A NOVEL APPROACH OF KNN CLASSIFIER OVER SEMANTICALLY SECURE ENCRYPTED DATA

Ms.Ashwini R. Garad

Student of ME, Computer Engineering Department, Pune University
JSPM's BSIOTR, Wagholi, Pune, Maharashtra , India.

Prof. Nitin Shivale

Department of Computer Engineering Bhivrabai Sawant Institute of Technology & ResearchWagholi, Pune

ABSTRACT

In the emerging cloud computing paradigm, data owners become increasingly motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. For the consideration of users' privacy, sensitive data have to be encrypted before outsourcing, which makes effective data utilization a very challenging task. 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.I focus on solving the classification problem over encrypted data. In particular, I propose a novel approach of KNN classifier over semantically secure 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.

Keyword : - *Classification, secure Encrypted data, Relational data*

1. Introduction

Recently, the cloud computing paradigm 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. 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.

1.1 Problem Definition:

To develop an application, for classification over secure encrypted data. There are different methods for classification; however each method has its own advantages and disadvantages. This paper uses support vector machine for classification.

Suppose User-1 have a database D of n records $t_1, t_{2, \ldots\ldots}$ t_n and m+1 attributes. Let t_{ij} denotes the j^{th} attribute value of record $t_{i.}$

User-1 encrypts the database attribute-wise. The encrypted database D' will be outsourced to the cloud.

Let User-2 be an authorized user who want to classify the input record $q=q_1,..q_m$ by applying the SVM classification based on D’.
This process use PPSVM protocol, it is defined as,
PPSVM(D’,q) → c_q
Where c_q denotes the class label for q after applying SVM classification on D’ and q

1.2 Literature Survey:

1. ***Public-Key Cryptosystems Based on Composite Degree Residuosity Classes [1]***

This paper [1] investigates a novel computational problem, namely the Composite Residuosity Class Problem, and its applications to public-key cryptography. This paper proposes a new trapdoor mechanism and derive from this technique three encryption schemes : a trapdoor permutation and two homomorphic probabilistic encryption schemes computationally comparable to RSA. The cryptosystems, based on usual modular arithmetic, are provably secure under appropriate assumptions in the standard model.

2. ***Fully Homomorphic Encryption Using Ideal Lattices [2]***

This paper propose a fully homomorphic encryption scheme – i.e., a scheme that allows one to evaluate circuits over encrypted data without being able to decrypt. This solution comes in three steps. First, it provide a general result – that, to construct an encryption scheme that permits evaluation of arbitrary circuits , it suffices to construct an encryption scheme that can evaluate (slightly augmented versions of) its own decryption circuit ; this call a scheme that can evaluate its (augmented) decryption circuit bootstrappable. Next, it describes a public key encryption scheme using ideal lattices that is almost bootstrappable. Lattice-based cryptosystems typically have decryption algorithms with low circuit complexity, often dominated by an inner product computation that is in NC1. Also, ideal lattices provide both additive and multiplicative homomorphisms (modulo a public-key ideal in a polynomial ring that is represented as a lattice), as needed to evaluate general circuits. Unfortunately, initial scheme is not quite bootstrappable – i.e., the depth that the scheme can correctly evaluate can be logarithmic in the lattice dimension, just like the depth of the decryption circuit, but the latter is greater than the former. In the final step, show how to modify the scheme to reduce the depth of the decryption circuit, and thereby obtain a bootstrappable encryption scheme, with-out reducing the depth that the scheme can evaluate. Abstractly, we accomplish this by enabling the en-crypter to start the decryption process, leaving less work for the de-crypter, much like the server leaves less work for the de-crypter in a server-aided cryptosystem.

3. ***Sharemind: a framework for fast privacy-preserving computations [3]***

Gathering and processing sensitive data is a difficult task. In fact, there is no common recipe for building the necessary information systems. This paper, present a provably secure and efficient general-purpose computation system to address this problem. The solution—SHAREMIND—is a virtual machine for privacy-preserving data processing that relies on share computing techniques. This is a standard way for securely evaluating functions in a multi-party computation environment. The novelty of our solution is in the choice of the secret sharing scheme and the design of the protocol suite. The protocols of SHAREMIND are information-theoretically secure in the honest-but-curious model with three computing participants. Although the honest-but-curious model does not tolerate malicious participants, it still provides significantly increased privacy preservation when compared to standard centralised databases.

4. ***Privacy Preserving Data Mining [4]***

This paper addresses the issue of privacy preserving data mining. Specifically, consider a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. This work is motivated by the need to both protect privileged information and enable its use for research or other purposes. The above problem is a specific example of secure multi-party computation and as such, can be solved using known generic protocols. However, data mining algorithms are typically complex and, furthermore, the input usually consists of massive data sets. The generic protocols in such a case are of no practical use and therefore more efficient protocols are required. Focus on the problem of decision tree learning with the popular ID3 algorithm. The protocol is considerably more efficient than generic solutions and demands both very few rounds of communication and reasonable bandwidth.

5. ***Data Privacy Through Optimal k-Anonymization [5]***

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 k-anonymization. A k-anonymized dataset has the property that each record is indistinguish-able from at least others. Even simple restrictions of optimized k-anonymity are NP-hard, leading to significant computational challenges. This paper present a new approach to exploring the space of possible anonymizations that tames the combinatorics of the problem, and develop data-management strategies to reduce reliance on expensive operations such as sorting. Through experiments on real census data, and show the resulting algorithm can find optimal k-anonymizations under two representative cost measures and a wide range of k. 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, the algorithm is used to explore the effects of different coding approaches and problem variations on anonymization quality and performance.

