ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017

Secure and Dynamic Query Service Provider in KNN Classification Using Clustering

A.Roslin Deepa[1], Dr. Ramalingam Sugumar[2]
Ph.D Scholar, Dept. of Computer Science, Christhu Raj College, Bharathidasan University, Trichy, Tamil Nadu-India[1].
Professor, Dept. of Computer Science, Christhu Raj College, Trichy, Tamil Nadu-India[2].

Abstract: Data mining has large variety of real time appliance in several fields such as financial, shopping, telecommunication, biological, and between government agencies. Classification is the one of the major task in data mining. For the past few years, due to the increase in various privacy problem, many theoretical and possible solution to the classification problem have been proposed below different sureness model. The data in the data mining are in encrypted form, existing privacy preserving classification system are not related. Since the data on the data mining is in encrypted form, existing privacy preserving classification technique is not relevant. In this paper, we focus on solving the classification problem over encrypted data. In exacting, we propose a secure k-NN classifier over encrypted data in the data mining approach is very efficient technique. The proposed k-NN protocol protects the privacy of the data, user's input query analysis, and data access pattern. Our work is the first to develop a secure k-NN classifier over encrypted data under the standard with XOR encryption algorithm. To provide enhanced security, a secure kNN protocol that protects the privacy of the data, user's input query analysis, and data access patterns. Also, we empirically analyze the efficiency of our protocols during various experiments. These results specify that our secure protocol is very efficient on the user end, and this light weight method allows a user to use any mobile device to perform the kNN query.

Keywords: KNN Classifier, Security, graph pattern matching, encryption, privacy preserving, secure protocol.

I. INTRODUCTION

Data mining is a influential new technique to discover knowledge within the huge amount of the data. Also data mining is the method of discover meaningful new connection, patterns and trend by passing large amounts of data stored in quantity, using pattern recognition technologies as well as numerical and mathematical techniques. The KNN-classification of time series is an significant domain of machine learning due to the extensive amount of time-series data in real-life application.

Newly, the cloud computing model is revolutionize N the organizations' way of operating their data mainly in the way they store, access and process data. As an rising computing model, cloud computing attracts many organization to consider critically regarding cloud potential in expressions of its cost-efficiency, flexibility, and offload of administrative overhead. Most often, organization delegate their computational operations in addition to their data to the cloud. Despite wonderful advantages that the cloud offers, privacy and security issues in the cloud are prevent companies to exploit those advantages. When data are extremely sensitive, the data need to be encrypted before outsourcing to the cloud. However, after data are encrypted, irrespective of the underlying encryption system, performing any data mining tasks become very demanding without ever decrypting the data.

There are added privacy concerns, confirmed by the following instance. *Example 1:* Suppose an insurance company outsourced its encrypted clients database and relevant data mining everyday jobs to a cloud. When an agent from the company desires to determine the risk level of a possible new customer, the agent can use a classification method to decide the risk level of the customer. First, the agent needs to create a data record q for the customer contain certain personal information of the customer, e.g., credit score, work, age, marital status, place, etc. Then this documentation can be send to the cloud, and the cloud will calculate the class make for q. however, since q contain sensitive information, to protect the customer's privacy, q must be encrypted before distribution it to the cloud. The above model shows that data mining over encrypted data on a cloud also needs to defend a user's record when the record is a part of a data mining procedure. Moreover, cloud can also obtain useful and responsive information about the actual data items by observes the data access pattern even if the data

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017

are encrypted. Therefore, the privacy/security requirements of the encryption problem on a cloud are threefold: (1) privacy of the encrypted data, (2) secrecy of a user's query record, and (3) thrashing data access patterns. Existing work on Privacy-Preserving Data Mining (either perturbation or secure multi-party computation based come close to) cannot solve the encryption difficulty. Troubled data do not possess semantic security, so data perturbation technique cannot be used to encrypt highly responsive data. Also the troubled data do not generate very correct data mining results. Secure multi-party computation (SMC) based move toward assumes data are circulated and not encrypted at each participate party. In addition, many transitional computations are performed based on non encrypted data.

