Resource Management and Reflection Optimization for Intelligent Reflecting Surface Assisted Multi-Access Edge Computing Using Deep Reinforcement Learning

Zhaoying Wang, Yifei Wei, *Member, IEEE*, Zhiyong Feng, *Senior Member, IEEE*, F. Richard Yu, *Fellow, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*— Multi-access edge computing (MEC) enables the computation-intensive and latency-critical application to be processed at the network edge, which reduces the transmission latency and energy consumption. The quality of the wireless channel seriously affects the performance of the edge network. Consequently, the performance of the edge network can be significantly improved from the perspective of communication. The recently advocated intelligent reflecting surface (IRS) intelligently controls the radio propagation environment to improve the quality of wireless communication links. This paper proposes an edge heterogeneous network with the assistance of intelligent reflecting surface. Specifically, the macro base station and small base stations are equipped with MEC servers, and IRS is adopted to provide an additional computation offloading link. The user association, computation offloading and resource allocation, as well as IRS phase shift design are optimized with the aim of minimizing the long-term energy consumption subject to the constraints imposed on quality of service (QoS) and available resources. The challenge of the optimization problem is rooted from the fact that update timescale of user association is different from others. Hence, a two-timescale mechanism is invoked by marrying tools from matching theory and deep reinforcement learning. More specifically, the user association decision takes place in the long timescale. In the short timescale, the computation offloading, resource allocation and IRS phase shift design strategy is performed. The effectiveness of the proposed two-timescale mechanism is verified by the simulation results.**

Manuscript received 2 March 2022; revised 14 July 2022; accepted 24 August 2022. Date of publication 14 October 2022; date of current version 13 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61871058, in part by the National Key Research and Development Program of China under Grant 2020YFA0711303, and in part by NSF under Grant CNS-2107216 and Grant CNS-2128368. The associate editor coordinating the review of this article and approving it for publication was S. Dey. *(Corresponding author: Yifei Wei.)*

Zhaoying Wang and Yifei Wei are with the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: wangzhaoying@bupt.edu.cn; weiyifei@bupt.edu.cn).

Zhiyong Feng is with the Key Laboratory of Universal Wireless Communications, Ministry of Education, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: fengzy@bupt.edu.cn).

F. Richard Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: Richard_Yu@carleton.ca).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: hanzhu22@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3202948.

Digital Object Identifier 10.1109/TWC.2022.3202948

***Index Terms*— Multi-access edge computing, intelligent reflecting surface, resource allocation, matching theory, deep reinforcement learning.**

I. Introduction

THE forthcoming sixth-generation (6G) network and the booming Internet of Things (IoT) technology contribute to an exponential growth of intelligent devices. The emergence of novel applications and services (e.g., autonomous vehicles, ultra-high-definition (UHD) video streams and augmented reality (AR), etc.) put forward higher requirements on bandwidth, latency, reliability, and energy consumption. Due to the shortfalls of high latency and high energy consumption caused by processing tasks in the remote cloud, the centralized cloud computing is incapable of ensuring the quality of service (QoS) for users [1]. To combat the above issue, a new paradigm multi-access edge computing (MEC) [2] is introduced to deploy computing, storage and control functions at the network edge (e.g., access points and base stations, etc.), which enables the resource-constrained mobile devices to execute the computation-intensive and latency-critical applications. Therefore, the computation tasks of terminal user equipment can be offloaded to the MEC server in the edge network for executing, thereby reducing the transmission latency and the energy consumption of user equipment, as well as alleviating the backhaul burden.

The computation offloading problem is formulated in the edge network with the consideration of whether to offload and which part to offload [3]. The authors in [4] consider the binary offloading scheme, where the application is offloaded to the MEC server as a whole or executed entirely on local equipment. The work in [5] and [6] considers partial offloading where the application consists of multiple procedures/components (e.g., AR application), and some components is executed on the user equipment and another part is executed at the network edge. Inter-user interference exists on both wireless communication links and edge computing nodes due to the limited resources, which impairs the overall performance of the edge network. Therefore, computation offloading and resource allocation are jointly considered in the recent literature. The transmission power allocation policy is proposed in [7] with the goal of minimizing the system energy consumption. The research in [8] formulates the bandwidth and computation resource optimization problem under QoS

1536-1276 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

guarantee constraints, and proposes an alternating direction multiplier based algorithm to solve the problem. However, due to the random channel fading characteristic, the quality of the computation offloading link between users and edge computing nodes cannot be guaranteed, which affects the data rate and cannot meet the needs of end users. There exist three typical methods to increase the data rate of wireless communication [9]. The first is to deploy more heterogeneous nodes (such as small cells) in the network to improve access availability and spectrum utilization. The second is to add more antennas at the base station to increase channel gain through massive Multiple Input Multiple Output (MIMO) technology. The third is to extend the available bandwidth with higher frequency bands such as mmWave. These promising technologies generate high hardware and energy costs, complex signal processing problem, and unable to intelligently adjust random channels while increasing wireless communication data rates.

Recently, a new paradigm Intelligent reflecting surface (IRS) is invoked to realize intelligent and reconfigurable wireless propagation environment in 6G wireless communication systems [10]. The surface is two-dimensional artificial electromagnetic material (namely metasurface), which consists of considerable passive reflection elements with special physical structures. The IRS controller implements intelligent control of the physical channel by adjusting the amplitude and phase shift of the passive reflective elements in a software-defined manner. Thus, the ideal multipath effect can be realized by adjusting the reflection amplitude and phase of the incident radio frequency (RF) signals. Subsequently, the received signal power can be enhanced through coherently adding the reflection RF signals and the interference can be mitigated via destructively combining signals [11]. Recent work focuses on integrating intelligent reflecting surface into traditional wireless networks to improve communication performance [12], such as channel modeling [13], channel estimation [14], [15], and passive reflection optimization in different scenarios [16], [17], [18], etc. In addition, IRS is utilized in novel scenarios, such as IRS-assisted physical layer security [19], [20] and IRS-aided wireless power transfer [21] to improve system performance.

IRS is expected to effectively enhance the communication and computation performance of edge network in recent research [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. By deploying IRS between users and edge servers, IRS provides auxiliary links for users through passive beamforming, which increases the wireless link capacity, thus the computation-intensive tasks can be offloaded to the edge servers without high computation latency and transmission energy consumption. Most of the existing work focuses on single-cell scenarios in IRS-assisted MEC systems [22], [23], [24], [25], [27], [28], [29], [30]. However, the multicell scenarios are considered in a paucity of the IRS-assisted MEC research work [26], [32]. For large-scale edge network with abundant users and edge servers, the deployment of IRS plays a crucial role in computation offloading and resource allocation strategies. IRS can be utilized to assist in offloading computation tasks to different MEC servers in order to achieve high resource utilization and low computation latency. Specifically, IRS can adjust the offloading channel of certain users to different servers with less computation burden, instead of offloading to the same adjacent server that would cause high computation latency. Therefore, this paper innovatively proposes edge heterogeneous network scenarios with IRS assistance to minimize energy consumption by optimizing user association, computation offloading and resource allocation, as well as IRS phase shift.

The formulation problems with coupled optimization variables are generally non-convex in the IRS-assisted MEC systems. Therefore, previous literature mainly employs alternate optimization to solve the radio and computation resource allocation subproblem and the IRS phase shift design subproblem separately [26], [32], [33]. The alternating optimization provides near-optimal solution with guaranteed convergence. However, due to the high computation complexity and execution time, the above solution may hinder the practical application of IRS in edge networks. Deep reinforcement learning (DRL) can solve complex optimization problems in the wireless communication system by adopting adaptive modeling and intelligent learning [34], [35], [36]. Few authors utilize the DRL algorithms to solve optimization problems in IRS-assisted MEC systems. The research work in [37] proposes the DRL algorithm to maximize total utilities of users in the IRS assisted wireless powered mobile edge computing network. An asynchronous actor-critic DRL based computation offloading scheme with reconfigurable intelligent surface assistance is designed in [9] to minimize the total latency of users for task execution. Therefore, the DRL algorithm is leveraged in this paper to learn resource management and reflection optimization strategy.

