Future Generation Computer Systems 37 (2014) 259–266

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Digital provenance: Enabling secure data forensics in cloud computing

Jin Li [a,*], Xiaofeng Chen [b], Qiong Huang [c], Duncan S. Wong [d]

[a] School of Computer Science, Guangzhou University, Guangzhou, China
[b] State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China
[c] College of Informatics, South China Agricultural University, Guangzhou, China
[d] The Department of Computer Science, City University of Hong Kong, Hong Kong

HIGHLIGHTS

- We propose a practical secure provenance scheme with fine-grained access control.
- A broadcast encryption technique is utilized to decrease the data owner's computational overhead.
- An attribute-based signature is applied to realize efficient anonymous authentication.

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form
10 August 2013
Accepted 4 October 2013
Available online 18 October 2013

Keywords:
Provenance
Cloud computing
Privacy
Attribute-based signature

ABSTRACT

Secure provenance that records the ownership and process history of data objects is vital to the success of data forensics in cloud computing. In this paper, we propose a new secure provenance scheme based on group signature and attribute-based signature techniques. The proposed provenance scheme provides confidentiality on sensitive documents stored in a cloud, unforgeability of the provenance record, anonymous authentication to cloud servers, fine-grained access control on documents, and provenance tracking on disputed documents. Furthermore, it is assumed that the cloud server has huge computation capacity, while users are regarded as devices with low computation capability. Aiming at this, we show how to utilize the cloud server to outsource and decrease the user's computational overhead during the process of provenance. With provable security techniques, we formally demonstrate the security of the proposed scheme under standard assumptions.

1. Introduction

Cloud computing is a promising next-generation computing paradigm which integrates multiple existing and new technologies such as virtualization and distributed computing. It provides unlimited "virtualized" resources to users as services across the Internet while abstracting the details from users. With the emergence of commercial cloud computing platforms such as Amazon's EC2 and S3 [1], Google's App Engine [2], and Microsoft's Azure [3], cloud computing has become more a reality than just a concept [4].

As in any existing application and system, security and privacy play an extremely important role for the success of cloud computing, and certainly raise a lot of challenges among the many others that cloud computing is confronted with. It is hard to imagine that a cloud customer, say a company, would like to store all its sensitive information on cloud computing platforms, e.g., Amazon's S3, and put the security protection of the information at the mercy of the cloud computing operator.

Besides the confidentiality of this sensitive information, the user's identity privacy, a fundamental right to privacy, is also expected in cloud computing. If the access to a cloud discloses a user's real identity, the user could still be unwilling to accept this paradigm. Thus, anonymous authentication [5] is desirable in cloud computing. Although anonymous authentication can provide privacy of a user's identity, it is required to only provide conditional anonymity. For example, when a group of users is authorized to access a document, if some dispute arises in a modification, the real user can be tracked by some designated party.

The provenance systems [6–8] have been developed to record provenance meta-data. Given its provenance, a data object can report who created and who modified its contents. Practical provenance systems use a specialized recording instrument to collect information about data processing at runtime. The instrument annotates data with information on the relevant

* Corresponding author.
E-mail addresses: jin71@gmail.com, jinli71@gmail.com (J. Li), xfchen@xidian.edu.cn (X. Chen), csqhuang@cityu.edu.hk (Q. Huang), duncan@cityu.edu.hk (D.S. Wong).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.10.006

operations performed on it. The ordered collection of provenance annotations becomes an unalterable record of data evolution called a provenance chain. Therefore, once a dispute arises in a document stored in a cloud, provenance is important for data forensics to provide digital evidences for post investigation. Provenance information has a wide range of critical application areas. For example, scientific data processing needs to keep track of data ownership and processing workflow to ensure the trust assigned to the output data. In business environments, provenance of documents is even more critical for regulatory and legal reasons. A company's financial reports are required to contain provenance information on the path the data took during various stages of processing and the principals who performed various actions on it.

Therefore, cloud computing should also provide provenance [9] to record the ownership and process history of data objects in the cloud in order to gain wide acceptance to the public. However, there are many challenges to provenance in cloud computing [8], in which we need protect the security of provenance information, i.e., to not violate the information confidentiality and user privacy in cloud computing. Specifically, these requirements [9] include confidentiality of documents, unforgeability of the provenance record, and conditional anonymity of the user's identity.

