IMPROVED INDEPENDENT ACCESS TO ENCRYPTED CLOUD DATABASES SUPPORTING MULTI-KEYWORD SYNONYM BASED QUERIES

Anitha P. B.[1], M. Nanthini[2]

[1,2]*Department of CSE, SSM College of Engineering, Anna university, (India)*

ABSTRACT

Placing critical data in the hands of a cloud provider should come with the guarantee of security and availability for data at rest, in motion, and in use. Several alternatives exist for storage services, while data confidentiality solutions for the database as a service paradigm are still immature. We propose a novel architecture that integrates cloud database services with data confidentiality and the possibility of executing concurrent operations on encrypted data. This is the first solution supporting geographically distributed clients to connect directly to an encrypted cloud database, and to execute concurrent and independent operations including those modifying the database structure.

The proposed architecture has the further advantage of eliminating intermediate proxies that limit the elasticity, availability, and scalability properties that are intrinsic in cloud-based solutions. The efficacy of the proposed architecture is evaluated through theoretical analyses and extensive experimental results based on a prototype implementation subject to the TPC-C standard benchmark for different numbers of clients and network latencies. We propose a semantic multi-keyword ranked search scheme over the encrypted cloud data, which simultaneously meets a set of strict privacy requirements. The proposed scheme could return not only the exact matching files, but also the files including the terms latent semantically associated to the query keyword.

Keywords: Cloud Computing, Cloud Databases, Database As A Service, Ranked Search, Depot, Encrypted Cloud, Homomorphic Encryption, SQL Queries

I INTRODUCTION

Cloud Databases provides a complete solution for customers demanding a high-performance, purpose-built infrastructure designed for relational databases backed and supported by engineers who specialize in MySQL workloads. A DBaaS promises to move much of the operational burden of provisioning, configuration, scaling, performance tuning, backup, privacy, and access control from the database users to the service operator, offering lower overall costs to users.

Database-as-a-service (DBaaS) is attractive for two reasons. First, due to economies of scale, the hardware and energy costs incurred by users are likely to be much lower when they are paying for a share of a service rather than running everything themselves. Second, the costs incurred in a well-designed DBaaS will be proportional to actual usage -this applies to both software licensing and administrative costs. The latter are often a significant expense because of the specialized expertise required to extract good performance from commodity DBMSs. By

centralizing and automating many database management tasks, a DBaaS can substantially reduce operational costs and perform well.

In this paper we define a new scheme named Latent Semantic Analysis (LSA)-based multi-keyword ranked search which supports multi-keyword latent semantic ranked search. By using LSA, the proposed scheme could return not only the exact matching files, but also the files including the terms latent semantically associated to the query keyword.

II RELATED WORKS

W. Jansen and T. Grance [1] presented a framework to realize highly customizable privacy-conscious composite services. Their approach can be summarized as follows: When a consumer provides data to a service provider, He/she wants to ensure that the information she provides will be used in a manner consistent with his/ her privacy preferences. To verify if this will be the case, the consumer requests a model of the service. The model summarizes the manner in which the composite service uses the consumer data.

P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin and M. Walfish [2] describe the design, implementation, and evaluation of Depot, a cloud storage system that minimizes trust assumptions. Depot does eliminate trust for updates: a client can always update any object for which it is authorized, and any subset of connected, correct clients can always share updates. Their evaluation suggests that the costs of these guarantees are modest and that Depot can tolerate faults and maintains good availability, latency, overhead, and staleness even when significant faults occur.

H. Hacigu¨mu¨ s¸, B. Iyer and S. Mehrotra [3] have developed and deployed a database service on the Internet, called NetDB2, which is in constant use. In a sense, a data management model supported by NetDB2 provides an effective mechanism for organizations to purchase data management as a service, thereby freeing them to concentrate on their core businesses.

C. Gentry [4] proposed a fully homomorphic encryption scheme -- i.e., a scheme that allows one to evaluate circuits over encrypted data without being able to decrypt. First, they provided a general result -- that, to construct an encryption scheme that permits evaluation of arbitrary circuits. Next, they describe a public key encryption scheme using ideal lattices that is almost bootstrappable.

H. Hacigu¨mu¨ s¸, B. Iyer, C. Li, and S. Mehrotra [5] Executing SQL over Encrypted Data in the Database-Service- Provider Model. It introduces several challenges, an important issue being data privacy. It is in this context that they specifically address the issue of data privacy.

