Multi keyword Ranked Search on Encrypted Cloud Data with User Revocation

Shivam Saurav[1], Manik Rustagi[2], Rinku Borole[3], Ravi Prasad[4]
Department of Comp. Engg
[1, 2, 3, 4] DYPCOE, Akurdi, Pune

Abstract-*Recently the attractiveness towards the cloud computing is growing tremendously. More clients now a days are opting to outsource data management on cloud because of its low cost in management and greater convenience. For sake of confidentiality, data which have sensitive should be encoded before uploading to the cloud server, which retrieve data using keyword based document retrieval .The system represent a novel technique which perform multikeyword ranked search on data which is encoded in cloud,these method concurrently supports active update actions like insert and delete the data on the cloud. Particularly, these systems joint two models, vector space and TF-IDF for index creation and query generation.These systems build a tree based index structure and intend a Greedy Depth first Search algorithm for providing effective multi-keyword ranked search. Most popular secure kNN algorithms is used for encoding the query and index vectors, and calculate precise relevance score between encoded query and index vectors. In index vector phantom terms are added for modifying search results and avoid statistical attacks. System also implement user revocation for improving the security.*

Keywords-Cloud Computing, User Revocation, Multi Keyword Ranked Search, Index Tree

I. INTRODUCTION

Internet hosting service is specially intended to host user files. The services may be cloud storage service, cyber locker, online file storage provider, file hosting service. By providing a password or other authentication, the same user or perhaps by other users permits to upload files on cloud that can be accessed over the internet from a different smart phone, tablet, computer or other network device.

A type of "network storage" for file access, individual backup, or file distribution meant at private individuals are permits personal file storage services [8]. Numerous remote file storage services are permitting users to share and synchronize all categories of files across all their devices. Users have right to upload their files keep them password-protected and share them publicly or and document-sharing services permits users to collaborate and share document files. To appear to be the same on all computer folders, a user must to generate special folders on each of their mobile devices or computers allowable by file syncing and sharing services and then synchronizes. Shared files are effortlessly available to other users for viewing or collaboration and also available through a mobile apps or website.

The system proposed here works on cloud data which is encrypted and implements the tree-based search. Dynamic operation and multi keyword ranked search are supported on collection of the documents. Characteristically, the vector space model and the majorly used "term frequency (TF) inverse document frequency (IDF) model are mixed in the query generation and index development to offer multi keyword ranked search. Prominent search efficiency is achieved using an index structure which is tree-based suggest a Greedy Depth-first Search algorithm based on the index tree. This system can adaptably achieve sub linear search time and also work with the insertion and deletion of documents because of special structure of tree-based index. The secure KNN algorithm is used to encrypt the query and index vectors [1].Also, guarantee precise relevance score calculation among encrypted query and index vectors. Author generate two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme is generated for the known cipher text model, and the enhanced dynamic multikeyword ranked search (EDMRS) scheme for the known background model to prevent diverse attacks in diverse threat models.

This system involves three different entities: cloud server, owner of data and data user. Owner of data have a collection of document, to be encrypted and stored on cloud. Data users are able to search on this encrypted data. In this system, data owner initially construct the secure searchable index tree and then generate encrypted document. Then this index tree and encrypted documents are store on server. Data owner also distributes the key to authorized uses, require for document decryption. Upon receiving query request for particular file from user, cloud server perform searching on the index tree and the collection of encrypted top k ranked results is provided. Finally data user decrypt the received documents using secret key received from data owner [2], [6].

The user revocation ability to data owners is provided in this system. When any of the user leave the organization,

data owner perform user revocation. System needs user revocation, because user who leave the organization, having secret key which may disturb the privacy of information. Therefore data owner modify the key of all users so that privacy is preserved.

II. LITERATURE SURVEY

In paper [1], author exhibit a secure multikeyword ranked search plan on encrypted data in cloud, which support dynamic upgrade processes such as reports insertion and deletion. The safe kNN algorithm is used to encrypt query and index vectors, and in the interim guarantee exact significance score computation between encoded index and query vectors. In particular, the vector space model and the broadly utilized TF IDF model are joined as a part of query generation and index construction. They build a unique tree-based index structure and intend a ”Greedy Depth first Search” algorithm to give effective multikeyword positioned look. With a specific end goal to oppose statistical attacks, apparition terms are add to the index vector for modifying result.

