IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 September 2023, accepted 29 September 2023, date of publication 31 October 2023,
date of current version 3 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3328957

RESEARCH ARTICLE

Low-Complexity Q-Learning for Energy-Aware Small-Cell Networks With Integrated Access and Backhaul

JUNSEUNG LEE[1], (Student Member, IEEE), HYUN-HO CHOI[2], (Senior Member, IEEE), SEUNG-CHAN LIM[2], (Member, IEEE), HYUNGSUB KIM[3], JEEHYEON NA[3], (Member, IEEE), AND HOWON LEE[1], (Senior Member, IEEE)

[1]School of Electronic and Electrical Engineering and Research Center for Hyper-Connected Convergence Technology, Hankyong National University, Anseong 17579, South Korea
[2]School of ICT, Robotic, and Mechanical Engineering and IITC, Hankyong National University, Anseong 17579, South Korea
[3]Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

Corresponding authors: Howon Lee (hwlee@hknu.ac.kr) and Hyun-Ho Choi (hhchoi@hknu.ac.kr)

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by the Korean Government through MSIT (5G Open Intelligence-Defined RAN (ID-RAN) Technique Based on 5G New Radio) under Grant 2018-0-01659.

ABSTRACT An integrated access and backhaul (IAB)-enabled small-cell network commonly utilizes frequency channels for access and backhaul links, and thus this network has a chance to utilize the frequency channels efficiently and optimally. However, there are still several problems with applying the IAB technology to practical small-cell networks, such as extremely high computational complexity caused by shared resource utilization and additional co-tier and cross-tier interference management. Therefore, we herein propose a multi-agent distributed Q-learning with pre-resource partitioning (MADQ-PRP) algorithm to solve the problem of frequency channel allocation and energy consumption. In MADQ-PRP, to reduce the computational complexity, each RL agent only considers its local state information to determine its following action. Nevertheless, by sharing and redistributing the rewards among agents, the overall reward can be maximized. Furthermore, we devise a pre-resource partitioning method depending on the variations in the number of SBSs per MBS and the numbers of MBS and SBS channels to reduce the computational complexity of the proposed MADQ-PRP algorithm. Through intensive simulations, we show the convergence of the proposed MADQ-PRP algorithm to the optimal solution obtained by the exhaustive search algorithm. Also, we demonstrate that the proposed MADQ-PRP algorithm outperforms several benchmark algorithms such as 'Random action,' 'SBS on-off,' 'SBS-only,' and 'MADQ-only' in IAB-enabled small-cell networks with non-uniform traffic distribution. Furthermore, it is confirmed that the proposed MADQ-PRP algorithm can reduce the CPU execution time by 9.1% and 97.9% compared to the distributed and centralized RL algorithms, respectively. The proposed algorithm based on the low-complexity RL and PRP could be one of the solutions to optimize the heterogeneous network performance from the perspective of the network operators when considering the coverage-capacity tradeoff.

INDEX TERMS Low complexity multi-agent Q-learning, pre-resource partitioning, network-wide energy efficiency, user outage, integrated access and backhaul, IAB-enabled small-cell networks.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Mobile data traffic is expected to exponentially increase in the forthcoming sixth generation (6G) cellular networks [1], [2], [3], [4], [5]. However, the limited frequency resources are insufficient to support this traffic demand. Furthermore, access and backhaul networks require a large amount of spectrum bandwidth as well as consume a vast amount of energy to support many kinds of real-time mobile application services. In particular, access technology accounts for a large portion of the total network energy consumption

The associate editor coordinating the review of this manuscript and approving it for publication was Lorenzo Mucchi.

