Beam Management in Ultra-dense Millimeter Wave Network via Federated Learning

Jian Wang[†], Qing Xue [†,*], Yao Sun[+], Gang Feng [†], Lun Tang[*], and Shaodan Ma[‡]

[†]*National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China*
[*]*Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Post and Telecommunications, China*
[+]*James Watt School of Engineering, University of Glasgow, UK*
[‡]*State Key Laboratory of Internet of Things for Smart City, University of Macau, China*
E-mails: xueq@cqupt.edu.cn, fenggang@uestc.edu.cn

***Abstract*—Millimeter wave (mmWave) communication is one of the key technologies in 5G and beyond systems to address the tremendous growth in mobile data traffic owing to the abundant spectrum resources. Ultra-dense network deployment is a promising solution to combat the limited coverage, high propagation loss and attenuation of mmWave signals. This study investigates the beam management, with focus on beam configuration of mmWave base stations, in the ultra-dense mmWave network. To fulfill adaptive and intelligent beam management while protecting user privacy, we employ a double deep Q-network under a federated learning to tackle the beam management problem which is formulated to maximize the long-term system throughput. Simulation results demonstrate the performance gain of our proposed scheme.**

***Index Terms*—Ultra-dense networks, millimeter wave (mmWave), federated learning, beam management.**

I. Introduction

Millimeter wave (mmWave) communication has been considered as a promising means to meet the projected requirements by dint of the abundant spectrum resources. However, compared with the traditional microwave communication networks, mmWave communication networks face two critical challenges. One is the limited coverage because of the serious propagation path loss. To tackle this issue, ultra-dense network (UDN) [1] is essential and promising, where various small cell base stations (SBSs) with different coverage are densely deployed, and thus the distance between users and SBSs becomes closer. The other is the susceptible to blockage due to the inherent directivity. An efficient strategy for enabling reliable transmissions and enhanced data rates is to employ the multi-connectivity technique [2], which enables a user to connect to several SBSs simultaneously.

For an ultra-dense mmWave system, the number of operating beams should be extremely large caused by the densely deployed mmWave SBSs (mSBSs). This leads to a more complex and critical beam management problem than that in conventional non-dense mmWave systems. In current mmWave systems, the term beam management is usually known as fine alignment of the transmitter and receiver beams to perform a variety of control tasks including initial access for idle users and beam tracking for connected users [3]. Some recent papers [3]–[6] have been published to provide overview of beam management for mmWave in 5G New Radio (NR) standard. Beam management procedures for handling mobility can be categorized into beam sweeping, beam measurement and reporting, beam determination, beam maintenance, and beam failure recovery [6]. To date, most of the efforts in beam management tackle the problem by resorting to beam training, sparse channel estimation, and location aided beamforming [7]–[9]. As mobile environments are increasingly complex, heterogeneous and evolving, machine learning (ML) techniques have attracted significant attention to optimize beam management in mmWave communication systems, especially in mobile applications or dynamic environments [10]–[14]. For example, the authors in [10] proposed a reinforcement learning based handoff policy to reduce the number of handoffs while maintaining user QoS requirements in mmWave HetNets. A deep learning based coordinated beamforming algorithm is proposed in [12] to reduce the training overhead. In [13], the authors proposed a deep neural network (DNN)-based beam management and interference coordination algorithm to reduce the interference and improve the sum-rate of dense mmWave network.

Although the existing ML-based mechanisms bring a number of benefits in beam management, they also expose some potential risks, most typically, in security and privacy protection. It is because that traditional centralized ML techniques normally require data collection and processing by a central controller, but the training data may be privacy sensitive in nature. This problem is becoming a bottleneck of large-scale implementation of traditional centralized ML schemes in daily life. Moreover, the overhead caused by centralized data aggregation and processing is often large in quantity. These reasons have led to a growing interest in a new ML model, namely federated learning (FL) [15]. In FL, participating learners collaboratively train a shared model by exploiting

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1806805, the National Natural Science Foundation of China under Grant 62001071 and 62061007, the China Postdoctoral Science Foundation under Grant 2020M683291, the Macao Young Scholars Program under Grant AM2021018, and the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN201900617. Q. Xue (the corresponding author) and J. Wang contributed equally to this work and should be regarded as co-first authors.

their local computation capability and training data, and thus only local model updates instead of raw data need to be transferred to a centralized parameter server. Thereby, FL has distinct privacy advantages and provides low latency, reduced power consumption. These natures of FL motivate this work, which is the first time to exploit FL to design an mmWave beam management scheme in the open literature.

