[1]Dr Prasad Rayi,
2Rayudu Vinay Kumar,
3Matta Venkata Durga Pavan Kumar,
4Mamatha B

Privacy-Preserving Federated Learning for Iot Devices

Abstract:

Home appliance manufacturers seek customer input to enhance their goods and services for the development of a smart home system. We create a federated learning (FL) system including a reputation mechanism to aid home appliance makers in training a machine learning model using consumer data.

Subsequently, producers may anticipate consumers' needs and consumption patterns in the future. The operational process of the system has two phases: in the first phase, customers train the preliminary model supplied by the manufacturer using both the mobile device and the mobile edge computing (MEC) server. Customers gather data from several household appliances via smart phones, then downloading and training the basic model with their local data. Upon generating local models, clients endorse their models and transmit them to the blockchain. To mitigate any malice from consumers or manufacturers, we use blockchain technology to substitute the centralized aggregator in the conventional federated learning system. Due to the immutable nature of blockchain data, the behaviors of hostile consumers or producers are traceable. In the second step, manufacturers designate consumers or organizations as miners to compute the averaged model using the received models from customers. Upon completion of the crowdsourcing work, one miner, designated as the interim leader, uploads the model to the blockchain. To safeguard client privacy and enhance test accuracy, we implement differential privacy on the collected features and provide a novel normalizing approach. We experimentally show that our normalizing approach surpasses batch normalization when features are subjected to differential privacy protection. Furthermore, to entice additional consumers to engage in the crowdsourced FL work, we have devised an incentive system to reward participants.

Index Terms—Blockchain, Crowdsourcing, Differential privacy, Federated learning, IoT, Mobile edge computing.

Introduction

Internet of Things (IoT)-enabled smart home solutions have surged in popularity in recent years due to its objective of enhancing quality of life. A Statista analysis projects that the global smart home market size will reach $53.3 billion by 2022. The smart home idea is primarily facilitated by IoT devices, smartphones, contemporary wireless communications, cloud and edge computing, big data analytics, and artificial intelligence (AI). Specifically, these modern technologies facilitate manufacturers in sustaining an uninterrupted connection among their smart home devices. The proliferation of smart home gadgets generates substantial data. Federated learning (FL) allows analysts to examine and use locally created data in a decentralized manner, eliminating the need to upload data to

[1] [1,2,3,4]international School Of Technology And Sciences For Women, A.P, India.

a centralized server; hence, the utility of the data is effectively retained while being stored locally. We develop a federated learning-based solution to assist home appliance makers in efficiently and simply using data supplied by consumers' products. Our solution treats home appliances of the same brand inside a household as a single unit, using a mobile phone to periodically gather data from these appliances and locally train the machine learning model [2]. Due of the restricted processing capacity and battery longevity of mobile phones, we delegate a portion of the training job to the edge computing server. The blockchain smart contract is used to create a global model by averaging the sums of locally trained models given by users. In this federated manner, source data is expected to uphold security and privacy.

Melis et al. [3] proved that gradient changes may provide substantial information about consumers' training data. Malefactors may get information from gradients sent by clients [4]. Furthermore, the federated training strategy for the model is vulnerable to model poisoning assaults [5]. Moreover, there are dangers of information leakage associated with the third party's mobile edge computing (MEC) server [6]. To mitigate the above described security and privacy concerns, we use blockchain technology and differential privacy. Apple is effectively using differential privacy in federated learning to enhance the privacy of its widely used voice assistant, Siri [7]. Manufacturers submit a first model with configured parameters. The model is accessible on the blockchain for users to download and train using their local data. The blockchain enables the crowdsourcing requester (i.e., manufacturer) to verify the absence of harmful updates from consumers. The conventional crowdsourcing system is operated by a third party that imposes significant service fees on clients, but our proposed solution utilizes blockchain technology to document crowdsourcing operations. Consequently, both clients and the requester may save substantial service prices while maintaining the functionality of the crowdsourcing system. Owing to the constraints of block size, we suggest using the InterPlanetary File System (IPFS) [8] as

