EFFECTIVE RETRIEVAL OF ENCRYPTED CLOUD DATA USING MULTI-KEYWORD RANKED SEARCH

G.Karthikeyan[1], V.Yuvaraj[2], A.Murugavel[3]

[1]*PG Scholar, SE, Anna University Regional Centre, Coimbatore, Tamil Nadu, India.*

[2]*Teaching Fellow, CSE Department, Anna University Regional Centre, Coimbatore, Tamil Nadu, India.*

[3]*PG Scholar, CSE, Anna University Regional Centre, Coimbatore, Tamil Nadu, India.*

***Abstract*— As Cloud Computing becomes prevalent, sensitive information are being increasingly centralized into the cloud. The traditional searching techniques are single keyword search and Boolean keyword search. In this project, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching" as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models.**

Index Terms—keyword Search, MRSE, Stemming, Top K-Query, RSA, Chunks;

I. INTRODUCTION

Cloud Computing often referred to as simply "The Cloud," is the delivery of on-demand computing data everything from applications to data centers over the Internet on a pay-for-use basis. Cloud Computing relies on restricting sharing of data to achieve coherence and economies of scale, similar to a utility like the electricity grid over a network [1] [2]. At the foundation of Cloud computing is the broader concept of converged infrastructure and shared services. Cloud storage are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating storage to users. For example, a cloud computer facility that serves European users during European business with a specific application e.g., email may reallocate the same storages to serve North American users during North America's business with a different application e.g., a web server. This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rack space, etc. are required for a variety of functions. With Cloud Computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications. Persons who are interrelated with the networking environment, Cloud Computing is a preferable platform for them. Therefore in recent days providing security has become a major challenging issue in Cloud Computing.

In the cloud environment, storages are shared among all of the servers, users and individuals. As a result files or data stored in the cloud become open to all. Therefore, data or files of an individual can be handled by all other users of the cloud. Thus the data or files become more vulnerable to attack. As a result it is very easy for an intruder to access, misuse and destroy the original form of data. An intruder can also interrupt the communication. Besides, cloud service providers provide different types of applications which are of very critical nature. Hence, it is extremely essential for the cloud to be secure. Another problem with the cloud system is that an individual may not have control over the place where the data needed to be stored. A cloud user has to use the storage allocation and scheduling, provided by the cloud service provider. Thus, it is also necessary to protect the data or files in the midst of unsecured processing. In order to solve this problem to apply security in Cloud Computing platforms.

Each provider serves a specific function, giving users more or less control over their cloud depending on the type. When you choose a provider, compare your needs to the cloud services available. The information housed on the cloud is often seen as valuable to individual with malicious intent. There is a lot of personal information and potentially secure data that people store on their computers, and this information is now being transferred to the cloud. This makes it critical for you to understand the security measures that your cloud provider has in place, and it is equally important to take personal precautions to secure your data. The multi-keyword retrieval [3] [4] over encrypted cloud data achieves high security and privacy.

II. RELATED WORK

[1] *Secured Multi-Keyword Ranked Search Over Encrypted Cloud Data*

Ankatha Samuyelu et al [5], main aim is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of “coordinate matching” (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically “inner product similarity”, i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm.

[2] *Providing Privacy Preserving in Cloud Computing*

Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust and needs to be considered at every phase of design. The [6] [9] paper tells the importance of protecting individual’s privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. Paper tells that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data.

[3] *Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data*

This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data [7] [11]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. For the first time, the paper has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality.

[4] *Cryptographic Cloud Storage*

When the benefits of using a public cloud infrastructure are clear, it introduces significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality and integrity of data. In [8] [10], an overview of the benefits of a cryptographic storage service, for example, reducing the legal exposure of both customers and cloud providers, and achieving regulatory compliance is provided. Besides this, cloud services that could be built on top of a cryptographic storage service such as secure backups, archival, health record systems, secure data exchange and e-discovery is stated briefly.

III. EXISTING SYSTEM

The enormous amounts of data are stored in remote, but not necessarily trusted servers. There are several privacy issues regarding to accessing data on such servers; two of them can easily be identified: sensitivity of i) keywords sent in queries and ii) the data retrieved; both need to be. Private Information Retrieval (PIR) enables the user to access public or private databases without revealing which data he is extracting. The large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results. So it gives undifferentiated results. Encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles.

