Volume 3, Issue 8, August 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Efficient Approach for Protecting Similarity Search on Metric data Using Cloud

Sumalatha.B *, Chandra Babu .J, Sumalatha.y
Department of CSE
India

Abstract In This paper we are using the cloud computing technology to provide security for the metric data .The data is to be revealed only to trusted users, not to the service provider or anyone else. Users query the server for the most similar data objects to a query example. Outsourcing offers the data owner scalability and a low-initial investment. The need for privacy may be due to the data being sensitive (e.g., in medicine), valuable (e.g., in astronomy), or otherwise confidential. Cloud computing technologies and the propagation of ocation-based services, research on outsourced spatial databases has been spotlighted. Here fore, the traditional spatial databases owners want to outsource their resources to a service provider so that they can reduce cost for storage and management. Existing privacy-preserving query processing algorithms encrypt spatial database and erform a query on encrypted data. Nevertheless, the existing algorithms may reveal the original database from encrypted database and the query processing algorithms fall short in offering query processing on road networks.

Keywords—Annaanimization ,Nearest Neighbour, security ,cloud ,Encryption, Integrity,protection

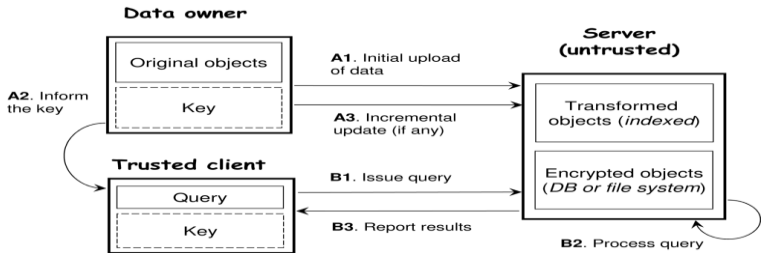
I. INTRODUCTION

Cloud computing services enable individuals and organizations to outsource the management of their data with ease and at low cost, even if they lack IT expertise. Cloud computing enables availability with respect to storage and computational resources as the number of service requests grows, without the need for costly investments in hardware and maintenance. Consider the example of a real-estate company that owns a large database with descriptions of properties and their locations. The company (i.e., the private data owner) wishes to allow authorized users (e.g., paying customers) to query for properties situated within a certain region. To save on hardware, investments and maintenance costs, the data owner outsources the management of its dataset to a service provider (SP) [4] that specializes in data storage and query processing. However, the SP may not be fully trusted, and could sell the data to a competitor. Furthermore, even if the SP is trusted, a malicious attacker can compromise the SP and gain unauthorized access to the data. To prevent such attacks, the data owner first encrypts the dataset according to a secret transformation and then uploads the encrypted data to the SP [4]. Only authorized users who know the transformation are able to learn the property locations.

II. EXISTING MODEL

This paper focuses on the outsourcing of metric datasets [2]. The main objective is to enforce the user authorization specified by the data owner, even when the service provider cannot be trusted. It presents techniques that protect location data from attackers, while allowing authorized users to issue queries that are executed efficiently by the SP [4]. In the literature, a number of concepts for securing databases have been studied. Private information retrieval technique [2]. In that technique gives query efficiency but never give the query privacy. Sometimes gives the query privacy while sacrificing the query efficiency and sometimes hide user query to the data owner. If user ask any query, in that technique just searching the query but never retrieve the query to the responding user. So there is no guarantee for retrieval and accuracy. It cannot prevent an attacker illegally copying the data from the data set.

Typically, cloud computing [1] providers attempt to solve any problem by offering the solutions only given to the data owner and are not to release outsourced data [2] to third parties. Nevertheless, even if the provider can compromise by anyone, the data is not guaranteed to be safe. Unintended leaks of data are reported regularly, and hackers may still exploit vulnerabilities to gain access to data. Therefore to believe that data owners will find it attractive to outsource encrypted rather than plain data.



Private information retrieval

Techniques hide the user's query, e.g., the data itemSearched for, but not the data being queried. To outsource Valuable data to an insecure server, such techniques areClearly not appropriate. Digital watermarking [2] establishes The data owner's identity on the data. Additional informationStored in the data helps prove ownership, but it cannotPrevent an attacker from illegally copying the data set.Anonymization techniques [25] secure data by releasing only

Related Work

We first introduce existing work on indexing and nearestNeighbor search techniques for metric data.

Scenario

From above depicts our scenario for outsourcing data. It consists of three entities: a data owner, a trusted query user, and an untrusted server. On the one hand, the data owner wishes to Upload his data to the server so that users are able to execute Queries on those data. On the other hand, the data owner Trusts only the users, and nobody else (including the server).The data owner has a set P of (original) objects (e.g.,Actual time series, graphs, strings), and a key to be used for transformation. First, the data owner applies transformation Function (with a key) to convert P into a set P0 of Transformed objects, and uploads the set P0 to the server (see step A1 in the figure). The server builds an index structure on the set P0 in order to facilitate efficient search.

