Security Enhanced Federated Learning Approach using Blockchain

S. Revathy
Department of Computer Science & Engineering,
Hindustan Institute of Technology and Science,
Chennai.
srevathy@hindustanuniv.ac.in

S. Sathya Priya
Department of Computer Science & Engineering,
Hindustan Institute of Technology and Science,
Chennai.
sathyap@hindustanuniv.ac.in

2022 International Conference on Computer, Power and Communications (ICCPC) | 979-8-3503-9784-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCPC55978.2022.10072091

***Abstract*— In traditional machine learning approach, data gathered from all the edge devices are sent to centralized server for training and prediction of the output. In the centralized approach, user has to compromise on the data privacy and integrity in sharing their own data to centralized server. To overcome this issue federated machine learning approach was introduced, in which model and data are decentralized and the machine learning model will be trained on the data in local devices and parameters will be sent to cloud server for consensus change, enhancing the data privacy of the users. But still authentication of the nodes to cloud server and vice versa is a major concern to be addressed in federated machine learning as malicious nodes can impersonate as authenticated node and communicate to cloud server. In the proposed model, node authentication is implemented using Ethereum based blockchain with smart contracts thereby enhancing security of Federated machine learning approach. The efficiency of the node authentication is measured and compared with machine learning algorithms which achieves 99% accuracy.**

***Keywords*—Blockchain, Centralized Server, Ethereum, Federated Learning, Security, Smart contracts.**

I. INTRODUCTION

The advancement in technologies play a vital role in improving the standards of the living. The technology improvement necessities developers and industries to innovate new applications with the convergence of latest technologies like Machine learning, Edge computing and Blockchain. The advanced connection capabilities, low cost and small size nature of the various devices impulses the developers to invent applications with advanced features. As the usage of these application increases, concerns of privacy and confidentiality of the data generated by these devices also increases.

The smart applications [9] that are developed with latest technologies requires large amount of user's data for training and deployment. Most of these applications share the collected data to the centralized server, which is used to aggregate and analyze the data collected by multiple devices and provide personalized services to the end users. The cloud service providers commercialize the data to third party applications who needs to publicize their products to the end users by selling their customer's data or share them with limited access through the Application Programming Interfaces which is prone to data inference attack.

The inference attack is known to be a data mining technique, in which the attacker is able to infer more sensitive information about the user without directly accessing it. It analyses the data and illegally gains knowledge about the individual user or the database thereby reducing the data privacy. The data collected by these devices are sensitive in nature like personal or behavioral information [15] including geographical information, age, gender, health information and touch screen inputs and hence it is crucial to handle the sensitive data by warranting the privacy of the users. The primary target of the attacker or intruder is to take control of the centralized cloud server and interrupt the service by initiating active and passive attacks.

The active attack is attack which directly causes damage to the resources involved in the network whereas passive attack is attack which indirectly eavesdrop the network and collects the traffic information that leads to the active attack. Some of the active attacks are
1. *DDOS attack*-In this attack malicious node sends continuous requests to authenticated node making its resources unavailable to the other nodes.
2. *Sybil attack*-The malicious node carries the identity of the authenticated node [10] and damage the resources.
3.*Worm hole attack*-In this attack the attacker places the malicious node in the network and make them to send and receive the information with authenticated nodes.

The passive attack does not damage any network resources but it eavesdrops the other authenticated nodes communication and analyzes the traffic information which leads to passive active attack.

To overcome the data leakage issues in centralized storage [1], federated machine learning approach is introduced where the centralized server has to send the copy of the global machine learning model to all the devices. The local devices have to download the copy of model and train them individually with their device data without sharing them to centralized storage. After training, the parameters will be shared to server for consensus change. This approach enhances the privacy of user's data

as they are not stored in centralized storage location. The Federated learning [2] also mitigate the centralized server attack and common server failure issues. Though this approach has various advantages there is possibility of active attacks like sybil attack, in which malicious node carries identity of authenticated node and makes all the other nodes communicating to them as vulnerable to attack.

To address this issue Blockchain based node authentication is introduced for federated machine learning in which the server and node communications are authenticated using smart contracts. The proposed system is developed with Ethereum blockchain which ensures trust and reliability in the network as the transactions in the blockchain are immutable and unalterable [5] which enhances the security of federated machine learning approach.

The remaining section of the paper is organized as follows, Section II describes about Blockchain and section III cover the related work of Federated learning approach with Blockchain. The section IV contains federated learning and comparison of traditional machine learning approach with federated learning approach and section V illustrates Blockchain based node authentication for federated learning and section VI discusses experimentation and results and finally section VII covers the conclusion.

