Chapter-1

INTRODUCTION

Cloud storage provides an on-demand remote backup solution. However, using a single-cloud storage provider raises concerns such as having a single point of failure and vendor lock-ins. As suggested in this system, a plausible solution is to stripe data across different cloud providers. By exploiting the diversity of multiple clouds, the fault tolerance of cloud storage can be improved. While striping data with conventional erasure, codes performs well when some clouds experience short-term transient failures or foreseeable permanent failures, there are real-life cases showing that permanent failures do occur and are not always foreseeable.

This work focuses on unexpected permanent cloud failures. When a cloud fails permanently, it is necessary to activate repair to maintain data redundancy and fault tolerance. A repair operation retrieves data from existing surviving clouds over the network and reconstructs the lost data in a new cloud. Today's cloud storage providers charge users for outbound data (see the pricing models in this system), so moving an enormous amount of data across clouds can introduce significant monetary costs. It is important to reduce the repair traffic (i.e., the amount of data being transferred over the network during repair), and hence, the monetary cost due to data migration.

To minimize repair traffic, regenerating codes have been proposed for storing data redundantly in a distributed storage system (a collection of interconnected storage nodes). Each node could refer to a simple storage device, a storage site, or a cloud storage provider. Regenerating codes are built on the concept of network coding, in the sense that nodes perform encoding operations and send encoded data. During repair, each surviving node encodes its stored data chunks and sends the encoded chunks to a new node, which then regenerates the lost data. It is shown that regenerating codes require less repair traffic than traditional erasure codes with the same fault-tolerance level. Regenerating codes have been extensively studied in the theoretical context. The important feature in this scheme is data migration, which stores the uploaded data between cloud servers, for retrieval of data if it is lost or corrupted.

However, the practical performance of regenerating codes remains uncertain. One key challenge for deploying regenerating codes in practice is that most existing regenerating codes require storage nodes to be equipped with computation capabilities for performing encoding operations during repair. On the other hand, to make regenerating codes portable to any cloud storage service, it is desirable to assume only a thin-cloud interface, where storage nodes only need to support the standard read/write functionalities. This motivates us to explore, from an applied perspective, how to practically deploy regenerating codes in multiple-cloud storage, if only the thin-cloud interface is assumed.

A proxy-based storage system is designed for providing fault-tolerant storage over multiple cloud storage providers. NC can interconnect different clouds and transparently stripe data across the clouds. On top of NC, we propose the first implementable design for the regenerating code.

The main disadvantage associated with the existing system is that:

1. Storage Repair
2. Fault tolerance.

The drawback of the existing system can be overcome using this proposed system. The main contribution of this paper is that Proposed system to identify the most representative fault tolerance and storage repair. There are several systems proposed for multiple-cloud storage. HAIL provides integrity and availability guarantees for stored data. DESPKY uses erasure coding to mitigate vendor lock-ins when switching cloud vendors. It retrieves data from the cloud that is about to fail and moves the data to the new cloud. Several studies propose efficient single node failure recovery schemes that minimize the amount of data read (or I/Os) for XOR-based erasure codes. For example, the authors of propose optimal recovery for specific RAID-6 codes and reduce the amount of data read by up to around 25 percent (compared to conventional repair that downloads the amount of original data) for any number of nodes. Note that our SR codes can achieve 25 percent saving when the number of nodes is four, and up to 50 percent saving if the number of nodes increases. NC can interconnect different clouds and transparently stripe data across the clouds, and has the same storage cost as in traditional erasure coding schemes based on RAID-6 codes, but uses less repair traffic when recovering a single-cloud failure. The main advantage of the proposed system is that:

1. Fault tolerance among clouds.

2. Data backup and recovery.
3. Regenerating codes.
4. Iterative Repairs.
5. Repair operations among the cloud

Cloud computing provides software, storage, computation, data access, resources without requiring cloud users to know the location and other details of the computing infrastructure. Cloud computing is the delivery of a service rather than a product where by software, information, shared resources are provided to a computer and other devices as a metered services over a network. Cloud based applications are accessed by the end users through a web browsers or mobile applications or desktop, while the business data and software are stored at a remote location on servers. Cloud application providers strive to give the better services and performances as the local end-user computers.

Figure 1.1 Cloud Architecture