User-centric Association in Ultra-dense mmWave Networks via Deep Reinforcement Learning

Qing Xue, Yao Sun, Jian Wang, Gang Feng, Li Yan, and Shaodan Ma

***Abstract*—For ultra-dense networks, user-centric architecture is regarded as a promising candidate to offer mobile users better quality of service. One of the main challenges of user-centric architecture is exploring efficient scheme for user association in the ultra-dense network. In this letter, we study dynamic user-centric association (UCA) problem for ultra-dense millimeter wave (mmWave) networks to provide reliable connectivity and high achievable data rate. We consider time-varying network environments and propose a deep Q-network based UCA scheme to find the optimal association policy based on the historical experience. Simulation results are presented to verify the performance gain of our proposed scheme.**

***Index Terms*—Ultra-dense mmWave network, user-centric, multiple association, deep learning.**

I. Introduction

Ultra-dense network (UDN) featured by the high density of small cells is promoted as one of the technology trends to meet the high capacity requirement of future mobile network. In UDN, various kinds of access points (APs) or small base stations (SBSs) are densely deployed to provide plenty of mobile access opportunities and flexible network connection choices. In [1], a reinforcement learning based power control scheme is proposed to suppress the downlink inter-cell interference and save energy for ultra-dense small cells. Deep Q-network (DQN) assisted resource allocation problem in ultra-dense cellular networks was studied in [2]. Moreover, to exploit the full potentials of UDN, the network architecture is shifted from traditional network/cell-centric to user-centric paradigm [3]. In a user-centric UDN, a user could ensure the satisfied quality of service (QoS) while moving around through dynamic AP/SBS grouping. Meanwhile, multi-connectivity scheme has been introduced to form the dynamic AP/SBS groups [4]. Thanks to the major advantages of large available bandwidth, immunity to interference, and effective integration with massive MIMO, millimeter wave (mmWave) technology [5], [6] can be effectively integrated with the UDNs requiring short-range high-rate communications. In return, the problems of large propagation path loss and high sensitivity to blockage of mmWave can be overcome. User association is identified as one of the critical issues to address in such networks since it affects the network performance principally.

Due to the extremely high density of SBSs, a user can receive multiple strong signals in UDNs. Thus, multiple association, which refers to the scenario where a user is allowed to connect to multiple SBSs, should be suitable. Although some literatures have studied the multiple association in UDNs [7]–[9], only a few of them are used for mmWave communications. Meanwhile, current methods for user association are generally based on the received power (*e.g.*, reference signal received power, RSRP, or received signal strength, RSS). However, due to denser topologies, such association would be highly inefficient in mmWave networks, as it may lead to unbalanced SBS loads and in turn affect the network fairness and may result in overly frequent handovers between the adjacent SBSs [10]. To improve the user experience in UDN, we introduce the user-centric association (UCA) paradigm [7] where the user is the central point in any considered association scheme. Different from traditional network-centric association, the UCA paradigm is supposed to provide seamless services for mobile users anywhere by "eliminating" the cell edge. This association paradigm would be a dominant factor in the network operation of future networks. Although the user-centric paradigm is a well-known concept, there are few solutions to UCA in the open literature.

In this letter, we first model UCA in mmWave UDN mathematically as a constrained optimization problem, and then solve it by applying a model-free reinforcement learning method. The main contributions can be summarized as follows. (1) To solve the UCA optimization problem, which is NP-hard due to the binary user association variables, we resort to DQN. In contrast to the existing literature, the action defined in DQN is executed only when the handover conditions are triggered, not immediately after it is selected. (2) Different from existing work, applying DQN to user association and network load balance of mmWave UDN is the first attempt, to the best of the authors' knowledge. Compared to the existing methods, simulation results demonstrate that the DQN-based UCA can effectively overcome the problem of unbalanced mSBS loads.

