A Privacy-Enhanced Multiarea Task Allocation Strategy for Healthcare 4.0

Xiaoding Wang, Mengyao Peng, Hui Lin, Yulei Wu, *Senior Member, IEEE*, and Xinmin Fan

***Abstract*—The continuous development of Healthcare 4.0 has brought great convenience to people. Through the Internet of Things technology, doctors can analyze patients' health data and make timely diagnosis. However, behind the high efficiency, the mobile crowdsensing technology used for data transmission still has the risk of leaking the privacy of task and patient information. To this end, this article proposes a privacy-enhanced multi-area task assignment strategy, named PMTA. Specifically, we use deep differential privacy to add noise to patient data, and then put the noise-added dataset into a deep Q-network for training, combined with a spectral clustering algorithm, to obtain an optimal classification strategy. Further, in order to address the problem of data silos, we adopt federated learning to jointly train the classification models of different hospitals to obtain a global model and realize data sharing among different hospitals. Finally, we use the optimal classification of patients for task deployment on the blockchain, and limit patients to only apply for tasks of the corresponding level through the smart contract technology, so as to protect task privacy. Experimental results show that our strategy can not only effectively protect task and patient privacy, but also achieve better system performance.**

***Index Terms*—Blockchain, deep differential privacy, federated learning, Healthcare 4.0, Internet of Things (IoT), mobile crowdsensing.**

I. Introduction

WITH the rapid development and innovation of Industry 4.0 technologies, the entire world is transitioning toward digital, fully automated, and cyber-physical systems [1]. Emerging technologies in Industry 4.0, including the Internet of Things (IoT), big data analysis, blockchain, cloud computing, and artificial intelligence, have been implemented in various other fields [2]. They have now been applied to the medical and health sector, making revolutionary changes to the field, and effectively promoting the emergence and development of Healthcare 4.0 [3]. The IoT technology is one of the most adopted technologies that promote the frontier development of Healthcare 4.0. It brings many key benefits, including providing efficient technical support for patients with chronic diseases, the elderly, and patients who need long-term health monitoring [4], [5]. The main purpose of the IoT technology in the field of digital health is to provide each participant with highly personalized, affordable, accessible, and timely Healthcare 4.0 services.

While the IoT technology brings high efficiency and convenience to Healthcare 4.0, it also faces some challenges. Among them, the mobile crowdsensing technology, as a commonly used computing mode in the IoT, can collect relevant information of participants scattered in various places [6]. In this process, it is inevitable that both the data collector and the data provider will face the problem of data privacy leakage [7]. Since there has been tremendous work devoted to solving the problem of privacy leakage of data providers, this article will focus more on protecting the privacy information of tasks published by data collectors [8]–[11]. This means that different hospitals cannot directly exchange medical data through data sharing, even for better treatment development. Thereby, the problem of data silos arises [12]. The challenge mentioned previously is the problem that need to be solved urgently in the field of Healthcare 4.0.

Based on the abovementioned challenges, this article proposes an effective privacy-enhanced multi-area task assignment strategy (PMTA) for Healthcare 4.0. The strategy mainly includes three important modules, namely data privacy protection module, data provider classification module, and smart contract design module. The first module provides privacy protection for data providers (i.e., patients). In this module, in order to prevent the sensitive personal information of data providers from being recovered by malicious attackers, a deep differential privacy protection method based on deep convolutional generative adversarial networks (DCGAN) [13] was proposed. The second module provides privacy protection for task data of data collectors (i.e., doctors). In this module, in order to protect the privacy information in tasks published by data collectors, tasks and data providers are classified by different machine learning techniques [14]. The last module implements the publishing and storage of data collection tasks. In this module, different levels of data providers request deployed tasks from the blockchain [15].

Manuscript received 26 December 2021; revised 10 March 2022 and 1 May 2022; accepted 29 June 2022. Date of publication 8 July 2022; date of current version 3 March 2023. This work was supported by the National Natural Science Foundation of China under Grant U1905211 and Grant 61702103 and in part by the Natural Science Foundation of Fujian Province under Grant 2020J01167 and Grant 2020J01169. Paper no. TII-21-5739. *(Corresponding authors: Hui Lin; Yulei Wu.)*

Xiaoding Wang, Mengyao Peng, and Hui Lin are with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China, and also with the Engineering Research Center of Cyber Security and Education Informatization, Fujian Province University, Fuzhou, Fujian 350117, China (e-mail: wangdin1982@fjnu.edu.cn; 18875857995@163.com; linhui@fjnu.edu.cn).

Yulei Wu is with the Department of Computer Science, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, EX4 4QF Exeter, U.K. (e-mail: y.l.wu@exeter.ac.uk).

Xinmin Fan is with the Network and Data Center, Fujian Normal University, Fuzhou 350117, China (e-mail: xmfan@fjnu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3189439.

Digital Object Identifier 10.1109/TII.2022.3189439

We then summarize the main contributions of this article as follows.

1) In order to protect the privacy of patients, i.e., to prevent their sensitive information from being stolen or leaked during data acquisition tasks, we propose a deep differential privacy protection method based on DCGAN. Specifically, a deep neural network is trained by adding Gaussian noise to the gradients of the parameters to provide differential privacy protection to patient data. Then, these data are generated through the generative adversarial mechanism of DCGAN to provide further privacy protection for patient data.
2) In order to protect the privacy of the task, that is, to prevent the patient from maliciously obtaining the privacy information contained in the data collection task issued by the doctor through collusion and other methods, we propose a patient classification method based on deep Q-learning, spectral clustering and federated learning. Specifically, considering the correlation between the basic information of patients, deep Q-learning (DQN) combined with the RatioCut algorithm in spectral clustering is used to classify patients, and assign a level to each patient according to the potential influence factor (PIF) w.r.t. information dissemination. Finally, through federated learning, a global classification model is generated for the areas where each hospital is located.
3) In order to protect task privacy and achieve efficient task deployment, we design a smart contract to limit patients' application to carry out tasks. Specifically, tasks are deployed on the blockchain, and in smart contracts, we set that patients can only apply for tasks corresponding to their individual levels, thereby greatly reducing the possibility of collusion between patients.
4) Extensive experiments are conducted on a real-world dataset MIMIC-IV. Experimental results and analysis show that the proposed strategy PMTA can not only prevent the patient's privacy information from being leaked or maliciously stolen, but also protect the privacy of task data. In addition, the blockchain-based task deployment system adopted by the strategy has excellent performance in terms of throughput and latency.

