Secure and Efficient Federated Learning for Smart Grid With Edge-Cloud Collaboration

Zhou Su, *Senior Member, IEEE*, Yuntao Wang, Tom H. Luan, *Senior Member, IEEE*, Ning Zhang, *Senior Member, IEEE*, Feng Li, *Member, IEEE*, Tao Chen, and Hui Cao, *Member, IEEE*

***Abstract*—With the prevalence of smart appliances, smart meters, and Internet of Things (IoT) devices in smart grids, artificial intelligence (AI) built on the rich IoT big data enables various energy data analysis applications and brings intelligent and personalized energy services for users. In conventional AI of Things (AIoT) paradigms, a wealth of individual energy data distributed across users' IoT devices needs to be migrated to a central storage (e.g., cloud or edge device) for knowledge extraction, which may impose severe privacy violation and data misuse risks. Federated learning, as an appealing privacy-preserving AI paradigm, enables energy data owners (EDOs) to cooperatively train a shared AI model without revealing the local energy data. Nevertheless, potential security and efficiency concerns still impede the deployment of federated-learning-based AIoT services in smart grids due to the low-quality shared local models, non-independently and identically distributed (non-IID) data distributions, and unpredictable communication delays. In this article, we propose a secure and efficient federated-learning-enabled AIoT scheme for private energy data sharing in smart grids with edge-cloud collaboration. Specifically, we first introduce an edge-cloud-assisted federated learning framework for communication-efficient and privacy-preserving energy data sharing of users in smart grids. Then, by considering non-IID effects, we design a local data evaluation mechanism in federated learning and formulate two optimization problems for EDOs and energy service providers. Furthermore, due to the lack of knowledge of multidimensional user private information in practical scenarios, a two-layer deep reinforcement-learning-based incentive algorithm is developed to promote EDOs' participation and high-quality model contribution. Extensive simulation results show that the proposed scheme can effectively stimulate EDOs to share high-quality local model updates and improve the communication efficiency.**

***Index Terms*—Artificial intelligence of things (AIoT), deep reinforcement learning (DRL), edge-cloud collaboration, federated learning, smart grid.**

I. Introduction

SMART grid, as the next generation of the power grid, holds significant promise to improve the reliability, flexibility, and efficiency of the power systems [1]–[4]. With the prevalence of Internet of Things (IoT) devices such as smart appliances and smart meters in smart grids, a tremendous amount of energy data is produced from individual smart devices [5], [6]. Driven by artificial intelligence (AI) technologies, energy service providers (ESPs) (e.g., utility companies) can make full use of such rich data for better power consumption prediction and increased profits and market penetration via dynamic pricing and customized energy strategies. Besides, users can enjoy personalized energy services with improved quality of experience (QoE). However, in such paradigms of AI of Things (AIoT) [7], the fine-grained smart metering data collection from users' smart devices that enables diverse intelligent smart grid services also imposes risks to users' privacy violation and data misuse [8]–[10], as the shared energy data inevitably contain users' private information. For example, the habits and lifestyle of users can be easily inferred from their real-time energy consumption data, abused by ESPs to facilitate business advertisements, and even results in criminal activities such as burglary. In particular, in AIoT, the private energy data need to be migrated to a central storage for knowledge extraction in conventional AI models [7], [11], [12].

Federated learning, as an appealing privacy-preserving AI paradigm, enables ESPs to extract knowledge and insights from users' private energy data, while each energy data owner (EDO) can keep their training data at their local devices [13]–[15]. As shown in Fig. 1, the EDOs only need to iteratively deliver the parameters of the local AI model trained on their local energy datasets in each round of the training process; the ESP aggregates a globally shared AI model based on EDOs' inputs and distributes it back to EDOs for their local training in the next round. This training process is repeated until a predefined desirable accuracy of the global model is attained [16]. Instead of

Manuscript received March 3, 2021; revised May 22, 2021; accepted June 27, 2021. Date of publication July 8, 2021; date of current version October 27, 2021. This work was supported in part by NSFC under Grant U20A20175 and Grant U1808207, the Fundamental Research Funds for the Central Universities. Paper no. TII-21-1042. *(Corresponding author: Tom H. Luan.)*

Zhou Su and Yuntao Wang are with the School of Cyber Science, Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zhousu@ieee.org; yuntao.wang@stu.xjtu.edu.cn).

Tom H. Luan is with the School of Cyber Information Engineering, Xidian University, Xi'an 710071, China (e-mail: Tom.luan@xidian.edu.cn).

Ning Zhang is with the Department of Electrical Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: ning.zhang@uwindsor.ca).

Feng Li and Tao Chen are with the State Grid Xinjiang Electric Power Company, Ürümqi 830002, China (e-mail: 252491552@qq.com; 303197127@qq.com).

Hui Cao is with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: huicao@mail.xjtu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3095506.

Digital Object Identifier 10.1109/TII.2021.3095506

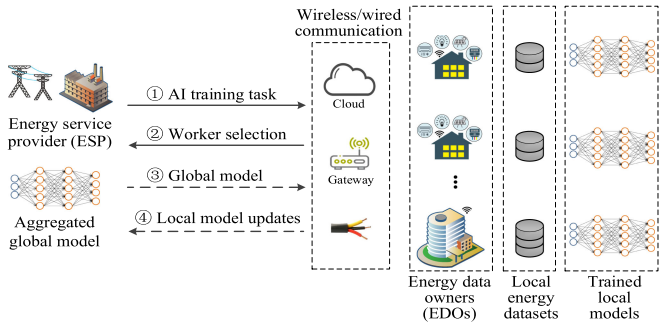


Fig. 1. Illustration of federated-learning-based AIoT in smart grids.

migrating all training data (e.g., users' private energy data) to a central server for model training, with federated learning, the raw private energy data are kept on users' local devices and only the intermediary model parameters are shared, thereby protecting users' data privacy in AIoT [17].

Despite the promising benefits, federated learning still faces a series of fundamental challenges when it comes to private energy data sharing in smart grids. First, as it incurs various costs during collaborative learning, users will be reluctant to participate in the learning process without proper compensations. Besides, selfish users may share low-quality local model updates (e.g., generated from uncalibrated data and fewer training samples) to save their costs and even cheat the ESP by delivering erroneous or meaningless local model updates, thereby deteriorating the accuracy of the global model. Second, different from the conventional AI models, the time for federated learning contains not only the computation time on model training but also the communication delay between users and the ESP for frequent gradients exchange. Due to the intermittent and unreliable wireless connections between users and the ESP, the efficiency of federated learning may be greatly degraded. Third, the training data owned by heterogeneous users (e.g., in residential, commercial, and industrial areas) is usually nonindependently and identically distributed (non-IID) in smart grid. Existing literature [15], [18] show that, compared with IID data, the accuracy can be deteriorated in federated learning with non-IID data. Accordingly, it increases the ESP's difficulty in evaluating its cost on accuracy loss in the learning process. Hence, how to design a secure and communication-efficient federated learning scheme in smart grids with a large number of users while considering the effects of non-IID data distribution needs to be investigated.

In the literature, many research works have been reported toward secure and efficient federated learning. For example, in [16], a contract theoretical incentive mechanism is designed to motivate mobile devices to participate in the federated learning process and contribute their computation resources by compensating their computation costs. Besides, an edge computing-based crowdsourcing framework is developed in [12] to facilitate federated learning services for IoT devices, where a Stackelberg game model is also employed to spur users' cooperation and high-quality AI model building. Meanwhile, in [19], an efficient synchronous federated learning mechanism is presented, where reinforcement learning (RL) is leveraged to derive the optimal pricing strategy of the aggregation server without knowledge of the accurate amount of users' training data. Notably, there exist multidimensional user private information (e.g., training cost, communication delay, and data distribution dissimilarity) in smart grids that are not readily available for the ESP during federated learning. However, existing works mainly consider 1-D private information, where the presence of multidimensional user private information in smart grids is absent. Besides, the related literature usually assumes that users' training data follows the ideal IID distribution, and few works focus on the non-IID data distribution of training samples. In addition, the free-riding users may cheat the ESP to gain more profits by disseminating low-quality, redundant, or meaningless local models, which is not sufficiently investigated in existing works. Therefore, it is still an open and vital issue to improve the security and efficiency of private data management in smart grids while stimulating users' high-quality model sharing under the federated learning framework.

In this article, to address the aforementioned issues, we propose a secure and efficient federated-learning-based AIoT scheme for smart grids with edge-cloud collaboration. Specifically, we first present an edge-cloud-assisted federated learning framework for *communication-efficient* and *privacy-preserving* energy data sharing in smart grids, where the orchestration of edge-cloud computing can promote proximally model training services with reduced communication latency and alleviated communication overhead in gradient exchange. Afterwards, we develop a local data evaluation model in federated learning with consideration of the non-IID distribution of heterogeneous users, where the IID distribution is regarded as a special case. Furthermore, the payoff functions of users and ESPs, as well as their optimization problems are formulated under the federated learning framework. Especially, the free riders are punished with a zero-payment mechanism to improve the fairness of learning. Finally, to deal with the lack of knowledge of multidimensional user private information, the interactions between users and ESPs are formulated as finite Markov decision processes (MDPs), and a two-layer deep reinforcement learning (DRL)-based incentive algorithm is devised to stimulate users' participation and high-quality model contribution. The main contributions of this article are threefold.

