DRL-Based AP Switch On/Off Scheme for Cell-Free Massive MIMO MEC Networks

Hyunjoon Suh, Jintaek Oh, Shinhyeok Kang and Taewon Hwang
School of Electrical and Electronic Engineering
Yonsei University
Seoul, Republic of Korea
Emails: hyunjoon.suh@gmail.com, jtoh55@yonsei.ac.kr, kaci0526@gmail.com, twhwang@yonsei.ac.kr

***Abstract*—In this paper, we consider a cell-free massive multiple-intput multiple-output (MIMO)-enabled mobile edge computing (MEC) network where AP switch on/off method is adopted for energy saving. We formulate a system-wise energy minimization problem, which jointly optimizes on/off mode of APs, uplink transmit power, and offloading ratio. The problem is non-convex and thus hard to solve using optimization methods. To solve the problem, we first reformulate it as a Markov decision process (MDP) and then propose a deep reinforcement learning (DRL)-based scheme. Simulation results show that our proposed scheme can reduce the energy consumption significantly.**

***Index Terms*—Mobile edge computing, cell-free massive MIMO, AP switch on/off, deep reinforcement learning**

I. Introduction

Recently, mobile users' demands for computation-intensive and latency-sensitive applications are increasing explosively. However, since user devices have capacity-limited battery and low computing power, it is challenging for them to compute the tasks efficiently. Mobile edge computing (MEC) [1], which enables users to offload their tasks to nearby edge servers with powerful computing capability, has emerged as a promising technology to address this problem. Enhancing uplink performance is essential for MEC since it directly affects offloading delay and the resulting total delay of computation task.

Cell-free massive multiple-input multiple-output (MIMO) [2], [3], where a large number of access points (APs) with a few antennas cooperatively serve a much smaller number of users simultaneously with the help of a central processing unit (CPU), is a promising technology to improve the coverage and throughput. Through the cooperation, the APs can act as a co-located massive MIMO BS and exploit the beamforming gain to increase the data rate. Also, owing to the densification of APs, the concept of "cell" disappears and all users can be provided with uniformly good communication service.

Based on these advantages, combining cell-free massive MIMO with MEC is expected to increase the uplink data rate for offloading and thus enhance the overall performance of MEC networks. Motivated by this, the study in [4] considered a cell-free massive MIMO MEC network and proposed a DRL-based resource allocation scheme to minimize the total energy consumption of users. The dense deployment of APs in cell-free massive MIMO raises concerns about huge energy consumption. To cope with this problem, AP switch on/off method, which turns off some APs to reduce the energy consumption, has been considered as an energy-efficient strategies for cell-free massive MIMO networks [5], [6].

In this paper, we consider a cell-free massive MIMO MEC network adopting AP switch on/off and propose a DRL-based joint AP switch on/off and resource allocation scheme. Specifically, we formulate an optimization problem of jointly controlling on/off mode of APs, uplink transmit power and offloading ratio to minimize the system-wise total energy consumption under the users' delay constraints. Then, to tackle the hardness of the problem and uncertainty of time-varying environment, we propose a deep deterministic policy gradient (DDPG)-based scheme to solve the problem. Different from the existing works, our work considers combining cell-free massive MIMO network adopting AP switch on/off with MEC network, and proposes a DRL-based AP switch on/off scheme.

II. System model and problem formulation

We consider a cell-free massive MIMO-enabled MEC network with M APs, each equipped with N antennas, and K single antenna users, such that $M \gg K$. All APs are connected to a CPU via fronthaul links. The CPU has an MEC server with computing capacity F [cycles/s]. Each AP can be either on (in active mode) or off (in inactive mode) for energy saving. Let $\alpha_m \in \{0,1\}$ denote the mode of AP m, where $\alpha_m = 1$ indicates AP m is on, otherwise $\alpha_m = 0$.

Assume that each user k has a computation task whose size is B_k [bits] and required number of cycle is C_k. Each user k locally computes a part of its task with computing capacity f_k^{loc} and offloads the remaining part to the MEC server with uplink transmit power $p_k \in [0, p_k^{\max}]$. Let $\theta_k \in [0,1]$ denote the offloading ratio of user k defined as the ratio of the amount of offloaded task to the amount of total task of user k.

A. Uplink rate

Channel between AP m and user k is given by $\mathbf{g}_{mk} = \sqrt{\beta_{mk}}\mathbf{h}_{mk} \in \mathbb{C}^{N\times 1}$, where β_{mk} is large-scale channel coefficient and $\mathbf{h}_{mk}$ is small-scale channel vector assumed to be i.i.d Rayleigh fading. According to the channel estimation in [3],

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00253249) and Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0020535, The Competency Development Program for Industry Specialist).

