International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

Ranking Efficient Attribute based keyword searching over encrypted data along with deduplication technique

Dr.Swapnaja.A. Ubale[1], Smita.V. Barkade[2], Dr.Sulbha.S.Apte[3]

[1] *Head of Department of Computer Engineering (II shift), Zeal college of engineering and research,Narhe,Pune*
[2] *Post Graduate student of Department of Computer Engineering, zeal college of engineering and research,Narhe,Pune*
[3] *Department of Computer Engineering, Walchand Institute of Technology, Solapur*

Abstract - *With the advancements in computer technology storage space have evolved day by day. And due to this you can store data anywhere and of whatever size. This kind of storage can be termed as cloud storage. As anything but important information such as personal details, banking details, health records of people, security information will be stored on cloud. Such information if revealed to every one might be misused. So such data should be encrypted and thus will be saved from cybercrimes, theft, and money laundering in online banking. But this encryption won't make the data as searchable. So the proposed system implements how the data can be encrypted and will give access to authorized users. And as the data storage will be huge on cloud, ranking will be used for increase searching efficiency. At the same time to save storage space of cloud deduplication technique is implemented. User revocation makes the system more robust*

***Key Words*: Cloud, ECC, Ranking, Deduplication Attribute based encryption, Attribute based keyword search**

1. INTRODUCTION

Computer science has always made inventions in making human life better and luxurious. So the cloud computing has played a vital role in that. Cloud allows you to store your data with such ease as you are actually storing your data in air cloud. It is as simple as create the account and store data on it. There are many cloud servers such as Amazon, IBM. As anyone can use and store data on cloud creates the security issues for the data. To secure the data, measures can be used such as allowing only authorized users to search the data and encrypt the stored data. This will secure the data in two ways. First it gives access to only authorized people, means the risk of misuse of data will be reduced greatly. And second as the data is not plaintext, many hackers won't be able to get what are the actual contents of the data.

There are many existing schemes on how to do searchable encryption which will make the data available to end user when required. Encryption can be divided on the basis of keys.This falls into two categories symmetric encryption and asymmetric encryption or public key cryptography. In this the major difference is in keys used for encryption and decryption will be same or different.

Assume that Google encrypted all the data and end user is not able to make use of that data. This makes the usability of data as zero. To address this issue we have implemented Attribute based encryption to give only access to authorized people. Elliptic cryptography curve algorithm used of encryption. To increase searching efficiency over the large cloud data ranking is implemented. And to avoid storage of similar data on cloud deduplication is also implemented.
By using attributes of user which will be unique when compared to others such as username, encryption of data will be done. When user will request for data file first check will be done whether this is authorized user or not. If this is authorized user, then check will be done whether this particular user have the access to that data.so this ensures authorization at fine grained level. User revocation plays important role, which stops searching by unauthorized user. Attribute based encryption falls in public key cryptography. And having its own advantages such as duplication of key won't happen in this. Uses different key for encryption and different for decryption.

When the data is stored on cloud first check will be done to check duplication of data. If the new data is not already present on cloud, data will be stored. If it's already present storage won't be done again.
At the end when the result is to be retrieved it will be done with ranking to increase user satisfaction for search. Because the data stored on cloud is huge and searching on it will take time.

Goals of Our proposed system are

1. Attribute based encryption.
2. Providing security to data.
3. Increasing efficiency of the system.
4. Improving search experience
5. Making storage efficient system.
6. User revocation

2. RELATED WORK

Searching over encrypted data can be accomplished by various secure search schemes which are actually based on either secret key cryptography (SKC) or public key cryptography (PKC). Curtmola et al. [1] presented an efficient single keyword encrypted data search scheme with the use of inverted index structure

For the first time ning cao,cong wang define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). Among various multi-keyword semantics, they choose the efficient similarity measure of as many matches as possible, to capture the relevance of documents to the search query.

User authorization should be in place to grant multiple users search access.Hwang and Lee[3] in the public-key setting presented a conjunctive keyword search scheme in multi-user multi-owner scenario. But this scheme is not scalable under dynamic cloud environment.

