Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH Open Access

BGFL: a blockchain-enabled group federated learning at wireless industrial edges

Guozheng Peng[1], Xiaoyun Shi[2*], Jun Zhang[1], Lisha Gao[3], Yuanpeng Tan[1], Nan Xiang[3] and Wanguo Wang[4]

Abstract

In the rapidly evolving landscape of Industry 4.0, the complex computational tasks and the associated massive data volumes present substantial opportunities for advancements in machine learning at industry edges. Federated learning (FL), which is a variant of distributed machine learning for edge-cloud computing, presents itself as a persuasive resolution for these industrial edges, with its main objectives being the mitigation of privacy breaches and the resolution of data privacy concerns. However, traditional FL methodologies encounter difficulties in effectively overseeing extensive undertakings in Industry 4.0 as a result of challenges including wireless communications with high latency, substantial heterogeneity, and insufficient security protocols. As a consequence of these obstacles, blockchain technology has garnered acclaim for its secure, decentralized, and transparent data storage functionalities. A novel blockchain-enabled group federated learning (BGFL) framework designed specifically for wireless industrial edges is presented in this paper. By strategically dividing industrial devices into multiple groups, the BGFL framework simultaneously reduces the wireless traffic loads required for convergence and improves the accuracy of collaborative learning. Moreover, to optimize aggregation procedures and reduce communication resource utilization, the BGFL employs a hierarchical aggregation strategy that consists of both local and global aggregation off-chain and on-chain, respectively. The integration of a smart contract mechanism serves to fortify the security framework. The results of comparative experimental analyses demonstrate that the BGFL framework enhances the resilience of the learning framework and effectively reduces wireless communication latency. Thus, it offers a scalable and efficient solution for offloading tasks in edge-cloud computing environments.

Keywords Federated learning, Blockchain, Edge-cloud cooperation, Wireless traffic

Introduction

Industry 4.0, which is also referred to as the fourth industrial revolution, is propelling revolutionary progress in intelligent manufacturing, virtual reality, and smart healthcare, among others [1–3]. The advent of this revolution gives rise to large data volumes and computationally demanding tasks, thereby creating an environment that facilitates the convergence of machine learning and the widespread adoption of distributed industrial edges [4–6]. In order to optimize data storage and processing, wireless industrial edges integrate with cloud computing and utilize edge computing in industrial environments. The hybrid edge-cloud model facilitates efficient data exchange and rapid connectivity between industrial machinery and computing nodes, thereby improving the ability to analyze data in real-time and adapt to the ever-changing requirements of Industry 4.0 applications.

Conventional machine learning approaches in such contexts frequently necessitate the manipulation of data across decentralized networks, potentially relying on external cloud servers for AI training purposes, which entails the handling of confidential industrial

*Correspondence:
Xiaoyun Shi
capyun007@tju.edu.cn
[1] China Electric Power Research Institute, Beijing 100192, China
[2] College of Intelligence and Computing, Tianjin University, Tianjin 300354, China
[3] State Grid Nanjing Power Supply Company, Nanjing 210019, China
[4] State Grid Intelligence Technology Co., Ltd., Jinan 250013, China

data [3, 7, 8]. This practice gives rise to substantial privacy and security concerns, underscoring the need for novel approaches that safeguard data confidentiality while capitalizing on the computational capabilities of cloud servers. By executing data processing locally at the edge, edge computing aids in the reduction of latency and bandwidth concerns [9]. In contrast, cloud computing provides scalable resources to support more intricate computations, thus striking a balance between security and efficiency.

In this edge-cloud cooperative framework, federated learning (FL) emerges as a viable strategy for addressing critical concerns regarding the security and privacy of data in industrial applications [10]. FL permits clients, or distributed devices, to train a model collaboratively while maintaining data on-premises [11–13]. The process of aggregating model updates via a central cloud server enables the implementation of a cooperative computation model that ensures data privacy remains uncompromised among various industrial participants.

Nevertheless, the implementation of FL in a hybrid edge-cloud setting encounters numerous obstacles, such as the management of conflicting network conditions, data distributions, and the diverse capabilities of FL clients [14–17]. These necessitate complex coordination mechanisms in order to facilitate effective learning. Additionally, robust security measures are required to safeguard data during transmission between edge devices and cloud servers [18–21].

Additionally, careful optimization is required due to the inherent trade-offs between communication efficiency and model accuracy in FL, especially in scenarios with constrained bandwidth. In addition to being adaptable to changing conditions, the dynamic nature of industrial environments necessitates that FL algorithms incorporate real-time feedback to preserve their efficacy.

- **Wireless communication latency**: In the context of the FL process, each training iteration mandates that participants transmit their local model parameters to a central server, which in turn disseminates an updated global model to all clients. This necessary bidirectional data flow introduces substantial wireless communication load and associated latency within the network [22–24]. To mitigate these issues, leveraging edge computing can be highly effective. By offloading the data processing tasks to edge servers located closer to the user terminals, latency is significantly reduced compared to processing on remote cloud servers. This proximity advantage of edge servers is crucial for real-time applications in IoT environments, where timely data processing is paramount.
- **Heterogeneity**: In the context of edge-cloud computing cooperation, a multitude of distributed devices is strategically positioned across diverse locations. Adding to the complexity, these distributed devices exhibit significant heterogeneity in terms of both computing power and data properties [25–27]. The amalgamation of local models, trained on such heterogeneous devices, into a singular overarching model presents a formidable challenge, potentially impeding the convergence of the learning process. The varied computational capabilities and distinct data characteristics inherent in these distributed devices introduce a layer of intricacy, emphasizing the need for nuanced approaches to effectively navigate and address the challenges associated with aggregating and consolidating diverse models within edge-cloud computing cooperation.
- **Security**: Significant security concerns exist in federated learning, including risks of malicious clients tampering with models and susceptibilities to single points of failure. The implementation of edge computing can bolster security measures through the local pre-processing and verification of data prior to its transmission to the cloud for additional aggregation. By early filtration of potentially malicious inputs, this local processing not only reduces the risk of transmitting sensitive data but also provides an additional layer of security. By combining edge and cloud computing in a collaborative framework, federated learning is implemented in an environment that is more secure and resilient, thereby safeguarding the integrity of the learning process across decentralized networks.

In addressing these challenges, ongoing research endeavors aim to enhance the robustness, security, and efficiency of federated learning in the context of industrial applications, thereby unlocking its full potential for fostering intelligent and privacy-conscious edge-cloud computing cooperation ecosystems.

Following recent triumphs, blockchain has emerged as a decentralized, open, and transparent data storage technology, offering a potential remedy for the challenges encountered in federated learning at wireless industrial edges [28–30]. Notably, some researchers have explored diverse blockchained federated learning (BFL) frameworks [31, 32]. These BFL architectures typically leverage blockchain to facilitate the exchange and verification of local model updates among devices, circumventing the single point of failure problem associated with central servers. Additionally, hybrid blockchain frameworks [33, 34], incorporating elements like consortium blockchain and direct acyclic graph (DAG), have been devised to enhance blockchain throughput and reduce communication complexity, particularly in the context of large-scale federated learning tasks [35]. However, these existing frameworks face limitations when applied to Industry 4.0 scenarios.

On one hand, the practice of storing all model parameters from industrial devices on a centralized blockchain may burden the system and lead to heavy wireless communication demands. On the other hand, maintaining multiple blockchains concurrently could undermine efficient model sharing.

In a departure from existing approaches, our proposed solution introduces a novel blockchain-enabled group federated learning (GFL) approach tailored for edge-cloud computing cooperation. The framework features a primary consortium blockchain, with various wireless industrial edges partitioned into distinct groups. These groups independently execute the consensus process without interference, while leveraging the InterPlanetary File System (IPFS) and blockchain for model parameter sharing. This collaborative learning framework, therefore, enables real-time cooperative services in distributed industrial scenarios. For instance, it facilitates quality inspection of identical equipment types across different factories or troubleshooting of charging piles in diverse regions. By doing so, it addresses issues such as high communication latency and security risks arising from data and geographical isolation. The primary contributions of this article can be succinctly summarized as follows.