$\mathbf{g}_{mk}$ is estimated as $\hat{\mathbf{g}}_{mk}$ by AP m. After channel estimation phase, user k deciding to offload transmits the data s_k to the all APs. Then, the received signal at active AP m is given by $\mathbf{y}_m = \sum_{l=1}^{K} \mathbf{g}_{ml}\sqrt{p_l}s_l + \mathbf{n}_m$, where $\mathbf{n}_m$ is noise at AP m.

Maximum ratio combining is used to detect s_k. Since only active APs (with $\alpha_m = 1$) calculate and send $\hat{\mathbf{g}}_{mk}^H\mathbf{y}_m$'s to the CPU, the CPU observes $r_k = \sum_{m=1}^{M} \alpha_m \hat{\mathbf{g}}_{mk}^H \mathbf{y}_m$, which can be decomposed into desired signal (DS_k), beamforming uncertainty (BU_k), inter-user interference ($\text{IUI}_{k,l}$), and noise:

$$\text{DS}_k = \mathbb{E}\left\{\sum_{m=1}^{M} a_m \hat{\mathbf{g}}_{mk}^H \mathbf{g}_{mk}\sqrt{p_k}\right\}$$

$$\text{BU}_k = \sum_{m=1}^{M} a_m \hat{\mathbf{g}}_{mk}^H \mathbf{g}_{mk}\sqrt{p_k} - \mathbb{E}\left\{\sum_{m=1}^{M} a_m \hat{\mathbf{g}}_{mk}^H \mathbf{g}_{mk}\sqrt{p_k}\right\}$$

$$\text{IUI}_{k,l} = \sum_{m=1}^{M} a_m \hat{\mathbf{g}}_{ml}^H \mathbf{g}_{ml}\sqrt{p_l}.$$

From above, the uplink SINR of user k is

$$\gamma_k = \frac{|\text{DS}_k|^2}{\mathbb{E}\left\{|\text{BU}_k|^2\right\} + \sum_{l\neq k}\mathbb{E}\left\{|\text{IUI}_{k,l}|^2\right\} + \sum_{m=1}^{M}\alpha_m \mathbb{E}\left\{\|\hat{\mathbf{g}}_{mk}^H\|^2\right\}}.$$

Then, uplink rate of user k is given by $R_k = W\log_2(1+\gamma_k)$, where W is bandwidth. Details of this subsection are in [3].

B. Computation and energy consumption model

Given offloading ratio $\theta_k \in [0,1]$, $(1-\theta_k)$ of user k's task is locally computed and θ_k of user k's task is offloaded. Then, local computing delay and energy of user k are given by $D_k^{\text{loc}} = \frac{(1-\theta_k)C_k}{f_k^{\text{loc}}}$ and $E_k^{\text{loc}} = \kappa(1-\theta_k)C_k\{f_k^{\text{loc}}\}^2 + p_k^{l,c}D_k^{\text{loc}}$, respectively, where κ is effective capacitance coefficient and $p_k^{l,c}$ is the power consumed by leakage currents.

Let F_k denote the MEC server's computing capacity allocated to user k's task, which is proportional to offloaded bits. Assuming the computing result size is much smaller than the input size, we ignore the result download delay. Then, offloading delay and energy of user k are given by $D_k^{\text{off}} = \frac{\theta_k B_k}{R_k} + \frac{\theta_k C_k}{F_k}$ and $E_k^{\text{off}} = (p_k + p_k^{t,c})D_k^{\text{off}}$, respectively, where $p_k^{t,c}$ is the circuit power consumed by user k.

Assuming parallel operation of local computing and offloading, the total delay of user k is given by $D_k^{\text{tot}} = \max\left\{D_k^{\text{loc}}, D_k^{\text{off}}\right\}$ and energy of user k are given by

$$E_k^{\text{tot}} = E_k^{\text{loc}} + E_k^{\text{off}}. \tag{1}$$

As in [5], power consumption model of AP m is given by

$$P_m = \begin{cases} (\xi_m^{\text{FH}} + \xi_m^{\text{AP}})\sum_{k=1}^{K} R_k + P_{m,\text{ON}}^{\text{FH,fix}} + P_{m,\text{ON}}^{\text{AP,fix}} & \alpha_m = 1 \\ P_{m,\text{OFF}}^{\text{FH,fix}} + P_{m,\text{OFF}}^{\text{AP,fix}} & \alpha_m = 0 \end{cases}, \tag{2}$$

