Enhanced Protection And Exploring Graded Keyword Search On Off-Shoring Cloud Data

M. Vinay Kumar*, M. Sailaja**, N. Krishna Vardhan**

**M.Tech(C.S.E), KTRMC Engg & Tech,Kondair,*

***M.Tech(C.S.E) ,IPCET(W),Kurnool*

***Associate Professor, KTRMC Engg & Tech,Kondair,*

Abstract

Now a days cloud computing plays an important role in data service outsourcing. However, to provide data security, the data in the cloud server has to be encrypted before sent to the commercial public cloud, which makes data utilization service, a very challenging task. The graded searchable methods provides a secured search over encrypted data through keywords, In this paper, we define and solve the problem of secure graded keyword search over encrypted cloud data. Graded search greatly enhances system usability by enabling search result relevance ranking instead of sending unwanted results, and further ensures the file retrieval accuracy. Order-preserving symmetric encryption (OPE) is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plain texts. This method will be able to provide efficient server-side grading without losing keyword privacy. We can say that the proposed solution enjoys a strong security compared to previous searchable schemes, while correctly realizing the goal of graded keyword search.

Index Terms—Cloud computing, Graded search, Searchable encryption, Order-Preserving symmetric encryption, secured data.

1. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network. Cloud computing entrusts remote services with a user's data, software and computation.

Cloud computing is an emerging technology which helps as an utility, through which clients are going to store their data in the cloud server and using applications from a set of computing resources. Here sensitive data is going to be stored in the server. Sometimes the hackers may try to access the cloud data, so the data has to be encrypted before outsourced to achieve privacy. The encryption techniques increase the data utilization from a large amount of data. To retrieve data files we introduced “Keyword search mechanism”. By this mechanism the users are going to retrieve the data files of their interest. In traditional search, encryption techniques the users are going to search data by using keywords without decrypting it, they support only Boolean keyword search only. In cloud computing graded keyword search enhances the system usability by displaying the matching files with the help of significant score. To achieve the security and usability we introduce advanced cryptographic and Information Retrieval (IR) techniques, and using one-to-many order preserving symmetric encryption.

Basically there are three types of public cloud Services:

Infrastructure as a service (IaaS): This is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components

Platform as a service (PaaS): Platform as a Service (PaaS) is a way to rent hardware, operating systems, storage and network capacity over the Internet. The service delivery model allows the customer to rent virtualized servers and associated services for running existing applications or developing and testing new ones.

Software as a service (SaaS): Software as a Service (SaaS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.

Cloud computing is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In cloud computing, the word cloud is used as a metaphor for "the Internet," so the phrase cloud computing means "a type of Internet-based computing," where different services such as servers, storage and applications are delivered to an organization's computers and devices through the Internet. Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are harnesses to solve problems to intensive for any stand-alone machine**.**

2. Background and Related Work

Now-a-days cloud servers get to store large amount of files. Here select and processing the files is the main problem. Whenever large number of files is available in cloud server under encryption some problems are generated. Totally all files are not encrypted. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. That's here there is no sufficient privacy and security in outsourcing. Some unauthorized users are entering and corrupt the content of information.

Previously user can selects the files in the form of a plain text files. This is ailing under access the files. There is no perfect decryption technique to access the files of representation process. Here we introduce encryption based secure keyword searching mechanism. It can provide efficient solution for accessing the data. It is a good usability to display the effective matching details files. These matching files are extracted with relevance score. This kind of matching files are retrieved with efficient mechanism. It can provide the results with guaranteed mechanism. All the files are collected with encryption format. All encrypted files are given weight in implementation process. These kinds of approaches show the better result in implementation.

Architecture:

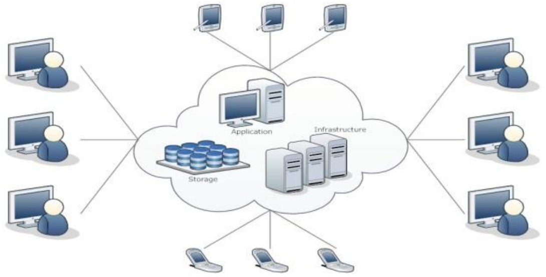


Fig.1: Architecture of cloud server

Advantages:

1. It can retrieve the results with less communication overhead.
2. It can provide the results with effective retrieval accuracy.

3. It can provide effective privacy and security application.

In the architecture we have three entities

1. Data owner: Data owner having collection of data files that he wants to outsource into the cloud server in encrypted format, this will increase effective data utilization.

2. Data user: When the data user wants to search the required files he enters a keyword in a secret form.

3. **Cloud server**: It is the place where a pool of data files and different applications can store.

The search result is displayed according to relevance score which improves file retrieval accuracy .In information retrieval process we maintain an inverted index to represent file ID's and relevance scores.

