BLOCKCHAIN-BASED FEDERATED LEARNING

Anonymous Authors[1]

ABSTRACT

Federated learning (FL) is a rapidly growing research field in machine learning, it has been a lot during the last years and will keep developping. However, the traditional FL framework still faces some problems that can affect the functioning of the whole system. In this report, we discuss recent advances and present a collection of works that have been done to solve security issues in FL and improve incentive mechanisms. Besides, we investigate the advantages of combining FL with blockchain technology to create a blockchain-based federated learning system. We will show that blockchain can offer more security to FL and provide a way to reward the workers of the system.

1 INTRODUCTION

1.1 Distributed Machine Learning

The constraints challenges of real-world federated learning settings involve many open problems for the distribution of a federated machine learning. As a consequence, most researchers working on federated learning problems will likely not be deploying production FL systems, nor have access to fleets of millions of real-world devices. Therefore, we need to distinct the practical settings that motivate the work and experiments conducted in simulation which provide evidence of the suitability of a given approach to the motivating problem. Thus FL research can be seen differently from an experimental perspective compared to ML researches in other fields. Hence, we will need additional considerations to conduct FL research as presented in (Kairouz et al.).

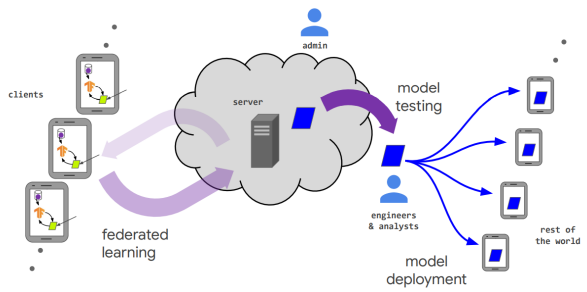

Figure 1. The lifecycle of an FL-trained model and the various actors in a federated learning system

1.2 Federated Learning

Federated learning (FL) is a distributed learning paradigm that aims to train machine learning models from scattered and isolated data. The special features of FL (compared to data center-based distributed training) are : statistical heterogeneity, system constraints and trustworthiness. To provide an answer to these 3 challenges we need to mix knowledge from different fields, including machine learning, wireless communication, mobile computing, distributed systems, and information security. Therefore FL is a truly interdisciplinary research field. FL has been studied well over the last years and is still developping. However, the traditional FL framework still faces some problems which reduce the reliability of the whole system. These problems detailed in (Wang & Hu) are the following : single point of failure, false data and the lalck of incentives.

- Single point of failure: this problem is based on the fiability of the aggregator which is the central server in a FL system. The structure of FL is presented Figure 2. This aggregator is employed to perform the integration of local training results and to update the global model. But, if the centralized aggregator is compromised, the whole FL system will go down. The reasons for a compromised aggregator are : an intentionally dishonest aggregation, an accidental network connection failure, or even an unexpected external attack.

-False data: despite the predefined protocols, and due to the huge number of clients in a FL mmodel we have to assume that not all the clients are honest and will use the model as expected. As a consequence, it may exist rogue clients who submit false data about their local training results. Therefore, the global performance of the model can be strongly affected. Besides, the whole FL system might be attacked by malicious clients via other means, such as training the local models using partial datasets.

-The lack of incentive methods: The lack of incen-

tives: in most of the traditional FL, clients contribute their computing powers without receiving any payments, this lead to the difficulty of encouraging clients to follow rigorously the protocol. Therefore the model will lack of reliable data. Moreover, the FL model will also lack of clients. Indeed, as FL requires multiple devices to work collaboratively, especially for the data-intensive training tasks where it needs a large number of participants.

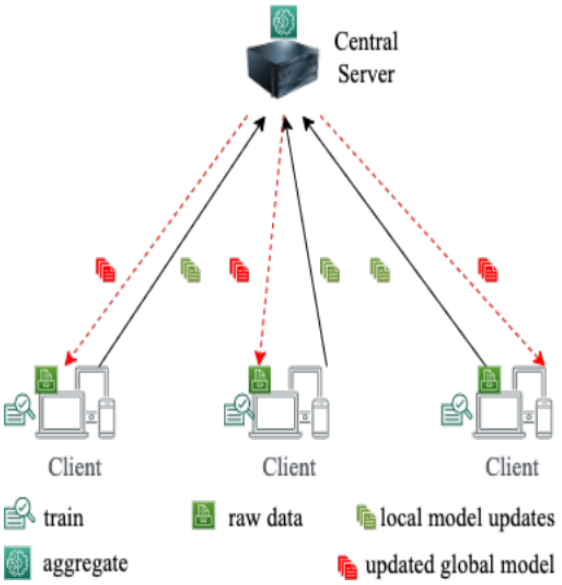

Figure 2. Topology of traditional FL

1.2.1 Attacks to Federated Learning

1. *Single point of failure attack*: the central server can jeopardize the security of the FL system in a number of ways, including (1) server instability that causes a system crash, (2) a compromised central server that produces a fake global model, and (3) maximum resource consumption.

2. *Denial of service and distributed denial of service attack*: malicious devices can overload the system and cause it to crash by persistently propagating incorrect model changes. Similar to this, if a FL server is compromised, it replicates this process and renders the FL system completely inoperable. Additionally, a malicious FL server may alter the initial global model with a new model that has a negligible difference in accuracy by injecting weak noise.

3. *Free-riding attack*: due to the high cost, dishonest participants will take advantage of incentives while not updating the local model. For instance, free-riders can immediately upload the untrained model and transmit noise-free false or comparable model updates. Therefore, this circumstance in FL systems may cause problems with fairness and reliability.

