ISSN: 2321-7782 (Online)
Impact Factor: 6.047

Volume 4, Issue 6, June 2016

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Efficient Synonyms Based MultiKeyWord Ranked Search with Cloud Security using Third Party Auditing

Khan Yasmeen[1]
Shree Ramchandra College of Engineering
SRCOE, Pune – India

Thombare B.H.[2]
Shree Ramchandra College of Engineering
SRCOE, Pune – India

Abstract: Cloud computing is based on internet that provides various services to users. Users store their data onto the cloud to remove burden of local data storage and maintenance. For security purpose data is stored in encrypted form. Thus Traditional plaintext search method proves to be useless. Therefore Challenge remains how to provide a practical efficient searching method that supports both Multikeyword ranked search and synonyms search of predefined keyword. Also data owner have no longer physical possession of data so the integrity and security of data become the major concern in the cloud computing. Data stored on the cloud server may be get corrupted and sometimes even the cloud service provider for his own benefit like for more space on data centre can discard the user data which is not used for a longer time. In order to maintain the integrity of data, data owner takes help of a Third Party Auditor (TPA). The TPA checks the integrity of data on user demand and the generate audit reports that help the owner to evaluate the risk of their services. Paper proposes practical efficient system that supports multi-keyword ranked search and synonyms based search also a small module implements auditing with help of TPA.

Keywords: Lucence index, Auditing, TPA, Multikeyword rank search, Hash algorithm.

I. INTRODUCTION

Cloud computing provides efficient resource management, economical cost, and fast deployment. Cloud computing has huge benefit due to this many companies have their own cloud center e.g. EC2 of Amazon, the Microsoft Azure, and IBM's Blue Cloud. Even cloud computing has a many benefits users are not willing to outsource their sensitive data like health records, financial data onto cloud server. Because data owners will lose direct control over these data. Encryption is an alternative way to solve problem and provide security. However on encrypted data traditional plaintext based search schemes not work well. Alternate solution is download all encrypted files from cloud and decrypts them locally to find the desired relevant document. However, this is impractical and time consuming method of searching which increase burden and computation cost of end users. So there is need of practical efficient search method that retrieve relevant encrypted document from cloud.

Cloud owner who store and use cloud data. Cloud service provider who manages and maintain storage information onto cloud. The third party auditor who audits the cloud file based on user requests. Auditing is necessary to check the data correctness in cloud. Auditing is performed by without affecting the client original data. TPA verifies user data integrity in cloud using hash algorithm.

II. PROPOSED SYSTEM

Existing system uses Searchable index tree in order to speedup search process. Index tree is balance binary created in bottom up manner. Documents form leaf node in tree. Data owner upload encrypted document along with index tree on to cloud server. When user send request to retrieve documents from cloud which contains keyword. Cloud service provider first search index tree travel till leaf node to find relevant document id. Once cloud service provider find document id then it retrieve and return encrypted document to user as search result. Existing system has various drawbacks it takes more time for search. As document increase tree size also grows which is difficult to handle. As own contribution Lucene indexing and searching method is implemented.

A. Lucene Indexing and Searching

It is possible to index and make searchable any data that can be converted to a text format. one can use Lucene to index and search data stored in files, web pages, on remote web servers, documents stored in local file systems, simple text files, Microsoft Word documents, HTML or PDF files, or any other format from which one can extract textual information. Lucene gives users full-text search capabilities. Indexing is done by analyzers. The unusable texts such as the stop words, word suffixes or prefixes are discarded at the analyzer stage. At the end of an indexing stage an index is created. A Lucene index consists of Lucene document class instances which defines the index documents. Each document contains a pair consisting of a Field name and a Field value.