Related Work

Blockchain and federated learning (FL) methodologies have been extensively used in the training of neural networks using distributed data [15]–[24]. Weng et al. [21] developed a system named DeepChain for collaborative learning. However, they did not delegate the training work to the edge server, nor did they suggest using differential privacy to safeguard the confidentiality of model parameters. Awan et al. [20] suggested a blockchain-based architecture for privacy-preserving federated learning, which safeguarded model updates using the immutability and decentralized trust attributes of blockchain technology. Li et al. [25] developed a decentralized blockchain architecture for crowdsourcing jobs, facilitating the execution of such activities without reliance on a centralized server. Lu et al. [16] advocated using blockchain, federated learning, and differentiated privacy for data sharing.

They directly included differential privacy noise into the original data rather than the gradients, potentially compromising accuracy significantly. Lyu et al. [22] conducted the first study on federated fairness inside a blockchain-supported decentralized deep learning framework and developed a local credibility mutual assessment system to ensure fairness.

They also devised a protocol for encryption to guarantee confidentiality and precision. Furthermore, federated learning (FL) has garnered significant interest lately [26]–[30], with privacy protection emerging as a critical concern, as examined in [31]–[35].

Li et al. [31] addressed the privacy concerns associated with sharing model changes in federated learning. They

suggested using sketch algorithms to develop sketching-based federated learning, which ensures privacy while preserving accuracy. Hao et al. [32] suggested a privacy-enhanced federated learning strategy to address the privacy concerns in federated learning. Their framework enables effective and privacy-preserving federated learning. Dolui et al. [33] used federated learning paradigms in recommender systems and matrix factorization, ensuring the functioning and privacy of these systems.

Nasr et al. [34] conducted an extensive privacy study using white-box inference attacks. Wang et al. [35] presented a method that integrates a generative adversarial network with a multitask discriminator to address user-level privacy leaks in federated learning against threats from a hostile server.

Moreover, several research concentrate on privacy-preserving crowdsourcing and the use of fog or edge computing to enhance performance, given their increasing popularity. Wu et al. [36] provided two general methods for assessing the privacy of mobile users and the usefulness of data in crowdsourced location-based applications, respectively. He et al. [38] developed a privacy model for the crowdsourced bus service, which utilizes models that are exclusively relevant to the conventional crowdsourcing method (i.e., customers send data to a centralized server) and does not account for federated learning crowdsourcing tasks that utilize locally trained models. Zhao et al. [40] introduced a privacy-preserving technique to thwart poisoning attacks in mobile edge computing. Users must transfer data to the MEC server on their system, which may compromise privacy; thus, we advise that users maintain their data locally.

Moreover, some research have integrated deep learning or federated learning with edge computing [44]–[47]. Lyu et al. [44] introduced a privacy-preserving deep learning architecture that employs a dual-layer protection technique, including Random Projection and Differentially Private Stochastic Gradient Descent to safeguard data privacy.

Jiang et al. [45] developed a collaborative training approach to safeguard feature privacy. Feature extraction is performed locally on devices such as smartphones, while categorization occurs in the cloud service. Nevertheless, they failed to use federated learning to safeguard the confidentiality of training data, nor did they provide a normalization method to enhance test accuracy. Wang et al. [46,47] suggested control techniques to address the issue of limited resources in IoT devices during participation in federated learning.

The Proposed System Model

The suggested system model is shown in Figure 1. The shown system paradigm consists of three tiers: federated learning, an optimization algorithm, and a blockchain network. Security and privacy of IoT devices may be attained with the use of federated learning and blockchain technology. This study employs Paillier encryption for security and privacy, as opposed to differential privacy or anonymization methods, which may complicate training data and auditing processes. This research indicates that IoT devices possess constrained resources and cannot simultaneously transmit data and perform both local and global training. This study aims to address future restrictions by reducing the total computational cost and transmission latency of the proposed system using robust machine learning and game theory approaches. In the course of local and global model training, blockchain enables federated clients, namely IoT devices, to interchange model parameters via uploading and downloading. Furthermore, every facet of IoT device functionality is safeguarded against internal and external attacks. The IoT devices in this research are responsible for local task initiation and model initialization.