4. *Poisoning attacks*: data poisoning and model poisoning are the two subtypes of poisoning episodes. The data poisoning attack is launched and the fraudulent model updates are spread by altering the training data for the model. Additionally, hostile participants have the ability to change dataset labels and apply predefined poisoned model updates, both of which hurt the performance of the global FL model. Also, the model is updated by randomly produced gradients in random and reverse poisoning attacks, while the training model is updated in the opposite direction (Chen et al., 2018) and (Li et al., 2021b).

5. *Main-in-the-middle attack*: occurs between the FL server and FL client's communication. To send fake model updates and manage the traffic, the attacker impersonates a FL server or client. Session hijacking and IP spoofing are the most typical forms.

6. *Eavesdropping attacks*: causes for FL participants' private information to leak. Similar to the broadcasted paradigm between the FL server and players, an adversary can alter, corrupt, remove, or intercept it. Specailly, the jammer attack against FL systems poses a greater risk because it has the potential to maliciously obstruct network connection on the server or client end through collisions or interference.

1.3 Blockchain

Blockchain, is an emerging technology that come with several attractive properties : decentralization, anonymity, and traceability. These properties has already shown their utilities in different fields. More recently, blockchain begin to be used to address the challenges faced by the acual FL. First, decentralization can be achieved by the deployment of blockchain in FL by replacing the central aggregator by the peer-to-peer blockchain system. On the other hand, the job of aggregating the global model can be handled by blockchain nodes to avoid the unreliability of the whole FL system usually caused by the centralized server's failure.. Figure 3 indicates the topology of block hain. Moreover, blockchain gives verification mechanisms to FL in the name of transaction verification. These mechanisms allows FL model to remove the unqualified or even malicious local model updates before the global model update. Furthermore, blockchain can sucessffully distribute rewards to FL clients to reward their participation and honest behaviors.

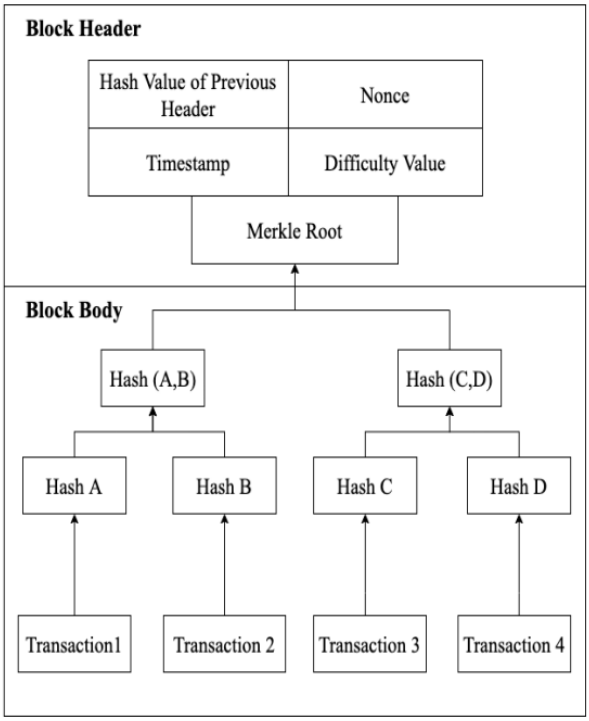

Figure 3. Structure of Block

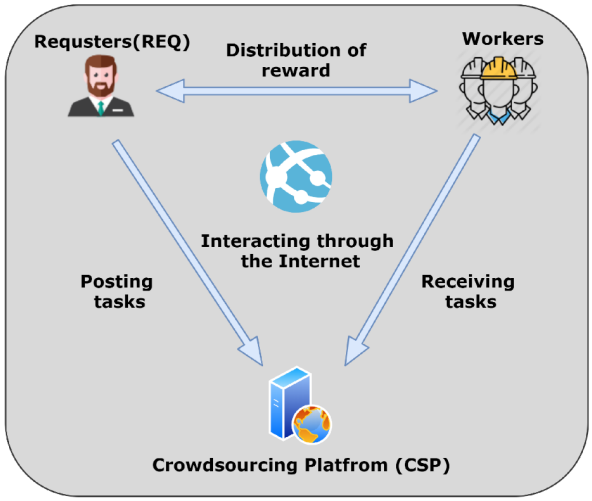

Figure 4. Traditional centralized crowdsourcing system

1.4 Crowdsourcing

Crowdsourcing involves obtaining work, information, or opinions from a large group of people who submit their data via the Internet, social media, and smartphone apps. People involved in crowdsourcing sometimes work as paid freelancers, while others perform small tasks voluntarily. For example, traffic apps like Waze encourage drivers to self-report accidents and other roadway incidents to provide real-time, updated information to app users. Today, crowd computing has different connotations depending on the perspectives. However, as a type of crowdsourcing, four basic components will always be required (also shown figure 4):

-Requester: typically, it is a server that hosts the crowd task. It is generally responsible for standardizing a task before releasing it. In some special tasks, it can even participate in the project as a unique worker.

-Workers: crowdworkers lend their devices to execute the jobs. In crowd computing, the entities of workers are generally idle devices contributed by crowdworkers, such as mobile phones or laptops. When a worker chooses to participate in a crowdsourcing task, their devices will execute the tasks according to an agreement. When the devices are found to be idle, the jobs are processed following the (CPU) cycle-stealing scheme.

-Crowdsourcing Platform: middleware for job management along with the server and client applications. The server application is responsible for creating tasks, discovering suitable workers, assigning tasks and scheduling tasks to the designated workers. Collecting the results from multiple workers, assembling them, and updating it in the server application, and for additional purposes. The platform will handle a large amount of work, so it usually needs to have strong computing performance or be deployed on the cloud.

