Privacy-Preserving Intelligent Resource Allocation for Federated Edge Learning in Quantum Internet

Minrui Xu , Dusit Niyato , *Fellow, IEEE*, Zhaohui Yang, *Member, IEEE*, Zehui Xiong , *Member, IEEE*, Jiawen Kang , *Member, IEEE*, Dong In Kim , *Fellow, IEEE*, and Xuemin Shen , *Fellow, IEEE*

Abstract—Federated learning (FL) is an emerging technology for empowering various applications that generate large amounts of data in intelligent cyber-physical systems (ICPS). Though FL can address users' concerns about data privacy, its maintenance still depends on efficient incentive mechanisms. For long-term incentivization to participants in data federation under dynamic environments, deep reinforcement learning as a promising technology has been extensively studied. However, the non-stationary problem caused by the heterogeneity of ICPS devices results in a serious effect on the convergence rate of existing single-agent reinforcement learning. In this paper, we propose a multi-agent learning-based incentive mechanism to capture the stationarity approximation in FL with heterogeneous ICPS. First, we formulate the secure communication and data resource allocation problem as a Stackelberg game in FL with multiple participants. Then, to tackle the heterogeneous problem, we model this multi-agent game as a partially observable Markov decision process. Particularly, a multi-agent federated reinforcement learning algorithm is proposed to learn the allocation policies efficiently by dwindling variances in policy evaluation caused by interaction among multiple devices without sharing privacy information. Moreover, the proposed algorithm is proved to attain convergence at an expected rate. Lastly, extensive experimental results demonstrate that our proposed algorithm significantly outperforms baselines.

Index Terms—Federated edge learning, quantum key distribution (QKD), resource allocation, deep reinforcement learning.

I. INTRODUCTION

ARTIFICIAL Intelligence (AI) enables a wide range of computing and networking applications in edge networks, e.g., smart cities [1], [2], [3], Internet of Vehicles [4], [5], and Metaverses [6], [7], [8]. As one of the critical technologies in AI, federated edge learning (FEL) is a novel paradigm of privacy-preserving machine learning (ML) for intelligent edge networks [9]. In FEL, multiple data owners (a.k.a., FEL workers) can train a global model collaboratively for a model owner without exposing their sensitive raw data. To ensure the security of data and models in FEL systems, many modern cryptographic schemes are applied [10], such as secure multi-party computation (MPC), trusted execution environment (TEE), and safe key distribution. For instance, a secure and trusted collaborative edge learning framework is proposed in [11], where homomorphic encryption (HE) and blockchain are leveraged to track and choke malicious behaviors. With the rapidly increasing computation power of quantum computers [12], novel techniques will be brought to empower FEL systems, including large-scale searching, optimization, and semantic communication. However, the FEL systems based on existing schemes are under serious security threats. For example, traditional key distribution schemes based on the hardness in computing certain mathematical problems are no longer considered to be safe in the post-quantum era [10]. Fortunately, based on quantum no-cloning theorem [13] and Heisenberg's uncertainty principle [14], quantum key distribution (QKD) [15] is promising for providing proven secure key distribution schemes for collaborative training between FEL workers and model owners by facilitating public key and model encryption against eavesdropping attacks.

Originated from classical QKD schemes such as Bennett-Brassard-1984 (BB84) [16] and Grosshans-Grangier-2002 (GG02) [17], some modern QKD schemes have paved the way for the Quantum Internet [18] in recent years. For example, the measurement device-independent QKD (MDI-QKD) [19] provides one of the practical QKD solutions by increasing the range of secure communications and filling the detection gaps with an untrusted relay by avoiding any eavesdropping attacks on the Quantum Internet. Although several existing works focus on the theoretical and experimental aspects of the deployments of MDI-QKD, the problems of QKD resource allocation in the Quantum Internet have been largely overlooked [20]. For example, in [21], a deterministic programming model and a heuristic approach based on the shortest path algorithm are proposed to optimize the deployment cost of QKD resources. However, the problem of optimal allocation of QKD resources for quantum-secured FEL systems with heterogeneous data and model owners remains open. In particular, the number of participating FEL workers at different locations and times is uncertain due to unpredictable node and device failures [22]. Therefore, different security levels might be required by cluster heads to encrypt local models during global aggregation. Specifically, the secret-key rate for reaching the information-theoretic security (ITS) requirement is dynamic to support the encryption of intermediate models and related information according to uncertainties in quantum-secured FEL systems.

To address these uncertainty issues, we propose a stochastic QKD resource (i.e., wavelength) allocation model to optimize the QKD deployment cost of the Quantum Internet. To protect FEL models and public keys from eavesdropping attacks, we propose a hierarchical architecture for quantum-secured FEL systems that includes the FEL layer, the control and management layer, and the QKD infrastructure layer. To handle the dynamics of security demands from the FEL layer, we model the QKD resource allocation of QKD managers and QKD controllers in the control and management layer as a stochastic programming model that allocates QKD resources from the QKD infrastructure layer to cluster heads in the FEL layer. However, the proposed stochastic model can hardly be applied in practice because it requires complete state information from FEL nodes and QKD nodes, which is infeasible for QKD managers and controllers to collect due to privacy concerns [23]. Fortunately, the independent QKD resource allocation problems can be addressed by the promising deep reinforcement learning (DRL) algorithms [24]. Nevertheless, the efficiency and stability of learning-based approaches still face the issues of "data islands" during their training and inference [23]:

- Initially, QKD managers and controllers configure QKD nodes to provide QKD resources based only on practical observation of the state of the FEL layer. However, the experiences, including observations, actions, and rewards, are kept in the local replay buffers due to privacy concerns. Therefore, the lack of collaboration between QKD managers and controllers makes QKD resource allocation problems more challenging to satisfy changing security demands in quantum-secured FEL systems.

- Furthermore, QKD managers are reluctant to share their rewards from the FEL layers directly with QKD controllers. Therefore, QKD controllers can only collect incomplete experiences, including states and actions, during their interaction with the FEL systems, which are insufficient for their local policy improvement. Therefore, QKD controllers can only use policies shared by QKD managers to instruct QKD resource allocation decisions independently to QKD nodes for the FEL layers.

These issues could lead to inadequate training efficiency and unstable inference performance for learning-based algorithms in privacy-preserving environments.

To overcome the aforementioned issues, in this paper, we propose a learning-based QKD resource allocation scheme for quantum-secured FEL systems, which is strengthened by federated reinforcement learning. In particular, we use the model-free off-policy soft actor-critic (SAC) [25] structure to learn the optimal QKD resource allocation strategy. For each QKD manager and each controller, a policy network is adopted to configure the QKD nodes by learning the allocation strategy during the interaction with quantum-secured FEL systems. Moreover, a Q-network is adopted as a critic of each QKD manager and controller to evaluate the state-action values of its local policy, i.e., the performance of the local policy. To avoid direct reward sharing, QKD managers encrypt the Q-networks and then share them with QKD controllers for their local policy evaluation and improvement. In this way, the incomplete experience issues of QKD controllers can be addressed, thus improving the training efficiency of agents in the control and management layer. In addition, to further improve convergence efficiency, the local policy of QKD controllers is aggregated as the global QKD resource allocation policy for QKD managers after improving the local policies of QKD controllers. Our contributions can be summarized as follows.

- We propose a new hierarchical architecture for quantum-secured FEL systems to resolve uncertain factors in global model aggregation while providing ITS transmission of public keys and models. This architecture is capable of protecting the transmission of FEL models from external and participant attacks.

- In the proposed architecture, unlike deterministic linear programming, we formulate the optimization problem as stochastic programming to resolve the uncertainty of the security demand in quantum-secured FEL systems, i.e., the required secret-key rates. Considering the dynamic factors in the global aggregation of FEL, such as the number of FEL workers, the proposed model aims to minimize the deployment cost of the systems.

- To solve the proposed stochastic model without complete information, we proposed a federated DRL scheme that allows QKD managers and controllers to make the optimal decision independently based only on their local partial observation. Specifically, the proposed scheme enables QKD managers and controllers to learn a global policy collaboratively while maintaining their experiences in local replay buffers. Therefore, the proposed scheme can learn a

TABLE I
ABBREVIATIONS AND DEFINITIONS

Abbreviations	Definitions
FEL	Federated Edge Learning
QKD	Quantum Key Distribution
AI	Artificial Intelligence
MPC	Multi-party Computation
TEE	Trusted Execution Environment
HE	Homomorphic Encryption
BB84	Bennett-Brassard-1984
GG02	Grosshans-Grangier-2002
MDI-QKD	Measurement Device-Independent QKD
ITS	Information-Theoretic Security
DRL	Deep Reinforcement Learning
SAC	Soft Actor-Critic
KM	Key Management
GKS	Global Key Server
LKM	Local Key Manager
QTs	Quantum Transmitters
QRs	Quantum Receivers
SIs	Security Infrastructures
MUX/DEMUX	Multiplexing/Demultiplexing
POMDP	Partially Observable Markov Decision Process
SAC	Soft Actor-Critic
QBN	QKD Backbone Networking
EVF	Expected Value Formulation
SIP	Stochastic Integer Programming

synthetic QKD resource allocation policy efficiently without prior knowledge while preserving the privacy of the learning agents.

- Extensive experiments demonstrate the effectiveness of stochastic and learning-based resource allocation schemes for quantum-secured FEL systems. The performance evaluation results illustrate that existing baselines in the Quantum Internet with average or random demand do not lead to acceptable solutions that are significantly inferior to the proposed schemes.

We organize the rest of this paper as follows. In Section II, we provide a review of the related works. In Section III, we discuss the system model. In Section IV, we discuss the proposed optimization solution approach. In Section V, we propose the federated deep learning-based algorithm. Finally, we conduct the simulation experiments in Section VI and conclude in Section VII. The abbreviations and definitions used in this paper are summarized in Table I.

II. RELATED WORKS

A. Federated Edge Learning

Due to the enormous volume of data generated at the network edge [26], [27], [28], FEL has emerged as a promising paradigm of distributed privacy-preserving learning to improve the efficiency and security of communication and sensing for edge networks [29], [30], [31]. To illustrate the effectiveness and efficiency of FEL, Xu et al. in [32] provided a systematic overview of the convergence of edge networks and learning. They highlight the potential benefits of learning-based communication systems, such as semantic communications, and the necessity of sustainable resource allocation for edge learning systems. For instance, Hardy et al. [33] proposed a two-stage federated end-to-end learning system with performance comparable to centralized learning systems, including privacy-preserving local dataset adaptation and federated logistic regression over intermediate results encrypted with additive homomorphic encryption (HE) schemes. To address the problems of multi-view sensing observations in distributed wireless sensing, Liu et al. [34] proposed a vertical FEL system for cooperative detection while preserving the data privacy of sensors. By considering the non-cooperative nature of FEL participants at the edge, Lim et al. [35] proposed a hierarchical framework, including the evolutionary game and the Stackelberg game, for edge association and resource allocation problems in FEL systems. In addition, to motivate data owners to participate in FL, Zhan et al. in [36] provided a comprehensive survey on how to design proper incentive mechanisms for different federated learning algorithms in heterogeneous edge networks. Considering that conventional key distribution schemes in secure FEL systems are no longer secure, Huang et al. [10] proposed a new architecture for federated learning systems in the Quantum Internet called StarFL, which uses satellite and quantum key distribution schemes to distribute public keys for FEL workers with provable security.