In this work, in order to improve beam utilization and reduce inter beam interference, we manage beams in a systematic manner instead of beam-by-beam basis. The systematic beam management mainly refers to the dynamic control of beam directions at the mSBS side (*i.e.*, mSBS beam configuration) based on periodically sensing instantaneous user distributions. Due to the preservation of private user data (*e.g.*, user locations), a decentralized learning approach under a FL framework is employed in the adaptive beam management, named BMFL. The key contributions can be summarized as follows.

- BMFL is underpinned by FL to intelligently tackle the complicated beam management problem aiming at maximizing the long-term system throughput.
- We address the issue of data privacy in BMFL by using FL to avoid any exchange of user private information (such as location, trajectory, behavior).
- To improve the learning convergence speed, we propose to employ data cleaning technique in our FL algorithm, by using only informative and valuable data for training.

The remainder of this paper is organized as follows. System model and systematic beam management problem are described in Section II. In Section III, algorithm of BMFL is presented. Performance of BMFL is evaluated in Section IV. Finally, Section V concludes the paper.

II. System Model and Problem Formulation

A. System Model

We consider a two-tier heterogeneous network, in which ultra-dense mSBSs are deployed randomly under the coverage of one macro BS (MBS) operating on conventional microwave band. Throughout the paper, we call this type of network as ultra-dense mmWave network (UDmmN). In order to communicate and exchange control information, the MBS and mSBSs are inter-connected via traditional backhaul X2 interfaces. We assume that all mSBSs share the total mmWave bandwidth W_{mm}. The mSBSs are denoted by $\mathcal{B} = \{1, ..., B\}$ and the users moving randomly within the UDmmN are denoted by $\mathcal{U} = \{1, ..., U\}$, where $B = |\mathcal{B}|$ and $U = |\mathcal{U}|$.

Due to the hardware limitation, we assume that mSBS b $(b \in \mathcal{B})$ can form up to M_b narrow transmit beams simultaneously by adopting beamforming technique. Meanwhile, we divide the small cell b $(b \in \mathcal{B})$ into S_b transmit sectors (or beam directions) with respect to the condition $0 < M_b \leq S_b$. We assume that the beams covering different sectors are mutually orthogonal in space, and each beam can serve multiple users within its coverage, for example, in a time division multiplexing manner. In order to improve beam utilization, beam management leveraging double deep Q-network (DDQN) under a FL framework is used in our work.

Meanwhile, we assume that beam management is performed in a synchronous time-slotted fashion [16] and the beams are static during each time slot. Within each slot, there are three main operations. (i) At the beginning of a slot, mSBSs calculate the accumulated network performance over previous slots, and then manage beams by adopting FL-based algorithm. (ii) Users choose suitable mSBSs to associate with. (iii) Transmit data at the selected beam/mSBS. In UDmmN, in addition to the MBS, user u $(u \in \mathcal{U})$ may receive data from several mSBSs surrounding it and can associate with up to $B_u^{\max}$ of them. Let $x_{u,b}(t) = \{0, 1\}$ denote the binary association indicator variable for user u $(u \in \mathcal{U})$ and mSBS b $(b \in \mathcal{B})$, where $x_{u,b}(t) = 1$ if user u is associated with mSBS b at time t, otherwise $x_{u,b}(t) = 0$. Denoting by $\mathcal{B}_u(t)$ $(\mathcal{B}_u(t) \subseteq \mathcal{B})$ the set of mSBSs associating with user u at time t, the number of these mSBSs is $B_u(t) = |\mathcal{B}_u(t)| = \sum_{b \in \mathcal{B}} x_{u,b}(t) \leq B_u^{\max}$. The signal-to-interference-plus-noise ratio (SINR) of user u receiving from mSBS b can be expressed as