[1] *Limitation of Existing System*

The existing system possesses several limitations and they are given as follows,

- Single-keyword search without ranking
- Boolean- keyword search without ranking
- Single-keyword search with ranking
- It provides less security
- Missing Data Integrity

IV. PROPOSED SYSTEM

In this dissertation, the challenging problem of privacy-preserving multi-keyword ranked each over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of coordinate matching. The data owner creates a search index for each document. The search index file is created using a secret key based trapdoor generation function where the secret keys are only known by the data owner. Then, the data owner uploads these search index files to the server together with the encrypted documents. Searchable index should be constructed to prevent server from performing such kind of association attack. Keyword Privacy As users usually prefer to keep their search from being exposed to others like cloud server, the most important concern is to hide what they are searching, Trapdoor Privacy Since only authorized users are allowed to acquire trapdoors for their search query, the server is not expected to have the ability to generate valid trapdoors from previous received ones., The data is divided into several Chunks and stored in different servers for effective secured data efficiency.

[1] *Advantages of Proposed System*

- Multi-keyword ranked search over encrypted cloud data (MRSE)
- Coordinate matching by inner product similarity.
- It provide high security

- Preserving data integrity and confidentiality.

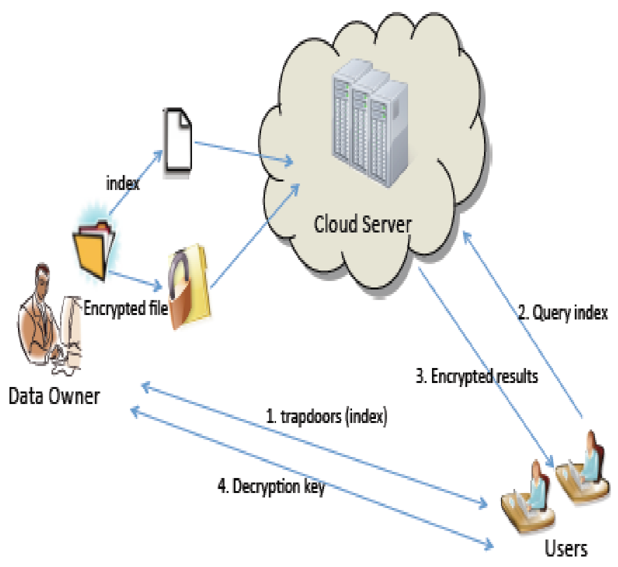

Fig 1.System Architecture

V. System Design

[1] *MRSE Framework*

For easy presentation, operations on the data documents are not shown in the framework since the data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data. With focus on the index and query, the MRSE [12] system consists of four algorithms as follows

1. ***Setup (ℓ)***

Taking a security parameter ℓ as input, the data owner outputs a symmetric key as SK.

2. ***BuildIndex(F, SK)***

Based on the dataset F, the data owner builds a searchable index I which is encrypted by the symmetric key SK and then outsourced to the cloud server. After the index construction, the document collection can be independently encrypted and outsourced.

3. ***Trapdoor(fW)***

With t keywords of interest in fW as input, this algorithm generates a corresponding trapdoor TfW.

4. ***Query(TfW, k, I)***

When the cloud server receives a query request as (TfW, k), it performs the ranked search on the index I with the help of trapdoor TfW, and finally returns FfW, the ranked id list of top-k documents sorted by their similarity with fW.

[2] *Cloud Setup*

Cloud servers are constructed with the files and the index information are maintained in the main cloud server. The data are added in each cloud servers, and network construction is made with the entire data index present in each cloud server. Query is given to the main cloud server, so that the main cloud server will verify the index information present in it & divert the query to the corresponding cloud servers.

[3] *Filtering Key words*

The words in the files are filtered and main keywords are filtered using Stemming Algorithm. The main keywords are extracted to filter the unwanted words. The Files names are updated in the corresponding cloud servers.

[4] *Encryption Module*

The query of the user is encrypted using RSA algorithm; this encryption process will prevent the data theft from the hackers. Data security is ensured using RSA encryption. **RSA** is an algorithm for public-key cryptography. It is the first algorithm known to be suitable for signing as well as encryption, and was one of the first great advances in public key cryptography. RSA is widely used in electronic commerce protocols, and is believed to be secure given sufficiently long keys and the use of up-to-date implementations. This algorithm is used to encrypt n decrypt file contents. It is an asymmetric algorithm. The RSA algorithm involves three steps: key generation, encryption and decryption.

Key generation

RSA involves a **public key** and a **private key.** The public key can be known to everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted using the private key. The keys for the RSA algorithm are generated the following way:

1. Choose two distinct prime numbers *a* and *b*.
2. Compute $n = ab$.
n is used as the modulus for both the public and private keys
3. Compute $\varphi(n) = (a - 1)(b - 1)$, where φ is Euler's totient function.
4. Choose an integer *e* such that $1 < e < \varphi(n)$ and greatest common divisor of $(e, \varphi(n)) = 1$; i.e., *e* and $\varphi(n)$ are co-prime.
e is released as the public key exponent.
Having a short bit-length.