B. The Quantum Internet

The Quantum Internet [18] connects quantum devices through quantum channels to provide long-term protection and future-proof security for the transmission of confidential information. It is expected that the Quantum Internet can provide new networking technologies for numerous critical applications by fusing quantum signal transmission with classical communication channels. Aiming to provide secure connectivity for mission-critical applications in the real world, Toudeh-Fallah et al. [37] established the first 800-Gbps quantum-secured optical channel up to 100 km, which is able to secure 258 data channels under AES-256-GCM with the quantum key update rate of one per second. By combining 700 fiber-based QKD links and two high-speed satellite-based QKD links, Chen et al. [38] developed an integrated space-to-ground quantum communication network with total coverage of 4,600 km. However, the costly quantum devices and the non-scalable routing schemes are obstacles to the large-scale deployment of the Quantum Internet. To address the scalability issues, Mehic et al. [39] proposed a routing protocol for the Quantum Internet that considers geographic distance and connection state to achieve high scalability by minimizing cryptographic key consumption. Meanwhile, For quantum-secured communications over backbone networks, Cao et al. [40] developed a programming-based resource allocation model and a heuristic algorithm to solve the deployment cost-minimized problem.
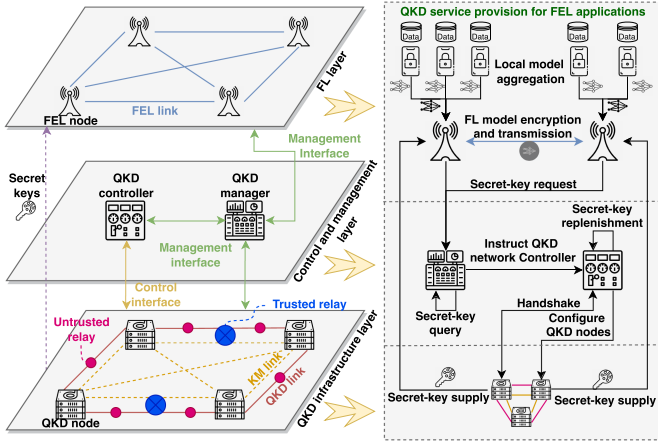
Fig. 1. An overview of quantum-secured federated edge learning in the Quantum Internet.

C. Learning-Based QKD Resource Allocation Schemes

Quantum key distribution is one of the mature applications of the Quantum Internet that has already been applied in some commercial scenarios [37]. However, QKD resources in the Quantum Internet still require efficient allocation schemes to bridge the gap between the low key generation rate of the Quantum Internet and the uncertain key demands of quantum-secured communication services. To fill this gap, Zuo et al. [41] proposed a reinforcement learning-based algorithm to learn the optimal QKD resource allocation strategy for the management of the quantum key pool in a cost-effective way. To address the dynamic arrival problem in multi-tenant QKD deployment, Cao et al. [42] proposed a reinforcement learning-based algorithm to reduce the deployment cost of QKD resources. However, these learning-based approaches consider QKD managers and controllers can share their observations and experiences without privacy concerns, which is infeasible in privacy-preserving systems. Therefore, in this paper, we propose a privacy-preserving learning-based resource allocation to minimize the deployment cost under uncertain security requirements of FEL workers and model owners in quantum-secured FEL systems.

III. SYSTEM MODEL

In this section, we first describe the proposed hierarchical architecture for quantum-secured FEL systems. We also describe the workflow in detail and give the complexity and security analysis. Based on the proposed architecture, we illustrate the network model, cost model, and uncertainty in quantum-secured FEL systems. As shown in Fig. 1, we consider a hierarchical architecture for quantum-secured FEL systems. In the FL layer, the cluster heads organize FL workers to join federated learning for model owners. The FL model encryption and transmission among cluster heads are secured by quantum cryptography. In the control and management layer, cluster heads initiate requests for quantum secret keys to the centralized QKD manager. Upon receiving these secret-key requests, each QKD manager first queries the secret-key status and then sends configurations to QKD controllers via the simple network management protocol

(SNMP). According to the received instructions, QKD controllers handshake with QKD nodes for detailed configurations. In the QKD infrastructure layer, there are three types of nodes (i.e., QKD nodes, trusted relays, and untrusted relays) and two types of links (i.e., key management (KM) and QKD links). We consider that the FEL nodes are co-located with the QKD nodes and that different types of links can be multiplexed within a single fiber [21]. Therefore, the topology of the QKD layer follows that of the FEL layer, which can be denoted by $G(\mathcal{V}, \mathcal{E})$. Here, $\mathcal{V}$ represents the set of FEL/QKD nodes, and $\mathcal{E}$ denotes the collection of fiber connections. Within a QKD node, there is a global key server (GKS), a local key manager (LKM), and one or more quantum transmitters (QTs). Between multiple QKD nodes, a QKD chain based on mixed relays can be used for global secret-key generation, where the trusted relays contain two or twice as many QTs [21], an LKM, and security infrastructures (SIs), while the untrusted relays consist of one or more quantum receivers (QRs). For links in the QKD layer, σ denotes the available wavelengths on QKD links and κ those on KM links.

A. Quantum-Secured Federated Edge Learning

As illustrated in Fig. 1, the proposed hierarchical architecture for quantum-secured FEL systems includes three layers [20], i.e., the FEL layer, the control and management layer, and the QKD infrastructure layer. In the FEL layer, a set of FEL workers denoted by $\mathcal{K} = \{1, \ldots, k, \ldots, K\}$ at the edge networks, i.e., end devices owning training datasets, participate in FEL tasks initialized by the set of model owners $\mathcal{O} = \{1, \ldots, o, \ldots, O\}$ with the aid of the set of cluster heads $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$ (e.g., base stations) [22]. Suppose there is a subset of FEL workers, i.e., K_j FEL workers [23], training a global model collaboratively with S_o data samples $\{x_i, y_i\}_{i=1}^{S_o}$ for the model owner o. Moreover, the feature space $x_i \in \mathbb{R}^{1 \times d}$ is distributed exclusively among FEL workers. The data samples in the local dataset of the FEL worker k within the cluster j can be denoted by $\{x_i^k \in \mathbb{R}^{1 \times d_k}\}_{k=1}^{K_j}$, where d_k is the dimension of its feature space. Without loss of generality, the number of FEL workers with labeled data is set to one in this paper, which is the FEL worker K_j. Therefore, the dataset with only features of FEL worker k can be denoted as $\mathcal{D}_i^k \triangleq \{x_i^k\}$ for $k = 1, \ldots, K_j - 1$. Meanwhile, let $\mathcal{D}_i^{K_j} \triangleq \{x_i^{K_j}, y_i^{K_j}\}$ denote the dataset of FEL worker K_j with features and labels. Let $\theta_k \in \mathbb{R}^{d_k}$ denote the model parameters of the FEL worker k and $\Theta_j = [\theta_1, \ldots, \theta_{K_j}]$ be the union of model parameters of all the FEL workers. The goal of FEL is to minimize inference loss through collaborative training of the optimal global model Θ^* for model owners that minimizes the loss function $\sum_{j=1}^{J} f_j(\sum_{k=1}^{K_j} \theta_k x^k, y^{K_j}; \Theta_j)$. As illustrated in Fig. 2, the training process of quantum-secured FEL systems in the proposed hierarchical architecture consists of five main steps [23], [43], [44]:

- *Step 1: Secure Communication Setup:* In quantum-secured FEL systems, the model owner o initializes its FEL task to the FEL worker k and distributes the initialized model parameters to the FEL workers via the cluster head j. First, the cluster head j needs to establish secure communication channels with the model owner o and the FEL worker k.
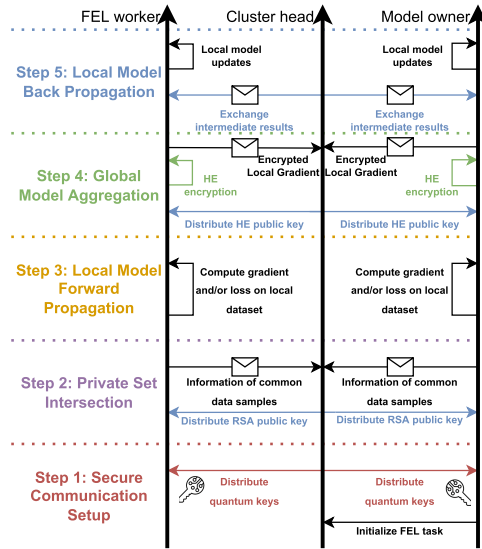
Fig. 2. The workflow of quantum-secured FEL. The black lines denote the classical communication links, the red lines denote the quantum links, and the blue lines denote the quantum-secured communication links.

To this end, the cluster head j sends QKD requests to QKD managers to establish quantum-secured communication channels with the model owner o and the FEL worker k, respectively. In detail, there are three main steps to distribute the quantum key $k_{(j,k)}^Q$ and $k_{(j,o)}^Q$ between the corresponding QKD nodes of the FEL worker k and model owner o via quantum channels, including transmitting the quantum bits, sifting the received bits, and estimating the error rate [16]. In this way, the cluster heads can establish quantum-secured communication channels using the quantum keys with model owners and FEL workers;

- *Step 2: Private Set Intersection:* In FEL systems, FEL workers have different data features and samples in their local datasets. Therefore, the model owner o must find a common set of data samples S_m for all participating FEL workers. The model owner can apply the *cross-database intersection* [45] to match the data samples among FEL workers while preserving their privacy. Specifically, the cluster head j applies the AES03 intersection protocol and blind signatures by generating a pair of RSA public and private keys $(e_j^{\text{RSA}}, d_j^{\text{RSA}}, n^{\text{RSA}})$ and using the encrypted public key $y_{j,k}(e_j^{\text{RSA}}) = E_{k_{j,k}^Q}(e_j^{\text{RSA}})$ to the FEL worker k and $y_{j,o}(e_j^{\text{RSA}}) = E_{k_{j,o}^Q}(e_j^{\text{RSA}})$ to the model owner o over quantum-secured communication channels. Then FEL workers and the model owner decrypt the public RSA key using their quantum keys for the intersection of the private set. By using the intersection of the private sets, the common data samples S_o are discovered by the model owner o without revealing the private information of FEL workers;

- *Step 3: Local Model Forward Propagation:* After determining the common data samples among FEL workers of the cluster head j, the FEL workers use their common local datasets $\mathcal{D}^{K_j} = \cup_{k=1}^{K_j} \mathcal{D}^k$ to train their local

models. The FEL workers $k = 1, \ldots, K_j - 1$ with features in their training samples can only obtain intermediate results $u_k = \theta_k x^k$ from forwarding propagation and cannot compute the loss for backward propagation. Meanwhile, the FEL worker K_j with labels can obtain intermediate inference results $u_{K_j} = \theta_{K_j} x^{K_j}$ and compute the local loss $f_j(\theta_{K_j} x^{K_j}, y^{K_j})$;

- *Step 4: Global Model Aggregation:* Once forward propagation is completed, the cluster head j creates an HE key pair $(e_j^{\text{HE}}, d_j^{\text{HE}})$ and distributes the public key e_j^{HE} to FEL workers and model owners via quantum-secured channels. Let $[[\cdot]]$ denote the operation of additive HE [23], the FEL workers first encrypt the intermediate results as $[[u_k]]$ and/or the encrypted loss $[[f_j(\theta_{K_j} x^{K_j}, y^{K_j})]]$ using the HE public key e_j^{HE}. Then, they transmit the encrypted intermediate inference results to model owners via the cluster headers. In this way, the FEL workers can share the intermediate inference results for global gradient and loss calculation without privacy leakage;

- *Step 5: Local Model Back Propagation:* The cluster head j decrypts the received intermediate results with the private key d_j^{HE}. Then, the cluster head sends the results to the model owners and FEL workers in the quantum-secured channels. Finally, the model owner o and the FEL workers decrypt and unmask the intermediate results and update their local model parameters accordingly.

