www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 11814-11819

Secure Scheme over private Cloud data supporting Semantic Relation

Sayli Bahekar[1], Jyoti NIkam[2]

[1] Dept. of Computer Engineering
MAEER'S MIT College of Engineering, Pune
Savitribai Phule Pune University
saylisbahekar@gmail.com

[1] Dept. of Computer Engineering
MAEER'S MIT College of Engineering, Pune
Savitribai Phule Pune University
Jnikam66@gmail.com

Abstract: ***Cloud becomes a necessity in the real IT world because of its appealing features It becomes a important tool to handle big data. but encryption makes it difficult to perform basic functionalities. Search scheme is cloud supports on exact keyword matching techniques. The synonym or semantic of the keyword is not considered. Hence a scheme is proposed which considers the semantics also of the submitted keyword using data structure like SRL and retrieves files in order of their relevance score.***

Keywords: Semantic, Cloud Data, Semantic Relationship Library (SRL), Inverted Index, Relevance Score

.

1. Introduction

Today's IT world has grown by heaps and bounds. A large amount of data is being generated on daily basis. This data needs to be managed in a proper way considering the confidentiality and security of the data. Cloud has evolved as a trust worthy and convenient infrastructure in today's IT era. Cloud becomes an important tool in managing such large amount of data .It provides facilities for managing data in a cost efficient and secure manner. Security of the data is assured as the data is encrypted and then stored on the cloud. Encryption may guarantee the security of the data but it makes many necessary and basic operations like search impossible to be performed.

Hence new searchable encryption schemes were developed. Using these schemes performing search operations on encrypted data was possible. Hence the problem of searching encrypted data was solved. But this is not a very efficient way to search because it supports only exact keyword matching, i.e only files containing the exact keyword which was submitted are retrieved. This results in inefficient output as the files of particular interest may not be retrieved and it may also lead to cost inefficiency in the pay per use cloud models. The solution to this is that the semantics of the query keyword should be considered along with the submitted query keyword.

In this paper, we propose a search scheme which supports secure semantic based search and similarity ranking. In the proposed scheme, a metadata set is constructed for each file. The metadata is being uploaded to the cloud. Using the metadata submitted by the owner, the cloud builds a structure called SRL (Semantic Relationship Library). SRL contains the strength of relationship between two words present in the file. Using SRL semantics of the keyword are taken into consideration. And files containing both the keyword and the semantically related words are retrieved.

2. Technology overview

A. Java

Java is an object oriented programming language and is platform independent i.e. compiled code can run on any machine supporting java. It follows WORA way i.e. "Write Once, Run Anywhere". Java was released in 1995 as a core part of Sun Microsystems' Java platform, developed by James Gosling at Sun Microsystems (which has since been acquired by Oracle Corporation).

B. NetBeans

NetBeans is a software development platform written in Java. The NetBeans Platform applications run under modules. Netbeans supports other languages as well for example PHP,C,C++etc. it provides a easy way for developers to write code and develop applications .NetBeans is cross-platform and runs on Microsoft Windows, Mac OS X, Linux, Solaris and other platforms supporting a compatible JVM.

C. MySQL

MySQL is a most widely used free open source relational database management system (RDBMS).The official MySQL

Workbench is a free integrated environment developed that makes easy users to graphically administer databases and visually design database structures.

3. Proposed scheme

A. SYSTEM ARCHITECTURE

There are 3 different elements in the proposed system model: 1.Data Owner, 2.Authorised Data User, 3.Private Cloud Server as shown in figure 1.

Data owner is the entity who owns the files and uploads them to the cloud. A corresponding metadata is build for every file and the metadata is also uploaded to the cloud.

Data User is the file retriever who wants to search a file of particular interest. Data User submits his interested keyword to the cloud. The cloud server initially constructs the SRL.

Upon receiving the keyword from the authorized user, the cloud server first expands the query keyword using SRL. Then the cloud server searches the index and returns matched files to the user in descending order of relevance score.

The user then decrypts the received files and gets the required data.

