International Journals of Advanced Research in Computer Science and Software Engineering
ISSN: 2277-128X (Volume-7, Issue-7)

Research Article | **July 2017**

Study of Dynamic Multikeyword Text Search Techniques over Encrypted Data in Cloud

Ankita Puri, Naveen Kumari
Punjabi University Regional Centre of Information and Technology Management, Mohali, Punjab, India

DOI: 10.23956/ijarcsse/V7I7/0110

Abstract — Day by Day ,with the advancement of modern technology over cloud computing motivating the data owners to outsource their data to the cloud server like Amazon, Microsoft, Azure etc .With the help of data outsourcing ,the organization can provide reliable data services to their user without any management of the overhead concern. Suppose, a large number of users that are on cloud and large number of documents on cloud, Its important for the service provider to allow multi-keyword query and provided the result that meet efficient data retrieval needs. In this paper, for the first time, we define and solve the challenging problem of privacy preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement.

Keywords— Cloud Computing, Multi Keyword Search, Ranked Search, Encryption method.

I. INTRODUCTION

CLOUD computing has been considered as another model of enterprise IT infrastructure, which can compose gigantic resource of computing, storage and applications, and empower users to appreciate pervasive, helpful and instant access to network through mutual pool of configurable computing resources with incredible efficiency and insignificant economic overhead [1]. Because of these engaging features, both individuals and enterprises are incited to outsource their data to the cloud, rather than buying software and hardware to deal with the data themselves. In spite of the different points of interest of cloud services, outsourcing delicate information, (for example, e-mail, individual health records, organization account information, government archives, and so forth.) to remote servers that generate issue with privacy concerns. The cloud service providers (CSPs) that keep the data for users may access users' sensitive information without authorization. A general way to deal with secure the data privacy is to encrypt the data before outsourcing [2]. On the other hand, this will bring about a huge expense in terms of data ease of use. For example, the current techniques on keyword-based information retrieval, which are mostly outlined on the basis of plaintext data, which can't be directly concerned on the encrypted data. Downloading all the data from the cloud and decrypt locally is clearly unrealistic. With a particular final objective to address the above issue, analysts have illustrated some all around helpful arrangements with totally homomorphic encryption [3] or missing RAMs [4]. In any case, these schedules are not down to earth in light of their high computational overhead for both the cloud server and user. In spite of what may be normal, more useful unique reason arrangements, for instance, searchable encryption (SE) plan have made specific responsibilities to the extent productivity, value and security. Searchable encryption scheme engage the user to store the encrypted data to the cloud and execute unequivocal word look for over ciphertext domain. As being what is indicated, complicated works have been proposed under assorted risk models to finish distinctive interest value. For instance single keyword search, closeness look, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multikeyword positioned quest finishes more thought for its pragmatic propriety. Starting late, some component arrangements have been proposed to reinforce embedding and erasing operations on archive gathering. These are colossal goes about as it is exceptionally possible that the data owner need to overhaul their data on the cloud server. Yet, few of the dynamic plan support successful multikeyword situated look. Inverse document recurrence (IDF)" model are joined in the list development and inquiry era to give multikeyword positioned seek. Keeping in mind the end goal to get high search Effectiveness, we develop a tree based list structure and based on this tree list we propose a "Greedy Depth –first Search" calculation. Because of the uncommon structure of our tree-based list, the proposed search scheme can flexibly accomplish sub-straight search time and manage the deletion and insertion of reports. To encrypt the index and query vectors we use the knn protected algorithm, and in the interim guarantee relevance score calculation between encrypted index and query vectors. Attempt to prevent such typical attacks on different threat models, we build two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known ciphertext model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.

II. MULTI KEYWORD RANKED SEARCH IN CLOUD

With a specific end goal to meet the down the requirements prerequisites, look over scrambled information that need to accompanying three capacities. In the first place, the searchable encryption plans to check multi-watchword seek, and

give a similar client encounter as looking in Google seek with various catchphrases; single-catchphrase hunt is a long way from acceptable by just returning extremely constrained and erroneous inquiry results.[1] Second, to rapidly recognize most applicable outcomes, the pursuit client would commonly incline toward cloud servers, so that it sort the returned list items in a sequential manner ,as the request positioned by the importance of the hunt demand to the records. Furthermore, demonstrating the positioned inquiry to clients can likewise dispense with the superfluous system activity by just sending back the most important outcomes from cloud to seek users.[1] Third, concerning the hunt effectiveness, since the quantity of the archives contained in a database could be remarkably vast, searchable encryption plans needs to be productive to rapidly react to the pursuit demands with least deferrals.

