International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

A Survey on Context based Search over Encrypted Cloud Data Techniques

Saurabhkumar Mistry[1], Purvi Tandel[2]

[1]*Computer Engineering, CGPIT, Uka Tarsadia University, Bardoli ,India, saur19792@gmail.com*
[2] *Computer Engineering, CGPIT, Uka Tarsadia University, Bardoli ,India, purvi.tandel@utu.ac.in*

Abstract— Cloud computing is one of the fast growing technology. Cloud computing support large scale infrastructure for high performance computing, so that large amount of data outsourced on cloud server. Data privacy is major issue in cloud computing because of cloud data server is not fully trusted there for data is encrypted and then stored on cloud. Due to encrypted data computation and searching of data is important tasks because traditional plaintext searching techniques can not directly be applied on encrypted cloud data. There for, Based on the user's requirement retrieving the encrypted data from cloud servers is very challenging task. Retrieving the encrypted data and processing the query over cloud server is very difficult. For searching of encrypted cloud data from cloud servers many searching techniques are available. This study concentrated on various searching techniques for retrieving the text data from cloud servers.

Keywords- Cloud Computing, Keyword Search, Encrypted Cloud Data, Data Retrieval

I. INTRODUCTION

Now a days a new generation of technology is transforming the world of computing. Today more and more users use the internet based services like data storage, processing and services collectively known as cloud computing. Many services like documentation, storage, emails and office applications like accounting, HR, purchase and CRM are delivered from cloud. Cloud computing provides pay per use, on demand service on internet. It offers organizations greater choice, agility and flexibility while also driving efficiency gains and lowering over all IT costs. Moving from a highly secure data center to internet based cloud model will require great emphasis on security and privacy.

II. CLOUD COMPUTING

Cloud computing provides elastic services, high performance and scalable data storage to a large and everyday increasing number of users. Cloud computing enlarged the arena of distributed computing systems by providing advanced internet services that complement and complete functionalities of distributed computing provided by the web, grid computing and peer-to-peer networks. In fact, cloud computing systems provide large-scale infrastructures for high-performance computing that are dynamically adapt to user and application needs [1].

Cloud computing can be defined on the basis of many aspects like processing, storage resources, the service-oriented interface and the exploitation of virtualization techniques etc. The National Institute of Standards and Technology (NIST) have given a complete reference definition. NIST defined "Cloud computing is a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[2]. Moreover, "Cloud model promotes availability and is comprised of five key characteristics, three delivery models, and four deployment models".

Five essential elements of cloud computing are:

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

In cloud computing there are main three service model are connecting with each other and create one cloud. The three service models are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The cloud computing architecture with its models is shown below figure 1.

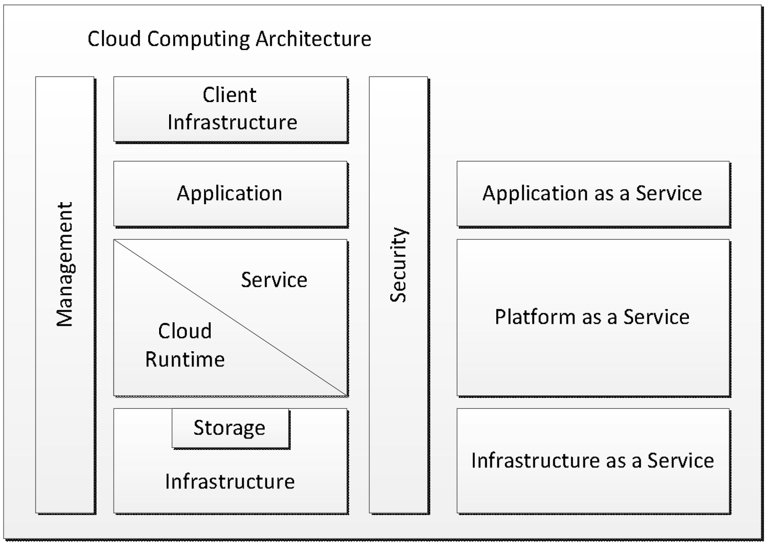

Figure 1. Cloud Computing Architecture

III. SEARCHING IN ENCRYPTED CLOUD DATA

Now a days cloud computing enables an economic paradigm in which more and more sensitive information are being centralized for data service outsourcing, such as emails, personal health record, government document, business confidential documents etc. By storing their data into the cloud, the data owners can be relieved from the burden of data storage and maintenance so as to enjoy the on-demand high quality data storage service. However, the fact that data owners and cloud servers may no longer be fully trusted. So that for sensitive data usually encryption is done then outsource the data for data privacy and combating unsolicited accesses.