II BLOCKCHAIN

Blockchain is defined as decentralized and distributed database or ledger technology that is shared among the nodes of the network.The datas are stored in the blockchain in the form of blocks.A block in the blockchain has the structure of header and body.The header contains hash of the previous block,timestamp and transaction data.The blocks are linked together forms chain once the transaction is succesfully validated using consensus.It is used to enhance the security,trust and transparency among the network.

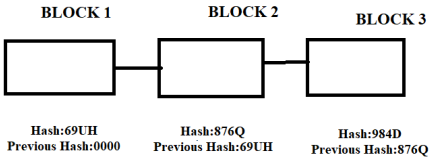

Fig 1. Structure of Blockchain

A. Characteristics

The blockchain is trustable and securable network which has the following characteristics.

Immutable:

The data in the blockchain are unalterable and permanent, hence they are called as immutable.

*Transparent***:**

The transaction in the network are accessible to all hence they are called as transparent.

*Distributed***:**

The information in the blockchain are distributed and each node maintains the copy of the transaction[11].This feature of the blockchain makes them resistant to centralized node failure.Hence they are called as distributed

Decentralized:

The blockchain does not have centralized server to store the information or it does not have centarlised governing authority.The transactions in the network are distributed to each nodes hence any anomaly or updation in network in broadcasted to all the participating nodes.

Verifiable:

Everyone in the network can verify the validity of the information,hence they are verifiable.

Secure:

The information stored in the blockchain are encrpted using cryptography algorithms and stored with hash values.Hence they are secure.

B.*Types of Blockchain:*

The blockchain is classified in to three types according to their nature of deployment[12].

- Public Blockchain
- Private Blockchain
- Consortium Blockchain

1.

Public Blockchain:

The blockchain with no restrictions to access is called public blockchain.Anyone can send transaction and participate as a validator.It is open to all.

Private Blockchain:

The blockchain which is used within the organisations or between the organisation with limited accesss to specified members is called private blockchain.

Consortium Blockchain

The blockchain which is the combination of both permissioned and permissionless is called consortium blockchain

Blockchain uses Consensus which is an algorithm that defines agreement that has to be agreed upon by all the nodes in the system.The popular consensus algorithm are PoW,PoS and PoA.It ensures the authenticity and trust among the peers in the network[13].Smart contracts are used in Ethereum based blockchain to is used to automate and validate transaction once condition is met. Blockchain is promising technology that is deployed in various applications like supply chain management,health care and smart grids[4] to protect network from the security attacks.

The convergence of two technologies federating learning with blockchain enhances the security framework of the federated learning approach.The federated learning approach is considered as boon to technology based industries as it unravels the data privacy issues of the users as it employs decentralized approach to send the model to

the local devices and train the model with global machine learning model.However it is uncertain to trust whether the server is sending the communication to authorized node or malicious node.The proposed Blockchain based node authentication ensures the centralized server is communicating with authenticated nodes using smart contracts.

III RELATED WORKS

The centralized servver in the networks acts as central hub to store the information received from the edge devices.The information shared by the devices are sensitive in nature like user's date of birth,identity information,home address and health conditions.In fewer cases it receives the confidential information of the organisation. The single point of failure causes disruption in the entire network. Nowadays most of the cyber attacks are aiming at centralized server.

The federated machine learning approach is considered to be the promising technology for privacy preserving.The researchers[8] who are analyzing the security threats and vulnerability against federated machine learning approach discusses that this approach is also vulnerable to poisoning attack where the updates can be poisoned by intruders and proposed techniques to improve the poisoning of local updates.The another attack was discussed by the author in his paper[6] in which model placement in server and the way to attack the model in the backdoor is discussed. According the author[7],the data poisoning attack not only poison the data of the device data but also poison the machine learning model.

This paper categorize adversary attackers[14] in the federated learning approach in the following context.

Honest but curious Federated Learning(FL) server:

The FL server is curious to receive the updates from the clients and infer the information about the private client's data set.

Malicious FL server:

This kind of servers have intention to launch the attacks on the private dataset of the clients.

Honest but curious client:

This kind of server are honest but curious to know about the successive global parameters from the server and tries to infer it.

Malicious Clients:

The malicious client tries to infer the updates of the other clients and construct its own global parameter and to get additional information about the other client's private data set.

