Review on Cluster Based Aggregate Key and Fine-Grained Index Searching Over Encrypted Cloud Data

Vrushali Kadam[1], Prof. V. B. More[2]

[1]ME Student, Department of Computer Engineering, MET's BKC IOE, Nasik, Maharashtra, India
[2]Professor, Department of Computer Engineering, MET's BKC IOE, Nasik, Maharashtra, India

---***---

Abstract - *With the convenience of cloud environment such as, storage, sharing, searching etc. data owner promote their data over cloud in encrypted format for preserving privacy of it. Sensitive information is encrypted before outsourcing it on cloud which can further utilized by other users using traditional approach known as plaintext keyword searching. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. In previous system, single keyword or Boolean keyword search is provided over an encrypted data. Define MRSE (multi keyword ranked searching technique for efficient searching over an encrypted data). For efficient utilization of MRSE technique refers "co-ordinate matching". In coordinate matching, multiple matches can possible by searching more relevant data documents to search query. As aggressive size increased in data documents searching phase reach to linear computational complexity. MRSE method can provide efficient search result than traditional search approach. As a part of contribution provide aggregate key searchable encryption on cloud stored data and also going to develop a practical and efficient multi-keyword search scheme which can support complicated logic search the mixed "AND", "OR" and "NO" operations of keywords.*

Key Words: **Cloud computing, cipher text search, ranked search, multi-keyword search, hierarchical clustering, and security.**

1. INTRODUCTION

Enterprise user having huge amount of data prefer to outsourced their data on cloud for proper data management and for privacy concern. Outsourcing of data have many benefits like, reduced data management cost and local storage space. Cloud service provider gives an assurance of security for end user data. Encryption is traditional approach for preserving privacy of an original data. Ciphertext search scheme is proposed that incorporates cryptography techniques (Fig.1). Generally, relationship of documents represents the attribute and properties of documents. Single keyword encryption technique needs to encrypt each word in the document uniquely therefore the cost of scanning whole document by reading each word is increased according to the size of document. Multiple keyword searchable encryptions are the way of secure searching based on vector space model. Vector search model analysis's lots of security information on search performance hence it cannon predicted as efficient technique for secure searching. The problem of searching is solved by existing system but increases the time as there is growth in documents. Further methods for better efficient search known as index search introduced but in this author ignored to show the relevance between documents with index document searching. With index searching approach user aims to retrieve more relevant documents with respect to given query which is not fulfilled at all. Merkel hash tree is another way introduced to search document over cloud environment. This technique only verifies the specific issues while it neglects privacy preserving competence. To overcome such problems in existing system there is requirement of efficient search technique with index search as well as having privacy concerned in it. As per analysis it is observe MRSE-HCI is technique for multi-keywords ranked search over an encrypted data and it is based on hierarchical clustering index. Usually users are concentrating on typical topics while retrieving documents form cloud. Hence MRSE-HCI is technique concentrates on specific field and increases the searching speed by calculating relevance score of query search documents. Due to calculation of relevance score among documents unnecessary fields are ignore which increases the searching speed.

On cloud storage generally user prefers to save documents in encrypted format to preserve privacy. Hence it is seems that efficient and reliable searching mythology is required for ciphertext search over cloud. MRSE-HCI strategy maintains the relationships between related documents. Clustering method is utilized for classification of documents into specific cluster domain with minimum relevance score among documents of given dataset. Cluster centre are dynamically constructed if any cluster is broken due to new cluster entry. In the process of dynamic cluster creation all documents reassigned and centre of each cluster is reselected. Therefore, number clusters are depending upon no. Of documents and their relationship with different documents i.e. plaintext documents. Ciphertext search technique is going to consider by concentrating on the problem of persevering relationship between plain texts documents over encrypted documents and enhance the performance of searching process. MRSE-HCI technique helps to retrieve online information as well as semantic search. Give best to retrieve correct search results in flexible time and with experimental results prove improvement in search efficiency, rank security, and the relevance between retrieved documents. Contribution in this system development is that aiming to develop such system that provides efficient secure searching of documents in cloud environment. Proposing a novel approach named as, Cluster Based Aggregate Key and Fine-Grained Index Searching over Encrypted Cloud Data. In this approach will implement multiple key aggregation tasks. To protect users' data privacy is a central question of cloud storage. With more mathematical tools, cryptographic schemes are getting more versatile and often involve multiple keys for a single application. Considering, how to "compress" secret keys in public-key cryptosystems which support delegation of secret keys for different ciphertext clusters in cloud storage. No matter which one among the power set of cluster, the delegate can always get an aggregate key of constant size.