However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Moreover, in cloud computing data owners may share their outsourced data with a large number of users. The individual users might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways is to selectively retrieve files through keyword-based search instead of retrieving all the encrypted files back which is completely impractical in cloud computing scenarios. Such keyword-based search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios, such as google search[5,6]. Unfortunately, data encryption restricts user's ability to perform keyword search and thus makes the traditional plaintext search methods unsuitable for cloud Computing. Besides this, data encryption also demands the protection of keyword privacy since keywords usually contain important information related to the data files. Although encryption of keywords can protect keyword privacy, it further renders the traditional plaintext search techniques useless in this scenario.

3.1. Boolean Symmetric Searchable Encryption (BSSE).

Boolean Symmetric Searchable encryption is most general model, which achieves all basic Boolean searches over encrypted data. The basic boolean operations are perform disjunction, the conjunction and the negation. The disjunctive search allows the user to search for encrypted documents containing "word1 or word2 ….or wordn". The conjunctive search allows the user to search for the encrypted documents containing: "word1 and word2 … and wordn" and finally the negative search allows the user to search for all encrypted documents which do not contain particular words[3].

The BSSE is worked on basic idea of considering keywords as vectors and using the Gram Schmidt process to orthogonalize and then orthonormalize them. It further makes use of a very efficient operation, the inner product, to perform searches at the server side. The inner product indeed leverages the orthonormalized keywords to efficiently test if a boolean expression query matches the label corresponding to an encrypted document or not. It is an application of mathematical and computing principles to practical security, especially in the searchable encryption scope[3].

In addition to that, our BSSE scheme presents new features for encrypted documents stored in the server. Indeed, the query sent for retrieving encrypted documents is randomized, which means that the server sees different queries even if the same boolean query is sent several times.

Apply Boolean Symmetric Searchable encryption (BSSE) [3] on a set of documents D perform some steps such that:

Step1: (K1, K2) → Gen (1k): a key generation algorithm, which takes a security parameter k as input and outputs two secret keys K1 and K2.

Step2: (X, C) → Enc(K1;K2,W,M,D): a probabilistic algorithm, which takes as input the collection of documents D, the keyword set W, the set of indexes M and the secret keys K1 and K2, and outputs the collection of encrypted documents C = {C1,….,Cn} and a set of labels X = {X1,….,Xn } where each label Xi is associated with the corresponding encrypted document Ci.

Step3: Q → Query(K1, B(W)): a probabilistic algorithm, which takes as input a boolean expression B(W) and the key K1, and outputs an encoded query.

Where, B(W) is Boolean expressions: Let W = {w1,….,wr} be a set of keywords. A boolean expression on W has the form: B(W) = +w1 *+w2 *…*+wr, where * is a binary logical operator * □ {□,□} and + is a unitary operator to indicate that the keyword can be complemented that contain null set.

Step4: L → Test(Q;X): a deterministic algorithm, which takes as input a query Q and the set of labels X and outputs the list of cipher texts L subset C matching the query.

In above steps the step1 and stpe2 is perform in data owner module and step3 and step4 is perform in user module who search the data.

3.2. 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[4].

This technique is providing some functionality[4].

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.

Above cloud data hosting service involving three different entities data owner (O), data user (U) and cloud server (CS). The stepwise algorithm is described below. Firstly setup phase is performed in which taking a security parameter p as input , data owner outputs a symmetric key as Sk.

Step1: Data owner (O) has a collection of n data files F = (f1,f2,…,fn) that he/she wants to outsource on the cloud server in encrypted form. For encrypting the data it uses symmetric encryption function and symmetric key Sk. The encrypted document collection stored in cloud server, denoted as C = (C1,C2,…,Cn).

Step2: To generate the searchable ranked based index associated with C, denoted as I = (I1, I2, …, In).Also generate sub index Ii for finding rank of document according to their similarity. For generating secure index we used order preserving symmetric encryption.

