ISSN: 0976-3104
SPECIAL ISSUE: (Emerging Technologies in Networking and Security (ETNS)
Sathya et al

ARTICLE OPEN ACCESS

A SURVEY OF MULTI KEYWORD SEARCH OVER THE ENCRYPTED DATA IN CLOUD

S. Sathya*, J. Gayathri, D. Radhika
Dept of Computer Science and Engineering, Vivekanandha College of Engineering for women, Namakkal, T. N., INDIA

ABSTRACT

***Aim*:** *Cloud computing [11] is a tremendous growth in every years and it can be utility the computing and large storage capability to the public users. The data owner can store the data in the cloud server is called data outsourcing and then the cloud data access for public users through the cloud server. The outsourced data are contains sensitive privacy information and it can be encrypted before uploaded to the cloud server and then the search user can access to the data through the cloud server is some difficulty of searching over the encrypted data in cloud. In this paper address this problem by developed the fine-grained multi-keyword search scheme over encrypted data in the cloud. There are three contribution of this paper. First one is, to provided relevance scores and preference factors upon keywords which enabled the precise keyword search. Second one is, to developed a complicated logic search the mixed AND, OR and NO operations of multi-keyword search scheme. And finally, auxiliary employ the classified sub-dictionaries technique to accomplish the index building, trapdoor generating and query. By using this experiments to the real-world dataset, so easily retrieve the result from dataset.*

***Corresponding author: Email:** sathyashree.mecse@gmail.com; **Tel.:** +91 9655883123

Published on: 2nd -December-2016

KEY WORDS

Fine grained multi keyword encryption, searchable encryption, and index and trapdoor generation.

INTRODUCTION

In cloud computing [11], data owners are moving their data to the cloud server to access the multiple public users. And then the cloud data open for public usage through the cloud server. The data owner uploaded the file to the cloud before that it can be encrypted is called as outsourcing data .The benefits of this kind of data outsourcing such as low-cost and flexible data access. It can also cause some privacy problems since the outsourced data had some sensitive information. It is necessary to encrypt the sensitive data before updating them to the cloud server. Moreover, data owners would be like to allow multiple search users to search over the encrypted data while access the control policies.

Multi Keyword Search

Multi keyword search technique [17] allows users to selectively retrieve files of interest and has been applied in plaintext search scenarios. Unfortunately, data encryption, which restricts users able to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Ranked search [17] greatly improves system usability by normal matching files in a ranked order regarding to certain relevance criteria.

Considering three different framework entities in **Figure-1**, Such as Data owner, Search user, and cloud server. Data owner have a collection of data documents to be sent to cloud server in the encrypted format. To activate the searching efficient over encrypted data, data owner, before sending data, data owner can encrypt the data and then outsource both the index and the encrypted document collection sent to cloud server. It can be search the document, an authorized user require a corresponding trapdoor through search mechanisms to receiving from data users, cloud server is take over to search the index and returns the matched set of encrypted documents.