During the training phase and the inference phase of FEL, there are four pillars of privacy for model owners and FEL workers. First, data privacy is the biggest concern of FEL workers. Next, model privacy, including model architecture privacy and model weight privacy, is essential for model owners that must not be stolen by any malicious participants. The compromise of model privacy makes the model owner have little motivation to improve the edge learning model performance, as their trained model can be easily sniffed or stolen by other participants. On the one hand, privacy in the input stage means that only the permitted FEL workers can input data to the model. On the other hand, privacy in the output stage means that the output of the model is only visible to the model owners. Following the literature definition in [10], we can have the following remarks.

Remark 1 (Privacy-preserving Distributed Machine Learning): The quantum-secured FEL systems are privacy-preserving distributed ML systems with no data and model privacy leakage during the input and output stages of the training phase and the inference phase.

During the input stage, FEL workers train their local models to fit their local datasets. They keep their training data local and then use the HE public key e^{HE} received from quantum-secured communication channels to encrypt the intermediate inference results. The local gradient encrypted with the public key can only be decrypted by the model owner who holds the private key. Therefore, in the input stage, the privacy of the data and models is well-protected.

During the output stage, the cluster heads decrypt the received intermediate results by using its private key of HE d^{HE}. On the one hand, the privacy of the data is preserved by the private set overlap in step 2 since no one can interfere with

the labels of FEL workers. On the other hand, the privacy of the models is enhanced since the public keys are distributed through quantum-secured communication channels so that no eavesdropper can intercept the local model updates and obtain the model architecture and model parameters from the encrypted models after encryption.

Complexity Analysis: Let $|k^Q|$ be the length of symmetric keys distributed via quantum channels. The complexity of generating symmetric keys for K FEL workers is $\mathcal{O}(K|k^Q|)$. During the private set intersection, the complexity to generate RSA key pairs for FEL workers is $\mathcal{O}(\log^2(n^{\text{RSA}}))$ [46]. Moreover, the complexity of the encryption and the decryption operations is $\mathcal{O}(K \log(|e^{\text{RSA}}|) \log^2(n^{\text{RSA}}) + \log^3(n^{\text{RSA}}))$ for K FEL workers. In FEL, the Paillier HE scheme [47] is usually adopted for additive HE. Therefore, the algorithmic complexity of HE encryption and decryption are both $\mathcal{O}(S \log(n^{\text{HE}}))$.

Security Analysis: Here, we can give external and participant attack examples, i.e., eavesdropping attacks and Sybil attacks, to show that the quantum-secured FEL systems can defend against common attacks in FEL. The first one is that the eavesdropping attacks during quantum-secured FEL training cannot pose any threat to the data and models in the system [48]. In other words, even if there is an eavesdropper in the system, the only information obtained during the eavesdropping attacks is encrypted with the unconditional secure scheme, i.e., one-time pad [49]. Moreover, the eavesdroppers have no way to decrypt or obtain valuable information from the ciphertext, i.e., the encrypted messages. Second, quantum-secured FEL systems can resist Sybil attacks [50] during the training process. Each FEL worker can only participate in FEL through one single quantum-secured channel. Therefore, FEL workers cannot forge their identities and submit multiple local models.

B. Networking Model for Quantum-Secured FEL Systems

Let ι denote the distance from a QT to its connected QR. Due to the symmetrical locations of two connected QTs, the distance between them can be approximated as $L \approx 2 \cdot \iota$. At length L between two connected QTs, the maximum attainable secret-key rate is denoted by K_L, which is in inverse proportion to the length L, i.e., an increment in L leads to a reduction in K_L [20]. In the proposed quantum-secured FEL model, $\mathcal{R}$ represents the set of quantum-secured FEL model transmission requests and $r(s_r, d_r, \rho_r) \in \mathcal{R}$ denotes one quantum-secured FEL model transmission request of FEL nodes, in which s_r and d_r represent the source node and the destination node of quantum-secured FEL model transmission request r, respectively. Let ρ_r be the amount of concurrent quantum-secured FEL links for satisfying the security demands (i.e., secret-key rates) of the FEL training between s_r and d_r [21], which can be calculated as $\rho_r = \left\lceil \frac{k_r}{K_L} \right\rceil$, where k_r denotes the required security demand between source node s_r and destination node d_r.

C. Cost Model and Provisioning Plans

Let $l_{(n,m)}$ denote the length of the physical link of the model transmission request between node n and node m.

1) Qts and Qrs: Since each MDI-QKD requires two QTs and one QR, the amount of QTs a_T^r and QRs a_R^r for the model transmission request r of quantum-secured FEL systems can be calculated as

$$a_T^r = \sum_{(n,m) \in E_r} 2 \cdot \rho_r \cdot \left\lceil \frac{l_{(n,m)}}{L} \right\rceil, \tag{1}$$

and

$$a_R^r = \sum_{(n,m) \in E_r} \rho_r \cdot \left\lceil \frac{l_{(n,m)}}{L} \right\rceil, \tag{2}$$

where E_r denotes the set of physical fibers along the transmission path of request r.

2) Lkms: The requested amount of LKMs a_{KM}^r for a quantum-secured FEL model transmission request r is

$$a_{KM}^r = \sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} + 1 \right\rceil. \tag{3}$$

3) Security Infrastructures: The requested amount of SI a_{SI}^r for a quantum-secured FEL model transmission request r is

$$a_{SI}^r = \sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} - 1 \right\rceil. \tag{4}$$

4) MUX/DEMUX Components: Finally, let a_M^r denote the number of MUX/DEMUX component pairs for a quantum-secured FEL model transmission request r is

$$a_M^r = \sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} \right\rceil + \sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} - 1 \right\rceil, \tag{5}$$

where the number of MUX/DEMUX component pairs required by QTs and QRs are represented by the first term $\sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} \right\rceil$ and the second term $\sum_{(n,m) \in E_r} \left\lceil \frac{l_{(n,m)}}{L} - 1 \right\rceil$, respectively.

5) QKD and KM Links: In the case of this study, three wavelengths are occupied by one QKD link and three wavelengths by one KM link [51]. For a quantum-secured FEL model transmission request r, the link cost can be calculated as

$$a_{Ch}^r = \sum_{(n,m) \in E_r} (3\rho_r l_{(n,m)} + l_{(n,m)}), \tag{6}$$

where the required length of QKD and KM links and FEL links can be represented by $3\rho_r l_{(n,m)}$ and $l_{(n,m)}$, respectively.

6) Provisioning Plans and Deployment Cost: When allocating resources for quantum-secured FEL applications in the proposed system model, the QKD manager needs to consider either a reservation plan or an on-demand plan, which is similar to the cloud and other online services [52], [53]. The reservation plan is for long-term allocation/subscription of QKD resources (specify what those are), while the on-demand plan is for short-term demand. If the QKD manager knows the demand of each FEL node in the FEL layer, it can provide the optimal reservation plan in the QKD layer. However, the demand of each FEL node is random as the number of workers in the FEL layers is uncertain. For example, some FEL nodes may require different secret-key rates in model transmission requests to protect their workers' models.

According to the above two types of subscription plans, each available QKD resource has two corresponding subscription costs, namely reservation costs and on-demand costs. We define the cost function as monetary units (e.g., dollars) per unit of QKD resources. In the reservation phase, $\beta_T^b, \beta_R^b, \beta_{KM}^b, \beta_{SI}^b, \beta_{MD}^b$, and β_{Ch}^b, are the reservation costs for the QTs, QRs, LKMs, SIs, MUX/DEMUX components, and QKD and KM links, respectively. Also, $\beta_T^o, \beta_R^o, \beta_{KM}^o, \beta_{SI}^o, \beta_{MD}^o$, and β_{Ch}^o, are the on-demand cost for the QTs, QRs, LKMs, SIs, MUX /DEMUX components, and QKD/KM links, respectively.

D. Uncertainty of Security Demands in Quantum-Secured FEL Systems

With the uncertainty of requests, the amount of required QKD resources by the FEL workers is not precisely known when the resources are reserved. In quantum-secured FEL, the encryption of public keys and intermediate inference results consume a certain amount of key resources. However, cluster heads choose different numbers of FEL workers to meet model owners' requirements for model accuracy and training error. Unfortunately, the model accuracy and training error are affected by various dimensional parameters such as data volume, algorithm quality, and FEL tasks. Therefore, cluster heads cannot accurately predict the model accuracy and thus require uncertain secret-key rates in the training process of FEL models. Let $\mathcal{K}_r = \{0, 1, \ldots, K\}$ represent the set of possible secret-key rates requirements of request $r \in \mathcal{R}$. The set of all possible secret-key rates of QKD nodes $\mathcal{K}$ in the QKD layer can be represented by the Cartesian product as

$$\mathcal{K} = \prod_{r \in R} \mathcal{K}_r = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_{|\mathcal{R}|}. \tag{7}$$

The probability distributions for both secret-key rates in $\mathcal{K}$ of all secure model transmission requests $\mathcal{R}$ are considered to be known. Note that statistical processes can be used to analyze historical data and that ML methods can predict the distribution of these demands.

IV. THE PROPOSED OPTIMIZATION MODELS

In this section, we first formulate the QKD resource allocation problem as a deterministic linear programming model. Furthermore, considering the uncertainty, i.e., the required secret-key rates, in quantum-secured FEL systems, we develop a stochastic programming model for QKD resource allocation. In the resource allocation model, the management signals are usually brief and can be encoded as quantum information and transmitted in QKD links, and thus the data transmission is secured by quantum cryptography.

A. Deterministic Integer Programming

Initially, consider the case where the actual FEL security requirements of FEL nodes are precisely known, and QKD resources can be subscribed to in a reservation plan, where each link has two decision variables.

