A RELIABLE RANKED SEARCH METHOD OVER ENCRYPTED CLOUD DATA

SHAIK MAHABOOB BASHA[1], A. EMMANUEL RAJU[2]
[1]PG Scholar, Dept of CSE, Dr. K.V. Subba Reddy Institute of Technology, AP, India,
[2]Assistant Professor, Dept of CSE, Dr. K.V. Subba Reddy Institute of Technology, AP.

ABSTRACT

The internet and the emergence of social networks produce terabytes of data every day. In this huge data scenario, the ability to outsource the data to a cloud storage facility saves the data management and storage facility cost. Some major challenges with this scheme are providing security and ensuring the privacy of the outsourced data. Although data security can be achieved through encryption searching on encrypted data become a complex task the proposed work suggests and efficient searching scheme for encrypted cloud data based on hierarchical clustering of documents. The hierarchical clustering method preserves the semantic relationship between the documents in the encrypted domain to speed up the search process. Consequently the proposed system has linear computational complexity during the search phase in response to an exponential increase in the number of documents. The system also ensures data privacy by providing only limited access of the documents to the different types of users by implementing access control mechanisms resulting in more secured data storage in the cloud. Furthermore, the weights of the keywords are taken into consideration in the ranking when generating the query result.

I. INTRODUCTION

In the Span of big data, huge amount of data is produced world-wide. Enterprises choose to outsource their large amount of data to cloud facility in order to reduce the cost of data management and storage facility spending. As a result of this, data volume in cloud storage facilities is experiencing a dramatic increase. Although cloud server providers claim that their cloud service is armed with strong security measures, security and privacy are major obstacles preventing the broad acceptance of cloud computing service. A traditional approach of reducing leakage in information is data encryption. However, this makes the server-side data utilization, such as searching on encrypted data, a very problematic task. In recent years, researchers have proposed many cipher-text search schemes by incorporating the techniques of cryptography. These methods have been proven with good security, but their methods need massive operations to be performed and also have high time complexity. Therefore, former methods are not suitable for the big data scenario where data volume is huge and applications require online processing of data. In addition, there is concealment in the relationship between documents in the above methods. The relationship between documents represents the properties of the documents and hence maintaining this relationship is necessary to fully express a document. For example, the relationship can be used to express its class. If a document is independent of any other documents except those documents that are related to business, then it is easy for us to assert this document belongs to the category of the business. Due to the blind encryption, this vital property has been concealed in the traditional methods. Therefore, proposing a method which can utilize and maintain this relationship to speed the search phase is desirable. Also, due to failure of software/hardware, and storage corruption, data search results may contain damaged data or may have been distorted by intruder. Therefore, a mechanism should be provided for users to verify the correctness as well as the completeness of search results. In this paper, every document is represented by a vector by using a vector space model, which means every document can be seen as a small point in a space. All the documents can be divided into several categories because of the relationship between different documents. In other words, the points with short distance in the high dimensional space can be classified into a specific category. The search time can be highly reduced by selecting the desired category and rejecting the irrelevant categories. The number of documents which user aims at is very small compared to the number of documents in the dataset. Due to the limited number of the desired documents, a specific category can be further divided into different sub-categories. Instead of using the traditional search methods, a backtracking algorithm is produced to search the targeted documents. Cloud server first searches the categories and gets the minimum desired sub-category. Then the cloud server selects the desired k documents from the minimum desired sub-category. The user decides the value of k and sends it to the cloud server. If current sub-category cannot satisfy the k documents, cloud server traces back to its parent and selects the desired documents from its brother categories. This process executes recursively until the desired k number of documents is either satisfied or the root node is reached. To verify the Integrity of the search result, hash function is used. All documents will be hashed and the hash result will be used to represent the document. The results of documents will be hashed again with the information of category that these documents belong to and the result will represent the current category. Similarly, every category can be represented by the hash result of the current category information as well as the sub-categories information. A virtual root is constructed to represent the data and categories. The virtual root is denoted

by the hash result of the concatenation of all the categories located in the first level. This virtual root will be signed so that it is verifiable. To verify the results of search, user now only needs to verify the virtual root, instead of verifying all the documents. Furthermore, the weights of the keywords are taken into consideration in the ranking when generating the query result.

II. RELATED WORK

Single Keyword Searchable Encryption

The notion of searchable encryption was first introduced by Song. The proposal was to encrypt the words in the document independently. This has a high searching cost due to the word by word scanning of the whole data. Cash et al. recently designed and implemented an efficient data structure. Due to the lack of rank mechanism, users require a lot of time to select the document when large number of documents contain the query keyword. Wang et al. used encrypted invert index to achieve secure ranked keyword search over the documents which were encrypted. In the search phase, the cloud server calculates the relevance score between documents and the query. In this way, related documents are ranked according to their score (relevance) and users can get top k relevant results. Boneh et al designed a searchable encryption construction, first of its type, where anyone can use public key to write to the data stored on server but private key is provided only to the authorized users and only these users can search. However, these methods mentioned above only support single keyword search.