-Network: all devices communicate through the Internet or a local area network. Traditional crowdsourcing applications generally uses the Internet as the network, and a platform as the center to control all devices. This paper innovatively proposes a blockchain-based crowdsourcing paradigm, and the essential explanation is described in Section 4.

2 Motivation

In the present time, the use of Machine learning (ML) is constantly increasing in every field. Thus, lots of data are generated and gathered from massive end users to train ML models which bring benefits in terms improve the serices offered to people. Therefore, ML framework usually requires end devices to transfer the collected data to the central server for training. But, this process causes two challenges : A large amount of communication resources is requirred to transfer data. Then, the submission of raw data increases the risk of privacy leakage, which making data owners unwilling to upload data to the central server by fear of security. FL paired with blockchain appeared to be a good solution

to solve this problem as well as the actual limits of FL mentionned earlier (single point of failure, false data and the lalck of incentives).

2.1 Security and privacy issues

In classical FL, the central FL server receives the trained local model updates, which are then combined to improve the overall FL model. Therefore, the entire model is dependent on the central FL server due to the model parameters aggregation scheme used in FL. Single Point of Failure (SPoF) and Distributed Denial of Service (DDoS) attacks result from the failure of a central server. Additionally, there is no visible means to record local model updates in the existing FL system.

2.2 Crowdsourcing issues

The development of crowdsourcing systems is often limited by the security weaknesses of the centralized nature of the systems, and there have been many research efforts to solve them. Encryption and differential privacy are used to protect the data privacy of all participants who involved in the communication process. Reputation-based mechanisms are proposed to address "false-reporting" and "free-riding" behavior.

3 Blockchain-based Federated Learning

Although Federated Learning allows for participants to contribute their local data without it being revealed, it faces issues in data security and in accurately paying participants for quality data contributions. Many researches have already been published in this fiels :

(Martinez et al.) propose an EOS Blockchain design and workflow to establish data security, a novel validation error based metric upon which (Martinez et al.) qualify gradient uploads for payment, and implement a small example of the blockchain Federated Learning model to analyze its performance. This workflow is designed for scalable recording and rewarding of gradients using both blockchain and off-chain databases of records at the same time. Their implementation on a small set of clients demonstrate that the blockchain does not interfere with the federated learning aggregation, while limiting the number of uploads and validating the claimed data cost per device.

(Awan et al., 2019) propose a blockchain-based privacy-preserving federated learning (BC-based PPFL) framework, which leverages the immutability and decentralized trust properties of blockchain to provide provenance of model updates. This framework is based on gradient aggregation over private data following a cryptographic protocol.

(Korkmaz et al., 2020) propose a decentralized federated learning approach named Chain FL that makes use of the blockchain to delegate the responsibility of storing the model to the nodes on the network instead of a centralized server. This methods doesn't effect results compared to traditional federated learning as they use the update steps the same way. However, results are slightly worse compared to classical machine learning.

(Sharma et al., 2020) propose a distributed computing defence framework for sustainable society using the features of blockchain technology and federated learning. This framework provides security against misbehavior detection in lightweight Internet of Medical Things (IoMT) devices, particularly in the artificial pancreas system (APS). The proposed approach employs privacy-preserving bidirectional long-short term memory (BiLSTM) and augments the security through the integration of Blockchain technology based on Ethereum smart contract environment.

In the work (Li et al., 2020) propose a crowdsourcing framework named CrowdSFL, that users can implement crowdsourcing with less overhead and higher security. Besides, to protect the privacy of participants, this framework include a new re-encryption algorithm based on Elgamal to ensure that interactive values and other information will not be exposed to other participants outside the workflow. As a result, this method is proved to be superior to some similar work in accuracy, efficiency, and overhead.

This paper (Chen et al., 2020) proposes a federated learning architecture based on the alliance chain, which defines a complete life cycle for the federated learning process based on blockchain technology. By combining it with the aggregation algorithm of federated learning, and by using knowledge distillation technology to extract network knowledge, the model compressed data before entering the block network for propagation, which reduces the load on the entire blockchain network while providing a better protection for data.

(Kumar et al., 2021) propose a framework that collects a small amount of data from different sources (various hospitals) and trains a global deep learning model using blockchain based federated learning. Additionally they collect real-life COVID-19 patients data, which is, open to the research community. The security of local parameters, the learning quality, and the varying computing and communication resources, are crucial issues that remain unexplored in federated learning schemes.

(Guo et al., 2020) propose a data sharing mechanism that

combines blockchain and federated learning over smart city. The security of local parameters, the learning quality, and the varying computing and communication resources, are crucial issues that remain unexplored in federated learning schemes. In response to the shortcomings of the original method, this paper designed the work nodes selection algorithm to enhance effectiveness of the federated learning task, and designed a consensus incentivemechanism to encourage the work node to more actively participate in the task. Combining this method with differential privacy technology is a good balance between privacy security and the practicality of the model

(Lu et al., 2021b) propose a general privacy-preserving federated learning scheme, which integrates blockchain with federated learning, for beyond 5G networks. Furthermore, this paper introduce potential application scenarios of the proposed scheme in beyond 5G. In the proposed scheme, the blockchain is used to maintain learning parameters and verify their accuracies, which can enhance the learning security and quality.

(Ma et al., 2021) propose a blockchain-based federated learning framework and a protocol to transparently evaluate each participant's contribution. The framework protects all parties' privacy in the model building phase and transparently evaluates contributions based on the model updates. Collaborative model development and privacy protection are critical considerations while training a global deep learning model. This method was tested on the handwriting digits dataset and demonstrate successful contributions evaluation.

