Cloud Data De-Duplication in Multiuser Environment: DeposM2

Mr. Kaustubh Borate, Prof. Bharti Dhote

Department of Computer Engineering, SIT Lonawala, SPPU, Pune, India

ABSTRACT

Nowadays, cloud computing produce a huge amount of sensitive data, such as personal Information, financial data, and electronic health records, social media data. And that causes duplication of data and that suffers to storage and performance of cloud system. Data De-Duplication has been widely used to eliminate redundant storage overhead in cloud storage system to improve IT resources efficiency. However, traditional techniques face a great challenge in big data De-Duplication to strike a sensible tradeoff between the conflicting goals of scalable De-Duplication throughput and high duplicate elimination ratio. De-Duplication reduces the space and bandwidth requirements of data storage services, and is most effective when applied across multiple users, a common practice by cloud storage offerings. I study the privacy implications of cross-user De-Duplication. Thus, an interesting challenging problem is how to deduplicate multimedia data with a multi user environment and propose an efficient system to overcome these types of problems. In this paper, I introduce a new primitive called Depos-M2 which gives a partial positive answer for these challenging problem. I propose two phases: De-Duplication and proof of storage, where the first one allows De-Duplication of data and letter one allows proof of storage that means give permission to respective user i.e. owner of that file.

KEYWORDS*: Cloud Computing; De-Duplication; Depos-M2; proof of storage*

INTRODUCTION

Nowadays, cloud computing produce a huge amount of sensitive data, cloud computing is becoming more and more attractive to both industry and academic due to the advantages of low cost, high accessibility, and easy sharing. And that causes duplication of data and that suffers to storage and performance of cloud system. As one of the storage outsourcing forms, cloud storage gains wide attention in recent years.

Cloud storage refers to scalable and elastic storage capabilities that are delivered as. Many companies, such as Amazon, Google, and Microsoft, provide their own cloud storage services, where users can upload their files to the servers, access them from various devices, and share them with the others.

A service using Internet technologies with elastic provisioning and use-based pricing that does not penalize users for changing their storage consumption without notice [1], [2]. The term data De-Duplication refers to techniques that store only a single copy of redundant data, and provide links to that copy instead of storing other actual copies of this data.1 with the transition of services from tape to disk, data De-Duplication has become a key component in the backup process. By storing and transmitting only a single copy of duplicate data, De-Duplication offers savings of both disk space and network bandwidth. For vendors, it offers secondary cost savings in power and cooling achieved by reducing the number of disk spindles. According to recent statistics, De-Duplication is considered to be the most-impactful storage technology

A. Approaches to De-Duplication
Data De-Duplication strategies can be categorized according to the basic data units they handle. In this respect there are two main data De-Duplication, strategies [1]: (1) File-level De-Duplication, in which only a single copy of each file is stored. Two or more files are identified as identical if they have the same hash value. This is a very popular type of Service offered in multiple products ;(2) Block-level De-Duplication , which segments files into blocks and stores only a single copy of each block. The system could either use fixed-sized blocks or variable-sized chunks. The discussion in this paper may be applied to both strategies.

LITERATURE SURVEY
A. DeyPOS: DE duplicable dynamic proof of storage and file sharing for multiuser environment
In this paper, we have discussed existing techniques for De-Duplication of cloud data for multi user environment. This system gives idea of avoid duplication of data that is going to store at cloud. This system compare data of file which is store already at cloud storage with newly uploaded file data. This system compare data only to avoid duplication.

B. ASE-PoW: a Proof of Ownership mechanism for cloud De-Duplication in hierarchical environments: A Survey
This paper provide Proof-of-Ownership (PoW) can be an effective De-Duplication technique to reduce storage requirements, by providing cloud storage servers the capability to guarantee that clients only upload and download files that they are in possession of.

C. Secure data Deduplication scheme for clod storage.
This paper aims at differentiation of data according to their popularity, and to security purpose this system gives idea of secure a data of file with using encryption technique with using AES algorithm.

D. Side channels in cloud services, the case of De-Duplication in cloud storage
This paper provide ideas about cross-user De-Duplication And risks due to the high savings Offered by cross-user De-Duplication, cloud storage providers are unlikely to stop using this technology.

E. Dynamic Data De-Duplication in Cloud Storage
This paper address the most things like storage of dynamic data and handling of dynamic data.

PROPOSED METHODOLOGY
This system model considers two types of entities: the cloud server and users. For each file, original user is the user who uploaded the file to the cloud server, while subsequent user is the user who proved the ownership of the file but did not actually upload the file to the cloud server.

