Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Dynamic user-centric multi-dimensional resource allocation for a wide-area coverage signaling cell based on DQN*

Zhou TONG†‡, Na LI†, Huimin ZHANG†, Quan ZHAO, Yun ZHAO, Junshuai SUN, Guangyi LIU

Future Research Lab, China Mobile Research Institute, Beijing 100053, China

†E-mail: tongzhou@chinamobile.com; linawx@chinamobile.com; zhanghuiminyjy@chinamobile.com

Received May 20, 2022; Revision accepted Aug. 30, 2022; Crosschecked Sept. 22, 2022

Abstract: The rapid development of communications industry has spawned more new services and applications. The sixth-generation wireless communication system (6G) network is faced with more stringent and diverse requirements. While ensuring performance requirements, such as high data rate and low latency, the problem of high energy consumption in the fifth-generation wireless communication system (5G) network has also become one of the problems to be solved in 6G. The wide-area coverage signaling cell technology conforms to the future development trend of radio access networks, and has the advantages of reducing network energy consumption and improving resource utilization. In wide-area coverage signaling cells, on-demand multi-dimensional resource allocation is an important technical means to ensure the ultimate performance requirements of users, and its effect will affect the efficiency of network resource utilization. This paper constructs a user-centric dynamic allocation model of wireless resources, and proposes a deep Q-network based dynamic resource allocation algorithm. The algorithm can realize dynamic and flexible admission control and multi-dimensional resource allocation in wide-area coverage signaling cells according to the data rate and latency demands of users. According to the simulation results, the proposed algorithm can effectively improve the average user experience on a long time scale, and ensure network users a high data rate and low energy consumption.

Key words: 6G; Wide-area coverage signaling cell; Multi-dimensional resource allocation; Deep Q-network (DQN)
https://doi.org/10.1631/FITEE.2200220 **CLC number:** TN929.5

1 Introduction

With the global commercialization of the fifth-generation wireless communication system (5G) network, mobile communication has risen to a new level, from the realization of "connection of people" to the establishment of "connection of things" between terminals in thousands of industries. Driven by the 5G network, the requirements of users are more differentiated, and the data rate and latency performance required by various new services and new applications are more extreme. Affected by the coverage of mainstream 5G network frequency bands (such as 3.5 GHz), to meet the extreme performance requirements of users, the deployment density of base stations (BSs) has to be greatly increased, which increases the 5G network construction cost and energy consumption.

The high energy consumption of the 5G network has also become a key issue of the sixth-generation wireless communication system (6G) network. To reduce the network power consumption caused by the dense deployment of high-frequency BSs and ensure the performance of network wide-area coverage, Liu

‡ Corresponding author
* Project supported by the National Key Research and Development Program of China (No. 2020YFB1806800)
ORCID: Zhou TONG, https://orcid.org/0000-0002-9469-9523

et al. (2022b) proposed a wide-area coverage signaling cell technical scheme. As shown in Fig. 1, in this scheme, the low-frequency (such as 700 MHz) control BSs/cells provide unified signaling coverage for a large geographical area, and are responsible for the transmission of radio resource control (RRC) messages and physical layer control signaling, thereby reducing the impact of high path loss caused by high-frequency bands and ensuring continuous and reliable connectivity and mobility. High-frequency (such as 62.5 GHz and above) data BSs/cells provide data transmission and a small amount of necessary signaling. These high-frequency data BSs have the characteristics of high capacity and on-demand activation, to reduce the interference between data cells and energy consumption of the entire network.

Resource allocation is also a key problem to be solved in wide-area coverage signaling cells, because resource allocation is related to both user experience and network efficiency. The application of artificial intelligence (AI) in 5G networks promotes the development of the mobile communication network and its application in vertical industries (Liu et al., 2022a). With the improvement of network automation and intelligence, AI has become one of the effective means of solving the problem of resource allocation in dynamic radio environments (Lin and Zhao, 2020). Ji et al. (2021) proposed an online bandwidth resource allocation algorithm based on deep reinforcement learning (DRL) to solve the resource allocation problem caused by operators by sharing network resources, which effectively improves the bandwidth resource utilization. Gang and Friderikos (2019) studied the bandwidth allocation and power allocation problems in 5G virtual network slicing and proposed an optimization framework for flexible inter-tenant resource sharing based on transmission power control. Luo et al. (2014) took the maximization of the average signal to interference plus noise ratio (SINR) as the goal of resource allocation, and used Q-learning to finish the channel assignment and power allocation at the same time. To overcome the excessive energy consumption problem in indoor wireless networks, Lü et al. (2021) proposed a deep Q-network (DQN) based transmission power allocation algorithm for home BSs. Ren et al. (2021) proposed a DRL-based approach to minimize long-term system energy consumption in a computation offloading scenario with multiple Industrial Internet of Things (IIoT) devices and multiple fog access points. In Zhao et al. (2015), a method based on the combination of K-means clustering and Q-learning was proposed to jointly optimize the spectrum allocation, load balancing, and energy saving in mobile broadband networks. The above research works were designed based on a traditional network architecture.

