ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

Fastened Dynamic Multi Keyword Searching Mechanism for Secure Data in Cloud

Sarun J. Varghese[1], Dr. M. Nithya[2]

M.E. Student, Department of CSE, VMKV Engineering College, Salem, Tamilnadu, India[1]

Professor & HOD, Department of CSE, VMKV Engineering College, Salem, Tamilnadu, India[2]

ABSTRACT: A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TFIDF model are combined in the index construction and query generation. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this paper. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

KEYWORDS:Keyword based document retrieval,hierarchical approach, Run time Estimation

I. INTRODUCTION

The goal of CloudAudit is to provide cloud service providers with a way to make their performance and security data readily available for potential customers. The specification provides a standard way to present and share detailed, automated statistics about performance and security. We consider data storage and sharing services in the cloud with three entities: the cloud, the third party auditor (TPA), and users who participate as a group. Users in a group include one original user and a number of group users. The original user is the original owner of data, and shares data in the cloud with other users. Based on access control policies, other users in the group are able to access, download and modify shared data. The cloud provides data storage and sharing services for users, and has ample storage space.The third party auditor is able to verify the integrity of shared data based on requests from users, without downloading the entire data. When a user (either the original user or a group user) wishes to check the integrity of shared data, she first sends an auditing request to the TPA. After receiving the auditing request, the TPA generates an auditing message to the cloud, and retrieves an auditing proof of shared data from the cloud. Then the TPA verifies the correctness of the auditing proof. Finally, the TPA sends an auditing report to the user based on the result of the verification. So we propose TPA only needs to hold an encrypted version of the client's secret key, while doing all these burdensome tasks on behalf of the client. The client only needs to download the encrypted secret key from the TPA when uploading new files to cloud. Besides, our design also equips the client with capability to further verify the validity of the encrypted secret keys provided by TPA. Encrypt the file using blowfish algorithm and then store into the cloud.the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi keyword ranked search and dynamic operation on the document collection. We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection. Due to

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic.

II. RELATED WORK

Ranked search can enable quick search of the mostrelevant data. Sending back only the top-k most relevantdocuments can effectively decrease network traffic. Some early works have realized the rankedsearch using order-preserving techniques, but they aredesigned only for single keyword search. Cao et al. realized the first privacy-preserving multi-keywordranked search scheme, in which documents and queriesare represented as vectors of dictionary size. With the"coordinate matching", the documents are ranked accordingto the number of matched query keywords.However, Cao et al.'s scheme does not consider theimportance of the different keywords, and thus is notaccurate enough. In addition, the search efficiency of thescheme is linear with the cardinality of document collection.Sun et al. presented a secure multi-keywordsearch scheme that supports similarity-based ranking.The authors constructed a searchable index tree basedon vector space model and adopted cosine measuretogether with TF×IDF to provide ranking results. Sun etal.'s search algorithm achieves better-than-linear searchefficiency but results in precision loss. Multi-keyword Boolean search allows the users toinput multiple query keywords to request suitable documents.Among these works, conjunctive keyword searchschemes only return the documents thatcontain all of the query keywords.The existing techniques on keyword-based information retrieval, whichare widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality.However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval.

Disadvantages:

- Support only exact matching in the context of keyword search.
- The ranked search does not differentiate documents with higher number of repeated terms than documents with lower number of repeated terms.

III. PROPOSED SYSTEM

A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data We explore the problem of maintaining the semantic relationship between different plain documents over the related encrypted documents and give the design method to enhance the performance of the semantic search. We also propose the Hierarchical Clustering Algorithm to adapt to the requirements of data explosion, online information retrieval and semantic search. At the same time, a Verifiable mechanism is also proposed to guarantee the correctness and completeness of search results. Built to evaluate the search efficiency, accuracy, and rank security. The experiment result proves that the proposed architecture not only properly solves the multi-keyword ranked search problem, but also brings an improvement in search efficiency, rank security, and the relevance between retrieved documents.

Advantages:

- We proposed the Hierarchical Clustering Algorithmto speed up server-side searching phase. Accompanying with the exponential growth of document collection, the search time is reduced to a linear time instead of exponential time.
- We design a search strategy to improve the rank privacy. This search strategy adopts the backtracking algorithm upon the above clustering method. With the growing of the data volume, the advantage of the proposed method in rank privacy tends to be more apparent.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

IV. SYSTEM ARCHITECTURE

Fig.1. Architecture of Multi keyword Ranked Search over Encrypted Cloud Data

The system model in this paper involves three differententities: data owner, data user and cloud server, asillustrated.

Data owner has a collection of documents F ={f} that he wants to outsource to the cloudserver in encrypted form while still keeping the capabilityto search on them for effective utilization. In ourscheme, the data owner firstly builds a secure searchabletree index I from document collection F, and thengenerates an encrypted document collection C for F.Afterwards, the data owner outsources the encryptedcollection C and the secure index I to the cloud server,and securely distributes the key information of trapdoorgeneration (including keyword IDF values) and documentdecryption to the authorized data users.1; f2; :::; fn Besides, the data owner is responsible for the updateoperation of his documents stored in the cloud server.While updating, the data owner generates the updateinformation locally and sends it to the server.

