share their data to other users (data consumers). In order to only share end-users' data to authorized users, it is necessary to design access control mechanisms according to the requirements of end-users. When outsourcing data into the cloud, end-users lose the physical control of their data. Moreover, cloud service providers are not fully-trusted by end-users, which makes the access control more challenging. For example, if the traditional access control mechanisms (e.g., Access Control Lists) are applied, the cloud server becomes the judge to evaluate the access policy and make access decision. Thus, end-users may worry that the cloud server may make wrong access decision intentionally or unintentionally, and disclose their data to some unauthorized users. In order to enable end-users to control the access of their own data, some attribute-based access control schemes are proposed by leveraging attribute-based encryption. In attribute-based access control, end-users first define access policies for their data and encrypt the data under these access policies. Only the users whose attributes can satisfy the access policy are eligible to decrypt the data. Although the existing attribute-based access control schemes can deal with the attribute revocation problem they all suffer from one problem: the access policy may leak privacy.

This is because the access policy is associated with the encrypted data in plaintext form. From the plaintext of access policy, the adversaries may obtain some privacy information about the end-user. For example, Alice encrypts her data to enable the "Psychology Doctor" to access. So, the access policy may contain the attributes "Psychology" and "Doctor". If anyone sees this data, although he/she may not be able to decrypt the data, he/she still can guess that Alice may suffer from some psychological problems, which leaks the privacy of Alice. To prevent the privacy leakage from the access policy, a straightforward method is to hide the attributes in the access policy. However, when the attributes are hidden, not only the unauthorized users but also the authorized users cannot know which attributes are involved in the access policy, which makes the decryption a challenging problem. Due to this reason, existing methods do not hide or anonymize the attributes. Instead, they only hide the values of each attribute by using wildcards, Hidden Vector Encryption and Inner Product Encryption. Hiding the values of attributes can somehow protect user privacy, but the attribute name may also leak private information. Moreover, most of these partially hidden policy schemes only support specific policy structures (e.g., AND-gates on multi-valued attributes).

The rise of cloud computing not only promotes changes in the information technology industry structure and operation mode but also produces a win–win situation for enterprises and individuals. Enterprises can outsource resources to cloud service providers, thus reducing hardware overhead and data maintenance costs. Moreover, customers can enjoy a variety of low-cost cloud services by the pay-per-use mode. The current cloud architecture can be further divided into public clouds and private clouds. Compared with a private cloud, a public cloud is more beneficial for realizing resource sharing, but it also faces more intractable security problems. Specifically, while the owner of cloud resources wants to prevent unauthorized users from accessing data stored in the cloud server, authorized users do not want their access behaviour to be monitored and tracked by the cloud service provider (SP).

Attribute-based encryption (ABE) is a one-to-many encryption scheme that makes it unnecessary for data owners to know the number and identity of users in the encryption stage. Data owners need only define unique access structures for sensitive data based on user attributes. Moreover, ABE combines attributes with ciphertexts and users' private keys so that users can access sensitive data only if the intersection of their attribute set and ciphertext attribute set satisfies the access structure defined in the encryption stage. ABE has become an important tool for meeting the needs of fine-grained access control in cloud storage applications.

II. LITERATURE SURVEY

Identify Relevant Keywords: Start by brainstorming and identifying keywords related to your topic, such as "big data access control," "privacy-preserving policy," "fine-grained access control," etc. These keywords will be useful when searching for relevant literature. Search Databases and Libraries: Utilize academic databases and libraries such as IEEE Xplore, ACM Digital Library, Google Scholar, or PubMed to search for research papers, conference proceedings, and journals related to your topic. Use the keywords identified in the previous step to refine your search. Refine Search Results: Review the titles, abstracts, and keywords of the search results to identify papers that closely match your topic. Exclude irrelevant papers and focus on those that specifically address an efficient and fine-grained big data access control scheme with privacy-preserving policy. Read the Selected Papers: Obtain access to the selected papers and thoroughly read them to understand the proposed access control scheme, the privacy-preserving techniques used, and the evaluation results. Take note of the strengths, weaknesses, and limitations of each scheme. Analyze Related Work: Pay attention to the related work sections of the selected papers. This will help you identify additional references and papers that might not have appeared in your initial search. These

