ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Secure K-Nearest Regional Query In Excess of Semantically Protected Encrypted Relational Data

S.Md Ismail, M.Tech
Associate Professor,
Department of CSE,
Al-Habeeb College of Engineering &Technology, Chevella, Hyderabad.

Mr.Mohd Anwar Ali, M.Tech
Associate Professor & HOD,
Department of CSE,
Al-Habeeb College of Engineering &Technology, Chevella, Hyderabad.

Vankudoth Saidulu
Post Graduate Student,
Department of CSE,
Al-Habeeb College of Engineering &Technology, Chevella, Hyderabad.

ABSTRACT:

Data Mining has wide applications in countless areas such as banking, medicine, scientific research and among supervision agencies. Arrangement is one of the generally worn tasks in data drawing out applications. For the precedent decade, due to ascend a range of privacy issues, many speculative and sensible solutions to the arrangement quandary have been projected beneath dissimilar protection models. However, with the recent reputation of obscure computing, users now have the occasion to subcontract their data, in encrypted form, as well as the data mining tasks to the cloud. Since the data on the cloud is in encrypted form, existing privacy-preserving categorization techniques are not appropriate. In this paper, I have spotlight on solving the cataloging problem over encrypted data. In particular, I have proposed a secure k-NN classifier over encrypted data in the cloud. The proposed protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. To the best of our knowledge, our work is the first to develop secure k-NN classifier over encrypted data under the semi-honest model. Also, we empirically analyze the efficiency of our proposed protocol using a real-world dataset under different parameter settings.

Index Terms:

Protection, k-NN classifier, outsourced databases, encryption.

1.INTRODUCTION:

A moment ago the cloud computing archetype is revolutionize the organization method of in commission their data predominantly in the way they stockpile, access and progression data.

As an budding computing paradigm, cloud computing attracts many organizations to consider seriously regarding cloud potential in terms of its cost-efficiency, flexibility, and divest of administrative overhead. Most often, organizations delegate their computational operations in addition to their data to the cloud. Regardless of 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 difficult without ever decrypting the records.

1.1.Objective of the Project:

Data Mining has wide applications in many areas such as banking, medicine, scientific research and among government agencies. Classification is one of the commonly used tasks in data mining applications. For the past decade, due to the rise of various privacy issues, many theoretical and practical solutions to the classification problem have been proposed under different security models. However, with the recent popularity of cloud computing, users now have the opportunity to outsource their data, in encrypted form, as well as the data mining tasks to the cloud. Since the data on the cloud is in encrypted form, existing privacy-preserving classification techniques are not applicable.

Efficient Public-Key Cryptosystems Provably Secure Against Active Adversaries:

This paper proposes two new public-key cryptosystems semantically secure against adaptive chosen-cipher text attacks. Inspired from a recently discovered trapdoor technique based on composite-degree residues, our

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

converted encryption schemes are proven, in the random oracle model, secure against active adversaries (NM-CCA2) under the assumptions that the Decision Composite Residuosity and Decision Partial Discrete Logarithms problems are intractable. We make use of specific techniques that differ from Bellare-Rogaway or Fujisaki-Okamoto conversion methods. Our second scheme is specifically designed to be efficient for decryption and could provide an elegant alternative to OAEP.Proposed two new public-key cryptosystems provably semantically secure against adaptive chosen-cipher text attacks i.e. secure in the sense of NM-CCA2. Computationally efficient for decryption, one of them could provide an alternative to OAEP. A typical research topic would be to ensure security against active adversaries relatively to the computational related problems CR and PDL. Another (independent) direction consists in improving their decryption throughputs by accelerating computations modulo p2, possibly using appropriate modular techniques.

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. The existing techniques are not applicable to outsourced database environments where the data resides in encrypted form on a third-party server. This paper proposed a novel privacy-preserving k-NN classification protocol over encrypted data in the cloud. Our protocol protects the confidentiality of the data, user's input query, and hides the data access patterns. We also evaluated the performance of our protocol under different parameter settings. Since improving the efficiency of SMINn is an important first step for improving the performance of our PPkNN protocol, we plan to investigate alternative and more efficient solutions to the SMINn problem in our future work.Also, we will investigate and extend our research to other classification algorithms.

2.Isolation Preserving Partition Data Mining:

Privacy-Preserving Data Mining – developing models without seeing the data is receiving growing attention. This paper assumes a privacy-preserving distributed data mining scenario: data sources collaborate to develop a global model, but must not disclose their data to others. Often, when legal/commercial reasons restrict sharing data, it may be imprudent to share models generated from the data. We have presented a method that bypasses this restriction. Space restrictions preclude a detailed analysis of the communication cost.