To present practically-efficient secure search functions over encrypted data Wenhai Sun, and Bing Wang present a verifiable privacy-preserving multi-keyword text search (MTS) scheme with similarity-based ranking to support Searching of more than one keyword and give the ranked results, [4] proposed building of index with Term frequency and vector space model is proposed by Sun,Wang.

3. PROPOSED SYSTEM

Proposed system comprised of variety of roles and modules as shown in below fig 1.

Roles in proposed system are as follows

Trusted Authority: Role played by trusted authority is distribution of keys to owner and user.
And this is assumed that it will be available and keys distribution will have done sole handedly from here.
Trusted authority should be online to serve the key request made at any time.

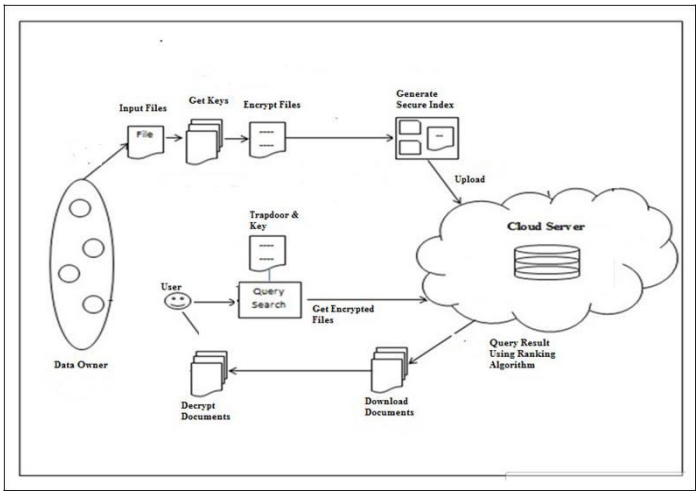

Fig -1: System Architecture of proposed system

Cloud Server: It is a Central system where all the data will be stored. These are semi trusted authority in this system. It provides the space for storage on pay per use basis.

Data Owner: It is the one who owns the data and wants to share that with end user. To avoid same data to be stored on cloud, data checks will be done. Proposed system reduced overhead of data owner from doing key management. And the data owner needs to have to be online all time.

Data User: These are the authorized users around whom the whole proposed system rotates.
They ask for access to particular data in order to make required work done. In short seeker of the information.

Modules in proposed system are as follows

Creation of roles: First of all, creation of data owner and data user will be done. And the list of these will be shared with server. So that cloud will have information on who can access what.

Duplication Check: Once the data owner is created, data upload part will come in picture and this check of redundant data will be done. To implement this technique, we have calculated the hash of every file.

File Upload: This module is heart of proposed system.
File preprocessing such as removal of stop words and stemming is done and then creation of index, encryption of index. Encryption of files and call to duplication check module and finally data uploading will be done

File retrieve: When user wants to have access to some data or file. He/she will generate the encrypted query called as trapdoor to search on encrypted data. And this will be send to cloud if the user is one who have authorization will be

allowed to search and based on access policy the data will be returned to end user.

Decrypt: In this module the end user who already has the keys will decrypt the data. As this is Attribute based encryption, data decryption will be done with the help of private key of user.

Ranking: Ranking is applied on returned result so searching experience is enhanced.

User Revocation: In this module user rights for searching over encrypted data will be revoked

4. ALGORITHM:

There are algorithms for every module. Module will represent the user level understanding of system and algorithm presents the system level implementation of proposed system. We have invented deduplication algorithm. Elliptic cryptography curve (ECC) algorithm is used for encryption. ECC have advantages over other encryption algorithm, key size is small but this small key size doesn't compromise the security of data.

With the use of small key size, it works same efficient way when compared with algorithm of large size.
Ranking algorithm used term frequency and inverse document frequency technique to do ranking on returned result from cloud server.

User revocation algorithm also plays important role in security.

