IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 1, May 2023

Securing Federated Learning Based on Blockchain Technology

Aswin K[1], Kamali R[2], Kurin Firathos S[3], Jeevitha G[4], Nandhini E[5]

Faculty, Department of Computer Science & Engineering[1]

Students, Department of Computer Science & Engineering[2,3,4,5]

Dhanalakshmi Srinivasan Engineering College, Perambalur, India

Abstract*:* *A blockchain-based federated learning approach with secure aggregation in a trusted execution environment for the Internet of Things (IoT). The proposed approach aims to address the privacy and security concerns associated with federated learning in IoT environments. The approach involves using a blockchain to store the learning model and to maintain a distributed ledger of transactions. The learning model is trained on local IoT devices using federated learning techniques, with each device contributing its local data. The aggregation of the model updates is performed securely within a trusted execution environment, using homomorphism encryption and secret sharing techniques. The proposed approach offers several advantages over traditional federated learning approaches, including improved privacy and security, increased scalability, and enhanced trustworthiness. It also enables the creation of a decentralized and democratic learning environment, where each device has an equal say in the learning process.*

The approach is evaluated using a real-world dataset, and the results demonstrate its effectiveness in terms of accuracy and privacy preservation. The paper concludes that the proposed approach has the potential to enable secure and scalable federated learning in IoT environments, with applications in healthcare, smart cities, and other domains.

This paper offers a blockchain-based federated learning (FL) framework with an Intel Software Guard Extension (SGX)-based trusted execution environment (TEE) for safely aggregating local models in the Industrial Internet of Things (IoT) Local models in FL can be modified with by attackers. As a result, a global model derived from manipulated local models may be incorrect. As a result, the proposed system makes use of a blockchain network to secure model aggregation. Each blockchain node contains an SGX-enabled CPU that secures the FL-based aggregating processes required to construct a global model. Blockchain nodes may validate the aggregated model's validity, perform a blockchain consensus method to secure the model's integrity, and add it to the distributed ledger for tamper-proof storage. . Before utilising the aggregated model, each cluster can acquire it from the blockchain and validate its fidelity. To assess the performance of the proposed system, we ran many experiments using various CNN models and datasets.

Keywords: Blockchain

I. INTRODUCTION

Blockchain-based federated learning with secure aggregation in Trusted Execution Environment (TEE) for Internet-of-Things (IoT) is a new approach to machine learning that aims to address some of the challenges associated with traditional machine learning techniques in IoT applications. IoT devices generate vast amounts of data, and traditional machine learning techniques typically involve centralizing this data for analysis, which can be problematic due to concerns about data privacy and security.

Federated learning is an alternative approach to machine learning that allows multiple IoT devices to collaborate on a machine learning task without sharing their data with a central server. Instead, each device performs local training on its own data, and the resulting model updates are sent to a central server for aggregation. This approach can help to preserve data privacy and security while still enabling effective machine learning. The traditional federated learning techniques are still susceptible to attacks on the aggregation server, as well as the potential for malicious devices to manipulate the model updates.

DOI: 10.48175/IJARSCT-9705

Blockchain-based federated learning with secure aggregation in TEE aims to address these issues by leveraging the security features of TEEs and the transparency and immutability of blockchain technology. TEEs provide a secure environment for computation and data storage, while blockchain technology provides a decentralized and transparent way of managing the machine learning process.

The expansion of the Internet of Things (IoT) has made it an essential component of a wide range of intelligent applications. Health care, industry, critical system infrastructure, agriculture, and transportation are examples of intelligent applications. Thanks to machine learning algorithms, IoT devices may collect a significant amount of data and behave independently in an intelligent system.

The vast amount of IoT data is critical in training a machine learning algorithm system. In general, IoT devices have limited resources and cannot run machine learning algorithms on their own. Edge computing technology is rapidly gaining recognition for forming intelligent networks in combination with IoT and machine learning.

A cluster is made up of an edge device (referred to as an edge server throughout the text) and IoT devices on the network. Edge devices in an intelligent system can house a machine learning algorithm that uses a locally created dataset to produce a trained model. Depending on the kind of IoT device, IoT devices create data and receive control instructions. Subsequently, the trained model may be employed in the system to make an informed choice.