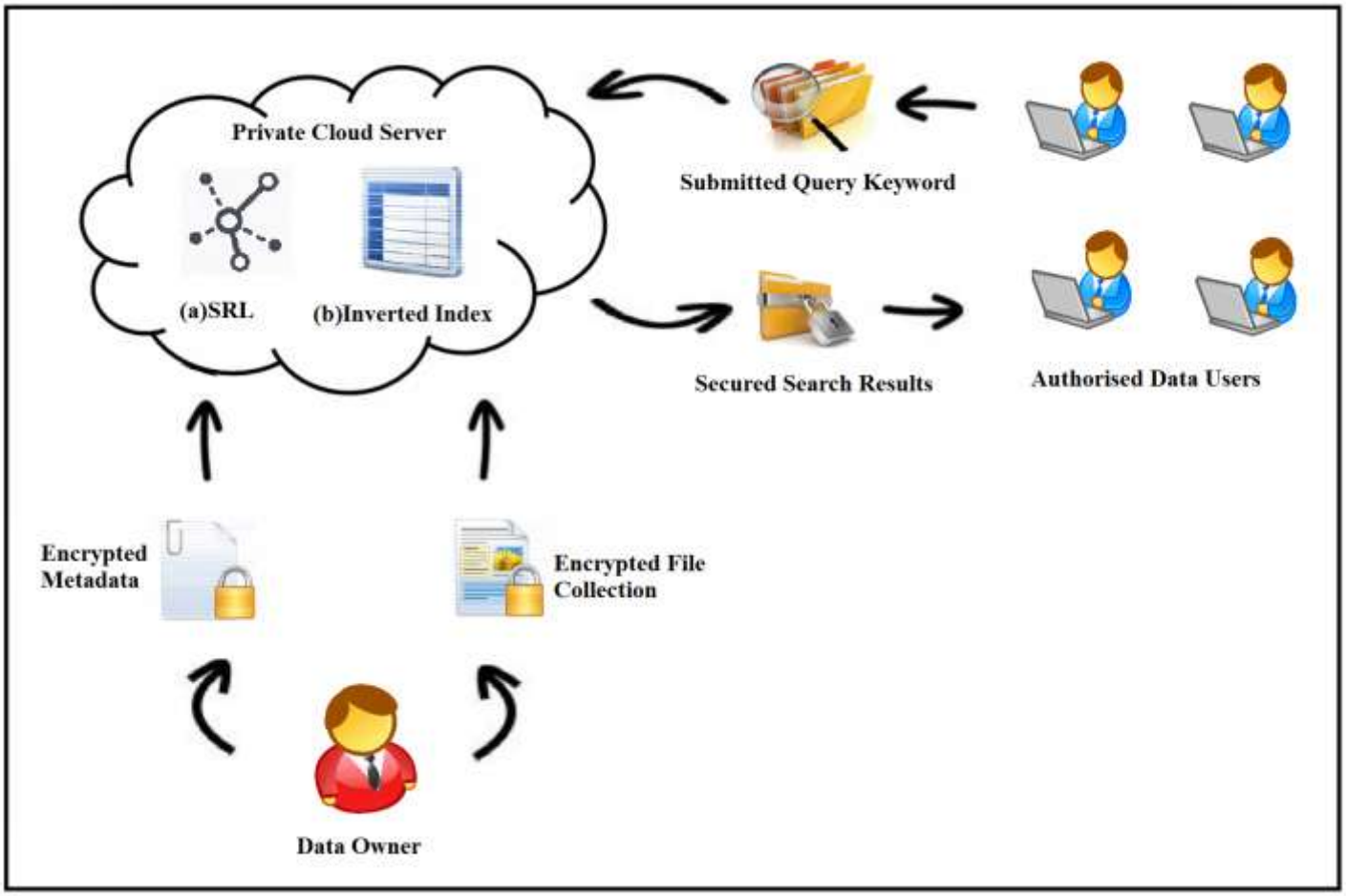

Fig1. *System Architecture for semantic expansion based search over encrypted private cloud data*

B. IMPLEMENTATION

The scheme consists of two phases

1. Setup Phase

2. Retrieval Phase

1. Setup phase

This phase consists of major operations like metadata creation, SRL generation etc.

a. Encryption and Key Sharing

The data owner initiates the scheme by generating the security keys to be shared among the authorized users. It takes the security parameters as inputs and generates random keys . Finally it outputs secret keys set used for later encryption, such as trapdoor generation and relevance score encryption.

b. Metadata Creation

Files are uploaded by the owner in encrypted form. The A piece of file-metadata is constructed for each file. The file-metadata consists of the file ID, keywords, and the relevance scores of keywords in response to the file. Consider the document to be uploaded to cloud.