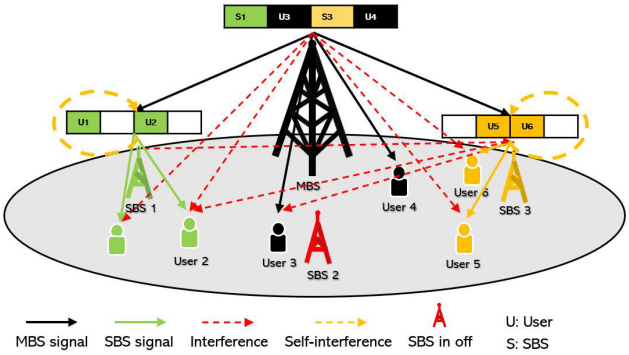


FIGURE 1. System model of IAB-enabled small-cell networks.

[6], [7], [8]. Also, to meet the technical requirements of the maximum data rate, ultra-low latency, and high reliability of 6G, cost-effective network operation and management are very important. Thus, spectrum-efficient and energy-efficient access and backhaul communications are one of the key challenges in 6G to reduce network energy consumption and carbon dioxide emissions [9], [10].

Also, the explosive installations of small cell base stations (SBSs) to support the increasing network traffic are accelerating network densification, and thus it may result in the need for more available frequency resources and an increase in network power consumption [7], [8]. Hence, network operators should efficiently utilize these limited frequency resources to support massive users within the entire network and optimally deploys many SBSs to reduce spectrum holes and network power consumption.

As one of the ways to resolve these problems, the concept of integrated access and backhaul (IAB) that commonly utilizes access and backhaul frequency resources was proposed for the forthcoming mobile communication networks [11], [12]. The IAB technology is getting a lot of attention as a promising solution to increase spectrum utilization efficiency with limited frequency resources. However, since the frequency bandwidth for user access links is shared by wireless backhaul links in the IAB architecture, this makes the resource allocation algorithm more complicated, resulting in increased network energy consumption [13], [14]. Thus, developing an advanced resource allocation algorithm with low complexity is of utmost importance to achieve an optimal solution in IAB.

B. RELATED WORKS

Reinforcement learning (RL)-assisted small cell networks have been studied in various aspects. In [15], Cheng et al. discussed how a double deep Q-network (DDQN) could be used to maximize the data rate of an IAB network with fixed user mobility.

Similarly, Lei et al. tried to maximize the data rate of the IAB network based on the actor-critic method [16], and E. Kim et al. proposed a multi-agent RL algorithm for maximizing energy efficiency (EE) in ultra-dense small-cell networks with non-uniform traffic distribution [17]. Since centralized RL has a huge number of states and actions when considering the large number of MBS or SBSs, the computational complexity incredibly increases, and thus finding optimal solution is nearly impossible.

In addition, the authors of [18] aimed at maximizing system utility in highly-dense small-cell networks by efficiently partitioning small cells into several clusters to reduce the inter-cell interferences. In [19], Lee et al. tried to maximize the network energy efficiency by using a deep neural network (DNN)-based optimal resource allocation in ultra-dense small cell networks. Furthermore, the authors of [20] considered dual access of small cell BSs using licensed and unlicensed bands to improve the energy efficiency in small cell networks based on a non-cooperative game theory. Also, [21] proposed a new RL framework considering unsupervised learning to maximize the sum rate of small cell networks using renewable resources. Even though these studies tried to maximize network performance, they still have limitations to optimally utilize multi-link common resources based on low-complexity RL operations.

C. CONTRIBUTIONS

Therefore, in this paper, we try to solve the above mentioned problem by using low-complexity multi-agent Q-learning in an IAB-enabled small-cell network. That is, through the reinforcement learning-based optimal channel allocation and transmit power control, we aim at maximizing network-wide energy efficiency in this network. Specifically, we propose multi-agent distributed Q-learning with a pre-resource partitioning technique (MADQ-PRP) to reduce the computational complexity of each agent. The main contributions of the proposed MADQ-PRP algorithm is summarized as follows.

