Two-Cloud Secure Database for Numeric-Related SQL Range Queries with Privacy Preserving

K.Deepak Rajkumar, M.Arvinth, M.Balagangatharan, Suguna Angamuthu
Final Year B.Tech-IT, Department of Information Technology, Nandha Engineering College, Erode
Assistant professor, Department of Information Technology, Nandha Engineering College, Erode
Email id: arvinthmano@gmail.com

Abstract: Industries and individuals outsource database to realize convenient and low-cost applications and services. In order to provide sufficient functionality for SQL queries, many secure database schemes have been proposed. However, such schemes are vulnerable to privacy leakage to cloud server. The main reason is that database is hosted and processed in cloud server, which is beyond the control of data owners. For the numerical range query ("">", "<", etc.), those schemes cannot provide sufficient privacy protection against practical challenges, e.g., privacy leakage of statistical properties, access pattern. Furthermore, increased number of queries will inevitably leak more information to the cloud server. In this paper, we propose a two-cloud architecture for secure database, with a series of intersection protocols that provide privacy preservation to various numeric-related range queries. Security analysis shows that privacy of numerical information is strongly protected against cloud providers in our proposed scheme.

I. INTRODUCTION

The growing industry of cloud has provide a service paradigm of storage/computation outsourcing helps to reduce users’ burden of IT infrastructure maintenance, and reduce the cost for both the enterprises and individual users [1], [2], [3].

One straightforward approach to mitigate the security risk of privacy leakage is to encrypt the private data and hide the query/access patterns. Unfortunately, as far as we know, few academia researches satisfy both properties so far. CryptDB [7] is the first attempt to provide a secure remote database application, which guarantees the basic confidentiality and privacy requirement, and provides diverse SQL queries over encrypted data as well. CryptDB uses a series of cryptographic tools to achieve these security functionality. Especially, order-preserving encryption [15] is utilized to realize numerical-related range query processes. From the perspective of query functionality, CryptDB supports most kinds of numerical SQL queries with such cryptology. However, such privacy leakage hasn’t been well addressed thoroughly, since OPE is relatively weak to provide sufficient privacy assurance.

Some specific purpose cryptography like order preserving encryption (OPE) will expose some private information to the cloud service provider naturally: As it is designed to preserve the order on ciphertexts so that it can be used to conduct range queries, the order information of the data, the statistical properties derived therefrom, such as the data distribution, and the access pattern will be leaked.

From the work in [16], the privacy can be preserved against the cloud, if the sensitive knowledge is partitioned into two parts, and distributed to two non-colluding clouds. In the literature [17], the authors also introduce a two-party system to design a secure knn query scheme, which enables the client to query k most similar records from the cloud securely. This divide-and-conquer mechanism can know any private information from one singe isolated part of the knowledge, and each of both clouds only knows its own part. In this paper, we introduce a secure two-cloud database service architecture, where the two clouds are non-colluding and both of them knows only part of knowledge. Based on this architecture, we further propose a series of interaction protocols for a client to conduct numeric-related query over encrypted data from remote cloud servers. The numeric-related query includes common query statements, such as greater than, less than, between, etc.. The main contribution of this paper can be summarized as follows: 1) We propose a non-colluding cloud architecture to conduct a secure database service, in which the data is stored in one cloud,
while the knowledge of query pattern is well partitioned into two parts, and knowing only one cannot reveal any private information; 2) We then present a series of intersection protocols to provide numeric-related SQL range query with privacy preservation, and especially, such protocols will not expose order-related information to any of the two non-colluding clouds. The rest of this paper is organized as follows: Section II discusses the related work. In Section III, the two-cloud architecture is presented, following with security assumption and security requirements. In Section IV, some definitions are given, including the preliminaries of cryptographic techniques. We provide the detailed scheme in Section V, and discuss the privacy and performance properties of the proposed scheme in Section VI and Section VII, respectively. Section VIII concludes this paper.