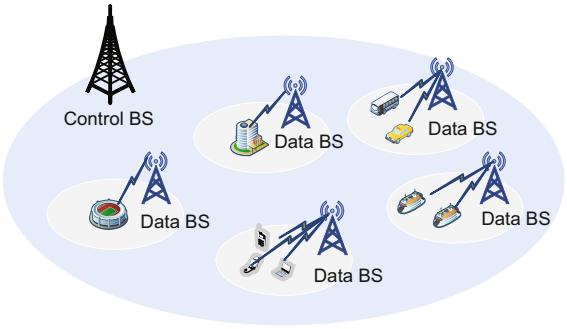

Fig. 1 Wide-area coverage signaling cell (BS: base station)

Different from traditional cells that are responsible for transmission of both signaling and data, the wide-area coverage signaling cell will primarily be in charge of the transmission of signaling messages as well as management of all data cell resources. For future wide-area signaling coverage scenarios, in this paper, the network side uses intelligent capabilities to summarize user characteristics, and uses AI tools to realize on-demand and dynamic resource allocation according to the differentiated requirements of users, which can improve the overall resource utilization of the network and greatly improve the user experience. In this paper, the user experience considered is the difference between the data rate revenue and the total delay loss.

The main contributions of this paper are summarized as follows:

1. Aiming at solving the problem of multi-dimensional resource allocation in wide-area coverage signaling cells, a user-centric dynamic allocation model is constructed for multi-dimensional wireless resources, in which more differentiated requirements of users in the future, such as rate and latency, are considered, and the actual limitations of network power and bandwidth are considered.

2. Considering the dynamic BS changes concerning the data queue, wireless channel state, and user service requirements, a user admission control

scheme is formulated to enable the on-demand on/off of data BSs.

3. A DQN-based dynamic allocation algorithm for wireless resources is proposed to realize user admission control and the dynamic and flexible allocation of physical resource blocks (PRBs) and power. According to the simulation results, the proposed algorithm can improve the average user experience on a long time scale, ensure a high data rate for users and low energy consumption of the network, and achieve real-time optimization of the overall network utility.

2 System model and problem formulation

2.1 System model

In this paper, we consider the wide-area coverage signaling cell scenario. The dynamic user-centric allocation model of multi-dimensional wireless resources is shown in Fig. 2. In this model, we assume that the network perceives each user that it serves, and that users regularly report their requirements to the network. Users in different industries have different quality of service (QoS) requirements, including the rate and latency. The network performs big data calculation on users through the data collection module, summarizes user characteristics, and customizes flexible and dynamic wireless resource allocation strategies according to user requirements. The resource allocation involved in the process of the BS providing services to users includes user admission control, PRB allocation, and power allocation.

In this model, we assume that there is a control BS and multiple data BSs in a specific area, $\mathcal{J}=\{1,2,\cdots,J\}$. The total bandwidth of W Hz is divided into multiple PRBs, $\mathcal{B}=\{1,2,\cdots,B\}$, which are shared by all BSs. Suppose that there are N users in the area, $\mathcal{N}=\{1,2,\cdots,N\}$. Due to the limitation of orthogonal frequency division multiple access (OFDMA), a user can access only one BS. Let $a_{j,n}(t)$ and $\phi_{j,n}^{b}(t)$ represent the binary user admission control factors, i.e., the user admission control of BS j and the allocation strategy of PRB b in time slot t, respectively. When user n accesses BS j in time slot t, $a_{j,n}(t)=1$; otherwise, $a_{j,n}(t)=0$. When BS j allocates PRB b to user n in time slot t, $\phi_{j,n}^{b}(t)=1$; otherwise, $\phi_{j,n}^{b}(t)=0$. $\phi_{j,n}^{b}(t)$ satisfies

$$\phi_{j,n}^{b}(t)\geq 0,\ \sum_{j\in\mathcal{J}}\phi_{j,n}^{b}(t)\leq B. \tag{1}$$

The channel state in each time slot is assumed to be fixed when a user requests access to each BS. The channel states among different time slots change randomly, and are independent of each other. The transmission rate provided by BS j to user n on PRB b in time slot t can be expressed as

$$r_{j,n}^{b}(t)=w_{j,n}^{b}\cdot\log_2\left(1+\frac{p_{j,n}^{b}(t)h_{j,n}^{b}(t)}{\sum\limits_{\forall j'\in\mathcal{J},j'\neq j}\sum\limits_{\forall n'\in\mathcal{N},n'\neq n}p_{j',n'}^{b}(t)h_{j',n'}^{b}(t)+\sigma^2}\right), \tag{2}$$

