EFFICIENT RETRIEVAL OVER DOCUMENTS ENCRYPTED BY ATTRIBUTES IN CLOUD COMPUTING

V.Savithri[1], P.Pushpa[2]

[1]Assistant Professor,Dept. of MCA,Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur, AP, India.

[2]PG Scholar,Dept. of MCA,Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur, AP, India.

ABSTRACT:

secure data storage and retrieval is the best research directions in cloud. Though lots of searchable encryption scheme have been proposed some of them support efficient retrieval over the documents. Which are encrypted based on their attributes. In this paper a hierarchical attribute based encryption scheme is designed for a data collection. A set of documents is encrypted together if they share an integrated access structure. Compared with the CP policy attribute based encryption schemes, both the cipher text storage space and time costs of encryption and decryption are saved. Then, an index structure named attribute based retrieval features tree is constructed for the document collection based on the TF-IDF model and the documents attributes. A depth first search algorithm for the attribute based retrieval features tree is designed to better the search efficiency which can be further improved by parallel computing. Except for the documents collections in our scheme can be applied to other data sets by modifying the attribute based retrieval features tree slightly. A thorough analysis and series of experiments performed to illustrate the security and efficiency of the proposed scheme.

1. INTRODUCTION

Lots of people and organizations are motivated to outsource their local document management systems to the cloud which is a promising information technique to process the explosive expanding of data. Cloud computing can collect and reorganize a huge amount of IT resources and evidently, the cloud servers can provide more secure, flexible, various, economic and customize services compared with the local management

systems. For all the advantages of cloud services, leaking the sensitive information, such as personal information, company financial data and government documents to the public is a big threat to the data owners. In addition to make full use of the documents on the cloud the data users has to access them flexibly and efficiently. Consequently, a big challenge of outsourcing the data to the cloud is how to protect the confidentiality of the data properly while maintaining their search ability.

An instinctual approach is encrypt the data first and then outsourcing the encrypted data to the cloud. A large number of searchable data encryption scheme have been proposed in the literatures including single keyword Boolean search scheme single keyword ranked search schemes and multi keyword Boolean search schemes. However, all these schemes cannot support effective, flexible and efficient data search because of their simple functionalities, Privacy- preserving multi-keyword ranked document search schemes are more promising and Practical. However, all the data in these scheme are organized by a coarse grained access control mechanism that is each permitted data user can organizations (e.g. the universities, school) at present and this can't satisfy the data owners and users in the future.

In this paper, a new circumstance is considered. A data user may be want to access part of the library (e.g. computers and data related papers etc.) and intuitively she wants to pay less money compare with the data users who want to access the whole library. In different words, in the data collection, each document can be accessed only by a set of specific data users. In this case, we need to design a fine grained access control mechanism for the data and it is more reasonable compared with the current method.

To make the data users able to access part of IEEE Explore Digital Library on demands, a possible approach is encrypting the documents through attribute-based encryption (ABE) schemes before outsourcing them to the cloud. Meanwhile, the permitted data users are assigned with a set of attributes. A data user can decrypt file if and only if her attributes match the files attributes. Recently, cipher text- policy attribute-based encryption (CP-ABE) is a hot research area and it can provide fine-grained, one to many and flexible access control. In these scheme each document is encrypted individual and their encryption efficiency can be better by employing hierarchical attribute based encryption schemes. However, these scheme can't be employed directly to solve our problem properly. First, existing schemes focus on encrypting a single access tree.

However, it is impossible that all the documents in IEEE Explore Digital Library share a single access tree and how to construct a set of optimized retrieve trees for the document collection is a big challenge. Second, in most existing schemes, when the documents are mapped to a set of shared retrieve trees, the data users need to store a huge number of secret keys which will be study in Section IV.B. Apparently, this is a heavy burden for the data users especially for an extremely large document collection and how to decrease the amount of secret keys for the data users is another challenge. Except for access control, document search efficiency is also a challenge for a large document collection. To our knowledge, most existing schemes can't support time-efficient retrieval over the documents which are organized under attribute-based access control mechanism.

