Information Sciences 432 (2018) 79–96

Contents lists available at ScienceDirect

Information Sciences

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

Executing multi-dimensional range query efficiently and flexibly over outsourced ciphertexts in the cloud

Zhuolin Mei[a], Hong Zhu[b,*], Zongmin Cui[a,*], Zongda Wu[c], Gang Peng[e], Bin Wu[a], Caicai Zhang[d]

[a] School of Information Science and Technology, Jiujiang University, Jiangxi, China
[b] School Of Computer Science and Technology, Huazhong University of Science and Technology, Hubei, China
[c] Oujiang College, Wenzhou University, Zhejiang, China
[d] School of Information engineering and Art Design, Zhejiang University of Water Resources and Electric Power, Zhejiang, China
[e] School of Information Science and Technology, Huizhou University, Guangdong, China

ARTICLE INFO

Article history:
Received 17 June 2016
Revised 27 November 2017
Accepted 30 November 2017
Available online 5 December 2017

Keywords:
Cloud computing
Encryption
Range search
Efficiency
Flexibility

ABSTRACT

Encryption is one of the most straightforward methods for ensuring the confidentiality of outsourced data on the cloud. However, encryption makes queries more difficult to perform. In recent years, new encryption schemes for facilitating queries have been proposed. However, for these schemes, some cannot support the scenario of multiple users, some are inefficient, and some are not sufficiently flexible (users must always ask the data owner for the tokens that are used for searching ciphertexts on the cloud). In this paper, we propose a scheme that supports efficient and flexible range search over ciphertexts in the scenario of multiple users. In our scheme, we construct an Encrypted Interval Tree (EIT) as the index for ciphertexts. The data owner outsources the EIT and ciphertexts to the cloud, and later distributes secret parameters (search keys, navigation paths and signatures) to users. By utilizing these secret parameters, users can generate tokens for the queried ranges without communication with the data owner and subsequently use the tokens to perform range search over ciphertexts on the cloud. Moreover, the signature technique is adopted in our scheme. Thus, the cloud can authenticate the identifiers of users and determine the legality of tokens. In this paper, we implement our scheme and conduct extensive experiments. The experimental results demonstrate the efficiency of our scheme. Finally, we analyze the security of our scheme.

1. Introduction

Driven by lower cost, higher reliability, better performance and faster deployment, data and services have been increasingly outsourced to the cloud [13,22]. However, outsourced data in a remote cloud are not under the direct control of data owners; thus, privacy concern naturally becomes a primary barrier to the adoption of cloud computing [19,23]. To protect the privacy of outsourced data, encryption before outsourcing is regarded as one of the most straightforward methods. However, conventional encryption schemes, such as block ciphers [11], do not directly support range queries without loss

* Corresponding authors.
E-mail addresses: meizhuolin@126.com (Z. Mei), hongzhuwh@163.com (H. Zhu), cuizm01@gmail.com (Z. Cui), zongda1983@163.com (Z. Wu), peng@hzu.edu.cn (G. Peng), zhangcc@zjweu.edu.cn (C. Zhang).

https://doi.org/10.1016/j.ins.2017.11.065
0020-0255/

of privacy. In recent years, although new encryption schemes [2,3,5,9,10,12,14,17,20] have been proposed to facilitate range queries, there still exist limitations.

In previous works [2,3], data are encrypted in an ordered manner. Thus, these methods can support efficient range search over ciphertexts. However, the ordering information reveals the privacy of ciphertexts. Adversaries can precisely estimate ciphertexts by exploiting the ordering information [7]. To reduce the leakage of ordering information, bucketization methods [9,10,12,20] have been proposed. In the bucketization methods, data are partitioned into several buckets, and different data in the same bucket are encrypted as a unit. In this way, the ordering information of ciphertexts in the same bucket can be protected. Thus, adversaries cannot precisely estimate the ciphertexts in buckets. However, these bucketization schemes do not support the scenario of multiple users. Namely, the only user in these schemes is the data owner himself/herself. This drawback limits the adoption of bucketization schemes in cloud computing.

