A Network Coding in Proxy based Storage System for Multiple Clouds

Asha. C M
Dept. of CSE,
RRCE,
Bangalore, India

Gayathri. K
Dept. of CSE
RRCE,
Bangalore, India

Arpitha. M S
Dept. of CSE
RRCE,
Bangalore, India

T. Auntin Jose
Associate Professor,
Dept. of CSE, RRCE,
Bangalore, India

***Abstract* — To avoid the failure for cloud storage, studies propose to stripe data supply across multiple cloud vendors. If a cloud server is failed permanently andloses all its data, we need to provide the lost data with the help of the other surviving clouds to preserve original data. So,here we a present a proxy-based storage system for fault-tolerance in multiple-cloud storage called NC Cloud, which achieves relative cost for repairing permanent single-cloud damaged. NC Cloud is built on top of a network-coding-based storage scheme called the functional minimum-storage regenerating codes, which maintain the same fault tolerance and data redundancy, but use less repair traffic and hence,less monetary cost because it charge only for out bound data. One key characteristics of FMSR is to free encryption operation during the failure of the cloud, whileprotect the benefits of network coding in repair.**

I. INTRODUCTION

CLOUD storage provides allows user to store data in cloud and provide access permission. It stores data in multiple clouds and splits file into different chunks .It provide fast accessing data stored in cloud. However, by using multiple clouds, we can minimize access time end user and improve the fault tolerance of cloud storage.When a cloud damaged permanently, it is necessary to maintain original data and then make stored available to end user. A repair operation retrieves data from existing surviving clouds over the network and reconstructs the lost data and stored in a new cloud. It is important to minimize cost during the repair ofclouds and hence, the monetary cost is lessdue to data migration.

To minimize repair traffic, regenerating codes have been proposed for storing the replicated data in a distributed cloudstorage system.Regenerating codes are built on the concept of network coding, in the sense that nodes perform encoding operations and send encoded data. During the repair operation, each surviving node encodes its stored data chunks and sends the encoded chunks to a new node.One key challenge for deploying regenerating codes is that most existing regenerating codes require storage nodes to be equipped with computation capabilities for performing encoding operation during repair.

In this paper, we present the design and implementation of NC Cloud, a proxy-based storage system designed for providing fault-tolerant storage over multiple cloud storage providers.NC Cloud can interconnect different clouds and transparently stripe data across the clouds. On top of NC Cloud, we propose the first implementable design for the functional minimum-storage regenerating (FMSR) codes.

Few contributions are summarized as follows:

- We present a design of FMSR codes,used to produce lost data. We show that in multiple-cloud storage, FMSR codes can minimize the cost during repair by 25 percent compared to RAID-6 codes when four storage nodes are used, and up to 50 percent as the number of storage nodes further increases. Simultaneously, FMSR codes maintain the same amount of storage overhead as RAID-6 codes but access speed of data is maximum.
- In particular paper, we propose a two-phase scheme, which ensures double-fault tolerance for stripe data across multiple cloud failure and recover of lost data is maintained in the current and next round of repair. By performing two-phase scheme, we ensure that double-fault tolerance is maintained after iterative rounds of repair of node damaged. We perform the simulations to validate the performance of two-phase scheme.
- We conduct monetary cost analysis to show that FMSR codes effectively reduce the cost of repair.

II. RELATED WORK

A. *Multiple-cloud storage.*

There are several systems proposed for multiple-cloud storage. HAIL provides integrity and availability guarantees for stored data. RACS uses erasure coding to mitigate vendor lock-ins when switching cloud vendors across multiple clouds. It retrieves data from the cloud that isabout to fail and regenerate that data in the proxy server, then moves that data to the new cloud.Unlike RACS, NC Cloud excludes the failed cloud in repair. Vukoli_c advocates using multiple independent clouds to provide Byzantine fault tolerance. DEPSKY addresses Byzantine faulttolerance

by combining encryption and erasure coding for stored data during the repair operation in the cloud server.

