ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Ensure Data Privacy in Back Propagation Neural Network Learning over Encrypted Cloud Data

M.T.Kiruthika, C.Selvi

PG Scholar, Department of CSE, Velalar College of Engineering and Technology, Erode, Tamilnadu, India

Assistant Professor, Department of CSE, Velalar College of Engineering and Technology, Erode, Tamilnadu, India

ABSTRACT: Computational resources and storage resources are shared under the cloud environment through the Internet. In cloud environment users' data are usually processed remotely in unknown machines that users do not own or operate. Neural network techniques are used for the classification process. Collaborative Back-Propagation Neural Network (BPNN) learning is applied over arbitrarily partitioned data. The participating parties and the cloud servers are involved in the privacy preserved mining process. Each participant first encrypts their private data and then uploads the cipher texts to the cloud. Cloud servers execute most of the operations in the learning process over the cipher texts. Secure scalar product and addition operations are used in the encryption and decryption process. 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 BPNN learning process is tuned with cloud resource allocation process.

KEYWORDS: Back Propagation, Collaborative learning, Computational resources, privacy preserved mining, Neural Network.

I.INTRODUCTION

Cloud computing is the long dreamed vision of computing as a utility, where users can remotely store the data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. Data outsourcing users can relieve from the burden of local data storage and maintenance. The fact that users no longer have physical possession of the possibly large size of outsourced data makes the data integrity protection in cloud computing a very challenging and excellent task, especially for users with constrained computing resources and computing capabilities. Cloud computing is sometimes viewed as a reincarnation of the classic mainframe client-server model.

On the other hand, assets are omnipresent scalable, highly virtualized. It include all the conventional threats, as well as novel ones. In mounting solution to cloud computing security issues it may be helpful to identify the problems and approaches in terms of Loss of control, Lack of trust, Multi-tenancy problems. To overcome these problems cryptographic techniques are enhanced by the user. Diffie-Hellman key agreement is not based on encryption and decryption, but instead relies on mathematical functions that enable two parties to generate a shared secret key for exchanging information confidentially online.

Essentially, each party agrees on a public value g and a large prime number p . Next, one party chooses a secret value x and the other party chooses a secret value y . Both parties use their secret values to derive public values, g x mod p and g y mod p, and they exchange the public values. Each party then uses the other party's public value to calculate the shared secret key that is used by both parties for confidential communications. A third party cannot derive the shared secret key because they do not know either of the secret values, x or y . Diffie-Hellman key exchange is widely used with varying technical details by Internet security technologies, such as IPSec and TLS, to provide secret key exchange for confidential online communications.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Neural Network learns by adjusting the weights so as to be able to correctly classify the training data and hence, after testing phase, to classify unknown data. Neural Network needs long time for training. Neural Network has a high tolerance to noisy and incomplete data. Back Propagation learns by iteratively processing a set of training data (samples).For each sample, weights are modified to minimize the error between network's classification and actual classification.

II.RELATED WORK

Neural Networks have been an dynamic study region for decades. Trained neural networks can predict capable outputs which might be tricky to find in the real world. Training neural network from the dispersed data is common: for example, making use of data from many hospitals to train the neural network to predict a certain disease, collecting datasets of purchased items from different grocery stores and training neural network from those to predict a certain pattern of purchased items. When training neural network from distributed data, confidentiality is a foremost burden.

The most admired techniques in machine learning is a multi-layer neural networks, in which the privacy protection trouble is far-off from being virtually solved. A preface approach is planned to facilitate privacy protection for gradient descent methods in common. In stipulations of multi-layer neural networks, the set of rules are inadequate as it is only for one simple neural network pattern with one node in the output layer and no hidden layers. Even though the procedure is well-designed in its simplification, it may be very controlled in practice for privacy preserving multi-layer neural networks.

Privacy preserving computation of activation function is a highly challenging problem. Because most of the recurrently used activation functions are immeasurable and incessant while cryptographic tools are defined in fixed fields. To conquer this complexity, first cryptographic method is used to securely compute sigmoid function, in which an existing piecewise linear approximation of the sigmoid function is used. Homomorphic encryption based approach and the ElGamal cryptographic scheme makes the algorithm practical. These algorithms are frivolous in conditions of computation and communication overheads.