6. *Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism [6]*

Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. This paper, propose a holistic and efficient solution that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism. The framework is scalable to large datasets by leveraging an index-based approach. Based on this framework, secure protocols for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index. Moreover, several optimization techniques are presented to improve the efficiency of the query processing protocol.

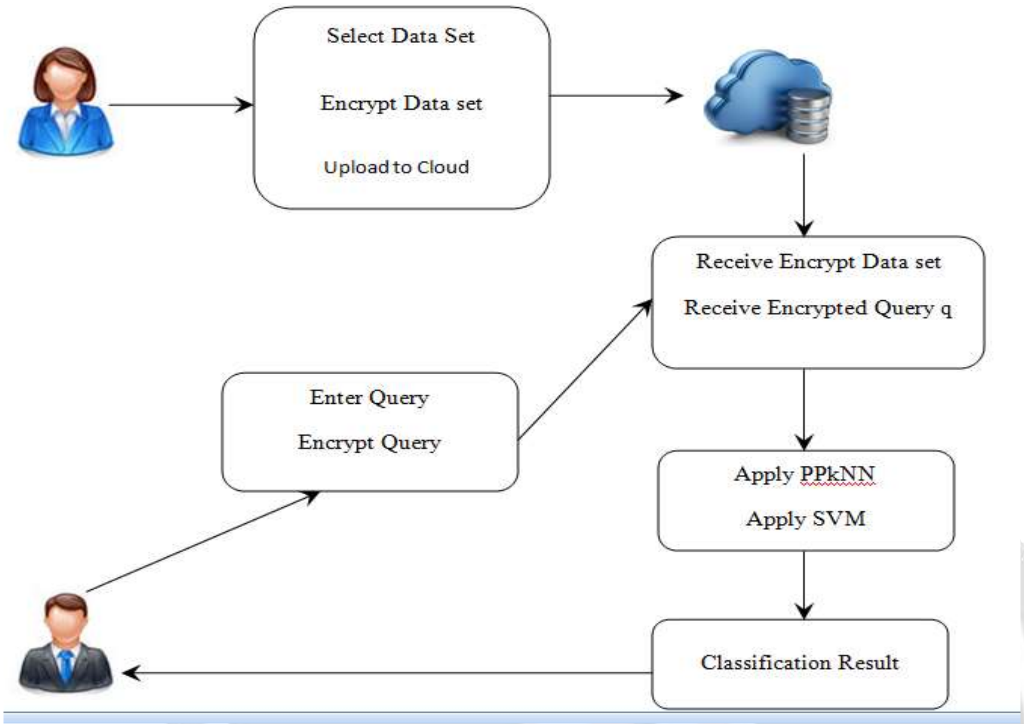
7. *Secure Multidimensional Range Queries over Outsourced Data [7]*

This paper studies the problem of supporting multidimensional range queries on encrypted data. The problem is motivated by secure data outsourcing applications where a client may store his/her data on a remote server in encrypted form and want to execute queries using server’s computational capabilities. The solution approach is to compute a secure indexing tag of the data by applying bucketization (a generic form of data partitioning) which prevents the server from learning exact values but still allows it to check if a record satisfies the query predicate. Queries are evaluated in an approximate manner where the returned set of records may contain some false-positives. These records then need to be weeded out by the client which comprises the computational overhead of our scheme. Author develop a bucketization procedure for answering multidimensional range queries on multidimensional data. For a given bucketization scheme we derive cost and disclosure-risk metrics that estimate client’s computational overhead and disclosure-risk respectively. Given a multidimensional dataset, its bucketization is posed as an optimization problem where the goal is to minimize the risk of disclosure while keeping query cost (client’s computational overhead) below a certain user- specified threshold value. Author provides a tunable data bucketization algorithm that allows the data owner to control the tradeoff between disclosure risk and cost.

8. *Secure KNN Computation on Encrypted Databases [8]*

Service providers like Google and Amazon are moving into the SaaS (Software as a Service) business. They turn their huge infrastructure into a cloud-computing environment and aggressively recruit businesses to run applications on their platforms. To enforce security and privacy on such a service model, we need to protect the data running on the platform. Unfortunately, traditional encryption methods that aim at providing \unbreakable" protection are often not adequate because they do not support the execution of applications such as database queries on the encrypted data. This paper discusses the general problem of secure computation on an encrypted database and proposes a SCONEDB (Secure Computation ON an Encrypted DataBase) model, which captures the execution and security requirements. As a case study, author focus on the problem of k-nearest neighbor (KNN) computation on an encrypted database. Author develops a new asymmetric scalar-product-preserving encryption (ASPE) that preserves a special type of scalar product. Author use APSE to construct two secure schemes that support KNN computation on encrypted data; each of these schemes is shown to resist practical attacks of a different background knowledge level, at a different overhead cost. To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade.

2.Architecture



3. CONCLUSIONS

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 approach of KNN classifier over semantically secure encrypted data in the cloud. Our protocol protects the confidentiality of the data, user's input query, and hides the data access patterns.

4. 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 Prof.Nitin Shivale and project coordinator Prof. Archana Lomte for providing special guidance. I would also like to thank our HOD Prof. 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.

5. REFERENCES

[1] P. Paillier, “Public key cryptosystems based on composite degree residuosity classes,” in Eurocrypt, pp. 223–238, 1999.

[2] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in ACM STOC, pp. 169–178, 2009

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

[4]Y. Lindell and B. Pinkas, “Privacy preserving data mining,” in Advances in Cryptology (CRYPTO), pp. 36–54, Springer, 2000

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

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

[7] 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

[8] 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.