To mitigate this security risks the integration of the technologies federated learning and blockchain is proposed [14] .This approach is used to protect the data in the mobile devives from leakage.The blockchain and federated learning integration approach to secure the fog computing is proposed[16] which warrants blockchain can be deployed with the federated learning approach to enhance the security of the network.This neccessites the need of security infrastructure to safeguard the federated learning approach from various active attacks which leads to data inference attack.To prevent the data leakage from the malicious server and malicious clients in the federated learning approach blockchain based node authentication is proposed for the federated learning approach which enhance the security of the server and clients with privacy preserving.

IV FEDERATED MACHINE LEARNING:

Federated Machine learning is the form of Machine learning which is used to train the central model on the Decentralized data[16].The copy of the central model is distributed to the nodes which trains the data on the central model and aggregates the result updates back to the Centralized server.It uses Fedavg algorithm for the optimization of the result updates.

A.Advantages of Federated Learning:

The advantages of Federated machine learning approach are

- Hyper personalized
- Privacy preserving
- Low latency

B. Classical Machine Learning Vs Federated Machine Learning

In classical machine learning approach,the data is gathered from all the edge devices and shared to centralized server for training.The machine learning model will be trained on the collected data and then output will be predicted for the test data or new data.This traditional machine learning approach pose threat to the privacy of user's data which is collected and stored in centralized server for training.To overcome this federated machine learning approach was introduced ,where the centralized machine learning model training will be carried out on decentralized data.The users are not required to share their data to the centralized server,instead the server will share copy of the machine learning model to the users devices ,in which training is carried out locally in each device independently and result is shared to the centralized server thereby reducing the burden of centrailzed server and at the same time preserving the privacy of user's data.The centralized server aggregates results from all the devices and updates the model to improve the performance.

The traditional machine learning approach requires all the data to be stored in centralized storage system.This affects the data privacy of the users by implementing data inference attacks.In federated machine learning approach the model is trained on decentralized data,eliminating the data inference attacks.The single point of failure is the major concern in the centralized machine learning approach,which affects the security and functions of the other nodes connected to them, if the server is attacked.In

federated machine learning approach the data is not shared to centralized server and connected nodes only share the parameters after training their local data with global model.

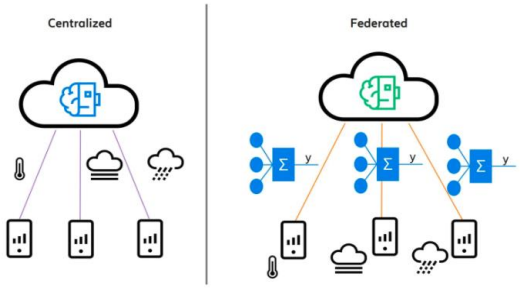

Fig 2. Traditional Machine Learning Vs Federated Machine Learning Training

V BLOCKCHAIN BASED NODE AUTHENTICATION FOR FEDERATED LEARNING

The Blockchain based node authentication is constructed using Ethereum blockchain.The Federated learning approach has centralized server which has global machine learning model.The Global machine learning model can be supervised or unsupervised based on the problem definition of the developers.

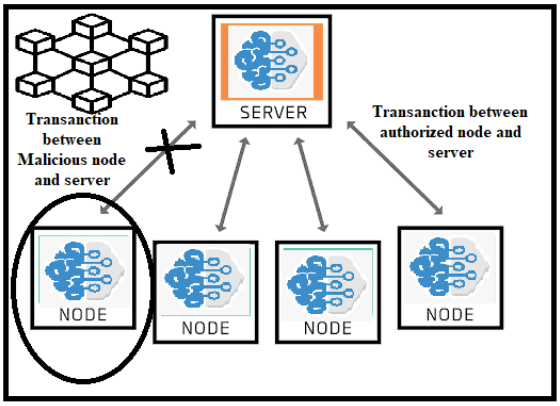

Fig 3. Node authentication for Federated learning

The Fig 1.3 ilustrates the node authentication for federated learning approach in which the server distribute the global machine learning model to the nodes.The nodes download the copy of the model and train with their device data.Before sending the copy of the machine learning model to the nodes ,the node authentication smart contract will be deployed which ensures the node is interacting with authenticated nodes.

Once the server starts to initiate communication with node block will be created.The smart contract validates whether Interaction is enabled between server and authenticated node.If it is valid then block will be added to the chain,otherwise the interaction will be aborted.