where ξ_m^{FH} and ξ_m^{AP} are the traffic-dependent power coefficients for fronthaul and AP m, respectively, $P_{m,l}^{\text{FH,fix}}$ is the traffic-independent power consumed by fronthaul when in mode $l \in \{\text{ON}, \text{OFF}\}$, and $P_{m,l}^{\text{AP,fix}}$ is the traffic-independent power consumed by AP m including RF chain power consumption when in mode l. Given AP operation time $T^{\text{AP}} \triangleq \max_k D_k^{\text{off}}$, total energy consumption of all APs is expressed as

$$E^{\text{AP}} = T^{\text{AP}} \cdot \sum_{m=1}^{M} P_m. \tag{3}$$

Finally, system-wise total energy consumption is given by

$$E(\{\alpha_m\}_{m=1}^{M}, \{p_k, \theta_k\}_{k=1}^{K}) = \sum_{k=1}^{K} E_k^{\text{tot}} + E^{\text{AP}}. \tag{4}$$

The problem of minimizing system-wise total energy consumption is formulated as

$$\begin{aligned} \underset{\{\alpha_m\}_{m=1}^{M}, \{p_k, \theta_k\}_{k=1}^{K}}{\text{minimize}} \quad & E(\{\alpha_m\}_{m=1}^{M}, \{p_k, \theta_k\}_{k=1}^{K}) \\ \text{subject to} \quad & D_k^{\text{tot}} \leq D_k^{\max}, \ \forall k, \end{aligned} \tag{5}$$

where $D_k^{\max}$ is the delay limit for user k's computation task. The problem (5) is a form of mixed-integer programming, which is non-convex, and therefore is hard to solve using optimization methods. To tackle this problem, we propose a DRL-based solution in the next section.

III. DRL-based joint AP switch on/off and resource allocation scheme

In this section, we reformulate (5) as a Markov-decision process (MDP) problem and propose a DDPG-based solution.

A. MDP formulation

1) State: The state at time slot t is defined as $s(t) \triangleq \{\mathbf{B}(t), \mathbf{C}(t), \mathbf{R}(t-1)\}$, where $\mathbf{B}(t) = [B_1(t), \cdots, B_K(t)]$ is task size at time slot t, $\mathbf{C}(t) = [C_1(t), \cdots, C_K(t)]$ is required cycles at time slot t, and $\mathbf{R}(t-1) = [R_1(t-1), \cdots, R_K(t-1)]$ is uplink rate of users at time slot $(t-1)$.

2) Action: The action at time slot t is defined as $\mathbf{a}(t) \triangleq \{\boldsymbol{\alpha}(t), \mathbf{p}(t), \boldsymbol{\theta}(t)\} \in \mathbb{R}^{M+2K}$, where $\boldsymbol{\alpha}(t) = [\alpha_1(t), \cdots, \alpha_M(t)]$ is on/off mode, $\mathbf{p}(t) = [p_1(t), \cdots, p_K(t)]$ is uplink transmit power, and $\boldsymbol{\theta}(t) = [\theta_1(t), \cdots, \theta_K(t)]$ is offloading ratio at time slot t.

3) Reward: To minimize total energy consumption $E(t)$, we consider a decreasing function of $E(t)$ as a reward. Also, for the delay constraints, we introduce the penalty C_{pen} which is proportional to the number of users not satisfying the delay constraint. As a result, we define the reward at time slot t as

$$r(t) \triangleq \frac{\sum_{k=1}^{K} B_k(t) \text{ [kbits]}}{E(t)} - C_{\text{pen}}. \tag{6}$$

B. DDPG-based solution

In DDPG [7], the critic network $Q(s, a|\boldsymbol{\varpi}^Q)$ approximates the state-value function of the MDP and the actor network $\mu(s|\boldsymbol{\varpi}^\mu)$ approximates the best action given state s. In our scheme, we adopt the sigmoid function whose value is in $[0,1]$ as the activation function of the output layer of the actor network. Let $\mathbf{o}(t) = [o_1(t), \cdots, o_{M+2K}(t)] \in \mathbb{R}^{M+2K}$ denote the output of actor network. To generate the discrete on/off actions $\alpha_m(t)$'s of the APs, we discretize $o_m(t)$, $m = 1, \cdots, M$,

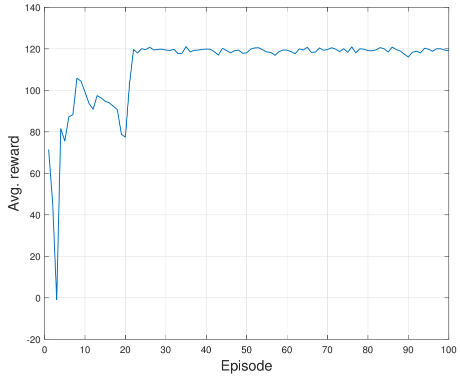

Fig. 1. Average reward versus episode index (M = 120 APs and K=12 users)

