Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

PRIVACY ENSURED HEALTHCARE DATA ANALYSIS UNDER CLOUDS

Ms. T. Vijaya Saratha[1], Ms. V. Deepa[2], Ms. K.K. Kavitha[3]
[1]M.Sc., MPhil., Assistant Professor (Computer Science)
[2]MPhil., Research Scholar,
[3]MCA., MPhil., Assistant Professor (Computer Science)
Selvamm Arts & Science College (Autonomous), Namakkal, Tamilnadu, India

Abstract - Computational and data resources are shared through the Internet under the Cloud computing environment. Demand based service provisioning is available on the cloud. User data values are passed into the computational tasks under the remote machines. Data control is limited in the remote data access process. Data categorization methods are used to identify the normal and anomaly values. Categorization tasks are performed using the neural network techniques. Neural network operations are carried out suing Back-Propagation Neural Network (BPNN) algorithm. Learning and testing operations are performed in the neural network process.

The Back-Propagation Neural Network (BPNN) learning operations are carried out using collaborative learning method. Privacy preserved data analysis is carried out with Data Owner, Trusted Authority (TA) and Cloud Server elements. Cryptographic key values are issued by the Trusted Authority. The learning process is initiated by the data owners. Cloud resources are provided by the cloud server for the learning process. Cloud environment processes the cipher text uploaded by the data owner. The encrypted results are passed to the data owners. Encrypted data values are decrypted by the data owners. Sensitive data values are protected using Boneh, Goh and Nissim (BGN) doubly homomorphic encryption algorithm. Secure scalar Product and addition protocols are adapted to protect the intermediate data values.

Health care data classification is performed without the Trusted Authority. Distributed key generation mechanism is applied in the system. User and data authentication tasks are handled by the cloud server. Resource scheduling mechanism is also integrated with the health data security scheme. Privacy preserved BPN learning process is tuned with cloud resource allocation process.

I. INTRODUCTION

Cloud computing as an emerging technology is expected to reshape information technology processes in the near future [4]. Due to the overwhelming merits of cloud computing, e.g., cost-effectiveness, flexibility and scalability, more and more organizations choose to outsource their data for sharing in the cloud. As a typical cloud application, an organization subscribes the cloud services and authorizes its staff to share files in the cloud. Each file is described by a set of keywords and the staff, as authorized users, can retrieve files of their interests by querying the cloud with certain keywords. In such an environment, how to protect user privacy from the cloud is a third party outside the security boundary of the organization becomes a key problem. With the recent adoption and diffusion of the data outsourcing paradigm, where data owners store their data on external servers, there have been increasing general demands and concerns for data confidentiality. Besides well-known risks of confidentiality and privacy breaks, threats to outsourced data include improper use of information: the server could use substantial parts of a collection of data gathered and organized by the data owner, potentially harming the data owner's market for any product or service that incorporates that collection of information. Traditional access control architectures assign a crucial role to the reference monitor [3] for ensuring

data confidentiality. The reference monitor is the system component responsible of the validation of access requests. The data outsourcing scenario challenges one of the basic tenets of traditional access control architectures, where a trusted server is in charge of defining and enforcing access control policies. This assumption no longer holds here, because the server does not even have to know the access control policy that is defined the data owner. We therefore need to rethink the notion of access control in open environments, where external servers take full charge of the management of the outsourced data and are not trusted with respect to the data confidentiality.

An important opportunity for a revision for the access control architecture can be based on the use of cryptography [2]. Cryptography can be considered as a tool that transforms information in a way that its protection depends only on the correct management of a compact secret. Cryptography is typically used when information is transmitted on a channel, with the assumption that the channel lies outside of the trust boundary of the system. The improvements in cryptographic algorithms, extremely reduces the cost for the use of cryptography for stored data, producing a continuous increase in its adoption. A simple application of cryptography to stored resources can then be based on the well-known correspondence between a network and storage service: both organize the information they have to transfer/store in discrete pieces. A more advanced solution takes into account that the nature of the storage service is different. For instance, in [1] the authors exploit cryptography to the aim of protecting the sensitive information plaintext represented in memory pages when a trusted process accesses it. Indeed, the application of cryptography for the protection of files is today available as an option in most modern operating systems, to make it impossible to access the information without access to the keys stored within the system. Encryption reduces the risk of loss of confidential information deriving from low-level access to the devices. The cryptographic protection can also be used to protect the swap area on disk, to reduce the risk that processes could access information they are not authorized to see by reading the content of swap pages released by processes managing confidential information. Nonetheless, the cryptography options offered by current operating systems have been designed to protect local resources and access control is still realized using the services of a reference monitor.