Multiple Keywords Searchable Encryption

To enhance search predicates, different conjunctive keyword search methods have been proposed. These methods have a large overhead. Pang et al. proposed a secure search technique based on vector space model. The efficiency and security of this technique is inefficient due to the lack of the security analysis for practical search performance. Cao et al. presented a novel method to solve the issue of multi-keyword ranked search over encrypted cloud data. But the drawback being that the search time of this technique grows exponentially accompanying with the exponential increase in the size of the document collections. Sun et al. gave a new architecture which achieves better search efficiency. However, the relevance between documents is ignored. As a result, expectations of the user cannot be fulfilled well. For example: given a query containing Cell and Phone, only the documents containing both these keywords will be retrieved by traditional methods. But by taking the semantic relationship between the documents into consideration, the documents containing Mobile and Phone should also be retrieved. As a result, the second result is better at meeting the expectations of the user.

III. PROBLEM STATEMENT

A traditional way to reduce information leakage is data encryption. However, this will make server-side data utilization, such as searching on encrypted data become a very challenging task. In the recent years, researchers have proposed many cipher text search schemes by incorporating the cryptography techniques. These methods have been proven with provable security, but their methods need massive operations and have high time complexity. Therefore, former methods are not suitable for the big data scenario where data volume is very big and applications require online data processing. In addition, the relationship between documents is concealed in the above methods. The relationship between documents represents the properties of the documents and hence maintaining the relationship is vital to fully express a document. If user are storing data in the cloud after they want any file it takes more time because they are not assign index. To overcome the security and searching problem on data a new scheme is needed.

IV. IMPLEMENTATION

We present the *MRSE-HCI* scheme. The vector space model adopted by the *MRSE-HCI* scheme is same as the *MRSE*, while the process of building index is totally different. The hierarchical index structure is used instead of the sequence index into the *MRSE-HCI*. In *MRSE-HCI*, all documents are indexed by a vector. All dimensions of the vector stand for a keyword and the value represents whether the keyword appears in the document or not. Similarly, the query is also represented by a vector. In the search phase, cloud server calculates the relevance score between the query and documents by computing the inner product of the query vector and document vectors and returns the target documents to user according to the top relevance score. Due to the fact that all the documents outsourced to the cloud server are encrypted, the semantic relationship is lost between plain documents over the encrypted documents. In order to maintain this relationship between plain documents over the encrypted documents, a clustering method is used by clustering the related index vectors of documents. Every document vector is viewed as a point in the high n-dimensional space. With the length of vectors being normalized, we know that the distance of cluster points in the n-dimensional space reflect the relevance of corresponding documents. In Other words, points of high relevant documents are very close to each other in the high n-dimensional space. As a result, we can cluster the documents based on the distance measure. As the volume of data in the data centers have experienced a dramatic growth, conventional sequence search approach will be very inefficient. To promote the search efficiency, a hierarchical clustering method is proposed. The proposed method clusters the documents based on the minimum relevance threshold at different stages, and then partitions the resulting clusters into sub-clusters until the constraints on the max size of cluster are reached. After receiving a legal request, cloud server will only search the related indexes layer by layer instead of scanning all indexes. Furthermore, the weights of the keywords are taken into consideration in the ranking when generating the query result.

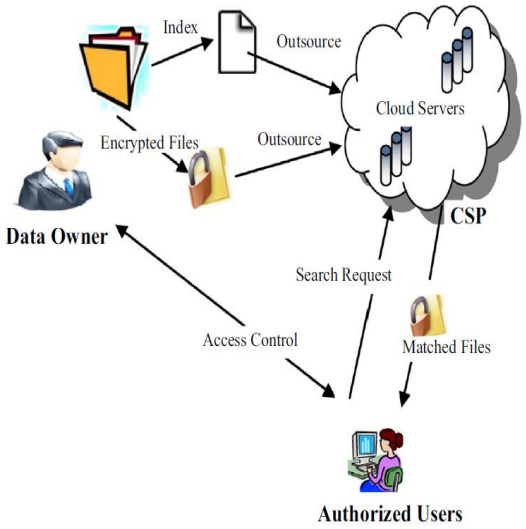

Figure: Architecture

V. MODULES DESCRIPTON

- Data Owner
- Cloud Service Provider
- Authorized User

DATA OWNER:

In this module, the data provider uploads their encrypted data in the Cloud server. For the security purpose the data owner encrypts the data file and then store in the server. The Data owner can have capable of manipulating the encrypted data file and performs the following operations Browse and encrypt and Uploads files, Grant Permission to cloud consumer / End user

CLOUD SERVICE PROVIDER:

The Cloud server manages which is to provide data storage service for the Data Owners. Data owners encrypt their data files and store them in the Server for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the Server and then Server will decrypt them. The server will generate the aggregate key if the end user requests for file authorization to access and performs the following operations such as View all User Files, Give privileges to user, View Search Transaction, View all attackers, View all End Users, View all Data Owners, Create Index on searched data and provide all related data related to corresponding keyword, View all android users.

