International Journal of Applied Engineering Research.
ISSN 0973-4562 Volume 10, Number 10 (2015), pp. 27413-27418

http://www.ripublication.com

Enriching utilization of Cloud storage system

Shraddha Ghogare and Ambika Pawar
Department of Computer Science and Engineering,
Symbiosis Institute of Technology,
Pune, MH, India.
E-mail: shraddha.ghogare@sitpune.edu.in

Ajay Dani
Symbiosis International University.

Abstract

As the technology has witnessed enormous growth in recent years, the need for storing and securing the data generated has ascended subsequently. Many organizations, therefore, began to utilize Cloud storage to fulfil this requirement. Furthermore, it was observed that cloud storage is mainly being used for Archive/Backup purpose. However, as the data being generated is of massive size, replicating it may incur some issues in near future. This paper, presents a solution which takes into account this problem, and gives a solution to it by using the concept of encoding.

AMS subject classification:
Keywords: Cloud Storage, Data replication, Encoding, Decoding, split, Data Security, Data integrity, Erasure codes.

1. Introduction

The word cloud computing essentially is used in many contexts now a days and it is not only used for performing computations on data but also for backup/archive purpose. As it is known that the data being generated now is huge as compared to it was in earlier days, therefore, even for storing such data cloud is being used, and the repository is called as Cloud Storage. Because of the numerous advantages like scalability, elasticity, pay as you go basis; cloud storage are popular among the industry and businesses.

However, the security of the data which is stored on cloud and effectual usage of the storage has been a concern. Three core aspects of security, i.e. Confidentiality, Integrity and Availability of data (CIA) are always desired in order to provide security

and privacy of data. Nevertheless, there is a lot of research going on to mitigate with the risks involved. Such research involves use of encryption techniques like AES, DES and now advanced technique of data colouring [3].

Furthermore, Access control schemes like Role-based Access method wherein every user is assigned specific roles and therefore corresponding privileges also helped to solve certain issues like insiders misusing access rights. SLAs(Service Level Agreements) are agreed by involved parties so as to deal with certain protocols to avoid data theft/data loss and access right violation.

Another approach to solve security breaches in cloud is to incorporate multi-cloud architecture, wherein, rather than storing flat files on single cloud, the data is divided into multiple parts and these parts are uploaded on multiple clouds. This approach helps to assure privacy of data up to certain extent. Even though these techniques offer security and privacy of the data on cloud storage, data integrity and availability are still important issues.

This paper proposes a scheme wherein data integrity and confidentiality along with effectual usage of cloud as storage are ensured by using concepts of erasure codes,hashing and muli-cloud architecture.

2. Related Work

Storage has always been a topic discussed since the evolution of technology. In its initial days, the capacity provided was in form of registers; and slowly and steadily it moved towards floppy disks, RAMs, hard disks, flash drives,etc. With the embellishment of technology, cloud computing came into existence and consequently, the cloud storage. Now a days the data that is getting generated is considerably large. For an instance, the data being generated in the form for flat files like text files,logs is massive. In order to maintain backup/archive,the need for storage is subsequent high as the data is generally replicated.

[9] says that rather than replicating the data over the clouds, encoding can be applied so as to save storage space. Moreover, there have been numerous efforts taken till now to secure data uploaded on cloud. [1] enlisted few security breaches and has given out some guidelines to follow to avoid them.

Also, to ensure integrity of data, usage of IDA algorithms was studied [5], wherein data in form of matrix of bytes is multiplied with a pivot matrix and resultant matrix is dispersed in such a manner that, out of n dispersed parts, only k can reconstruct the original byte matrix. However, data loss is still a core concern.

The utilization of RAID(redundant array of independent disks) like system has been done in [8], wherein a redundant array of clouds has been developed. This approach makes use of multi-cloud architecture to meet CIA and is termed as RAIC. Further it was optimized by [4] to develop a Desktop application called as NubiSave. The architecture developed is called as RAOC(Redundant Array of Optimized Clouds).

This paper implements a scheme wherein, encoding decoding techniques,Hashing and RAOC concepts are used to provide integrity, confidentiality and availability to data

uploaded on cloud and to save cloud storage space by avoiding replication of entire file on clouds.

Further this paper is organized as- Section 3 delineates Proposed system, section 3.1 shows the results and Section 4 concludes the paper.

3. Proposed Scheme

To provide security and privacy to data uploaded on cloud,the proposed scheme makes use of multiple cloud service providers over single one and in order to improve utilization of cloud storage, the techniques of file splitting and encoding are used. Figure 1 delineates this mechanism.

In order to provide a complete security solution, to ensure integrity of files uploaded on cloud, hashing algorithm (SHA-256) has been implemented as the first step of algorithm. The hash code is generated at the beginning for the file which is compared when the user asks to download the file. If hash code differs, it means integrity of the file has been lost.