Author of [2] consider the issue of construction a safe cloud storage service. In the open cloud framework client usually can’t completely trust service provider, To come up with a safe cloud data storage service, a few architectures that join later and non-standard cryptographic primitives keeping in mind the end goal to achieve our objective.

Paper [3] focuses the issue of seeking on data that is encoded utilizing a public key framework. Let client Bob who sends email to client Alice encoded under Alice’s open key. An email portal needs to verify that email contain the catchphrase “urgent” so it could course the email as needs be. Alice, then again does not wish to give the passage the capacity to decode every one of her messages. Authors characterize and develop a mechanism that empowers

Author of [4] demonstrate to make an open key encryption scheme for Alice that permits PIR looking over encoded records. Our answer gives a theoretic answer for an open issue postured by Boneh et al. giving the primary plan that does not uncover any fractional data in regards to client’s search in general public key setting and with non-trivially minor communication complexity. The principle method of our answer likewise takes into consideration Single-Database PIR composing with sub linear communication complexity, by consideration of autonomous hobby.

Author portrays their cryptographic plans for the issue of looking on encoded data and give evidences of security to the subsequent crypto systems in [5]. Our systems have various critical favorable circumstances. They are irrefutably secure: they give irrefutable mystery to encryption, as untrusted server can’t read anything related which relates to plain text, when just known the figure content; for searches they give query isolation, which indicates about untrusted server as it can’t read anything else mostly the plain text rather than query item; they give meticulous looking, so untrusted server can’t scan for a discretionary words without having the client’s permission; they likewise support secreted queries, so client might approach the untrusted server to hunt down a mystery word without uncovering the word to server.

Paper [6] offer answers for this issue under all around characterized security prerequisites. Their plans are effective as in no public key crypto system is included. To be sure, we have encryption free methodology technique decided for the distant files. They are additionally incremental, in that you can upload new documents which are protected beside past queries at the same time searchable against future queries.

Author of [7] start by looking into previous ideas of security and developed novel and high security descriptions. They present two developments that signify secure under their novel definitions. Fascinatingly, despite satisfying high security guaranties, their improvements are more proficient than each single past improvement.

III. PROPOSED SYSTEM

3.1 Problem Statement

Maintainability, accessibility are the functions of cloud, also the main motivation is the privacy attract towards the distinct organization for storing their data on the cloud server. Sensitive data is encrypted before putting on cloud to maintain the confidentiality. To retrieve multikeyword query and ranking result, we use tree based search scheme. One of the boring tasks is to sustain the thousands of records encoded keys of the multiple users and recovering the sensitive document when the client is revoked is one of the challenging tasks for the organization.

3.2 System Architecture

The following figure 1 describes the system which contains three distinct modules such as data owner, user of data and cloud. Owner of data contains the huge collection of documents, which have to store on cloud in the encoded format. Data users are able to search on this encoded data. In this system, data owner first build the secure searchable index tree and then generate encrypted document. Then this index tree and encrypted documents are store on server. Data owner

also distributes the key to authorized uses, require for document decryption. Upon receiving query request for particular file from user, cloud server searching the tree index and proceeds the set of top k ranked encoded results. Finally data user decrypt the received documents using secret key received from data owner.

This work presents the user revocation capability to data owners. When any of the user leave the organization, data owner perform user revocation. System needs user revocation, because user who leave the organization, having secret key which may disturb the privacy of information. Therefore data owner modify the key of all users so that privacy is preserved.

