International Journal of Engineering Science Invention Research & Development; Vol. III, Issue IX, March 2017
www.ijesird.com, e-ISSN: 2349-6185

MANAGING THE CLOUD STORAGE USING DE-DUPLICATION AND SECURED FUZZY KEYWORD SEARCH FOR MULTIPLE DATA OWNERS

Abinaya.R[1]; Saradhambal.G.Ms[2]
UG Student[1] , ,Assistant Professor[2] Dept. of CSE, IFET College of Engineering, Villupuram
Email:abinayajayasri07@gmail.com[1];saradhadivya@gmail.com[2]

ABSTRACT- **Cloud compute, is trouble-free for data owners to outsource their data to public cloud servers and it allows data users to retrieve the data. Cloud servers to perform secure search without knowing actual data of both keywords and trapdoors. When the number of files continue to increase, the condition of every suitcases section node cannot be assured by the manager. High volumes of files will result in wasted hardware property, increased control difficulty of the data center, and a less efficient cloud storage space system. To reduce workloads due to replica files, the index name servers are used to supervise not simply file storage and also data de-duplication.**

Keywords: **Eavesdropping,Trapdoors,INS, De-Duplication**

1. INTRODUCTION

Cloud computing air force can be classified as any computing or storage space. As far as data storage space is worried, although numerous scheme have been presented to recover file chunk and data compression, the waste of possessions caused by modification or changes is often ignored.
it has develop into more and more popular for data owners to delegate their data to unrestricted cloud servers even as allowing data users to recover this data. For isolation concern, a secure search over encrypted cloud data has annoyed several research works under the single holder model. though, nearly all cloud servers in perform do not just give out one owner; as an alternative, they support multiple owner to share the benefits bring by cloud computing. In this paper, we proposition schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM).and dissipate of resources may occur as the system processes duplicate and unnecessary data, despite the flexibility and haste of the cloud storage space system.

2. EXISITING WORK

They propose the conception of searchable encryption, which is a cryptographic primal that enables users to perform a keyword-based search on an encrypted dataset; just as on a plaintext dataset extending these techniques for ranked multi-keyword search will incur heavy multiplication and storage costs. These research not only decrease the working out and storage space charge for protected keyword investigate over encrypted cloud data, but also enrich the group of hunt function, together with protected rank multi-keyword investigate, fuzzy keyword search, and similarity search.

3. PROPOSED SYSTEM

Approximate string matching is done by proposing fuzzy keyword search Data de-duplication is a specialized data compression procedure for eliminate duplicate copies of repeating data in storage**.**
To decrease workloads due to duplicate files, the index name servers are used to handle not only file storage space and also data de-duplication.

4. MODULES

- Multiple owners
- Outsourcing data
- De duplication
- Fuzzy search

4.1 MODULES DESCRIPTION

4.1.1 Multiple Owners

To allow the cloud server to perform safe search between multiple owners' data encrypted with unusual secret keys, we methodically build a novel secure search protocol. To rank the look for results and preserve the privacy of significance scores between keywords and files.

4.1.2 Outsourcing data

The data owner have a proper registration, and have to give the file to upload with the keywords (multi-keywords). The data owner has a collection of n files to outsource on top of the cloud server in encrypted form and expect the cloud server to offer keyword retrieval service to data owner himself or supplementary authorized users.

To achieve this, the data owner needs to build a searchable index from a collection of keywords extract absent of files, and then outsources mutually the encrypted index and encrypted files on top of the cloud server**.**

4.1.3 De-duplication

De-duplication is a method used for eliminate replacement copy of data all the way through a de-duplication scanning development, which improve the system performance and decreases the bandwidth occupied by data transmission.

The techniques divide a file into chunks and calculates a single 128-bit hash code of each lump by MD5, i.e., the only signature of the chunk.Current de-duplication-related techniques and do research have all aimed at deleting duplicate data at the server side, but none has been proposed to discuss data de-duplication and unneeded data elimination at the client side.

4.1.4 Fuzzy Search

Fuzzy keyword search greatly enhances system usability by frequent the similar files when users' thorough inputs exactly match the predefined keywords or the closest possible equal files based on keyword relationship semantics, when exact match fails.

The key idea behind our secure fuzzy keyword search is two-fold:

[1] Building up and about fuzzy keyword sets that include not only the exact keywords but also the ones reverse to some extent due to minor typos, design inconsistency, etc.

[2] Manipulative an professional and secure penetrating move toward for file retrieval based on the resulted fuzzy keyword sets.

