ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

Privacy Saving Multi-Keyword Ranking Search Anonymous ID on the Cloud Data that has been encrypted

Rameshwar Murlidhar Indoriya[1], Pasuladi Santosh[2]

Abstract— **Cloud computing market is growing rapidly in the recent years. It provides on-demand, high scalability computing resources with high availability and reliability. However, security and privacy of user's data is considered as one of the biggest obstacles. Cloud computing provide great economic savings. Actually protecting data privacy, sensitive data must to be encrypted before outsourcing to untrusted cloud. So enabling encrypted cloud data searching is very important. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE).For achieving solution for above problem we choose the "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure. In order to provide more security to the data on the cloud server, the allocation of the anonymous ID to the user is done. To improve the search experience data retrieval services, further expansion of the two methods will be made to support multiple search semantics.**

***Index Terms*—Cloud Computing, Privacy Preserving, MRSE, Anonymous ID.**

I. INTRODUCTION

According to [1], "Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services)".This pool of resources is typically exploited by a pay- per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.

To protect data privacy in the cloud, sensitive data, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before they are getting to outsourcing to the commercial public cloud [2].In [3] Information Retrieval systems ranked documents by their estimation of the usefulness of a document for a user query.

Manuscript received May, 2016.

***Rameshwar Indoriya**, Department Of Computer Science and Engineering, St.Mary's Engineering College, Jawaharlal Nehru Technology University (JNTUH) Hyderabad, India.*

***Pasuladi Santosh**, Department Of Computer Science and Engineering, St.Mary's Engineering College, Jawaharlal Nehru Technology University (JNTUH) Hyderabad, India.*

[13] Proposed a secure and scalable fine-grained data access control scheme for cloud computing. User secret keys are defined to reflect their access structures so that a user is able to decrypt a cipher text if and only if the data file attributes satisfy his access structure. To maintain integrity of outsourced data third party auditor (TPA) [14] play important role.

Ranked search [15] utilizes system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. "Multi-owner" data sharing, where the encrypted data are contributed by multiple owners and can be searched By multiple users so there is scalable framework for Authorized Private Keyword Search (APKS) over encrypted cloud data [16].Attribute-based encryption (ABE) and predicate encryption (PE) for inner products [18]. In a predicate encryption scheme, secret keys are associated with predicates, and cipher texts are associated with attributes. A user should be able to decrypt a cipher text if and only if their private key predicate evaluates to 1 when applied to the cipher text attribute. In this paper we used **"anonymous user ID"** for proving security to users as well owners data with **"co-ordinate matching"** and **"Inner Product Similarity"** [3] as for resolving problem of multi-Keyword Ranked Searching over cloud data that has been encrypted.

The rest of the paper is organized as follows. Section 2 gives the Literature Survey. Then we provide the framework for Multi-Keyword Ranked Search in Section 3, followed by Section 4, which gives the detailed description of our System modules. Section 5 shows the Result of system. Finally, Section 6 gives the concluding remark of the whole paper.

II. LITERATURE SURVEY

Many Authors worked on to ensure privacy and security is for storing as well as retrieving data from untrusted cloud storage.

L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner [1] proposed advanced definition of cloud storage. According to [1] clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). They also focuses on three scenarios of cloud computing as *Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS).*They also proposed, virtualization is the key enabler technology of clouds, as it is the basis for features such as, on demand sharing of resources, security by isolation, etc.

Usability is also an important property of Clouds. Also, security enhancements are needed so that enterprises could rely sensitive data on the Cloud infrastructure.

S. Kamara and K. Lauter[2] focused on types of cloud such as private, public cloud as well they light on, how cryptography used to store data on untrusted public cloud? [2] Also introduced cryptographic storage service architecture such as customer architecture and enterprise architecture. Customer architecture has basically three elements like Data processor, data verifier and token generator. Enterprise architecture has four elements as credential generator, Data processor, data verifier and token generator. Also they explain various advantages of "cryptographic cloud storage" include confidentiality, integrity, Non-repudiation.

A. Singhal[3]have introduced key advances in the field of Information Retrieval(IR).[3] said early IR systems were boolean systems which allowed users to specify their information need using a complex combination of Boolean ANDs, ORs and NOTs. Most IR systems assign a numeric score to every document and rank documents by this score. Several models have been proposed for this process like the three most used models in IR research are the vector space model, the probabilistic models, and the inference network model. In the vector space model text is represented by a vector of *terms*. The definition of a term is not inherent in the model, but terms are typically words and phrases. *Probabilistic* model family of IR models is based on the general principle that documents in a collection should be ranked by decreasing probability of their relevance to a query. This is often called *the probabilistic ranking principle* (PRP). In inference network model, document retrieval is modeled as an inference process in an inference network. Most techniques used by IR systems can be implemented under this model.

