Survey On Multi-Keyword Search Over Encrypted Data In Cloud Computing

Varshini.B.V

Geethapriya.J

Assistant Professor,
R.M.D Engineering College, Tamilnadu

***Abstract:** Cloud storage is very popular in recent trend as it provides more benefits over the traditional storage solutions. To ensure security in cloud, encryption techniques play a major role when data are outsourced to the cloud. The problem of retrieving the encrypted data over the cloud is complex. Many search techniques are used for retrieving the encrypted data from cloud. This paper focuses on a set of Multi-keyword Search mechanisms over encrypted data, which provides secured data retrieval with high efficiency. It concludes that, Multi-keyword search is meant to be best methodology for searching the encrypted data in the Cloud. It gives more efficiency than single keyword search.*

***Keywords:** Cloud, Multi-keyword search, Outsourced data, Encrypted data.*

I. INTRODUCTION

Cloud computing provides various facilities to the users and enterprises to outsource the sensitive data such as emails, personal health records, company finance data, and government documents, etc. Since data owners outsourced the non-encrypted sensitive data, it causes risks. The cloud server may leak data information to unauthorized entities or even be hacked. So, for the protection of sensitive data, information has to be encrypted before it is outsourced to the public cloud. Various searchable encryption schemes are available. It allows users to securely search over the encrypted data through single keyword. This single keyword scheme does not provide any relevant information. Multi-keyword search greatly enhance the system usability by enabling search result gives relevant data instead of sending undifferentiated data. It increases the file retrieval accuracy. Searchable Secure index is created for information retrieval. Multi-keyword search over outsourced cloud data improve the efficiency of searching encrypted cloud data.

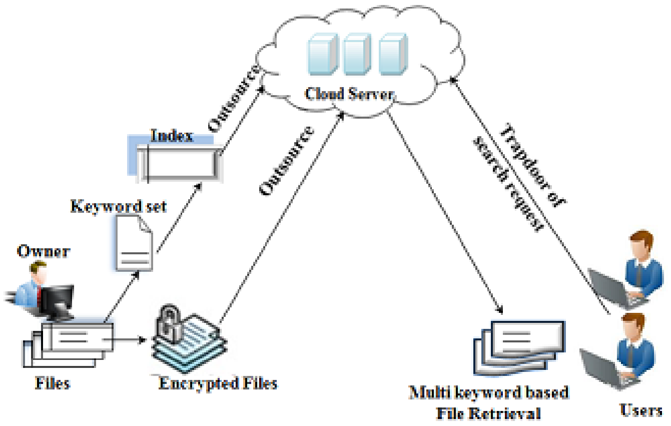

Figure 1.1: Architecture for multi-keyword based file retrieval over encrypted data cloud

Fig 1.1 shows the architecture for multi-keyword based search over encrypted cloud data. The architecture has three entities data owner, cloud server and end users. Data owner have the collection of files that is to be outsourced in an encrypted format. User search the required file with the help of keyword in secure manner, cloud server search the index file and return the corresponding list of files. Before outsourcing, data owner first create a secure searchable index from a collection of distinct keywords extracted from the file collection, and store both the index and the encrypted file collection on the cloud server. Authentication between user and data owner is done. For the search operation, user creates a secure search request in the form of trapdoor to the server. On receiving the search request, server search on the index file and issue the corresponding files to the user. Multi-keyword search greatly enhances system usability by returning the matching files with the help of some relevance criteria.

II. CLOUD COMPUTING MODELS

Cloud gives resources to user through different models. It is provided by the service providers and hosted by cloud vendors to users. Fig 2.1 and 2.2 explains the various cloud services and layered architecture of cloud services. The various Service models in cloud are explained as follows:

- ✓ SAAS: Software as a Service provides the required software, network and operating system to the users. Users don't need to install them in their hardware. It is an application that can be accessed from anywhere in the world only if we have a computer with an internet connection.