To support the previously discussed service, we first design an algorithm to generate hierarchical retrieve trees for the document collection. The proposed algorithm take on the greedy strategy to build the access trees incrementally and each access tree grow by continuously splitting the nodes in the tree. Then we design a cipher text policy attribute based hierarchical document collection encryption scheme called CP-ABHE. In the suggested scheme, a set of documents can share a same integrated access tree and be encrypted together rather than being encrypted individually. In this way, both the cipher text storage space and time costs of the encryption/decryption are saved. The security of the proposed scheme is proved theoretically, and its capability is also evaluated by simulation.

To support exact and efficient document search over the encrypted documents, a complicated index structure is then constructed for the document collection. We first map the documents to document vectors based on the TF-IDF model and in addition, the attributes of the documents are also taken into thought. The similar function between the document vectors is thoroughly design and the vector are organize based on their relative similarity in the attribute based retrieval features tree. Specifically, the similar vectors compose micro cluster which are then, aggregated with each other to generate macro clusters until all the vectors belong to one cluster. The attribute based retrieval features vector of the node in the tree are used to describe the inherent properties of the cluster represented by the node. At last a depth first search algorithm for the attribute based retrieval features tree is designed to both the search efficiency and accuracy.

The main contributions of this paper are summarized as follows:

- A practical hierarchical attribute-based document muster encryption scheme is proposed in which the documents are organized and

controlled based on attributes. The proposed scheme can greatly decrease the storage and computing load.

- We map the documents to vectors in which both the keywords and associated attributes are considered. The ARF tree is proposed to organize the document vectors and support time-efficient document accessible. In addition, a depth-first search algorithm is designed.

- A partial simulation is performed to illustrate the security, efficiency and effectiveness of our scheme. Specifically,

The proposed encryption scheme perform well in both time and storage efficiency. In addition, our scheme also provides efficient and accurate data retrieval method.

2. LITERATURE SURVEY

Our approach is mainly related with two research fields of cloud computing, i.e., ciphertext-policy attribute-based document encryption and encrypted document retrieval. The related work in these two fields is provided in the following. Since Sahai et al. proposed the identity-based encryption (IBE) scheme [19], many ABE schemes have been proposed in which CP-ABE schemes are very promising because of their flexibility and scalability. In these CP-ABE schemes, the documents with different access structures need to be encrypted individually. To improve the encryption/ decryption efficiency and scalability hierarchical attribute- based encryption has been widely researched in which a set of documents may share a common access structure and can be encrypted together. Wang et al. Propose a hierarchical attribute-based encryption scheme named FHCP- ABE and have proved its security theoretically. An advantage of the scheme is that the data users can decrypt all the authorized documents by computing the secret key once. Therefore, both the time costs of encryption and decryption are saved. Wang et al. design a scheme named HABE with the traits of high performance, fine-grained access control, scalability and full delegation. HABE is a combination of hierarchical identity-based encryption and CP-ABE. Wan et al. propose hierarchical attribute-set-based encryption scheme(HASBE) by extending ciphertext-policy attribute-set based encryption (ASBE) with a hierarchical structure of the data users. The HASBE scheme can be seamlessly incorporated with a hierarchical structure of system users by applying a delegation algorithm to ASBE.

Deng et al. Extend ABE to CP-HABE to support hierarchically distributing and delegating the secret keys which can be used in large organizations. Guo et al. propose a resilient-leakage hierarchical attribute-based encryption

scheme to defend against the auxiliary input leakage attack and the security of the scheme is detailedly analyzed. In addition to encrypting the documents, we also attempt to search the encrypted document efficiently and accurately. Consequently, multi-keywords ranked document retrieval over encrypted document collections is also strongly related with our scheme. In [17], Cao et al. first propose a basic privacy-preserving multi-keyword ranked search scheme based on secure kNN algorithm. A set of strict privacy requirements are established and then two schemes are proposed to improve the security and search experience. However, an apparent drawback of this scheme is that the search efficiency is linear with the cardinality of the document collection and consequently, it cannot be used to process extremely large document databases. Xia et al. design a keyword balanced binary (KBB) tree to organize the document vectors and propose a "Greedy Depth-First Search" algorithm to improve the search efficiency. Moreover, the index tree can be updated dynamically with an acceptable communication burden.