- Our proposition introduces a groundbreaking blockchain-enabled group federated learning framework strategically deployed at wireless industrial edges. The distinctive feature of the BGFL framework lies in its innovative approach of segmenting wireless industrial edges into well-defined groups. This segmentation serves a dual purpose: firstly, it endeavors to elevate the precision of collaborative learning initiatives, and secondly, it addresses the critical issue of latency inherent in wireless communication, a pivotal factor for achieving efficient convergence in federated learning processes. By structuring wireless industrial edges into cohesive groups, BGFL not only enhances the accuracy of collaborative learning endeavors but also ensures a streamlined approach to wireless communication, thereby contributing to the overall effectiveness and convergence of the federated learning framework in the dynamic landscape of wireless industrial environments.
- Within the BGFL framework, we embrace a hierarchical aggregation approach that seamlessly integrates both local aggregation off-chain and global aggregation on-chain mechanisms. This strategic incorporation is meticulously designed to alleviate the inherent complexity associated with the aggregation process, thereby serving as a catalyst for the improved learning efficiency of the diverse industrial devices interconnected within the network. The hierarchical nature of the aggregation methodology not only optimizes the coordination of local and global aggregation processes but also strategically positions itself as a means to streamline the complexities involved, ensuring a more efficient and effective federated learning experience across the spectrum of industrial devices within the network.
- The integration of smart contracts assumes a crucial role within the BGFL framework, serving as a linchpin for updating the global aggregation group and consolidating local model parameters, thereby fortifying the security of the entire process. Empirical evidence gleaned from rigorous experimentation substantiates that BGFL not only propels learning performance to new heights but also reinforces the overall robustness of the learning framework. The judicious use of smart contracts not only contributes to the accuracy and security of the federated learning process but also attests to the practical efficacy of BGFL in real-world scenarios, affirming its position as an advanced and secure framework for collaborative learning in wireless industrial environments.

The subsequent sections of this article are structured as follows: "Blockchain-enabled collaborative learning framework at wireless industrial edges" section provides a comprehensive overview of the proposed BGFL framework. In "Collaborative learning procedure" section, we delineate the detailed procedural aspects of BGFL for computing cooperation. Subsequent to this, "Simulation results and discussions" section presents experimental evaluations of our proposed scheme. The article is concluded in "Conclusion and future works" section.

Blockchain-enabled collaborative learning framework at wireless industrial edges

We first describe the BGFL framework, followed by its key components, which is as shown in Fig. 1.

Framework

We illustrate the framework and use cases of BGFL at wireless industrial edges. In the BGFL framework, a group of wireless industrial edges is selected to participate in FL. Specifically, BGFL consists of one main consortium blockchain formed by multiple groups. Here, there are two types of groups in the blockchain: local training group and global aggregation group. For the

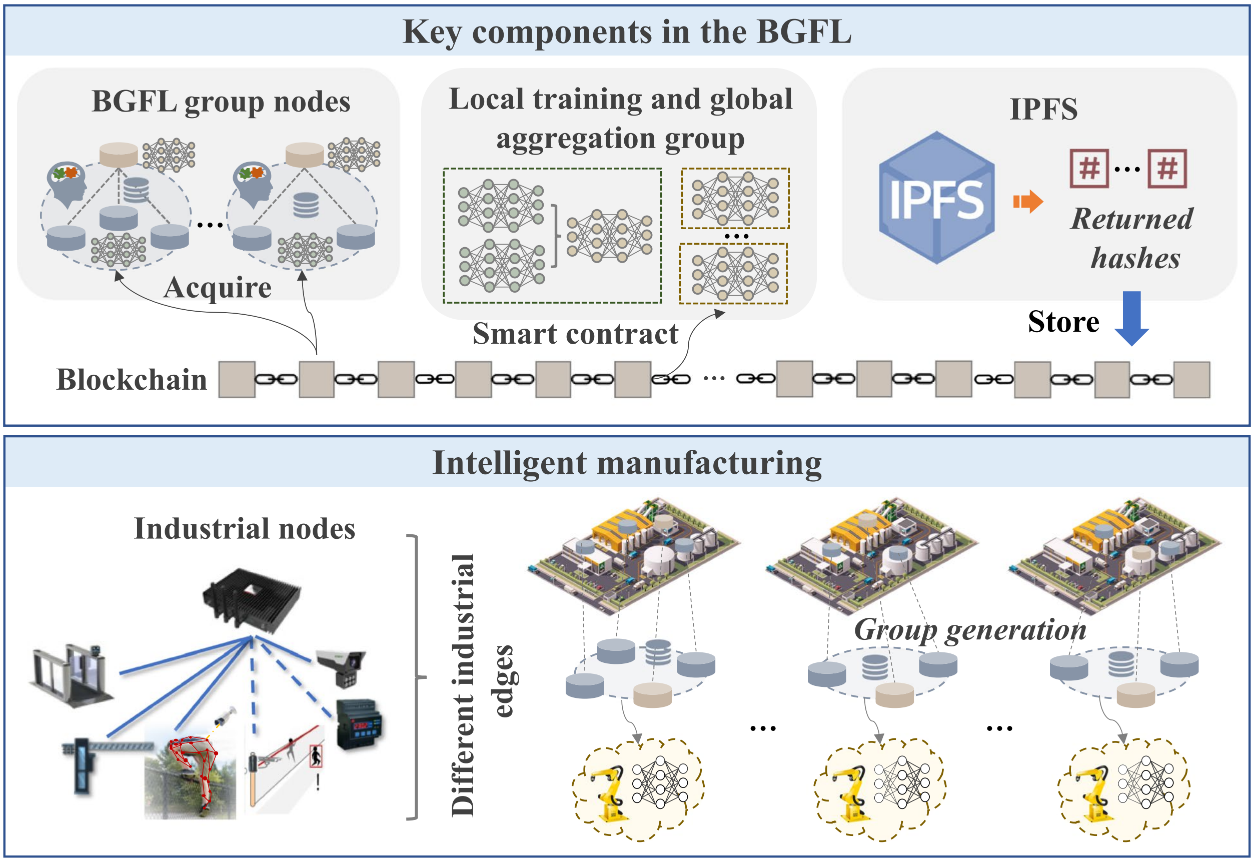
Key components in the BGFL
BGFL group nodes
Acquire
Local training and global aggregation group
Smart contract
IPFS
IPFS
#
#
Returned hashes
Store
Blockchain
Intelligent manufacturing
Industrial nodes
Different industrial edges
Group generation

Fig. 1 Blockchain-enabled group federated learning approach at wireless industrial edges

local training group, a set of edge nodes would be partitioned into multiple independent groups to train AI tasks locally. Besides, the global aggregation group is composed of the group leaders from the local training group to validate transactions and aggregate local parameters in a decentralized manner. Moreover, the IPFS is leveraged to store the model parameters to mitigate the storage load of the blockchain. The detailed design components of the BGFL are illustrated as follows.

Key components

Then we illustrate the detailed design components of the BGFL, including blockchain, BGFL group nodes, local training, and global aggregation group, and IPFS as follows.

Blockchain

Within the BGFL framework, we have established a consortium blockchain dedicated to recording and overseeing the credit and learning parameters associated with edge nodes. This consortium blockchain incorporates group partitioning to realize data isolation and confidentiality within the same chain, utilizing the cluster dimension. This innovative approach allows for the extension of the blockchain architecture from a single-chain single-ledger to a single-chain multi-ledger paradigm. Essentially, this means that the multitude of blockchain nodes can be categorized into distinct groups, each catering to diverse services. Consequently, BGFL not only enhances the efficiency of collaborative learning among heterogeneous nodes but also ensures heightened protection of individual privacy.

- Blockchain Group: The Blockchain Group in the BGFL framework exploits group partitioning to categorize blockchain nodes into multiple groups, thereby catering to diverse requirements. Each group maintains its own ledger for transaction and data management, conducting the consensus process autonomously, independent of other group's consensus mechanisms. This strategic partitioning ensures that groups can operate independently, enhancing privacy protection. Notably, a blockchain node has the capability to concurrently participate in multiple groups, engaging in diverse multi-party collaboration services. In the context of BGFL, the blockchain nodes comprising the global aggregation group also function as nodes within the local training group. Specifically, the global aggregation group is composed of leaders, denoting blockchain nodes with the highest credit, from each local training group.

- Smart Contracts: In order to guarantee the security of the BGFL framework, we have devised several smart contracts, namely the task release contract and the global aggregation contract. The task release contract encompasses crucial elements such as the testing dataset, initial global model parameters, and evaluation metrics. Notably, the accuracy performance serves as an evaluation metric, assessing the training quality of the BGFL model, and subsequently, is employed for calculating the credit of the group nodes. Additionally, the global aggregation contract plays a pivotal role in verifying and aggregating the parameters of multiple models from distinct training groups, monitoring the local training group in the process. This intricate system of smart contracts ensures the robustness and integrity of the BGFL framework.

