Knowledge-Based Systems 79 (2015) 18–26

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

L-EncDB: A lightweight framework for privacy-preserving data queries in cloud computing

Jin Li [a,b,*,1], Zheli Liu [c,*,1], Xiaofeng Chen [b], Fatos Xhafa [d], Xiao Tan [e], Duncan S. Wong [e]

[a] School of Computer Science and Educational Software, Guangzhou University, Guangzhou, PR China
[b] State Key Laboratory of Integrated Service Networks (ISN), Xidian University, Xi'an, PR China
[c] College of Computer and Control Engineering, Nankai University, Tianjin, PR China
[d] Department of Languages and Informatics Systems, Technical University of Catalonia, Spain
[e] Department of Computer Science, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Available online 19 April 2014

Keywords:
Data query
Outsourcing
Privacy
Format-preserving encryption
Cloud Computing

ABSTRACT

With the advent of cloud computing, individuals and organizations have become interested in moving their databases from local to remote cloud servers. However, data owners and cloud service providers are not in the same trusted domain in practice. For the protection of data privacy, sensitive data usually have to be encrypted before outsourcing, which makes effective database utilization a very challenging task. To address this challenge, in this paper, we propose L-EncDB, a novel lightweight encryption mechanism for database, which (i) keeps the database structure and (ii) supports efficient SQL-based queries. To achieve this goal, a new format-preserving encryption (FPE) scheme is constructed in this paper, which can be used to encrypt all types of character strings stored in database. Extensive analysis demonstrates that the proposed L-EncDB scheme is highly efficient and provably secure under existing security model.

1. Introduction

The ever-increasing amount of valuable digital data both at home and in business needs to be protected, since its irrevocable loss is unacceptable. The advent of cloud storage motivates enterprises and organizations to outsource data storage to third-party cloud providers, as evidenced by many real-life case studies. Cloud storage services promise to be a solution for this problem. In recent years, their popularity has increased dramatically. They offer user-friendly, easily accessible and costsaving ways to store and automatically back up arbitrary data, as well as data sharing between users and synchronization of multiple devices.

As in any existing application and system, security and privacy play an extremely important role for the success, and certainly raise new challenges among the many others that cloud storage is confronted with. Specifically, when entrusting data to the cloud, data owner also releases control over the data, resulting that their trust is put in the cloud service provider's integrity and in the security of its access control mechanisms. However, individuals and especially businesses hesitate to entrust their data to cloud storage services since they fear that they will lose control over it. Recent successful attacks on cloud storage providers have exacerbated these concerns. The providers are trying to alleviate the situation and have taken measures to keep their customers' data secure. The simple and popular solutions adopted for data privacy are traditional encryption techniques such as public key encryption or symmetric key encryption. Through these encryption methods before outsourcing, the security of users's data can be protected.

However, traditional database encryption will change the data structure of original data, and results in the impracticability of database application for various kinds of SQL operations. If the data structure is changed, it cannot support data operations over ciphertext such as range query and fuzzy query. Especially, there has been considerable recent interest in the paradigm of data mining-as-service: a company (data owner) lacking in expertise or computational resources can outsource its mining needs to a third party service provider. However, both the outsourced database and the knowledge extracted through data mining are considered private property of the data owner in many applications. Thus, to protect data privacy while realizing data mining and knowledge extraction, the data owner is required to transform its data without changing its structure before outsourcing.

Contribution. To realize effective data utilization after secure outsourcing, we propose a lightweight encrypted database

* Corresponding author at: School of Computer Science and Educational Software, Guangzhou University, Guangzhou, PR China. Tel.: +86 20 39366631 (J. Li).
E-mail addresses: jinli71@gmail.com (J. Li), liuzheli1978@163.com (Z. Liu).
[1] Jin Li and Zheli Liu contribute to this work equally.

http://dx.doi.org/10.1016/j.knosys.2014.04.010
0950-7051/

mechanism denoted by L-EncDB. This new mechanism is able to protect sensitive information while keeping the data structure in outsourcing service for big database application. In the proposed L-EncDB system, the encryption and query are based on SQL. Through only one interface, all SQL sentences for database can be interpreted.

Furthermore, based on format-preserving encryption (FPE) technique and a new character string FPE scheme, L-EncDB can be implemented to preserve data type and length in ciphertext. It enables (i) to encrypt data and store them without changing original database structure, (ii) to perform SQL operations on all kinds of databases, including text database such as SQLite and Access and (iii) to support SQL-based operations including advanced fuzzy and range queries.

