NextGen Scientific Publication
American International Journal of Computer Science and Technology
Volume 1 Issue 1, Pg. No. 1-11, AIJCST-V1I1P101, 2019
https://doi.org/10.63282/3117-5481/AIJCST-V1I1P101

Original Article

A Blockchain-Integrated Computational Framework for Secure Multi-Party Machine Learning in Cloud-Edge Collaboration

***Dr. Rajan Krishna**
Department of Computer Applications, Kongundu Engineering College, Tamil Nadu, India.

Abstract:

The swift cloud and edge computing development promoted the implementation of distributed learning models as different organizations can train them cooperatively and share sensitive data without the need to access it. There is however a serious challenge of making sure that the data privacy, model integrity and secure collaboration is ensured. The paper introduces a blockchain-enhanced framework to conduct computations based on secure multi-party machine learning (MPML) over cloud-edge collaborative settings. The framework proposed uses blockchain technology to build a sense of trust, immutability and verifiability during data sharing as well as training the model. The framework also includes the principles of secure multi-party computation (SMPC) and federation of learning, designed to maintain the privacy of data, but to make models optimize the performance on a heterogeneous set of nodes. We also give a step by step methodology of the system architecture, consensus protocols, encryption mechanisms, and collaborative learning algorithms. Experimental testing illustrates the effectiveness of the framework in regard to security, scale, and the accuracy of the model. Indeed, our findings reveal that blockchain coupled with MPML will help to solve security threats, accountability, and trust among cooperating entities substantially. The framework offers the solid solution to the real-world cloud-edge collaborative applications, such as healthcare, finance, and smart cities

Keywords:

Blockchain, Multi-Party Machine Learning (MPML), Cloud-Edge Collaboration, Federated Learning, Secure Multi-Party Computation (SMPC), Data Privacy, Model Integrity.

Article History:

Received: 23.11.2018

Revised: 07.12.2018

Accepted: 18.12.2018

Published: 06.01.2019

1. Introduction

1.1. Background

The combination of cloud computing with edge computing has changed the scenery of distributed data processing and machine learning (ML), as it can now be more efficient, scalable, and responsive with the design of the intelligent systems. Cloud computing has essentially endless capacity to compute, store data, as well as be accessed all over the globe, which is suitable to train large-scale models and coordinate them centrally. On the flipside, edge computing brings data to these devices like sensors, IoT devices, and mobile endpoints in close proximity to reduce the latency, bandwidth requirements, and reliance on continuous connectivity to the cloud are diminished. This combination has enabled the emergence of the cloud-edge collaborative learning, which harnesses the power of the two areas to enable real-timely, data-driven decision making across distributed infrastructures. Multi-Party Machine Learning (MPML) is now believed to be an attractive model within this developing ecosystem, which allows several independent actors,

Copyright @ 2019 by the Author(s). This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-NC-ND 4.0) License (https://creativecommons.org/licenses/by-sa/4.0/)

including organizations, institutions, or devices, to jointly train machine learning models without exchanging raw data. It is done by using privacy-preserving methods, such as federated learning (FL) and Secure Multi-Party Computation (SMPC), enabling local model training and aggregation of updates of updates safely. The problem is that MPML not only ensures better privacy of data, but also leads to collaborative intelligence sharing on the distributed networks so that the others can enjoy the benefits of shared knowledge without raising up to lost control over the proprietary datasets. Nevertheless, in spite of such benefits, security, trust, and accountability in MPML processes are a significant challenge. Distributed features of cloud-edge collaboration pose a threat to the system in terms of tampering with data, model poisoning, and unauthorized involvement. In addition, lack of transparent and verifiable system to monitor contributions and ensure the integrity of the model can form a dent within the level of trust of the participating entities. Those difficulties signify the necessity of the safe, transparent, and decentralised system that could ensure the data integrity, confirm the actions of participants, and provide the fair cooperation. The incorporation of blockchain technology into MPML provides a favorable route to deal with these issues, which will provide the platform to a reliable, privacy-sensitive, and verifiable collaborative learning in cloud-edge settings.

