International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Survey on Multi keyword Ranked Search Scheme over Encrypted Data

Vishal Jalindar Gondil, Prof. H. A. Hingoliwala
Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Pune, Maharashtra, India

Article Info
Volume 8 Issue 2
Page Number : 439-445

Publication Issue :
March-April-2021

Article History
Accepted : 15 April 2021
Published : 30 April 2021

ABSTRACT

In recent years, the advancements and the fame of cloud computing are increasing which is actuating the data owners to keep their personal and professional data on public cloud servers like Amazon, Microsoft, Google, Apple, etc with the help of data outsourcing. The other advantage of outsourcing the data over cloud servers is for high benefit and lesser cost in managing the data and the data can be accessed from anywhere and at any time. However, for privacy concerns, the data that are highly sensitive should be encrypted before outsourcing. Taking into consideration the huge amount of data users and files that are present in the cloud, it is important that multiple keywords should be allowed in the searching request and retrieve the files relevant to those keywords. There are some methods and solutions offered to provide privacy and security for the data over the cloud server. Since the document vector's dimension is equal to the dictionary's size, traditional searchable encryption schemes based on the bag-of-words model require a lot of space to store the document set's index. The bag-of-words model often ignores semantic information between keywords and documents, resulting in potentially meaningless search results for users. The natural language processing (NLP) model can be used as it extracts document features from word and paragraph context information. The features can be used to assess document similarity and provide latent semantics information. The NLP model was used to construct a semantic-conscious multi keyword graded search scheme in this survey on dynamic semantic aware multi keyword ranked search.

Keywords : Cloud Computing, Data Outsourcing and security, Natural Language Processing, Multi-keyword Search.

I. INTRODUCTION

Consumer-centric cloud computing, which has evolved in recent years, is a new model for enterprise-level IT that provides on-demand high-quality software and services from a shared group of computing resources. Data storage is a basic service provided by cloud system. By making use of the cloud, the users can be completely released from the troublesome local data storage and maintenance. Also, it also has a significant risk to the confidentiality of those stored files. Specifically, the cloud servers managed by cloud providers are not trusted totally by users while the data files stored in the cloud may be sensitive and confidential, such as business plans. To provide data privacy, as basic solution is to encrypt data files, and then upload the encrypted data into the cloud. Since the Cloud Service Provider (CSP) has full control over the outsourced data, it is likely that it may learn more information from it, which may lead to issues such as data privacy. As a result, confidential data must be encrypted before being sent to a cloud server. Current plaintext search methods, on the other hand, are made useless by the encrypted data. Since cloud users just need to search for the data they're interested in, not any of it, the simple and cumbersome method of downloading all data and decrypting it locally appears to be impractical. As a consequence, finding a reliable and effective search service for encrypted outsourced data is important.

Cloud customers can easily find the most relevant data using proven search methods such as ranked search and multi-keyword search. It also reduces network traffic by sending only the most critical information in response to the user's request. Due to a lack of detailed details about the data, it's likely that in a real-world search situation, the user looks for synonyms of predefined keywords rather than exact or fuzzy keywords. These approaches only support exact or fuzzy keyword search methods. That is, synonym substitution and/or semantic variation, both of which are normal user searching activities, are not tolerated. As a result, multi-keyword ranked search for encrypted cloud data based on semantics remains a difficult issue. Figure 1. Shows the general scenario of multi-keyword search over encrypted data. It contains 3 main entities namely data owner, data receiver and untrusted cloud server.

