Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

October 2017, 24(5): 23–34
www.sciencedirect.com/science/journal/10058885

The Journal of China Universities of Posts and Telecommunications

http://jcupt.bupt.edu.cn

Hybrid cloud approach for block-level deduplication and searchable encryption in large universe

Liu Zhenhua, Kang Yaqian (✉), Li Chen, Fan Yaqing

School of Mathematics and Statistics, Xidian University, Xi'an 710071, China

Abstract

Ciphertext-policy attribute-based searchable encryption (CP-ABSE) can achieve fine-grained access control for data sharing and retrieval, and secure deduplication can save storage space by eliminating duplicate copies. However, there are seldom schemes supporting both searchable encryption and secure deduplication. In this paper, a large universe CP-ABSE scheme supporting secure block-level deduplication are proposed under a hybrid cloud mechanism. In the proposed scheme, after the ciphertext is inserted into bloom filter tree (BFT), private cloud can perform fine-grained deduplication efficiently by matching tags, and public cloud can search efficiently using homomorphic searchable method and keywords matching. Finally, the proposed scheme can achieve privacy under chosen distribution attacks block-level (PRV-CDA-B) secure deduplication and match-concealing (MC) searchable security. Compared with existing schemes, the proposed scheme has the advantage in supporting fine-grained access control, block-level deduplication and efficient search, simultaneously.

Keywords block-level deduplication, searchable encryption, large universe, BFT

1 Introduction

Cloud computing is available that data owners can outsource their data to cloud without leaking their sensitive information and data user with a certain feature can access these data [1–3]. As a kind of 'one-to-many' public key encryption, attribute-based encryption (ABE) is considered as one of the most suitable encryption technology for cloud system. In ABE system introduced by Sahai and Waters [4], data stored in the cloud can be encrypted according to access control policies and decrypted only by the users with certain attributes. Ostrovsky et al. [5] divided ABE into two types: key-policy ABE (KP-ABE) and CP-ABE. In CP-ABE mechanism, the ciphertext is related to attribute access structure and user's private key is associated with an attribute set, while KP-ABE mechanism is inverse. Here, we mainly discuss CP-ABE. General CP-ABE systems [6–8] only support for 'small universe' attribute, in which attribute space size is limited in the security parameter of setup, and public parameter's size increases linearly with the size of attribute set. Rouselakis et al. [9] expanded 'small universe' CP-ABE to 'large universe' attribute, in which the attribute universe's size can be large enough.

ABE provides the data confidentiality and expressive fine-grained access control. Nevertheless, encryption would prevent search on the encrypted data. To solve this problem, Song et al. [10] proposed a keyword searchable encryption scheme, which belongs to searchable symmetric encryption. However, their scheme cannot realize the private retrieval for third-party data. In 2004, Boneh et al. [11] put forward a searchable non-symmetric or public key encryption scheme. Although Boneh et al. solved the problem of private retrieval for third-party data, their scheme has low efficiency. The reason is to use many bilinear pair operations when data owner encrypts all keywords and then sends a message. Furthermore, public key searchable encryption (PEKS) [12] enables one to retrieve keywords ciphertext without compromising the original data security. And fine-grained access control can make a accurate search in cloud. In order to achieve

Received date: 20-07-2017
Corresponding author: Kang Yaqian, E-mail: 2393340199@qq.com
DOI: 10.1016/S1005-8885(17)60230-9

fine-grained access control in PEKS, Xiong et al. [13] proposed a searchable encryption of CP-ABE scheme, which combines CP-ABE algorithm and a simple and effective searchable ciphertext algorithm of homomorphic encryption, and not only fine-grained access control but also effective ciphertext retrieval. Later, Li et al. [14] proposed a homomorphic CP-ABSE scheme in 'large universe' based on large integer factoring hard problem, which makes more convenient search by using homomorphic encryption.

