Privacy Preserving Multi-Keyword Ranked Search For Multiple Data Owners In Cloud Computing

A.Saffron[1], Dr.J.Grasha Jacob[2]
[1]Dept of Computer Science
[2]Assistant Professor, Dept of Computer Science
[1, 2] Rani Anna Govrnt College for Women, Tirunelveli

Abstract- *Cloud computing is the most important service for many industries and companies. Cloud computing is used for storing their large amount for data. Through this opportunity user can upload their data and download that data from the cloud. The number of people using the cloud storage is increasing due to its easiness of use. Most cloud server do not serve single user, it give service to multiple user at the same time. This paper consist of functions in which user can search multiple file and send the file to multiple user at the same time. This can be define and focus on the problem of ranked search over encrypted data. It captures the relevance of data documents to the search query. A dynamic secret key is generated which prevent others for stealing data. To increase confidentiality in the case of multiple data owners a tree-based ranked multi-keyword search scheme can be used these use the frequently used rank searching algorithm for present the output for multi-keywords. They consider a large amount of data in the cloud. This paper is used to search multiple keywords and share files to multiple users. This consist of searchable technique which gives the facility to search over cloud.*

Keywords- Cloud computing, Ranking Search, Multiple Data Owners, Multi-keyword Search.

I. INTRODUCTION

Cloud computing is more popular. To protect the data privacy in cloud, the data has to be encrypted first by the user before uploading on the cloud and after that the data is decrypted by the key and download data on the local system. When the user wants to search a particular file then the cloud server can perform search without knowing the exact keyword. This searching technique gives rank-wise result to the user. This ranking technique is based on the frequently search the file name by the user. It also contains the encryption and decryption technology. Encryption is done at the time of uploading the data to cloud. RSA algorithm is used to file encryption. Decryption is that time provide the privacy and show the result in ranking form to make easy. As a new model of computing, cloud computing provides abundant benefits including easy access, decreased costs, quick deployment and flexible resource management, etc. Enterprises of all sizes can leverage the cloud to increase innovation and collaboration. Despite the abundant benefits of cloud computing, for privacy concerns, individuals and enterprise users are reluctant to outsource done at the time of downloading the data from cloud. When the user wants to share the file to others then a key is generated. This key is generated to protect the system from attackers. By using this key, user can download data from cloud. A tree based multi keyword search scheme is used to provide an efficient search. The words that are seemed as keywords for a document are identified and an index is formed. All the indexes such formed are then merged into one. A depth first search is used to identify the corresponding data file of the user with each search request. The TF-IDF model is used to return the top-k results. An efficient search is used to perform a depth first search. In this paper, we suggest when search multiple owner multiple keywords their sensitive data, including emails, personal health records and government confidential files, to the cloud. This is because once sensitive data are outsourced to a remote cloud; the corresponding data owners lose direct control of these data. To preserve their privacy, they will encrypt their own health data with their secret keys. In this scenario, only the authorized organizations can perform a secure search over this encrypted data contributed by multiple data owners.

II. LITERATURE SURVEY

With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "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. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

With the increasing adoption of cloud computing for data storage, assuring data service reliability, in terms of data correctness and availability, has been outstanding. While redundancy can be added into the data for reliability, the problem becomes challenging in the “pay-as-you-use” cloud paradigm where we always want to efficiently resolve it for both corruption detection and data repair. Prior distributed storage systems based on erasure codes or network coding techniques have either high decoding computational cost for data users, or too much burden of data repair and being online for data owners. In this paper, we design a secure cloud storage service which addresses the reliability issue with near-optimal overall performance. By allowing a third party to perform the public integrity verification, data owners are significantly released from the onerous work of periodically checking data integrity. To completely free the data owner from the burden of being online after data outsourcing, this paper proposes an exact repair solution so that no metadata needs to be generated on the fly for repaired data. The performance analysis and experimental results show that our designed service has comparable storage and communication cost, but much less computational cost during data retrieval than erasure codes-based storage solutions. It introduces less storage cost, much faster data retrieval, and comparable communication cost comparing to network coding-based distributed storage systems. We consider the problem of building a secure cloud storage service on top of a public cloud infrastructure where the service provider is not completely trusted by the customer. We describe, at a high level, several architectures that combine recent and non-standard cryptographic primitives in order to achieve our goal. We survey the benefits such architecture would provide to both customers and service providers and give an overview of recent advances in cryptography motivated specifically by cloud storage. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In this paper, we offer solutions for this problem under well-defined security requirements. Our schemes are efficient in the sense that no public-key cryptosystem is involved. Indeed, our approach is independent of the encryption method chosen for the remote files.

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud. For the protection of data privacy, sensitive data usually have to be encrypted before outsourcing, which makes effective data utilization a very challenging task. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. That is, there is no tolerance of minor typos and format inconsistencies which, on the other hand, are typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in Cloud Computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficacy very low. In this paper, for the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. 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. In our solution, we exploit edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search.

