ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 10, April 2015

Improve Fault Tolerance and Dependable Data Integrity Protection of Storage System in Cloud of Clouds

Neha Sharma (P.G Student, CSE Department, MVJCE, Bangalore)
Sharavana K (Associate professor, CSE Department, MVJCE, Bangalore)

Abstract*— To protect outsourced data in multiple cloud storage environment, we provide fault tolerance and dependable data integrity for cloud storage. Suppose if cloud storage experiences the permanent failure and loses all its data then, there is a need to regenerate the lost outsource data with help of other surviving cloud storages. We propose a Proxy based Storage for fault tolerant multiple cloud storage environment called as Network coding based storage. Functional minimum-storage regenerating (FMSR) codes are used for regenerating lost outsourced data in multiple cloud storage environment. We also validate that FMSR codes reduce repair traffic and incur less monetary cost during data transfer than traditional erasure [RAID-6].

***Index Terms*——* **Fault Tolerance, Regenerating Codes, Repair Traffic, Network-Coding**

I. INTRODUCTION

Cloud storage is a service model in which data is maintained, managed and backed up remotely and made available to client or end users over a network. In a case cloud providers uses single cloud storage or a system is designed with single cloud storage raises concerns of failure and vender lock ins. To provide a solution to this concern data can be distributed across different cloud providers by using multiple cloud storage we can improve fault tolerance and integrity of cloud storage.

In cloud storage we consider two types of failure: Transient and permanent failure. Suffers short term failure or predictable permanent failures. There are live cases which shown that permanent failures do occur and are not always predictable [8], [6] Conventional erasure codes were used for transient failure. So in this paper focuses on unexpected permanent failure. When cloud storage experiences permanent failure, it is necessary to active repair in order to maintain fault tolerance and redundancy that leverages the benefits of cloud storage. Since data striped across different cloud providers the repair operation fetch data from other surviving cloud over network and rebuilt the lost data in new cloud storage called cloud of cloud, so moving an excessive amount of data across cloud can introduce significant monetary costs due to data migration. In this paper we propose a proxy based design for multiple cloud providers which provide fault tolerant storage and propose implementable design for FMSR (functional minimum storage regenerating) codes[6], [7].

II. NEED OF REPAIR OPERATION IN MULTIPLE CLOUD ENVIRONMENT

In this section, we describe the need of repair operation in multiple cloud storage, especially in permanent cloud failure. we consider two types of cloud storage failure : Transient and permanent failure. Short term failure is consider as transient failure. In multiple cloud environments, failed cloud will return to normal after some time and no data are loss. Table 1 show several live examples of transient failure where the duration of the failures ranges from several minutes to several days.

Unexpected Permanent failure

A long term failure is called as permanent failure, means that the data on a failed cloud will become unavailable. So outsourced data cannot be recoverable in unexpected permanent failure as compare to transient failure. Although we expect that a unexpected permanent failure rarely to happen. There are several situations where permanent failure are possible - unavailability of data center in disasters, threatened attacks, loss of data and corruption. AFCOM [9] found that 65 % of data centers have no procedure to handle with cyber criminals. So there is a need to repair the lost data from permanent cloud failure and retrieve data from other surviving cloud.

FMSR Codes

In this section we propose a proxy based storage design that connects multiple cloud storage as shown in Fig. 1a. We consider if a cloud encounters an unexpected permanent failure, the proxy based storage activates the repair operation shown in Fig. 1b. The proxy based storage design takes lost data pieces from other surviving cloud storage and rebuilt new data pieces and stores in to new cloud which is called as cloud of cloud. One important note is repair operation does not involve direct interactions among cloud storage. Here we considered fault tolerant storage which is based on a type of MDS (maximum distance separable) codes. Given a file of size M , we divide a file object into equal size of native chunks, which are combined linearly to form code chunks. when an (n,k) MDS codes used, the native/code chunks are then distributed over n nodes, each storing chunks of total size M/k, means that original file may be rebuilt from the chunks contained in any k of the n nodes. Thus it can tolerates the failure of any n-k nodes. The additional feature of FMSR codes is that rebuilt the native chunks stored in a failed node

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 10, April 2015

can be achieved by downloading up to 50 % less data from the surviving nodes than rebuilt the complete file.