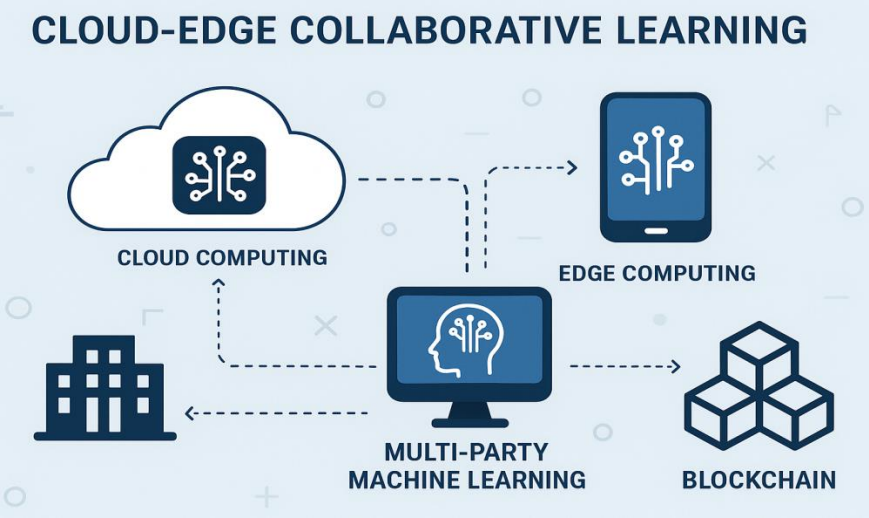

Figure 1. Background

1.2. Importance of Blockchain-Integrated Computational Framework

Introduction of blockchain technology into computational and machine learning platforms is an immature step towards the establishment of secure, transparent, and decentralized collaborative systems. Blockchain offers a layer of trust in cloud-edge models where several parties are involved in the training and data processing of a model; the layer should be integrity-assuring, accountable, and tamper resistant. The inclusion of a blockchain database in a computational system not only makes the system more secure but also creates more fairness and transparency in distributed machine learning systems like Multi-Party Machine Learning (MPML) and Federated Learning (FL). The subsequent sub-sections identify the critical areas in which blockchain integration would bring high value.

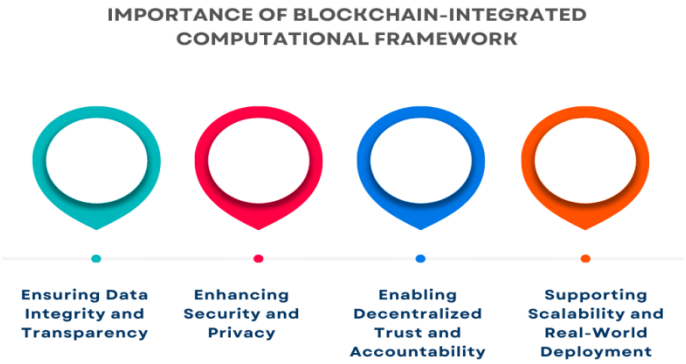

Figure 2. Importance of Blockchain-Integrated Computational Framework

1.2.1. Ensuring Data Integrity and Transparency:

The ledger of blockchain records all the transactions, all updates of the models, and all the contributions made by the participants, which results in a transparent and auditory track of actions. The collaborative learning process cryptographically connects each update made in the process to the records made before it, and thus it is impossible to alter or delete any document. This will help to ensure the integrity of data and enable the participants to check that their contributions are properly reflected in the global model. Blockchain provides the transparency, which enhances trust among the nodes that do not have complete trust in each other, which is usually the case in cross-organizational partnerships.

1.2.2. Enhancing Security and Privacy:

According to the traditional cloud-edge architecture, central aggregation nodes may fall prey to insecurity attacks to include data leakage and model poisoning, or parameter manipulation. The framework can remove the use of one trusted authority by using blockchain with Secure Multi-Party Computation (SMPC) and homomorphic encryption. Numerous consensus mechanisms are used to validate updates to the model prior to them being recorded on the ledger so that only legal and confirmed information is added to the global model. This security design works on a two-tiered protection of insider and external threats as well as user privacy.

1.2.3. Enabling Decentralized Trust and Accountability:

The key functions that are automated by smart contracts in the blockchain include participant authentication, validation of model update, and reward distribution. Such self-executing agreements provide equitable involvement and eliminate unscrupulous conduct without the need to have centralized supervision. Moreover, blockchain offers an open source framework of trust, in which a consensus mechanism (e.g., Practical Byzantine Fault Tolerance (PBFT)) can provide agreement among nodes despite the faultiness or maliciousness of certain nodes. This makes various participants responsible and cooperate in a secure and transparent and verifiable way.

1.2.4. Supporting Scalability and Real-World Deployment:

Computational frameworks that are built on blockchain are of particular importance to scalable cloud-edge systems, with many devices and organizations interacting in real time. The framework allows applications in crucial areas, including healthcare, finance, smart manufacturing, and IoT-based systems, by decentralizing data control and making them verifiably collaborative. Such industries require various standards of data privacy, regulation adherence, and trust, and the blockchain implementation is a strong solution to all of them.

1.3. Secure Multi-Party Machine Learning in Cloud-Edge Collaboration

Cloud-edge secure multi-party machine learning (MPML) is a revolutionary paradigm of distributed artificial intelligence balancing privacy, efficiency, and security of data. Traditional centralized machine learning requires all data to be hosted and processed in a centralized location data usually in the cloud, which casts great doubts on privacy, ownership of data, and adherence to regulatory measures such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Privacy Act (HIPAA). MPML aims to overcome these issues and enable several organizations or end devices to build a shared model without revealing their local data. Every individual does local training on its own data and submits encrypted model updates which are safely combined to create a global model. The latter is especially suitable in cloud-edge architectures, wherein data are constantly generated at distributed edge nodes, and sending the unprocessed information to the cloud is both inefficient and insecure.