BGFL group nodes

Assume that there are N edge nodes. All candidate edge nodes are denoted as $Z = \{n_1, n_2, \cdots, n_N\}$. In Fig. 2, each node n_k executes model training with their local dataset D_k, where computation resource for local model training such as CPU cycle frequency from the worker n_k is f_k. These edge nodes will be selected to the BGFL training group based on their contribution, which is related to the contribution of data size and the computing power of the edge node. Specifically, the contribution of edge nodes n_k is described as follows:

$$Contribution_k = log \frac{D_k}{\sum_{k=1}^{N} D_k} + f_k. \tag{1}$$

In the Industry 4.0 landscape, the computational and wireless communication capabilities of devices within different groups vary significantly due to diverse hardware and physical factors. Compounding this variability, the grouped devices are often geographically dispersed, leading to the creation of heterogeneous datasets with distinct characteristics. The amalgamation of such diverse devices into a single model can markedly impact the convergence of collaborative learning initiatives.

To address these challenges, the Blockchain-enabled Group Federated Learning proposed in this paper introduces the use of partition clustering methods, aiming to minimize the distance between devices within each group. The objective function is as follows:

$$\min \sum_{G_k \in \mathcal{G}} \sum_{i,j \in s_k} d_{ij} \quad s.t. \bigcup_{G_k \in \mathcal{G}} G_k = Z, \ \bigcap_{G_k \in \mathcal{G}} G_k = \emptyset, \tag{2}$$

where d_{ij} is the distance between edge nodes n_i and n_j, and G_k and $\mathcal{G}$ are the sets of the k-th group and all groups, respectively.

To achieve the aforementioned objective, this paper employs algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and K-means. These methods play a pivotal role in categorizing the heterogeneous group devices into distinct training

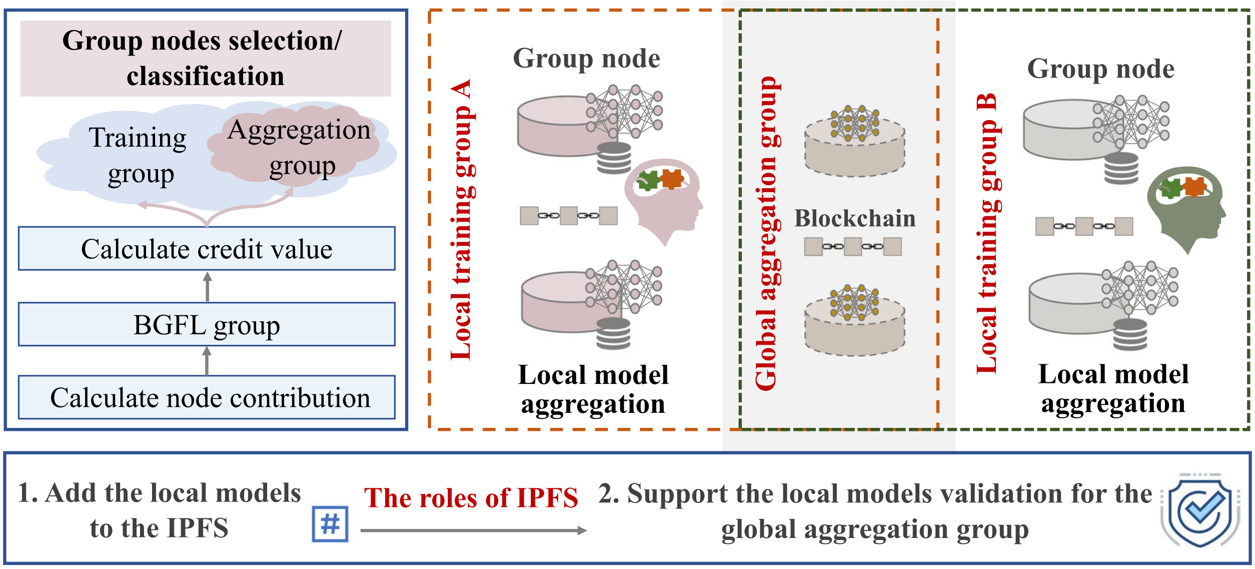

Fig. 2 The group partition in the BGFL

groups post server node selection [36]. Subsequently, the determination of the group leader for each training group is based on their respective credit. The credit of each edge node is computed considering its contribution and training performance. The appointed group leaders of all training groups collectively form a global aggregation group responsible for aggregating the received local updates from the individual training groups.

Local training and global aggregation group

Within the BGFL framework, two types of groups coexist: local training groups and global aggregation groups. The intricate relationship process between these groups is elucidated in Fig. 3. This delineation serves as a foundational framework for the subsequent stages of collaborative learning, encapsulating the dynamics of local training within distinct groups and the subsequent global aggregation of their contributions. This approach not only accommodates the diverse capabilities and geographical dispersion of devices in Industry 4.0 scenarios but also ensures a cohesive and efficient federated learning process.

In detail, the local training group, comprising edge nodes with analogous contributions, takes on the responsibility of orchestrating the collaborative efforts of devices within the same group to accomplish the training task. Given the similarity in performance among group nodes, there is minimal disparity in the completion times of local model training. Consequently, in the BGFL framework, we adopt a synchronous approach to local training. The edge nodes within each training group initially retrieve the initial model parameters from the main blockchain via a smart contract and subsequently engage in local training activities.

Let's assume there are m edge nodes selected to participate in the training group G, denoted as the node set $\{d_1, d_2, \cdots, d_m\}$ with corresponding datasets $\{n_1, n_2, \cdots, n_m\}$. The objective of edge node j is to minimize the expectation of the loss function for mini-batches sampled from D_j. Each group conducts the aggregation

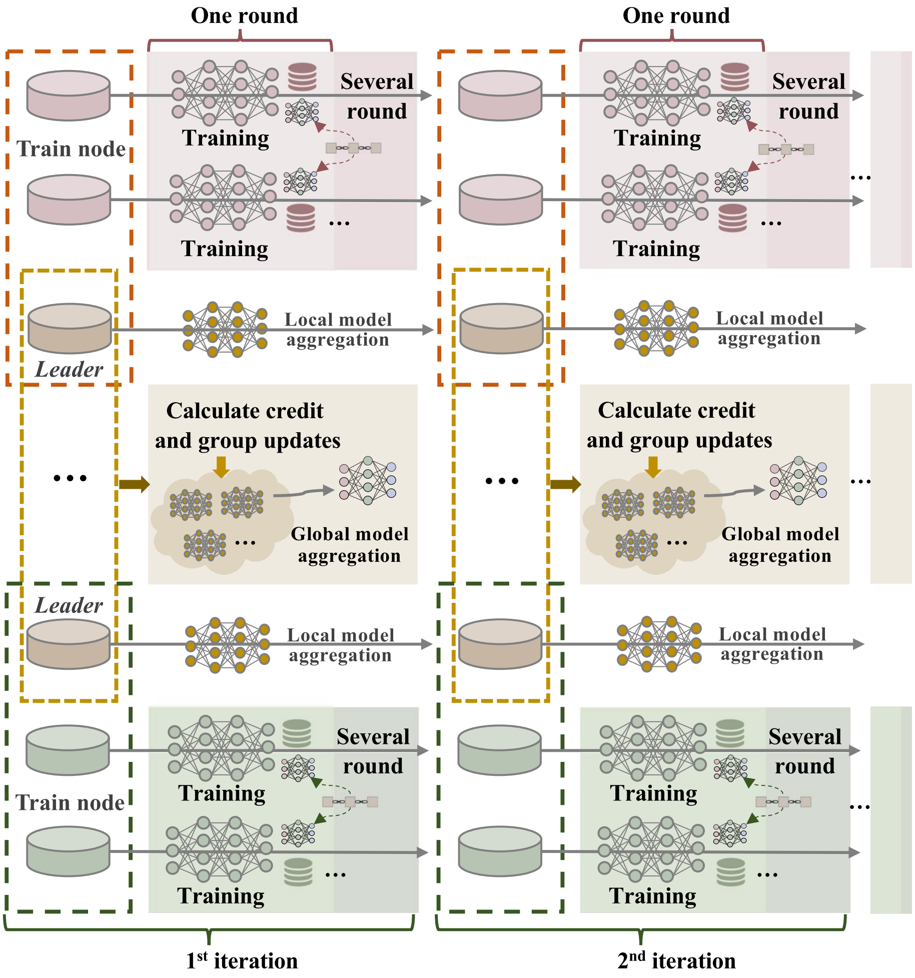

Fig. 3 The specific relationship process between local training and global aggregation group

of local model parameters from individual edge nodes after several rounds of local training. The aggregation weight assigned to each node is determined by its credit, reflecting both its contribution and training performance. Upon the completion of a local training iteration, the resultant model parameters are uploaded to the IPFS. Simultaneously, the returned model hash is stored in the main blockchain, ensuring traceability and accountability in the federated learning process. This synchronous and credit-based approach not only optimizes the training coordination within the local group but also enhances the overall efficiency and reliability of the collaborative learning framework in BGFL.