System Design

This section introduces the system designed for smart home appliance makers seeking to develop a machine learning model that utilizes data from customers' appliances to assess consumer habits and enhance their services and goods.

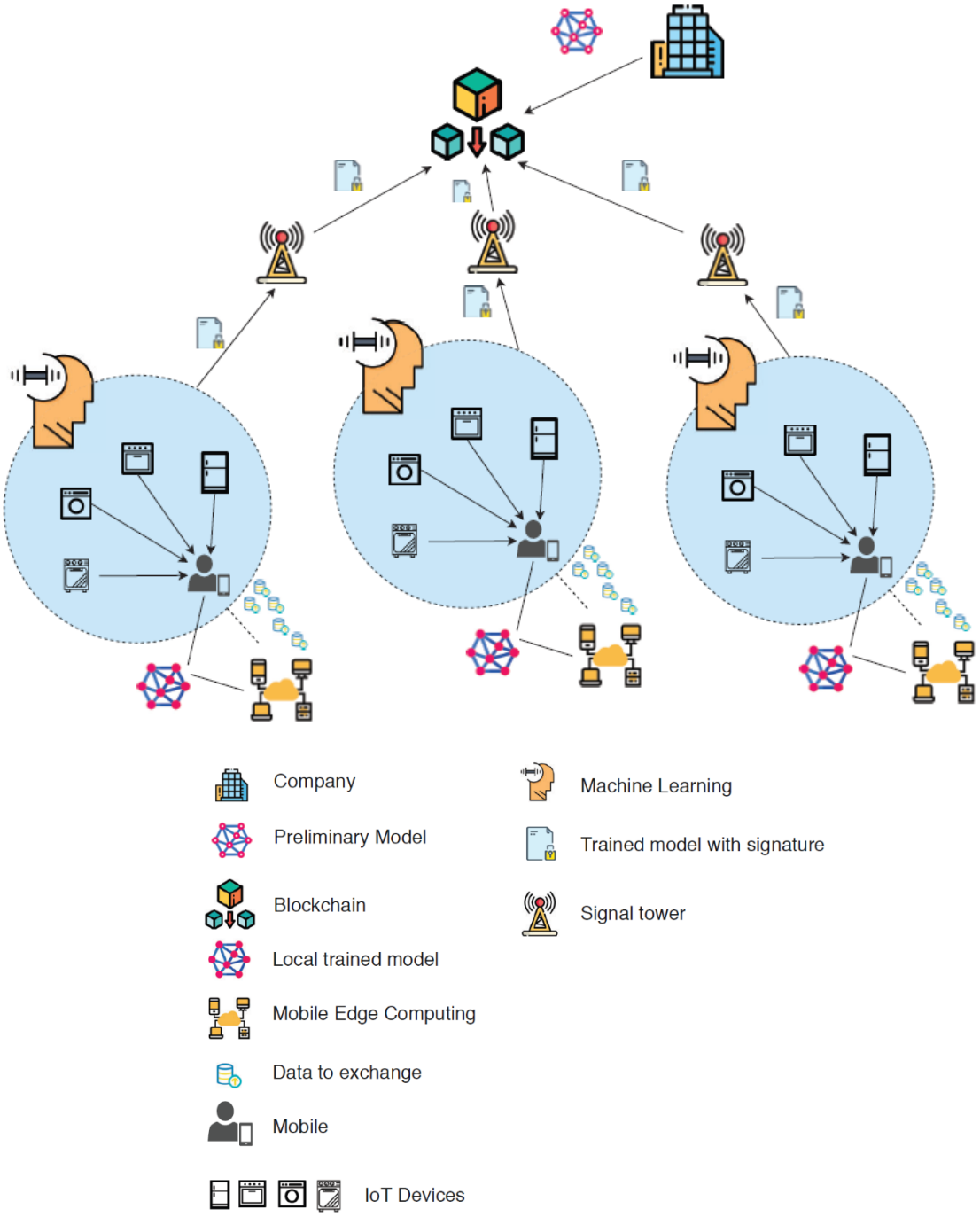

Figure 1: An overview of our system.

