IJCSN International Journal of Computer Science and Network, Volume 5, Issue 4, August 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Enhancement in Classification of Semantically Secure Encrypted Data

[1] Chaitali Shewale, [2] S.M. Sangve

[1,2] Computer Department, ZCOER
SPI Pune University
Pune, India

Abstract - Techniques which are already existing are not having the high clustering to the encrypted data on cloud, therefore, to solve the classification problem on encrypted data. A k-NN algorithm for classification of encrypted data in the cloud, is also proposed for security purpose. In this paper we proposed the the new DBSCAN algorithm which will provide the better clustering than the K-NN clustering of the data on the cloud, provides In day by day the popularity of web services are attracting with their rapid development, As a result, there is a huge amount of heterogeneous data. The data needs to be mine for various applications in many organizations like scientific research, medicine and among government agencies. Data Mining is perspective is on a very large scale. Classification of outsourced data is One of the most important tasks in data mining applications. So in data mining area many practical as well as theoretical solutions to the classification problem have been proposed. As as a privacy issues solution, Different security models used in different solutions. The recent needs of IT make Cloud Computing to exist. Always users can outsource encrypted form of their data in and the data mining tasks to the cloud. Privacy preserving classification security to user's input query and hiding the data accessing patterns on the cloud. As well as it will provide better clustering solution for the outsourced data.

Keywords - *Security, DBSCAN, Outsourced Databases, Encryption.*

1. Introduction

In early days of life web services and their popularity are increasingly day by day With the rapid development of web services, As a result, there is a huge amount of heterogeneous data. The data needs to be mine for various applications in various organizations like scientific research, medicine and government among All agencies .In these days, in cloud computing revolution in the organizations way of operating their data is changed as the way they can process and access data. As an emerging computing paradigm, cloud computing includes many organizations to con-sider seriously related cloud potential in terms of its flexibility, cost-efficiency, off load of admin work overhead. Often, organizations allot their computational operations of data to the cloud. Despite tremendous privacy and security issues advantages that the cloud offers, in the cloud are preventing companies to utilize those advantages. When data are very important, the data encryption is necessary before outsourcing to the cloud. However, when data are encrypted, data mining tasks of encrypted data becomes very challenging when data gets encrypted without ever decrypting the data.

There are many security & privacy issues, discussed by the following example. Suppose an insurance company outsourced its encrypted consumers database and relevant data mining tasks to a cloud, and agent wants to determine the risk level of a potential, the agent need to use a classification method to determine the risk level of the customer. The agent needs to generate a data record M for the customer who have certain customer's personal information of the, e.g., credit score, age, marital status, etc. The record can be sent to the cloud, and the cloud will compute the class label for M. M contains very important information, to protect the customer's privacy, M should be encrypted before sending it to the cloud. The example as discussed above is the data mining of encrypted data (denoted by DMED) on a cloud it is also necessary to protect a user's record and data mining process record is a part of a cloud computing. Moreover, cloud also derive can useful and very important information about the actual data contents by observing the data access patterns as if the encrypted data are [2], [3]. so, the privacy/security requirements of the DMED problem on a cloud have three modules: hiding data access patterns. confidentiality of the encrypted data Confidentiality of a user's query record, As an emerging computing areas, cloud computing provides the cloud potentials to the

different organizations in considering many parameters likewise its cost-efficiency, flexibility, and offload of administrative overhead. often, as always organizations presenting their computational operations of their data on the cloud. Cloud provides advantages in different areas to the organizations, but with the privacy and security issues to get privacy and security advantages. The highly important data should be encrypted before outsourcing to the cloud. As the data is encrypted, Data mining tasks becomes very difficult not including ever decrypting the data.

There are other privacy issues, As discussed above problem suppose an insurance company trying to outsource database of customers which is in encrypted format and data mining tasks given to the cloud. An agent of the company wants to guess the risk level of a potential , the agent can use a classification method to estimate the risk level of the customer. For measuring risk level of customer , the agent should to create a data record M which contains some personal information of the customer, Then the created record is sent to the cloud, and the cloud computing includes finding class label for r. as, r contains highly important information, to lock the user privacy before sending it to the cloud., r should be encrypted The above scenario shows that data mining over data which is encrypted data (denoted by DMED) on a cloud. As a record is a part of the data mining process it must be protected, cloud can also be trained useful and sensitive information about the actual data. The data access patterns yet if data provided is it encrypted [2], [3]. Therefore, the confidentiality/safety necessities of the DMED issue on a cloud are: (1) hiding data access patterns (2) confidentiality of the encrypted data, (3) confidentiality of a user's query record, and.

