ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Efficient &Secure Searching, Ranking and Documents Integrity Checking over Encrypted Cloud Data Storage

Rupali Nehete, Prof.Y.B.Gurav

M. E Student, Department of Computer Engineering, TSSM's Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

Professor, Department of Computer Engineering, TSSM's Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

ABSTRACT: From last few years many of the users are using cloud computing and it has been a very popular in users. Cloud computing is a way to provide a computing service over the network. Cloud services give a user as well as to organization to utilize the resources such as software's or hardware which are managed by third party at somewhere else remotely. Due to this the productivity is increased but at the same time while storing the delegate data somewhere else will rise the matters like security and privacy of the data for outliers. In proposed system TPA is used to check the integrity of the data which is stored on the cloud. KASE system is developed to take care of the security issues. This system provides concept of key aggregation and develops a tree structure depending on user query. This can give an easy display of the result to the user. Aggregate key can decrypt the file at the particular node. We have also modified the way of outsourcing by encrypting the file before outsourcing and in same manner they can be searched on the cloud when user makes request. Trapdoor generation is used for searching the encrypted files on the cloud. The proposed system also provides the ranking of results generated at cloud server with similarity. These encrypted ranked results are decrypted at user side, to provide the security. Experimental results prove that the system provide efficient and accurate search over encrypted data and reduced the overhead for auditing of documents.

KEYWORDS:Multi cloud, TPA, Public auditing, Data integrity, Encryption, secure hash.

I. INTRODUCTION

Cloud computing is one of the fast growing technology in every aspects. It is known by several names such as utility computing, grid computing etc. The concept of cloud computing is concentrated on one basic goal which is providing services over the internet, on the user request, for remote location regardless of the personal device or even on a server of an specific organization. The main idea behind the use of cloud computing is to make computing flexible regardless of devices and location. In short the location of information storage and the location of processing of the information are not important any more. By which it is possible to retrieve and processing anywhere and anytime form any device as long as internet is available. The cloud concept additionally implies that, for people and associations, computing will progressively be seen as resource which does not have any limits. It provides various facilities such as network bandwidth, storage, resources etc. on demand. But while using cloud computing, one must take care of data security. For example, suppose, some data is stored of particular organization, then only people belongs to that same organization can access those information. But there is one more condition is that, only classified information is access by users. There are several methods available, use to provide security to the data which is stored on cloud, remotely. But for sharing the part of data on cloud is not sorted yet. At that time role of data owner takes place which is responsible for providing the access permission to other users. One way to provide the security for shared data is to encrypt it, while storing by data owner itself, so that the information will be hidden from outliers as well as cloud provider. But during the data sending process, the data can be altered by hackers. So data owner must conform that data stored on cloud must be in original form and not altered. For taking care of this issues, our system has implanted a concepts called Third Party Auditor (TPA), which is taking care of integrity checking, on the request of client after storing data on the cloud.

Multiple Users of same organization can access the data stored on server in encrypted format, only using the key provided by data owner. In previous system, multiple key are generated by data owner for all user to share the documents. But such task leads to management of large number of keys with key overhead. Instead of this, in our propose system, we are implementing the concept of aggregate key. According to this concept, a common aggregate key is generated to share a group of documents among users. This single key is used by all users. But still there is one more problem regarding documents accessing. As the documents are in encrypted format, users can not get the original information. So to overcome this problem, concept of trapdoor is used. Using this trapdoor, user can search inside encrypted data. Whenever user requests or sends a query to the server for a document cloud server search them using Trapdoor.

To provide the efficient and accurate search, another technique is used by proposed system, which is called as ranking of encrypted results. According to this, when the match results are found, system ranked then according to similarity that is with cosine similarity. Finally this ranked results are provided to users which are decrypted by using symmetric key.

In rest of the paper we will see Related Work related to the topic in section II, Proposed methodology and its architectural view is explained in Section III. Section IV presents implementation details. In Section V we will discuss the Result and discussions, followed by Section VI that presents the Conclusion, at the end we have mentioned various references used to implement our system.

II. RELATED WORK

In some recent work [1], authors propose a new idea of key aggregate searchable encryption (KASE) and implement the idea through a concrete KASE scheme. In this scheme, a single key is distributed to all users by data owner for sharing a big file, and a single trapdoor has to submit by users to the cloud for querying the shared file.

The idea of searchable encryption under multi-user setting is proposed in paper[2]. They focus on the concept of the dynamical user insertion and removing users by keeping up a list of complementary key which is on the public cloud.