Combination of MPML and cloud-edge collaboration boost real time intelligence and scalability. On-site data processing and initial model training are done by the edge nodes to minimize network congestion and latency whereas complex aggregation and optimisation are done on the cloud servers. The Secure Multi-Party Computation (SMPC) methods of additive secret sharing and homomorphic encryption allow combining model updates in a secure manner- so no individual participant/aggregator of has access to sensitive data. Moreover, the risk of data leakages, model inversion, and poisoning attacks prevalent in the distributed learning settings are addressed with this structure. To increase the levels of trust and accountability, MPML could be supplemented by the blockchain technology, which tracks all transactions, model updates, and contributions by the participants in a permanent register. This provides openness and verifiable network between parties that might be mistrusted. On the whole, privacy-preserving, scalable, and trustful distributed learning is made possible because of secure MPML in cloud-edge collaboration, with privacy as the key feature in sensitive areas of healthcare, finance, smart cities, and factory IoT, where the contexts of data confidentiality and reliability of the system are essential.

2. Literature Survey

2.1. Cloud-Edge Collaborative Learning

Cloud-edge collaborative learning is a new paradigm and uses both cloud and edge computing to effectively improve responsiveness and efficiency of machine learning (ML) systems. At edge nodes, which include: IoT devices, gateways, or local servers in this architecture, pre-processing of data, and feature extraction and partial model training occur near the source of data, thus, minimizing latency and a reduction in the utilisation of bandwidth. The cloud that has strong computational force will summarize the updates in the model, or conduct a global optimization to provide a convergence and accuracy. This separation of labor is not only doing real-time analytics and adaptive learning possible, but also contributes to data privacy preservation through reducing data transmission. In spite of these advantages, there are numerous challenges such as communicating with heterogeneous edge devices that have different computational performances, communication overhead of distributed node etc. and security threats that come with decentralized data processing.

2.2. Multi-Party Machine Learning (MPML)

Multi-Party Machine Learning (MPML) can be used to help multiple organizations or stakeholders train models collaboratively without the necessity of sharing raw data. This strategy is especially useful in fields where data security and security regulations are paramount, including medical care, finances, and intelligent cities. The methods such as Federated Learning (FL) and Secure Multi-Party Computation (SMPC) are the bases of MPML since they allow them to update the models together and preserve the privacy of individual datasets. Nevertheless, some of the unanswered questions of MPML, such as the creation of cross-organizational trust, model integrity in resistance to adversarial manipulation, and resistance to data poisoning or backdoor attacks that can undermine shared models, remain. Moreover, varying data distribution, computational concerns, and network situations among involved parties are major obstacles to uniform performance and equality among the parties.

2.3. Blockchain for Secure Collaboration

The blockchain technology is a de-centralized and unwriteable system that enhances trust and transparency over collaborative machine learning systems. Having an unchanging registry of the transactions, blockchain also guarantees the authenticity and hereditability of all data exchange, model changes, and training assistance. Its combination with MPML gives good assurances to integrity of data, accountability and secure aggregation of distributed learning outcomes. Smart contracts have the potential to further automate management of trust and apply rules of protocol without the use of centralized intermediaries. Nevertheless, current blockchain based collaborative learning models are usually limited concerning scalability, efficiency of communication and latency particularly when implemented to resource limited edge cloud computational setups. Furthermore, the security assurances of blockchain and the computational and energy limitations of edge devices are also a fundamental research problem in the development of realistic and scalable blockchain-enhanced collaborative Learning Systems.

3. Methodology

3.1. System Architecture

3.1.1. Edge Nodes:

The first layer of the proposed framework is edge nodes and they work near data sources, i.e. IoT devices, sensors, or local gateways. The tasks of them include gathering the raw data produced on-the-fly, the preprocessing of the raw data, such as noise removal and feature extraction, and the localized model training. Edge nodes lower communication latency, usage of bandwidth, and privacy as sensitive data does not go off-site because the processing happens locally. Moreover, the locally trained models or their gradients are safely transferred to the cloud layer or blockchain layer to aggregate and validate the models.

3.1.2. Cloud Nodes:

Cloud nodes are the coordination point in the system, which handles large-scale computation and world-wide model aggregation. They are fed model changes via several edge nodes by means of which these changes are combined to create a global model and redistributed to the edges through which the smoothed model is reused further training processes. Responsibilities of the cloud layer also cover scheduling the resources, aligning data, and optimization in performance in order to stabilize learning process and in heterogeneous environments, the learning process remains efficient. Its capacity to perform complex computations makes it possible to undertake the process of model fusion, hyperparameter optimization, and performance monitoring.

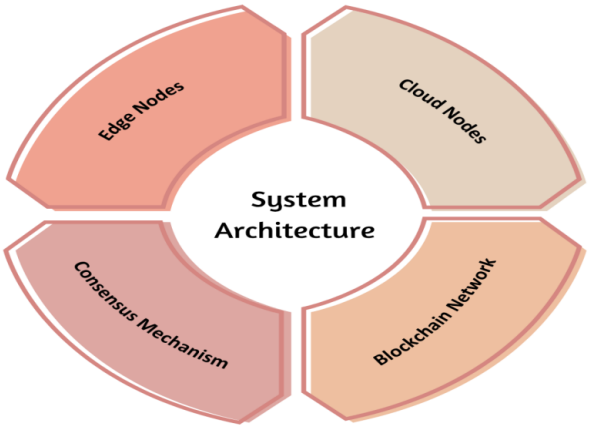

Figure 3. System Architecture

3.1.3. Blockchain Network:

The blockchain network constitutes the trust system in the suggested architecture as it is a decentralized and immutable registry of every model update and transaction. It captures the contribution of every edge node, being transparent, accountable and traceable in collaborative learning. Smart contracts embedded into the blockchain automatically impose rules in the system which can include access control and verifying that the update was made and can also provide rewards to participating nodes. This decentralized system will get rid of any central authority which will lessen single points of failure and other malaise manipulations.