A Lucene index is an inverted index. An inverted index means that the content of the documents that are analyzed has their important terms indexed as a pair consisting of a field name and a field value. Documents are searched in the fields and in their values. A Lucene index consists of many segments. A segment is created every time when new documents are created and indexed. Hence each segment has many documents stored in it. The documents consist of indexed Fields. An indexed field is a pair of a field name and a field value pairs. Fields are used for calculating weights and ranking search results. Once the text has been extracted from a document content has to be indexed. A new document has to be created. Searching is the process takes the user queries and they are parsed using the searcher parser. The results of a search method consist of hits from the index.

Main search classes in Lucene are IndexSearcher, Query, Term, and Hits.An IndexSearcher searches a document from an index. An index is opened in read only mode and uses its methods to return the search results Query A query class help users to defining user queries. Various type of query supported by Lucence are BooleanQuery, FilteredQuery, MultiTermQuery, PhrasePrefixQuery, PhraseQuery, PrefixQuery, RangeQuery, SpanQuery and TermQuery. The QueryParser class can automatically understand which type the user query belongs to. Term class represents the text in a document while searching. The term class takes two parameters, (a field and a text) the field is one in which the text will be searched. The term class is used for constructing the user query. Hits After the construction of a query, the IndexSearcher class searches the documents from the index. The results of the IndexSearcher class are pointed by the Hits class.

Scoring Lucene scoring is fast and it hides almost all of the complexity from the user. Lucene scoring uses a combination of the Vector Space Model (VSM) of Information Retrieval and the Boolean model to determine how relevant a given Document is to a User's query. In VSM the more times a query term appears in a document relative to the number of times the term appears in all the documents in the collection, the more relevant that document is to the query. Document is a collection of Fields. Each Field has semantics about how it is created and stored tokenized, raw data, compressed.

B. TPA

The third party auditor is able to verify the integrity of shared data based on requests from users, without downloading the entire data. Often data owner provide a pre-computed MD5 checksum for the documents, so that a TPA can compare the checksum of the downloaded file to it to find integrity of data. When owner wishes to check the integrity of shared data, they first send an auditing request to the TPA. After receiving the auditing request, the TPA generates an auditing message to the cloud, and calculates checksum using MD5 algorithm and compare this value with original hash value if both hash values are same integrity is maintain. Finally, the TPA sends an auditing report to the owner based on the result of the verification.

III. EXPERIMENTAL RESULT (SCREEN SHOTS)

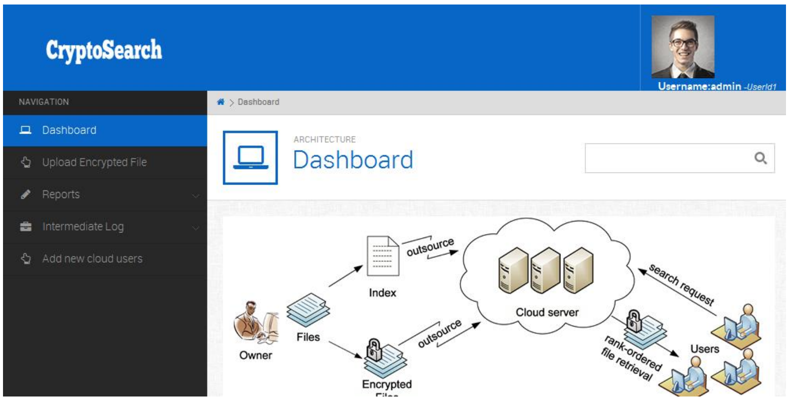

Fig. 1 Home page.

