© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : 10.32628/IJSRSET

Efficient and Enhanced Data Encryption on Skyline Queries

K. Kalaivani[1], K. Karthikeyan[2]

[1]PG Scholar, Department of Computer Science and Engineering, SNS College of Engineering, Coimbatore, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, SNS College of Engineering, Coimbatore, Tamil Nadu, India

ABSTRACT

Cloud computing is used to reduce the cost of large-scale data storage. User can outsource his data to cloud servers. Security and privacy are the major concern while outsourcing the data to unauthorised cloud servers. Medical records face these same security and privacy issues. So we can encrypt the medical data and outsource the data to cloud server. Query processing is done on encrypt the medical data. The challenging task is to query process the encrypt the medical data without revealing the original content. In this project we present the efficient query process and result retrieval. Skyline query protocol is applied for encrypted data for sematic information processing.

Keywords : Cloud Computing, Query Processing, Sematic Information Processing.

I. INTRODUCTION

Data mining is process of extracting useful information from large amount of databases. Data mining plays a major role in an exploratory analysis because of information in large volumes of data. The data mining techniques are useful for predicting the various diseases in the medical field. The diagnosis of this disease is intricate process. It should be diagnosed accurately and correctly. Due to limitation of the potential of the medical experts and their unavailability at certain places put their patients at high risk. Normally, it is diagnosed using intuition of the medical specialist. It would be highly advantageous if the techniques will be integrated with the medical information system.

Healthcare organizations can reduce costs by accomplishment of computer based data and/or decision support systems. Healthcare services data is very huge as it incorporates patient records, resource management information and updated information. Human services associations must have capacity to break down information. Treatment records of many patients can be stored away in computerized way; furthermore data mining methods may help in finding out a few vital and basic inquiries related with healthcare organizations. Diseases can be predicted through the analysis made on some attributes like age, sex, chest pain type, blood pressure, cholesterol, fasting blood sugar, Maximum heart rate achieved. Based on the values of the attributes, we make indexes for all associated frequent item sets. The presence of these item sets depends on the threshold value specified.

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 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. 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. 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.

Figure1 : Overall Architecture

II. RELATED WORKS

There are three different supervised machine learning algorithms for heart disease prediction. They are Naive Bayes, K-nearest neighbour, and Decision tree. These algorithms have been used for analyzing the heart disease. Tanagra is the data mining tool used for classifying these medical data and these data are calculated using 10 fold cross validation. Naive Bayes algorithm performs well when compared to other algorithms. Genetic algorithm has been used in, to reduce the definite data size to obtain the best possible subset of attribute which is essential for heart disease prediction. Classification is supervised learning method to extract models relating main classes of data. Decision Tree, Naive Bayes and Classification via clustering are the three classifiers used to analyze the occurrence of heart disease for the patients. Shekar et al proposed new algorithm to mine association rules from medical data based on digit sequence and clustering for heart attack prediction the entire data base is divided into partitions of equal size, each partition will be called cluster. This approach reduces main memory requirement since it consider only a small cluster at a time and it is scalable and efficient.

Fully homomorphic encryption schemes enable arbitrary computations on encrypted data. Even though it is shown that we can build such encryption schemes with polynomial time, they remain far from practical even with the state of the art implementations. Many techniques have been proposed to support specific queries or computations on encrypted data with varying degrees of security guarantee and efficiency. We are not aware of any formal work on secure skyline queries over encrypted data with semantic security. Bothe et al. and their demo version illustrated an approach about skyline queries on so-called “encrypted” data without any formal security guarantee. Another work studied the verification of skyline query result returned by an untrusted cloud server. The closely related work is secure kNN queries which we discuss in more detail here. Wong et al. proposed a new encryption scheme called asymmetricscalar-product-preserving encryption. In their work, data and query are encrypted using slightly different encryption schemes and all clients know the private key. Hu et al. proposed a method based on provably secure privacy homomorphism encryption scheme. However, both schemes are vulnerable to the chosen-plaintext attacks.

