ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

A Review on Correspondence Techniques for Effective Keyword Search over Encrypted Data in Cloud

Nitin B. Pawar [1], Prof. Dinesh D.Patil [2]

P.G. Student, Department of Computer Science and Engineering, HSMSSGB, COET, Bhusawal [MH], India [1]

Associate Professor, Department of Computer Science and Engineering, HSMSSGB, COET, Bhusawal [MH], India [2]

ABSTRACT: Various techniques of keyword based searching are used for retrieving the encrypted data from cloud servers. Keyword based searching over encrypted data is necessary for accessing the outsourced sensitive data in cloud computing. In this paper we present a survey on, keyword based searching techniques. This survey work involves a comparative study of these keyword based searching techniques on the basis of accuracy of search results and efficiency in data updating. The existing techniques depend upon a global dictionary, which not only affects the accuracy of search results but also causes inefficiency in data updating. It concludes that till now semantic-based compound keyword search (SCKS) techniques is the best methodology, achieves not only semantic-based search but also multi-keyword search and ranked keyword search. Additionally, SCKS also eliminates the predefined global library and can efficiently support data update. We would like to extend our work by enhancing security and preserving privacy by modifying k-Nearest Neighbor (kNN) algorithm in SCKS.

KEYWORDS: Keyword search techniques, Searchable encryption, Semantic-based keyword search

I. INTRODUCTION

Cloud storage becomes more and more popular in the recent trend since it provides various benefits over the traditional storage solutions.Cloud storage services enable users to remotely access data in a cloud anytime and anywhere, in a pay-as-you-go manner using any device. One of the most popular services of cloud computing is data outsourcing. For reasons of cost and convenience, public as well as private organizations can now outsource their large amounts of data to the cloud and enjoy the benefits of remote storage. However, allowing a cloud service provider (CSP), whose purpose is mainly for making a profit, to take the custody of sensitive data, raises underlying security and privacy issues. Outsourcing unencrypted data to cloud by the owner is not much secure because server may leak information to cyberpunks. Hence To preserve privacy, users opt to encrypt data before outsourcing. Thus, the traditional keyword Search cannot be directly executed on the encrypted data, which limits the utilization of data. To address this problem, proposed the idea of searchable encryption (SE) [1] that allows users to search on encrypted data through a keyword. Afterwards, various searchable encryption schemes were proposed to meet different requirements of keyword based searching such as fuzzy keyword search, multi-keyword search, ranked keyword search and semantic-based keyword search.

II. TECHNIQUES FOR KEYWORD BASED SEARCH

Many searchable techniques [9] have been proposed on the basis of keyword search. Discussion is made on the existing techniques that are been intend by many authors. This study analyses the algorithms for searching the encrypted content. Survey is made on these algorithms based on the working principle, merits and demerits. It also compares the complexity, efficiency overhead of various algorithms and shows which technique is better to handle while retrieving the encrypted content. It includes working of encryption algorithm, how searching is done on the encrypted content, advantages and disadvantages of each technique.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

A. *Fuzzy Keyword Search*

Although traditional searchable encryption schemes [9] allow a user to securely search over encrypted cloud data through keywords and selectively retrieve files of interest, these techniques support only the exact keyword search. That is, there is no tolerance of minor typos and format inconsistencies, which on the other hand, are typical user searching behavior and happen very frequently e.g. Diamond and Daimond. Fuzzy keyword search solves the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy [10]. Fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. It uses edit distance to quantify keywords similarity and develop a novel technique i.e., a wild-card based technique, for the construction of fuzzy keyword sets. The technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzy keyword set is significantly reduced.

B. *Multi-Keyword Search*

Early works mostly only support single keyword search. Later, several multi-keyword search schemes were proposed [6], [10], [12]. Multi-keyword search is when a user searches and lists for multiple variations of same keyword. Among various multi-keyword semantics, we choose the efficient principle of coordinate matching, i.e. to find as many matches as possible, to capture the similarity between search query and data documents. Data owner store data in encrypted form and data user generate trapdoors [1], [2], [4] to send query request in encrypted form. Re-encryption of keyword index and trapdoors [1], [10], [12] used to increase more security from attackers.

C. *Ranked Keyword Search*

Ranked Keyword Search [9], [12], [14] is a very useful technique that allows a user to securely search encrypted documents on cloud through a single keyword and retrieve documents of interest. However application of this approach to large scale cloud data would not be suitable, as it cannot accommodate high service level requirements like system usability, user searching experience and information discovery. Although various techniques 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.

D. *Multi-Keyword Ranked Search*

Multi-keyword ranked search (MRSE) [9], [12] has been proposed to overcome the problem of Ranked keyword search. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in the context of Cloud Computing [8].