For nominal attributes, assuming k classes and r values for the attributes, protocol 1 is O(rkn); this is reasonable for small values of r and k (where Na¨ıve Bayes is most effective), while building the tree for numeric attributes is O(kn). Evaluating the tree requires an operation for each attribute, where the operations are constant (although non-trivial, dominated by the cost of the secure in protocol). Future work will address the practical cost of this method, using tools such as hardware cryptographic accelerators. This paper is based on the semi-honest model.While the components can be extended to the malicious model, doing so efficiently is an interesting research problem. In general, the efficiency of privacy-preserving protocols is open – most are significantly more expensive than non-privacy-preserving protocols for the same problem. Progress in this area will enable application of data mining to opportunities that are currently unexplored due to privacy and security concerns.

2.1 Privacy Preserving Mining of Association Rules:

A framework for mining association rules from transactions consisting of categorical items where the data has been randomized to preserve privacy of individual transactions. While it is feasible to recover association rules and preserve privacy using a straightforward \uniform" randomization, the discovered rules can unfortunately b e exploited to and privacy breaches. We analyze the nature of privacy breaches and propose a class of randomization operators that are much more elective than uniform randomization in limiting the breaches. We derive formulae for an unbiased supp ort estimator and its variance, which allow us to recover item set supports from randomized datasets, and show how to incorporate these formulae in to mining algorithms. Finally, we present experimental results that validate the algorithm by applying it on real dataset.

2.2 Data Privacy through Optimal k-Anonymization:

Data de-identification reconciles the demand for release of data for research purposes and the demand for privacy from individuals. This paper proposes and evaluates an optimization algorithm for the powerful de-identification procedure known as -anonymization. A -anonymized dataset has the property that each record is indistinguishable from at least others.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Even simple restrictions of optimized -anonymity are NP-hard, leading to significant computational challenges. We present a new approach to exploring the space of possible anonymizations that tames the combinatory of the problem, and develop data-management strategies to reduce reliance on expensive operations such as sorting. Through experiments on real census data, we show the resulting algorithm can find optimal anonymizations under two representative cost measures and a wide range. We also show that the algorithm can produce good anonymizations in circumstances where the input data or input parameters preclude finding an optimal solution in reasonable time. Finally, we use the algorithm to explore the effects of different coding approaches and problem variations on anonymization quality and performance. To our knowledge, this is the first result demonstrating optimal -anonymization of a non-trivial dataset under a general model of the problem.

2.3 Query Processing over Encrypted Data:

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 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 protocol under the semi-honest model.

2.3.1 Homomorphism addition

$$D_{sk}(E_{pk}(a+b)) = D_{sk}(E_{pk}(a) * E_{pk}(b) \bmod N^2)$$

2.3.2 Homomorphism multiplication

$$D_{sk}(E_{pk}(a*b)) = D_{sk}(E_{pk}(a)^b \bmod N^2)$$

2.3.3 Semantic security

The encryption scheme is semantic call secure. Briefly, given a set of cipher texts, an adversary cannot deduce any additional information about the plaintext(s).

3. PRIVACY-PRESERVING PRIMITIVES:

3.1. Introduction:

The Systems Development Life Cycle (SDLC), or Software Development Life Cycle in systems engineering, information systems and software engineering, is the process of creating or altering systems, and the models and methodologies that people use to develop these systems. In software engineering the SDLC concept underpins many kinds of software development methodologies. These methodologies form the framework for planning and controlling the creation of an information system the software development process.

3.2. Existing System:

Existing work on privacy-preserving data mining (PPDM) (either perturbation or secure multi-party computation (SMC) based approach) cannot solve the DMED problem. Perturbed data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly sensitive data. Also the perturbed data do not produce very accurate data mining results. Secure multi-party computation based approach assumes data are distributed and not encrypted at each participating party.

3.2.1 Disadvantages:

•Perturbed data do not possess semantic security.

3.3. Proposed System:

Focus on solving the classification problem over encrypted data. In particular, we propose a secure k-NN classifier over encrypted data in the cloud. The proposed protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. To the best of our knowledge, our work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model. Also, we empirically analyze the efficiency of our proposed protocol using a real-world dataset under different parameter settings.Proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. Specifically, we focus on the classification problem since it is one of the most common data mining tasks. Because each classification technique has their own advantage, to be concrete, this paper concentrates on executing the k-nearest neighbor classification method over encrypted data in the cloud computing environment.

