Future Generation Computer Systems 62 (2016) 66–75

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Secure searches in the cloud: A survey

Fei Han [a], Jing Qin [a,*], Jiankun Hu [b]

[a] School of Mathematics, Shandong University, Jinan, China

[b] School of Engineering and Information Technology, University of New South Wales Defence Force Academy Canberra, Australia

HIGHLIGHTS

- Comprehensive survey on searchable encryption for cloud security.
- A new classification of searchable encryption applicable to cloud security.
- Open and future research directions and questions.

ARTICLE INFO

Article history:
Received 30 May 2015
Received in revised form
13 November 2015
Accepted 13 January 2016
Available online 2 February 2016

Keywords:
Cloud security
Cloud storage
Information retrieval
Searchable encryption
Secure keyword search

ABSTRACT

Cloud security is a huge concern when users and enterprises consider to deploy cloud computing services. This article mainly presented an important aspect of cloud security which is called searchable encryption. Searchable encryption is a scheme that enable users to secure search on encrypted data stored in cloud. To guarantee privacy of users' data, users encrypt their data files before uploading, then when they need to fetch some files, they can use searchable encryption to execute keyword search on encrypted data. Searchable encryption can be classified by three models, each is corresponding with different application scenario. Two of them are suitable with cloud architecture. They are called searchable public key encryption and searchable symmetric encryption respectively. Searchable public key encryption is mainly for data sharing scenario. A user wants to share some data files with other users privately. He can encrypt these files with receivers' public key. And receivers can securely search on it with their private keys. Searchable symmetric encryption can enable users securely using their online storage. They can upload data files which are encrypted with their secret keys. And then they can securely search on these encrypted data with the secret key no matter when and where they are.

1. Introduction

With the rapid development of information technology, the data of the Internet is an explosive growth. The emergence of Cloud Computing [1–3] turns to be a promising paradigm for massive data storage, more specially, cloud Storage service. Cloud computing has been the hottest word in the IT area. It provides services in a pay-per-use model to users who can access the network. Similarly cloud storage provides users on-demand storage service, such as Dropbox, Amazon Simple Storage Services (S3), and Google Drive. Using these services, users can upload their data to the cloud storage server, and access their online data over the network regardless of when and where they are. For business users, rather than building their own data center, the company can leverage cloud storage service to store their data to the cloud storage server. As shown above, cloud services provided us very huge convenience. However, the security of the cloud is far from satisfying. Cloud security [4,5] has become a challenging problem towards data of users and companies when they adopt cloud services.

Cloud security can be classified into storage security and computation security [6]. Storage security means the users' data privacy of online storage, such as avoiding data leakage, data integrity, and assured data deletion. Since in cloud services data is stored online, data owners lost the ultimate control of their data, then they cannot physically protect data from attackers' interception or manipulation. For cloud computation aspect, when users delegate the cloud server to do some computation, such as keyword search and scientific computation, cloud may not perform a secure and exact computation for the sake of saving computing resources, hence users also need to guarantee data privacy during cloud computation.

* Corresponding author.
E-mail addresses: hanf1987@163.com (F. Han), qinjing@sdu.edu.cn (J. Qin), J.Hu@adfa.edu.au (J. Hu).

http://dx.doi.org/10.1016/j.future.2016.01.007
0167-739X/

Here we will talk about the most common operation in the network services, information retrieval. We all have searched some keyword in Google. After we input some keywords, Google will return the result that is relevant to the target keywords. These operations are all done in plaintext form, Google or some guys that can eavesdrop on the network can easily know the data users want to know. If these data contains some private information of users, then their data privacy is violated. To solve this security issue, cryptographers proposed a notion called "searchable encryption" to protect the data privacy while users executed some keyword searches. Briefly speaking, Searchable encryption is a cryptosystem whose ciphertext is searchable. This system can provide data privacy by data encryption, moreover it can enable the users securely search a keyword and nobody can get information of keyword and search result.

In this paper, we provide a systematic survey on searchable encryption in three aspects—security requirements, search functionalities and deployment model. Searchable encryption is an encryption system that supports keyword search. Hence security is a key concern of the construction. We firstly demonstrate security requirements for searchable encryption schemes and then discuss how to deal with these security issues. On the other hand, searchable encryption is more than an encryption scheme, it provides users for the ability to privately search on encrypted data. So we demonstrate the search functionalities that searchable encryption scheme supports. For deployment model, we classify searchable encryption into three models. Consequently, we discuss different application scenarios for each model and provide the corresponding constructions.

