KNN CLASSIFICATION SCHEME BASED PRIVACY PRESERVING POLICY OVER SEMANTICALLY SECURE ENCRYPTED DATABASE

Ms.Megha D.Savekar.[1], Prof.Sonali A.Patil[2]

[1] *Student,ComputerEngineering , BSIOTR Pune,Maharashtra, India*
[2] *Professor, ComputerEngineering , BSIOTR Pune,Maharashtra, India*

ABSTRACT

Data mining is used in variety of fields such as banking as well as medicine.Data mining not only used in scientific research but also in government agencies. Our ceter of research is to solve the classification problem in enciphered data. In particular, we recommend a secure k-NN classifier in encrypted data over the cloud. The recommended protocol preserve data confidentiality.It also preserve privacy in case of input query of user.It also shield the patterns of data access. We use secure k-NN classifierin case of encrypted data using the semi-honest model. our proposed novel method solve the DMED problem in efficient way.It estimates that the enciphered data are redistributed in a cloud. we give more attention on the classification problem.This paper focus on executing the k-nearest neighbor classification method in enciphered data in case of environment of cloud computing.

Keyword : - *Security, KNN Classifier, Outsourced Database*

1. Introduction

Data mining analyse the data in different view. It summarizes data in the form of information. It increases the revenue as well as cuts cost. Data mining is known as analytical a tool which is used for analysing data. It permits the applicant to evaluate data from variety of dimensions as well as angles. Next step is to categorize it, and at last recap the relationships identified. Data mining not only search correlations but also patterns in variety of fields in huge number of relational databases. Data is in the form of facts, numbers, and text. This can be developed by a computer. Now a day, companies are assembling enormous as well as increasing number of data not only in multiple formats but also in variety of databases. *Information* can be catered by the patterns. It can also supplied by association and relationships. It is possible to change Information into *knowledge* related to historical patterns as well as future trends. Cloud computing, also called as on-demand computing. It is one of the type of Internet-based computing, which supply shared resources as well as data to computers and other devices which is based on demand. It is a model for implementing universal, on-demand access to a shared pool of arranging computing resources. Users and enterprises get different types of solution from Cloud computing as well as storage with various capabilities for storing and processing their data in third-party data canters. It depends on of resources sharing for achieving the purpose of coherence as well as economies of scale. At the foundation of cloud computing is the expansive concept of converged infrastructure as well as shared utility. Secure multi-party computation is a subcategory of cryptography. Its main goal is to create methods which jointly compute a function for their inputs by keeping those inputs private. The two party setting is interesting from two points of view, first is an applications perspective and second is, it can be used in the two party setting which is not used in the multi-party case. Secure multi-party computation was subpart of the two-party setting.

There are no special roles of parties in the secret sharing based methods. Instead of that there is sharing of the data associated with each amongst the parties. Also each gate is evaluated by a protocol. The function is called as a “circuit” over GF(p), which is as opposed to the binary circuits applicable to Yao. This type of circuit is known as an arithmetic circuit in the literature. It is made up of addition as well as multiplication “gates” in which the values operated on are defined over GF(p). Recent work on privacy-preserving data mining (PPDM) cannot solve the issue related to DMED. It works on either perturbation or protected multi-party computation. Semantic protection is not applied to perturbed information. Information perturbation techniques may not be useful to secure highly confidential information. There is no precise information mining outcomes using perturbed information. Secure multi-party computations based strategy shows that information is spreaded as well as not secured at every taking involving party. Non-encrypted information is used to conduct many advanced calculations.

2. LITERATURE SURVEY

In this paper [1], a new realistic procedure for remote data storage space with efficient accessibility pattern comfort and correctness is introduced.

In paper [2], a completely homomorphic security plan is recommended – i.e., a plan that allows one to assess circuits over secured information without being able to decrypt.

In paper [3], collecting and handling delicate data is a challenging work. In fact, there is no common formula for building the necessary computer.

In paper [4], a structure for mining association rules from dealings made up of particular products where the information has been randomized to protect comfort of personal dealings.

In paper [5], the capability of databases to arrange and work together often improves comfort issues. Data warehousing along with data mining, providing data from several resources under a single authority, improves the risk of comfort offenses.

In paper [6], allocated privacy preserving data mining methods are crucial for mining several databases with a lowest information disclosure.

In our most recent work [7], we proposed a novel secure k-nearest neighbor query protocol over encrypted data that protects data confidentiality, user’s query privacy, and hides data access patterns.

In this paper [8], they display how to divide data D into n items in such a way that D is quickly reconstruct able from any k items, but even finish details of k - 1 items shows definitely no details about D.