On the other hand, secure deduplication is an available technique that can save storage space, reduce backup data storage, storage capacity and energy consumption through eliminating redundant copies of ciphertext stored in the cloud. Chen et al. [15] showed two main deduplication strategies: file-level deduplication and block-level deduplication. For file-level deduplication, duplicate information can be checked as a file, and only one copy is associated with each file stored in the cloud. For block-level deduplication, every file can be split into many blocks, and the repeated data can be checked in the form of blocks. It means that block-level deduplication can support fine-grained operation. Cui et al. [16] proposed a secure file-level deduplication scheme based on CP-ABE in a hybrid cloud, where private cloud performs duplicate detection through judging whether access policy is mutually contained and public cloud implements storage. Moreover, Zhao et al. [17] proposed an updatable block-level message-locked encryption scheme, in which deduplication computation is only $O\left(\mathrm{lb}\left|F\right|\right)$, where $\left|F\right|$ is the size of file.

CP-ABSE and secure deduplication have been applied extensively in cloud, and thus it is interesting to construct a scheme supporting fine-grained access control, keyword search and secure block-level deduplication. Li et al. [18] proposed firstly a secure deduplication storage scheme supporting keyword search. They presented a file-level deduplication general scheme, and then modified into a block-level deduplication general scheme. However, Li et al. did not construct concrete schemes and their general schemes cannot consider fine-grained access control.

In this paper, we try to solve the above problems. Our contributions are summarized as following:

1) Our scheme combines block-level secure deduplication and large universe CP-ABSE under a hybrid cloud mechanism. Private cloud performs secure block-level duplicate detection which achieves fine-grained deduplication, and public cloud implements data retrieval using homomorphic searchable method.

2) Using a BFT in hybrid cloud, our scheme enhances deduplication efficiency of private cloud and retrieval efficiency of public cloud.

3) Based on large integer factoring hard problem and some assumptions, we prove that our block-level deduplication is PRV-CDA-B secure deduplication and MC searchable security.

In this paper, the remainder is organized as follows. In Sect. 2, we describe some definitions and security assumptions. Outline of our scheme is showed in Sect. 3. In Sect. 4, we present the details of our CP-ABE searchable scheme supporting block-level deduplication in large universe. The performance and security analysis is given in Sect. 5. In Sect. 6, we draw conclusions.

2 Preliminaries

2.1 Definitions

Definition 1 (Bilinear parings [6]) G and G_1 are two multiplicative cyclic groups with prime order p. g is a generator of G. We note that $e: G\times G\rightarrow G_1$ is a bilinear map if the following properties hold.

1) Bilinearity: for all $a,b\in Z_p$ and $g\in G$, we have $e(g^a,g^b)=e(g,g)^{ab}$.

2) Non-degeneration: $e(g,g)\neq 1$.

3) Computability: there is an efficient algorithm to compute $e: G\times G\rightarrow G_1$.

Definition 2 (Linear secret sharing scheme (LSSS)) Denote that U is a universe of possible attributes, $\boldsymbol{A}_{l\times l}$ is a matrix, and ρ:$\{1,2,...,l\}\rightarrow U$ is a function which maps each row of $\boldsymbol{A}_{l\times l}$ to U. A secret sharing scheme $\prod$ over access matrix $\boldsymbol{A}_{l\times l}$ and map ρ is an LSSS.

1) Share$_{(A,\rho)}$: Select a number $s\in Z_P$ and a vector $\boldsymbol{v}=\left(s,y_2,...,y_l\right)\in Z_P^l$, where $y_2,y_3,...,y_l$ are used to share S. Compute $\lambda_i=\boldsymbol{A}_i\boldsymbol{v}^{\mathrm{T}}$, $i=1,2,...,n$, where $\boldsymbol{A}_i$ is the ith row vector of $\boldsymbol{A}_{l\times l}$.

2) Reconstruct$_{(A,\rho)}$: $S\in U$ is an attribute set which satisfies access strategy $(\boldsymbol{A},\rho)$. Denote a set $I=\left\{i\mid\rho\left(i\right)\in S\right\}$. There exists a set of constants $\left\{w_i\in Z_P\right\}_{i\in I}, I=\left\{i:\rho\left(i\right)\in L\right\}$ which makes $\sum\limits_{i\in I}w_iv_i=s$.