After this, the shuffle algorithm, as described in [7] shuffles bytes of the file based on the limit specified. The limit of how many bytes to shuffle can be given out by the user as per his requirements. For an instance if limit is specified as 3, groups of 3 bytes each are shuffled with one another.

Split algorithm is then applied to form chunks of the file. Furthermore, Figure 2 demonstrates this technique. The central application server consists of two implicit layers as - i. Encoder-Decoder, ii. Cloud providers connectivity.

The first part takes in the flat file as input and produces encoded chunks of it by using either Apache Base64 or Java Base64 encoding. Then these encoded parts of file are uploaded to multiple cloud in a way that all even chunks are send to Dropbox and

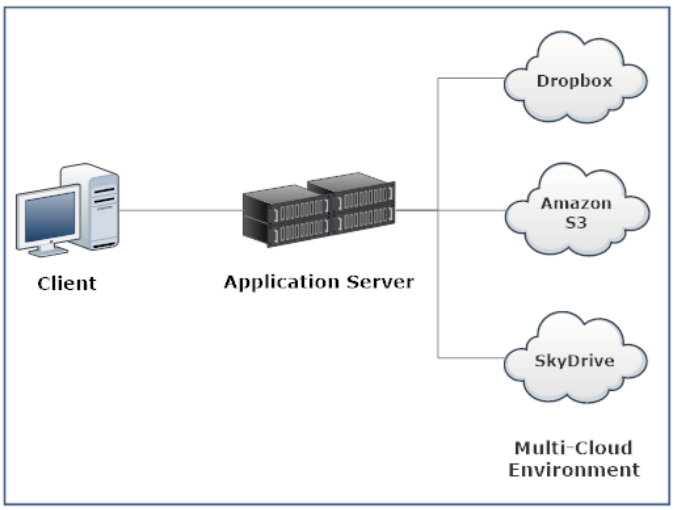

Figure 1: Utilization of Multi-Cloud environment in proposed System.

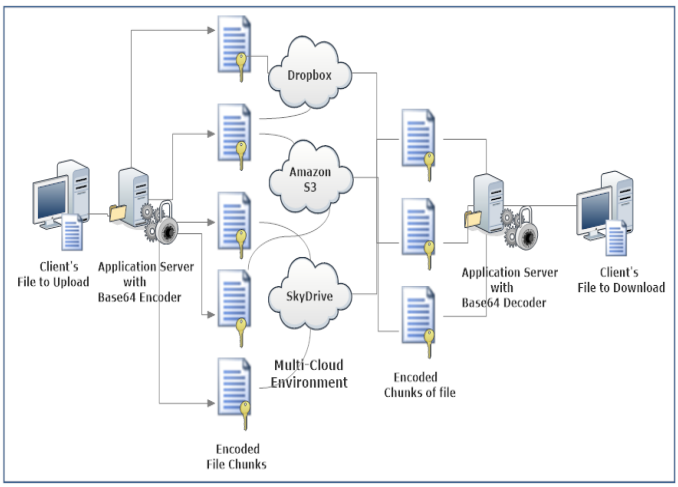

Figure 2: Detailed Architecture of proposed Scheme.

odd ones to SkyDrive. Advantage here is that no single cloud has entire file, hence confidentiality is achieved. Also, as the part of files are encoded ones, error correction is ensured; which leads to data reliability. Also replicas of chunks are maintained on several clouds so as to maintain availability of file.

On the other hand, to get back the original file, reverse procedure is applied. Chunks of files are decoded and are merged together and reverse shuffling algorithm is applied. Lastly, a hash code is generated for the file and is compared with previously computed one, to ensure data integrity. Also, in order to reduce time taken for this process, concepts of multi-threading and distributed systems are to be used in time to come.

3.1. Results

The aforementioned technique has been implemented using two libraries. i. Sun Java Base64 ii. Apache Base64. Both libraries have advantages and drawbacks. Figure 3 depicts results observed so far.

A comparison is made based on the results to find out which library suits where. It is observed that Java Base64 takes significantly less time to encode and decode the data than that of Apache Base64. However, [6] explains why not to use Java Base64 over Apache. Furthermore, another library called MigBase64 is under implementation which claims to be faster than java codecs [2].

Figure 4 shows overall performance of the system. Uploading speed is a little better than the downloading speed. However, performance of the scheme depends upon the Internet speed user has. Today, as the data growth is huge and files are of TBs and PBs size, the storage needed to store and backup these files is considerably high. With the proposed scheme, it has been reduced drastically. For an instance, the storage needed to store and backup 2MB file on single cloud is 2MB and another 4 MB- 8MB is required to replicate it on backup servers (considering file is replicated on 2 or 4 backup servers).