To overcome computational complexity and privacy problems, (Durga & Poovammal, 2022) proposes the blockchain empowered federated framework to enhance the perception of multiple sources of heterogeneous CT images. It is based on sharing the data among the hospitals while maintaining privacy and security. Also, an ensemble of capsule networks and extreme learning machines are used for effective feature extraction and classification to detect the COVID-19 among the different sources of publicly available heterogenous CT image datasets. Besides, federated learning is adopted for the collaborative training of hospitals backed with blockchain technology. In addition, the combination of chaotic encryption keys in the process of data retrieval and sharing process has added more trust in terms of maintaining privacy and security.

In the work (Lu et al., 2020a), the problem of edge data sharing among vehicles in an internet of vehicles (IoV) framework is studied. This work proposes a hybrid blockchain mechanism that includes the permissioned blockchain and the local directed acyclic graph in IoV. Based on the hybrid blockchain mechanism, it proposed the asynchronous federated learning scheme and further improved the learning efficiency by using DRL to select the optimized participating nodes. By integrating learning parameters into the blockchain, the qualities of learned models can be then be verified through the two-stage verification.

(Lee & Kim, 2021), gaves a short overview of blockchain and federated learning and showed how blockchain technology can enhance and solve privacy issues. Moreover, it presents an application of a blockchain federated learning framework in the industrial, vehicle network, and healthcare sectors.

3.1 Impact on machine learning

3.1.1 Training

Blockchain can be used to train machine learning models in a privacy-compliant way (Ladia, 2019)), using a decentralized advanced-Proof-of-Work (aPoW) algorithm (Kusi et al., 2020), and that blockchain can be used to solve issues with traditional machine learning models, such as the lack of data privacy and the need for continuous training (Shengwen Ding & Chenhui Hu, 2022). (Xu et al., 2022) suggests that blockchain-based crowdsourcing can be used to collect and annotate data for machine learning models.

3.1.2 Sharing

Blockchain can also be used to share machine learning models while preserving privacy. (Ladia, 2019)), (Lu et al., 2020a), and (Lee & Kim, 2021) all suggest that blockchain can be used for this purpose. (A. B. Kurtulmus & K. Daniel, 2018) suggests that blockchain can be used to facilitate the exchange of machine learning models. It is suggested that blockchain can be used to share machine learning models while preserving privacy.

3.2 Characteristics

1. *Decentralization*: it has numerous decentralized servers that can store model changes with high resilience to single point of failure attacks.

2. *Traceability*: all block histories are kept connected in a chain. The authors of model changes cannot be disputed by participants.

3. *Immutabililty*: the server can detect and prevent record tampering. To make each block permanent and unchangeable, it has a distinct hash value.

4. *Incentives*: participants are engaged through incentive or reward systems, and as a result, they contribute with high-quality data model updates that lead to an accurate global model.

5. *Integrity and realiability*: since every block is connected through cryptography, any data changes can be quickly identified. Blockchain demonstrates its inherent reliability and security.

6. *Trust*: a consensus method is used to build trust between participants. Participants that accept the contract are permitted to take part in training sessions.

3.3 Architecture

Blockchain-based FL system and its core elements:

1. *Federated Learning participants*: participants engage in model training and submit local model changes for verification and aggregation throughout the next phase.

2. *FL integration with blockchain*: the integration serves as middleware to enable communication between blockchain and FL players.

3. *Miners working*: the FL participants communicate local model updates to the miners. Each FL participant and data holder has a direct line of communication with the miner, ensuring ongoing collaboration.

4. *Smart contract*: employed by researchers (Khan et al., 2021) for a variety of purposes, including participant registration, model training coordination, local model update aggregation, participant evaluation, and reward administration.

5. *Consensus algorithm*: he formation, verification, and acceptance of a new block on a blockchain network are all governed by a shared agreement reached by all stakeholders. The miners agree to create a completely converged global model.

6. *Blockchain network*: The blockchain network is expanded with verified new blocks. Until the required learning rate is reached, the FL model process continues.

3.4 Workflow

Until the global model has reached the proper learning rate or has reached full convergence, these stages (part of one-epoch operation) are repeated.

1. *Local model training*: FL clients first train the local model updates using their own local datasets, then upload the model for additional steps like verification and aggregation.

2. *Smart contract execution*: for engagement among parties in the blockchain network. For instance, when FL participants meet the requirements for the FL model training process, they register through the smart contract. Following the successful registration of FL participants, the miners receive local model upgrades.

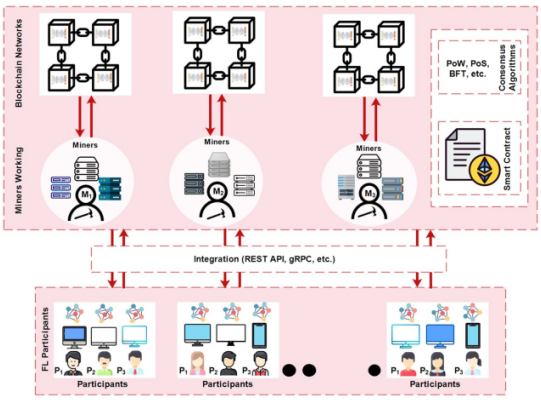

Figure 5. Blockchain-based federated learning architecture

3. *Local model upload*: the local model updates are sent to the blockchain miners for upload. On the basis of the consensus mechanism, the miners authenticate and verify local model modifications.

4. *Start mining process*: the registered FL participants send local model updates to the linked miners. The local model updates are then aggregated and verified by the miners.

5. *Run consensus algorithm*: the consensus algorithm is executed by each miner up until it receives a freshly created block from other miners. The network's miners then broadcast the new block.

6. *Add block into blockchain*: the blockchain network now has a fresh block added to it.