II. LITERATURE SURVEY

This paper proposes a robust block chained multi-layer decentralized federated learning (RBML-DFL) framework to ensure the federated learning's robustness. Firstly, by adopting the three-layered framework, the blockchain connects the federated learning components to secure the privacy and data safety of federated learning

A blockchain-based data acquisition scheme during the pandemic in which federated learning (FL) is employed to assemble privacy-sensitive data as a form of the trained model instead of raw data.

With the rapid development of the Internet of Things (IoT), more and more data are generated by smart devices to support various edge services. Since these data may contain sensitive information, security and privacy of data aggregation has become a key challenge in IoT. To tackle this problem, a blockchain-based secure data aggregation strategy, namely (BSDA), is proposed for edge computing empowered IoT.

Blockchain and SDN are two top innovations utilized to create secure network architectures and provide trustworthy data transmission. They offer a strong and trustworthy platform to deal with dangers and problems, including security, privacy, adaptability, scalability, and secrecy.

In recent years, traditional logistics systems are developing toward intelligence based on the Internet of Things (IoT). Sensing devices throughout the logistics network provide strong support for smart logistics. However, due to the insufficient local computing and storage resources of IoT devices

Data collaboration with cloud technologies is becoming more popular for personal use as well as business applications. Due to the increasing data protection regulations worldwide, different cryptographic techniques have been designed to enable secure data sharing for a useror a group of users.

Blockchain-based federated learning (BFL) has attracted intensive research attention due to that the training process is auditable and the architecture is server less avoiding the single point failure of the parameter server in vanilla federated learning (VFL).

Blockchain and Deep Learning (DL) are two of the most revolutionary concepts in the field of Computer Science.

Both have made astounding leaps in research and application areas such as Finance, Healthcare, Internet of Things, and many more.

Federated learning is a popular privacy-enhanced distributed machine learning method that solves the problem of local data privacy by gathering the training results insteadof the raw data to generate a global model

Federated learning facilitates the collaborative training of a global model among distributed clients without sharing their training data. Secure aggregation, a new security primitive for federated learning, aims to preserve the confidentiality of both local models and training data

DOI: 10.48175/IJARSCT-9705

III. EXISTING SYSTEM

Blockchain-based federated learning with secureaggregation in a trusted execution environment (TEE) for the Internet-of-Things (IoT) is a relatively new approach that can enhance the privacy and security of existing IoT systems. This approach combines the benefits of blockchain, federated learning, secure aggregation, and TEEs to create a robust and secure system for IoT devices. Blockchain-based federated learning with secure aggregation in a TEE can mitigate these issues by allowing IoT devices to collaborate in a decentralized manner, without compromising the privacy of the data they generate.

In this approach, a blockchain network is used to facilitate communication between IoT devices, while federated learning isused to train machine learning models on data generated by thesedevices. Secure aggregation ensures that the data remains private and is not exposed to unauthorized parties. TEEs are used to provide a secure execution environment for the machine learningmodels, ensuring that they cannot be tampered with orcompromised by malicious actors.

The benefits of this approach include increased privacy andsecurity for IoT devices and their data, reduced risk of data breaches and cyber-attacks, and improved machine learning model accuracy and performance. However, implementing this approach requires significant technical expertise and resources, and may not be feasible for all IoT systems. The blockchain-based federated learning with secure aggregation in a TEE is a promising approach for enhancing the security and privacy of existing IoT systems. It combines several advanced technologiesto create a robust and secure system that can be used for a wide range of IoT applications.

A blockchain network particularly created for IoT that seeks to secure data sharing and payments between devices. The decentralised data exchange platform that leverages blockchain to safeguard data sharing while retaining data owners' control. An open-source platform that enables safe and efficient data exchange across IoT devices as well as decentralised data processing at the edge. The collective endeavour to provide safe and trustworthy data processing in multi-party contexts.

IV. PROPOSED SYSTEM

The proposed system of Blockchain-Based Federated Learning with Secure Aggregation in Trusted Execution Environment (TEE) for Internet-of-Things (IoT) is a novel approach to addressing the challenges of data privacy and security in IoT environments. The system utilizes a blockchain- based federated learning approach, where multiple IoT devices collaborate to train a machine learning model without sharing their raw data with each other or with a central server. Instead, each device trains a local model on its own data and only shares encrypted model updates with the other devices.

To ensure the security of the model updates, the system employsa secure aggregation protocol that utilizes a TEE. A TEE is a secure and isolated environment within a device that can protectsensitive data and code from external threats.