Conventional methods on Privacy-Preserving Data Mining cannot solve the Data Mining of Encrypted Data problem. Because many middle part of computing while mining, are based on is the data is encrypted or not data. As a result, this paper proposes a newly proposed techniques are used for solving effectively the DMED problem. It is considered that the outsourced encrypted data to a cloud. To be more precise, we focused on the classification of encrypted data as it is one of the most common data mining farm duties. This paper proposes DBSCAN on executing the process over encrypted data in the cloud computing environment.

The paper is arranged in following way: Section I contains Introduction Section II contains the literature survey including the related work. Section III, contains proposed system is. Finally, the section IV concludes the paper.

2. Literature Survey

In this section, we have published previous paper briefly which we have studied research papers related to the maintaining privacy data mining (Privcy Preserving Data Mining) and query processing over encrypted data.

Paper Title: Fully homomorphic encryption using ideal lattices
Author: C. Gentry
Published year: 2009.

C. Gentry presented a fully Homomorphic cryptosystems in which DMED problem is solved. The system presented solution in which It allowed a third-party (that hosts the encrypted data) randomly to execute functions over encrypted data deprived of ever decrypting them. The problem with this system is, using techniques are very costly and these are not yet practically explored. For example, C. Gentry and S. Halevi [5] showing for insufficient security parameters one "Loading" operation on the homomorphic system it is observed that it takes a time for performing on ahigh performance machine. It takes at least 30 seconds to complete the task.

Paper Title: How to share a secret
Author: A. Shamir
Published year: 1979.

In this paper Author proposed a scheme which is secret sharing in secured multiparty computation (SMC), for developing a PPkNN protocol. SMC based approach assumes data are divided and data not encrypted for each participation party, intermediate computations are performed on non-encrypted data.

Paper Title: Sharemind: A framework for fast privacy-preserving computations
Author: D. Bogdanov, S. Laur, and J. Willemson
Published year: 2008.

In this paper Author proposed the constructions based on Share mind, a well-known SMC framework, it assumes that the number of participating parties are three. Thus, our work is orthogonal to Share mind and other secret sharing based schemes.

Paper Title: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise
Author: Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu
Published year: 1996.

2.1 Privacy-Preserving Data Mining (PPDM)

R. Agrawal and R. Srikant [8], Y. Lindell and B. Pinkas [9] introduces the symbol of maintaining privacy under data mining applications. The traditional PPDM techniques broadly classified into two types: (i) data demonstration and (ii) data circulation. Agrawal and Srikant [8] proposed the first data perturbation method to build a decision-making tree classifier, later many other methods were proposed (e.g., [10]–[12]). However, data agitation methods for encrypted data cannot be applicable for which data is semantically protected. Due to the addition of statistical disturbances to the data, agitation methods not to be produced accurate data mining results. On the other hand, Lindell and Pinkas [9] presenteded the very first resolution tree classification under the two or more than one -party setting. The third user do not have real key is to describe the queries considering the data were distributed between them. Since then much work has been published using SMC techniques (e.g., [13]–[15]). It is found that the PPkNN problem solving is difficult using the previously discussed data distribution techniques as the data in our case is in encrypted format but not distributed among multiple revelries. Hence, we also unable to consider proved as secured k-NN methods densed data distributed in two class parties (e.g., [16]).