I.H. Witten, A. Moffat, and T.C. Bell [4] broadly explained document databases, compression, indexes, images of documents and the mg system. They also briefly introduced models of the data that needs to be compressed and includes adaptive models. Huffman codes are discussed along with algorithms and data structures for dealing with them. Arithmetic coding is discussed along with techniques for implementing it. Symbol-wise models are introduced and four data compression techniques based on them are discussed. They are Prediction by Partial Matching (PPM), block-sorting compression, Dynamic Markov Compression (DMC) and word based compression. Dictionary-based compression models such as LZ77, LZ78, and the LZW variant of LZ78 are discussed here.The process of creating indexes to aid the process of searching text efficiently by means of keywords. Subject of querying indexes, Memory-based and sort-based inversion approaches for index construction are described. Overall techniques for compressing and indexing are discussed by [4].

D. Song, D. Wagner, and A. Perrig, [5] describe cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Their techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the cipher text; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. Their [5] algorithms are simple, fast (for a document of length the encryption and search algorithms only need o(n) stream cipher and block cipher operations and introduce almost no space and communication overhead, and hence are practical to use today.

Y.-C. Chang and M. Mitzenmacher, [6] provided solutions for the problem of "Privacy Preserving Keyword Searches on Remote Encrypted Data". Their schemes are efficient in the sense that no public- key cryptosystem is involved. Indeed, approach is independent of the encryption method chosen for the remote file storage.

R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, [7] Searchable symmetric encryption (SSE) allows a party to outsource the storage of its data to another party (a server) in a private manner, while maintaining the ability to selectively search over it. They first present a definition of a multi-user searchable encryption scheme (MSSE) and some of its desirable security properties, followed by an efficient construction which, in essence, combines a single-user SSE scheme with a broadcast encryption (BE) scheme.

D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, [8] study the problem of searching on data that is encrypted using a public key system. [8] Defined the concept of a public key encryption with keyword search (PEKS) and gave two constructions. Constructing a PEKS is related to Identity Based Encryption (IBE), though PEKS seems to be harder to construct.

Jianhua Yu, **Jin Li**, **Xueli Wang, Wei Gao [9]** formalize and solve the problem of efficient conjunctive fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. Conjunctive fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. More specifically, [9] used edit distance to quantify keywords similarity, for the construction of fuzzy keyword sets. Based on the constructed fuzzy keyword sets, they propose an efficient conjunctive fuzzy keyword search scheme.

D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, [10] interested in finding solutions that are communication-efficient and, at the same time, respect the complete privacy of Alice. In [10], Alice creates a public key that allows arbitrary senders to send her encrypted e-mail messages. Each such message M is accompanied by an "encoded" list of keywords in response to which M should be retrieved. These email messages are collected for Alice by Bob, along with the "encoded" keywords. When Alice wishes to search in the database maintained by Bob for e-mail messages containing certain keywords, she is able to do so in a communication efficient way and does not allow Bob to learn *anything* about the messages that she wishes to read, download or erase. In particular, Alice is not willing to reveal what particular messages she downloads from the mail database, from which senders these emails are originating and/or what the search criterion is, including the access pattern.

N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, [11] defined and solve the problem of privacy-preserving graph

query in cloud computing (PPGQ). To reduce the times of checking sub graph isomorphism, [11] adopt the efficient principle of "filtering-and verification" to prune as many negative data graphs as possible before verification. A feature-based index is firstly built to provide feature-related information about every encrypted data graph. Then, they choose the efficient inner product as the pruning tool to carry out the filtering procedure. To achieve this functionality in index construction, each data graph is associated with a binary vector as a sub-index where each bit represents whether the corresponding feature is sub graph isomorphic to this data graph or not. The query graph is also described as a binary vector where each bit means whether the corresponding feature is contained in this query graph or not. The inner product of the query vector and the data vector could exactly measure the number of query features contained in the data graph, which is used to filter negative data graphs that do not contain the query graph.

W.K. Wong, D.W. Cheung, B. Kao, and N. Mamoulis, [12] discussed the general problem of secure computation on an encrypted database and propose a SCONEDB (Secure Computation ON an Encrypted DataBase) model, which captures the execution and security requirements. They focus on the problem of k-nearest neighbor (kNN) computation on an encrypted database. We develop a new asymmetric scalar-product-preserving encryption (ASPE) that preserves a special type of scalar product. We use APSE to construct two secure schemes that support kNN computation on encrypted data.