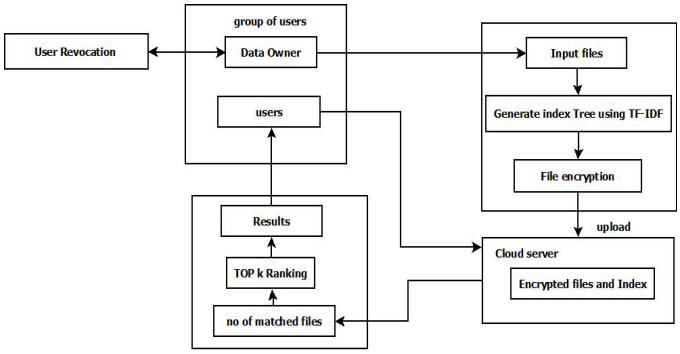

Fig.1 System Architecture

IV. TECHNIQUES USED

4.1 Algorithm

In the existing system for generating the index tree, index tree algorithm is used. Also the BDMRS[1] and EDMRS[1] algorithm is implemented in the existing system. To provide security, it is necessary for client to not to have the capability for accessing the data when they leave the organization. Because of this motive, the owners owner modify the set of attribute also send updated data to the users.

User Revocation:

1: Choose a novel value of Si, Snew ϵ Zq.
2: The novel value of Snew is obtained by changing the entry of vector Vnew.
3: Evaluate λx=Rx.ϑnew, for row Xi correspond to leaf Iu.
4: Evaluate C1,x for Xi.
5: Novel value of C1,x is privately broadcasting to the cloud.
6: C0 = Me(g, g)Snew is evaluated and store it into cloud.
7: Whoever wants to decrypt the data Novel value of C1,x is transferred but not stored with the data.

Where Zq: Cyclic group of prime order q. C is the ciphertext, g is generator of Zq.

4.2 Mathematical Model

Let S, be a system such that,
S = {I, IDF, E, IT, UP, MK, K, R, UR}

1) Input files

I is a set of all input files of data owners,

I= {I1, I2,...., In},

Where, I1, I2,.. are the number of input filesof data owners wants store on cloud.

2) Assign file identifier

IDF is a set of all file identifiers,

IDF= {IDF1, IDF2,...., IDFn },

Where, IDF1, IDF2,.. are the number of fileidentifier assign by system.

3) Index tree generation

IT is a set of all index tree generated bysystem of each encrypted file,

IT= {IT1, IT2, ..., ITn },

where, IT1, IT2, ...are all index tree of documents.

All files are signified by the vector in thevector support model,in which normalized TF values of keyword are element in document.The nodes of the tree are signified by u, indexvector is evaluated in the child node of u.

If the node(u) is a leaf node, then termfrequency is calculated.

4) File encryption

E is a set of all encrypted files store at server,

E= {E1, E2,...., En},

Where, E1, E2,.. are all encrypted store atserver.

5) Upload to server

UP is a set of all files uploaded to server,

UP= {UP1, UP2, ...,UPn},

Where, UP1, UP2,.... are all encrypted filesand index tree uploaded to server.

6) Multi keyword query input

MK is a set of all queries requested by userto system,

MK= {MK1, MK2, ...,MKn},

Where, MK1, MK2,.... are all multi keywordquery requested to server.

7) Top k ranking

K is set of all top k ranking of requestedquery,

K= {k1, K2, ..., Kn},

Where KR1, KR2,... are all ranking files.

8) Result

R is set of all results,

R= {R1, R2, ..., Rn},

Where R1, R2,... are all output files.

9) User Revocation

UR= {ur1, ur2,....., urn}

Where, UR is the set of users who leave theorganization.In this step, data owner modify the key of allusers when any one of them leave organization.Because of this, security of organizationis increases.

V. RESULT ANALYSIS

5.1 Experimental Setup

For implementation the system required the software as: JDK with version 1.8, window platform is used. IDE tool is used, for the development. There is no need of any hardware in this framework.

5.2 Dataset Used

Multiple files are used as a dataset, the file should be 1 KB to 100 KB memory.

5.3 Results

Fig. 2 shows time graph; in above graph X-axis shows number of documents in collection while Y-axis show time required for generating index tree in ms, with increase in number of documents the time required to generate index tree is also increase.

