International Journal of Knowledge Based Computer Systems
Volume 5 Issue 2 December 2017
ISSN.: 2321-5623

Secure Ranked Keyword Search Over Encrypted Cloud Data

K. Mariyammal[1], A. Padmapriya[2]
[1]M.Phil. Scholar, Department of computer Application, Alagappa University, Karaikudi, Tamil Nadu, India.
[2]Associate Professor, Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India.

Abstract: Distributed computing is a system of remote servers facilitated on the Internet and used to store, oversee and prepare information set up of neighborhood servers. In the existing system three techniques used to extract information from the cloud. In this paper, a vector space show is utilized and each record is spoken to by a vector, which implies each report can be viewed as a point in a high dimensional space. The interdependence between varieties of documents are grouped into several categories. The pursuit time can be to a great extent decreased by choosing the coveted class and deserting the insignificant classifications. Cloud server will first search the categories and get the minimum desired sub-category. At that point the cloud server will select the desired k documents from the least possible desired sub-category. The value of k is to be decided by the user earlier and sent to the cloud server. To verify the search result, user has to verify the virtual root, instead of verifying every document. Furthermore, the proposed technique has favorable position over the traditional method in the rank privacy and relevance of retrieved documents.

Keywords: Cloud computing, Cloud services, Keyword search, Raking.

I. Introduction

Cloud computing is a type of internet based registering that gives normal PC preparing assets and information to PCs and Most distributed computing administrations drop into three general classifications: framework as an administration (IaaS), stage as an administration (PaaS) and programming as an administration (Saas). Infrastructure-as-a-service (IaaS**).**

Iaas is a is a type of distributed computing that gives virtualized processing assets over the web.

Platform-as-a-service (PaaS) refers to distributed computing services that supply an on-demand environment for developing, testing, delivering and managing software applications. PaaS is intended to make it less demanding for engineers to rapidly make web or versatile applications, without disturbing about setting up or managing the underlying infrastructure of servers, storage, network and databases needed for development.

Cloud computing allows multiple cloud users called residents to share a common physical computing infrastructure. Using quick implementation of the thought of Programming as a Service (SaaS) and Service Oriented Architecture (SOA), the Internet has evolved into an important service delivery infrastructure instead of only providing host connectivity

1. Software as a Service (SaaS)

Programming as-a-benefit (SaaS) is a method for delivering software applications over the Internet, on demand and typically on a contribution basis. With SaaS cloud suppliers have and deal with the product application and hidden foundation and handle any support, comparable programming overhauls and security fixing. Clients interface with the application over the Internet, typically with a web program on their telephone, tablet or PC.

Fig. 1: Cloud Computing

2. Cloud Security

Cloud computing security or, more simply, cloud security refers to an expansive arrangement of strategies, advancements, and controls sent to ensure information, applications, and the related foundation of cloud computing. It is a sub-space of computer security, network security, and, more broadly, information security.

3. Productivity

On location datacenters commonly require a great deal of "racking and stacking"— equipment set up, programming fixing and other tedious IT administration errands. Distributed computing evacuates the requirement for a large number of these assignments, so IT groups can invest energy in achieving more important business goals.

4. Privacy

Providers confirm that every single basic data (Mastercard numbers, for instance) are masked or, on the other hand encoded and that lone approved clients approach information in its entirety.

5. Data Security

Various security risks are related with cloud information administrations: not just customary security dangers, for example, arrange listening stealthily, illicit assault, and refusal of administration assaults**.**

II. Background Study

In [1], C. Wang, N. Cao, K. Ren, and W. J. Lou, described about that one-to-many request safeguarding mapping technique. It is used in Positioned look incredibly improves framework ease of use by supporting item pertinence positioning as opposed to sending undifferentiated outcomes, and further ensures the file retrieval accuracy. Specifically, this technique explore statistical measure the approach i.e. relevance score, from data recovery to fabricate a safe accessible file, and build up a one-to-many request safeguarding mapping system to legitimately ensure those touchy score information. The difficulties are Outsource index file cannot be maintain properly and most importantly there is a possibility of data leakage.