B. *Empirical studies on regenerating codes.*

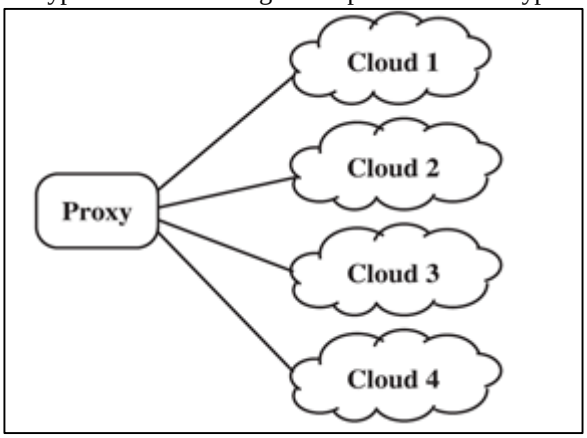
Existing studies on regenerating codes mainly focus on theoretical analysis. Several studies empirically evaluate random linear codes for peer-to-peer storage. The authorspropose reconstruction of codes to minimize the number of surviving nodes to contact during recovery with a tradeoff of incurring a higher storage cost, and evaluate the codes on a cloud storage simulator. The authors alsoevaluate the encoding/decoding function performance regenerating codes.

C. *Follow-up studies on FMSR codes.*

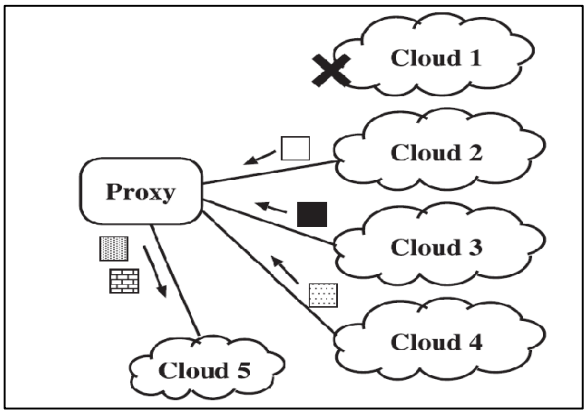
We extend NCCloud to support integrity checking of FMSR-codeddata against Byzantine attacks. We also theoreticallyprove that our two-phase scheme can preserve the MDS property of the stored data after iterative repairs. Inthis work, we focus on the practical deployment of regenerating codes for producing lost data.

III. IMPLEMENTATION

We implement NC Cloud, a proxy based storage system that acts as bridges between the user applications and multiple clouds. The file system layer presents NCCloud as mounted drive, which can then be easily interfaced with general user applications. The coding layer deals with the encrypt the data during the upload and decrypt during



(a) Normal operation



(b) Repair operation

Fig: 1.Proxy-based design for multiple-cloud storage: (a) normal operation and (b) repair operation when cloud node 1 fails. During repair, the proxy server regenerates the lost data and send to the new cloud.

download functions. The storage layers deals with reading/writing requests which different of other surviving clouds. Each file is associated with a metadata object that holds the file details and coding operation, which is replicated in proxy server. The coding layer implements both traditional RAID-6 and FMSR code techniques. FMSR codes generate multiple chunks to be stored in the cloud.Multiple chunks destined for the same repository are then combined before uploading the data to the cloud.We make use of NC Cloud; propose a proxy-based storage system design thatinterconnects multiple cloud repositories, as shown in Fig: 1a. The proxy serves as an interface or an intermediate between client applications and the clouds. If the cloud is permanently failed, the proxy analyses about the failure of cloud and issues for the repair operation as shown in Fig: 1b. The proxy reads the required data pieces from other existing clouds, reconstructs the lost data pieces in proxy server where replication is also done, and writes these new pieces of regenerated data to a new cloud. This repair operation does not involve direct interactions among the clouds during the failure where user does not know about the failure of cloud. This paper considers a cloud-of-clouds setting with two levels of reliability: fault tolerance to stripe data across multiple clouds and recovery to regenerate the lost data from the clouds. First, we assume that the multiple-cloud storage is mainly used to provide double-fault tolerant. Second, we consider single-fault recovery in multiple-cloud storage, that allows a permanent cloud failure is less frequent but possible. Our primary objective is to reduce the cost of storage node repair occurred due to the migration of data over the clouds for a permanent single-cloud failure. In this paper, we focus on comparing two codes: traditional RAID-6 codes and our FMSR codes for double-fault tolerance.We define the repair traffic as the amount of retrieval of data being downloaded from the other surviving clouds during the single-cloud failure recovery. We seek to minimize the repair traffic for cost-effective repair. Here, we do not consider the cost for inbound data stored in the cloud (i.e., the data beinguploaded and stored to a cloud), as it is free of charge for many cloud providers.