II. RELATED WORK

A considerable amount of methods for privacy preservation in data mining use cryptography techniques from the Secure Multi-party Computation (SMC) area based on the seminal works by Yao and Goldreich et al. Several relevant operations for SMC were defined. These operations are applied, amongst others, in the following data mining algorithms to ensure privacy presservation: Decision tree induction was enhanced for vertically and horizontally partitioned data to generate decision trees without data disclosure. Privacy preserving association rule mining was proposed clustering methods [8]. A summary of data mining applications and their privacy preserving solutions is given by Vaidya et al. [9]. In the area of neural networks, the aspect of privacy preservation is mostly disregarded.

Wan et al. present a generic formulation for secure computation of gradient descent methods [10]. The authors discuss a multi-party-protocol for vertically partitioned data that can be used to train a neural network. To ensure privacy, the target function is defined as a composition of two functions. Thus, the weights can be adapted locally. A second protocol for a secure summation of two scalar products is also suggested as a part of the overall process.

A privacy preserving version of self organizing maps (SOM) is presented by Han et al. [11]. SOMs belong to the class of unsupervised learning techniques and are applied, e.g., for dimension reduction. The authors present a two-party protocol to adapt the network weights iteratively for vertically partitioned data. Barni et al. address in [12] a two-party privacy preserving protocol for neural network based computation. In their setting, the first party owns the confidential data and the second party owns the confidential model that is applied on the first party's data. Both, the data and the network

model, are kept private. The approach does not consider how the used network model is actually trained but assumes that it already exists. We present our approach which extends the presented works by proposing a multi-party protocol for privacy-preserving neural network learning on horizontally partitioned data.

III. PRIVACY PRESERVED DATA PROCESSING UNDER CLOUD

Despite the potential benefits, one crucial issue pertaining to the Internet-wide collaborative neural network learning is the protection of data privacy for each participant. In particular, the participants from different trust domains may not want to disclose their private data sets, which may contain privacy or proprietary information, to anybody else [6]. In applications such as healthcare, disclosure of sensitive data, for example, protected health information (PHI), is not only a privacy issue but of legal concerns according to the privacy rules such as Health Insurance Probability and Accountability Act (HIPAA). To embrace the Internet-wide collaborative learning, it is imperative to provide a solution that allows the participants, who lack mutual trust, to conduct neural network learning jointly without disclosing their respective private data sets [7]. The solution shall be efficient and scalable enough to support an arbitrary number of participants, each possessing arbitrarily partitioned data sets.

Secure multiparty computation (SMC) can be used to solve problems of this kind. But the extremely high computation and communication complexity of SMC, due to the circuit size, usually makes it far from practical even in the two-party case [5]. To provide practical solutions for privacy preserving Back-Propagation neural (BPN) network learning, three main challenges need to be met simultaneously: 1) To protect each participant's private data set and intermediate results generated during the BPN network learning process, it requires secure computation of various operations, for example, addition, scalar product and the nonlinear sigmoid function, which are needed by the BPN network algorithm; 2) to ensure the practicality of the proposed solution, the computation/ communication cost introduced to each participant shall be affordable. In particular, it shall be able to support an arbitrary number of participants without introducing tremendous computation/communication costs to each participant; 3) for collaborative training, the training data sets may be owned by different parties and partitioned in arbitrary ways rather than a single way of partition.

Collaborative BPN network learning is applied over arbitrarily partitioned data. A trusted authority (TA), the participating parties (data owner) and the cloud servers entities are involved in the privacy preserved mining process. TA is only responsible for generating and issuing encryption/decryption keys for all the other parties. Participating party is the data owner uploads the encrypted data for the learning process. Cloud server is used to compute the learning process under cloud resource environment. Each participant first encrypts their private data with the system public key and then uploads the ciphertexts to the cloud. Cloud servers execute most of the operations in the learning process over the ciphertexts. Cloud server returns the encrypted results to the participants. The participants jointly decrypt the results with which they update their respective weights for the BPN network. Boneh, Goh and Nissim (BGN) doubly homomorphic encryption algorithm is used to secure the private data values. Data splitting mechanism is used to protect the intermediate data during the learning process. Random sharing algorithm is applied to randomly split the data without decrypting the actual value. Secure scalar product and addition operations are used in the encryption and decryption process. The following issues are identified from the privacy preserved data process in cloud environment. Centralized key distribution model, Malicious party attacks are not handled, Noisy data upload is not controlled and Resource allocation and data distribution is not optimized.