The rest of this article is organized as follows. Related work is presented in Section II. The system model is introduced in Section III. The implementation details of the proposed strategy PMTA is elaborated in Section IV. Experimental results and analysis are given in Section V. Finally, Section VI concludes this article.

II. Related Work

Privacy and security issues in healthcare 4.0 have attracted significant attentions, and many research works have been proposed.

Elmisery *et al.* [16] enhanced user privacy by utilizing end-user personal gateways as intermediate fog nodes between IoT devices and cloud healthcare services. To protect the patient's electronic health records, Hathaliya *et al.* [17] developed a biometric-based authentication and key agreement scheme against known and unknown attacks. Considering the security, privacy, and interoperability problems of existing telesurgery systems, Gupta *et al.* [18] combined the immutability and interoperability of smart contracts, and proposed a blockchain-based, safe and flawless, interoperable telesurgery framework. Then, aiming at the problems of safety, reliability, delay, and storage cost of existing drone systems, Gupta *et al.* [19] proposed an outdoor medical supply mechanism based on the Ethereum blockchain using drones. The mechanism provides reliable communication between drones and entities, ensuring the early delivery of needed medical supplies to critically ill patients. In order to protect data security and privacy under the condition that both the transmission medium (such as cloud server) and the key are compromised, Qiu *et al.* [20] designed a secure data storage by combining selective encryption algorithms with fragmentation and dispersion. Gupta *et al.* [21] addressed security and privacy issues through Ethereum smart contracts, and published storage costs through the InterPlanetary File System. In addition, they demonstrated a real-time smart contract written in Solidity and deployed in the Truffle suite, and tested for security vulnerabilities in the MyThril open-source tool. To improve data accessibility among healthcare providers, Tanwar *et al.* [22] developed an access control policy algorithm. The algorithm used the concept of chain code to realize the sharing of electronic medical records based on Hyperledger by simulating the environment. Bhattacharya *et al.* [23] built a blockchain-based deep learning-as-a-service framework for sharing EHR records among multiple healthcare users.

These state-of-the-art strategies provided efficient and secure solutions to the privacy issues in Healthcare 4.0. They used various technologies to protect patient data, either on the user side, or on the medical database side, or by controlling access rights. Compared with these strategies, this article proposes an effective PMTA strategy, named PMTA, for Healthcare 4.0, which can protect the privacy of both patients and tasks.

III. System Model

The PMTA system model proposed in this article is shown in Fig. 1. To address the problem of data silos, we construct a global model across different hospitals through federated learning. Among them, data silos mean that each organization has its own data, and the data between different organizations are often stored and customized independently, so the data of each organization are like an island, unable to connect and interact with other data. Federated learning is essentially a distributed machine learning technology, and its goal is to achieve global modeling and improve model effects on the basis of ensuring data privacy security and legal compliance [24]. Therefore, using federated learning not only protects the privacy of patient data in various hospitals, but also enables data sharing among different hospitals. In this system model, we treat each hospital as an independent area, data collectors usually refer to the doctors in the hospital, and data providers refer to the patients who participate in the data collection task.

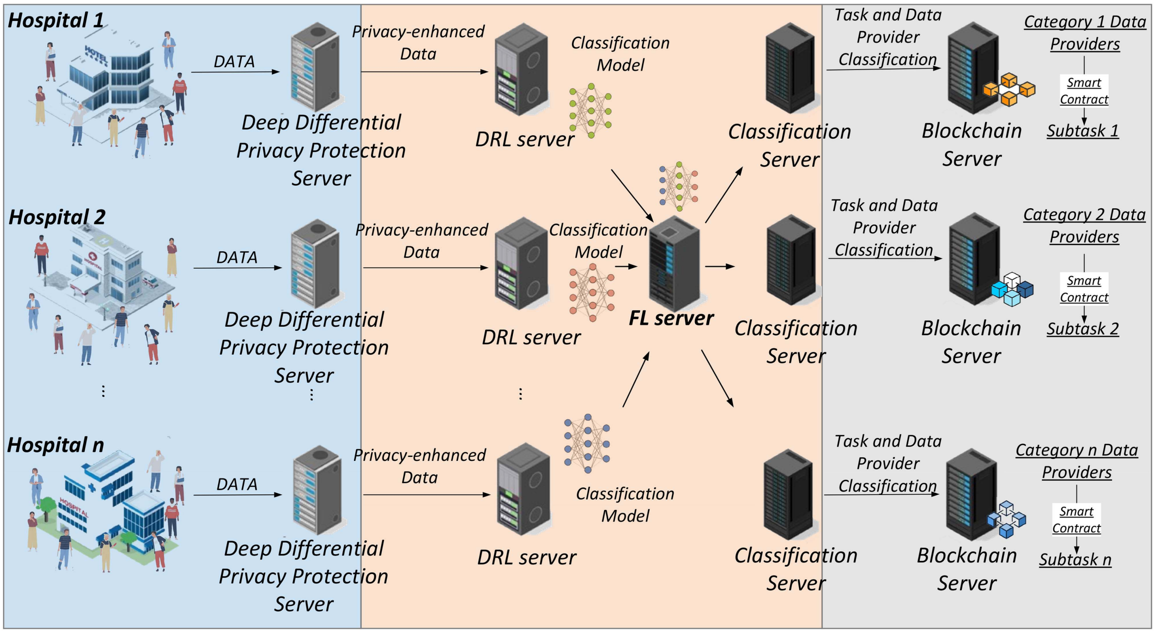


Fig. 1. System model of PMTA.