The use of blockchain technology in the system provides additional security and transparency. The blockchain serves as a distributed ledger that records all transactions in a tamper-proof manner, ensuring that the model updates are not modified or tampered with.

The proposed system of Blockchain-Based Federated Learning with Secure Aggregation in Trusted Execution Environment for Internet-of-Things offers a promising solution to the challenges of data privacy and security in IoT environments. By utilizing a federated learning approach, secure aggregation protocols, and a TEE, the system can enable IoT devices to collaboratively train machine learning models while preserving the privacy and security of their data.

Local Model Generation

At each cluster, the local model generation (LMG) step is carried out to build a locally trained model, similar to the initialization phase of the original FL. Figure 4 depicts an overview of the LMG phase. In the proposed system, we assume that the edge servers of various clusters train models for image classification using convolutional neural network (CNN)-based image classification, with model parameters received from the global model stored in tamper-proof storage.

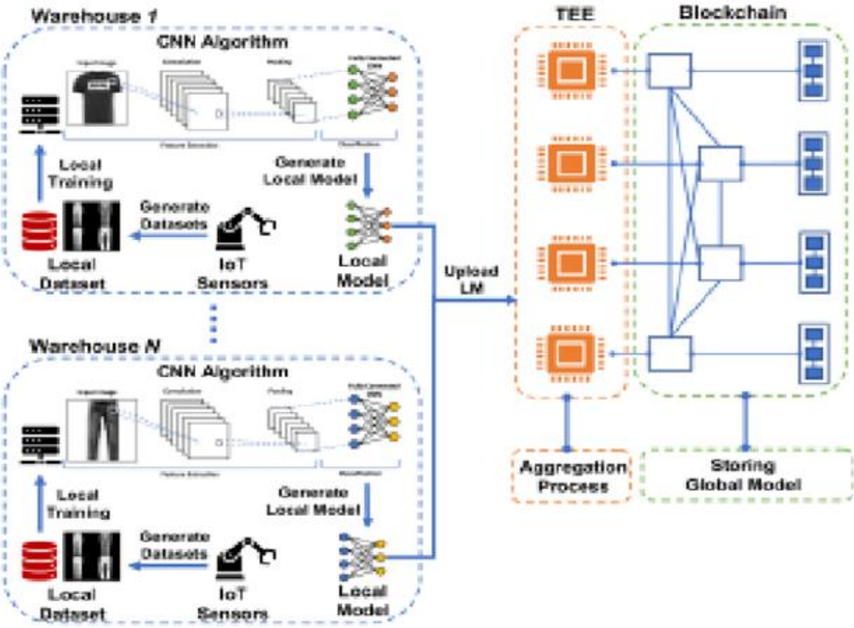
Local model generation is a key component of the proposed approach for blockchain-based federated learning with secure aggregation in a trusted execution environment for the Internet ofThings (IoT). In this approach, local models are generated on each IoT device using federated learning techniques, with each device contributing its local data to the training process. The localmodel generation process involves several steps. First, the IoT devices download the current model from the blockchain and useit as a starting point for training. Next, each device trains the model using its local data, while preserving the privacy of the data using techniques such as differential privacy or federated learning. They are sent to the trusted execution environment (TEE) for secure aggregation. The TEE uses homomorphic encryption and secret sharing techniques to combine the local model updates in a secure and privacy-preserving manner, resulting in a global model that reflects the contributions of all the participating IoT devices.

This approach offers several advantages over traditional approaches to machine learning, including improved privacy, reduced communication costs, and increased scalability. The generating models locally on each device, the approach minimizes the amount of data that needs to be transferred over the network, reducing the risk of data breaches and other security threats. Local model generation is a critical component of the proposed approach, enabling secure and efficient federated learning in IoT environments.

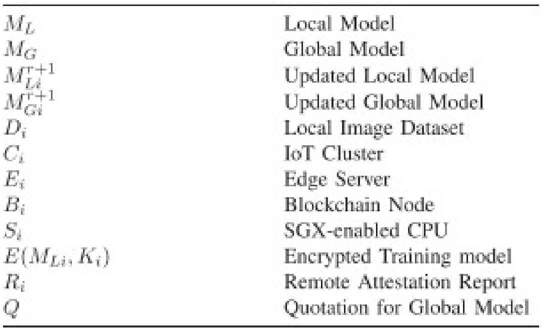
V. SYSTEM ARCHITECTURE

We describe the suggested blockchain-based FL with asafe aggregation mechanism driven by TEE. To begin, we will provide an overview of the system architecture. Following that, we go over the various components of our proposed framework in depth**.**

In this system, we investigate a FL-based collaborative learning model that uses TEE for safe aggregation and blockchain for tamper-proof data exchange and storage. We presume that there are numerous warehouses outfitted with multiple IoT cameras capable of scanning product images and recognizing the type of goods. Because IoT cameras have limited resources for executing machine learning algorithms, each warehouse hosts and executes a machine learning algorithm on an edge server. First, IoT cameras and the edge server create a cluster; the edge server trains a model based on the local dataset and generates a trained model known as a local model, abbreviated as ML.