3.1.4. Consensus Mechanism:

The consensus system means that all nodes within the blockchain systems will agree on the correctness of the model updates and their order of procedures to be put permanent before being added. It is very important in ensuring that data is not tampered with by unauthorized personnel. Consensus algorithms in use can include Proof of Stake (PoS), Practical Byzantine Fault Tolerance (PBFT), or Delegated Proof of Stake (DPoS) depending on the needs of the system to obtain a secure and efficient validation. This ensures that only authentic and trusted updates of the model are added to the blockchain, which will promote a sense of reliability and trust between all the interacting parties.

3.2. Secure Multi-Party Computation (SMPC)

Secure Multi-Party Computation (SMPC) is an encryption method that allows two or more parties to execute a calculation on their confidential data without revealing the original information to each other. Within the suggested framework, SMPC makes edge and cloud nodes able to collectively train and combine machine learning models and guarantee data confidentiality. The system uses a hybrid of additive secret sharing and homomorphic encryption in order to safely carry out model aggregation. Such a model prevents data leaks and malicious inference, as well as permits trusted cooperation between distributed and potentially untrusted nodes.

3.2.1. Edge Nodes:

The first layer of the proposed framework is edge nodes and they work near data sources, i.e. IoT devices, sensors, or local gateways. The tasks of them include gathering the raw data produced on-the-fly, the preprocessing of the raw data, such as noise removal and feature extraction, and the localized model training. Edge nodes lower communication latency, usage of bandwidth, and privacy as sensitive data does not go off-site because the processing happens locally. Moreover, the locally trained models or their gradients are safely transferred to the cloud layer or blockchain layer to aggregate and validate the models.

3.2.2. Cloud Nodes:

Cloud nodes are the coordination point in the system, which handles large-scale computation and world-wide model aggregation. They are fed model changes via several edge nodes by means of which these changes are combined to create a global model and redistributed to the edges through which the smoothed model is reused further training processes. Responsibilities of the cloud layer also cover scheduling the resources, aligning data, and optimization in performance in order to stabilize learning process and in heterogeneous environments, the learning process remains efficient. Its capacity to perform complex computations makes it possible to undertake the process of model fusion, hyperparameter optimization, and performance monitoring.

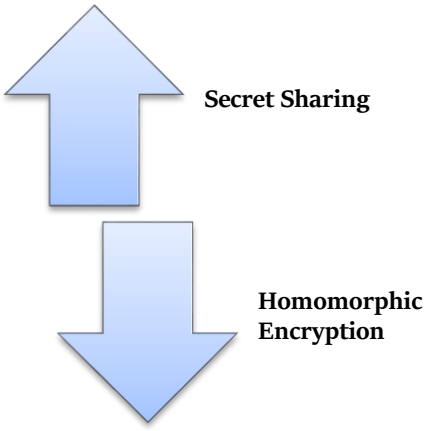

Figure 4. Secure Multi-Party Computation (SMPC)

3.2.3. Secret Sharing:

New Technology Secret sharing splits a private input value (such as a model parameter x) into several random values or shares, which are spread out over n nodes. The concept is that the original value cannot be reassembled by any individual node making it private. Mathematically the secret x can be divided as:

$$x = s_1 + s_2 + s_3 + \ldots + s_n$$

Given that s I, s 2, and so forth, s n are the random shares of each node. The local calculation is performed on each word with the help of its own share and the original value x could be restored only after the synthesis of all the shares. The mechanism enables distributed computation of model parameters without exposing sensitive information to one participant hence increasing privacy and confidence in collaborative learning models.

3.2.4. Homomorphic Encryption:

Homomorphic encryption allows mathematical operations to operate on the encrypted data directly, and yields an encrypted result, which on decryption will be identical to the result of doing mathematical operations on the plaintext. That is, we can compute, without ever bringing to light the underlying data. e.g. assuming Enc(x) and Enc(y) denote the encrypted values of x and y, then:

$$\mathbf{Enc(x) + Enc(y) = Enc(x + y)}$$

This property enables the safe aggregation of model updates by various nodes with the data of the respective nodes remaining confidential. It is possible to add encrypted model parameters across many different sources to the cloud or blockchain, but it does not need to decrypt them to guarantee the end-to-end privacy of the training and aggregation. This is what renders homomorphic encryption an essential support of privacy-sensitive distributed and multi-party machine learning.

3.3. Federated Learning Integration

3.3.1. Local Training on Edge Nodes:

In the suggested architecture, federated learning (FL) starts with training at the edge nodes. Local models The edge devices each train a local machine learning model based on their respective datasets, which are then kept locally to support privacy. The decentralized method guarantees that sensitive data, including user information, or sensor readings are never transferred to the edge environment. Raw data are not shared but instead only model parameters or gradients are shared to be aggregated. This highly minimizes the risk of data exposure and also reduces the communication cost that may be caused by transmitting the large datasets instead of compact model updates.