Document Name: Cloud.txt

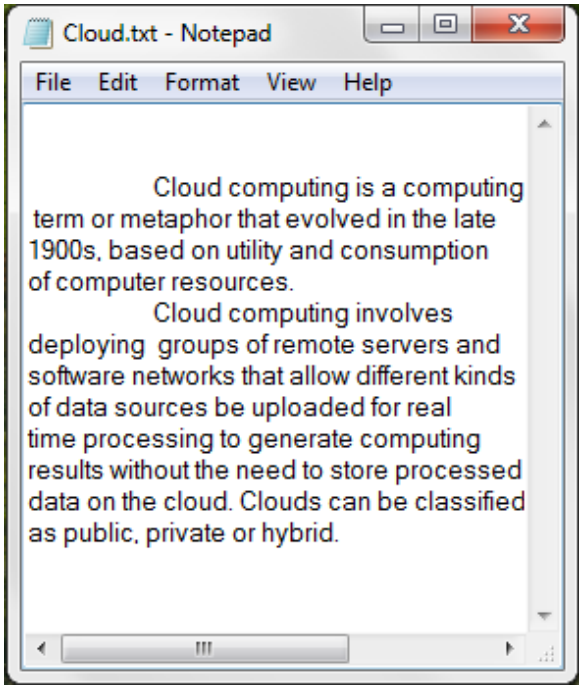

Fig.2. Input Document

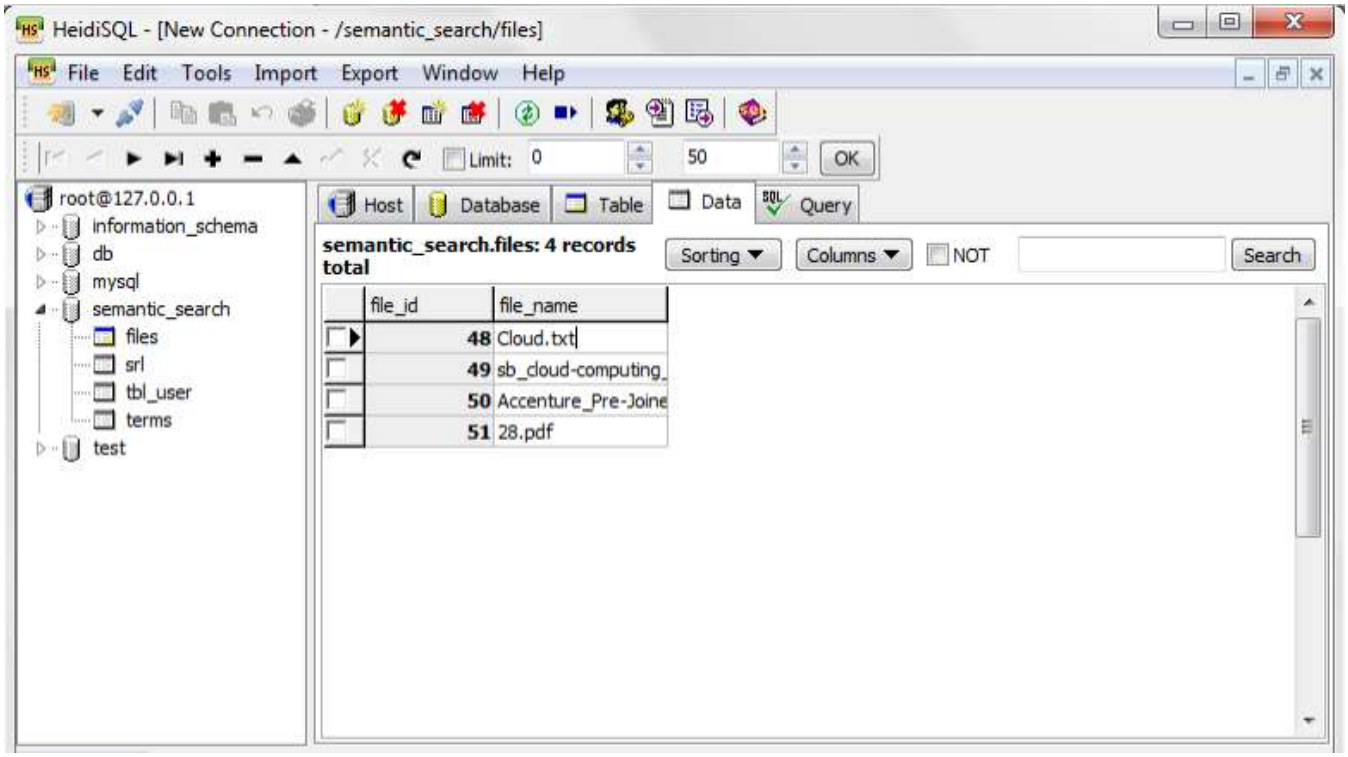

Fig.3 Uploaded Files

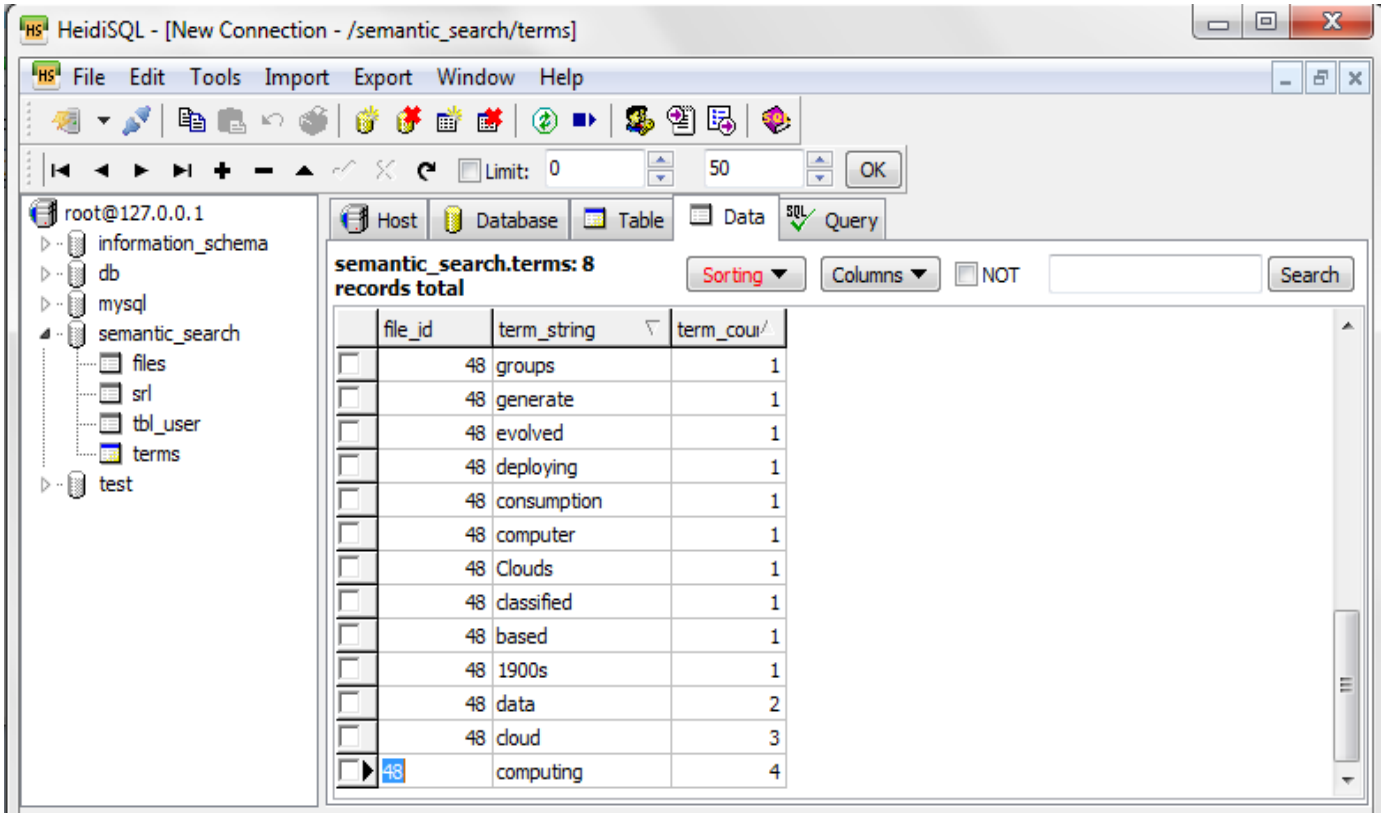

Fig.4 Metadata

Figure 4 shows the metadata created for the text file cloud.txt wherein the stop words are omitted and the count of remaining words in the file is maintained as shown in the figure.

c. SRL(Semantic Relationship Library) generation

SRL is a data structure which stores the probability of connection between two words on the basis of their occurrence in the file.