- MADQ-PRP can achieve the EE optimal solution when considering the practical uneven traffic distribution and random user mobility in IAB-enabled small-cell network environments.
- MADQ-PRP can reduce the computational complexity of the reinforcement learning by effectively designing a multi-agent reinforcement learning framework maximizing the network-wide energy efficiency as well as minimizing the number of outage users. Specifically, each agent determines its action based on its local information so that it can significantly reduce the computational overhead arising from the centralized RL approaches.
- Despite the complex architecture of IAB-enabled networks, MADQ-PRP can converge to an optimal solution due to the action space adjustment based on the proposed pre-resource partitioning method. By comparing the performance with an exhaustive search-based optimal solution, we prove the optimal convergence of the proposed MADQ-PRP algorithm.
- Through the simulation results, we show the extensibility and flexibility of the proposed MADQ-PRP algorithm and its performance excellency compared to

TABLE 1. Notation Summary.

Parameter	Description	Parameter	Description
$\mathbb{K}$	Total BS set including MBSs and SBSs	$\mathbb{M}$	MBS set
$\mathbb{S}$	SBS set	$\mathbb{C}$	Frequency resource set
$\mathbb{K}_0$	Co-tier BS set using the same frequency channel	$\mathbb{K}_c$	Cross-tier BS set using the same frequency channel
$\mathbb{U}$	The total users set	U_{out}	Number of outage users
P_k^t	Transmit power of BS k	P_m^t	Transmit power of MBS m
P_s^t	Transmit power of SBS s	$P_{i,k}^r$	Received power of node i from BS k
ξ_i	Achievable data rate of node i	ξ_m	Achievable data rate of MBS m
ξ_s	Achievable data rate of SBS s	σ_i^2	Noise power density of node i
$\varsigma(i,k)$	Channel gain between node i and BS k	$\Gamma_{i,k}^n$	SINR of node i using channel n of BS k.
W_i^n	Bandwidth size of frequency channel n allocated to node i	$\mathbb{S}_M(t)$	MBS's state at time step t
$\mathbb{S}_S(t)$	SBS's state at time step t	$\mathbb{C}_{M,S}$	Frequency channel set assigned to MBS-to-SBS link
C_M^n	Node index using MBS's nth channel	C_S^n	Indication whether SBS's nth channel is used or not
$a_{i,M}$	Action set of MBS i	$a_{j,S}$	Action set of SBS j
$\pm\Delta C_{S_n}^n$	Transition of SBS assigned to nth channel of MBS	$\pm\Delta C_u^n$	Transition of SBS's nth channel availability
$\pm\Delta p_{t,M}$	Amount of transmit power adjustment of MBS	$\pm\Delta p_{t,S}$	Amount of transmit power adjustment of SBS

the various benchmark algorithms such as 'Random action', 'SBS on-off', 'SBS-only', and 'MADQ-only' with respect to average energy efficiency and the average number of outage users. Specifically, the conventional on-off power control algorithm is too simple, and thus it has great advantages when applied to a practical system. However, because the action set is simple, there are limitations in finding the optimal solution. On the other hand, although the centralized RL approach can find the optimal solution, its computational complexity is very high, rendering it challenging to apply it to the actual systems.

D. PAPER ORGANIZATION

The remainder of this paper is organized as follows: Section II introduces the system model and channel model of IAB-enabled small-cell networks, and Section III proposes the low-complexity multi-agent Q-learning algorithm with PRP for maximizing the network-wide energy efficiency. In Section IV, we demonstrate various simulation results of the benchmark and proposed algorithms. Also, in Section V, we make conclusions. Finally, the notations and symbols used in this paper are listed in Table 1.

II. SYSTEM MODEL

A. CHANNEL MODEL AND NETWORK-WIDE ENERGY EFFICIENCY CALCULATION

As shown in Fig. 1, we consider two-tier downlink IAB-enabled small-cell network environments consisting of MBSs (IAB donor), SBSs (IAB-node), and users. Here, in accordance with IAB concept, MBSs and SBSs utilize the same frequency resources ($\mathbb{C}_N = [1, \ldots, N]$), and SBSs are randomly placed around MBSs. Also, assume that each MBS can allocate one frequency channel to one user or one SBS, and each SBS can assign one frequency channel to each user. In addition, if the number of users to be served in each cell is greater than the cell capacity, MBSs and SBSs preferentially do users with higher signal to interference plus noise ratio (SINR).