III. RELATED STUDY

Chi Chen et al. [1] developed the searchable encryption for multi-keyword ranked search over the storage data. Specifically, by taking the huge number of outsourced documents (data) on the cloud, algorithm utilize the k-nearest neighbor and relevance score techniques to design an efficient multi-keyword search scheme that can give back the ranked search results based on the accuracy. This system improve the search efficiency by supporting efficient index, and hide access pattern of the search user by adopt the blind storage system.

Keerthana G et al. [2]fabricated an application for enhancing cloud security utilizing partition and encryption technique which will enhance the cloud security. First of all record from client were taken and partition it into number of parts. After partition we encrypt the all record parts. At that point we send record parts to various cloud servers. At the point when client need that information back we take that information from cloud servers and decrypt that information. After decrypting, merging of that information is done and offer it to client. Our objective is that the application ought to have straight forward client interface for clients adaptability. The proficient method for giving security is the utilization of hybrid cryptography for more secured sending and receiving of information.

K.Ramadevi and L.Sunitha Rani [3]proposed schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors and systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel Additive Order and Privacy Preserving Function family. To prevent the attackers from eaves dropping secret keys and pretending to be legal data users submitting searches, we propose a novel dynamic secret key generation protocol and a new data user authentication protocol. Furthermore, PRMSM supports efficient data user revocation.

Khnd Sri Sandhya and K. Venkat Rao [4]scheme is based on multi-keyword ranked search which supports dynamic update operations. The data owner generates an exceptional tree-based catalog composition together with “Greedy Depth-first Search” criteria to make successful multi-keyword search. Achieving parallelism is the limitation of the existing system. The described technique extended the existing scheme with secure Dynamic Key generation along with the vector space model and it also includes TF_IDF model for index development as well for query generation. The dynamic key generation favors parallel search process by allowing multiple users to retrieve the same encrypted cloud data

Veerraju Gampala and Sreelatha Malempati [5]employed probabilistic public key encryption algorithm for encrypting the data and invoke ranked keyword search over the encrypted data to retrieve the files from the cloud. This technique aimed to achieve an efficient system for data encryption without sacrificing the privacy of data. Further, this ranked keyword search greatly improves the system usability by enabling ranking based on relevance score for search result, sends top most relevant files instead of sending all files back, and ensures the file retrieval accuracy. Thorough security and performance analysis, we prove that our approach is semantically secure and efficient.

Shreejit Pillai et al. [6]proposed schemes to deal with privacy preserving ranked multi keyword search in a multi owner environment to enable cloud servers to perform secure search without knowing the actual data of both keywords systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance course between keywords and files, and proposed a novel additive order and privacy preserving function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data owners submitting searches proposed a dynamic secret key generation key protocol and a new data user authentication protocol.

Shrilakshmi Prasad and B. S. Mamatha [7] defined and solve the problem of association attack by encrypting the index file using Paillier cryptographic algorithm. So cloud will have the challenge of searching the index file with the search query where both will be in an encrypted format. Hence privacy of the document will be preserved. Cosine similarity search is used to retrieve the top matching documents based on their relevance score and the beauty of the proposed system is the user can give multiple keywords in their search query. The enterprises are interested in storing their data in the public cloud. Before uploading the data on to the cloud, it needs to be encrypted to preserve privacy. In order to ease searching, the index file should be built for each document. The index file contains the keyword and its count in the particular document. The unencrypted index file leads to association attack since with the keywords and their count, the content of the document can be known.

Seema Ranga and Ajay Jangra [8]aimed of Intrusion detection System is to defend the security of the Computer system by a layer over the defense system. IDS systems sense the misuse, breach in the security system and also the malicious or unauthorized access to the system. Although Firewalls works for the same reason but the major difference between firewalls and the IDS is IDS suspect the source of the attack and signals the alarm to the system but a firewall directly stops the communication without informing the system. These attacks requires true concerns as they harm the data stored in system and also effect the network traffic, data packet etc.

Wei Zhang et al. [9]proposed scheme to deal with privacy preserving ranked multi-keyword search in a multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, algo systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files and proposed a novel additive order and privacy preserving function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data users submitting searches, this approachdeveloped a novel dynamic secret key generation protocol and a new data user authentication protocol.