In paper[3]-[5], authors gives an idea for MRSE depending on secure inner product computation, and for improve different stringent privacy requirements in two dissimilar threat models authors developed a two improved MRSE methods. In paper [5], author demonstrates how CSP can misdirection the SLA to cut off their cost and turn out to be more aggressive, by hiding from TPAs. In this paper authors presented crowd sourced TPA model to keep CSPs under observation and simultaneously find any deception.

A novel multi-user searchable symmetric encryption technique is developed in paper[6]. This novel technique is combination of single user searchable symmetric encryption scheme and identifier-based encryption algorithm; they have also created a new type of data structure according to their needs. This method also provides secrecy of searched keyword and gets the access controllability of file. It is multi user searchable technique.

Authors found two issues with cloud data those are, intensity of keywords which is sent in queries and data which is retrieved as a result of those query in paper[7]. Each one of them must not be visible. It must be encrypted before we store it on the cloud as a privacy of the document must be maintained. This is done by utilizing symmetric key cryptography algorithm. While picking the file of user interest, multi-keyword query must be generated which then comes as the accruing the top k ranked files. In this paper by analyzing, they have implemented Lucene indexing algorithm. In paper [8], author given a number of schemes for Patient Controlled Encryption, each of which is suitable for diverse setting. SSE construction is proposed in [9]. Also created the scheme and calculated its performance. Test results shows the system is efficient. In paper [10], authors' concentrates on public-key encrypted data and developed a method from multi-keyword searches which supports ranked results. Solution provide in this paper makes use of simple indexing structure also uses influence homomorphic encryption and private information retrieval (PIR) protocols to work on the user queries inside a privacy-preserving manner.

III. PROPOSED SYSTEM

In our system a method of searchable encryption, key aggregation system, third party auditor and lastly raking of search document is introduced by which we are able to improve the performance of system in terms of time and memory. To make sure of integrity of the stored data, auditing of documents are performed by TPA on receiving data owner request.

A. Problem Definition

Propose a novel approach that allows the users to search over encrypted cloud storage data and get the ranking of encrypted results. Also implement the Key Aggregate Searchable encryption (KASE) approach, in which only single aggregate key and single aggregate trapdoor is share among the network entities to make an efficient search over encrypted data and to reduce the key management overhead. A method of searchable encryption, key aggregation system, third party auditor and raking of search document is proposed which improves the performance of system in terms of time and memory. To provide document integrity checking, also propose a third party authentication scheme.

B. System Architecture:

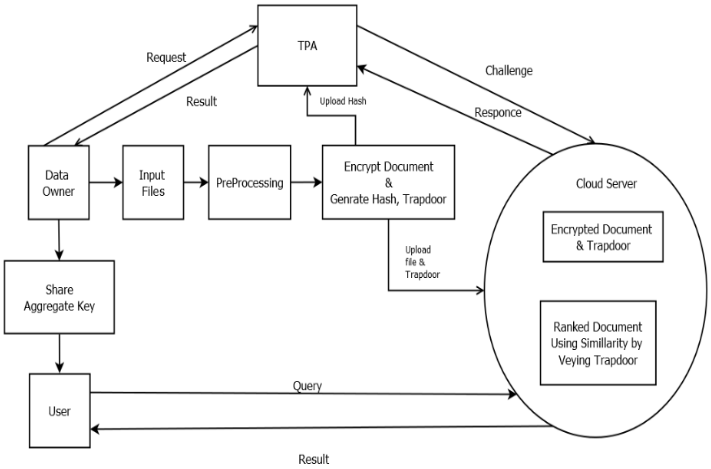

Fig.1. System Architecture

Fig. 1 depicts the architectural view of our proposed system. The system is consisting of four basic entities such as, data owner, users, cloud servers and end users. In this, data owner wants to store their files on cloud server. Before storing the documents, it is preprocessed with stemming and stopwords. Then for each word of all preprocess documents, TF-IDF is generated. Data owner also generates the trapdoor, which is used for users to search over encrypted data. This trapdoor is consisting of aggregated key and keywords. After all this procedure, documents and trapdoor are encrypted using aggregated key and symmetric key and store on cloud servers. Along with this, hash of encrypted document is created and send it to the TPA, which will further used by TPA for checking integrity of documents, store at cloud server.