III. Multi-keyword Ranked Search

Consider a cloud computing broadly involve three different entities, as illustrated in Fig. 1: ***data owner*** **(O)**, ***data user*** **(U)**, and ***cloud server*** **(CS)**. Data owner has a collection of *n* data files *C* = (*F*1, *F*2, . . . , *Fn*) that he wants to upload on the cloud server in encrypted form To do so, before outsourcing, data owner will first build a secure searchable index *I* from a set of *m* distinct keywords *W* = (*w*1, *w*2, ...,*wm*) extracted from the file collection *C*, and store both the index *I* and the encrypted file collection *C* on the cloud server. Authorization between the data owner and users is appropriately done. To search the file collection for a given keyword *w*, an authorized user generates and submits a search request in a secret form called as a trapdoor *Tw* of the keyword *w* to the cloud server. Upon receiving the search request *Tw*, the cloud server is responsible to search the index *I* and return the corresponding set of files to the user.

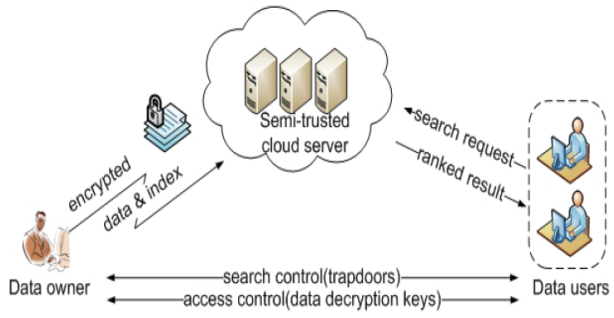

Fig 1. System Architecture

We consider the secure ranked keyword search problem as follows: the search result should be returned according to certain ranked relevance criteria (e.g., keyword frequency based scores, as will be introduced shortly), to improve file retrieval accuracy for users without prior knowledge on the file collection *C*. However, cloud server should learn nothing or little about the relevance criteria themselves as they exhibit significant sensitive information against keyword privacy. To reduce bandwidth, the user may send an optional value *k* along with the trapdoor *Tw* and cloud server only sends back the top-*k* most relevant files to the user's interested keyword *w*.

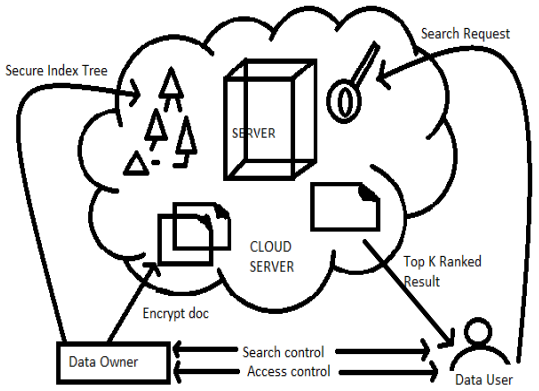

Fig 2. Privacy Preserving Over Cloud Data

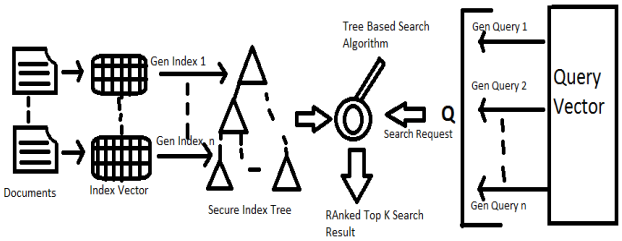

Fig 3. Searching Scheme over Cryptographic Cloud Data

To efficiently achieve multi-keyword ranked search, we apply "inner product similarity" [6] to quantitatively evaluate the efficient similarity measure "coordinate matching." and then we proposed anonymous user Id for securing users data.

IV. System Modules

There are four modules are proposed in this system as *A) Encrypt Module B) Client Module C) Multi-Keyword Module D) Admin Module.*

A. Encrypt Module

This module help the server to encrypt the document using RSA Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

B. Client Module

This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required

file and register the user details and get activation code in mail from the "customerservice404" email before enter the activation code. After that user can download the Zip file and extract that file.

C. Multi-Keyword Module

This module is used to help the user to get the accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list.

D. Admin Module

This module is used to help the server to view details and upload files with the security. Admin uses the log key to the login time. Before the admin logout, change the log key. The admin can change the password after the login and view the user downloading details and the counting of file request details on flowchart. The admin can upload the file after the conversion of the Zip file format.

