Auditing with Achieving Effective Cloud Search over Encrypted Cloud Data

Mr. Dinesh R. Somvanshi[1], Dr. S.T. Singh[2]
Department of Computer Engineering
[1]PK Technical Campus, ChakanPune, India
[2]Campus Director PK Technical Campus, Chakan Pune, India

Abstract-*The Cloud server allows a user to store their data on a cloud without worrying about correctness & integrity of data. Cloud data storage has many advantages over local data storage. User can upload their data on the cloud and can access those data anytime anywhere without any additional burden. The User will use a cloud service as like local one. At very same time it very major concern that whatever data stored on cloud through internet is safe and not tampered annoying users or hackers. While accessing the data user must assure that his data in same as it is stored by cloud service provider. This is done with help of third party auditor which doesn't have any knowledge of the data. While searching the data on cloud authorized user can found the result efficiently by entering the keyword. A practically efficient and flexible searchable scheme which supports both multi-keyword ranked search and synonym based search, Vector Space Model is used to build document index, and each document is expressed as a vector where each dimension value is the Term Frequency weight of its corresponding keyword. The new vector has the same dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure can be used to compute similarity of one document to the search query.*

Keywords-Cloud Computing, Third Party Auditor, Cloud Service provider, keyword search and ranked search.

I. INTRODUCTION

The applications such as e-commerce, search, music on internet, gaming or website on which dating uses the techniques to mine large volume of data for better result as match. In existing system for collaborative filtering which is playing the important role for variety of internet service on e-commerce website or applications such as amazon, flipkart etc. The social networking having ratings, review and passing comments on particular posts or quote these activities are common in day today life. For example, one friend recommends the shoes to his friend. The symbolized or personalized recommendation having important role of applications for their customers because learning the alternatives is not an easy task.

Recommender system technique is one approach to information overload problem. The big achievement of the recommender system is to provide the user with recommendations that reflection taken place to the user's personal interest. There are two big approaches for decision making or for the personalized recommendation, these two approaches are collaborative filtering and content based filtering, the collaborative filtering uses the other user's opinions which is having similarity to the login user or active user. Collaborative based uses only the user's preferences of the active user as a filter or uses as a priority criteria for the results. The collaborative filtering and content based filtering approaches having some disadvantages or some weakness; with the help of those we try to invent new approach that is hybrid approach. Different techniques have been tried to combine two approaches into a new techniques, many taken from the field of artificial intelligence

II. RELATED WORK

CSP is a separate entity, which controls and manages cloud server. The User can store their large amount of data on the cloud. The risk of correctness of data, Cloud server has threats of data integrity in terms of insiders & outsiders. Wang[2] implemented third party auditor with checking the authorization of users in privacy preserving public auditing scheme. That third party auditing check the integrity of cloud data store by users. It is implemented by different algorithm for setup phase, audit phase etc. To achieve the effective cloud services like ranked search with multi-keyword on encrypted data various cloud data supports synonym queries by Xingming Sun, Fu Zhangjie, Lu Zhou[1], proposed the system to solve the issue of searching over encrypted data on cloud. This scheme specifies with two aspects which are ranked search for similarity and synonym based search.

Li et al. [7] firstly proposed a fuzzy keyword search scheme over encrypted cloud data, which combines edit distance with wildcard-based technique to construct fuzzy keyword sets, to address problems of minor typos and format inconsistence. Cloud computing is a technology that uses the internet and central remote servers to maintain data and

applications. . Using fuzzy search the exact keywords are displayed along with similarity keywords, which solve the problems faced by the cloud users. This paper concentrates on solving the problems of the user who search the data with the help of fuzzy keyword on cloud. Cabarcos[3] implemented middleware architecture as new scheme which allows transmitting without interruptions from one device to other in cloud services for customers. Searchable encryption in a cloud, in that the authorized user can search the encrypted data which stored over cloud data efficiency.

Seung Gwan Lee et al [5] proposed a personalized DTV Program Recommendation system under a cloud computing environment. The system can analyze and use the viewingpattern of consumers to personalize the program recommendations. A wide deployment of Internet Protocol Television (IPTV), Cable Television (CATV), Internet, User Created Contents (UCC), and Digital Television (DTV) enabled the rapid increase of channels and programs which can be selected by consumers. This was not expected when we consider the conventional television program technologies and policies. Due to these paradigm changes, hundreds of channels and programs are now available to consumers.