1) *Edge-cloud integrated federated learning framework*. We develop an edge-cloud-assisted federated learning framework to facilitate private data sharing in the smart grid, where two optimization problems are formulated based on the local data evaluation model and user payoff model with non-IID data distribution. With the aforementioned models, both the communication efficiency and privacy preservation of users can be improved.
2) *DRL-based optimal strategy decision*. Due to the presence of multidimensional user private information and the large state space, each ESP can apply deep Q-network (DQN) to derive the optimal payment strategy for users to tradeoff its accuracy loss and expenditure in federated learning. Meanwhile, each user can apply the DQN to derive its optimal training strategy of local model (i.e.,

data quality and data size) to maximize its payoff in the dynamic network.

3) *Extensive simulations-based performance evaluation*. Extensive simulations are conducted to evaluate the performance of the proposed scheme. The simulation results demonstrate that the proposed scheme can effectively stimulate high-quality local model sharing of users, acquire optimal strategies for participants, and reduce task delays.

The remainder of this article is organized as follows. Section II reviews the related work. Section III introduces the system model. Section IV formulates the optimization problems. The DRL-based solution for optimal strategy decision is proposed in Section V. Performance evaluation is given in Section VI. Section VII concludes this article.

II. Related Works

A. Edge-Cloud Cooperation Mechanisms in Smart Grids

Recently, the edge-cloud cooperation mechanisms in smart grids have attracted wide attention from both academia and industry. Zhang *et al.* [2] present a cost-effective building demand response control framework under edge-cloud collaboration to enable automated RL control execution with little human configuration in smart grids. Jiang *et al.* [3] develop an edge-cloud-assisted automatic control strategy for massive users' air conditioners to eliminate the rebound peak of the power grid, where edge servers can collaboratively migrate the heavy tasks from the cloud based on the cooperative game model. Zhao *et al.* [9] investigate a privacy-preserving data aggregation mechanism with dynamic pricing in edge computing-based smart grids, where edge nodes are deployed at substations to facilitate the information flows between users' smart meters and the control center. Liu *et al.* [20] propose a collaborative edge-cloud computing-based architecture for a secure and efficient function query and aggregation communication in smart grids, where users' private usage data are encrypted and outsourced to a cloud to avoid data misuse while the deployment of edge nodes enables low-latency communications. Ruan *et al.* [5] design a hierarchical residential energy management mechanism with edge-cloud cooperation to enable real-time response and improve the system reliability in power systems, where a load priority model is also studied for user satisfaction evaluation. One can observe that the implementation of edge-cloud cooperation mechanisms in smart grids under the federated learning framework is seldom studied in the existing literature.

B. Incentives of Federated Learning

In recent times, there have been a number of efforts on incentives for the emerging federated learning (as a distributed collaborative AI paradigm). Du *et al.* [21] survey existing works on the efficiency, security, and privacy concerns of federated learning and investigate its potentials and challenges in vehicular IoT applications. Jiao *et al.* [22] devise a reverse auction-based incentive mechanism to optimize the social welfare and motivate data owners' participation in federated learning. Besides, a DRL-based auction algorithm integrated with graph neural networks is developed to further improve the system efficiency. Lim *et al.* [23] propose a hierarchical incentive framework in mobile networks where the federation formation among multiple mobile users is modeled as a coalition game and the optimal contract design for data owners with different quality and quantity of local data within a federation is modeled by the contract game. Tran *et al.* [11] investigate the optimized federated learning problems in wireless networks to cope with the heterogeneous power limits and local training samples of mobile devices, where a variable decomposition approach is used to split the nonconvex optimization problem and obtain the closed-form solutions of subproblems. Yu *et al.* [24] design a fair and sustainable incentive mechanism for dynamic revenue sharing among data owners in federated learning to maximize the collective utility while ensuring contribution fairness, expectation fairness, and regret distribution fairness. By modeling the interactions between the model owner and mobile devices as a Stackelberg game with budget limit, Sarikaya *et al.* [25] present a game-based incentive algorithm to optimize the CPU power allocation among workers for faster convergence rate of federated learning. Nonetheless, the presence of multidimensional user private information and the free-riding data owners in designing optimal incentive mechanisms are still underexplored. Besides, to realize existing works in practice, the edge-cloud integrated architecture for the implementation of efficient federated learning should be further studied.

Different from existing works, our work studies the cost-effective edge-cloud integrated federated learning for smart grids. The free-riding EDOs, multidimensional private information of users, and effects of non-IID data distribution are jointly considered under the edge-cloud integrated architecture to improve the security and efficiency of federated learning in smart grids. In addition, a DRL-based algorithm with two layers is employed to acquire the optimal strategies of both EDOs and ESPs in the dynamic networks without being aware of accurate network parameters and participants' private parameters.

III. System Model

In this section, we introduce the system model including the network model, federated learning model, data utility model, and threat model. Table I summarizes the key notations.

A. Network Model

Fig. 2 shows the scenario of edge-cloud integrated federated learning for smart grids, which includes a cloud, multiple geographically distributed aggregators, ESPs, and a large number of EDOs.

1) Energy Data Owners (EDOs): Let $\mathbb{I} = \{1, \ldots, i, \ldots, I\}$ denote the set of individual EDOs in the investigated area, e.g., residential areas and commercial areas. Each EDO $i \in \mathbb{I}$ is the owner of the smart home or smart building, and possesses a variety of heterogeneous energy data generated by his smart meters, home appliances, solar panels, electric vehicles, etc. Let $\mathbb{S}_i$ denote the private energy dataset (i.e., a collection of ordered energy usage records collected from various data sources) owned

TABLE I
KEY NOTATIONS

Notation	Description
$\mathbb{I}, \mathbb{J}, \mathbb{M}$	Sets of EDOs, ESPs, and aggregators.
$\mathbb{I}_m$	Set of EDOs in the the coverage of aggregator m.
$\Upsilon_{j,k}$	k-th published federated learning task of ESP j.
K_j	Total number of tasks of ESP j.
$\mathbb{I}_{j,k}$	Set of EDOs that join the learning process of task $\Upsilon_{j,k}$.
$\Theta^n_{i,j,k}$	Local model update of EDO i for task $\Upsilon_{j,k}$ at n-th epoch.
$\widetilde{\Theta}^n_{m,j,k}$	Edge aggregation by aggregator m for task $\Upsilon_{j,k}$ at n-th epoch.
$\Theta^n_{j,k}$	Aggregated global model by cloud for task $\Upsilon_{j,k}$ at n-th epoch.
$S_{i,j,k}$	Contributed training data size of EDO i for task $\Upsilon_{j,k}$.
$Q_{i,j,k}$	Data quality of training samples of EDO i for task $\Upsilon_{j,k}$.
$\Delta_{i,j,k}$	Earth mover's distance (EMD) of EDO i for task $\Upsilon_{j,k}$.
$\overline{\Delta}_{j,k}$	Average EMD value of task $\Upsilon_{j,k}$.
$\Psi_{i,j,k}$	Accuracy of local model (AoLM) of EDO i in task $\Upsilon_{j,k}$.
$\Lambda_{j,k}$	Accuracy of global model (AoGM) of task $\Upsilon_{j,k}$.
$p_{i,j,k}$	Payment for the highest AoLM of EDO i in task $\Upsilon_{j,k}$.
$u_{i,j,k}, c_{i,j,k}$	Payoff/cost function of EDO i in performing task $\Upsilon_{j,k}$.
$\pi_{j,k}$	Payoff function of ESP j in k-th task.
$\boldsymbol{\sigma}_{i,j,k}$	Training strategy of EDO i in task $\Upsilon_{j,k}$.
$\mathbf{s}^t_{j,k}$	AoLM state vector of EDOs in task $\Upsilon_{j,k}$ at time slot t.
$\mathcal{Q}(\mathbf{s}^t_{j,k}, \mathbf{p}^t_{j,k})$	Q-function at state $\mathbf{s}^t_{j,k}$ with action $\mathbf{p}^t_{j,k}$.
$\hat{s}^t_{i,j,k}$	Payment state to EDO i in task $\Upsilon_{j,k}$ at time slot t.
$\hat{\mathcal{Q}}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k})$	Q-function at state $\hat{s}^t_{i,j,k}$ with action $\boldsymbol{\sigma}^t_{i,j,k}$.

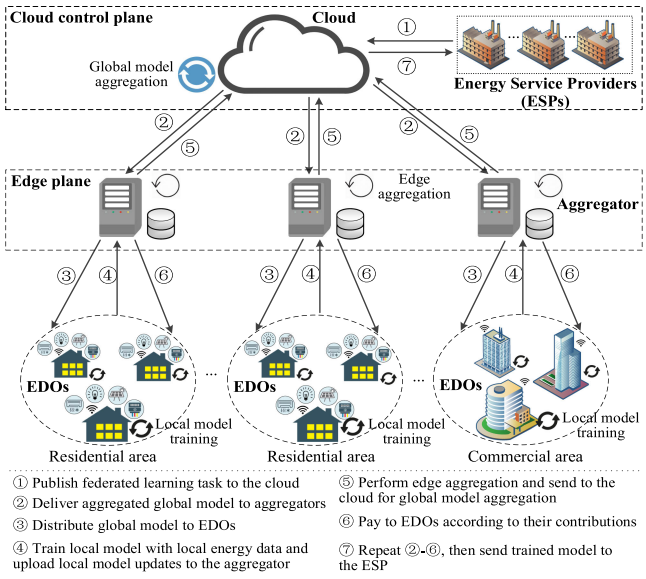


Fig. 2. System overview of the edge-cloud integrated federated learning for personal energy data analysis in the smart grid.

by EDO $i \in \mathbb{I}$, and $S_i = |\mathbb{S}_i|$ is the size (i.e., the number of data samples) of dataset $\mathbb{S}_i$.