Privacy protection methods for knowledge detection can be categorized into two groups, data perturbation methods and cryptographic methods. Methods of the first group use data deformation. Methods of the second group are used for mutual model learning: $P \geq 2$ parties put in their data for the learning of a mutual model according to protocols that prevent the disclosure of the contributed data. Cryptographic methods have been projected for supervised learning, among else with decision trees, support vector machines and Bayesian networks, for unsupervised model learning, e.g., with K-means and for association rule discovery. Here the method focus on supervised learning and discuss privacy protection for neural network training. The data contributed between these parties are horizontally partitioned, in the logic that the mutual model is built upon the unification of the contributed datasets.

III.PROBLEM FORMULATION

Collaborative BPN network learning is applied over arbitrarily partitioned data. A trusted authority (TA), the participating parties 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 cipher texts to the cloud. Cloud servers execute most of the operations in the learning process over the cipher texts. 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 problems are identified from the existing system. They are 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. The following drawbacks are identified from the existing system.

- Centralized key distribution model

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

- Malicious party attacks are not handled
- Noisy data upload is not controlled
- Resource allocation and data distribution is not optimized

IV.PROPOSED METHODOLOGY

The mutual learning progression is handled without the Trusted Authority (TA). Mutual BPN network learning is functional over randomly partitioned data. Key production and providing operations are conceded out in a disseminated manner. Key aggregation Mechanism is used for generating keys. In this process the key values are exchanged by using Diffie-Hellman key exchange method. This method relies on mathematical functions that enable the data owners to generate a shared secret key for exchanging information confidentially. Each party agrees on a public value and a large prime number in this exchanging method. The parties can use their secret keys with a symmetric key algorithm to conduct confidential online communications. The secret key is identified only to each party and is not at all noticeable on the others. Data is encrypted and decrypted with this key which makes the datasets more secure. Cloud server is improved to confirm the user and data level details. The noisy data upload is controlled by the server. It executes most of the learning operations over cipher texts. Privacy preserved Back Propagation Neural Network learning process is tuned with cloud resource allocation process.

A. Cloud Server

The cloud server manages the user and resource details. User authentication is performed in the cloud server. The cloud server aggregate resources from different resource providers. Resource scheduling process is used to allocate computational resources to the training process.

B. Data Provider

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.

C. Key Management

Data owners are involved in this process. The participants have no prior knowledge of each other to jointly establish a shared secret key. This key is used to encrypt the data.

D. Upload Process

Shared data values are uploaded from the data provider to the cloud server. The perceptive datasets are protected by the encryption process. Boneh, Goh and Nissim (BGN) doubly homomorphic algorithm is used for the data encryption. Two types of key generation models are used in this process. The 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.

E. Training Process

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 deflect to the data provider.

F. Data Classification

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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

V.PROPOSED ALGORITHM

A. Multiparty Neural Network Learning With Privacy

In multi-party neural network learning the input data sets owned by the parties can be arbitrarily partitioned. The computational and communicational costs on each party shall be practically efficient and the system shall be scalable. Specifically, we consider a 3-layer neural network for simplicity but it can be easily extended to multilayer neural networks. The learning data set for the neural network, which has N samples, is arbitrary partitioned into $Z(Z \geq 2)$ subsets. Each party Ps holds $x_1^m, x_2^m, ..., x_a^m$ and has

$x_{11}^m + x_{12}^m + ... + x_{1Z}^m = x_1^m$

.....

$x_{a1}^m + x_{a12}^m + ... + x_{a1Z}^m = x_{a.1}^m$

Each attribute in sample $\langle x_1^m, x_2^m, \ldots, x_a^m \rangle, 1 \leq m \leq N,$ is possessed by only one party—if Ps possesses $x_k^m, 1 \leq k \leq a,$ then $x_{ks}^m = x_k^m$ otherwise $x_{ks}^m = 0$. In this paper, w_{jk}^h denotes the weight used to connect the input layer node k and the hidden layer node j; w_{ij}^o denotes the weight used to connect the hidden layer node j and the output layer node i, where 1 ≤ k ≤ a, 1 ≤ j ≤ b, 1 ≤ i ≤ c and a, b, c are the number of nodes of each layer. For collaborative learning, the main task for all the parties is to jointly execute the operations defined in the Feed Forward stage and the Back-Propagation stage. During each learning stage, except for the final learned network, neither the input data of each party nor the intermediate results generated can be revealed to anybody other than TA.

To achieve the above goals, the main idea of our proposed scheme is to implement a privacy preserving equivalence for each step of the original BPN network learning algorithm. Different from the original BPN network learning algorithm, our proposed scheme lets each party encrypt her/ his input data set and upload the encrypted data to the cloud, allowing the cloud servers to perform most of the operations, i.e., additions and scalar products. To support these operations over cipher texts, we adopt and tailor the BGN "doubly homomorphic" encryption for data encryption. Nevertheless, as the BGN algorithm just supports one step multiplication over cipher text, the intermediate results shall be first securely decrypted and then encrypted to support consecutive multiplication operations. For privacy preservation the decrypted results known to each party cannot be the actual intermediate values.