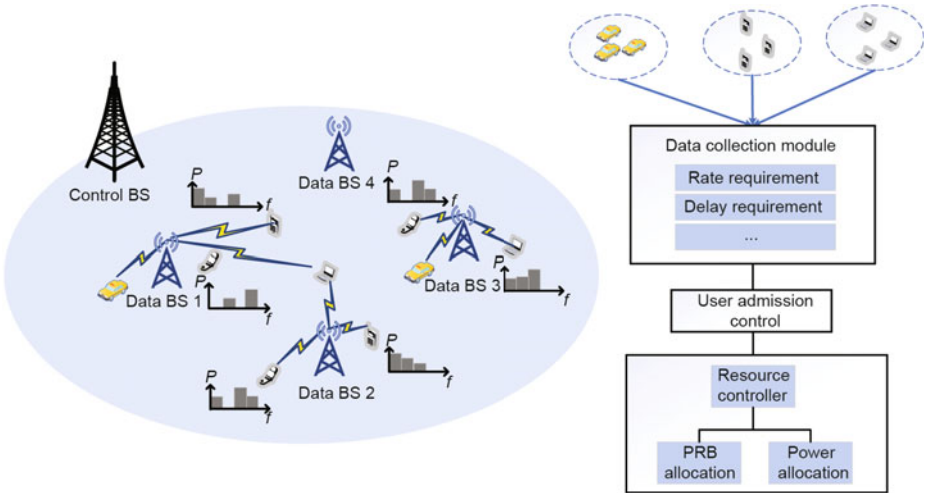

Fig. 2 Dynamic user-centric resource allocation model in a wide-area coverage signaling cell (BS: base station; PRB: physical resource block)

where $w_{j,n}^{b}$ is the bandwidth allocated by BS j to user n on PRB b, and σ^2 is the noise power. The noise power is the same on all PRBs of all BSs for all users. $p_{j,n}^{b}(t)$ represents the power allocated by BS j to user n on PRB b in time slot t. Let $\mathcal{H}$ be a finite set of channel states. When user n accesses BS j in time slot t, $h_{j,n}(t)$ is the channel gain, where $h_{j,n}(t) \in \mathcal{H} = \{h_1, h_2, \cdots, h_H\}$ (here, H is the number of different channel states in this model).

Therefore, the total transmission rate provided by BS j for all users accessing the BS in time slot t is

$$r_j(t) = \sum_{b\in\mathcal{B}} \sum_{n\in\mathcal{N}} a_{j,n}(t)\, \phi_{j,n}^{b}(t)\, r_{j,n}^{b}(t). \tag{3}$$

The total rate of all BSs in time slot t in the whole network is

$$r(t) = \sum_{j\in\mathcal{J}} r_j(t). \tag{4}$$

The long-term average total rate of the whole network is

$$\bar{r} = \lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} E\left[r(t)\right]. \tag{5}$$

Consider a discrete-time queuing system, in which the length of each time slot is fixed. Denote the number of data packets arriving at BS j accessed by user n in time slot t as $X_{j,n}(t)$. The number of arriving data packets follows the Poisson distribution with parameter $\lambda_{j,n}$ and is independent and identically distributed between different time slots. This model constructs a corresponding queue for the data packets of the services to be processed by each BS. At the beginning of time slot t, the queue length of BS j is $Q_j(t)$, $Q_j(t) = \sum_{n\in\mathcal{N}} Q_{j,n}(t)$, where $Q_{j,n}(t)$ is the queue length of user n accessing BS j.

The dynamic update process of $Q_j(t)$ is described as follows:

$$Q_j(t+1) = \max\left\{Q_j(t) - D_j(t), 0\right\} + X_j(t), \tag{6}$$

where $D_j(t) = \varepsilon_j(t)\, wA_j(t)/S$ represents the number of data packets leaving the queuing of BS j in time slot t, $\varepsilon_j(t)$ represents the spectral efficiency in time slot t, w is the bandwidth of each PRB, $A_j(t)$ is the number of PRBs allocated by BS j to users in time slot t, S is each data packet's size in the BS queue, and $X_j(t) = \sum_{n\in\mathcal{N}} X_{j,n}(t)$ is the number of data packets arriving at BS j in time slot t. Let $\mathcal{Q}(t) = \{Q_1(t), Q_2(t), \cdots, Q_J(t)\}$ represent the global queue state information of the network in time slot t. The global channel state information in time slot t can be expressed as $H(t) = \left\{\bar{h}_1(t), \bar{h}_2(t), \cdots, \bar{h}_J(t)\right\}$, where $\bar{h}_j(t)$ $(j = 1, 2, \cdots, J)$ represents the average channel gain of users accessing BS j in time slot t.

2.2 Optimization problem

The objective of this study is to maximize the overall user experience on a long time scale, that is, the difference between the data rate revenue and the total delay loss.

The total radio interface delay considered in this study includes the processing delay d_{proc}^{n} and the transmission delay d_{tran}^{n} of user n. After the BS receives the data request from the corresponding user, the time required to process the data packets is defined as the processing delay. The data processing delay of user n accessing BS j is expressed as

$$d_{\text{proc}}^{n}(t) = \frac{X_{j,n}(t)\, S_{j,n}}{R_{j,n}}, \tag{7}$$

where $R_{j,n}$ is the rate at which BS j processes the data packets of user n, and $S_{j,n}$ is the data packet size of user n accessing BS j.

