Computers and Electrical Engineering 000 (2016) 1–14

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Cellular automata based secure distributed storage scheme with integrity proof[☆]

Yousheng Zhou[a,b], Feng Wang[c], Fei Tang[a,*], Xiaojun Wang[d]

[a] College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[b] School of Computer Science, Chongqing University, Chongqing 400044, China
[c] College of mathematical sciences, De Zhou University, Shandong 253023, China
[d] School of Electronic Engineering, Dublin City University, Dublin, Ireland

ARTICLE INFO

Article history:
Received 9 September 2015
Revised 16 September 2016
Accepted 2 November 2016
Available online xxx

Keywords:
Cellular automata
Threshold storage
Integrity proof
Cloud security

ABSTRACT

The cloud storage service, widely used in daily life due to its convenience, can sometimes suffer from availability and confidentiality problems. For instance, data in cloud storage can be damaged by hardware failures or malicious destruction, or even be exposed to unauthorized parties, and this poses a big risk for user data stored in cloud. To overcome these problems and challenges, a Cellular Automata based secure Distributed storage scheme with Integrity Proof, termed CAD-IP, is proposed in this paper. CAD-IP leverages threshold based storage service to provide robustness and confidentiality of user private data. Homomorphic hashing is integrated into the proposed scheme, which facilitates the verifier to check the integrity of the data on servers. A sampling strategy that greatly reduces the computation and communication cost is adapted in the proposed scheme. Analysis demonstrates that the proposed scheme can not only achieve perfectness, confidentiality and unforgeability, but also enable the verifier to effectively detect any modification or deletion of their file shares. Meanwhile, the proposed CAD-IP scheme supports dynamic update over the stored file shares without downloading and re-uploading the entire file shares.

1. Introduction

Cloud computing is seen as the next wave of information technology for individuals and organizations, which treats computing as a service rather than a product, enabling users to access and share a wide variety of applications, data, and resources through an universal interface [1]. This new economic computing model is commonly referred to as cloud computing and includes various types of services such as: infrastructure as a service (IaaS), where a customer makes use of a service providers computing, storage or networking infrastructure; platform as a service (PaaS), where a customer leverages the providers resources to run custom applications; and finally software as a service (SaaS), where customers use software that is run on the providers' infrastructure [1]. Cloud computing has been widely applied into many fields, such as e-health wireless sensor networks [2], the management for mobile devices and so on [3]. Cloud computing has experienced exponen-

[☆] Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Z. Wang.
[*] Corresponding author.
E-mail address: tangfei@cqupt.edu.cn (F. Tang).

http://dx.doi.org/10.1016/j.compeleceng.2016.11.004
0045-7906/© 2016 Elsevier Ltd. All rights reserved.

tial growth over the last few years due to its convenience and pay-as-you-go charging model, and the growth is expected to increase over the next few years worldwide [4].