Using encryption as a way to complete data privacy may cause another issue during the query processing step in the cloud. In common, it is very difficult to procedure encrypted data without ever have to decrypt it. The question here is how the cloud can perform the queries over encrypted data however the data stored at the cloud are encrypted at all times. In the literature, various technique related to query processing over encrypted data have been proposed, including range queries and other collective queries. However, these techniques are either not applicable or inefficient to answer advanced queries such as the k-nearest neighbor (kNN) query.

This proposed system focus on the classification problem because it is one of the most frequent data mining tasks. Because each classification method has their own advantage, to be actual, this paper concentrate on executing the k-nearest neighbor classification method over encrypted data in the cloud computing environment with more security based XOR encryption algorithm.

II. Privacy-Preserving Data Mining (PPDM)

Privacy Preserving Data Mining (PPDM) is definite as the procedure of extracting/deriving the information about data without compromise the privacy of data. In the past decade, many privacy-preserving classification techniques have been proposed in the literature in order to defend user privacy. In particular to privacy preserving classification, the objective is to make a classifier in order to expect the class label of input data record base on the distributed training dataset without compromise the privacy of data.

Classification is one essential task in many applications of data mining such as health-care, shopping, colleges and business. Recently, performing data mining in the cloud involved significant awareness. In cloud computing, data holder outsources his/her data to the cloud. Though, from user's perception, privacy becomes a significant issue when responsive data needs to be outsourced to the cloud. The shortest way to protector the outsourced data is to relate encryption on the data ahead of outsourcing.

Unfortunately, as the hosted data on the cloud is in encrypted form in our trouble domain, the existing privacy preserve classification techniques are not satisfactory and applicable to PPkNN owed to the following reasons. (i) In existing methods, the data are partition among at least two parties, while in our case encrypted data are hosted on the cloud. (ii) Because some quantity of information is loss due to the calculation of statistical noises in order to conceal the sensitive attribute, the existing methods are not accurate. (iii) Leakage of data admittance patterns: the cloud can easily obtain useful and sensitive information about users' data items by minimally observing the database entrée patterns. For the similar reasons, in this paper, we do not believe secure k-nearest neighbor technique in which the data are discrete between two parties.

III. System Architecture

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017

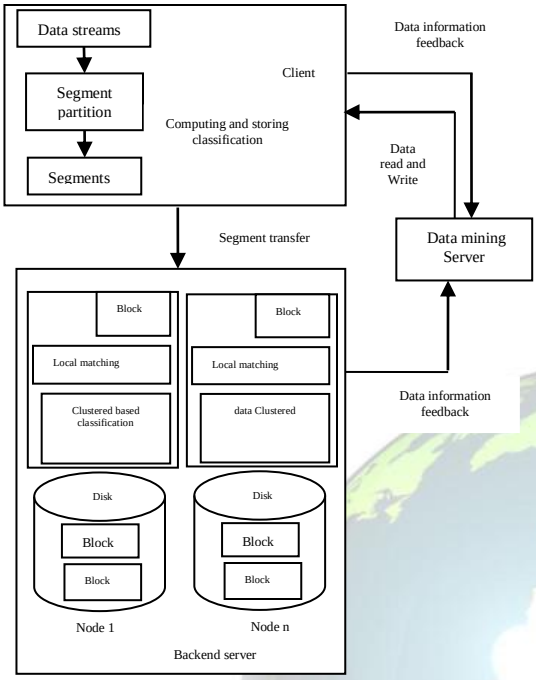


Fig.1. The system architecture consists of three practical components in above figure, Client, Data Server, and Backend Server.

• Client is answerable for gather backup datasets and communicates with storage node and data Server to replace information. At the same time, the client process segmenting, data classifications, storing graph data's by clustering, and distribute segments to storage node.

• Data Server is to store and look up all graph index pattern matching of files and segments.

• Backend Server is to remove duplicate data and store backup data. The system requests several storage nodes for parallel deduplication.

Our classification implementation process, for a received backup stream, files are segmented to be distributed to the nodes. When one data segment comes, it is check for similarity in local graph index. If it is the similar, the segment data is not stored. If it is high similar, the system calls the equivalent block. When the data segment is low similar, it is check for similarity in graph index if the data segment have a high similarity, the method assigns the segment to the related node. If it is not a high similarity, home node stores the segment data graph and updates clustered graph database server of other nodes.