Definition. A searchable encryption (SE) cryptosystem consists of three parties, a cloud storage server, data owners (DO) and receivers (searchers). The cloud storage server is the party that stores the data uploaded by DO, executes the test algorithm, and then return the result to receivers. DO is the party that owns the data files and encrypts them before uploading. Receiver is the party who want to execute some keyword search and get the result. The framework can be concluded as followed:

(1) DO encrypts his data files and the associated keywords index, then uploads encrypted files to storage server.
(2) If a receiver wants to issue a keyword search, he first computes the trapdoor responding to the keyword and sends this trapdoor to server.
(3) When server gets the search request, he computes the trapdoor with encrypted index to find if there exists some match. If so, send the data files which contain this keyword to the receiver.

Remark. In some scenarios, the role of DO and receiver is played by one user.

Framework of searchable encryption: A general searchable encryption scheme consists of four algorithms:

Setup$(1^\lambda) \rightarrow K$: The setup algorithm takes as inputs a security parameter 1^λ, and then it outputs the keys of the scheme.

Enc$(K, D) \rightarrow (I, C)$: The encryption algorithm takes as inputs the data files collection along with keys generated above, encrypts the data files and associated keywords index.

Trapdoor$(K, w) \rightarrow td_w$: The trapdoor algorithm takes as inputs the target keyword and secret key. It generates the trapdoor by encrypting keyword with secret key.

Test$(td_w, I) \rightarrow \{0, 1\}$: The test algorithm takes as inputs the encrypted keyword index and trapdoor of target keyword, by computing it returns 1 if success, otherwise return 0.

Users and storage server can apply the above four algorithms to complete the search.

The Setup algorithm generates keys for running whole encryption system.

The user generates associated keywords index for data files, and encrypts index by running Enc algorithm. Then user uploads the ciphertext and secure index to the storage server.

The receiver picks a target keyword and generate trapdoor for this keyword by running Trapdoor algorithm, then sends it to the server.

As the server receives the search query, it runs Test algorithm with encrypted index and trapdoor. If success, then return the files associated with the matching keyword.

The rest of this paper is organized as follows. In Section 2, we list the security requirement of searchable encryption. The state-of-the-art search functionalities that have been studied is present in Section 3. In Section 4, we give a formal definition of searchable encryption and classify searchable encryption into three classes based on the deployment model. For each model, we provides the application scenarios and demonstrates the corresponding works. In Section 5, we demonstrate the applications of searchable encryption. Section 6 will give conclusion and future research directions.

2. Security requirements for searchable encryption

As a encryption system, searchable encryption system must meet some security properties. It is an encryption system, hence it should guarantee the data privacy, i.e. ciphertext does not reveal any private information of the data plaintext. Except this, there are other security requirements as follows:

(1) **Search pattern and Access pattern**

These two definitions is key security requirements of searchable encryption, Abdalla et al. [7] gave formal definitions as follows:

Search pattern is induced by a search query. The search pattern of a query for keyword W is defined as the information whether a data file contains this keyword. This definition is only about the relation of inclusion, but irrelevant to the information of keyword and index.

Access pattern refers to the information of search result. It is defined as which files user gets, however, the information of files is not revealed, since they encrypted.

For both search pattern and access pattern, the trivial methods for protecting them is using PIR [8,9] or OKS [10] scheme, but owing to inefficiency of them, it is not the best solution.

The leakage of search pattern has attracted some attentions. Shen et al. [11] studied the predicate privacy of encryption systems. By using inner product encryption, the security of scheme is improved that nothing is leaked except for the access pattern. So the search pattern is protected. Bösch et al. [12] adopted inner product encryption. Meanwhile they combine somewhat homomorphic encryption [13] to protect the search pattern. They also noted that with PIR technique, the access pattern is also protected. Access pattern is more possible to be leaked than search pattern, since it contains the information of search result. Islam et al. [14] presented the risk about access pattern disclosure and showed how to protect that. Liu et al. [15] presented that many existing scheme are vulnerable to search pattern leakage and this will cause security issue, then, the authors develop a grouping based construction to construct a search pattern hiding searchable encryption from existing searchable encryption. The construction will reduce the leakage from search pattern leakage to group pattern leakage.

(2) **Keyword privacy**

In searchable encryption system, the privacy of the keyword is fundamental requirement, so keywords in index are encrypted as ciphertext. This property guarantees that an attacker cannot guess