This work was supported in part by the NSFC under Grant 62001071, China Postdoctoral Science Foundation under Grant 2020M683291 and 2019TQ0270, Macao Young Scholars Program under Grant AM2021018, Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN201900617, Science and Technology Development Fund, Macau SAR (File no. 0036/2019/A1 and no. SKL-IOTSC-2021-2023), and the Research Committee of University of Macau under Grant MYRG2018-00156-FST. *(Corresponding author: Yao Sun.)*

Q. Xue is with the Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing 400065, China, and also with University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China. The UESTC and CQUPT are the co-first affiliations. (e-mail: *xueq@cqupt.edu.cn*). Y. Sun is with James Watt School of Engineering, University of Glasgow, UK (e-mail: *Yao.Sun@glasgow.ac.uk*). J. Wang and G. Feng are with the National Key Laboratory of Science and Technology on Communications, UESTC, Chengdu 611731, China (e-mails: *wangjians@std.uestc.edu.cn*, *fenggang@uestc.edu.cn*). L. Yan is with the Key Lab of Information Coding & Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: *liyan@swjtu.edu.cn*). S. Ma is with the State Key Laboratory of Internet of Things for Smart City and the Department of Electrical and Computer Engineering, University of Macau, Macao SAR, China (e-mail: *shaodanma@um.edu.mo*).

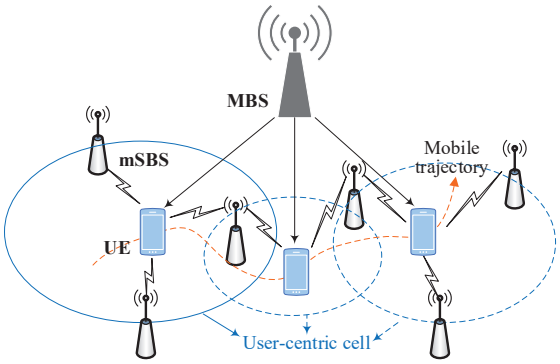

Fig. 1: An illustration of ultra-dense mmWave HetNet.

II. System Model and Problem Formulation

As shown in Fig. 1, we consider a two-tier heterogeneous network (HetNet), in which ultra-dense mmWave SBSs (mSBSs) are deployed randomly under the coverage of one macro BS (MBS) operating in conventional microwave band. The MBS and mSBSs are inter-connected via traditional backhaul X2 interfaces. For supporting robust communications, multi-connectivity technique, which enables user equipments (UEs) to simultaneously maintain multiple possible connections to different cells, is considered in the HetNet. Supposing that the low frequency link with high reliability between the MBS and a UE has been established, we focus on the optimum UCA policy, which is to determine which mSBSs the UE is associated with, to maximize the long-term rate performance of the *user-centric cell*. There may exist multiple potential serving mSBSs around a typical UE in the ultra-dense mmWave HetNet, but only a few of them can be selected to serve this UE. Meanwhile, the UE may switch the serving BSs, especially mSBSs, to meet its own QoS requirements as the HetNet environment (*e.g.*, the channel conditions) is time-varying. For the dynamic association management, as in [11], a time-slotted decision-making process is adopted. Specifically, the UE first chooses its serving mSBSs at the beginning of a slot and then starts data transmission with the associated mSBSs. We assume that a UE (e.g., UE u) equipped with M_u^{ant} antennas and M_u^{RF} radio frequency (RF) chains is associated with up to $M_u^{\max}$ mSBSs simultaneously. The UE and each serving mSBS communicate via one transmit-receive beam pair (one data stream), such that $M_u^{\max} \le M_u^{\text{RF}} \le M_u^{\text{ant}}$. Let $\mathcal{B}$ and $\mathcal{U}$ denote the set of mSBSs and UEs in the HetNet respectively, $x_{u,b}$ denotes the binary association indicator variable for UE $u \in \mathcal{U}$ and mSBS b ($b \in \mathcal{B}$), we have

$$x_{u,b} = \begin{cases} 1, & \text{if UE } u \text{ is associated with mSBS } b, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