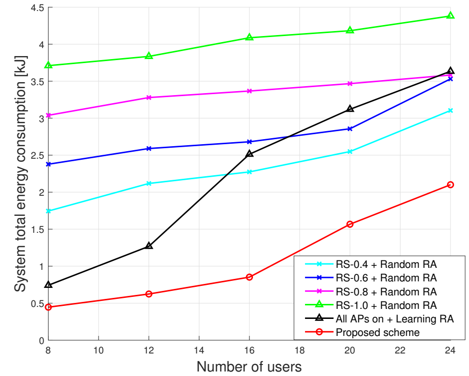

Fig. 2. System-wise total energy consumption versus the number of users (M = 120 APs)

into $\alpha_m(t)$'s using a threshold value of 0.5. Uplink transmit power and offloading ratio are given by $p_k(t) = p_k^{\max} o_{M+k}(t)$ and $\theta_k(t) = o_{M+K+k}(t), k = 1, \cdots, K$, respectively.

IV. Simulation Results

In our simulation, we consider a square area of 1km $\times$ 1km. We assume that $M = 120$ APs with $N = 4$ antennas are deployed in 12×10 grid pattern in the area and K users are uniformly distributed at random. We employ the model in [2] for β_{mk}. MEC server's computing capacity is set to $F = 25$ GHz. Maximum uplink transmit power and computing capacity of the users are set to $p_k^{\max} = 0.5$ W and $f_k^{\text{loc}} = 0.8$ GHz, $\forall k$, respectively. Task size is sampled by a uniform distribution $B_k \sim U[4, 8]$ Mbits, and the required cycle is $C_k = 10^3 B_k$. We use the same parameters for AP power consumption in (2) as [5]. In the DDPG algorithm, both actor and critic networks are fully-connected DNNs with 4 hidden layers of 128 neurons and use ReLU for activation function.

For comparison, we consider following benchmark schemes:

- All APs ON + Learning resource allocation (RA): All APs are always on, i.e., $\alpha_m = 1, \forall m$, and resource allocation $[\mathbf{p}(t), \boldsymbol{\theta}(t)]$ is learned by using DDPG.
- Random switch (RS)-p_{on}+Random RA: APs are randomly switched on with probability p_{on} and $[\mathbf{p}(t), \boldsymbol{\theta}(t)]$ is randomly chosen.

Fig. 1 shows the average reward of episodes in training. Each episode consists of 250 time steps. We can see that average reward is low and fluctuating in the early stages. But, as the training progresses, it converges to a steady value.

Fig. 2 shows the system-wise total energy consumption performance of the various schemes versus the number of users K. We can observe that the proposed scheme obviously outperforms benchmark schemes. This directly shows that by smartly choosing which APs to turn on and allocating resources, the proposed scheme can reduce the energy consumption significantly. As we can see from the performance gap between the proposed scheme and "All APs on + Learning RA" scheme, even though the resource allocation is good, turning all APs on is not an energy-efficient choice. Besides, as we can see from the performance gap between "RS-1.0 + Random RA" scheme and "All APs on + Learning RA" scheme, the resource allocation is also very crucial for reducing energy consumption. In addition, it can be seen that total energy consumption increases as the number of users increases in all schemes. One simple and obvious reason for this is that the number of energy consumer has increased. Another reason is an increase in offloading delay due to an increase in inter-user interference and a decrease in the allocated computing capacity per user from the MEC server.

V. Conclusion

In this paper, we have proposed a joint AP switch on/off and resource allocation scheme for energy-efficient cell-free massive MIMO MEC networks. We have formulated a system-wise energy minimization problem, which is challenging to solve using optimization methods. Then, we have proposed a DDPG-based scheme to solve the problem. By the simulation results, we have demonstrated that the proposed scheme can significantly reduce the system-wise total energy consumption of cell-free massive MIMO MEC networks.

References

[1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322-2358, 4th Quart., 2017.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[3] M. Rahmani et al., "Deep reinforcement learning-based sum rate fairness trade-off for cell-free mMIMO," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6039-6055, May 2023.

[4] F. D. Tilahun, A. T. Abebe and C. G. Kang, "DRL-based distributed resource allocation for edge computing in cell-free massive MIMO network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Rio de Janeiro, Brazil, 2022, pp. 3845-3850.

[5] G. Femenias, N. Lassoued, and F. Riera-Palou, "Access point switch ON/OFF strategies for green cell-free Massive MIMO networking," *IEEE Access*, vol. 8, pp. 21788–21803, 2020.

[6] M. Ito et al., "Joint AP on/off and user-centric clustering for energy-efficient cell-free massive MIMO systems," *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, London, United Kingdom, 2022, pp. 1-5.

[7] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, arXiv:1509.02971.