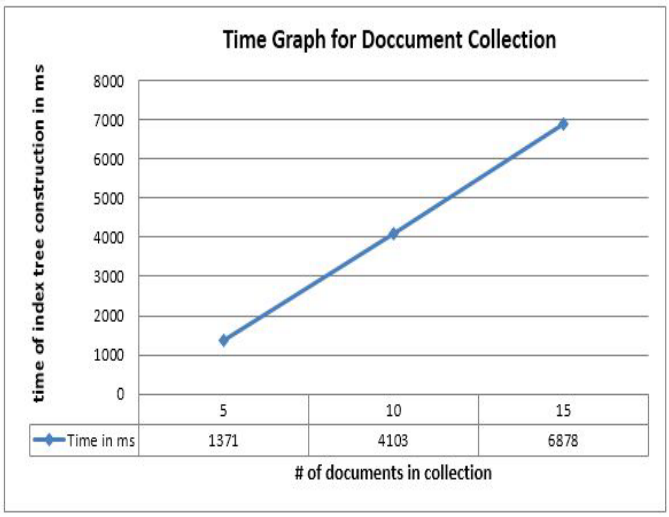

Fig. 2. Time Graph for Document Collection

Fig. 3 shows time graph; in above graph X-axis shows number of keywords in dictionary while Y-axis show time required for generating index tree in ms, with increase in number of keywords the time required to generate index tree is also increase.

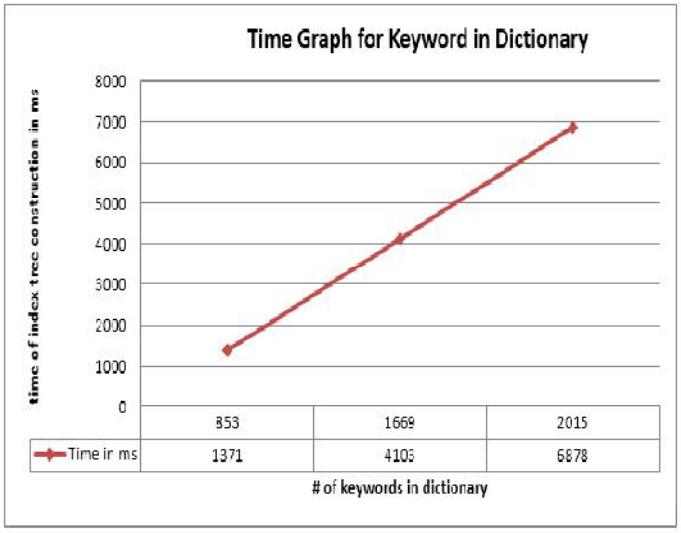

Fig. 3. Time Graph for Keyword in Dictionary

VI. CONCLUSION

This scheme introduced the method which is more protected, proficient and dynamic approaches are used. System provides the accurate multi keyword ranked search, the dynamic insertion, deletion of documents and the user revocation. In user revocation, data owner modify the key of all users when any one of them leave organization. It increases the security of organization. Here we build a index using keyword balanced binary tree, and to gain better efficiency than linear search. System proposes a greedy depth first Search algorithm. To protect the security against two threat model secure kNN algorithm is used. Finally results show the accuracy evaluation of the proposed protocol.

ACKNOWLEDGMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the college powers for giving the obliged base and backing.

REFERENCES

[1] Zhihua Xia, Xinhui Wang, Xingming Sun, “A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data”, IEEE transactions on parallel and distributed systems VOL: PP NO: 99 YEAR 2015.

[2] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in Financial Cryptography and Data Security. Springer, 2010, pp. 136- 149.

[3] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology- Eurocrypt 2004. Springer, 2004, pp. 506-522.

[4] D. Boneh, E. Kushilevitz, R. Ostrovsky, andW. E. Skeith III, "Public key encryption that allows pir queries," in Advances in Cryptology-CRYPTO 2007. Springer, 2007, pp. 50-67.

[5] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44-55.

[6] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proceedings of the Third international conference on Applied Cryptography and Network Security. Springer-Verlag, 2005, pp. 442-455.

[7] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 79-88.

[8] C.Orencik, M. Kantarcioglu, and E.Savas, "A practical and secure multi-keyword search method over encrypted cloud data," in Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on. IEEE, 2013, pp. 390397.