Ranking Equation (1) for generating ranked of searchable keyword is shown below figure 2 [4].

$$Score(Q,F_d) = \sum_{t\in Q} \frac{1}{|F_d|} \cdot (1+\ln f_{d,t}) \cdot \ln(1+\frac{N}{f_t}) \qquad \text{.........................}\ (1)$$

Here Q denotes the searched keywords; $f_{d,t}$ denotes the TF of term t in file F_d; ft denotes the number of files that contain term t; N denotes the total number of files in the collection; and $|F_d|$ is the length of file F_d, obtained by counting the number of indexed terms, functioning as the normalization factor.

Step3: After performing step1 and step2 the encrypted data and generated index is stored on cloud server.

Step4: Generating trapdoor(W) with t keywords of interest in set of keywords W as input, this algorithm generates users searchable keyword trapdoor Tw. This trapdoor is work as communicator between user and cloud data server.

Step5: At last the when cloud server receives a query request as trapdoor (Tw, k), it performs the ranked search on the index I with the help of trapdoor Tw and finally returns list of all documents according to ranked id with similar searched keywords(W).

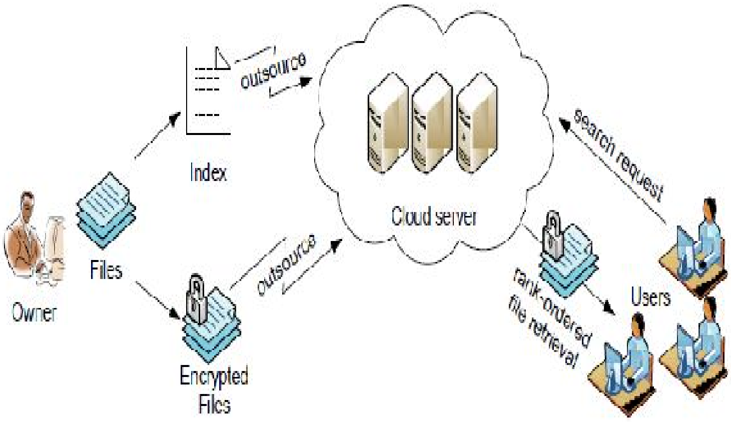

Figure 2. Architecture of the search over encrypted cloud data

3.3. Multi Keyword Ranked Search

In this method 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 eliminate unnecessary traffic and improve search accuracy.

There are four modules of searching are performed over encrypted cloud data. The four modules are Encrypt module, client module, multi-keyword module and admin module.

Initially In data owner are upload the data files on the cloud server so data files are encrypted first and along with generate the encrypted index with the use of encrypted module in which perform some steps [5,9].

- Setup: In setup phase taking a security parameter p as input and data owner outputs symmetric key as Sk.
- BuildIndex(D,Sk): Based on data set D, data owner builds a searchable index I which is encrypted by the symmetric key Sk and then outsourcing to a cloud server.
- Enc(D, Sk): In this step using symmetric key encryption algorithm data set D = (d1,d2,…,dn) that contain n number of documents are encrypted and generate encrypted data set C=(C1,C2,…,Cn).

In second phase the user or clients wants to search the documents from the data set it searching multi-keyword. In this the client module is work for searching keyword.

- Trapdoor(W): In the trapdoor the n numbers of keywords of interest in W as input, this algorithm generates a trapdoor Tw.

In last phase the searching of the keywords related data is searched using privacy-preserving Multi-keyword Ranked Search (MRSE). In MRSE its work on a "coordinate matching" that 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). This algorithm is concerns some privacy requirements are keyword privacy, trapdoor privacy, search pattern, access patterns.

- Query(Tw,k,I): When cloud server receives a query request as (Tw, k), it performs the ranked search on the index I with the help of trapdoor Tw and finally returns Dw, the ranked id list of top-k documents stored by their similarity with keywords in search request.

This scheme introduces a lower overhead on both the computation and communication. Also explore other multi-keyword semantics over encrypted data, integrity check of rank order in search result and privacy guarantees in stronger threat model.

3.4. Fuzzy Keyword Search

In this method searching the encrypted text based cloud data using fuzzy keyword searching. The techniques which are used to obtain the fuzzy keyword search are wild-card based technique, gram based technique and symbol based Trie-traverse search scheme. In fuzzy keyword search it performs some steps that are shown in figure 3 below [7].

