THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Query Optimization in Encrypted Databases for Providing DaaS

Sunil B. Mane
Research Scholar, University of Pune, Pune, India
Pradeep K. Sinha
Senior Director, Corporate Strategy and R&D, C-DAC, Pune, India

Abstract:
Nowadays, data is getting generated tremendously in fields like consumer or end user generated content, financial service industry, broadcast media and service industry, manufacturing, transportation, retail, wholesale and utility industry, government and healthcare industry, oil and gas industry. Digital revolution has transformed present society from an industrial society in to an information society. Maintaining database (at data owner side) requires purchasing the required hardware, implementing database products, establishing network connectivity, and hiring the professionals etc. This traditional solution, becoming increasingly expensive and impractical as the database systems and problems become complicated and more larger. Security and cost concerns force many companies to outsource their databases. Data outsourcing has various benefits like decrease the cost 5-10 times from initial calculated cost, increase database availability.

Key words: *Encrypted Databases, Query Processing, Database outsourcing, Query optimization*

1. Introduction

Database as a Service (DaaS) enables client to use database without maintaining it at their site. In DaaS model data owner stores their confidential data at potentially insecure service provider. Hence it opens new security challenges such as data confidentiality, data security, user privacy, authentication etc. Data privacy is addressed by using encrypting the database which is stored at external service provider site. Different encryption techniques can be used to fulfill this requirement.
This paper explores techniques to store data in encrypted form and execute queries on it. Optimization of such encrypted queries to decrease execution time is also addressed in this paper.

2. Database as a Service (DaaS)

Database as a Service is a structural and functional approach enabling IT providers to deliver database operations as a service to one or more customers. Data owners have to encrypt their data before outsourcing it to external servers in order to maintain data confidentiality. At the same time clients must still be able to execute their queries over encrypted data. In the DaaS model, data owners store their data to potentially un-trusted service providers. Outsourcing the databases emerged as an affordable data management model for data owners with limited capabilities to host and maintain large in-house data centers of potentially significant resource footprint. Since most of the data management and query execution load is incurred by the external service provider and not by the data owner, this is intuitively advantageous and significantly more affordable for customers with less experience, resources or trained man-power.

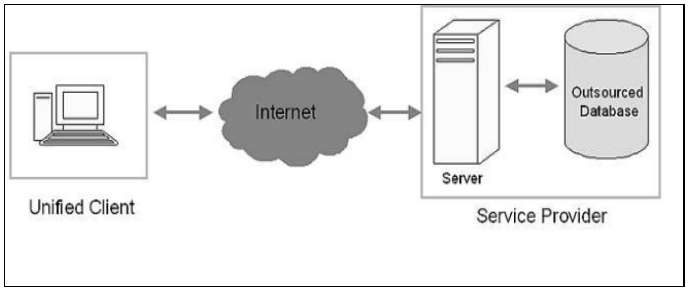

Figure 1: Database as a Service Model

- Challenges
  For successful database outsourcing, there should be appropriate solution for significant challenges which affect overall performance, usability, scalability. Such problems are listed below; they are mainly related to security of client data.

3. Problem Statement

For assuring data confidentiality and data security data owner relies on different encryption schemes. Data encryption deals with use of encryption schemes to transform plaintext data into unreadable form. First objective of paper is data stored is in encrypted form so that service provider could not tamper it. Encryption technique should not be vulnerable to any cryptographic attack.
Second objective is minimizing execution time of encrypted SQL queries. Hence choosing a good query execution plan amongst various different plans plays important role in query optimization.

4. Proposed System Design

For achieving query optimization in encrypted databases we have built a system with following components in it.