3.3.2. Secure Aggregation Using SMPC:

Following the local training, the model updates generated by every edge node are felicitously amalgamated by the means of Secure Multi-Party Computation (SMPC). By using schemes like additive secret sharing and Homomorphic encryption, the process of

aggregation is carried out by summing local model parameters without disclosing local updates to the other players or the central aggregator. This makes it impossible to deduce sensitive information using the shared gradients with the help of a single party, including the cloud server. The outcome is a globally revised model that gains out of the shared knowledge and at the same time, preserves privacy and trust between all actors.

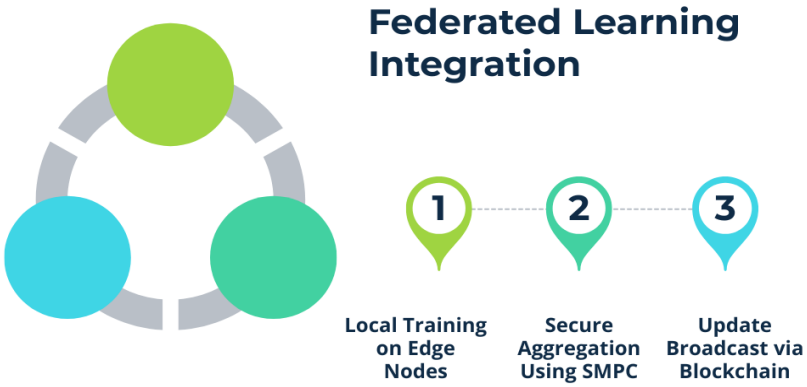

Figure 5. Federated Learning Integration

3.3.3. Update Broadcast via Blockchain:

After the secure aggregation is done, the global model update is stored, and the update is transmitted via the blockchain network. All the updates are recorded as part of the immutable transactions which guarantee traceability, accountability and resistance to tampering. The process is controlled by smart contracts that verify the authenticity of updates and control the accessibility of the participating nodes. This Decentralized broadcasting approach avoids the risk of the introduction of malicious changes, enables transparent cooperation and secures that all edge nodes have received up-to-date and valid versions of the model to be used in the next round of training. The framework enhances data integrity and creates trust in a distributed environment by using blockchain in federated learning.

3.4. Blockchain Consensus Mechanism

The proposed system incorporates the Practical Byzantine Fault Tolerance (PBFT) consensus-mechanism in an attempt to provide way to secure, reliable, and efficient updates to the model across the blockchain network. PBFT most effectively assumes the use of consortium or permissioned blockchain (where only a trusted set of participants, comprising edge nodes, cloud nodes, etc.) operate in a partially trusted environment. In contrast to PoW (Proof of Work) or PoS (Proof of Stake), a token-based validity method and based on high-caliber computation, PBFT does not demand significant computational effort, instead a multi-phase voting scheme is used to agree on consensus between interacting nodes, even when malicious or faulty nodes are present. Each round nodes communicate with one another to perform updates to the model, which is verified by all the honest nodes at least uniquely over a series of communication rounds (pre- prepare, prepare and commit) before all the honest nodes come into agreement with the same version of the model and at this point, the model is added to the blockchain ledger. The smart contracts are incorporated into the consensus process in order to ensure fairness, accountability, and transparency. Such contracts automatically establish established rules and conditions, including the rate of updating of the models, validation of the contributions of the nodes and reward or incentive distribution to contributing nodes based on the contributions that have been validated. This automation minimizes human intervention, there is a lot less overhead in the operation and there would be no requirement to have centralized operation. Moreover, the PBFT-based mechanism offers high-speed validation (that is low-latency) which is appropriate when using edge-cloud environments that need real-time or near-real-time learning. It can resist the attack of unwanted attacks, such as data manipulation and model poisoning, as it can survive up to one-third of malicious or faulty nodes without affecting the integrity of the entire system. The framework has struck a balance between performance, performance, and security by integrating PBFT and smart contracts. It will guarantee that each update of a model is transparently verified, that no submissions or tampering can occur and that a reliable and auditable account of joint learning activities is created. This consensus mechanism is therefore very critical in ensuring reliability and integrity of the decentralised federated learning ecosystem.