In the scheme of Lalit Kumar & Abhishek Mehta [6] proposed the scheme which contains model in which User, third party auditor (TPA) and scheme for retrieving the file, encryption, and decryption of file. The integrity of the data from cloud service provider (CSP) & control give to the third party auditor. Shrinivas[9], implement the scheme which user random masking and the homomorphic non linear authentication for storing cloud data. It ensures the user that the third party auditor will not knowing about content stored by user over cloud during the auditing process, which eliminate the burden on cloud for auditing and the data security for the user. In this scheme, the TPA can concurrently handle the multiple auditing requests and that perform in the batch manner for better efficiency.

Author in [9] proposed the system for the auditing the cloud data the extended feature of the scheme is not allow to unauthorized data access for the user to maintains the integrity of data stored by authorizes users. This system monitors the user specified parameters to check whether the user is an existing user or new user. In the check of a new user if it matches the parameters for checking it give the permission by Audit protocol for access or else blocking that specific user automatically by the system.

III. SYSTEM ARCHITECTURE

In our system, the Effective cloud storage auditing with achieving effective cloud search includes different Steps which are Setup Phase, Verifiable Data Updating and Challenge, Proof Generation and Verification, Index identification, Query confidentiality and Query Unlikeability. The auditing of the data store in the cloud storage server with data size is not fixed

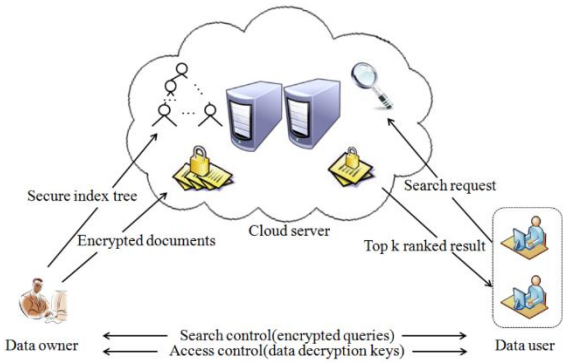

Fig 3.1 System Architecture

The data owner, the data user and the cloud server, as illustrated in Fig.1. The data owner, individual or enterprise, has a document collection DC which will be outsourced into the cloud. The data owner encrypts DC in the form of C before outsourcing to the cloud. And for the purpose of searching interested data, the data owner will also generate a searchable index I based on a set of distinct keywords W extracted from DC. Then, the encrypted file collection C and searchable index I will be outsourced to the cloud together by the data owner.

In the search stage, the system will generate an encrypted search trapdoor based on the keywords or the synonyms of the predefined keywords entered by the user. Given the trapdoor, the cloud server will search the index I and then return search results to the user. The search result is a set of encrypted documents containing the entered keywords, and they are well-ranked according to similarity measures. An additional feature provided by the system is that it can return a certain number of documents instead of all relevant documents. By sending a parameter k together with the search query, the user can get top-k most relevant documents. In the setup phase the keys for encryption and decryption while storing and retrieving the data in the auditing process. For searching we implemented different algorithms, functions, methods just like Tree based algorithm, Rank based function, similar keywords expansion etc. Using the above steps we can implement efficient ranked search scheme for the searching the encrypted data by the authorized cloud users. In the Verifiable Data Updating phase, the data is divided into the blocks. These blocks were updated in the cloud storage by different operations like partial modifying, complete change, delete a block, add new block etc. The verification of the update of the data is done. In the phase of Challenge, Proof

Generation and Verification token, generated by the Third Party Auditor for the authentication of the user in the system by the handshaking mechanism. The TPA generates the proof for auditing the data for the user and the cloud service provider. The verification is done by the TPA. In this system also introduced the effective methods for searching keywords on the encrypted cloud data which is outsourced.

A. Mathematical Model

Notations:

• W – dictionary of keywords, W={w| w1, w2,…,wn}
• $f_{d,j}$ – the document having term frequency of w_j;
• f_j–keyword present in the documents;
• M - number of documents present in the collection of document;
• N - keyword dictionary having number of keywords;
• $w_{d,j}$ - term frequency weight calculated from $f_{d,j}$,
• $w_{q,j}$, the inverse document frequency weight calculated from N and f_j;

The function of similarity is defined as:

$$SC\ (Q\ , D_d) = \frac{\sum_{j-1}^{N} w_{q,j} . w_{d,j}}{\sqrt{\sum_{j-1}^{N}(w_{q,j})^2\ .\ \sum_{j-1}^{N}(w_{q,j})^2}}$$