Furthermore, the global aggregation group coalesces leaders from various training groups, characterized by robust computational capabilities and high credit scores. This group plays a pivotal role in orchestrating the collaboration among training group devices to aggregate intermediate parameters effectively. Recognizing the inherent heterogeneity among different groups, we opt for an asynchronous global update strategy to enhance the model's generalization performance while maintaining a certain level of test accuracy.

In this asynchronous approach, we retrieve all local model parameters based on the model hash, initiating the global aggregation process through a smart contract. This triggers the amalgamation of contributions from diverse training groups, and a new model is generated as the culmination of this collaborative effort. Subsequently, this newly aggregated model is securely placed on the main blockchain, ensuring transparency, accountability, and accessibility for all participating nodes. This methodology not only accommodates the diversity and varying capabilities of distinct groups but also facilitates a dynamic and efficient evolution of the federated learning model in the BGFL framework.

IPFS

IPFS, recognized as a distributed storage system and a foundational element of Web 3.0, serves as the backbone for decentralized file sharing. It operates by storing a unique hash of files, making them universally accessible to all network nodes while efficiently eliminating redundancy in the wireless network. Notably, IPFS employs a content-addressed block storage model, ensuring the security of stored transactions and contributing to a high-throughput storage mechanism. This dual functionality makes IPFS a robust choice for facilitating secure and decentralized data storage.

In the context of the proposed federated learning framework, leveraging IPFS proves advantageous in multiple aspects. By uploading the model parameters to IPFS and storing the corresponding hashes in the blockchain, a significant reduction in storage load is achieved without compromising on security. This dual integration optimally utilizes the strengths of both technologies: IPFS streamlines data storage and accessibility, while the blockchain ensures the immutability and transparency of the stored information. Consequently, the use of IPFS in storing model parameters represents a substantial advancement, leading to a substantial decrease in the storage space requirements of the blockchain ledger. This innovative approach not only enhances the efficiency of data storage but also contributes to the overall scalability and sustainability of the proposed BGFL framework.

Design advantages of BGFL

Through the above design principles, the main advantages of the BGFL framework can be summarized as follows.

- Efficiency: Generally, the heterogeneity of edge nodes in Industry 4.0 could result in significantly inefficient collaborative learning performance. The proposed BGFL divides a multitude of similar edge nodes into a group to lighten the wireless communication latency, while combining IPFS to reduce the storage load. Apart from communication and storage, the BGFL can improve model generalization, while reducing the aggregation complexity by aggregating a grouping of similar local client models, thus providing an efficient collaborative learning mechanism.
- Scalability: The bottleneck of blockchain in the edge scenario is the scalability due to the complicated consensus and the proliferation of transactions. The BGFL introduces the multiple groups to execute the learning training task collaboratively Each group maintains its own ledger and reaches the consensus independently. As such, the framework can address the problems of limited scalability, further supporting large-scale collaborative learning tasks.
- Reliability: Traditional collaborative learning relies on the central aggregator to generate a new global model for the next iteration, which is prone to the problem of a single point of failure and targeted attacks. The proposed BGFL enables edge devices to collaboratively train models in a decentralized manner, while storing credits of the edge devices by blockchain, thus guaranteeing the reliability of data and models.

Use case: quality inspection in intelligent manufacturing

Let's delve into a typical use case within the context of BGFL. In the realm of advanced smart manufacturing factories aligned with Industry 4.0 principles, the pervasive integration of artificial intelligence technologies plays a pivotal role in augmenting production efficiency, curbing labor costs, and averting potential disasters. Take, for instance, the conventional challenge of manually sorting defective products, which is both inefficient and costly in traditional manufacturing processes. The integration of AI-based intelligent recognition technology presents a transformative solution, significantly expediting the sorting process and improving overall efficiency.

Furthermore, AI-based image recognition applications extend to critical areas such as fire detection and smoke detection, offering a broader range of detection capabilities compared to traditional temperature sensors. This expanded scope enhances early warning systems, contributing to a safer and more secure manufacturing environment.

In the industrial manufacturing landscape, the demand for low latency is paramount. Efficient decision-making on production lines is crucial to maintaining optimal production efficiency, and any delay in communication can lead to significant setbacks. Given the intolerability of prolonged communication delays in industrial manufacturing, the adoption of BGFL becomes pivotal.

Moreover, the inherent heterogeneity of edge-embedded devices in industrial manufacturing settings, driven by diverse production tasks, introduces variations in the quantity and quality of data available for model training. This diversity underscores the need for flexible and adaptive federated learning frameworks that can accommodate and leverage the varied data sources effectively.

System security stands out as a top priority in industrial manufacturing. The emphasis on robust security systems is crucial to prevent economic losses resulting from potential system failures. The integration of BGFL not only addresses the challenges posed by communication latency and heterogeneous data but also aligns with the stringent security requirements of industrial manufacturing scenarios, thus establishing a comprehensive and efficient solution for intelligent and secure Industry 4.0 applications.

In advanced smart manufacturing factories under Industry 4.0, the edge inference devices are generally isolated by tens of kilometers or more, with a large number of edge test nodes. Apparently, the model performance based on collaborative learning between these heterogeneous edge nodes is inefficient due to the high communication cost. Specifically, the intelligent manufacturing with BGFL works as follows:

- Determining whether an edge device can participate in the BGFL by calculating its contribution.
- The edge nodes in all the quality inspection agencies will be divided into several small groups.
- All the training groups will train their local quality inspection models and aggregate the model parameters of local group nodes after several synchronous rounds. Then the global aggregation group aggregates the local inspection models asynchronously and obtains the model updates for the novel local model training.
- The quality inspection agencies will determine whether there are quality problems with the equipment based on the collaborative learning results.

Collaborative learning procedure

To perform AI tasks in a decentralized manner, we develop a novel collaborative learning framework: BGFL. More details about the operational procedure of BGFL are given as follows, which is also as shown in Fig. 4.

Task publishment

In the operational flow of the proposed BGFL framework, the initiation of a collaborative learning task is facilitated through the execution of a smart contract, duly signed by the task publisher. This smart contract encapsulates essential task information, including the initial model parameters and the termination condition for the task.

Simultaneously, each smart contract encompasses additional task specifications, denoted as F (required computing power), C (acceptable communication latency), and R (credit rating). These parameters play a pivotal role in the node selection process. Specifically, during the participant selection phase, the combined computing power of the chosen nodes must surpass the task's computing power requirement. Additionally, the credit rating and communication latency of each participant must meet or exceed the criteria stipulated by the task contract.

The initiation of the collaborative learning task commences with the creation of the genesis transaction by the smart contract. Subsequently, the 0-th round of training in every local training group is triggered by this smart contract, employing the initial model parameters denoted as P_0. As the iterative training progresses through the k-th iteration, the model parameters undergo updates, transitioning to P_k and reflecting the cumulative learning insights gained during the collaborative process. This iterative and contract-based approach ensures a dynamic and adaptive learning process within the BGFL framework, driven by the collaborative efforts of diverse and qualified edge nodes meeting the specified task criteria.

Group nodes selection and division

The edge nodes that would like to join the task will send a join request with the node information (i.e., identity and contribution information) to the task publisher. Then the edge nodes' information will be validated by the task