For mmWave transmission, the path loss in dB can be modeled as [12] $L(d_{u,b}) = \alpha + 20\log_{10}(f_c) + 10\beta\log_{10}(d_{u,b}) + \chi$, where f_c is the carrier frequency in GHz, $d_{u,b}$ is the link distance in meters, α and β are the attenuation value and path loss exponent respectively, and χ represents the shadowing. The signal-to-interference-plus-noise ratio (SINR) of UE u when associated with mSBS b at time slot t is

$$\text{SINR}_{u,b} = \frac{p_{u,b} g_{u,b}^T g_{u,b}^R \Big/ L(d_{u,b})}{\gamma_{\text{mm}} W_{u,b} + I_u}, \tag{2}$$

where $p_{u,b}$ and $W_{u,b}$ are the transmit power and bandwidth allocated to UE u by mSBS b respectively, $g_{u,b}^T$ and $g_{u,b}^R$ are the transmit and receive antenna gain of link $\ell_{u,b}$ respectively, γ_{mm} is the noise power spectral density at the UE, and $I_u = \sum_{i \in \mathcal{I}_u} p_i \cdot g\left(\xi_i^T\right) \cdot g\left(\xi_i^R\right) \cdot L_i^{-1}$ is the interference received by $\ell_{u,b}$ from simultaneous mmWave links, where $\mathcal{I}_u$ is the set of these simultaneous links, p_i is the transmit power of link i ($i \in \mathcal{I}_u$), ξ_i^T and ξ_i^R are respectively the angles of link i's transmit beam and UE u's receive beam offset from the axis of UE u and the mSBS of link i, $g(\xi)$ denotes the mmWave antenna gain pattern with respect to the relative angle ξ to its boresight, and L_i is the path loss of link i. The high directivity and short distance of mmWave transmission enable few intra-/inter-cell interference among simultaneous links. Furthermore, the low interference can be eliminated by appropriate spatial precoding. Hence, all mSBSs can be assumed to employ the same frequency band, and UEs can access to the full available bandwidth W_{mm} of each mSBS. That is, $W_{u,b} = W_{\text{mm}}$ for $\forall u \in \mathcal{U}$, $\forall b \in \mathcal{B}$. The achievable rate of UE u from mSBS b can be given as $R_{u,b} \triangleq W_{\text{mm}} \log_2(1 + \text{SINR}_{u,b})$. We propose to use the total achievable rate $R_t^u = \sum_{\forall b \in \mathcal{B}_t^u} R_{u,b}$ delivered by all serving mSBSs to UE u as the metric to gauge the QoS at time slot t, where $\mathcal{B}_t^u$ is the set of the serving mSBSs. In this study, we formulate the UCA problem mathematically as a constrained optimization problem, which is to maximize the long-term network performance in terms of user achievable rate and network load balance, *i.e.*,

$$\begin{aligned} P1: \max_{\mathbf{x}_u} & \left\{ \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \frac{\sum_{b\in\mathcal{B}} x_{u,b} R_{u,b}(t)}{\sum_{b\in\mathcal{B}_t^u} \sum_{v\in\mathcal{U}} x_{v,b}} \right\} \\ s.t. & \begin{cases} C1: x_{u,b} \in \{0,1\}, \\ C2: 0 \le \sum_{b\in\mathcal{B}} x_{u,b} \le M_u^{\max}, \\ C3: \text{SINR}_{u,b} \ge \phi, \\ C4: 0 \le \sum_{b\in\mathcal{B}} x_{u,b} p_{u,b} \le p_{\max}, \end{cases} \end{aligned} \tag{3}$$

where ϕ is a given SINR threshold, $p_{\max}$ is the maximum transmit power for each UE, and $\mathbf{x}_u \triangleq [x_{u,b}]_{b\in\mathcal{B}}$ denotes the association matrix. In P1, C2 represents the constraint on the number of serving mSBSs for UE u, C3 represents the QoS constraint which ensures that each link can achieve the minimum SINR to meet QoS requirement, and C4 represents the total power constraint. A learning-based UCA algorithm for solving $P1$ is proposed in the following section.

III. DQN-based Association Scheme for User-centric Cell

For the time-varying HetNet environment that lacks accurate information and model, the decision-making process of user association can be defined as a partially observable Markov decision process (POMDP) due to (i) the Markov property, *i.e.*, the behavior of this stochastic process at times in the future

depends on only the present network state, and (ii) no other than local state can be observed as the strategies and states of other UEs are impossible to be known without information interaction. We apply DQN to provide the intelligent scheme for UCA. Each UE learns the best association decision from its past experience automatically and independently. In this study, the essential components of DQN are defined as follows.

1) *Agent*: Each UE in $\mathcal{U}$ acts as an independent agent.