$$SINR_{u,b} = \frac{p_{u,b} \cdot G_{u,b}^T \cdot G_{u,b}^R / PL(d_{u,b})}{p_N + \sum_{k \in \mathcal{B}, k \neq b} \frac{p_{u,k} \cdot G_{u,k}^T \cdot G_{u,k}^R}{PL(d_{u,k})}}, \quad (1)$$

where $p_{u,b}$ is the allocated transmit power of mSBS b to user u, $G_{u,b}^T$ and $G_{u,b}^R$ are the transmit and receive antenna gain respectively, $d_{u,b}$ is the transmission distance between user u and mSBS b and $PL(d_{u,b})$ is the corresponding propagation path loss, p_N is the noise power, and the right part of the denominator represents the total power of interfering signals.

B. Problem Formulation

For the case that the user is associated with multiple small BSs, the achievable rate of user u at time t should be the sum of data rate received from all the associated BSs. Thus, the data rate can be expressed as

$$r_u(t) = \begin{cases} \sum_{b \in \mathcal{B}_u(t)} W_{mm} \log_2(1 + SINR_{u,b}(t)), & \text{if } \mathcal{B}_u(t) \neq \emptyset, \\ \frac{W_{mbs}}{N_{mbs}(t)} \log_2(1 + SINR_{u,mbs}(t)), & \text{if } \mathcal{B}_u(t) = \emptyset, \end{cases} \quad (2)$$

where W_{mbs} is the total available bandwidth of the MBS, $N_{mbs}(t)$ is the number of users served by the MBS at time t, and $SINR_{u,b}(t)$ $(SINR_{u,mbs}(t))$ is the obtained SINR of user u from mSBS b (the MBS) at time t. Here we assume that both the bandwidth and transmit power are evenly allocated to the serving users in each beam. Hence, the system throughput at time t is

$$R(t) = \sum_{u \in \mathcal{U}} r_u(t). \quad (3)$$

Denote the optimization variable $\Pi_b(t)$ as the set of sectors covered by mSBS b at time t, and the beam management policy for the whole system at time t is denoted by $\Pi(t) = \left[\Pi_1(t), \Pi_2(t), ..., \Pi_{|\mathcal{B}|}(t)\right]$. Taking a suitable policy can let more sectors be covered by mSBSs, and thus improve the system throughput. To this end, we formulate the beam management problem as follows with the objective of maximizing the long-term system throughput.

$$\mathbf{P1}: \max_{\Pi(t)} \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t} R(t)\right] \tag{4}$$

$$\text{s.t. } 0 \le |\Pi_b(t)| \le M_b, \forall t \in T, \tag{4-1}$$

$$SINR_{u,b}(t) \ge \chi,$$

$$\forall (u,b) \in \{(u,b)|x_{u,b}(t)=1\}, t \in T, \tag{4-2}$$

$$r_u(t) \ge \hat{r}_u, \forall u \in U, t \in T, \tag{4-3}$$

$$\sum_{b\in B} x_{u,b}(t) \le B_u^{\max}, \forall u \in U, t \in T, \tag{4-4}$$

$$x_{u,b}(t) = \{0,1\}, \forall u \in U, b \in B, t \in T \tag{4-5}$$

where $\mathbb{E}[\cdot]$ is the expectation of the variable, T with cardinality T is the set of time slot for adjusting beam management policy, χ is the SINR threshold that users can correctly receive and decode the information, and $\hat{r}_u$ is the minimum requirement on data rate of user u. In problem **P1**, Constraint (4-1) ensures that the maximum number of beams for mSBS b is M_b. Constraints (4-2) and (4-3) guarantee that the SINR of the link between users and the serving mSBSs should be greater than the threshold χ and the achieved transmission rate of user u needs to exceed the minimum requirement $\hat{r}_u$. (4-4) and (4-5) are the constraints on user association, where the number of associated mSBSs for user u cannot exceed the access capability $B_u^{\max}$.

