Open access Journal International Journal of Emerging Trends in Science and Technology

IC Value: 76.89 (Index Copernicus) Impact Factor: 2.838 **DOI:** https://dx.doi.org/10.18535/ijetst/v3i12.11

A Review on Enabling Synonym Based Fined-grained Multi-keyword Search Using Hierarchical Clustering

Authors

Miss.Pawar Pratima Prakash[1], K. N. Shedge[2]

[1]ME Computer, Computer Engineering Department
Email: *pratimapawar92@gmail.com*

[2]ME Computer, Computer Engineering Department
Email: *kishorshedge2007@gmail.com*

Abstract

Cloud is huge platform that provides various services its users such as, storage, management, maintain, searching, sorting of data etc. With this advantages data owner outsourced their data in cloud. Before outsourcing of data on cloud user encrypt it as per security concern. Other users can use cloud data by downloading and decrypting it from cloud. Before performing download and decryption operation on cloud data user have to perform searching for required data. Traditionally, plaintext keyword search approach is used to perform 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. MRSE is the multi keyword ranked searching technique for efficient searching over an encrypted data. For efficient utilization of MRSE technique we referco-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.

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

INTRODUCTION

To preserve relationship between original and encrypted documents over cloud environment to improve search efficiency MRSE-HCI technique is discussed. MRSE-HCI is multi-keyword ranked search over encrypted data based on hierarchical clustering index. In MRSE-HCI, search time is increases linearly as aggressive growing size of data collection. Idea of this paper was derived from observation of user retrieval needs. In this paper authors aim to increase document searching speed by calculating relevance score between user query and documents. Due to relevance score evaluation, user gets the documents related with users query. Therefore, irrelevant fields get ignored which tends to increase the searching speed. Main aim of maintaining relationship between different plain documents and encrypted document can achieved using clustering method. Relevance score metric is used to calculate relationship between different documents. Problem in this technique is constraint on the cluster may break if any document added to the cluster. Cluster center is created dynamically and then number of clusters is further decided by attributes and property of dataset. Hierarchical method is utilized to get better clustering result within large amount of data collection. In hierarchical clustering method no. of clusters and minimum relevance score increases as increase in the level of maximum size of cluster reduced. If cluster exceeded its size level, it will be further divided into several sub-clusters. Ranked privacy preservation strategy also followed in this paper. Searching user query document is an iterative process in which system evaluates the relevance score between query and document included in the small cluster. If the document in small cluster does not satisfies user query document then system again search back to its parent cluster. After whole searching procedure there is one more classification is required for the most frequent document extraction hence user query documents are ranked by system to make searching efficient and flexible.

Therefore, ranked privacy enhanced system performance. For further improvement, a verifiable tree is constructed upon hierarchical clustering method. It helps to verify integrity of the search result. MR-MHT and cryptographic signature is used for authentication of tree structure. In verification process each document will be hashed and the hashed result used as document representative.

Finally they were contributing themselves to make investigation to maintain relationship between plain documents over encrypted documents by proposing clustering method. They utilized MRSE-HCI mechanism to speed up the searching operation. In this they designed a backtracking algorithm to improve searching strategy with ranked privacy. Overall searching process is based on ciphertext search scenario. In this paper, efficiency of search and security under two popular threat models is analyzed.

Basically, relevance measure is the ideal concept or base for the concept of " co-ordinate matching. To quantify relevance of document query and different document relevance measure is used. QHC algorithm is used for hierarchical cluster. It is basically dynamic ' k' -means. In this clustering approach for clusters minimum relevance threshold is defined for compact & dense cluster. It is iterative process until k is stable.

RELATED WORK

In this section we are going to discussed related work about Efficient Privacy-Preserving Ranked Keyword Search Method. As per the tightness temporal constraints there are three types classified into trajectory patterns.

They are explain as following:

A. *Single keyword Searchable encryption*

Mihir Bellare, Alexandra Boldyreva and Adam O Neill [1] represents public key encryption scheme to achieve deterministic privacy in encryption algorithm. For secure deterministic encryption scheme encrypt-with-hash technique is introduced by author. This technique replaces the coins used by a standard encryption schemes with the hash of messages. To permit more efficient search ability ESE i.e. efficient searchable encryption primitive defined in this paper. A deterministic encryption provides the solution for linearly scan queries. Other than ESE, CCA used to provide security proofs; it is extension to the hash function. Song et al. [2] proposed the notion of searchable encryption. They described remote searching methodology on encrypted data using untrusted server. In this encryption methodology each word in encrypted independently. As each word is encrypted independently the required cost of scanning complete data collection word by word is more. Number of advantages existed for this system due to its provable security and support for hidden search and query isolation.