However, the document vectors are chaotically organized in the tree and the search efficiency can be further improved. Chen et al. take the relationships of documents into consideration and a hierarchical-clustering-based index structure is designed to improve the search efficiency. In addition, a verification scheme is also integrated into their

scheme to guarantee the correctness of the results. Though the index structure can obtain sub-linear search efficiency, it cannot return the accurate search results. Fu et al. [16] present a personalized multikeyword ranked search scheme in which an interest model of the data users is integrated into the document retrieval system to support personalized search and improve users' search experience. Specifically, the interest mode of a data user is built based on her search history with the help of WordNet [38] in order to depict her behaviors in fine grit level. However, this scheme cannot support dynamic update operations, because the document vectors are constructed based on the statistical information of all the documents in the collection. In addition, though a MDB-tree is employed to improve the search efficiency, the effectiveness of the tree is hard to predict. Li et al. propose a new attribute-based encryption scheme (KSF-OABE) which can implement keyword search function. Though the design goal of KSFOABE is some similar with our scheme, it cannot hierarchically encrypt a document collection and support efficient multi-keyword document retrieval.

3. PROPOSED WORK

In this paper, we attempt to design a fine-grained access control mechanism for the encrypted

documents which also support efficient document search. The search result of a query is defined as the top-k relevant encrypted documents with legal attributes. The process of executing a document query is presented in Fig. 1 and it is mainly composed of five stages:

Stage. 1: The data owner is responsible for collecting and pre-processing the documents, and then obtains a set of high quality files F. He sets the attributes for each document and then hierarchically encrypts the document collection based on attributes. In addition, an index vector is extracted from each document based on the document's content and attributes. An index structure I is constructed based on the index vectors of the documents. At last, both the encrypted documents C and encrypted index structure are sent to the cloud server. The cloud server is responsible for storing the encrypted documents and executing document search based on the index structure.

Stage. 2: When a data user wants to search a set of interested documents, she first needs to register herself as an authorized data user at the certificate authority (CA) center. Then, if possible, several attributes selected from A are assigned to the data user by CA and a corresponding secret key associated with these attributes is sent to the data user. At last, the data user can send a query request Q to the cloud server.

Stage. 3, 4, 5: Once a query is received from a data user, the cloud server first communicates with the CA to check the legality of the data user and her attributes. If the data user is authorized, the cloud server searches the index structure to obtain the search result SR. Then the corresponding encrypted documents are extracted from the encrypted document collection C and sent to the data user. At last, the data user decrypts the documents by her secret key. Note that, the legality checking functionality is optional which can be employed to improve the security level of the whole system. With legality checking, the data users who didn't register themselves in the CA center cannot search the interested documents through the cloud server. However, the security of the system doesn't greatly decrease without this functionality and it can be explained by the fact that the illegal data users cannot decrypt the documents returned by the cloud server because they don't have the secret keys

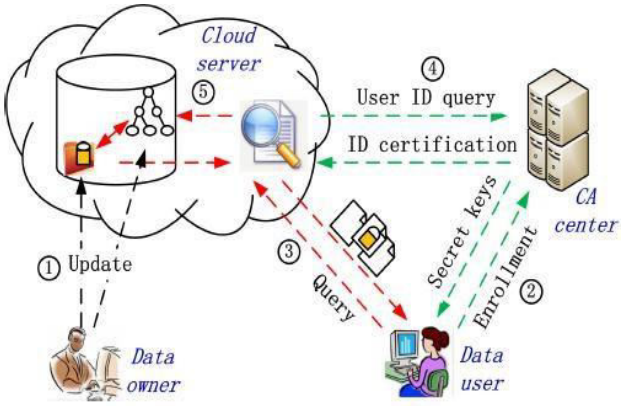

Fig.1. System Overview.

In this paper, we assume that the CA center and the cloud server are trustable. Specifically, the CA center can distribute proper attributes to the data users and the cloud server can execute all the instructions honestly. We further assume that the data users are greedy and attempt to obtain as many plaintext files as possible. The data users try to collude with other users to decrypt the encrypted documents. We mainly restrict our attention to the process of encryption, document search and decryption.

4. CONCLUSIONS

In this paper, we consider a new encrypted document retrieval scenario in which the data owner wants to monitor the documents in fine-grained level. To help this service, we first design a novel classified attribute-based document encryption scheme to encrypt a set of documents jointly that share an integrated access structure. Further, the ARF tree is proposed to organize the document vectors based on their parallels. At last, a depth-first search algorithm is designed to improve the search efficiency for the data users which is extremely important for large document collections. The performance of the approach is completely calculated by both abstract analysis and experiments.