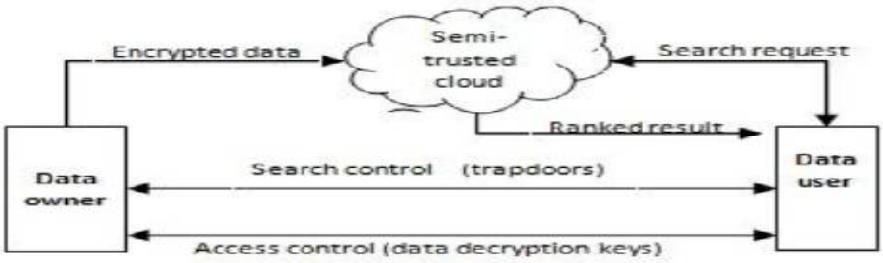

Fig.1 - Architecture of the multi-keyword ranked search over encrypted cloud data [9].

MRSE [9] involves three entities: the data owner, the data user and the cloud server, as shown in fig. 1. The data owner has a collection of data documents that are to be outsourced to the cloud after performing encryption. Before outsourcing the documents to the cloud data owner will build an encrypted searchable index *I* and then outsource both index *I* and encrypted documents. Data user acquires a corresponding trapdoor *T* to search document collection for *t* given keywords. On receiving *T* the cloud server is responsible to search the index *I* and return the corresponding set of

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

encrypted documents. Ranking is done according to some search criteria in order to improve the document retrieval accuracy.

E. *Semantic-based Keyword Search*

Semantic-based keyword search [1], [2] not only is convenient for users but also exactly expresses users' intentions. Specifically, in some circumstances, users might not be familiar with the encrypted documents stored in cloud storage or might only want the semantically related results; therefore, the search keywords are usually semantically related to the document rather than via an exact or fuzzy match. So, the semantic-based keyword search is of practical importance and has attracted much attention. However, the existing approaches [1], [13] must rely on a predefined global dictionary whose quality greatly influences the accuracy of the search result. the predefined dictionary is constructed based on all documents in the dataset, the update of a single document can cause the reconstruction of the dictionary and even all document indexes, which is inefficient [11].

F. *Semantic-based Compound Keyword Search*

Semantic-based compound keyword search (SCKS) [1], [2] scheme uses a topic set in a field and Vector Space Model (VSM) to express the semantic information of keywords. Each element of the keyword vector corresponds to a field topic, and the value is the semantic similarity between the keyword and the topic. Because the keywords and field topics can be compound concepts, we initially propose an ontology-based compound concept semantic similarity (CCSS) calculation method to measure their semantic similarity. In CCSS, the compound is decomposed into subject headings and auxiliary words, and the relationships between them are used to measure the similarity. Locality-Sensitive Hashing (LSH) function is able to hash similar items to the same bucket with high probability. Hence, we construct the document index by using LSH to map multiple keyword vectors into only one vector. Compared with the existing schemes [2], [3], in which the vector value is only 0 and 1, SCKS can express more semantic information of the document. Another advantage of SCKS is that it can support data update efficiently because no global dictionary need be predefined and each document is individually indexed.

III. LITERATURE REVIEW

In the paper of 'Semantic-based Compound Keyword Search over Encrypted Cloud Data' existing schemes depend upon a global dictionary, due to this affects the accuracy of search results, causes inefficiency in data updating and only process keyword as single words, which split the original semantics and achieve low accuracy. To overcome these limitations, Bo Lang [1] proposed initially a compound concept semantic similarity (CCSS) calculation method to measure the semantic similarity between compound concepts. Next, by integrating CCSS with Locality-Sensitive Hashing function and the secure *k*-Nearest Neighbor scheme, a semantic-based compound keyword search (SCKS) scheme is proposed. SCKS achieves not only semantic-based search but also multi-keyword search and ranked keyword search.

In 'Compound Keyword Search of Encrypted Cloud Data by using Semantic Scheme' paper Harsh Gupta [2] proposed same schemes as Bo Lang [1] proposed but Harsh Gupta [2] measure the semantic sameness of compound concepts, put forth a novel approach that considers the concept constituent features and several other factors influencing similarity. additionally Security Enhanced SCKS, allowed to submit queries adaptively.

In 'Synonymous Keyword Search Over Encrypted Data in Cloud' paper Avani Konda [3] proposed the index construction and query generation of the vector space model and TF_IDF model that is widely used are combined. "Greedy Depth-first Search", a special tree-based index structure algorithm to provide multi-keyword ranked search that is efficient is proposed. To encrypt the index and query process, the secure KNN algorithm is used. Due to the use of special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

In 'EFFECTIVE DATA SEARCH FOR ENCRYPTED RELATIONAL DATA IN CLOUD USING K-NEAREST NEIGHBOR ALGORITHM' Paper Abirami Swetha.L [4] proposed protocol is to develop a privacy preserving k-NN classifier over encrypted data under the semi –honest model. Also empirically analyze the efficiency of the proposed protocol using a real world data set under different parameter settings. Specifically, focus on the classification problem since it is one of the most common data mining tasks. Because each classification technique has their own advantage, this protocol concentrates on executing the k-nearest neighbor classification method over encrypted data in a cloud.

