ARTICLE IN PRESS

Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Future Generation Computer Systems

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

Privacy-preserving data utilization in hybrid clouds☆

Jingwei Li [a], Jin Li [b], Xiaofeng Chen [c], Zheli Liu [a], Chunfu Jia [a,*]

[a] College of Information Technical Science, Nankai University, Tianjin, China

[b] School of Computer Science, Guangzhou University, Guangzhou, China

[c] State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China

HIGHLIGHTS

- We propose a hybrid architecture for privacy-preserving data utilization.
- We propose a system for exact keyword search and access control over encrypted data.
- We show how to extend our system to support fuzzy keyword search.
- We demonstrate approaches for outsourcing cryptographic access control.

ARTICLE INFO

Article history:
Received 28 December 2012
Received in revised form
19 May 2013
Accepted 6 June 2013
Available online xxxx

Keywords:
Privacy-preserving keyword search
Fuzzy keyword search
Fine-grained access control
Attribute-based encryption
Cloud computing

ABSTRACT

As cloud computing becomes prevalent, more and more sensitive data is being centralized into the cloud, which raises a new challenge on how to utilize the outsourced data in a privacy-preserving manner. Although searchable encryption allows for privacy-preserving keyword search over encrypted data, it could not work effectively for restricting unauthorized access to the outsourced private data. In this paper, aiming at tackling the challenge of privacy-preserving utilization of data in cloud computing, we propose a practical hybrid architecture in which a private cloud is introduced as an access interface between the data owner/user and the public cloud. Under this architecture, a data utilization system is provided to achieve both exact keyword search and fine-grained access control over encrypted data. Security and efficiency analysis for the proposed system are presented in detail. Then, further enhancements for this system are considered in two steps. (1) We show how to extend our system to support efficient fuzzy keyword search while overcoming the disadvantage of insignificant decryption in the existing privacy-preserving fuzzy keyword search scheme. (2) We demonstrate approaches to realize an outsourcing cryptographic access control mechanism and further reduce the computational cost at the data user side.

1. Introduction

Cloud computing is capable of providing seemly unlimited "virtualized" resources to users as services across the whole Internet while hiding platform and implementation details [1]. Today's cloud service providers are able to offer both highly available storage and massively parallel computing resources at relatively low costs.

As cloud computing becomes prevalent, more and more sensitive data is being centralized into the cloud with the desire to be shared by users with specified privileges. Due to the fact that data owners and cloud storage are no longer in the same trusted domain, it always follows the custom that sensitive data should be encrypted prior to outsourcing. In this case, the problem of data searching and utilizing becomes a challenge.

Searchable encryption that was first initiated by Song et al. [2] to allow privacy-preserving keyword searches on encrypted data has been intensively researched in recent years [3–9]. Although it can keep data both confidential and searchable for a single user, considering a keyword search with fine-grained access control in multi-user setting—a more common scenario in cloud computing, existing searchable encryption may not work effectively. A naive approach to achieving this goal is through sharing the secret query key among a group of multiple users. Nevertheless, this not only increases the risk of key exposure, but also makes it hard to revoke a user's searching ability. Curtmola et al. [10] proposed another approach based on the broadcast encryption technique, but such a solution [10] only works in a broadcasting pattern (i.e., only one data owner and multiple data users). Furthermore, broadcast

☆ A preliminary version of this paper has been presented at the 6th International Conference on Network and System Security (NSS 2012).

* Corresponding author. Tel.: +86 13512208668.
E-mail addresses: lijw1987@gmail.com (J. Li), jin71@gmail.com (J. Li), xfchen@xidian.edu.cn (X. Chen), liuzheli1978@163.com (Z. Liu), cfjia@nankai.edu.cn (C. Jia).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.06.011

encryption in general is a quite expensive primitive and the data owner may not execute it easily.

1.1. Our contribution

In this paper, aiming at efficiently solving the problem of privacy-preserving data utilization in cloud computing we consider a hybrid architecture consisting of a public cloud and a private cloud. Unlike the existing data utilization system, the private cloud is involved as a proxy to allow data owner/users to securely outsource/query their data to/from the public cloud. Actually, this type of architecture is reasonable and has attracted a lot of attentions recently. For example, an enterprise might use a public cloud service, such as Amazon S3 for archived data but continue to maintain in-house cloud for managing operational customer data.

Under the hybrid architecture, we design a practical data utilization system which supports both privacy-preserving keyword search and fine-grained access control over encrypted data. In the proposed system, the operation of trapdoor generation is securely delegated to the private cloud but leaves behind only the file encryption and decryption at the data owner/user side. Compared with the recent work [11] which just described a keyword search scheme in sketch, we present a concrete and practical data utilization system over the hybrid cloud architecture. Moreover, rigorous analysis and experimental simulation are also provided to demonstrate the security and efficiency of the proposed system.

Furthermore, we consider to enhance our system in two steps. (1) On functionality, we present an advanced scheme to support efficient fuzzy keyword search. With the help of the private cloud, the overhead computation of generating a fuzzy keyword set and insignificant decryption are both eliminated at the data user side. (2) On efficiency, we discuss the issue of outsourcing attribute-based encryption (ABE) to a private cloud to further relieve the computational cost at the user side.

