Applied Mathematics and Computation 314 (2017) 31–43

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Privacy-preserving certificateless provable data possession scheme for big data storage on cloud

Debiao He[a,b], Neeraj Kumar[c], Huaqun Wang[d], Lina Wang[e], Kim-Kwang Raymond Choo[f,g,*]

[a] Co-Innovation Center for Information Supply & Assurance Technology, Anhui University, Hefei, China
[b] State Key Lab of Software Engineering, Computer School, Wuhan University, Wuhan, China
[c] Department of Computer Science and Engineering, Thapar University, Patiala, India
[d] Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, School of Computer School, Nangjing University of Posts and Telecommunications, Nanjing, China
[e] Key Laboratory of Aerospace Information Security and Trusted Computing Ministry of Education, Computer School, Wuhan University, Wuhan, China
[f] Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78258, USA
[g] School of Information Technology & Mathematical Sciences, University of South Australia, Adelaide, SA 5095, Australia

ARTICLE INFO

Keywords:
Public cloud storage
Cloud data integrity
Privacy-preserving
Provable data possession

ABSTRACT

Cloud storage is generally regarded as one of the most promising technologies to address various big data challenges (e.g. secure storage for big data), due to the capability to provide scalability and functional diversity. However, how to efficiently audit the integrity of outsourced data remains a research challenge. Provable data possession (PDP) scheme can potentially be used to verify the integrity of outsourced data without downloading such data. However, existing PDP schemes suffer from either certificate management or key escrow problems. A number of certificateless PDP (CLPDP) schemes for the public cloud storage have been designed to address the above problems. However, most of them do not offer privacy protection from the verifier (i.e. verifier could obtain the data stored in the cloud when verifying their integrity). In this paper, we propose a privacy-preserving CLPDP (PP-CLPDP) scheme to address certificate management and key escrow problems, as well as ensuring privacy protection. We also prove the security and evaluate the performance of our proposed PP-CLPDP scheme.

1. Introduction

The increasing popularity and widespread adoption of digital devices (e.g. mobile devices and Internet-of-Things devices) and communication technologies (e.g. social networks) have resulted in a significant amount of data captured, stored and disseminated in electronic only form (also known as big data). For example, IDC and EMC predicted that the amount of data will increase from 2837 exabytes (EB) in 2012 to 40,000 EB in 2020 [1].

One big data related challenge is secure storage [2]. Traditional data storage approach where data are stored on user managed hardware is no longer viable. For example, it was estimated that the cost of managing data storage system

[*] Corresponding author at: Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78258, USA.
E-mail address: raymond.choo@fulbrightmail.org (K.-K.R. Choo).

http://dx.doi.org/10.1016/j.amc.2017.07.008
0096-3003/© 2017 Elsevier Inc. All rights reserved.

is between 5 and 7 times that of buying hardware devices [3], which is a significant cost. This is a challenge faced by private sectors and government agencies. For example, Quick and Choo highlighted the big forensic data challenges faced by international law enforcement and intelligence agencies [4,5].

It is, therefore, unsurprising that cloud storage is considered a viable solution to providing cost-effective secure storage for big data, due to inherent characteristics such as on-demand self-service, and location independent resource polling [6–10]. Examples of popular cloud services include Amazon Simple Storage Service (S3) and Google Cloud.

While public cloud storage provide users with benefits (e.g. convenience and cost savings), there are associated security and privacy challenges such as losing control of data outsourced to the cloud and ensuring the integrity of outsourced data (no unauthorized modification, deletion, etc) [6,11–13]. Despite recent advances and the vested interest by cloud storage services to ensure the security and privacy of outsourced data, public cloud storage providers still suffer from data leakage and other security breaches [14]. Ensuring the integrity of data has also attracted more and more attentions of researchers and remains an ongoing research area, which is the focus of this paper.

1.1. Related literature

Traditional cryptographic methods are generally not suitable for checking the integrity of outsourced data, as it will not be realistic (e.g. costly and may present another security risk) to download data from the cloud server for verification. Thus, provable data possession (PDP) was proposed to facilitate remote integrity verification, where the integrity of the data can be verified without downloading [15].

In the seminal work by Ateniese et al., two PDP schemes based on the large integer factoring problem were proposed [15]. A year later in 2008, Ateniese et al. [16] extended one of their previous schemes to a dynamic setting. However, Ateniese et al.'s scheme does not support the insert operation. Since then, a number of PDP schemes [17–21] have been proposed for different applications. We also remark that schemes [17–21] are designed for a traditional public key cryptography (PKC) setting, where certificate authority generates a certificate for each user which binds his/her identity to the public key. Management of public key certificates has proven to be a harder task than was initially realized, and PDP schemes designed for a traditional PKC setting (e.g. [15–21]) suffers from the same inherent key management issue.

More than three decades ago in 1984, the identity-based (ID-based) cryptography was presented by Shamir [22]. In this approach, the need for certificates is removed since the identity of the owner is the public key. Leveraging benefits afforded by ID-based cryptography, Wang et al. [23] defined a security mode for ID-based provable data possession (ID-based PDP) schemes. They also presented a concrete ID-based PDP scheme and proved the security of the scheme. Wang [24] also presented an ID-based PDP scheme for multi-public cloud storage. A number of ID-based PDP schemes [25–29] have also been presented in recent times, although some of the schemes have the key escrow limitation.

In the certificateless public key cryptography (CL-PKC) approach [30], the user's private key comprises a user selected secret value and the KGC generated partial private key. In the literature, there are a number of certificateless cryptography schemes [31–33]. For example, Wang et al. [34] proposed a certificateless provable data possession (CLPDP) scheme, as well as a robust security model. Despite the scheme having a claimed security proof, Al-Riyami and Paterson [30] revealed that the scheme of Wang et al. is vulnerable against Type I adversary. A Type I adversary can use some randomly chosen value to replace the user's public key. He et al. [35] then presented another CLPDP scheme based on bilinear pairings. However, the CLPDP scheme does not preserve data privacy because the verifier can extract the user's data by solving the system of linear equations.

1.2. Our contributions and paper organization

In this paper, we put forward a CLPDP scheme for public cloud storage, which is partially inspired by Wang et al.'s PDP scheme [23]. Our main contributions are summarized as below.

- We present a security model for CLPDP scheme, which builds on the security model of Huang et al. [36].
- We present a privacy-preserving CLPDP scheme for public cloud storage, based on bilinear pairings. We then prove the security of the scheme in our security model.

In the next section, we present relevant background materials. We will then present our proposed scheme in Section 3 and its security analysis in Section 4. In Section 5, we evaluate the performance of the proposed scheme. Section 6 concludes this paper.

2. Background

2.1. Bilinear pairings

Suppose G_1 and G_2 are two cyclic groups with the order q, where q is a prime. Let e: $G_1 \times G_1 \to G_2$ be a rational function. We say that e is a bilinear map when it meets the following three conditions.

(1) *Bilinear*: there exists $S, T \in G_1$ and $a, b \in Z_q^*$, such that $e(aS, bT) = e(S, T)^{ab}$ holds.