

SURVEY ON CLOUD BACKUP SERVICES OF PERSONAL STORAGE

Nidhi Prakash Panpaliya¹, Prachi Sorte²

¹M.E Student, Dept. of Information Technology, RMD Sinhgad school of Engineering, Maharashtra, India

²Asst. Professor, Dept. of Information Technology, RMD Sinhgad school of Engineering, Maharashtra, India

Abstract

In widespread cloud environment cloud services is tremendously growing due to large amount of personal computation data. Deduplication process is used for avoiding the redundant data. A cloud storage environment for data backup in personal computing devices facing various challenge, of source deduplication for the cloud backup services with low deduplication efficiency. Challenges facing in the process of deduplication for cloud backup service are-1)Low deduplication efficiency due to exclusive access to large amount of data and limited system resources of PC based client site.2)Low data transfer efficiency due to transferring deduplicate data from source to backup server are typically small but that can be often across the WAN.

Keywords- Cloud computing, Deduplication, cloud backup, application awareness

1. INTRODUCTION

Cloud computing is more striking a service having a great potential to alter the large part of the IT industry. Cloud computing is the centralized storage for the data and it also provides the online access to various computer services and resources. Cloud computing broadly focuses on maximizing the efficiency of shared resources. Clouds can be classified according to users as private, public or Hybrid. Public cloud is the cloud which is made offered in pay-as-you-use manner to the general public, and their service being sold is utility computing. Private cloud will not be available to general public, they include control over the company's data and it ensures the security and also having the greater potential risk for data loss. Cloud user faces various security threads from inside and outside the cloud. And they are also responsible for the application level security [5]. To remove the bugs in distributed system in a cloud environment is one of the difficult challenges.

Cloud backup for end user's is nothing but an unlimited amount of data storage space which is secure and highly available for backup data from personal computing devices. Data storage space on the cloud is made accessible in pay-as-you-use behavior for the backup data. For data security of personal computing devices cloud backup service has become a cost-effective alternative. Due to increasing the large volume of data in corporate world, the market of cloud backup data services is growing up. As in the traditional backup data process there will be requirement of higher network bandwidth connectivity between the client and the server [3]. A cloud backup service provides the appropriate environment with restricted bandwidth and low data change rate.

Data deduplication an effective technology for eliminating the redundant data in backup data. In the web data there will be less focus on use of redundancy to reduce cost. Deduplication facing different challenges is that it requires computation rate and segmentation of the data.

Deduplication helps to identify the redundancy in data and transparently eliminating it. Deduplications of sub-files are more effective than whole file deduplication [2]. Sub-file is nothing but the chunk level deduplication. Efficiency of the deduplication process is based on the size of the chunks. As smaller sizes of chunk higher the duplication detection. There are various advantages of the deduplication process such as low CPU utilization, overall RAM usage reduced, save cloud storage space. An application aware deduplication process helps to reduced the computational overhead by using data chunking scheme and adaptive hash function which is based on the application awareness. A framework for video deduplication which is based on the application level view of redundancy at content level relatively than at the byte level [2]. This framework provides the similarity detection by content checking rather than byte level with the help of application awareness deduplication scheme.

The propose approach of application aware Local-Global source deduplication it not only formulate use of application awareness but also combines the local and global duplicate data detection. ALG-Dedupe scheme helps to achieve not only higher deduplication efficiency by reducing deduplication latency but also saves the cloud storage cost. Application awareness adapts different types of applications independently during the local and global duplicate check process, which helps to reduce the system. The proposed system combines local deduplication and global deduplication to balance the effectiveness and latency of duplicate data.

2. LITERATURE SURVEY

2.1 Deduplication Analysis

The propose takes an investigation on personal data by investigate how the space consumption efficiency of chunk data, data redundancy and decrease the computation overhead and also the functioning of the hash function. The study Application aware deduplication motivates the design

by following some interpretation of deduplication for cloud backup services in the personal computing environment [4]:

- The most of storage space is filled by a small number of compressed files with low sub-file redundancy after file-level deduplication.
- The maximum arrangement of hash finger printing and chunking method which can helps to reduce the system overhead on resources.
- Data shared between various types of applications is insignificant due to difference in data content and format.
- To achieve high deduplication efficiency with various application datasets there will be the best choices of chunking methods and deployment of chunk level redundancy.

3. THE SYSTEM ARCHITECTURE

The main purpose of the local and global deduplication scheme is to utilizing not only low overhead but also to utilize high overhead cloud assets to reduce the computational transparency by using an intelligent data chunking scheme. In this system there will be the adaptive use of the hash function based on the application awareness [1]. To advance the efficiency of the system and low system overhead on client side it combines the local and global source deduplication with application awareness.