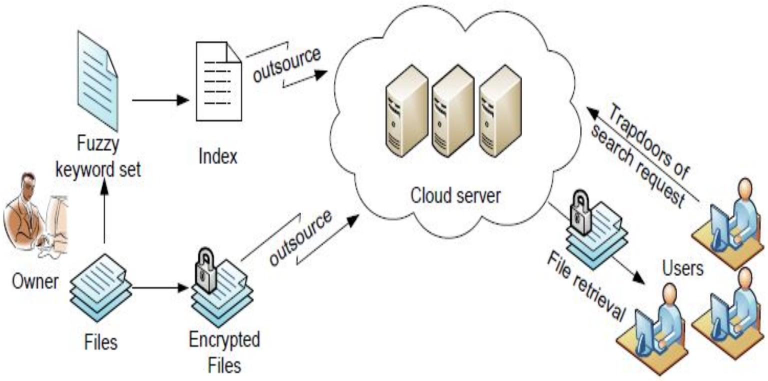

Figure 3. Architecture of Fuzzy Keyword Search

Step1: In first step the data owner is encrypted the data files using the symmetric key encryption technique and outsource over the cloud data server.

Step2: After encryption of data set data owner are generate the fuzzy keyword sets using fuzzy methods that describe above and on that fuzzy keyword set the index is generated and this index and encrypted files are outsource over cloud server.

Step3: In this step users are search the keyword using the trapdoors of search request. These trapdoors are passing search request keyword to fuzzy keyword search techniques.

Step4: After receive the request the fuzzy keyword search technique is generate the query that query is search the keyword in index and give the appropriate data files from data sets.

In next section explain the fuzzy search techniques in brief that are wild-card based search, Gram based search and Trie-Traverse search scheme.

3.4.1. Wild-Card based Search

Wild card based technique is used to solve the problem of edit operation at the same position keywords. In this edit operation problem is solved by edit distance with include substitution, deletion and insertion. First of all we generate Wildcard based fuzzy set.

In the above straightforward approach, all the variants of the keywords have to be listed even if an operation is performed at the same position. Based on the above observation, we proposed to use a wildcard to denote edit operations at the same position. The wildcard-based fuzzy set of wi with edit distance d is denoted as Swi,d = {S' wi,0, S'wi,1, …, S'wi,d}, where S' wi,t denotes the set of words wi' with t wildcards. Note each wildcard represents an edit operation on wi. For example, for the keyword CASTLE with the pre-set edit distance 1, its wildcard-based fuzzy keyword set can be constructed as S_{RIMZIM},1 = {RIMZIM, * RIMZIM, * IMZIM, *IMZIM, R*MZIM, …, RIMZI*M, RIMZI*, RIMZIM*}. The total number of variants on RIMZIM constructed in this way is only 13 + 1, instead of 13 × 26 + 1 as in the above exhaustive enumeration approach when the edit distance is set to be 1 [10].

3.4.2. Gram based Search

To generate the fuzzy keyword set, we use the concept of k-grams index, which is used to perform wildcard queries on plain-text files. K-grams is a sequence of k characters. For example, "any", "nyw", "ywa" and "way" are all the 3-grams of the word "anyway". We use the character $ to denote the beginning or the end of a word. Thus, the set of 3-grams generated is: "$an", "any", "nyw", "ywa", "way" and "ay$". In a k-grams index, our dictionary contains all the k-grams of every word in the collection. For each k-grams, we create a posting list of all the words in the collection that contain all the characters in the gram. For instance, in figure 4 the 3-gram "emp" would point to all the words such as employable and employee[8].

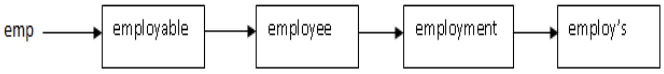

Figure 4. Gram based Search example

During the indexing process, our system first constructs the dictionary of all the k-grams in the collection. Posting list for each k-grams are then generated. All of these posting lists compose the k-gram index, that we called safe index. This predefined index will be used to generate fuzzy keyword set.

3.4.3. Symbol based Trie-traverse Search Scheme

To enhance the search efficiency, we now propose a symbol based trie-traverse search scheme, where a multiway tree is constructed for storing the fuzzy keyword set over a finite symbol set. The key idea behind this construction is that all trapdoors sharing a common prefix may have common nodes. The root is associated with an empty set and the symbols in a trapdoor can be recovered in a search from the root to the leaf that ends the trapdoor. All fuzzy words in the trie can be found by a depth first search.

