Chapter 36
Data Backup and Recovery With a Minimum Replica Plan in a Multi-Cloud Environment

Mohammad M. Alshammari
International Islamic University Malaysia, Malaysia

Ali A. Alwan
https://orcid.org/0000-0003-3279-9366
International Islamic University Malaysia, Malaysia

Azlin Nordin
International Islamic University Malaysia, Malaysia

Abedallah Zaid Abualkishik
American University in the Emirates, UAE

ABSTRACT

Cloud computing has become a desirable choice to store and share large amounts of data among several users. The two main concerns with cloud storage are data recovery and cost of storage. This article discusses the issue of data recovery in case of a disaster in a multi-cloud environment. This research proposes a preventive approach for data backup and recovery aiming at minimizing the number of replicas and ensuring high data reliability during disasters. This approach named Preventive Disaster Recovery Plan with Minimum Replica (PDRPMR) aims at reducing the number of replications in the cloud without compromising the data reliability. PDRPMR means preventive action checking of the availability of replicas and monitoring of denial of service attacks to maintain data reliability. Several experiments were conducted to evaluate the effectiveness of PDRPMR and the results demonstrated that the storage space used one-third to two-thirds compared to typical 3-replicas replication strategies.

DOI: 10.4018/978-1-7998-8954-0.ch036

1. INTRODUCTION

Cloud computing delivers numerous benefits including reduced costs for data storage backup and data accessibility (Li et al., 2012; Li et al., 2015; Li et al., 2016; Attiya & Zhang, 2017). The essential cloud characteristic is its ability to store data while ensuring its availability, which is an important feature when storing sensitive information. Despite the success of cloud computing and its immediately observable benefits, however, the rapid development of the scale and complexity of today's cloud services and infrastructures has also revealed important challenges regarding the design of fundamental cloud computing architectures, specifically concerning high data reliability requirements and storage costs.

Many surveys conducted over recent years have shown that enterprises and critical business organizations are moving from the single-cloud to the multi-cloud (Tebaa et al., 2014; Sengupta et al., 2014; Liu & Shen, 2017; Alshammari et al., 2017; Alshammari et al., 2018). Moreover, using a minimum of two clouds (or more) is a way to reduce the risk of failure with regard to service availability, data loss, and compromised privacy. Also, using multiple clouds simultaneously can reduce the risk when using a public cloud for applications and data (Sengupta & Annervaz, 2012; Prazeres & Lopes, 2013; Lenk & Tai, 2014; Alhazmi, 2016; Sabbaghi et al., 2017; Liu & Shen, 2017; Choo & Chung 2018). Keeping data in one single-cloud environment may not be prudent as any damage to the datacenters of this single-cloud in the case of disaster will result in permanent data loss (Tebaa et al., 2014; Sengupta et al., 2014; Gu et al., 2014; Alshammari et al., 2017; Alshammari et al., 2018). Other solutions for developing a data backup and recovery plan involve multi-cloud providers in which multiple data replicas are generated for several remote CPs (Sengupta et al., 2014; Gu et al., 2014; Sulochana & Dubey, 2015). These approaches guarantee high data reliability and minimize the risk of data loss in case of disasters, thereby ensuring that user data are recoverable in the event of catastrophic failure. Most proposed solutions assume that the data should be replicated into at least three copies (three replicas) to ensure high reliability (Lei et al., 2007; Li et al., 2012; Gu et al., 2014; Li et al., 2015; Li et al., 2016; Du et al., 2017). These copies may be stored in one location or distributed over multiple locations. However, these solutions incur high storage costs and consume a significant amount of storage space, which leads to high network traffic, particularly for data-intensive applications in the cloud.

This article presents a low-cost strategy named Preventive Disaster Recovery Plan with Minimum Replica (PDRPMR) to manage data recovery in multi-cloud context before the disaster. The strategy attempts to manage the data recovery process with the lowest possible number of replicas maintaining low-cost storage without affecting the data reliability requirements. Furthermore, we use different cloud scheduler strategies. The PDRPMR strategy is inspired by the PRCR mechanism proposed in (Li et al., 2016). PRCR has been developed to work in the single cloud environment. Nevertheless, PDRPMR has been developed to determine the minimum number of replicas with the intention of reducing the storage space, cost of storage and Recovery Time Objective (RTO) in the multi-cloud environment. To facilitate the replication plan in determining the required number of replicas, we request the user to specify (1) the storage duration and (2) the importance of data. The expected storage duration is either short-term or long-term. For short-term data, we identify the short-term to be for several days (up to one week), storage in a single replica is enough to ensure data reliability. Long-term data (more than a week) may require a higher level of data reliability than the reliability assurance provided by one replica, thus two replicas are stored and periodically and proactively checked. It should be noted that there is no common consent that short-term and long-term has any specific duration. The duration of the short and long terms could be determined by the policy and the legal requirements of the company set by the policy