1) $X = \{X_{(n,m)} | (n,m) \in \mathcal{E}\}$ indicates the set of numbers of wavelengths allocated in each KM link, e.g., $X_{(n,m)} = 1$ means that the QKD manager reserves one wavelength for the model transmission request.
2) $F = \{F_{(n,m)} | (n,m) \in \mathcal{E}\}$ indicates the set of the numbers of wavelengths allocated in each QKD link.

Based on the cost model and demand, the deterministic integer programming for the reservation plan can be formulated to minimize the total cost of the QKD problem as follows:

$$\min_{X^b, F^b} C^b\left(X^b, F^b\right) = \sum_{(n,m) \in \mathcal{E}} \left[\frac{F_{(n,m)}^b}{3} \left(a_T^b \beta_T^b + a_R^b \beta_R^b\right) \right.$$

$$+ X_{(n,m)}^b \left(a_{KM}^b \beta_{KM}^b + a_{SI}^b \beta_{SI}^b + a_{MD}^b \beta_{MD}^b\right)$$

$$\left. + l_{i,j} \left(F_{(n,m)}^b + X_{(n,m)}^b\right) \beta_{Ch}^b \right], \tag{8}$$

subject to:

$$\sum_{j \in \mathcal{V}} \sum_{p \in \kappa} x_{(n,m),p}^r - \sum_{j \in \mathcal{V}} \sum_{p \in \kappa} x_{(m,n),p}^r$$

$$= \begin{cases} 1 & \text{if } i = s_r \\ -1 & \text{if } i = d_r \\ 0 & \text{otherwise} \end{cases}, \forall r \in R, \forall (n,m) \in \mathcal{E} \tag{9}$$

$$X_{(n,m)}^b = \sum_{r \in R} \sum_{p \in \kappa} x_{(n,m),p}^r, \forall (n,m) \in \mathcal{E} \tag{10}$$

$$F_{(n,m)}^b = \sum_{r \in R} \sum_{q \in \sigma} f_{(n,m),q}^r, \forall (n,m) \in \mathcal{E} \tag{11}$$

$$\sum_{q \in \sigma} f_{(n,m),q}^r = 3\rho_r \sum_{p \in \kappa} x_{(n,m),p}^r, \forall r \in R, (n,m) \in \mathcal{E} \tag{12}$$

$$\sum_{j \in \mathcal{V}} f_{(n,m),q}^r = \sum_{j \in \mathcal{V}} f_{(j,i),q}^r, \forall r \in R, i \in \mathcal{V}, q \in \sigma \tag{13}$$

$$\sum_{j \in \mathcal{V}} x_{(n,m),p}^r = \sum_{j \in \mathcal{V}} x_{(j,i),p}^r, \forall r \in R, i \in \mathcal{V}, p \in \kappa \tag{14}$$

$$\sum_{r \in R} \sum_{p \in \kappa} x_{(n,m),p}^r \leq |\kappa|, \quad \forall (n,m) \in \mathcal{E} \tag{15}$$

$$\sum_{r \in R} \sum_{q \in \sigma} f_{(n,m),q}^r \leq |\sigma|, \forall (n,m) \in \mathcal{E} \tag{16}$$

$$\sum_{r \in R} x_{(n,m),p}^r \leq 1, \quad \forall (n,m) \in \mathcal{E}, p \in \kappa \tag{17}$$

$$\sum_{r \in R} f_{(n,m),q}^r \leq 1, \quad \forall (n,m) \in \mathcal{E}, q \in \sigma \tag{18}$$

The flow conservation constraint (9) guarantees that the QKD path flows between two distant FEL nodes is one in source/destination QKD nodes and zero in other QKD nodes. That is due to specifying a QKD path for a quantum-secured FEL request with a dedicated source and destination QKD nodes. The constraints in (10) and (11) guarantee that the demand is satisfied where $\sum_{p \in \kappa} x_{(n,m),p}^r$ and $\sum_{q \in \sigma} f_{(n,m),q}^r$ are the requested numbers of KM and QKD wavelengths in

link (n, m), respectively. The constraint in (12) specifies the number of wavelength channels requested by the QKD and the KM links of each quantum-secured FEL model transmission request, where the QKD/KM links and the FEL links are required to take up $3\rho_r$ and one wavelength channels, respectively. The same wavelength channel constraints in (13) and (14) are the constraints on wavelength continuity, which ensures that the identical wavelength channel is allocated to the link on the chosen path of each secure model transmission request. The wavelength capacity constraints (15) and (16) guarantee that the total wavelength channels of the QKD/KM links should not exceed the available wavelength channels. The wavelength uniqueness constraints (17) and (18) guarantee either zero or one wavelength channel can be allocated for the secure model transmission requests.

B. Stochastic Integer Programming

The deterministic integer programming developed in (8)–(18) is no longer applicable if the demand for the resources is unknown. Therefore, we describe stochastic integer programming (SIP), which optimizes the overhead of the QKD resources allocated to all quantum-secured FEL model transmission requests. The first phase includes all reservations that must be determined before the requirements can be implemented and analyzed. It is critical for the QKD manager to reserve the

number of QKD resources to be utilized before the demand is observed. In the second phase, allocations are made to accommodate real-time demand. After observing the real-time demand, if the reserved QKD link resources are less than the demand, FEL nodes have to pay for the cost of the additional QKD resources required.

The variables $X^o = \{X^o_{(n,m)}|(n,m) \in \mathcal{E}\}$ and $F^o = \{F^o_{(n,m)}|(n,m) \in \mathcal{E}\}$ denote the sets of the number of requests served in KM links and QKD links in the second stage, respectively. The expected overhead in the second stage is formulated as a function $\mathbb{E}_\Omega[L(X^o, F^o, \omega)]$, where $\omega \in \Omega = \mathcal{K}$ denotes the set of possible secret-key rates (called realizations, in general) observed in the second stage.

Thus, the total objective of this SIP model under uncertainty [52] is

$$\min_{X^b, F^b, X^o, F^o} C\left(X^b, F^b\right) + \mathbb{E}_\Omega\left[L\left(X^b, F^b, \omega\right)\right], \quad (19)$$

where

$$L\left(X^b, F^b, \omega\right) = \min_{Y = \{X^o(\omega), F^o(\omega)\}} L(Y) \quad (20)$$

is the cost function in the second phase for the given realization ω. The deterministic equivalent SIP of (8)–(18) for QKD resource allocation is expressed as (21)–(31) shown at the bottom of this page. In the optimization objective (21), there are probabilities $p(k)$, each denoting the probability of demand $k \in \mathcal{K}$

$$\min_{X^b, F^b, X^o(k), F^o(k)} \quad C^b(X^b, F^b) + \sum_{k \in \mathcal{K}} p(k) \left[C^o\left(X^o(k), F^o(k)\right)\right] \quad (21)$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{V}} \sum_{p \in \kappa} x^r_{(n,m),p}(k) - \sum_{j \in \mathcal{V}} \sum_{p \in \kappa} x^r_{(j,i),p}(k) = \begin{cases} 1 & \text{if } i = s_r \\ -1 & \text{if } i = d_r \\ 0 & \text{otherwise} \end{cases}, \quad \forall r \in R, k \in \mathcal{K} \quad (22)$$

$$X^b_{(n,m)} + X^o_{(n,m)}(k) \geq \sum_{r \in R} \sum_{p \in \kappa} x^r_{(n,m),p}(k), \quad \forall (n,m) \in \mathcal{E}, k \in \mathcal{K} \quad (23)$$

$$F^b_{(n,m)} + F^o_{(n,m)}(k) \geq \sum_{r \in R} \sum_{q \in \sigma} f^r_{(n,m),q}(k), \quad \forall (n,m) \in \mathcal{E}, k \in \mathcal{K} \quad (24)$$

$$\sum_{q \in \sigma} f^r_{(n,m),q}(k) = 3\rho_r(k) \sum_{p \in \kappa} x^r_{(n,m),p}(k), \quad \forall r \in R, (n,m) \in \mathcal{E}, k \in \mathcal{K} \quad (25)$$

$$\sum_{j \in \mathcal{V}} f^r_{(n,m),q}(k) = \sum_{j \in \mathcal{V}} f^r_{(j,i),q}(k), \quad \forall r \in R, i \in \mathcal{V}, q \in \sigma, k \in \mathcal{K} \quad (26)$$

$$\sum_{j \in \mathcal{V}} x^r_{(n,m),m}(k) = \sum_{j \in \mathcal{V}} x^r_{(j,i),p}(k), \quad \forall r \in R, i \in \mathcal{V}, p \in \kappa, k \in \mathcal{K} \quad (27)$$

$$\sum_{r \in R} \sum_{p \in \kappa} x^r_{(n,m),p}(k) \leq |\kappa|, \quad \forall (n,m) \in \mathcal{E}, k \in \mathcal{K} \quad (28)$$

$$\sum_{r \in R} \sum_{q \in \sigma} f^r_{(n,m),q}(k) \leq |\sigma|, \quad \forall (n,m) \in \mathcal{E}, k \in \mathcal{K} \quad (29)$$

$$\sum_{r \in R} f^r_{(n,m),q}(k) \leq 1, \quad \forall (n,m) \in \mathcal{E}, q \in \sigma, k \in \mathcal{K} \quad (30)$$

$$\sum_{r \in R} x^r_{(n,m),p}(k) \leq 1, \quad \forall (n,m) \in \mathcal{E}, p \in \kappa, k \in \mathcal{K} \quad (31)$$

being realized. Eq. (22) is the flow conservation constraint. Eqs. (23) and (24) are demand satisfaction constraints. Eq. (25) is the wavelength channel number constraint. (26) and (27) are the constraints to ensure each secure model transmission requests the same wavelength channel. Eqs. (28) and (29) are wavelength capacity constraints, while (30) and (31) are wavelength uniqueness constraints.

V. THE PROPOSED LEARNING-BASED SCHEME

In this section, we formulate the above QKD resource allocation problem as a learning task. In particular, we model the stochastic model as a partially observable Markov decision process (POMDP) for QKD managers and controllers. Based on the properties of the POMDP, we design a learning-based QKD resource allocation scheme based on federated reinforcement learning [54], [55] to explore the optimal solution without sharing the private experiences. Moreover, we give the convergence analysis and the complexity analysis of the proposed scheme.

A. POMDP for Quantum-Secured FEL Systems

1) State Space: The state space $s(t) \in \mathcal{S}$ of the QKD networks can be represented by the reservation strategy and the on-demand deployment of the QKD manager and QKD controllers at tth slot. Therefore, the state of QKD networks is defined as $s(t) := [X^b(t-l), F^b(t-l), \ldots, X^b(t-1), F^b(t-1)]$. This indicates that the state of the QKD network is composed of the past l reservation strategies and on-demand deployment. Therefore, for each QKD controller, their observation is a part of the global state, which can be denoted as $s_n(t) := [X_n^b(t-l), F_n^b(t-l), \ldots, X_n^b(t-1), F_n^b(t-1)]$, $n = 1, \ldots, N$. Without loss of generality, we consider that one QKD manager is co-located with one QKD controller, but only some of the QKD controllers are QKD managers. Therefore, the observation of QKD manager m can be denoted as $s_m(t) := [X_m^b(t-l), F_m^b(t-l), \ldots, X_m^b(t-1), F_m^b(t-1)]$, for $m = 1, \ldots, M$.