A. System Overview

Figure 1 presents a comprehensive overview of our system design. The system has three primary components: manufacturers, customers, and blockchain. Manufacturers submit a proposal for a crowdsourced federated

learning job. Subsequently, consumers interested in the crowdsourced federated learning projects upload their learned models to the blockchain. Ultimately, the blockchain functions as the centralized server for aggregating customer models, while a designated miner computes and produces the global federated learning model for home appliance makers. Subsequently, we shall explain each component in depth.

Producers. Manufacturers seek the development of a machine learning model to forecast client consumption behaviors and enhance home appliances, constituting a crowdsourced federated learning challenge. Individuals possessing household appliances are eligible to engage in the FL task. To further federated learning, we use blockchain technology to record the original model with randomly chosen parameters. Alternatively, producers must provide the model to all parties or store it in a third-party cloud service. Furthermore, both producers and consumers cannot refute documented contributions or actions. Ultimately, producers will develop a machine learning model as an increasing number of consumers engage in the crowdsourced federated learning activity.

Producers. Manufacturers seek the development of a machine learning model to forecast client consumption behaviors and enhance home appliances, constituting a crowdsourced federated learning assignment. Individuals possessing household appliances are eligible to engage in the FL task. To further federated learning, we use blockchain technology to record the baseline model with randomly chosen parameters. Alternatively, producers must provide the model to all parties or store it in a third-party cloud service. Furthermore, both producers and consumers cannot refute documented contributions or actions. Ultimately, producers will develop a machine learning model as an increasing number of consumers engage in the crowdsourced federated learning activity.

Clients. Customers with home appliances that meet crowdsourcing criteria may apply to participate in the FL task. Nonetheless, because to the different storage and processing capabilities of household appliances, it is challenging to facilitate the training of the deep model on each IoT device.

To resolve this problem, we use the partitioned deep model training methodology [45,48]. We use a cell phone to gather data from household equipment and extract attributes.

To safeguard privacy, we include differential privacy noise into the features. Subsequently, clients proceed to train the fully linked layers on the MEC server. We delineate the consumers' obligations in four distinct phases as follows.

Step 1: Customers get the preliminary model from the blockchain. Customers eager to engage in the FL work verify and download the first model submitted by the manufacturers and accessible on the blockchain.

Step 2: Customers derive features on the mobile device. The mobile phone regularly gathers data from all connecting home equipment. Subsequently, users may start the training of the model using the gathered data. Due to the MEC server being supplied by a third party, there is a potential for information leakage. Consequently, we partition the local training process into two phases: mobile training and MEC server training. To avoid compromising the model's accuracy by directly perturbing the original data, we use the convolutional neural network (CNN) layers as feature extractors to derive features from the original data on the mobile device. Subsequently, we include -DP noise into the features prior to transferring them to the fully linked layers inside the MEC.

Step 3: Clients train fully linked layers on the mobile edge computing server. The mobile device transmits the privacy-preserving characteristics and original labels to the mobile edge computing server, which assists in training the fully connected layers. The training loss is sent to the mobile device to update the front layers.

Step 4: Customers upload models into the blockchain. Upon completion of model training, clients authenticate

model hashes using their private keys and then upload the models to the blockchain using cellphones. However, if miners ascertain that the signature is incorrect, the transaction fails due to the potential for an attacker to compromise the learning process using fabricated data. Upon confirmation of the transaction by miners, clients may use the transaction history as an invoice to assert rewards and repute. Section IV-B delineates the specifics of reputation computation. The unchangeable characteristic of the blockchain ensures that both producers and buyers cannot refute transactions recorded on it.