2) Energy Service Providers (ESPs): Let $\mathbb{J} = \{1, \ldots, j, \ldots, J\}$ denote the set of ESPs (e.g., utility companies) in the smart grid. Distinguished from the centralized AI models, each ESP $j \in \mathbb{J}$ can publish a set of federated learning tasks, denoted by $\Upsilon = \{\Upsilon_{j,1}, \ldots, \Upsilon_{j,k}, \ldots, \Upsilon_{j,K_j}\}$, to the cloud and derive knowledge and insights learned from EDOs' shared personal energy data while keeping the training data on EDO's local devices, thereby preserving the user privacy. K_j is the total number of tasks of the ESP j. By processing the aggregated personal energy data of EDOs, each ESP $j \in \mathbb{J}$ can deliver customized energy services to customers (i.e., individual EDOs) and enhance its quality of service for increased profits and market penetration. Besides, each EDO can enjoy improved QoE and make personalized energy plans. Let $\mathbb{I}_{j,k} = \{1, \ldots, i, \ldots, I_{j,k}\}$ be the set of EDOs that join the learning process of task $\Upsilon_{j,k}$.

3) Cloud: The cloud can be publicly accessed and is provisioned with powerful computing, communication, and storage capacities. A group of federated learning tasks are hosted and managed by the cloud in the cloud control plane.

4) Aggregators: A group of aggregators geographically distributed in the network form the edge plane, the set of which is denoted as $\mathbb{M} = \{1, \ldots, m, \ldots, M\}$. Each aggregator $m \in \mathbb{M}$ acts as the edge computing node and can offer edge computing and wireless communication services for multiple smart homes and smart buildings within a specific area [1], [5], [9]. As aggregators are in proximity of energy users, the high latency and intermittent connection involved in remote data transmissions in federated learning can be mitigated. Moreover, the deployment of aggregators can facilitate proximal model aggregations at the edge of the network, thereby alleviating the heavy data traffic to the cloud. The set of EDOs in the coverage of the aggregator m is defined as $\mathbb{I}_m = \{1, \ldots, i, \ldots, I_m\}$.

B. Federated Learning Model

An illustrative example that applies federated learning in the smart grid is short-term load forecasting (STLF) [13]. As the related energy data in the ESP are limited, federated-learning-based STLF relies on a collection of vast high-quality energy data from households to attain an accurate STLF prediction for improved power delivery efficiency. During federated learning for task $\Upsilon_{j,k}$, each involved EDO $i \in \mathbb{I}_{j,k}$ can determine the set $\mathbb{S}_{i,j,k} \subseteq \mathbb{S}_i$ of energy data with data size $S_{i,j,k} = |\mathbb{S}_{i,j,k}|$ in training global model $\Theta_{j,k}$. $\mathbb{S}_{i,j,k}$ can be further denoted by a collection of input–output pairs $\{x_a, y_a\}_{a=1}^{S_{i,j,k}}$, where x_a is the input (e.g., a household electricity usage record) and y_a is the label (e.g., the actual individual STLF). In machine learning, the target is to learn the model parameters $\Theta_{j,k}$ that predicts the label y_a based on the input x_a. The loss function, denoted as $f(x_a; \Theta_{j,k})$, characterizes the gap between the prediction value and the actual label on each training sample $\{x_a, y_a\}$. For each EDO $i \in \mathbb{I}_{j,k}$, its loss function is the average prediction loss of all data samples in $\mathbb{S}_{i,j,k}$, i.e.,

$$\mathscr{L}_i(\Theta_{j,k}) = \frac{1}{S_{i,j,k}} \sum_{a=1}^{S_{i,j,k}} f(x_a; \Theta_{j,k}). \tag{1}$$

The global loss function, denoted as $\mathscr{L}(\Theta_{j,k})$, is the weighted average of all EDOs' local loss functions. The objective of federated learning is to find the optimal global model parameters $\Theta^*_{j,k}$ for task $\Upsilon_{j,k}$ that minimize the global loss function, i.e.,

$$\begin{aligned}\Theta^*_{j,k} &= \arg\min_{\Theta_{j,k}} \mathscr{L}(\Theta_{j,k}) \\ &= \frac{1}{\sum_{i=1}^{I_{j,k}} S_{i,j,k}} \arg\min_{\Theta_{j,k}} \sum_{i=1}^{I_{j,k}} S_{i,j,k} \mathscr{L}_i(\Theta_{j,k}).\end{aligned} \tag{2}$$

Under edge-cloud integrated smart grids, to solve formula (2) with federated learning, as shown in Fig. 2, the following phases need to be undertaken at the nth global training epoch ($n = 0, 1, 2, \ldots$).

1) *Task publishing and model initialization.* After the ESP publishes its federated learning task (step ①), the cloud initializes a global model $\Theta_{j,k}^{n}$, $n=0$, with random weights and delivers it to all participating EDOs via the aggregators in the edge plane (step ②∼③).
2) *Local update.* After receiving the global model $\Theta_{j,k}^{n-1}$, $n \geq 1$, each EDO $i \in \mathbb{I}_{j,k}$ calculates its local model update $\Theta_{i,j,k}^{n}$ using its local energy data samples, and then, sends $\Theta_{i,j,k}^{n}$ to the corresponding aggregator for further processing (step ④). In particular, the EDO i splits its local dataset $\mathbb{S}_{i,j,k}$ into minibatches with size ς_B and trains each minibatch via stochastic gradient descent (SGD), i.e.,
$$\Theta_{i,j,k}^{n} \leftarrow \Theta_{j,k}^{n-1} - \eta \nabla \mathscr{L}_i \left(\Theta_{j,k}^{n-1}\right) \tag{3}$$
where η is the learning rate of SGD, and $\nabla \mathscr{L}_i$ is the gradient of $\mathscr{L}_i$ on the minibatch.
3) *Edge aggregation.* Each aggregator $m \in \mathbb{M}$ performs edge aggregation by aggregating the local model updates of all EDOs within its coverage, and then, it sends the intermediate aggregation results $\widetilde{\Theta}_{m,j,k}^{n} = \sum_{i \in \mathbb{I}_m} S_{i,j,k} \Theta_{i,j,k}^{n}$ and $\widetilde{S}_{m,j,k} = \sum_{i \in \mathbb{I}_m} S_{i,j,k}$ to the cloud (step ⑤).
4) *Global aggregation.* The cloud updates the current global model $\Theta_{j,k}^{n}$ by synthesizing the edge aggregation results retrieved from aggregators, and then, it calculates the payments to the involved EDOs according to their contributions (step ⑥). Specifically, the classical federated averaging algorithm [14] is applied for global model aggregation, i.e.,
$$\Theta_{j,k}^{n} \leftarrow \frac{1}{\sum_{m \in \mathbb{M}} \widetilde{S}_{m,j,k}} \sum\nolimits_{m \in \mathbb{M}} \widetilde{\Theta}_{m,j,k}^{n}. \tag{4}$$
5) *Learning ends.* Until the global model attains a predefined accuracy, the learning process ends and the trained model is delivered to the corresponding ESP (step ⑦).

In federated learning, a *global training epoch* is composed of the distributed local update phase on EDOs followed by the parallel edge aggregation phase on aggregators and a global aggregation phase on the cloud.

C. Data Utility Model

Based on works [22] and [23], the higher data availability or data utility of EDOs' local training data in performing federated learning tasks can result in higher accuracy of EDOs' local model updates and a faster convergence rate of the trained global model. Accordingly, EDOs can make distinct contributions to task completion. To quantify EDOs' local data utilities and contributions in each task, an *inference accuracy of local model (AoLM)* metric is defined in federated learning. Considering the unique characteristics of federated learning in smart grids, three critical attributes of EDO's local energy data are considered in AoLM evaluation, i.e., the *data size*, the *data quality*, and the *data distribution*. Based on [23] and the experimental results in [26], large training data samples or higher data quality (e.g., a higher level of data cleaning and annotation with anomaly detection and redundancy elimination) generally contribute to improved model accuracy and better prediction performance. Besides, different from the assumption of IID distribution of training data in conventional AI models, the local energy data of EDOs are usually heterogeneous and non-IID under federated learning scenarios.