Therefore, propose approach can be more flexible than hierarchical key assignment which can only save spaces if all key-holders share a similar set of privileges. As a part of contribution an aggregate key searchable encryption on cloud stored data is provided. Secondly, to support for complicated logical search the mixed "AND", "OR" and "NO" operations of keywords.

2. RELATED WORK

D.X. Song, D. Wagner et al. [1], discussed about public key encryption techniques. They addressed the problem of encrypted data searching. Generally data owner outsourced their data to the cloud into encrypted format; it can be used by other user of cloud. Cryptographic techniques are also proposed by them. It provides the proof of security. They show that their technique provides the data confidentiality over searching approach. The proposed technique in this paper required fundamental primitives from traditional symmetric key cryptography because it uses well-defined security related notions. In this paper, they defined that their system is more probably secure as hidden search & query isolation is also provided. There are two constructions related to IBE have been discussed. Though proposed PEKS utilized IBE scheme but it has converse problem. It required extra properties to be exploited. It outputs less efficiency as it is based on basic trapdoor permutations.

Moni Naor and Kobbi Nissim [2], addressed the problem of certificate revocation. The certificate revocation is arranged by authenticated dictionaries. Author also considered the scenario in which certificates does not involved. Certificate verification is the variant process of memory checking. The proposed technique is closer to CRL and CRT rather than CRS. Proposed techniques can easily verify that the specific certificate number is in the list or not. It does not require whole list. Also communication cost is also less. Many challenging issues are solved by integrating ABE, PRE & lazy re-encryption techniques. The proposed scheme has salient features such as, confidentiality of user access privileges as well as accountability of user secret key. The proposed approach is based on the observation of practical scenario of data files that have different attributes.

E. J. Goh [3], proposed secure index search IND-CKA. It is also known as, "Z-IDX" using pseudo-random function and Bloom filter. The proposed scheme is efficient for data searching. An Index construction technique that uses hash table seems to be unsuitable for encrypted index documents because there may chances of information leak. For construction of IND-CKA secure indexes there is requirement of analyzing, bloom filters, pseudo random functions & pseudo-random generators. Pseudo random functions are computationally undefined from random functions whereas, pseudo random generator produced a string which also undefined by random string.

P. Golle, J. Staddon1 et al. [4], proposed a protocol for conjunctive keyword query on encrypted data. Web searching on conjunctive keywords is the crucial task. It does not compass all possible and feasible search criteria. The proposed protocol is complex server to separate out the encrypted keywords of documents. It allows small conjunctive search with less capability. This work partially solves the Boolean search problem over an encrypted data. For complete solution there is need of disjunctive search. The challenging issue of information leak isn't addressed by proposed approach. The proposed scheme to perform conjunctive search on encrypted data. It utilized Shamir Secret Sharing. It provides efficient searching. In this scheme trapdoors are sequential in the large number of documents. Bilinear pairing is used as alternative to produced fixed size of trapdoors. As it has symmetric settings, it incurs significantly less overhead.

A. Swaminathan, Y. Mao et al [5], proposed technique for integration of relevance score. They developed a framework to preserve confidentiality of ranked search over in large scale document collections. There additional focused is on security specific point. The challenging security issues include protecting communication traffic analysis and the communication links. It required efficient algorithm design.

