International Journal of Innovative Research in Information Security (IJIRIS)
Issue 2, Volume 5 (May 2015)
ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

Data Deduplication with Security in Cloud Data Centres

R.Karthikeyan
Multimedia technology,
KSR College of Engineering
Namakkal, India

Mr. R.Velumani
Asst.Professor,
KSR College of Engineering
Namakkal, India

Abstract--- Cloud data centers that can be used for storing and maintain the personal data backups. Nowadays increase in the data storage value there is more space needed. The data storage space needed more and more For that deduplication is one of the storage technology that can be used in the data center. Because deduplication eliminates the redundant data and allow only the unique data in the cloud backup. And it can improves the storage efficiency and also the network bandwidth.

Keywords— cloud data backup, deduplication,

I. INTRODUCTION

Cloud computing technology is a utility oriented computing service through the internet delivering the IT services on demand. There are different types of services that are provided by the cloud computing .There is

1. *Infrastructure as a service.*
2. *Platform as a service.*
3. *Software as a service.*

These are services provided by cloud providers for the cloud users. These services are accessible only through the internet. Example Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) is the popular cloud service providers.

There are the two challenges in the cloud backup services. There are the storage space and the network bandwidth. Nowadays the large volume of the digital data has been increased. The data redundancy is one of the storage technologies which can the stores the multiple copies of the same data. The data backup that could be helpful at the disaster recovery .There is lot of storage space needed for storing the data content. This results increase in the storage cost especially for the large data storage IT companies. The data backup became a cost could be effective approach. The other challenge is large backup window, due to the low network bandwidth between user and service provider constraining the data transmission. The backup window is represented by the time spent on sending specific dataset to backup destination.

The data deduplication is one of the compression techniques which can eliminate the duplicate copies of the data. Data deduplication allows only the unique content to the cloud storage. And also increase the network bandwidth.

Data deduplication is an approach that reduces the storage area needed to store data. The amount of data that has to transfer over a network. The processes are partitioned large data objects called chunks or blocks. For each block generate the unique key by the cryptographic hash function called fingerprints. And then replace the duplicate chunks with their hash fingerprints by the index lookup table. And finally transfer the unique chunks for the communication or to store data in the cloud.

Data deduplication that follows two approach to implementation. There are finger prints and delta based deduplication approaches. The delta deduplication approach is the oldest method performs the chunking, but it is not searching similar, but necessarily identical data blocks. And the fingerprint base data deduplication approaches all the chunks are fingerprinted using the cryptographic hash function. And then it can search in the index lookup table for the identical file. If there is identical file it cannot be stored. It is not identical means it can be stored in the index lookup table and then it can be transfer over the network.

The data that can be stored at the cloud backup can be secure. Because the personal data that cannot be visible to the others and cannot be easily downloaded. So the data after deduplication that performs encryption and decryption process for secure backup.

II.RELATED WORK

In the traditional storage stack comprising applications, file systems and storage hardware, each of the layers contains different kinds of information about the data they manage and such information in one layer is typically not available to any other layers. Codesign for storage and application is possible to optimize deduplication based storage system when the lower-level storage layer has extensive knowledge about the data structures and their access characteristics in the higher-level application layer.

III.EXISTING SYSTEM

Application –aware Local-Global source deduplication scheme, in this data deduplication by combining can perform local and global deduplication.it can save the cloud storage capacity and reduce deduplication time.

It is a combination of local-global source deduplication with application awareness which improves effectiveness of deduplication with low system overhead at the client side.

ALG dedupe consist a file size filter where small files are first filtered out by for efficiency reasons, and backup data Streams are broken into chunks by an intelligent chunker using an application aware chunking strategy .

Application aware deduplicator generating chunk fingerprints in hash engine and performing data redundancy check in application-aware indices in both local client and remote cloud for the data chunks. the fingerprints are indexed for the local redundancy check.

The meta data for the file containing chunk is updated to the point to the location of the existing chunk if the similar data is found. If there is no similar data further it can checks for the global deduplication. if a match is found it will be the duplicate chunks otherwise it can be updated in the index and it will be the unique data.

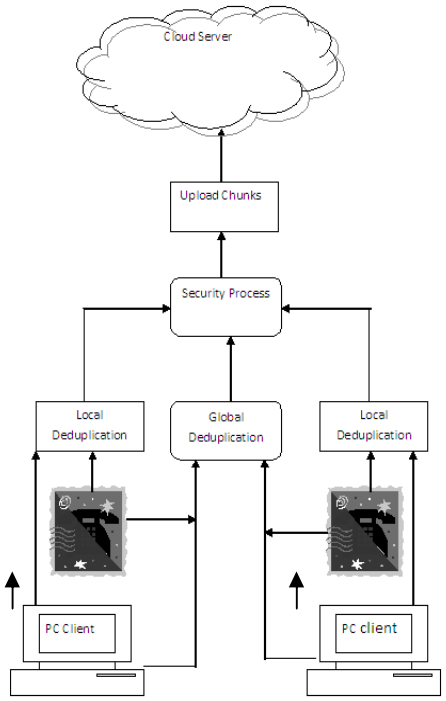
To improve the network bandwidth efficiency over WAN, at client side the fingerprints will be transferred in batch and new data chunks will be packed into large units called segments in the segment store module with tiny files before their transfers to reduce cloud computing latency. The local duplicate detection in ALG Dedupe significantly reduces the number of global fingerprint lookup requests. ALG-Dedupe are shown to improve the deduplication efficiency with very low system overhead. It has reduced the backup window size and improves power-efficiency.