Existing system

The need of secure big data storage service is more desirable than ever to date. The basic requirement of the service is to guarantee the confidentiality of the data. However, the anonymity of the service clients, one of the most essential aspects of privacy, should be considered simultaneously. Moreover, the service also should provide practical and fine-grained encrypted data sharing such that a data owner is allowed to share a cipher text of data among others under some specified conditions. Cloud computing attracts increasing attention from both research and industry communities. Outsourcing data and computation to cloud server provides a cost efective 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. To protect the confidentiality of the query from cloud server, authorized clients also send encrypted queries to the cloud server. 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.

III. PROPOSED SYSTEM

Proposed system focus on the problem of secure skyline queries on encrypted data, another type of similarity search important for multi-criteria decision making. Proposed system outsource its electronic health records to the cloud and the data is encrypted to ensure data confidentiality. 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 fp1; p2g, 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. We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the relevance of data documents to the search query.

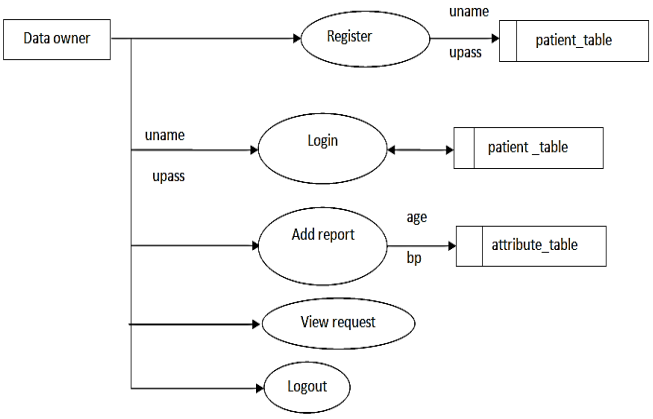

Figure 2 : Dataowner module

In cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval. Keyword- based retrieval is a typical data service and widely applied in plaintext scenarios, in which users retrieve relevant files in a file set based on keywords. However, it turns out to be a difficult task in cipher text scenario due to limited operations on encrypted data. Besides, to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to users. on-demand high-quality applications and services from a shared pool of configurable computing resources.

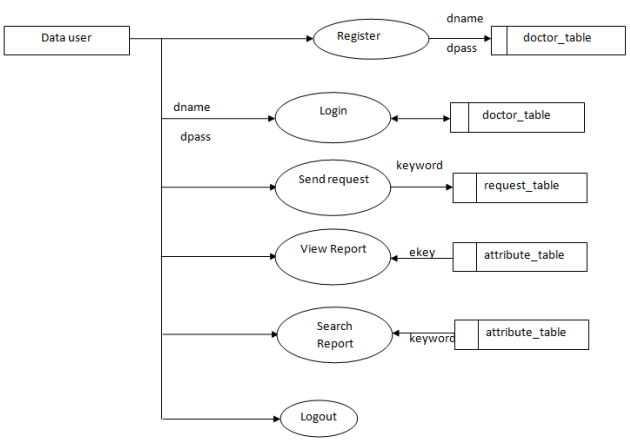

Figure 2 : Datauser module

Figure 3 : Send Request

In this project, 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.

Dataowner register their details like name, email, phone, username and password. Dataowner logs into the system with registered username and password.

Dataowner enter their disease history attribute details like age, gender, whether they are overweight, whether they are mentally stressed, whether they have pin in chest, blood pressure detail, blood sugar details. All the dataowner files gets encrypted and gets stored in cloud. The dataowner files has attributes like report name, report keyword etc.

In this module Dataowner can view the file access request send by the data user. File name, report name will be displayed to Dataowner. Dataowner can grant the access to Datauser and access key will notified to the Datauser.

Datauser registers their details like name, email, phone, username and password.

Datauser logs into the system with username and password.

In this module Datauser can search the report file by specifying report keyword. The file is shown based on report keyword given by Dataowner. Datauser can send the file access request to Dataowner. After Dataowner approval Datauser can view the file.

In this module Datauser can view report file by specifying the access key given by the data owner. If correct access key is given encrypted file will be decrypted and shown to user. If wrong access key is given, error message is thrown to user.