references can provide further insights and comparisons for your literature survey. Identify Common Approaches: Look for common approaches, methodologies, or techniques used in the papers you've reviewed. Identify recurring themes or trends in the literature, such as the use of attribute-based access control, encryption, or anonymization techniques for privacy preservation. Summarize and Compare: Summarize the key findings, methodologies, and results from each paper. Compare the different access control schemes, highlighting their similarities, differences, and contributions. Identify the strengths and weaknesses of each scheme and how they address the challenges of efficient and fine-grained big data access control with privacy preservation. Provide a Critical Analysis: Offer a critical analysis of the literature, discussing the gaps, limitations, and open research questions that you have identified. This analysis can serve as a basis for future research and potential enhancements to the existing schemes

In [1], Li et al. constructed a multi-authority access control scheme that offers robustness and verifiability. In this scheme, multiple authorities can jointly maintain the attribute universe, and no single authority can completely control a particular attribute.

In [2], Yang et al. proposed a privacy-preserving access control system in which all attributes are hidden in the access policy.

In [3], Xue et al. proposed a system that allows access control for both data owners and the cloud. The system allowed servers to authenticate users when they request downloads, thus preventing resource consumption attacks by malicious users.

In [4], Belguith et al. proposed an attribute-based fine-grained access control framework, thus realizing two layers of access control: one for fine-grained access control and one for anonymous data access.

In [5], Li et al. proposed an access control scheme against user collusion. This scheme uses the idea of the attribute group to realize user attribute revocation and outsource the operation of attribute revocation to a third party.

In [6], Wang et al. proposed a distributed access control scheme that supports outsourced decryption. The pseudonym technique can be used to realize user anonymity, but a pseudonym used for a long time cannot ensure user privacy.

In [7], Deng et al. proposed an attribute-based proxy re-encryption scheme, enabling a user with decryption ability to share ABE-encrypted data with another user of an identity-based encryption scheme.

In [8], Zhang et al. proposed a multi-authority patient health record-sharing system that supports fine-grained access control. This system requires users and cloud servers to perform

lightweight anonymous authentication, thus ensuring the integrity of outsourced data and protecting user privacy simultaneously.

III. PROPOSED SYSTEM

We propose an efficient and fine-gained big data access control scheme with privacy-preserving policy, where the whole attributes are hidden in the access policy rather than only the values of the attributes. 2) We also design a novel Attribute Bloom Filter to evaluate whether an attribute is in the access policy and locate the exact position in the access policy if it is in the access policy. 3) We further give the security proof and performance evaluation of our proposed scheme, which demonstrate that our scheme can preserve the privacy from any LSSS access policy without employing much overhead.

In this paper, we aim to hide the whole attribute instead of only partially hiding the attribute values. Moreover, we do not restrict our method to some specific access structures. The basic idea is to express the access policy in LSSS access structure. Without the attribute matching function r, it is necessary to design an attribute localization algorithm to evaluate whether an attribute is in the access policy and if so find the correct position in the access policy.

To this end, we further build a novel Attribute Bloom Filter to locate the attributes to the anonymous access policy, which can save a lot of storage overhead and computation cost especially for large attribute universe.