Smart contract for Node Authorization: The Pseudocode for node authorization checks whether the interaction is initiated between authenicated server and the authenicated node.

```
// Define a contract ServerRole to manage this role - add, remove, checkcontract ServerRole {

 using Roles for Roles.Role;

 // Define 2 events, one for Adding, and other for Removing

 event NodeAdded(address indexed account);

 event nodeRemoved(address indexed account);

 // In the constructor to add node

 constructor() public {

  _addNode(msg.sender);

 }

 // Define a modifier

 modifier onlyNode(address OnlyNode ID) {

  require(isValidNode(OnlyNode ID));;}

 // Define a function 'isServer' to check this role

 function isServer(address account) public view returns (bool) {

  return nodes.has(account);

 }

 // Define a function 'addServer' that adds this role

 function addServer(address account) public {

  _addServer(account);

 }

 // Define a function
```

```
function renounceServer() public {

  _removeServer(msg.sender);

}

// Define an internal function '_addServer' to add this role, called by 'addServer'

function _addServer(address account) internal {

  server.add(account);

  emit serverAdded(account);

}  // Define an internal function '_removeServer' to remove this role, called by 'removeServer'

function _removeServer(address account) internal {

  server.remove(account);

  emit serverRemoved(account);

}}
```

Once node authentication is completed ,the nodes sends only the updates of training model to the server ,instead of sharing whole data.The weighted average of the updates collected from the nodes and model will be improved by the server using the FedAvg algorithm.In the next epoch,the copy of improved model will be send to all the nodes.

VI EXPERIMENTAL SETUP AND RESULTS

The Blockchain based node authentication is constructed using Ethereum framework with truffle and Ganache is used to provide ethers for testing. The consortium blockchain is used,which is a combination of permissionless and permissioned.The roles of the centralized server and nodes are predefined by the administrators, hence they are permissioned.On the other hand any nodes can interact with server hence they are permissionless.The transaction once they are validated by the smart contracts written in solidity language.

The Ethereum dataset[3] which contains the address of normal and anomaly transaction is used for the evaluation of node authentication where their roles are defined in smart contracts.The gas consumption is represented in ether (ETH) .

The data set contains the features of Hash,Nonce,Transaction index,Block number,Time stamp,from address and toaddress.The first 500 hash is taken with 80% of authenticated address and 20 % of anomaly address .

The distribution of the normal and anomaly address in the dataset are illustrated in Fig 4.

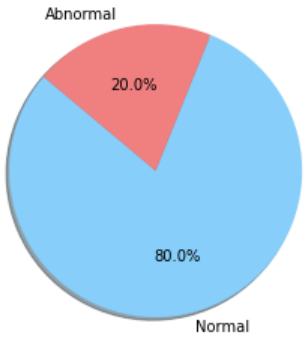

Fig 4. Normal Vs Abnormal data distribution

The node authentication performance is evaluated using K-Nearest Neighbours(KNN) and Random Forest algorithm which are known to be the best classification algorithm for classification problem.The KNN algorithm classifies the data based on how its neighbours classifies.The Random forest algorithm uses ensemble approach for better classification.

The classification accuracy of KNN and Random forest are 95.7 and 99.6 respectively.In this Random forest outperforms KNN algorithm.

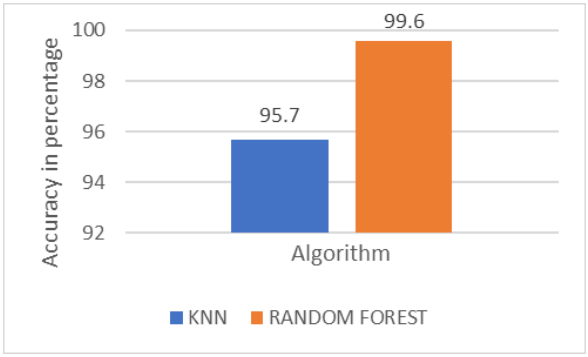

Fig 5. Performance comparison KNN and Random forest

The gas consumption for the node authentication for each transaction is depicted in Fig 6.

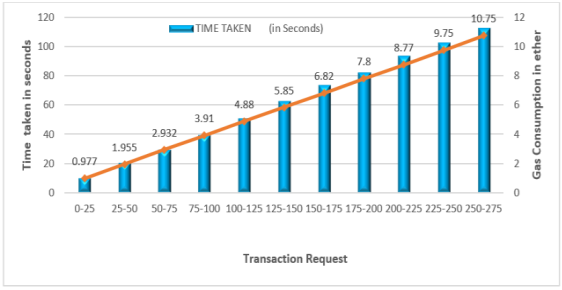

Fig 6. Gas Consumption for Node Authentication

The experiment is done using 275 sample data with authorized and anomaly address.The gas consumption is measured for transaction request .The authorization of authenticated and anomaly node is evaluated with KNN and Random forest algorithm which provides 95.7 and 99.6 respectively .