II. RELATED WORK

Fuzzy query over encrypted data is becoming a popular topic, since in practical scenarios, some query requests usually want to retrieve data with similar, rather than exactly same indexes[18], [19]. Fuzzy searchable encryption has been introduced for cloud computing in many literatures, such as [20], [21], [22], [23], [24], [25]. These schemes deal with the issue that search keywords allows small-scaled distinction in character/numeric level. Specifically for numerical keywords, the query predicate can get numerical records within a range. Some schemes targeted at spatial query, especially knn [17], [26], [27], [28], [29], which focus on the distance between the query vector and the data. They usually inquire about certain spatial objects (or several numerical attributes) related to the others within a certain distance. Range query [30], [31], [32] has been proposed for that purpose. However, such existing range query schemes are not suitable for practical secure database due to high storage overhead to maintain the corresponding ciphertext.

III. SYSTEM ARCHITECTURE,
SECURITY ASSUMPTION AND SECURITY REQUIREMENTS A. System Architecture Our proposed secure database system includes a database administrator, and two non-colluding clouds. In this model, the database administrator can be implemented on a client’s side from the perspective of cloud service. The two clouds (refer to Cloud A and Cloud B), as the server’s side, provide the storage and the computation service.

The two clouds work together to respond each query request from the client/authorized users (availability). For privacy concerns, these two clouds are assumed to be non-colluding with each other, and they will follow the intersection protocols to preserve privacy of data and queries (privacy). In our scheme, the knowledge of stored database and queries is partitioned into two parts, respectively stored in one cloud. The mechanism guarantees that knowing either of these two parts cannot obtain any useful privacy information. As shown in Fig. 2(a), to conduct a secure database, data are encrypted and outsourced to be stored in one cloud (Cloud A), and the private keys are stored in the other one (Cloud B). For each query, the corresponding knowledge includes the data contents and the relative processing logic. We utilize a prototype of knowledge partition, dividing application logic into two parts, which is firstly proposed by Bohli et al. in [16]. The application logic, as a secret knowledge, is partitioned into two parts, each of which is only known to one cloud. This prototype is shown in Fig. 2(b). Intuitively, this two-cloud architecture increases some complexity to some extent, and we will analyze and point out that this overhead is acceptable in Section B.

B. Security Assumption Following the general assumption of many related works in public cloud, we assume the clouds to be honest-but-curious: On one hand, both of the two clouds will respond with correct information in the interactions of our proposed scheme (honest); on the other hand, the clouds try their best to obtain private information from the data that they process (curious). From the perspective of privacy assurance, here the data not only include permanently stored information (i.e., database), but also each temporary query request (i.e., queries). Additionally and importantly, as the assumption in some existing works [16], [38], we assume that the two clouds A and B are non-colluding: Cloud A follows the protocol to add required obfuscation to protect privacy against cloud B, so that cloud B cannot obtain additional private information in the interactions with Cloud A. No private information is delivered beyond the scopes of protocols.

C. Potential Threats and Privacy Requirements This section describes the potential threats and the privacy requirements when the database is outsourced to public cloud. The stored data contents and the query processes. Although there are many data encryption schemes, some fail to provide sufficient privacy preservation after statistical analysis: Repeated and large-amount query processes not only leak the access patterns but also disclose the stored...encrypted data progressively.

Data contents. The privacy of data contents includes (1) the definition and description of each column (column name) in the table of the stored database, and (2) the values of each record in the table. Some related works have mainly focused on this issue, in which the column names are blinded (such as CryptDB [7]) and meanwhile the values are encrypted with some other encryption techniques (such as Order Preserving Encryption[33]) and some deterministic encryption schemes[7], so that the adversaries cannot easily and directly guess the meaning of the column, or the values of the data. However, in an outsourced database, utilizing encryption alone, without other mechanisms, is far from being enough to preserve the privacy of the data contents. With the development of data analysis, by extracting features from data and queries, classification technique can help understand the definition of columns, and then breach of confidentiality of data contents.