The main contributions and innovations of this work are summarized below:

- This paper proposes an IRS-assisted edge heterogeneous network including the macro base station and multiple small base stations equipped with MEC servers. The IRS provides auxiliary links for users and intelligently controls the channel status to enhance the communication performance between users and base stations, and achieve efficient resource utilization. With the aim of minimizing the long-term energy consumption of all users while guaranteeing the QoS (e.g., latency requirements) of users, a two-timescale mechanism is invoked to optimize the user association, computation offloading and resource allocation, as well as IRS phase shift in this paper.
- For the long timescale user association problem, matching theory with low complexity is adopted to perform one-to-many matching based on two sides' preferences between users and BSs. Since the interference between users matched with SBSs affects the transmission rate, we utilize swap matching to deal with the interdependence among users' preferences (externalities).
- Markov decision process (MDP) is applied to model the short timescale optimization problem which can be solved through the reinforcement learning (RL) algorithm. To deal with the high-dimension state space, the value functions in RL are approximated by deep

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

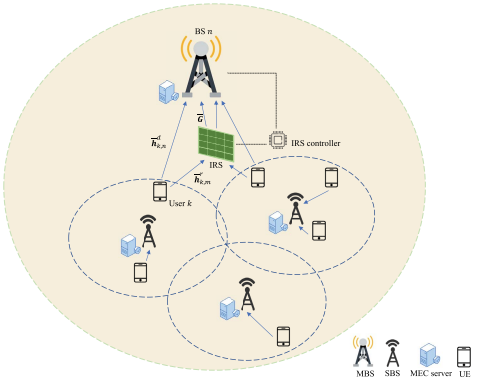

Fig. 1. System model.

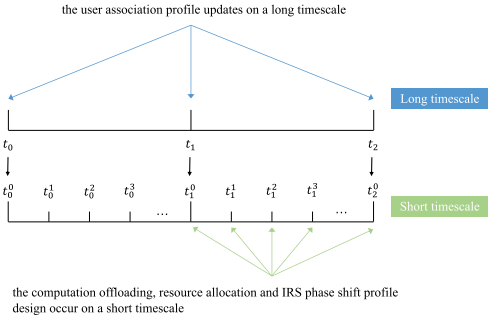

Fig. 2. Graphical illustration of two-timescale model.

neural network (DNN), as well as experience replay and independent target network techniques to speed up the convergence of DRL algorithm. Specifically, the Deep Q-network (DQN) algorithm is introduced to learn computation offloading, resource allocation and IRS phase shift design policy.

- The simulation result validates the convergence of the proposed two-timescale algorithm. In contrast with the benchmark schemes, the performance of energy consumption is demonstrated in different simulation environments. The proposed algorithm shows that a suitable IRS phase shift design can provide the passive beamforming gain, thereby reducing the energy consumption of edge network.

The organization of this paper is listed as follows. The system models are showed in Section II. The Section III introduces the optimization problem. The two-timescale mechanism for resource management and reflection optimization is proposed in Section IV. Section V give the simulation parameters and results. Finally, this paper is concluded in Section VI.

Notation: In this paper, italic letters represent scalars. Vectors and matrices are indicated by boldface lowercase and uppercase letters, respectively. The superscript $(\cdot)^T$ and $(\cdot)^H$ represent transpose operation and Hermitian transpose operation, respectively. $\mathbb{R}^{M\times N}$ represents real matrices with the space of $M\times N$. $\mathbb{C}^{M\times N}$ represents complex matrices with the space of $M\times N$.

II. System Models

This paper considers multiple single-antenna base stations (BSs) in edge heterogeneous networks, as shown in Fig. 1. $\mathcal{N}=\{0,1,\ldots,N\}$ denotes the set of BSs, and the symbol n represents the nth base station, where $n=0$ represents the macro base station (MBS) and $n\in\{1,\ldots,N\}$ denotes the small base stations (SBSs). Each BS is equipped with a multi-core MEC server and the number of CPU cores of BS n is C_n, which can simultaneously serve at most C_n users. The set of single-antenna user equipment (UE) is expressed as $\mathcal{K}=\{1,\ldots,K\}$, and the symbol k is used to denote the kth UE. The set of users associated with BS n is denoted as $\mathcal{K}_n=\{1,2,\ldots,K_n\}$, where K_n is the number of users associated with BS n and k_n refer to the kth user associated with the BS n. Each user has latency-critical applications including multiple procedures/components with dependency. Each component of the task is computed on the local equipment or on the MEC server. Since the computing resource of the SBSs are limited, and to avoid frequent handovers, users can associate with the MBS for computation offloading. As the cell-edge users are far away from the MBS, the quality of the channel is poor. Therefore, it is assumed that the IRS is utilized to assist the users to associate with MBS. There exist M IRS reflection elements and the mth reflection element is represented by symbol m. A smart controller connected to the IRS dynamically adjusts the reflection elements and also exchanges control information with the MBS via a separate wireless link. The base station controller connected to all BSs is responsible for resource management and reflection optimization [38]. The MBS is considered as a centralized controller in this paper [39].

Since the update timescale of user association is larger than the timescale of computation offloading, resource allocation and IRS phase shift design, the IRS-assisted edge heterogeneous network scenario is modeled as a two-timescale edge computing model. Hence, the process of the IRS-assisted edge heterogeneous network occurs on two different timescales: the user association profile updates on a long timescale, and computation offloading, resource allocation and IRS phase shift profile design occur on a short timescale, as shown in Fig. 2. The basic time unit of the long timescale is defined as epoch. $\mathcal{I}=\{0,1,..,I\}$ is adopted to represent the index set of user association profile starting at each epoch $t_i, \{t_i|i\in\mathcal{I}\}$. Each epoch can be divided into a set of time slots which is denoted as $\mathcal{J}=\{0,1,..,J\}$. Computation offloading, resource allocation and IRS phase shift design are executed at each time slot $t_i^j, \{t_i^j|i\in\mathcal{I}, j\in\mathcal{J}\}$, where t_i^J is the maximum task execution latency. Without loss of generality, we omit (t_i) and (t_i^j) in the following expressions, unless epoch t_i and time slot t_i^j are emphasized.

Next, we describe the system model that includes the communication model in Subsection II-A, and the computation model in Subsection II-B.

A. Communication Model

In this paper, $x_{k,n}\,(t_i)\in\{0,1\}$ represents the user association variable, where $x_{k,n}=1$ indicates that user k associates with BS n, otherwise 0. At each epoch, each user can be served by only one BS: $\sum_{n=0}^{N}x_{k,n}=1$. On the condition

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

of $x_{k,0} = 1$, user k associates with the MBS with two links: user-BS direct link and user-IRS-BS reflection link. If $x_{k,n} = 1, \forall n \in \mathcal{N}, n \neq 0$, user k is associates SBS with user-BS direct link. The direct channel of user k and BS n (baseband equivalent time-domain channel) is denoted as $\bar{\boldsymbol{h}}_{k,n}^{d} \in \mathbb{C}^{L_{k,n}^{d} \times 1}$. The time-domain channel of the MBS and IRS reflection elements is define as $\bar{\boldsymbol{G}} = [\bar{\boldsymbol{g}}_1, \ldots, \bar{\boldsymbol{g}}_m] \in \mathbb{C}^{L_0 \times M}$. The time-domain channel of user k and IRS reflection element m is expressed as $\bar{\boldsymbol{h}}_{k,m}^{r} \in \mathbb{C}^{L_k \times 1}$. $L_{k,n}^{d}$, L_0 and L_k are the number of delayed taps of the corresponding link, respectively. The above channels are assumed to remain approximately constant at each time slot. Large-scale path loss and small-scale fading are taken into account in the communication model. Furthermore, the direct channel of kth user and nth BS is assumed to follow the exponential power-delay feature for each multipath channel: $\left[\bar{\boldsymbol{h}}_{k,n}^{d}\right]_l = \sqrt{\varrho_{k,n}^{d} \frac{1-\alpha}{1-\alpha^{L_{k,n}^{d}}}} \alpha^{l/2} \nu_l, \forall l = 0, .., L_{k,n}^{d}$, where $\varrho_{k,n}^{d}$ denotes the large-scale path loss, and $0 < \alpha < 1$. Small-scale fading ν_l follows the complex Gaussian distribution with zero mean and unit variance $\nu_l \sim \mathcal{CN}(0, 1)$ [40]. The above expression is also applicable to the channel of the MBS and the IRS, and the channel of the users and the IRS, which will not be described here. The path loss is defined as $\varrho = \varrho_0 (d/d_0)^{-\beta}$, where β denotes the path loss exponent. ϱ_0 denotes the reference path loss at reference distance d_0. The IRS phase shift matrix is expressed as $\boldsymbol{\phi}\left(t_i^j\right) = \left[\phi_1\left(t_i^j\right), \ldots, \phi_M\left(t_i^j\right)\right]^T \in \mathbb{C}^{M \times 1}$, where $\phi_m = \beta_m e^{j\theta_m}$, $\beta_m \in [0, 1]$ denotes the amplitude and $\theta_m \in [0, 2\pi]$ denotes the phase. In this paper, the amplitude is set to a maximum value of 1, and the discrete phase shift design with ρ phases is considered. The phase set is denoted as $\theta_m \in \{0, \Delta\theta, \ldots, (\rho - 1)\Delta\theta\}, \Delta\theta = \frac{2\pi}{\rho}$. The time-domain effective reflection channel through reflection element m is denoted as the convolution of the user-IRS channel, the IRS reflection coefficient and the IRS-MBS channel: $\bar{\boldsymbol{h}}_{k,m}^{r} * \phi_m * \bar{\boldsymbol{g}}_m = \phi_m \bar{\boldsymbol{h}}_{k,m}^{r} * \bar{\boldsymbol{g}}_m \in \mathbb{C}^{L_k^r \times 1}$, where $L_k^r = L_0 + L_k - 1$ is the corresponding number of delayed taps and $*$ denotes the convolution operation. We adopt Orthogonal Frequency-Division Multiple Access (OFDMA) in our work. The number of equally divided sub-bands is B, and the set is denoted as $\mathcal{B} = \{1, \ldots, B\}$, the bth sub-band is represented by symbol b [27], [40]. The orthogonal frequency spectrum is assumed among users associated with the same BS, as well as users associated with the MBS and SBS. $\mathcal{B}_0$ and $\mathcal{B}_n, \forall n \in \{1, \ldots, N\}$ are represented as the set of sub-bands allocated to the MBS and SBSs, respectively where $\mathcal{B}_0 \cap \mathcal{B}_n = \emptyset, \forall n \in \{1, \ldots, N\}$. Therefore, we consider the interference between SBSs. The sub-band allocation variable $y_{k,b}\left(t_i^j\right) \in \{0, 1\}$ indicates whether sub-band b is allocated to the user k, where $y_{k,b} = 1$ denotes that sub-band b is allocated to user k, otherwise 0. At each time slot, each sub-band can only be allocated to at most one user associated with any BS: $\sum_{k=1}^{\mathcal{K}_n} y_{k,b} \leq 1, \forall n \in \mathcal{N}, \forall b \in \mathcal{B}$. Each user should be allocated to at least one sub-band: $\sum_{b=1}^{B} y_{k,b} \geq 1, \forall k \in \mathcal{K}$. The transmit power on sub-band b is expressed as $p_{k,b}\left(t_i^j\right)$, where the transmit power is assumed to be a discrete set: $p_{k,b}\left(t_i^j\right) \in \{0, p_1, p_{2,\ldots,} P_k^{max}\}$. And the relationship between the transmit power with the sub-band allocation variable $y_{k,b}$ is defined as,