Between the BS and the user, the time required to transmit data packets over the air interface is defined as the transmission delay. The data transmission delay of user n is expressed as

$$d_{\text{tran}}^{n}(t) = \frac{X_{j,n}(t)\, S_{j,n}}{\sum\limits_{b\in\mathcal{B}} \phi_{j,n}^{b}(t)\, r_{j,n}^{b}(t)}. \tag{8}$$

The total radio interface delay of user n is

$$d^{n}(t) = d_{\text{proc}}^{n}(t) + d_{\text{tran}}^{n}(t). \tag{9}$$

The total air interface delay of the whole network is

$$\begin{aligned} d(t) &= \sum_{j\in\mathcal{J}} d_j(t) \\ &= \sum_{j\in\mathcal{J}} \sum_{n\in\mathcal{N}} a_{j,n}(t) \left(d_{\text{proc}}^{n}(t) + d_{\text{tran}}^{n}(t)\right). \end{aligned} \tag{10}$$

The long-term average total air interface delay of the whole network is

$$\bar{d} = \lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} E\left[d(t)\right]. \tag{11}$$

The average network benefit and the average network cost of the system can be expressed as

$$g_{\mathrm{r}}=\overline{r}\delta_{\mathrm{r}}, \tag{12}$$

$$l_{\mathrm{d}}=\bar{d}\delta_{\mathrm{d}}, \tag{13}$$

where δ_{r} and δ_{d} refer to the unit prices of the data rate and delay, respectively.

The overall average user experience is

$$\bar{u}=g_{\mathrm{r}}-l_{\mathrm{d}}. \tag{14}$$

Therefore, the optimization problem is

$$\max_{a(t),\phi(t),p(t)}\ \bar{u}\quad \text{s.t.} \tag{15a}$$

$$\mathrm{C1}:\sum_{b\in\mathcal{B}}\phi_{j,n}^{b}\left(t\right)r_{j,n}^{b}\left(t\right)\geq r_{\min}^{n},\ \forall n\in\mathcal{N}, \tag{15b}$$

$$\mathrm{C2}:d^{n}\left(t\right)\leq d_{\max}^{n},\ \forall n\in\mathcal{N}, \tag{15c}$$

$$\mathrm{C3}:\sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}a_{j,n}(t)\phi_{j,n}^{b}(t)p_{j,n}^{b}(t)\leq p_{\max}^{j},\ \forall j\in\mathcal{J}, \tag{15d}$$

$$\mathrm{C4}:\sum_{n\in\mathcal{N}}\phi_{j,n}^{b}\left(t\right)\leq 1,\ \forall j\in\mathcal{J},\ \forall b\in\mathcal{B}, \tag{15e}$$

$$\mathrm{C5}:\sum_{j\in\mathcal{J}}a_{j,n}\left(t\right)=1,\ \forall n\in\mathcal{N}, \tag{15f}$$

$$\mathrm{C6}:\sum_{n\in\mathcal{N}}a_{j,n}\left(t\right)R_{j,n}\leq R_{j},\ \forall j\in\mathcal{J}, \tag{15g}$$

$$\mathrm{C7}:r_{j}\left(t\right)\leq W_{j},\ \forall j\in\mathcal{J}. \tag{15h}$$

C1 indicates that user admission control and resource allocation should meet the minimum data rate requirements of users. C2 indicates that user admission control and resource allocation should meet the user delay limit. C3 means that the total power allocated to users by each BS should not exceed its maximum transmission power limit $p_{\max}^{j}$. C4 means that each PRB can be assigned to only one user. C5 indicates that each user can be associated with only one BS. C6 means that the data processing rate required by each user on any BS should not exceed the total data processing rate of the BS, where R_j represents the total data processing rate of BS j. C7 represents that the total allocated bandwidth of BS j is not greater than the upper limit of the available bandwidth W_j of BS j.

3 Dynamic resource allocation algorithm based on DQN

In traditional resource allocation problems, the Q-learning algorithm is often used. The problem of the Q-learning algorithm is that when the state space and action space are discrete and the dimension is not high, a Q-table can be used to store the Q value of each state-action pair. However, when the state space and action space are high-dimensional and continuous, the action space and state space are too large, and it is very difficult to use a Q-table. As an algorithm based on value iteration which is similar to Q-learning, DQN is a concrete implementation of the combination of a deep learning multi-layer convolution neural network (CNN) and Q-learning. When the state space and action space are high-dimensional and continuous, DQN can transform the update of Q-table into a function-fitting problem. By fitting a function instead of the Q-table to generate the Q value, similar states can obtain similar output actions. Therefore, we propose a DQN-based dynamic allocation algorithm for wireless resources to solve our optimization problem and dynamically allocate wireless resources in the access network.

3.1 Reconstruction of constrained Markov decision process (CMDP) based on DQN

The optimization problem in this study can be formulated as a CMDP problem (Xu et al., 2021). CDMP is closely related to reinforcement learning. CDMP uses a time-varying random variable to simulate the state of the system, and its state transition depends on the current state and the action vector applied to the system. A Markov decision process is used to calculate the action strategy, which will maximize the utility related to the expected reward. In this model, user admission control, PRB allocation, and power allocation are formulated as a CDMP problem, which can be denoted as a quadruple $\langle\mathcal{C},\mathcal{A},p_a\left(c'|c\right),R_a\left(c'|c\right)\rangle$, where $\mathcal{C}$ represents the finite set of states in the network and $\mathcal{A}$ represents the finite set of possible actions. When action a is taken in state c during the current time slot t, $p_a\left(c'|c\right)$ is the probability that the state will transition to c' from c. When the system transitions to state c' after performing action a in state c, $R_a\left(c'|c\right)$ is the reward function, indicating the immediate cost/reward, which reflects the learning objective. The basic elements include the system state, resource allocation behavior, state transition probability, and cost function.