In [2], Pang, J. Shen, and R. Krishnan, described about that similitude based content recovery conspire. It is used when Users of online services are increasingly vary that their activities could release confidential information on their business or personal activities. It would be alluring for an online archive administration to perform content recovery for clients, while securing the protection of their exercises. In this article,they used a security saving, closeness based content recovery conspire. This scheme did not support many practical engines which are built using the vector space model. The retrieval rate is low in this scheme.

In [3], H. Pang and K. Mouratidis, described about threshold-based algorithms. The documents are searched by similarity in properties and based on these properties ranking is done. This is not reasonable for large dataset and the text query mechanism is not producing proper results.

In [4], N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou H. D. X. D, described about that multi-catchphrase positioned look over scrambled cloud information (MRSE) while protecting strict framework astute security in the distributed computing worldview. The creators proposed an essential thought for the MRSE utilizing secure inward item calculation, which is adjusted from a safe k-closest neighbor (kNN) procedure. This strategy will give two essentially enhanced MRSE plans in a step-by-step method. This will address the inflexible protection prerequisites in two threat models with increased attack capability. It encounters the dataset overhead problem.

In [5], D. X. D. Song, D. Wagner, and A. Perrig, described about the Single Keyword Searchable Encryption. It is used to encrypt each word in the archive independently. This method formally characterized a secure index structure and formulate a security model for index known as semantic security against adaptive chosen keyword attack (ind-cka).

In [6], D. Boneh and B. Waters, described about that Multiple Keywords Searchable Encryption. It is used to secure hunt plot in view of vector space model because of the absence of the security analysis for frequency information and practical search performance, it proposed a secure hunt plot in view of vector space model.

In [7], C. Martel, G. Nuckolls, P. Devanbu, M. Gertz, A. Kwong, and S.Stubblebine, described about Verifiable Search Based on Authenticated Index. The idea of data verification has been well studied in the area of databases. In a plaintext database scenario, a assortment of strategies have been have been produced. Most of these works are light the original work and refinements by Naor and Nissm for certificate revocation. Merkle hash tree and cryptographic mark methods are utilized to develop verified tree structure whereupon end clients can check the rightness and completeness of the query results.

The Limitations of the existing work is a high searching cost due to the scanning of the whole data collection word by word. Due to the time lack of ranking The ordering will give a quick access to the Cloud Service records where as positioning will orchestrate the rundown as indicated by the need of these methods grows exponentially accompanying with the exponentially increasing size of the document collections. They only focus on the verification-specific issues ignoring the search privacy preserving capabilities.

III. Proposed System

1. Ranking Keyword Search

The proposed work is about to optimize the topic based Cloud Service crawling process with the concept of exclusion of replicated documents. For this a new architecture is proposed, this architecture is using the rank based keyword search approach.

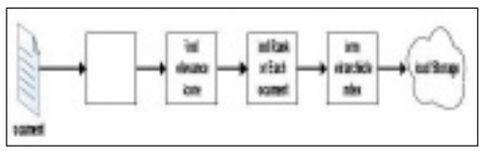
In this work the ranking is done with respect to the main criteria called keyword search Analysis. The user will interact to the Cloud benefit with this topic based question to recover the Cloud Service documents. In light of the input the data is retrieved from the Cloud Service. For the documents collection clouds generally use concepts like indexing and the ranking.

The ordering will give a quick access to the Cloud Service records whereas positioning will orchestrate the rundown as indicated by the need. Presently as a Cloud benifit documents are fetched, the proposed approach will retrieve the keywords from the archive in light of relevancy which is calculated by performing the match of service keywords with user query.

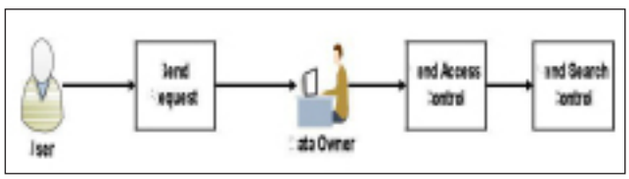
Presently as a new document is retrieved it will generate the hash tree and perform a hash tree based comparison to dissect the relevancy ratio. In light of this factor the initial ranking is assigned to the cloud benefit.

Step 1: The data owner selects the document to be uploaded in the cloud. Then using key generation algorithm, it generate key for each document. After generating key for each document, it encrypts the document with key and store in cloud.