4. Results and Discussion

4.1. Experimental Setup

The proposed blockchain-enabled federated learning system combined with Secure Multi-Party Computation (SMPC) is tested on the basis of the experimental setup to measure the performance, efficiency, and safety of the proposed framework. To evaluate the generalization capabilities of the model to both simple and complex visual recognition, two bench mark datasets, MNIST, and CIFAR-10, are used. A simple system of 60,000 training images and 10,000 testing images of handwritten digits known as MNIST dataset is a small dataset that can be used to test model convergence and bare learning integrity. Conversely, CIFAR-10 has 60,000 color images with ten object classes to gain a more challenging situation that challenges the framework with increased computation requirement, its capability, and communication efficiency. The simulation model of the cloud-edge architecture has ten edge nodes and one cloud node. The edge nodes will be configured to have dissimilarity in terms of computational resources, which is reflective of a realistic distributed environment where a device can be different in terms of processing power, memory storage and network bandwidth. The local model training is done on each edge node by training on a part of the dataset, data is processed by local feature extraction, and encrypted model changes are sent via the SMPC protocol. The cloud node which is installed on a high-performance server conducts aggregation of the global model, coordinates the process of synchronization between edge devices and interacts with the blockchain network to maintain integrity and traceability of the updates. Key metrics, which are model accuracy, communication overhead, latency and security verification, are used to evaluate performance. Model accuracy is a measure that holds how well the collaborative learning works between distributed nodes, and the costs that are associated with data transmission are measured by communication overhead. Latency measures system responsiveness, which is a key characteristic of cloud-edge coordination efficiency. Lastly, the verification of security is the means of making sure that the incorporation of the SMPC and blockchain aims at securing the data privacy, avoiding manipulation, and ensuring the reliable interaction of all parties involved. Such an arrangement gives a complete evaluation of how the proposed framework can be able to meet the goals of federated learning, secure and efficient, and scalable.

4.2. Model Accuracy

Table 1. Model Accuracy

Dataset	Centralized Accuracy	Proposed Framework Accuracy
MNIST	99.2%	98.7%
MNIST	87.5%	86.8%

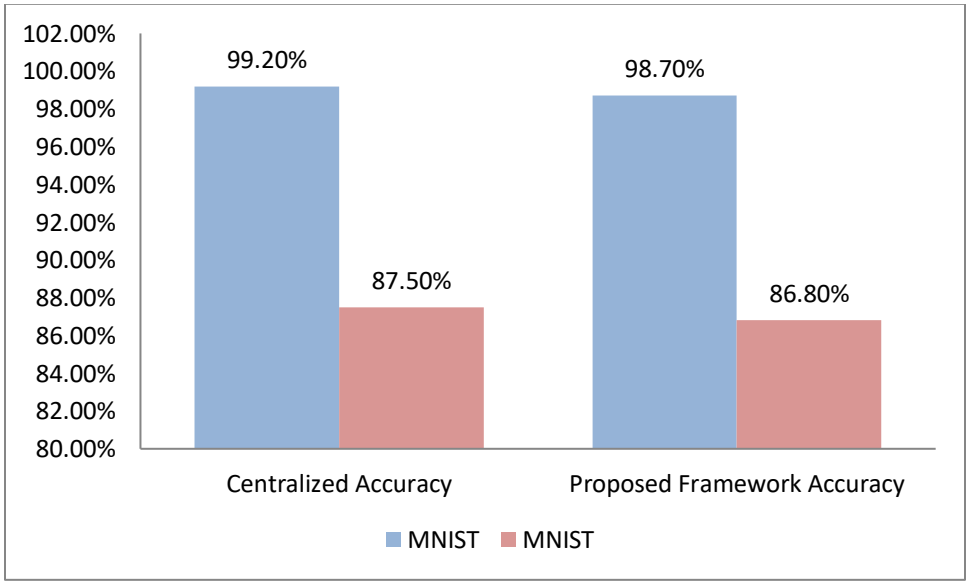

Figure 6. Graph representing Model Accuracy

4.2.1. MNIST Dataset:

In the case of MNIST data the presented blockchain-based federated learning model reached an accuracy of 98.7 which is just slightly lower than the accuracy 99.2 acquired through centralized training. This slight decrease of the accuracy is mostly related to the

safe aggregation procedure that has been achieved by combining both SMPC and blockchain. Regardless of this minor consideration, the framework effectively shows that distributed and privacy-protecting learning can provide performance comparable to centralized models without actual data exchange. The large accuracy in the case of MStNT also demonstrates that the framework effectively manages small picture image classification, despite the heterogeneity of the edges nodes and encrypted model transfer.

4.2.2. CIFAR-10 Dataset:

The proposed system produced an accuracy of 86.8 on the more complicated CIFAR-10 dataset, which was 87.5 in the centralized condition. The minor variance of 0.7% can be explained by the extra computing and communication costs imposed by the secure model aggregation and validation of consensus in the blockchain layer. However, the findings show that the framework has good generalization performance and scalability when it comes to processing data of high quantities and color images. The low accuracy loss indicates that the implementation of federated learning, SMPC, and blockchain offers a good tradeoff between privacy and security in addition to learning efficiency. On the whole, the findings of the experiment support the idea that the suggested decentralized model is capable of attaining the level of accuracy comparable to the centralized models with the promise of improved data confidentiality and trustworthiness of the system in place.