2) Action Space: The action space $a(t) = [a_1(t), \ldots, a_N(t)] \in \mathcal{A}$ is the reservation strategy profile of QKD resource allocation deployment at time slot t, i.e., $a_n(t) = [X_n^b(t), F_n^b(t)]$, $n = 1, \ldots, N$.

3) State Transition Probability Function: The state transition probability function $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ represents the changing rules of uncertain factors. The training environment transitions to the next moment of state based on the current state of the environment and the action input by the training agent, calculated by the probability function for the state transition. Therefore, we can assume $s(t+1) = s'$ and the probability of transitioning to this state is $P(s'|s, p)$, where $s(t) = s$ and $p(t) = p$. The agent's policy can change the trajectory of the state transition.

4) Reward: In the deployment cost minimization problem defined in (21), the reward in time slot t is the ratio of near-minimal deployment cost to current deployment cost. However, the reward is only observable by the QKD manager in the Quantum Internet. Therefore, the reward function is always zero for all QKD controllers, i.e.,

$R_n(s(t), a(t)) = 0, n = 1, \ldots, N$. Moreover, the reward function can be expressed as the normalized cost $R_m(s(t), a(t)) = \frac{C^b(X^b, F^b) + \sum_{k \in \mathcal{K}} p(k)[C^o(X^o(k), F^o(k))]}{C^b(X^{b,*}, F^{b,*}) + \sum_{k \in \mathcal{K}} p(k)[C^{o,*}(X^{o,*}(k), F^{o,*}(k))]}$ for the QKD manager m.

5) Q-Function: Let γ be the discounting factor that determines how much QKD managers care about future cumulative returns versus immediate rewards. Therefore, the Bellman equation for Q^π w.r.t the policy π is

$$Q^\pi(s(t), a(t)) = \mathbb{E}_{s(t+1), a(t+1) \sim \mathcal{P}^\pi} \left[R(s(t), a(t)) + \gamma(Q^\pi(s(t+1), a(t+1)) + \tau H(\pi(\cdot|s(t+1))) \right], \quad (32)$$

where $H(\pi(\cdot|s(t+1))) = -\log \pi(\cdot|s(t+1))$ is the expected entropy for policy π in state $s(t+1)$. To determine the relative importance of the entropy term, τ is the additive temperature parameter that help affects the optimal policy to determine its stochasticity. Therefore, the Q-function in the SAC algorithm [25] can be expressed as

$$Q^\pi(s(t), a(t)) = \mathbb{E}_{s(t+1), a(t+1) \sim \mathcal{P}^\pi} \left[R(s(t), a(t)) + \gamma(Q^\pi(s(t+1), a(t+1)) - \tau \log(\pi(\cdot|s(t+1))) \right]. \quad (33)$$

B. Federated DRL-Based QKD Allocation Scheme

To help each QKD manager and controller evaluate its QKD resource allocation function $\pi(\cdot)$ parameterized by φ, a Q-function $Q(\cdot, \cdot; \vartheta)$ parameterized by a neural network ϑ is developed. After fitting the Q-network to the allocation policy, the policy network can be trained using the Q-network via the policy gradient algorithm. The Q-network can specify the expected returns for the actions performed by the policy network. In this way, the probability of actions leading to lower expected QKD deployment costs is increased. The probability of actions leading to higher expected costs decreases until the policy converges to the Q-network. The learning agent needs to repeat this process until it converges to the optimal policy. At each local iteration of the learning agents, training experiences are sampled randomly from the replay buffers of QKD managers or QKD controllers to update the network parameters of policy or Q-networks, respectively.

1) Local Policy Iteration for QKD Managers: The parameters $\vartheta_{m,i}, i = 1, 2$ in double Q-networks of the QKD manager m is updated by minimizing the difference between the output of Q-networks and the target is performed via gradient descent as follows:

$$\epsilon_{m,i}^{e+1} = \arg\min_{\vartheta_{m,i}} \frac{1}{|B_m|} \sum_{(s(t), a(t)) \sim B_m} \left[Q(s(t), a(t); \vartheta_{m,i}) - y_m(s(t), a(t)) \right]^2, \quad (34)$$

where the target $y_m(s(t), a(t))$ is given by

$$
\begin{aligned}
y_m(s(t), a(t)) = {} & R_m\left(s(t), a(t)\right) \\
& + \gamma\Bigg[\min_{i=1,2} Q(s(t+1), \tilde{a}(t+1); \vartheta_{m,i,targ}) \\
& - \tau \log \pi(\tilde{a}(t+1)|s(t+1); \varphi_m)\Bigg], \quad (35)
\end{aligned}
$$

for $i = 1, 2$ and $\tilde{a}(t+1) \sim \pi(\cdot|s(t+1); \vartheta_m)$.

Moreover, the policy of the learning agent in the proposed scheme should, in each state, maximize the expected future return plus expected future entropy. To this end, by utilizing the min-double-Q trick [56] and the entropy loss term, the policy network of QKD manager m is updated to maximize the trained Q-network as follows:

$$
\begin{aligned}
\varphi_m^{e+1} = {} & \arg\max_{\varphi_m} \frac{1}{|B_m|} \sum_{s \in B_m} \Bigg[\min_{i=1,2} Q(s, \tilde{a}(s; \varphi_m); \vartheta_{m,i}) \\
& - \tau \log(\pi(\tilde{a}(s; \varphi_m)|s; \varphi_m))\Bigg], \quad (36)
\end{aligned}
$$

where $\tilde{a}$ is reparameterized from a squashed Gaussian policy w.r.t. the mean $\epsilon(s; \varphi_m)$ and variance $\sigma(s; \varphi_m)$ output from the policy network φ_m as follows:

$$
\tilde{a}(s, \xi; \varphi_m); \varphi_m) = \tanh\left(\epsilon(s; \varphi_m) + \sigma(s; \varphi_m) \odot \xi\right), \quad (37)
$$

and ξ is an input noise vector sampled from some fixed distribution.

Finally, the target Q-networks are updated with $\varphi_{m,i,targ}^{e+1} \leftarrow \zeta\vartheta_{m,i,targ}^e + (1 - \zeta)\vartheta_{m,i}^{e+1}$, for $i = 1, 2$.

2) Local Policy Iteration for QKD Controllers: Since there is no replay buffer in the QKD controller that can be trained with Q-network, QKD controllers cannot train their policy networks independently. Fortunately, QKD controllers can borrow the trained Q-network from QKD managers to facilitate the update of its local policy, i.e. $\vartheta_i^{e,fed} \leftarrow \vartheta_{m,i}^e$, for $i = 1, 2$. We then update the policy network of QKD controller n by maximizing the objective via gradient ascent as follows:

$$
\begin{aligned}
\varphi_n^{e+1} = {} & \arg\max_{\varphi_n} \frac{1}{|B_n|} \sum_{s \in B_n} \Bigg[\min_{i=1,2} Q(s, \tilde{a}(s; \varphi_n); \vartheta_i^{e,fed}) \\
& - \tau \log(\pi(\tilde{a}(s; \varphi_n)|s; \varphi_n))\Bigg], \quad (38)
\end{aligned}
$$

where $\tilde{a}(s; \varphi_n)$ is a sample from $\pi(\cdot|s; \varphi_n^e)$ which is differentiable w.r.t. φ_n^e via the reparametrization trick [25].

3) Global Policy Aggregation: Without exposing the private information, the global policy is aggregated from the local QKD allocation policy while maintaining their private experiences in their local replay buffers. The global aggregation process via FedAvg [57] of local policy networks can be presented as:

$$
\varphi^{e+1} = \frac{\sum_{i=1}^N B_n \varphi_n^{e+1}}{B}, \quad (39)
$$

where $B = \sum_{n=1}^N B_n$ is the number of total incomplete experiences in the replay buffers of QKD controllers. And the updated global policy φ^{e+1} is used for next-episode training. Algorithm 1

Algorithm 1: The Proposed Intelligent Resource Allocation Scheme Based on Federated DRL.

---

1   Initialize $\varphi_m, \vartheta_{m,1}, \vartheta_{m,2}, \vartheta_1^{fed}, \vartheta_2^{fed}, B_n, B_m$.
2   **for** *Episode $e \in 1, \ldots, E$* **do**
3   **for** *each decision slot t* **do**
4   The QKD manager m and QKD controller n observe $s_m(t)$ and $s_n(t)$, respectively.
5   Input $s_m(t)$ and $s_n(t)$ to policy network and obtain the reservation strategy profile $a_n(t)$ of QKD resource allocation.
6   The QKD manager m store $(s_m(t), a_m(t), R_m, s_m(t+1))$ in B_m and the QKD controller n store $(s_n(t), a_n(t), s_n(t+1))$ in B_n.
7   **end**
8   **for** *each local policy iteration* **do**
9   The QKD managers update the policy networks according to Eq. (36) and the Q networks according to Eq. (38).
10   The QKD managers updates target Q-networks with $\varphi_{m,i,targ}^{e+1} \leftarrow \zeta\vartheta_{m,i,targ}^e + (1-\zeta)\vartheta_{m,i}^{e+1}$, for $i = 1, 2$.
11   The QKD controllers borrow the trained Q-network from QKD managers to facilitate the update of its local policy, i.e. $\vartheta_i^{e,fed} \leftarrow \vartheta_{m,i}^e$, for $i = 1, 2$.
12   The QKD controllers update the local policy networks according to Eq. (40).
13   The local policy networks of QKD controllers are aggregated in Eq. (41) and then sent to QKD managers.
14   **end**
15   **end**

---

summarizes the proposed intelligent resource allocation scheme based on federated DRL for quantum-secured FEL systems.

C. Convergence Analysis and Complexity Analysis

Denote φ^* as the corresponding critic under an optimal QKD resource allocation policy. We have the following assumptions [58]:

Assumption 1: For all $C_n(\varphi_n)$,
- $C_n(\varphi_n)$ is convex;
- $C_n(\varphi_n)$ is ε-smooth, i.e., $C_n(\varphi_n') \leq C_n(\varphi_n) + \nabla C_n(\varphi_n) \cdot (\varphi_n' - \varphi_n) + \frac{\varepsilon}{2}\|\varphi_n - \varphi_n'\|^2$, for $\forall\varphi_n$ and φ_n'.

Assumption 1 is used to provide the feasibility of solution space and guarantee the update rule of the federated reinforcement learning-based scheme. Then, we can have the lemma as follows:

Lemma 1: $C(\varphi) = \sum_{n=1}^N B_n C_n(\varphi_n)/B$ is ϵ-strongly convex and ε-smooth.

Proof: According to Assumption 1, the global cost function $C(\varphi)$ is the weighted finite-sum of local cost function $C_n(\varphi_n)$. Furthermore, by applying the triangle inequality and the definition of convex, $C(\varphi)$ is ϵ-strongly convex and ε-smooth.