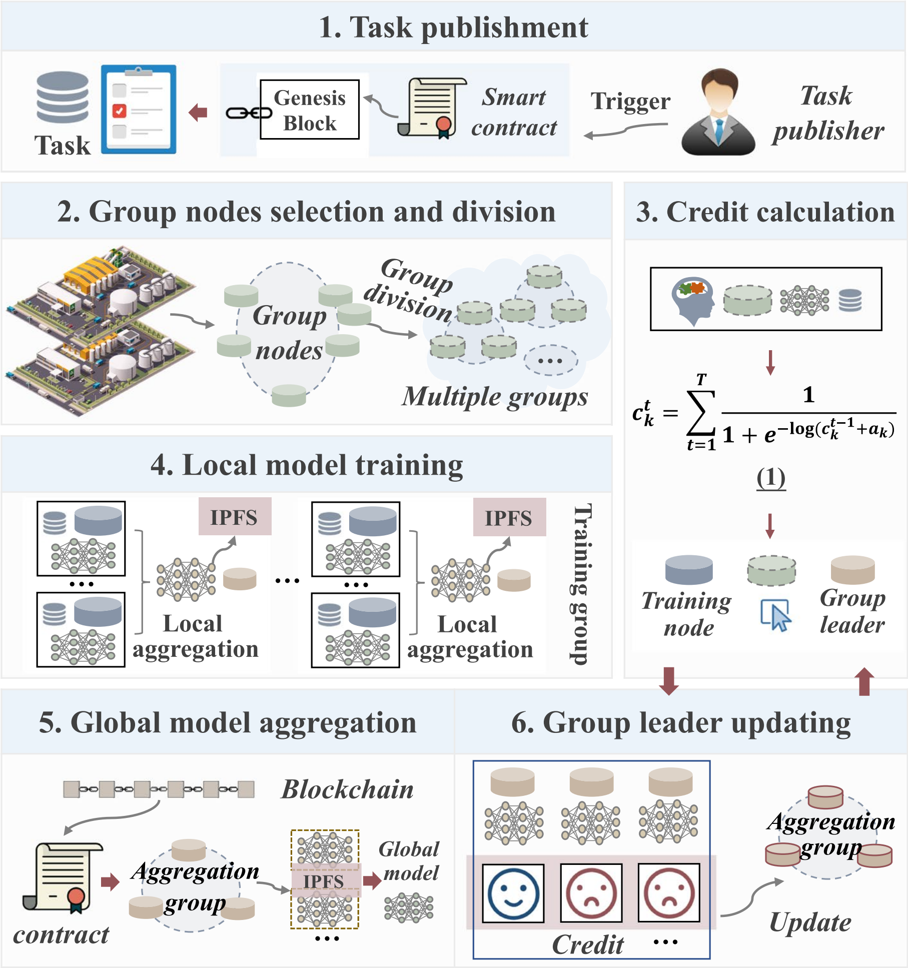

Fig. 4 The operational procedure of BGFL

publisher, and the edge nodes with high contribution can be allowed to join the BGFL framework. We define the set of all selected edge nodes as N, and we must satisfy:

$$\sum_{k \in N} f_k \geq F. \tag{3}$$

At the same time, for each node k in the set of edge nodes N, we must also ensure:

$$credit_k \geq R. \tag{4}$$

$$latency_k \geq C. \tag{5}$$

The process of node selection within the BGFL framework involves the subdivision of chosen nodes into multiple training groups using sophisticated partitioning clustering methods. Within each training group, the credit of every edge node is meticulously calculated, with the node boasting the highest credit bestowed with the role of a group leader responsible for aggregating local models. This strategic approach ensures that every training group benefits from the participation of a designated node in the global aggregation process, substantially escalating the cost of potential attacks and thereby fortifying the overall security and stability of the federated learning system.

Crucially, it is noteworthy that the group leaders themselves are not directly involved in local model training, focusing instead on the critical task of aggregating models. This segregation of responsibilities contributes to a more efficient and secure federated learning environment within BGFL. Following this, the global aggregation group is meticulously composed, encompassing all the appointed group leaders. This collective group of leaders collaboratively contributes to the generation of the global model update, a process that encapsulates the diverse insights gained from the various training groups. Through this orchestrated mechanism, BGFL ensures not only the robustness of its collaborative learning paradigm but also the resilience and security needed for effective federated learning in Industry 4.0 scenarios.

Credit calculation

To prevent malicious nodes from providing fake quality assessment results, each node will be given a credit rating at the end of each iteration. What is noteworthy is that the latest credit is related to its learning the performance and accumulative credits stored in the blockchain, a sigmoid normalization function (i.e., Equation in the third procedure in Fig. 4) is leveraged to normalize the calculation of credit value. Specifically, we define the credit rating calculated by a node in the k-th iteration as c_k^t, where

$$c_k^t = \sum_{t=1}^{T} \frac{1}{1+e^{-\left(c_k^{t-1}+a_k\right)}}. \tag{6}$$

In this way, if a node has performed well in previous training iterations, it will not be discarded just because it performed poorly in one iteration, e.g., due to equipment or network factors. When the edge nodes in every group send their model parameters to the group leader, the local updates within the test dataset will be validated. Afterward, the node's credit will be obtained by the validation accuracy and its historical credit. Significantly, the initial credit of the edge node is determined by its contribution.

Local model training

It is worth noting that the initial model parameter is stored in the distributed ledger of the main blockchain, while it will be issued to the edge nodes by triggering a smart contract.

- Local model updates: Drawing from the initially initialized model parameters, the training nodes seamlessly access the latest model and embark on the execution of local training procedures. In a meticulous process, each edge node within every training group engages in the training of the local model, utilizing its raw data to minimize the expectation of mini-batches sampled from its local dataset [11]. Subsequently, the locally trained model parameters are transmitted to the designated group leader. Following the completion of several rounds of local training, the group leader takes charge by conducting local weight aggregation, harmonizing the contributions from individual nodes within the training group. This strategic orchestration of local training and subsequent aggregation by the group leader forms a pivotal step in the collaborative and federated learning process within the BGFL framework, ensuring the continual refinement and evolution of the global model.
- Local model aggregation: The group leader will perform weight model aggregation operation locally. Here, the model aggregation weight of the edge node is related to its credit. Significantly, the initial credit of the edge node is determined by its contribution, while every node's credit will be recalculated in every new iteration. Assume that the local aggregation is synchronous. After enough edge nodes upload their local models to the group leader, the aggregation will be triggered. To prevent malicious nodes from prolonging training time and degrading training performance, we specify a time threshold beyond which the training parameters of the node in this round will be abandoned. Once the local model aggregation is completed, all the local aggregation parameters will be uploaded to the global model aggregation group to obtain the global update. Specifically, the aggregation uses a credit score-based method, and its aggregation formula is as follows:

$$w_k^t = \sum_{i=0}^{N} c_{k,i}^t w_i, \tag{7}$$

where w_k^t represents the global model parameters in the k-th iteration, and $c_{k,i}^t$ represents the credit score of the i-th edge device in that iteration.

Global model aggregation

Upon the conclusion of the training group's requisite rounds within the ongoing iteration, the parameters derived from local aggregation undergo an upload process to the IPFS. Subsequently, the corresponding model hashes are committed to the main blockchain, where they undergo thorough verification. The validated hashes are then systematically appended to newly created update blocks. Once the comprehensive process of uploading and verifying the model has been successfully executed, the smart contract is activated to initiate the global aggregation phase.

This pivotal step in the federated learning process results in the creation of a novel global model seamlessly incorporated into the blockchain, marking the culmination of a synchronized and secure update cycle within the BGFL framework.

Group leader updating

At the culmination of each iteration within the BGFL framework, a reselection process occurs for nodes within the global aggregation group, with the current training group leader being subject to potential change. This mechanism ensures that every eligible node possesses the opportunity to ascend to the role of training group leader, contingent upon possessing superior credit. Notably, credit serves as a comprehensive metric, considering both the physical factor of contribution and training performance.

For effective cooperative learning, the group leader assumes the crucial responsibility of aggregating local node models. However, if a leader behaves maliciously, the overall performance of the global model can be significantly compromised. To mitigate such risks, BGFL employs a dynamic approach to electing the training group leader. The edge node with the highest credit is elected as the new training group leader, while all leaders collectively form the renewed global aggregation group. This transition initiates the conversion of the new group leader to the consensus node, and the preceding group leader assumes the role of the observation node.

Given that the credit of an edge node incorporates both the physical factor (contribution) and training performance, the process of updating group leaders serves as a mechanism to weed out lazy or unreliable workers, as well as workers with malicious intent. This deliberate curation of group leaders not only increases the cost of potential attacks but also bolsters the security and stability of the BGFL framework, contributing to a resilient and adaptive federated learning environment within the realm of Industry 4.0.

Simulation results and discussions

In this section, we present an extensive experimental analysis to evaluate the performance of the BGFL framework.

System implementation

To elucidate the performance of the BGFL framework, we set up the main blockchain based on the FISCO BCOS[1]. The FISCO BCOS platform is an open-source blockchain system project with a convenient pre-compiled contract framework at wireless industrial edges where the contract logic is fixed and computationally intensive. Factory occupancy detection is a popular issue in Industry 4.0. As such, we perform it on the real-world occupancy detection dataset[2], while simulating 2/4 groups with various industrial nodes, where each group has equal training nodes and a group leader with the highest credit. Moreover, we use Python 3.8.10 and Tensorflow 2.6.0 to realize the BGFL on an Intel Core CPU i7 with a clock rate of 2.8 GHz with 4 cores, while simulating the multi-nodes training in each blockchain group through the thread pool.

Numerical results

Parameter analysis

Specifically, we leverage the convolutional neural network to train the task while letting the local epoch be 9 and batch size be 50. Next, we tested the robustness of some parameters of the learning model. Figure 5 shows the impact of the learning rate on the prediction accuracy of the BGFL model. At the beginning of training, the model with the highest learning rate converges the fastest, as shown in Fig. 5 (1). As the epoch increases, the accuracy value of the model prediction with the higher learning rate is lower than that of the low learning rate. This is because the smaller the learning rate, the slower the loss gradient decreases and the longer it takes to converge. Meanwhile, too large a gradient descent step may cross the optimal value, causing the gradient to oscillate back and forth around the minimum, thus deteriorating performance. Here, we set the learning rate is 0.001, the loss function is shown in Fig. 5 (2).