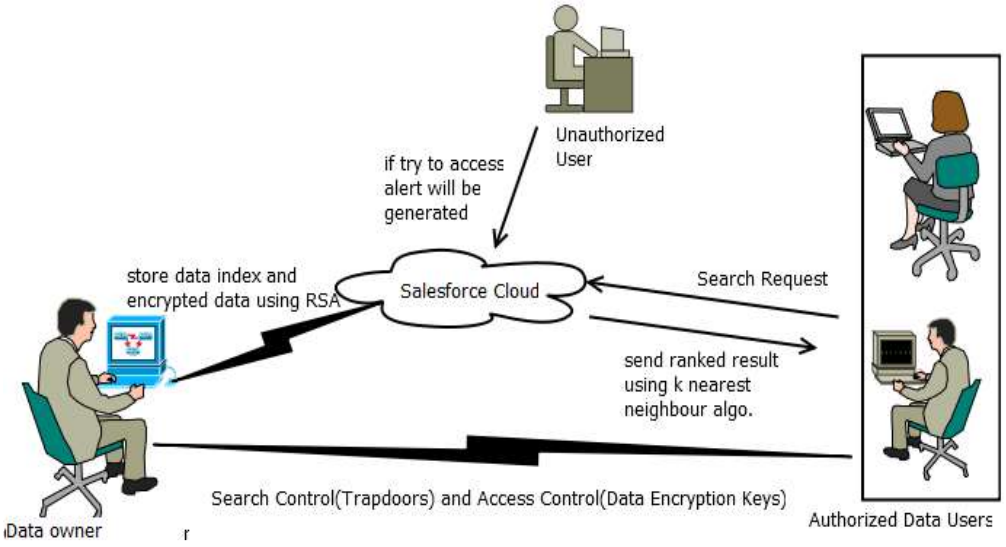


Fig. 1: An example of search over encrypted cloud data in sales force cloud

To improve the search efficiency, the searchable encryption schemes must support multi-keyword search [12]. And search users would like the cloud server to return results in a specific order, so they can achieve the more relevant results quickly. Furthermore, to make the searchable encryption schemes suitable for more practical situations, such as the situation that the data is contributed from many data owners and can be searched by many search users. And this scheme to support search authorization, It means the cloud server would only return the authorized results to the search users.

System Model

To provides the relevance scores and the preference factors of keywords for searchable encryption over the encrypted data. The relevance scores of keywords can enable more precise returned results and the preference factors of keywords denote the importance of keywords in the search keyword set specified by search users. To realize the AND, OR and NO operations in the multi-keyword search for searchable encryption over the encrypted cloud data and it compared with schemes in [16], [3] and [4], the proposed scheme can achieve more broad functionality and lower query complication.

Data Owner

The data owner outsources the data to the cloud for reliable data access to the search users. To protect that data privacy and it means unauthorized person doesn't modify the content of original data. And the data owner encrypts the outsourced data through symmetric encryption. To improve the search efficiency, the data owner generates some keywords for each outsourced data. The corresponding index is then created according to the keywords and a secret key.

The data owner has a collection documents F= {f1, f2, f3……fn}. And the data owner wants to outsource the data to the cloud in encrypted format at the same time he wants to keep the capability to search on them for effective utilization. What is all data owner's process? The Data owner first builds a secure searchable keyword index I from document collection Fand then generates an encrypted document collection C for Afterwards, the data owner outsources the encrypted collection C and the secure keyword index I to the cloud server. Securely distributes the key information of trapdoor generation (including IDF values) and document decryption to the authorized data users. Data owner is responsible for the update operation of his documents stored in the cloud server while updating. The data owner generates the update information and sends it to the server.

Cloud server

Cloud server stores encrypted documents collection C and encrypted searchable keyword indexes I that are received from the data owner. Data owner provides data access and search services to search users upon receiving the trapdoor TD from the data user, the cloud server excites the search over the keyword index I. Finally returns the corresponding collection of collection of matching documents based on certain operations such as AND, OR and NO operation of keywords.

Search user

Search users are authorized one's to access the document from cloud server with following three steps. First step is, the search user receives both the secret key and symmetric key from the data owner. Second step is, according to the search keywords, the search user using the secret key to generate trapdoor according to search control mechanisms and sends it to the cloud server. Finally, the search user receives the matching document collection from the cloud server and decrypts the encrypted data by using symmetric key.