Theorem 1: Considering that $C(\varphi)$ is ε-smooth and ϵ-strongly, let $\nu = 1/C$ and $\varphi^* = \arg\min_\varphi C(\varphi)$, we have

$$
\|\varphi^e - \varphi^*\| \leq \left(1 - \frac{\epsilon}{\varepsilon}\right)^e \|\varphi^1 - \varphi^*\|, \quad (40)
$$

so the gradient dispersion can be derived as $O(\bar{\varphi}) = \frac{\varepsilon}{\epsilon} \log(\|\varphi^1 - \varphi^*\|/\bar{\varphi})$, which is used to illustrate how the parameters φ^t are distributed in each worker.

Proof: According to the ϵ-strongly convexity of $C(\varphi)$, we have

$$\nabla C(\varphi)(\varphi - \varphi^*) \geq C(\varphi) - C(\varphi^*) + \frac{\epsilon}{2}\|\varphi - \varphi^*\|^2. \quad (41)$$

Thus, we can obtain the following:

$$\left\|\varphi^{e+1} - \varphi^*\right\|^2 = \left\|\varphi^e - \nu\nabla C(\varphi^e) - \varphi^*\right\|^2$$
$$= \|\varphi^e - \varphi^*\| - 2\nu\nabla C(\varphi^e)(\varphi^e - \varphi^*) + \nu^2\|\nabla C(\varphi^e)\|^2$$
$$\leq \|\varphi^e - \varphi^*\| - 2\nu(C(\varphi) - C(\varphi^*)$$
$$+ \frac{\epsilon}{2}\|\varphi^e - \varphi^*\|^2) + \nu^2\|\nabla C(\varphi^e)\|. \quad (42)$$

By smoothing $C(\varphi)$, the gradient bound can be obtained as

$$C(\varphi^*) \leq C(\varphi - \frac{1}{\varepsilon}\nabla C(\varphi))$$
$$\leq C(\varphi) - \|\nabla C(\varphi)\|^2 + \frac{1}{2\varepsilon}\|\nabla C(\varphi)\|^2$$
$$\leq C(\varphi) - \frac{1}{2\varepsilon}. \quad (43)$$

Combining (42), (43) can be reformulated as

$$\left\|\varphi^{e+1} - \varphi^*\right\| = \left\|\varphi^e - \nu\nabla C(\varphi^e) - \varphi^*\right\|^2$$
$$\leq \|\varphi^e - \varphi^*\| - \nu\epsilon\|\varphi^{e+1} - \varphi^*\| + 2\nu(\nu\varepsilon - 1)(C(\varphi - C(\varphi^*)))$$
$$\leq \left(1 - \frac{\epsilon}{\varepsilon}\right)\|\varphi^e - \varphi^*\| \leq \left(1 - \frac{\epsilon}{\varepsilon}\right)\|\Delta^*(\varphi)\|, \quad (44)$$

where ν is set as the last iteration. $\qquad\square$

The expected convergence bound can be estimated as

$$[C(\varphi^e) - C(\varphi^*)] \leq \bar{\varphi}^e[\Delta^e(C(\varphi^*))], \quad (45)$$

where $C(\varphi)$ is proven to be bounded as $\Delta^e(C(\varphi^*)) = C(\varphi^1) - C(\varphi^*)$. Our proposed learning-based scheme can find an optimal QKD resource allocation strategy without sharing information and experience. As for the complexity of the proposed learning-based QKD allocation algorithm, each QKD manager and controller maintains its local policy and make independent decisions independently during the QKD resource allocation phase. Moreover, the input and output dimensions are constant and determined by the dimensions of the observation and action spaces. Therefore, the computation complexity is $O((N + M)UL)$ in each decision slot, where L is the number of hidden layers and U is the number of hidden neurons in each hidden layer. The proposed learning-based resource allocation scheme for the Quantum Internet in shown in Fig. 3 and the overall workflow is summarized in Algorithm 1.

VI. EXPERIMENTAL EVALUATION

This section evaluates the effectiveness of the proposed SIP scheme and the learning-based QKD resource allocation scheme in different system settings and topologies. First, we describe the setting of the experiments. Then, we analyze the cost structure of the proposed SIP scheme. Moreover, we compare the proposed
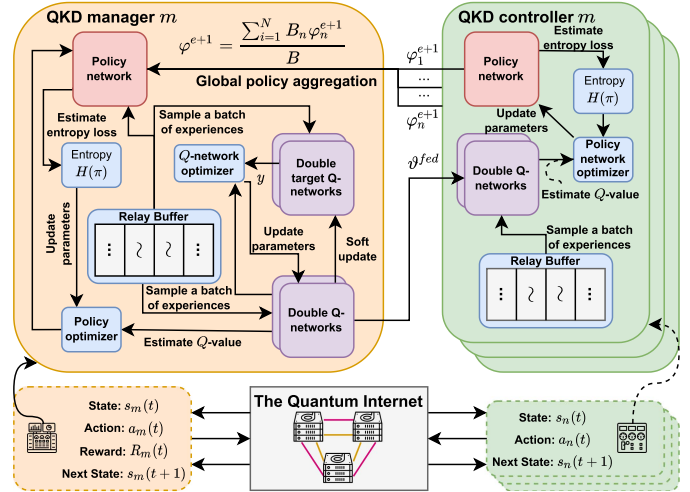

Fig. 3. The proposed learning-based resource allocation scheme for the Quantum Internet.

TABLE II
RESERVED AND ON-DEMAND COST VALUES IN EXPERIMENTAL
EVALUATION (UNITS) [21]

Cost	β_{Tx}^r	β_{Rx}^r	β_{KM}^r	β_{SI}^r	β_{MD}^r	β_{Ch}^r
Reserved	1500	2250	1200	150	300	1
On-demand	6000	9000	3000	500	900	4

scheme with other baseline schemes to demonstrate its effectiveness in resolving uncertainty in QKD resource allocation. Finally, a convergence analysis is performed to illustrate the performance of the proposed learning-based scheme.

A. Experiment Settings

Similar to [21], experiments are performed on two well-known topologies (i.e., the 14-node NSFNET topology and the 24-node USNET topology). The distance L between two QT is set to be 160 km. The required secret-key rate is considered to be the same for all secure model transmission requests. To streamline the evaluation, the uncertainty $\mathcal{K}$ is reformulated by $\acute{\mathcal{K}}$ as

$$\acute{\mathcal{K}} = \{(k_1, k_2, \ldots, k_{|\mathcal{R}|})|0 \leq k_1, k_2, \ldots, k_R < |\Omega|,$$
$$k_1 = k_2 = \cdots = k_{|\mathcal{R}|}\}, \quad (46)$$

where $|\Omega|$ denotes the total number of scenarios. For each scenario, the probability distribution of secret-key rate requirements obeys a Poisson process, in which the average secret-key rate requirements are set to $\lfloor|\Omega|/3\rfloor$. The default number of scenarios is set to $|\Omega| = 10$. The default numbers of quantum-secured FEL model transmission requests $|\mathcal{R}|$ are set to 100 and 200 for NSFNET and USNET, respectively. Finally, for the performance evaluation and analysis, the cost values are given in Table II, with a unit representing a normalized monetary unit. As for the proposed learning-based approach, the learning rate is set to 0.003 for Q-networks and 0.0001 for policy networks. The temperature coefficient τ is set at 0.01. The reward discounting factor is set to 0.95, and the update coefficient of target Q-networks is set to
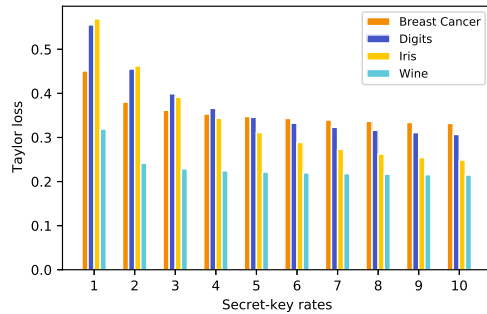
Fig. 4. Performance of the global model in quantum-secured FEL systems with various secret-key rates.

(a) NSFNET. (b) USNET.

Fig. 5. The cost structure of SIP for quantum-secured FEL over the NSFNET and the USNET topologies.

(a) NSFNET. (b) USNET.

Fig. 6. Deployment costs of QKD resources versus numbers of quantum-secured FEL model transmission requests.

0.01. The number of training epochs is set to 100. Finally, the size of the replay buffer is set to 20000 for each learning agent. The SIP model is formulated and solved by Gurobipy 9.5.2 and the DRL algorithms are implemented via PyTorch 1.9.0.

B. Global Model Performance Under Different Secret-Key Rates

We conduct our experiments on four datasets from scikit-learn[1], including breast cancer, digits, iris, and wine. As shown in Fig. 4, the Taylor loss of the global model decreases as the required secret-key rates increase. The model owner can hire more FEL workers if the cluster head has higher secret-key rates to maintain the quantum-secured channels with the FEL workers. Thus, the performance of the global model can be improved by including more knowledge from the datasets of FEL workers. For the complex Digits dataset, the Taylor loss [59] of the classification model decreases from about 0.55 to about 0.3 as the secret-key rates increase. However, the simple wine dataset only requires lower secret-key rates to converge. The reason is that the secret-key rates affect the number of concurrent quantum-secured channels that the cluster heads can maintain, which affects the number of FEL workers in global iterations.

C. Cost Structure Analysis

Initially, the cost structure of the SIP for quantum-secured FEL applications is studied. To simplify the presentation of the cost structure, equal resources are reserved for each QKD request on each link. In Fig. 5, the costs in the first stage, the second state, and overall are illustrated by varying the number of reserved requests for each link. As expected, the cost of the first-stage resource reserve increases with the increase in secret-key rates. However, after the realized demand is known, the costs of the second stage decrease because the reserved QKD wavelengths increase as the requests require more minor on-demand compensation. Here, the optimized plan can be determined (e.g., when the reserved number of wavelengths equals three, as shown in Fig. 5). The analysis of the cost structure
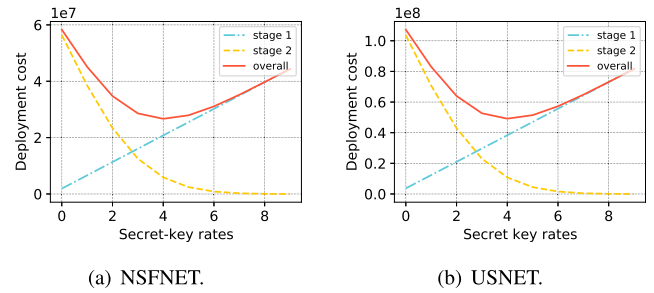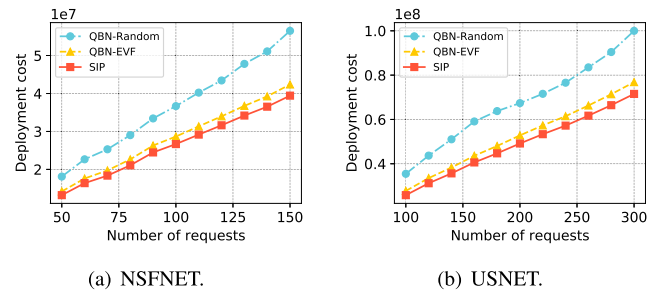
[1][Online]. Available: https://scikit-learn.org/stable/datasets/toy_dataset.html

shows that the optimal solution is not easy to obtain due to the uncertainty of the system in terms of security. For instance, the optimal QKD resource allocation plan is not at the point where the cost of the second stage exceeds the cost of the first stage. Thus, it is necessary to formulate the SIP of the QKD layer in such a way that the deployment cost is optimized.