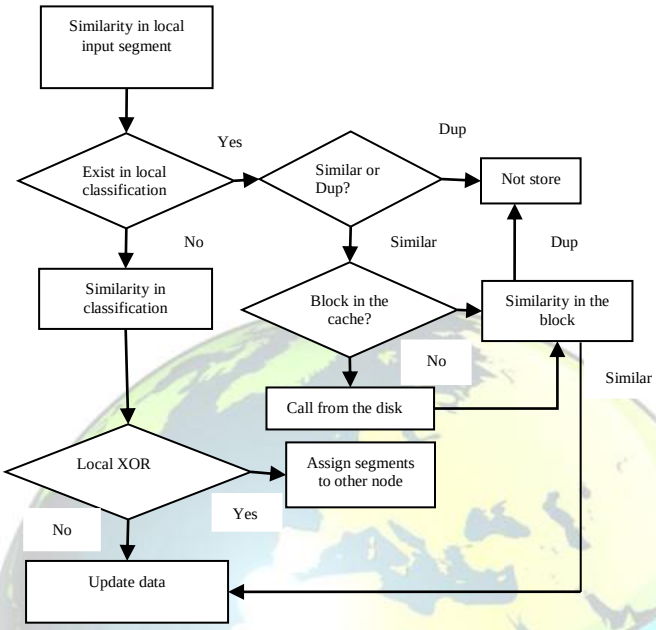
Data Server

In this module, the data server uploads their data in the server. For the security reason the data owner encrypts the data file and after that store in the server. The Data owner can have proficient of manipulating the encrypted data file.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017

IV. IMPLEMENTATION



.

EVALUATION

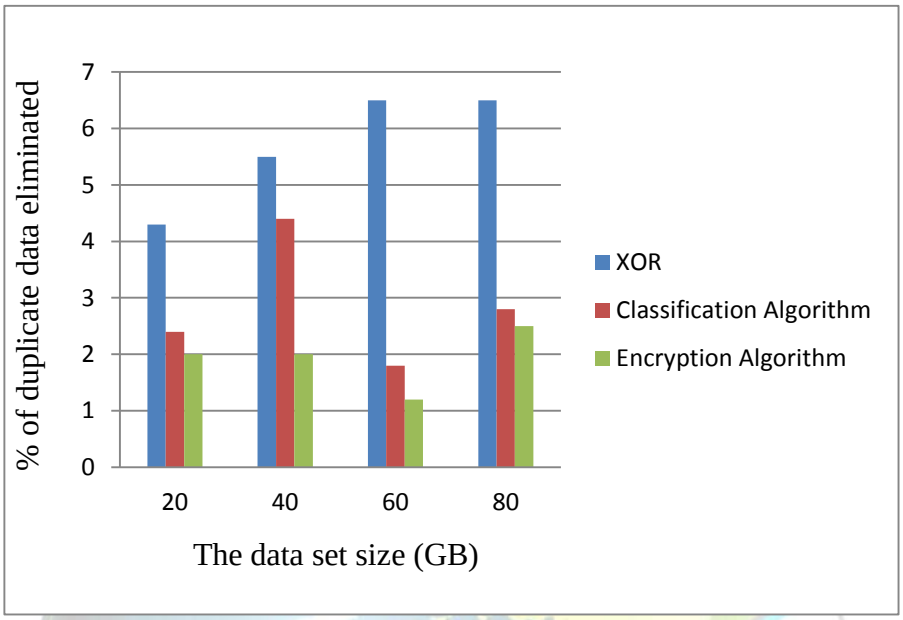
Our system is tested for its removal ratio on memory and throughput. The systems accept a small probability of false positive. The data sets are collected of files from a sequence of backups. For elimination ratio, we compare our system with two additional situational systems, the Overlapping clustering algorithm and the slicing algorithm. The incremental backup is used to test duplicate elimination ratio in memory. For data classification throughput, it is experimental throughput of cluster as nodes increase. The experiment is gain to estimate our system performance.