To study the non-IID effects on accuracy loss, the widely adopted earth mover's distance (EMD) metric is leveraged to quantify the dissimilarity (i.e., weights divergence) of local data distributions among different EDOs. A large EMD indicates a higher dissimilarity that adversely influences the accuracy of the global model. We consider a Y-class classification task in which EDO i's data samples $\{\mathbf{x}_i, y_i\}$ are distributed over $\mathbb{X} \times \mathbb{Y}$ following the distribution $\mathcal{P}_{i,j,k}$, where $\mathbb{X}$ is a compact space and $\mathbb{Y} = \{1, \ldots, Y\}$ is a label space. Given the distribution $\mathcal{P}_{j,k}$ for the whole population in task $\Upsilon_{j,k}$, the EMD of EDO i, denoted as $\Delta_{i,j,k}$, is defined as
$$\Delta_{i,j,k} = \sum\nolimits_{u=1}^{Y} \|\mathcal{P}_{i,j,k}(y=u) - \mathcal{P}_{j,k}(y=u)\|. \tag{5}$$
In the aforementioned equation, $\mathcal{P}_{i,j,k}$ and $\mathcal{P}_{j,k}$ mean the proportions of energy data with label u in EDO i's local energy data and all EDOs involved in task $\Upsilon_{j,k}$, respectively. Let $S_{i,j,k}^{\max}$ be EDO i's maximum amount of training data that can be contributed to task $\Upsilon_{j,k}$, then we have $0 \leq S_{i,j,k} \leq S_{i,j,k}^{\max}$. Here, $S_{i,j,k}=0$ means that EDO i does not participate the task $\Upsilon_{j,k}$. Let $Q_{i,j,k}$ denote the data quality of training samples of EDO i for task $\Upsilon_{j,k}$, which meets the constraint
$$Q_{i,j,k} = \begin{cases} 0, & \text{EDO } i \text{ is cheating} \\ (0,1], & \text{otherwise} \end{cases}. \tag{6}$$
Here, $Q_{i,j,k}=0$ means that EDO i is free riding and cheats the ESP j by delivering fake, redundant, or meaningless local training samples. $Q_{i,j,k}=1$ represents the highest data quality of EDO i's local training data. Based on the experimental validation in [22] and [27], the AoLM of EDO i in task $\Upsilon_{j,k}$, denoted as $\Psi_{i,j,k} \in [0,1]$, can be modeled by the following function via curve fitting approaches, i.e.,
$$\begin{aligned} \Psi_{i,j,k} &= \varphi\left(Q_{i,j,k}, S_{i,j,k}, \Delta_{i,j,k}\right) \\ &= \rho\left(\Delta_{i,j,k}\right) - \lambda_1 e^{-\lambda_2 (\lambda_3 Q_{i,j,k} S_{i,j,k})^{\rho\left(\Delta_{i,j,k}\right)}} \end{aligned} \tag{7}$$
where $\rho(\Delta_{i,j,k}) = \lambda_4 e^{-(\lambda_5 + \frac{\Delta_{i,j,k}}{\lambda_6})^2}$. Here, λ_u $(1 \leq u \leq 6)$ are positive curve fitting parameters. The first term $\rho(\Delta_{i,j,k})$ captures the degradation of the model performance when the EMD $\Delta_{i,j,k}$ increases. The exponential term $\lambda_1 e^{-\lambda_2 (\lambda_3 Q_{i,j,k} S_{i,j,k})^{\rho(\Delta_{i,j,k})}}$ reflects that the higher data quality and data size, the better model accuracy, and the diminished marginal returns. Furthermore, the *inference accuracy of global model (AoGM)* metric is defined to quantify the quality of global model $\Theta_{j,k}$. Accordingly, the AoGM of task $\Upsilon_{j,k}$, denoted as $\Lambda_{j,k}$, can be expressed as
$$\Lambda_{j,k} = \rho\left(\overline{\Delta}_{j,k}\right) - \lambda_1 e^{-\lambda_2 \left(\frac{\lambda_3}{I_{j,k}} \sum\limits_{i \in \mathbb{I}_{j,k}} Q_{i,j,k} S_{i,j,k}\right)^{\rho\left(\overline{\Delta}_{j,k}\right)}} \tag{8}$$

where $\overline{\Delta}_{j,k} = \frac{1}{I_{j,k}} \sum_{i \in \mathbb{I}_{j,k}} \Delta_{i,j,k}$ is the average EMD value of task $\Upsilon_{j,k}$. The IID distribution of EDOs' local energy data can be regarded as a special case when $\overline{\Delta}_{j,k} = 0$.

D. Threat Model

In the system, both the cloud and aggregators are supposed to be *honest-but-curious*, who will follow the federated learning protocol honestly but attempt to learn from all its received messages. The following two types of threats are considered to undermine the security of federated learning services in smart grids.

1) Low-Quality Local Model Update Attack: Due to the selfishness and the constrained resources, EDOs may share low-quality local model updates (e.g., generated from uncalibrated data and fewer training samples) to reduce their costs if without sufficient compensations. As a consequence, the accuracy of the trained model will be deteriorated, causing a performance degradation of the federated learning service.

2) Free-Riding Attack: The self-interested EDOs may launch the free-riding attack by cheating the ESPs and delivering fake, meaningless, or redundant local model updates while enjoying personalized energy services from ESPs via federated learning. Accordingly, the enthusiasm of honest participants can be damped.

IV. Problem Formulation

In this section, we analyze the payoff functions of ESPs and individual EDOs, and then, formulate their optimization problems during federated learning services in smart grids.

A. Payoff Function of EDOs

To compensate EDOs' cost in federated learning process, a payment strategy $p_{i,j,k}$ is determined by the ESP j to each involved EDO i for task $\Upsilon_{j,k}$ as a reward to his contribution. Besides, the zero-payment strategy is adopted by ESPs for EDOs that do not participate the learning process (i.e., $S_{i,j,k} = 0$) or cheat the ESP (i.e., $Q_{i,j,k} = 0$). The payoff function of EDO $i \in \mathbb{I}_{j,k}$ is the difference between its revenue minus its cost, i.e.,

$$u_{i,j,k}\left(Q_{i,j,k}, S_{i,j,k}\right) = \varpi \kappa_p p_{i,j,k} \Psi_{i,j,k} - (1-\varpi)\, \mathcal{C}_{i,j,k} \quad (9)$$

where ϖ is the weight parameter to balance the revenue and cost. κ_p is the adjustment parameter. $p_{i,j,k}$ is the payment for the highest AoLM of EDO i in task $\Upsilon_{j,k}$, i.e., $\Psi_{i,j,k} = 1$. $\mathcal{C}_{i,j,k}$ is the total cost of EDO i in executing task $\Upsilon_{j,k}$, which includes the local data cost $c^{\text{data}}_{i,j,k}$, the local computational cost $c^{\text{comp}}_{i,j,k}$, and the communication cost $c^{\text{comm}}_{i,j,k}$, i.e.,

$$\mathcal{C}_{i,j,k} = c^{\text{data}}_{i,j,k} + c^{\text{comp}}_{i,j,k} + c^{\text{comm}}_{i,j,k}. \quad (10)$$

Let η^1_i be the unit cost of local energy data with the highest data quality (i.e., $S_{i,j,k} = 1$ and $Q_{i,j,k} = 1$). Then, the local data cost $c^{\text{data}}_{i,j,k}$ of EDO i can be defined as [23]

$$c^{\text{data}}_{i,j,k} = \eta^1_i Q_{i,j,k} S_{i,j,k}. \quad (11)$$

Next, the local computational cost $c^{\text{comp}}_{i,j,k}$ of EDO i is in proportion to its data size $S_{i,j,k}$ [15], i.e.,

$$c^{\text{comp}}_{i,j,k} = \eta^2_i S_{i,j,k} \xi^l_{i,j,k} \chi \quad (12)$$

where η^2_i is EDO i's unit computational cost, χ is the model size, and $\xi^l_{i,j,k}$ is the number of local training epoches. Finally, EDO i's communication cost $c^{\text{comm}}_{i,j,k}$ is dominated by the uplink transmission cost [16] and is denoted as: $c^{\text{comm}}_{i,j,k} = \varrho_i \frac{\chi}{\gamma^{\text{up}}_{i,m}}$, where ϱ_i is the transmit power of EDO i, and $\gamma^{\text{up}}_{i,m}$ is the uplink transmission rate between EDO i and the corresponding aggregator m.

B. Payoff Function of ESPs

The payoff function of the ESP $j \in \mathbb{J}$ in kth task is the obtained model accuracy minus the total payment to all involved EDOs and the cloud, i.e.,

$$\pi_{j,k}\left(\mathbf{p}_{j,k}\right) = \mu \nu_a \Lambda_{j,k} - (1-\mu)\left(\sum_{i \in \mathbb{I}_{j,k}} p_{i,j,k} \Psi_{i,j,k} + \lambda_c \Xi^{\text{comp}}_{j,k}\right) \quad (13)$$

where $\mathbf{p}_{j,k} = (p_{i,j,k})_{1 \le i \le I_{j,k}}$, and $0 \le p_{i,j,k} \le p^{\max}_{j,k}$. Here, $p^{\max}_{j,k}$ is the maximum payment of the ESP j to any EDO in task $\Upsilon_{j,k}$. $\Lambda_{j,k}$ is the AoGM indicating the performance of the trained global model of the task $\Upsilon_{j,k}$. μ is the weight parameter to balance the model accuracy and total payment. ν_a and λ_c are positive adjustment parameters. $\Xi^{\text{comp}}_{j,k}$ is the payment of the ESP j to the cloud for the cost in global model aggregation, which is proportional to the model size χ and the number of EDOs $I_{j,k}$ [22], i.e., $\Xi^{\text{comp}}_{j,k} = p_{\text{cloud}} \chi I_{j,k}$, where p_{cloud} is the unit payment of ESPs to the cloud.

C. Optimization Problems

The target of each EDO $i \in \mathbb{I}_{j,k}$ is to maximize its payoff function by deciding its training strategy (i.e., the data quality $Q_{i,j,k}$ and data size $S_{i,j,k}$) in task $\Upsilon_{j,k}$, and its optimization problem is given by

$$\begin{aligned} \text{Problem 1}: \max_{Q_{i,j,k}, S_{i,j,k}} \; & u_{i,j,k}\left(Q_{i,j,k}, S_{i,j,k}\right) \\ \text{s.t. } & 0 \le S_{i,j,k} \le S^{\max}_{i,j,k},\ 0 \le Q_{i,j,k} \le 1. \end{aligned} \quad (14)$$

The target of each ESP $j \in \mathbb{J}$ is to maximize its payoff function by deciding its payment strategy vector $\mathbf{p}_{j,k}$ for all involved EDOs in each task $\Upsilon_{j,k}$, $1 \le k \le K_j$, and its optimization problem is given by

$$\begin{aligned} \text{Problem 2}: \max_{\mathbf{p}_{j,k}} \; & \pi_{j,k}\left(\mathbf{p}_{j,k}\right) \\ \text{s.t. } & 0 \le p_{i,j,k} \le p^{\max}_{j,k},\ 1 \le i \le I_{j,k}. \end{aligned} \quad (15)$$

V. Two-Layer DQN-Based Optimal Strategy

In practice, as EDOs' training strategies are private and only known by themselves, ESPs usually have difficulty in accurately estimating the private training strategies of EDOs to determine the optimal payment strategies. Alternatively, based on the previous observations, each ESP can apply reinforcement learning (RL) technology to design its optimal payment strategy to

Algorithm 1: DQN-Based Payment Strategy of the ESP.