Specifically, in order to prevent the patient data from being subjected to inference attacks and other attacks that can threaten patient privacy during the model training process, the data of patients participating in data collection are sent to the server with noise addition. This process adopts DCGAN-based deep differential privacy protection algorithm. After adding noise to the patient data, an optimal classification strategy of the data provider for each area is calculated by the deep reinforcement learning technology [25], which is implemented using the DQN algorithm [26]. After the training of the classification model of data providers of each area is completed, these models are combined to form a global model through the federated averaging algorithm [27].

In order to prevent malicious data providers from stealing the privacy in the tasks issued by doctors through collusion attacks, data providers will be classified by the classification server, and on this basis, tasks will also be classified. This is because task classification can prevent malicious data providers from obtaining complete information in tasks at one time, thereby preventing direct leakage of task privacy. In contrast, data provider classification is to prevent malicious data providers from piecing together the complete task information through collusion, thereby preventing indirect leakage of task privacy.

After completing the classification of tasks and data providers, in order to ensure that data providers of different levels can only select tasks of the corresponding level, data collectors publish tasks in the blockchain and use the smart contract technology to limit the task selection of data providers. The smart contract designed on the basis of the classification of tasks and data providers can effectively prevent malicious data providers from stealing private information in tasks through collusion attacks.

Therefore, PMTA not only improves the privacy protection of patient data through the deep differential privacy technology and federated learning technology, but also ensures that sensitive information in the tasks issued by the hospital is not stolen by implementing data provider and task classification on the blockchain.

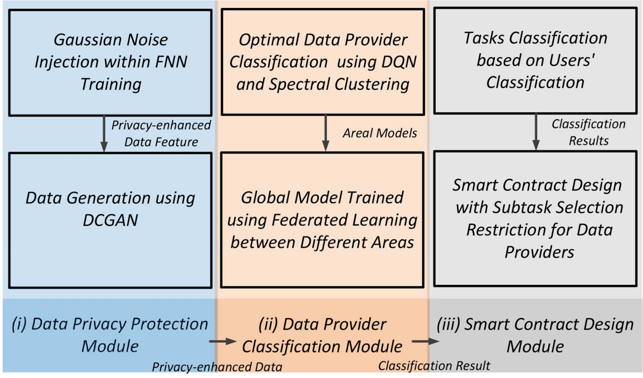

Gaussian Noise Injection within FNN Training
Privacy-enhanced Data Feature
Data Generation using DCGAN
Optimal Data Provider Classification using DQN and Spectral Clustering
Areal Models
Global Model Trained using Federated Learning between Different Areas
Tasks Classification based on Users' Classification
Classification Results
Smart Contract Design with Subtask Selection Restriction for Data Providers
(i) Data Privacy Protection Module
(ii) Data Provider Classification Module
(iii) Smart Contract Design Module
Privacy-enhanced Data
Classification Result

Fig. 2. Flowchart of PMTA.

IV. Implementation of the PMTA

The proposed PMTA consists of three important modules, which are as follows:

1) data privacy protection module,
2) data provider classification module, and
3) smart contract design module (see Fig. 2).

A. Data Privacy Protection

In order to prevent attackers from using GAN to restore the data in the training dataset during the application of the deep learning model, and to protect the sensitive information of data providers in the training dataset, we proposes a DCGAN-based differential privacy protection method.

First, we explain why DCGAN is employed in data privacy protection. GAN is mainly composed of a generator G and a discriminator D. G and D are two sides of the game, the generator G captures the distribution of the sample data, and the discriminator is a binary classification, which is used to judge whether the input result comes from the training data (rather than the generated data). Therefore, in the GAN training process, the goal of the generator G is to generate results close to the original data to fool the discriminator D, and the goal of D is to distinguish the results generated by G from the real data as much as possible. Since G and D are difficult to balance, GAN optimizes the following objective function:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_{\text{data}}(x)}[\log_2 D(x)] + E_{z\sim P_z(z)}[\log_2(1-D(G(z)))] \quad (1)$$

where P_z is the input noise distribution of G, and P_{data} is the actual data distribution.

However, if G and D are not well balanced, G may eventually collapse to a saddle point. To avoid this shortcoming, Radford *et al.* [13] proposed DCGAN. This method first uses batch normalization to solve the initialization problem, then removes the fully connected layer to improve the convergence speed, and finally uses strided convolution and fractionally strided convolution, instead of pooling layers, reducing spatial sampling; then, the training process becomes stabilized. Thereby, we adopt DCGAN combined with the deep differential privacy mechanism to provide privacy protection for data providers.

1) Deep Differential Privacy Implementation: Providing differential privacy protection to data providers can be achieved by adding differential privacy noise to the process of minimizing the parameter loss function of a deep learning model, i.e., feedforward neural network (FNN), using stochastic gradient descent with the following steps.

1) Use the stochastic gradient descent algorithm to randomly select a number of training samples S and calculate the gradient of each sample, i.e., $g_t(x_i) \leftarrow \nabla_{\theta_t} L(\theta_t, x_i)$.
2) Check whether $g_t(x_i)$ is lower than the threshold C; if not, adjust $g_t(x_i)$ by $\bar{g}_t(x_i) \leftarrow g_t(x_i)/\max(1, \|g_t(x_i)\|_2/C)$.
3) Add Gaussian noise to the gradient, i.e., $\bar{g}_t(x_i) \leftarrow \frac{1}{S}\sum_{i=1}\bar{g}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I)$, and then update the parameters by $\theta_{t+1} \leftarrow \theta_t - \eta\bar{g}_t$, where η represents the learning rate.

The abovementioned steps are iteratively executed T rounds. Abadi *et al.* [28] suggested that the training of the deep neural network can be based on lot. Thereby, we first calculate the gradient value of each batch. Then, we randomly select a group of batches to form a lot, calculate its gradients, and add noise to it, where each lot follows an independent distribution with probability $q = L/N$ and N denotes the size of the input dataset. Finally, we calculate the average gradient of lot for parameter update. Each iteration of the algorithm consists of N/L computations of lots.