3.3.1 Advantages:

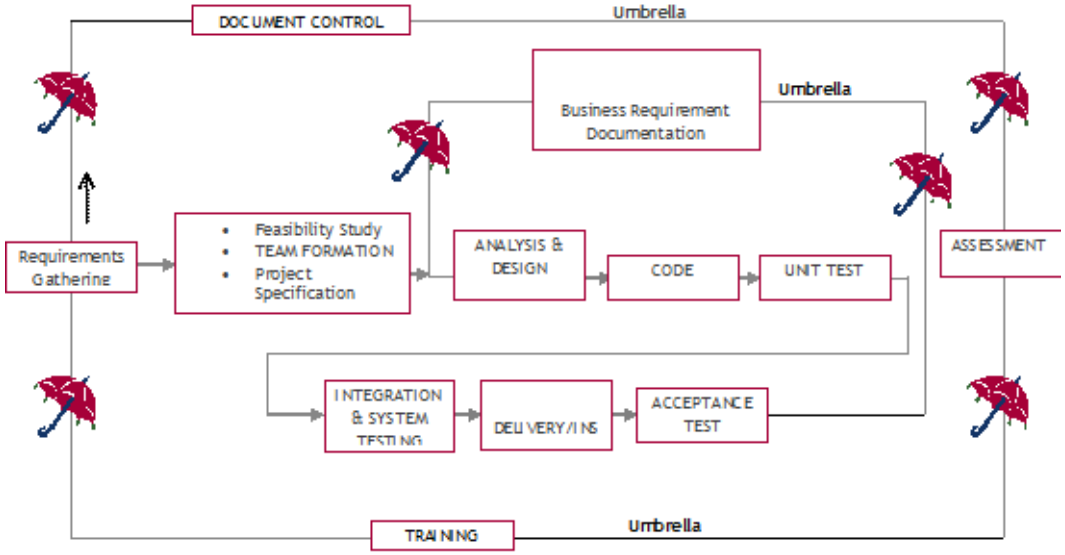
•It protects the confidentiality of data, privacy of user's input query.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

•Provide Hidden data access patterns.

•Data records correspond to the k-nearest neighbors and the output class label are not known to the cloud.

3.4. Process Model Used With Justification
SDLC (Umbrella Model):



SDLC is nothing but Software Development Life Cycle. It is a standard which is used by software industry to develop good software.

3.4.1 Stages in SDLC:

•Requirement Gathering
•Analysis
•Designing
•Coding
•Testing
•Maintenance

4.ASSESSMENT
4.1 Functional Test Plan

Tests Cases: Totally 4 Tests are performed to know different functionalities of k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data. Each Test case has the following parameters

•Test Description
•Input Parameters
•Action
•Expected Results.

Test Case ID: 1
Test Description: Cloudseerver1
Input Parameters: uploads any dataset
Action: Press “Upload” on the web page
Expected Results: Dataset is uploaded.

Test Case ID: 2
Test Description: cloudserver2
Input Parameter: No parameters
Action: keys are generated and automatically sends to server1
Expected Results: Server1 receives keys.

Test Case ID: 3
Test Description: User
Input Parameter: Select buying, maintaintence, door, safty
Action: Click “Submit query” Button.
Expected Results: Entered query is submitted.

Test Case ID: 4
Test Description: Search
Input Parameter: Based on users query it searches.
Action: When clicks “Submit query” button.
Expected Results: Results are displayed.

4.2 Functional Requirements:

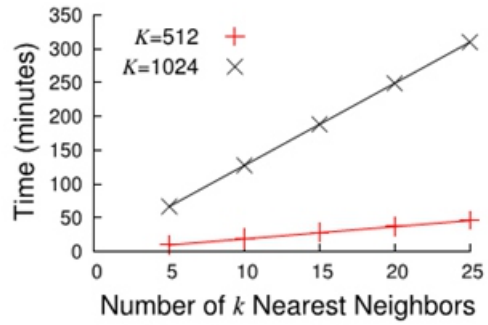
A requirement specifies a function that a system or component must be able to allow the user to perform some kind of function.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

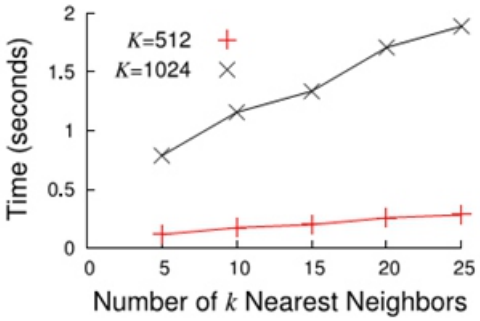
No	Requirement	Requirement Description	Priority
1.	Cloudserver1	This function will upload the dataset and stores the data.	Mandatory Requirement.
2	Cloud server2	This function will generate key values and send it to server1.	Mandatory Requirement.
3	User	Here user enters the query and submits to cloud. From cloud it searches and displays the result.	Mandatory Requirement.