$$y_{k,b} = \begin{cases} 0, & \text{if } p_{k,b} = 0, \\ 1, & \text{if } p_{k,b} > 0. \end{cases} \tag{1}$$

And $\sum_{b=1}^{B} p_{k,b} \leq P_k^{max}, \forall k \in \mathcal{K}$, where P_k^{max} denotes the maximum transmit power of kth user.

The zero-padded concatenated IRS-BS channel and user-IRS channel of IRS reflection element m is denoted as $\boldsymbol{h}_{k,m} = \left[\left[\bar{\boldsymbol{h}}_{k,m}^{r} * \bar{\boldsymbol{g}}_m\right]^T, 0, \ldots, 0\right]^T \in \mathbb{C}^{B \times 1}$. Thus, the zero-padded concatenated channel between user k and IRS, and the MBS can be uniformly expressed as $\boldsymbol{H}_k = [\boldsymbol{h}_{k,1}, \ldots, \boldsymbol{h}_{k,M}] \in \mathbb{C}^{B \times M}$. And we denote the zero-padded concatenated user-BS channel as $\boldsymbol{h}_{k,n}^{d} = \left[\bar{\boldsymbol{h}}_{k,n}^{d}, 0, \ldots, 0\right] \in \mathbb{C}^{B \times 1}$. Consequently, the superposed channel impulse response (CIR) is derived as,

$$\boldsymbol{h}_{k,n}^{CIR} = x_{k,0} \boldsymbol{H}_k \boldsymbol{\phi} + \boldsymbol{h}_{k,n}^{d}, \tag{2}$$

where $L_{k,n} = max\left(L_{k,n}^{d}, L_k^r\right)$ stands for the number of delayed taps. To eliminate the inter-symbol interference (ISI), the number of cyclic prefixes is assumed to be not less than the maximum number of delayed taps. The channel frequency response (CFR) on sub-band b of user k associated with BS n is defined as,

$$h_{k,n,b}^{CFR} = x_{k,0} \boldsymbol{f}_b^H \boldsymbol{H}_k \boldsymbol{\phi} + \boldsymbol{f}_b^H \boldsymbol{h}_{k,n}^{d}, \tag{3}$$

where $\boldsymbol{f}_b^H$ is the bth row of discrete Fourier transform (DFT) matrix $\boldsymbol{F}_B$. The $B \times B$ DFT matrix $\boldsymbol{F}_B$ is defined as $[F_B]_{\imath,\jmath} = e^{-j\frac{2\pi \imath \jmath}{B}}, 0 \leq \imath, \jmath \leq B - 1$. It is assumed that the perfect knowledge of channel $\boldsymbol{h}_{k,n}^{d}$ and $\boldsymbol{H}_k$ are available at BSs.[1] The achievable rate (bps) of user k associated with BS n is defined as (4), shown at the bottom of the next page, where Γ is the gap between a specific modulation and coding scheme and the channel capacity. The receiver noise at each sub-band is modeled as an independent circularly symmetric complex Gaussian (CSCG) random variable with zero mean and variance σ^2.

B. Computation Model

In fact, mobile applications consist of multiple procedures/components (such as face recognition and AR applications), and so it is necessary to partially offload the users' computation tasks to the MEC servers. Task models of partial offloading consist of data partition model [41] and task partition model [42]. The task input size of the data partitioning model is bit-wise independent. The task can be divided into groups of any size, and then are computed on the user equipment and the MEC server concurrently. Whereas, the dependencies between the components of application cannot be ignored in certain applications. Hence, a typical directed acyclic graph (DAG) task-call graph is utilized in the task partition model. $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is denoted as the task-call graph, where the set of vertices $\mathcal{V}$ stands for the set of component and

[1] Naturally, the assumption is idealistic. Therefore, the algorithm proposed in this paper can be regarded as representing the best-case bound for the energy performance of realistic scenarios.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

the set of edges $\mathcal{E}$ represents the dependency between subtasks. Typical dependency models of subtasks include sequential dependency, parallel dependency, and general dependency. The sequential dependency task model is considered in this paper (e.g., immersive applications [43] or deep neural network models [44], [45]). The subtasks can be computed locally on the user equipment, or offloaded to the MEC servers for computation.

Tuple $\mathcal{J}_k^v \triangleq (v, \chi_k^v, d_k^{u,v}, d_k^{v,w})$ represents subtask v of user's application $\mathcal{G}_k$, where $d_k^{u,v}$ stands for the input data size of subtask v and u is the previous task of subtask v. $d_k^{v,w}$ expresses the output data size of subtask v and w is the next task of subtask v. χ_k^v (cycles/byte) denotes the number of clock cycles performed by the microprocessor per byte of data. The maximum tolerable latency of user k is expressed as T_k^{max}.

BS n manages a virtual task queue in each time slot t_i^j to store the computation requests of users associated with it and the queue is represented by $\boldsymbol{Q}_n\left(t_i^j\right) = \{\boldsymbol{q}_{1,n}\left(t_i^j\right), \ldots, \boldsymbol{q}_{l,n}\left(t_i^j\right)\}$, where l is the maximum length of the task queue, $\boldsymbol{q}_{i,n} = \{k, \mathcal{J}_k^v\}$ denotes the parameter vector of element i in the task queue. The set of computation offloading decision for the task queue at time slot t_i^j is expressed as $\boldsymbol{z}_n\left(t_i^j\right) = \{z_{1,n}\left(t_i^j\right), \ldots, z_{l,n}\left(t_i^j\right)\}$, where $z_{i,n} \in \{0, 1\}$ denotes the computation offloading variable of element $\boldsymbol{q}_{i,n}$. When the subtask i is offloaded to the MEC server: $z_{i,n} = 1$, if the subtask i is executed locally on user equipment: $z_{i,n} = 0$. At time slot t_i^j, the total execution time of the task queue $\boldsymbol{Q}_n\left(t_i^j\right)$ is derived as

$$T_n^{exe}\left(t_i^j\right) = \sum_{i=1}^{l} T_{i,n}^{exe}\left(t_i^j\right), \tag{5}$$

where $T_{i,n}^{exe}\left(t_i^j\right)$ is defined as

$$T_{i,n}^{exe}\left(t_i^j\right) = \begin{cases} \frac{d_k^{u,v}\chi_k^v}{f_k^l}, & \text{if } z_{i,n} = 0, \\ \left(1 - z_{i,n}\left(t_i^{j-1}\right)\right)\frac{d_k^{u,v}}{r_{k,n}} + \frac{d_k^{u,v}\chi_k^v}{f_n^c}, & \text{if } z_{i,n} = 1, \end{cases} \tag{6}$$

where f_k^l denotes the CPU frequency of user k and f_n^c expresses the CPU core frequency of the MEC server associated with BS n. At each time slot, the computation capability of the MEC server is limited: $\sum_{i=1}^{l} z_{i,n} \leq C_n$.