The received signal strength indicator (RSSI) of user i (or SBS i, $i \in \mathbb{I}$) from SBS k (or MBS k, $k \in \mathbb{K}$) ($P_{i,k}^r$) can be obtained as

$$P_{i,k}^r = P_k^t \varsigma(i,k). \tag{1}$$

Here, P_k^t is the transmit power of BS k, and $\varsigma(i,k)$ means channel gain between node (user or SBS) i and BS (SBS or MBS) k. From equation (1), the SINR of node i from BS k ($\Gamma_{i,k}^n$) can be represented as

$$\Gamma_{i,k}^n = \frac{P_{i,k}^r}{(\sum_{k_0 \in \mathbb{K}_0} P_{i,k_0}^r) + (\sum_{k_c \in \mathbb{K}_c} P_{i,k_c}^r) + \sigma_i^2}, \tag{2}$$

where $\mathbb{K}_0$ represents a co-tier BS set using the same frequency channel, and $\sum_{k_0 \in \mathbb{K}_0} P_{i,k_0}^r$ in the equation (2) denotes co-tier interference. Also, $\mathbb{K}_c$ represents a cross-tier BS set using the same frequency channel, the equation $\sum_{kc \in \mathbb{K}_c} P_{i,k_c}^r$ represents cross-tier interference, and σ_i^2 is the noise power density of node i.

From equation (2), the achievable data rate of node i (ξ_i) can be expressed as

$$\xi_i = \sum_{n \in \mathbb{C}} W_i^n \log_2(1 + \Gamma_{i,k}^n), \tag{3}$$

where n represents the index of currently utilizing frequency channel, and $\mathbb{C}$ is the entire set of frequency channels. Also, W_i^n is the bandwidth size of frequency channel n allocated to node i. With this, we can calculate the network-wide

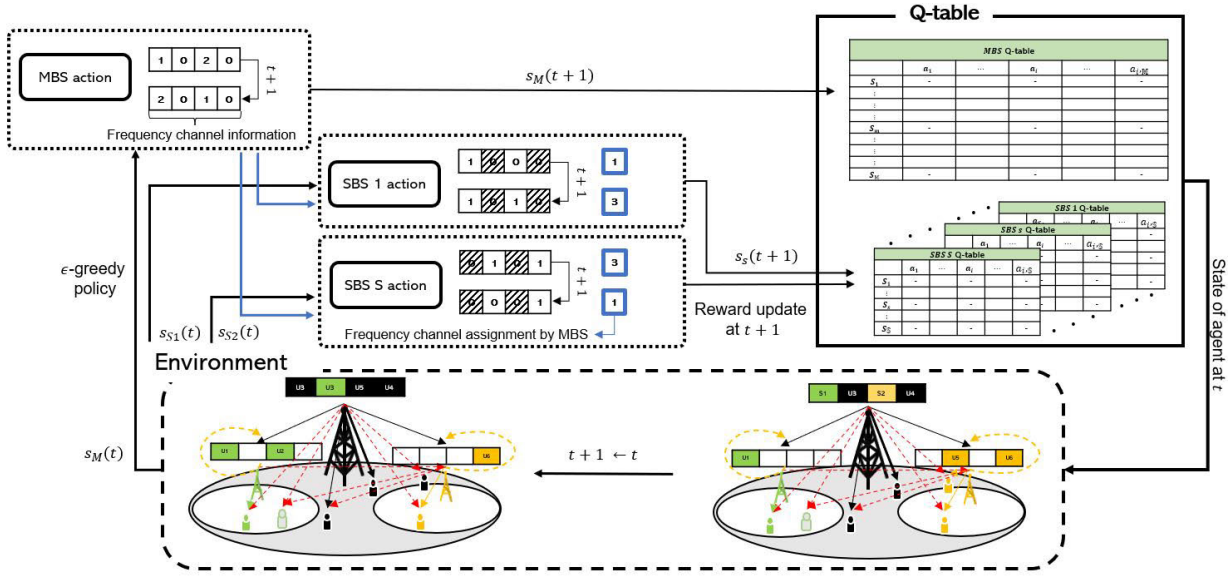


FIGURE 2. Proposed MADQ-PRP framework for IAB-enabled small-cell networks.

