# Deep Reinforcement Learning-Based Sum rate Maximization in Tethered UAV-Aided IAB Network

Yerin Lee
*Electronic & Electrical Eng.*
*Hankyong National Univ.*
Anseong, South Korea
qkek0379@hknu.ac.kr

Heejung Yu
*Electronic & Information Eng.*
*Korea Univ.*
Sejong, South Korea
heejungyu@korea.ac.kr

Howon Lee
*Electronic & Electrical Eng.*
*Hankyong National Univ.*
Anseong, South Korea
hwlee@hknu.ac.kr

**Abstract**—Tethered unmanned aerial vehicle base station (TUBS), which receives constant power from the ground, is a promising technology to solve the problem of UBS performance degradation and is suitable for use as base stations. In addition, the concept of integrated access and backhaul (IAB) has been introduced for efficient frequency resource utilization. Herein, we propose a double deep Q-network-based TUBS positioning and resource allocation to maximize the sum-rate in TUBS-aided IAB networks.

*Keywords—TUAV, IAB, DDQN, optimal resource allocation, optimal TUAV control*

## Ⅰ. Introduction

Tethered unmanned aerial vehicle base station (TUBS), which receives constant power from the ground, is a promising technology to solve the problem of UBS performance degradation and is suitable for use as a base station [1]-[3]. In addition, the concept of integrated access and backhaul (IAB) has been introduced and standardized for efficient frequency resource utilization [4]. Herein, we propose a double deep Q-network (DDQN)-based TUBS positioning and resource allocation to maximize the sum rate in an aerial TUBS-aided IAB network.

## Ⅱ. Proposed DDQN-based Joint Optimization Algorithm

We consider a two-tier IAB network consisting of MBS, TUBSs, and ground users. The MDP of the proposed algorithm is defined as shown in Table 1.

| Agent | MBS | TUBS |
|---|---|---|
| State | $\boldsymbol{B^{A_1}_{(i,k)}(\tau), T\mathrm{x}_i(\tau)}$ | $\boldsymbol{B^{A_2}_{(j,k')}(\tau), T\mathrm{x}_j(\tau), D_j(\tau)}$ |
| Action | $\boldsymbol{B^{A_1}_{(i,k)}, \pm\Delta_{T\mathrm{x}_i}, \Delta_{T\mathrm{x}_i} = 0}$ | $\boldsymbol{B^{A_2}_{(j,k')}, \pm\Delta_{T\mathrm{x}_j}, \Delta_{T\mathrm{x}_j} = 0, \pm\Delta_r, \Delta_r = 0, \pm\Delta_\theta, \Delta_\theta = 0, \pm\Delta_\phi, \Delta_\phi = 0}$ |
| Reward | $\boldsymbol{\frac{BW}{N}\{\Sigma_i\Sigma_k\Sigma_c\log_2(1+\Gamma^c_{i,k}) + \Sigma_j\Sigma_{k'}\Sigma_c\log_2(1+\Gamma^c_{j,k'})\}}$ | $*$ $\boldsymbol{BW}$: Total bandwidth<br>$*$ $\boldsymbol{N}$: Number of subchannels |

**Table 1.** Markov Decision Process (MDP) design

where $\boldsymbol{B^{A_1}_{(i,k)}(\tau)}$ , $\boldsymbol{T\mathrm{x}_i(\tau)}$ denote subchannels and transmission power information allocated by the transmitter MBS to the 1st-tier receiver $\boldsymbol{k}$ (user, TUBS), respectively. $\boldsymbol{B^{A_2}_{(j,k')}(\tau), T\mathrm{x}_j(\tau)}$ denote subchannels and transmission power information allocated by the transmitter TUBS $\boldsymbol{j}$ to the 2nd-tier receiver $\boldsymbol{k'}$ (user), respectively, and $\boldsymbol{D_j(\tau) \in \{r_j(\tau), \theta_j(\tau), \phi_j(\tau)\}}$ denote the position of TUBS $\boldsymbol{j}$ in the spherical coordinates. Both the MBS and TUBSs have the common unified goal of maximizing the network-wide sum rate. Accordingly, the common reward can be represented as the sum of MBS and TUBSs. $\boldsymbol{\Gamma}$ denote signal-to-noise ratio (SINR). Assuming received power $\boldsymbol{P}$, subchannel allocation vector element $\boldsymbol{b}$, interference $\boldsymbol{I}$ from another base station, and noise $\boldsymbol{\sigma^2}$, the SINR is calculated as $\boldsymbol{\Gamma = \frac{Pb}{I+\sigma^2}}$.

## Ⅲ. Simulation Results & Conclusion

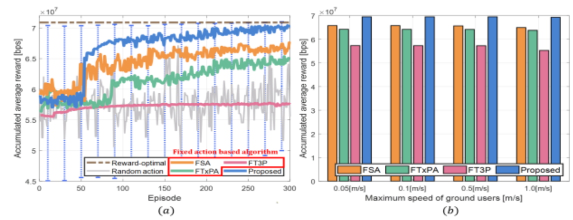


**Fig. 2.** (a) Accumulated average reward vs. episode (b) Test results across 200,000 trials

We consider fixed TUBS 3D positioning (FT3P), fixed transmission power allocation (FTxPA), fixed subchannel allocation (FSA), random action, and reward-optimal as benchmarks to analyze the performance of the proposed algorithm in a three-agent scenario. The $\boldsymbol{BW}$ is 20 MHz and $\boldsymbol{N}$ is 3. As shown in Fig. 2(a), the proposed algorithm converged closely to the optimal and outperformed the other benchmark algorithms. Specifically, the proposed demonstrated gains of 8.97 %, 13.54 %, and 37.2 % compared with the FSA, FTxPA, and FT3P algorithms, respectively. Fig. 2(b), shows that the performances of the algorithms deteriorated slightly as the moving speed increases. Nevertheless, the proposed algorithm consistently outperformed the benchmark algorithms and maintained relatively high performance levels across all moving speeds.

### Acknowledgment

This work was supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00794, Development of 3D Spatial Mobile Communication Technology).

### References

[1] M. Kishk et al., “Aerial base station deployment in 6G cellular networks using tethered drones: The mobility and endurance tradeoff,” IEEE Veh. Tech. Mag., vol. 15, no. 4, pp. 103–111, Dec. 2020.

[2] H. Lee et al., "Towards 6G hyper-connectivity: Vision, challenges, and key enabling technologies," in Journal of Communications and Networks, vol. 25, no. 3, pp. 344-354, Jun. 2023.

[3] S. Lim, H. Yu and H. Lee, "Optimal Tethered-UAV Deployment in A2G Communication Networks: Multi-Agent Q-Learning Approach," in IEEE Internet of Things Journal, vol. 9, no. 19, pp. 18539-18549, 1 Oct.1, 2022.

[4] W. Lei et al., “Deep reinforcement learning-based spectrum allocation in integrated access and backhaul networks,” IEEE Trans. on Cog. Commun. and net., vol. 6, no. 3, pp. 970–979, Sep. 2020.