V. RESULTS

Fig 4. Home Page

Fig 5. Admin Login Page

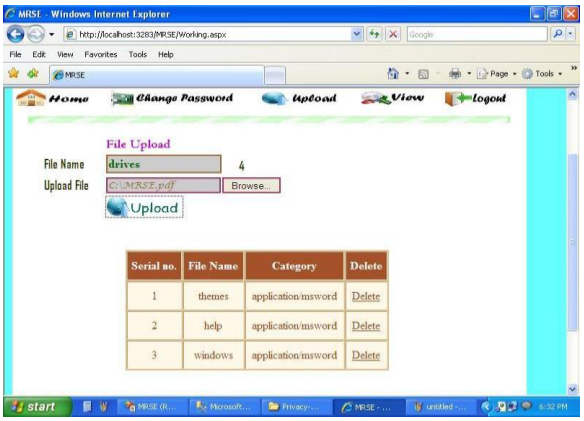

Fig 6. Working Page

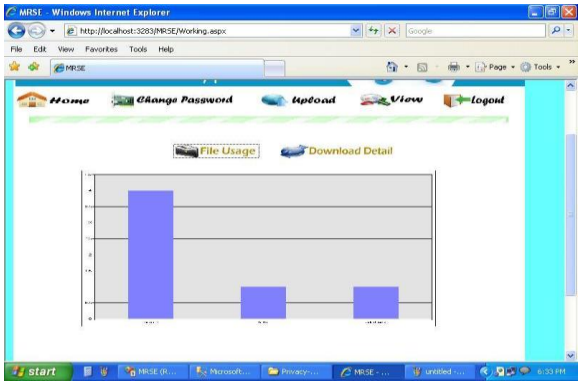

Fig 7. File Usage Page

Fig 8. Search Page

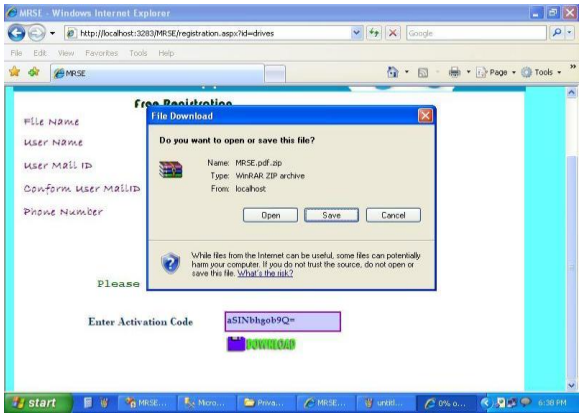

Fig 8. Download Page

VI. CONCLUSION

The multi-keyword ranking search with previous studies primarily used to provide privacy to the cloud data. Previous work, In addition, by using a secure inner product operation has proposed the basic idea of MRSE. This paper was needed to provide the actual privacy than presentation. In this system, stringent privacy is provided by assigning a unique ID to the cloud user. This user ID is, in order to protect the data of users on the cloud from the CSP and the third party of the user, will cloud service providers, as well as hidden from the third party of the user is maintained. As a result, by hiding the identity of the user, confidentiality of the data of the user is maintained.

REFERENCES

[1] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "ABreak in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.
[2] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.
[3] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.
[4] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.
[5] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
[6] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.
[7] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.
[8] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.
[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.
[10] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.
[11] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "Privacy preserving Query over Encrypted Graph-Structured Data in Cloud Computing," Proc. Distributed Computing Systems (ICDCS), pp. 393-402, June, 2011.
[12] W.K. Wong, D.W. Cheung, B. Kao, and N. Mamoulis, "Secure kNN Computation on Encrypted Databases," Proc. 35th ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 139-152, 2009.
[13] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[14] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[15] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, "IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467- 1479, Aug. 2012.
[16] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.
[17] E. Shen, E. Shi, and B. Waters, "Predicate Privacy in Encryption Systems," Proc. Sixth Theory of Cryptography Conf. Theory of Cryptography (TCC), 2009.
[18] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.

Mr. Rameshwar Murlidhar Indoriya has completed his B.E. in Information Technology Department from Anuradha Engineering College, Sant Gadgebaba Amaravati University (SGBAU). Presently he is pursuing his Masters in Computer Science and Engineering in St.Mary's Engineering College, Hyderabad, India.

Mr. P. Santosh has completed B.Tech (CSIT) under JNTU Hyderabad, and M.Tech (CSE) from Osmania University Hyderabad, Telangana. Having experience of 12 years in Academic and presently working as Assistant Professor in St.Mary's Engineering College, Hyderabad.