At time slot t_i^j, the total energy consumption of the task queue $\boldsymbol{Q}_n\left(t_i^j\right)$ is expressed as

$$E_n^{exe}\left(t_i^j\right) = \sum_{i=1}^{l} E_{i,n}^{exe}\left(t_i^j\right), \tag{7}$$

where $E_{i,n}^{exe}\left(t_i^j\right)$ is obtained as

$$E_{i,n}^{exe}\left(t_i^j\right) = \begin{cases} \zeta_{mob}\left(d_k^{u,v}\right)\chi_k^v\left(f_k^l\right)^2, & \text{if } z_{i,n} = 0, \\ \left(1 - z_{i,n}\left(t_i^{j-1}\right)\right)\sum_{b=1}^{B} p_{k,b}\frac{d_k^{u,v}}{r_{k,n}} & \\ +\zeta_e d_k^{u,v}\chi_k^v\left(f_n^c\right)^2, & \text{if } z_{i,n} = 1. \end{cases} \tag{8}$$

ζ_{mob} and ζ_e denote the effective capacitance coefficients that are determined by the chip architecture of user equipment and MEC server, respectively [46], [47]. It is worth noting that we only consider the uplink execution latency and energy consumption, and the downlink latency and energy consumption are ignored in this paper [44].

III. Problem Formulation

In this paper, the optimization problem is formulated to minimize the system energy consumption over the entire time horizon while satisfying the QoS of users, i.e.,

$$\min_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{P},\boldsymbol{\phi},\boldsymbol{Z}} \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{n=0}^{N} E_n^{exe}\left(t_i^j\right) \tag{9}$$

$$\textit{s.t.} \quad \sum_{n=0}^{N} x_{k,n}\left(t_i\right) = 1, \forall k \in \mathcal{K}, \tag{9a}$$

$$\sum_{k=1}^{\mathcal{K}_m} y_{k,b}\left(t_i^j\right) \leq 1, \forall m \in \mathcal{N}, \forall b \in \mathcal{B}, \tag{9b}$$

$$\sum_{b=1}^{B} y_{k,b}\left(t_i^j\right) \geq 1, \forall k \in \mathcal{K}, \tag{9c}$$

$$\sum_{b=1}^{B} p_{k,b}\left(t_i^j\right) \leq P_k^{max}, \forall k \in \mathcal{K}, \tag{9d}$$

$$\sum_{i=1}^{l} z_{i,n}\left(t_i^j\right) \leq C_n, \forall n \in \mathcal{N}, \tag{9e}$$

$$T_{k,n}^{exe}\left(t_i\right) \leq T_k^{max}, \forall k \in \mathcal{K}, \tag{9f}$$

Here $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K] \in \mathbb{R}^{(N+1)\times K}$ denotes the user association matrix where $\boldsymbol{x}_k = [x_{k,0}, x_{k,1}, \ldots, x_{k,N}]^T \in \mathbb{R}^{(N+1)\times 1}$. $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K] \in \mathbb{R}^{B\times K}$ stands for the sub-band allocation matrix where $\boldsymbol{y}_k = [y_{k,1}, \ldots, y_{k,B}]^T \in \mathbb{R}^{B\times 1}$. $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K] \in \mathbb{R}^{B\times K}$ expresses the power allocation matrix where $\boldsymbol{p}_k = [p_{k,1}, \ldots, p_{k,B}]^T \in \mathbb{R}^{B\times 1}$. $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_M]^T \in \mathbb{C}^{M\times 1}$ represents the phase shift matrix. $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_l] \in \mathbb{R}^{(N+1)\times l}$ shows the computation offloading matrix. Constraint (9a) restricts that only one BS can be associated with each user. Constraints (9b) and (9c) guarantee that each sub-band can only be allocated to at most one user associated with any BS at each time slot and each user should be allocated to at least one sub-band, respectively. Constraint (9d) reflects the power of all sub-bands allocated

$$r_{k,n} = \sum_{b=1}^{B} x_{k,n} y_{k,b} W log_2\left(1 + \frac{p_{k,b}\left|h_{k,n,b}^{CFR}\right|^2}{\sum_{m=1,m\neq n}^{N}\sum_{l\in\mathcal{K}_m} p_{l,b}\left|\boldsymbol{f}_b^H \boldsymbol{h}_{l,n}^d\right|^2 + \Gamma\sigma^2}\right), \tag{4}$$

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

to each user cannot exceed the maximum transmit power. Constraint (9e) reveals that the MEC server connected to BS n can serve at most C_n users at the same time. Constraint (9f) ensures that the execution time of user's application should meet the requirement of the maximum tolerable latency, where $T_{k,n}^{exe}(t_i) = \sum_{j=1}^{J} T_{k,n}^{exe}\left(t_i^j\right)$.

The optimization problem is a nonlinear integer programming with variables of different timescales, which indicates the problem is generally NP-hard. To combat the above issue, this paper adopts a two-timescale mechanism for solving the long timescale variables and the short timescale variables separately. Specifically, the matching theory is employed to obtain the user association decision in the long timescale. In the short timescale, the computation offloading, sub-band and power allocation, as well as IRS phase shift strategy can then be learned using deep reinforcement learning. The reason for employing the matching theory to address the user association problem depends on the advantage of low complexity in comparison with the traditional solutions, e.g., the exhaustive search. To be specific, the complexity is $\mathcal{O}\left(N^K\right)$ in the exhaustive search method, which leads to the exponential growth in terms of the users' number, while the complexity is $\mathcal{O}\left(K^2\right)$ in the matching theory method [48]. Since the dynamics of the environment which including wireless channels, computation requests, and resource states can affect the computation offloading, resource allocation and IRS phase shift design decisions, the short timescale optimization problem is viewed as the sequential decision problem. Therefore, the short timescale problem is modeled as MDP and solved by reinforcement learning.

IV. The Proposed Two-Timescale Mechanism for Resource Management and Reflection Optimization

First, we introduce an outline of the proposed resource management and reflection optimization scheme. The details of proposed scheme will be described in the Subsections IV-A and IV-B. In the beginning of each epoch, we carry out a user association scheme. Since each user is allowed to associates with only one BS, and each BS can serve multiple users, we develop a one-to-many matching algorithm to associate each BS with multiple users. Furthermore, due to the high-dimension state space of the short timescale RL-based framework, the DRL approach is then employed to learn the computation offloading, resource allocation and IRS phase shift design scheme in each time slot.

A. User Association Using Matching Theory

In this paper, the two-sided matching game is utilized to model the long-timescale user association problem where there exist two disjoint sets of players, the user set $\mathcal{K}$, and the BS set $\mathcal{N}$. In the proposed game, each user can be matched with one BS while each BS can be matched with multiple users. Thus, a one-to-many matching is taken into account and defined as follows.

Definition 1: The proposed one-to-many matching game consists of two sets of players, $\mathcal{K}$ and $\mathcal{N}$, and the matching x is defined as a function from $\mathcal{K} \times \mathcal{N}$ to the set of all subsets of $\mathcal{K} \times \mathcal{N}$ with

$$\begin{aligned} |x(k)| &= 1, \forall k \in \mathcal{K}, \\ |x(n)| &\leq K, \forall n \in \mathcal{N}, \\ n &= x(k) \Leftrightarrow k \in x(n). \end{aligned}$$

Therefore, the user association indicator can be specified from a matching x,

$$x_{k,n}(t_i) = \begin{cases} 1, & \text{if } n = x(k), \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Each user aims to be associated with the BS which enables the user to achieve its maximum utility. Hence, the user's matching preference over the BSs is sorted in the descending order based on the achievable rates. The preference profile of the BS is defined over all users which minimizes the energy consumption. Therefore, the matching preference of the BS over the users is based on the negative energy consumption in the descending order. The preference profile of user k is represented by a vector of the utility $\psi_k(x)$ which is defined as follows:

$$\psi_k(x) = r_k(\boldsymbol{x}_k, \boldsymbol{Y}, \boldsymbol{P}, \boldsymbol{\phi}), \tag{11}$$

and the preference profile of BS n is represented by a vector of utility $\psi_n(x)$ which is defined as:

$$\psi_n(x) = -E_n^{exe}(\boldsymbol{X}, \boldsymbol{z}_n, \boldsymbol{Y}, \boldsymbol{P}, \boldsymbol{\phi}), \tag{12}$$

where $z_n, \boldsymbol{Y}, \boldsymbol{P}, \boldsymbol{\phi}$ are obtained by the short-timescale computation offloading, resource allocation and IRS phase shift design scheme in Subsection IV-B.

Since the rate is affected by interference between the users which are associated with the SBSs, the preference of user also hinges on the association situation of other users. Therefore, the preference dynamically changes with the matching state of other players and the interdependency between the preferences of players is defined as externalities [49], [50]. Therefore, one-to-many matching with externalities can be solved by swap matching [51].