IV. CONCLUSION

The proposed work presents NC Cloud, a proxy-based storage system, multiple-cloud storage system that practically addresses the reliability and availability of cloud for on- demand backup storage. NC Cloud not only provides fault tolerance in storage, but also allows reducing the cost during repair when cloud fails permanently during attack of the hacker in cloud servers. NC Cloud implements a practical version of the FMSR codes, which reconstructs lost chunks and sends to the new cloud. Our FMSR code eliminates the encoding requirement of storage nodes (or cloud) during repair, while ensuring that the new

set of data chunks stored after each round of repair operation preserves the required fault tolerance. NC Cloud prototype shows the effectiveness of FMSR codes in the cloud on demand remote backup usage and storage in terms of monetary costs and time taken to response. Here cost is charged for outbound data that are retrieved from the cloud and not for storing the data in cloud.

V. REFERENCES

[1] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: ACase for Cloud Storage Diversity," Proc. ACM First ACM Symp.Cloud Computing (SoCC '10), 2010.

[2] K.D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availabilityand Integrity Layer for Cloud Storage," Proc. 16th ACM Conf.Computer and Comm. Security (CCS '09), 2009.

[3] B.Calder et al., "Windows Azure Storage: A Highly AvailableCloud Storage Service with Strong Consistency," Proc. 23rd ACMSymp. Operating Systems Principles (SOSP '11), 2011.

[4] B. Chen, R. Curtmola, G. Ateniese, and R. Burns, "Remote DataChecking for Network Coding-Based Distributed Storage Systems,"Proc. ACM Workshop Cloud Computing Security Workshop(CCSW '10), 2010.

[5] H.C.H. Chen and P.P.C. Lee, "Enabling Data Integrity Protectionin Regenerating-Coding-Based Cloud Storage," Proc. IEEE 31stInt'l Symp. Reliable Distributed Systems (SRDS '12), 2012.

[6] A.G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A Survey onNetwork Codes for Distributed Storage," Proc. IEEE, vol. 99, no. 3,pp. 476-489, Mar. 2011.

[7] A.Duminuco and E. Biersack, "A Practical Study of RegeneratingCodes for Peer-to-Peer Backup Systems," Proc. IEEE Int'l Conf.Distributed Computing Systems (ICDCS '09), 2009.

[8] Y. Hu, H.C.H. Chen, P.P.C. Lee, and Y. Tang,"NCCloud:Applying Network Coding for the Storage Repair in a Cloud-of-Clouds," Proc. 10th USENIX Conf. File and Storage Technologies(FAST), 2012.

[9] Y. Hu, Y. Xu, X. Wang, C. Zhan, and P. Li, "Cooperative Recoveryof Distributed Storage Systems from Multiple Losses withNetwork Coding," IEEE J. Selected Areas in Comm., vol. 28, no. 2,pp. 268-276, Feb. 2010.

[10] Y. Hu, C.-M.Yu, Y.-K. Li, P.P.C. Lee, and J.C.S. Lui, "NCFS: Onthe Practicality and Extensibility of a Network-Coding-BasedDistributed File System," Proc. Int'l Symp. Network Coding(NetCod '11), 2011.

[11] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li,and S. Yekhanin, "Erasure Coding in Windows Azure Storage,"Proc. USENIX Conf. Ann. Technical Conf. (ATC '12), 2012.