Author Eu-Jin Goh [3], developed an efficient IND-CKA secure index construction called Z-IDX using pseudo-random functions and Bloom filters, and represent how to use Z-IDX to implement searches on encrypted data. This scheme is more efficient for encrypted data search. It requires O (1) search time per document. It is capable of handling compressed data, variable length words and regular expression queries. Search indexes are the natural extension for the problem of constructing data structures with privacy guarantee. In [4]-[5], a searchable symmetric encryption is used to outsource symmetric private information retrieval. Authors S. Jarecki, C. Jutla, al, extended OXT protocol from MC-SSE settings In order to obscure any relation in original clear-text database randomization of data locations is called in EDB i.e. encrypted database.

Privacy-preserving similarity based text retrieval [6] is proposed by W. Sun, B. Wang, N. Cao et al. A vector space model is used having cosine measure is utilized in this paper for secure searching result with an accuracy. To achieve privacy meets two secure index schemes are used in two threat model. Tree based index search scheme is implemented for refinement of search efficiency. Finally, author represents the performance of their system with BMTS and EMTS in terms of search effectiveness, efficiency and privacy.

B. Multiple Keywords Searchable Encryption

F. Li, M. Hadjieleftheriou al. [7] discussed about outsource database i.e. ODB. They proposed a comprehensive evaluation of authenticated index structure based on variety of cost metrics. They extended their work to dynamic environment. Query formulation technique is used for query formulation. In future work they are planning to extend their idea to multidimensional structure with more types of queries. A secure KNN computation on encrypted databases is introduced in [8]. SCONEDB model is introduced in this paper to capture activities of users and attacker on encrypted database. Existing techniques such as, OPE for range queries also introduced in it. Symmetric encryption approach ASPE is used to preserve special type of scalar products. Security goal are also included as another component into SCONEDB model for future evaluation. Author C. Gentry represents a complete homomorphic encryption using ideal lattices [9]. In this paper author proposed a randomize algorithm due to lack of space. Introduced lattice based cryptosystem is typically used for decryption as it have ability of decryption. This scheme is not boots trappable i.e. i.e., the depth that the scheme can correctly evaluate can be logarithmic in the lattice dimension, just like the depth of the decryption circuit, but the latter is greater than the former. D. Boneh and G. Crescenzo proposed PEKS scheme for public key encryption with keyword search [10]. This scheme is related to IBE scheme i.e. Identity Based Encryption. But the problem is PEKS scheme is complicated to design. Author showed that PEKS implies Identity Based Encryption, but the converse is currently an open problem. There PEKS scheme required IBE construction to prove its security by exploiting more attributes.

Confidentiality-Preserving Rank-Ordered Search approach discussed in [11] by A. Swaminathan and Y. Mao. They construct a framework to maintain data confidentiality in ranked order search in large scale document. The proposed mechanism extracts the most relevant document from an encrypted collection based on the encrypted search queries. This technique attempts to bring together advanced information retrieval capabilities and secure search capabilities. In this paper author plan to focus on securing indices, other important security issues include protecting communication links and combining traffic analysis in future work. A secret key can produce tokens for testing any promoted query signify. Without analyzing any other information about plaintext the query significance token anyone test the predicate on a given ciphertext [12].

Analyzing security of searching on encrypted data a general framework is represented in this paper. There are protocols introduced in [13], for conjunctive search which is provably difficult for server to differentiate between encrypted keywords.

PROBLEM DEFINITION

"To maintain semantic relationship between different plain documents over related encrypted document and improve the semantic search performance. Also provide fined-grained and synonym search over an encrypted documents."