1: **Initialize:** $\mathcal{Z}_{j,k}$, φ_j, $\Omega_{j,k}=\emptyset$, W, ς.
2: Initialize the Q-network with random weights ϕ.
3: **for** $t=1,2,\ldots,T$ **do**
4: Set $\mathbf{s}_{j,k}^t=\{\Psi_{i,j,k}^{t-1}\}_{i=1}^{I_{j,k}}$.
5: **if** $n\leq W$ **then**
6: Choose $\mathbf{p}_{j,k}^t\in(\mathcal{Z}_{j,k})^{I_{j,k}}$ at random.
7: **else**
8: Construct $\alpha_{j,k}^t$ via (16) and input to the CNN.
9: Attain the Q-function $\mathcal{Q}(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k})$, $\forall\mathbf{p}_{j,k}\in(\mathcal{Z}_{j,k})^{I_{j,k}}$, from the CNN output via (18).
10: Opt $\mathbf{p}_{j,k}^t$ via the ϵ-greedy policy and send $p_{i,j,k}^t$ to EDO $i\in\mathbb{I}_{j,k}$.
11: **end if**
12: Observe and evaluate EDOs' AoLM sequence $\{\Psi_{i,j,k}^t\}_{i=1}^{I_{j,k}}$.
13: Calculate payoff $\pi_{j,k}^t$ via (13).
14: Update $\Omega_{j,k}$ with the new experience $\delta_{j,k}^t$ in (17).
15: Select a minibatch $\mathbb{O}_{j,k}$ from $\Omega_{j,k}$ at random.
16: Update ϕ^t with the gradient vector in (20).
17: **end for**

stimulate EDOs' high-quality local model sharing without fully knowing EDOs' local training policies [28]. Besides, each EDO can design its optimal training strategy in federated learning by adopting RL technology to maximize its payoff function via trial and error.

A. DQN-Based Payment Strategy

An overpaid policy tends to cause overtraining, and thus, reduces the ESP's long-term rewards. Meanwhile, an underpaid task usually results in undertraining or even a failure of the learning task. Besides, EDOs usually select their training strategies according to the historical payments. Thereby, ESP's current payment strategy impacts the future federated learning results and the long-term payoffs. The payment process of the ESP j can be modeled as an MDP with finite states under repeated interactions [29], [30], and the optimal payment policy can be derived via RL approaches. The DQN-based payment for the ESP j is shown in Algorithm 1. In this article, we assume that the criteria for evaluating the AoLMs of EDOs are publicly known for all the participants and the AoLM evaluation is accurate. In RL, the current payment strategy vector $\mathbf{p}_{j,k}^t$ of the ESP j in task $\Upsilon_{j,k}$ at tth time slot is selected according to the current system state $\mathbf{s}_{j,k}^t$ and the Q-function $\mathcal{Q}(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t)$. In line 4, the state $\mathbf{s}_{j,k}^t$ is composed of the prior AoLM sequences of EDOs in performing task $\Upsilon_{j,k}$, i.e., $\mathbf{s}_{j,k}^t=\{s_{i,j,k}^t\}_{i=1}^{I_{j,k}}=\{\Psi_{i,j,k}^{t-1}\}_{i=1}^{I_{j,k}}$. For simplicity, the feasible payments of the ESP j for different AoLM levels are uniformly discretized into $Z_{j,k}$ levels, i.e., $p_{i,j,k}^t\in\mathcal{Z}_{j,k}=\{\frac{z}{Z_{j,k}-1}\cdot p_{j,k}^{\max}\}_{0\leq z\leq Z_{j,k}-1}$.

Note that for the ESP j, its size of state space can be denoted as $(I_{j,k})^{A\times B}$ [31], which increases with both the number of involved EDOs (i.e., $I_{j,k}$) and the size of action space of every EDO (i.e., $A\times B$), where A and B are the numbers of data quality and data size levels of EDOs in task $\Upsilon_{j,k}$, respectively. Thereby, traditional RL methods such as Q-learning need to address the curse of dimensionality for efficient payment decision making. In our DQN-based payment strategy, a deep convolutional neural network (CNN) is exploited to cope with the curse of dimensionality and reduce the convergence time. Specifically, to learn from historical experiences efficiently, each ESP j constructs its state sequence $\alpha_{j,k}^t$ for kth task in DQN (line 8), which contains the recent $W+1$ states, i.e.,

$$\alpha_{j,k}^t=\left\{\mathbf{s}_{j,k}^{t-W},\mathbf{s}_{j,k}^{t-W+1},\ldots,\mathbf{s}_{j,k}^t\right\}. \tag{16}$$

Besides, a replay memory, denoted as $\Omega_{j,k}=\{\delta_{j,k}^d\}_{1\leq d\leq D}$, is employed to store the latest D experiences and improve the learning performance. Here, each experience $\delta_{j,k}^t$ is composed of the state sequence transition, the current payment, and the obtained payoff, i.e.,

$$\delta_{j,k}^t=\left\{\alpha_{j,k}^t,\mathbf{p}_{j,k}^t,\pi_{j,k}^t,\alpha_{j,k}^{t+1}\right\}. \tag{17}$$

Here, ESP j's payoff $\pi_j(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t)$ is rewritten as $\pi_{j,k}^t$ in short.

In the CNN, two convolutional (Conv) layers are included, followed by two fully connected (FC) layers. As summarized in Table II, the first Conv layer uses 20 filters each with stride 1 and size 4×4, and the second Conv layer has 40 filters each with stride 1 and size 2×2. We employ the rectified linear unit (ReLU) as activation functions in both Conv layers. The first FC layer contains 320 ReLUs, and the second FC layer has Z outputs for each payment strategy $p_{i,j,k}$. Via state-space compression, the state sequence $\alpha_{j,k}^t$ can be reshaped into a 8×8 matrix, and then, input to the CNN. The Q-function for each state-action pair, as the expected discounted long-term reward, can be estimated by $\mathcal{Q}(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t)\approx\mathcal{Q}(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t;\phi^t)$, where ϕ^t is a vector of Q-network parameters at tth time slot. The Q-function can be approximated by the output of the CNN (line 9) as

$$\begin{aligned}&\mathcal{Q}\left(\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t\right)\\&=\mathbb{E}_{\mathbf{s}_{j,k}^{t+1}}\left\{\pi_{j,k}^t+\varphi_j\max_{\mathbf{p}'_{j,k}}\mathcal{Q}\left(\mathbf{s}_{j,k}^{t+1},\mathbf{p}'_{j,k}\right)|_{\mathbf{s}_{j,k}^t,\mathbf{p}_{j,k}^t}\right\}\end{aligned} \tag{18}$$

where the payoff $\pi_{j,k}^t$ is the immediate reward of the ESP j, $\varphi_j\in[0,1]$ is the discount factor indicating the myopic view of the ESP j, and $\mathbf{s}_{j,k}^{t+1}$ is the new system state transformed from the current state $\mathbf{s}_{j,k}^t$ with action $\mathbf{p}_{j,k}^t$. Based on the current system

TABLE II
ARCHITECTURE PARAMETERS OF THE CNN FOR ESPS AND EDOS IN THE DQN

Layer	Conv 1	Conv 2	FC 1	FC 2 of ESP	FC 2 of EDO
Input	8×8	$5\times5\times20$	640	320	320
Filter Size	4×4	2×2	/	/	/
Stride	1	1	/	/	/
Num. of Filters	20	40	320	Z	$A\times B$
Activation	ReLU	ReLU	ReLU	ReLU	ReLU
Output	$5\times5\times20$	$4\times4\times40$	320	Z	$A\times B$

Algorithm 2: DQN-Based Training Strategy of the EDO.

1: **Initialize:** $\mathcal{A}_i$, $\mathcal{B}_i$, φ_i, $\hat{\Omega}_{i,j,k} = \emptyset$, W, ς.
2: Initialize the Q-network with random weights $\hat{\phi}$.
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Set $\hat{s}^t_{i,j,k} = p^{t-1}_{i,j,k}$.
5: **if** $n \leq W$ **then**
6: Choose $Q^t_{i,j,k} \in \mathcal{A}_i$ and $S^t_{i,j,k} \in \mathcal{B}_i$ at random.
7: **else**
8: Construct $\hat{\alpha}^t_{i,j,k}$ via (21) and input to the CNN.
9: Attain the Q-function $\hat{\mathcal{Q}}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}_{i,j,k})$, $\forall \boldsymbol{\sigma}_{i,j,k} \in \mathcal{A}_i \times \mathcal{B}_i$, from the CNN output via (23).
10: Opt $\boldsymbol{\sigma}^t_{i,j,k}$ via the ϵ-greedy policy.
11: **end if**
12: Observe the payment $p^t_{i,j,k}$.
13: Calculate payoff $u^t_{i,j,k}$ via (9).
14: Update $\hat{\Omega}_{i,j,k}$ with the new experience $\hat{\delta}^t_{i,j,k}$ in (22).
15: Select a minibatch $\hat{\mathbb{O}}_{j,k}$ from $\hat{\Omega}_{i,j,k}$ at random.
16: Update $\hat{\phi}^t$ with the gradient vector in (25).
17: **end for**

state and the output Q-values, in line 10, the current payment $\mathbf{p}^t_{j,k}$ of the ESP j can be determined via the ϵ-greedy mechanism to balance the exploration and exploitation, in which the greedy action that results in the highest Q-value is opted with a high chance ϵ_j (i.e., *exploitation*), while other actions are opted with a small probability $1 - \epsilon_j$ at random (i.e., *exploration*).