The following formula is used to calculate the semantic relation between two words.

$$I(x, y) = \log_2 \frac{P(x,y)}{p(x)p(y)}$$

Here, P(x,y) is the probability of occurrence of x and y together, p(x) and p(y) are the probabilities of individual occurrence of x and y in the file collection respectively.

Lets see how the formula works for the document Cloud.txt. .Consider the 2 words 'cloud ' and 'computing'. In our file

P(x,y) = 2/n

P(x) = 3/n

P(y)= 4/n

Where n = 35 is total no of terms in the file.

I(x,y) = 0.765.

Similarly the semantic relation for other words is calculated.

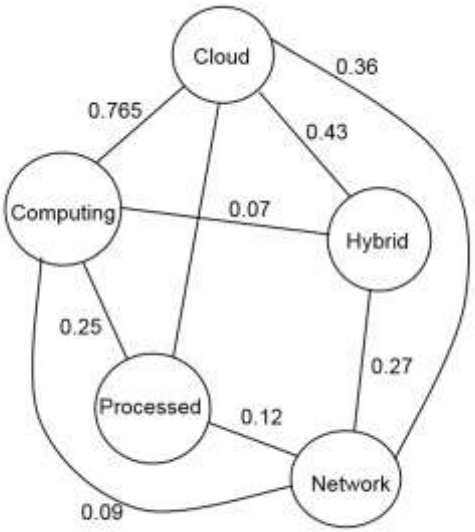

Fig 4. Semantic Relationship Library

2. *Retrieval Phase*

The actual task of searching and data retrieval is performed in this phase.

The authorized user submits a query keyword. The keyword is expanded on basis of the SRL and considering its synonyms.