S. Zerr, D. Olmedilla et al [6], discussed similar approach in Zerber+R ranking model. Relevance transformation function can calculate relevance score between various undefined terms. It helps to untrusted servers to do not reveal the information about indexed data. IR technique is used to enable fast query execution over large indexes which make system scalable. Top-k retrieval utilized IR technique for accessing index control information. Inverted index is represented as sequence of posting elements. Main aim of this paper is to developed Zerber+R ranking model that allows retrieving top-k inverted index without information leakage.

N. Cao, C. Wang, et al. [7], proposed MRSE scheme. It is based on computation of secure inner product. In this author discussed about Boolean keyword search and single keyword search technique. They addressed the problem with this technique. The proposed MRSE is Multi-keyword Ranked Search Encryption in which they established strict policies for securing cloud data. "Coordinate matching" is implemented for many search matches. Multi-keyword semantic refines the relevance of result. This technique is widely used in information retrieval community. KNN algorithm is used for secure inner product computation. It calculates the Euclidean distance between data records and query vector. It uses 'k' nearest database records. In this algorithm a secret key constructed from one bit vector. The proposed techniques provide the guarantee of search efficiency and privacy.

C. Wang, N. Cao, et al. [8], proposed statistical measure approach i.e. calculation of relevance score from the retrieval of information. The calculated relevance score is then used to construct secure searchable index. To protect information of sensitive score order preserving mapping technique is used. Order-preserving symmetric encryption (OPSE) is implemented to provide provable security under pseudorandom functions of security framework. Another technique, RSSE is used for document index construction. The relevance score of document is attached with index at each entry point. The proposed technique required additional mapping cost, bits for encrypted score and whole encryption entry cost. The additional encrypted score for bit is the primary issue due to cheap or less cost of storage. Order preserving mapping technique is based on presampling and relevance score training which is not efficient. Binary Search () algorithm is used to inherit the importance of OPSE. Reverse mapping algorithm is also proposed for the completeness of proposed approach.

W. Sun, B. Wang [9], proposed MTS i.e. privacy preserving multi-keyword text search approach. It supports multi-keyword search and ranked search result. A term frequency and vector space model is used to build index. Higher search accuracy is evaluated using cosine search similarity. To improve search efficiency, tree-based index structure and multi-dimensional algorithm is implemented. Also author explained efficiency and effectiveness of the proposed approach through experimental evaluation. The proposed technique avoids the data leak of sensitive information. For refinement of search efficiency, tree based search approach is used. Three -efficiency related factors are identified by proposed method which can improve the search algorithm efficiency on index tree based approach.

C. Chen, X. Zhu et al [10], proposed Privacy-Preserving Ranked Keyword Search approach. Proposed approach maintains the relationship between original and encrypted documents. To fulfill this task they have proposed MRSE-HCI technique. It is multi-keyword ranked search over encrypted data based on hierarchical clustering index technique maintains the semantic relationship between documents over cloud. It helps to enhance efficiency of searching to retrieve documents from cloud. They have utilized clustering approach for similar document grouping. The proposed approach overcomes the problems in data explosion, online document retrieval and semantic search.

3. SYSTEM ARCHITECTURE

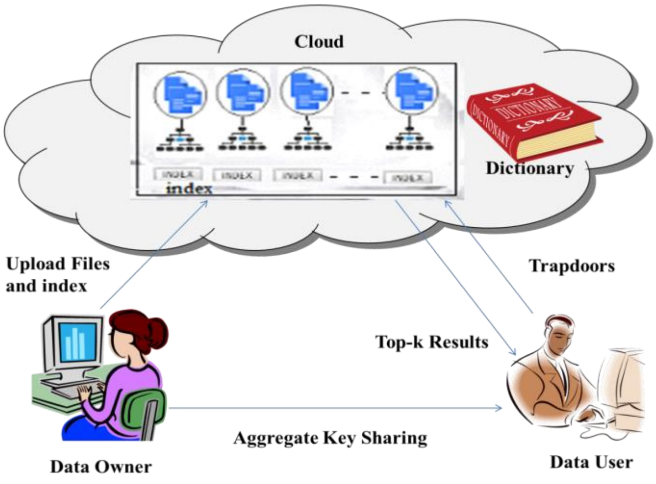

Fig.1: System Architecture

1. Data Owner:

- Owner of data is responsible to upload data and index in encrypted form on cloud.
- Before uploading data on cloud data owner have to collect documents, and then apply proper indexing on them.
- Before performing encryption master key is generated by system which is secured and data owner can used this key for the documents that belonging to same cluster.
- Data owner provides access permission i.e. sharing for end user.