2) Privacy Loss Calculation: Although the algorithm can provide differential privacy protection for data, it still brings privacy loss, and this value reflects the privacy protection effect of the deep learning model. Therefore, we need to calculate the privacy loss throughout the training process. The privacy loss is defined as follows. For two adjacent datasets $d, d' \in D$ and the mapping mechanism M, introducing an auxiliary input variable aux and an output $o \in R$, the privacy loss of the mapping mechanism M at the output o, denoted by $c(o; M, \text{aux}, d, d')$, is defined as

$$c(o; M, \text{aux}, d, d') \triangleq \log\frac{Pr[M(\text{aux}, d) = o]}{Pr[M(\text{aux}, d') = o]}. \quad (2)$$

The training process usually requires the use of gradient descent for many times, resulting in the accumulation of the privacy budget. According to the composability of differential privacy, the moments accountant method can be used to minimize the privacy loss. Note that the parameters of each layer of the neural network are closely related to the differential privacy mechanism in each iteration, so for a given mapping mechanism M, the privacy loss in the λth iteration, denoted by $\alpha_M(\lambda; \text{aux}, d, d')$, is defined as

$$\alpha_M(\lambda; \text{aux}, d, d') \triangleq \log E_{o\sim M(\text{aux},d)}[\exp(\lambda c(o; M, \text{aux}, d, d'))]. \quad (3)$$

Further, the privacy loss boundary value of the mapping mechanism M, denoted by $\alpha_M(\lambda)$, is defined as

$$\alpha_M(\lambda) \triangleq \max_{\text{aux},d,d'} \alpha_M(\lambda; \text{aux}, d, d'). \quad (4)$$

It has been proved that $\alpha_M(\lambda)$ satisfies the following properties: 1) Given a mechanism M consisting of a set of submechanisms $M_1, M_2, \ldots, M_k$, satisfying $M_i : \prod_{j=1}^{i-1} R_j \times D \rightarrow R_i$, the privacy loss bound satisfies $\alpha_M(\lambda) \leq \sum_{i=1}^{k}\alpha_{M_i}(\lambda)$; 2) $\forall\epsilon > 0$, the mapping mechanism M is (ϵ, δ)-differentially private if and only if $\delta = \min\exp(\alpha_M(\lambda) - \lambda\epsilon)$ [28]. The abovementioned two properties determine the privacy loss of each iteration of the deep neural network algorithm and the maximum number of iterations that can achieve the tolerance of data privacy violation. In particular, in the case of adding Gaussian noise, let μ_0 and μ_1 be the probability density functions of $N(0, \sigma^2)$ and $N(1, \sigma^2)$, respectively, and μ be the mixture probability density function of these two. That is, $\mu = (1-q)\mu_0 + q\mu_1$. It can be deduced that $\alpha(\lambda) = \log\max(E_1, E_2)$, where $E_1 = E_{z\sim\mu_0}[(\mu_0(z)/\mu(z))^\lambda]$, $E_2 = E_{z\sim\mu}[(\mu(z)/\mu_0(z))^\lambda]$, such that the privacy loss boundary equals to $q^2\lambda(\lambda+1)/(1-q)\sigma^2 + O(q^3/\sigma^3)$. That suggests the proposed DCGAN-based method is (ϵ, δ)-differentially private for any δ and $\epsilon < c_1 q^2 T$, if we choose

$$\sigma \geq c_2\frac{q\sqrt{T\log(1/\delta)}}{\epsilon} \quad (5)$$

where c_1 and c_2 are constants.

To sum up, in FNN training, the privacy budget of the deep network is calculated, and Gaussian noise is added to the stochastic gradient descent to minimize the overall privacy budget. Then, based on the data feature processed by the differential privacy empowered FNN, DCGAN is used to generate privacy protected data (see Fig. 3).

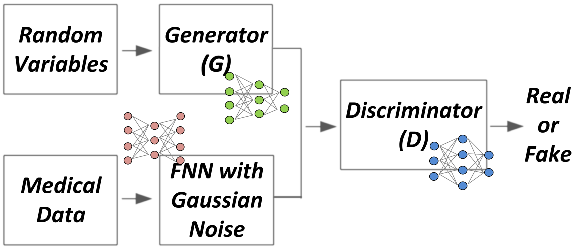


Fig. 3. The framework of data privacy protection based on DCGAN.

B. Data Provider Classification

1) Spectral Clustering-Based Data Provider Classification: We implement the classification of each areal data provider through a spectral clustering algorithm. To be specific, we treat all data providers in the area as nodes of the network, so that we can classify data providers according to their attributes. First, we construct a network $\mathcal{G}$ of potential relationships for each area. In $\mathcal{G}$, each node represents a data provider in the area, while each edge is associated with the similarity value $f_{\text{sim}}(.,.)$ between the end points of that edge, i.e., for edge $v_i v_j$, we have

$$f_{\text{sim}}(v_i, v_j) = e^{-\frac{\| x_i - x_j \|^2}{2\sigma^2}}. \tag{6}$$

In order to ensure the balance between classes, that is, the number of nodes in each class needs to be roughly the same, so as to effectively avoid the impact of the large gap between the number of data providers in different categories [29]. To this end, we use the RatioCut algorithm to partition the graph $\mathcal{G}$. When the classification of data providers is completed, we define the (PIF w.r.t. information dissemination of each data provider v_i, considering node degree $\mathcal{D}(.)$, betweenness centrality $\mathcal{BC}(.)$, and local clustering coefficient $\mathcal{L}c(.)$ on v_i, as

$$\text{PIF}_i = \mathcal{D}(v) + \mathcal{L}c(v) + \mathcal{BC}(v). \tag{7}$$

According to PIF_i, each data provider v_i is given an initial level $\mathcal{L}_i$ for task application, i.e., $\mathcal{L}_i \propto \text{PIF}_i$. That is, the larger the PIF, the higher the level. However, with the execution of data collection tasks, the level of each data provider will change according to the degree of task completion and the degree of task privacy leakage as well. This means that malicious data providers are likely to be unable to apply for tasks due to their low level.

2) DQN-Based Optimal Classification Mechanism: In PMTA, we adopt a DQN-based method, which provides the optimal classification for data providers. Different from Q-learning, DQN uses deep neural networks instead of the traditional Q-table, where these deep neural networks can be represented by two action-value functions $Q(.)$ and $\hat{Q}(.)$ with parameters ω and ω^-, respectively. Due to the nonuniformity between high-dimensional state space and low-dimensional action space, both neural networks only use the state as input and the output is the Q-value of the action-value function about each possible action.