7. *Download global model*: the global model can be downloaded upon request by devices. Since FL participant devices used their own resources to train the model, they are free to download it. External devices, on the other hand, must pay fees in order to access the global model. The entire community can gain from fully trained models in this way.

4 Research Lines

4.1 Security issues: privacy

There are many security problems neglected in federated learning, Indeed, FL allows for participants to contribute their local data to a common goal, thus, it faces issues in data security. To answer this challenge, FL provides an attractive

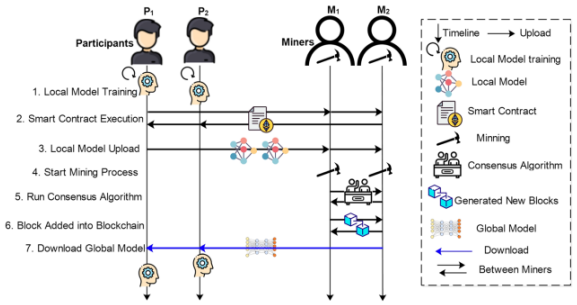

Figure 6. Blockchain-based federated learning system's single-epoch workflow

structure (presented in (Kairouz et al.)) for decomposing the overall machine learning workflow into the approachable modular units we desire. One of the primary attractions of the federated learning model is that it can provide a level of privacy to participating users through data minimization: the raw user data never leaves the device, and only updates to models (e.g., gradient updates) are sent to the central server. These model updates are more focused on the learning task at hand than is the raw data (i.e. they contain strictly no additional information about the user, and typically significantly less, compared to the raw data), and the individual updates only need to be held ephemerally by the server. To solve this accuracy problem, (Weng et al., 2021) present a distributed, secure, and fair deep learning framework. (Nguyen et al., 2020) provide a state-of-art survey on the integration of blockchain with 5G networks and beyond. (Kumar et al., 2020) introduce a secure and decentralized training for distributed data. (Lu et al., 2020b) propose a new architecture based on federated learning to relieve transmission load and address privacy concerns of providers. (Otoum et al., 2020) introduce a solution that integrates both federated learning and blockchain to ensure both data privacy and network security. The security of federated learning is increasingly being questioned, due to the malicious clients or central servers' constant attack to the global model or user privacy data. To address these security issues (Li et al., 2021a) propose a decentralized federated learning framework based on blockchain, i.e., a Blockchain-based Federated Learning framework with Committee consensus (BFLC). (Liu et al., 2020b) propose a blockchain-based secure FL framework to create smart contracts and prevent malicious or unreliable participants from involving in FL. (Lu et al., 2021a) introduce the digital twin wireless networks (DTWN) by incorporating digital twins into wireless networks, to migrate real-time data processing and computation to the edge plane. In further (Xu et al., 2021) propose the concept of device's score and use entropy weight method to measure the quality of model update. Other influential work includes (Lu et al., 2020a).

4.1.1 Impact on security and privacy

Blockchain can be used to improve the security of machine learning models. (Chen et al., 2020) found that the blockchain technique can be used to design a decentralized privacy-preserving and secure machine learning system. (Fadaeddini et al., 2019) found that the proposed framework uses the Stellar Blockchain infrastructure for secure decentralized training of the deep models. (Goel et al., 2019) found that the proposed approach provides security to both deep learning model and the biometric template. (Wang et al., 2020) found that the proposed system obtains comparable performance with that of conventional distributed systems, and bounded performance in the case of Byzantine attacks. Therefore, blockchain appears to be a promising solution for improving the security of machine learning models.

Blockchain also can be used to train machine learning models while preserving privacy. (Ladia, 2019), (Chen et al., 2018), and (Weng et al., 2021) all propose different ways to do this. These papers suggest that blockchain is a promising solution for training machine learning models while preserving privacy.

4.2 Quality management: incentive

There are many security problems neglected in federated learning, for example, the participants may behave incorrectly in gradient collecting or parameter updating, and the server may be malicious as well. In FL processing, the data quality shared by usersdirectly affects the accuracy of the federated learning model, and how to encourage more data owners to share data is crucial. Inother words, how to design a good incentive mechanism is the key problem in FL.

(Weng et al., 2021) present a distributed, secure, and fair deep learning framework named *DeepChain* to solve these problems. This framework provides a value-driven incentive mechanism based on Blockchain to force the participants to behave correctly. This paper (Kang et al., 2019) introduces reputation as the metric to measure the reliability and trustworthiness of the mobile devices. Then, it designs a reputation-based worker selection scheme for reliable federated learning by using a multiweight subjective logic model. (Liu et al., 2020a) propose FedCoin, a blockchain-based peer-to-peer payment system for FL to enable a feasible Shapley Value (SV) based profit distribution. In FedCoin, blockchain consensus entities calculate SVs and a new block is created based on the proof of Shapley (PoSap) protocol. (Lin et al., 2022) propose a novel Wirelessly Powered Edge intelliGence (WPEG) framework, which aims to achieve a stable, robust, and sustainable edge intelligence by energy harvesting (EH) methods. Besides, by constructing a two-stage Stackelberg