Examining problem **P1** we realize that the problem is hard to solve by using traditional optimization method. The rational behind is that the long-term optimization objective with unknown user movement behavior is formulated. Thus, the network environment (including user locations, channel quality, network resources, etc.) of future time slot cannot be obtained or even mathematically modeled at the beginning. An efficient and promising way to solve **P1** is to resort to ML algorithms. As mentioned above that the raw data in term of user locations is quite private and should be carefully protected rather than being exchanged among multiple mSBSs like that in most reinforcement learning algorithms, FL, is next adopted to derive the optimal beam management policy of **P1**.

III. FL-based Beam Management in UDmmN

In this section, we propose a novel beam management mechanism for mSBS beam configuration based on FL in UDmmN, called BMFL, with the aim to maximize the long-term throughput while enforcing the protection of user location privacy.

A. Markov Decision Process Model for UDmmN

We formulate the beam management problem as a markov decision process (MDP) model, where a specific mSBS b ($b \in B$) makes a decision (action) on beam directions at each time slot to maximize long-term throughput and the network state may be changed by these sequential actions. We define the state, action and reward as follows.

State: Current operating beam sectors, serving users of mSBSs and the available bandwidth of the MBS are used to describe the system state. Specifically, S_t is the set of all network states for all BSs at time t. For a specific mSBS b, the state is $s_t^b = \{U_b, \Pi_b(t), W_b^{mbs}, S_t^K\} \in S_t$, where U_b represents the set of serving users and $\Pi_b(t)$ represents the corresponding beam sectors occupied by these users, W_b^{mbs} represents the available bandwidth of MBS for users in U_b. Moreover, $S_t^K = \{\Pi_k(t)\}_{k=1,2,\dots,|B|,k\neq b}$ represents the available sectors of all mSBSs except for mSBS b at time t.

Action: Let A_t be the set of actions for all mSBSs at time t. Note that an mSBS is an agent which trains local model independently. For a specific mSBS b, let $a_t^b = \{\Pi_b(t)\} \in A_t$ be the action, which means that mSBS b serves users in U_b with covered beams in $\Pi_b(t)$ at time t.

Reward: In order to maximize the long-term system throughput, we define the reward as $R_t = R(t)$, where $R(t)$ is the optimization objective of **P1**.

In the MDP for beam management, a large number of mSBSs and users result in a large state space and action space. Specifically, the state space for beam management is a discrete space with $M_b \cdot |B| \cdot |U|$ dimensions and action space is a discrete space with $M_b \cdot |B| \cdot B_u(t)$ dimensions. Moreover, as the state and action of a specific mSBS will affect that of others, it is unrealistic to obtain the transition probability. Combined the above issues as well as the protection of user location privacy, we exploit FL based on deep reinforcement learning (DRL).

B. FL-based Beam Management in UDmmN

As shown in Fig. 1, the proposed beam management scheme BMFL consists of two steps, *i.e.*, data cleaning and model training (including local model updating, local model training, global model aggregating).

Data Cleaning: Indeed, a specific mSBS can obtain all location information of its serving users. However, it is unrealistic for mSBSs to choose all the users in their coverage range to participate in local training as: 1) The computing resources occupied for training on a specific mSBS may be inadequate. 2) The location of some certain users needs to be protected. Therefore, to solve the issue mentioned-above while increasing sample diversity as much as possible, mSBS will clean data according to: 1) The coverage of mSBS. The users that are not located in the coverage range will not be chosen to participate in local training. 2) The frequency of participating training. If some users have not participated in local model training for a long time, mSBS will choose them as the participants in the training for next global model updating and thus to increasing the sample diversity. Therefore, the users will be chosen to train local model on the mSBS once both condition (5) and (6) are met.