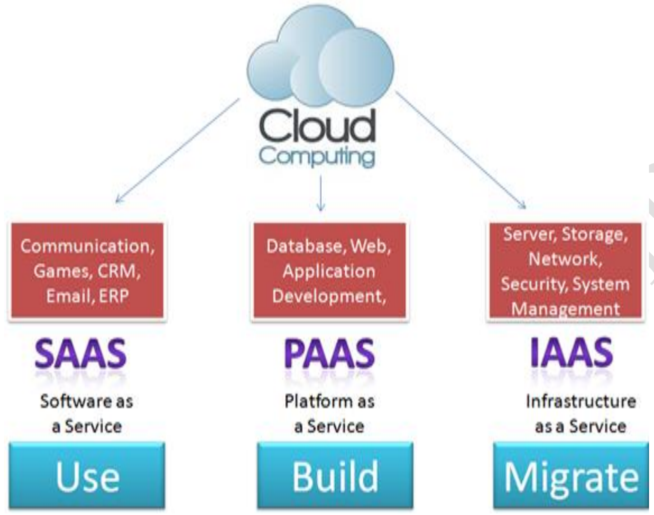
Cloud Computing
Communication, Games, CRM, Email, ERP
SAAS
Software as a Service
Use
Database, Web, Application Development,
PAAS
Platform as a Service
Build
Server, Storage, Network, Security, System Management
IAAS
Infrastructure as a Service
Migrate

Figure 2.1: Different cloud services

- ✓ PAAS: Platform as a Service provides network and operating system to users. It is a platform for the developers to create their own application. At this layer user don't need to manage their virtual machines and no need to manage an operating system.
- ✓ IAAS: Infrastructure as a Service also known as *"hardware as a service"*. It's about the physical environment of cloud where it provides the storage space, networking and other needed resources. The user has the control on storage, network.

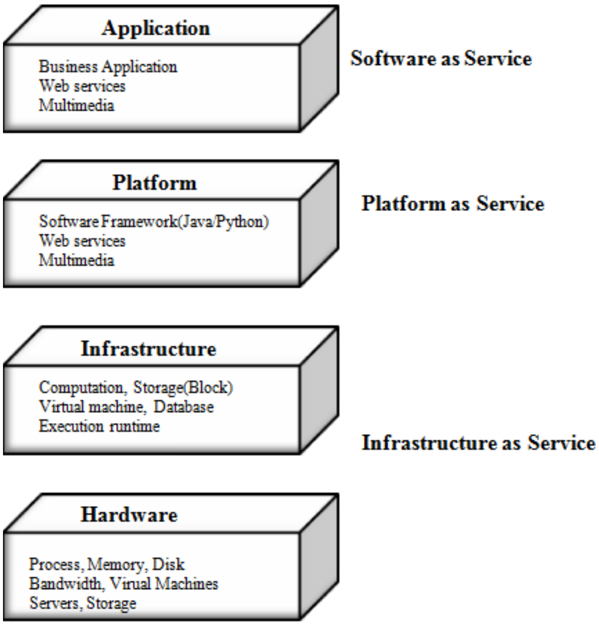

Figure 2.2: Layered Architecture of cloud services

Application layer is the most utilized layer of cloud computing where the users deploy their applications. Platform layer is useful to run applications for the user. Infrastructure layer enables user request for computing resources by accessing suitable resources and deploy huge numbers of virtual machines (*VMs)* on hardware. The hardware layer is referred to as server layer. It represents the physical hardware that provides actual resources. Hardware resources are of low cost.

III. DEPLOYMENT MODELS IN CLOUD

- ✓ *PRIVATE CLOUD:* The infrastructure is dedicated to a particular organization and not shared with other organizations. Private cloud is more expensive and more secure when comparing to public cloud. Examples Oracle, Sun, IBM.
- ✓ *PUBLIC CLOUD:* The infrastructure can be shared by many organizations through internet. The infrastructure is hosted by cloud vendor at the vendor's premises. Examples Google, Amazon, Microsoft.
- ✓ *COMMUNITY CLOUD:* Within a community resources are shared between organizations, like same company of different branches.
- ✓ *HYBRID CLOUD:* Its combination of Private cloud, Community cloud and Public cloud. For sensitive data, user uses private cloud and other data user uses public cloud.