5. EXPERIMENTAL RESULTS:

We have used java for implementing this system. The results show that the implemented system performance in terms of storage and time.
It is proved that the proposed system performed better.

Below chart 1 and chart 2 represents the time and memory comparison of the proposed system with the existing system. Time and storage required in existing system without deduplication is more than the time required for proposed system with deduplication

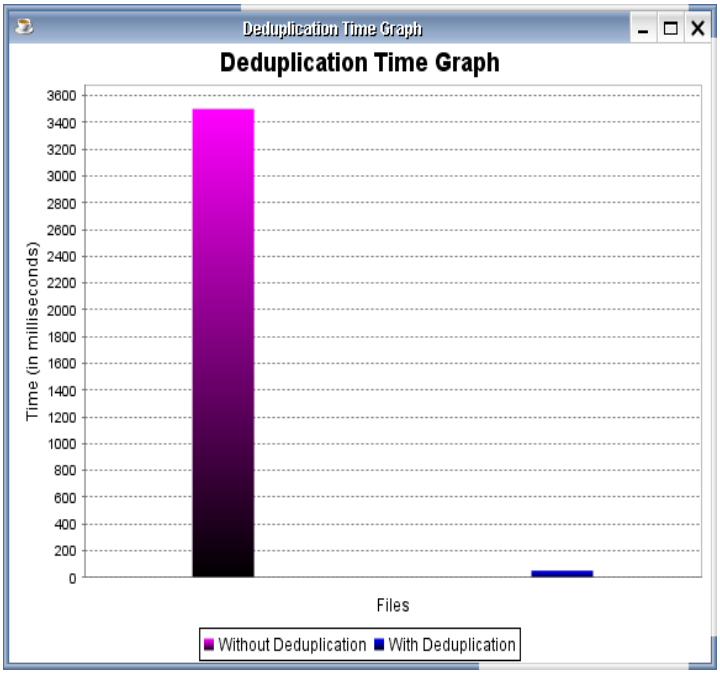

Chart -1: Time Graph

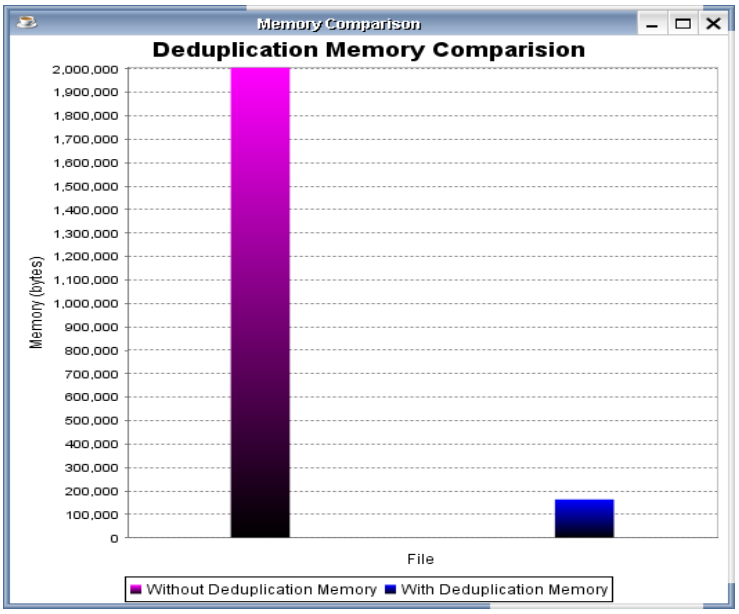

Chart -2: Memory Graph

Time and Memory required for data storage in existing system is large as clearly shown by graphs.

Below chart 3 and chart 4 shows the search result without ranking and with ranking respectively.
Ranking will increase the search satisfaction for the end user.