SYSTEM ARCHITECTURE

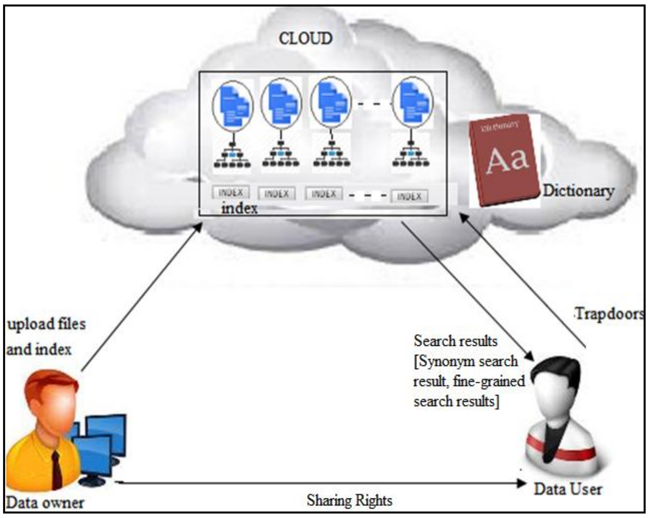

Figure 1: System Architecture

Above figure 1 represents the system architecture. It mainly consists of six phases given as below:

1) **Key Gen:** In this this phase sk i.e. secret key generated which is used for index as well as document encryption
2) **Index Gen:** Using secret key sk, encrypted index is generated in this phase and at the same time clustering process also begins.
3) **Encrypt:** Symmetric encryption algorithm is used for document collection encryption. Semantic security is achieved with this algorithm.
4) **Trapdoor:** In this, with the help of users input keywords and secret key encrypted query vector is generated.
5) **Search:** Searching is performed by cloud entity by comparing trapdoor and encrypted index top-k results are given to the user in the form of result.
6) **Decrypt:** Encrypted documents are retrieved by user and decrypted using key which generated in the first step.

CONCLUSION

Basically, data owner preffered cryptography techniques for data encryption before uploading it on cloud. It is for data privacy concern. There are several data encryption techniques available to protect revealing of data and user' s identity. While uploading data data owner' s defines the access/sharing rights for end user. At the other user end they retrieves the data using keyword searching mechanism. Several existing keyword search methods are available such as, PEKS, SCONEDB, Rank-Ordered Search, OPE etc. In this review paper we did analysis of exsting keyword seach techniques. According to our analysis ther are some techniques namely, Tree based index search is efficient refinement of search technique are beneficial for search approach whereas, the techniques like PEKS, IBE etc have some limitations like, protecting communication links and combining traffic analysis, have some difficulties in conjective search. As per our deep analysis in previous methods does not support to relationship mantainance between original and encrypted documents while retrieving them using keyword search approach over cloud environament. Hence, our analysis in this review paper from literature survey conclude that there is need of appropriate approach to maintain relation-nship between original and encrypted documents.

REFERENCES

1. M. Bellare, A. Boldyreva, and A. O' Neill, Deterministic and effi- ciently searchable encryption," in Proc. 27th Annu. Int. Cryptol. Conf. Adv. Cryptol., Santa Barbara, CA, 2007, pp. 535– 552.
2. 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.
3. E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol. 2003, pp. 216. 2003.
4. S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, " Outsourced symmetric private information retrieval," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., Nov. 2013, pp. 875– 888.
5. D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M. C. Rosu, and M. Steiner, Dynamic searchable encryption in very large databases: Data structures and implementtation," in Proc. Netw. Distrib. Syst. Security Symp., vol. 14, 2014, Doi: http://dx.doi.org/10.14722/ndss.2014.23264.
6. 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.
7. F. Li, M. Hadjieleftheriou, G. Kollios, and L. Reyzin, " Dynamic authenticated index structures for outsourced databases," in

Proc. ACM SIGMOD, Chicago, IL, 2006, pp. 121– 132.

8. W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, “ Secure kNN computation on encrypted databases,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, Providence, RI, 2009, pp. 139– 152.
9. G. Craig, “ Fully homomorphic encryption using ideal lattices,” in Proc. 41st Annu. ACM Symp. Theory Comput., 2009, vol. 9, pp. 169– 178
10. Y. H. Hwang and P. J. Lee, “ Public key encryption with conjunctive keyword search and its extension to a multi-user system,” in Proc. 1st Int. Conf. Pairing-Based Cryptography, Tokyo, JAPAN, 2007, pp. 2– 22.
11. 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.
12. D. Boneh and B. Waters, “ Conjunctive, subset, and range queries on encrypted data,” in Proc. 4th Conf. Theory Cryptography, Amsterdam, NETHERLANDS, 2007, pp. 535– 554
13. 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.