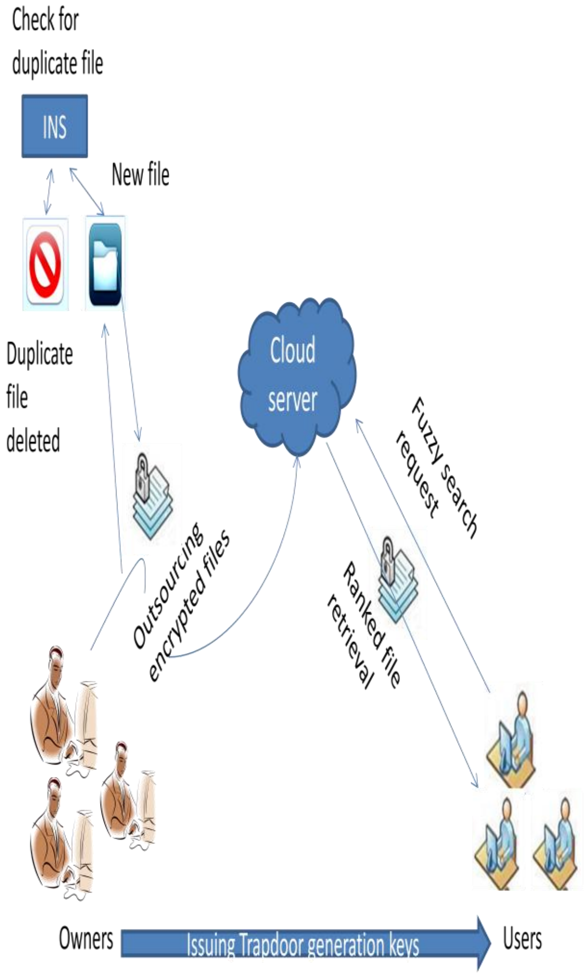

Figure 4.1 system architecture

4.2 ADVANTAGES OF THE PROPOSED SYSTEM

- It reduces the computation and storage cost.
- Secured fuzzy keyword search over encrypted cloud data.
- It provides excellent security and scalability
- Duplicate copies of data are eliminated through de-duplication scanning process.

5. CONCLUSION

This paper proposed the INS to process not only file density, lump identical, data de-duplication, real-time reaction control, IP in order, and busy level index monitoring, but also file storage, optimized node collection, and server load matching Three major contributions of this paper include the subsequent. Thus the authentication for data users and notice attackers who steal the secret key and carry out illegal searches, a novel dynamic secret key generation protocol and a new data user authentication protocol. To enable the cloud server to perform secure search among numerous owners data encrypted with unlike secret keys, a novel secure search protocol is constructed.

REFERENCES

[1] Wei Zhang, Yaping Lin, Sheng Xiao, Jie Wu and Siwang Zhou" **Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing"** in IEEE journal,Jan 2015.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacypreserving public auditing for secure cloud storage," Computers, IEEE Transactions on, vol. 62, no. 2, pp. 362–375, 2013.

[4] D.Song, D.Wagner, and A.Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE International Symposium on security and ,Privacy (S&P'00), Nagoya, Japan, Jan. 2000, pp. 44–55

[5] Y.-M. Huo, H.-Y. Wang, L.-A. Hu, and H.-G. Yang, "A cloud storage architecture model for data-intensive applications," in *Proc. Int. Conf. Comput. Manage.*, May 2011, pp. 1–4.

[6] L. B. Costa and M. Ripeanu, "Towards automating the configuration of a distributed storage system," in *Proc. 11th IEEE/ACM Int. Conf. Grid Comput.*, Oct. 2010, pp. 201–208

[7] H. Ohsaki, S. Watanabe, and M. Imase, "On dynamic resource management mechanism using control theoretic approach for wide-area grid computing," in *Proc. IEEE Conf. Control Appl.*, Aug. 2005, pp. 891–897.

[8] H. Dezhi and F. Fu, "Research on self-adaptive distributed storage system," in *Proc. 4th Int. Conf. Wireless Commun. Netw. Mobile Comput.*, Oct. 2008, pp. 1–4.

[9] J. Wang, P. Varman, and C.-S. Xie, "Avoiding performance fluctuation in cloud storage," in *Proc. Int. Conf. High Performance Comput.*, Dec. 2008, pp. 1–9.

[10] C.-Y. Chen, K.-D. Chang, and H.-C. Chao, "Transaction pattern based anomaly detection algorithm for IP multimedia subsystem, *IEEE Trans. Inform. Forensics Security*, vol. 6, no. 1, pp. 152–161, Mar. 2011.