Take state c as the input to the DQN algorithm. After the neural network analysis, the DQN

algorithm outputs the corresponding action. The main idea behind the algorithm is to approximate the distribution of Q values using the neural network training function f_{ap}. The Q value can be denoted as

$$Q\left(c,a\right)\approx f_{\mathrm{ap}}\left(c,a,\theta\right), \tag{16}$$

where Q denotes the main network's weight, and $Q\left(c,a\right)=\left[Q\left(c,a_1\right),Q\left(c,a_2\right),\cdots,Q\left(c,a_K\right)\right]$ (here, K is the maximum number of actions that can be taken in $\mathcal{A}$).

The target Q-network is updated only once in a period, while the main network is updated after each iteration. The target Q value can be denoted as

$$Q'=r\left(c,a\right)+\gamma\left[\max_{a'\in\mathcal{A}}Q\left(c',a',\theta^{-}\right)\right], \tag{17}$$

where the discount factor $\gamma\in[0,1)$ represents the decay degree of the reward function value, indicating the impact of the future reward on the current behavior choice, and θ^{-} is the target Q-network's weight. To improve the network prediction performance, it is required to learn and train the weight function repeatedly to fit complicated environmental data.

Fig. 3 depicts the DQN training procedure. In this training model, the optimization of weight θ is achieved by minimizing the loss function between the main network and the target Q-network, which can be described as

$$L\left(\theta\right)=E\left[\left(Q'-Q\left(c,a,\theta\right)\right)^2\right]. \tag{18}$$

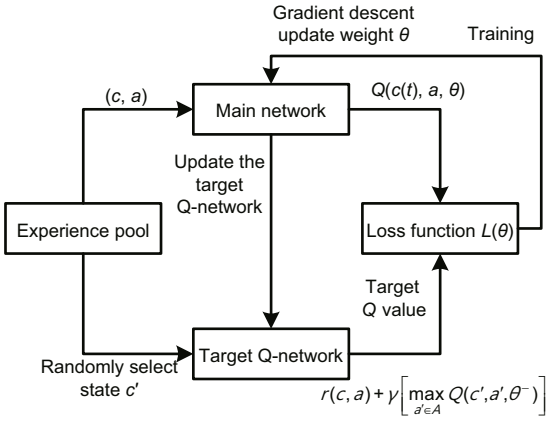

Fig. 3 Deep Q-learning network training model

The optimal allocation strategy for wireless resources can be found using the trained main network of the DQN algorithm after the main network has been trained. The process of the dynamic wireless resource allocation algorithm is organized as follows: in time slot t, the system state is specified as $c_t=\left(Q\left(t\right),H\left(t\right)\right)\in\mathcal{C}$, and the action is defined as $a_t=\left(a\left(t\right),\phi\left(t\right),p\left(t\right)\right)\in\mathcal{A}$. $\pi:\mathcal{C}\rightarrow\mathcal{A}$, which is a stability policy and can be expressed as $a=\pi\left(c\right)$, is the process of mapping the state space to the action space. According to the initial state c and the strategy $\pi\in\Pi$, where Π represents the set of all possible strategies, in time slot t, the expected cumulative network sum rate can be denoted as

$$\begin{aligned}\bar{r}^{\pi}\left(c\right)&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma_t r\left(c_t,a_t\right)|c_0=c\right\}\\&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma_t\sum_{j\in\mathcal{J}}r_j\left(t\right)|c_0=c\right\}.\end{aligned} \tag{19}$$

The expected cumulative sum delay of the total network radio interface is

$$\begin{aligned}\bar{d}^{\pi}\left(c\right)&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma_t d\left(c_t,a_t\right)|c_0=c\right\}\\&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma_t\sum_{j\in\mathcal{J}}d_j\left(t\right)|c_0=c\right\}.\end{aligned} \tag{20}$$

3.2 Algorithm implementation

The proposed algorithm's state, action, and reward are specifically defined as follows:

State: Define the state of the network system of the access network as $c_t=\left(Q\left(t\right),H\left(t\right)\right)\in\mathcal{C}$, including the global queue state information $Q\left(t\right)$ and the global channel state information $H\left(t\right)$.

Action: Action set a_t^* is defined as a series of vectors. Each vector represents user admission control, PRB, and power allocation on all BSs, satisfying $\left[a^*\left(t\right),\phi^*\left(t\right),p^*\left(t\right)\right]=\arg\max_{a(t),\phi(t),p(t)}\left(\bar{r}^{\pi}\left(c\right)\delta_{\mathrm{r}}-\bar{d}^{\pi}\left(c\right)\delta_{\mathrm{d}}\right)$, where $a^*\left(t\right)$, $\phi^*\left(t\right)$, and $p^*\left(t\right)$ represent the user admission control scheme, PRB, and power allocation strategy that satisfy the user experience maximization in time slot t, respectively.