A. Architecture
There are five phases in a deduplicatable dynamic PoS system: Pre-processing, Uploading, De-Duplication, Updating, and Proof of Storage. In the first pre-processing phase, users intend to upload their local files. The cloud server decides whether these files should be uploaded. If the upload process is granted, go into the upload phase; otherwise, go into the De-Duplication phase. In the upload phase, the files to be uploaded do not exist in the cloud server. The original users encodes the local files and upload them to the cloud server. In the duplication phase, the files to be uploaded already exist in the cloud server. The subsequent users possess the files locally and the cloud server stores the authenticated structures of the files. Subsequent users need to convince the cloud server that they own the files without uploading them to the cloud server. Note that, these three phases (pre- process, upload, and De-Duplication) are executed only once in the life cycle of a file from the perspective of users. That is, these three phases appear only when users intend to upload files. If these phases terminate normally, i.e., users finish uploading in the upload phase, or they pass the verification in the De-Duplication phase, I say that the users have the ownerships of the files.

Advantages of Proposed System
➢ The duplicated files are mapped with a single copy of the file data by mapping with the existing file data in the cloud
➢ The comprehensive requirements in multi-user cloud storage systems and introduced the model of deduplicatable dynamic PoS.
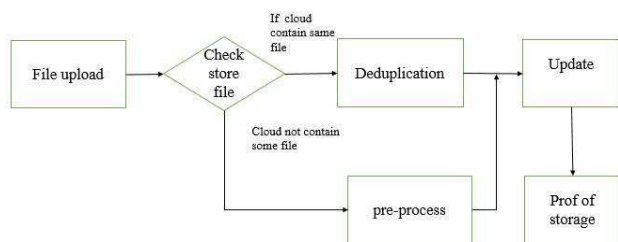
B. Algorithms
Let D be the Whole system which consists,
D= {I, P, O}

Where,
Q- Users Query {q1, q2..., qN}
P- Procedure,
F-Files set of {f1, f2,...,fn}
I-Input,
I-{F, Q},
O- Output.

Where:
F = Represents the file,

m1, m2, m3, m4= representing the ith block of the file, e = encryption key

Phase 1: Pre-process Phase
In the pre-processing phase,
$e \leftarrow H(F)$, $id \leftarrow H(e)$.

Then, the user announces that it has a certain file via id.

If the file does not exist, the user goes into the upload phase.

Otherwise, the user goes into the De-Duplication phase.

Phase2 The Upload File

$(C, T) \leftarrow Encoding(e, F)$
Let the file $F = (m_1, \ldots, m_n)$.

The user first invokes the encoding according

Phase3. The De-Duplication Data(file) res $\in \{0, 1\} \leftarrow$ De-Duplication $\{U(e, F), S(T)\}$

If a file announced by a user in the pre-process phase exists in the cloud server, the user goes into the De-Duplication phase and runs the De-Duplication protocol

Phase 4: The Update File
res $\in \{he*, (C*, T*)i, \perp\} \leftarrow Updating\{U(e, i, m, OP), S(C, T)\}$

In this phase, a user can arbitrarily update the file by invoking the update protocol

Phase 5: The Proof of Storage to Owner res $\in \{0, 1\}$ $\leftarrow Checking\{S(C, T), U(e)\}$

At any time, users can go into the proof of storage phase if they have the ownerships of the files. The users and the cloud server run the checking protocol.

RESULT AND DISCUSSIONS
User can upload, download update on cloud server and provide data De-Duplication.

CONCLUSIONS
Thus I have study about dynamic data De-Duplication and problems occur due to with cloud system and tries to resolve these problems with introducing DeposM2 system. That can positively respond to De-Duplication and Proof of Storage to original user in multiuser environment

---

REFERENCES

[1] Side channels in cloud services, the case of De-Duplication in cloud storage {Danny Harnik(IBM Haifa Research Lab,dannyh@il.ibm.com),Benny Pinkas (Barlian university, benny@pinkas.net), Alexandra Shulman-Peleg (IBM Haifa Research Lab,shulmana@il.ibm.com)}

[2] DEDUPLICABLE DYNAMIC PROOF OF STORAGE AND FILE SHARING FOR MULTI USER ENVIRONMENT (Mangesh Todkari, Saif Khan, Mayur Kadam, Kaustubh Borate (BE Dept. of IT, SKNSITS, Lonavala, India) *Vol-3 Issue-3 2017-IJARIIE-ISSN(O)-2395-4396*

[3] ASE-PoW: a Proof of Ownership mechanism for cloud De-Duplication in hierarchical environments Lorena Gonz_alez-Manzano1, Jose Maria de Fuentes1, and Kim-Kwang Raymond Choo2 , October 2016

[4] Secure and Efficient Cloud Data De-Duplication With Randomized Tag Tao Jiang, Xiaofeng Chen, *Senior Member, IEEE*, Qianhong Wu, *Member, IEEE*, Jianfeng. VOL. 12, NO. 3,

[5] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Proc. of FC*, pp. 136–149, 2010.

[6] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 340–352, 2016.

[7] Z. Xiao and Y. Xiao, "Security and privacy in cloud computing," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 843– 859, 2013.

[8] Ardagna, R. Asal, E. Damiani, and Q. H. Vu, "From Security to Assurance in the Cloud: A Survey," *ACM Comput. Surv.*, vol. 48, no. 1, pp. 2:1–2:50, 2015.MARCH 2017