Based on the experience replay method [32], a minibatch $\mathbb{O}_{j,k} \subseteq \Omega_{j,k}$ with size $O \leq D$ is randomly sampled from the replay memory $\Omega_{j,k}$ for Q-network training (line 15). In particular, the Q-network parameters ϕ^t at tth time slot is updated by minimizing the loss function via minibatch SGD (line 16), where the loss function is defined as the mean-squared error (MSE) of the target optimal Q-function, i.e.,

$$\mathscr{L}\left(\phi^t\right) = \mathbb{E}_{\delta^t_{j,k}}\left[\left(\pi^t_{j,k} + \varphi_j \max_{\mathbf{p}'_{j,k}} \mathcal{Q}\left(\alpha^{t+1}_{j,k}, \mathbf{p}'_{j,k}; \phi^{t-1}\right) - \mathcal{Q}\left(\alpha^t_{j,k}, \mathbf{p}^t_{j,k}; \phi^t\right)\right)^2\right]. \tag{19}$$

Differentiating the $\mathscr{L}(\phi^t)$ with respect to the weights, we attain the gradient as follows:

$$\nabla_{\phi^t}\mathscr{L}\left(\phi^t\right) = \mathbb{E}_{\delta^t_{j,k}}\left[\left(\pi^t_{j,k} + \varphi_j \max_{\mathbf{p}'_{j,k}} \mathcal{Q}\left(\alpha^{t+1}_{j,k}, \mathbf{p}'_{j,k}; \phi^{t-1}\right) - \mathcal{Q}\left(\alpha^t_{j,k}, \mathbf{p}^t_{j,k}; \phi^t\right)\right) \cdot \nabla_{\phi^t}\mathcal{Q}\left(\alpha^t_{j,k}, \mathbf{p}^t_{j,k}; \phi^t\right)\right]. \tag{20}$$

B. DQN-Based Model Training Strategy

The training strategy-making process of the EDO i can be modeled as a finite MDP, and the EDO's optimal training policy can be derived via the DQN. The DQN-based training strategy for the EDO i is shown in Algorithm 2. At tth time slot, the current training strategy $\boldsymbol{\sigma}^t_{i,j,k} = (Q^t_{i,j,k}, S^t_{i,j,k})$ of the EDO i in the task $\Upsilon_{j,k}$ in the DQN is decided based on the current system state $\hat{s}^t_{i,j,k}$ and the Q-function $\hat{\mathcal{Q}}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k})$. In line 4, the observed system state $\hat{s}^t_{i,j,k}$ is composed of the prior payment for the task $\Upsilon_{j,k}$, i.e., $\hat{s}^t_{i,j,k} = p^{t-1}_{i,j,k}$. For simplicity, the data quality and data size levels of the EDO i in the task $\Upsilon_{j,k}$ are uniformly discretized into A_i and B_i levels, respectively. We have $Q^t_{i,j,k} \in \mathcal{A}_i = \{\frac{a}{A_i - 1}\}_{0 \leq a \leq A_i - 1}$, and $S^t_{i,j,k} \in \mathcal{B}_i = \{\frac{b}{B_i - 1} S^{\max}_{i,j,k}\}_{0 \leq b \leq B_i - 1}$. In the DQN-based training process for the task $\Upsilon_{j,k}$, each EDO i builds its state sequence $\hat{\alpha}^t_{i,j,k}$ in line 8, which is composed of the recent $W + 1$ states, i.e.,

$$\hat{\alpha}^t_{i,j,k} = \left\{\hat{s}^{t-W}_{i,j,k}, \hat{s}^{t-W+1}_{i,j,k}, \ldots, \hat{s}^t_{i,j,k}\right\}. \tag{21}$$

Then, the state sequence is reshaped it into a 8×8 matrix as an input to the CNN. Besides, each EDO i stores the latest D experiences into a replay memory, denoted as $\hat{\Omega}_{i,j,k} = \{\hat{\delta}^d_{i,j,k}\}_{1 \leq d \leq D}$, for better learning performance. Each experience $\hat{\delta}^t_{i,j,k}$ includes the state sequence transition, the current training action, and the reward, i.e.,

$$\hat{\delta}^t_{i,j,k} = \left\{\hat{\alpha}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}, u^t_{i,j,k}, \hat{\alpha}^{t+1}_{i,j,k}\right\}. \tag{22}$$

Here, the payoff $u_{i,j,k}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k})$ is the immediate reward of the EDO i and is rewritten as $u^t_{i,j,k}$ in short.

As shown in Table II, the architecture of the CNN for each EDO is similar to ESPs except that it has $A \times B$ outputs. The outputs of the CNN model are the estimation of the Q-function for each state-action pair (line 9), i.e.,

$$\begin{aligned}\hat{\mathcal{Q}}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}) &\approx \hat{\mathcal{Q}}(\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}; \hat{\phi}^t) \\ &= \mathbb{E}_{\hat{s}^{t+1}_{i,j,k}}\left\{u^t_{i,j,k} + \varphi_i \max_{\boldsymbol{\sigma}'_{i,j,k}} \hat{\mathcal{Q}}\left(\hat{s}^{t+1}_{i,j,k}, \boldsymbol{\sigma}'_{i,j,k}\right) \mid_{\hat{s}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}}\right\}\end{aligned} \tag{23}$$

where $\varphi_i \in [0, 1]$ is the discount factor indicating the myopic view of the EDO i, and $\hat{s}^{t+1}_{i,j,k}$ is the new system state transformed from the current state $\hat{s}^t_{i,j,k}$ with action $\boldsymbol{\sigma}^t_{i,j,k}$. In addition, EDO i's current training strategy $\boldsymbol{\sigma}^t_{i,j,k}$ is decided in line 10 through the ϵ-greedy mechanism by choosing the greedy action that maximizes the Q-function with a high chance ϵ_i and randomly selecting other nongreedy actions with a small chance $1 - \epsilon_i$.

At each time slot, the parameters $\hat{\phi}^t$ of the Q-network is updated in line 16 similarly by minimizing $\mathscr{L}(\hat{\phi}^t)$, i.e., the MSE of the target optimal Q-function, via SGD with minibatch updates, i.e.,

$$\mathscr{L}\left(\hat{\phi}^t\right) = \mathbb{E}_{\hat{\delta}^t_{i,j,k}}\left[\left(u^t_{i,j,k} + \varphi_i \max_{\boldsymbol{\sigma}'_{i,j,k}} \hat{\mathcal{Q}}\left(\hat{\alpha}^{t+1}_{i,j,k}, \boldsymbol{\sigma}'_{i,j,k}; \hat{\phi}^{t-1}\right) - \hat{\mathcal{Q}}\left(\hat{\alpha}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}; \hat{\phi}^t\right)\right)^2\right] \tag{24}$$

$$\nabla_{\hat{\phi}^t}\mathscr{L}\left(\hat{\phi}^t\right) = \mathbb{E}_{\hat{\delta}^t_{i,j,k}}\left[\left(u^t_{i,j,k} \varphi_i \max_{\boldsymbol{\sigma}'_{i,j,k}} \hat{\mathcal{Q}}\left(\hat{\alpha}^{t+1}_{i,j,k}, \boldsymbol{\sigma}'_{i,j,k}; \hat{\phi}^{t-1}\right)\right.\right.$$

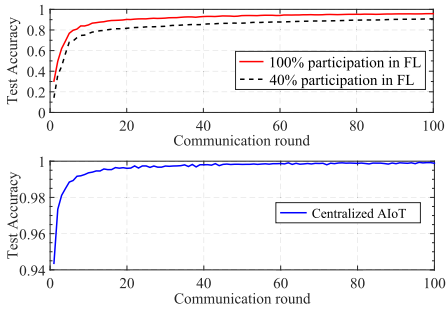


Fig. 3. Comparison of prediction accuracy of the CNN on MNIST under federated learning (FL) and centralized AIoT with different participation levels and communication rounds.

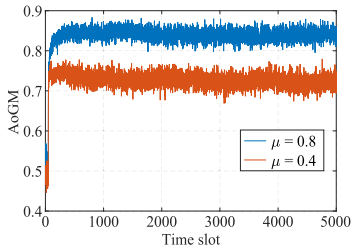


Fig. 4. Evolution of the AoGM of the ESP with different values of μ.

$$- \hat{\mathcal{Q}}\left(\hat{\alpha}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}; \hat{\phi}^t\right)\Big) \cdot \nabla_{\hat{\phi}^t} \hat{\mathcal{Q}}\left(\hat{\alpha}^t_{i,j,k}, \boldsymbol{\sigma}^t_{i,j,k}; \hat{\phi}^t\right)\Big]. \quad (25)$$

Based on the experience replay method, in line 15, a minibatch $\hat{\mathbb{O}}_{j,k} \subseteq \hat{\Omega}_{i,j,k}$ is randomly selected from the memory pool $\hat{\Omega}_{i,j,k}$ to train $\hat{\phi}^t$.

VI. Performance Evaluation

In this section, we evaluate the performance of the proposed scheme by using a simulator in Python. We first give the security analysis of our system. Then, the simulation setup is introduced, followed by the numerical results and findings.