We design a secret sharing algorithm that allows the parties to decrypt only the random shares of the intermediate values. The random shares allow the parties to collaboratively execute the following steps without knowing the actual intermediate values. Data privacy is thus well protected. The overall algorithm is described in Algorithm 1 is the privacy preserving equivalence. To support the operations defined in Algorithm 1, we propose three other cloud-based algorithms for secure scalar product and addition, secure random sharing and sigmoid function approximation process. After the entire process of the privacy preserving learning, all the parties jointly establish a neural network representing the whole data set without disclosing any private data to each other.

begin

Input : each p_s 's data set for N data Samples, $x_{1s}^v, x_{2s}^v, \ldots, x_{as}^v, 1 \leq v \leq N, w_{jks}^{hv}$ and w_{ijs}^{ov} for N Samples iteration $_{max}, \eta$,target value t_i

Output : Network with final weights : $w_{jk}^h, w_{i,j}^{\sigma}, 1 \leq k \leq a, 1 \leq j \leq b, 1 \leq i \leq C$

1 **for** iteration = 1,2,…, iteration $_{max}$ **do**

2 **for** v = 1,2,,…,N do

3 // **Feed Forward Stag :** for j=1,2,…b do

4 each P^s obtain

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

5 random shares φ_{vs} for $\sum_{k=1}^{a}(x_{k1}^{v})+x_{k2}^{v}+.....+x_{kZ}^{v})*\left(w_{jk2}^{hv}+w_{jk2}^{hv}+...+w_{jk}^{hv}\right)$

6 all the parties compute the sigmoid function and obtain the random shares $h_{vjs,}$

7 $\sum_{s=1}^{Z}h_{vjs}=h_{vj}$ and $h_{vj}=f\left(\sum_{s=1}^{z}\varphi_{vs}\right)$ where

8 f() is the approximation for the sigmoid function

9 **for** i = 1,2…, c do

10 all the parties and cloud calculate

11 Error = $\frac{1}{2}\sum_{=1}^{c}(t_i-o_i)^2$

12 **if** Error $>$ *threshold* **then**

13 **// Back – Propagation Stage :**

14 **for** i = 1,2…, c do //(step 1)

15 each P_s

16 obtain random shares Δw_{ijs}^{ov} for

17 $\Delta w_{ij}^{av}=\left(-\left(t_{vi}-\sum_{s=1}^{z}o_{vis}\right)*\left(\sum_{s=1}^{Z}h_{vjs}\right)\right)$

18 for j=1,2,…,b do //(step 2)

19 **each P_s**

20 Obtain random shara μ_s^v for

21 $\sum_{i=1}^{c}[\left(\sum_{s=1}^{Z}o_{vis}-t_{vi}\right)*\left(\sum_{s=1}^{Z}w_{ijs}^{ov}\right)]$ //(step3)

22 each P_s

23 Obtains random share κ_s^v for

24 $\sum_{s=1}^{Z}x_{ks}^{v}*\sum_{s=1}^{z}\mu_s^v$ //(step 4)

25 each P_s

26 Obtains random share ϑ_s^v for

27 $\sum_{s=1}^{Z}h_{vjs}*(1-\sum_{s=1}^{Z^s}h_{vjs})$ // **(step 5)**

28 each P_s

29 Obtain random share Δw_{jks}^{hv} for

30 $\Delta w_{jk}^{hv}=\sum_{s=1}^{Z}\kappa_s^v*\sum_{s=1}^{Z}\vartheta_s^v$

31 Each P^s updates $w_{jks}^{hv}-\eta*\Delta w_{jks}^{hv}$

32 **else**

Learning Finished;

Algorithm 1: Privacy Preserving Multi-Party BPN Network Learning Algorithm

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

VI. EXPERIMENTAL RESULTS

A. SECURITY ANALYSIS

The algorithm that allows the multiple parties to perform secure scalar product and homomorphic addition operations on cipher texts using cloud computing. Party encrypts his data with the secret key and uploads the cipher text to the cloud. The cloud server computes the original messages based on their cipher texts.