In this paper [9], the problem of privacy preserving data mining is addressed. Particularly, a situation in which two parties having private databases wish to run a data mining algorithm on the partnership of their databases, without exposing any needless details.

In this paper [10], another practical method for remote information storage roo m with proficient availability example solace and rightness is presented. A storage room client can set up this methodology to issue secured read, composes, and embeds to a possibly inquisitive and unsafe storage room administration office, without uncovering data or openness sorts. The supplier is inadequate to set up any association between consequent gets to, or even to separate between a read and a compose. Besides, the buyer is given solid rightness ensures for its capacities – illicit organization conduct does not go unnoticed. We grew first sensible framework requests of greatness faster than present usage that can perform over different questions every second on 1 Tbyte+ databases with full computational solace and accuracy.

In paper [11], a totally homomorphic security arrangement is prescribed – i.e., an arrangement that permits one to survey circuits over secured data without having the capacity to decode. Our cure comes in three activities. Starting, we offer a typical result that, to assemble a security plan that permits appraisal of unessential circuits, it suffices to make a security plan that can survey (marginally upgraded releases of) its own unscrambling circuit; we contact an arrangement that can evaluate its (increased) decoding circuit boots trappable. Forthcoming, we clarify an open key security arrangement utilizing immaculate cross sections that is just about boots trappable. Grid based cryptosystems for the most part have unscrambling calculations with low circuit multifaceted nature, frequently secured with an internal thing calculation that is in NC1. Additionally, culminate cross sections offer both additive and multiplicative homeomorphisms (modulo an open key impeccable in a polynomial band that is appeared as a grid), as required to survey normal circuits.

In this paper [12], they show how to separation information D into n things in a manner that D is rapidly reproduce capable from any k things, however even complete points of interest of k - 1 things indicates unquestionably no insights about D. This procedure permits the improvement of successful key administration methods for cryptographic systems that can work securely and viably notwithstanding when setbacks harm 50 percent the things and assurance breaks uncover everything except one of the staying things.

3. SYSTEM ARCHITECTURE

Privaccy preserving KNN system architecture is shown in figure 1.it consist of input query and dataset.then it consist of Privacy Preserving KNN.it is divided into three subsectuions, first it consist of privacy preserving policy.second subsection is secure retrieval KNN.third subsect ion is secure computation of majority class.at last it generate output. The proposed PPkNN protocol mainly consists of the following two stages:

Stage 1: Secure Retrieval of k-Nearest Neighbors (SRkNN):
 In this stage, User initially sends his query q (in encrypted form) to C1.
 After this, C1 and C2 involve in a set of sub-protocols to securely retrieve (in encrypted form) the class
 Labels corresponding to the k-nearest neighbors of the input query q.
 At the end of this step, encrypted class labels of k-nearest neighbors are known only to C1.

Stage 2: Secure Computation of Majority Class (SCMCk):
 C1 and C2 jointly compute the class label with a majority voting among the k-nearest neighbors of q.
 At the end of this step, only User knows the class label corresponding to his input query record q.

PRIVACY-PRESERVING PRIMITIVES

It consist of secure multiplication, secure squared Euclidean distance, secure bit – decomposition, secure minimum, secure bit OR, secure frequency primitives. These proposals ensure the protection of user privacy together with the authentication, integrity and non-repudiation of transmitted messages during communication. the security and cryptographic protocols used in communication systems are usually designed according to the specific security requirements of the systems. Furthermore, the cryptographic designers have to consider the computational capabilities of intermediate and end nodes, bandwidth, communication delay, the number of users and other aspects as well.