TABLE1. Example of Transient Failures in Different Cloud Storage

Cloud Service	Failure Reason	Duration	Date
Google Gmail	Software bag [4]	4 Days	Feb 27-Mar 2,2011
Google Search	Programming error [12]	40 Mins	Jan 31,2009
Amazon S3	Gossip protocal blowup [3]	6-8 Hours	July 20,2008
Microsoft Azure	Malfunction in Windows Azure [11]	22 Hours	Mar 13-14,2008

In this paper we consider a multiple cloud storage setting with two reliability factor: Fault tolerance and recovery of lost data. First assume that the multiple cloud storage environment is double fault tolerant means end user can always access their data as long as no more than two clouds storage experience transient failure (examples in Table 1). Second one single fault recovery in multiple cloud environment given that a permanent cloud of storage failure is less frequent but expected. Our main purpose is to minimize the cost of repair operation for unexpected permanent failure and compare two codes traditional RAID-6 codes and FMSR (functional minimum regenerating) codes with double -fault tolerance. We consider the repair traffic as the amount of outbound data being downloaded from the surviving cloud storage during single cloud storage failure recovery. We are trying to minimize the repair traffic for cost effectiveness. We do not consider inbound traffic means the data being written to new cloud storage. Let us assume that we store a file of size M on four cloud storage. First consider conventional RAID-6 codes, which are double fault tolerant. RAID-6 code implementation is based on the Reed Solomon code [12] as shown in Fig.2 (a)

In RAID-6 a file object is divided in tow native chunks (i.e., A and B). Each native chunks size will be M / 2. After that add two more code chunks those are formed by linear combination of native chunks. Suppose if a node 1 experiences any failure then proxy storage node must download the equal no of chunks as the original file from two other nodes (i.e., Node 2 and node 3 with data pieces B and A+B). Then rebuilt the and stores the lost data chunks A in to new node. So the total storage size is 2M, while the repair traffic is M. Regenerating codes are introduced to minimize the repair traffic. One class of regenerating codes is called as exact minimum storage regenerating (EMSR) codes. EMSR codes maintain the same storage size as RAID-6 codes, while the storage node send encoded chunks to proxy based storage node so as minimize the repair traffic. In Fig. 2b represents the double fault tolerant implementation of EMSR codes. Here we divide the file into four native chunks and allocate the native chunks and code chunks. Let us assume Node 1 experiences any failure, to repair it each surviving node send XOR summation of data chunks to proxy based storage, then it rebuilt the lost chunks. Now we can see that in EMSR codes the storage size is 2M same as RAID-6, and the repair traffic is 0.75 M, which represents 25 percent of saving as compared to RAID-6 codes.

