Review on Classification over Semantically Secure Encrypted Data

VarshaSalunkhe
Assistant Professor,
IT Dept, Atharva College of Engineering,
Mumbai, Maharashtra,India

Komal Gothwal
Assistant Professor,
IT Dept, Atharva College of Engineering,
Mumbai, Maharashtra,India

***Abstract*— Data Mining is a broad area now days due to the need of knowledge discovery on a very large scale. One of the techniques used in data mining applications is the Classification. The rapid development of web services made web users to use these services on larger extent. As a result, there is a huge amount of heterogeneous data. This data needs to be extracted for various real time and other type of applications like bio-logical research, scientific research, banking and among government sector. Cloud Computing has been an emerging trend to store, access and retrieve data on a large scale. This mining of data over cloud uses encryption algorithms to maintain security. This article studies various encryption methods and classification techniques over semantically secure encrypted data.**

***Keywords*— Security, k-NN Classifier, Classification, Encryption.**

I. INTRODUCTION

Recently, the cloud computing paradigm [1] is revolutionizing the organizations' way of operating their data particularly in the way they store, access and process data. As an emerging computing paradigm, cloud computing attracts many organizations to consider seriously regarding cloud potential in terms of its cost-efficiency, flexibility, and offload of administrative overhead. Vast amount of data is being generated that needs to be mined, stored and accessed whenever required. Cloud offers huge storage as well as various facilities to access the stored data at any time. Most often, organizations delegate their computational operations in addition to their data to the cloud. Despite tremendous advantages that the cloud offers, privacy and security issues in the cloud are preventing companies to utilize those advantages. When data are highly sensitive, the data need to be encrypted before outsourcing to the cloud. However, when data are encrypted, irrespective of the underlying encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data. Classification is one of the widely applied works in data mining applications. As increasing popularity of cloud computing, users now able to outsource their information as well as the data management tasks to the cloud.

II. LITERATURE SURVEY

P Williams et al[2], introduced a new practical system for remote data storage with efficient access pattern privacy and correctness. A storage client can set up this system to assign encrypted reads, writes, and inserts to a potentially inquisitive and malicious storage service provider, without any knowledge of information or access patterns.

Pascal Paillier [3], recognizes the major computational problem called Composite Residuosity Class Problem. To handle this problem author proposed a new trapdoor mechanism and generated three encryption scheme such as a trapdoor permutation and two homomorphic probabilistic encryption schemes which are providing secure result under the assumption of intractability.

C Gentry[4] proposed a fully homomorphic encryption scheme that introduced a scheme which allows to evaluate circuits over encrypted data without being able to decrypt. This scheme was studied to solve the DMED problem. It permits a third-party (that hosts the encrypted data) to perform random functions over encrypted data without decrypting them. Such techniques are very costly and they are not yet practically explored.

Proposed technique resolves the DMED issue since it permits a third-party to implement random functions over encrypted data without ever decrypting them. This technique is very expensive and their usage in practical applications has yet to be explored.

In this paper they introduced a new number-theoretic problem and a related trapdoor mechanism using the concept of composite degree residue. Three new cryptosystems are obtained based on proposed technique, which are providing secure result under the assumption of intractability.

Table I: Summary of Literature Survey

Sr.no	Paper	Proposed	Advantage
1	Building Castles out of Mud: Practical Access Pattern Privacy and Correctness on untrusted Storage [2]	Introduced new and efficient practical scheme with well-organized access pattern privacy for remote Data storage with exactness.	Proposed mechanism is faster and also present Privacy as well as correctness.
2	k-Nearest Neighbor Classification Over Semantically Secure Encrypted Relational Data [4]	Projected a secure k-NN classifier for encrypted data in the cloud.	The proposed KNN (k Nearest Neighbor) protocol provides security for the users input query, privacy of the data and data access patterns.
3	Fully Homomorphic Encryption Using Ideal Lattices [5]	Proposed method solve the DMED difficulty as it permits a third-party to carry out random function over encrypted data without ever decrypting them.	Decryption overhead is reduced.
4.	Sharemind: a framework for fast privacy- preserving Computations [6]	Projected a new approach for developing privacy-preserving applications, that is SHAREMIND.	The SHAREMIND structure is considered to be a proficient and effortlessly programmable platform for developing and testing many privacy- preserving algorithms. . By using SHAREMIND anyone can develop secure multi-party protocols without the past knowledge of all implementation details

R. Dey, C. Tang, K. Ross, and N. Saxena [5], paper mainly focus on solving the difficulty of encrypted data classification. Paper proposed a new PPkNN protocol, a secure k-NN classifier over semantically secure encrypted data. Author projected a secure k-NN classifier for encrypted data in the cloud. Commonly used method in data mining is classification which is used in health-care and business. The proposed KNN (k Nearest Neighbor) protocol gives protection for the users input query, privacy of the data and data access patterns. Efficiency of proposed technique gives improved result. In proposed protocol once the encrypted data are handover to the cloud, Alice does not contribute in any computations. Thus, no information is exposed to Alice which ultimately achieves privacy.

