Executing SQL over Encrypted Data in the Database-Service-Provider Model

Hakan Hacigumus, Bala Iyer, Chen Li and Sharad Mehrotra

Abhishek Srivastava
Student ID: 861307778

March 2, 2017
CS 236, Winter 2017

The problem:
The paper discusses a "Database as a Service" model where applications or users can store their data and perform any operations on them through Internet. But two kind of Data Security problem arises for the data stored, protection from outsiders and protection from service provider. Paper mostly focuses on the second issue.

The contribution:
The authors proposes an architecture where *client* sends queries, also stored some metadata of the database and *server* stores data in encrypted format. There are multiple ways of handling the execution of queries. Server maintains additional information about encrypted database to provide execution of queries on the server side. Queries sent are encrypted using algebraic framework.

The method:
The Authors presents mechanism to encrypt the data and store them at server side. These mechanism includes: *Partition Functions* which partitions the domain of attribute values in some specific way, *Identification Functions* which assigns an attribute value to some partition value from previous step, *Mapping Functions* maps a value in domain of attribute to identifier partition to which the value belongs either preserving ordering or randomly, *Storing Encrypted Data* provides way of storing the tuple, additional column *etuple* is added whose value is encryption of all the attributes values and *Decryption Functions* which tries to decrypt the data stored which can be applied to obtain original values.

Authors proposes mapping conditions which defines ways to translate a specific client-side query to encrypted server-side query. Different grammar rules on query conditions were:

- Attribute operation Value
- Attribute1 operation Attribute2
- (Condition1 $\vee$ Condition2) OR (Condition1 $\wedge$ Condition2) OR ($\neg$ Condition)

Here operations which are allowed in the queries are $<$,$>$,$=$,$\leq$ and $\geq$. These operations are handled in their own way by appropriately mapping between the unencrypted data and encrypted data.

Author proposes to divide the operator computation between client and server. At server superset of values will be generated using some operations and upon those superset values operations will be performed at client-side to give the result of the operator. Author discusses following operators and way to divide their computations: *Selection*, *Join*, *Grouping & Aggregation*, *Sorting*, *Duplicate-Elimination*, *Set-Difference*, *Union* and *Projection* Operators. Author suggested to use pull some heuristics strategy to minimize the computation cost at client end.

Comments:
The paper proposes a novel method to do encrypted query on sotred encrypted data on server. Different experiments are performed to evaluated query execution of different kind of queries with different operations.

However, Some drawbacks of the proposed methods were:

- Encrypting large amount of tuples with large number of attributes can be a costly operations.
- Pipelining mechanism is not supported using this model because final results are being calculated at client-end.