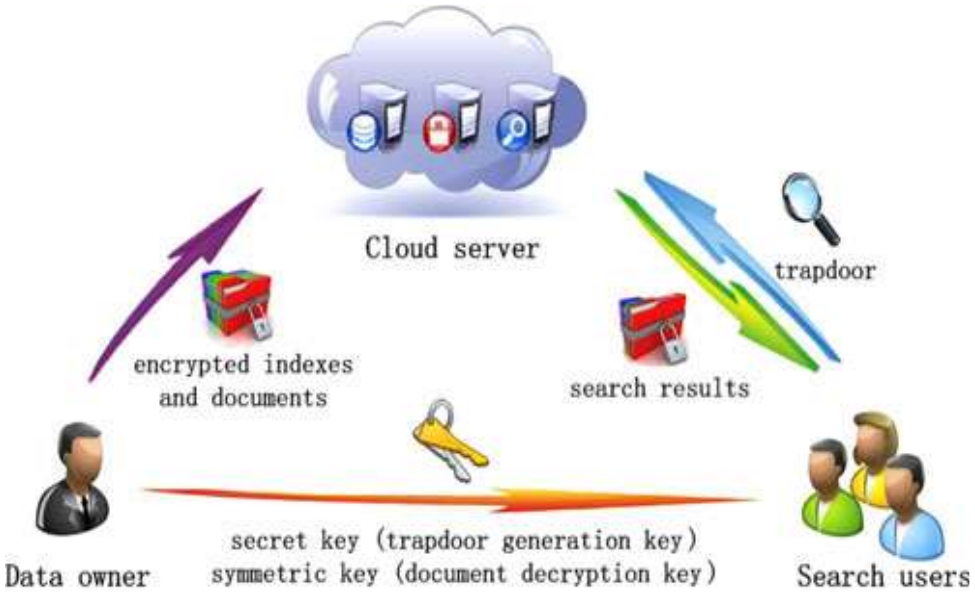


Fig. 2: Key generate of both data owner and search user over the cloud

LITERATURE SURVEY

It deals with the information by using multi keyword search scheme [2] to identify the original encrypted data and its major role for the user to collect original data.

In [1]authors D. X. Song, D. Wagner, and A. Perrigin the paper "**A Practical and Secure Multi-Keyword Search Method over Encrypted Cloud Data"** presented a well-organized privacy-preserving multi keywords search method over encrypted cloud data by used minhash functions. A multi-keyword search technique can be combining of several keywords in a single query. By dint of increasing the search constraints and also fetched the most relevant items returned to the search user. Since a multi-keyword search method that returns the matching encrypted data in a ranked ordered manner and it can hold three steps is like, First step is, to present a minhash based privacy-preserving multi keyword search method that provides high precision rates. Second step is, to provides the security requirements and formally prove that the proposed method satisfies adaptive semantically security. Third step, to use a ranking method depends on term frequencies and inverse document frequencies (tf-idf) of keywords and demonstrate that it is efficient and effective by providing the implementation results. One of the main advantages of this paper is the enable of multi keyword search in a single query.

In [2] authors H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen in this paper **"Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage"** describes the searchable

encryption for multi-keyword ranked search over the storage data. There are large numbers of outsourced documents (data) in the cloud. In this paper utilize the relevance score and *k*-nearest neighbor techniques to develop an effective multi-keyword search scheme and it can be returned to the ranked search results based on the accuracy. Security analysis established that scheme can achieve confidentiality of documents and index, trapdoor privacy and concealing access pattern of the search user. Finally, using extensive simulations can achieve much improved efficiency in terms of search functionality and search time.

In [3] authors BoyangWang,YantianHou, Ming Li in this paper "**Practical and Secure Nearest Neighbor Search on Encrypted Large-Scale Data"** presented a new searchable scheme which can effectively and securely enable nearest neighbor search over encrypted data in clouds. Particularly modify the search algorithm of nearest neighbors with tree structures (R-trees), where the modified algorithm adapts to lightweight cryptographic primitives (Order-Preserving Encryption) without affecting the linear search complication. Moreover, which can be used for secure k-nearest neighbor search and it is compatible with another similar tree structures. In this paper results on Amazon EC2 show that scheme is extremely practical over massive datasets.

In [4] authors H. Li, D. Liu, K. Jia, and X. Lin in this paper "**Achieving Authorized and Ranked Multi-keyword Search over Encrypted Cloud Data"** presented an authorized and ranked multi-keyword search scheme (ARMS) over encrypted cloud data by included the cipher text policy attribute-based encryption(CP-ABE) and SSE techniques. To motivate the research on the searchable encryption technique, it can be allows the search user to search over the encrypted data in cloud. In this paper, particularly focus the symmetric searchable encryption (SSE) techniques. However, they do not conceive the search authorization problem that requires the cloud server to return the search results to authorized users. Security analysis demonstrates that the ARMS scheme can achieve confidentiality of documents collusion resistance and trapdoor unlinkability.

In [5] authors Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen in this paper "**Secure Dynamic Searchable Symmetric Encryption with Constant Document Update Cost"** presented to leverage the secure k-nearest neighbor to proposed a secure dynamic searchable symmetric encryption scheme. In this scheme can achieve two important security features is backward privacy and forward privacy which are very challenging in Dynamic Searchable Symmetric Encryption(DSSE) area. And also to evaluate the performance of proposed scheme compared with other DSSE schemes. The comparison results demonstrate the efficiency of the scheme in terms of the storage, search and update complexity.

