International Journal On Engineering Technology and Sciences – IJETS™
ISSN(P): 2349-3968, ISSN (O): 2349-3976
Volume III, Issue V, May- 2016

An Efficient Keyword Searching Based On HierarchicalIndex By Using Cluster Management

[1] DITTU T.CHRISTYANA, [2] K.LOHESWARAN

[1]PG Scholar, [2]Assistant Professor
[1, 2] Department of Computer Science and Engineering
[1, 2]Sasurie College of Engineering
Vijayamangalam, India
[1]dittuchristyana@gmail.com
[2] loheswaran.k@gmail.com

Abstract – Abstract In cloud computing the data owners are highly motivated in migrating their local sites to the public commercial cloud. This particular process is commonly called as outsourcing. It is very important to notice that while migrating the data management systems to the commercial sites certain private sites should be encrypted. Thus implementing the confidentiality and security. The most important aspect of the cloud data management is to meet effective data retrieval. Due to its flexibility and scalability it allows large number of users and documents in cloud. Through high efficiency retrieval is made more rapid.

Index Terms- Semantic relationship, Plain document, clustering and relevance score.

1. INTRODUCTION

Cloud computing become one of the monster technique. They are producing enormous amount of data like terabyte per day. The important aspect to be considered is cloud security various limitations are observed in the available solutions. When considering about security measure the high massive complex operations are involved. They include integrity of verifiable search result and hash results of the document. In recent years various mechanismshave been upgraded.

2. EXISTING SOLUTIONS

Now a day they adopt the searchable encryption method for searching the document the particular result is obtained or otherwise the relevant document based on the keyword is obtained. When considering about public key encryption authorization of the person is not considered. Anyone can unlock the relevant details based on their search on keyword. They include different types based on the key word searchable encryption. They are

- SINGLE KEYWORD SEARCHABLE ENCRYPTION.
- MULTI KEYWORD SEARCHABLE ENCRYPTION.
- VERFIABLE SEARCH BASED ON AUTHENTICATED INDEX.

2.1 SINGLE KEYWORD SEARCHABLE ENCRYPTION

The name itself implies that keyword encryption. They are encrypting each and every word in the document independently. This leads to high cost and increases complexity in managing the data structure. They areformulating model of semantic security against chosen adaptive attack.

2.2 MULTI KEYWORD SEARCHABLE ENCRYPTION.

The searchable encryption includes multiple numbers of words.Consider an example of

mobile service independently. Thesearch results for mobile and service is obtained the relevance between plain text are ignored. But the semantic relationship is also considered. They are the building stage process. They need excess amount for searching multiple numbers of words. Here the efficiency is not achieved. Efficiency is achieved only when the architecture needs limited amount of time to search relevant document.

2.3 VERFIABLE SEARCH BASED ON AUTHENTICATED INDEX

The data verification is considered as important in this sector in the databases. Here the refinements are considered to avoid the unwanted text. They include revocation hash tree techniques and crypto graphic signature and variousmanagement are considered. Hash chain is particularly used here. The limitations applied over here are that they are not directly used in the architecture. The various cryptographic signatures are applied over here to implement the authenticated process. Thus they are referred as verifiable search index. The limitations existed is that they are not directly involved in the architecture and they are only with in the long process.

3. LIMITATIONS OF EXISTING SOLUTIONS

Even though the hash treeand cryptographic signature are used. But they are not implemented directly in the architecture. Due to the implementation bilinear map they increases communication cost. This disrupts the efficiency of the key word searching in the database management system. More over the plaintext is ignored by considering semantic relationship. The dynamic classification of the document is absent. They increases the communication cost when implementing authenticated structure by implementing bilinear map.the search time is also increased than expected. The end user always suggest for the quick and efficient architecture. The existed probably lacks this character.

4. PROPOSED SYSTEM

Proposed system introduces Multi keyword rank search method. They are adopted over the encrypted data in the hierarchical clustering index. The search time is also upgraded by using linear growth. This speed up the relationship is provided between documents. In case search strategy is qualified by the rank method. Confidentiality is strongly implemented by using rank privacy.

Cloud is separated in to different sub cluster. This secures relevance score between the query and the search result. The clusters grouped together to form a child cluster. When the key word is searched on cluster the process are continued until the relevant details are obtained. The search process is also carried at the nearest child cluster is searched. They are continued until the search result is obtained. Due to the absent of bilinear map communication cost is reduced. There is no more exponential search time. Clustering method is able to solve the problem avoiding the relevance between plain documents. Various tree structures are implemented over here to improve efficiency there by providing the correctness and completeness of the architecture. Correctnessand completeness and limit the search time exponential rate to linear rate indexing of the documents is necessary for the encryption. When the request is send by the user at first the indexing documents are searched. When the relevance documents are obtained. Then the requested result can be sent to the user. Here the authentication verified while sending the encrypted document .Due to the method of hierarchical search index the documents are together as cluster as they are classified as parent cluster and various children clusters are

referred as sub cluster. Thus they form a tree structure. The algorithm mostly used to implement the process. The particular algorithm used here is back tracking algorithm. Day to day data volume is increase enormously .Various methods are applied to improved efficiency.one of the method applied over here is virtual root. Virtual root is applied to search the relevant document quickly.