Where $w_{q,j} = 1 + \ln f d, j\ ,\ w_{d,j} = \ln (1 + \frac{N}{fj})$

IV. SYSTEM IMPLEMENTATION

4.1 System Implementation

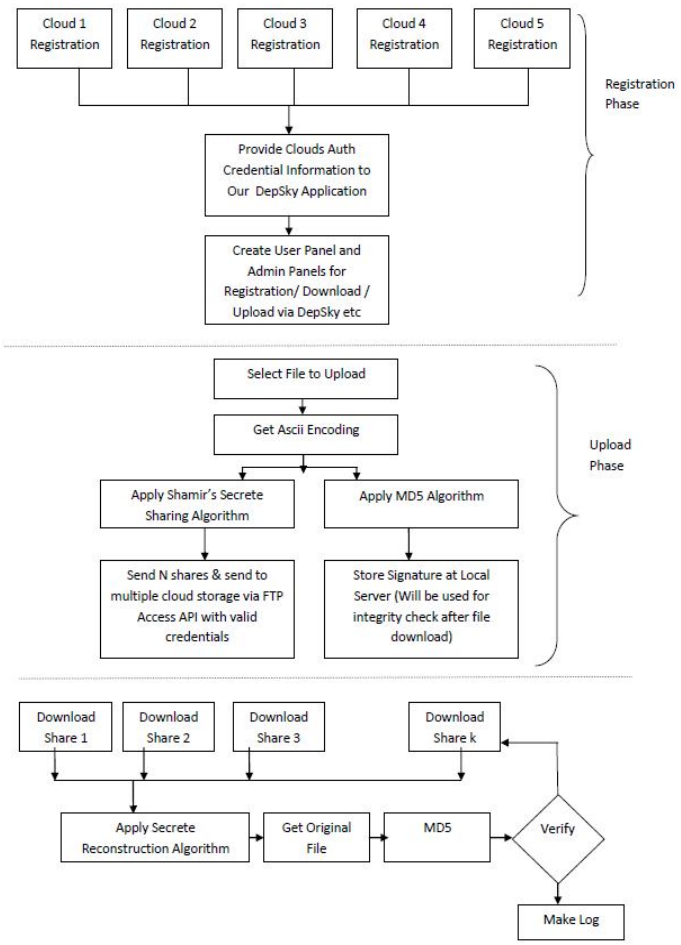

Fig 4.1 System Workflow

4.2 Screenshots

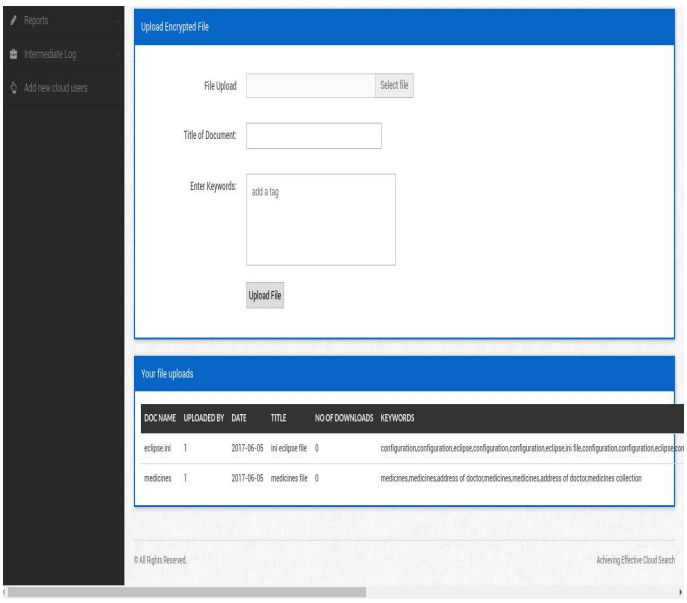

Fig 4.2 File upload

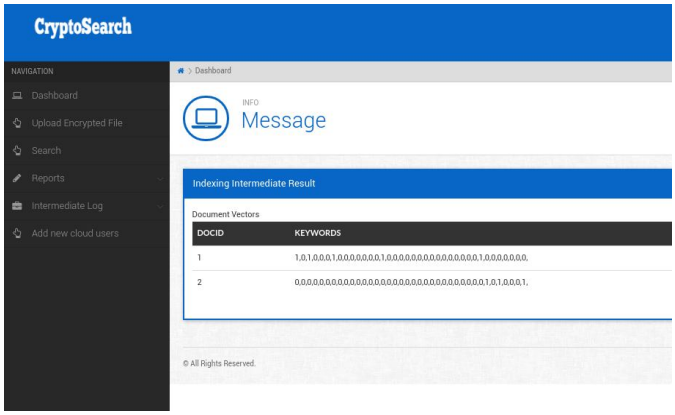

Fig 4.3 Indexing Intermediate Results

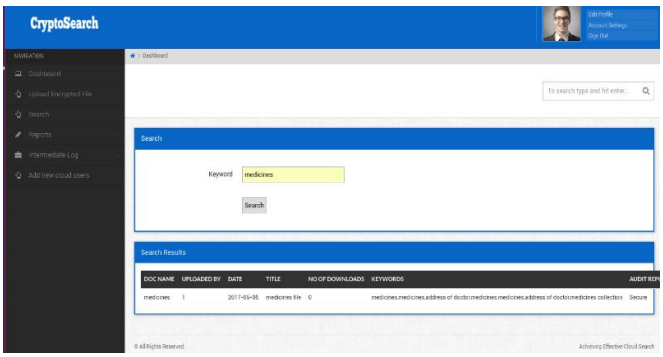

Fig 4.4 Keyword Search Result

The final screen shot shows the indexing of the search results with the indexing the document containing the keyword present in the encrypted cloud data. The table in fig 4.4 showing the actual outcome for the system by making the indexing the cloud stored encrypted data.

V. CONCLUSION

This system achieved the search result when the authorized cloud users input the synonyms of predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution without having exact knowledge of data. The combination of vector space model and cosine measure is well known in information retrieval to evaluate the similarity between search request and search documents. The performance of our system is analyzed in detail which includes search accuracy, search efficiency by experiment on the real dataset. The result shows that our system giving the solution which is very efficient and effective in supporting synonym based searching. The aim of our project is that cloud users can search the most relevant product or data. Our proposed system provides security to cloud storage by maintaining data integrity and privacy preserving with effective cloud data search on the encrypted data.

REFERENCES

[1] Fu Zhangjie, Xingming Sun & 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] C Wang, and Sherman S. M. Chow K Ren and W. Lou, 'Privacy-Preserving Public Auditing for Secure Cloud Storage', IEEE Transaction on Computers I, vol.62, no. 2, pp.362-375, February 2013.