$$d_{u,b} \le \rho_b, \tag{5}$$

$$\frac{n_u}{N_{total}} \le \eta, \tag{6}$$

where ρ_b and η are the coverage radius threshold of mSBS b

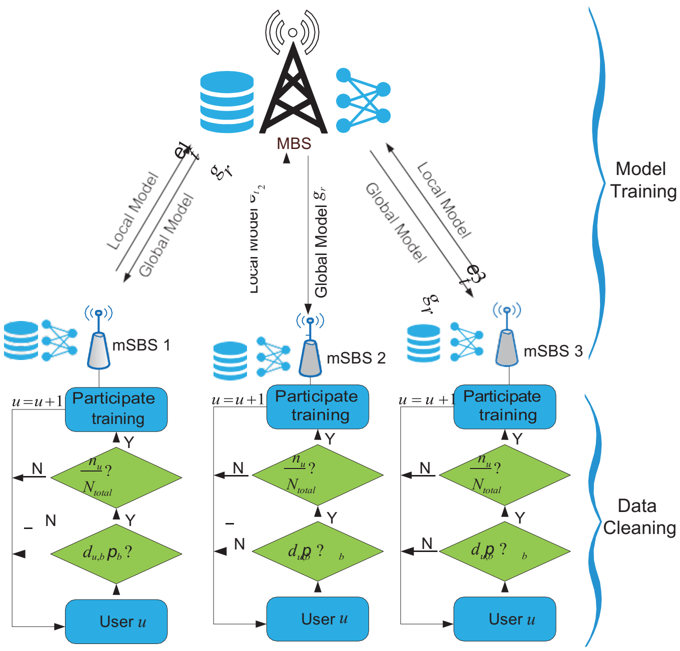

Fig. 1: Beam management based on FL (BMFL) in UDmmN.

and the frequency threshold of participating training respectively. For a specific user u, $\frac{n_u}{N_{total}}$ represents the frequency of participating training, where n_u is the number of participating training rounds of user u and N_{total} is the total training rounds of the relevant mSBS so far. Here we assume U^*_b ($U^*_b \in U_b$) represents the set of the users that participate in local model training on mSBS b.

Model Training: Once finishing the data cleaning, mSBS begins to train local model including local model updating, local model training, and global model aggregating, which are shown as follows.

1) Local model updating: We assume that each communication round consists of τ time slots. We denote the local beam management model on mSBS b at time t and the global model at communication round r by θ^b_t and g_r respectively. During each time slot, each mSBS performs local training once. At the begin of communication round r, mSBSs will receive global model g_r from the MBS to update θ^b_t according to

$$\theta^b_{t+1} = g_r - \frac{\lambda}{K_b}\sum_{b=1}^{|B|}\nabla L(\theta^b_t), 1 \le t \le T, \tag{7}$$

where λ is the learning rate, K_b is the total amount of training data of mSBS b, and $L(\theta^b_t)$ is the loss function which will be given in *Local model training*.

2) Local model training: For a specific mSBS, once all training data is cleaned and the local model is updated, the mSBS begins to train local beam management model based on the location information of participants within its coverage range. As mentioned, a large number of mSBSs and users result in large state space and action space. Therefore, during each communication round, we employ the discrete-action DRL algorithm, DDQN, to train the local beam management model on individual mSBSs. DDQN can tackle the issue of large state/action space by introducing the experience pool and decoupling the selection from the evaluation to reduce the correlation among data. DDQN evaluates the greedy policy according to the Q-network with weight θ and estimates state-action value Q according to the target network $\hat{Q}$ with weight $\hat{\theta}$.The update in DDQN is the same as that in deep Q-Network, but the target is replaced by

$$y^b_t = R_{t+1} + \gamma Q(s^b_{t+1}, \arg\max_{a^b_t} Q(s^b_{t+1}, a^b_t; \theta^b_t); \hat{\theta}^b_t), \tag{8}$$

where

$$\pi = \arg\max_{a^b_t} Q(s^b_{t+1}, a^b_t; \theta^b_t), \tag{9}$$

is an ϵ-greedy policy used to manage beam sectors, θ^b_t and $\hat{\theta}^b_t$ are the weight vectors of Q-network and $\hat{Q}$-network for mSBS b respectively, and $\gamma \in [0, 1]$ is the discount factor representing the discounted impact of future reward.