Performance analysis

The examination of the influence of the number of industrial nodes on wireless communication latency, as depicted in Fig. 6, sheds light on insightful observations. Notably, the industrial nodes within each group are equitably distributed. It becomes evident that BGFL consistently outperforms the conventional FedAvg approach, with the advantageous trend becoming more pronounced as the number of nodes increases. This phenomenon can be attributed to the localized group aggregation in BGFL, typically situated at the wireless industrial edges. Here, the data transfer rates are inherently higher than those observed in the transmission from clients to the centralized server. Additionally, the correlation between the number of groups and wireless communication latency unveils a compelling trend: an increase in the number of groups corresponds to a reduction in wireless communication latency. This intriguing relationship underscores the efficacy of BGFL in optimizing communication dynamics, particularly in scenarios with varying numbers of industrial nodes and groups, presenting a promising avenue for enhancing the efficiency of collaborative learning in wireless industrial environments.

In a follow-up experiment, we simulate 4 groups with 20 industrial nodes, where each group has 4 training nodes and 1 group leader. Assuming that the geographic location is evenly distributed between each client and the parameter data decays by 1% per 100m during transmission, we compare the wireless communication latency of our proposed algorithm (with 2 and 4 groups) to the FedAvg approach. The results are shown in Fig. 7.

Analysis of Fig. 7 reveals an intuitive outcome: the BGFL consistently outperforms the FedAvg baseline

[1] https://github.com/FISCO-BCOS

[2] http://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+

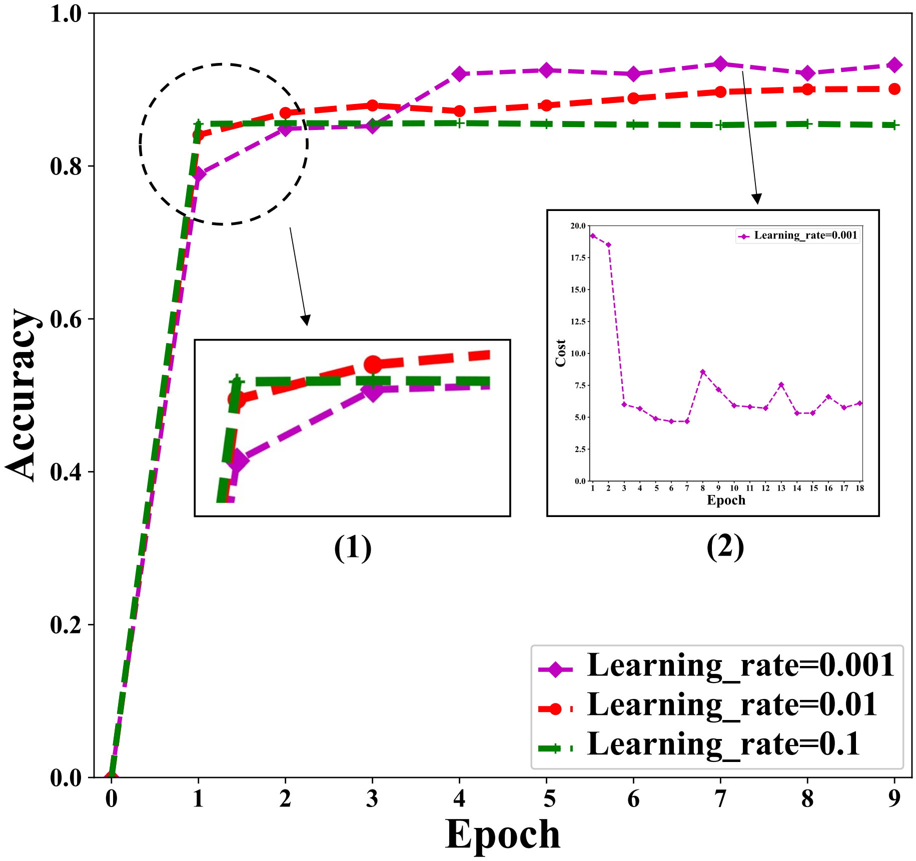

Fig. 5 The operational procedure of BGFL

across all scenarios. Notably, a discernible trend emerges wherein a larger blockchain group correlates with diminished wireless communication latency for the proposed algorithm. This can be attributed to the distinctive approach employed in BGFL, where clients are initially aggregated locally within the blockchain groups. This strategic maneuver significantly mitigates the escalation in data transfer size caused by data decay. Furthermore, it becomes evident that the collaborative learning performance is intricately linked to the geographical distance between industrial nodes. As this distance increases, the collaborative learning process experiences heightened wireless communication latency, owing to the augmented size of the requisite transmissions. The nuanced observations from this analysis underscore the efficacy of BGFL in optimizing wireless communication dynamics and enhancing collaborative learning performance compared to the baseline FedAvg model.

In Fig. 8, a comparative analysis of model prediction accuracy across different collaborative learning frameworks, namely FedAvg [37] and Random, is presented. The Random collaborative learning framework deviates from the conventional average aggregation by aggregating weights in a random manner. Notably, BGFL emerges as the frontrunner in terms of prediction accuracy, boasting an impressive value of 0.9327, while maintaining an acceptable convergence rate compared to the other two frameworks. The exceptional performance of BGFL can be attributed to its strategic allocation of higher weights to training nodes with elevated credit values in each training iteration, thereby significantly enhancing overall training performance.

Delving into the specifics, our framework exhibits training accuracy percentages that are 3.73% and 33.03% higher than those observed in the traditional FL and Random frameworks, respectively. This substantial improvement underscores the efficacy of BGFL in optimizing the training process by dynamically assigning weights based on credit values. The discernible advantage in accuracy positions BGFL as a robust and efficient collaborative learning framework, demonstrating its potential to outperform traditional methods and random aggregation approaches in real-world scenarios.

In the realm of traditional FL architectures, each client independently trains its local model based on local data, thereby exposing a vulnerable attack surface for potential