EE considering the number of outage users in IAB-enabled small-cell networks. The network-wide EE can be calculated as

$$EE = \frac{\xi_m + \sum_{s\in\mathbb{S}} \xi_s}{\sum_{s\in\mathbb{S}} P_s^t + \sum_{m\in\mathbb{M}} P_m^t}. \tag{4}$$

Here, $\mathbb{S}$ and $\mathbb{M}$ denote the set of SBSs and the set of MBSs, respectively. ξ_m and ξ_s represent the data rate of MBS and the data rate of SBS, respectively. Also, P_m^t and P_s^t mean the transmit power of MBS and SBS, respectively.

B. USER MOBILITY MODEL

In this paper, the user mobility was modeled based on the random walk model [23]. The random walk model assumes that users move on 2-dimensional x-y plane, and user's moving speed can be obtained within $[V_{min}, V_{max}]$. Here, V_{max} and V_{min} are the maximum and minimum moving speed of users. Moreover, the direction θ_u is randomly determined in the range of $[0 - 2\pi]$. In time-step t, the user's moving speed can be described as

$$\mathbf{V_u}(t) = [V_u cos(\theta_u(t)), V_u sin(\theta_u(t))] \tag{5}$$

In equation(5), the moving speed and direction of a certain user are initialized in every episode and set to a random value within the set range. In this paper, we assume that users can move within the coverage area of the MBS. Since there are multiple SBSs within one MBS, this means that users can move in and out of the SBS coverage area. Accordingly, users are associated with the MBS or SBSs providing the highest SINR, and the associated BS may be changed according to the user's movement.

C. TRAFFIC MODEL

In this paper, we assume that data traffic of IAB-enabled small-cell networks is generated based on the actual measurements proposed by [22]. This traffic model has been obtained through practical experiments and basically considers a geographically disproportionate traffic distribution. Specifically, half of cell sites take only 15% of the entire data traffic, whereas 5% of cell sites carry 20% of the total data traffic. Along with this, data traffic growth tends to increase the most in places where traffic is already high, while traffic growth is slower in areas where the load is already low.

III. PROPOSED LOW-COMPLEXITY MULTI-AGENT Q-LEARNING (MADQ-PRP) FOR MAXIMIZING NETWORK-WIDE ENERGY EFFICIENCY

We herein propose a MADQ-PRP framework to maximize network-wide energy efficiency while minimizing the number of outage users for IAB-enabled small-cell networks. As shown in Fig. 2, each agent (MBS or SBS) has its individual Q-table. The Q-table of MBS contains information about the channel usage status that MBS allocates and its transmit power. Also, in the proposed MADQ-PRP framework, state, action, reward, policy, and Q-function are defined as follows.

A. STATE

The goal of the proposed MADQ-PRP framework is to maximize network-wide energy efficiency and minimize the number of outage users in the entire IAB-enabled small-cell networks. The MBS's state consists of channel allocation status and the amount of its current transmit power, and the SBS's state describes channel availability and the strength of its current transmit power. The states of MBS and SBS ($\mathbb{S}_M(t)$