For a specific mSBS b, if it is in state s^b_t with action a^b_t at time slot t, we will get the corresponding state-action value, which is given by

$$Q(s^b_t, a^b_t) = E[\sum_{k=t} \gamma^k R_t | s^b_t, a^b_t], \tag{10}$$

The objective of DDQN is to minimize the gap between Q and $\hat{Q}$, *i.e.*, loss function. Therefore, DDQN running on each mSBS can be trained by minimizing the loss function, which is given by

$$L(\theta^b_t) = E[(y^b_t - Q(s^b_t, a^b_t; \theta^b_t))^2], \tag{11}$$

Moreover, when DDQN is used to approximate the value function using the neural network, gradient descent method is employed to update the parameter value θ^b_t. Therefore, the update scheme in DDQN is given by

$$\theta^b_{t+1} = \theta^b_t + \alpha\left[y^b_t - Q\left(s^b_t, a^b_t; \theta^b_t\right)\right]\nabla Q\left(s^b_t, a^b_t; \theta^b_t\right), \tag{12}$$

where α is the step size.

After training local data for τ time slots, mSBSs will send training parameters θ^b_t ($b \in B$) to the MBS to update the global model.

3) Global model aggregating: Once receiving all local models (*i.e.*, θ^b_t for $\forall b \in B$) at the end of communication round r, the MBS updates the global model by

$$g_r = \frac{\sum_{b \in B} K^b \theta^b_t}{K}, t = \tau, \tag{13}$$

where $K = \sum_{b \in B} K_b$ is the total amount of training data. After updating the global model g_r, the MBS will broadcast the global mode g_r to all mSBSs to update their local models.

The BMFL algorithm for beam management is presented as **Algorithm 1**.

IV. Simulation and Discussions

In this section, we conduct numerical simulations to first evaluate the convergence of the proposed BMFL algorithm, and

Algorithm 1 BMFL Algorithm for Beam Management

Input: $U, B, U_b^* = \varnothing, s^b, a^b, \eta, \rho_b, \gamma, C, K_b, \lambda, \tau$

output: Beam sectors π^b.

```
Initialize experience relay pool D_b, ∀b ∈ B;
Initialize the global weights g_0;
for communication round r = 1, 2, ... do
    ↓Data Cleaning
    for b = 1, 2, ..., |B| do
        for u = 1, 2, ..., |U| do
            if d_{u,b} ≤ ρ_b and n_u / N_total ≤ η then
                U_b^* = {U_b^*, u};
                u = u + 1;
            else
                U_b^* = {U_b^*};
                u = u + 1;
            end if
        end for
    end for
    Collect data from U_b^* for mSBS b, ∀b ∈ B;
    ↓Update Local Model
    if r == 1 then
        Initialize θ_0^b, ∀b ∈ B;
    else
        θ_0^b = g_{r−1} − (λ / K_b) Σ_{b=1}^{|B|} ∇L(θ_t^b), ∀b ∈ B.
    end if
    ↓Train Local model
    Let θ̂_0^b = θ_0^b, initialize target action-value function Q̂(·) according to θ̂_0^b;
    for t = 1 to τ do
        Receive the initial state s_t^1, s_t^2, ..., s_t^{|B|};
        Select a_t^b = argmax_a Q() using ε-greedy policy;
        Execute action a_t^b;
        Obtain R_t^b and s_{t+1}^b;
        Store (s_t^b, a_t^b, R_t^b, s_{t+1}^b) into D_b, ∀b ∈ B;
        Randomly select a sample (s_j^b, a_j^b, R_j^b, s_{j+1}^b) from D_b, 1 ≤ j ≤ t, ∀b ∈ B;
        Calculate y_t^b according to equation (8);
        Perform a gradient descent step on (y_t^b − Q(s_t^b, a_t^b; θ_t^b))^2;
        Update the parameter θ_t^b according to equation (12);
        Every C steps, reset Q̂ = Q;
    end for
    ↓Update global model
    g_r = Σ_{b=1}^{|B|} K_b θ_τ^b / K
end for
Obtain beam sectors π^b.
```

then compare the performance in term of network throughput with the following two beam management algorithms.