2.2 Query Processing over Encrypted Data

Advanced Internet technologies in networking materialization is a fueled of the "software as a service" modeling of computing enterprise. Commercially viable software services includes renting electronic mail services, general storage services, spreadsheet disaster protection services. User access provided by "Database as a Service" model for creating, storing, modifying, and retrieving data from anywhere in the world, provided that Internet is available. It produces several challenges, to data privacy issue. It is in this context specifically address the issue of data privacy. There are two main privacy issues. First, owner data should be assured that the data storage service-provider site is to be protected against data stolen problem from outsiders. Second, data should be secured even from the service providers, if the providers themselves cannot be trusted. if the providers themselves not be confident about data security so for some techniques are to adopted to execute SQL queries over encrypted data. So as to process without having decrypt the data of the query at the service providers' site is challenging, Decryption with the remainder of the query processing are processed at the client site. So here we will discuss techniques proposed, related to query processing over encrypted data in [14]–[15]. It is proved that PPkNN is comparatively more complex than the running of simple kNN queries over encrypted data [16], [17]. The transitional k-nearest neighbors in the classification process, should not be reveal to the cloud or any other users [21] which presents the k-nearest neighbors to the user. And, even we know the k-nearest neighbors, still very hard to find that the query is belonging from which class label between these neighbors because these are encrypted at the first phase to prevent the cloud from learning very important information. Third, these not overcome the access pattern privacy issue which is a privacy key need from the user's perspective.

Y. Elmehdwi, B. K. Samanthula [18], proposed a secure k-nearest neighbor query protocol over encrypted data. The protocol paper proposes in the protection data confidentiality, user's query privacy, and disabling access of data patterns. Even if, as mentioned above, PPkNN is a more composite problem, it cannot be solved directly using as the existing secure k-nearest neighbor techniques over encrypted data. Therefore, this paper proposed a newly praposed solution to the PPkNN classifier sloving problem over encrypted data.

This paper is unlike from the all existing work [18] in this paper three aspects presented. First, paper, introduced new security primitives, secure minimum (SMIN), secure minimum out of n numbers (SMINn), secure frequency (SF), and found new solutions. Second, the proposed work in [18] did not provide any formality of security analysis of the underlying sub-protocols. On the other hand, this paper extended the discussion about proofs of the security underlying sub-protocols and the PPkNN protocol under the semi-honest model. Third, the preliminary proposed work in [18] reports only secured kNN query which is like as to Stage 1 of PPkNN. However, Stage 2 in PPkNN is entirely new.

3. Proposed System

Existing systems of privacy preserving classification techniques are not valid for the encrypted data on cloud. Therefore, this paper presenting solution to the classification problem on encrypted data. A spatial clustering density-based applications with noise (DBSCAN) DBSCAN clustering algorithm over encrypted data in the cloud, is proposed here for clustering as well as security purpose. The system proposes a protocol to provide security with the proper cluster labels to the encrypted data on the cloud, maintains privacy with proper cluster head /labels of user (third party's query) for input query and hides the data access patterns on the cloud. A reliable DBSCAN cluster is the first best cluster over the encrypted data under the semi-honest model. Privacy Preserving in kNN is a more composite problem and it cannot be solved using the already existed secure k-nearest neighbor method over encrypted data. Therefore,

in some paper proposed a new solution to the Privacy Preserving in kNN classifier problem over encrypted data. For this, proposed system uses a traditional set of nonspecific sub-protocols that will be used in building the proposed k-NN classifier. The security primitives using by the protocols are : i) secure minimum (SMIN), ii) secure minimum out of n numbers (SMINn) and iii) secure frequency (SF). Presented in paper system offers solution to each of these primitive. So the protocol proposed in the paper [11] is advantageous in protection of the confidentiality of the data maintaining Privacy of user's input query and Hiding the data access patterns. So all above discussed paper are related to the security preserving techniques.

Paper focused on the better clustering of outsourced encrypted data on cloud so the our paper focused on firstly proper classification or proper clustering of the encrypted data and also we can achieve the data security but mainly this paper focus on the clustering as the improved clustering is achieved data access for the system will very easy and fast even the access is very fast the security of the data is also taken into consideration

Density-based spatial clustering of applications with noise (DBSCAN) is used here for the data clustering this algorithm firstly proposed by the Peter Kriegel , Martin Ester Hans-,Jörg Sander and Xiaowei Xu in the year of 1996. Actually It is a density-based clustering algorithm to higher density data also in large density data access is the major issue and it is applicable to the given a set of points in some space, it makes groups together to all the points that are closely packed together (points are with many nearly neighborhoods), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are not the far away). DBSCAN is the best clustering algorithms and also most cited in scientific literature. In 2014, the algorithm was received award the test of time (an award given to algorithms which have received considering in theory and practice) at the proceeding data mining

There is problem of clustering of points in data is actually challenging process when the clusters are of different size, density and data-shape. Some of these issues are significant when the data is of very high
dimensionality and when it includes noise and outliers.