A. Security Analysis

Our proposed approach can defend against adversaries defined in Section III-D. Besides, existing differential privacy (DP) schemes [33] and cryptography-based secure aggregation schemes [34] can be further applied in our system to prevent the inexplicit privacy leakage from aggregators and the cloud. Specifically, the proposed approach can protect the system from the following attacks.

1) *Low-quality local model update attack.* Through the DQN-based incentive mechanism, the selfishness of EDOs can be suppressed by dynamically assigning proper compensations to them. Meanwhile, simulation results in Figs. 4 and 8 validate that our approach can motive EDOs to contribute high-quality local model updates.

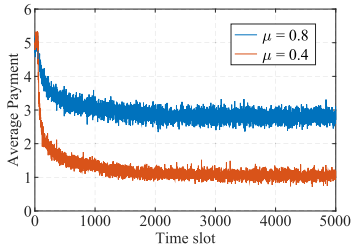


Fig. 5. Evolution of average payment of EDOs with different values of μ.

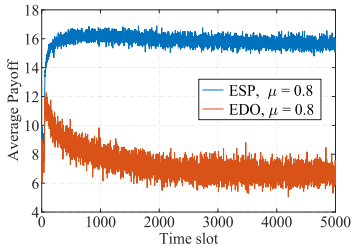


Fig. 6. Evolution of average payoff of ESP and EDOs.

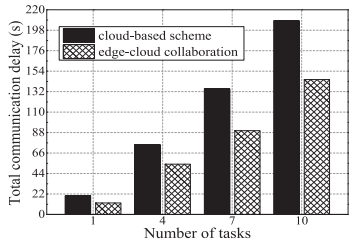


Fig. 7. Total communication delay versus number of tasks in two schemes.

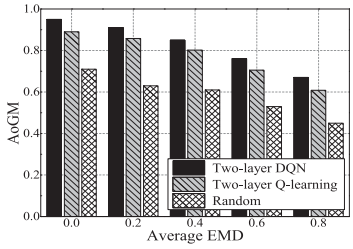


Fig. 8. AoGM of ESP versus average EMD in three schemes with the existence of free-riding EDOs.

2) *Free-riding attack.* Via quality estimation of local model updates in the two-layer MDP process, the zero-payment strategy is utilized to punish free-riding EDOs that cheat the ESP by delivering fake, meaningless, or redundant local model gradients to promote fairness in federated learning. Besides, Fig. 9 shows that our approach can mitigate the effect of the free-riding attack by attaining a desirable payoff for the ESP in the existence of free riders.

B. Simulation Setup

We consider a simulation scenario of federated-learning-based private data sharing in the smart grid with one aggregator,

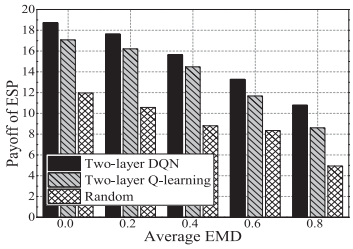


Fig. 9. Payoff of ESP versus average EMD in three schemes with the existence of free-riding EDOs.

TABLE III
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
ν_a	28	χ	0.5MB [22]
p_{cloud}	1.4	λ_c	0.002
ϖ	0.5	κ_p	8
η_i^1	0.26	η_i^2	0.04
$S_{i,j,k}^{\max}$	4000	$p_{j,k}^{\max}$	10
ς_B	32	η	0.01 [16]
W	12 [31]	D	1000
Z	11	A, B	3, 3

four ESPs, and 50 individual EDOs. Each ESP publishes the federated digit classification tasks trained by the CNN on the MNIST dataset, which includes 60 000 training samples and 10 000 testing samples in ten digit labels. Each EDO's training samples are compressed proportionally to its data quality $S_{i,j,k}$, which follows the uniform distribution in [0,1]. The number of global/local training epoches are set as $\xi_{j,k}^g = 10$ and $\xi_{i,j,k}^l = 5$. The EMD values of EDOs are varied by altering the number of digit labels and are uniformly distributed within [0,0.8]. The maximum waiting time in each global training epoch is fixed at $t_{j,k}^{\max} = 2.5$s. The uplink transmission rate between the cloud and EDO is within [0.5,10]Mb/s, while that between the aggregator and EDO is within [2,9]Mb/s. Two types of ESPs are considered, i.e., the accuracy-sensitive ESP with $\mu = 0.8$ and the cost-sensitive ESP with $\mu = 0.4$. Based on [22], we set $\lambda_1 = 0.361$, $\lambda_2 = 4.348$, $\lambda_3 = 0.001$, $\lambda_4 = 0.993$, $\lambda_5 = 0.178$, and $\lambda_6 = 1.743$. For the DQN, we set high discount factors (i.e., $\varphi_j = \varphi_i = 0.8$) to strive for the long-term high reward [29], and adopt the high greedy parameters (i.e., $\epsilon_j = 0.92$, $\epsilon_i = 0.95$) for fast exploitation [17]. Other parameters in the simulation are listed in Table III.

Two conventional schemes are compared to evaluate the performance of the proposed scheme as follows.

1) *Two-layer Q-learning scheme [29]:* The Q-learning algorithms are utilized by both ESPs and EDOs to find the optimal pricing and training strategies in federated learning, respectively. Here, the discount factors and greedy parameters remain unchanged.
2) *Random scheme:* Both ESPs and EDOs randomly choose their strategies during the repeated interactions, respectively.

C. Numerical Results

Fig. 3 compares the prediction accuracy of the CNN model on the MNIST dataset under FL and centralized AIoT paradigms with different participation levels and communication rounds. From Fig. 4, we can observe that the global model trained under the FL paradigm can coverage after about 30 communication rounds (i.e., global training epochs). Besides, compared with the 99.78% prediction accuracy in the centralized AIoT approach, our FL-based AIoT approach can attain 95.8% accuracy under full participation at 100th communication round. It indicates that our FL approach only suffers a small accuracy loss in exchange for users' privacy preservation in model training. Moreover, as seen in Fig. 4, the prediction accuracy in the FL approach under 40% participation at 100th communication round is 91.75%, which is smaller than that under full participation. The reason is that the smaller participation level, the fewer training data samples, and correspondingly, the relatively lower model accuracy.

Figs. 4 and 5 depict the evolutions of the AoGM of the ESP and the average payment of EDOs in the DQN over time, respectively. As seen from the two figures, the accuracy-sensitive ESP (i.e., $\mu = 0.8$) attaches more importance to the model accuracy and prefers higher payment to motivate EDOs' higher quality of local model updates, which results in a higher AoGM. Besides, the cost-sensitive ESP (i.e., $\mu = 0.4$) cares more about the payments to EDOs, and will require less data or lower data quality from EDOs to reduce its payment, thereby resulting in a decline of the AoGM. In addition, from these two figures, the proposed two-layer DQN algorithm can converge after about 2000 time slots.

Fig. 6 shows the evolutions of the average payoff of the ESP and EDO in the DQN when the time slot changes, where the ESP type is set as $\mu = 0.8$. As seen from Fig. 6, both the average payoffs of the ESP and EDO converge after about 2000 time slots. In Fig. 6, after observing the high model quality of EDOs motivated by the initial high payment, the ESP intends to decrease its payment to increase its utility for the future high reward. Meanwhile, with the decline of the payment, the EDO tends to decrease its payoff to seek high long-term reward.

Fig. 7 shows the total communication delay in two schemes when the number of tasks varies from 1 to 10. In the conventional cloud-based scheme, the global/local model updates are directly exchanged between the cloud and EDOs in federated learning, where the aggregators are not considered. From Fig. 7, we can see that our proposal with edge-cloud collaboration can efficiently reduce the communication delay given different number of tasks. The reason is that the aggregators in the edge plane can improve the network connectivity and capacity to reduce the effect of unreliable and intermittent wireless connections between users and the cloud. Meanwhile, the aggregators can operate cooperatively with the cloud in federated learning via edge aggregation operations. Thereby, a higher communication efficiency can be achieved under our edge-cloud collaborative framework.

Figs. 8 and 9 compare the proposed scheme with two conventional schemes in terms of the AoGM and the average payoff of ESP, respectively, when the average EMD $\overline{\Delta}_{j,k}$ varies from 0 to 0.8. In both simulations, the ratio of free-riding EDOs is set as $\zeta \cdot \overline{\Delta}_{j,k}$. We set $\mu = 0.8$ and $\zeta = 0.25$. Here, $\overline{\Delta}_{j,k} = 0$ indicates the IID distribution of data. From Fig. 8, we can observe that the AoGM decreases with the increase

of the average EMD in three schemes, which is in accordance with the theoretical analysis in (8). Especially, when $\overline{\Delta}_{j,k} = 0$, the highest AoGM and the highest payoff of the ESP can be attained in three schemes compared with non-IID distributions. Besides, as seen from these two figures, the proposed scheme outperforms the other two schemes in achieving a higher AoGM and an improved average payoff of the ESP. It is because in the two-layer Q-learning scheme, due to the large state-space and the corresponding curse of dimensionality, the ESP cannot efficiently seek its optimal pricing strategy to stimulate EDOs' high AoLM contributions, thereby causing the relatively lower AoGM of users and smaller payoff of the ESP. Besides, in the random scheme, as the strategies of both ESPs and EDOs are determined at random, they cannot attain the maximized payoffs as well as the optimal AoGM. In the proposed scheme, the cheating behaviors of free riders can be efficiently mitigated through the zero-payment mechanism to ensure the high AoGM and ESP's payoff. Additionally, via state-space compression and Q-function estimation in the DQN, both EDOs and ESPs can efficiently acquire their optimal training and pricing strategies for maximizing long-term payoffs in the highly dynamic network.