4.3. Communication Overhead

The communication overhead in the suggested blockchain-based federated learning system is a vital factor in considering the overall performance of the system with special emphasis on distributed clouds-edge systems where several nodes are in contact with each other on a regular basis. Secure Multi-Party Computation (SMPC) with blockchain technology integration also creates an extra communication latency based on the model parameters encrypted and consensus validation. The experimental results have shown that the framework incurs an estimated 10-15% of communication delay on top of the legacy federated learning configurations, which have not implemented security improvements. This overhead is largely due to two factors: the computational cost of the encryption and decryption of data in SMPC-based secure aggregation, and the network traffic of blockchain synchronization, block verification and block propagation. Nonetheless, the latency that has been achieved could still be considered acceptable during cloud-edge collaborative learning, as the trade-off between security, privacy, and performance is, nonetheless, crucial in this case. The edge nodes performing local computation send only model updates but not raw data, thus saving a lot of bandwidth usage at the expense of the traditional centralized systems. In addition, the blockchain layer competently handles the flow of updates using the optimized consensus protocols including Practical Byzantine Fault Tolerance (PBFT) which reduces unnecessary communication and guarantees the quick acceptance of various nodes. This scheme will make the cost of communication to grow in line with the number of participants making sure that the cost does not grow exponentially as the system size increases. Moreover, the model compression and update sparsification are additional techniques used to reduce further the load of communication in the framework to ensure that only the vital model parameters are shared. With these optimizations and asynchronous updates, they can keep the system responsive even when there is great heterogeneity of the nodes or even when there is an unstable network. Within the general context, as the application of cryptographic functions and blockchain validation induce a certain amount of quantified communication overhead, the enhancement of the data confidentiality, traceability, and trust made to counterbalance its increased latency is worthwhile. The structure, therefore, has a trade-off between secure cooperation and efficiency of communication in large scale edge-clouds.

4.4. Security Analysis

The safety of the suggested blockchain-based federated learning system is one of the most vital factors, as it can guarantee that the confidential information is stored securely, and that collaborative model training is credible in the context of distributed edges-cloud infrastructure. Using a combination of Secure Multi-Party Computation (SMPC), homomorphic encryption, and a blockchain-based consensus mechanism, the framework can be safely used to avoid data leakage, unauthorized model updates, and tampering. SMPC also guarantees that raw data does not exit the edge nodes, and giving the opportunity to calculate (like model aggregation) going through encrypted shares. This ensures that, where one of the nodes or communication channels is compromised it cannot extract any meaningful information about the underlying data. Homomorphic encryption makes this defense even more robust by allowing arithmetic operations to be done on encrypted model parameters so that all operations of aggregating and updating remain hidden. Blockchain network is an open decentralized register that will be used to store all model updates and the activity of participants permanently and irreversibly. Using Practical Byzantine Fault Tolerance (PBFT) in offering consensus, the system will guarantee that only authentic updates submitted by-honest nodes can be incorporated and attached to the blockchain. This will guard against malicious agents who may seek to offer falsified or poisoned model updates since any conflicting or invalid transaction will be automatically faced out during the consensus process. Smart contracts include rules within the system, including the frequency of

updating, verification of contribution and distribution of rewards, which decrease the possibility of human error or manipulation. Security of the framework was tested by penetration and integrity testing and the framework was attacked by simulating the attacks like unauthorized access, interception and updates with malicious updates. The tests have ensured that the framework has managed to identify and prevent these attacks without affecting the performance of the models or stability of the systems. Besides, integrity checking was done to ensure that the transactions recorded were consistent, tampered and could be traced back to their originating node. On the whole, such a combination of cryptographic schemes, blockchain integrity, and consensus-based validation creates a strong security stance, which can result in reliable, privacy preserving and auditable collaborative learning in heterogeneous and may have totally untrusted edge-cloud networks.

4.5. Discussion

The outcomes of the experiment imply the effectiveness and feasibility of deploying blockchain with Multi-Party Machine Learning (MPML) in a cloud-edge collaborative learning setup. The federated learning approach, the Secure Multi-Party Computation (SMPC) framework, and the use of a blockchain-based consensus mechanism make this framework effective to tackle the major issues of distributed machine learning such as data privacy, model integrity, and trust between heterogeneous participants. The comparison between the MNIST and CIFAR-10 data collections provides evidence that the framework is as accurate in the model as the conventional centralized training, which means that the gains of the security measures and decentralization do not impair the learning process significantly. It proves that privacy-preserving aggregation and encrypted computations are effectively adoptable with significant loss in model generalization or convergence. Although the integration of encryption and blockchain entails quantifiable communications and computational costs that are estimated at 1015 percent in terms of latency, this cost is justified by the possession of security, transparency, and accountability. The blockchain record guarantees all model changes are stored and verified, which prevents hackers who attempt to tamper with the model, breach, and model poisoning. Smart contracts replace authorities and human control by enforcing rules in the system, including checking updates and incentives based on visiting other individuals, which is automatically checked by other participants in the system. SMPC and homomorphic encryption also eliminate the risk of data leakage as the updates to models may be combined between the encrypted messages, and the sensitive data in edge nodes is confidential. The other observation that is important is the scalability of the framework. With the extra cryptographic and consensus functions, the system does not impede low-latency communication and efficient coordination between heterogeneous edge nodes and shows promise of being deployed in real-world IoT and edge-cloud infrastructures. The findings indicate that the MPML, when incorporated with blockchain, does not just enhance the security, but also the trust and accountability among the participants, which has become a key issue when the scope of multi-organization or cross-domain collaborations is considered, where the sensitivity of data and adherence is ranked high. Altogether, the suggested solution provides a sufficient balance between security, effectiveness, and performance by providing a scalable and decentralized solution to privacy-protecting, collaborative learning in cloud-edge architectures.