Cloud computing brings incredible convenience for both individuals and organizations to store their data in cloud, and users can access their data freely via some interface, for instance a browser, at any time anywhere, while they need not to care about the operating, monitoring and maintenance behind the data [5,6]. While the benefits of using a cloud infrastructure are clear, it introduces significant security and privacy risks. In most occasions, however, data are under the full control of the cloud service providers once users outsource their storage to the cloud. In reality, the services sometimes inevitably suffer from confidentiality, availability and robustness problems, and this renders customers cannot use them for reasons, such as service down, network connection banned, or malicious attack [7–9], which greatly reduces the availability of the cloud services and becomes one of the hurdles hindering wider adaptation of cloud computing. Although cloud storage has enormous promises, unless the issues of confidentiality and reliability are properly addressed many potential customers will be reluctant to make the move [10–12]. How to prevent private data in cloud accessed by unauthorized parties is a challenge for any cloud storage providers. Access control seems to be a suitable solution to this problem[13], however, data stored in cloud may be shared by multiple users and the cloud servers are not always trusted. Merely using access control is not effective, and cryptographic methods can be used to ensure confidentiality [14–16]. Wang et al. [17] proposed a hierarchical attribute based encryption scheme for cloud storage, Zhou et al. [18] proposed a role-based encryption (RBE) scheme that allows Role-Based-Access-Control (RBAC) policies to be enforced for the encrypted data stored in public clouds. Users who violate the rules of the shared data should be revoked by the data owner, re-encryption can be adopted to address it [19–22]. Do et al. [20] proposed a proxy re-encryption scheme to resolve it by dividing a data file into header and body. However, re-encryption commands can be intercepted when they are transferred over an untrusted network, Liu et al. [21,22] solved this problem by proposing a time-based re-encryption scheme, which enables the cloud servers to automatically re-encrypt data based on their internal clocks. Although the confidentiality of the data can be ensured using encryption, the user has to download the entire data from the server if the user wants to search the data. To address this problem, research has been conducted on keyword search over encrypted data [23,24,26]. With the construction of a special tree-based index structure, Xia et al. [23] presented a multi-keyword ranked search scheme over encrypted cloud data, which supports dynamic update operations. Fu et al. [24] proposed a searchable encryption scheme which supports both multi-keyword ranked search and parallel search based on Vector Space Model (VSM). In addition, Fu et al. [25,26] proposed some improved schemes based on the semantic relationship between concepts and the uni-gram to improve efficiency and accuracy in multi-keyword search.

The efforts stated above only focus on confidentiality and availability of stored data in cloud, however, robustness is not provided. Lin et al. [28] considered the problem of robustness and confidentiality with erasure code-based method, however, their scheme cannot provide integrity proof. Data loss can happen occasionally due to either natural reasons or malicious attacks. Though, backup technology is adopted by cloud service providers, the confidentiality cannot be ensured, since all the backups are the same and encrypted by the same private key, once the only secret key used for encryption is leaked or destroyed, the confidentiality of all the backups cannot be assured anymore. Provable data possession (PDP) [39] can efficiently address the problem of integrity, and it allows a user who has stored data to an untrusted server to verify whether the server possesses the original data without retrieving it. Many efforts on data integrity have been made in recent years, such as Zhu et al. [37] presented a cooperative PDP scheme based on homomorphic verifiable response and hash index hierarchy, which was claimed to have the properties of completeness, knowledge soundness, and zero-knowledge. However, Wang et al. [35] pointed out the scheme in [37] fails to provide knowledge soundness and it is vulnerable to cheating attacks. To facilitate the PDP in multicloud storage, Wang et al. [38] proposed an efficient identity-based distributed provable data scheme using the bilinear pairings. In order to deal with the issue of burdensome certificate management in Public Key Infrastructure(PKI) based cloud storage, Wang et al. [36] constructed a certificate-based remote data integrity checking model, which enables the cloud servers to detect the malicious clients. In addition, some other research focused on the extension of PDP [41–43].

Motivated by the characteristic distributed property of threshold secret sharing [29], a novel distributed storage scheme with integrity proof based on cellular automata [30] for cloud storage, i.e. the CAD-IP scheme, is constructed in this paper, which owns desirable security and performance features as follows,

(1) *Security*. Our CAD-IP scheme can simultaneously provide correctness, correctness, confidentiality, and reliability. Due to easy hardware implementation and its pseudorandom behavior, cellular automata have been widely used in cryptography [34,44–49]. Although there exist some research [34,44,45] on cellular automata based secret sharing, these schemes are not the desired solutions for distributed cloud storage since they cannot provide integrity proof. In our construction, threshold storage of the file ensures confidentiality and robustness, and it tolerates Byzantine failures [31], where a storage server may fail in arbitrary ways.

(2) *Performance*. By utilizing the homomorphic token [32], the presented scheme achieves dynamic operation on random sampled blocks, not all data blocks, to provide validation of integrity, which means only several bytes of hashing value would be transferred over the communication channel, and the client needs less storage to complete validation compared to previous threshold based storage schemes. Furthermore, no additional encryption/decryption or encoding/decoding operations when to provide confidentiality in our scheme. Above all, computation and communication cost are reduced significantly in our proposed CAD-IP.