Statistical properties. Besides the static properties can disclose the private information of data contents, such properties themselves are already sensitive and private for the client. Order Preserving Encryption(OPE), which is widely used in constructing the secure database, with support of range queries, directly exposes the statistical information in the encryption field. Furthermore, the leakage of statistic properties is part of the nature of outsourced cloud database.
service: the cloud can learn the statistical properties (like order) by repeated query requests. As an example, Fig. 3 describes such an attack: After two simple queries over one same column, the order relationship of some data in certain column can be determined. There are also some other direct and indirect scenarios to leak statistical properties. In this way, even though the order property is not exposed to the semi-trusted cloud at the beginning, the cloud can gradually find out the order information after many query requests.

Query pattern. The query pattern also contains privacy information, as they can reveal the client’s purpose of the query. Even worse, such pattern can leak some statistical properties, as discussed above.

Based on the above discussion, we assert that an outsourced secure database providing numeric-related queries should prevent the following private information from being obtained by the honest-but-curious clouds:

Data contents. The data contents includes item values and column names, which are the raw data that should be protected against any potential adversaries

Statistical properties. It includes the order of data and their probability distributions, some of which include “>”.

Query pattern. Each query should be kept private against the honest-but-curious clouds and any unauthorized parties. The secrecy of such pattern should be well preserved even after many query processes.

IV. PRELIMINARIES AND SOME DEFINITIONS

A. Paillier Cryptographic Algorithm There are various cryptographic techniques to support numeric-related operations (e.g., addition, multiplication, XOR) upon the encryption field. Paillier cryptosystem [41] is one of the most popular techniques that provides addition homomorphic, which means: if two integers a and b are encrypted with a same key k into two ciphertexts (be denoted as Ek(a) and Ek(b)), there exists an operation (refer to as “⊗”), such that Ek(a) ⊗ Ek(b) = Ek(a + b). Paillier cryptographic algorithm is composed of the following phases: key generation, encryption and decryption. Key generation. Two large and independent prime numbers p and q are randomly selected. Then we compute n = p · q and μ = λ −1 mod n, where λ is the least common multiple of p and q, and commonly λ = lcm(p−1, q−1). The public key (PK) is n, and the private key (SK) is (λ, μ). Encryption. Let m be the integer to be encrypted. Firstly, we select a random number r ∈ Z ∗ n2 , and then the ciphertext of m can be computed as follows: E(m; r) = (n + 1)m · r mod n 2 . Decryption. Let the ciphertext c = E(m; r). The plaintext m can be recovered as follows: m = (c · μ mod n 2 ) − 1 n · μ mod n. Paillier cryptosystem holds additive homomorphic in group Z + n , which corresponds to the multiplication operation in the encryption field in Zn2 . The following equation illustrates the homomorphic property of Paillier cryptosystem. E(m1; r1) · E(m2; r2) = (n + 1)m1 r1 n 1 · (n + 1)m2 r2 n 2 = (n + 1)m1+m2 (r1 · r2) n = E(m1 · m2; r m2 1 ). As the random number r does not affect the result of decryption in Paillier encryption, Eq. (4) can be seen as the product of m1 and m2 in the encryption field. In the rest of this paper, we use E(m, PK) to denote the encryption result of the plaintext m with PK, and D(X, SK) to denote the decryption result. We use capital letters like “X” to denote encrypted results (ciphertext), and lowercase letters like “x” to denote unencrypted results (plaintext). The random number r ∈ Z ∗ n is omitted in the discussion of our scheme. For number comparison, the sign of a plaintext number in Paillier cryptosystem is defined as follows: each participated plaintext integer x is assumed to be x < n/2. Then for clarity, the sign of x is defined to be positive if 0 < x < n/2, and the sign is defined to be negative if x > n/2. As a result, the arithmetic subtraction of arbitrary two integers (xi − xj ) will not exceed the threshold n/2 if xi > xj , and the subtraction will exceed n/2 if xi < xj .