III ENCRYPTED HIERARCHICAL INDEX BASED SEARCH

Encrypted hierarchical index search , this module is always gives a privacy with query efficiency and query is guarantee to be accurate. my database only have medical related data such as fever, headache and diabetes disease related information. Here all the data'sare stored in the hierarchical order in a subject wise or age wise or disease wise .For example two or more people affected by fever so they are asking fever related queries to the data owner. Then the data owner goes for similarity searching operation with help of the EHI algorithm. In First step is searching for fever related queries is available in the database or not. Here indexing is very important it is mainly used for efficient searching of data in the database. The searching operation is finished successfully. Then goes for fetching operation. The fever related queries [7] available in the database so the data is fetched and finally retrieve the data to the trusted users alone. Encrypted hierarchical index search , this module is always gives a privacy with query efficiency and query is guarantee to be accurate. my database only have medical related data such as fever, headache and diabetes

IV. METRIC PRESERVING TRANSFORMATION

The same EHI type of operation is also do here (searching, indexing, fetching, retrieving).Metric preserving transformation, for evaluating the NN [9] query, after that MPT gives the final result at two rounds of communication during the query phase. Here we use distance bounding phase and candidate retrieval phase by using that two phases it gives the final result at single rounds of communication. Distance bounding phase main function is to filter the keyword in the database list and candidate retrieval phase is also filter the number in the database list. Here both are using the optimization method is mainly used for reducing the processing overhead and increasing the efficiency. How to reduce the processing overhead, first step is largest database split up into smallest database finally merge the database we get the result in single rounds The transformation key consists of an encryption Key CK, an integer A, and A pairs of the form (ai,ri) where ai is an(anchor) object and ri is a distance value.

V. FLEXIBLE DISTANCE-BASED DYNAMIC HASHING

Here we propose A new hashing-based technique is called flexible distance-based dynamic hashing, for processing the NN [9] query. The main advantage of this technique is that the server always returns a constant-sized candidate set. Candidate set is nothing but total number of item set we are used in our transaction. The client then refines the candidate set to obtain the final result. Even though FDH is not guaranteed to return the exact result, the final result is very close to the actual NN in practice. During query processing, FDH allows the client to specify an integer parameter Θ for increasing the accuracy of a query result, without rebuilding the transformed data stored at the server. In addition, our FDH method employs a novel technique for conceptually linking similar [3] hash buckets, in order to maximize the utility of the transformed data for answering queries. The transformation key consists of an encryption Key CK, an integer A, and A pairs of the form (ai,ri) where ai is an object and ri is a distance value. The query processing strategy is to apply a similarity search on the above metric space index .The pseudocode of the searching algorithm for FDH. The client specifies an additional integer parameter and requests the server to retrieve the tuples whose bitmaps are the closest to the query bitmap BM. After receiving the result tuples from the server, the client decodes them into original objects and computes their distances from q.The client refines the candidate set to obtain the final result. The FDH is not guaranteed to return the exact result but the final result is very close to the actual result. This parameter provides a trade-off between the query cost and accuracy [2]. It always gives flexibility to the user. The FDH gives a final result at single rounds of communication.

Conclusions

Existing solutions either offer query efficiency at no privacy, or they offer complete data privacy while sacrificing query efficiency. It is attractive to be able to maintain data confidentiality with respect to untrusted parties, including the service provider. The paper presents methods to encode a dataset such that only authorized users can access the content, while the service provider "blindly" evaluates queries [4], without seeing the actual data. It is important for the data owner to choose an appropriate transformation method that best matches the requirements. We are proposing three

transformation methods. The first method is encrypted hierarchical index search algorithms gives the final result multiple rounds of communication. The second method is Metric Preserving Transformation method guarantees correctness of the final search result, but at the cost of two rounds of communication. The third proposed method is Flexible Distance-based Hashing methods finishes in just a single round of communication, but does not guarantee retrieval of the exact result. But actual result is very close to the exact result. This transformation methods achieve different trade-offs between the data privacy and query efficiency.

Acknowledgment

This research was supported by Basic Science Research program through the National Research Foundation of orea(NRF) funded by the Ministry of Education, Science and Technology(grant number 2010-0023800)

REFERENCES

1 P. Ciaccia, M. Patella, and P. Zezula, "M-Tree: An Efficient Access Method for Similarity Search in Metric Spaces," Proc. Very Large Databases (VLDB), pp. 426-435, 1997.

2 . M.L. Yiu, G. Ghinita, C.S. Jensen, and P. Kalnis, " Outsourcing Search Services on Private Spatial Data," Proc. IEEE 25th Int'l Conf. Data Eng. (ICDE), pp. 1140-1143, 2009.

3. G.R. Hjaltason and H. Samet, "Index-Driven Similarity Search in Metric Spaces," ACM Trans. Database Systems, vol. 28, no. 4, pp. 517-580, 2003.

4. G. Aggarwal, T. Feder, K. Kenthapadi, S. Khuller, R. Panigrahy, D.Thomas, and A. Zhu, "Achieving Anonymity via Clustering,"Proc. 25th ACM SIGMOD-SIGACT-SIGART Symp. Principles ofDatabase Systems (PODS), pp. 153-162, 2006.

5, Sion, R. Query Execution Assurance for Outsourced Databases. VLDB, 2005.

6. Yang, Y., Papadopoulos, S., Papadias, D., Kollios, G. Spatial Outsourcing for Location-based Services. ICDE, 2008.

7. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," Int'lJ. Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5,pp. 557-570, 2002.

8. W.K. Wong, D.W. Cheung, B. Kao, and N. Mamoulis, "Secure k-NN Computation on Encrypted Databases," Proc.35th ACM SIGMOD Int'l Conf. Management of Data, pp. 139-152, 2009

