Journal of Network and Computer Applications 43 (2014) 121–141

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

A review on remote data auditing in single cloud server: Taxonomy and open issues

Mehdi Sookhak [a,*], Hamid Talebian [a], Ejaz Ahmed [a], Abdullah Gani [a], Muhammad Khurram Khan [b]

[a] Mobile Cloud Computing Research Lab, Faculty of Computer Science and IT, University of Malaya, Kuala Lumpur, Malaysia
[b] Center of Excellence in Information Assurance (CoEIA), King Saud University, Saudi Arabia

ARTICLE INFO

Article history:
Received 25 October 2013
Received in revised form
27 January 2014
Accepted 25 April 2014
Available online 6 May 2014

Keywords:
Cloud computing
Remote data auditing
Provable data possession
Proof of retrievability
Proof of ownership

ABSTRACT

Cloud computing has emerged as a computational paradigm and an alternative to the conventional computing with the aim of providing reliable, resilient infrastructure, and with high quality of services for cloud users in both academic and business environments. However, the outsourced data in the cloud and the computation results are not always trustworthy because of the lack of physical possession and control over the data for data owners as a result of using to virtualization, replication and migration techniques. Since that the security protection the threats to outsourced data have become a very challenging and potentially formidable task in cloud computing, many researchers have focused on ameliorating this problem and enabling public auditability for cloud data storage security using remote data auditing (RDA) techniques. This paper presents a comprehensive survey on the remote data storage auditing in single cloud server domain and presents taxonomy of RDA approaches. The objective of this paper is to highlight issues and challenges to current RDA protocols in the cloud and the mobile cloud computing. We discuss the thematic taxonomy of RDA based on significant parameters such as security requirements, security metrics, security level, auditing mode, and update mode. The state-of-the-art RDA approaches that have not received much coverage in the literature are also critically analyzed and classified into three groups of provable data possession, proof of retrievability, and proof of ownership to present a taxonomy. It also investigates similarities and differences in such framework and discusses open research issues as the future directions in RDA research.

Contents

** Corresponding author.
E-mail addresses: m.sookhak@ieee.org (M. Sookhak), talebian@siswa.um.edu.my (H. Talebian), ejazahmed@ieee.org (E. Ahmed), abdullah@um.edu.my (A. Gani), mkhurram@ksu.edu.sa (M.K. Khan).

http://dx.doi.org/10.1016/j.jnca.2014.04.011
1084-8045/

1. Introduction

Cloud computing is a new model of computing in contrast to conventional desktop computing. Today's, this new paradigm became popular and received increasing attention by researchers (academia) and industry. According to The National Institute of Standards and Technology (NIST) Cloud Computing is

> "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (network, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort*" (Mell and Grance, 2011).

This technology allows users to outsource their data to a remote server operated by a third party called cloud service provider (CSP) (Zhibin and Dijiang, 2012). In addition, computing resources such as memory, disk storage, processor, and bandwidth are virtualized and clients are able to access them using the Internet (Kumar and Yung-Hsiang, 2010). The term cloud refers to a thousand of virtualized servers distributed over a set of data centers with different geographical locations connected together through telecommunication links. The services on the cloud are delivered to the users as pay-as-you-go pricing model. This means users are only charged for the amount of service they have used similar to water and electricity bills.

Adopting cloud computing offers various advantages to both end users and CSP. For end users the advantages include rapid elasticity, measured service, minimal capital investment, lower maintenance cost, and location-independent access to the services (Kumar and Yung-Hsiang, 2010; Wang et al., 2010). On the other hand, CSP achieves a higher level of resource utilization and thus saves energy consumption.

Despite several benefits, some security concerns inhibit users to fully adopt this new technology and shift from traditional computing to cloud computing (Zhibin and Dijiang, 2012). By storing data to a remote server, user loses his physical control over data and instead delegates management of data to an un-trusted party (Cong et al., 2010; Wei et al., 2013). Even though cloud resources are very powerful and reliable comparing to that of client, the data on the cloud is still vulnerable to many threats from inside or outside the cloud (Wang et al., 2010). These threats might compromise confidentiality, integrity, and availability of data. An unfaithful provider might delete less frequently accessed data to free up disk space or hide data loss to protect his reputation (Yang and Jia, 2012a). In addition, security attacks, Byzantine failure, server failure and power outage are likely to happen. Amazon S3 breakdown (Team, 2008), Gmail email mass deletion (Arrington, 2006), Sidekick Cloud Disaster (Cellan-Jones, 2009) Breakdown of Amazon EC2 2010 (Miller, 2010) are example of such events.

Cloud users need to make sure their data remain intact after uploading to the remote server. Traditional integrity checking techniques such as hash functions and signatures require a local copy of the entire data. Unfortunately, these techniques are not well suited for the cloud environment because downloading possibly large files is impractical due to its high communication cost. This even becomes worse in case of mobile computing devices with limited power, storage capacity, and connectivity. As a result, devising a proper audit service which can remotely check the integrity of outsourced data in the cloud is deemed as a crucial need.

Remote data auditing (RDA) refers to a group of protocols to securely, frequently, and efficiently verify the correctness of the data over a cloud managed by untrustworthy provider without having to retrieve the data (Ateniese et al., 2008). The RDA protocols are able to check a small fraction of entire data, called spot checking, and give a probabilistic guarantee for the data integrity. To design a remote data audit mechanism the following important criteria must be taken into account: (1) Efficiency: audit the data with the minimum computational cost over the server and particular client. The auditing service is also reasonable for the communication overhead between client and server, (2) Public verifiability: delegate the audit task to a trusted third party auditor rather than a client in order to reduce the computation cost over the client, (3) Frequency: number of times that user is able to verify the integrity of outsourced data by generating a challenge message, (4) Probability of detection: probability by which a protocol detects data corruption, (5) Recovery: ability to recover data in case of data corruption, and (6) Dynamic update: enabling the cloud user to update the outsourced data by using insert, delete, modify, and append operation without requiring to download the whole data.

This paper reviews the state-of-the-art remote data auditing efforts that are used to check the integrity of outsourced data in a single cloud server. We study and classify the characteristics of remote data auditing approaches by thematic taxonomy into five groups, namely security requirements, security objective, performance metrics, auditing mode, and update mode. The impacts of RDA in cloud and mobile cloud computing are also presented. The main contribution of the paper is to review the state-of-the-art RDA methods, categorize current RDA mechanisms into three classes of provable data possession, proof of retrievability, and proof of ownership based on the implications, requirements, and critical characteristics. To the best of our knowledge, this is the first effort to categorize data storage strategies applied in single cloud computing. Furthermore, we identify the issues in existing solutions for data auditing and challenges to cloud based application processing and mobile device limitations. This paper lists some challenges and open issues to guide researchers to choose the appropriate domain for future research and acquire ideas for further investigations.

The rest of the paper is organized as follows. Section 2 presents the fundamental concepts of cloud computing and mobile cloud computing. It also explains data auditing and its requirements.