J. Parallel Distrib. Comput. 112 (2018) 97–107

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

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

Achieving verifiable, dynamic and efficient auditing for outsourced database in cloud

Tao Xiang [a,*], Xiaoguo Li [a], Fei Chen [b], Yuanyuan Yang [c], Shengyu Zhang [d]

[a] College of Computer Science, Chongqing University, Chongqing 400044, China
[b] Department of Computer Science and Engineering, Shenzhen University, Shenzhen 518060, China
[c] Department of Electrical & Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA
[d] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong

HIGHLIGHTS

- We propose a novel auditing scheme for outsourced database without TPA.
- We propose a new cuckoo filter without false positive ratio.
- We propose a controllable Paillier encryption to support various database operations.
- We prove that our scheme protects the privacy and query integrity of outsourced data.

ARTICLE INFO

Article history:
Received 1 November 2016
Received in revised form 7 May 2017
Accepted 12 October 2017

Keywords:
Cloud computing
Database outsourcing
Outsourcing security
Integrity auditing

ABSTRACT

With the help of cloud computing, users can outsource their data to a cloud service provider (CSP) and enjoy the high-quality on-demand services from the cloud. On the other hand, once owners no longer have physical possession of the outsourced data, the protection of data integrity in cloud computing becomes a critical and challenging task. In this paper, we propose a novel and verifiable auditing scheme for outsourced database without a third party auditor (TPA). It can simultaneously authenticate the correctness and completeness of query results, preventing the dishonest CSP from returning fake or incomplete results to the users. Our scheme also supports partial attribute retrieval and flexible data dynamics. The security proof and the performance analysis show that our proposed scheme is secure and efficient for practical deployment.

1. Introduction

With the continuing growth of cloud computing, it becomes prevalent to outsource the management of large-sized data from local sites to commercial public clouds. In this way, organizations or companies can concentrate on their core tasks and operate other business applications via Internet, without incurring substantial hardware, software, and personnel costs in maintaining them in-house. The database-as-a-service (DAS) model, introduced in [14], becomes an important database outsourcing paradigm. In this model, the cloud service provider (CSP) is responsible for supplying not only plenty of software or hardware resources to host the clients' databases, but also mechanisms for clients to create, access, and update their outsourced data.

However, the database outsourcing comes with the challenge on the security of users' data, and it is very important to provide adequate security measures for protecting the outsourced data from both malicious outside attackers and the CSP itself. Security in database outsourcing mainly implies data privacy and data integrity [28]. First, to protect data privacy, traditional encryption schemes can ensure that the CSP learn nothing about the outsourced data. Second, the data integrity includes two aspects, storage integrity and query integrity. Storage integrity refers to the ability to check whether the outsourced data are lost or corrupted without retrieving it, many existing techniques such as provable data possession (PDP) [1] and proofs of retrievability (POR) [24] aim to solve this challenge. Query integrity refers to the ability to verify the *correctness* and *completeness* [16,29] of the query results returned from CSP. Specifically, correctness means that the owner must be able to check the validity of the returned results, i.e., the returned records do exist in the outsourced database and have not been modified in any way. Completeness means that the query results should include all valid records that satisfy the query conditions.

* Corresponding author.
E-mail address: txiang@cqu.edu.cn (T. Xiang).

https://doi.org/10.1016/j.jpdc.2017.10.004
0743-7315/

To ensure the query integrity, in recent years, many solutions have been proposed to verify the correctness and the completeness in DAS model. Many verifiable outsourced database solutions that based on Merkle hash tree (MHT) [8,18] are proposed to verify the correctness of search results, but most of them cannot achieve completeness verification. To address this issue, some solutions are explored based on aggregated signature and signature chaining [18,19,21,34], however these constructions fail to verify the case when an empty set is returned by the cloud. As an alternative way, probabilistic verification methods achieve the completeness by inserting some fake records into the outsourced database beforehand, however it requires that the fake records must be known by all authenticated users. Recently, Wang et al. [29] presented a novel verifiable auditing scheme based on the Bloom filter that overcome the above problems, but their proposal only works in the static setting. In [30], they improve their work with a invertible Bloom filter (IBF) such that it works in the dynamic setting.

However, the existing solutions fulfilling correctness and completeness verification at the same time usually suffer from the following two weaknesses. First, they require high communication overhead. Because the existing solutions are based on the *record level* of database table (e.g. [18,29,30]), the cloud server needs to return all the attributes of a record for verifiability when answering a query. The communication cost between cloud server and the client is expensive even if only a single attribute of database table is retrieved in the query, especially when database tables contain a great number of attributes. Second, a trusted third party auditor (TPA) is usually needed in the frameworks of existing solutions (e.g. [29,30]). On the one hand, TPA is the bottleneck of TPA-based auditing architectures and it also induces extra overhead. On the other hand, the assumption of trusted TPA may be not available in practical deployment.

