Auditing Cloud Data Storage with Effective keyword Search Strategy

Mr. Dinesh R. Somvanshi[1]
[1] Department of Computer Engineering
[1] PK Technical Campus, Chakan, Pune, India

Abstract- *The cloud computing is mechanism in which users stores their data using high quality application and services provided by cloud service provider. Cloud computing allow a user to use different services through the Internet on their demand. Because of cloud computing user can store their data without worrying about local storage space. The user does not need the physical storage of the large outsource data. The user can store their data on to the cloud efficiently. At a very same time the user is worried about the data storage security, thus public audit ability is enabled for data storage security which is critically important. The effective third party auditor (TPA) is introducing with distinct characteristics of auditing the cloud storage without demanding the actual copy locally and no additional online burden to the cloud users.*

The data is stored in encrypted format, so typical search by keyword method is not efficient. Nowadays using rough keyword search method, hacker can search encrypted data over cloud. In our system we are using the multikeyword ranked search method to search get approximate results.

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

I. INTRODUCTION

The hardware and software are the resource for computing used in cloud computing to get the services over internet. That has benefits of frequent access of network without depending upon the resources location. Cloud computing allows sharing of resources to increase the economy similar to a utility over a network.

Without worrying about space and maintenance, the user of cloud can save their data over cloud. When a user stores their data on the cloud, it provides relief of burden for managing data storage, access data anywhere irrespective of locationand expenditures can be on hardware and software and personal maintenance. By configuring the resources of computing, it gives the different facilities & quality softwares over internet. Local storage has number of disadvantages as compare to the cloud data storage. The cloud service provider only provides data storage services to the user. Hence how the user will get a confirmation about stored data correctness and integrity. CSP uses third party auditor (TPA) concept where TPA will audit user data, and it will give a confirmation to the user that data stored in the cloud is secure.

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 interms of insiders & outsiders.Wang[1] 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[2], 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.

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.

In the SMDS: Secure Model for Cloud Storage by the KrunalSuthar, Permalink Kumar, Hitesh Gupta [5] in this system different operational algorithms are used for effective security option using Modern symmetric encryption schemes, with this user's sensitive data is confident with maintaining the integrity verification with the low cost of computation. The symmetric key sharing is handled with the public key cryptography.

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[7], 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 the actual 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. For new user prompts the parameters to match requirements specified during the user creation. 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. PROPOSED SYSTEM

In our proposed 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.

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}
- fd,j – the document having term frequency of wj;
- fj–keyword present in the documents;
- M - number of documents present in the collection of document;
- N - keyword dictionary having number of keywords;
- wd,j - term frequency weight calculated from fd,j,
- wq,j, the inverse document frequency weight calculated from N and fj;

The function of similarity is defined as:

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

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

IV. PERFORMANCE ANALYSIS AND RESULTS

The binary search and method of keywords extraction are used to evaluate the performance our system.

1. Extraction of Keywords:

In our system, term frequency, inverse documents frequency, similarity of test document and nearest neighbor (KNN), test document similarity are use to make effective the system. The accuracy is measure by the function, which is use as macrofuction.
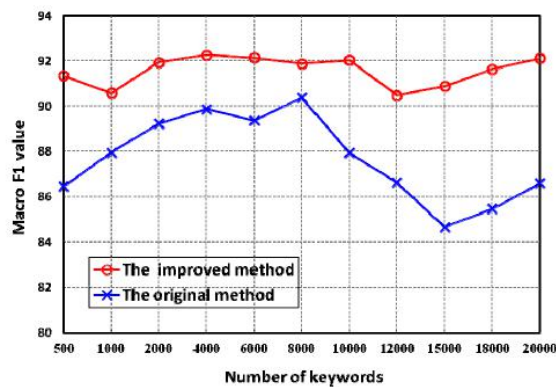
Figure 1.

It shows that, the previous method with nearest neighbor algorithm is showing the less accuracy of search as compare to the method to combine of nearest neighbor with macro fuction algorithm. This method extracted more keywords from text file.

2. Search Efficiency

Calculated with the help of calculating the scores of result of rank based and relevant document similarity.
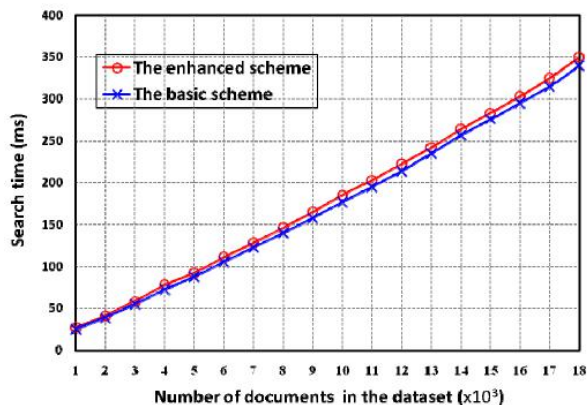

Figure 2.

This graph shows the time for search in proposed scheme and basic scheme.

V. CONCLUSION

The regularity and audit, which are need for enterprises, can fulfill by cloud computing mechanism to ensure the privacy. Cloud service provider need to address this issues regarding cloud data security and privacy the status is attain by cloud computing domain. Our proposed system provides security to cloud storage by maintaining data integrity and privacy preserving with effective cloud data search on the encrypted data. We are using Secrets sharing

algorithm to store secret data on multiple clouds. Along with this, we are using index searching algorithm for more accurate search. In this, the semantics-based search and similarity rank search over encrypted cloud data achieved results by cloud users which are authorized by cloud service provider with input like predefined keywords, random matching keywords which are not an exact match, because of the possibility of synonym substitution knowing about the 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] KrunalSuthar, Permalik Kumar and Hitesh Gupta, 'SMDS: secure Model for Cloud Data Storage', International Journal of Computer applications, vol56, October 2012.

[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] 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

[8] KGovinda, 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] 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.

[10]Balkrishnan. S, Saranya. G, Shobana. S and Karthikeyan S, 'Introducing EffectiveThird Party Auditing (TPA) for DatStorageSecurity in Cloud', International Journal of computer science and Technology, vol. 2, no. 2, ISSN 2229-4333 (Print) ISSN: 0976-8491(Online), June 2012

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