Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Controllable keyword search scheme supporting multiple users

Jun Ye [a,b,*], Yong Ding [c]

[a] School of Mathematics and Statistics, Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Sichuan, China
[b] Guangxi Key Laboratory of Cryptography and Information Security, Guangxi, China
[c] School of Computer Science and Information Security, Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, China

HIGHLIGHTS

- A novel keyword search scheme with file existence protection is proposed.
- The ciphertexts cannot be searched if the users are not permit to access to them.
- The valid query cannot be transformed into another valid query for different keywords by an adversary.
- The keyword search scheme is flexible and can achieve access control.

ARTICLE INFO

Article history:
Received 28 April 2017
Received in revised form 24 August 2017
Accepted 10 September 2017
Available online xxxx

Keywords:
Keyword search
Controllable
Data privacy

ABSTRACT

In recent years, with the rapid growth of data, data processing in local place is becoming more and more difficult. With the help of cloud computing, more and more people store their data in the remote cloud servers, so that the local storage space can be saved. In order to protect the data privacy, data will be encrypted before uploading. Keyword search technique enables user to retrieve the corresponding files with the required keywords. In traditional keyword search scheme, all the files will be returned to users if they contain the required keywords. However, it is insecure in some cases. If the user has no right to read the files, he can also get the files, although he cannot decrypt them. The user obtains extra information of the files, i.e., the files contain the same required keyword though he/she cannot decrypt them. In this paper, a controllable keyword search scheme is proposed to address this issue. The files which the user has no right to access to cannot be retrieved, even if they contain the required keyword. This scheme is useful and practical especially under the circumstances with hierarchical user groups.

1. Introduction

Cloud computing gathers a lot of resources together and provides us with a lot of convenient services. The service architecture of cloud computing are the three main aspects, i.e., Infrastructure as a Service, Platform as a Service and Software as a Service. These three service modes are realized in the way of outsourcing. The high cost parts, e.g. the cost of purchasing hardware and software, the cost of management and maintenance, are outsourced to cloud servers. Clients just need to pay for the service. In this way, clients can obtain a lot of service with a little cost.

Cloud server is powerful, and it provides huge storage space for users. Cloud storage is a type of network storage service, which gathers a large number of different types of storage devices in network together and provides some kinds of storage and access service through various application software and application programming interface (API). The physical storage spans multiple servers and the physical environment is managed by cloud storage providers, who are responsible for keeping the data available and accessible and protecting the physical environment.

The structure of cloud storage consists of four parts. Storage part, the basic part of cloud storage, provides large amounts of storage space for clients with a huge number of equipments distributed in many different areas. Management part, the core part of cloud storage, coordinates the execution and work among multiple storage devices, such that the different devices can provide the same kind of service. API part, the flexible part of cloud storage, developments various application service interfaces and application service according to the different kinds of business types. Access control part, the controllable part of cloud storage, allows the authorized clients log in the cloud storage system through the

[*] Corresponding author at: School of Mathematics and Statistics, Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Sichuan, China.
E-mail addresses: yejun@suse.edu.cn (J. Ye), stone_dingy@126.com (Y. Ding).

http://dx.doi.org/10.1016/j.future.2017.09.030
0167-739X/© 2017 Published by Elsevier B.V.

standard public application interface to obtain the cloud storage service.

In recent years, due to the fast growth of data, people cannot afford the burden of data storage and processing. In order to save the storage resources of their own, more and more users store their private data to the remote cloud servers. However, the cloud servers are untrusted. It may steal some sensitive information of users. So the data will be encrypted when uploading to the cloud server for preserving privacy. This makes it quite challenging for people to retrieve the data.

Searchable encryption helps people retrieve the required files in a practicable way. However, most of the existing schemes are limited to the single-user settings where the owner who generates the encrypted data on the cloud is also the single user to perform encrypted keyword searches on it [1]. Unfortunately, many practical applications, such as, medical care join research, require a database to support search operations by multiple users. Some keyword search schemes supporting multiple users are proposed in recent years, in which keyword privacy, query privacy and access control are considered. However, most of them do not consider the security of the existence of files.