D. Performance Evaluation Under Various Parameters

In the performance evaluation, the proposed SIP scheme is compared with two baseline schemes, i.e., the QKD backbone networking (QBN) scheme with expected value formulation (QBN-EVF) allocation and the QBN with random (QBN-Random) allocation, presented in [21]. In the case of QBN-EVF, the secret-key rates in the first stage are determined by the demanded average, which represents an approximate solution. On the other hand, in the QBN-Random scheme, the values of the decision variables are generated uniformly from zero to $|\Omega|$, indicating a random scheme.

The deployment costs of QBN-Random, QBN-EVF, and SIP for different numbers of quantum-secured FEL model transmission requests are shown first. In Fig. 6, we can observe that as the number of secure model transmission requests increases, the deployment cost of all three solutions increases accordingly. Moreover, the difference in deployment cost between the three schemes also increases with the increase in the number of quantum-secured FEL model transmission requests. Specifically, the deployment cost of QBN-EVF is slightly higher than that of SIP, while the deployment cost of QBN-Random is 50% higher than that of SIP. A similar situation arises in
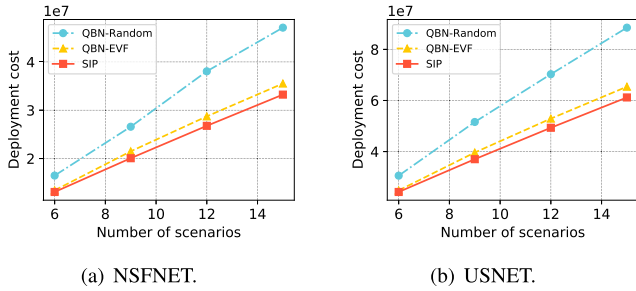
Fig. 7. Deployment costs of QKD resources versus numbers of scenarios in quantum-secured FEL.
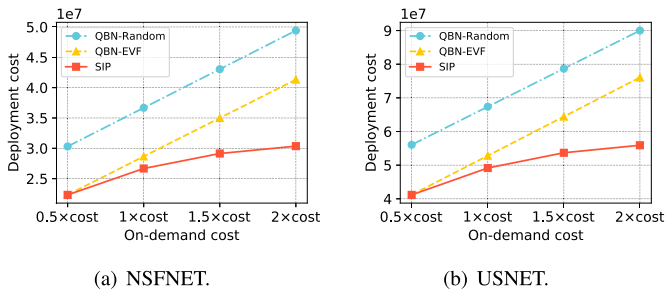
Fig. 8. Deployment costs of QKD resources versus different on-demand cost schemes.

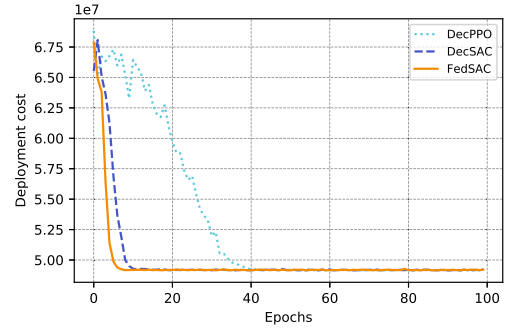Fig. 9. Convergence analysis of the proposed scheme.

the comparison of different numbers of scenarios, as shown in Fig. 7, namely, the deployment cost of QBN-EVF is slightly higher than that of SIP, while the deployment cost of QBN-Random is twice as expensive as that of SIP. Fig. 8 shows the impact of On-demand costs on the deployment costs of each scenario. For the QBN-Random and QBN-EVF scenarios, the deployment cost increases exponentially with the On-demand cost. In detail, at *0.5×cost*, QBN-Random's deployment cost is 50% more than SIP's, while QBN-EVF's deployment cost is very close to SIP's. However, when the On-demand cost rises to *2×cost*, the advantage of QBN-EVF decreases, and the deployment cost rises to about 1.5 times that of SIP, while the deployment cost of QBN-Random rises to about twice of SIP.

E. Convergence Efficiency of the Proposed Schemes

In the convergence efficiency analysis of the proposed federated reinforcement learning (FedSAC), the decentralized proximal policy optimization (DecPPO) [60], and the decentralized SAC (DecSAC) schemes are used as the baselines. The DecPPO scheme leverages policy gradient to train the local policy of QKD managers to maximize long-term advantages. The DecSAC scheme adopts soft-Q-learning to optimize the performance of QKD managers while maximizing the entropy of action probability. Both the DecPPO and DecSAC schemes train the QKD managers in a decentralized manner while they cannot utilize the incomplete experiences in QKD controllers' replay buffers. As shown in Fig. 9, the proposed scheme and the baseline scheme converge to the optimal QKD resource allocation policy in both network topologies. However, there is

a difference in the efficiency of their convergence. For example, the PPO scheme needs about 80 epochs to converge to the optimal QKD resource allocation in the NSFNET topology, while SAC needs about 20 epochs. With the incomplete experience of QKD controllers, FedSAC has a significant improvement in convergence efficiency, requiring only about ten epochs to converge. In the USNET topology, each scheme takes less time to converge, e.g., 40 epochs for PPO, ten epochs for SAC, and five epochs for FedSAC. The reason for this could be that the deployment cost in the USNET topology is slightly higher than in the NSFNET topology, which is a clearer signal for the DRL agents to learn.

VII. CONCLUSION

In this paper, we have studied the QKD resource allocation problem for secure federated edge learning. To protect the transmission of FEL models from eavesdropping attacks, we have proposed a hierarchical architecture to facilitate quantum-secured FEL systems. Based on this, we have formulated the optimization of the QKD resource allocation scheme as a SIP model and obtained a cost-effective scheme under uncertainty. To allocate resources in a decentralized and privacy-preserving manner, we proposed a learning-based QKD allocation scheme empowered by federated deep reinforcement learning. The proposed model and scheme have achieved a lower deployment cost of QKD resources compared to the baseline schemes as they can accommodate probabilistic variations in FEL security demand adequately. In the future, we plan to study the resource allocation problem for quantum-secured semantic communication systems in edge networks.

REFERENCES

[1] H. Wu, Z. Zhang, C. Guan, K. Wolter, and M. Xu, "Collaborate edge and cloud computing with distributed deep learning for smart city Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8099–8110, Sep. 2020.

[2] Z. Yang et al., "Energy-efficient wireless communications with distributed reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 665–679, Jan. 2022.

[3] K. Choi, L. Bedogni, and M. Levorato, "Enabling green crowdsourced social delivery networks in urban communities," *Sensors*, vol. 22, no. 4, 2022, Art. no. 1541.

[4] D. A. Chekired, M. A. Togou, L. Khoukhi, and A. Ksentini, "5G-slicing-enabled scalable SDN core network: Toward an ultra-low latency of autonomous driving service," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1769–1782, Aug. 2019.

[5] N. Cheng et al., "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.

[6] T. Huynh-The, Q.-V. Pham, X.-Q. Pham, T. T. Nguyen, Z. Han, and D.-S. Kim, "Artificial intelligence for the metaverse: A survey," 2022, *arXiv:2202.10336*.

[7] M. Xu et al., "Quantum-secured space-air-ground integrated networks: Concept, framework, and case study," *IEEE Wireless Commun.*, early access, Apr. 19, 2022, doi: 10.48550/arXiv.2204.08673.

[8] M. Xu et al., "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Commun. Surv. Tut.*, early access, Nov. 10, 2022, doi: 10.1109/COMST.2022.3221119.

[9] W. Y. B. Lim et al., "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 22, no. 3, pp. 2031–2063, Jul.–Sep. 2020.

[10] A. Huang et al., "StarFL: Hybrid federated learning architecture for smart urban computing," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 4, pp. 1–23, 2021.

[11] X. Tang et al., "Secure and trusted collaborative learning based on blockchain for artificial intelligence of things," *IEEE Wireless Commun.*, vol. 29, no. 3, pp. 14–22, Jun. 2022.

[12] W. Xuan, Z. Zhao, L. Fan, and Z. Han, "Minimizing delay in network function visualization with quantum computing," in *Proc. IEEE 18th Int. Conf. Mobile Ad Hoc Smart Syst.*, 2021, pp. 108–116.

[13] V. Bužek and M. Hillery, "Quantum copying: Beyond the no-cloning theorem," *Phys. Rev. A*, vol. 54, no. 3, 1996, Art. no. 1844.

[14] H. Weyl, "Quantenmechanik und gruppentheorie," *Zeitschrift für Physik*, vol. 46, no. 1, pp. 1–46, 1927.

[15] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Modern Phys.*, vol. 81, no. 3, 2009, Art. no. 1301.

[16] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theor. Comput. Sci.*, vol. 560, pp. 7–11, 2014, doi: 10.1016/j.tcs.2014.05.025.

[17] F. Grosshans and P. Grangier, "Continuous variable quantum cryptography using coherent states," *Phys. Rev. Lett.*, vol. 88, no. 5, 2002, Art. no. 057902.

[18] S. Wehner, D. Elkouss, and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, 2018, Art. no. eaam9288.

[19] H.-K. Lo, M. Curty, and B. Qi, "Measurement-device-independent quantum key distribution," *Phys. Rev. Lett.*, vol. 108, no. 13, 2012, Art. no. 130503.

[20] Y. Cao, Y. Zhao, Q. Wang, J. Zhang, S. X. Ng, and L. Hanzo, "The evolution of quantum key distribution networks: On the road to the Qinternet," *IEEE Commun. Surv. Tut.*, vol. 24, no. 2, pp. 839–894, Apr.–Jun. 2022.

[21] Y. Cao, Y. Zhao, J. Li, R. Lin, J. Zhang, and J. Chen, "Hybrid trusted/untrusted relay based quantum key distribution over optical backbone networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 9, pp. 2701–2718, Sep. 2021.

[22] W. Y. B. Lim et al., "Decentralized edge intelligence: A dynamic resource allocation framework for hierarchical federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 3, pp. 536–550, Mar. 2022.

[23] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[24] Z. Li, M. Xu, J. Nie, J. Kang, W. Chen, and S. Xie, "NOMA-enabled cooperative computation offloading for blockchain-empowered Internet of Things: A learning approach," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2364–2378, Feb. 2021.

[25] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.

[26] Z. Yang, Y. Hu, Z. Zhang, W. Xu, C. Zhong, and K.-K. Wong, "Reconfigurable intelligent surface based orbital angular momentum: Architecture, opportunities, and challenges," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 132–137, Dec. 2021.