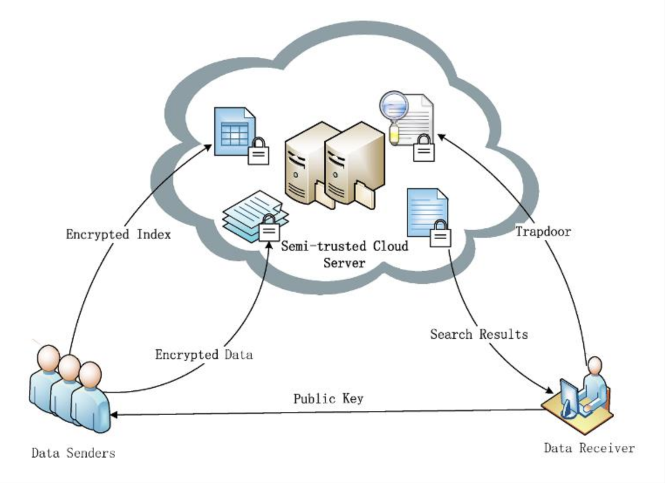

Fig.1 General scenario of multi-keyword search over encrypted data

The data owner and user registers himself at cloud server and then do login with valid username and password in to system. After login, user generates public and private keys for data encryption. The owner encrypts the files using the public keys and uploaded these files at cloud server. Data owner also upload the index tree for searching task. When user want to search keywords / data, he /she generates query and send it to server in return the related document to query gets to user with the help of index tree. After receiving the documents user decrypt, it using private key.

To address this issue of an effective search framework, authors X. Dai, H. Dai, C. Rong, G. Yang and F. Xiao, at el in [1] proposes a system for searchable scheme that includes both multi-keyword ranked search and semantic based search. Multi-keyword search and result ranking are discussed using the Vector Space Model. VSM creates a document index for each document, i.e., each document is represented as a vector, with each dimension value representing the Term Frequency (TF) weight of each corresponding keyword. During the query process, a new vector is created. It has the same dimensions as a document index and the Inverse Document Frequency (IDF) weight as a dimension value. The similarity between the text and the search query is then calculated using the cosine calculation. We expand the keyword collection with semantic terms or natural language words for each keyword to improve the usefulness of the search process. Data retrieval via semantic query will be possible in the future. Even if the consumer is unfamiliar with the exact or synonyms of encrypted data keywords, he or she may try searching for it using natural language. This makes semantic search more effective, and users won't have to worry about the keyword created for each individual term on the cloud. By incorporating this method into the structure, data will be extracted from the cloud in a safe manner, and costs will be reduced.

Z. Xia, X. Wang, X. Sun and Q. Wang at el in [14] proposed the vector space model. They used TF x IDF model for the index construction and query generation. They construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results.

Considering the huge amount of data users and files available in the cloud servers it has been very important to permit keyword search techniques for the data retrieval as requested by the client.

Here in this paper in section II we will study the related work on the multi keyword search in cloud, in section III various technique and algorithm used by different authors are stated., in section IV we conclude the paper.

II. LITERATURE SURVEY

These days, an ever-increasing number of individuals are persuaded to re-appropriate their nearby information to public cloud workers for incredible comfort and diminished expenses in information the board. However, regarding security issues, delicate information ought to be encoded prior to rethinking, which obsoletes conventional information use like catchphrase based archive recovery. Xingming Sun et al [1] present a protected and proficient multi watchword positioned search plot over encoded information, which also upholds dynamic update tasks like erasure and addition of records. In particular, we develop a list tree dependent on vector space model to give multi catchphrase search, which in the interim backings adaptable update tasks. In addition, cosine similitude measure is used to help precise positioning for query output. To improve search proficiency, we further propose a hunt calculation dependent on "Greedy Depth first Traverse Strategy". Besides ensure the pursuit protection, propose a safe plan to meet different security prerequisites in the known code text danger model. Tests on this present reality dataset show the adequacy and effectiveness of proposed plot.