2) *State*: Let s_t^u denote the state observed by UE $u \in \mathcal{U}$ at time slot t for characterizing the HetNet environment. For a fully distributed self-learning association scheme, the observation of UE u includes its current serving mSBSs chosen at the last time slot, *i.e.*., $\mathcal{B}_{t-1}^u$, the public information of the HetNet, and the candidate serving mSBSs. The public information (*e.g.*, the number of served UEs of each mSBS) can be obtained from the MBS at the beginning of each time slot. The number of served UEs for mSBS b ($\forall b \in \mathcal{B}$) is $U_t^b = \sum_{u \in \mathcal{U}} x_{u,b}(t)$, which can be uploaded to the MBS via X2 interface. Thereby, the public information at time slot t is $\mathcal{P}_t = \left\{U_t^b; \forall b \in \mathcal{B}\right\}$. In terms of resource (*e.g.*, power) allocation, generally, the less the number of served UEs, the more the power available for each UE, and the better the corresponding QoS. Each mSBS $b \in \mathcal{B}$ broadcasts reference signals to UEs at the start of each time slot, and then UE u determines its candidate serving mSBSs by measuring the received signal strength of the reference signals. Denoting the set of candidate serving mSBSs at time t as $\mathcal{B}_{t,\text{can}}^u$ ($\mathcal{B}_{t,\text{can}}^u \cap \mathcal{B}_{t-1}^u = \emptyset$), we define $s_t^u = \left\{\mathcal{P}_t, \mathcal{B}_{t,\text{can}}^u, \mathcal{B}_{t-1}^u\right\}$.

3) *Action*: In the POMDP, the network states change with the actions which are defined as the association policies of the user-centric cell in this work. Let $\mathcal{A}_t = \left\{a_{t,0}^u, a_{t,1}^u\right\}$ denote the action space of UE u at time slot t, where $a_{t,0}^u$ denotes a silent action and $a_{t,1}^u$ denotes a handover to the target mSBSs. When UE u chooses $a_{t,0}^u$, no association request is sent. It means that the association with the current serving mSBSs is the best choice or the handover trigger conditions are not met. The handover trigger conditions can be expressed as

$$\begin{cases} \mathcal{B}_{t,\text{can}}^u \neq \emptyset, & \text{(4a)} \\ \exists \tau_0 \in \left[(t-T)\,\tau, t\tau\right], s.t.\ R_t^u < \eta, & \text{(4b)} \\ R_t^{u*} - R_t^u \geq R_\Delta > 0, & \text{(4c)} \end{cases}$$

where τ is the length of each time slot, τ_0 is the maximum endurance time that the UE allows the achievable rate from the serving mSBSs to be lower than the threshold η in the past T slots, R_t^{u*} is the estimated transmission rate of the target mSBSs, and R_Δ is a given threshold for rate increment between the target mSBSs and the serving mSBSs. Specifically, Eq. (4a) indicates that there are some candidate serving mSBSs for UE u, (4b) indicates that the UE's time in the receiving rate below the threshold in the past T slots exceeds the tolerance time limit, and (4c) indicates that the estimated transmission rate of the target mSBSs is higher than the rate of the current serving mSBSs. In our scheme, the action $a_{t,1}^u$ is executed only when the handover conditions are triggered, not immediately after it is selected. It is because that, although performing $a_{t,1}^u$ may get better rate performance, there will be execution overhead in terms of time and radio resource, and the overhead cost may be higher than the performance improvement.

4) *Reward*: In order to optimize the UE achievable rate and the network load balance, we model the reward function when UE u selects $a_t^u \in \mathcal{A}_t$ as

$$r_t^u\left(s_t^u, a_t^u\right) = \sum_{b \in \mathcal{B}_t^u} \int_{t+\frac{\Delta\tau}{\tau}}^{t+1} x_{u,b} R_{u,b}\left(t\right) dt - \lambda \sum_{b \in \mathcal{B}_t^u} \mathcal{P}_t\left(b\right), \quad (5)$$

where $\Delta\tau$ ($\Delta\tau < \tau$) denotes the interruption time used for association, $\mathcal{P}_t\left(b\right) = U_t^b$, λ is a weight factor, and $x_{u,b} = 1$ if the association with the serving mSBS b is retained or the association request to the target mSBS b is accepted. The reward function represents a tradeoff among the user achievable rate and the network load balance and provides a reward signal for the UE to evaluate the impact of new association policy. Specifically, the first term of the reward is the total achievable rate of UE u adopting new association policy, and the second term is the user load of the network.