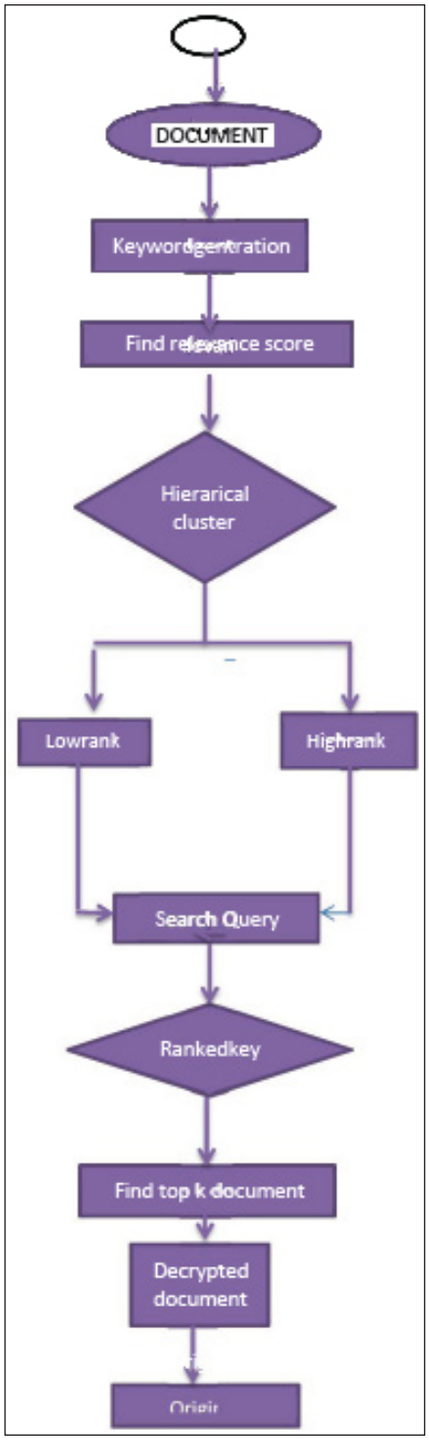
Step 2: Each document is assigned with keyword. To find the relevance score, first find the frequency of keyword in the document. After finding the relevance score, rank the document based on relevance score. If the document has high relevance score, it consider as high rank. Based on the relevance scores, form hierarchical index. It contains keyword and key with the ranking. Store the hierarchical index in the cloud.



Step 3: The users search the document in the server. Based on the search result, it send search request to the data owner. The data send the access control and search control to the user. The access control means, it give access to the cloud storage. The search control means to give access to index for decryption.



Step 4: The user access the cloud using access control key. In the trapdoor function, It generates encrypted query with users input keywords and secret key. The cloud returns the top-k document to the user. The user decrypts the document with decryption key. Original document is retrieved using decryptkey**.**



IV. Conclusion

The proposed multi keyword ranked search-Hierarchical clustering index (MRSE-HCL) will adapt to the requirements of data explosion, online information retrieval and semantic search. The theoretical overview of the proposed methodology is presented in the paper. It will address the issues in the existing methods can improve the search efficiency and rank security.

Reference

[1] C. Wang, N. Cao, K. Ren, and W. J. Lou, "Enabling secure and efficient ranked," vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "An efficient privacy ranked keyword search over encrypted cloud data," In *Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst.,* G Genova, Italy, pp. 253-262, 2010.

[3] H. Pang, J. Shen, and R. Krishnan, "Privacy-preserving similarity based text retrieval," *ACM Trans. Internet Technol.,* vol. 10, no. 1, pp. 39, Feb. 2010.

[4] N. Cao, C. Wang, M. Li, K. Ren, and J. Lou, "Privacy-preserving multi keyword ranked search over encrypted cloud data," In *Proc. IEEE INFOCOM,* Shanghai, China, pp. 829-837, 2011.

[5] H. Pang, and K. Mouratidis, "Authenticating the query results of text search engines," In *Proc. VLDB Endow.,* vol. 1, no. 1, pp. 126–137, Aug. 2008.

[6] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," In *Proc. IEEE INFOCOM,* San Diego, CA, 2010,

[7] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, "Highly scalable searchable symmetric encryptionth support for Boolean que-ries," *Advances in Cryptology – CRYPTO* pp. 353-373, 2013.