International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 24th Feb 2018. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-02

Title: A REVIEW ON SECURE AND EFFICIENT SKYLINE QUERIES ON ENCRYPTED DATA

Volume 08, Issue 02, Pages: 133–137.

Paper Authors

S.VENKATESH , MR.Y.SRINIVASA RAO

Vignan's Lara Institute of Technology & Science

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

A REVIEW ON SECURE AND EFFICIENT SKYLINE QUERIES ON ENCRYPTED DATA

S.VENKATESH [1], MR.Y.SRINIVASA RAO[2]

Assistant Professor[1], Department of M.C.A ,Vignan's Lara Institute of Technology & Science

M.C.A Student[2],Department of M.C.A ,Vignan's Lara Institute of Technology & Science

Abstract:

Outsourcing data and computation to cloud server provides a cost-effective way to support large scale data storage and query processing. However, due to security and privacy concerns, sensitive data (e.g., medical records) need to be protected from the cloud server and other unauthorized users. One approach is to outsource encrypted data to the cloud server and have the cloud server perform query processing on the encrypted data only. It remains a challenging task to support various queries over encrypted data in a secure and efficient way such that the cloud server does not gain any knowledge about the data, query, and query result. In this paper, we study the problem of secure skyline queries over encrypted data. The skyline query is particularly important for multi-criteria decision making but also presents significant challenges due to its complex computations. We propose a fully secure skyline query protocol on data encrypted using semantically-secure encryption. As a key subroutine, we present a new secure dominance protocol, which can be also used as a building block for other queries. Furthermore, we demonstrate two optimizations, data partitioning and lazy merging, to further reduce the computation load. Finally, we provide both serial and parallelized implementations and empirically study the protocols in terms of efficiency and scalability under different parameter settings, verifying the feasibility of our proposed solutions

Introduction

As an emerging computing paradigm, cloud computing attracts increasing attention from both research and industry communities. Outsourcing data and computation to cloud server provides a cost- effective way to support large scale data storage and query processing. However, due to security and privacy concerns, sensitive data need to be protected from the cloud server as well as other unauthorized users.A common approach to protect the confidentiality of out- sourced data is to encrypt the data . To protect the confidentiality of the query from cloud server, authorized clients also send encrypted queries to the cloud server. Figure1illustrates our problem scenarios of secure query processing over encrypted data in the cloud. The data owner outsources encrypted data to the cloud server. The cloud server processes encrypted queries

from the client on the encrypted data and returns the query result to the client. During the query processing, the cloud server should not gain any knowledge about the data, data patterns, query, and query result.

Fully homomorphic encryption schemes ensure strong security while enabling arbitrary computations on the encrypted data. However, the computation cost is prohibitive in practice. Trusted hardware such as Intel's Software Guard Extensions (SGX) brings a promising alternative, but still has limitations in its security guarantees. Many techniques have been proposed to support specific queries or computations on encrypted data with varying degrees of security guarantee and efficiency (e.g., by weaker encryptions). Focusing on similarity search, secure k-nearest neighbor (kNN) queries, which return k most similar (closest) records given a query record, have been extensively studied. In this paper, we focus on the problem of secure skyline queries on encrypted data, another type of similarity search important for multi-criteria decision making. The skyline or Pareto of a multi-dimensional dataset given a query point consists of the data points that are not dominated by other points. A data point dominates another if it is closer to the query point in at least one dimension and at least as close to the query point in every other dimension. The skyline query is particularly useful for selecting similar (or best) records when a single aggregated distance metric with all dimensions is hard to define. The assumption of kNN queries is that the relative weights of the attributes are known in advance, so that a single similarity metric can be computed between a pair of records aggregating the similarity between all attribute pairs. However, this assumption does not always hold in practical applications. In many scenarios, it is desirable to retrieve similar records considering all possible relative weights of the attributes (e.g., considering only one attribute, or an arbitrary combination of attributes), which is essentially the skyline or the "pareto-similar" records.

Motivating Example. Consider a hospital who wishes to out- source its electronic health records to the cloud and the data is encrypted to ensure data confidentiality. Let P denote a sample heart disease dataset with attributes ID, age, trestbps (resting blood pressure). We sampled four patient records p1,...,p4 from the heart disease dataset of UCI machine learning repository as shown in Table 1(a) and Figure 2. Consider a physician who is treating a heart disease patient q = (41,125) and wishes to retrieve similar patients in order to enhance and personalize the treatment for patient q. While it is unclear how to define the attribute weights for kNN queries (p1 is the nearest if only age is considered while p2,p3 are the nearest if only trestbps is considered), skyline provides all pareto similar records that are not dominated by any other records. Skyline includes all possible 1NN results by considering all possible relative attribute weights, and hence can serve as a filter for users. Given the query q, we can map the data points to a

new space with q as the origin and the distance to q as the mapping function. The mapped records ti[j] = |pi[j]−q[j]|+q[j] on each dimension j are shown in Table 1(b) and also in Figure 2. It is easy to see that t1 and t2 are skyline in the mapped space, which means p1 and p2 are skyline with respect to query q. Our goal is for the cloud server to compute the skyline query given q on the encrypted data without revealing the data, the query q, the final result set {p1,p2}, as well as any intermediate result (e.g., t2 dominates t4) to the cloud. We note that skyline computation (with query point at the origin) is a special case of skyline queries.