Nevertheless, if the size of the local dataset is tiny, ML accuracymay be low. As a result, the edge server of a cluster Ci participates in FL, which involves numerous clusters ofcomparable warehouses, by submitting its ML to produce an aggregated model known as a global model, designated as MG. For the global model in our suggested scenario, we use the federated averaging (FedAVG) technique, which will be presented. A typical FL method consists of three steps: startup, aggregation, and updating. Unlike the typical FL technique, which aggregates ML in a centralised server (e.g., a cloud server), our suggested framework aggregates ML on a blockchain platform. A blockchain network is made up of several nodes, and each node gets all ML and separately aggregates it to create their own copy of an MG. Each blockchain node is assumed to have a TEE host. To maintain security throughout the aggregation process, each blockchain node aggregates in its TEE host and generates an MG. To guarantee that all nodes have the same MG, blockchain nodes use a consensus method. After consensus is obtained, each blockchain node saves the MG in its own blockchain. Ultimately, the blockchain network distributes MG to all of the edge servers. When MG is received, edge servers validate it and update their basic model with it. The MG is used by edge servers in the warehouse for product recognition. The suggested system, which is divided into three stages: local model development, secure TEE-enabled aggregation, and blockchain- based global model storage.



M_L	Local Model
M_G	Global Model
$M_{Li}^{\tau+1}$	Updated Local Model
$M_{Gi}^{\tau+1}$	Updated Global Model
D_i	Local Image Dataset
C_i	IoT Cluster
E_i	Edge Server
B_i	Blockchain Node
S_i	SGX-enabled CPU
$E(M_{Li}, K_i)$	Encrypted Training model
R_i	Remote Attestation Report
Q	Quotation for Global Model

VI. MODULE LIST

- Admin and login
- Blockchain-Based Federated Learning
- TEE-Based Secure Aggregation
- Model Evaluation and Deployment

6.1 Admin and Login

The admin module would include functionalities for managing user accounts, such as adding new users, modifying user information, and deleting user accounts. It may also involve features like password reset, account recovery, and user role management.

The login module would handle user authentication and authorization. It would require users to provide valid credentials, such as username and password, to access the system. Once authenticated, users would be granted appropriate authorization levels based on their roles and permissions to perform various actions within the systemhe raw data collected from IoT devicesis processed to extract relevant features. The extracted features are then encrypted using homomorphic encryption and sent to theblockchain.

6.2 Blockchain-Based Federated Learning

This module consists of a blockchain network that manages the federated learning process. The IoT devices communicate with each other and the blockchain to train a global model. The blockchain network securely aggregates the encrypted gradients from the devices and updates the global model.

6.3 TEE-Based Secure Aggregation

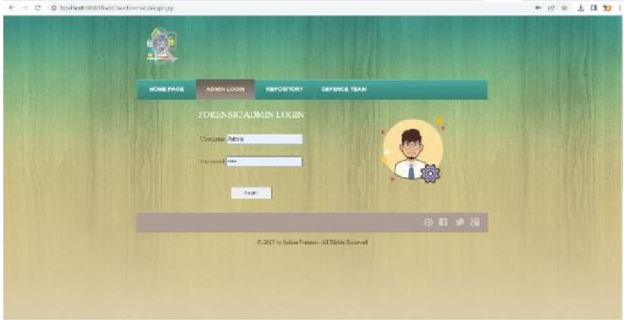
This module uses a TEE to protect the privacy of device owners. The TEE runs on the device and performs secure aggregation of the encrypted gradients. The encrypted gradients are decrypted and aggregated within the TEE, ensuring that the device owner's data remains private.

6.4 Model Evaluation and Deployment

This module evaluates the performance of the trained model and deploys it to the IoT devices for inference. The deployed model is optimized for the IoT devices, considering their limited resources.