Based on the classification results obtained from the previous decision, we take the largest proportion of malicious data providers in each category as the current state s_j, and set the classification strategy as action a_j, and execute action a_j in state s_j will be rewarded r_j. Since the classification of data providers directly affects the execution of tasks, the reward should be calculated w.r.t. the privacy protection degree of the task and the task completion degree of the data provider. The reason for that is as follows. We argue that nonmalicious data providers can provide good privacy protection for each class, while malicious data providers are highly likely to sabotage the task. Therefore, the reward is given by

$$r = \alpha \prod_{k=1}^{n} (1 - P_k) + (1 - \alpha) \frac{\sum_{k=1}^{n} c_k (1 - P_k)}{N} \tag{8}$$

where α and $1 - \alpha$ represent the proportion of privacy protection degree and task completion degree in actual demands, respectively, n represents the number of classifications, P_k denotes the percentage of malicious data providers of the kth class, c_k represents the number of data providers of the kth class, and N is the total number of data providers.

To get optimal classification, we train the deep neural network using supervised learning with the target Q-value. Specifically, we perform a gradient descent with respect to the parameter ω according to the loss function $(y_j - Q(s_j, a_j; \omega))^2$, where

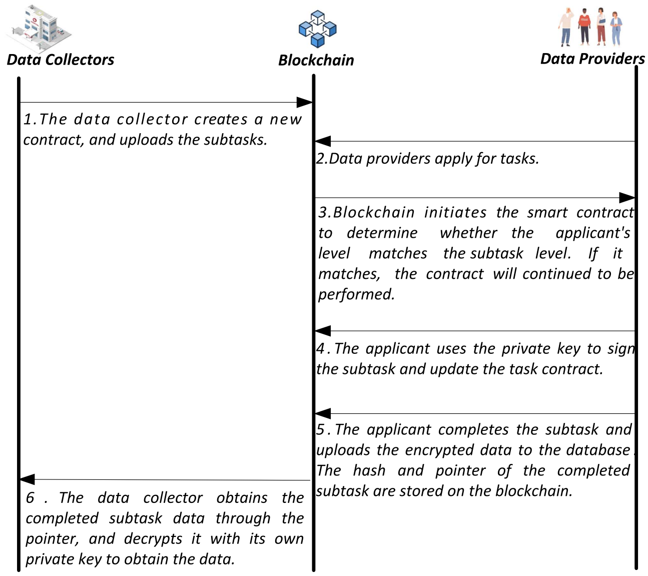
$$y_j = r_j + \gamma \text{max}_{a'} \hat{Q}(s_{j+1}, a'; \omega^-). \tag{9}$$

At every $\mathcal{T}$ time steps, we update ω^- by $\omega^- = \omega$. Then, the optimal data provider classification is obtained, when the DQN algorithm converges.

3) Model Training Based on the Federated Average Algorithm: Through the algorithm in the previous section, we achieve the optimal classification of each areal data provider, and then, we need to train these local submodels into a global model to solve the problem of data silos, which will help doctors to provide patients a more comprehensive diagnosis. Considering that federated learning can complete the training of a global model under the premise of ensuring data privacy [30], this article uses the federated average algorithm to generate a global DQN-based classification model. We assume that there are K participants (i.e., K areal data providers) joining in the federated learning with totally n samples, where the number of samples of the kth participant is denoted by n_k. Then, the federated average algorithm aggregates the parameters of each participant ω^k by weighted averaging to update the parameters of the global model $\tilde{\omega}$ in each training round t, i.e.,

$$\tilde{\omega}_{t+1} = \sum_{t=1}^{K} \frac{n_k}{n} \omega_{t+1}^{k} \tag{10}$$

where ω_{t+1}^{k} is learned by the kth participant according to the parameter of the global model $\tilde{\omega}_t$ obtained in the previous round. When the federated average algorithm converges, the global data provider classification model is obtained.



1.The data collector creates a new contract, and uploads the subtasks.

2.Data providers apply for tasks.

3.Blockchain initiates the smart contract to determine whether the applicant's level matches the subtask level. If it matches, the contract will continued to be performed.

4.The applicant uses the private key to sign the subtask and update the task contract.

5.The applicant completes the subtask and uploads the encrypted data to the database The hash and pointer of the completed subtask are stored on the blockchain.

6 . The data collector obtains the completed subtask data through the pointer, and decrypts it with its own private key to obtain the data.

Fig. 4. Smart contract execution process diagram.

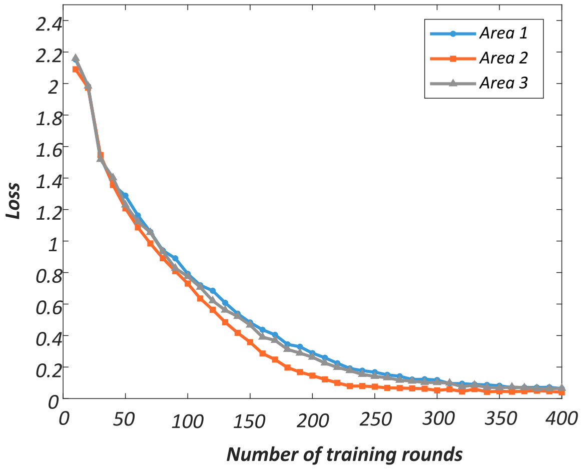


Fig. 5. Training loss for three different areas.

C. Smart Contract Design

After classifying data providers, data collectors classify relevant tasks as needed, and these tasks will be deployed on the blockchain. Similar to literature [31], we modify the block header and add a level attribute to it to ensure that only tasks of the same level can be stored in the block, and also limit the permissions of data providers to apply for tasks in the block.

Specifically, in the task publishing process, data collector inputs the corresponding task level as the attribute of the task, and publishes it to the block of the corresponding level for storage. Then, when data provider requests a task, the smart contract matches its level with the level of the requested task. If the match is successful, data provider can view the corresponding task content; otherwise, the request is rejected. The specific task request and access control process are shown in Fig. 4. As we have analyzed before, classification of tasks and data providers not only prevents data providers receiving tasks from obtaining complete task information at one time, but also prevents them from obtaining privacy information contained in tasks through collusion attacks.