To allow the access over encrypted documents, for all authorized users, data owner generates and distributes the common aggregate key. Using this key user can get particular information by querying to cloud server. Then cloud server verifies each query user by verifying the aggregate key and trapdoor. The results are generated if and only if user is authenticated. After this, to provide the accurate results and to improve the search efficiency, cloud server makes ranking of all results using cosine similarity. This ranked results are also in encrypted form and send it to users. At user side, these results are decrypted using symmetric key.

IV. MATHEMATICAL MODEL

Algorithm 1: AES Algorithm

1. Key Expansion: - Use the Rijndael's key schedule Round keys which are derived from the cipher key.
2. If Dist_to_tree(u) >Dist_to_tree(DCM) and First-Sending(u) then
3. Initial Round: - Add_Round_Key where every byte of the state is combined with the round key utilizing bitwise XOR.
4. Rounds
   I. Sub_Bytes : non-linear substitution step.
   II. Shift_Rows : transposition step.
   III. Mix_Columns : columns mixing operation
   IV. Add_Round_Key
5. Final Round: It contain SubBytes, ShiftRows and Ad-d_Round_Key

Algorithm 2: ECC Encryption
Elliptic curve cryptography (ECC) is a way to deal with public key cryptography in light of the logarithmic structure of elliptic curves over finite fields.

- Key Generation

Key generation is an important step in which we have to generate both public key and private key pair.
Q = d * P
d be any random number within the range of (1 to n-1).
P is the point on the curve.
Q is the public key.
d is the private key / secrete key.

- Encryption:

Let 'M' be the message that Alice wants to send to the bob. For this Alice has to represent this message on the curve. This has in-depth implementation details. let 'm' has the point 'M' on the curve 'E'.

Select 'k' be any Random of range [1 - (n-1)].

The output of the step is two cipher texts which are CT1 and CT2.
CT1 = k*P
CT2 = M + k*Q
CT1 and CT2 will be sent to bob.

- Decryption:

Bob wants the original 'm' that was send by Alice, Bob performs following steps to get original message 'm'.

M = CT2 - d * CT1

Working Principle

- *Setup Phase*:-In this step we build complete UI framework required for the application which include complete authentication system allowing different entities to sign in to the application.
- *Aggregate key and trapdoor generation Phase*:-In this phase, we allow data owner to generate the hierarchy of cipher text class and to generate aggregate key for the specific level. This step include the trapdoor generation of file that means extracting most important keywords of the file automatically through the TF-IDF concept.
- *Searchable encryption and decryption*:-This phase basically deals with the encryption of files and trapdoors using AES algorithm. This stepl upload the cipher text on cloud storage and enable the end user to download the files using aggregate key.
- *Third party auditor*:-*Third Party Auditor (TPA): A*s shown in Fig.2. TPA performs auditing of large number of files simultaneously and on multiple cloud servers. It checks the data integrity of files.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

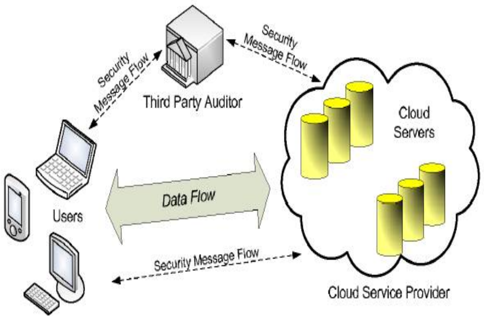

Fig.2.Third Party Auditing

- *Analysis, Testing & Deployment:*-This phase basically deals with the conclusion section and performs testing of the application.

V. RESULTS

A. Experimental Setup

The system is built using Java (Version JDK 8) to evaluate the efficiency, effectiveness. The development tool used is NetBeans (Version 8). The experiments performed on Core2Duo Intel processor, 2GB RAM under Windows XP Professional.

B. Results

Fig. 3 shows the comparison of Existing system and proposed system in the matter of time where we can see clearly the proposed system consumes less time compared with the time taken by existing system. Where as in figure 3 we can see the comparison of the existing and proposed system in memory consumption where it is clearly seen that our system uses less memory compared with existing system.

With introducing TPA in proposed system, we reduced the memory overhead also. Figure 4 shows the comparison between the existing system and proposed system, in existing system the document are retrieved after query search are not rank so the similarity of retrieved document is less as in proposed system the documents are rank , so the similarity of document to query search is more as compare to existing system.