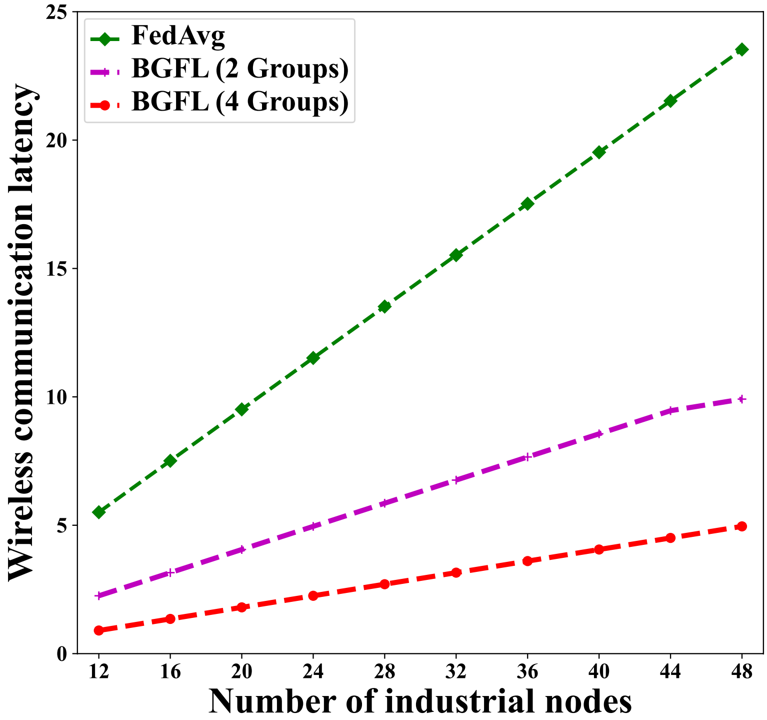

Fig. 6 Wireless communication latency VS numeber of industrial nodes

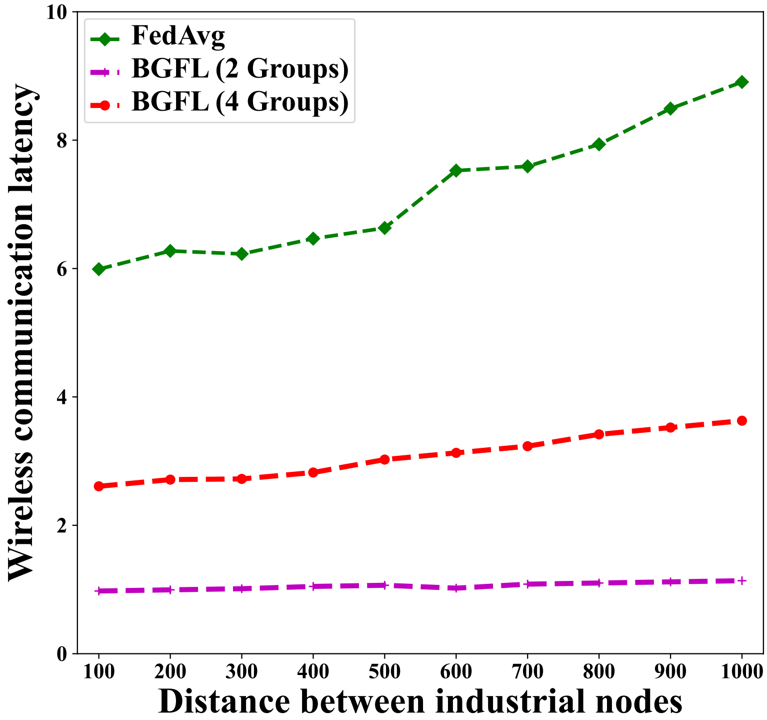

Fig. 7 Wireless communication latency VS distance of industrial nodes

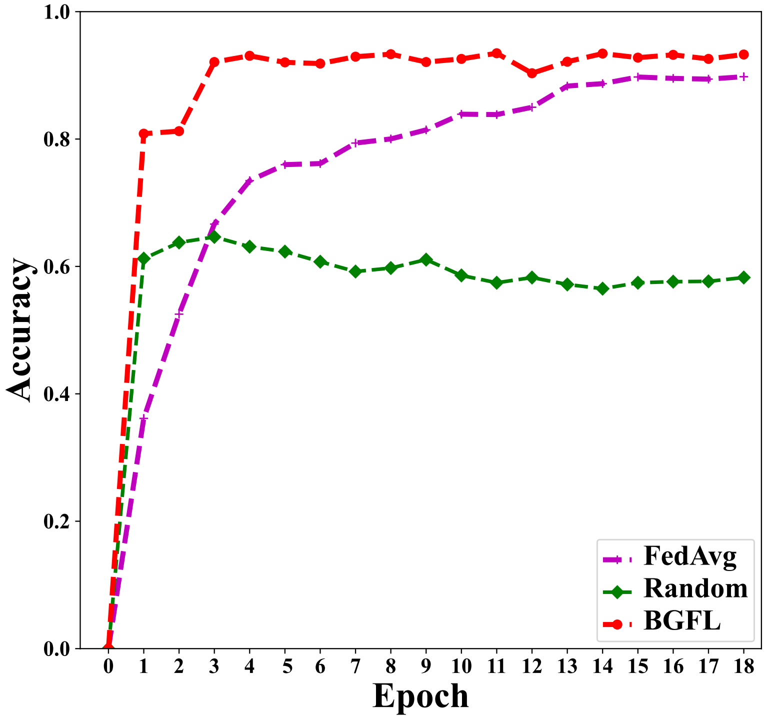

Fig. 8 Accuracy comparison

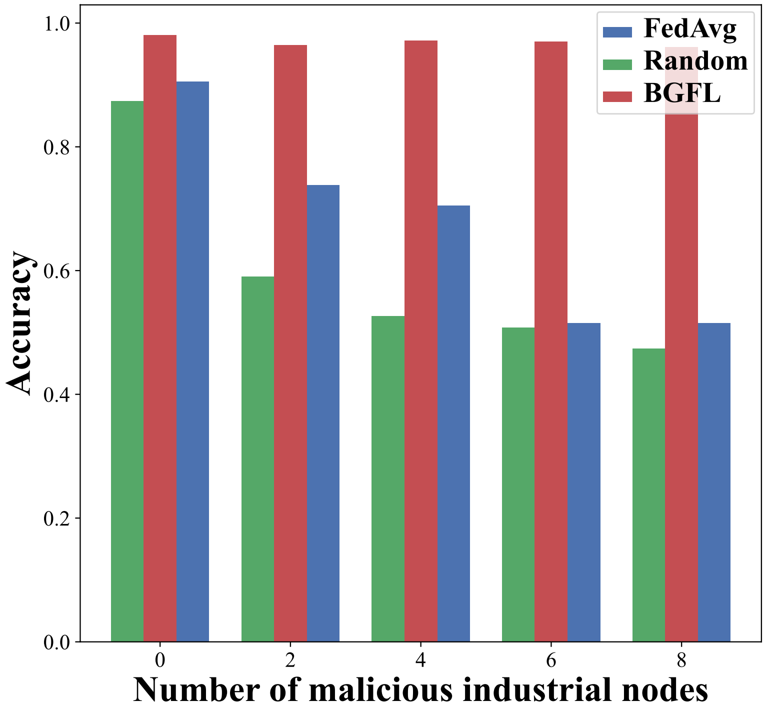

Fig. 9 Security analysis

malicious actors. To assess the robustness of the proposed BGFL, FedAvg, and Random frameworks against malicious node attacks, simulations were conducted, and the outcomes are illustrated in Fig. 9. Intriguingly, BGFL consistently outperforms the other two frameworks in terms of training performance, even in the presence of varying numbers of malicious nodes. Impressively, the average training accuracy of the BGFL framework maintains a high level at 0.9694 under such conditions.

The superior performance of BGFL can be attributed to its unique ability to address the vulnerability inherent in existing FL frameworks. In traditional FL setups, there is a lack of mechanisms to validate the correctness of locally learned models, thereby providing an avenue for clients to submit malicious models. When a node uploads malicious model parameters, the performance of cooperative learning rapidly degrades. BGFL, however, proves resilient in the face of such attacks, showcasing a remarkable 42.53% and 46.82% improvement in accuracy compared to the other two collaborative learning frameworks when confronted with eight malicious nodes in the edge scenario. This underscores the enhanced security and robustness embedded in the BGFL framework, making it a promising solution for collaborative learning in scenarios where security concerns are paramount.

Conclusion and future works

In the scope of our research, we have introduced the blockchain-enabled group federated learning (BGFL) framework, which is specifically developed for wireless industrial edges connected to the Internet of Things (IoT). This framework aims to effectively harness the combined capabilities of edge and cloud computing. Our research fills significant voids in current methodologies and introduces an advanced algorithmic structure that serves as the foundation for the BGFL framework. By conducting extensive simulations, we have successfully illustrated that BGFL substantially improves precision and protection in comparison to conventional methods, especially when implemented in practical situations. Our framework demonstrates exceptional robustness and adaptability, thereby establishing its efficacy in a wide range of collaborative learning tasks across multiple institutions. The ability to adapt is of the utmost importance in enabling the smooth implementation of extensive Federated Learning (FL) assignments, thus propelling the development of collaborative learning in the era of Industry 4.0.

We intend to further develop the capabilities of BGFL in the future by investigating the flexible blockchain group method, which has the potential to significantly enhance the framework's adaptability in real-world scenarios. Furthermore, our objective is to optimize the efficacy of the collaborative learning procedure through the integration of smart contracts for asynchronous parameter aggregation. This strategic progression is in line with our continuous dedication to enhancing and progressing the BGFL framework, guaranteeing its sustained effectiveness and pertinence in the swiftly changing domains of mobile edge computing and cloud computing.

Authors' contributions

Author #1 Guozheng Peng: Author 1 plays a key role in the methodology of research, responsible for designing research methods and frameworks and ensuring the methodological foundation and effectiveness of the research. In addition, he is responsible for planning experimental designs and methods to ensure the reproducibility and scientificity of the research. Author #2 Xiaoyun Shi: Author 2 plays an important role in software development and is responsible for developing and implementing software tools or platforms that support blockchain-enabled group federated learning for industrial edge applications. He leads the overall design, coding, and testing of the software to ensure its stability and performance in practical applications. Author #3 Jun Zhang: Author 3 is responsible for verifying the effectiveness of the study, covering the design of experimental and testing plans, to ensure that the results and methods of the study are feasible and effective in practical applications. In addition, he analyzes and interprets experimental data to support the conclusions of the study. Author #4 Lisha Gao: Author 4 is responsible for collecting, organizing, and managing data to ensure the quality and consistency of the datasets used. She performs data cleaning and preprocessing to ensure the availability and analyzability of the data. Author #5 Yuanpeng Tan: Author 5 is responsible for conducting field investigations and experimental work to collect data and information for research analysis and validation. He collects actual data in industrial edge environments to support empirical analysis of the research. Author #6 Nan Xiang: Author 6 undertook most of the survey work, including collecting relevant literature and understanding existing technologies and solutions, in order to provide a comprehensive review of the background and related work sections of the paper. Author #7 Wanguo Wang: Author 7 plays a crucial role in the initial draft of the paper, integrating various aspects of the research and presenting the main findings and methods.