Definition 2: Given a one-to-many matching x with $k \in x(n)$, and $k' \in x(n'), k, k' \in \mathcal{K} \quad n, n' \in \mathcal{N}$, a swap matching is defined as $x_{kn}^{k'n'} = \{x \backslash \{(k,n),(k',n')\} \cup \{(k,n'),(k',n)\}\}$.

A swap matching allows one pair of users (k, k') to switch their matched BSs (n, n') while keeping other user-BS matchings unchanged.

Definition 3: For the matching x, (k, k') is a swap-blocking pair if and only if [50].

1) $\forall u \in \{k, k', n, n'\}, \psi_u(x_{kn}^{k'n'}) \geq \psi_u(x)$,
2) $\exists u \in \{k, k', n, n'\}, \psi_u(x_{kn}^{k'n'}) > \psi_u(x)$,
3) the constraint (9f) is satisfied.

Hence, two users exchange their respective matched BSs on condition that after the swap matching operation between a swap-blocking pair, 1) the utilities of both users and BSs will not decrease, 2) the utility of at least one increases, 3) the latency constraint of each user is not violated.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

Definition 4 (Two-Sided Exchange Stability): A matching x^* is two-sided exchange stable if swap-blocking pairs don't exist [51].

Algorithm 1 One-to-Many Matching with Externalities based User Association

1. **Initialization:** Choose a random matching x while the constraint (9f) is satisfied. And calculate (11) and (12).
2. **repeat**
3. Choose user $k \in \mathcal{K}$,$x(k) = n$ and user $k' \in x(n')$.
4. **if** the pair of users (k, k') is a swap-blocking pair in the current matching
5. Update $x \leftarrow x_{kn}^{k'n'}$;
6. Calculate (11) and (12).
7. **until** There exist no swap-blocking pairs in the current matching.
8. The two-sided exchange-stable matching x^* is obtained and then the user association indicator is obtained according to (10).
9. **return** Stable one-to-many matching results.

The matching based algorithm is summarized as follows. Firstly, a matching x is randomly initialized under the condition that the QoS of users is satisfied, and the utility of the user and BS is calculated according to (11) and (12). Then, the iterative process is looped to find swap-blocking pair, update swap matching and calculate utility until two-sided exchange stable matching is reached, thus determining the user association strategy. Algorithm 1 presents the one-to-many matching with externalities based user association algorithm. Each swap operation reduces the system energy consumption strictly and generates a new matching. After finite number of iterations, the algorithm will converge to a stable matching owing to the limited number of users and BSs, which ensures the convergence of the Algorithm 1. Consequently, the exchange between any two users will not reduce the system energy consumption, which achieves a local optimal solution.

B. Deep Q-Network Based Computation Offloading, Resource Allocation and IRS Phase Shift Design

The user association strategy $\boldsymbol{X}^*$ is obtained using matching theory in the Subsection IV-A. Subsequently, Markov decision process is employed to model the short timescale optimization problem. The MDP consists of a five-elements tuple $\langle S, A, P, R, \gamma \rangle$, where S is the state space, A is the action space, P is the state transition probability, R is the reward, and γ is the discount factor which is used to calculate cumulative returns. The goal of reinforcement learning algorithm is to learn an optimal policy given an MDP, where the policy refers to the mapping from state to action: $\pi(s|a) = P[A_t = a|S_t = s]$. Whereas, the state of the next time slot cannot be obtained in the original optimization problem, which means that the state transition probability of the MDP framework is unknown in advance. Therefore, this paper applies the model-free RL to address the above issue. Specifically, through the continuous interaction with the environment, the agent evaluates the actions according to the feedback of the environment (reward), and aims to continuously improve the policy, until the optimal solution of actions in each state is found. The definitions of the state, action and reward in the RL-based framework are given below.

State: At time slot t_i^j, the state of the agent is defined as $\boldsymbol{S}\left(t_i^j\right) = \{\boldsymbol{S}_0(t_i^j), \boldsymbol{S}_1(t_i^j), \ldots, \boldsymbol{S}_N(t_i^j)\}$. The state includes channel state, virtual task queue and current computing resources. Thereinto, the state of the SBS is defined as $\boldsymbol{S}_n = \left\{\bar{\boldsymbol{H}}_n^d, \boldsymbol{Q}_n, c_n\right\}, \forall n \in \{1, \ldots, N\}$, where $\bar{\boldsymbol{H}}_n^d = \{\bar{\boldsymbol{h}}_{1,n}^d, .., \bar{\boldsymbol{h}}_{K,n}^d\}$. $c_n\left(t_i^j\right)$ indicates the number of CPU cores available to the server at time slot t_i^j, $c_n\left(t_i^j\right) = c_n\left(t_i^{j-1}\right) - \sum_{i=1}^{k} z_{i,n}\left(t_i^{j-1}\right)$ and $c_n\left(t_i^0\right) = C_n$. The state of the MBS is defined as $\boldsymbol{S}_0 = \left\{\bar{\boldsymbol{H}}_0^d, \bar{\boldsymbol{G}}, \bar{\boldsymbol{H}}_m^r, \boldsymbol{Q}_0, c_0\right\}$, where $\overline{\boldsymbol{H}}_m^r = \{\bar{\boldsymbol{h}}_{1,m}^r, .., \bar{\boldsymbol{h}}_{K,m}^r\}$.

Action: At time slot t_i^j, the action of the agent is defined as $\boldsymbol{A}\left(t_i^j\right) = \{\boldsymbol{Y}\left(t_i^j\right), \boldsymbol{P}\left(t_i^j\right), \boldsymbol{\phi}\left(t_i^j\right), \boldsymbol{Z}\left(t_i^j\right)\}$, which denote the sub-band allocation, power allocation, IRS phase shift design and computation offloading actions, respectively.

Reward: The optimization goal of this work is to minimize the system energy consumption under the latency constraint, thus the reward is defined as the weighted sum of negative energy consumption and latency penalty. At time slot t_i^j, the reward is defined as follows,

$$\begin{aligned} R\left(t_i^j\right) = &-\beta_1 \left[\sum_{n=0}^{N} E_n^{exe}\left(t_i^j\right) \middle| \boldsymbol{X}^*(t_i)\right] \\ &+ \mathbf{1}\{j = J\}\beta_2 \sum_{k=1}^{K}\sum_{n=0}^{N} \min \\ &\times \left\{\left(T_k^{max} - T_{k,n}^{exe}(t_i)\right), 0\right\}. \end{aligned} \tag{13}$$

The two terms of the reward have different units. Accordingly, a weighting factor β is added to the reward for normalization, where $\beta_1 + \beta_2 = 1$ and $\beta_i \geq 0, \forall i \in \{1, 2\}$. $\mathbf{1}\{j = J\}$ denotes an indicator function whose value is 1 when $j = J$, otherwise 0.

It is worth noting that the action space of this work is discrete. Therefore, the policy can be optimized according to the action-value function $Q^{\pi}(s, a)$ (Q function). Traditional RL algorithms such as Q-learning [52] store the value function in the Q-table. However, the state space of our work is high-dimension, it is difficult to store and calculate value functions in the table since the computation time and complexity of the RL algorithms can increase exponentially, making it hard to converge. To address the above issue, deep neural networks are utilized for approximating the estimated value functions. The neural networks are trained using the training sample obtained by the interaction between the agent and the environment to approximate the value function, which improves the estimation accuracy, thereby accelerating the convergence speed of the RL. In this paper, the Deep Q-Network (DQN) algorithm [53] that introduces DNN into Q-learning is leveraged to learn the computation offloading, resource allocation and IRS phase shift design strategy.

Reinforcement learning is considered unstable or even difficult to converge when the value functions are approximated using nonlinear functions such as DNNs. The reasons are as follows. Firstly, there exist correlations between the data

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

collected through a series of observations, and a tiny change of Q function will significantly change the policy and thus change the data distribution. Secondly, there exist correlations between the Q function and the target value. The DQN algorithm employs the experience replay and independent target network techniques to deal with the algorithm instability. Specifically, the experience replay technique randomly samples data to break the correlation between data and smooth the change of data distribution. Independent target network indicates that the target value and the Q function are represented by different parameters, and the parameter update frequency is set to be different to reduce the correlation between the two networks.

Algorithm 2 The DQN Based Computation Offloading, Resource Allocation and IRS Phase Shift Design Algorithm

Initialization: replay buffer D with capacity N, Q network Q with random weight θ, target network $\hat{Q}$ with weight $\theta^- = \theta$.