IV. LITERATURE SURVEY

SINGLE KEYWORD SEARCHABLE ENCRYPTION

A Single keyword searchable encryption method is generally built by creating an encrypted searchable index. The index content is hidden to the server. When the server gives the correct trapdoors that are generated via a secret key(s), the information will be revealed to the user. The main problem of single keyword based search is that it is not comfortable enough for complex information needs. This drawback of single keyword search is overcome by multi-keyword searchable encryption.

MULTI-KEYWORD SEARCHABLE ENCRYPTION

The different kinds of searchable encryption approaches have been developed to provide the ability for retrieving the encrypted documents through a keyword search. In general, these systems build a secure index structure and outsource it along with the encrypted documents to the remote server. Authorized users submit their requests as secret trapdoors that are included with the stored indexing information. The received trapdoor is used by the server to search over the stored index, and the matching encrypted documents are retrieved. However, the previous searchable encryption schemes are not practical for real world cloud computing scenario because these systems are developed to handle a single keyword search. The various multi-keyword searching schemes are discussed below:

A. PRIVACY PRESERVING RANKED MULTI-KEYWORD SEARCH FOR MULTIPLE DATA OWNERS IN CLOUD COMPUTING

Wei Zhang et.al [3] (2016) proposed new protocols such as novel dynamic secret key generation protocol and a new data user authentication protocol. It enabled the cloud server to perform secure search among multiple owners' data, which is in encrypted form with different secret keys. A novel additive order and privacy preserving method is to rank the search results and preserve the privacy of relevance scores between keywords and files.

ADVANTAGES

- ✓ It's more efficient on large data and keyword sets.

DISADVANTAGES

- ✓ It does not support secure fuzzy keyword search in a multi-owner data

B. A SECURE AND DYNAMIC MULTI-KEYWORD RANKED SEARCH SCHEME OVER ENCRYPTED CLOUD DATA

Zhihua Xia et.al [7] (2016) proposed a secure multi-keyword ranked search over encrypted cloud data. In this scheme, data owner can do dynamic update operations like deletion and insertion of documents and sending them to the cloud server. The vector space model and the TF X IDF model are combined to create index file and query generation for encrypted documents. Tree based index structure is created to achieve sub-linear search time and it deals with deletion and insertion of documents. The secure KNN algorithm is used to encrypt the index and query vectors and it is ensuring accurate relevance score calculation between encrypted index and query vectors. In order to avoid statistical attacks, in the index vector phantom terms are added for blinding search results.

ADVANTAGES

- ✓ Its supports dynamic update operations like deletion and insertion of documents.
- ✓ Balanced Binary Tree as index and Greedy Depth-first Search algorithm is used to obtain better efficiency than linear search.
- ✓ Parallel search process is carried out to reduce the time cost.

DISADVANTAGES

- ✓ All the users keep the same secure key for trapdoor generation in a symmetric search encryption (SE) scheme.
- ✓ The revocation of the user is difficult. If it's needed to revoke a user in this scheme, we need to rebuild the index and distribute the new secure keys to all the authorized users.
- ✓ Dishonest data user may distribute his/her secure keys to the unauthorized user.

C. ENABLING FINE-GRAINED MULTI-KEYWORD SEARCH SUPPORTING CLASSIFIED SUB-DICTIONARIES OVER ENCRYPTED CLOUD DATA

Hongwei Li et.al [2] (2016) proposed new search scheme called as Fine Grained *Multi-keyword* Search (FMS). FMS introduced the relevance scores and preference factors upon keywords. It enables the correct keyword search and personalized user experience. Then, they developed a very efficient fine grained multi-keyword search scheme which supports complicated logic search the mixed “AND”, “OR” and “NO” operations of keywords. The classified sub-dictionaries technique is used to achieve better efficiency on index building and trapdoor generation. Finally, they have analyzed the security in terms of confidentiality of documents, privacy protection of index and trapdoor, and *unlink* ability of trapdoor. They have validated the performance of the proposed schemes using the real-world data set.