Data users are authorized ones to access the documentsof data owner. With t query keywords, theauthorized user can generate a trapdoor TD accordingto search control mechanisms to fetch k encrypted documentsfrom cloud server. Then, the data user can decryptthe documents with the shared secret key.

Cloud server stores the encrypted document collectionC and the encrypted searchable tree index I for dataowner. Upon receiving the trapdoor TD from the datauser, the cloud server executes search over the index treeI, and finally returns the corresponding collection of topkranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection Caccording to the received information. We analyse the BDMRS schemeaccording to the three predefined privacy requirementsin the design goals:

1) Index Confidentiality and Query Confidentiality: In theproposed BDMRS scheme, Iand TD are obfuscatedvectors, which mean the cloud server cannotinfer the original vectors Duand Q without thesecret key set SK. The secret keys

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

Mu1and MareGaussian random matrices. According to, theattacker (cloud server) of COA cannot calculate thematrices merely with cipher text. Thus, the BDMRSscheme is resilient against cipher text-only attack(COA) and the index confidentiality and the queryconfidentiality are well protected.

2) Query Unlinkability: The trapdoor of query vectoris generated from a random splitting operation,which means that the same search requests willbe transformed into different query trapdoors, andthus the query unlinkability is protected. However,the cloud server is able to link the same searchrequests according to the same visited path and thesame relevance scores.

3) Keyword Privacy: In this scheme, the confidentialityof the index and query are well protected that theoriginal vectors are kept from the cloud server. Andthe search process merely introduces inner productcomputing of encrypted vectors, which leaksno information about any specific keyword. Thus,the keyword privacy is protected in the knowncipher text model. But in the known backgroundmodel, the cloud server is supposed to have moreknowledge, such as the term frequency statistics ofkeywords. This statistic information can be visualizedas a TF distribution histogram which revealsshow many documents are there for every TF valueof a specific keyword in the document collection.Then, due to the specificity of the TF distributionhistogram, like the graph slope and value rangethe cloud server could conduct TF statistical attackto deduce/identify keywords. In theworst case, when there is only one keyword in thequery vector, i.e. the normalized IDF value for thekeyword is 1, the final relevance score distribution is exactly the normalized TF distribution of thiskeyword, which is directly exposed to cloud server.Therefore, the BDMRS scheme cannot resist TFstatistical attack in the known background model.

- **Hierarchical Clustering Algorithm**

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. The quality of a pure hierarchical clustering method suffers from its inability to perform adjustment, once a merge or split decision has been executed. Then it will neither undo what was done previously, nor perform object swapping between clusters. Thus merge or split decision, if not well chosen at some step, may lead to some-what low-quality clusters. One promising direction for improving the clustering quality of hierarchical methods is to integrate hierarchical clustering with other techniques for multiple phase clustering. So in this paper, we describe a fewimproved hierarchical clustering algorithms that overcome the limitations that exist in pure hierarchical clustering algorithms.Cluster Analysis (data segmentation) has a variety of goals that relate to grouping or segmenting a collection of objects (i.e., observations, individuals, cases, or data rows) into subsets or clusters, such that those within each cluster are more closely related to one another than objects assigned to different clusters. Central to all of the goals of cluster analysis is the notion of degree of similarity (or dissimilarity) between the individual objects being clustered. There are two major methods of clustering: hierarchical clustering and k-means clustering. For information on k-means clustering, refer to the k-Means Clustering section. In hierarchical clustering, the data is not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters that each contains a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by a series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings.

V. SIMULATION AND RESULT

WORK TO BE DONE IN PHASE II:

- We proposed the Hierarchical Clustering Algorithmto speed up server-side searching phase.
- Accompanying with the exponential growth of document collection, the search time is reduced to a linear time instead of exponential time.
- We design a search strategy to improve the rank privacy.
- This search strategy adopts the backtracking algorithm upon the above clustering method.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

- With the growing of the data volume, the advantage of the proposed method in rank privacy tends to be more apparent.