game, the underlying energy-knowledge trading incentive mechanisms are also proposed with the optimal economic incentives and power transmission strategies. (Wang et al., 2021) propose SFAC, a secure federated learning framework for UAV-assisted mobile crowdsensing (MCS). This framework is composed of a two-tier reinforcement learning-based incentive mechanism that promote UAVs' high-quality model sharing when explicit knowledge of network parameters are not available in practice. A new ecosystem of ML model trading over a trusted Blockchain-based network is proposed by (Nguyen et al., 2021). Extensive experimental evaluation of the proposed approach shows a competitive runtime performance, with a 15% drop in the cost of execution, and fairness in terms of incentives for the participants. Problems, however, can arise if there is a lack of quality data for AI-model training, scalability, and maintenance. (Chaabene et al., 2022) propose a data-centric federated learning architecture leveraged by a public blockchain and smart contracts to overcome this significant issue. The proposed approach employs privacy-preserving bidirectional long-short term memory (BiLSTM) and augments the security through the integration of Blockchain technology based on Ethereum smart contract environment (Rahmadika et al., 2022). While several works focus on strategic incentive designs and client selection to overcome this problem, there is a major knowledge gap in terms of an overall design tailored to the foreseen digital economy, including Web 3.0, while simultaneously meeting the learning objectives. To address this gap (Pandey et al., 2022) propose a contribution-based tokenized incentive scheme, namely `FedToken`, backed by blockchain technology that ensures fair allocation of tokens amongst the clients that corresponds to the valuation of their data during model training.

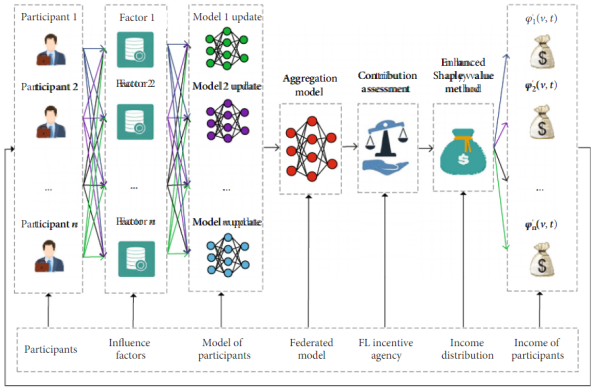

Figure 7. FL incentive model

In his work, (Yang et al., 2022) propose an incentive mechanism based on the enhanced Shapley value method for FL. In this mechanism presented figure 7, the enhanced Shapley value method is proposed to measure income distribution, which takes multiple influence factors as weights. The analytic hierarchy process (AHP) is used to find the corresponding weight value of the influence factors. Finally, the numerical experiments are carried to verify the performance of the proposed incentive mechanism. Other influential work includes (Lu et al., 2020a).

4.2.1 *Multifactor incentives*

In order to provide a better experience for both the workers and the requester of the model, we should investigatee the possibility to provide a multifactor incentive mechanism. A multifactor incentive is based on three parameters : the workers binding, their reputation and the estimation of the quality of their recent data. From these parameters we can define a Analytic Hierarchy Process (AHP) based on the following points :
-Objective level: winning workers
-Criteria level: parameters criteria
-Alternative level: workers available
In order to evaluate this multifactor incentive approach, we use the following formula which apply weights to the different parameters of the AHP:

$$\theta_i = \omega_1 B_i + \omega_2 R_i + \omega_3 Q_i \tag{1}$$

Where

$$\omega_i \geq 0 \tag{2}$$

and

$$\sum_{\omega_i=1}^{3} \omega_i = 1 \tag{3}$$

4.2.2 *Impact on incentive models*

Blockchain can be used to incentivize federated learning. (Kumar et al., 2020), (Pandey et al., 2022), (Zhilin Wang et al., 2022) and (Martinez et al.) all found that blockchain can be used to create a value-driven incentive mechanism, a contribution-based tokenized incentive scheme, an incentive mechanism to allocate resources, and establish data security and accurately pay participants, respectively. This incentivizes clients to participate in federated learning, which improves the efficiency of the training. Consequently, blockchain does have an impact on incentive models of federated learning.
Blockchain also can be used to create new incentive models for federated learning. (Kumar et al., 2020) found that blockchain technology can be used to create a value-driven incentive mechanism for federated learning. (Zhilin Wang et al., 2022) found that a two-stage Stackelberg game can be used to allocate resources for clients in blockchain-based federated learning. (De Brito Goncalves & Da Silva Villaca, 2022) found that a blockchain-based federated learning

framework can be used to increment the system security using an incentive mechanism. Together, these papers suggest that blockchain can be used to create new incentive models for federated learning.

5 Conclusion

In this report, we have seen the benefits of consortium blockchain technology that begin with, a resistant system to the single point of failure problems of traditional FL. Then, blockchain offers more transparency to the system by trusting the provenance of data and helps also to preserve th privacy of the workeers that contribute to the system with their datas. Finally, the cooperative management of blockchain reduces cost and enhances the flexibility of the system. As presented, the use of hybrid incentive mechanism also provide benefits to FL systems. Indeed, it encourages workers to contribute valuable data and penalizes malicious ones thanks to more accurate rewards. Furthermore, hybrid incentives ensures favorables short-term and long-term incentives for workers which is beneficial to the whole federated learning system.

References

A. B. Kurtulmus and K. Daniel. Trustless Machine Learning Contracts; Evaluating and Exchanging Machine Learning Models on the Ethereum Blockchain, 2018.

Awan, S., Li, F., Luo, B., and Liu, M. Poster: A reliable and accountable privacy-preserving federated learning framework using the blockchain. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, pp. 2561–2563, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367479. doi: 10.1145/3319535.3363256. URL https://doi.org/10.1145/3319535.3363256.

Chaabene, R. B., Amayed, D., and Cheriet, M. Leveraging centric data federated learning using blockchain for integrity assurance, 2022. URL https://arxiv.org/abs/2206.04731.

Chen, X., Ji, J., Luo, C., Liao, W., and Li, P. When Machine Learning Meets Blockchain: A Decentralized, Privacy-preserving and Secure Design. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 12 2018. doi: 10.1109/bigdata.2018.8622598. URL http://dx.doi.org/10.1109/BigData.2018.8622598.