This paper presents density-based clustering algorithms: DBSCAN and SNN. Here, the role of the clustering DBSCAN algorithms is to find clusters of Points of Interest (POIs) and then use the clusters to automatically characterize geographical different regions. In these densed datasets, each dataset record represents a POI. The major characteristics of the DBSCAN algorithm

3.1 DBSCAN Algorithm

The Ester, et al. [in Ester1996] invented DBSCAN algorithm firstly, and it is density-based cluster head symbols of clusters. Clusters are recognized by looking at the point density. Regions with a high density of points show the existence of clusters while regions with a low density of points indicate clusters of noise or outliers clusters. This algorithm is suitable for dealing with large datasets, with noise, and is able to identify clusters which are of various and shapes sizes.

3.1.1 The Algorithm

The key idea of the DBSCAN algorithm is that, for every point of a cluster, the neighborhood of in a given area has to contain at least a minimum number of points, that is, the densities of the neighborhood should be exceed some predefined threshold. This algorithm needs three different input parameters are: k, the neighbor list size;

Eps, the radius that delimitate the neighborhood area of a point (Eps- neighborhood);

MinPts, the minimum number of points that must exist in the Eps-neighbourhood.

The clustering process is based points classification of the dataset as basic points, border points and noise points, and useing density relationship of points (directly density-reachable, density-reachable, density-connected [Ester1996]) to form the clusters.

4. Conclusion

This paper presents various existing methods from different papers used for the privacy preserving query processing in data mining and over encrypted data can mention as (PPDM). To protect user data privacy, with the discussion of various proposals of privacy-preserving classification techniques are presented over the past papers. The existing techniques are applicable to outsourced database environments where the data resides in encrypted form on a third-party server and also the classification or clustering of the data with proper label is discussed here This paper proposed a new privacy-preserving clustering protocol DBSCAN is used over encrypted data of the cloud. This protocol protects the confidentiality of the data, as a user's input query, and hides the data access patterns. As well as provide fast access to data as better clustering is done using DBSCAN algorithm. Future research can be focus on more efficient solutions to the DBSCAN problem, because the performance of the DBSCAN protocol security results comparatively not better than KNN so we have focused

on better clustering even though we will get security Also, this paper can be extended towards the other clustering algorithms with better security.

References

[1] M. Kantarcioglu and C. Clifton, "Privately computing a distributed k-nn classifier," in PKDD, pp. 279–290, 2004.

[2] 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.

[3] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: practical access pattern privacy and correctness on untrusted storage," in ACM CCS, pp. 139–148, 2008.

[4] Y. Qi and M. J. Atallah, "Efficient privacy-preserving k-nearest neighbor search," in IEEE ICDCS, pp. 311–319, 2008.

[5] C. Gentry and S. Halevi, "Implementing gentry's fullyhomomorphic encryption scheme," in EUROCRYPT, pp. 129–148, Springer, 2011.

[6] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, pp. 612–613, Nov. 1979.

[7] D. Bogdanov, S. Laur, and J. Willemson, "Sharemind: A framework for fast privacy-preserving computations," in ESORICS, pp. 192–206, Springer, 2008.

[8] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in ACM Sigmod Record, vol. 29, pp. 439–450, ACM, 2000.

[9] L. Xiong, S. Chitti, and L. Liu, "K nearest neighbor classification across multiple private databases," in CIKM, pp. 840–841, ACM, 2006.

[10] P. Zhang, Y. Tong, S. Tang, and D. Yang, "Privacy preserving naive bayes classification," ADMA, pp. 744–752, 2005.

[11] Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu, " A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", The Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, 1996

[12] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in IEEE ICDE, pp. 217–228, 2005.

[13] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing private queries over untrusted data cloud through privacy homomorphism," in IEEE ICDE, pp. 601–612, 2011.

[14] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in ACM SIGMOD, pp. 563–574, 2004.

[15] B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, "Secure multidimensional range queries over outsourced data," The VLDB Journal, vol. 21, no. 3, pp. 333–358, 2012.

[16] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in ACM SIGMOD, pp. 139–152, 2009.

[17] X. Xiao, F. Li, and B. Yao, "Secure nearest neighbor revisited," in IEEE ICDE, pp. 733–744, 2013.

[18] Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k- nearest neighbor query over encrypted data in outsourced environments," in IEEE ICDE, pp. 664–675, 2014