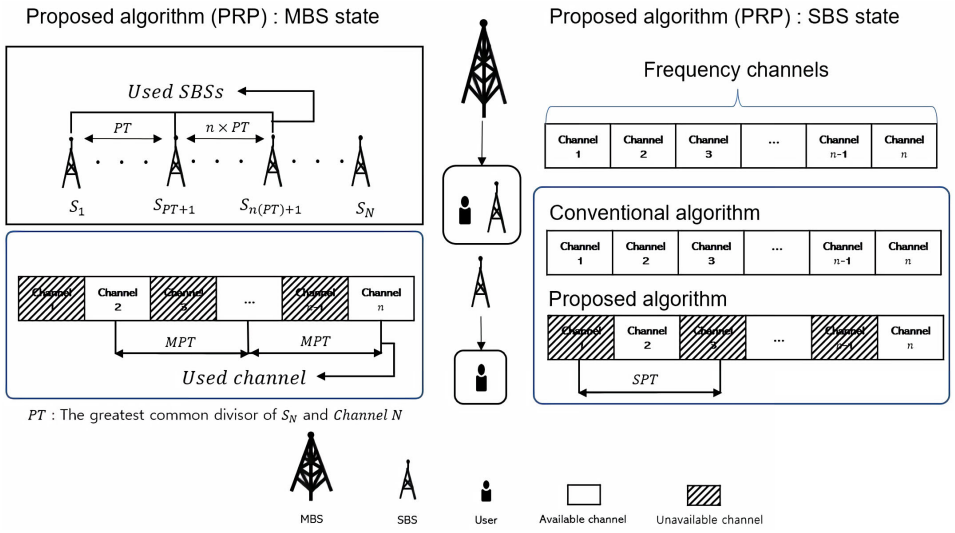


FIGURE 3. Detailed procedure of proposed PRP method in IAB-enabled small-cell networks.

and $\mathbb{S}_S(t)$) can be represented as follows.

$$\mathbb{S}_M(t) = [C_M^1(t), \cdots, C_M^n, \cdots, C_M^{|\mathbb{C}|}(t), P_m^t(t)], \quad (6)$$

$$\mathbb{S}_S(t) = [C_S^1(t), \cdots, C_S^n(t), \cdots, C_S^{|\mathbb{C}|}(t), P_s^t(t), \mathbb{C}_{M,S}]. \quad (7)$$

Here, $\mathbb{S}_M(t)$ in the equation (6) means the state of the MBS at time step t, and $C_M^n(t)$ represents the allocation status of the nth channel at time step t. Also, $|\mathbb{C}|$ is the total number of channels, and $P_M^t(t)$ denotes MBS's transmit power at time step t. In equation (7), $\mathbb{S}_M(t)$ means the state of SBS at time step t, and $C_S^n(t)$ indicates whether the nth channel is used or not. $P_S^t(t)$ represents the transmit power of SBS S at time step t, and $\mathbb{C}_{M,S}$ is a frequency channel set assigned to MBS-to-SBS link.

B. ACTION

MBS's action is to allocate $|\mathbb{C}|$ channels to SBSs and users and adjust the transmit power between the minimum to the maximum. Also, SBS's action is to decide whether or not to use frequency channels and adjust the transmit power between the minimum to the maximum. Actions of MBS i and SBS j can be expressed as follows.

$$a_{i,M} = [\pm\Delta C_{S_n}^1, \cdots, \pm\Delta C_{S_n}^n, \cdots, \pm\Delta C_{S_n}^{|\mathbb{C}|}, \pm\Delta p_{t,M}] \quad (8)$$

$$a_{j,S} = [\pm\Delta C_u^1, \cdots, \pm\Delta C_u^n \cdots, \pm\Delta C_u^{|\mathbb{C}|}, \pm\Delta p_{t,S}] \quad (9)$$

In equation (8), $\pm\Delta C_{S_n}^n$ denotes the transition of SBS assigned to nth channel of MBS. When it becomes 0, the MBS allocates the corresponding frequency channel to its associated user. In addition, $\pm\Delta p_{t,M}$ describes the amount of transmit power adjustment of MBS. In equation (9), $\pm\Delta C_u^n$ determines the transition of SBS's nth channel availability, and $C_u^n \in [0, 1]$. Moreover, $\pm\Delta p_{t,S}$ is the amount of transmit power adjustment of SBS.