5) *Q-Function Approximation*: For $u \in \mathcal{U}$, $a_t^u \in \mathcal{A}_t$, the Q-function that calculates the quality of a given state-action pair $\left(s_t^u, a_t^u\right)$ is approximated as

$$Q\left(s_t^u, a_t^u; \theta\right) \triangleq f\left(s_t^u, a_t^u; \theta\right), \quad (6)$$

where $f(\cdot)$ is a function which can be represented by DNN parameterized by θ. A target Q-function $\hat{Q}\left(s_t^u, a_t^u; \theta^-\right)$, with parameters θ^-, is the same as $Q\left(s_t^u, a_t^u; \theta\right)$ except that θ^- is copied every C training steps from θ. In DQN, the estimated Q-value $Q\left(s_t^u, a_t^u; \theta\right)$ is used for training and $\hat{Q}\left(s_t^u, a_t^u; \theta^-\right)$ is used for evaluating the real value of selecting action a_t^u under state s_t^u. The target used by DQN is then

$$Y_t^{\text{DQN}} = r_{t+1}^u + \rho \max_{a_t^u \in \mathcal{A}} \hat{Q}\left(s_{t+1}^u, a_t^u; \theta^-\right), \quad (7)$$

where ρ ($0 < \rho < 1$) is the discount factor for future rewards.

6) *Loss Function*: The loss function $Loss\left(\theta\right)$ is defined as the mean squared error between Y_t^{DQN} and $Q\left(s_t^u, a_t^u; \theta\right)$, *i.e.*,

$$Loss\left(\theta\right) = \mathbb{E}_{d \sim D}\left[\left(Y_t^{\text{DQN}} - Q\left(s_t^u, a_t^u; \theta\right)\right)^2\right], \quad (8)$$

where d is a mini-batch data sampled from the experience replay memory $D = \left\{..., \left(s_t^u, a_t^u, r_t^u, s_{t+1}^u\right), ...\right\}$. During the learning procedure, θ is updated at each iteration to minimize the loss function, in general, by utilizing gradient descent.

Based on this model, the DQN-based UCA scheme can be summarized as **Algorithm 1**, where M_u and p_u are the number of serving mSBSs for UE u and the total transmit power of these mSBSs for the UE respectively, η_1 is a given threshold of $R_{u,b}$, υ is the learning rate, N_c is the number of training steps to the convergence of DQN. The computational complexity of the DQN-based UCA denoted by $\mathcal{O}_{\text{UCA}}$ is evaluated by jointly considering the computational complexity of (re-)association/handover and DQN, which is given by $\mathcal{O}_{\text{UCA}} = \mathcal{O}_{\text{handover}} + \mathcal{O}_{\text{DQN}} = \mathcal{O}\left(\min\left\{M_u^{\max}, \left|\mathcal{B}_{t,\text{can}}^u\right|\right\}\right) + \mathcal{O}\left(N_e N_c\right)$, where N_e is the overall operation epochs in the training phase. In fact, for DQN, the expensive computational complexity can be done at the offline training phase. Therefore, the proposed algorithm is implementation-friendly with low computational

complexity for solving the problem $P1$. For the traditional association based on maximum rate (i.e., choosing the mSBSs with highest achievable rate to be associated with), the computational complexity is $\mathcal{O}\left(\sum_{i=1,2,\ldots,M_u^{\max}} C_{|\mathcal{B}|}^{i}\right) + \mathcal{O}_{\text{handover}}$ which is usually much higher than that of the proposed algorithm. To meet QoS requirements, UE u first associates with mSBS b that meets $rss \leq RSS_b = \max_{\forall i \in \mathcal{B}_{t,\text{can}}^u} RSS_i$, where RSS_b is the received signal strength of reference signals from mSBS b and rss is a given threshold. Then the UE establishes high-quality directional mmWave link with mSBS b by beam training [5], [6]. In general, the QoS constraint $\text{SINR}_{u,b} \geq \phi$ in Eq. (3) can be satisfied and, moreover, can be ensured by beam tracking when the UE moves.