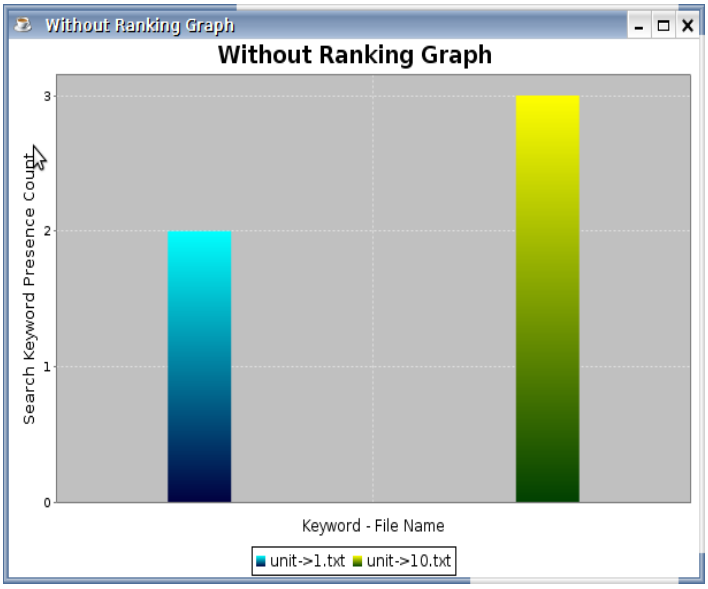

Chart-3: Search without ranking

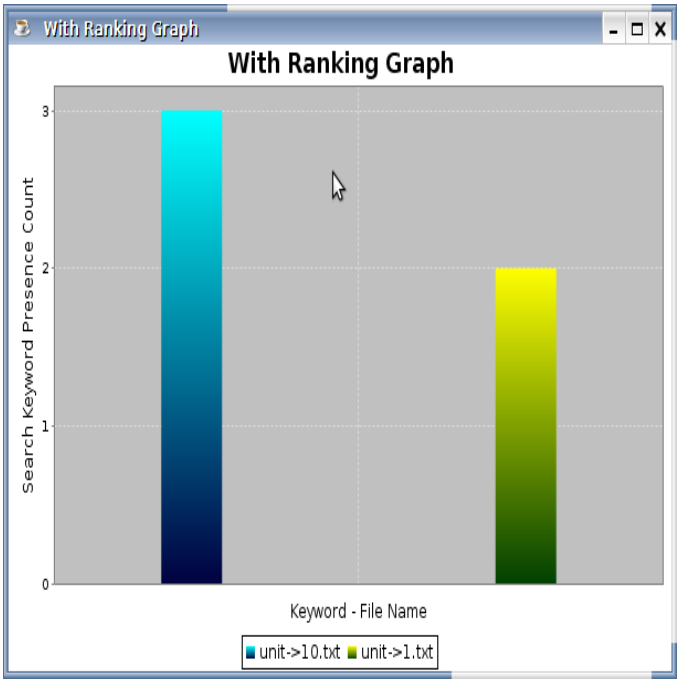

Chart-4: Search with ranking

For showing the result analysis we have used the file as parameter against the time and memory required.
In ranking result the file having more no of occurrence of searched keyword will be returned.

6. CONCLUSIONS

We have introduced a system based on attribute based keyword search and attribute based encryption.
It proves to be efficient and secure based on the implementation details and illustration with real time examples. Ranking helps to give better results. Deduplication technique gives way to efficient storage system. Implementation of User revocation makes proposed system more robust.

REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 79–88.
[2] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE Conf. Comput. Commun., 2011, pp. 829–837.
[3] Y. H. Hwang, and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Proc. 1st Int. Conf. Pairing, 2007, pp. 2–22.
[4] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 11, pp. 3025–3035, Nov. 2014.
[5] Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 89–98.
[6] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Attribute-based keyword search with fine-grained owner enforced search authorization in the cloud," in Proc. IEEE Trans on Parallel Distrib. Syst., vol. 27, no. 4, Apr. 2016.
[7] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable,and fine-grained data access control in cloud computing," in Proc. IEEE Conf. Comput. Commun., 2010, pp. 1–9.
[8] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE Trans. Parallel Distrib. Syst,vol. 24, no. 1, pp. 131–143, Jan. 2013.