B. Numeric-Related SQL Queries The Structured Query Language (SQL) is a specified purpose programming language, which is used to manage data in a relational database system, which has became a standard of the ANSI and ISO in 1986[42] and 1987[43] respectively. A query operation can request arbitrary data with a statement to describe the desired data. The requested data can be several columns of one or more tables in the database, and it can also be aggregated results from the original data (such as sum, average, count of the data.). To obtain the desired data, the query contains some statements to describe the requirement, e.g., some numeric-related .

<table>
<thead>
<tr>
<th>Column</th>
<th>Ti</th>
<th>Query Processes</th>
<th>Guessed Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T(1)</td>
<td>T(1) &gt; A1</td>
<td>T(2) &lt; T(0) &lt; T(3)</td>
</tr>
<tr>
<td></td>
<td>T(2)</td>
<td>T(1) &gt; A1</td>
<td>T(3)</td>
</tr>
<tr>
<td></td>
<td>T(3)</td>
<td>T(1) &gt; A1</td>
<td>T(2)</td>
</tr>
</tbody>
</table>

V. OUR PROPOSED TWO-CLOUD SCHEME In this section, we firstly give an overview of our proposed two-cloud scheme, and then present the detailed interaction protocols to realize range query with privacy preservation on outsourced encrypted database. A. Overview In our scheme, two clouds (refer to Cloud A and Cloud B, respectively) have been assigned distinct tasks in the database system: Cloud A provides the main storage service and stores the encrypted database. Meanwhile, Cloud B executes the main computation task, to figure out whether each numerical record satisfies the client’s query request with its own security key. With the assumption of no collusion between two clouds, the knowledge of application logic can be partitioned into two parts in our proposed scheme, where each one part is only known to one cloud. As we will analyze in this paper, one single part of knowledge cannot reveal privacy of the data and the query.

Based on the two-cloud architecture, our scheme provides an approach to query numeric-related data with privacy preservation. The client can retrieve the desired data from the cloud.

The proposed mechanism can preserve the privacy of data and query requests against each of the two clouds. Specifically, Cloud A only knows the query request type and the final indexes, but due to dummy items appending, Cloud
A cannot accurately understand the finally satisfied index set for each single request. Meanwhile, in order to prevent Cloud A from launching multiple specific-purpose query requests to deliberately to seek more knowledge about the data, we introduce a token based scheme, which can restrict the number of items and the range of columns that Cloud A can only process. For Cloud B, it knows the satisfied indexes of each single request, but after the proposed operations, it does not know the relationship of the corresponding items. Moreover, Cloud B can hardly distinguish whether two received columns are generated from one or more columns in the original database.

The basic idea to realize this type of complex query request is intuitive: Firstly, we run separately and independently the procedures in Section V-B and Section V-C1 to generate a index set for each simple condition (e.g. Ti1 > a1 or Ti2 < a2). Then, if the logic gate is “\lor”, we can compute the union of two index sets; Otherwise, for “\land”, we can compute the intersection of them. For instance, if S1, S2 and S3 are the index sets respectively for three simple conditions Ti1 > a1, Ti2 > a2, and Ti3 < a3. Then the final index S of the combination query request above is as follows:

VIII. CONCLUSION

In this paper, we presented a two-cloud architecture with a series of interaction protocols for outsourced database service, which ensures the privacy preservation of data contents, statistical properties and query pattern. At the same time, with the support of range queries, it not only protects the confidentiality of static data, but also addresses potential privacy leakage in statistical properties or after large number of query processes. Security analysis shows that our scheme can meet the privacy-preservation requirements. Furthermore, performance evaluation result shows that our proposed scheme is efficient. In our future work, we will consider to further enhance the security while ensuring practicality, and we will extend our proposed scheme to support more operations, such as “SUM/AVG”.

REFERENCES