In [6] authors Neelam S. Khan,Dr. C. Rama Krishna,Anu Khuranain this paper "**Secure Ranked Fuzzy Multi-Keyword Search over Outsourced Encrypted Cloud Data"** to solve the problem of effective Secure Ranked Fuzzy Multi-Keyword Search over Outsourced Encrypted Cloud Data (RFMS).RFMS improves user searching experience by returning the matching files when users input query to retrieve any exactly matches the predefined keyword dictionary or closest possible keywords in the dictionary depended on similarity semantics when exact match fails. Information discovery has been made effective by searching with multiple keywords with ranking so as to eliminate false positives. Keyword dictionary has been made dynamically. Overhead of updating the dictionary when new files need to be uploaded has been minimized. And also by using one-to-many mapping between plaintext and cipher text, the method guarantees security.

In [7] authors J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li in this paper "**Towards Secure Multi-Keyword Top-k Retrieval over Encrypted Cloud Data"** an addressing data privacy problems by using searchable symmetric encryption (SSE). For the first time to formulated the privacy issues from the aspect of similarity relevance and schemes robustness. To observe that server-side ranking depended on order-preserving encryption (OPE) unavoidably leaks data privacy. To eliminate the leakage, to provides a two-round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval. In TRSE, service a vector space model and homomorphic encryption. The vector space model helps to deliver sufficient search accuracy and the homomorphic encryption enables users to include in the ranking and it's done on the server side by operations on cipher text. The result, information leak can be eliminated and data security is ensured.

In [8] authors W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li in this paper "**Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking"** it presents the privacy-preserving multi-keyword text search (MTS) scheme used similarity-based ranking over encrypted data in cloud. For supporting the multi-keyword search and search result ranking to construct the search index based on term frequency and the vector space model with cosine similarity measure to suggest higher search result

accuracy. To increase the search efficiency for developed the tree-based index structure and different adaptive methods for multi-dimensional (MD) algorithm .They enhanced the search privacy scheme to construct the two secure index schemes to meet the difficult privacy requirements under strong threat models is called cipher text and background model. Advantage of this scheme depends upon the index tree structure to enable authenticity check over the returned search results.

In [9] author Zhihua Xia in this paper "**A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data"** describes a secure multi-keyword ranked search scheme over encrypted cloud data, which consecutively supports dynamic update operations like deletion and insertion in the documents. Particularly the vector space model and the broadly used TF×IDF model are combined in the index construction and query generation. In this paper construct a special tree-based index structure and suggest a Greedy Depth-first Search algorithm to provide well-organized multi-keyword ranked search. The secure kNN algorithm is exploited to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. Advantage of this paper is to use the special tree-based index structure, the proposed scheme can succeed sub-linear search time and deal with the deletion and insertion of documents flexible in encrypted cloud.

In [10] authors B. Zhang and F. Zhang in this paper "**An efficient public key encryption with conjunctive-subset keywords search"** presented a Public Key Encryption Keyword Search (PEKS) scheme with Subset keywords search and it means that the receiver could query the subset keywords embedded in the cipher text. In this paper to solve the problem of conjunctive with subset keywords search function, deliberate the demerits about the existed schemes and then give out a more effective construction of Public Key Encryption with Conjunctive-Subset Keywords Search (PECSK) scheme.

COMPARATIVE ANALYSIS OF DIFFERENT MKS SCHEMES

This section presents the comparison of different Multi Keyword Search (MKS) schemes the user can identify the information are reviewed in the comparative analysis section.

Table: 1. Comparison of different MKS schemes

Papers	Search Techniques	Advantage	Disadvantage
[1]	Privacy Preserving Multi Keyword Search	1)Most relevant items are retrieved by the user 2)Preventing unnecessary Communication	1)Low efficiency compared Randomized 2)Order Preserving Low privacy for securing data.
[2]	Multi keyword Ranked Search scheme	1)It enable accurate and secure search Over encrypted mobile cloud data. 2) Security analysis can effective documents and index and trapdoor privacy access pattern of the search user.	1)It not give accurate results 2) This scheme doesn't consider the importance of the different keywords.
[3]	Nearest neighbor search *(or)k*-nearest neighbor scheme	Very secure nearest neighbor search practical on massive Datasets.	To retrieve the nearest data from datasets.
[4]	Authorized And Ranked Multi Keyword Search Scheme(ARMS)	ARMS scheme can achieve confidentiality of documents, trapdoor unlink ability and collusion resistance.	ARMS is not explore the dynamic searchable Encryption in dataset.
[5]	Similarity Based Ranking Multi Keyword Search	This search scheme achieves better than linear search efficiency	The results in precision loss
[6]	Ranked Fuzzy Multi-Keyword Search	It can be more efficient by reducing the searching time	1)Un trusted server to search for a secret word 2)Low bandwidth