Cloud computing is creating parcel important to give answer for information rethinking and great information administrations. Increasingly more establishment, associations and partnerships are investigating the chance of having their applications, information and their IT resources in cloud. As the information and there cloud's size expands looking of the pertinent information is relied upon to be a test. To defeat this test, search file is made to help in quicker hunt. In any case, search Index creation and calculation has been intricate and tedious, prompting cloud down time there by obstructing the quickness in responding to information demand for strategic prerequisites. Focal point of Kavitha R et al [2] is to clarify how reusability of search record is assisting with diminishing the intricacy of search list com put activity. Search record is proposed to be made utilizing boundaries like closeness importance, client positioning and plan strength. Client positioning assists with ensuring a catchphrase is utilized oftentimes in the transferred information [2]. The proposed framework characterized that the reusability of search record idea decreases cloud devouring time while keeping up the security utilizing accessible symmetric encryption (SSE). The document mentioned from client is gotten from the cloud, utilizing Two-round accessible encryption (TRSE) plot that upholds top k multi-catchphrase recovery.

The benefit of capacity as a help numerous ventures are moving their significant information to the cloud, since it costs less, effectively versatile and can be gotten to from anyplace any time. The client fulfillment in benefit augmentation is also considered as the expense of the cloud. Under the expense, both the leasing cost just as energy use cost additionally considered. To build the benefit there is should diminish the expense. To limit the expense need to arrange the worker precisely. At the hour of worker arrangement, registering is done throughout the expected holding up time just as administration charge. Existing cloud suppliers was used a solitary long haul technique to arrangement cloud stage. In any case, this single long haul leasing technique has the issue of incapable to offer the assistance with the excellent and also drives squandering the assets. To settle this issue, a framework called Double asset Renting (RR) is created by Poonam P. Khot, S.D.Satav et al [4]. This idea incorporates the both present moment just as long haul leasing philosophies. Twofold asset leasing approach guarantees the nature of administration and limits the wastage of assets. The trust between cloud client and supplier is principal. Here security as a boundary is utilized to set up trust. Cryptography is one method of building up trust. Accessible encryption is a cryptographic technique to give security.

In writing numerous scientists have been chipping away at creating proficient accessible encryption plans. Prasanna B T et al [3] investigates some powerful cryptographic procedures dependent on information structures like CRSA and B-Tree to expand the degree of safety. It attempted to execute the pursuit on scrambled information utilizing Azure cloud platform.

Slawomir Grzonkowski et al [5] suggested a security test for the CE cloud administrations' confirmation convention. The convention's latest shortcoming, defeat by Zero Knowledge Proof (ZKP). The client secret phrase is secured using the ZKP system. There is also a SeDici 2.0 elective convention for ZKP depicted here. SeDici 2.0 is a ZKP based on a third-party trusted(TTP) convention. The ZKP's main goal is to have a superior anti-phishing arrangement. The ZKP strategy focuses on phishing and common validation.

A hierarchical clustering for cipher text search in a big data environment is presented by C. Chen et al [6]. The progressive bunching groups the report dependent on least similitude edge. At that point apportioning the resultant group into sub-bunch until a requirement of greatest size of group is reached. During the pursuit stage, this methodology can arrive at a straight computational intricacy against dramatic size of record assortment. The progressive bunching is utilized for better grouping result. In this way enormous assortment of record can be effectively grouped accordingly, improving proficiency of the inquiry. The proposed framework acquires improvement search proficiency, rank security, and the similitude between recovered archives. Secure conjunctive watchword positioned search over scrambled cloud information [7] pick the standard of arrange coordinating with that is utilized to distinguish the likeness among question and information archive. The calculation that are used in the conjunctive watchword search is pailler cryptosystem, Rijndael calculation, cosine comparability search.

A new model, known as multi keyword query encrypted data, was proposed by R. X. Li et al [8]. (MKQE). Use inner product similarity in MKQE to test coordinate matching quantitatively. For each file, MKQE creates an index vector based on the keywords it contains. For index vector encryption and trapdoor creation, two invertible matrices and a bit vector are also used. When a multi-keyword query is sent, MKQE creates a query vector based on the collection of requesting keywords.

When exact match fails, J. Li and Q. Wang [9] use the Fuzzy keyword search approach to improve device compatibility by returning only matching files that have an exact match of the predefined keywords or the closest possible matching files based on keyword similarity semantics steps. They create an advanced technique for constructing fuzzy keyword sets that significantly reduces storage and representation overheads by using the edit distance to compute keyword similarity.

