International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Secure Skyline Queries over the Encrypted Data

Nimmy K R

Student, Dept. of Computer Science Engineering, Thejus Engineering college, Kerala, India

---

Abstract - *Outsourcing data to cloud server provides better usability of data. But to preserve the security and privacy these outsourced data need to be protected from the cloud server and other unauthorized users. The encryption of data can protect the data. For query processing, the skyline query is used and it is particularly important for multi-criteria decision making. As a part of the work, the work can be categorized into four parts mainly. And they are Data Transformation, where data to be transformed into different form by performing RSA encryption and the noise appending process where random noise values are generated and appended to the data. Then comes the Auditing where it checks whether there is an occurrence of attack or not. Skyline Computations specifies the Secure Skyline computations over the encrypted data. And the Query retrieval which is to submit the query and should return the efficient result by performing various calculations such as skyline computations by TPA(Third Party Authority).*

Key Words: **Skyline Query**, **Multi-criteria decision making**, **RSA Encryption, Auditing, Noise Appending, Skyline Computations, Query Retrieval etc** ...

1.INTRODUCTION

As an arising trend, most of the users uploads their datasets to the external clouds like Google cloud, AWS, IBM cloud etc...This outsourcing of information and computation to cloud server provides many benefits like value edges, time edges etc. However privacy problems are to be thought of. To preserve the security and privacy, these outsourced information have to be compelled to be encrypted. Figure 1 illustrates the case of secure query processing over encrypted information within the cloud. The data owner outsources encrypted data to the cloud server. The cloud server processes queries from the user on the encrypted data and returns the query result to the user. Throughout the query processing, the cloud server shouldn't gain any data concerning the informations, information patterns, query, and query result[1].

Figure 1: Secure similarity queries

1.1 Secure Skyline Query

Here introduces the secure skyline queries on encrypted information, that is another kind of similarity search where similarity queries are queries that retrieve objects that are just like the query object. There are mainly three types of similarity queries and they are as follows[2].

- Similarity range
- Similarity nearest-neighbour
- Similarity join

And these skyline queries are for multi-criteria decision making. That is , if over one attributes are there and we have to be compelled to get results satisfying all the attribute criteria, we will directly use skyline queries.

1.2 Organization

The rest of the paper is organized as follows. Section two presents the related works. Section three introduces the background definitions and the problem definitions. The proposed methodology and its algorithms are introduced in section four. Finally in section five, concludes the paper.

2. RELEATED WORKS

S. B¨orzs¨onyi et al.[4] extend database systems by a Skyline operation. This operation filters out a collection of attention grabbing points from a probably massive set of data points. A point is attention grabbing if it is not dominated by any other point. [5] has the challenge that the CSPs can't be absolutely trusted, which gives fake query results for various bad cases. [6] introduces the fully homomorphic encryption schemes that performs arbitrary computations on encrypted information however that is not sensible. [7] proposed a new encryption method called asymmetric scalar-product-preserving encryption. But the drawback is it is less secure since all users know the private key. [8] proposes the work on secure kNN queries over the encrypted database outsourced to a cloud. A user issues an encrypted query record to the cloud, and the cloud returns the k similar results to the user. but such methods has the drawback that relative weights should be known earlier only then the single similarity metric can be computed. [1] proposes efficient skyline querying but the security is restricted since the keys used for encryptions are stored at c2 cloud server which is not secure. Decryption can be done after obtaining these keys from c2 when there is a communication possible with

c1 and c2. There are many other works which is based on kNN queries which is in the secure multi-party computation (SMC) setting [9] where the data is given between two parties who want to cooperatively find the answers without revealing to each other their sensitive data. [10] proposes private information retrieval (PIR) setting where the query is private but the information is public, which are different from our settings.

3. BACKGROUND DEFINITIONS AND PROBLEM DEFINITION

In this section, first illustrates the various notations and then the various definitions used in this paper and then discusses the problem definition.