PERFORMANCE

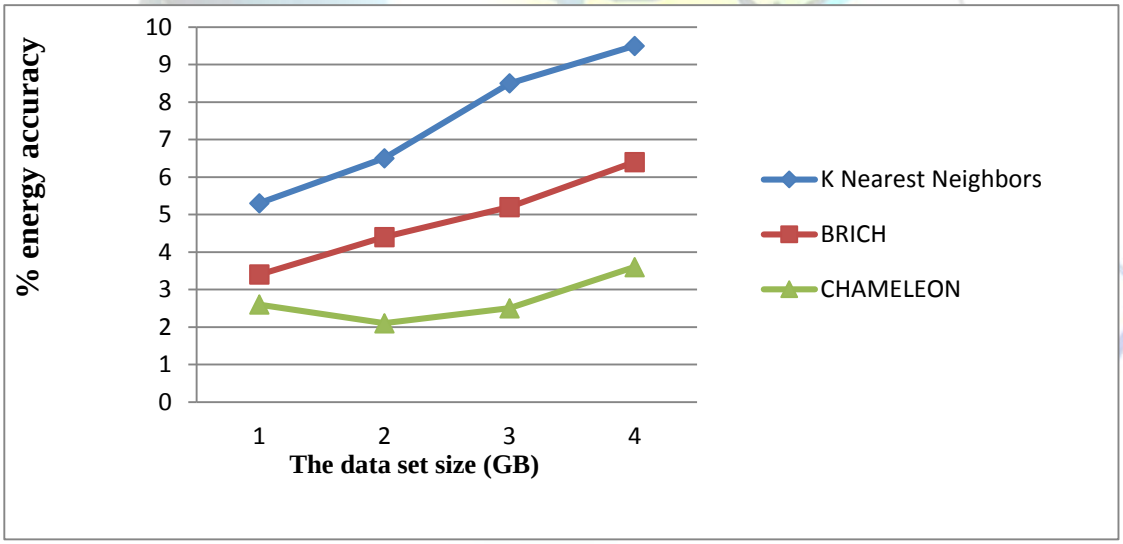
We call our system XOR encryption as the backup size increases; KNN with XOR improve about 60%~68% performance of the data. The average data rate ratio is 64.5%. The average elimination rates of classification algorithm and encryption algorithm are respectively 28.38% and 23.23%.Compared with classification and encrypted algorithm, our system has a high elimination performance. Overlapping clustering only exploit file similarity, but it miss some duplicate data in the file and ignore locality of files. Encryption algorithm exploits the inbuilt locality in a backup stream. But when the data sets do not have place, it will lose its advantages and locate little duplicate data. Our system finds more duplicate data in the similar data set. The system check local graph pattern table of using pattern matching to find similar segments, and at the same time to use multi graph algorithm to detect duplicate segments in the node.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017



Figue 2 The percentage of performance



V. CONCLUSION

Classification is a significant task in many data mining applications such as discovery of fraud by credit card companies and calculation of tumor cells levels in blood. To protect user privacy, various privacy-preserving classification techniques has been proposed in the literature for the earlier period decade. However, the existing techniques are not appropriate in outsourced database location where the data resides in encrypted form on a third-party server. Beside this direction, this paper proposed a novel privacy-preserving k-NN classification protocol over encrypted information in the cloud. Our protocol protects the privacy of the data, user's input query, and hides the data contact patterns. We also evaluate the performance of our protocol under different stricture settings. Since improving the efficiency of data's is an important first step for humanizing the performance of our encrypted protocol, we

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 2, February 2017

plan to investigate alternative and more resourceful solutions to the encryption problem in our future work. Also, in this paper, we used the well-known k-NN classifier and urban a privacy-preserving protocol for it over encrypted data. As a future work, we will inspect and extend our investigation to other classification algorithms.

REFERENCES

[1]. C. C. Aggarwal and P. S. Yu. A general survey of privacy-preserving data mining models and algorithms. Privacy-preserving data mining, pages 11–52, 2008.

[2]. Y. Aumann and Y. Lindell. Security against covert adversaries: Efficient protocols for realistic adversaries. Journal of Cryptology, 23(2):281–343, Apr. 2010.

[3]. R. Agrawal and R. Srikant. Privacy-preserving data mining. In ACM Sigmod Record, volume 29, pages 439–450. ACM, 2000

[4]. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order preserving encryption for numeric data. In ACM SIGMOD, pages 563–574, 2004.26

[5]. P. Mell and T. Grance, "The nist definition of cloud computing (draft)," NIST special publication, vol. 800, p. 145, 2011

[6]. S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in CRiSIS, pp. 1 –9, 2012.