C. Wang et al. [10] proposed the Ranked search, which enhances device usability by returning only matching files to customers in a ranked order based on relevance criteria such as keyword frequency. It demonstrates the inefficiency of the state-of-the-art searchable symmetric encryption (SSE) security concept by providing a simple yet ideal construction of ranked keyword search. They provide a description for graded searchable symmetric encryption and an efficient design by properly leveraging existing cryptographic primitives and order-preserving symmetric encryption to achieve better practical efficiency (OPSE).

N. Cao, et al. [11] proposed a scheme that addresses the problem of privacy preservation in multi-keyword ranked search for encrypted cloud data (MRSE), as well as a collection of strict privacy standards for a safe cloud data utilization system. With different multi-keyword semantics, they extract the similarity score between search query and data documents using the efficient principle of "coordinate matching," i.e., as many matches as possible, and then use "inner product similarity" to quantitatively formalize such a principle for similarity calculation.

W. Sun et al [12] presented multi-keyword text search (MTS) scheme by using similarity-based ranking to resolve the issue of privacy. They deliver two stable index schemes to meet the strict privacy criteria while also incorporating strong threat models to further enhance search privacy. They also proposed a method to construct the search index based on the vector space model to accommodate multi-keyword queries and search result ranking functionalities.

R. P. Rashmi and S. M. Sangve at el in [15] proposes an improved remote data possession checking protocol (RDPC) based on homomorphic hash algorithm. Their proposed system supports secure and efficient dynamic operations at block level. Dynamic operation includes insert, delete, update, and modify. To find the location of each data Merkle Hash Tree (MHT) is used. A third party auditor be called as trusted party auditor is used who checks the user's data stored in cloud storage for its correctness and integrity.

III. TECHNIQUES AND ALGORITHMS

1. THE SSE SCHEME

The SSE scheme stands for searchable symmetrical encryption strategy. The dynamic searchable symmetric encryption (SSE) which consists of a setup with algorithms and the protocols such as the search and update between the server and the client. The static SSE is similar to that of the dynamic SSE but with no update protocol.

2. THE CIPHER TEXT

The cipher text, which is used to provide the security for an encrypted data. The cloud server can access only the encrypted documents and the indexes which are secure.

3. THE TREE BASED ALGORITHM

The tree based algorithm is accommodated from the MDBtree based MD algorithms, which enables multi-keyword ranked search. The key factors which affects the search efficiency is identified and certain schemes in constructing the index tree to efficiently increase the search is provided.

4. THE LSH

The Locality Sensitive Hashing (LSH) is an algorithm for the estimation for closer neighbor searching in the higher dimensional spaces. The LSH is used for mapping hash functions that are present as a set to the objects to many buckets so that the homogeneous objects shares the bucket with higher probabilities, while the non-homogeneous one does not. The LSH utilizes the locality sensitive function family for this purpose.

5. THE KNN SCHEME

In the KNN scheme, the eucledian distance between the data records and the query vectors is used for selecting the K-nearest neighbor database records. The inner product computation is done using the secure KNN scheme.

6. PUBLIC KEY SEARCH ENCRYPTION TECHNIQUE

The public key search encryption technique is a technique which allows user to encrypt data and send it to the cloud server. The data owner provides decryption key which might be different.

7. THE SECURE INDEX SCHEME

The secure index scheme constructs a secure index for files. The secure index permits the data user for searching an encrypted file which is having a keyword without decrypting the file.

8. RANKED SEARCHABLE SYMMETRIC ENCRYPTION

The ranked searchable symmetric encryption enhances the usability of the system by giving back the similar files in an order that is ranked in regard to certain criteria that is relevant.

9. Doc2Vec MODEL

The Doc2Vec model uses features extracted from a document set to perform semantic search. It also uses Natural Language Processing to understand the meaning of words.