III. PROBLEM FORMULATIONS

Searchable Encryption

Searchable Encryption (SE) plans keep up the classification and security of proprietor's information by encouraging looking watchwords specifically on scrambled information. The members of a safe inquiry model in a cloud, normally includes information proprietor, information client and cloud server. Information proprietor scrambles the documents and using so as to relate catchphrases based file records any known cryptographic calculations. The trapdoors (encoded catchphrases) are utilized to inquiry scrambled records by cloud server in cloud database. To securely search over encrypted data, Searchable encryption schemes usually build up an index for each keyword of interest and associate the index with the files that contain the keyword. By integrating the trapdoors of keywords within the index information, effective keyword search can be realized while both file content and keyword privacy are well-preserved. Considering the large number of data users and huge amount of outsourced data files in cloud, this problem is particularly challenging as it is extremely difficult to meet also the practical requirements of performance, system usability, and high-level user searching experiences. We focus here two searching methodologies which is retrieving text based data files from encrypted cloud data The general architecture of search over encrypted cloud data is shown below.

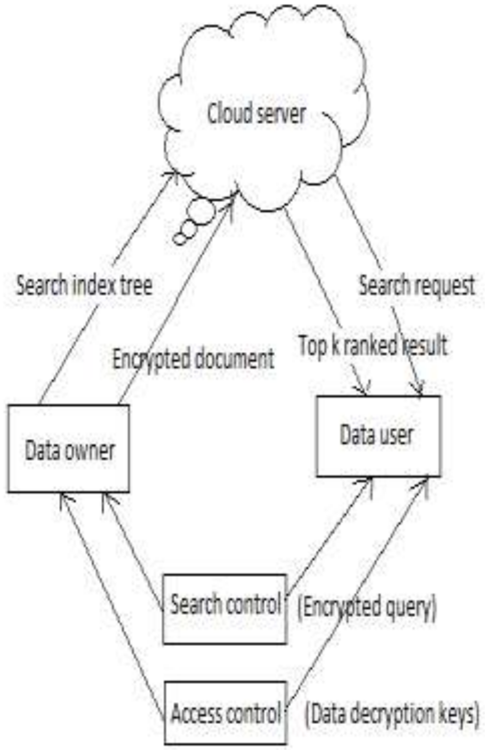

Figure 1: Searchable Encryption

FRAMEWORK OF EFFICIENT RANKED KEYWORD SEARCH

The data owner generates the searchable index terms from the unique words which was extracted from file collection. They contain the sample words and index terms which were extracted from file collection with the rank calculated. The index terms are published on cloud server with encrypted file for the identification of files easily.

Score Calculation

For calculating the score for each file, term frequency, document frequency, the length of files and the number of documents that the data owner has in his collection needs to be measured. The term frequency calculated based on how many times the keywords occurs in the same document, and for each file and for each term this needs to be calculated. The document frequency calculated based on how many times a particular keyword exists in the different document. The score or rank of the file is calculated using below equation

$$\text{Score}=\left(\frac{1}{filelength}\right)*(1+\log(termfrequency))*$$
$$(1+\log\left(no\,\frac{ofdocument}{document\ frequency}\right))$$

Ranked Keyword Search

Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., Keyword frequency), so achieve the privacy preserving data hosting service in context of cloud computing. Ranked keyword search method protect the relevance score of keyword to leaking the information about keyword for that integrate the new crypto primitive order preserving symmetric encryption and properly modify it for purpose of protect the sensitive weight information. These can be used Ranking Technique Order preserving mapping technique algorithm.

This technique is providing some functionality

1. It provides effective protocol, which fulfills the secure ranked search functionality with little relevance score information leakage against keyword privacy.
2. Ranked searchable symmetric encryption scheme is provide as-strong-as-possible security guarantee compared to previous Searchable symmetric encryption schemes. Searched using query and searched data is shown to the user.

Multi Keyword Ranked Search

In this searching of cloud data using Privacy Preserving Multi keyword Ranked Search (MRSE). Here basic concept is used is co-ordinate matching. Co-ordinate matching obtains the similarity between search query and documents. Inner product similarity is also used to describe the multi keyword ranked search over encrypted cloud data (MRSE). The features of this method are, multi-keyword ranked search, privacy preserving, high efficiency is eliminating unnecessary traffic and improve search accuracy.

Algorithm

RSA Algorithm

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 φ(n) = (a – 1)(b – 1), where φ is Euler's totient function.
4. Choose an integer e such that 1 < e < φ(n) and greatest common divisor of (e, φ(n)) = 1; i.e., e and φ(n) are co prime.e is released as the public key exponent having a short bit-length.