- **Client:** Client is any computer that wants results from server. Client sends its plane query to server. We have assumed three types of clients as beginner, proficient and expert and created GUI according to it. After all background processes like parsing, encryption/decryption, database connectivity, sql query execution client receives required result.
- **Data Owner Server:** This Server is having all the rights on the database since it owns the data. This server can be in same LAN as that of client or it can be on other remote network. Data owner server mainly does parsing of query and encryption of that query before sending it to the third party database service provider. Connection between data owner server and third party service provider is done with help of either oracle 10g database or mysql database connectivity. Data owner server will then receive encrypted results from third party database service provider and then data owner server will decrypt this data according to previous encryption strategies.
- **Encryption/Decryption module:** We have kept this module separate to make encryption and decryption of the query as well as results obtained from third party service provider make as much time as we want. Various encryption algorithms like AES, Triple DES, Blowfish have been implemented for testing the working model of the proposed scheme. These techniques have used since there is no any cryptanalytic attacks reported so far.
- **Third Party Service Provider:** Concept of this project that is query optimization in encrypted databases for providing database as a service is implemented using third party service provider. It is a server located in the premises of the service provider other than client or server. All the data stored in database of third party service provider is in encrypted format. Third party service provider accepts query requests from data owner, clients of data owner and updates its database accordingly or returns requested query results by data owner. As third party service provider has completely encrypted database data confidentiality is achieved. And it increases database availability in much cheaper price.

In this proposed design as shown in Figure 2, data owner encrypts data and stores it at service provider side which is managed by an application. The encrypted database is augmented with some additional information such as index that helps in some amount of query processing to occur at the external server without violating data privacy. The data owner also maintains metadata which is used by a query translator for translating the user query into different portions, i.e., a query over encrypted data, for execution on the service provider side, and a query over decrypted data. The service provider generates encrypted intermediate results set, which is transferred to the client and stored as temporary results. The client application has a query executor module that decrypts the temporary results and performs the query execution over decrypted data and generates actual final results which is an answer to the query, for display to the requested client.
In this environment, the data owner maintains the needed encryption key(s), and the data is encrypted by him/her before it is sent to the service provider for insertion in the encrypted database. The data is always remaining encrypted when it is stored on or processed by the service provider.

5. Implementation

We have implemented the above proposed model using java socket programming. Database connectivity is created using jdbc-odbc driver with oracle as database. Various components of the proposed model such as Graphical user interface, Query parsing and processing, Encryption and decryption mechanism, and Query optimization are implemented in java language.
Following issues considered and addressed during implementation.

5.1. Query Optimization

Query optimization is very important part in query processing over encrypted database systems. The overall cost of the information system is cost of DBMS and cost of the user working with system. Query optimization module has to choose among various existing execution plans to resolve query. For minimizing response time of query requested by client we have chosen various paths. Query is a language expression that describes the data to be retrieved from the database. In the context of query optimization it is often assumed that queries are expressed in a content based manner, giving the optimizer sufficient choices among alternative evaluation procedures. Query optimization tries to minimize response time for a given query.

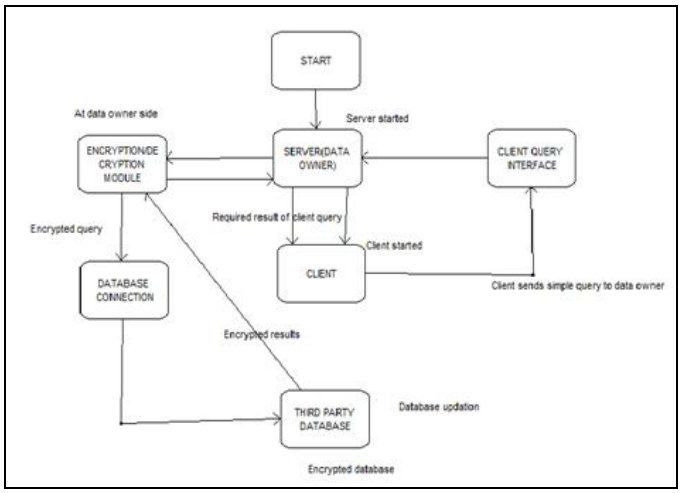

Figure 2: Proposed System Design

Following issues considered and addressed during implementation.