Encryption

Alice transmits her public key *(n, e)* to Bob and keeps the private key secret. Bob then wishes to send message **M** to Alice.He first turns **M** into an integer m, such that $0<m<n$ by using an agreed-upon reversible protocol known as a padding scheme. He then computes the cipher text *c* corresponding to

$$c=m^e(mod\ n)$$

This can be done quickly using the method of exponentiation by squaring. Bob then transmits *c* to Alice. Note that at least nine values of m could yield a cipher text c equal to *m*, but this is very unlikely to occur in practice.

Decryption

Alice can recover *m* from c by using her private key exponent d via computing

$$m=c^d(mod\ n)$$

Given *m*, she can recover the original message **M** by reversing the padding scheme.

[5] Cloud *Server*

Cloud Server is the major main server which contains the index data of the entire data present in all sub Cloud Servers. The Cloud Server will act as the main server to receive the query from the user. The user query is encrypted using RSA algorithm, and sends to the main Cloud Server. The main Cloud Server decrypts the query and match with the index data present in it. The main Cloud Server will find the best match file using ranking algorithm.

[6] *Ranking Algorithm Module*

In this module we rank the best file by calculating the ratio between term frequencies with the total number of keywords. The value is calculated and compared with the rest of the values. The maximum valued files are ranked in order. The files are retrieved to the user as the index data of all the files are maintained in the index of the main cloud server. K-nearest neighbor search identifies the top k nearest neighbors to the query. This technique is commonly used in predictive analytics to estimate or classify a point based on the consensus of its neighbors. K-nearest neighbor graphs are graphs in which every point is connected to its k nearest neighbors.

The basic idea of our new algorithm: The value of dmax is decreased keeping step with the ongoing exact evaluation of the object similarity distance for the candidates. At the end of the step by step refinement, dmax reaches the optimal query range Ed and prevents the method from producing more candidates than necessary thus fulfilling the r-optimality criterion.

```
Nearest Neighbor Search (q, k) // optimal algorithm
1. Initialize ranking = index.increm-ranking (F(q), df)
2. Initialize result = new sorted-list (key, object)
3. Initialize dmax = w
4. While o = ranking.getnext and d,(o, q) I d,,, do
5. If do@, s> s dmax then result.insert (d,(o, q) , o)
6. If result.length 2 k then dmax = result[k].key
7. Remove all entries from result where key > dmax
8. End while
Report all entries from result where key I dmax
```

[7] *Best File Identification*

The best file identification is achieved using Top k Query Algorithm. The maximum ranked values are obtained using Term frequency calculation. The files are kept in the ascending order. The best files are given as output to the main cloud server. The main cloud server retrieves top files and given as output to the user.

VI. Experimental Results

To activate ranked search for effective utilization of outsourced cloud data, our system design should simultaneously achieve security and performance guarantees as follows.

[1] *Secured Multi-keyword Ranked Search*

To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.

[2] *Privacy*

To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements file. We compare the four user to fetch and modify same file at different time scale.

[3] *Effectiveness with high performance*

Above goals on functionality and privacy should be achieved with low communication and computation overhead.

VII. Conclusion

In this dissertation, the motivate and solve the problem of efficient and secure ranked multi-keyword search on remotely stored encrypted database model where the database users are protected against privacy violations .We appropriately increase the efficiency of the scheme by using symmetric-key encryption method rather than public- key encryption for document encryption. We also propose to use the blinded encryption technique in accessing the contents of the retrieved documents without revealing them to other parties. We prove that our proposed method satisfies the security requirements. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. We implement the entire scheme and extensive experimental results on the implementation demonstrate the effectiveness and efficiency of our solution.

REFERENCE

[1] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A Break in the Clouds: Towards a Cloud Definition,” ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.

[2] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, “LT Codes-Based Secure and Reliable Cloud Storage Service,” Proc. IEEE INFOCOM, pp. 693-701, 2012..

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011

[4] Ning Cao,, Cong Wang, Ming Li, Kui Ren, Wenjing Lou,” Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data” IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 1, JANUARY 2014

[5] Ankatha Samuyelu Raja Vasanthi ,” Secured Multi keyword Ranked Search over Encrypted Cloud Data”, 2012

[6] Jain Wang, Yan Zhao, Shuo Jaing, and Jaijin Le,"Providing Privacy Preserving in Cloud Computing", 2010.

[7] Y. Prasanna, Ramesh. "Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data", 2012.Practice and Theory in Public Key Cryptography, pages 332–350, 2010.

[8] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," *Proc 14th Int'l Conf. Financial Cryptograpy and Data Security*, Jan 2010.

[9] Shiba Sampat Kale, Shivaji R Lahane, "Privacy Preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over Encrypted Cloud Data" (IJCSIT) Vol. 5 (6) , 2014

[10] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM,

[11] Y. Hwang and P. Lee, "Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System," Pairing, vol. 4575, pp. 2-22, 2007.

[12] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure RankedKeyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int' lConf. Distributed Computing Systems (ICDCS '10), 2010.