5. Conclusion

This paper introduces a unified computational system inspired by blockchain and intended to implement secure multi-party machine learning (MPML) on cloud-edge collaborative systems. The suggested framework tackles the increasingly high demand based on privacy-sensitive and reliable collaborative learning wherein customarily distributed nodes or multiple organizations can collectively train machine learning models without sharing raw data with each other. The system incorporates the ability, with the use of Secure Multi-Party Computation (SMPC), federated learning, and blockchain technology, to maintain sensitive data confidential on edge nodes, and to achieve secure aggregation and updates to global model. SMPC enables computations to be done with encrypted data shares, so that no single node has access to full datasets, and homomorphic encryption ensures that an aggregation of model updates can be achieved at the cost of no loss of privacy. The federated learning could be used to train the models on local edge devises and reduce the latency and bandwidth consumption because model parameters are sent without transmitting raw data. Blockchain is a second level of security and transparency which offers an irreversible registry of model updates, smart contract-related implementation of regulations, and consensus algorithms which check in contributions and resist interference and mischief.Experimental assessments of both MNIST and CIFAR-10 datasets show the framework has an extremely high model performance equal to centralized training, and there are relatively small performance losses due to secure aggregation and consensus algorithms. Communication overhead, latency, and security checks demonstrate that the extra computational and networking expense of encryption and blockchain integration are reasonable and allow the framework to be applicable to the cloud-edge deployment of real-world applications. The security test, penetration testing and integrity check confirm the power of the system to stop data leakage, unauthorized updates, and malicious concerns confirming that the combination of the SMPC and blockchain will have effective assurances of privacy, trust, and accountability. Scalability of the framework is also promising and the framework is well suited in

heterogeneous coordination of edge devices with the framework being convergent in the model and synchronization with low latency. It proves that it can be utilized in the real world in various fields including healthcare, finance, and smart city IoT networks, in which data sensitivity and regulatory compliance are the most significant values. Future research will involve additional efforts in optimizing the communication efficiency by using the methods like model compression, update sparsification, and increased support of other machine learning architectures, such as deep neural networks and graph-based models. Also, the implementation of the framework in the conditions of real operations will enable testing its strength in the environment of dynamism of the network and the conditions of different reliability of nodes and involvement in the activities of the large scale. All in all, this research can provide a safe and open and scalable paradigm of collaborative learning in cloud-edge ecosystems that strikes the right balance between privacy, efficiency, and trust and forms a basis of future developments in decentralized and privacy-preserving artificial intelligence.

References

[1] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge Computing: Vision and Challenges. IEEE Internet of Things Journal, 3(5), 637–646.
[2] Satyanarayanan, M. (2017). The Emergence of Edge Computing. Computer, 50(1), 30–39.
[3] Wang, S., Zhang, X., Zhang, Y., Wang, L., Yang, J., & Wang, W. (2017). A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications. IEEE Access, 5, 6757–6779.
[4] Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2017). Convergence of Edge Computing and Deep Learning: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 22(2), 869–904.
[5] Li, E., Zhou, Z., & Chen, X. (2018). Edge Intelligence: On-Demand Deep Learning Model Co-Inference with Device-Edge Synergy. Proceedings of ACM HotEdge, 2018.
[6] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. AISTATS 2017.
[7] Bonawitz, K., et al. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning. ACM CCS 2017.
[8] Konečný, J., McMahan, B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated Learning: Strategies for Improving Communication Efficiency. arXiv:1610.05492.
[9] Smith, V., Chiang, C.-K., Sanjabi, M., & Talwalkar, A. (2017). Federated Multi-Task Learning. NeurIPS 2017.
[10] Shokri, R. & Shmatikov, V. (2015). Privacy-Preserving Deep Learning. CCS 2015.
[11] Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. IEEE Access, 4, 2292–2303.
[12] Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. IEEE Security and Privacy Workshops (SPW), 180–184.
[13] Swan, M. (2015). Blockchain: Blueprint for a New Economy. O'Reilly Media.
[14] Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). Blockchain for IoT Security and Privacy: The Case Study of a Smart Home. IEEE PerCom Workshops 2017.
[15] Kaur, K., Garg, S., & Buyya, R. (2018). Security and Privacy Challenges in Cloud-Edge Collaboration. IEEE Cloud Computing, 5(3), 64–72.
[16] Mohassel, Payman & Zhang, Yupeng. "SecureML: A System for Scalable Privacy-Preserving Machine Learning." (2017)
[17] SecureML: A System for Scalable Privacy-Preserving Machine Learning — Payman Mohassel & Yupeng Zhang, 2017.
[18] Takada, Toshiyuki; Hanada, Hiroyuki; Yamada, Yoshiji; Sakuma, Jun; &-others. "Secure Approximation Guarantee for Cryptographically Private Empirical Risk Minimization." (2016)
[19] Garg, Sanjam; Miao, Peihan; Srinivasan, Akshayaram. "Two-Round Multiparty Secure Computation Minimizing Public Key Operations." (2018)
[20] (Survey) "An Exhaustive Survey on Privacy Preserving Machine Learning using Homomorphic Encryption and Secure Multiparty Computation Techniques." (2017) — provides background though full details span beyond 2017.