In our real lives, the search privilege may be categorized according to the hierarchies of users. For instance, there are two hospitals in a city, Hospital A and Hospital B. A doctor in Hospital A wants to study the "SARS" virus, so he searches the patient records in the cloud. We assume 10 records returned and only 4 of them belong to Hospital A. So the doctor can only decrypt the 4 records. In this process, the other records seem to be secure, while some information has been leaked. Though the doctor in Hospital A just studies the 4 records he has the right to obtain, the doctor gets the knowledge that there are other 6 records tagged "SARS" virus in Hospital B. In many cases, if the user does not have the right to decrypt the ciphertext, he cannot obtain any information about this file even if it contains the required keyword. Hence, in keyword search systems, we need new techniques to control the search manner of users according to their hierarchies.

Attribute-based encryption (ABE) is a delicate technique for access control, which allows the encryptor to share some messages to a series of users who have certain attributes, without public key certificate. In many cases, people may not care whom the visitor is. Sometimes it may be impossible to obtain such information. In attribute-based access control system, attributes are used to determine the access permission which get rid of the restrict of identity-based policy.

Our contributions. In this paper, by utilizing the advantage of attribute-based encryption technique, we propose an attribute-based controllable keyword search scheme supporting multiple users. We consider the more practical scenarios, in which users are divided into different groups according to their hierarchies. And users in different groups have different rights to read the files. The main contributions are as follows.

- A controllable keyword search scheme is proposed. Only the files, which accord with the contents of search query and satisfy the access authority of users, can be found in the search process. If the users have no rights to access to the file, it can never be found in the search process, even if it contains the required keywords.
- The existence of files is protected. Any information of the files will be revealed, if the users are not allowed to know about them.
- A secure search manner is proposed. The cloud server cannot get any information about the keywords from the search queries. And an adversary cannot generate a valid search query from an intercepted search query.

1.1. Related work

Attribute-based encryption. Cloud computing provides a fast and convenient service for people, and service-oriented computing has appeared to become a promising research area. There are a lot of research works on cloud computing and its practical applications, such as outsourcing computation [2,3] and data sharing [4]. In order to guarantee the security of the outsourced data leads to the research on access control, such as [5]. Sahai and Waters [6] proposed the fuzzy identity-based encryption scheme in 2005, which provides an error-tolerance property for IBE. Attribute-based encryption is an important application of fuzzy IBE, which can be divided into two categories, ciphertext-policy ABE and key-policy ABE [7,8]. Based on secret sharing scheme, a attribute-based encryption scheme is proposed by Sahai. Goyal [9], in which a tree access structure was used. This scheme can be used to construct fine-grained access control. The first ciphertext-policy ABE scheme is proposed by Bethencourt [10], which is defined through the tree access structure and can deal with AND and OR gates. With the development of cloud computing, the techniques of attribute-based encryption are used in searchable encryption.

Keyword search. Cloud storage [11–13], which is like broader cloud computing [14], provides huge storage space for people to store and share their data. It brings much convenience for people. And many cloud storage schemes are proposed, such as, [15,16]. Due to the cloud server is untrusted, people have to encrypt their data before upload to the cloud server. In order to retrieve the encrypted data easily, the searchable encryption comes out. Song et al. [17] proposed the first keyword search scheme in 2000. However, the required keywords will be revealed under the statistical analysis attack. Chang et al. [18] built an encrypted hash function to improve the efficiency. Then the formal security notion of searchable encryption is proposed by Curtmola et al. [19]. In the existing keyword search schemes, most schemes are symmetric encryption based [20–22], and a lot of existing schemes are limited to the single-user setting. In addition, verifiable keyword search schemes have been proposed, such as [23–25]. Then Zhang et al. [26] proposed an efficient, and flexible keyword search scheme which supports both multi-keyword ranked search and parallel search. And then a "Greedy Depth-first Search" algorithm is proposed in [27] to improve the efficiency of multi-keyword ranked search.

In 2008, Bao et al. [28] proposed a searchable encryption keyword search scheme in multi-user settings. Every user can upload the encrypted data and every user can also search the required files over all the ciphertexts. Then there are also several schemes for multi-user search [29,30]. Zhao et al. [1] proposed an attribute-based keyword search scheme. Firstly, server checks whether the user can decrypt the ciphertext or not by the attribute-based signature, then, searches the files in the valid ciphertexts set. However, the required keyword has nothing to do with user's attributes, and if the user queries the same keyword twice, it will be identified by the server. Cao et al. [31] used the attribute-based encryption to control the decryption of ciphertexts. The ciphertexts which cannot be decrypted by the requesting user will be exposed to the server, since the search work has nothing to do with user's attributes.

1.2. Organization

The organization of this paper is as follows. Some preliminaries are given in Section 2. The system model of keyword search