V. Performance Evaluation

A. Experiment Setup

We use Python to verify the performance of the proposed PMTA on data provider classification on a computer with an Intel Core i7 processor with a frequency of 1.50 GHz, 16G running memory, and a 64-bit Windows 10 operating system. In addition, we run Hyperledger Fabric 1.2 in Ubuntu system with VMware Workstation 14 Pro, 4 GB RAM, two processors to evaluate the performance of the blockchain in PMTA.

This experiment is conducted using a real-world dataset MIMIC-IV.[1] As a relational database, MIMIC-IV contains comprehensive information, such as laboratory measurements, medications used, recorded vital signs, and more for each patient in hospitals at the tertiary academic medical center in Boston, MA, USA. This database is designed to support a variety of research in healthcare. In the experiments, we use the basic attributes of patients for classification of patients (data providers), asking patients to provide their treatment data to the doctors (data collectors) as data collection tasks. At the same time, we assume that there is a certain ratio of malicious patients who may provide incomplete treatment and may leak sensitive information of the task. The deep learning network model used in the experiment is an FNN with a depth of 3, the nodes in the hidden layer are 1000, the activation function is ReLU, and a softmax classifier with cross entropy.

In addition, we evaluate the performance of the proposed strategy PMTA by metrics, such as training loss, training accuracy, average blockchain latency, throughput, and CPU utilization. Generally speaking, the performance of a strategy can be reflected by the number of training rounds required to improve training accuracy and reduce training loss. At the same time, low latency, low CPU usage, and high throughput imply the optimization provided by the strategy for blockchain performance.

B. Experimental Results

1) Training Results of DQN and Federated Learning: We divided patient data into three different areas, i.e., Area 1, Area 2, and Area 3, to verify the performance of PMTA in data provider classification using DQN. Fig. 5 shows the loss value of the model training in each area. From the observation of the figure, we find that with the continuous increase in the number of training rounds, the corresponding loss value first drops rapidly and then gradually stabilizes, getting closer to 0. Although there are some differences among the data providers from three different areas, the overall trend is basically the same. In addition, it is clear that the algorithm is close to convergence around 400 rounds, this is because we use the RatioCut algorithm, which can balance class division, and the DQN algorithm, which can provide optimal classification, for data provider classification.

[1] [Online]. Available: https://physionet.org/content/mimiciv/0.4/

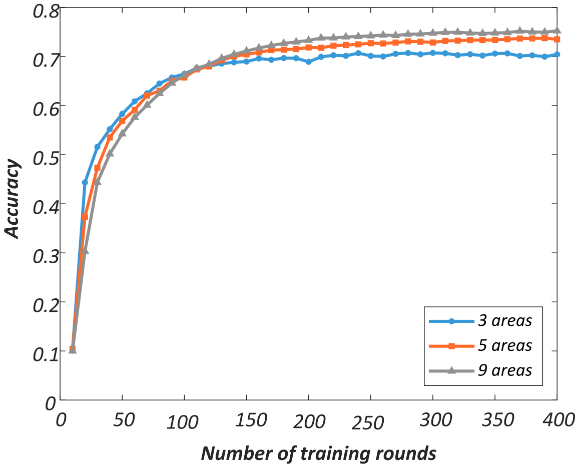
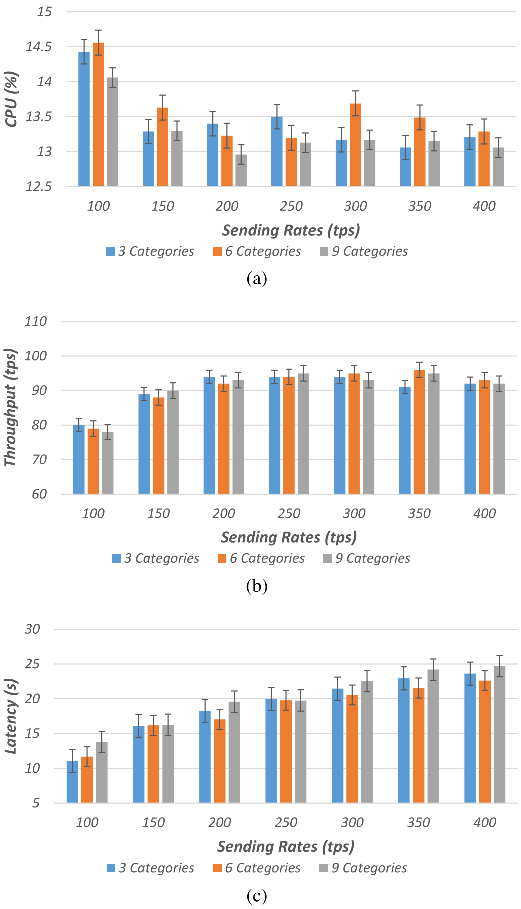


Fig. 6. Global model training accuracy.



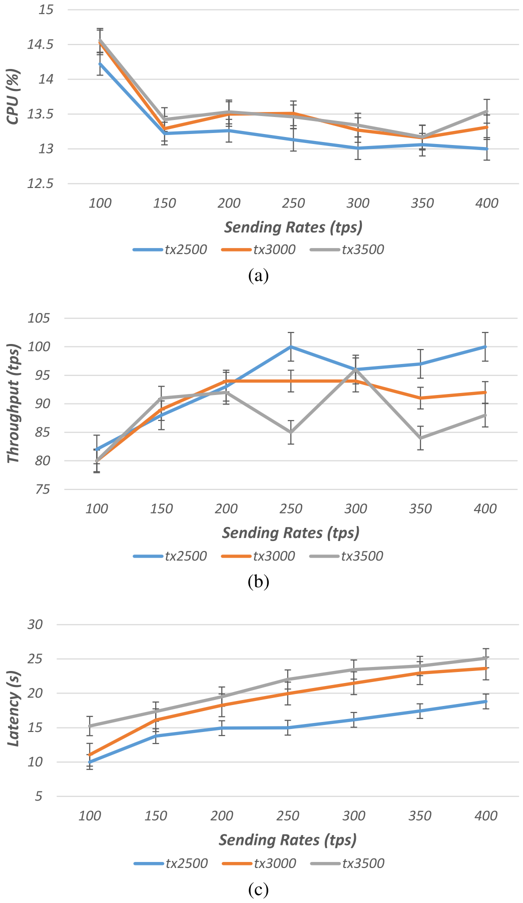
(a)

(b)

(c)

Fig. 7. Blockchain performance in average (a) CPU utilization, (b) throughput, and (c) latency based on three different classification scenarios.