Chen, Y., Chen, Q., and Xie, Y. A methodology for high-efficient federated-learning with consortium blockchain. In *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 3090–3095, 2020. doi: 10.1109/EI250167.2020.9347025.

De Brito Goncalves, J. P. and Da Silva Villaca, R. A Blockchained Incentive Architecture for Federated Learning. In *2022 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 8 2022. doi: 10.1109/blockchain55522.2022.00074. URL http://dx.doi.org/10.1109/Blockchain55522.2022.00074.

Durga, R. and Poovammal, E. Fled-block: Federated learning ensembled deep learning blockchain model for covid-19 prediction. *Frontiers in Public Health*, 10, 2022.

Fadaeddini, A., Majidi, B., and Eshghi, M. *Privacy Preserved Decentralized Deep Learning: A Blockchain Based Solution for Secure AI-Driven Enterprise*, pp. 32–40. Springer International Publishing, 2019. doi: 10.1007/978-3-030-33495-6_3. URL http://dx.doi.org/10.1007/978-3-030-33495-6_3.

Goel, A., Agarwal, A., Vatsa, M., Singh, R., and Ratha, N. Securing CNN Model and Biometric Template using Blockchain. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 9 2019. doi: 10.1109/btas46853.2019.9185999. URL http://dx.doi.org/10.1109/BTAS46853.2019.9185999.

Guo, S., Xiang, B., Xia, X., Yan, Z., and Li, Y. Blockchain and federated learning based data security sharing mechanism over smart city. 11 2020. doi: 10.21203/rs.3.rs-104012/v1.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. Advances and open problems in federated learning. URL http://arxiv.org/abs/1912.04977. version: 3.

Kang, J., Xiong, Z., Niyato, D., Xie, S., and Zhang, J. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal*, 6(6): 10700–10714, 2019. doi: 10.1109/JIOT.2019.2940820.

Khan, S. N., Loukil, F., Ghedira, C., Benkhelifa, E., and Bani-Hani, A. I. Blockchain smart contracts: Applica-

tions, challenges, and future trends. *Peer-to-Peer Networking and Applications*, 14:2901 – 2925, 2021.

Korkmaz, C., Kocas, H. E., Uysal, A., Masry, A., Ozkasap, O., and Akgun, B. Chain fl: Decentralized federated machine learning via blockchain. In *2020 Second International Conference on Blockchain Computing and Applications (BCCA)*, pp. 140–146, 2020. doi: 10.1109/BCCA50787.2020.9274451.

Kumar, R., Khan, A. A., Kumar, J., Zakria, Golilarz, N. A., Zhang, S., Ting, Y., Zheng, C., and Wang, W. Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging. *IEEE Sensors Journal*, 21(14):16301–16314, jul 2021. doi: 10.1109/jsen.2021.3076767. URL https://doi.org/10.1109%2Fjsen.2021.3076767.

Kumar, S., Dutta, S., Chatturvedi, S., and Bhatia, M. Strategies for enhancing training and privacy in blockchain enabled federated learning. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 333–340, 2020. doi: 10.1109/BigMM50055.2020.00058.

Kusi, G. A., Xia, Q., Cobblah, C. N. A., Gao, J., and Xia, H. Training Machine Learning Models Through Preserved Decentralization. In *2020 16th International Conference on Mobility, Sensing and Networking (MSN)*. IEEE, 12 2020. doi: 10.1109/msn50589.2020.00080. URL http://dx.doi.org/10.1109/MSN50589.2020.00080.

Ladia, A. *Privacy Centric Collaborative Machine Learning Model Training via Blockchain*, pp. 62–70. Springer International Publishing, jun 25 2019. doi: 10.1007/978-3-030-23813-1_8. URL http://dx.doi.org/10.1007/978-3-030-23813-1_8.

Lee, H. and Kim, J. Trends in blockchain and federated learning for data sharing in distributed platforms, 2021. URL https://arxiv.org/abs/2107.08624.

Li, Y., Chen, C., Liu, N., Huang, H., Zheng, Z., and Yan, Q. A blockchain-based decentralized federated learning framework with committee consensus. *IEEE Network*, 35(1):234–241, jan 2021a. doi: 10.1109/mnet.011.2000263. URL https://doi.org/10.1109%2Fmnet.011.2000263.

Li, Z., Liu, J., Hao, J., Wang, H., and Xian, M. Crowdsfl: A secure crowd computing framework based on blockchain and federated learning. *Electronics*, 9(5), 2020. ISSN 2079-9292. doi: 10.3390/electronics9050773. URL https://www.mdpi.com/2079-9292/9/5/773.

Li, Z., Yu, H., Zhou, T., Luo, L., Fan, M., Xu, Z., and Sun, G. Byzantine resistant secure blockchained federated learning at the edge. *IEEE Network*, 35(4):295–301, 2021b. doi: 10.1109/MNET.011.2000604.

Lin, X., Wu, J., Bashir, A. K., Li, J., Yang, W., and Piran, M. J. Blockchain-based incentive energy-knowledge trading in iot: Joint power transfer and ai design. *IEEE Internet of Things Journal*, 9(16):14685–14698, 2022. doi: 10.1109/JIOT.2020.3024246.

Liu, Y., Ai, Z., Sun, S., Zhang, S., Liu, Z., and Yu, H. *FedCoin: A Peer-to-Peer Payment System for Federated Learning*, pp. 125–138. Springer International Publishing, Cham, 2020a. ISBN 978-3-030-63076-8. doi: 10.1007/978-3-030-63076-8_9. URL https://doi.org/10.1007/978-3-030-63076-8_9.