1) Brute-Force Search (BFS): Find the optimal beam coverage by searching all the possible beam sectors. This algorithm can reach the optimal solution of beam management with extremely high computational complexity.

2) Evenly Deployed Beam (EDB): Deploy the beams in a uniform manner. In this algorithm, we only need to optimize the direction of one beam for each mSBS, and the direction of the other beams can thus be determined as the rule of uniform deployment.

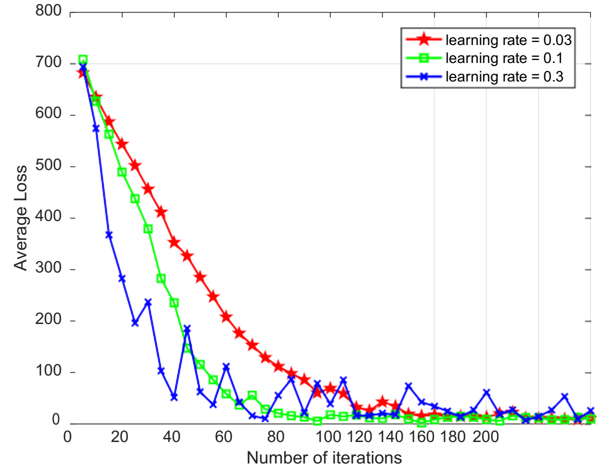

Fig. 2: Convergence of the BMFL in UDmmN.

We consider a square area with the size $100m \times 100m$, where multiple mSBSs and users are randomly distributed, and an MBS is located at the center. For the coverage of each mSBS, we uniformly divide it into 8 sectors (i.e., each sector is with the coverage of 45°). Each mSBS generates three beams covering different sectors. We set the transmit frequency as 28 GHz with available bandwidth 2 GHz, and the transmit power of mSBS is set to 37 dBm. For the MBS, the settings are given as that the transmit frequency is 2.1 GHz, the available bandwidth is 100 MHz, and the transmit power is 50 dBm. Each user can be associated with up to 3 mSBSs. The SNR threshold is set to -20dB.

In the first experiment, we evaluate the convergence of the proposed BMFL algorithm under three typical learning rates 0.03, 0.1 and 0.3 as shown in Fig. 2. From this figure, we find that all the three curves reach the convergence after around 200 iterations. Specifically, the BMFL algorithm reaches the convergence after around 80 iterations when learning rate is 0.1 while around 130 iterations of learning rate 0.03 and nearly 200 iterations for learning rate 0.3. These convergence results clearly demonstrate the effectiveness and rationality of BMFL.

Next, we compare network throughput of BMFL with the varying density of user under three different mSBS densities, shown in Fig. 3. As expected, we find that the network throughput has little difference for all the three mSBS densities when the user density is lower than 900 per km^2. Moreover, with the increase of user density the achieved network throughput is the highest under the mSBS density of 900 per km^2 which is because of the abundant beam resource.

Finally, we compare the network throughput of BMFL with BFS and EDB, which are defined at the beginning of this section. Fig. 4 shows the network throughput of the three beam management algorithms under the varying user density from 500 per km^2 to 3000 per km^2. We fix the mSBS density as 600 per km^2 in this simulation. As expected that the BFS beam management algorithm achieves the highest throughput as all the potential solutions have been searched and tested. Importantly, we find that BMFL achieves the second highest network throughput with relatively small difference of that in BFS but much higher than that of EDB. These results further demonstrate the performance gain of the proposed BMFL algorithm.