ADVANTAGES

- ✓ Provides good security.
- ✓ Better performance in terms of functionality, query complexity and efficiency.

DISADVANTAGES

- ✓ It does not support highly *scalable* efficient search on large database.

D. TOWARD EFFICIENT MULTI-KEYWORD FUZZY SEARCH OVER ENCRYPTED OUTSOURCED DATA WITH ACCURACY IMPROVEMENT

Zhangjie Fu et.al [1] (2016) proposed an efficient Multi-keyword fuzzy ranked search scheme based on Wang et.al scheme for multi-keyword fuzzy search. Wang et.al scheme was vulnerable to server out-of-order problems during ranking process, which was addressed by Zhangjie Fu et.al. The traditional techniques were focused on multi-keyword exact search and single keyword fuzzy search. But, those techniques had a very less significance in real-search scenarios encrypted data, compared with multi-keyword fuzzy search. Zhangjie Fu et.al developed a method of keyword transformation based on a uni-gram, which is tolerable to misspelling of one letter and for other spelling mistakes in search terms. Keyword weight is considered to construct the ranked list of the results to achieve high accuracy. So, the files which are more relevant to the keywords will have greater chances to appear first on the list.

ADVANTAGES

- ✓ Pre-defined fuzzy set (MFSE) of keywords is not required.
- ✓ Tolerant to single and multiple misspelling of letters in keywords.
- ✓ Efficient and accurate.
- ✓ Result is ranked according to relevance score.
- ✓ Handles the problems occurring in bigram method.

DISADVANTAGES

- ✓ The unigram approach removes the order dimension, so keywords like anagrams cannot be distinguished.

E. SECURE MULTI-KEYWORD SEARCH SUPPORTING DYNAMIC UPDATE AND RANKED RETRIEVAL

Jingbo Yan et.al [8] (2016) proposed a new multi-keyword dynamic search scheme with result ranking to make the search over large-scale encrypted data more secure. To enhance security, they used a powerful function-private inner product encryption. This scheme evaluates the inner product in function-private way and prevents the leakage of search pattern. The data owner first generates a secure searchable index on the basis of keyword set extracted from a document collection. First, the index has to be loaded to the disk and then loaded into the memory gradually according to the query. This requires a high I/O efficiency. They used B^+ tree to reduce the number of I/O operations. If the degree of B^+ tree is larger, the fewer nodes will be generated which results in less construction time.

ADVANTAGES

- ✓ Prevents data owner from learning additional information about document collection or query during its search.
- ✓ Reduces communication cost by returning Top-k relevant results to the users.
- ✓ Reduces computation costs because, users don't have to decrypt every document to choose the desired ones.
- ✓ Provides better I/O efficiency.

DISADVANTAGES

- ✓ When document collection is too large, the collection will be divided into sub- collections and stored in different servers, which makes the ranking process to be delayed.

F. PRIVACY-PRESERVING MULTI-KEYWORD SIMILARITY SEARCH OVER OUTSOURCED CLOUD DATA

Chia-Mu Yu et.al [9] (2015) proposed three solutions to overcome a problem of Privacy-Preserving Multi-keyword Similarity Search (PPMKSS) over outsourced cloud data. Consider multiple keywords are specified by the user to search text data in cloud. The cloud returns the files contain more than a threshold number of input keywords or similar keywords, where the similarity is defined based on the edit distance metric. Chia-Mu Yu [9] proposed three solutions namely PPMKSS-1, PPMKSS-2, and PPMKSS-3, which deals with search problems. First solution is blind signature is to provide user access privacy, second solution is novel use of Bloom filter's bitten pattern, which provides the speedup of task search at the cloud side. The third solution is to achieve secure search against insider threats and efficient in terms of the search time at the cloud side.