3.1 Notations

For references, a summary of notations is given in Table 1.

Table 1: Notations and its details

NOTATION	DEFINITIONS
P	Dataset of n tuples
q	Query tuple of client
$p_i[j]$	The j^{th} attribute of p_i
n	Number of points in P
m	Number of dimensions
Pk	Public key
Sk	Secret key

3.2 Definitions

Skyline : Given a dataset P = {p1,p2.....pn} in m dimensional space. Let Pa and pb denotes two different points in P and if pa dominates pb, then pa<=pb and for at least one j, pa[j]<pb[j] where j lies between 1 and m. The skyline points are the data points that are not dominated by any other data points in P.

Dynamic Skyline Query : Given a dataset P = {p1,p2.....pn} and a query point q in m dimensional space. Let pa and pb denotes two different points in P, we say that pa dynamically dominates pb with regard to the query point q, denoted by pa<pb, if for all j, |pa[j]-q[j]|<=|pb[j]-q[j]| and for at least one j, |pa[j]-q[j]|<|pb[j]-q[j]|,where pi[j] is the jth dimension of pi and 1<=j<=m. The skyline points are the data points that are not dominated by any other data points in P[1].

3.3 Problem Definitions

Designing a skyline protocol over the encrypted data in cloud environment for multi-criteria decision making without learning any information about data.

4. PROPOSED METHODOGY

Due to wide variety of methods in query processing ,some of the feasible ideas are integrated into a protocol along with some extra ideas for preserving better security. The project work is aimed to provide a comprehensive solution which combines privacy of data, proper storage of data and efficient query retrieval. Secure skyline queries on encrypted data, which is another type of similarity search where similarity search is nothing but queries that retrieve objects that are similar to query object. Skyline queries are important for multi-criteria decision making.

4.1 System Architecture

The Figure 2 illustrates the abstract system architecture of the secure skyline protocol. Skyline protocol illustrates the concept of skyline query which is useful for multi-criteria decision making. And the skyline queries retrieves efficient results on each search process. The architecture consists of mainly four modules and they are Data Owner, Client, TPA, Cloud servers such as C1 and C2. Data Owner is the one who encrypts and stores the data. TPA performs skyline computations as well as noise appending and stores several information on C1 and C2. The C1 contains noise added encrypted data and C2 contains the noise details. And TPA returns the skyline computation results after both skyline and noise appending computations to the client who submits the query.

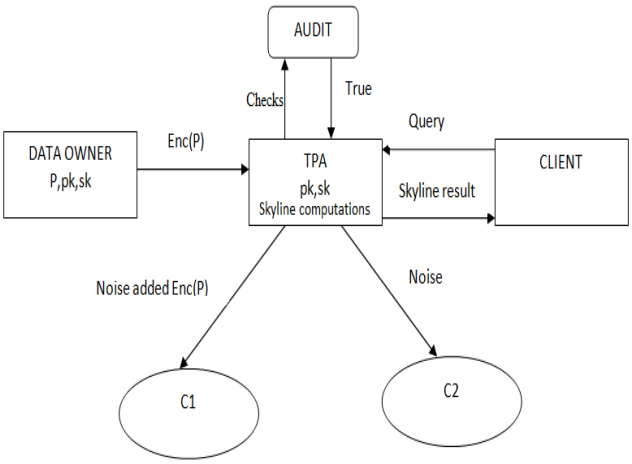

Figure 2: System Architecture

4.2 System Functionalities

A. Encryption

Initially the data owner loads the dataset and encrypts the dataset using the RSA encryption algorithm. Then loads the encrypted data to cloud.

B. Auditing
Auditing checks whether there is an occurrence of attack or not. If data undergoes some attack, the auditing result will return false. Else if no attack is detected during the auditing, the auditing results return true. Skyline computations are performed only when the auditing result is true. That is no data is modified.