III SYSTEM DESIGN

We propose a novel architecture that integrates cloud database services with data confidentiality and the possibility of executing concurrent operations on encrypted data.

This is the first solution supporting geographically distributed clients to connect directly to an encrypted cloud database, and to execute concurrent and independent operations including those modifying the database structure.

The system model can be considered as three entities, as depicted in Fig .1: the data owner, the data user and the cloud server. Data owner has a collection of data documents and a set of distinct keywords is extracted from the

data collection D. The data owner will firstly construct an encrypted searchable index I from the data collection D. Then, the data owner uploads both the encrypted index I and the encrypted data collection C to the cloud server. Data user provides t keywords for the cloud server. The cloud server only sends back top-l files that are most relevant to the search query.

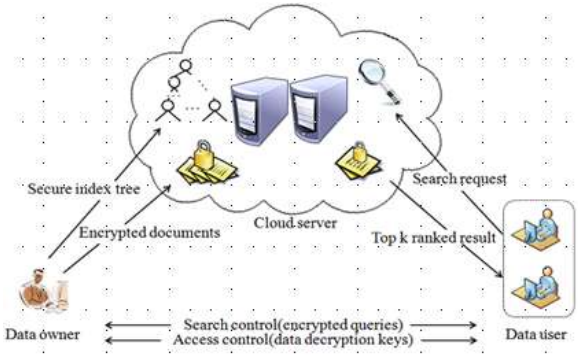

Figure 1- System Architecture

3.1 Advantages of Proposed System

- The proposed architecture does not require modifications to the cloud database, and it is immediately applicable to existing cloud DBaaS, such as the experimented PostgreSQL Plus Cloud Database, Windows Azure and Xeround.
- There are no theoretical and practical limits to extend our solution to other platforms and to include new encryption algorithm.
- It guarantees data confidentiality by allowing a cloud database server to execute concurrent SQL operations (not only read/write, but also modifications to the database structure) over encrypted data.
- It provides the same availability, elasticity, and scalability of the original cloud DBaaS because it does not require any intermediate server.

3.2 Methodology

A) Setup Phase

Here describe how to initialize a secure DBaaS architecture from a cloud database service acquired by a tenant from a cloud provider. Here assume that the DBA creates the metadata storage table that at the beginning contains just the database metadata, and not the table metadata.

B) Metadata Module

In this module, we develop Meta data. So our system does not require a trusted broker or a trusted proxy because tenant data and metadata stored by the cloud database are always encrypted. In this module, we design such as Tenant data, data structures, and metadata must be encrypted before exiting from the client.

C) Sequential SQL Operations

The first connection of the client with the cloud DBaaS is for authentication purposes. Secure DBaaS relies on standard authentication and authorization mechanisms pro-vided by the original DBMS server. After the authentication, a user interacts with the cloud database through the Secure DBaaS client.

D) Concurrent SQL Operations

The support to concurrent execution of SQL statements issued by multiple independent (and possibly geographically distributed) clients is one of the most important benefits of Secure DBaaS with respect to state-of-the-art solutions.

3.2.1 Multi-Keyword Ranked Search

The existing systems like exact or fuzzy keyword search, supports only single keyword search. These schemes doesn't retrieve the relevant data to users query therefore multi-keyword ranked search over encrypted cloud data remains a very challenging problem. To meet this challenge of effective search system, an effective and flexible searchable scheme is proposed that supports multi-keyword ranked search. To address multi-keyword search and result ranking, Vector Space Model (VSM) is used to build document index, that is to say, each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. A new vector is also generated in the query phase. The vector has the same dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure can be used to compute similarity of one document to the search query.

The equation for finding the F- Measure is:

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Multi-keyword Ranked Search: It supports both multi-keyword query and support result ranking.

Privacy-Preserving: Our scheme is designed to meet the privacy requirement and prevent the cloud server from learning additional information from index and trapdoor.