ADVANTAGES

- ✓ PPMKSS-3 is highly efficient in terms of storage, computation, and communication overhead.
- ✓ A Blind Signature method ensures the user access privacy.

DISADVANTAGES

- ✓ It does not support dynamic update operations like deletion and insertion of documents.
- ✓ It does not support secure search index.

G. PRIVACY-PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA:

Ning Cao et.al [12] (2014) proposed a variety of privacy requirements for secure *Multi-keyword* Ranked Search over Encrypted (MRSE) cloud data. Similarity measure of "coordinate matching" method is introduced to capture the relevance of data documents to search query efficiently and use "inner product similarity" method is to evaluate such similarity measure. The two improved MRSE schemes are to achieve various strict privacy requirements in two different models i.e., System model and Threat Model. To improve search experience of the data search service, two schemes such as *TF X IDF*, and dynamic data operations is used to support more search semantics.

ADVANTAGES

- ✓ It supports Dynamic data operation in cloud environment.

DISADVANTAGES

- ✓ It does not support Integrity check in rank order in the search result, when the cloud server is *untrusted.*

H. VERIFIABLE PRIVACY-PRESERVING MULTI-KEYWORD TEXT SEARCH IN THE CLOUD SUPPORTING SIMILARITY-BASED RANKING:

Wenhai Sun et.al [10] (2014) proposed a verifiable privacy-preserving multi-keyword text search (MTS) scheme with similarity-based ranking to provide efficient and secure search functions over encrypted data. To support multi-keyword search and search result ranking, construct the search index tree based on term frequency and the vector space model with cosine similarity measure to obtain higher search result accuracy. Tree-based index structure and various adaptive methods for multi-dimensional (MD) algorithm are used to improve the search efficiency. To improve the search privacy, two secured index schemes such as BMTS for known cipher text model and EMTS for known background model are developed to meet the privacy requirements under strong threat models. The proposed index tree structure is to enable authenticity check over the returned search results.

ADVANTAGES

- ✓ Its efficiency better than linear search in cloud environment.
- ✓ The leakage of sensitive frequency information is avoided.

DISADVANTAGES

- ✓ It doesn't supports Dynamic data operation in cloud environment.

I. ACHIEVING EFFECTIVE CLOUD SEARCH SERVICES: MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA SUPPORTING SYNONYM QUERY

Zhangjie Fu et.al [11] (2014) proposed a data search technique which combines multi-keyword ranked search and synonym based search to attain more accurate search results and to support synonym queries because, the existing search approaches cannot support requirements such as ranked search, multi-keyword search and semantic-based search. It is quite common that the cloud customers searching an input might be synonyms of predefined keywords instead of the exact keywords. The traditional search schemes have no tolerance to synonym substitution. To attain multi-keyword search and result ranking, they used Vector Space Model (VSM) to build the document index [10], where each document is expressed as a vector in which each dimension values is the Term Frequency (TF). With these document index vectors, an index tree has been constructed. Thus, the related documents can be found by traversing the tree. The E-TFIDF algorithm was used to extract the most representative keywords from outsourced text documents.

ADVANTAGES

- ✓ Supports both multi-keyword ranked search and Synonym-based search, so more suited for real search scenarios.
- ✓ The E-TFIDF algorithm improves the accuracy of search results.
- ✓ The index data consume only a small amount of storage space.

DISADVANTAGES

- ✓ The search time depends on the number of documents in the data set.

J. TOWARD SECURE MULTI-KEYWORD TOP-K RETRIEVAL OVER ENCRYPTED CLOUD DATA

Jiadi Yu et.al [13] (2013) proposed the two round search able encryption (TRSE*)* search is to eliminate data leakage, which supports top-k *multi-keyword* retrieval. In this scheme they have employed in vector space model and *homomorphic* encryption. The vector space model supports to provide sufficient search accuracy. The *homomorphic* encryption, which enables users to involve in the ranking scheme while the majority of computing works, is done on the server side and its operations only on cipher text.