VII CONCLUSION

The technology improvement widens the security threat among the growing network.The architecture of IoT and banking sectors mainly depends on their centralized server for data storage and decision making.However single point of failure disrupts the entire economy and data loss.The one promising solution is to adapt decentralized approach like federated learning which preserves privacy of the data prventing data loss.In our proposed model the major concern of server to node authentication is addressed by implementing blockchain based node authentication with smart contracts.The performance evaluation also produces satisfactory results of 95.7% and 99.6 % with KNN and Random forest algorithms.Though accuracy rate is higher the time taken to validate and gas consumption is higher for node authentication which will be concentrated in future work to reduce the same.Thus the security enhanced federated learning approach with blockchin based node authentication can be used for the network to enhance the privacy preserving and as server shield.

REFERENCES

[1] Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, Sweta Mishra, “Detection and mitigation of fraudulent resource consumption attacks in cloud using deep learning approach”, Journal of Information Security and Applications, Volume 56,2021,102672,ISSN 2214-2126.

[2] Alberto Blanco-Justicia, Josep Domingo-Ferrer, Sergio Martínez, David Sánchez, Adrian Flanagan, Kuan Eeik Tan,”Achieving security and privacy in federated learning systems: Survey, research challenges and future directions”,Engineering Applications of Artificial Intelligence,Volume 106,2021,104468,ISSN 0952-1976.

[3] Al-E’mari, S., Anbar, M., Sanjalawe, Y., Manickam, S. (2021). A Labeled Transactions-Based Dataset on the Ethereum Network. In: Anbar, M., Abdullah, N., Manickam, S. (eds) Advances in Cyber Security. ACeS 2020. Communications in Computer and Information Science, vol 1347. Springer, Singapore. https://doi.org/10.1007/978-981-33-6835-4_5

[4] Blockchaincouncil,walnut,https://www.blockchain-council.org/blockchain/top-10-promising-blockchain-use-cases/ available from 2019.

[5] Dharmin Dave, Shalin Parikh, Reema Patel, Nishant Doshi,A Survey on Blockchain Technology and its Proposed solutions,ProcediaComputer Science,Volume 160,2019,Pages 740-745,ISSN 1877-0509.

[6] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, “How To Backdoor Federated Learning,” in Proc. of AISTATS, 2020, pp. 1-10.

[7] G. Sun, Y. Cong, J. Dong, Q. Wang, L. Lyu and J. Liu, "Data Poisoning Attacks on Federated Machine Learning," in *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 11365-11375, 1 July1, 2022, doi: 10.1109/JIOT.2021.3128646.

[8] N. Bhagoji, S. Chakraborty, E. Mittal, and S. Calo, “Analyzing Federated Learning through an Adversarial Lens,” [Online]. Available: https://arxiv.org/abs/1811.12470., 2018.

[9] N. C. Will, "A Privacy-Preserving Data Aggregation Scheme for Fog/Cloud-Enhanced IoT Applications Using a Trusted Execution Environment," *2022 IEEE International Systems Conference (SysCon)*, 2022, pp. 1-5, doi: 10.1109/SysCon53536.2022.9773838.

[10] Pim Otte, Martijn de Vos, Johan Pouwelse,TrustChain: A Sybil resistant scalable blockchain, Future Generation computer Systems,2017,ISSN 0167-739X.

[11] S. Revathy and S. S. Priya, "Blockchain based Producer-Consumer Model for Farmers," *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2020, pp. 1-5, doi: 10.1109/ICCCSP49186.2020.9315214.

[12] The truth about block chain”,Harvard business review,https://hbr.org/2017/01/the-truth-about-blockchain.

[13] T. M. Fernández-Caramés and P. Fraga-Lamas, "A Review on the Use of Blockchain for the Internet of Things," in IEEE Access, vol. 6, pp.32979-33001,2018.

[14] Weilong Wang, Yingjie Wang, Yan Huang, Chunxiao Mu, Zice Sun, Xiangrong Tong, Zhipeng Cai,Privacy protection federated learning system based on blockchain and edge computing in mobile crowdsourcing,Computer Networks,Volume 215,2022,109206,ISSN 1389-1286.

[15] X. Gong, Y. Chen, Q. Wang, M. Wang and S. Li, "Private Data Inference Attacks against Cloud: Model, Technologies, and Research Directions," in IEEE Communications Magazine, doi: 10.1109/MCOM.004.2100867,2022

[16] Y. Qu *et al*., "Decentralized Privacy Using Blockchain-Enabled Federated Learning in Fog Computing," in *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5171-5183, June 2020, doi: 10.1109/JIOT.2020.2977383