Performance Analysis

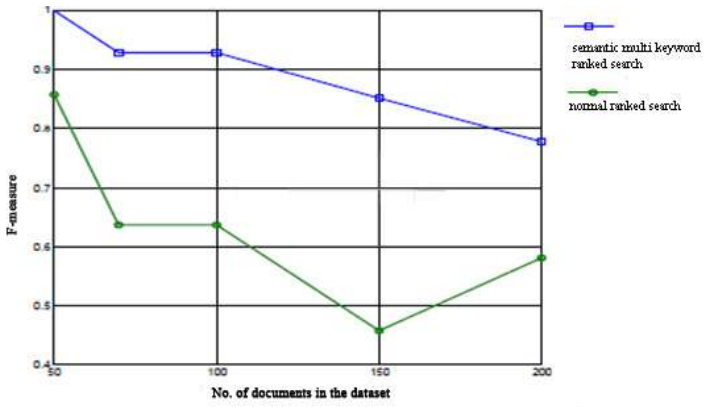

Fig. 2. Comparison of two searches

1) *Index Confidentiality.* The TF values of keywords are stored in the index. Thus, the index stored in the cloud server needs to be encrypted;

2) *Trapdoor Unlinkability.* The cloud server should not be able to deduce relationship between trapdoors.

3) *Keyword Privacy.* The cloud server could not discern the keyword in query, index by analyzing the statistical information like term frequency.

IV CONCLUSION

We propose an innovative architecture that guarantees confidentiality of data stored in public cloud databases. Unlike state-of-the-art approaches, our solution does not rely on an intermediate proxy that we consider a single point of failure and a bottleneck limiting availability and scalability of typical cloud database services. A large part of the research includes solutions to support concurrent SQL operations (including statements modifying the database structure) on encrypted data issued by heterogeneous and possibly geographically dispersed clients. The proposed architecture does not require modifications to the cloud database, and it is immediately applicable to existing cloud DBaaS, such as the experimented PostgreSQL Plus Cloud Database [23], Windows Azure [24], and Xeround [22]. There are no theoretical and practical limits to extend our solution to other platforms and to include new encryption algorithms. It is worth observing that experimental results based on the TPC-C standard benchmark show that the performance impact of data encryption on response time becomes negligible because it is masked by network latencies that are typical of cloud scenarios. These performance results open the space to future improvements that we are investigating.

REFERENCES

[1] A. Arasu, S. Blanas, K. Eguro, R. Kaushik, D. Kossmann, R. Ramamurthy, and R. Venkatesan, January 2013, "Orthogonal security with cipherbase," in *Proc. of the 6th Conference on Innovative Data Systems Research*.

[2] A. Boldyreva, N. Chenette, and A. O'Neill, August 2011, "Order-preserving encryption revisited: Improved security analysis and alternative solutions," in *Proc. of the Advances in Cryptology – CRYPTO 2011, Springer*.

[3] "Amazon Elastic Compute Cloud (Amazon Ec2)," Amazon Web Services (AWS), http://aws.amazon.com/ec2.

[4] J. L. Dautrich Jr and C. V. Ravishankar, March 2013, "Compromising privacy in precise query protocols," in *Proc. of the 16th ACM International Conference on Extending Database Technology*.

[5] L. Ferretti, M. Colajanni and M. Marchetti, December 2012, "Supporting Security and Consistency for Cloud Database,", in *Proc. Fourth Int'l Symp. Cyberspace Safety and Security*.

[6] Luca Ferretti, Michele Colajanni, and Mirco Marchetti, February 2014, "Distributed, Concurrent, and Independent Access to Encrypted Cloud Databases", VOL. 25, NO. 2.

[7] M. Hadavi, E. Damiani, R. Jalili, S. Cimato, and Z. Ganjei, September 2013, "AS5:A secure searchable secret sharing scheme for privacy preserving database outsourcing," in *Proc. of the 5th International Workshop on Autonomous and Spontaneous Security. Springer,*.

[8] "Postgres Plus Cloud Database," EnterpriseDB, http://enterprisedb.com/cloud-database.

[9] M. Yabandeh and D. G´omez Ferro, April 2012 , "A critique of snapshot isolation," in *Proc. of the 7th ACM european conference on Computer Systems,*.

[10] "Xeround: The Cloud Database," Xeround.

[11] Zhangjie Fu, Xingming Sun, Nigel Linge and Lu Zhou, "Achieving Effective Cloud Search Services: Multikeyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query", *IEEE Transactions on Consumer Electronics*, Vol. 60, No. 1, February 2014

BIOGRAPHY

[1] **Anitha P. B**. received her B.E. in Computer Science and Engineering From Anna University, India in 2007. She is pursuing M. E. in Computer Science and Engineering at SSM College of Engineering, Erode, affiliated to Anna University, India. Her area of research interest is in Cloud Computing.

[2] **M. Nanthini** is presently working as Assistant Professor in Dept. of Computer Science and Engineering, SSM College of Engineering, Erode. Her area of research interest is in Cloud Computing.