Reward: Considering that the objective of this algorithm is to maximize the overall average user experience, the reward function is defined as the sum of user experience gained after all users associate BSs and allocate their PRB and power when constraints

C1–C7 are satisfied. Otherwise, it is defined as a negative feedback:

$$r(c,a)=\begin{cases}\sum\limits_{n\in\mathcal{N}}u^n, & \text{s.t. C1}-\text{C7},\\ -1, & \text{otherwise.}\end{cases}\tag{21}$$

The specific flow of the algorithm is shown in Algorithm 1. At step 3, the optimal action a_t under state c_t according to the output result of the latest main network is obtained. At step 4, the PRB and power allocation of the access network are jointly adjusted according to a_t, to ensure the QoS in real time and obtain the final user admission control scheme $a_{j,n}(t)$, power allocation strategy $p^b_{j,n}(t)$, and PRB allocation strategy $\phi^b_{j,n}(t)$. Then the resource allocation process ends.

Algorithm 1 DQN-based dynamic allocation

Input: system initial state c and the corresponding reward $r(c,a)$

1: **for** $t=1,2,\cdots,T$ **do**

2: In current time slot t, monitor the global state c_t of the access network, including the global channel state information $H(t)$ and the global queue state information $Q(t)$

3: Calculate the optimal power and PRB allocation actions, $a_t=\arg\max\limits_{a\in\mathcal{A}}Q(c_t,a,\theta)$

4: Adjust the power and PRB allocation depending on the optimal action a_t

5: $t=t+1$

6: **end for**

Output: user admission control scheme $a_{j,n}(t)$, power allocation strategy $p^b_{j,n}(t)$, and PRB allocation strategy $\phi^b_{j,n}(t)$

4 Simulation results and analysis

In this section, the overall user experience of the system and the average user experience of a single user are used as the performance evaluation indices to evaluate the feasibility of the built model and the effectiveness of the proposed algorithm. The algorithm proposed in this study is compared with the heuristic algorithm (Kalil et al., 2017) and the minimum distance allocation (MDA) algorithm (Zhang et al., 2021). In the heuristic algorithm, the weight of each user is calculated according to the queue state and channel state of each BS in the current time slot and the minimum resource requirement of each user. Based on the calculated user weight, network resources are allocated to the corresponding users according to the weight in each discrete resource scheduling time slot. In the MDA algorithm, each BS associates users according to the shortest distance, and each PRB allocates the same amount of power for users.

4.1 Simulation environment

In the simulations, we assume that four BSs are distributed uniformly in a 2 km × 2 km area. The coordinates are (0.5, 0.5), (0.5, 1.5), (1.5, 0.5), and (1.5, 1.5) km, and users are randomly distributed in the area. Assuming that there are three types of services required by users, the minimum rate requirements and the total radio interface delay requirements of different users are different, and the arrival process of user data packets follows an independent and identically distributed Poisson distribution. In addition, set the noise power $\sigma^2=10^{-7}$ mW. The optional power level on the PRB is $\{0,0.5,1\}$ dBm. The service rate unit price and the delay unit price are 5 per Mb/s and 1 per ms, respectively.

In the DQN-based dynamic allocation algorithm, a multi-layer CNN is used in the main network and target Q-network, including three convolution layers and two fully connected layers. The relevant information of each layer includes the size of the convolution kernel, the size of the convolution step, and the number of convolution kernels. The queue length of each BS is discretized into a finite number of equally spaced intervals, and each interval represents the current queue state. Therefore, the system state space in the constrained Markov problem is a finite state set. The parameters of the target Q-network are updated every 200 iterations. In the training process, the capacity of the DQN experience playback pool is set to 10 000. $\varepsilon=0.7$ is the probability value of an ε-greedy strategy. The remaining parameters are shown in Table 1.

4.2 Performance evaluation

Fig. 4 shows the changes of the user experience of the system of the three resource allocation algorithms with the advancement of time series when the number of users is 30 and the maximum transmission power of the BS is 39 dBm. The figure shows that the user experience of the proposed algorithm and the heuristic algorithm tends to be stable over time, while as a static resource allocation algorithm, the

user experience obtained by the MDA algorithm does not change with time. Compared with the heuristic and MDA algorithms, the proposed algorithm can obtain superior user experience on a long time scale.

Fig. 5 illustrates the relationship between the average user experience and the number of users when the maximum transmission power of the BS is 39 dBm on a long time scale. Fig. 5a shows the average user experience of all the users in the system, and Fig. 5b shows the average user experience of a single user. The simulation results show that compared with the heuristic and MDA algorithms, the proposed algorithm can obtain the maximum average user experience and has the greatest optimal effect on the user experience. Because the heuristic algorithm considers the user's minimum demand for resources, the heuristic algorithm can guarantee the service rate, but cannot achieve the optimal user experience. In the MDA algorithm, each PRB allocates the same amount of power for users, and resources cannot be flexibly and dynamically allocated according to the user's needs.

In Fig. 5a, when the number of users is small, the average user experience obtained by the heuristic algorithm is similar to that obtained by the proposed algorithm, because the network resources are relatively sufficient. With the increase in the number of users, the increase of the data rate revenue is greater than the total delay loss in the whole network, so the average user experience of all the users in the system increases. In addition, it can be seen from Fig. 5b that the average user experience of a single user decreases with the increase in the number of users, due to the limitation of radio resources in the network.