C. Noise Appending

Randomly generated noise values are appended to the data and appended noise values and its index locations where these noise values are appended are stored in the cloud server C2. These noise values are added to make data more secure, that is some sort of data transformation are performed so that no one can obtain the actual data. So the data privacy is strictly followed.

D. Skyline computations

The skyline computations are performed by the Third party Authority. And the skyline computations are performed based on the skyline computation algorithm which is discussed in section 4.3.

4.3 Algorithms

The main algorithms used here are RSA algorithm and Skyline computation algorithm.

A. RSA Algorithm

RSA algorithm is asymmetric cryptography algorithm. Asymmetric trully means that it works on two totally different keys and they are Public Key and Private Key. As the name describes that the Public Key is shared among everyone and Private key is not shares among everyone that is it was kept private.
An example of asymmetric cryptography :

1. A user sends its public key to the server and requests for some information.
2. The server encrypts the information using client's public key and sends the encrypted information.
3. User receives this information and decrypts it.

Since this is asymmetric, nobody else except browser, the user can decrypt the data even if a third party has public key of browser[3].

B. Skyline Computation Algorithm

Input : A dataset and a query
Output: Skyline query results
Steps :-

1. Compute values for each n tuples and m attributes using the equation

$$t_i[j] = p_i[j] - q[j] + q[j]$$

2. Find the attribute sum
3. Choose the tuple with smallest sum S (ti) as a skyline
4. Add corresponding tuple to the skyline pool
5. Delete those tuples dominated by min tuple
6. Delete min tuple from dataset
7. Return skyline pool

5. CONCLUSIONS

The paper implements a method to achieve the following objectives, skyline queries for multi-criteria decision making, should obtain efficient results on each search, should be useful when a single aggregated distance metric with all dimensions is hard to define. So to achieve all these objectives, the paper proposes the Skyline queries over the encrypted data which particularly important for multi-criteria decision making. As a part of the work, the work can be categorized into four parts mainly. And they are Data Transformation, where data to be transformed into different form by performing RSA encryption and the noise appending process where random noise values are generated and appended to the data. And noise added data and noise details are stored in C1 and C2 respectively. Then comes the Auditing where it checks whether there is an occurrence of attack or not. Skyline Computations specifies the Secure Skyline computations over the encrypted data. And the Query retrieval which is to submit the query and should return the efficient result by performing various calculations such as skyline computations by TPA(Third Party Authority).

REFERENCES

[1] Liu, Jinfei, et al. "Secure and Efficient Skyline Queries on Encrypted Data." IEEE Transactions on Knowledge and Data Engineering (2018).

[2] https://en.wikipedia.org/wiki/Similarity_search

[3] https://en.wikipedia.org/wiki/RSA_(cryptosystem)

[4] Borzsony, Stephan, Donald Kossmann, and Konrad Stocker. "The skyline operator." Proceedings 17th international conference on data engineering. IEEE, 2001.

[5] Chen, Wenxin, et al. "Secure outsourced skyline query processing via untrusted cloud service providers." IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE, 2016.

[6] Gentry, Craig. "Fully homomorphic encryption using ideal lattices." Stoc. Vol. 9. No. 2009. 2009.

[7] Wong, Wai Kit, et al. "Secure kNN computation on encrypted databases." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

[8] Elmehdwi, Yousef, Bharath K. Samanthula, and Wei Jiang. "Secure k-nearest neighbor query over encrypted data in outsourced environments." 2014 IEEE 30th International Conference on Data Engineering. IEEE, 2014.

[9] Y. Qi and M. J. Atallah. E_cient privacy-preserving k-nearest neighbor search. In ICDCS 2008.

[10] S. Papadopoulos, S. Bakiras, and D. Papadias. Nearest neighbor search with strong location privacy. PVLDB, 2010.

BIOGRAPHIES

Nimmy K R is a master student in APJ Abdul Kalam Technological University, Kerala. Her research interest includes computer security, information security and Cloud security.