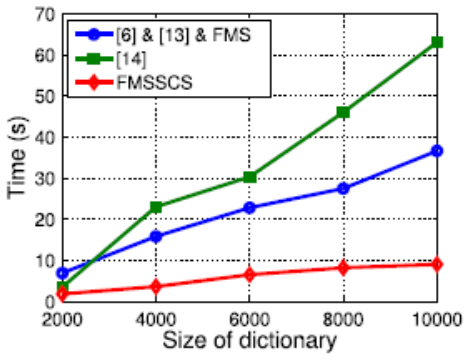
[7]	Top-k multi-keyword retrieval search	To retrieve the result very effective from cloud	1)Low level protections only 2)Computation cost high
[8]	Privacy-Preserving Multi-Keyword Text Search	It depends upon the index tree structure to enable authenticity check over the returned search results	Index problem is occurred during multiple search user access the same data
[9]	Dynamic Multi-Keyword Ranked Search (DMRS) Scheme	It special tree-based index structure can achieve sub-linear search time and deal with the deletion and insertion of documents flexible	1) Low efficiency 2)Dynamic search scheme doesn't realize the multi keyword ranked search functionality
[10]	Conjunctive-Subset Keywords Search	It can only return the results Which match all the keywords simultaneously	It cannot Provides acceptable results ranking functionality

Performance evaluation

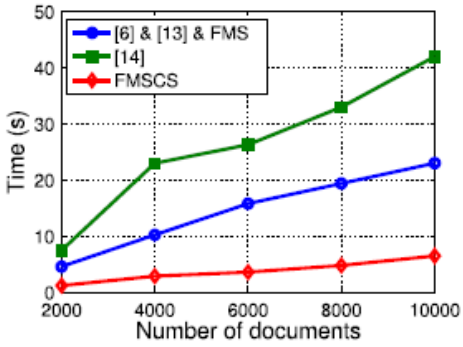
The outsourced documents are encrypted by the symmetric encryption algorithm. In also, the data owner generated the secret key and it sent to the search user through a secure model. Since symmetric encryption algorithm is secure, unauthorized can't recovery the encrypted documents without the secret key. So it can be achieved by confidentiality of encrypted documents. Compare with other multi keyword search scheme, the fine grained multi keyword search is best performance to retrieving matching results from datasets. In below graphs represented the storage and communication to the overall process model in encrypted retrieval results from the cloud server. And also, in (b) number of documents increasing, the retrieving process time will be increased. In (c) the number will be increasing, the searching process is best compare with previous techniques by using AND, OR and no operation. Every keyword is encrypted by using symmetric encryption algorithm to store in cloud server and it takes lot of space. Avoid this issue; in future to developing the compression techniques and it means first encrypted the keyword after that before storing the encrypted data in the cloud by using compression techniques to reduce the space in cloud.

POSSIBLE SOLUTION

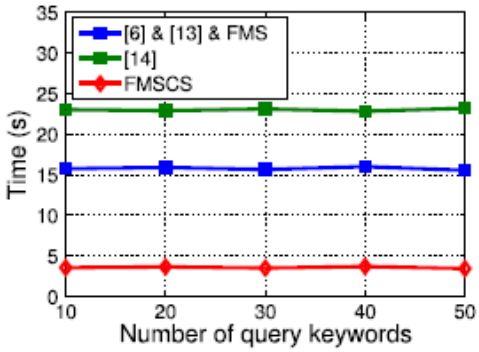
The relevance score and k-nearest neighbor techniques [3] to develop an effective multi keyword search scheme that can return the ranked search results depended on the accuracy. Within fine grained multi keyword search leverage an efficient index to further improve the search efficiency and adopt the blind storage system to conceal access of the search user. An authorized and ranked multi-keyword search scheme (ARMS) [4] over encrypted cloud data by exploit the cipher text policy attribute-based encryption (CP-ABE) and SSE techniques. Security analysis demonstrates that the proposed ARMS scheme can achieve collusion resistance. In this paper presents the FMS schemes which not only support multi-keyword search over encrypted data, but also achieve the fine-grained keyword search to investigate the relevance scores and the preference factors of keywords and the logical rule of keywords.