D. Bogdanov, S. Laur, and J. Willemson [6], author proposed a new method for implementing privacy-preserving applications, namely SHAREMIND. SHAREMIND is based on share computing methods. SHAREMIND is basically a virtual machine for privacy-preserving data processing.
Performance is improved by using proposed method when compared to other similar frameworks. Application development interface developed by SHAREMIND is easy which primarily focus on implementation of data mining algorithm not on privacy problems.

The SHAREMIND structure is considered to be a proficient and effortlessly programmable platform for developing and testing many privacy-preserving algorithms. By using SHAREMIND anyone can build up secure multi-party protocols without the previous knowledge of all implementation details.

Agrawal and Srikant, Lindell and Pinkas [7], introduced the idea of privacy-preserving under data mining applications. Aim of paper is to develop a classifier in order to forecast the class label of input data record on the basis of distributed training dataset without compromising the confidentiality of data.

Author proposed first data perturbation method to build a decision-tree classifier. To precisely calculate the distribution of original data values author proposed a new reconstruction process. By using these reconstructed distributions, we are able to build classifiers whose accuracy is equivalent to the accuracy of classifiers built with the original data.

H. Hu, J. Xu, C. Ren, and B. Choi [8], studied the difficulty of processing private queries on indexed data for mutual privacy protection in a cloud environment. Author projected an encryption technique based on privacy homomorphism and an efficient solution that comprises a secure traversal structure. The proposed framework is scalable to large datasets by using an index-based method.

Secure protocols based on this framework is invented for processing usual queries such as k-nearest-neighbor queries (kNN) on R-tree index. Proposed framework is scalable to large datasets. This approach shows many benefits such as feasibility, efficiency and robustness.

In paper [9], the ability of databases to orchestrate and cooperate frequently enhances solace problems. Information warehousing along with information mining, giving information from a few assets under a solitary power, enhances the danger of solace offenses. Security ensuring information mining shows a technique for organizing this problem, mainly if information mining is done in a way that doesn't reveal data past the result. This paper gives a method to freely preparing k-nn class from apportioned assets without revealing any insights about the assets or their information, except that uncovered by the last classification result.

In paper [10], allotted protection safeguarding information digging routines are vital for mining a few databases with a least data divulgence. We give a structure along a general model and in addition multi-round calculations for study side to side parceled databases utilizing a solace ensuring k Nearest Neighbor (kNN) classifier.

III. CONCLUSION

This paper presented an comprehensive survey of classification data mining that are used to classify the data and a various methods to defend the information leakage. The key features, the pros and cons of each recommendation algorithm are described.

Classification has features to enhance the reach and benefits of data mining technology. As per survey, a strong need to develop secure classification technique to outsourced data to the cloud.

REFERENCES

[1] P. Mell and T. Grance, “The nist definition of cloud computing
[2] (draft),” NIST special publication, vol. 800, p. 145, 2011.
[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] P. Paillier, “Public key cryptosystems based on composite degree residuosity classes,” in Proc. 17th Int. Conf. Theory Appl. Cryptographic Techn., 1999, pp. 223–238.
[5] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, “k-nearest neighbor classification over semantically secure encrypted relational data,” eprint arXiv:1403.5001, 2014.
[6] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in ACM STOC, pp. 169–178, 2009
[7] D. Bogdanov, S. Laur, and J. Willemson, “Sharemind: A framework for fast privacy-preserving computations,” in Proc. 13th Eur. Symp. Res.
[8] Comput. Security: Comput. Security, 2008, pp. 192–206.
[9] R. Agrawal and R. Srikant, “Privacy-preserving data mining,” ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.
[10] H. Hu, J. Xu, C. Ren, and B. Choi, “Processing private queries over untrusted data cloud through privacy homomorphism,” in Proc. IEEE 27th Int. Conf. Data Eng., 2011, pp. 601–612.
[11] R. J. Bayardo and R. Agrawal, ―Data privacy through optimal k-anonymization,‖ in IEEE ICDE, pp. 217– 228, 2005.
[12] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, ―Privacy preserving mining of association rules,‖ Information Systems, vol. 29, no. 4, pp. 343–364, 2004.