K-Nearest Neighbor

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 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.
```

IMPLEMENTATION MODULES

- Data User
- Data owner
- File Upload with Encryption
- Ranked Search
- File Download with Decryption

Data User

Data users are users on this system, who will be able to download files from the cloud that are uploaded by the data owners. The cloud server is stored on the file could be in huge numbers, there is a search facility provided to the user. The user should be able to do a multi-keyword search on the cloud server. The result will appears for the specific search; these users should be able to send a request to the respective data owners of the file through the system (also called trap-door request) for downloading these files. The request will also be provided a data user where it will notify if the data owner has accepted or rejected the request. If the request has been accepted, the users should be able to download the decrypted file**.**

Data Owner

The data owners should be able to upload the files. The encrypted files are before the files are uploaded to the cloud. They uploaded file to the server must be data owners are provided an option to enter the keywords for the file. These keywords are used for the indexing purpose which helps the search return values very quickly. When once these file available on the cloud, the data users should be able search using the keywords. The data owners will also be provided with a request approval screen so they are able to approve or reject the request that is received by the data users.

File Upload With Encryption

Files are uploaded to the server after file is encrypted by the encryption method. This encryption is done by RSA algorithm and generate key. This is used to encrypt the data files. This converts the plain text into the cipher text. This Encrypted Data is in the form of Binary and stored in Cloud.

Ranked Search

Rank Search allows the data users to search the files with multi-keyword rank searching. These use the frequently used rank searching algorithm for present the output for multi-keywords. Coordinate matching principle will be adopted for the multi-keyword searching. This also takes care of creating an index for faster search.

The step of ranked search is shown below

1. Data owner collects the file and generate the index by extracting the keyword from data files and published index and data files on cloud.
2. After outsourced the data files user is enable to search and download the data files from cloud server.
3. User can search through only single keyword that is encrypted and using this keyword one trapdoor is generated.
4. Using trapdoor the relevant keyword data files is searched using query and searched data is shown to the user.

File Download with Decryption

User needs decryption key to download the data files. When the user wants to access the data files then the server send the decryption key. Through this decryption key, the user who wants to access the data file, uses this decryption key to decrypt the files with the help of private key sent at the time of file sharing. This is used to decrypt the data files. This converts the cipher text into the plain text. This uses the RSA algorithm.

IV. CONCLUSIONS

In this research paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of “coordinate matching,” i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use “inner product similarity” to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we propose a basic idea of MRSE using secure inner product computation. Then, we give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. We also investigate some further enhancements of our ranked search mechanism, including supporting more search semantics, i.e., TF_IDF, and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduce low overhead on both computation and communication. In our future work, we will explore checking the integrity of the rank order in the search result assuming the cloud server is untrusted.

REFERENCES

[1] Y. Lin, X. Yao and W. Zhang, "An Efficient Ranked Multi-Keyword Search for Multiple Data Owners Over Encrypted Cloud Data*," in* IEEE Access*, vol. 6, pp. 21924*- 21933, 2018.

[2] Wei Zhang, Student Member, IEEE, Yapping Lin, Member, IEEE, Shang Xiao, Member, IEEE. Fellow, IEEE, and Zhou,” Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing”.

[3] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, “Secure ranked multi-keyword search for multiple data Owners in cloud computing,” in *Proc. 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN2014)*. Atlanta, USA: IEEE, June 2014, pp. 276–286.

[4] C. Wang, S. S. Chow, Q. Wang, K. Ran, and W. Lou, “Privacy preserving public auditing for secures cloud storage,” Computers, IEEE Transactions on, vol. 62, no.2, pp. 362–375, 2013.[4] T. Jung, X. Y. Li, Z. Wan, and M. Wan, “Privacy preserving cloud data access with multi-authorities,” in Proc. IEEE INFOCOM’13, Turin, Italy, Apr. 2013, pp.

[5] A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Communication of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[6] C. Wang, S. S. Chow, Q. Wang, and W. Lou, “Privacy preserving public auditing for secure cloud storage,” *Computers, IEEE Transactions on*, vol. 62, no. 2, pp. 362–375, 2013.

[7] D.Song, D.Wagner, and A.Perrig, “Practical techniques for searches on encrypted data,” in *Proc. IEEE International Symposium on Security and Privacy (S&P’00)*, Nagoya, Japan, Jan. 2000.

[8] Ning Cao, Cong Wang, Ming Li, KuiRen, Wenjing Lou , “Privacy Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data” IEEE Transactions on parallel and Distributed Systems, Vol. 25,January 2014.

[9] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, “Privacy preserving public auditing for secure cloud storage,” *Computers, IEEE Transactions on*, vol. 62, no. 2, pp. 362–375, 2013.

[10] D.Song, D.Wagner, and A.Perrig, "Practical techniques for searches on encrypted data," in *Proc. IEEE International Symposium on Security and Privacy (S&P'00)*, Nagoya, Japan, Jan. 2000, pp. 44–55.

[11] E. Goh. (2003) Secure indexes. [Online]. Available: http://eprint.iacr.org/.

[12] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proc. ACM CCS'06*, VA, USA, Oct. 2006, pp. 79–88.

[13] D. B. et al., "Public key encryption with keyword search secure against keyword guessing attacks without random oracle," *EUROCRYPT*, vol. 43, pp. 506–522, 2004.