Izzat Alsmadi and Ikdam Alhami [10] defined that information users depend heavily on emails' system as one of the major sources of communication. Its importance and usage are continuously growing despite the evolution of mobile applications, social networks, etc. Emails are used on both the personal and professional levels. They can be considered as official documents in communication among users. Emails' data mining and analysis can be conducted for several purposes such as: Spam detection and classification, subject classification, etc. In this paper, a large set of personal emails is used for the purpose of folder and subject classifications. Algorithms are developed to perform clustering and classification for this large text collection. Classification based on NGram is shown to be the best for such large text collection especially as text is Bi-language

S. No	Paper	Methods
1	"A Secure and Dynamic Multi Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, February 2016 [16]	Scheme supports dynamic update operations like deletion of documents and insertion of documents. Tree-based index structure and "Greedy Depth First Search" algorithms are use to provide efficient multi-keyword ranked search.
2	"Enabling Fine-Grained Multi Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, May/June 2016 [17]	Relevance scores and preference factors of keywords use to enable precise keyword search and personalized user experience. Support complicated logic search by using the mixed "AND", "OR" and "NO" operations of keywords. Classified sub-dictionaries technique is used to achieve better efficiency on index building, trapdoor generating and query.
3	"Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement", IEEE Transactions on Parallel and Distributed Systems, September 2016 [18]	By using the user search history, a user interest model is build for individual user with the help of semantic ontology WordNet. The user interest model is use to realize automatic evaluation of the keyword priority and it solved the limitation of the artificial method of measuring.
4	"An Efficient File Hierarchy Attribute Based Encryption Scheme in Cloud Computing", IEEE Transactions on Information Forensics and Security, June 2016 [19]	Uses Ciphertext-policy attribute-based encryption (CP-ABE) encryption technology to solve the challenging problem of secure data sharing in cloud computing. Efficient file hierarchy attribute-based encryption scheme is proposed in cloud computing.
5	"Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Information Forensics and Security, January 2014 [20]	Propose two MRSE schemes based on the similarity measure of "coordinate matching" to provide as many matches as possible to effectively capture the relevance of outsourced documents to the query keywords while meeting different privacy requirements. "Inner product similarity" is used to quantitatively evaluate similarity measure.

Comparative Study for various mylti-keyword searching techniques [15]

IV. METHODS FOR STRATEGIES AND CALCULATION OF KEYWORD EXTRACTION

A portion of the models, strategies and calculations being utilized as a part of the current framework are talked about and concise as takes after. [13]

Vector Space Model: This model is utilized to speak to the content by a vector of capacities. The terms are the words and expressions. On the off chance that words are considered as terms, each word turns into a free measurement in a high measurement vector space. On the off chance that term speaks to a content, it gets a non-zero an incentive in the content vector along the measurement comparing to the term. Content vectors are exceptionally space and no term is prior rank as negative esteem.

Probabilistic Model: The guideline of probabilistic model is that the archives in a gathering support to be positioned by diminished likelihood to inquiry significance. This guideline is called as the probabilistic positioning guideline. The positioning paradigm is monotonic under log-odd changes. Each probabilistic model that is proposed depends on an alternate probabilistic estimation system.

Inference Network Model: A model that is utilized for a report to instantiate a term. The credit from different terms is collected given to figure what might as well be called a numeric score for information.

Term Weighting: Term weighting is a procedure that depends upon the better estimation of different probabilities. The fundamental three variables play in term weight plan is:
Term Frequency - Words that rehash various circumstances in a record.
Record Frequency - Words that show up in numerous records are viewed as normal.
Record Length - When accumulation have archives of differing lengths, longer reports impact to score higher since they contain more words and more reiteration.

Searchable Encryption Algorithm: A calculation that comprises of the polynomial time randomized calculations. They are:
KeyGen(s) - s is a security parameter taken and used to produce a key match either open or private.
PEKS (Apub, w) - Apub is an open key and w is a word which are utilized to deliver a searchable encryption.
Trapdoor (Apriv, w) - Apriv is a private key and w is a word which are utilized to deliver a trapdoor Tw.

Cipher content Security: It is a system that is utilized to give security to the encoded information. A figure content aggressor could without much of a stretch break semantic security by reordering the watchwords and submitting the subsequent figure content for decoding. A standard procedure is utilized to break this and this method is known as the figure content security.

Private Key Searchable Encryption: A model called private key searchable encryption is utilized to look on a private key encoded information. The client himself encodes information, in order to compose in a discretionary way.

Public Key Searchable Encryption: Open key searchable encryption is a model that permits client to encode information and send it to the server. The proprietor gives unscrambling key might be distinctive.

Algorithm	Constraint	Speed	Time Taken
Greedy DFS	NA	Fast	Less
MHR Tree	NA	Faster	Much Less
MDB Search	NA	Average	Average
Sequential search	Data should be limited	Comparative slow	Comparative more
Binary search	Sorted data needed	Comparative slow	Comparative more

Comparison between Searching Techniques [14]

V. CONCLUSION

In this paper a review on various multikeyword setext search technique over encrypted cloud is studied. We assemble a special keyword balanced binary tree as the index, and intend a "Greedy Depth-first Search" algorithm to acquire preferable effectiveness over linear search. Likewise, the parallel search procedure can be completed to further lessen the time cost. Among the above defined technique Sequential and binary searach are least effective where as the Greedy DFS and MHR trees are efficient methods for Multi Keyword Ranked Searching. IN the multi-keyword ranked search some of the algorithm use K-Means clustering technique which has many drawbacks like efficient centroid selection, outliers. In the proposed work the drawback of K-Mean clustering may be removed for an efficient search algorithm.