[27] Y. C. Eldar, A. Goldsmith, D. Gündüz, and H. V. Poor, *Machine Learning and Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[28] C. Xu, S. Liu, Z. Yang, Y. Huang, and K.-K. Wong, "Learning rate optimization for federated learning exploiting over-the-air computation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3742–3756, Dec. 2021.

[29] J. Kang et al., "Communication-efficient and cross-chain empowered federated learning for artificial intelligence of things," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 2966–2977, Sep./Oct. 2022.

[30] S. Li et al., "A novel ISAC transmission framework based on spatially-spread orthogonal time frequency space modulation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1854–1872, Jun. 2022.

[31] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing," *IEEE Commun. Surv. Tut.*, vol. 24, no. 2, pp. 967–993, Apr.–Jun. 2022.

[32] W. Xu et al., "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," 2022, *arXiv:2206.00422*.

[33] S. Wang et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.

[34] P. Liu, G. Zhu, W. Jiang, W. Luo, J. Xu, and S. Cui, "Vertical federated edge learning with distributed integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2091–2095, Sep. 2022.

[35] W. Y. B. Lim, J. S. Ng, Z. Xiong, D. Niyato, C. Miao, and D. I. Kim, "Dynamic edge association and resource allocation in self-organizing hierarchical federated learning networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3640–3653, Dec. 2021.

[36] Y. Zhan, J. Zhang, Z. Hong, L. Wu, P. Li, and S. Guo, "A survey of incentive mechanism design for federated learning," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 2, pp. 1035–1044, Apr.–Jun. 2022.

[37] F. Toudeh-Fallah et al., "Paving the way towards 800 Gbps quantum-secured optical channel deployment in mission-critical environments," 2022, *arXiv:2202.07764*.

[38] Y.-A. Chen et al., "An integrated space-to-ground quantum communication network over 4,600 kilometres," *Nature*, vol. 589, no. 7841, pp. 214–219, 2021.

[39] M. Mehic et al., "A novel approach to quality-of-service provisioning in trusted relay quantum key distribution networks," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 168–181, Feb. 2020.

[40] Y. Cao, Y. Zhao, J. Wang, X. Yu, Z. Ma, and J. Zhang, "Cost-efficient quantum key distribution (QKD) over WDM networks," *J. Opt. Commun. Netw.*, vol. 11, no. 6, pp. 285–298, Jun. 2019.

[41] Y. Zuo, Y. Zhao, Y. Xiaosong, A. Nag, and J. Zhang, "Reinforcement learning-based resource allocation in quantum key distribution networks," in *Proc. Asia Commun. Photon. Conf., Int. Conf. Inf. Photon. Opt. Commun.*, 2020, pp. 1–3.

[42] Y. Cao, Y. Zhao, J. Li, R. Lin, J. Zhang, and J. Chen, "Multi-tenant provisioning for quantum key distribution networks with heuristics and reinforcement learning: A comparative study," *IEEE Trans. Netw. Serv. Manage.*, vol. 17, no. 2, pp. 946–957, Jun. 2020.

[43] K. Cheng et al., "SecureBoost: A lossless federated learning framework," *IEEE Intell. Syst.*, vol. 36, no. 6, pp. 87–98, Nov./Dec. 2021.

[44] Y. Liu et al., "FedBCD: A communication-efficient collaborative learning framework for distributed features," *IEEE Trans. Signal Process.*, vol. 70, pp. 4277–4290, 2022.

[45] G. Liang and S. S. Chawathe, "Privacy-preserving inter-database operations," in *Proc. Int. Conf. Intell. Secur. Inform.*, 2004, pp. 66–82.

[46] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[47] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptographic Techn.*, 1999, pp. 223–238.

[48] Y.-A. Xie et al., "Securing federated learning: A covert communication-based approach," *IEEE Netw.*, early access, Jul. 25, 2022, doi: 10.1109/MNET.117.2200065.

[49] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, 1949.

[50] C. Fung, C. J. Yoon, and I. Beschastnikh, "Mitigating sybils in federated learning poisoning," 2018, *arXiv:1808.04866*.

[51] A. Wonfor et al., "Field trial of multi-node, coherent-one-way quantum key distribution with encrypted 5×100 G DWDM transmission system," in *Proc. 45th Eur. Conf. Opt. Commun.*, IET, 2019, pp. 1–4.

[52] S. Chaisiri, B.-S. Lee, and D. Niyato, "Optimization of resource provisioning cost in cloud computing," *IEEE Trans. Serv. Comput.*, vol. 5, no. 2, pp. 164–177, Apr.–Jun. 2012.

[53] D. Chen, X. Zhang, L. Wang, and Z. Han, "Prediction of cloud resources demand based on hierarchical Pythagorean fuzzy deep neural network," *IEEE Trans. Serv. Comput.*, vol. 14, no. 6, pp. 1890–1901, Nov./Dec. 2021.

[54] M. Xu et al., "Multi-agent federated reinforcement learning for secure incentive mechanism in intelligent cyber-physical systems," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 22095–22108, Nov. 2022.

[55] X. Wang, R. Li, C. Wang, X. Li, T. Taleb, and V. C. M. Leung, "Attention-weighted federated deep reinforcement learning for device-to-device assisted heterogeneous collaborative edge caching," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 154–169, Jan. 2021.

[56] H. Hasselt, "Double Q-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, vol. 23, pp. 2613–2621.

[57] 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.

[58] X. Wang, C. Wang, X. Li, V. C. M. Leung, and T. Taleb, "Federated deep reinforcement learning for Internet of Things with decentralized cooperative edge caching," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9441–9455, Oct. 2020.

[59] S. Hardy et al., "Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption," 2017, *arXiv:1711.10677*.

[60] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

Zhaohui Yang (Member, IEEE) received the B.S. degree from Southeast University, Nanjing, China, in June 2014, and the Ph.D. degree in communication and information system with the National Mobile Communications Research Laboratory, Southeast University, in May 2018. From 2018 to 2020, he was a Postdoctoral Research Associate with King's College London, London, U.K. From 2020 to 2022, he was a Research Fellow with University College London, London. He is currently a ZJU Young Professor with the College of Information Science and Electronic Engineering Zhejiang Key Lab of Information Processing Communication and Networking, Zhejiang University, Hangzhou, China, and Research Scientist with Zhejiang Lab, Hangzhou. His research interests include joint communication, sensing, and computation, federated learning, reconfigurable intelligent surface, UAV, and NOMA. He is also an Associate Editor for IEEE COMMUNICATIONS LETTERS, *IET Communications*, and *EURASIP Journal on Wireless Communications and Networking*. He has guest edited a feature topic of IEEE COMMUNICATIONS MAGAZINE and a special issue of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was the Co-Chair of international workshops with more than 10 times, including ICC, GLOBECOM, WCNC, PIMRC, and INFOCOM. He was an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and 2020.

Zehui Xiong (Member, IEEE) received the Ph.D. degree with Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Singapore University of Technology and Design, Singapore, and an Honorary Adjunct Senior Research Scientist with Alibaba-NTU Singapore Joint Research Institute, Singapore. He was the Visiting Scholar with Princeton University, Princeton, NJ, USA, and University of Waterloo, Waterloo, ON, Canada. His research interests include wireless communications, Internet of Things, blockchain, edge intelligence, and Metaverse. He has authored or coauthored more than 180 research papers in leading journals and flagship conferences and many of them are ESI Highly Cited Papers. He was the recipient of more than 10 Best Paper Awards in international conferences and is listed in the World's Top 2% Scientists identified by Stanford University, Stanford, CA, USA. He is the Editor or Guest Editor of many leading journals, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He was the recipient of IEEE Early Career Researcher Award for Excellence in Scalable Computing, IEEE Technical Committee on Blockchain and Distributed Ledger Technologies Early Career Award, IEEE Internet Technical Committee Early Achievement Award, IEEE TCI Rising Star Award, IEEE Best Land Transportation Paper Award, IEEE CSIM Technical Committee Best Journal Paper Award, IEEE SPCC Technical Committee Best Paper Award, IEEE VTS Singapore Best Paper Award, Chinese Government Award for Outstanding Students Abroad, and NTU SCSE Best Ph.D. Thesis Runner-Up Award. He is also the Associate Director of Future Communications R&D Programme.

Minrui Xu received the B.S. degree from Sun Yat-sen University, Guangzhou, China, in 2021. He is currently working toward the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include quantum communication, federated learning, deep reinforcement learning, and mechanism design.

Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from the King Mongkuts Institute of Technology Ladkrabang, Bangkok, Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

Jiawen Kang (Member, IEEE) received the Ph.D. degree from the Guangdong University of Technology, Guangzhou, China, in 2018. From 2018 to 2021, he was a Postdoc with Nanyang Technological University, Singapore. He is currently a Professor with the Guangdong University of Technology. His research interests include blockchain, security, and privacy protection in wireless communications and networking.

Dong In Kim (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1990. He was a Tenured Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. Since 2007, he has been a SKKU-Fellowship and Distinguished Professor with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. He is currently a Fellow of the Korean Academy of Science and Technology, Seongnam, South Korea, and a Member of The National Academy of Engineering of Korea, Seoul. Since 2014, he has been a first recipient of the NRF of Korea Engineering Research Center in Wireless Communications for RF Energy Harvesting. He has been listed as a 2020 Highly Cited Researcher by Clarivate Analytics. Since 2001, he has been the Editor, Editor-at-Large, and Area Editor of *Wireless Communications* I for IEEE TRANSACTIONS ON COMMUNICATIONS. From 2002 to 2011, he was the Editor and Founding Area Editor of Cross-Layer Design and Optimization for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. From 2008 to 2011, he was the Co-Editor-in-Chief of IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS. From 2012 to 2015, he was the Founding Editor-in-Chief of IEEE WIRELESS COMMUNICATIONS LETTERS. He was selected the 2019 recipient of the IEEE Communications Society Joseph LoCicero Award for Exemplary Service to Publications. He was the General Chair of IEEE ICC 2022 in Seoul.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers university, Piscataway, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular ad hoc and sensor networks. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society. He was the recipient of the R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) presents in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015, and Education Award in 2017 from the IEEE Communications Society, Technical Recognition Award from Wireless Communications Technical Committee in 2019, and AHSN Technical Committee in 2013. He was also the recipient of the Excellent Graduate Supervision Award in 2006 from the University of Waterloo, and the Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada. He was the Technical Program Committee Chair/Co-Chair for the IEEE Globecom'16, IEEE Infocom'14, IEEE VTC'10 Fall, IEEE Globecom'07, Symposia Chair of IEEE ICC'10, and Chair of the IEEE Communications Society Technical Committee on *Wireless Communications*. He is the elected IEEE Communications Society Vice President for Technical and Educational Activities, Vice President for Publications, Member-at-Large on the Board of Governors, Chair of the Distinguished Lecturer Selection Committee, and a Member of IEEE Fellow Selection Committee. He was/is the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, *IET Communications*, and *Peer-to-Peer Networking and Applications*.