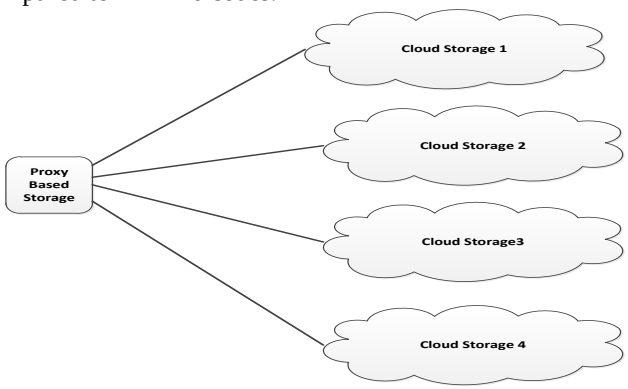

Fig. 1 (a) Normal operation

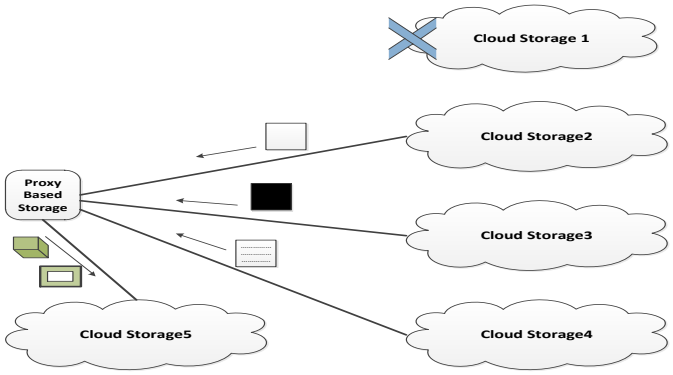

Fig. 1 (b) Repair operation

We now presents the double fault tolerant implementation of FMSR codes which represents in Fig.2(b). Here we are dividing the whole file into four chunks, and built 8 distinct code chunks(i.e., P1.....P8) formed by linear combinations of native chunks. The size of each code chunks is M/4 . Any 2 nodes can be used to reconstruct the original four native chunks. Let us assume a node 1 experiences failure, then proxy based storage fetch one code chunk from each surviving nodes. Hence proxy based storage downloads three code chunks, and chunk size will be M/4 considered, after that proxy storage reconstruct two code chunks P'1 and P'2 formed by linear combination of collected three code chunks, thereafter proxy storage writes P'1 and P'2 into a new node.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 10, April 2015

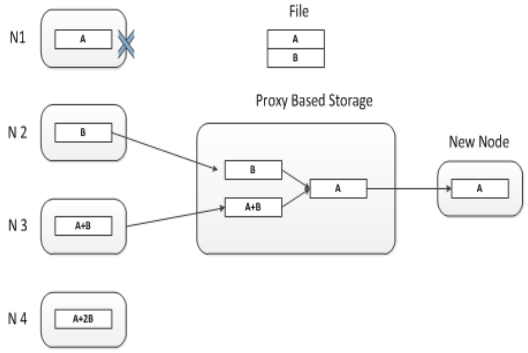

Fig.2(a) RAID-6

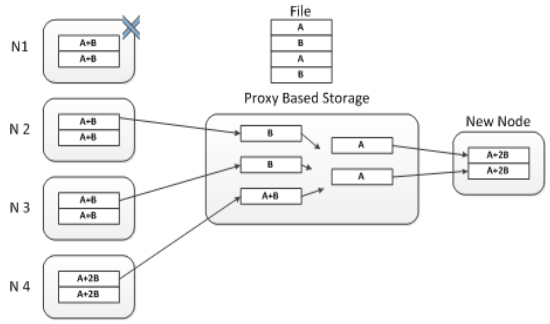

Fig. 1 (b) Repair operation

In FMSR codes, the storage size is 2M which is same as RAID - 6 but the repair traffic is 0.75 M which is same as in EMSR codes. The main key feature of FMSR codes is that nodes does not perform encoding at the time of repair. To generalize double fault tolerant, FMSR codes for n nodes, divide the file M in to 2(n-2) native chunks and use those native chunks for generating 2n code chunks. After that each node can able to store two code chunks of size M divide by 2(n-2) each. Thus the total storage size of the node is Mn divides by n-2. For repairing failed node ,we have to download one chunk from each other surviving (n-1) nodes, so the repair traffic can be M(n-1) divides by 2(n-2). In contrast for RAID-6 codes, total size is Mn divides by n-2, where the traffic repair is M. If n is large FMSR codes can Reduce the repair time by close to 50 percent. Important note is that FMSR codes are non -systematic, so they keep only code chunks not the native chunks. To access a single chunk, we have to download and decode the entire file object for the specific chunk. In systematic FMSR codes actually those are traditional RAID storage kept only native chunks instead of code chunks.

III. EXISTING SYSTEM

There are many systems proposed for multiple cloud storage. HAIL provides data integrity and availability that gives insurance for the stored data [7]. RACS uses erasure coding scheme to mitigate vendor locks in at the point of switching cloud providers and also fetch the data from the failed cloud, and move the data into a new one cloud. In contrast RACS, proxy based storage excludes the failed cloud storage in repair time. Vokolic [5] advocates using multiple independent clouds to provide Byzantine fault tolerance. DEPSKY [3] address Byzantine fault tolerance by combing encryption and erasure coding for stored data. All the above systems uses erasure codes for providing fault tolerance, while proxy based storage consider regenerating codes , which focused on fault tolerance and storage repair. Minimizing I/Os- There are several studies represents efficient single storage node failure recovery process which minimize the amount of data read for XOR based erasure code, for example there are few papers propose optimal recovery for RAID 6 codes and minimize the amount of data read up to 25% for any number of storage node, but In FMSR codes can achieve 25% for 4 nodes and up to 50% if number of nodes increases.