1. **For** Episode $= 1, \cdots, M$ **do**
2. Initialize state s_1.
3. **For** each step $t = 1, \cdots, T$ **do**
4. According to the ε-Greedy strategy, the action a_t is randomly choosed with the probability of ε, and is choosed based on $a_t = \text{argmax}_a Q(s_t, a; \theta)$ with the probability of $(1-\varepsilon)$.
5. execute action a_t, transit to state s_{t+1} and receive reward r_t.
6. store (s_t, a_t, r_t, s_{t+1}) in replay buffer $\mathcal{D}$.
7. sample a minibatch of samples (s_j, a_j, r_j, s_{j+1}) from $\mathcal{D}$ randomly.
8. set the target of TD-error:
$$Y_j = \begin{cases} r_j, & \text{if Episode} = J+1, \\ r_j + \gamma \max_{a'} Q\left(s_{j+1}, a'; \theta_i^-\right), & otherwise. \end{cases}$$
9. Perform gradient descent for $\left(Y_j - Q(s_j, a; \theta_i)\right)^2$, update network parameter θ.
10. Update target network parameter every C steps $\theta^- = \theta$.
11. **End For**
12. **End For**

For the DQN algorithm, DNN is applied to approximate the action-value function $Q(s, a; \theta_i)$ (Q network), where θ_i is the parameter of Q network at the ith iteration. The input of the DNN is the state, followed by two fully connected layers, and the output is the action-value function corresponding to all actions in the input state. The data set that stores the experience of the agent is required as the replay buffer $D_t = \{e_1 \cdots, e_t\}$, where the experience of each step $e_t = (s_t, a_t, r_t, s_{t+1})$ consists of the current state, action, reward and next state. DQN applies the update method of Q-learning, randomly and uniformly samples a minibatch data in the replay buffer, and adopts the following loss function to update the neural network parameters:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s') \sim U(D)} \times \left[\left(r + \gamma \max_{a'} \hat{Q}\left(s', a'; \theta_i^-\right) - Q(s, a; \theta_i) \right)^2 \right]. \tag{14}$$

Specifically, the Q network $Q(s, a; \theta_i)$ and the target network $\hat{Q}\left(s', a'; \theta_i^-\right)$ need to be updated, where θ_i^- denotes the target network parameter at the ith iteration, and the update time of parameters θ_i and θ_i^- are different. The target network parameter θ_i^- is updated every C steps with the Q network parameter θ_i and is fixed at other times, while θ_i is updated every step. Training the Q network is the process of updating the parameter θ with the goal of minimizing the loss function. $Y_i = r + \gamma \max_{a'} Q\left(s', a'; \theta_i^-\right)$ denotes the target of temporal difference (TD). Therefore, utilizing the target Y_i with another parameter set of delayed update adds a delay between the time when the Q network is updated and the time when the TD target Y_i is updated, which solves the problem of oscillation or non-convergence in RL. Gradient descent method is adopted to update parameters with the aim of minimizing the loss function $L(\theta_i)$:

$$\nabla_{\theta_i} L(\theta_i) = \mathbb{E}_{(s,a,r,s')} \left[\left(Y_i - Q(s, a; \theta_i) \right) \nabla_{\theta_i} Q(s, a; \theta_i) \right], \tag{15}$$

$$\theta_{i+1} = \theta_i + \alpha \nabla_{\theta_i} L(\theta_i), \tag{16}$$

where α is the learning rate.

Algorithm 2 demonstrates the DQN based computation offloading, resource allocation and IRS phase shift design algorithm. First, the replay buffer D and the two networks are initialized. Episode represents the process from the initial state of agent to the final state. The following steps are performed for each episode. The initial state s_1 of each episode is initialized. For each step t of episode m, action a_t is chosen according to the ε-Greedy strategy, then the agent performs action a_t, observes reward r_t and transits to the next state s_{t+1}. The experience (s_t, a_t, r_t, s_{t+1}) is stored in the replay buffer D, and a minibatch of data is uniformly and randomly sampled from D in order to update the Q network parameter θ according to (16) using the gradient descent method. The target network updates parameter θ^- every C steps. The complexity of learning procedure for the DQN based algorithm is denoted as $\mathcal{O}\left(T\left(\sum_{l=0}^{L_{\text{DQN}}} n_{\text{DQN}}^{(l)} n_{\text{DQN}}^{(l+1)}\right)\right)$, where L_{DQN} is the number of hidden layers in the DNN, and n_{DQN} is the number of neurons in each layer. We adopt a fully-connected DNN that has two hidden layers of 100 neurons and adopt the activation function of ReLU.

The proposed two-timescale mechanism for resource management and reflection optimization is shown in Algorithm 3. Specifically, in each epoch i, the matching theory algorithm is performed. And then at the current epoch, the DRL algorithm is performed for each timescale slot j.

Algorithm 3 The Proposed Two-Timescale Mechanism for Resource Management and Reflection Optimization

Initialization: The number of epochs I and the number of time slots J.

1. **For** $i = 0, \cdots, I$ **do**
2. At epoch t_i, the user association scheme is obtained based on Algorithm 1.
3. **For** $j = 1, \cdots, J$ **do**
4. At time slot t_i^j, the computation offloading, resource allocation and IRS phase shift design scheme is learned based on Algorithm 2.
5. **End For**
6. **End For**

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Delayed taps $L^d_{k,n}$, L_0 and L_k	4, 2, 3
β of MBS-IRS, user-IRS, user-BS channel	3.5, 2.2, 2.2
Gap Γ	8.8dB
Noise variance σ^2	-110dBm
Number of sub-bands B	32
Sub-band bandwidth W	180KHz
Maximum transmit power P_k^{max}	20dBm
Number of discrete power levels	10
Input data size of the subtask $d_k^{u,v}$	U[100, 500]KB
Number of clock cycles χ_k^v	U[4000, 12000]cycles/byte
Maximum tolerable latency T_k^{max}	U[1.5, 3]s
Computation capability of UE f_k^l	10^9cycles/s
Computation capability of MEC server f_n^c	5×10^{10}, 10^{10}cycles/s
Effective capacitance coefficients ζ_{mob}, ζ_e	10^{-27}, 10^{-29}
Learning rate α	0.01
Discount factor γ	0.9
Capacity of replay buffer and minibatch	600, 128

V. NUMERICAL SIMULATIONS

The performance of the proposed two-timescale mechanism for the resource management and reflection optimization is simulated by Python 3.6 and TensorFlow 1.12.0. The numbers of MBS and SBSs are set to 1 and 2, and the number of users is considered as 10. In addition, the number of CPU cores for the MEC server connected to the MBS is set to 16, and the number of CPU cores for the MEC server connected to the SBS is set to 8. The distance between the IRS and the MBS is set to 200m. Users are located in a semicircular area within 50m around the IRS, and the distance between users and the SBS is within 50m. For each multipath channel, $\alpha = 0.5$. The reference path loss ϱ_0 at reference distance $d_0 = 1$m is -30dB. The computation task is considered as augmented reality application, which consists of 3 separable computation-intensive subtasks [47]. The specific simulation parameters of this paper are given in Table I.

Two benchmark algorithms are adopted for comparison with the proposed algorithm. 1) random phase shift scheme: the user association, computation offloading and resource allocation strategy is optimized according to Algorithm 3. The IRS phase is randomly selected which obeys the uniform distribution with the range of $[0, 2\pi]$ instead of optimizing based on the proposed algorithm. 2) without IRS scheme: the user association, computation offloading and resource allocation strategy is optimized in the edge heterogeneous network without IRS assistance.

Fig. 3 illustrates the average convergence performance of the proposed algorithm and Q-learning based benchmark algorithm for each epoch. The reward gradually increases as the training continues. It is worth noting that due to the large state space and action space, the environment is complex in the proposed IRS-assisted edge heterogeneous network, about 1,250 training episodes are required for the proposed algorithm to converge properly. While the Q-learning based algorithm requires about 1800 iterations to converge. The proposed DQN based scheme learns much faster than the Q-leaning based scheme. Since the DRL based scheme can improve the efficiency and accuracy for estimating the Q value by means

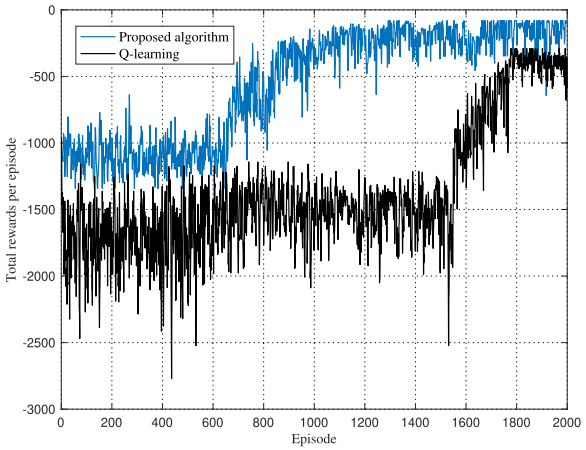

Fig. 3. The convergence performance of the proposed algorithm.