10. SEMANTIC-AWARE MULTI-KEYWORD RANKED SEARCH

A privacy-preserving searchable encryption scheme based on the LDA topic model and the query likelihood model which extract the feature keywords from the document using the LDA-based Information Gain (IG) and Topic Frequency-Inverse Topic Frequency (TF-ITF) model.

IV. CONCLUSION

In this survey paper we have summarized different kinds of multi keywords searching methods for encrypted cloud data. A study on the data privacy techniques and issues in various searching techniques are covered such as efficiently result ranking, query privacy, etc. Several searching techniques have been studied for the efficient retrieval of data or files or documents over the encrypted cloud data. From above survey we can say that semantic aware techniques work better compare to traditional encrypted searching techniques.

V. REFERENCES

[1]. Xingming Sun, Xinhui Wang, Zhihua Xia, Zhangjie Fu and Tao Li Jiangsu ,"Dynamic Multi-keyword Top-k Ranked Search over Encrypted Cloud Data Engineering Center of Network Monitoring", Nanjing University of Information Science & Technology, Nanjing, 210044, China sunnudt@163.com, wxh_nuist@163.com, xia_zhihua@163.com, wwwfzj@126.com

[2]. Kavitha R1, R J Poovaraghan," Reusability of Search Index over Encrypted Cloud Data on Dynamic update" SRM University, Chennai, India1 Assistant Professor (OG), Department of

Computer Science, SRM University, Chennai, India2

[3]. Prasanna B T, C B Akki, “Dynamic Multi-Keyword Ranked Searchable Security Algorithm Using CRSA and B-Tree”, Department of ISE, EPCET Associate Professor, Bengaluru, INDIA-560049, Department of ISE, SJBIT Professor, Bengaluru, INDIA-560060

[4]. Khot, Poonam P. and S. Satav. “A Profit Maximization Scheme for Enhancing Quality of Service (QoS) in Cloud Computing.” International Journal of Computer Applications 145 (2016): 35-39.

[5]. S. Grzonkowski, P. M. Corcoran, and T. Coughlin, “Security analysis of authentication protocols for nextgeneration mobile and CE cloud services,” in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp. 83-87.

[6]. C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, “A hierarchical clustering method For big data oriented ciphertext search,” in Proc. IEEE INFOCOM, Workshop on Security and Privacy in Big Data, Toronto, Canada, 2014, pp. 559-564.

[7]. 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.

[8]. R. X. Li, Z. Y. Xu, W. S. Kang, K. C. Yow, and C. Z. Xu, “Efficient Multi-keyword ranked query over encrypted data in cloud computing, Futur. Gener. Comp. Syst., vol. 30, pp. 179-190, Jan. 2014

[9]. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," Proceedings of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, pp. 1-5, Mar. 2010.

[10]. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 253-262, 2010.

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

[12]. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou, "Privacy preserving multi-keyword text search in the cloud supporting similarity based ranking," ASIACCS 2013, Hangzhou, China, May 2013, pp. 71-82, 2013.

[13]. X. Dai, H. Dai, C. Rong, G. Yang and F. Xiao, "Enhanced Semantic-Aware Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data," in IEEE Transactions on Cloud Computing, doi: 10.1109/TCC.2020.3047921.

[14]. Z. Xia, X. Wang, X. Sun and Q. Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340-352, 1 Feb. 2016, doi: 10.1109/TPDS.2015.2401003.

[15]. R. P. Rashmi and S. M. Sangve, "Public auditing system: Improved remote data possession checking protocol for secure cloud storage," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015, pp. 75-80, doi: 10.1109/ICATCCT.2015.7456858.

Cite this article as :

Vishal Jalindar Gondil, Prof. H. A. Hingoliwala, "Survey on Multi keyword Ranked Search Scheme over Encrypted Data", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 2, pp. 439-445, March-April 2021.

Journal URL : https://ijsrset.com/IJSRSET2182112