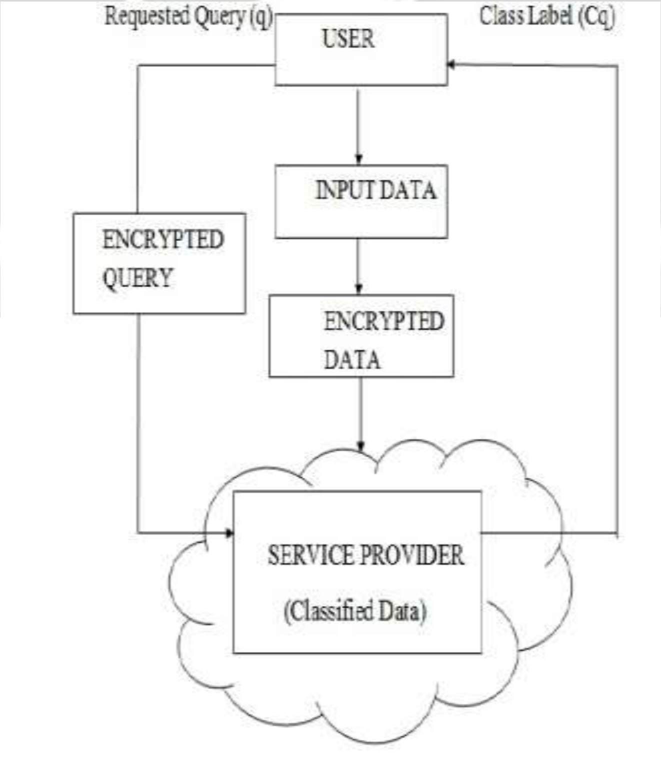

Fig. 1 Privacy Preserving KNN

4. IMPLEMENTATION DETAILS

4.1 Data Perturbation:

We proposed the first data perturbation technique to build a decision-tree classifier, and many other methods were proposed later. However, as mentioned earlier, data perturbation techniques cannot be applicable for semantically secure encrypted data. Also, they do not produce accurate data mining results due to the addition of statistical noises to the data. On the other hand proposed the first decision tree classifier under the two party setting assuming the data were distributed between them. Since then much work has been published using SMC techniques. We claim that the PPkNN problem cannot be solved using the data distribution techniques since the data in our case is encrypted and not distributed in plaintext among multiple parties. For the same reasons, we also do not consider secure k-NN methods.

4.2 Data Distribution:

In this paper, we introduced new security primitives, namely secure minimum (SMIN), secure minimum out of n numbers (SMINn), secure frequency (SF), and proposed new solutions for them. Second, the work in did not provide any formal security analysis of the underlying sub-protocols. On the other hand, this paper provides formal security proofs of the underlying sub-protocols as well as the PPkNN protocol under the semihonest model. Additionally, we discuss various techniques through which the proposed PPkNN protocol can possibly be extended to a protocol that is secure under the malicious setting. Third, our preliminary work in addresses only secure kNNquery which is similar to Stage 1 of PPkNN. However, Stage 2 in PPkNN is entirely new. Finally, our empirical analyses in Section 6 are based on a real dataset whereas the results in are based on a simulated dataset. Furthermore, new experimental results are included in this paper.