We assigned patient data into three, five, and nine different areas, respectively, to verify the performance of PMTA for data provider classification using federated learning. Fig. 6 depicts the training accuracy of the federated learning model. The three curves in this figure represent different training cases. Observed from this figure, it is obviously that no matter how many areas are involved in the federated learning, training accuracy is constantly rising with the increase in the number of training rounds. Similar to Fig. 5, the global model constructed by federated averaging also stabilizes at around 400 rounds to converge, this is because the federated average algorithm fully takes into account the difference in the number of samples in each area to accelerate the convergence of the algorithm.



(a)

(b)

(c)

Fig. 8. Blockchain performance in average (a) CPU utilization, (b) throughput, and (c) latency with different number of transactions, when the number of classes equal to 3.

The experimental results shown in Figs. 5 and 6 suggest that the DQN-based classification mechanism of the proposed PMTA can classify data providers from areas of different characteristics efficiently and accurately within limited training rounds.

2) Performance Evaluation of Blockchain: In Fig. 7, we compare the blockchain performance under different classification scenarios in terms of average throughput, latency, and CPU utilization. Specifically, these performance results are all test results when the transaction number is set to 3000. As shown in Fig. 7(b) and (c), it is clear that the throughput and latency of the system increase significantly with the increase in the sending rate, while the CPU utilization shown in Fig. 7(a) is relatively stable, basically staying between 13% and 14%. Among them, in Fig. 7(b), when the number of classifications equals 6 and the sending rate increases to 350tps, system throughput reaches the highest value, which is nearly 96tps. Moreover, in Fig. 7(c), when the number of classifications equals 3 and the sending rate reaches 100tps, system latency is as low as 11.06 s.

Fig. 8 shows the comparison of blockchain performance in terms of average throughput, latency, and CPU utilization with three classes and different numbers of transactions. We use tx2500, tx3000, and tx3500 to denote the number of transactions to be 2500, 3000 and 3500, respectively. Fig. 8(a) shows the CPU utilization for three transaction numbers. It is obvious that the CPU utilization is relatively stable, always staying between 13% and 14%. As shown in Fig. 8(b), when the number of transactions equals 2500 and the sending rate increases to 400tps, system throughput reaches the highest value of 100tps. As shown in Fig. 8(c), the average latency goes up significantly with the increase in the sending rate, i.e., the more the transactions, the longer the latency.

The experimental results shown in Figs. 7 and 8 indicate that the proposed PMTA, through data provider and task classification based on DQN, spectral clustering and federated learning, and the corresponding smart contract design, can efficiently and securely process large-scale transactions. This suggests that the PMTA can effectively solve the problems of patient privacy leakage and data silos, thereby promoting the rapid development of Healthcare 4.0.

VI. Conclusion

In this article, we proposed a PMTA for Healthcare 4.0 using IoT, blockchain, and AI technologies. In PMTA, the mobile crowdsensing technology in IoT was used to issue tasks and receive data. In this process, in order to protect the sensitive information of data providers from being leaked, the strategy used deep differential privacy technology to add noise to the provided data. In addition, in order to protect sensitive information in tasks, we employed deep Q-networks combined with the spectral clustering technique and federated learning technique to train optimal classification strategies to classify tasks and data providers. Finally, in order to achieve privacy protection and improve task deployment efficiency, we combined data provider classification and task classification with task storage and publishing on the blockchain, and limited the choice of data providers through the design of smart contracts on different levels of tasks. Experimental results and performance analysis clearly showed that our proposed strategy performs well in terms of data privacy protection, task privacy protection, and system performance.

References

[1] X. Zhou et al., "Intelligent small object detection for digital twin in smart manufacturing with industrial cyber-physical systems," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 1377–1386, Feb. 2022.

[2] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 912–921, May/Jun. 2021.

[3] C. Huang, G. Zhang, S. Chen, and V. Albuquerque, "Healthcare industry 4.0: A novel intelligent multi-sampling tensor network for detection and classification of oral cancer," *IEEE Trans. Ind. Inform.*, to be published, doi: 10.1109/TII.2022.3149939.

[4] X. Fan, H. Wang, F. Xu, Y. Zhao, and K.-L. Tsui, "Homecare-oriented intelligent long-term monitoring of blood pressure using electrocardiogram signals," *IEEE Trans. Ind. Inform.*, vol. 16, no. 11, pp. 7150–7158, Nov. 2020.

[5] X. Zhou, X. Yang, J. Ma, and K. I.-K. Wang, "Energy efficient smart routing based on link correlation mining for wireless edge computing in IoT," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2021.3077937.

[6] X. Zhou, X. Xu, W. Liang, Z. Zeng, and Z. Yan, "Deep-learning-enhanced multitarget detection for end–edge–cloud surveillance in smart IoT," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12588–12596, Aug. 2021.

[7] C. Guo, P. Tian, and K.-K. R. Choo, "Enabling privacy-assured fog-based data aggregation in e-healthcare systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 3, pp. 1948–1957, Mar. 2021.

[8] J. Ni, A. Zhang, X. Lin, and X. S. Shen, "Security, privacy, and fairness in fog-based vehicular crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 146–152, Jun. 2017.

[9] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2019–2032, Oct. 2018.

[10] J. Ni, K. Zhang, Q. Xia, X. Lin, and X. S. Shen, "Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1317–1331, Jun. 2020.

[11] H. Wu, L. Wang, G. Xue, J. Tang, and D. Yang, "Enabling data trustworthiness and user privacy in mobile crowdsensing," *IEEE/ACM Trans. Netw.*, vol. 27, no. 6, pp. 2294–2307, Dec. 2019.