The suggested scheme can be further increased in several aspects: First, in this paper, we assume that each node in the access trees represent an "AND" gate and this limits the springiness of assigning the attributes to the documents. In the future, we will attempt to introduce "OR" gates into the access trees. Second, the access structure of the document collection is generated in a greedy manner and we will check whether it can be further improved to reduce the number of access trees. In addition, the withdrawal method of the data users' attributes needs to be designed. Third, the update strategy of the ARF tree should be proposed. Though the ARF tree naturally supports adding new nodes to the tree, the method of erasing a node from the tree did not provided. Fourth Part, a new document collection, in which each file is associated with a set of proper characteristics, should be developed and a methodical experiment should be conducted on the collection to test the love of issue γ on the approach.

5. REFERENCES

[1] K. Ren, C. Wang, and Q. Wang, "Security challenges for the public cloud," IEEE Internet Computing, vol. 16, pp. 69– 73, Jan. 2012.

[2] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted

data," in Security and Privacy, 2000. SandP 2000. Proceedings. 2000 IEEE

Symposium on, pp. 0–44, 2002.

[3] E. J. Goh, "Secure indexes," Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216., 2003.

[4] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in ACM Conference on Computer and Communications Security, pp. 79–88, 2006.

[5] J. Li, Y. Shi, and Y. Zhang, "Searchable ciphertext- policy attribute-based encryption with revocation in cloud storage," International Journal of Communication Systems, vol. 30, no. 1, 2017.

[6] Y. Miao, J. Ma, X. Liu, X. Li, Q. Jiang, and J. Zhang, "Attributebased keyword search over hierarchical data in cloud computing," IEEE Transactions on Services Computing, vol. PP, no. 99, pp. 1–1, 2017.

[7] A. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. L. Varna, S. He, M. Wu, and D. W. Oard, "Confidentiality- preserving rank-ordered search," in ACM Workshop on Storage Security and Survivability, Storagess 2007, Alexandria, Va, Usa, October, pp. 7–12, 2007.

[8] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 23, pp. 1467–1479, Aug. 2012.

[9] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber +r : topk retrieval from a confidential index," in International Conference on Extending Database Technology: Advances in Database Technology, pp. 439– 449, 2009.

[10] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," Lecture Notes in Computer Science, vol. 3089, pp. 31–45, 2004.

[11] B. Dan and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Theory of Cryptography Conference, pp. 535–554, 2007.

[12] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption," in International Conference on Theory and Applications of Cryptographic Techniques, pp. 62–91, 2010.

[13] Y. Miao, J. Ma, X. Liu, X. Li, Z. Liu, and H. Li, "Practical attributebased multi-keyword search scheme in mobile crowdsourcing," IEEE

Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2017.

[14] Y. Miao, J. Ma, X. Liu, Q. Jiang, J. Zhang, L. Shen, and

Z. Liu, "Vcksm: Verifiable conjunctive keyword search over mobile e-health cloud in shared multi-owner settings," Pervasive and Mobile Computing, vol. 40, pp. 205–219, 2017.

[15] C. Chen, X. Zhu, P. Shen, J. Hu, S. Guo, Z. Tari, and A. Zomaya, "An efficient privacy-preserving ranked keyword search method," IEEE Transactions on Parallel and Distributed Systems, vol. 27, pp. 951–963, Apr. 2016.

[16] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," IEEE Transactions on Parallel and Distributed Systems, vol. 27, pp. 2546–2559, Sep. 2016.

[17] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy- preserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, pp. 222–233, Jan. 2014.

[18] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 27, pp. 340–352, Jan. 2016.

[19] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in International Conference on Theory and Applications of Cryptographic Techniques, pp. 457–473, 2005.

[20] J. Hur and K. N. Dong, "Attribute-based access control with efficient revocation in data outsourcingystems," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 7, pp. 1214–1221, 2010.

AUTHOR'S PROFILE:

Ms.V. SAVITHRI currently she is working as Assistant professor in Audisankara college of Engineering andTechnology (AUTONOMOUS), NH-5,BypassRoad,Gudur, Tirupati (Dt.), Andhra Pradesh, India.

P.PUSHPA is pursuing MCA from Audisankara College of Engineering and

Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.