IV. PRIVACY ENSURED HEALTHCARE DATA ANALYSIS

The collaborative learning process is handled without the Trusted Authority (TA). Key generation and issue operations are carried out in a distributed manner. Cloud server is enhanced to verify the user and data level details. Privacy preserved BPN learning process is tuned with cloud resource allocation process. The cloud data analysis process is designed to utilize the cloud resources for the training process. Key aggregation process is used to generate and share the key values. Training is performed under the cloud server with privacy. The system is divided into six major modules. They are cloud server, trusted authority, data provider, upload process, training process and data classification. The cloud server module is designed to provide resources for the clients. Trusted authority module is designed to manage key distribution process. Data provider is designed to share the data in the cloud. Data encryption and upload process are managed under upload process module. Neural network learning process is carried out under the training process module. Data classification module is designed to classify the client data values.

The cloud server manages the user and resource details. User authentication is performed in the cloud server. The cloud server collects resources from different resource providers. Resource scheduling process is used to allocate computational resources to the training process. Trusted Authority (TA) application is used for key management process. Public key and secret key values are generated in the trusted authority. Key values are issued to the data providers. User accounts are verified with cloud server environment. Data provider maintains the shared data values. Noise removal process is applied on the data values. Multiple data providers are involved in the data classification process. Data providers are referred as data owner or parties. Shared data values are uploaded from the data provider to the cloud server. Encryption process is carried out to secure the sensitive attributes. Boneh, Goh and Nissim (BGN) doubly homomorphic algorithm is used for the encryption process. The data provider uses two types of key generation models. They are Trusted Authority (TA) based key model and Distributed key model. Trusted Authority generates and issues the key value to the data provider. Aggregation based key generation mechanism is used in distributed key model. Labeled transaction data values are collected and updated in the cloud server.

Resource scheduling process is initiated in the cloud server for the training process. Back Propagation Neural network (BPN) algorithm is used for the training process. Random sharing algorithm is used in the data splitting process to secure the intermediate data values. Training process results are redirected to the data provider. Trained data values are collected from the cloud server. Data provider decrypts the trained data values. Data encryption/decryption tasks are carried out using secure scalar product and addition mechanism. Test data values are compared with the trained data values for the class assignment process.

V. EXPERIMENTAL ANALYSIS

The cloud data security system is constructed to share computational resources and data values with security and privacy features. The system is build with three major components. They are the cloud server, trusted authority and the data owner applications. The cloud server is deployed to provide computational resources to the data owners. The trusted authority is employed to provide the key values for the data owners. The data owner's application is constructed to handle the data share and computational tasks. The data owner uploads the shared data values to the cloud server with security and privacy. The cloud server performs the computational tasks on the data values that are received from the data owners. The computational tasks are carried out on the encrypted data values. The cloud server returns the results to the data owners in encrypted form. The data owner decrypts the processed data values and applies it to the analysis process.

Data mining methods are applied to discover the hidden knowledge from large databases. Classification techniques are used to assign the labels for the transactions. Learning and testing operations are carried out under the classification process. Learning or training process is called to identify the patterns from the labeled transactions. The testing process is called to identify the class labels using the learned patterns. The Back Propagation Neural Network (BPNN) algorithm is used for the classification process. The BPNN operations are divided into two phase. The learning or training phase is carried out under the cloud server. The testing phase is performed under the data owner environment. Multiple data owners share their data values to the cloud server. The data upload process is secured with homomorphic cryptography methods. All the security operations are carried out using the key values.