Simulation Result

This section assesses the efficacy of the proposed EINFO optimization algorithm and contrasts its performance with that of the current INFO algorithm [54].

We define the population size p N as 30 and the maximum iteration y max as 500. The values of other parameters used in this investigation are shown in Table 2. The paper first examines Griewank's role as an optimization tool for evaluating the algorithms. Additionally, alternative optimization goal functions are taken into account in the assessments. The suggested system is constructed using Python 3.10, equipped with 8 GB of RAM and a CPU capacity of 1.60 GHz.

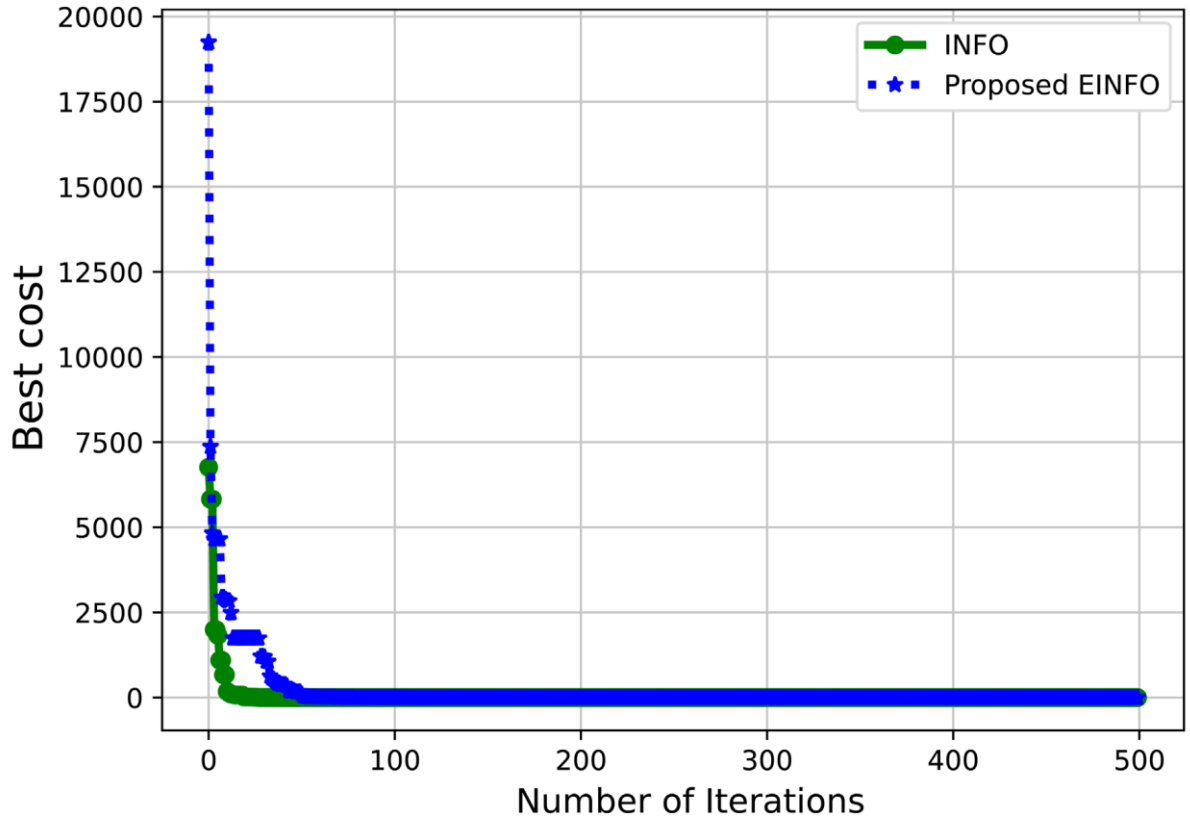

Figure 2. Evaluation based on convergence.