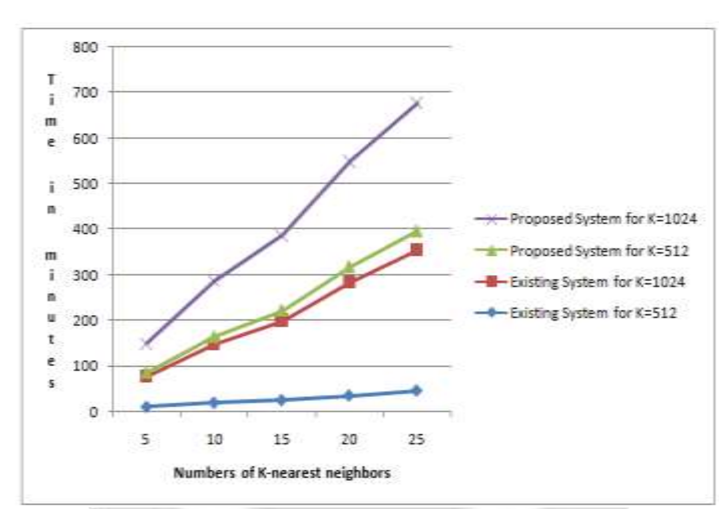

Fig 4.1. Total Cost of Stage 1

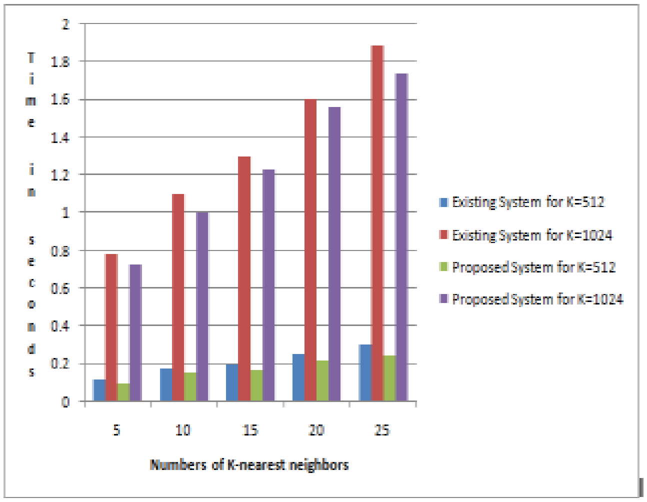

Fig 4.2 Total Cost of Stage 2

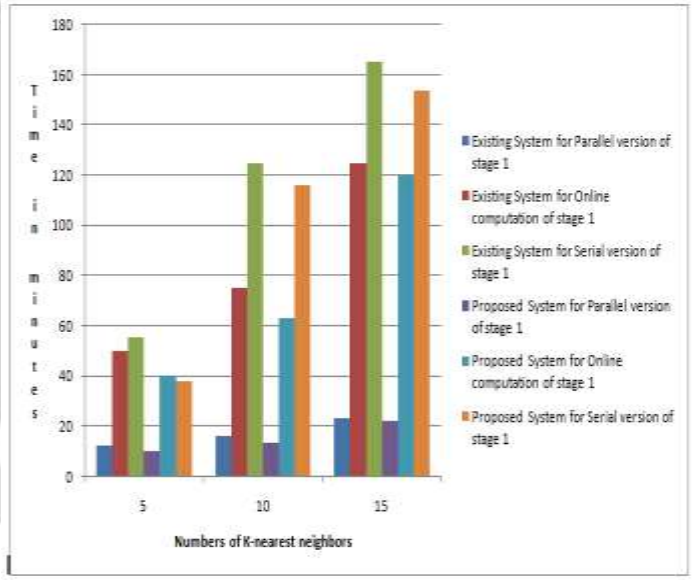

Fig 8.2.3: Efficiency gains of Stage 1

5. CONCLUSIONS

A novel security safeguarding k- NN grouping convention over scrambled information in the cloud. Our convention ensures the secrecy of the information, client's info inquiry, and conceals the information access designs. We likewise assessed the execution of our convention under diverse parameter settings. Since enhancing the productivity of SMINn is an imperative initial step for enhancing the execution of our PPkNN convention, we plan to explore elective what's more, more productive answers for the SMINn issue in our future work. Likewise, we will explore and extend our examination to other grouping calculations.

6. ACKNOWLEDGEMENT

I take this special opportunity to express my sincere gratitude towards professor and all the people who supported me during my entire project work. I would like to express my gratitude to my guide and also the project coordinator Prof. Sonali A.Patil for providing special guidance. I would also like to thank our HOD Dr. G. M. Bhandari who always has enough time to solve student's problems at any time. Finally thanks to all teachers who are always supportive at us.

7. REFERENCES

[1]. P. Williams, R. Sion, and B. Carbunar, ―Building castles out of mud: Practical access pattern privacy and correctness on Untrusted storage,‖ in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.
[2]. B. K. Samanthula, Y. Elmehdwi, and W. Jiang, ―**k-nearest neighbor classification over semantically secure encrypted relational data**,‖ e-print arXiv: 1403.5001, 2014.
[3]. D. Bogdanov, S. Laur, and J. Williamson, ―Sharemind: **A framework for fast privacy-preserving computations**,‖ in Proc. 13th Eur. Symp. Res. Compute. Security: Compute. Security, 2008, pp. 192–206.
[4]. A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, ―**Privacy preserving mining of association rules**,‖ Inf. Syst., vol. 29, no. 4, pp. 343–364, 2004
[5]. M. Kantarcioglu and C. Clifton, ―**Privately computing a distributed k-nn classifier**,‖ in Proc. 8th Eur. Conf. Principles Practice Know. Discovery Databases, 2004, pp. 279–290.
[6]. L. Xiong, S. Chitti, and L. Liu, ―**K nearest neighbor classification across multiple private databases**,‖ in Proc. 15th ACM Int. Conf. Inform. Know. Manage. 2006, pp. 840–841.
[7]. R Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "**Secure k-nearest neighbor query over encrypted data in outsourced environments**," in Proc. IEEE 30th Int. Conf. Data Eng., 2014, pp. 664–675.
[8]. A. Shamir, ―**How to share a secret**,‖ Common. ACM, vol. 22, pp. 612–613, 1979.
[9]. Y. Lindell and B. Pinkas, ―**Privacy preserving data mining**,‖ in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
[10]. P. Paillier, ―**Public key cryptosystems based on composite degree residuosity classes**,‖ in Eurocrypt, pp. 223–238, 1999.
[11]. C. Gentry, ―**Fully homomorphic encryption using ideal lattices**,‖ in ACM STOC, pp. 169–178, 2009.
[12]. A. Shamir, ―**How to share a secret**,‖ Commun. ACM, vol. 22,pp. 612–613, Nov. 1979.