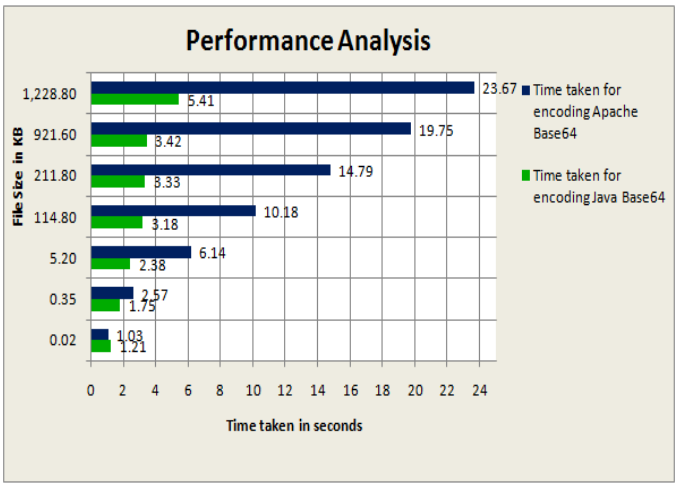

Figure 3: Performance Analysis for Encoding with Java and Apache Base64.

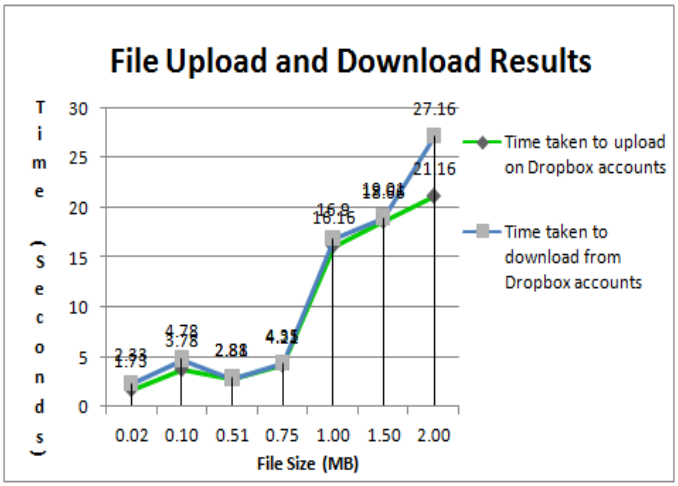

Figure 4: Performance analysis of file upload and download with proposed scheme.

Alternative to this approach is to use proposed scheme which makes effectual use of multi-cloud architecture to improve utilization of cloud storage.

Also, level of integrity,confidentiality, reliability and availability is higher than the previous approaches. As of now, two accounts of Dropbox are in use for forming multi-cloud architecture. The implementation of SkyDrive is in progress and will be incorporated in this scheme in near future. With this addition, CIA may achieve higher level of security. Also, in time to come, concept of erasure codes (IDA) is to be utilized in this scheme.

4. Conclusion

This scheme takes into consideration problem of data loss, confidentiality integrity and mitigates over them by using encoding techniques and a combination of them. Also, the system gives a complete solution by using hashing and usage of multiple clouds. The implementation as of now focused mainly on encoding, however, in time to come, use of erasure codes will be embedded into the system. Also, to reduce processing time, distributed system will be used along with central application server. In addition to this, to provide data confidentiality, encryption algorithms can be incorporated in the system.

References

[1] Google. Google's Approach to IT Security. https://cloud.google.com/files/Google-CommonSecurity-WhitePaper-v1.4.pdf/, 2012. [Online; accessed 26-May-2014].

[2] Mikael Grev. MiGBase64. http://migbase64.sourceforge.net/.

[3] Kai Hwang and Deyi Li. Trusted cloud computing with secure resources and data coloring. *Internet Computing, IEEE*, 14(5):14–22, 2010.

[4] Johannes Muller Josef Spillner and Alexander Schill. Creating optimal cloud storage systems. In *Future Generation Computer Systems*, page 1062âŁ“1072, 2013.

[5] Sian-Jheng Lin and Wei-Ho Chung. An efficient (n, k) information dispersal algorithm for high code rate system over fermat fields. *Communications Letters, IEEE*, 16(12):2036–2039, 2012.

[6] Oracle. Why Developers Should Not Write Programs That Call 'sun' Packages. http://www.oracle.com/technetwork/java/faq-sun-packages-142232.html.

[7] Ambika Vishal Pawar and Ajay A Dani. Design of privacy model for storing files on cloud storage. *Journal of Theoretical & Applied Information Technology*, 67(2), 2014.

[8] Stephan Grob Ronny Seiger and Alexander Schil. Seccsie: A secure cloud storage integrator for enterprises. In *Commerce and Enterprise Computing (CEC)*.

[9] Eric Slack. Big file storage scales for large data applications. http://searchstorage.techtarget.com/feature/Big-file-storage-scales-for-large-data-applicationss, 2014. [Online; accessed 23-Aug-2014].