2. Data User:

- Authorized user request for document from cloud.
- While requesting document users have to specify some keywords related to requested documents.
- Implementing key aggregation hence authorized user will get aggregate for downloading required documents.

3. Server:

- Server is nothing but cloud storage which is responsible to maintain documents uploaded by data owner.

- When other users request for documents, server provide service to them.
- Cloud server search for encrypted index document to retrieve user requested documents.
- Finally returns top-k document searched results.
- System aims to protect data from revealing the information to the cloud server. So, that efficiency of search can be improved.

3. CONCLUSIONS

In this paper, we have reviewed some previous techniques of document searching on cloud server. Multiple methods proposed in literature survey have some limitations such as, single keyword search i.e. Boolean keyword search etc. Also existing techniques does not capable of maintaining relationship between plaintext documents and encrypted documents in which user simply add search keyword in plain text format. There have some privacy issues over searching mechanism i.e. revealing of user identity or data. Hence, according to our analysis and observation from the study of literature survey there is need of an efficient technique which can preserve relationships between documents and support for complicated logical search with privacy concern. We analyzed that the MRSE-HCI can be better solution for earlier discussed problem. It also supports multi-keyword searching to retrieve documents from cloud.

ACKNOWLEDGEMENT

We are glad to express our sentiments of gratitude to all who rendered their valuable guidance to us. We would like to express our appreciation and thankful to Dr. K. V. Metre, ME Coordinator and Dr. M.U. Kharat, Head of Department of Computer Engineering, MET's BKC IOE, Nasik. We also thank the anonymous reviewers for their comments.

REFERENCES

[1] D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44–55.

[2] M. Naor and K. Nissim, "Certificate revocation and certificate update," IEEE J. Sel. Areas Commun., vol. 18, no. 4, pp. 561–570, Apr. 2000

[3] E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol. 2003, pp. 216. 2003.

[4] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. Proc. 2nd Int. Conf. Appl. Cryptography Netw. Security, Yellow Mt, China, 2004, pp. 31–45.

[5] A.Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, "Confidentiality-preserving rank-ordered search," in Proc. ACM ACM Workshop Storage Security Survivability, Alexandria, VA, 2007, pp. 7–12.

[6] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+R: Topk retrieval from a confidential index," in Proc. 12th Int. Conf. Extending Database Technol.: Adv. Database Technol., Saint Petersburg, Russia, 2009, pp. 439–449.

[7] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829–837.

[8] C. Wang, N. Cao, K. Ren, and W. J. Lou, "Enabling secure and effi- cient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.

[9] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inform., Comput. Commun. Security, Hangzhou, China, 2013, pp. 71–82.

[10] C. Chen, X. Zhu,. P. Shen, J. Hu, "An Efficient Privacy-Preserving Ranked Keyword Search Method," IEEE Trans. Parallel Distrib. Syst., vol. 27, no.4, pp., Apr. 2016.

BIOGRAPHIES

MS. Vrushali Kadam received the B.E. degree in Information Technology from MET's BKC IOE, Nasik in 2015. She is currently pursuing her Maters degree in Computer Engineering From MET's BKC IOE, Nasik, Savitribai Phule Pune University. This paper is published as a part of the research work done for the degree of masters.
(e-mail:vrushalirk94@gmail.com)

Prof. V. B. More is a professor at MET's BKC IOE, Nasik. Dept. of Computer Engineering, Savitribai Phule Pune University.(e-mail: vbmore2005@rediffmail.com)