Journal of Computer and System Sciences 90 (2017) 14–27

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

A query privacy-enhanced and secure search scheme over encrypted data in cloud computing

Hui Yin [a,b], Zheng Qin [a,*], Lu Ou [a], Keqin Li [c]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan, 410082, China
[b] Department of Mathematics and Computer Science, Changsha University, Changsha, Hunan, 410022, China
[c] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form 4 December 2016
Accepted 16 December 2016
Available online 6 January 2017

Keywords:
Cloud computing
Secure index
Secure search
Privacy protection

ABSTRACT

With the emerging of the cloud computing, secure search over encrypted cloud data has become a hot research spot. Previous schemes achieve weaker query privacy-preserving ability due to the limitations of query trapdoor generation mechanisms. In these schemes, a data owner usually knows fully well the query contents of data users and a data user can also easily analyze query contents of another data user. In some application scenarios, the data user may be unwilling to leak their query privacy to anyone else except himself. We propose a privacy-enhanced search scheme by allowing the data user to generate random query trapdoor every time. We leverage Bloom filter and bilinear pairing operation to construct secure index for each data file, which enables the cloud to perform search without obtaining any useful information. We prove that our scheme is secure and extensive experiments demonstrate the correctness and practicality of the proposed scheme.

1. Introduction

With the rapid development of cloud computing, more and more organizations and individual users are beginning to outsource their private data to cloud for enjoying IT cost savings, quick deployment, excellent computation performance, and on-demand high quality services. But, the cloud, as a semi-trusted entity [1], is not fully trusted by its customers usually due to many reasons [2]. Thus, cloud customers are usually reluctant to outsource their sensitive data to cloud in the form of plaintext. An effective solution is to encrypt data before outsourcing.

However, encryption makes effective data retrieval and utilization a very challenging task. Song et al. first introduced a practical technique that allows users to securely search over encrypted data through keywords in [3]. Later, many searchable encryption schemes have been proposed based on symmetric key and public-key setting [4–9] to strengthen security and improve query efficiency. Recently, with the growing popularity of cloud computing, how to securely and efficiently search over encrypted cloud data becomes a research focus. Some approaches have been proposed in [10–18] based on traditional searchable encryption schemes which aim to protect data users' access privacies and query privacies with better query efficiency for cloud computing.

Although existing secure query techniques allow the cloud server to perform effective search over encrypted data without knowing any useful information of data files and user query contents, most of these schemes that are designed based on

* Corresponding author.
E-mail address: zqin@hnu.edu.cn (Z. Qin).

http://dx.doi.org/10.1016/j.jcss.2016.12.003
0022-0000/

the symmetric encryption setting will straightway leak user query privacies to some internal entities, besides cloud, due to the limitations of the mechanism of query trapdoors generation. More precisely, to obtain the query trapdoors, authorized data users have to ask the data owner for encrypted query keyword(s) or keyword encryption key(s), which not only either requires the data owner to always keep online or brings heavy key management overhead, but also causes the data owner to know fully well the query contents of data users. In fact, in many application scenarios, data users are reluctant to disclose their query contents to any other entities, including data owners, for example, a suspected patient may be not willing to public his query contents to anyone else when he searches electronic medical record data for knowing pathogeneses and symptoms.

On the other hand, obviously, different data users have the same query trapdoor forms when they use identical query keywords to query because all authorized data users have the same query trapdoors returned by data owners or adopt the same keys to encrypt their query keywords. Thus, it is inevitable that a data user explicitly knows other data users' query contents by comparing query trapdoor literally. Moreover, we cannot exclude the possibilities that cloud attempts to collude with a compromised data user to expose query privacies of other data users in an open cloud environment.

To release the participation of the data owner and eliminate query key management cost in the process of query, some searchable encryption schemes based on the public-key setting have been proposed such as [4,19], which allow the data user to generate query trapdoor using own private key. However, these schemes still cannot achieve strong privacy protect if they are applied directly in the cloud environment, because secure indexes are encrypted using public key and, in the process of matching between secure indexes and encrypted query keywords, the cloud is able to obtain query contents easily by dictionary attack. Moreover, every time the data user uses the same private key to encrypt query keywords, thus the cloud can obtain data user's query keywords by analyzing the previous query results and submitted encrypted query keywords.

1.1. Our contributions

In this paper, we propose a query privacy-enhanced secure search scheme over encrypted cloud data based on secure index technique by letting the data user generate query trapdoor using randomly chosen secret keys every time. In our scheme, the query contents of a data user cannot be obtained or inferred by any other entities, including the cloud server, the data owner, and the other data users, except the user data himself. We mainly make three key contributions as follows. First, we present an efficient secure search scheme with strong query privacy protection. Our scheme allows a data user to generate random query trapdoor every time by randomizing query keyword encryption key while enables the cloud server to correctly query over encrypted secure index. Second, security analysis and proof show that our scheme is secure and query privacy-enhanced. Lastly, we implement our scheme, evaluate and compare performances on a real data set with the representative searchable encryption schemes SSE [5] and secure KNN [9].

The rest of our paper is organized as follows. We first review related work in Section 2. In Section 3, we define our system model, threat model, security definition, and several necessary techniques. We define and construct our secure scheme in detail and analyze its correctness in Section 4. In Section 5, we analyze the complexity and security of our proposed scheme. We prove that our scheme is strong privacy protective and secure in Section 6 and evaluate our scheme through practical experiments in Section 7. Lastly, we conclude this paper in Section 8.

2. Related work

2.1. Conventional searchable encryption

Song et al. first introduced a practical technique [3] that allows a data owner to use an unconventional encryption method to encrypt each word of a document and a server to perform secure search by going through the whole encrypted document using a specified encrypted keyword. To improve the efficiency and system availability, in [6], Goh et al. made use of Bloom filters and pseudo-random functions to construct secure searchable index for each encrypted data file and defined search semantic security model against adaptive chosen keyword attack. The construction allows data owners to dynamically update new files without requiring rebuilding existing indexes, but the query privacy may be revealed if keywords have been searched before. To further improve security and search efficiency, in [5], Curtmola et al. adopted the inverted index [20] technique and hash table to design a novel searchable encryption scheme named SSE (Searchable Symmetric Encryption) and formally presented new and stronger security definitions. But, this scheme requires predefining a global keywords dictionary which incurs indexes rebuilding when updating data files. Other schemes based symmetric-key such as [7,8] had also been proposed to improve searchable encryption techniques. In [4] Boneh et al. first constructed a searchable encryption scheme under public-key setting. To improve user query experiences and enrich search functionalities, multi-keywords conjunctive and disjunctive search schemes over encrypted data were proposed in [19,21,22].

2.2. Secure search in cloud computing

The data outsourcing service paradigm promotes the further study on secure privacy-preserving search for cloud computing. Based on SSE [5], Wang et al. [13] first used keyword relevance score to implement top-k secure search over encrypted