Table 1 Simulation parameters

Parameter	Value
Number of PRBs, Z	50
System bandwidth, W	10 MHz
Maximum transmission power of BS j, $p^j_{\max}$	20, 25, 30, 35, 39 dBm
Minimum data rate limit of user n, $r^n_{\min}$	5 Mb/s, 1 Mb/s, 51 kb/s
Maximum delay limit of user n, $d^n_{\max}$	10, 7.5, 1 ms
Pathloss from a BS to a user	37.6lg[d (km)]+128.1 dB
Noise power spectral density	−174 dBm/Hz
Data packet size, S	4 kb/packet
Maximum number of iterations	3000
Discount factor, γ	0.9
Time slot length	1 ms
Learning rate	0.0001

PRB: physical resource block; BS: base station

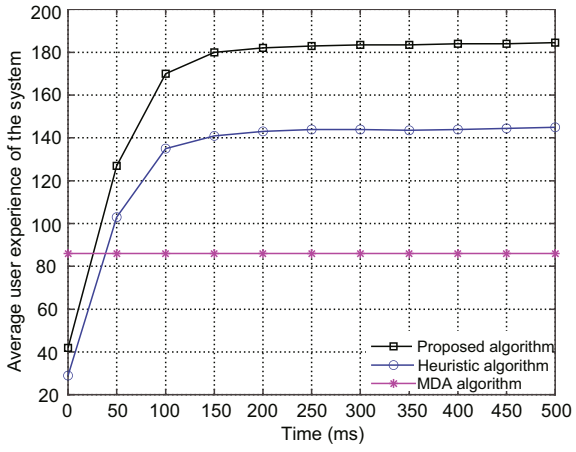

Fig. 4 Changes of the user experience of the system over time when the number of users is 30 and the maximum transmission power of the base station is 39 dBm

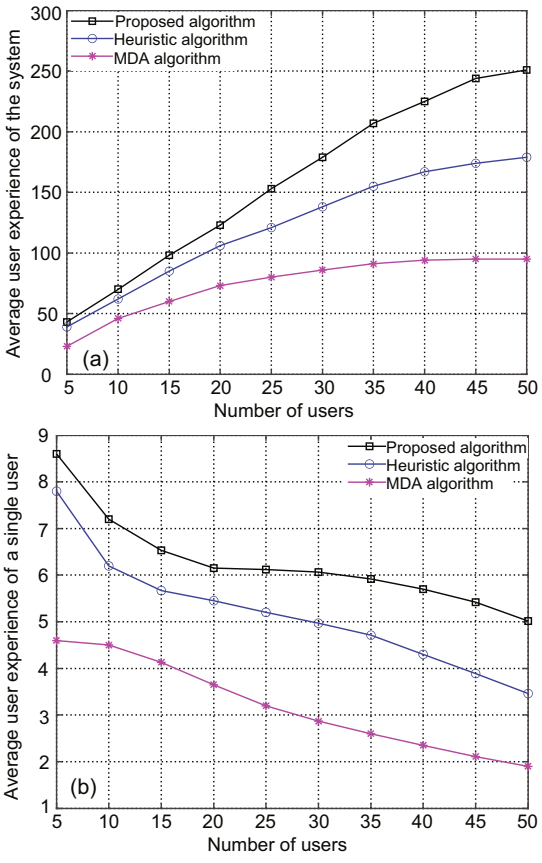

Fig. 5 Average user experience varying with the number of users when the maximum transmission power of the base station is 39 dBm: (a) average user experience of the system; (b) average user experience of a single user

When the number of users in the system is small, the network resources are relatively sufficient, and a single user can obtain a high data rate and a low delay. With the increase in the number of users in the system, the available resources are limited. When the number of users reaches a certain scale, the proposed algorithm can maintain only the user's minimum requirements for rate and delay. Therefore, the average user experience of a single user gradually decreases as the number of users increases. From the simulation results, it can be concluded that the proposed algorithm can maintain the optimal performance and maximize the average user experience regardless of the overall user experience of the system or the average user experience of a single user.

Fig. 6 shows the relationship between the average user experience of the system and the maximum transmission power of the BSs when the number of users is 30. It can be seen from Fig. 6 that the user experience of the three algorithms all increases with the increase in the maximum transmission power of the BSs. An increase in the transmission power of the BSs will boost the data rate revenue and improve the overall user experience of the system. When the maximum transmission power of the BSs is small, the average user experience of the MDA algorithm is negative, because the transmission power of the BSs is too small to guarantee the service rate and latency requirements of the surrounding users. By comparing these three algorithms, it can be concluded that the proposed algorithm can guarantee the maximum average user experience and has the best performance.