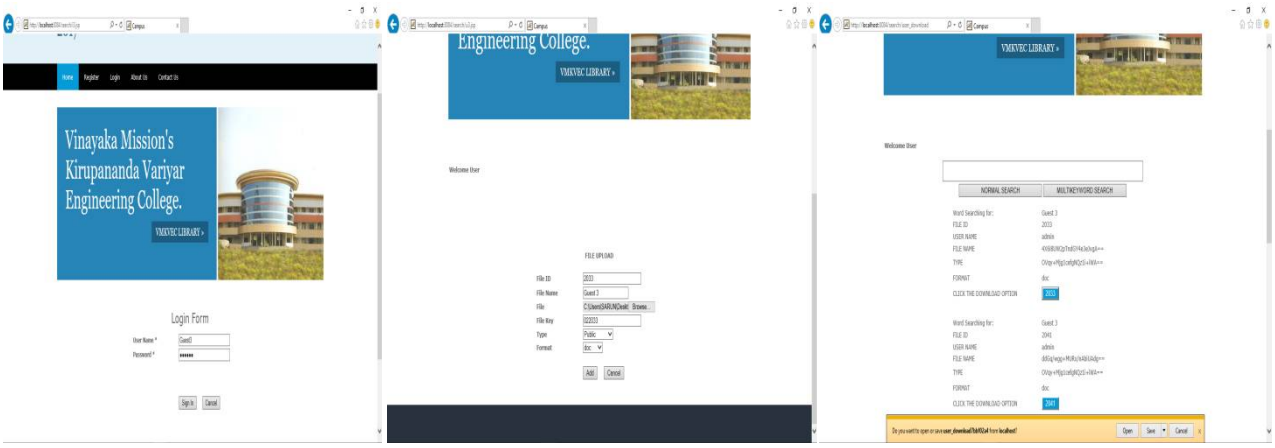

Fig2. Login, Upload, Search and Download Files

VI. CONCLUSION AND FUTUREWORK

A secure, efficient and dynamic searchscheme is proposed, which supports not only the accuratemulti-keyword ranked search but also the dynamicdeletion and insertion of documents. We constructa special keyword balanced binary tree as the index,and propose a "Greedy Depth-first Search" algorithm toobtain better efficiency than linear search. In addition,the parallel search process can be carried out to furtherreduce the time cost. The security of the scheme isprotected against two threat models by using the secure KNN algorithm. Experimental results demonstrate theefficiency of our proposed scheme.There are still many challenge problems in symmetricSE schemes. In the proposed scheme, the data owneris responsible for generating updating information andsending them to the cloud server. Thus, the data ownerneeds to store the unencrypted index tree and the informationthat are necessary to recalculate the IDF values.Such an active data owner may not be very suitable forthe cloud computing model. It could be a meaningfulbut difficult future work to design a dynamic searchableencryption scheme whose updating operation can becompleted by cloud server only, meanwhile reservingthe ability to support multi-keyword ranked search. Inaddition, as the most of works about searchable encryption,our scheme mainly considers the challenge from thecloud server. Actually, there are many secure challengesin a multi-user scheme. Firstly, all the users usuallykeep the same secure key for trapdoor generation in asymmetric SE scheme. In this case, the revocation of theuser is big challenge. If it is needed to revoke a user in this scheme, we need to rebuild the index and distributethe new secure keys to all the authorized users. Secondly, symmetric SE schemes usually assume that all the datausers are trustworthy. It is not practical and a dishonestdata user will lead to many secure problems. For example,a dishonest data user may search the documents anddistribute the decrypted documents to the unauthorizedones. Even more, a dishonest data user may distributehis/her secure keys to the unauthorized ones. In thefuture works, we will try to improve the SE scheme tohandle these challenge problems.

REFERENCES

[1] K. Ren, C. Wang, Q. Wang et al., "Security challenges for the publiccloud," IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," inFinancial Cryptography and Data Security.Springer, 2010, pp. 136–149.
[3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D.dissertation, Stanford University, 2009.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

[4] O. Goldreich and R. Ostrovsky, "Software protection and simulationon oblivious rams," Journal of the ACM (JACM), vol. 43, no. 3,pp. 431–473, 1996.
[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Publickey encryption with keyword search," in Advances in CryptologyEurocrypt2004. Springer, 2004, pp. 506–522.
[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. SkeithIII,"Public key encryption that allows pir queries," in Advances inCryptology-CRYPTO 2007. Springer, 2007, pp. 50–67.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques forsearches on encrypted data," in Security and Privacy, 2000. S&P2000.Proceedings.2000 IEEE Symposium on.IEEE, 2000, pp. 44–55.
[8] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive,vol. 2003, p. 216, 2003.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keywordsearches on remote encrypted data," in Proceedings of the Third internationalconference on Applied Cryptography and Network Security.Springer-Verlag, 2005, pp. 442–455.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchablesymmetric encryption: improved definitions and efficient constructions,"in Proceedings of the 13th ACM conference on Computerand communications security. ACM, 2006, pp. 79–88.

BIOGRAPHY

SARUN J. VARGHESE is a **M.E** student in the department of CSE, VMKV Engineering College, Salem, India. He received his **B.E** degree in Computer Science and Engineering from Vinayaka Missions University, Salem, India. His research interests are Cloud Computing and Networking.

Dr. M. NITHYA, M.E., Ph.D.is a Professor &Head of the Department of CSE in VMKV Engineering College, Salem, India. Her research interests are Data mining, frequent item set mining and networking. She has published several papers in different national and international conferences and journals.She has guided several students at UG and PG level.