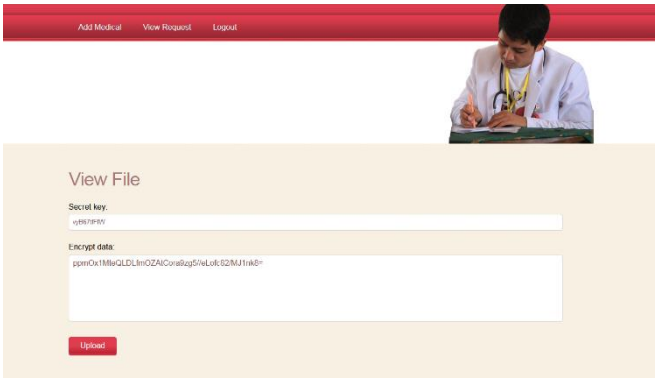

Figure 4 : File Encryption

We summarize our contributions as follows.

We study the secure skyline problem on encrypted data with semantic security for the first time. We assume the data is encrypted using cryptosystem which provides semantic security and is partially homomorphic. We propose 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. We present two secure skyline query protocols. The first one, served as a basic and effcient solution, leaks some indirect data patterns to the cloud server. The second one is fully secure and ensures that the cloud gains no knowledge about the data including indirect patterns. The proposed protocols exploit the partial (additive) homomorphism as well as novel permutation and perturbation techniques to ensure the correct result is computed while guaranteeing privacy. We provide security and complexity analysis of the proposed protocols.

Compared with our conference version, we present two new optimizations, data partitioning and lazy merging, to further reduce the computation load. For the data partitioning, we theoretically analyze the optimal number of partitions given the number of points, the expected number of output skyline points, the number of decomposed bits, and the number of dimensions. In addition, we propose a lazy merging scheme that aims to reduce computation overhead due to the smaller partition sizes at the later stage of the partitioning scheme. We also provide a complete implementation including both serial and parallelized versions which can be deployed in practical settings. We empirically study the efficiency and scalability of the implementations under different parameter settings, verifying the feasibility of our proposed solutions. The skyline computation problem was first studied in computational geometry field where they focused on worst-case time complexity.

Figure 5 : File Decryption

IV. CONCLUSION AND FUTURE WORK

In this project, we proposed a fully secure skyline protocol on encrypted data. It ensures semantic security in that the cloud servers 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, data encryption 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.

V. REFERENCES

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

[2]. Cao .N, Yu .S, Yang .Z, Lou .W, and Hou. Y, "LT Codes- Based Secure and Reliable Cloud Storage Service," Proc.

[3]. Chang .Y.-C and Mitzenmacher .M, "Privacy Preserving Keyword Searches on Remote Encrypted Data,"

[4]. Curtmola. R, Garay J.A, Kamara. S, and Ostrovsky .R, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc.

[5]. 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

[6]. E. Dellis and B. Seeger. Efficient computation of reverse skyline queries. In VLDB, pages 291–302, 2007

[7]. I Kala , Efficient Fault Detection Mechanism For Reliable Transmission In Mobile Adhoc Networks ,the International Journal of Applied Engineering Research, ISSN, 0973-4562,2015

[8]. I Kala, LS Jayashree Survey of Routing Protocols and Attacks for Mobile Ad hoc Networks, Networking and Communication Engineering, pages: 999-1002, 2011

[9]. I Kala, N Karthikeyan, S Karthik Region Based AODV Geographic Routing Protocol for Quasi MANET- Asian Journal of Information Technology, ISSN 5994, 2016.

[10]. 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

[11]. Song. D, Wagner. D, and Perrig . A, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[12]. Y. Elmehdwi, B. K. Samanthula, andW. Jiang. Secure k-nearest neighbor query over encrypted data in outsourced environments. In ICDE 2014

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

Cite this article as :

K. Kalaivani, K. Karthikeyan, "Efficient And Enhanced Data Encryption on Skyline Queries", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 659-664, March-April 2019.

Journal URL : http://ijsrset.com/IJSRSET207145