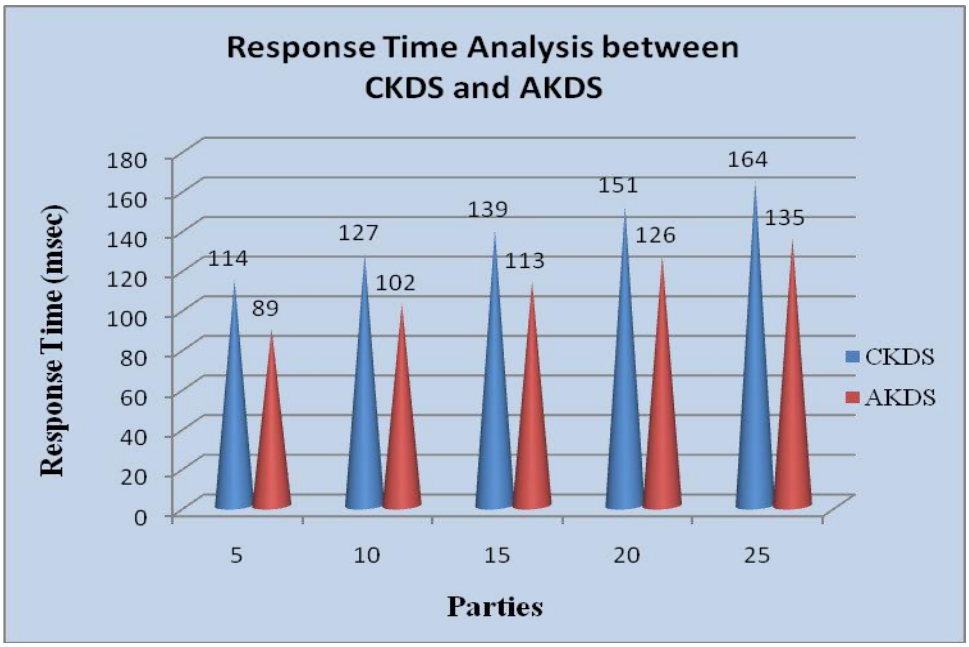

Figure No: 5.1. Response Time Analysis between CKDS and AKDS

The cloud data security is provided using Boneh, Goh and Nissim (BGN) algorithm. The BGN algorithm is doubly homomorphic algorithm. The secure scalar product and secure scalar sum methods are also used to secure the intermediate values. The data values are uploaded from multiple data owners. Security and privacy are provided for the sensitive attributes. Secure Hash Algorithm (SHA) is employed to verify the data integrity values. The Centralized Key Distribution Scheme (CKDS) is adapted to manage the key generation and distribution process in the existing system. Trusted Authority (TA) is used for the key management process in the Centralized Key Distribution Scheme (CKDS). The proposed system is constructed with the Aggregation based Key Distribution Scheme (AKDS). In the Aggregation based Key Distribution Scheme (AKDS) the key generation and distribution operations are carried out with the data owner involvement. Trusted Authority (TA) is eliminated in the Aggregation based Key Distribution Scheme (AKDS).

The cloud data security scheme is enhanced with resource scheduling methods. The system also handles the anonymous and malicious attacks. User authentication process is integrated with the cloud server to verify the user data upload process. Anonymous and malicious requests are rejected with reference to the authentication process. The classification operations are carried out with different parties. Heart disease diagnosis data values are used in the system performance analysis. Each data

owner or party uploads diagnosis data for different country. The sensitive values are encrypted and uploaded for the training process. The classification operations are carried out under the data owner environment.