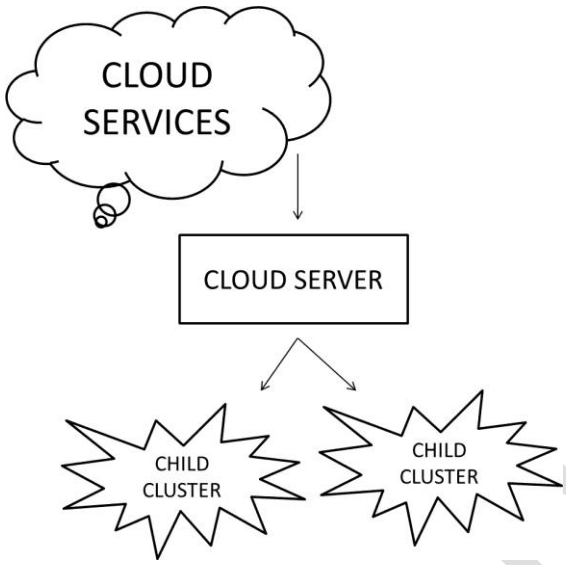

Fig1. Architecture of Sub cluster

The data owner is responsible for various processes which include collecting document, building the document index and outsourcing them in an encrypted format. The data owner is located in the cloud server.In the following diagram the concept is explained about the data owner and the data user. The searching process is carried on the hierarchical clustering index. Query request have been sent to the clod server for requesting the document. The various control implied here is access control and search control.As already explained about the giant growth of data sets the server should provide huge space. when the user sends data request .freshness should be implemented by retrieving the new details of latest information about the particular request. There should not be any constraint about the space. Cloud server should have enough capability for maintaining more number of data sets.at each time the data sets should be updated.

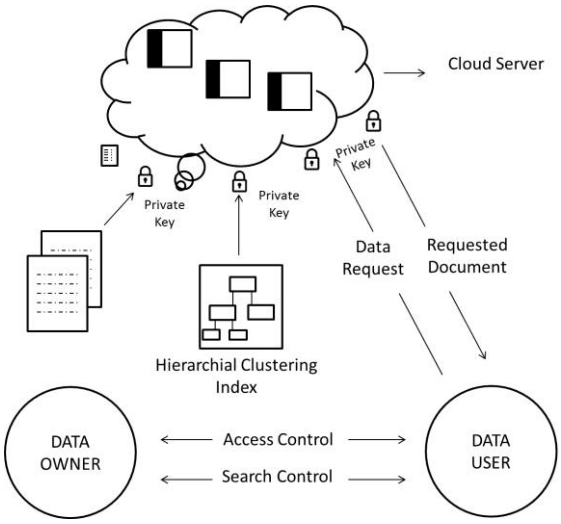

Fig 2.Architecture of MRSE-HCI

In MRSE-HCI Architecture vector is used to represent the keyword. The vector indicates the searches related to the document.There obtaining the relevance score between the query and document. In fig .2.data user sends request to the server. The data which is located in the cloud server are encrypted. When the request reaches the server they checks for the user's authentication. Private Key is maintained for the encrypted document. And they checks for the verification of the user authentication for sending the requested document. When the author's authentication is verified .then the requested is send to the user.Thus they are considered as trusted architecture.HCI is shown where the documents are maintained in the tree structure.

Here the vector is introduced between the query and the document. By considering the search phase segment the calculations are obtained by calculation the query vector and document vector. There should not be any compromise when regarding to secure like leaking of information .By considering this various cryptographic signatures are implemented.

The clustering method is very important to maintain the relationship between plain

document and encrypted document. Here the relevant documents are grouped together to make this made the search phase very of easy simple. Data owner and data user are the most important participants of this structure. The various controls implemented are access control and search control. They further maintain the efficiency.

5. TERMS IN ARCHITECTURE

In this MRSE – HCI the retrieval process is based on two factors they are relevance between query and the documents. Different search results are integrated includes three aspects they are Correctness, Completeness no qualified documents are omitted from the search result. Correctness implies that data should be uploaded by the data owner and they should be remained unmodified. Data privacy should be maintained properly in order implement confidentiality. These are the factors which are in accordance with quality measure data management. Symmetric is also a conventional way to achieve the privacy of the system. Cloud server can be commonly consider as public where they are partially considered as trusted. Because they store huge amount of datasets of different users. Data privacy should be well implemented by improving the confidentiality. The data owner plays vital role by encrypting the document. They are using various fragments for encrypting the document. When the data user request for the particular document. They send in encrypted form. Then the particular document is decryptedand they are used by the user.In the cloud storage various servers are located if server leads to failure the other functioning server should take over its function. Thus they should provide the property of fault tolerance. The scalability should also encourage for maintaining large number of user.Correctness should be maintained while retrieving the document and when new data set is added to the server. There should not be any alterations in the in the further document while adding the new data set. There should not be any constraints for maintaining the space between the documents. Since they need logarithmic growth to meet the need big data source.Hash values and root nodes are maintained. These are particular terms to be noted to improve efficiency.

6. CONCLUSION

In this we know about the particular and how they are involved in the efficient transformation of desired document .the important key note mentioned here is search efficiency and maintaining the legal document. And sending the legal document to the authorized user .the techniques have been improved from the existed for the data user.

The rank privacy confidentiality added further advantage to the proposed model

ACKNOWLEDGMENT

We wish to thank our Professor Dr.P.Sampath Sasurie College of Engineering for his support and encouragement.

REFERENCES

1. M.Bellare , A.Boldyreva and A.O.Neil “Deterministic and Efficient searchable Encryption” 2007.
2. C.Chen, X.J Zhu, P.S. Shen and J.K Hu” A Hierarchical Clustering Method For Big Data Oriented Cipher text Search “ Canada 2014.
3. Y.C .Chang ,and M.Mitzen Matcher “Privacy Preserving keyword for searches on Encrypted Data.
4. S.GrzonKwoski,P.M.Corcan “Security Analysis of Authentication Protocols “Germany 2011.
5. I.H. Witten , A.Mofat and T.C Bell” Managing Giga bytes Compressing and Indexing Documents” Kauffman1999.

6. S.C.Yu,C.Wang,K.Ren and W.J. Lou"Acheiving Secure Scalable and fine grained data access San Diego 2010.