Inverted index maintains a list of mappings which corresponds to set of files that contain the keywords .*Ranking functions* are used to calculate relevance score of matching files for a given keyword .To calculate the relevance score we use a statistical measurement i.e. TF*IDF ,where TF-term frequency and IDF –inverse document frequency.TF is number of times a given keyword find within a file and IDF is calculated by dividing the total number of files by the number of files containing the keyword.

$$score(Q,Fd) = \sum_{t \epsilon Q} \frac{1}{Fd} . (1 + \ln fd,t). \ln\, (1 + \frac{N}{ft})$$

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

3. Graded Keyword Search

Here we are going to implement graded searchable symmetric encryption based on existing searchable symmetric encryption schemes. Searchable symmetric encryption allows data owner to outsource data in encrypted manner. In this data outsourcing, there may be loss of data by less security. To overcome this we have graded searchable symmetric encryption mechanisum(GSSE). This algorithm is based on four individual algorithms, they are "KeyGen, BuildIndex, TrapdoorGen and SearchIndex" which contributed the two step process.

Setup-Here data owner initialized public and private parameters by KeyGen algorithm and pre-process the data file C, to generate inverted index based on random keywords. The owner encrypts data file C and makes an index with relevance scores.

Retrieval-The user enters a keyword and makes a trapdoor by using TrapdoorGen, in the server. Then the cloud server will find the matching files, and displays file IDs based on Search Index algorithm.

In preserving symmetric encryption schema, numerical ordering of plaintext is preserved. In this, random order preserving will be considered. Thus high rate of information leakage is present. To avoid this we go for one-to-many Order preserving mapping, which preserves the data security and plaintext order. It incorporates random plaintext-to-bucket mapping of OPSE, and implemented in below Algorithm 1. TapeGen( ) is a random coin generator, HYGEINV( ) is used instance of HGD( ). By the use of the OPM algorithm, server can rank the files efficiently.

Algorithm 1: One-to-many Order-preserving Mapping-OPM

```
procedure OPMK(D,R, m,id(F))
While |D|! = 1 do
    { D,R } ← BinarySearch(K,D,R,m);
end while
coin $\xleftarrow{R}$ TapeGen(K,(D,R,1||m,id(F)));
c $\xleftarrow{coin}$ R;
```

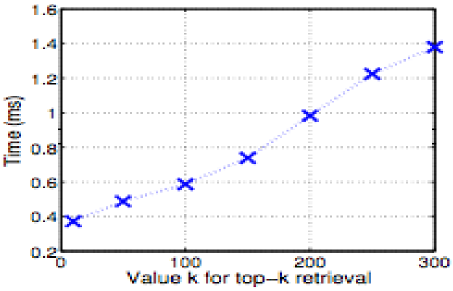
```
return c;
end procedure
procedure
BinarySearch(K,D,R,m);
$M \leftarrow D; N \leftarrow R;$
$d \leftarrow \min(D)-1;\ r \leftarrow \min(R)-1;$
$y \leftarrow r + [N/2];$
coin $\xleftarrow{R}$ TapeGen(K,(D,R,0||y));
$x \xleftarrow{R} d$ + HYGEINV(coin,M,N,y-r);
if $m \le x$ then
    $D \leftarrow \{d + 1, \ldots\ldots., x\};$
    $R \leftarrow \{r + 1, \ldots\ldots\ldots, y\};$
else
    $D \leftarrow \{x + 1, \ldots\ldots\ldots, d + M\};$
    $R \leftarrow \{y + 1, \ldots\ldots\ldots, r + N\};$
end if
return { D,R };
end procedure
```

4. Performance Analysis

Performance of our proposed system is calculated based on the effectiveness and efficiency. This includes time and cost for searching the keywords. Effectiveness of one-to-many order preserving mapping is determined by relevance score R. search time includes fetching the posting list in the index, decrypting and rank-ordering each entries. We considered to retrieve top-K retrievals. Fig.1 list our search time cost against the value of k increases, for the index as in table.1.

Number of files	Per.keyword list size	Per.list build time
1000	12.414Kb	5.44.S

Table.1. Per keyword index construction overhead for 1000 RFC files.



5. Conclusion

In this paper we solve the security problems which may occur during outsourcing the data. In previously, there is no security for the data because for leakage. Any third party may hack the leaked data and also there is a burden in accessing the data items from different vast retrieved searches. To avoid this by implementing an OPSE (ordered preserving symmetric encryption) with GSSE mechanism. Present solution is best for effective data utilization and provide security for outsourced cloud data compared to previous one. Hence accuracy is also increased.

6. Reference

1. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data,” in Proc. of ICDCS’10, 2010.
2. D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. of IEEE Symposium on Security and Privacy’00, 2000.
3. A.Boldyreva, N. Chenette, Y. Lee, and A. O’Neill, “Order preserving symmetric encryption,” in Proc. of Eurocrypt’09, volume 5479 of LNCS. Springer, 2009.
4. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” in Proc.of INFOCOM’11, 2011.
5. C.Wang, Q.Wang, K.Ren, and W.Lou, “Toward Secure and Dependable Storage Services in Cloud Computing”,IEEE Transactions on service Computing