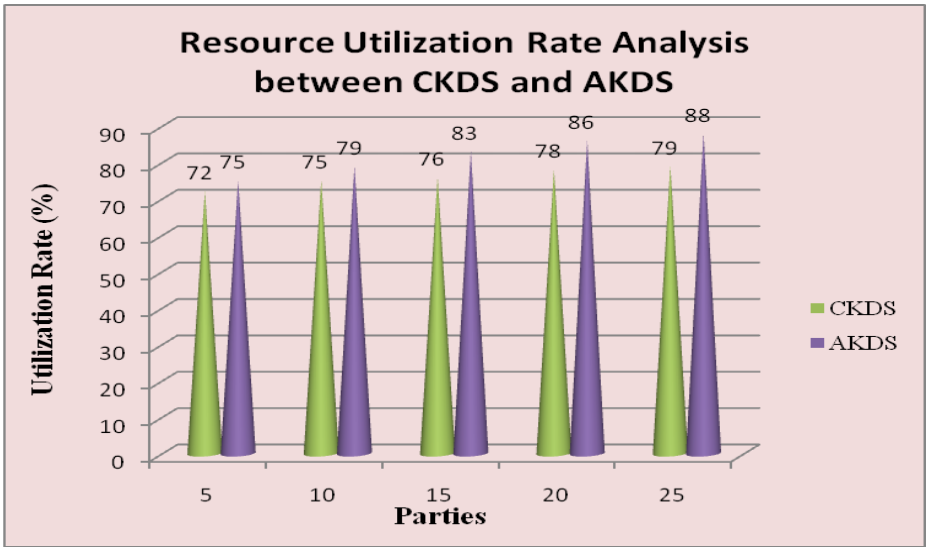

Figure No:5.2. Resource Utilization Rate Analysis between CKDS and AKDS

The Centralized Key Distribution Scheme (CKDS) and Aggregation based Key Distribution Scheme (AKDS) are used in the testing process. The performance analysis is carried out with different resource level and data owner count. The system uses the response time and resource utilization rate for the performance process. The response time analysis for the Centralized Key Distribution Scheme (CKDS) and Aggregation based Key Distribution Scheme (AKDS) techniques are shown in figure 5.1. The AKDS reduces the average response time 20% than the Centralized Key Distribution Scheme (CKDS). The resource utilization rate analysis for the Centralized Key Distribution Scheme (CKDS) and Aggregation based Key Distribution Scheme (AKDS) are shown in figure 5.2. The resource utilization rate in Aggregation based Key Distribution Scheme (AKDS) is 10% increased than the Centralized Key Distribution Scheme (CKDS).

VI. CONCLUSION AND FUTURE WORK

The cloud resource management and privacy preserved data classification scheme is designed for the cloud environment. Multi party based collaborative leaning scheme is used for privacy preserved Back Propagation Neural network. Data privacy is ensured with encrypted data learning process using cloud resources. Privacy preserved BPN learning scheme is enhanced without using the Trusted Authority for key management process. The system also handles the malicious party attacks in the learning process. Collaborative learning model improves the classification accuracy level. The system reduces the computational and communication cost in privacy preserved data classification process. Data privacy is improved in all parties. Key generation and issue load is minimized in the aggregation based cryptographic model.

- The privacy preserved cloud data sharing scheme can be enhanced to handle storage space and shared data process.
- The system can be adapted to support wireless cloud platforms.
- The system can be improved with fault tolerant resource scheduling mechanism.
- The system can be upgraded to support cost based resource sharing mechanism.

REFERENCES

I. Ankur Bansal, Tingting Chen and Sheng Zhong, "Privacy Preserving Back-Propagation Neural Network Learning Over Arbitrarily Partitioned Data" IEEE Transactions On Parallel And Distributed Systems, 2013
II. Huiqi Xu and Keke Chen, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation" IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014
III. Luca Ferretti and Mirco Marchetti, "Distributed Concurrent and Independent Access to Encrypted Cloud Databases" IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014
IV. P. Mell and T. Grance, ''The NIST Definition of Cloud Computing (Draft),'' in NIST Special Publication. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2011.
V. Qin Liu, Chiu C. Tan, Jie Wu and Guojun Wang, "Towards Differential Query Services in Cost-Efficient Clouds" IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, June 2014
VI. Ron C. Chiang and H. Howie Huang, "TRACON-Interference-Aware Scheduling for Data-Intensive Applications in Virtualized Environments", IEEE Transactions On Parallel And Distributed Systems, May 2014
VII. Ke Li, Weiming Zhang, Ce Yang and Nenghai Yu, "Security Analysis on One-to-Many Order Preserving Encryption-Based Cloud Data Search", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 9, September 2015
VIII. Thijs Veugen, Robbert de Haan, Ronald Cramer and Frank Muller, "A Framework for Secure Computations With Two Non-Colluding Servers and Multiple Clients, Applied to Recommendations", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 3, March 2015.
IX. Vaidya, J., Clifton, C., Zhu, M.: Privacy Preserving Data Mining. Volume 19 of Advances in Information Security. Springer, 2006.
X. Wan, L., Ng, W.K., Han, S., Lee, "V.C.S.: Privacy-Preservation For Gradient Descent Methods", In Berkhin, P., Caruana, R., Wu, X., eds.: KDD, New York, NY, USA, ACM Press, 775-783, 2007.
XI. Han, S., Ng, W.K.: Privacy-Preserving Self-Organizing Map. In Song, I.Y., Eder, J., Nguyen, T.M., eds.: DaWaK. Volume 4654 of Lecture Notes in Computer Science., Springer, 428-437, 2007.
XII. Barni, M., Orlandi, C., Piva, A.: A Privacy-Preserving Protocol For Neural-Network-Based Computation. In: MM&Sec '06: Proceeding of the 8th workshop on Multimedia and security, New York, NY, USA, ACM Press, 146-151, 2006.