IV. PROPOSED SYSTEM

We propose a proxy based, a multiple cloud storage system that provides the reliability of cloud backup storage, also provides the cost effective repair when any of the cloud experiences the failure .This proxy based system uses FMSR codes which regenerates new parity chunks at the time of repair. FMSR codes eliminates the need of encoding requirement of cloud storage nodes during repair operation. and multiple clouds. The design is based on three layers file system layer, coding layer and storage layer. In file layer ,a file is tagged with meta data sub components, which is replicated in each repository . Meta data holds the file details. The coding layer responsible for coding and decoding functions and storage layer maintain the read and write request. The coding layer implements both concepts RAID 6 and FMSR codes.

V. CONCLUSIONS

We propose a proxy based, a multiple cloud storage system that provides the reliability of cloud backup storage, also provides the cost effective repair when any of the cloud experiences the failure .This proxy based system uses FMSR codes which regenerates new parity chunks at the time of repair. FMSR codes eliminates the need of encoding requirement of cloud storage nodes during repair operation.

VI. ACKNOWLEDGMENT

I convey my sincere thanks to our college Principal and other teaching faculty of computer science department. Special thanks to my guide for allowing me to carry out the project.

REFERENCES

[1] B.chen, R.curtmola, G Ateniese, and R. Burns, "Remote Data Checking for Network Coding Based Distributed Storage Systems," Proc .ACM Workshop Cloud Computing Security Workshop (CCSW '10), 2010.

[2] A. Kermarrec, N.L. Scouarnec, and G.Straub, "Repairing Multiple Failure with Coordinated and adaptive regenerating

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 10, April 2015

Codes," Proc. Int'l Symp. Network Coding (NetCod'11), June 2011.

[3] Amazon Web Services, "Amazon S3 Availability Event: July 20 2008," http://status.aws.amazon.com/s3-20080720.html, July 2008.

[4] B. Treynor, "Gmail Back Soon for Everyone," Event: 2013. http://gmailblog.blogspot.com/2011/02/gmail-back-soon-for-everyone.html.

[5] M. Vulolic, "The Byzantine Empire in the Inter Cloud," ACM SIGACT News, vol. 41, pp.105-111, Sept 2010.

[6] E. Naone, "Are We Safeguarding Social Data?" http://www.technologyreview.com/blog/editor/22924/, Feb.2009.

[7] K. D Bowers, A. Juels, and A. Opera," HAIL: A High Availability And Integrity Layer for Cloud Storage," Proc. 16 ACM Conf. Computer and Comm. Security (CCS'09) ,2009.

[8] A.G. Dimakis, P.B. Godfrey, Y. Wu, M.Wainwright, and K.Ramchandran, "Networking Coding for Distributed Storage Systems," IEEE Tans. Information Theory, vol.56, no.9, pp.4539-4551, Sept.2010.

[9] H. Blodget, "Amazon's Cloud Crash Disaster Permanently," www.businessinsider.com/amazon-lost-data 2011-4/, Apr.2011.

[10] C. Preimesberger, "Many Data Centers Unprepared for Disasters-Industry Grou ,"www.eweek.com/c/a/ITManagment/Many Data Centers Unprepared-for-Disasters-Industry Group -772367/, Mar.2011.

[11] N. kolakowski," Microsoft's Cloud Azure Service Suffers Out-Age," http://eweekeurope.co.uk/news-solutionapplications/microsoft-cloud-azure-service-suffers-outage-395,2013.

[12] M. Mayer," This Site May Harm Your Computer on Every SearchResult,"http://googleblog.blogspot.com/2009/01/thissite-may-harm-your-computer-on.html,2013.

[13] I. Reed and G. Solomon, "Polynomial Codes Over Certain Finite fields," j. the Soc. Industrial and Applied Math.,vol. 8,no.2,pp.300-304,1960.

AUTHOR'S PROFILE

Sharavana .K presently working as an Associate Professor in the department of Computer Science & Engineering of MVJ College of Engineering Bangalore. He received his B.E and M.Tech degree from Visvesvaraya Technology University and Sathyabama University in CSE. Life Member of ISTE. His research interests are Cloud Computing, Adhoc Networks, and Next generation protocol in the Linux Standard.

Neha Sharma completed B.Tech (IT) from Uttar Pradesh Technical University and pursuing M.Tech (CSE) in MVJ College of Engineering Bangalore. Her main research interests are Cloud Computing, Automata.