REFERENCES

[1] C. Chen, X. Zhu, P. Shen, J. Hu, S. Guo, Z. Tari, A. Y. Zomaya, "An Efficient Privacy-Preserving Ranked Keyword Search Method", IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 4, 2016

[2] G. Keerthana, S. Prabu, P. Swarnalatha, "An Efficient Data Security in Cloud Computing using Cryptography", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 5, 2016, pp: 654-660

[3] K. Ramadevi, L. S. Rani, "Ranked keyword search over Cloud Storage by several owners using dynamic hidden keys", International Journal of Computer Science, ISSN: 2348-6600, Volume 4, Issue 2, No 7, 2016, pp: 894-900

[4] K. S. Sandhya, K. V. Rao, "Privacy-Preserving and Dynamic Multikey Generation over Encrypted Cloud Data", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 10, 2016, pp: 78-82

[5] S. K. Pasupuleti, S. Ramalingam, R. Buyya, "An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing", Journal of Network and Computer Applications, ISSN: 1084-8045, Vol: 64, 2016, pp: 12-22

[6] S. Pillai, G. Ransing, N. Ransing, S. Markad, N. Sable, "Survey on Privacy Preserving Multi Keyword Search in Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering, ISSN (Online) 2278-1021 ISSN (Print) 2319 5940, Vol. 5, Issue 11, 2016

[7] S. Prasad, B. S. Mamatha, "Retrieving documents from encrypted cloud data in a secured way using cosine similarity search with multiple keyword search support", International Journal of Advance Research in Computer Science and Management Studies, ISSN: 2321-7782 (Online), Volume 4, Issue 5, 2016, pp: 108-115

[8] S. Ranga, A. Jangra, "A Study of IDS Technique Using Data Mining", International Journal of Technical Research & Science, ISSN No.: 2454- 2024 (online), Volume 1 Issue 6, 2016, pp: 152-158

[9] W. Zhang, Y. Lin, S. Xiao, J. Wu, S. Zhou, "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing", IEEE transaction on Computers, ISSN: 0018-9340, Vol: 65, Issue: 5, 2016, pp: 1566-1577

[10] I. Alsmadi, I. Alhami, "Clustering and classification of email contents", Journal of King Saud University – Computer and Information Sciences,ISSN: 1319-1578, Vol: 27, 2015, pp: 46-57

[11] C. Liu, Z. Peng, L. Wu, "Role of Time-Domain Based Access Control Model", Journal of Software Engineering and Applications, Volume: 9, 2016, pp: 57-62

[12] A. Shaheen, A. Slei, "Comparing between different approaches to solve the 0/1 Knapsack problem", International Journal of Computer Science and Network Security, Volume: 16, No: 7, July 2016, pp: 1-10

[13] Jaikishan Tindwani, Aruna Gupta, "A Survey on Multi-Keyword Ranked Query Search over Encrypted Cloud Storage", International Journal of Science and Research, ISSN (Online): 2319-7064, Volume 4 Issue 11, November 2015, pp: 2366-2370

[14] SonamDarda, Manasi. K. Kulkarni, "Study of Multi-keyword Ranked Searching and Encryption Technique over Cloud",) International Journal of Computer Science and Information Technologies, ISSN: 0975-9646, vol: 6, issue: 6, 2015, pp: 5417-5420

[15] Kalyani Sonawane, Rahul Dagade, "A Survey on Multi-Keyword Ranked Search over Encrypted Cloud Data with Multile Data Owners", International Journal of Computer Applications, ISSN: 0975–8887, vol: 162, no: 11, 2017, pp: 9-12

[16] Zhihua Xia, Xinhui Wang, Sun Xingming, Qian Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, vol: 27, issue: 2, 2015

[17] Hongwei Li, Yi Yang, Tom H. Luan, Xiaohui Liang , Liang Zhou, Xuemin Sherman Shen, "Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, Vol: 13, Issue: 3, 2016, pp: 312 – 325

[18] Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, Fengxiao Huang, "Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement", IEEE Transactions on Parallel and Distributed Systems, Vol: 27, Issue: 9, 2016, pp: 2546-2559

[19] Shulan Wang, Junwei Zhou, Joseph K. Liu, Jianping Yu, Jianyong Chen, Weixin Xie, "An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing", IEEE Transactions on Information Forensics and Security, Vol: 11, Issue: 6, 2016, pp: 1265-1277

[20] Ning Cao, Cong Wang, Ming Li, Kui Ren, Wenjing Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Vol: 25, Issue: 1, 2014, pp: 222-233