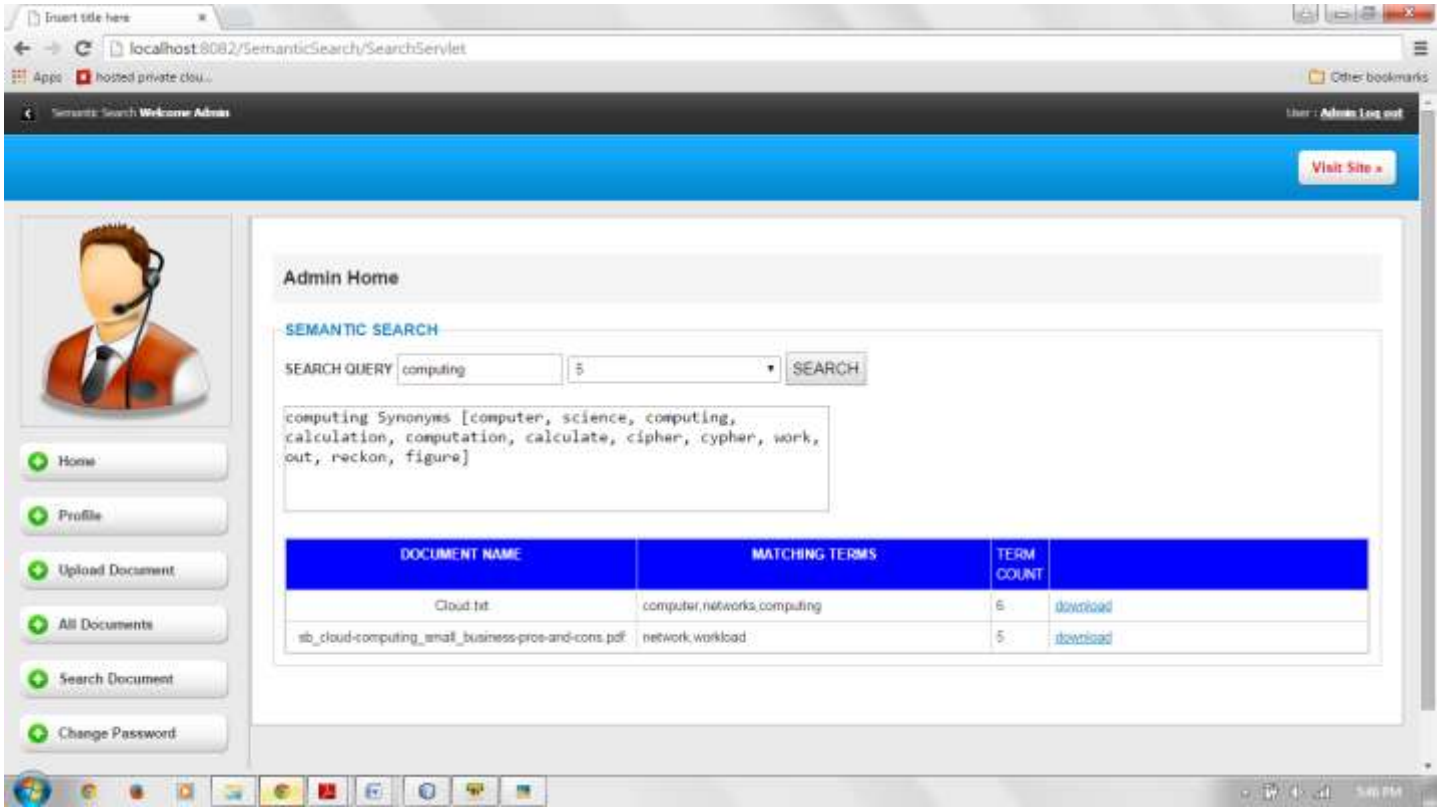

Fig 5.Retrieved Files

Conclusion

The proposed scheme implements cost effective search scheme over encrypted private cloud data in this pay-as-per-use cloud paradigm. It not only returns the exactly matched files but also the files which include terms which are semantically related to the query keyword. Thus, it reduces overhead on the data user.

Hence, in this paper an attempt is made to solve the problem of efficient rank search over encrypted private cloud data. Here, we strengthen the security factor by using asymmetric encryption algorithm. Thus, the proposed scheme is secure and privacy preserving while it correctly realizes the goal of ranked keyword search.

References

[1] Xia et al.: Secure semantic expansion based search over encrypted cloud data supporting similarity ranking. Journal of Cloud Computing: Advances, Systems and Applications 2014 3:8.

[2] Prof C. R. Barde, "Secured Multiple-keyword search over encrypted Cloud data," International Journal of Emerging Technology and Advanced Engineering, Volume 4,Issue 2,Feb 14.

[3] Xingming Sun, Yanling Zhu, Zhihua Xia and Lihong Chen, "Privacy preserving keyword based Semantic Search over Encrypted cloud data," International journal of Security and its Applications, Volume 8, No.2 (2014).

[4] Zhangjie Fu, Xingming Sun, Senior, 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.

[5] Song DX, Wagner D, Perrig A, "Practical techniques for searches on encrypted data," Proceedings of IEEE Symposium on Security and Privacy.IEEE, Berkeley, California, pp 44–55(2000)

[6] Cao N, Wang C, Li M, Ren K, Lou W, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," Proceedings of IEEE INFOCOM. IEEE, Shanghai, China, pp 829–837(2011)

[7] Yang C, Zhang W, Xu J, Xu J, Yu N "A Fast Privacy-Preserving Multi-keyword Search Scheme on Cloud Data,"

International Conference on Cloud and Service Computing (CSC). IEEE, Shanghai, China, pp 104–110(2012)

[8] Stefanov E, Papamanthou C, Shi E, "Practical Dynamic Searchable Encryption with Small Leakage," NDSS '14, San Diego, CA, USA

[9] Liu C, Zhu L, Li L, Tan Y, " Fuzzy keyword search on encrypted cloud storage data with small index," IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE, Beijing, China, pp 269–273(2014)

[10] Wang C, Ren K, Yu S, "Urs KMR Achieving usable and privacy-assured similarity search over outsourced cloud dat," Proceedings of IEEE INFOCOM. IEEE, Orlando, Florida, USA, pp 451–459(2012)

[11] Li J, Wang Q, Wang C, Cao N, Ren K, Lou W, "Fuzzy keyword search over encrypted data in cloud computing," Proceedings of IEEE INFOCOM. IEEE, San Diego, CA, USA, pp 1–5(2014)

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 11814-11819