

## MANAGING THE CLOUD STORAGE USING DEDUPLICATION AND SECURED FUZZY KEYWORD SEARCH FOR MULTIPLE DATA OWNERS

Dr.E.kodhai<sup>1</sup>, Divakar krishna.S.V<sup>2</sup>, Natarajan. R<sup>3</sup>, Vikneshr. K<sup>4</sup>, Allwin Andrews. C<sup>5</sup>, P. Yellamma<sup>6</sup>, Dr. Jafar A. Alzubi<sup>7</sup>

<sup>1,2,3,4,5</sup>Dept of Computer Science and Engg, Sri Manakula Vinayagar Engineering College, Puducherry.

<sup>6</sup>Bharathiar University R & D Center, Coimbatore, India.

<sup>7</sup>School of Engineering, Al-Balqa Applied University, Jordan.

**Abstract:** Cloud Computing has been envisioned as the next generation architecture of IT Enterprise. Cloud Computing moves the application software and databases to the large data centers, here the management of the data and services may not be fully trustworthy. Cloud offers many key features such to provide the user with the ease of use. Cloud offers two main purposes, for computing and storage. It optimizes the cloud storage by reducing the amount of space used to store the files. We are also implementing secure fuzzy keyword search that uses fuzzy logic to securely search files from the cloud storage based on approximate keyword values from user.

The proposed solution consists of a data de-duplication which is a technique that eliminates copies of the same file thereby improving the system performance and decreases the bandwidth used in data transmission. The fuzzy keyword search provides a user-friendly interface to search for files by using fuzzy logic which approximates user keywords to matching keywords to retrieve relevant files. Here, we are improving the efficiency of the cloud storage using de-duplication techniques and privacy preserving fuzzy keyword search technique.

### 1. Introduction

Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). Some vendors use terms such as IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) to describe their products, but we eschew these because accepted definitions for them still vary widely. The line between “low-level” infrastructure and a higher-level “platform” is not crisp.

The cloud also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically re-allocated per demand. This can work for allocating resources to users. Figure.1 provides an approach which should maximize the use of computing powers thus

reducing environmental damage as well since less power, air conditioning, rackspace, etc. is required for a variety of functions.

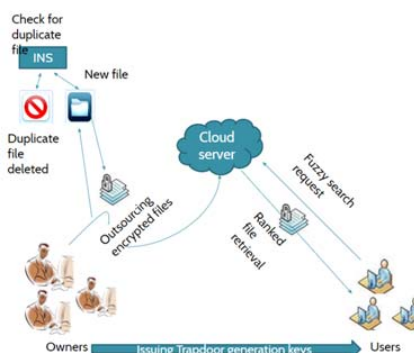


Figure.1 Working of Cloud

**De-duplication** is the process of eliminating the redundant files or information such that only one unique instance of the data is retained on the storage. It is an intelligent compression technique which improves the usability of the cloud storage. It reduces the space used by the same files having multiple copies and thereby improves performance of the cloud server.

### 2. Related Works

Himshai Kamboj and Bharati Sinha [<sup>1</sup>]. This paper presents a deduplication system named “DeDup App” that runs on service hardware. At front end, DeDup App has a user interface. At back end, HDFS is being used as storage system and quick indexing. Basically, there are two problems to be addressed. Firstly, identifying the duplication in data files and secondly managing the ownership challenge for that data file. For above problems, we are using MD512 algorithm for making fingerprint of each data file being uploaded for initial checking of duplicated files and 256-bits AES for encryption decryption.

Shunrong Jiangy, Tao Jiang and Liangmin Wangy [2]. In this paper, we propose a secure data deduplication scheme with efficient PoW process for dynamic ownership management. Specially, our scheme supports both cross-user file-level and inside-user blocklevel data deduplication. During the file-level deduplication, we construct a new PoW scheme to ensure the tag consistency and achieve the mutual ownership verification. Moreover, we design a lazy update strategy to achieve efficient ownership management.

P.Priya ponnusamy, Dr..R.Vidhyapriya, S.Uma Maheswari [3]. This paper proposes a searching strategy over the encrypted data, which encourages multi-keywords ranked searching and dynamic functions on the collection of files. We have summarized different kinds of searching methods for encrypted cloud data. A study on the data privacy and utilization issues are covered here for various searching techniques. Some of the important issues, efficiently result ranking, query privacy etc. Some of the searching techniques focuses on the data security and some on the data utilization.

Wang Jie, Yu Xiao, Zhao Ming, Wang Yong [3]. In this paper, we solve the problem of full-scale fuzzy keyword set construction according to the input keyword, and construct the feedback scheme to produce pointer vector including fuzzy keyword, edit distance and keywords dynamic score, which is feasible in hybrid cloud model. Thus, different vectors form the character vector database within its data structure. It can go to the trapdoor construction procedure after access to the database with its edit distances to construct fuzzy keyword set, which makes the fullest use of the retrieval history and statistical misspelled keywords, precisely and quickly realizing the aim of ranked fuzzy keyword search over cloud encrypted data.

### 3. Research Direction

We can improve the de-duplication efficiency by dividing each file into chunks of data and designate a unique fingerprint for each chunk

MD5 algorithm is used to generate unique fingerprints for which the index name server (INS) looks for matching fingerprints to determine whether a file already exists

### 4. Conclusion

At present, most de-duplication related techniques and research have all aimed at eliminating duplicate data at the server side. In such cases, the user needs to upload the file to the server where it is checked whether it already exists or has a unique signature. In our proposed

system, we are creating chunks of data and comparing their fingerprints with the index database to check if the data already exists. This process is done at the client side itself so it provides more efficiency in terms of bandwidth usage. This method divides a file into chunks of data and generate a unique fingerprint for each chunk. The current system does not support fuzzy keyword search. The fuzzy keyword search is used in our system which provides added usability to the user by allowing a user to search for data efficiently even with minor typos and mistakes in the keywords. It generates a trapdoor for the search terms to search the encrypted data directly

### References

- [1] Himshai Kamboj and Bharati Sinha, "DEDUP: Deduplication system for Encrypted Data in Cloud", International Conference on Computing, Communication and Automation ICCCA , pp. 28 – 35, 2017
- [2] Shunrong Jiangy, Tao Jiang and Liangmin Wangy, "Secure and Efficient Cloud Data Deduplication with Ownership Management", IEEE, pp.1-1, 2017
- [3] P.Priya ponnusamy, Dr..R.Vidhyapriya, S.Uma Maheswari, "A Survey on Multi-Keyword Ranked Search Manipulations over Encrypted Cloud Data", International Conference on Computer Communication and Informatics , pp. 63-64, 2017.
- [4] Wang Jie, Yu Xiao, Zhao Ming, Wang Yong, "A Novel Dynamic Ranked Fuzzy Keyword Search Over Cloud Encrypted Data", IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, pp.11969-11973, 2014