Though secure provenance is vital to the success of data forensics in cloud computing, before its deployment in cloud computing, two critical issues have to be addressed, namely, (1) fine-grained access control: when a document is being created, the data owner can specify a fine-grained access control policy for the documents stored remotely in the cloud servers; (2) low computation and communication overhead at the data owner/user side: in cloud computing, the computational ability is not required to be high except for the cloud server. Actually, the devices are always assumed to be devices with low computational capability. Thus, a provenance system with low computation for data owners and users is preferred in cloud computing.

Aiming at this, we propose a practical secure provenance scheme with fine-grained access control based on the bilinear pairing technique in this paper, which can provide trusted evidence for data forensics in cloud computing. Our contribution in this paper is as follows.

(1) The computation and communication overhead for the data owner is low. Compared with the previous work [5], two new techniques are utilized here to decrease the data owner's computational overhead. The first is broadcast encryption, which is used by the cloud server to control the user's access. The other is the attribute-based signature, which is computed by users, instead of data owners, as part of their access requests.
(2) The computational overhead for the data owner/user has been significantly reduced by outsourcing the cryptographic operation of exponentiation in a bilinear group. More specifically, the computation is moved from the data owner/user side to the cloud server by using the following two techniques. The first is to use the two-server model [10] to compute the exponentiation cooperatively. The second is to use the proxy re-signature method [11]. As a result, we significantly reduce the complexity at the user/data owner side with respect to the computation of modular exponentiation from $O(k)$ to $O(1)$ in terms of the number of modular multiplications required [12], where k represents the number of bits of the exponent.

1.1. Organization

The rest of the paper is organized as follows. In Section 2, we present the related work for a secure provenance system. In Section 3, the architecture and the security model for a secure provenance system are given. In Section 4, we show some basic tools which will be used in this paper, which include the attribute-based signature scheme and the group signature scheme. In Section 5, a new and efficient secure provenance scheme is given, as well as its security analysis. We also discuss how to provide fine-grained access control and better efficiency in this section. Finally, we draw our conclusions in Section 6.

2. Related work

Provenance has been studied extensively in archival theory for the purpose of asserting authenticity. In recent years, provenance has also gained importance in digital realms and e-Science [13,14]. However, most schemes require trustworthiness of the server. Provenance systems that do not rely on a trusted server have also been developed [15].

Although provenance of workflow and documents has been studied extensively in the past, very little work has been done on securing the provenance information. To address such security issues in provenance, Hasan et al. [8] first formally defined the security and privacy issues of a provenance system, including confidentiality and user privacy.

Recently, Lu et al. [5] proposed a new secure provenance system to achieve user privacy and message confidentiality. The basic tool they used is the group signature technique. For the first time, they showed how to achieve both user anonymity and message confidentiality efficiently. However, the system can only support a simple access policy, that is, only one attribute is issued to each user. It is critical to achieve a fine-grained access control system because such a system facilitates granting differential access rights to a set of users and allows flexibility in specifying the access rights of individual users.

One naive approach to support provenance in cloud computing is as follows. Each user registers to a third party and is issued a certificate of the group signature to achieve conditional anonymous authentication. Each data owner encrypts each of his/her document sets with a broadcast encryption by including users owning corresponding privilege. The data owner also sends a broadcast decryption key to each valid user for each document set. As a result, when there is a user to be revoked, the data owner has to update by re-computing all the broadcast encryptions associated with this revoked user. Thus, such a naive approach is not practical, especially in cloud computing.

The notion of attribute-based encryption (ABE) [16,17], which stemmed from fuzzy identity-based encryption proposed by Sahai and Waters [18], enables for the first time public key based one-to-many encryption with fine-grained access control. Therefore, it is envisioned as a highly promising public key primitive for realizing scalable and fine-grained access control systems, where differential yet flexible access rights can be assigned to individual users. To address a complex and general access policy, two kinds of ABE have been proposed [16]: key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE). In KP-ABE, the access policy is assigned in an attribute private key, whereas, in CP-ABE, the access policy is specified in the ciphertext.

Attribute-based signature (ABS) [19–22] is also proposed to realize fine-grained access control in anonymous authentication systems, in which a signer is defined by a set of attributes instead of a single string representing the signer's identity. Compared with ABE, it does not require interaction between two participants to realize the access control. In ABS, a user obtains a certificate for a set of attributes from an attribute-certification authority known as the attribute authority. An attribute-based signature assures the verifier that a signer, whose set of attributes satisfies a (possibly) complex predicate, has endorsed the message.

Though both ABE and ABS can be used to provide a fine-grained access control system, they cannot revoke the user anonymity and find out the user's identity in a provenance record when a dispute arises. Thus, the provenance system constructed directly from ABE