In 'Identity-Based Encryption with Outsourced Revocation in Cloud Computing' by Jin Li [5] proposed Identity-Based Encryption (IBE) which simplifies the public key and certificate management at Public Key Infrastructure (PKI) is an important alternative to public key encryption. However, one of the main efficiency drawbacks of IBE is the overhead computation at Private Key Generator (PKG) during user revocation. Furthermore, Jin Li [5] proposed another construction which is provable secure under the recently formulized Refereed Delegation of Computation model. Finally, provide extensive experimental results to demonstrate the efficiency of our proposed construction.

In Mikhail Strizhov [6] 'Multi-keyword Similarity Search Over Encrypted Cloud Data' paper proposed a novel secure and efficient multi-keyword similarity searchable encryption (MKSim) that returns the matching data items in a ranked ordered manner. Unlike all previous schemes, search complexity is sublinear to the total number of documents that contain the queried set of keywords. Analysis demonstrates that proposed scheme is proved to be secure against adaptive chosen-keyword attacks. Mikhail Strizhov [6] show that there approach is highly efficient and ready to be deployed in the real-world cloud storage systems.

In Y. Elmehdwi [7] 'Secure k-nearest neighbor query over encrypted data in outsourced environments' paper proposed focus on solving the k-nearest neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud, and the cloud returns the k closest records to the user. First present a basic scheme and demonstrate that such a naive solution is not secure. To provide better security, Y. Elmehdwi [7] proposed a secure kNN protocol that protects the confidentiality of the data, user's input query, and data access patterns. Also, empirically analyze the efficiency of our protocols through various experiments. These results indicate that secure protocol is very efficient on the user end, and this lightweight scheme allows a user to use any mobile device to perform the kNN query

In Neelam S. Khan [9], 'A Survey on Secure Ranked Keyword Search over Outsourced Encrypted Cloud Data' paper a study of keyword based searching algorithms. In which various searching techniques are used for retrieving the encrypted data from cloud servers. This survey work involves a comparative study of these keyword based searching algorithms which are most helpful to me in my literature review. Neelam S. Khan [9] concludes in paper multi-keyword ranked search MRSE scheme is the best methodology for searching the encrypted data.

IV. SYSTEM DESIGN

A. The *Existing System*

In Semantic-based Compound Keyword Search (SCKS) [1], [2] proposed a compound concept semantic similarity (CCSS) calculation method to measure the semantic similarity between compound concepts. Next, by integrating CCSS with Locality-Sensitive Hashing function and the secure k-Nearest Neighbor scheme. secure k-Nearest Neighbor (SkNN) to encrypt the index and query.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

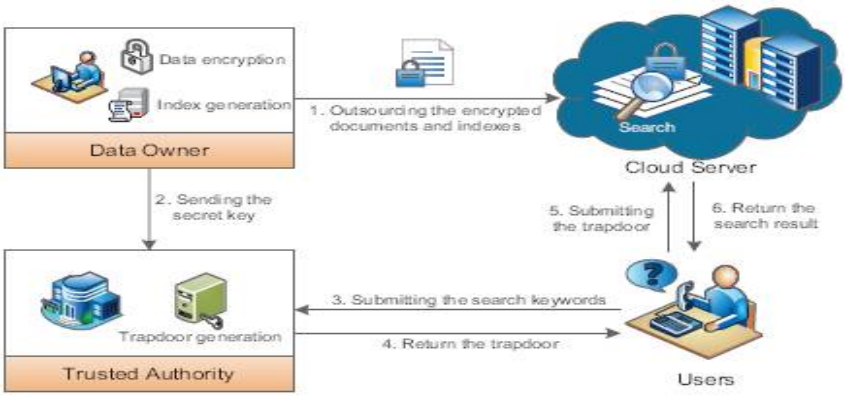

Fig. 2. The model of keyword search [1]

In existing system SkNN adopts asymmetric scalar-product-preserving encryption (ASPE) [1], [2], [15] scheme which preserves scalar product between the query and tuples in database. Due to this rise of various privacy issues. Privacy-preserving classification techniques [4], [16] are not applicable for Encrypted form in Cloud. So we need to focus on solving the classification problem over encrypted data, protects the confidentiality of data, privacy of user's input query and hides the data access patterns.

B. *Proposed System*

In our proposed system we focus on solving the classification problem over encrypted data. In particular, we propose a secure k-nearest neighbor query protocol over encrypted data that protects data confidentiality, user's query privacy, and hides data access patterns. The proposed system is uses SCKS technique and enhancing secure kNN algorithm over encrypted cloud data by developing the semi –honest model.