5.2. Query Optimization

Query optimization is very important part in query processing over encrypted database systems. The overall cost of the information system is cost of DBMS and cost of the user working with system. Query optimization module has to choose among various existing execution plans to resolve query. For minimizing response time of query requested by client we have chosen various paths. Query is a language expression that describes the data to be retrieved from the database. In the context of query optimization it is often assumed that queries are expressed in a content based manner, giving the optimizer sufficient choices among alternative evaluation procedures. Query optimization tries to minimize response time for a given query.

5.3. Client Usability

Client must be provided with sufficient software, hardware and network resource to use its data from the databases those are outsourced. There should be efficient mechanisms for the client to create, update and access the outsourced data. User-friendly Java GUI is provided to client with instruction for building query and handling its own private data.

5.4. Data Confidentiality

Data confidentiality is main challenge for data outsourcing. Data confidentiality is the protection of private information from leaks when it is stored, or transmitted across insecure networks such as the Internet. The data and information must be encrypted by using strong encryption algorithm that decreases the probability of data and information to be compromised during the transition of data and information between the client and server, even if the server is malicious. We used a symmetric algorithms and symmetric keys like AES, DES etc. this encryption/decryption algorithm is performed**.** Twice to store more secured data at database service provider.

5.5. User Privacy

The database outsourcing implementation, is the malicious server shouldn't data access privacy is simply at be able to analyze and learn anything about the client's query patterns by performing such as query statistical attacks . To avoid this server get encrypted query so it’s difficult for un-trusted server to obtain query pattern and client’s data.

5.6. Query Correctness

Query Correctness is a technique or mechanism makes the client able to verify the integrity and completeness of the query result. Query correctness aims to insure the outsourced data could not temper by the un-trusted server (Database Service Provider), and to insure the query executed successfully and the result does not truncated by the server. We build SQL PARSER component for handling this problem.

6. Conclusion

This paper provides a way to secure and efficient communication between data owner and external service provider in a Database As a Service (DaaS) model. Proposed scheme demonstrates the idea of query processing on encrypted databases which uses query parsing, query optimization using efficient query execution plans to answer client query in optimum cost. The main goal of design of architecture is to redistribute the functionality of outsourced database and trusted server. Proposed architecture of the model supports three main security issues in the outsourced database that is query correctness, data confidentiality, and user privacy.

7. References

1. Gultekin Ozsoyoglu, David A. Singer and Sun S. Chung "Anti-Tamper Databases: Querying Encrypted Databases," IFIP (International Federation for Information Processing), Volume 142, pp. 133-146, 2004.
2. Ahmed M.A. Al thneibat, BahaaEldin M. Hasan ,Abd El Fatah .A. Hegazy and NermineHamza, "Secure Outsourced Database Architecture," IJCSNS International Journal of Computer Science and Network Security, Vol.10 No.5, May 2010.
3. Luc Bouganim, Yanli Guo**,** Inria Rocquencourt Le Chesnay and France "Database Encryption," Journal Proceedings of the VLDB Endowment VLDB ,Vol. 3 Issue 1-2, pp. 25-35, September 2010.
4. Einar Mykletun and Gene "Aggregation Queries in the Database-As-a-Service Model," Proceeding DBSEC'06 Proceedings of the 20th IFIP WG 11.3 working conference on Data and Applications Security, pp. 89-103, 2006.
5. BogdanCarbunar and Radu Sion2, "Joining privately on outsource of data," Proceeding SDM'10 Proceedings of the 7th VLDB conference on Secure data management, pp. 70-86, 2010.
6. Matthias Jarke, Jurgen Koch "Query optimization in database systems," Journal ACM Computing Surveys (CSUR) Surveys Homepage archive, Vol. 16 Issue 2, pp. 111-152, June 1984