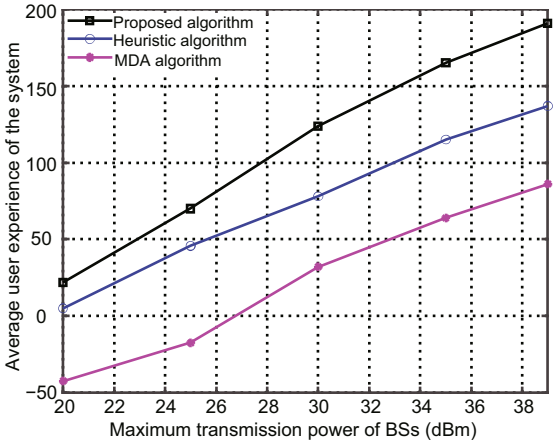

Fig. 6 Average user experience of the system varying with the maximum transmission power of the base stations (BSs) when the number of users is 30

5 Conclusions and future work

Considering the future wide-area coverage signaling cell scenario, we proposed a dynamic user-centric multi-dimensional resource allocation method. Considering the different QoS requirements of users in different industries, we constructed a dynamic allocation model for wireless resources. A DQN-based dynamic allocation algorithm for wireless resources was proposed to maximize the overall user experience. In the model, the network fully perceived its state through various measurements reported by the terminal. The proposed algorithm realized on-demand user admission control and dynamic resource allocation according to the requirements of rate and latency reported by users. The simulation results showed that the proposed algorithm can effectively improve the average user experience on a long time scale, while ensuring the user's minimum data rate requirements and latency constraints and ensuring low energy consumption of the network in the process of resource allocation, thus achieving the goals of optimizing the overall network utility in real time and realizing on-demand wireless resource allocation.

In the future research work, more types of resources can be considered in this paper's model, including communication resources, computing resources, and cache resources, to enable deeper integration of data, information, and communication technologies.

Contributors

Zhou TONG, Na LI, Junshuai SUN, and Guangyi LIU designed the research. Zhou TONG and Huimin ZHANG conducted the simulations. Zhou TONG drafted the paper. Na LI helped organize the paper. Zhou TONG, Quan ZHAO, and Yun ZHAO revised and finalized the paper.

Compliance with ethics guidelines

Zhou TONG, Na LI, Huimin ZHANG, Quan ZHAO, Yun ZHAO, Junshuai SUN, and Guangyi LIU declare that they have no conflict of interest.

Data availability

Data not available due to commercial restrictions. Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

References

Gang J, Friderikos V, 2019. Inter-tenant resource sharing and power allocation in 5G virtual networks. *IEEE Trans Veh Technol*, 68(8):7931-7943. https://doi.org/10.1109/TVT.2019.2917426

Ji CY, Bi MH, Zhou Z, et al., 2021. Online bandwidth resources allocation algorithm for multi-tenancy PON based on deep reinforcement learning. *Opt Commun Technol*, 45(9):36-39 (in Chinese). https://doi.org/10.13921/j.cnki.issn1002-5561.2021.09.009

Kalil M, Al-Dweik A, Sharkh MFA, et al., 2017. A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks. *IEEE Access*, 5(1):20814-20827. https://doi.org/10.1109/access.2017.2746666

Lin MT, Zhao YP, 2020. Artificial intelligence-empowered resource management for future wireless communications: a survey. *China Commun*, 17(3):58-77. https://doi.org/10.23919/JCC.2020.03.006

Liu GY, Deng J, Zheng QB, et al., 2022a. Native intelligence for 6G mobile network: technical challenges, architecture and key features. *J China Univ Posts Telecommun*, 29(1):27-40. https://doi.org/10.19682/j.cnki.1005-8885.2022.2004

Liu GY, Li N, Deng J, et al., 2022b. The SOLIDS 6G mobile network architecture: driving forces, features, and functional topology. *Engineering*, 8(1):42-59. https://doi.org/10.1016/j.eng.2021.07.013

Luo Y, Shi ZP, Zhou X, et al., 2014. Dynamic resource allocations based on Q-learning for D2D communication in cellular networks. 11th Int Computer Conf on Wavelet Active Media Technology and Information Processing, p.385-388. https://doi.org/10.1109/ICCWAMTIP.2014.7073432

Lü YP, Jia XD, Lu Y, et al., 2021. A deep Q-learning based resource allocation algorithm in indoor wireless networks. *Comput Eng Sci*, 43(7):1250-1255 (in Chinese).

Ren YJ, Sun YH, Peng MG, 2021. Deep reinforcement learning based computation offloading in fog enabled Industrial Internet of Things. *IEEE Trans Ind Inform*, 17(7):4978-4987. https://doi.org/10.1109/TII.2020.3021024

Xu H, Tong Z, Shen H, et al., 2021. Dynamic communication and computation resource allocation algorithm for end-to-end slicing in mobile networks. 3rd Int Conf on Artificial Intelligence for Communications and Networks, p.251-267. https://doi.org/10.1007/978-3-030-90196-7_22

Zhang TK, Wang XF, Yang LW, et al., 2021. A SFC deployment and computation resource allocation joint algorithm in mobile networks. *J Beijing Univ Posts Telecommun*, 44(1):7-13 (in Chinese). https://doi.org/10.13190/j.jbupt.2020-035

Zhao QY, Grace D, Vilhar A, et al., 2015. Using K-means clustering with transfer and Q learning for spectrum, load and energy optimization in opportunistic mobile broadband networks. Int Symp on Wireless Communication Systems, p.116-120. https://doi.org/10.1109/ISWCS.2015.7454310