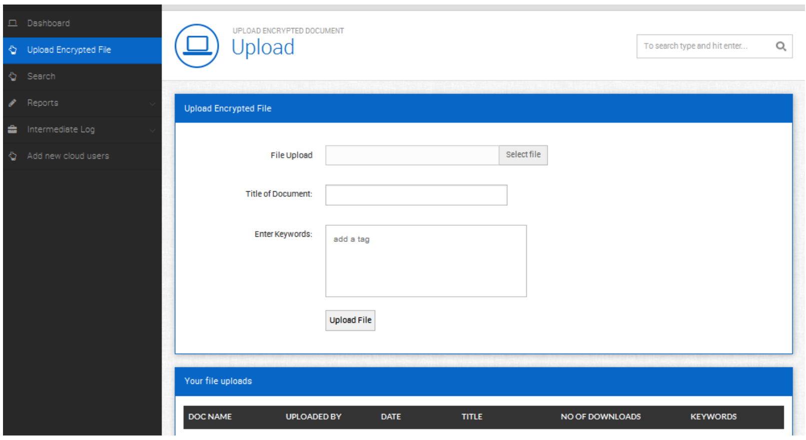

Fig. 2 Uploading of encrypted documents.

IV. PERFORMANCE ANALYSIS

Performance of existing and proposed system is compared based on various measures like precision, recall, Fmeasures. Precision is ratio of relevant document retrieve by total no of documents retrieve. Fig 3 shows Precision graph plotted based on no of documents retrieved verses precision.

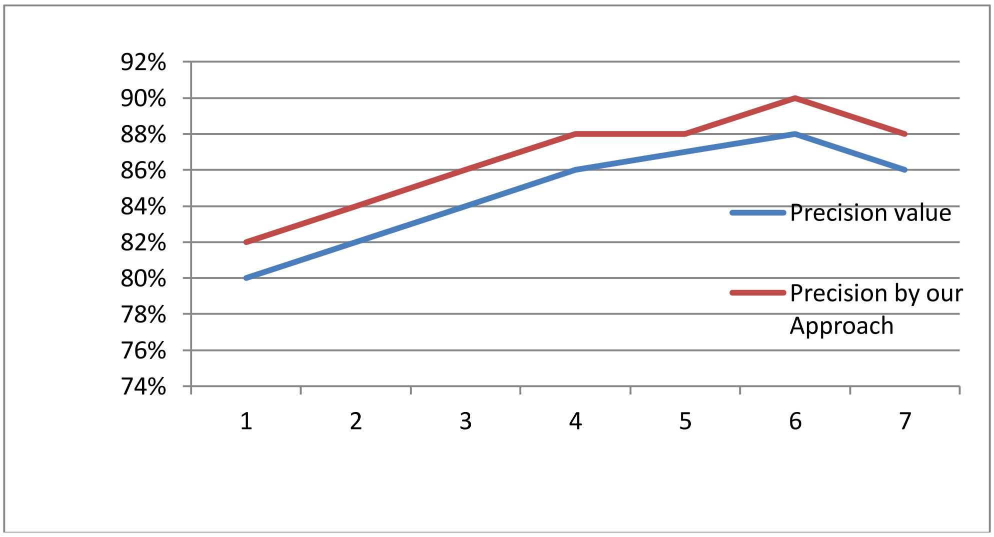

Fig. 3 Precision graph.

Recall is ratio of relevant documents retrieve by total relevant document. F1 macro graph is plotted based on no of keyword. Proposed system has high F1 macro value means it return relevant result.