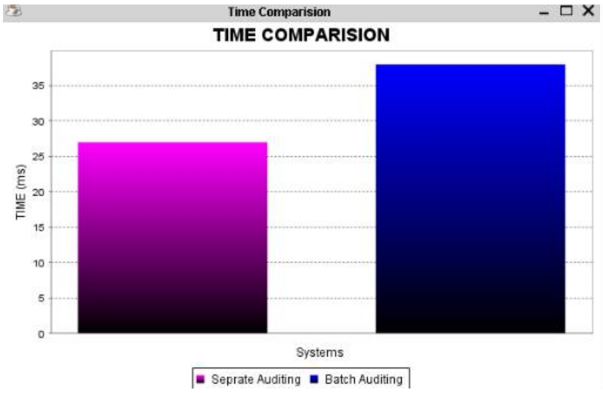

Fig.3. Time Comparison between present and proposed system

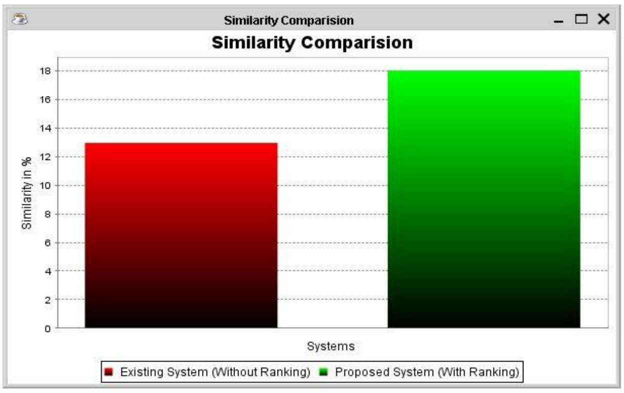

Fig 4. Similarity Graph

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed key aggregate searchable encryption scheme with privacy-preserving public auditing for secure multi cloud storage. This system simultaneously performs the auditing of multiple user requests. To provide the security of files, system encrypts the files using AES algorithm and then store on cloud server. For data integrity checking, TPA is introduced with the concept of SHA1 algorithm. TPA reduces the key management problem on client side. As system uses hash for integrity checking, there is no need to provide the original data file to the TPA for auditing purpose. Because of this the data confidentiality is increased. This will reduce the cost required for personal auditing process conducted by users themselves. Experimental results prove that the system is highly secure and efficient with TPA and reducing the cost of auditing at user side.This system will be enhanced by allowing the uploading of image and multimedia data on cloud server.

REFERENCES

[1] Cui, Baojiang, Zhe Liu, and Lingfeng Wang. "Key-aggregate searchable encryption (KASE) for group data sharing via cloud storage." IEEE TRANSACTIONS ON COMPUTERS VOL: PP NO: 99 YEAR 2015.

[2] Q. Wang, Y. Zhu and X. Luo, "Multi-user Searchable Encryption with Coarser-Grained Access Control without Key Sharing," International Conference on, Wuhan, 2014, pp. 119-125.

[3] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," in IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222-233, Jan. 2014.

[4] S. Rizvi, A. Razaque and K. Cover, "Cloud Data Integrity Using a Designated Public Verifier," IEEE 17th International Conference on, New York, NY, 2015, pp. 1361-1366.

[5] M. Hussain and M. B. Al-Mourad, "Effective Third Party Auditing in Cloud Computing," 28th International Conference on, Victoria, BC, 2014, pp. 91-95.

[6] Z. Yaling, J. Zhipeng and W. Shangping, "A Multi-user Searchable Symmetric Encryption Scheme for Cloud Storage System", 5th International Conference on, Xi'an, 2013, pp. 815-820.

[7] D. D. Rane and V. R. Ghorpade, "Multi-user multi-keyword privacy preserving ranked based search over encrypted cloud data," International Conference on, Pune, 2015, pp. 1-4.

[8] J. Benaloh, M. Chase, E. Horvitz, and K. Lauter, “Patient Controlled Encryption: Ensuring Privacy of Electronic Medical Records”, ACM Workshop on Cloud Computing Security (CCSW 09). ACM, 2009, pp. 103114.[2LIT]

[9] S. Kamara, C. Papamanthou, T. Roeder. “Dynamic searchable symmetric encryption”, ACM conference on Computer and communications security (CCS), ACM, pp. 965-976, 2012.

[10] S. Buyrukbilen and S. Bakiras, "Privacy-Preserving Ranked Search on Public-Key Encrypted Data," IEEE 10th International Conference on, Zhangjiajie, 2013, pp. 165-174.