SYSTEM ARCHITECTURE

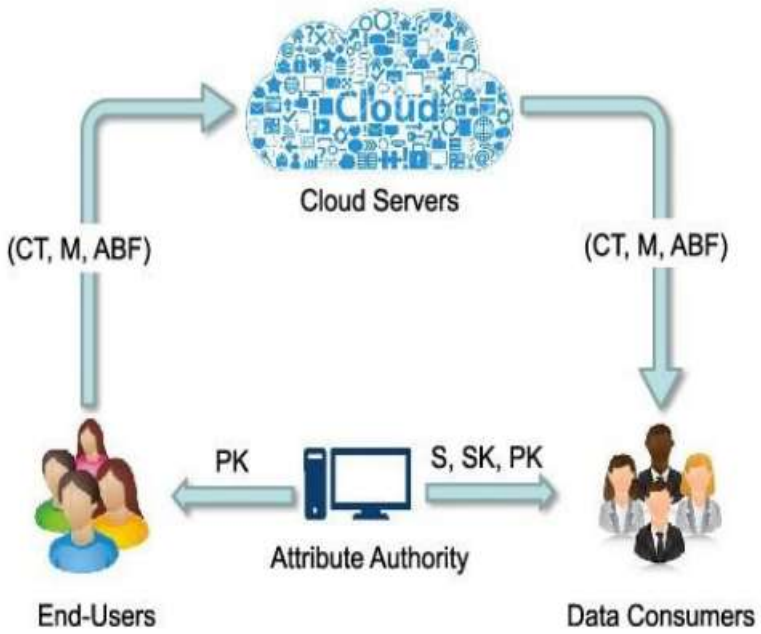

Fig.1 System architecture

Implementing an efficient and fine-grained big data access control scheme with privacypreserving policy requires careful consideration of various components and techniques. In this expanded explanation, we will provide a detailed overview of the implementation process for such a scheme.

• At the core of the implementation, there needs to be a robust access control mechanism that can handle the scale and complexity of big data. One possible approach is to adopt attribute-based access control (ABAC), which allows access decisions to be based on various attributes associated with the data, users, and the environment. ABAC provides fine-grained control by considering attributes such as user roles, data sensitivity, time of access, and location.

• To implement ABAC, a policy management component is required. This component enables administrators to define and manage access control policies that govern data access. The policies can be expressed using a policy language or rule-based system, allowing administrators to define complex access rules and conditions. These policies specify who can access what data under what circumstances, ensuring that fine-grained control is enforced.

• To preserve privacy while enforcing access control, encryption techniques play a crucial role. The data should be encrypted at rest and during transmission. Depending on the requirements and security level needed, different encryption methods such as symmetric encryption, asymmetric encryption, or homomorphic encryption can be employed.

• Symmetric encryption involves using the same key for both encryption and decryption and is typically faster and more efficient. Asymmetric encryption, on the other hand, uses different keys for encryption and decryption, providing stronger security but at a higher computational cost. Homomorphic encryption allows computations to be performed on encrypted data without decrypting it, enabling privacy-preserving data analysis.

• In the implementation, a secure key management system should be established to securely generate, store, and distribute encryption keys. Key management protocols and technologies such as Key Management Interoperability Protocol (KMIP) or Hardware Security Modules (HSMs) can be utilized to ensure the confidentiality and integrity of encryption keys.

• To support fine-grained control, the scheme should provide dynamic policy enforcement. This can be achieved through the use of policy engines that evaluate access requests in real-

time based on the defined access policies. These policy engines should be scalable and capable of efficiently processing large volumes of access requests.

• To enable dynamic policy enforcement, user attributes and contextual information need to be considered. This can be achieved through user attribute retrieval mechanisms that gather information such as user roles, data sensitivity levels, time of access, and location. This information is then used by the policy engine to evaluate access requests and determine if the requested access should be granted or denied.

• For privacy-preserving policy enforcement, techniques like secure multi-party computation (SMPC) can be employed. SMPC allows multiple parties to perform computations on encrypted data without revealing the actual data to any individual party. This enables collaborative analysis while preserving data privacy

Authorization mechanisms can leverage the access control policies defined in the scheme to determine whether a user is granted access to specific data based on their attributes and permissions. This can involve evaluating the user's attributes against the policy conditions and making access control decisions accordingly.

• To ensure accountability and auditability, the scheme should include logging and monitoring capabilities. Access logs can be maintained to track who accessed which data and when. Additionally, advanced technique.