ADVANTAGES

- ✓ TRSE scheme guarantees high data privacy in cloud.
- ✓ Data leakage is eliminated in cloud environment.
- ✓ Data security is ensured and reduces communication overhead

DISADVANTAGES

- ✓ It does not support effective search able index update.

V. CONCLUSION

In this survey paper, various Multi-keyword searching techniques for encrypted data in cloud are discussed. Some of the important issues to be handled by the searching technique are Data utilization, Data privacy, keyword privacy, Fine-grained Search, Scalability, Efficiency, Index privacy, Query Privacy, Ranking Result, Index confidentiality, Query confidentiality and Semantic security. Data security can be provided by some different methods like binary balanced tree as an Index and security can be provided by the Public-Key Encryption. From the above survey, we can justify that Multi-keyword search over encrypted data in cloud is more efficient than Single keyword search. The advantages and disadvantages of various Multi-keyword searching techniques are also discussed in this paper.

REFERENCES

[1] Zhangjie Fu, Xinle Wu, Chaowen Guan, Xingming Sun, Kui Ren, "Toward Efficient Multi-Keyword Fuzzy Search Over Encrypted Outsourced Data With Accuracy Improvement" IEEE Transactions On Information Forensics And Security, Vol. 11, No. 12, December 2016.

[2] Hongwei Li, Yi Yang,Tom H. Luan, Xiaohui Liang, Liang Zhou, Xuemin (Sherman) Shen, "Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions On Dependable And Secure Computing, Vol. 13, No. 3, May/June 2016.

[3] Wei Zhang, Yaping Lin, Sheng Xiao, JieWu, Siwang Zhou, "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing", IEEE Transactions On Computers, Vol. 65, No. 5, May 2016.

[4] Hui Cui, Zhiguo Wan, Robert H. Deng, Guilin Wang, Yingjiu Li, "Efficient and Expressive Keyword Search Over Encrypted Data in Cloud", IEEE Transactions on Dependable and Secure Computing Journal Of , Vol. , No., 2016.

[5] Chi Chen, Xiaojie Zhu, Peisong Shen, Jiankun Hu, Song Guo, Zahir Tari, Albert Y. Zomaya, "An Efficient Privacy-Preserving Ranked Keyword Search Method", IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 4, April 2016.

[6] Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, Fengxiao Huang", Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement," IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 9, September 2016.

[7] Zhihua Xia, Xinhui Wang, Xingming Sun,Qian Wang, Member, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, February 2016.

[8] Jingbo Yan, Yuqing Zhang, Xuefeng Liu, "Secure Multi-keyword Search Supporting Dynamic Update and Ranked Retrieval", Services and applications, China Communications, 2016.

[9] Chia-Mu Yu, Chi-Yuan Chen, and Han-Chieh Chao, "Privacy-Preserving Multi-keyword Similarity Search Over Outsourced Cloud Data", 1932-8184 © 2015 IEEE Systems Journal.

[10] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, "Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking", IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 11, November 2014.

[11] Zhangjie Fu, Xingming Sun, Nigel Linge, Lu Zhou, "Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query", IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014.

[12] 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.

[13] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Minglu Li, "Toward Secure Multi-keyword Top-k Retrieval over Encrypted Cloud Data", IEEE Transactions On Dependable And Secure Computing, Vol. 10, No. 4, July/August 2013 239.

[14] Cong Wang, Ning Cao, Kui Ren, Wenjing Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE Transactions On Parallel And Distributed Systems, Vol. 23, No. 8, August 2012.

[15] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou, "Fuzzy Keyword Search over Encrypted Data in Cloud Computing", INFOCOM, 2010 Proceedings IEEE.