Algorithm 1 DQN-based UCA scheme

Input: Network topology (mSBS and UE distributions).

Initialize DQN, including θ, $\theta^- = \theta$, and D;
Initialize initial state, υ, ρ, C, N_c;
for dynamic user-centric cell **do**
  Find the current state $s_t^u = \left\{\mathcal{P}_t, \mathcal{B}_{t,\text{can}}^u, \mathcal{B}_{t-1}^u\right\}$.
  Choose action a_t^{u*} according to ε-greedy strategy [13]:
  Set the value of random selection probability ε;
  Generate a random value in $(0,1)$, record it with ε_0;
  If $\varepsilon_0 > \varepsilon$, $a_t^{u*} = \max_{a_t^u \in \mathcal{A}_t} \hat{Q}\left(s_{t+1}^u, a_t^u; \theta^-\right)$;
  Otherwise, randomly select an action in $\mathcal{A}_t$ as a_t^{u*}.
  if $a_t^{u*} = a_{t,0}^u$ **then**
    No association request is sent, and $\mathcal{B}_t^u = \mathcal{B}_{t-1}^u$.
  else if association trigger conditions in (4) are met **then**
    Set $\mathcal{B}_t^u = \mathcal{B}_{t-1}^u$.
    Switch from the m_u serving mSBSs that $R_{u,b}(t) < \eta_1$ to the m_u best candidate mSBSs in $\mathcal{B}_{t,\text{can}}^u$, and update $\mathcal{B}_t^u$ accordingly.
    Remove the selected candidate mSBSs out of $\mathcal{B}_{t,\text{can}}^u$.
    Calculate $M_u = |\mathcal{B}_t^u|$.
    repeat
      Request to associate with the best mSBS in $\mathcal{B}_{t,\text{can}}^u$, *e.g.*, b, and remove it out of $\mathcal{B}_{t,\text{can}}^u$.
      if $x_{u,b} = 1$ **then**
        Record mSBS b into $\mathcal{B}_t^u$.
        Update $M_u = M_u + 1$.
      end if
    until $M_u = M_u^{\max}$ or $\mathcal{B}_{t,\text{can}}^u = \emptyset$.
    Perform beam training/alignment to establish high quality mmWave links between the UE and the serving mSBSs.
  end if
  Get the reward r_t^u and new state s_{t+1}^u.
  Store transition $\left(s_t^u, a_t^u, r_t^u, s_{t+1}^u\right)$ in D.
  for training step $n_c = 1$ to N_c **do**
    Sample random mini-batch d from D.
    Perform a gradient descent step on $Loss(\theta)$ with respect to θ, *i.e.*, $\theta \leftarrow \theta - \upsilon \cdot \nabla_\theta Loss(\theta)$.
    Update $\hat{Q} = Q$ every C steps by resetting $\theta^- = \theta$.
  end for
end for

Output: Association decision, *i.e.*, the serving mSBS set $\mathcal{B}_t^u$.

TABLE I: DQN structure in the simulation.

Layer	Generate
1	nn.Sequential(nn.Linear(in_dim, n_hidden_1), nn.ReLU())
2	nn.Sequential(nn.Linear(n_hidden_1, n_hidden_2), nn.ReLU())
3	nn.Sequential(nn.Linear(n_hidden_2, n_hidden_3), nn.ReLU())
4	nn.Sequential(nn.Linear(n_hidden_3, out_dim))