Innovation. In this paper, a novel FPE scheme with the method of "multi-radix modular addition" is proposed to support the L-EncDB lightweight framework for privacy-preserving outsourced database. The new proposed FPE can preserve both length and storage size of character strings, which cannot be efficiently achieved in the traditional FPE schemes. Based on the FPE scheme, data operations such as data mining and SQL-based queries can be directly executed over ciphertexts in the proposed L-EncDB framework. Furthermore, L-EncDB framework can be extended to text database (such as SQLite used in mobile) and NoSQL databases, which have not been considered in the previous related work.

1.1. Organization

The rest of this paper proceeds as follows. In Section 2, we give a survey for the related work to ours. In Section 3, we propose the system architecture and construction method for the L-EncDB system. In Section 4, we propose a practical construction of FPE for character string. Its security and performance analysis is also given in this section. In Section 5, we present the implementation of prototype for L-EncDB with the proposed FPE, and in Section 6, we present an extension of the L-EncDB to NoSQL database encryption. Finally we draw conclusion and show the future work in Section 7.

2. Related work

We briefly discuss FPE technique and privacy-preserving database encryption solutions in this section.

2.1. FPE

The notion of FPE [1–4] has been proposed to generate ciphertext with the same format as plaintext while encrypting sensitive information. More specifically, FPE can keep data type and length in the ciphertext, therefore, without changing database structure and field type. Thus, the use of FPE enables upgrading database security in a transparent way. The goal of FPE is to generate ciphertext which falls in the same *domain* as the plaintext. Some practical FPE schemes have been proposed for simple domains such as integer [5], character data [3] and datetime [6]. Character data is the common data type in database, which appears in the form of character strings, i.e., the finite sequences of characters from some character sets. However, there is no suitable character FPE solution to preserve both length and storage size of strings above. For a string with character from iso-8859-1 or ASCII, where the storage size of each character is 1 byte, the length of string is equal to its storage size, and FFX is also suitable in this case.

However, most of character sets are represented using more powerful encoding formats, and different characters may require different byte counts to represent. In this paper, such a character set is called "multi-byte character set". Consider a character string of length n with each character in UTF-8, its storage size will be from n bytes to $4n$ bytes. In this case, FFX is unsuitable. In 2012, Li et al. [7] proposed a solution based on cycle-walking [8]. However, cycle-walking cannot guarantee stable efficiency, which is impractical in most applications. In this paper, we develop a new FPE method in Section 4.

2.2. Privacy-preserving database encryption

A number of research results [9–13] were proposed for privacy-preserving database encryption. However, most of them cannot provide the complete solution for general SQL-base operations over encrypted data. To support query over encrypted numerical data, Hakan et al. [9] firstly presented a basic framework of how to ensure data security in "Database-As-Service" (DAS) model, in which a coarse query is executed by the database service provider. Based on this basic framework, Wu et al. [12] described a solution for query over encrypted character strings.

One of the most typical database encryption solutions is CryptDB [14], which explores an intermediate design point to provide confidentiality for applications that use database management systems (DBMSes). As shown in Fig. 1, CryptDB works by intercepting all SQL queries in a database proxy, which rewrites queries to execute over encrypted data (CryptDB assumes that all queries go through the proxy). The proxy encrypts and decrypts all the data, and changes some query operators, while preserving the semantics of the query. However, CryptDB is not designed for existing database applications and the DAS model of cloud storage. In cloud computing, users are able to store, modify and retrieve data from anywhere in the world, as long as they have access to the Internet. CryptDB changes the database structure and stores the ciphertexts generating by different encryption methods.

2.3. Other related work

The notion of order preserving encryption (OPE) [15–18] is another important encryption method in database to achieve the confidentiality while keeping the order of underlying plaintexts. Such a property allows users to perform comparison and range query over encrypted data without decrypting them. Another notion related is searchable encryption (SE) [19,20], which provides functionalities to perform keyword search over encrypted data without decrypting them. There are also some other related

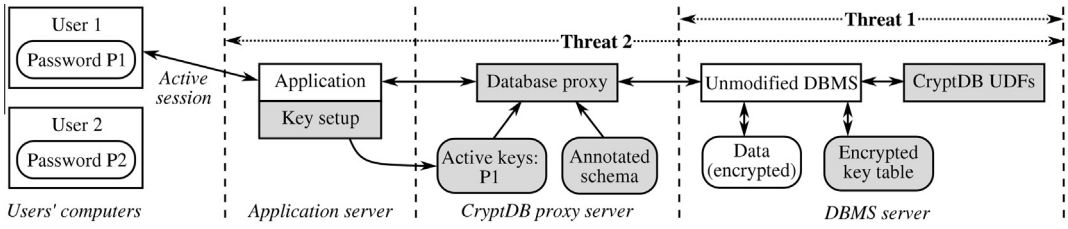

Fig. 1. CryptDB architecture.