V. CONCLUSION

In this paper, survey is made on keyword based search used for data retrieval from the outsourced encrypted cloud data. We studied correspondence techniques for keyword based search over encrypted cloud data such as single keyword search, multi-keyword search, multi-keyword ranked search and semantic-based compound keyword search. Our survey concludes that existing techniques usually depends upon a global dictionary, which affects the accuracy of search results and causes inefficiency in data updating except semantic-based compound keyword search (SCKS) technique. SCKS is most efficient technique, which achieves not only semantic-based keyword search but also multi-keyword search and ranked keyword search and eliminates the predefined global library with efficiently support data update. So we propose our work in keyword based search with SCKS technique by enhancing secure kNN algorithm that secure data legacy, queries of user and hide access patterns of data.

REFERENCES

1. Bo Lang, (*Member*, *IEEE*), Jinmiao Wang, Ming Li, and Yanxi Liu, 'Semantic-based Compound Keyword Search over Encrypted Cloud Data', IEEE TRANSACTIONS ON SERVICES COMPUTING, Vol. No.10, Page No. 1 – 14,2018.
2. Harsh Gupta, Kartik Ahirrao, Noopur Sonaje, S.N. More, 'Compound Keyword Search of Encrypted Cloud Data by using Semantic Scheme', International Research Journal of Engineering and Technology (IRJET), Vol. No.05,Issue 12, 2018.
3. Avani Konda, Sai Praneeth Gudimetla, Balaji T, Gopi Krishna Subramanyam, Usha Kiruthika, 'Synonymous Keyword Search Over Encrypted Data in Cloud', International Journal of MC Square Scientific Research Vol.9, No.2, 2017.
4. Abirami Swetha.L,Gokila.R,Mr.Vijaya Ragavan.P, 'EFFECTIVE DATA SEARCH FOR ENCRYPTED RELATIONAL DATA IN CLOUD USING K-NEAREST NEIGHBOR ALGORITHM', International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE), Volume 20, Issue 3, 2016.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

5. Jin Li, Jingwei Li, Xiaofeng Chen, Chunfu Jia, Wenjing Lou (Senior Member, IEEE), 'Identity-Based Encryption with Outsourced Revocation in Cloud Computing', IEEE TRANSACTIONS ON COMPUTERS, Vol. No.64, Paper no.02, Page No.425-437, 2015.
6. Mikhail Strizhov, Indrajit Ray, 'Multi-keyword Similarity Search Over Encrypted Cloud Data', IFIP International Information Security Conference, pp 52-65, 2014.
7. Y. Elmehdwi, B. K. Samanthula, W. Jiang, 'Secure k-nearest neighbor query over encrypted data in outsourced environments', *2014 IEEE 30th International Conference on Data Engineering*, pp. 664–675, 2014.
8. B. Yao, F. Li, X. Xiao, 'Secure nearest neighbor revisited', IEEE 29th International Conference on Data Engineering (ICDE), pp. 733–744, 2013.
9. Neelam S. Khan, Dr. C. Ramakrishna, 'A Survey on Secure Ranked Keyword Search over Outsourced Encrypted Cloud Data', International Conference on Computer Networks and Information Technology (ICCNIT), 2014
10. B. Wang, S. Yu, W. Lou, Y. T. Hou, 'Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud', *IEEE International Conference on Computer Communications*, pp. 2112–2120, 2014.
11. M. Kuzu, M. S. Islam, M. Kantarcioglu, 'Efficient similarity search over encrypted data', IEEE 28th International Conference on Data Engineering (ICDE), pp. 1156–1167, 2012.
12. R. Li, Z. Xu, W. Kang, K. C. Yow, and C. Z. Xu, 'Efficient multikeyword ranked query over encrypted data in cloud computing', *Future Generation Computer Systems*, vol. 30, no. 1, pp. 179–190,2014.
13. Z. Fu, F. Huang, X. Sun, A. Vasilakos, C. N. Yang, 'Enabling semantic search based on conceptual graphs over encrypted outsource data', *IEEE Transactions on Services Computing*, vol. PP,no. 99, pp. 1–1, 2016.
14. C. Wang, N. Cao, J. Li, K. Ren, W. Lou, 'Secure ranked keyword search over encrypted cloud data', *IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, pp.253–262,2010.
15. W. K. Wong, D. W.-l. Cheung, B. Kao, N. Mamoulis, 'Secure kNN computation on encrypted databases', Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 139–152, 2009.
16. H. Hu, J. Xu, C. Ren, B. Choi, 'Processing private queries over untrusted data cloud through privacy homomorphism', Proc. IEEE 27th Int. Conf. Data Eng., pp. 601–612, 2011.