The convergence analysis depicted in Figure 2 demonstrates that the proposed EINFO circumvents premature convergence by effectively searching both locally and globally within the solution space, identifying solutions with a high density near the global optimum and a low density away from it. The suggested EINFO successfully identifies optimum solutions by analyzing areas in the search space that exhibit the most favorable costs.

Conclusion And Future Work

This study presents a concept for a blockchain-based crowdsourced federated learning system aimed at enabling IoT device makers to get deeper insights into user behavior. We use many cutting-edge technologies to develop the system, including mobile edge computing servers, blockchain, distributed storage, and federated learning. Furthermore, our system implements differential privacy to safeguard the confidentiality of consumer data. To enhance the accuracy of the federated learning model, we have developed a novel normalizing approach that has been shown to surpass batch normalization when the privacy of features is safeguarded by differential privacy. Implementing an effective reward structure for the crowdsourcing job increases the likelihood of consumer participation. The blockchain will verify all consumer revisions throughout the federated training to ensure that model updates are responsible, hence deterring malevolent customers or manufacturers.
In the future, we want to do further experiments and evaluate our system using authentic household appliance datasets.

Furthermore, we will endeavor to identify the deterministically optimum equilibrium between local epochs and global epochs to enhance test accuracy.

References

[1] S. R. Department, “Smart home - Statistics & Facts,” 2020. [Online]. Available: https://www.statista.com/topics/2430/smart-homes/

[2] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, “When edge meets learning: Adaptive control for resourceconstrained distributed machine learning,” in IEEE Conference on Computer Communications (INFOCOM), 2018, pp. 63–71.

[3] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, “Exploiting unintended feature leakage in collaborative learning,” in IEEE Symposium on Security and Privacy (S&P), 2019.

[4] B. Hitaj, G. Ateniese, and F. Perez-Cruz, “Deep models under the GAN:information leakage from collaborative deep learning,” in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 603–618.

[5] C. Fung, C. J. Yoon, and I. Beschastnikh, “Mitigating sybils in federated learning poisoning,” arXiv preprint arXiv:1808.04866, 2018.

[6] Y. Zhang, T. Gu, and X. Zhang, “Mdldroid: a chainsgd-reduce approach to mobile deep learning for personal mobile sensing,” in 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, 2020, pp. 73–84.

[7] K. Hao, “How Apple personalizes Siri without hoovering up your data,” 2019. [Online]. Available: https://www.technologyreview.com/2019/12/ 11/131629/apple-ai-personalizes-siri-federated-learning/

[8] J. Benet, “IPFS-content addressed, versioned, P2P file system,” arXiv preprint arXiv:1407.3561, 2014.

[9] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in Theory of Cryptography Conference (TCC), 2006, pp. 265–284.

[10] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, “Our data, ourselves: Privacy via distributed noise generation,” in International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2006, pp. 486–503.

[11] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, “Privacy loss in Apple’s implementation of differential privacy on macOS 10.12,” arXiv preprint arXiv:1709.02753, 2017.

[12] U´ . Erlingsson, V. Pihur, and A. Korolova, “RAPPOR: Randomized aggregatable privacy-preserving ordinal response,” in ACM Conference on Computer and Communications Security (CCS), 2014, pp. 1054– 1067.

[13] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Artificial Intelligence and Statistics, 2017, pp. 1273–1282.

[14] J. Koneˇcn`y, H. B. McMahan, F. X. Yu, P. Richt´arik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” in NIPS Workshop on Private Multi-Party Machine Learning, 2016.

[15] X. Qu, S. Wang, Q. Hu, and X. Cheng, “Proof of federated learning: A novel energy-recycling consensus algorithm,” arXiv preprint arXiv:1912.11745, 2019.

[16] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, “Blockchain and federated learning for privacy-preserved data sharing in industrial IoT,” IEEE Transactions on Industrial Informatics, 2019.

[17] P. Ramanan, K. Nakayama, and R. Sharma, “Baffle: Blockchain based aggregator free federated learning,” arXiv preprint arXiv:1909.07452, 2019.