Funding

The work is funded by the science and technology project of SGCC (State Grid Corporation of China): Research on Key Technologies and Applications of Intelligent Edge Computing for Transmission Line Defect Sensing(5700-202318309A-1-1-ZN).

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 7 May 2024 Accepted: 26 August 2024
Published online: 08 October 2024

References

1. Alsamhi SH, Shvetsov AV, Hawbani A, Shvetsova SV, Kumar S, Zhao L (2023) Survey on federated learning enabling indoor navigation for industry 4.0 in B5G. Future Gener Comput Syst 148:250–265
2. Yang Y, Feng L, Sun Y, Li Y, Zhou F, Li W, Wang S (2024) Decentralized cooperative caching and offloading for virtual reality task based on gan-powered multi-agent reinforcement learning. IEEE Trans Serv Comput 17(1):291–305

3. Otoum S, Ridhawi IA, Mouftah HT (2023) A federated learning and blockchain-enabled sustainable energy trade at the edge: a framework for industry 4.0. IEEE Internet Things J 10(4):3018–3026
4. Duan Q, Huang J, Hu S, Deng R, Lu Z, Yu S (2023) Combining federated learning and edge computing toward ubiquitous intelligence in 6g network: Challenges, recent advances, and future directions. IEEE Commun Surv Tutor 25(4):2892–2950
5. Chai S, Huang J (2024) Dependent task scheduling using parallel deep neural networks in mobile edge computing. J Grid Comput 22(1):27
6. Li A, Song SL, Chen J, Li J, Liu X, Tallent NR, Barker KJ (2020) Evaluating modern GPU interconnect: Pcie, nvlink, nv-sli, nvswitch and gpudirect. IEEE Trans Parallel Distrib Syst 31(1):94–110
7. Ranathunga T, McGibney A, Rea S, Bharti S (2023) Blockchain-based decentralized model aggregation for cross-silo federated learning in industry 4.0. IEEE Internet Things J 10(5):4449–4461
8. Aloqaily M, Ridhawi IA, Kanhere SS (2023) Reinforcing industry 4.0 with digital twins and blockchain-assisted federated learning. IEEE J Sel Areas Commun 41(11):3504–3516
9. Lakhan A, Grønli T, Bellavista P, Memon S, Alharby M, Thinnukool O (2024) IoT workload offloading efficient intelligent transport system in federated ACNN integrated cooperated edge-cloud networks. J Cloud Comput 13(1):79
10. Du M, Zheng H, Gao M, Feng X (2024) Adaptive decentralized federated learning in resource-constrained IoT networks. IEEE Internet Things J 11(6):10739–10753
11. Kaur G, Grewal SK (2024) Aggregation techniques in wireless communication using federated learning: a survey. Int J Wirel Mob Comput 26(2):115–126
12. Pfeiffer K, Rapp M, Khalili R, Henkel J (2023) Federated learning for computationally constrained heterogeneous devices: A survey. ACM Comput Surv 55(14s):334:1–334:27
13. Kar B, Yahya W, Lin Y, Ali A (2023) Offloading using traditional optimization and machine learning in federated cloud-edge-fog systems: A survey. IEEE Commun Surv Tutor 25(2):1199–1226
14. Sun X, Yang S, Zhao C (2023) Lightweight industrial image classifier based on federated few-shot learning. IEEE Trans Ind Inform 19(6):7367–7376
15. Bugshan N, Khalil I, Rahman MS, Atiquzzaman M, Yi X, Badsha S (2023) Toward trustworthy and privacy-preserving federated deep learning service framework for industrial internet of things. IEEE Trans Ind Inform 19(2):1535–1547
16. Yang W, Xiang W, Yang Y, Cheng P (2023) Optimizing federated learning with deep reinforcement learning for digital twin empowered industrial IoT. IEEE Trans Ind Inform 19(2):1884–1893
17. Qiu W, Ai W, Chen H, Feng Q, Tang G (2023) Decentralized federated learning for industrial IoT with deep echo state networks. IEEE Trans Ind Inform 19(4):5849–5857
18. Moudoud H, Cherkaoui S (2023) Multi-tasking federated learning meets blockchain to foster trust and security in the metaverse. Ad Hoc Netw 150(103):264
19. Guo X (2022) Implementation of a Blockchain-enabled Federated Learning Model that Supports Security and Privacy Comparisons. In: 5th IEEE International Conference on Information Systems and Computer Aided Education (ICISCAE) 2022. IEEE, Dalian, p 243–247
20. Bodagala H, Priyanka H, (2022) Security for IoT using federated learning. In: 2022 International Conference on Recent Trends in Microelectronics. Automation, Computing and Communications Systems (ICMACC), pp 131–136
21. Zhao L, Tang X, You Z, Pang Y, Xue H, Zhu L (2020) Operation and Security Considerations of Federated Learning Platform Based on Compute First Network. In: 2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops). Chongqing, p 117-121
22. El Houda ZA, Nabousli D, Kaddoum G (2022) Cost-efficient federated reinforcement learning- based network routing for wireless networks. In: 2022 IEEE Future Networks World Forum (FNWF). Montreal, p 243-248
23. Behmandpoor P, Patrinos P, Moonen M (2022) Federated learning based resource allocation for wireless communication networks. In: 2022 30th European Signal Processing Conference (EUSIPCO). Belgrade, p 1656–1660
24. Shaheen M, Farooq MS, Umer T (2024) AI-empowered mobile edge computing: inducing balanced federated learning strategy over edge for balanced data and optimized computation cost. J Cloud Comput 13(1):52
25. Giagkos D, Tzenetopoulos A, Masouros D, Soudris D, Xydis S (2023) Darly: Deep reinforcement learning for QoS-aware scheduling under resource heterogeneity optimizing serverless video analytics. 16th IEEE International Conference on Cloud Computing, CLOUD 2023, Chicago, IL, USA, July 2-8, 2023 pp 1–3
26. Xiong J, Zhu H (2024) Privmaskfl: A private masking approach for heterogeneous federated learning in IoT. Comput Commun 214:100–112
27. Razaque A, Khan M, Yoo J, Alotaibi A, Alshammari M, Almiani M (2024) Blockchain-enabled heterogeneous 6G supported secure vehicular management system over cloud edge computing. Internet Things 25:101115
28. Qu Y, Pokhrel SR, Garg S, Gao L, Xiang Y (2021) A blockchained federated learning framework for cognitive computing in industry 4.0 networks. IEEE Trans Ind Inform 17(4):2964–2973
29. Qiu C, Yao H, Wang X, Zhang N, Yu FR, Niyato D (2020) AI-chain: Blockchain energized edge intelligence for beyond 5G networks. IEEE Netw 34(6):62–69
30. Masood AB, Hasan A, Vassiliou V, Lestas M (2023) A blockchain-based data-driven fault-tolerant control system for smart factories in industry 4.0. Comput Commun 204:158–171
31. Huang X, Han L, Li D, Xie K, Zhang Y (2023) A reliable and fair federated learning mechanism for mobile edge computing. Comput Netw 226(109):678
32. Li Y, Chen C, Liu N, Huang H, Zheng Z, Yan Q (2021) A blockchain-based decentralized federated learning framework with committee consensus. IEEE Netw 35(1):234–241
33. Ayepah-Mensah D, Sun G, Boateng GO, Anokye S, Liu G (2024) Blockchain-enabled federated learning-based resource allocation and trading for network slicing in 5G. IEEE/ACM Trans Netw 32(1):654–669
34. Huang X, Wu Y, Liang C, Chen Q, Zhang J (2023) Distance-aware hierarchical federated learning in blockchain-enabled edge computing network. IEEE Internet Things J 10(21):19163–19176
35. Wan Y, Qu Y, Gao L, Xiang Y (2022) Privacy-preserving blockchain-enabled federated learning for B5G-driven edge computing. Comput Netw 204:108671
36. Zhang Z, Yue S, Zhang J (2024) Towards resource-efficient edge AI: from federated learning to semi-supervised model personalization. IEEE Trans Mob Comput 23(5):6104–6115
37. Aboueleneen N, Alwarafy A, Abdallah M (2023) Secure and energy-efficient communication for internet of drones networks: a deep reinforcement learning approach. In: IEEE International Wireless Communications and Mobile Computing, IWCMC 2023, Marrakesh, Morocco, June 19-23, 2023, pp 818–823

Publisher’s Note