DOI: 10.48175/IJARSCT-9705

VII. RESULT



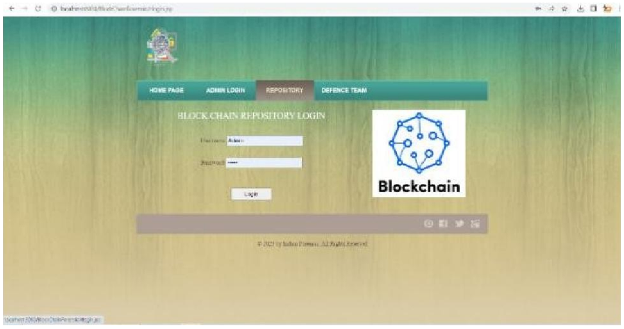
FORENSIC ADMIN LOGIN

Username Admin

Password ****

Login

© 2023 by Indian Forensic. All Rights Reserved



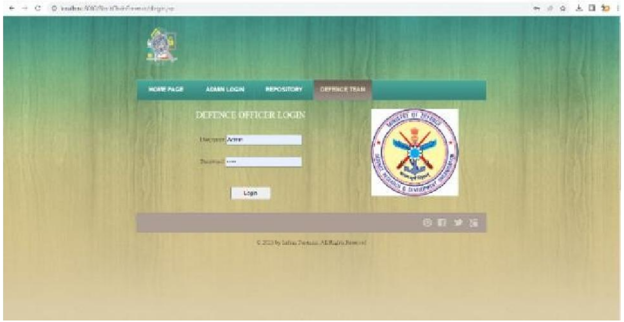
BLOCK CHAIN REPOSITORY LOGIN

Username Admin

Password *****

Login

© 2023 by Indian Forensic. All Rights Reserved



DEFENCE OFFICER LOGIN

Username Admin

Password ****

Login

© 2023 by Indian Forensic. All Rights Reserved

VIII. CONCLUSION

A blockchain and TEE-enabled FL framework for IoTwas presented in this paper. The primary goal of this framework was to assure the reliable aggregation of local models in order togenerate a global model. The aggregation took place on theblockchain network. To provide safe aggregation, the suggested system used the Intel SGX-based TEE, with each blockchain node doing the aggregating process. Each blockchain node in thissystem is equipped with an SGX-enabled processor that securelybuilds a global model to assure trustworthiness. Then, before being put to the blockchain, the global model was confirmed by the blockchain network via a consensus method, ensuring tamperproof storage. Only the blockchain network allowed usersto view and verify the global model's integrity. We generated

REFERENCES

[1]. M. Alazab, S. P. RM, M. Parimala, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, "Federated learning for cybersecurity: Concepts, challenges and future directions," IEEE Trans. Ind. Inform., vol. 18, no. 5, pp. 3501–3509, May 2022.

[2]. L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature feakage in collaborative Slearning," in Proc. IEEE Symp. Secur. Privacy, 2019, pp. 691–706.

[3]. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, May 2020.

DOI: 10.48175/IJARSCT-9705

[4]. L. Yin, J. Feng, H. Xun, Z. Sun, and X. Cheng, “A privacy-preserving federated learning for multiparty data sharing in social IoTs,” IEEE Trans. Netw. Sci. Eng., vol. 8, no. 3, pp. 2706–2718, Jul.–Sep. 2021.

[5]. Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim, and C. Miao, “Federated learning in the sky: Aerial-ground air quality sensing framework with UAV swarms,” IEEE Internet Things J., vol. 8, no. 12, pp. 9827–9837, Jun. 2021.

[6]. Lin Tao, Kong Lingjing, Sebastian U Stich and Martin Jaggi, "Ensemble Distillation for Robust Model Fusion in Federated Learning", Advances in Neural Information Processing Systems, 2020.

[7]. Swaroop Ramaswamy, Rajiv Mathews, Kanishka Rao and Françoise Beaufays, "Federated learning for emoji prediction in a mobile keyboard", arXiv preprint, 2019.

[8]. Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos and Yasaman Khazaeni, "Federated Learning with Matched Averaging", Inter- national Conference on Learning Representations, 2020.

[9]. M. Alazab, S. P. RM, M. Parimala, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, “Federated learning for cybersecurity: Concepts, challenges and future directions,” IEEE Trans. Ind. Inform., vol. 18, no. 5, pp. 3501–3509, May 2022

[10]. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, May 2020.