(a)



(b)



(c)

CONCLUSION

In this paper, to investigated on the fine-grained multi-keyword search (FMS) problem in encrypted data and it can be describes the two FMS schemes. First scheme is FMS I take in both the relevance scores and preference factors of keywords to give efficient search to the search users. Second schemes is FMS II achieve secure and effective search with practical functionality like as AND, OR and NO operations of keywords. In this scheme, multiple user can access the same encrypted data at same time is very hard situation and it means data owner to send the symmetric and secret key to multiple search user is very difficult. For the future work, can be consider the extensibility of the file set and the multi-user cloud environments. One more future is developing the highly scalable searchable encryption to enabling effective search on large databases.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] DX Song, D Wagner, A Perrig.[2000] Practical techniques for searches on encrypted data, in Proc. S&P, IEEE, pp. 44–55.

[2] H Li, D Liu, Y Dai, TH Luan, X Shen .[2014] Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage, IEEE

Trans. Emerging Topics Comput., 2014, DOI:10.1109/TETC.2014.2371239.

[3] Boyang Wang,Yantian Hou, Ming Li. [2016]Practical and Secure Nearest Neighbor Search on Encrypted Large-Scale Data, The 35th Annual IEEE International Conference on Computer Communications, IEEE INFOCOM.

[4] [4] H. Li, D. Liu, K. Jia, and X. Lin. [2015] Achieving authorized and ranked multi-keyword search over encrypted cloud data, in *Proc IEEE Int. Conf Commun.*, , to be published.

[5] Y Yang, H Li, W Liu, H Yang, M Wen. [2014]Secure dynamic searchable symmetric encryption with constant document update cost," in Proc. IEEE GLOBECOM, 775–780.

[6] Neelam S Khan, C Rama Krishna, Anu Khurana. [2014] Secure Ranked Fuzzy Multi-Keyword Search over Outsourced Encrypted Cloud Data, 2014 5th International Conference on Computer and Communication Technology, 978-1-4799-6758-2/14/$31.00 ©2014 IEEE.

[7] J Yu, P Lu, Y Zhu, G Xue, M Li. [2013]Towards secure multi keyword top-k retrieval over encrypted cloud data, *IEEE Trans.Dependable Secure Comput*., 10(4): 239–250, Jun..

[8] W Sun, B Wang, N Cao, M Li, W Lou, YT Hou, H. Li. [2014] Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, *IEEE Trans. Parallel Distrib Syst*, 25(11): 3025–3035

[9] Zhihua Xia.[2016] A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data," IEEE transactions on parallel and distributed systems, 2016 to be published.

[10] B. Zhang and F. Zhang,"An efficient public key encryption with conjunctive-subset keywords search," J. Netw. Comput. Appl., vol. 34, no. 1, pp. 262–267, 2011.

[11] H Liang, LX Cai, D Huang, X Shen, D Peng. [2012] "An smdp-based service model for inter domain resource allocation in mobile cloud networks," *IEEE Transactions on Vehicular Technology*, 61(5): 2222–2232.

[12] W Sun, S Yu, W Lou, YT Hou, H Li. [2014] "Protecting your right: Attribute based keyword search with fine-grained owner-enforced search authorization in the cloud," in *Proceedings of INFOCOM*. IEEE, 2014.

[13] Y Yang, H Li, W Liu, H Yang, M Wen.[2014] Secure dynamic searchable symmetric encryption with constant document update cost "in *Proceedings of GLOBECOM, USA*

[14] Y Yang, H Li, M Wen, H Luo, R Lu. [2014] Achieving ranked range query in smart grid auction market," in *Proceedings of ICC*, pp.951–956.

[15] N Cao, C Wang, M Li, K Ren, W Lou. [2014] "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., 25(1): 222–233

[16] https://support.google.com/websearch/answer/173733?hl=en,2014.

[17] CR Barde, Pooja Katkade, Deepali Shewale, Rohit Khatale. [2014] Secured Multiple-keyword Search over Encrypted Cloud Data,Department of Computer, Pune University, GESRHSCOE, Nashik, India,2014.