B. NUMERICAL ANALYSIS

It is hard to guarantee that the final results are always small enough for the pollards lambda method to efficiently decrypt. This is also because the numbers enclosed in the vectors are too large. To conquer this restriction we propose to let the data holders divide the numbers into smaller chunks. And the cloud then decrypts the smaller chunks with which the final results can be recovered. The decryption process can be parallelized for efficiently. The following table represents the learning rate of each datasets.

Dataset	Sample	Class	Learning Rate
Cancer	180	3	0.1
Diabetes	150	2	0.1
Iris	230	3	0.1

Table1: Experiment Datasets and Parameters

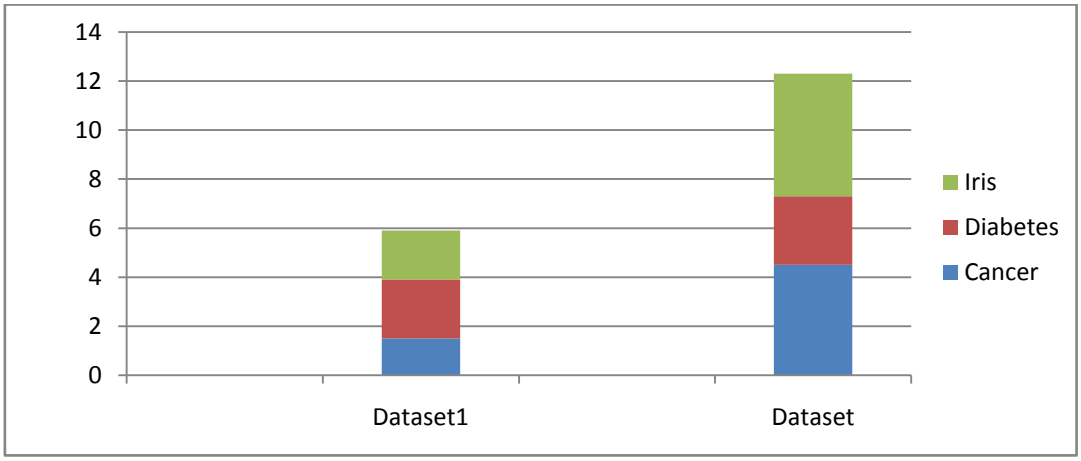

Fig1: Encryption Cost for Different Datasets

VII. CONCLUSION AND FUTURE WORK

The cloud resource management and privacy preserved data classification scheme is designed for the cloud environment. Various party based collaborative leaning schemes are used for privacy conserved Back Propagation Neural network. Data privacy is assured with encoded data learning process using cloud resources. Privacy conserved BPNN learning scheme is improved without using the Trusted Authority for key management process. The system also holds the malicious party attacks in the learning process. Collaborative learning model improves the classification accuracy level. The system scales down 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

In our future work, we will improve the resource allocation process for data classification techniques and to allow the surfacing of other key management techniques to come up with highly proficient and secure key management system in terms of complexity, and authentication overhead.

REFERENCES

1. P. Mell and T. Grance, ‘‘The NIST Definition of Cloud Computing (Draft),’’ in NIST Special Publication. USA: National Institute of Standards and Technology, 2011.
2. R. Curtmola, J. Garay, S. Kamara and R. Ostrovsky, ‘‘Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,’’ in Proc. ACM CCS, 2006.
3. Bethencourt and Waters, ‘‘New Techniques for Private Stream Searching,’’ ACM, Jan. 2009.
4. Qin Liu and Guojun Wang, “Towards Differential Query Services in Cost-Efficient Clouds”, IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, June 2014
5. Q. Liu, J. Wu and G. Wang, ‘‘Efficient Information Retrieval for Ranked Queries in Cost-Effective Cloud Environments,’’ in Proc. IEEE INFOCOM, 2012.
6. J. Bethencourt, Song and B. Waters, ‘‘New Techniques for Private Stream Searching,’’ ACM Trans. Inf. Syst. Security, 2009.
7. Q. Liu, C. Tan, J. Wu and G. Wang, ‘‘Cooperative Private Searching in Clouds,’’ J. Parallel Distrib. Comput., Aug. 2012.
8. X. Yi and E. Bertino, ‘‘Private Searching for Single and Conjunctive Keywords on Streaming Data,’’ in Proc. ACM Soc., 2011.
9. Hore and S. Mehrotra, ‘‘Indexing Encrypted Documents for Supporting Efficient Keyword Search,’’ in Proc. Secure Data Manage., 2012.
10. M. Finiasz and K. Ramchandran, ‘‘Private Stream Search at the Same Communication Cost as a Regular Search: Role of LDPC Codes,’’ in Proc. IEEE ISIT, 2012.