Liu, Y., Peng, J., Kang, J., Iliyasu, A. M., Niyato, D., and El-Latif, A. A. A. A secure federated learning framework for 5g networks. *IEEE Wireless Communications*, 27(4):24–31, aug 2020b. doi: 10.1109/mwc.01.1900525. URL https://doi.org/10.1109%2Fmwc.01.1900525.

Lu, Y., Huang, X., Dai, Y., Maharjan, S., and Zhang, Y. Blockchain and federated learning for privacy-preserved data sharing in industrial iot. *IEEE Transactions on Industrial Informatics*, 16(6):4177–4186, 2020a. doi: 10.1109/TII.2019.2942190.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles. *IEEE Transactions on Vehicular Technology*, 69(4):4298–4311, 2020b. doi: 10.1109/TVT.2020.2973651.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Low-latency federated learning and blockchain for edge association in digital twin empowered 6g networks. *IEEE Transactions on Industrial Informatics*, 17(7):5098–5107, 2021a. doi: 10.1109/TII.2020.3017668.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Blockchain and federated learning for 5g beyond. *IEEE Network*, 35(1):219–225, 2021b. doi: 10.1109/MNET.011.1900598.

Ma, S., Cao, Y., and Xiong, L. Transparent contribution evaluation for secure federated learning on blockchain. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, apr 2021. doi: 10.1109/icdew53142.2021.00023. URL https://doi.org/10.1109%2Ficdew53142.2021.00023.

Martinez, I., Francis, S., and Hafid, A. S. Record and reward federated learning contributions with blockchain. In *2019*

International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 50–57. doi: 10.1109/CyberC.2019.00018.

Nguyen, D. C., Pathirana, P. N., Ding, M., and Seneviratne, A. Blockchain for 5g and beyond networks: A state of the art survey. *Journal of Network and Computer Applications*, 166:102693, sep 2020. doi: 10.1016/j.jnca.2020.102693. URL https://doi.org/10.1016%2Fj.jnca.2020.102693.

Nguyen, L. D., Pandey, S. R., Beatriz, S., Broering, A., and Popovski, P. A marketplace for trading ai models based on blockchain and incentives for iot data, 2021. URL https://arxiv.org/abs/2112.02870.

Otoum, S., Al Ridhawi, I., and Mouftah, H. T. Blockchain-supported federated learning for trustworthy vehicular networks. In *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020. doi: 10.1109/GLOBECOM42002.2020.9322159.

Pandey, S. R., Nguyen, L. D., and Popovski, P. Fedtoken: Tokenized incentives for data contribution in federated learning, 2022. URL https://arxiv.org/abs/2209.09775.

Rahmadika, S., Astillo, P. V., Choudhary, G., Duguma, D. G., Sharma, V., and You, I. Blockchain-based privacy preservation scheme for misbehavior detection in lightweight iomt devices. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2022. doi: 10.1109/JBHI.2022.3187037.

Sharma, P. K., Park, J. H., and Cho, K. Blockchain and federated learning-based distributed computing defence framework for sustainable society. *Sustainable Cities and Society*, 59:102220, 2020. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2020.102220. URL https://www.sciencedirect.com/science/article/pii/S2210670720302079.

Shengwen Ding and Chenhui Hu. Survey on the Convergence of Machine Learning and Blockchain, 2022.

Wang, Q., Guo, Y., Wang, X., Ji, T., Yu, L., and Li, P. Ai at the Edge: Blockchain-Empowered Secure Multiparty Learning With Heterogeneous Models. *IEEE Internet of Things Journal*, 7(10):9600–9610, 10 2020. doi: 10.1109/jiot.2020.2987843. URL http://dx.doi.org/10.1109/JIOT.2020.2987843.

Wang, Y., Su, Z., Zhang, N., and Benslimane, A. Learning in the air: Secure federated learning for uav-assisted crowdsensing. *IEEE Transactions on Network Science and Engineering*, 8(2):1055–1069, 2021. doi: 10.1109/TNSE.2020.3014385.

Wang, Z. and Hu, Q. Blockchain-based federated learning: A comprehensive survey. URL http://arxiv.org/abs/2110.02182. version: 1.

Weng, J., Weng, J., Zhang, J., Li, M., Zhang, Y., and Luo, W. Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive. *IEEE Transactions on Dependable and Secure Computing*, 18(5):2438–2455, 2021. doi: 10.1109/TDSC.2019.2952332.

Xu, C., Qu, Y., Eklund, P., Xiang, Y., and Gao, L. *BAFL: An Efficient Blockchain-Based Asynchronous Federated Learning Framework*. Deakin University, January 2021. ISBN 9781665427449. URL https://dro.deakin.edu.au/articles/conference_contribution/BAFL_An_Efficient_Blockchain-Based_Asynchronous_Federated_Learning_Framework/20622720/1.

Xu, H., Wei, W., Qi, Y., and Qi, S. Blockchain-Based Crowdsourcing Makes Training Dataset of Machine Learning No Longer Be in Short Supply. *Wireless Communications and Mobile Computing*, 2022:1–13, jul 26 2022. doi: 10.1155/2022/7033626. URL http://dx.doi.org/10.1155/2022/7033626.

Yang, X., Tan, W., Peng, C., Xiang, S., Niu, K., and Ying, J. Federated learning incentive mechanism design via enhanced shapley value method. *Wirel. Commun. Mob. Comput.*, 2022, jan 2022. ISSN 1530-8669. doi: 10.1155/2022/9690657. URL https://doi.org/10.1155/2022/9690657.

Zhilin Wang, Qin Hu, Ruinian Li, Minghui Xu, and Zehui Xiong. Incentive Mechanism Design for Joint Resource Allocation in Blockchain-based Federated Learning, 2022.