IV. PROPOSED SYSTEM

Security ensured Application aware Local-Global source Deduplication (SALG-Dedupe) scheme is proposed to perform deduplication with security. Encrypted cloud storage model is used to secure personal data values. Deduplication scheme is adapted to control data redundancy under Smart Phone environment. File level deduplication scheme is designed for global level deduplication process.

ARCHITECTURE DIAGRAM

The deduplication system is adapted for the Computer and Smart phone clients. The system provides security for the backup data values. Small size files are also included in the deduplication process. The system is divided into six major modules. They are Cloud Backup Server, Chunking Process, Block level Deduplication, File level Deduplication, Security Process and Deduplication in Smart Phones



The cloud backup server module is designed to maintain the backup data for the clients. Chunking process module is designed to split the file into blocks. Block signature generation and deduplication operations are carried out in block level deduplication module. File level deduplication module is designed to perform deduplication in file level. Data security module is designed to protect the backup data values. Deduplication process is performed in the mobile phones in Deduplication under Smart phones module.

CLOUD BACKUP SERVER

The cloud backup server module is designed to maintain the backup data for the clients. Chunking process module is designed to split the file into blocks. Block signature generation and deduplication operations are carried out in block level deduplication module. File level deduplication module is designed to perform deduplication in file level. Data security module is designed to protect the backup data values. Deduplication process is performed in the mobile phones in Deduplication under Smart phones module.

CHUNKING PROCESS

File size filter is used to separate the tiny files. Intelligent chunker is used to breakup the large size files into chunks. Backup files are divided into compressed files, static uncompressed files and dynamic uncompressed files. Static files are un-editable and dynamic files are editable. Compressed files are chunked with Whole File Chunking (WFC) mechanism. Static uncompressed files are partitioned into fix-sized chunks by Static Chunking (SC). Dynamic uncompressed files are broken into variable-sized chunks by Content Defined Chunking (CDC).

BLOCK LEVEL DEDUPLICATION

Chunks finger prints are generated in the hash engine. Rabin hash functions with 12 bytes are used as chunk fingerprint for local duplicate data detection for compressed files. Message Digest MD5 algorithm is used for global deduplication process in compressed files. Secure Hash Algorithm (SHA1) is used for deduplication in uncompressed static files. Chunks finger prints are generated in the hash engine. Rabin hash functions with 12 bytes are used as chunk fingerprint for local duplicate data detection for compressed files.

Message Digest MD5 algorithm is used for global deduplication process in compressed files. Secure Hash Algorithm (SHA1) is used for deduplication in uncompressed static files. Dynamic uncompressed files are hashed using Message Digest (MD5) algorithm. Deduplicate detection is carried out in the local client and remote cloud. Fingerprints are indexed in local and global level. Deduplication is performed by verifying the finger print index values.

FILE LEVEL DEDUPLICATION

Tiny files are maintained under segment store environment. File level deduplication is performed on files with the size less than 10 KB. File level fingerprints are generated using Rabin hash Function. Deduplication is performed with file level fingerprint index verification mechanism.

SECURITY PROCESS

The backup data values are maintained in encrypted form. Modified Advanced Encryption Standard (MAES) algorithm is used in the encryption/decryption process. Encryption process is performed after the deduplication process. Local and global keys are used for the data security process. Deduplication in Smart Phones.

DEDUPLICATION IN SMART PHONES

The deduplication process is tuned for smart phone environment. Smart phones are used as client for cloud backup services. File level and block level deduplication tasks are supported by the system. Data security is also provided for the smart phone environment.

V. CONCLUSION

The cloud data center allows only the redundant data to the cloud storage using the above proposed system with security. And now the data can be stored safe and the secure in the data center.

REFERENCES

[1] Yinjin Fu, Hong Jiang, Nong Xiao, Lei Tian, Fang Liu and Lei Xu, “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] Keke Chen, James Powers, Shumin Guo and Fengguang Tian, “CRESP- Towards Optimal Resource Provisioning for MapReduce Computing in Public Clouds” IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, June 2014.
[3] Cheng-Kang Chu, Sherman S.M. Chow, Wen-Guey Tzeng, Jianying Zhou and Robert H. Deng, “Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage” IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014.
[4] Xuyun Zhang, Laurence T. Yang, Chang Liu and Jinjun Chen, “A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using MapReduce on Cloud” IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014.
[5] Manel Bourguiba, Kamel Haddadou, Korbi and Guy Pujolle, “Improving Network IO Virtualization for Cloud Computing” IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 3, March 2014.
[6] Dinil Mon Divakaran and Mohan Gurusamy, “An Online Integrated Resource Allocator for Guaranteed Performance in Data Centers” IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, 2014.
[7] A. Katiyar and J. Weissman, ‘‘ViDeDup: An Application-Aware Framework for Video De-Duplication,’’ in Proc. 3rd USENIX Workshop Hot-Storage File Syst., 2011.
[8] F. Douglis, H. Qian and P. Shilane, ‘‘Content- Aware Load Balancing for Distributed Backup,’’ in Proc. 25th USENIX Conf. LISA, Dec. 2011
[9] S. Kannan, A. Gavrilovska and K. Schwan, ‘‘Cloud4HomeV Enhancing Data Services with @Home Clouds,’’ in Proc. 31st ICDCS, 2011.
[10] P. Anderson and L. Zhang, ‘‘Fast and Secure Laptop Backups With Encrypted De-Duplication,’’ in Proc. 24th Int’l Conf. LISA, 2010.
[11] Y. Tan,H. Jiang, D. Feng, L. Tian, Z. Yan and G. Zhou, ‘‘SAM: A Semantic-Aware Multi-Tiered Source De-Duplication Frame-Work for Cloud Backup,’’ in Proc. 39th ICPP, 2010, pp. 614-623.