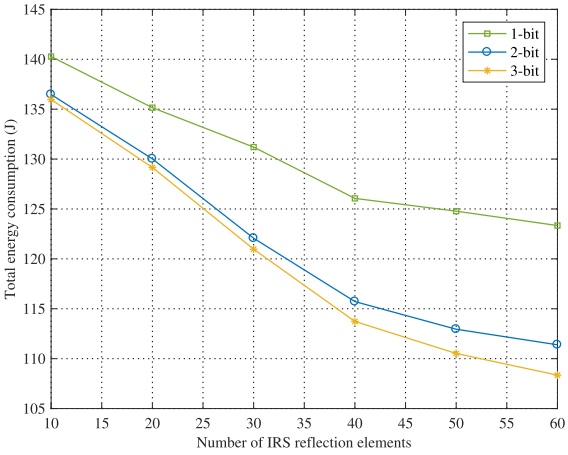

Fig. 4. The influence of different IRS discrete phase shift levels on total energy consumption.

of DNNs, thereby allowing the agent to obtain the optimal strategy faster. During the training process, the rewards of subsequent episodes will fluctuate due to the utilization of ε-greedy strategy for exploitation and exploration. Thus, the result verifies the convergence of the proposed method.

Fig. 4 demonstrates the influence of the discrete phase shift level number on energy consumption. The 1-bit phase shift stands for $\rho = 2$; the 2-bit phase shift represents $\rho = 4$; the 3-bit phase shift denotes $\rho = 8$. When increasing the phase shift level from 1-bit to 2-bit, the reduction in energy consumption varies from 2.7% to 9.69%. However, when the phase shift level increases from 2-bit to 3-bit, the energy consumption decreases from 0.37% to 2.7%. The results demonstrate that the adoption of 1-bit phase shift has a greater impact on energy consumption compared to the adoption of 2-bit and 3-bit, while the performance loss of adopting 2-bit and 3-bit is acceptable. Considering the difficulty of designing IRS, 2-bit phase shift is generally applied in practical system [54]. In addition, the increasing of IRS reflection elements can compensate the loss generated through the low-precision discrete phase shift. Therefore, 2-bit reflection array is adopted to obtain good performance in this work.

Fig. 5 depicts the total energy consumption obtained based on the proposed algorithm and benchmark algorithms under a

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

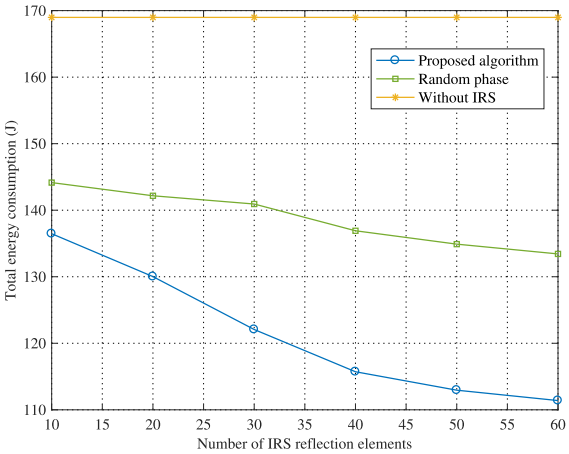

Fig. 5. The total energy consumption versus the number of IRS reflection elements.

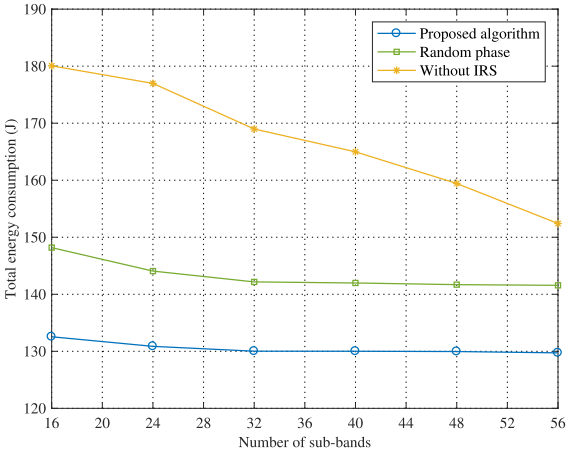

Fig. 6. The total energy consumption versus the number of sub-bands.

set of various IRS reflection elements number. The gap in total energy consumption between the without the IRS scheme and the random phase shift scheme grows as the reflection element increases, which indicates that even without careful design of the IRS reflection coefficient, the energy consumption can be reduced through the IRS assistance. Furthermore, the proposed algorithm outperforms the random phase shift scheme, which reflects that the passive beamforming gain is provided to reduce the communication burden through the refined IRS reflection coefficient design.

Fig. 6 shows the trend of total energy consumption as the number of sub-bands changes. The energy consumption reduces as the sub-bands number augments. This can be explained as the sum of channel gains of each user equipment increases with the escalation of sub-band. Furthermore, when M increases, the sub-bands typically have greater channel gain as it is assumed that the sub-bands are independent. For the proposed algorithm, the total energy consumption is down slightly when the number of sub-bands increases from 16 to 32. Moreover, the figure depicts the insignificant reduction in energy consumption when the number of sub-bands is larger than 32. This can explained as the energy consumption is mainly generated through the process of computation offloading when the communication resources are

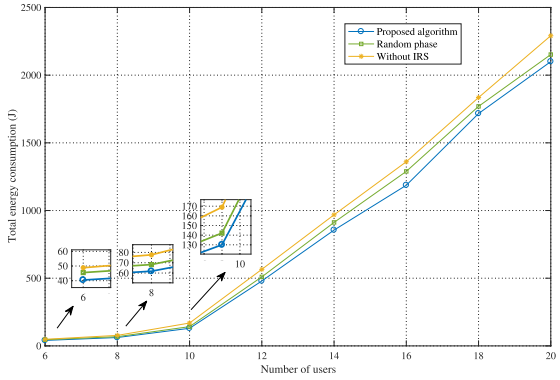

Fig. 7. The total energy consumption versus the number of users.

not enough, while the energy consumption is mainly generated by the process of task computation when the communication resources are sufficient.

Fig. 7 illustrates the relationship between the total energy consumption and the number of users. As the number of users increases, the energy consumption of the proposed and benchmark schemes augments rapidly. Compared with the benchmark schemes, the energy consumption generated by the proposed algorithm is the lowest since the user association, computation offloading, resource allocation and IRS phase shift design are jointly optimized by the proposed algorithm. To be specific, compared with the proposed algorithm, the energy consumption obtained by random phase shift scheme is severely affected by the inability to optimize the phase shift. Besides, the energy consumption generated by the without IRS scheme is the highest in contrast with the proposed and random phase shift schemes. Because channel condition cannot be improved with IRS assistance, which results in the energy consumption of the communication link increases. Whereas, the proposed algorithm and random phase shift scheme utilize the assistance of IRS to enhance the wireless link capacity and reduce energy consumption.

VI. Conclusion

This paper proposes resource management and reflection optimization scheme for IRS assisted edge heterogeneous network. Specifically, a scenario composed of the MBS and SBSs which are equipped with MEC servers is considered, in which the IRS assists the users in offloading computation tasks to the MBS. The optimization objective of our work is to minimize the long-term total energy consumption while guaranteeing the quality of service for the users. The optimization problem is formulated as two-timescale mechanism since the update timescale of user association is larger than the timescale of computation offloading, resource allocation and IRS phase shift design. For the long timescale, the matching theory based user association algorithm is proposed. For the short timescale, we put forward the DQN-based computation offloading, resource allocation, and IRS phase shift design algorithm. The simulation results validate the convergence performance of the two-timescale mechanism and illustrate that limited IRS discrete phase shift levels can achieve good performance. Furthermore, by quantifying the

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

energy consumption of the IRS-assisted edge heterogeneous network in different simulation environments, the proposed algorithm demonstrates phase-shift design can provide the passive beamforming gain in comparison with the benchmark schemes, which enables the edge network to reduce energy consumption.

REFERENCES

[1] Q. Luo, S. Hu, C. Li, G. Li, and W. Shi, "Resource scheduling in edge computing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2131–2165, 4th Quart., 2021.

[2] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing-A key technology towards 5G," *ETSI White Paper*, vol. 11, no. 11, pp. 1–16, Sep. 2015.

[3] Y. Zhang, H. Liu, L. Jiao, and X. Fu, "To offload or not to offload: An efficient code partition algorithm for mobile cloud computing," in *Proc. IEEE 1st Int. Conf. Cloud Netw. (CLOUDNET)*, Paris, France, Nov. 2012, pp. 80–86.

[4] A. Ndikumana et al., "Joint communication, computation, caching, and control in big data multi-access edge computing," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1359–1374, Jun. 2020.

[5] M. Qin et al., "Service-oriented energy-latency tradeoff for IoT task partial offloading in MEC-enhanced multi-RAT networks," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1896–1907, Feb. 2021.

[6] Y. Ding, C. Liu, X. Zhou, Z. Liu, and Z. Tang, "A code-oriented partitioning computation offloading strategy for multiple users and multiple mobile edge computing servers," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4800–4810, Jul. 2020.