IV. RESULTS

Fig.2 Home page

Fig.3 Data consumer login

Fig.4 Data consumer home

Fig.5 view files

Fig.6 Download file

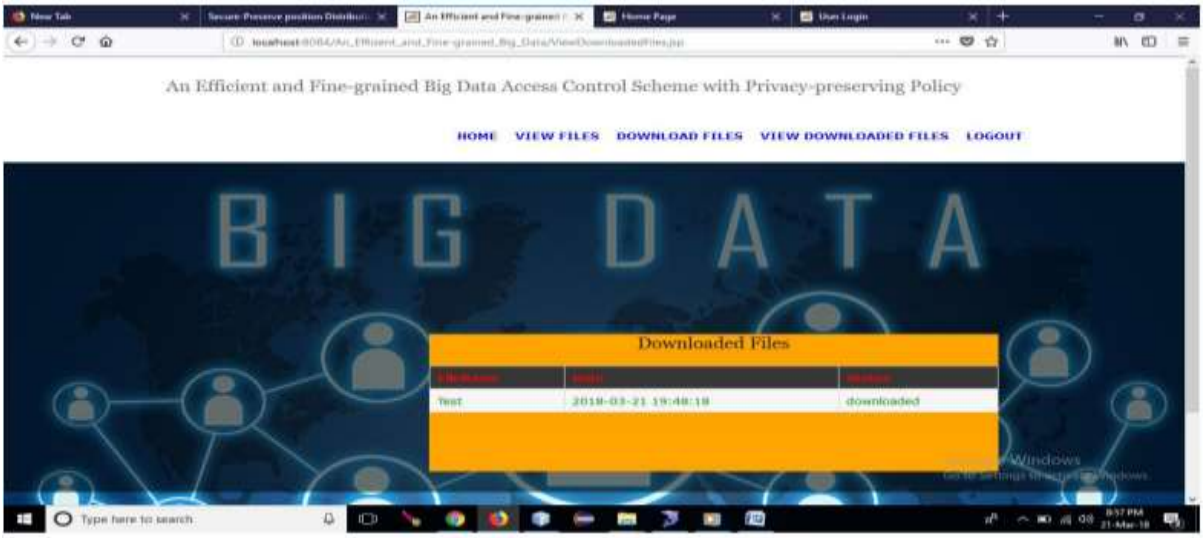

Fig.7 All downloaded file(consumer)

Fig.8 File upload by data owner

Fig.9 Uploaded files view

Fig.10 Cloud server page

Fig.11 All consumers downloaded files

V. CONCLUSION

In this paper, we have proposed an efficient and fine-grained data access control scheme for big data, where the access policy will not leak any privacy information. Different from the existing methods which only partially hide the attribute values in the access policies, our method can hide the whole attribute (rather than only its values) in the access policies. However, this may lead to great challenges and difficulties for legal data consumers to decrypt data. To cope with this problem, we have also designed an attribute localization algorithm to evaluate whether an attribute is in the access policy. In order to improve the efficiency, a novel Attribute Bloom Filter has been designed to locate the precise row numbers of attributes in the access matrix. We have also demonstrated that our scheme is selectively secure against chosen plaintext attacks. Moreover, we have implemented the ABF by using MurmurHash and the access control scheme to show that our scheme can preserve the privacy from any LSSS access policy without employing much overhead. In our future work, we will focus on how to deal with the offline attribute guessing attack that check the guessing “attribute strings” by continually querying the ABF.

REFERENCES

[1] P. Mell and T. Grance, “The NIST definition of cloud computing,” [Recommendations of the National Institute of Standards and Technology-Special Publication 800-145], 2011.

[2] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, “Toward efficient andprivacy-preserving computing in big data era,” IEEE Network, vol. 28,no. 4, pp. 46–50, 2014.

[3] K. Yang and X. Jia, "Expressive, efficient, and revocable data accesscontrol for multiauthority cloud storage," IEEE Trans. Parallel Distrib.Syst., vol. 25, no. 7, pp. 1735–1744, July 2014.

[4] H. Li, D. Liu, K. Alharbi, S. Zhang, and X. Lin, "Enabling fine-grainedaccess control with efficient attribute revocation and policy updatingin smart grid," KSII Transactions on Internet and Information Systems(TIIS), vol. 9, no. 4, pp. 1404–1423, 2015.

[5] K. Yang, Z. Liu, X. Jia, and X. S. Shen, "Time-domain attribute-basedaccess control for cloud-based video content sharing: A cryptographicapproach," IEEE Trans. on Multimedia (to appear), February 2016.

[6] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive,efficient, and provably secure realization," in Proc. of PKC'11. Berlin,Heidelberg: Springer-Verlag, 2011, pp. 53–70.

[7] H. Lin, Z. Cao, X. Liang, and J. Shao, "Secure threshold multi authorityattribute based encryption without a central authority," in Proc. ofINDOCRYPT'08. Springer, 2008, pp. 426–436.

[8] T. Nishide, K. Yoneyama, and K. Ohta, "Attribute-based encryptionwith partially hidden encryptor-specified access structures," in Appliedcryptography and network security. Springer, 2008, pp. 111–129.

[9] J. Li, K. Ren, B. Zhu, and Z. Wan, "Privacy-aware attribute-basedencryption with user accountability," in Information Security. Springer,2009, pp. 347–362