[3] Cabarcos P.A., F.A. Mendoza, R.S. Guerrero, and D. Diaz-Sanchez, 'SuSSo: seamless and ubiquitous single sign-on for cloud service continuity across devices', IEEE Trans. Consumer Electron, vol. 58, no. 4, pp. 1425-1433, 2012.

[4] Wang Shao-hu, Chen Dan-we & Chang Su-qin, 'Public auditing for ensuring cloud data storage security with zero knowledge Privacy' College of Computer, Nanjing University of Posts and Telecommunications, China, 2009

[5] S. G. Lee, D. Lee, and S. Lee. 'Personalized DTV program recommendation system under a cloud computing environment.'2010.

[6] Abhishek Mehta and Lalit Kumar Awasthi, 'Cloud Data Security while using Third Party Auditor', International Journal of Scientific Engineering Research, Volume 3, Issue 6, ISSN 2229-8 June 2012.

[7] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou. 'Fuzzy keyword search over encrypted data in cloud computing.'2010.

[8] K Govinda, V. Gurunathprasad & H. Sathish Kumar, 'Third Party Auditing for Secure Data Storage in Cloud Through Digital Signature Using RSA', International Effective Cloud Auditing With Maintaining Integrity Journal of Advanced science and Technical Research, vol 4, no. 2, ISSN: 2249-9954,4August 2012.

[9] D. Shrinivas, 'Privacy-Preserving Public Auditing in Cloud Storage security', International Journal of computer science and Information Technologies, vol 2, pp. 2691-2693, ISSN: 0975-9646, 2011

[10] Kiran Kumar, K.Padmaja & P.Radha Krishna 'Automatic Protocol Blocker for Privacy-Preserving Public Auditing in Cloud Computing', IEEE 2012 Transaction on Cloud Computing, Volume: PP, Issue:99.

[11] Dr. P. K. Deshmukh, Prof. R. A. Deshmukh, 'Investigation of TPA (Third Party Auditor Role) for Cloud Data Security', International Journal of Scientific and Engineering Research, vol. 4, no. 2,ISSN 229-5518, Feb 2013..