[7] Z. Kuang, L. Li, J. Gao, L. Zhao, and A. Liu, "Partial offloading scheduling and power allocation for mobile edge computing systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6774–6785, Aug. 2019.

[8] Y. Wang, X. Tao, Y. T. Hou, and P. Zhang, "Effective capacity-based resource allocation in mobile edge computing with two-stage tandem queues," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6221–6233, Sep. 2019.

[9] Y. Dai, Y. L. Guan, K. K. Leung, and Y. Zhang, "Reconfigurable intelligent surface for low-latency edge computing in 6G," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 72–79, Dec. 2021.

[10] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[11] S. Gong et al., "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2283–2314, 4th Quart., 2020.

[12] M. A. El Mossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[13] W. Tang et al., "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.

[14] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021.

[15] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

[16] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems via joint transmit and reflective beamforming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[17] M. Hua, Q. Wu, D. W. K. Ng, J. Zhao, and L. Yang, "Intelligent reflecting surface-aided joint processing coordinated multipoint transmission," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1650–1665, Mar. 2021.

[18] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

[19] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[20] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[21] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, Aug. 2020.

[22] Y. Liu et al., "Intelligent reflecting surface meets mobile edge computing: Enhancing wireless communications for computation offloading," Jan. 2020, *arXiv:2001.07449*.

[23] F. Zhou, C. You, and R. Zhang, "Delay-optimal scheduling for IRS-aided mobile edge computing," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 740–744, Apr. 2021.

[24] Y. Cao, T. Lv, Z. Lin, and W. Ni, "Delay-constrained joint power control, user detection and passive beamforming in intelligent reflecting surface-assisted uplink mmWave system," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 482–495, Jun. 2021.

[25] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.

[26] S. Hua and Y. Shi, "Reconfigurable intelligent surface for green edge inference in machine learning," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[27] T. Bai, C. Pan, H. Ren, Y. Deng, M. Elkashlan, and A. Nallanathan, "Resource allocation for intelligent reflecting surface aided wireless powered mobile edge computing in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5389–5407, Aug. 2021.

[28] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for Internet of Things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619–623, Mar. 2021.

[29] X. Hu, C. Masouros, and K.-K. Wong, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3709–3725, Jun. 2021.

[30] S. Huang, S. Wang, R. Wang, M. Wen, and K. Huang, "Reconfigurable intelligent surface assisted mobile edge computing with heterogeneous learning tasks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 2, pp. 369–382, Jun. 2021.

[31] Y. Chen, M. Wen, E. Basar, Y.-C. Wu, L. Wang, and W. Liu, "Exploiting reconfigurable intelligent surfaces in edge caching: Joint hybrid beamforming and content placement optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7799–7812, Dec. 2021.

[32] Z. Li et al., "Energy efficient reconfigurable intelligent surface enabled mobile edge computing networks with NOMA," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 427–440, Jun. 2021.

[33] B. Li, W. Wu, Y. Li, and W. Zhao, "Intelligent reflecting surface and artificial-noise-assisted secure transmission of MEC system," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 11477–11488, Jul. 2022, doi: 10.1109/JIOT.2021.3127534.

[34] Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 680–692, Jan. 2018.

[35] Y. Wei, F. R. Yu, M. Song, and Z. Han, "Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor-critic deep reinforcement learning," *IEEE Internet Things J.* vol. 6, no. 2, pp. 2061–2073, Apr. 2019.

[36] S. Shrivastava, B. Chen, C. Chen, H. Wang, and M. Dai, "Deep *Q*-network learning based downlink resource allocation for hybrid RF/VLC systems," *IEEE Access*, vol. 8, pp. 149412–149434, 2020.

[37] X. Zhang, Y. Shen, B. Yang, W. Zang, and S. Wang, "DRL based data offloading for intelligent reflecting surface aided mobile edge computing," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–7.

[38] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4692–4707, Oct. 2019.

[39] S. Zhang et al., "Energy-efficient massive MIMO with decentralized precoder design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15370–15384, Dec. 2020.

[40] C. Pradhan, A. Li, L. Song, J. Li, B. Vucetic, and Y. Li, "Reconfigurable intelligent surface (RIS)-enhanced two-way OFDM communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16270–16275, Dec. 2020.

[41] O. Muñoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4738–4755, Oct. 2015.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.

[42] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[43] S. Yu, R. Langar, X. Fu, L. Wang, and Z. Han, "Computation offloading with data caching enhancement for mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11098–11112, Nov. 2018.
[44] S. Yu, X. Chen, Z. Zhou, X. Gong, and D. Wu, "When deep reinforcement learning meets federated learning: Intelligent multitimescale resource management for multiaccess edge computing in 5G ultradense network," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2238–2251, Feb. 2021.
[45] G. Guo and J. Zhang, "Energy-efficient incremental offloading of neural network computations in mobile edge computing," in *Proc. IEEE Global Commun. Conf.*, Taipei, Dec. 2020, pp. 1–6.
[46] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.
[47] Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, "Joint computation offloading and user association in multi-task mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12313–12325, Dec. 2018.
[48] Y. Deng, Z. Chen, X. Chen, and Y. Fang, "Throughput maximization for multiedge multiuser edge computing systems," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 68–79, Jan. 2022.
[49] K. Bando, "Many-to-one matching markets with externalities among firms," *J. Math. Econ.*, vol. 48, no. 1, pp. 14–20, Jan. 2012.
[50] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, and M. Elkashlan, "Many-to-many matching with externalities for device-to-device communications," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 138–141, Feb. 2017.
[51] E. Baron, C. Lee, A. Chong, B. Hassibi, and A. Wierman, "Peer effects and stability in matching markets," in *Proc. Int. Symp. Algorithmic Game Theory (SAGT)*, Amalfi, Italy, Oct. 2011, pp. 117–129.
[52] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, May 1992.
[53] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[54] Q. Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

Zhaoying Wang received the B.Eng. degree from the School of Computer Science and Information Engineering, Hefei University of Technology, Anhui, China, in 2017, and the Ph.D. degree from the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2022. Her research interests include resource allocation, mobile edge computing, and deep reinforcement learning.

Yifei Wei (Member, IEEE) received the B.Sc. and Ph.D. degrees in electronic engineering from the Beijing University of Posts and Telecommunications (BUPT), China, in 2004 and 2009, respectively. He was a Visiting Ph.D. Student with Carleton University, Canada, from 2008 to 2009. He was a Post-Doctoral Research Fellow with Dublin City University, Ireland, in 2013. He was the Vice Dean of the School of Science, BUPT, from 2014 to 2016. He was a Visiting Scholar with the University of Houston, USA, from 2016 to 2017. He is currently a Professor and the Vice Dean of the School of Electronic Engineering, BUPT. His current research interests include energy-efficient networking, heterogeneous resource management, machine learning, and edge computing. He has served on the technical program committee members of numerous conferences. He received the Best Paper Awards at the ICCTA 2011 and ICCCS 2018. He also served as a Symposium Co-Chair for IEEE GLOBECOM 2020, and a Track Co-Chair for International Conference on Artificial Intelligence and Security (ICAIS) 2019, 2020, 2021, and 2022. He is the Guest Editor of a Special Issue in the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING in 2021.

Zhiyong Feng (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. She is currently a Professor at the BUPT, and the Director of the Key Laboratory of the Universal Wireless Communications, Ministry of Education, China. Her main research interests include wireless network architecture design and radio resource management in 5th generation mobile networks (5G), spectrum sensing and dynamic spectrum management in cognitive wireless networks, and universal signal detection and identification. She is the Vice Chair of the Information and Communication Test Committee of the Chinese Institute of Communications (CIC). She is serving as an Associate Editor-in-Chief for *China Communications*. She is a Technological Advisor for International Forum on NGMN.

F. Richard Yu (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of British Columbia (UBC) in 2003. His research interests include connected/autonomous vehicles, artificial intelligence, cybersecurity, and wireless systems. He has been named in the Clarivate Analytics list of "Highly Cited Researchers" since 2019. He is a fellow of the Canadian Academy of Engineering (CAE), Engineering Institute of Canada (EIC), and IET. He received several best paper awards from some first-tier conferences. He is an Elected Member of the Board of Governors of the IEEE VTS and the Editor-in-Chief of the IEEE VTS Mobile World Newsletter. He is a Distinguished Lecturer of IEEE in both VTS and ComSoc.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer of JDSU, Germantown, MD. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an Assistant Professor at Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as the Computer Science Department, University of Houston, Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS) in 2016, and several best paper awards in IEEE conferences. He is also the Winner of the 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks." He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018, AAAS Fellow since 2019, and ACM Distinguished Member since 2019. He is a 1% Highly Cited Researcher since 2017 according to Web of Science.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 14:18:09 UTC from IEEE Xplore. Restrictions apply.