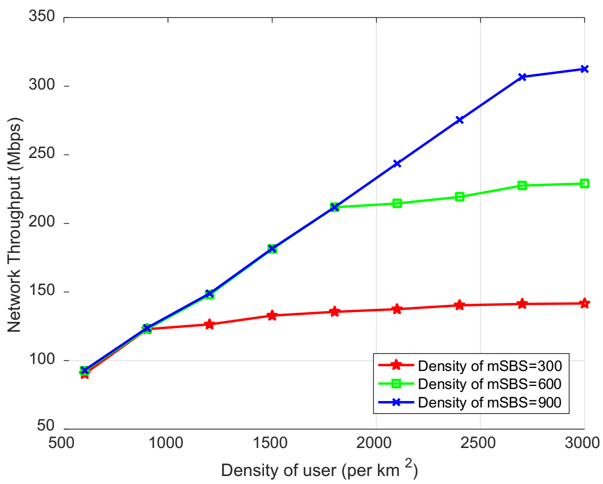

Fig. 3: Comparisons of network throughput of BMFL under different mSBS densities.

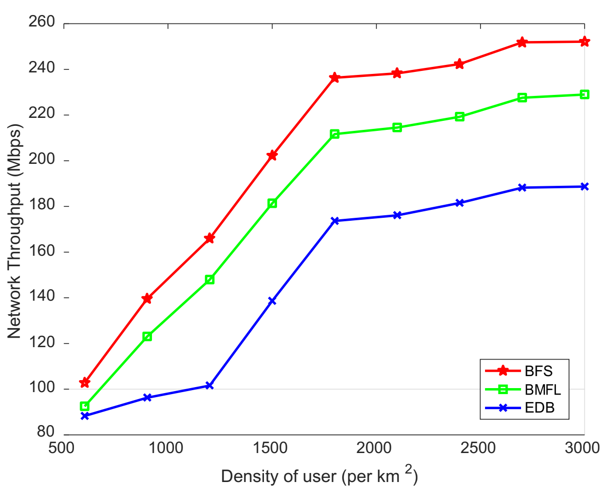

Fig. 4: Comparisons of network throughput for the three beam management algorithms.

V. Conclusions

Due to the directional transmission and dense network deployment, complexity of beam management problem in mmWave communication systems becomes a real challenge. To address the complex and dynamic control issues, in this paper we have proposed a federated DRL-based adaptive beam management mechanism, (i.e., BMFL). We formulated the mSBS beam configuration problem as a MDP model and introduced DDQN into FL to cope with large state-action space issues while enhancing the protection of user privacy. Simulation results have shown that the BMFL provides a better tradeoff between computational complexity and network throughput.

References

[1] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[2] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. Santos Filho, and G. Fettweis, "How reliable and capable is multi-connectivity?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1506–1520, 2019.

[3] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 1, pp. 173–196, 2019.

[4] E. Onggosanusi, M. S. Rahman, L. Guo, Y. Kwak, H. Noh, Y. Kim, S. Faxer, M. Harrison, M. Frenne, S. Grant, R. Chen, R. Tamrakar, and a. Q. Gao, "Modular and high-resolution channel state information and beam management for 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 48–55, 2018.

[5] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Standalone and non-standalone beam management for 3GPP NR at mmWaves," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 123–129, 2019.

[6] Y. R. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5G and beyond," *IEEE Access*, vol. 8, pp. 13 282–13 293, 2020.

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

[8] 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.

[9] D. C. Arajo and A. L. F. de Almeida, "Beam management solution using Q-learning framework," in *IEEE 8th International Workshop on CAMSAP*, 2019, pp. 594–598.

[10] Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S. P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Trans. Mobile Comput*, vol. 17, no. 6, pp. 1456–1468, 2017.

[11] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5519–5533, 2020.

[12] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.

[13] P. Zhou, X. Fang, X. Wang, Y. Long, R. He, and X. Han, "Deep learning-based beam management and interference coordination in dense mmWave networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 592–603, 2019.

[14] S. Moon, H. Kim, and I. Hwang, "Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications," *Journal of Commun. and Networks*, vol. 22, no. 3, pp. 177–184, 2020.

[15] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, 2020.

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