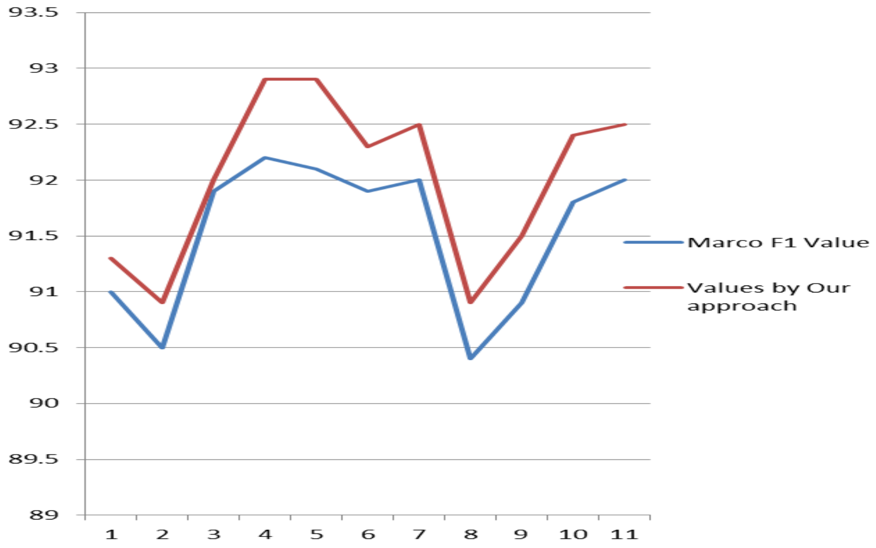

Fig. 4 F1 measure

A. Efficiency Calculation

1. Index size:

Lucene create index of small size i.e. only 1MB heap required. Compared to existing system where index size increases as no of document increases.

2. Index construction time:

Lucene required less time for index construction. Also it has facilities to create index in batches. It can merge multiple remote indexes at query time.

3. Searching Efficiency:

Lucence support powerful, Accurate and efficient search algorithms. It support many powerful query types like phrase queries, wildcard queries, proximity queries, range queries fielded searching (e.g., title, author, contents), date-range searching, sorting by any field. Multiple index searching with merged results. Also allows simultaneous update and searching.

4. Relevancy:

By default lucence return ranked result i.e best results returned first.

V. Conclusion

Thus our proposed system provides efficient searching over encrypted cloud data. Also it supports auditing which is necessary to resolve the cloud integrity issues. In our system TPA has learn no information about data content stored on the cloud server during the efficient auditing process, which not only eliminates the load of cloud user from the tedious and possibly expensive auditing task, but also reduces the users fear of their outsourced data leakage.

Acknowledgement

We would like to thank all the people involved in research in Achieving effective cloud search services. We would like to thank Prof. Thombare B.H our project guide and Prof. Deepti Varshney, Head of the Department for their valuable guidance and for providing all the necessary facilities.I also thankful to all the staff members of Computer Engineering Department for their valuable time, support, comments, suggestions and persuasion. Also I like to thank the institute for providing the required facilities, Internet access and important books.

References

1. Zhangjie Fu, Xingming Sun, Nigel Linge and Lu Zhou, "Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query", IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014.
2. Liqiang Nie, Meng Wang, Member, IEEE, Yue Gao,Zheng-Jun Zha, Member, IEEE, and Tat-Seng Chua, Senior Member, IEEE "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 2,FEB 2013
3. Sara Paiva, "A Fuzzy Algorithm for Optimizing Semantic Documental Searches", International Conference on Project Management / HCIST 2013.
4. Wenhai Sun, Ahucheng Yu, Wenjing Lou and Y. Thomas Hou, "Verifiable Attribute-based Keyword Search with Finegrained Owner-enforced Search Authorization in the Cloud" ,in IEEE Journal, 2013.
5. R. Sanchez, F. Almenares, P. Arias, D. Diaz-Sanchez, and A. Marin ",Enhancing privacy and dynamic federation in IdM for consumer cloud computing", IEEE Trans. Consumer Electron., vol. 58, no. 1, pp. 95-103,2012.
6. S. Grzonkowski, and P. M. Corcoran, "Sharing cloud services: user authentication for social enhancement of home networking", IEEE Trans. Consumer Electron., vol. 57, no. 3, pp. 1424- 1432, 2011.
7. 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
8. Qian Wang and Cong Wang and Kui Ren, Wenjing Lou, Jin Li "Enabling Public Auditability And Data Dynamics For Storage Security in Cloud Computing" in IEEE transactions on parallel and distributed systems, 2011, vol. 22, no. 5.
9. Cong Wang and Kui Ren and Wenjing Lou and Jin Li,"Toward Publicly Auditable Secure Cloud Data Storage Services" in IEEE, 2010.