IV. Performance Evaluation

In this section, we conduct numerical simulations to compare the performance of proposed DQN-based UCA with the other two traditional association schemes, namely choosing the mSBSs with highest achievable rate to be associated with (*i.e.*, maximum rate association, Rate-max) and being associated with the nearest mSBSs (*i.e.*, minimum distance association, Dis-min). The Dis-min method usually corresponds to the association based on the maximum received power (*e.g.*, RSRP or RSS). We consider that mSBSs are uniformly distributed in a rectangular area of 500×500 m^2 and set the transmit frequency as 28 GHz with available bandwidth 2 GHz. In the simulation, we use a set of random values to represent the number of users in each small cell, to simulate the problem of user density. Meanwhile, an mSBS is able to admit at most 6 UEs simultaneously, and the maximal number of mSBSs that a UE is allowed to be associated with at a time is set to 9. Moreover, the transmit power is set to $p_{u,b} = 30$ dBm, the parameters of path loss model $L(d_{u,b})$ are respectively set to $\alpha = 32.4$ dB, $\beta = 2.1$ and $\chi \sim \mathcal{N}\left(0, 4^2\right)$, the SINR threshold is set to $\phi = -32$ dB, and the noise power spectral density is set to $\gamma_{\text{mm}} = -10$ mW/GHz. To simplify our simulations, the transmit beam gain and the receive beam gain are assumed to be $g_{u,b}^T \approx g_m^T = 20$ dB and $g_{u,b}^T \approx g_m^R = 10$ dB respectively. Meanwhile, as mmWave networks have been shown to be noise-limited rather than interference-limited, the SINR can be approximated by the SNR for mmWaves which is quite different than the trend in sub-6GHz networks [14]. For the learning part settings, we mainly use the nn.Module, nn.Sequential, and nn.Linear of PyTorch to build a four layer DQN, as given in Table I, where in_dim =2 $|\mathcal{B}|$+1, n_hidden_1 = 90, n_hidden_2 = 120, n_hidden_3 = 90, and out_dim = 1. Specifically, the two parameters in nn.Linear() are the size of each input sample and the size of each output sample respectively. Meanwhile, we set the size of experience replay pool is 1000, the batch-size is 32, and the value of random selection probability $\varepsilon_0 = 0.1$. All the other settings such as reward, action, state keep consistent with those in our modeling part.

In the first experiment, we evaluate the convergence of the proposed DQN-based UCA algorithm under three typical learning rates 0.03, 0.06 and 0.1. Fig. 2 compares the average loss of DQN-based UCA with the relationship of the number of iterations under the three learning rates. We find that all the three curves reach the convergence after dozens or hundreds of iterations. As expected, the higher learning rate, the faster convergence speed. For example, the algorithm can be converged after around 70 iterations when learning rate is 0.1 while around 220 iterations of learning rate 0.03. These convergence results clearly demonstrate the

effectiveness and rationality of UCA. We set the learning rate to 0.06 in the following experiments. In the second experiment, we compare load balance of the network under the three association schemes, *i.e.*, DQN-based UCA, Rate-max and Dis-min. Fig. 3 shows the variance of UE load on each mSBS under different number of mSBSs. We find that the proposed DQN-based UCA outperforms the other two schemes when the number of mSBSs exceeds 9. This is because the designed DQN-based algorithm helps each UE to better understand the load of the surrounding mSBSs after the interactions with the network environment. Finally, we evaluate the transmission rate of DQN-based UCA with comparisons of the other two traditional schemes, Rate-max and Dis-min. Fig. 4 shows the transmission rate of the three association schemes under varying number of mSBSs from 4 to 18. It can be found that Rate-max achieves the highest transmission rate due to that the UEs always select the mSBSs with high data rate provisioned at each decision time. Moreover, we find that the DQN-based UCA achieves the second highest transmission rate with relatively small difference of that in Rate-max but much higher than that of Dis-min. Essentially, the DQN-based UCA provides a better tradeoff between UE load balance and transmission rate. For example, when the number of mSBSs is 18, the variance of UE load on each mSBS is about 46 and the transmission rate is 115 Mbps for the DQN-based UCA, while around 52 and 119 Mbps respectively for the Rate-max. We see that the DQN-based UCA can overcome the problem of unbalanced mSBS loads which in turn affect the network fairness and may result in overly frequent handovers between the adjacent mSBSs.

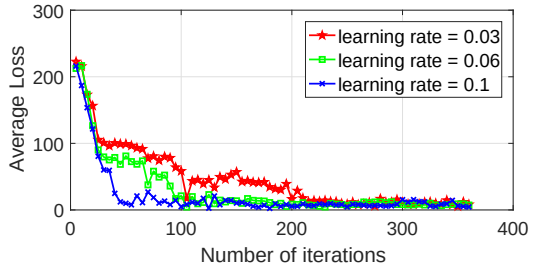

Fig. 2: Convergence of the DQN-based UCA.