Existing system:

A common approach to protect the confidentiality of outsourced data is to encrypt the data. To protect the confidentiality of the query from cloud server, authorized clients also send encrypted queries to the cloud server. Illustrates our problem scenario of secure query processing over encrypted data in the cloud. The data owner outsources encrypted data to the cloud server. The cloud server processes encrypted queries from the client on the encrypted data and returns the query result to the client. During the query processing, the cloud server should not gain any knowledge about the data, data patterns, query, and query result.Fully homomorphic encryption schemes ensure strong security while enabling arbitrary computations on the encrypted data. However, the computation cost is prohibitive in practice.Trusted hardware such as Intel's Software Guard Extensions(SGX) brings a promising alternative, but still has limitations inits security guarantees. Many techniques have been proposed to support specific queries or computations on encrypted data with varying degrees of security guarantee and efficiency (e.g., by weaker encryptions).

Proposed system:

In this, we focus on the problem of secure skyline queries on encrypted data, another type of similarity search important for multi-criteria decision making. The skyline or Pareto of a multi-dimensional dataset given a query point consists of the data points that are not dominated by other points. A data point dominates another if it is closer to the query point in at least one dimension and at least as close to the query point in every other Dimension. The skyline query is particularly useful for selecting similar (or best) records when a single aggregated distance metric with all dimensions is hard to define. The assumption of kNN queries is that the relative weights of the attributes are known in advance, so that a single similarity metric can be computed between a pair of records aggregating the similarity between all attribute pairs. However, this assumption does not always hold in practical applications. In many scenarios, it is desirable to retrieve similar records considering all possible relative weights of the attributes (e.g., considering only one attribute, or an arbitrary combination of attributes), which is essentially the skyline or the "pare to-similar" records.

Modules:

Skyline Computation Illustrate an iterative skyline computation algorithm which will be

used as the basis of our secure skyline protocol. We note that this is not the most efficient algorithm to compute skyline for plaintext compared to the divideand-conquer algorithm. secure skyline protocol based on this algorithm for two reasons: 1) the divideand-conquer approach is less suitable if not impossible for a secure implementation compared to the iterative approach, 2) the performance of the divide-and-conquer algorithm deteriorate with the "curse of dimensionality".

Secure Dominance Protocol

A fully secure dominance protocol, which can be used as a building block for skyline queries as well as other queries, e.g., reverse skyline queries and k-skyband queries

Secure Skyline Protocol

The basic protocol clearly reveals several information to C1 and C2 as follows. When selecting the skyline tuple with minimum attribute sum, C1 and C2 know which tuples are skyline points, which violates our result privacy requirement. When eliminating dominated tuples, C1 and C2 know the dominance relationship among tuples with respect to the query tuple q, which violates our data pattern privacy requirement.

Conclusion:

We proposed a fully secure skyline protocol on encrypted data using two non-colluding cloud servers under the semi-honest model. It ensures semantic security in that the cloud server knows nothing about the data including indirect data patterns, query, as well as the query result. In addition, the client and data owner do not need to participate in the computation. We also presented a secure dominance protocol which can be used by skyline queries as well as other queries. Furthermore, we demonstrated two optimizations, data partitioning and lazy merging, to further reduce the computation load. Finally, we presented our implementation of the protocol and demonstrated the feasibility and efficiency of the solution. As for future work, we plan to optimize the communication time complexity to further improve the performance of the protocol.

References

1. F. Baldimtsi and O. Ohrimenko. Sorting and searching behind the curtain.In FC 2015, pages 127–146, 2015.2

2. A. Beimel. Secret-sharing schemes: a survey. In International Conference on Coding and Cryptology, pages 11–46. Springer, 2011.

3. J. L. Bentley. Multidimensional divide-and-conquer. Commun. ACM, 23(4):214–229, 1980.

4. J. L. Bentley, H. T. Kung, M. Schkolnick, and C. D. Thompson. On the average number of maxima in a set of vectors and applications. J. ACM, 25(4):536–543, 1978.

5. S. B¨orzs¨onyi, D. Kossmann, and K. Stocker. The skyline operator. In ICDE 2001.

6. S. Bothe, A. Cuzzocrea, P. Karras, and A. Vlachou. Skyline query processing over encrypted data: An attribute-order-preserving-free approach. In PSBD@CIKM, pages 37–43, 2014.

7. S. Bothe, P. Karras, and A. Vlachou. eskyline: Processing skyline queriesover

encrypted data. VLDB, 6(12):1338–1341, 2013.

8. C. Y. Chan, H. V. Jagadish, K.-L. Tan, A. K. H. Tung, and Z. Zhang. Finding k-dominant skylines in high dimensional space. In SIGMOD Conference, pages 503–514, 2006.

9. W. Chen, M. Liu, R. Zhang, Y. Zhang, and S. Liu. Secure outsourced skyline query processing via untrusted cloud service providers. In INFOCOM 2016.

10. V. Costan and S. Devadas. Intel sgx explained. Technical report, Cryptology ePrint Archive, Report 2016/086, 20 16. http://eprint. iacr.org.