VII. Conclusion

In this article, we have proposed a novel federated-learning-enabled AIoT solution for secure and efficient personal energy data sharing in smart grids with edge-cloud collaboration. First, an edge-cloud integrated federated learning framework was presented for privacy-preserving and communication-efficient energy data analysis in smart grids. Second, with consideration of the non-IID distribution of heterogeneous users, a local data evaluation model for cost modeling and two optimization problems for ESPs and EDOs were developed under the proposed framework. Third, a DRL-based incentive algorithm with two layers was devised in the presence of multidimensional user private information and the large state space to search the optimal pricing strategies of EDPs and optimal training strategies of EDOs. At last, extensive simulations had demonstrated that the proposed scheme can efficiently motivate EDOs' high-quality local model sharing, improve the ESP's payoff, and reduce task latencies. For the future work, the blockchain-based reliable federated learning and the local model evaluation model under DP-based gradient perturbation in AIoT will be investigated.

References

[1] Y. Wang, Z. Su, Q. Xu, T. Yang, and N. Zhang, "A novel charging scheme for electric vehicles with smart communities in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8487–8501, Sep. 2019.

[2] X. Zhang, D. Biagioni, M. Cai, P. Graf, and S. Rahman, "An edge-cloud integrated solution for buildings demand response using reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 420–431, Jan. 2021.

[3] A. Jiang, H. Wei, J. Deng, and H. Qin, "Cloud-edge cooperative model and closed-loop control strategy for the price response of large-scale air conditioners considering data packet dropouts," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4201–4211, Sep. 2020.

[4] Y. Wang, Z. Su, and N. Zhang, "BSIS: Blockchain-based secure incentive scheme for energy delivery in vehicular energy network," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3620–3631, Jun. 2019.

[5] L. Ruan, Y. Yan, S. Guo, F. Wen, and X. Qiu, "Priority-based residential energy management with collaborative edge and cloud computing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1848–1857, Mar. 2020.

[6] C. Feng, Y. Wang, K. Zheng, and Q. Chen, "Smart meter data-driven customizing price design for retailers," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2043–2054, May 2020.

[7] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 7789–7817, May 2021.

[8] Y. Wang *et al.*, "SPDS: A secure and auditable private data sharing scheme for smart grid based on blockchain and smart contract," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2020.3040171.

[9] S. Zhao *et al.*, "Smart and practical privacy-preserving data aggregation for fog-based smart grids," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 521–536, 2020.

[10] Y. Liu, W. Guo, C. Fan, L. Chang, and C. Cheng, "A practical privacy-preserving data aggregation (3PDA) scheme for smart grid," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1767–1774, Mar. 2019.

[11] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 1387–1395.

[12] S. R. Pandey, N. H. Tran, M. Bennis, Y. K. Tun, A. Manzoor, and C. S. Hong, "A crowdsourcing framework for on-device federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3241–3256, May 2020.

[13] A. Taik and S. Cherkaoui, "Electrical load forecasting using edge computing and federated learning," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

[14] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," *in Proc. Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[15] N. Ding, Z. Fang, and J. Huang, "Optimal contract design for efficient federated learning with multi-dimensional private information," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 186–200, Jan. 2021.

[16] J. Kang, Z. Xiong, D. Niyato, S. Xie, and J. Zhang, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10700–10714, Dec. 2019.

[17] Y. Wang, Z. Su, N. Zhang, and A. Benslimane, "Learning in the air: Secure federated learning for UAV-assisted crowdsensing," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1055–1069, Apr.–Jun. 2021.

[18] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, *arXiv:1806.00582*.

[19] Y. Zhan, P. Li, Z. Qu, D. Zeng, and S. Guo, "A learning-based incentive mechanism for federated learning," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6360–6368, Jul. 2020.

[20] J. Liu, J. Weng, A. Yang, Y. Chen, and X. Lin, "Enabling efficient and privacy-preserving aggregation communication and function query for fog computing-based smart grid," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 247–257, Jan. 2020.

[21] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular Internet of Things: Recent advances and open issues," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 45–61, 2020.

[22] Y. Jiao, P. Wang, D. Niyato, B. Lin, and D. I. Kim, "Toward an automated auction framework for wireless federated learning services market," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2020.2994639.

[23] W. Y. B. Lim *et al.*, "Hierarchical incentive mechanism design for federated machine learning in mobile networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9575–9588, Oct. 2020.

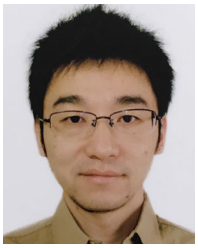
[24] H. Yu *et al.*, "A sustainable incentive scheme for federated learning," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 58–69, Jul./Aug. 2020.

[25] Y. Sarikaya and O. Ercetin, "Motivating workers in federated learning: A Stackelberg game perspective," *IEEE Netw. Lett.*, vol. 2, no. 1, pp. 23–27, Mar. 2020.

[26] Y. Jiao, P. Wang, S. Feng, and D. Niyato, "Profit maximization mechanism and data management for data analytics services," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2001–2014, Jun. 2018.

[27] Q. Liu, S. Huang, J. Opadere, and T. Han, "An edge network orchestrator for mobile augmented reality," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 756–764.

[28] Z. Su, Y. Wang, Q. Xu, and N. Zhang, "LVBS: Lightweight vehicular blockchain for secure data sharing in disaster rescue," *IEEE Trans. Dependable Secure Comput.*, to be published, doi: 10.1109/TDSC.2020.2980255.
[29] Q. Xu, Z. Su, and R. Lu, "Game theory and reinforcement learning based secure edge caching in mobile social networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3415–3429, 2020.
[30] X. Chen *et al.*, "Multi-tenant cross-slice resource orchestration: A deep reinforcement learning approach," 2018, *arXiv:1807.09350*.
[31] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, "A secure mobile crowdsensing game with deep reinforcement learning," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 1, pp. 35–47, Jan. 2018.
[32] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.
[33] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," 2017, *arXiv:1712.07557*.
[34] K. Bonawitz *et al.*, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 1175–1191.



Zhou Su (Senior Member, IEEE) research interests include wireless networking, mobile computing, and network security.

Mr. Su is an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL, the IEEE OPEN JOURNAL OF COMPUTER SOCIETY, *IET Communications*, etc. He is the Chair of the Multimedia Services and Applications over Emerging Networks Interest Group of the IEEE Comsoc Society and the Multimedia Communications Technical Committee. He was the recipient of the Best Paper Award of the IEEE International Conference on Communications 2020, the 13th IEEE International Conference on Big Data Science and Engineering 2019, the IEEE Cyber Science and Technology Congress, etc.

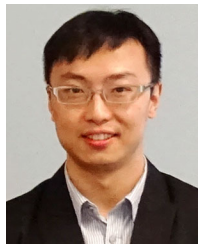
Yuntao Wang is currently working toward the Ph.D. degree with the School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an, China.

His research interests include security and privacy in wireless network architecture and vehicular networks.

Tom H. Luan (Senior Member, IEEE) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2004, the M.Phil. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2007, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2012.

He is currently a Professor with the School of Cyber Engineering, Xidian University, Xi'an. He has authored or coauthored more than 40 journal articles and 30 technical articles in conference proceedings. His research interests include content distribution and media streaming in vehicular ad hoc networks, peer-to-peer networking, and the protocol design and performance evaluation of wireless cloud computing and edge computing.

Dr. Luan was the recipient of one U.S. patent. He was a TPC Member of the IEEE Global Communications Conference, the IEEE International Conference on Communications, and the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, and the Technical Reviewer for multiple IEEE Transactions, including the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.



Ning Zhang (Senior Member, IEEE) received the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015.

He was a Postdoc Research Fellow with the University of Waterloo and the University of Toronto, Toronto, ON. He is currently an Associate Professor with the University of Windsor, Windsor, ON.

Dr. Zhang is an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL, the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, the IEEE ACCESS, and the *IET Communications, and Vehicular Communications*, and a Guest Editor for several international journals, such as the IEEE WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He has also been a Track Chair for several international conferences and a Co-Chair of several international workshops. He was the recipient of the Best Paper Awards from the IEEE Globecom in 2014, the IEEE International Conference on Wireless Communications and Signal Processing in 2015, the *Journal of Communications and Information Networks* in 2018, the IEEE International Conference on Communications in 2019, the IEEE Technical Committee on Transmission Access and Optical Systems in 2019, and the IEEE/CIC International Conference on Communications in China in 2019.

Feng Li (Member, IEEE) received the M.S. degree from North China Electric Power University, Beijing, China, in 2010.

Since 2010, he has been with Electric Power Research Institute, State Grid Xinjiang Electric Power Company Ltd., Wulumuqi, China, working on information security related. He has authored or coauthored more than 20 papers. His research interests include information security, industrial control security, and data security.

Tao Chen received the master's degree from the North China Electric Power University, Beijing, China, in 2009.

He is currently a Senior Engineer with State Grid Xinjiang Electric Power Company Ltd., Wulumuqi, China. His research interests include network security, Internet-of-Things security, and data security.

Hui Cao (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 2000, 2004, and 2009, respectively.

He is currently a Professor with the School of Electrical Engineering, Xian Jiaotong University. From 2014 to 2015, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. He has authored or coauthored more than 50 scientific and technical papers in recent years. His research interests include knowledge representation and discovery.

Dr. Cao was the recipient of the Second Prize of the National Technical Invention Award.