V. Conclusion

In a dense mmWave HetNet, the user association is a quite challenging yet of paramount importance issue which affects not only users but the whole network performance. In this letter, we break the conventional rule of network-centric while coming up with the scheme of user-centric association. Specifically, we exploit deep reinforcement learning to let user intelligently choose multiple mSBSs to access at a time. Thus, the long-term system load balance and network throughput can be improved which are also verified by our simulations. In general, this work can be seen as a pioneer of using artificial intelligence (AI) to solve user association problem under a UDN via the view of user-centric.

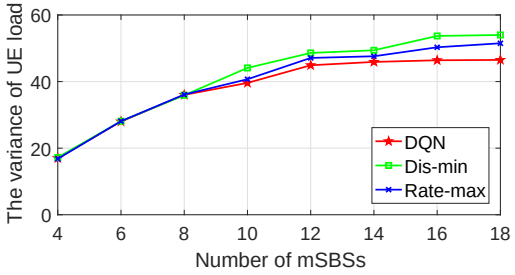

Fig. 3: Comparisons of load balance for the three schemes.

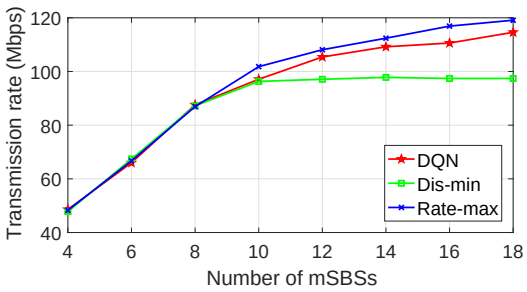

Fig. 4: Comparisons of transmission rate for the three schemes.

References

[1] L. Xiao, H. Zhang, Y. Xiao, X. Wan, S. Liu, L.-C. Wang, and H. V. Poor, "Reinforcement learning-based downlink interference control for ultra-dense small cells," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 423–434, 2020.

[2] X. Liao, J. Shi, Z. Li, L. Zhang, and B. Xia, "A model-driven deep reinforcement learning heuristic algorithm for resource allocation in ultra-dense cellular networks," *IEEE Trans. Veh. Technol*, vol. 69, no. 1, pp. 983–997, 2020.

[3] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78–85, 2016.

[4] H. Zhang, W. Huang, and Y. Liu, "Handover probability analysis of anchor-based multi-connectivity in 5G user-centric network," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 396–399, 2019.

[5] Q. Xue, X. Fang, and C.-X. Wang, "Beamspace SU-MIMO for future millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1564–1575, 2017.

[6] Q. Xue, X. Fang, M. Xiao, S. Mumtaz, and J. Rodriguez, "Beam management for millimeter-wave beamspace MU-MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 205–217, 2019.

[7] M. Kamel, W. Hamouda, and A. Youssef, "Performance analysis of multiple association in ultra-dense networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3818–3831, 2017.

[8] X. Wang, L. Li, J. Li, C. Yang, L. Wang, and D. Niyato, "Traffic-aware multiple association in ultradense networks: A state-based potential game," *IEEE Syst. J*, vol. 14, no. 3, pp. 4356–4367, 2020.

[9] R. Liu, G. Yu, and G. Y. Li, "User association for ultra-dense mmWave networks with multi-connectivity: A multi-label classification approach," *IEEE Wireless Commun. Lett*, vol. 8, no. 6, pp. 1579–1582, 2019.

[10] A. S. Cacciapuoti, "Mobility-aware user association for 5G mmWave networks," *IEEE Access*, vol. 5, pp. 21 497–21 507, 2017.

[11] C. Shen and M. van der Schaar, "A learning approach to frequent handover mitigations in 3GPP mobility protocols," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[12] A. Maltsev, V. Erceg, E. Perahia, C. Hansen, R. Maslennikov, A. Lomayev, A. Sevastyanov, A. Khoryaev, G. Morozov, M. Jacob, S. Priebe, T. Kurner, S. Kato, H. Sawada, K. Sato, and H. Harada, "Channel models for 60 GHz WLAN systems," IEEE 802.11-09/0334r8, May 2010.

[13] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 2014.

[14] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6244–6258, 2016.