4.3 Proof of Security for SMIN:

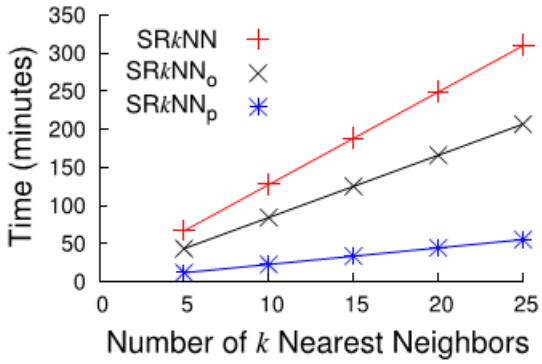
As mentioned in Section 2.3, to formally prove that SMIN is secure under the semi-honest model, we need to show that the simulated image of SMIN is computationally indistinguishable from the actual execution image of SMIN. An execution image generally includes the messages exchanged and the information computed from these messages.



(a) Total cost of Stage 1



(b) Total cost of Stage 2



(c) Efficiency gains of Stage 1

Fig: Computation costs of PPkNN for varying number of k nearest neighbors and encryption key size K

5.Literature survey:

Managing and Accessing Data in the Cloud: Privacy Risks and ApproachesEnsuring proper privacy and protection of the information stored, communicated, processed, and disseminated in the cloud as well as of the users accessing such information is one of the grand challenges of our modern society. As a matter of fact, the advancements in the Information Technology and the diffusion of novel paradigms such as data outsourcing and cloud computing, while allowing users and companies to easily access high quality applications and services, introduce novel privacy risks of improper information disclosure and dissemination. In this paper, we will characterize different aspects of the privacy problem in emerging scenarios. We will illustrate risks, solutions, and open problems related to ensuring privacy of users accessingServices or resources in the cloud, sensitive information stored at external parties, and accesses to such information Effective and efficient solutions for protecting the privacy of the parties interacting in a cloud infrastructure as well as of the data stored and processed are of paramount importance for enabling a widespread exploitation of the cloud technology.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Privacy is however far from being a trivial problem to address and represents a big challenge for all parties that use and develop cloud technology. In this paper, we discussed the main privacy risks that arise in a cloud scenario and illustrated some solutions for addressing them.

Building Castles out of Mud: Practical Access Pattern Privacy and Correctness on Untrusted Storage:

We introduce a new practical mechanism for remote data storage with efficient access pattern privacy and correctness. A storage client can deploy this mechanism to issue encrypted reads, writes, and inserts to a potentially curious and malicious storage service provider, without revealing information or access patterns. The provider is unable to establish any correlation between successive accesses, or even to distinguish between a read and a write. Moreover, the client is provided with strong correctness assurances for its operations. Illicit provider behavior does not go undetected. We built a first practical system – orders of magnitude faster than existing implementations – that can execute over several queries per second on 1Tbyte+ databases with full computational privacy and correctness. In this paper we introduce a first practical oblivious data access protocol with correctness. The key insights lie in new constructions and sophisticated reshuffling protocols that yield practical computational complexity (to O(log n log log n)) and storage overheads (to O (n)). We also introduce a first practical implementation that allows a throughput of sever all queries per second on 1Tbyte+ databases, with full computational privacy and correctness , orders of magnitude faster than existing approaches.

ACKNOWLEDGMENTS:

The authors wish to thank the anonymous reviewers for their invaluable feedback and suggestions. This work has been partially supported by the US National Science Foundation under grant CNS-1011984.

6.CONCLUSIONS:

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. The existing techniques are not applicable to outsourceddatabase environments where the data resides in encrypted form on a third-party server.

This paper proposed a novel privacy-preserving k-NN classification protocol over encrypted data in the cloud. Our protocol protects the confidentiality of the data, user's input query, and hides the data access patterns. We also evaluated the performance of our protocol under different parameter settings. Since improving the efficiency of SMINn is an important first step for improving the performance of our PPkNN protocol, we plan to investigate alternative and more efficient solutions to the SMIN n problem in our future work. Also, we will investigate and extend our research to other classification algorithms.

7. Reference :

[1] P. Mell and T. Grance, "The NIST definition of cloud computing (draft)," NIST Special Publication, vol. 800, p. 145, 2011.
[2] S. De Capitani di Vimercati, S. Foresti, and P. Samarati,"Managing and accessing data in the cloud: Privacy risks and approaches," in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.
[3] 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. Comput. Commun. Security, 2008, pp. 139–148.
[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 Proc. 41st Annu. ACM Sympos. Theory Comput., 2009, pp. 169–178.
[7] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techn.: Adv. Cryptol., 2011, pp. 129–148.
[8] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, pp. 612–613, 1979.
[9] D. Bogdanov, S. Laur, and J. Willemson, "Sharemind: A framework for fast privacy-preserving computations," in Proc. 13th Eur.Symp. Res. Comput. Security: Comput. Security, 2008, pp. 192–206.
[10] R. Agrawal and R. Srikant, "Privacy-preserving data mining,"ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.