To support range search over ciphertexts in the scenario of multiple users, public-key-based schemes [5,14,17] are proposed. However, compared with [2,3,9,10,12,20], these public-key-based schemes are very inefficient because they require many complex calculations. Additionally, they are not flexible enough. We assume that a data owner has distributed a token to a user u (the token is used to search the range $[a, b]$, which indicates u's search privilege), but now u only wants to search the sub-range $[c, d]$ ($[c, d] \subset [a, b]$). In this situation, u must ask the data owner for a new token for the queried range $[c, d]$. Thus, the data owner must frequently respond to the query requests from different users, check their privileges, and generate and distribute tokens to them. Consequently, the data owner needs a large amount of bandwidth and powerful devices to perform these cumbersome tasks. This reduces the significance of the cloud, because data owners hope the powerful will to do all the cumbersome tasks for them.

Considering the limitations of the above schemes, we propose a scheme that can support efficient logarithmic complexity search over encrypted multi-dimensional data in the scenario of multiple users. Our scheme can protect the privacy of outsourced data to a great extent, and adversaries cannot precisely estimate the ciphertexts on cloud. Moreover, our scheme is very flexible: if a user maintains the secret parameters of a range $[a, b]$, which indicates his/her search privilege, the user himself/herself can generate a token for the range $[c, d]$ ($[c, d] \subset [a, b]$) without communication with the data owner, and then use the token to search ciphertexts on the cloud. Thus, our scheme alleviates the data owner's workload of privilege check, token generation and distribution, and reduces the bandwidth consumption between the data owner and users.

We summarize our approach in the following four aspects.

(1) A numeric column of a database table can be viewed as one dimension in a multi-dimensional coordinate system. Thus, entries in the database table can be viewed as data points in the multi-dimensional coordinate system. To protect the privacy of these data points, the data owner could encrypt them and then outsource the ciphertexts to the cloud. To facilitate range search, data owner could build an index for these encrypted data points and then outsource the index to the cloud. We refer to this index as the Encrypted Interval Tree (EIT). In EIT, (i) each node is associated with a unique Multi-Dimensional Range (MDR); (ii) the MDR of the root covers all outsourced data points; (iii) the MDR of an internal node is the union of the MDRs of its children; (iv) the MDR of a leaf node has the smallest size, and encrypted data points that are covered by the MDR are stored under the leaf.

(2) The data owner can choose any secure encryption scheme to encrypt the outsourced data points. Thus, the security of the encrypted data points only relies on the encryption scheme that has been chosen by the data owner. The data owner can choose any secure encryption scheme that supports k-Nearest Neighbor (kNN) search over ciphertexts to encrypt the MDRs of nodes in EIT.

(3) In the nodes of EIT, MDRs are encrypted using different encryption keys. According to the range that a user can search, the data owner distributes secret parameters (navigation paths, signatures and search keys) to the user. According to the structure of EIT, we design the one-way derivation for search keys: if a user has the secret parameters of MDR[1], the user can derive the search key for the range R ($R \subset MDR$). Then, the user can generate a token for R by utilizing the navigation paths, signatures and search key of R. Finally, the user sends the token to the cloud to search the range R.

(4) After receiving the token from a user, the cloud authenticates the identifier of the user and checks the validity of the token. If the token is legal, the cloud performs range search on EIT and returns the encrypted data points that are covered by R to the user as the search results.

The contributions of this paper are listed as follows:

(1) We propose Encrypted Interval Tree (EIT) as the index for the encrypted data on cloud.

(2) We propose two key derivation methods on EIT. If the data owner distributes the secret parameters of R to a user, the user can derive the search key for any sub-range Q ($Q \subset R$) and then generate the token for Q. After receiving the token from the user, the cloud can efficiently perform range search on EIT.

(3) We conduct extensive experiments. The experimental results show that our scheme has good performance on token generation and range search. Finally, we analyze the security of our scheme.

The remainder of this paper is organized as follows. Section 2 discusses the related work. Section 3 presents the architecture of our scheme and preliminaries. Section 4 presents the tasks that are performed by data owner, including the

[1] MDR is the abbreviation for multi-dimensional range and *MDR* denotes a specific multi-dimensional range.