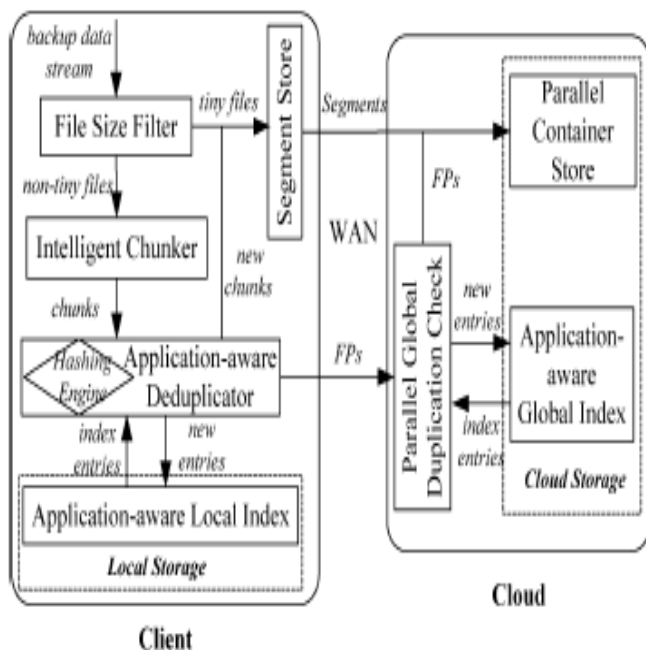


Fig-1: Application aware local -Global deduplication scheme

The propose system contains four main components:

3.1 File Size Filter for Backup Data Stream

A personal computing device contains the most of the tiny files which holds a negligibly small percentage of the storage capacity. To decrease the data overhead the planned system filters out these tiny files in file size filter before

stating the deduplication process. And make the group of all tiny files together into large unit in the segment store and that will be stored as segment in segment store. Efficiency of the data transfer over WAN is increase due to segment store.

3.2 Intelligent Chunker

Data chunking scheme having the great impact on the efficiency of data deduplication. There will be the contrary relationship among the deduplication ratio and the average chunk size. Chunking can be prepared on the basis of frequency based or content based. CBC is the stateless chunking algorithm which divides the long stream of data into smaller units by removing duplicates. To achieve a better mean balance between deduplication ratio and deduplication overhead, we deduplicate compressed files with WFC for its low sub-file redundancy.

3.3 Applications-Aware Deduplicator

After performing the data chunking in Chunker, the deduplication of the data chunks will be performed in application aware, which generates the FP in hash engine of the data chunks and detecting replica chunks in both local client and remote cloud. Application aware local and global deduplication strikes the superior stability between computation overhead on client side and hash collision will be avoided to keep data reliability. For performing deduplication on client side and on global cloud it requires two types of application aware indices such as local index on client side and global index on cloud side. ALG Deduplication performs the periodic synchronization to backup application aware index and to protect the data integrity of PC backup datasets.

3.4 Segment Management

To reduce the cost of cloud storage and avoiding higher overhead of network protocol due to small file transfer, Application aware deduplication will group the deduplicate data of various smaller files and chunk into large units that will be known as segments. These group of deduplicate data will be stored in the segment store before transferring these data over WAN.

3.5 Container Management

A container is nothing but the self describing data structure in chunk descriptors for stored chunk. Container is maintained for each arriving backup data stream. These backup data is nothing but the segments send over the cloud and which will be routed to store node over the cloud with its respective fingerprints.

4. CONCLUSION

In this paper we combine two kind of deduplication process namely local and global deduplication process. This can help us to balance the efficiency and latency of deduplication. We propose ALG Dedupe for improving the deduplication efficiency. This approach is also used for

minimizing computational overhead and maximizes deduplication effectiveness, along with the computational overhead this approach improves power efficiency, reduces cloud cost without enlarging a backup window size.

REFERENCES

- [1]. Y.Fu, H.Jiang, "Application-Aware Local-Global Source Deduplication for Cloud Backup Services of Personal Storage," IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO 5, MAY 2014.
- [2]. A. ElShimi, R. Kalach, A. Kumar, J. Li, A. Oltean, and S. Sengupta, "Primary Data Deduplication Large Scale Study and System Design," in Proc. USENIX ATC, 2012, pp. 285-296.
- [3]. A. Katiyar and J. Weissman, "ViDeDup: An Application-Aware Framework for Video De-Duplication," in Proc. 3rd USENIX Workshop Hot-Storage File Syst., 2011, pp. 31-35.
- [4]. Y. Fu, H. Jiang, N. Xiao, L. Tian, and F. Liu, "AA-Dedupe: An Application-Aware Source Deduplication Approach for Cloud Backup Services in the Personal Computing Environment," in Proc. 13th IEEE Int'l Conf. CLUSTER Comput., 2011, pp. 112-120.
- [5]. M. Armbrust, A.Fox, I.Stoica, M.Zaharia, "A view of cloud computing," Communication of the acm, April 2010,vol.53,no.4.

BIOGRAPHIES



Nidhi Prakash Panpaliya has completed bachelor degree from sant. Gadage baba Amravati University and now M.E Student at RMD Sinhgad school of Engineering, Pune, Maharashtra, India



Prachi Sorte, Dept. of Information Technology, Working at RMD Sinhgad school of Engineering, Warje Savitribai Phule, Pune university, Maharashtra, India