[12] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, Jul./Aug. 2020.

[13] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[14] S. Messaoud, A. Bradai, O. B. Ahmed, P. T. A. Quang, M. Atri, and M. S. Hossain, "Deep federated q-learning-based network slicing for industrial IoT," *IEEE Trans. Ind. Inform.*, vol. 17, no. 8, pp. 5572–5582, Aug. 2021.

[15] Y. Sun, L. Zhang, G. Feng, B. Yang, B. Cao, and M. A. Imran, "Blockchain-enabled wireless Internet of Things: Performance analysis and optimal communication node deployment," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5791–5802, Jun. 2019.

[16] A. M. Elmisery, S. Rho, and M. Aborizka, "A new computing environment for collective privacy protection from constrained healthcare devices to IoT cloud services," *Cluster Comput.*, vol. 22, no. 1, pp. 1611–1638, 2019.

[17] J. J. Hathaliya, S. Tanwar, S. Tyagi, and N. Kumar, "Securing electronics healthcare records in healthcare 4.0: A biometric-based approach," *Comput. Elect. Eng.*, vol. 76, pp. 398–410, 2019.

[18] R. Gupta, S. Tanwar, S. Tyagi, N. Kumar, M. S. Obaidat, and B. Sadoun, "Habits: Blockchain-based telesurgery framework for healthcare 4.0," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst.*, 2019, pp. 1–5.

[19] R. Gupta et al., "VAHAK: A blockchain-based outdoor delivery scheme using UAV for healthcare 4.0 services," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2020, pp. 255–260.

[20] H. Qiu, M. Qiu, M. Liu, and G. Memmi, "Secure health data sharing for medical cyber-physical systems for the healthcare 4.0," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2499–2505, Sep. 2020.

[21] R. Gupta, A. Shukla, and S. Tanwar, "Aayush: A smart contract-based telesurgery system for healthcare 4.0," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.

[22] S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic healthcare record system for healthcare 4.0 applications," *J. Inf. Secur. Appl.*, vol. 50, 2020, Art. no. 102407.

[23] P. Bhattacharya, S. Tanwar, U. Bodkhe, S. Tyagi, and N. Kumar, "BinDaaS: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1242–1255, Apr.–Jun. 2019.

[24] B. Yin, H. Yin, Y. Wu, and Z. Jiang, "FDC: A secure federated deep learning mechanism for data collaborations in the Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6348–6359, Jul. 2020.

[25] Z. Yan, J. Ge, Y. Wu, L. Li, and T. Li, "Automatic virtual network embedding: A deep reinforcement learning approach with graph convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 6, pp. 1040–1057, Jun. 2020.

[26] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

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

[28] M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 308–318.

[29] U. V. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[30] X. Qu, J. Wang, and J. Xiao, "Quantization and knowledge distillation for efficient federated learning on edge devices," in *Proc. IEEE 22nd Int. Conf. High Perform. Comput. Commun., IEEE 18th Int. Conf. Smart City, IEEE 6th Int. Conf. Data Sci. Syst.*, 2020, pp. 967–972.

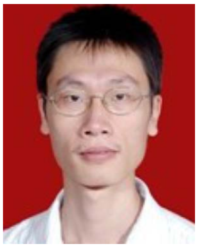
[31] H. Lin, S. Garg, J. Hu, G. Kaddoum, M. Peng, and M. S. Hossain, "A blockchain-based secure data aggregation strategy using sixth generation enabled network-in-box for industrial applications," *IEEE Trans. Ind. Inform.*, vol. 17, no. 10, pp. 7204–7212, Oct. 2021.

Xiaoding Wang received the Ph.D. degree in applied mathematics from the College of Mathematics and Informatics, Fujian Normal University, Fuzhou, China, in 2016.

He is currently an Associate Professor with the College of Computer and Cyber Security, Fujian Normal University. His main research interests include network optimization and fault tolerance.

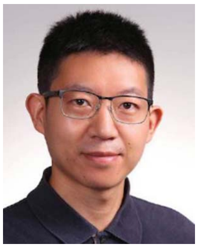
Mengyao Peng received the bachelor's degree in network engineering from the Lishui University, Lishui, China, in 2019. She is currently working toward the the master's degree in cyberspace security with the School of Computer and Cyberspace Security, Fujian Normal University, Fuzhou, China.

Her research interests include mobile crowdsensing, blockchain, and privacy security.



Hui Lin received the Ph.D. degree in computing system architecture from the College of Computer Science, Xidian University, Xi'an, China, in 2013.

He is currently a Professor with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China, where he is also an M.E. Supervisor with the College of Computer and Cyber Security. He has authored or coauthored more than 50 papers in international journals and conferences. His research interests include mobile cloud computing systems, blockchain, and network security.



Yulei Wu (Senior Member, IEEE) received the B.Sc. (first-class Hons.) degree in computer science and the Ph.D. degree in computing and mathematics from the University of Bradford, Bradford, U.K., in 2006 and 2010, respectively.

He is currently a Senior Lecturer with the Department of Computer Science, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, U.K. His main research interests include networking, Internet of Things, edge intelligence, AI and ethics, and privacy and trust.

Dr. Wu is an Associate Editor for the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, as well as an Editorial Board Member of *Computer Networks* and *Future Generation Computer Systems.* He is a Senior Member of the ACM, and a Fellow of the Higher Education Academy (HEA).

Xinmin Fan received the master's degree in software engineering from Fuzhou University, Fuzhou, China, in 2006.

He is currently the Deputy Director and a Researcher with the Network and Data Center (Network Information Office), Fujian Normal University, Fuzhou, China. He is also the Secretary General of the Fujian University Online Education Alliance, the Vice Chairman and Secretary General of the Fujian University Education Technology Research Association, and a Member of the "MOOC Construction Specification and Application Guidelines Drafting Group" of the Teaching Informatization and Teaching Method Innovation Steering Committee of the Ministry of Education. He has been engaged in the research and application practice of education informatization, online education, and higher education management for a long time.