1.2. Related work

Symmetric searchable encryption. Symmetric searchable encryption (SSE) was proposed by Song et al. [2], in which a user stores his encrypted data on a semi-trusted server and later searches with a certain keyword. In [2], each keyword is independently encrypted under a specified two-layered encryption. Subsequently, Goh [12] introduced a bloom filter to construct secure indexes for the keyword search, which allows a server to check if a file contains a keyword without decrypting the entire file. A formal treatment of SSE was presented by Curtmola et al. [10]. They provided the security notions for SSE and presented the "index" approach, in which an array and a look-up table are built for the entire file collection. Each entry of the array is used to store an encryption of the file identifier set associated with a certain keyword, while the look-up table enables one to locate and decrypt the appropriate element from the array. Recently, aiming at providing SSE with efficient search and update, Liesdonk et al. [6] presented two schemes: the first one transforms each unique keyword to a searchable representation such that user can keep track of the set of associated keywords via appropriate trapdoors. The second one deployed a hash chain by repeatedly applying a hash function to an initial seed. Since only the user knows the seed, he/she is able to traverse the chain forward and backward, while the server is able to traverse the chain forward only.

Public-key searchable encryption. Boneh et al. [13] firstly proposed and suggested a public-key encryption with keyword search (PEKS) construction. Such a primitive can be widely applied in a store-and-forward system, such as an email system, in which a receiver can search for data that is encrypted under the receiver's public key. Subsequently, several improved constructions at different security levels have been presented [14,8]. Recently, many relevant extensions on keyword search have also been proposed, such as conjunctive keyword search [9,15], fuzzy keyword search [5], etc.

Twin cloud architecture. Recently, Bugiel et al. [16] provided an architecture consisting of twin clouds for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. Actually, their work is the inspiration for this paper; based on their twin cloud architecture, we consider addressing the privacy-preserving utilization problem over encrypted data in a public cloud. Moreover, the adversary model in this paper is stronger than that in [16]. Specifically, the private cloud in Bugiel et al.'s work [16] is required to be fully trusted, while it only needs to be semi-trusted in our system.

1.3. Organization

The rest of this paper is organized as follows. In Section 2, we briefly revisit some preliminaries of this paper. In Section 3, we propose the system model for our data utilization system. In Section 4, we propose a practical data utilization system in cloud computing. The security and efficiency analysis for the proposed system are respectively presented in Section 5 and Section 6. In Section 7, we consider further enhancements of our system in functionality and efficiency. Finally we draw conclusions in Section 8.

2. Preliminaries

In this section, we will introduce some background about searchable encryption, computable bilinear maps and ABE.

Trapdoors of keywords. Trapdoors of the keywords can be realized by applying a one-way function f which is defined as follows: given a keyword w and a private key k, the trapdoor of w can be defined as $T_{k,w} = f(k, w)$.

Bilinear map. Let $\mathbb{G}$ and $\mathbb{G}_T$ be two multiplicative cyclic groups of prime order p. Let g be a generator of $\mathbb{G}$ and e be a bilinear map $e : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$. The bilinear map e has the following properties:

- *Bilinearity*: for all $u, v \in \mathbb{G}$ and $x, y \in \mathbb{Z}_p$, we have $e(u^x, v^y) = e(u, v)^{xy}$.
- *Non-degeneracy*: $e(g, g) \neq 1$.
- *Computability*: There is an efficient algorithm to compute $e(u, v)$ for $\forall u, v \in \mathbb{G}$.

We say that $\mathbb{G}$ is a bilinear group if the group operation in $\mathbb{G}$ and the bilinear map $e : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ are both efficiently computable.

Attribute-based encryption. ABE was introduced by Sahai and Waters [17] and later formalized in [18] to construct fine-grained access control over encrypted data. Two flavors of ABE are classified, namely KP-ABE (key-policy ABE) and CP-ABE (ciphertext-policy ABE). In KP-ABE, attributes are used to describe the encrypted data and policies are built into users' keys, while it is inverse in CP-ABE. In this paper, we utilize the CP-ABE to facilitate key management and cryptographic access control in an expressive and efficient way. Let Ω and $\mathcal{AP}$ denote the attribute list and access policy. The CP-ABE scheme $\mathcal{ABE}$ consisting of four algorithms is described as follows:

- $\mathsf{Setup}_{\mathrm{ABE}}(1^\lambda)$: The setup algorithm takes as input a security parameter 1^λ. It outputs the public key pk_{ABE} and the master key msk_{ABE}.
- $\mathsf{KeyGen}_{\mathrm{ABE}}(\Omega, msk_{\mathrm{ABE}})$: The key extraction algorithm takes as input an attribute list Ω and the master key msk_{ABE}. It outputs the user's private key $sk_{\mathrm{ABE}}[\Omega]$.

Please cite this article in press as: J. Li, et al., Privacy-preserving data utilization in hybrid clouds, Future Generation Computer Systems (2013), http://dx.doi.org/10.1016/j.future.2013.06.011