In this section, consider a natural extension from the previous single user setting to multiuser setting, where a data owner stores a file collection on the cloud server and allows an arbitrary group of users to search over his file collection [7].

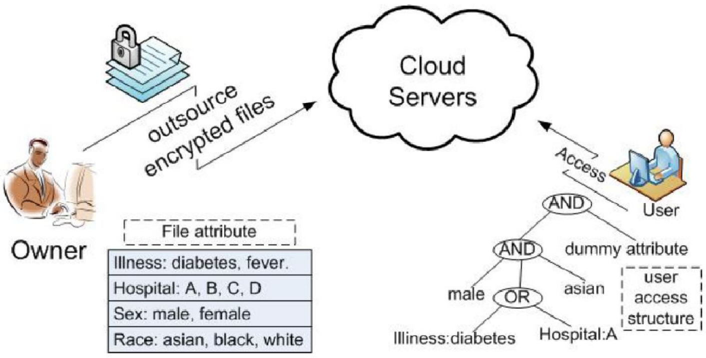

Figure 5. Symbol based Trie-Traverse Search example

IV. COMPARISONS OF ENCRYPTED CLOUD DATA SEARCHING TECHNIQUES

In above section discuss various searching techniques for encrypted cloud data which have its own advantage and disadvantage and also uses different algorithms for searching and encryption of data files. So the comparisons of these techniques are shown below table.

Table 1. Comparisons of Encrypted Cloud Data Searching Techniques

Sr. No.	Searching Technique	Process or Algorithm used	Advantage	Disadvantage
1	Boolean Symmetric Searchable Encryption (BSSE)	GramSchmidt, Orthogonalization process, Labeling and inner products searching of data	Linear Search, Randomized search, Focus on simple keyword matching	-Used only for searching Boolean queries. -Longer computation phase.
2	Ranked Keyword Search	Ranking Technique, Order preserving mapping technique	Highly Efficient, Search according rank and history of searching keywords	-Network Traffic occurs -Large amount of Post-Processing of encrypted files
3	Multi Keyword Ranked Search	Co-ordinate matching, Inner product similarity.	Eliminate traffic, Improve search accuracy, Privacy preserving multi keyword used	-Not suitable for large scale data -Not search single keyword
4	Fuzzy Keyword Search	Wildcard based technique, Gram based technique, Symbol based trie-traverse search scheme	Edit distance can be implemented, Highly Efficient, Increase search effectiveness	-Large storage complexities, Support only Boolean keyword search not support ranked search problem

CONCLUSION

In this study deep analysis is made on Encrypted Cloud Data Searching techniques which are used to retrieving the original data from the encrypted data on cloud servers. Many searchable techniques have been analysed based on single keyword search, multiple keyword search, boolean symmetric search and fuzzy search. The main goal is to search text based data on the encrypted cloud storage in privacy preserving manner and retrieve original data from large scale and distributed cloud storage.

REFERENCES

[1] M. Armbrust, et al., "A view of cloud computing", Communications of the ACM, vol. 53, no. 4,pp. 50-58, April 2010.

[2] "The NIST Definition of Cloud Computing". National Institute of Standards and Technology. Retrieved 24 July 2011.

[3] Tarik Moataz, Abdullatif Shikfa, "Boolean Symmetric Searchable Encryption", ASIA CCS '13 Proceedings of the 8th ACM SIGSAC symposium on Information computer and communications security, .pp. 265-276, NY, USA , 2013.

[4] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data", International conference on distributed computing systems, 2010.

[5] Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Louy, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" Proc. IEEE Symp. Security and privacy, 2013.

[6] Ming Li, Shucheng Yu, Ning Cao and Wenjing Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing", 31st International conference on distributed computing systems, 2011.

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

[8] Wei Zhou, Lixi Liu, He Jing, Chi Zhang, Shaowen Yao, Shipu Wang, "K-Gram Based Fuzzy Keyword Search over Encrypted Cloud Computing", Journal of Software Engineering and Applications, vol. 6,page. 29-32, 2013.

[9] Cengiz Orencik, Erkay Savas, "Efficient and secure Ranked Multi-Keyword Search on Encrypted cloud data", ACM ISSN no 978-1-4503-1143-4,2012.

[10] Saeed Sedghi, Peter van Liesdonk, Svetla Nikova, Pieter Hartel, and Willem Jonker, "Searching Keyword with Wildcards on Encrypted Data".