To address these issues, in this paper, we propose a verifiable database auditing scheme to efficiently verify both the correctness and the completeness of query results. For saving the communication costs, we introduce *partial attribute retrieval*, which means only the retrieved attributes are returned from the CSP in the process of a query. Specifically, we utilize the aggregated signature to realize the partial attribute retrieval, which is performed on an *element level* rather than a *record level*. For eliminating the TPA, we propose a dynamically adjustable-capacity cuckoo filter (DACF) without false positive ratio. Our contribution can be summarized as follows.

- We propose a verifiable and efficient auditing scheme for privacy preserving outsourced database in the secret-key setting without a third party auditor (TPA). It checks the integrity of database at element level and thus supports partial attribute retrieval, which saves substantial communication overhead. To the best of our knowledge, our scheme is the first one to support verifiable partial attribute retrieval.
- We propose a dynamically adjustable-capacity cuckoo filter (DACF) without false positive ratio and a controllable Paillier encryption to enable our auditing scheme to support various dynamic operations flexibly and efficiently, including data insertion, deletion, and update. Compared with many existing solutions, our scheme achieves correctness and completeness simultaneously in the semi-honest model.
- We prove that the proposed scheme not only protects the privacy of outsourced data, but also ensures the integrity, including correctness and completeness of search results in the semi-honest adversary model. Furthermore, the thorough experimental evaluation demonstrates that our scheme is quite efficient in verification and ready for practical use.

The rest of the paper proceeds as follows: Section 2 reviews the related work. Section 3 formulates the auditing problem in database outsourcing. Section 4 introduces some useful preliminaries. Section 5 presents two novel supporting mechanisms serving as the foundation of our scheme. Section 6 presents our verifiable auditing scheme. Section 7 and Section 8 provide the analysis and performance evaluation results, respectively. Finally, Section 9 concludes the paper.

2. Related work

In this section, we briefly review the closely related work on privacy-preserving verifiable outsourced database at first. Then we explain the difference between our proposed partial attribute retrieval and the existing work on taking queries over arbitrary dimensions.

2.1. Privacy-Preserving verifiable outsourced database

To achieve privacy-reserving verifiable data outsourcing, we can leverage some off-the-shelf cryptographic primitives to ensure the data integrity. However, the resulting solutions are either computation extensively or loopholes existentially, and does not suitable for practical deployment. For example, in [13] and [2], the solutions based on oblivious RAMs and verifiable computation were proposed to address the security problems in remote data storage. Although their protocols are pretty beautiful in theory, the resulting cloud storage auditing protocol is also far beyond practical deployment due to its inefficiency. For practical deployment, many well-designed solutions have been proposed in recent years and they can be categorized into three categories as follows according to their approaches.

The first approach is based on authenticated data structures (e.g., Merkle hash tree (MHT), B-Tree) to provide the integrity of search results [8,16,22,29,35]. The main idea is to generate an index based on MHT for the whole database, and the integrity auditing can be achieved by re-computing the signature on the root of MHT. The security of MHT is based on the collision-resistance of the hash function used. In order to achieve the verifiability of each single data tuple, a set of $\log(n)$ sibling nodes from a specific leaf to the root must be computed and sent to the user, which is extensive in communication complexity. Furthermore, most of MHT-based schemes can ensure the completeness of search results.

The second approach is based on aggregated signature and signature chaining [19,21,34]. This approach reduces communication and computation overhead for query verification. In [19], the authors extend the aggregated signatures to the immutability version, preventing the adversary from computing a new valid aggregated signature even if some aggregated signatures have been possessed. In [34], a new verifiable aggregation query scheme was proposed for outsourced database. In their work, each tuple is assigned an authentication tag based on a polynomial, which can be used to check the integrity of query results for certain aggregation queries. However, their scheme cannot achieve full completeness.

The third approach is probabilistic integrity verification [25,31]. The main idea is that the data owner inserts some fake tuples to the database beforehand. However, an implicit assumption of this method is that the fake tuples must be known by all users. Trivially, a dishonest CSP can collude with any compromised user and then definitely know the fake tuples. Besides, this approach requires the CSP to return all attributes of a tuple and thus cannot support some common database operations such as projection.