AUTHORIZED USER:

In this module, the user can only access the data file with the secret key. The user can search the file for a specified keyword. The data which matches for a particular keyword will be indexed in the cloud server and then response to the end user.

VI. CONCLUSION

The problem of searching and securely accessing the encrypted data in the cloud is analyzed. It is understood that maintaining the semantic relationship between the documents reduce the search time for a document. The proposed work is based on multi keyword ranked search over encrypted data. The use of hierarchical clustering method to cluster the documents preserves the semantic relationship between the documents. The experimental results prove that the proposed system has a linear growth in time complexity when the size of the documents collection increased exponentially. It also implements a dedicated module named cloud manager to ensure the privacy of cloud data by granting only limited access to the documents collection to different classes of users. Furthermore, the weights of the keywords are taken into consideration in the ranking when generating the query result.

VII. REFERENCES

[1] S. Grzonkowski, P. M. Corcoran, and T. Coughlin, ”Security analysis of authentication protocols for next-generation mobile and CE cloud services,” in Proc. ICCE, Berlin, Germany, 2011, pp. 83-87.
[2] D. X. D. Song, D. Wagner, and A. Perrig, ”Practical techniques for searches on encrypted data,” in Proc. S & P, BERKELEY,CA, 2000, pp. 44-55.
[3] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano,”Public key encryption with keyword search,” in Proc. EUROCRYPT,Interlaken, SWITZERLAND, 2004, pp. 506-522.
[4] Y. C. Chang, and M. Mitzenmacher, ”Privacy preserving keyword searches on remote encrypted data,” in Proc. ACNS, Columbia Univ, New York, NY, 2005, pp. 442-455.
[5] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, ”Searchable symmetric encryption: improved definitions and efficient constructions,” in Proc. ACM CCS, Alexandria, Virginia, USA,2006, pp. 79-88.
[6] M. Bellare, A. Boldyreva, and A. O’Neill,”Deterministic and efficiently searchable encryption,” in Proc. CRYPTO, Santa Barbara,CA, 2007, pp. 535-552.
[7] D. Boneh, and B. Waters, ”Conjunctive, subset, and range queries on encrypted data,” in Proc. TCC, Amsterdam, NETHERLANDS, 2007, pp. 535-554.
[8] D. X. D. Song, D. Wagner, and A. Perrig, ”Practical techniques for searches on encrypted data,” in Proc. S & P 2000, BERKELEY,CA, 2000, pp. 44-55.
[9] E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol.2003, pp. 216. 2003.
[10] C. Wang, N. Cao, K. Ren, and W. J. Lou, Enabling Secure andEfficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467-1479, Aug. 2012.
[11] A. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, ”Confidentiality-Preserving Rank-Ordered Search,” in Proc. ACM StorageSS, Alexandria, VA, 2007, pp. 7-12.

[12] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, ”Zerber+R: top-k retrieval from a confidential index,” in Proc. EDBT, Saint Petersburg, Russia, 2009, pp. 439-449.
[13] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, ”Secure Ranked Keyword Search over Encrypted Cloud Data,” in Proc. ICDCS, Genova, ITALY, 2010.
[14] P. Golle, J. Staddon, and B. Waters, ”Secure conjunctive keyword search over encrypted data,” in Proc. ACNS, Yellow Mt,China, 2004, pp. 31-45.
[15] L. Ballard, S. Kamara, and F. Monrose,”Achieving efficient conjunctive keyword searches over encrypted data,” in Proc.ICICS, Beijing, China, 2005, pp. 414-426.
[16] R. Brinkman, Searching in encrypted data: University of Twente, 2007.
[17] Y. H. Hwang, and P. J. Lee, ”Public key encryption with conjunctive keyword search and its extension to a multi-user system,” in Proc. Pairing, Tokyo, JAPAN, 2007, pp. 2-22.
[18] H. Pang, J. Shen, and R. Krishnan, Privacy-Preserving Similarity-Based Text Retrieval, ACM Trans. Internet. Technol., vol. 10, no. 1, pp. 39, Feb. 2010.
[19] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, ”Privacy- Preserving Multi-keyword Ranked Search over Encrypted Cloud Data,” in Proc. IEEE INFOCOM, Shanghai, China, 2011,
pp. 829-837.
[20] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, ”Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,” in Proc. ASIACCS,Hangzhou, China, 2013, pp. 71-82.