C. REWARD AND Q-FUNCTION UPDATE

In this paper, the reward of MADQ-PRP is defined as the combination of the network-wide energy efficiency and the number of outage users. The reward is represented as

$$R = \left(\frac{\xi_m + \sum_{s\in S}\xi_s}{\sum_{s\in\mathbb{S}} P_s^t + \sum_{m\in\mathbb{M}} P_m^t}\right) \times e^{\left(-\frac{U_{out}}{|\mathbb{U}|}\right)}, \quad (10)$$

where U_{out} is the number of outage users and $|\mathbb{U}|$ denotes the total number of users in the entire network. Also, from equation (10), the Q-value of agent j is updated as follows.

$$\begin{aligned} Q(s_j(t), a_j(t)) = {} & (1-\alpha)\cdot Q(s_j(t), a_j(t)) \\ & + \alpha[R_j(s_j(t+1), a_j(t)) \\ & + \eta \cdot max_{a_{j'}\in A} Q(s_j(t+1), a_j')], \end{aligned} \quad (11)$$

where, α is the learning rate, and β is the discount factor.

D. POLICY

To choose the action, we utilize a decayed ϵ-greedy policy. This policy helps finding the optimal solution by taking a random action with probability ϵ that gradually decreases as the episode progresses. The decayed ϵ-greedy policy can be represented as

$$a_t = \begin{cases} \text{Random action} & \text{with Probability } \epsilon, \\ \arg\max\limits_{a_t\in A} Q(s_t, a_t) & \text{with Probability } 1-\epsilon. \end{cases} \quad (12)$$

E. PRE-RESOURCE PARTITIONING (PRP) METHOD

To reduce the computational complexity of the multi-agent distributed Q-learning, we propose a pre-resource partitioning (PRP) method which limits the number of frequency channels assignable to users. If MBS uses all available frequency channels, the number of states can be calculated as $(|\mathbb{S}| + |\mathbb{M}|)^{|\mathbb{C}|} \times P_{tx,M}$. Here, $|\mathbb{S}|$, $|\mathbb{M}|$, and $|\mathbb{C}|$ denote the number of SBSs, the number of MBSs, and the number of

TABLE 2. Computational complexity of benchmark and proposed algorithms.

Algorithm	Centralized QL	Distributed QL	MADQ-PRP
Computational complexity ($O(\cdot)$)	$O(\lvert\mathbb{S}_\mathbb{M}\rvert\lvert\mathbb{S}_\mathbb{S}\rvert^N\lvert\mathbb{A}_\mathbb{M}\rvert\lvert\mathbb{A}_\mathbb{S}\rvert^N)$	$O(\lvert\mathbb{S}_\mathbb{M}\rvert\lvert\mathbb{A}_\mathbb{M}\rvert + N\lvert\mathbb{S}_\mathbb{S}\rvert\lvert\mathbb{A}_\mathbb{S}\rvert)$	$O(\frac{\lvert\mathbb{S}_\mathbb{M}\rvert\lvert\mathbb{A}_\mathbb{M}\rvert}{\psi_m} + N\frac{\lvert\mathbb{S}_\mathbb{S}\rvert\lvert\mathbb{A}_\mathbb{S}\rvert}{\psi_s})$

Algorithm 1 Proposed MADQ-PRP Algorithm for Maximizing Network-Wide Energy Efficiency While Minimize the Number of Outage Users in IAB-Enabled Small-Cell Networks

Place MBS M and SBSs S in network
Place the user U non-uniform in the cell range of SBS.
Partition the bandwidth to form a channel Ch_N.
Apply PRP method to MBS and SBS.
for Every episodes (t) **do**
  Calculate $\epsilon_{init} \times (1 - \epsilon_{init})^{\frac{t}{t_{max}}}$
  $V_{u,t}$ and $\theta_{u,t}$ are determined by random walk model for each user. ▷ $u \in U$
  for Every iterations (i) **do**
    MBS action is determined according to the ϵ-greedy policy.
    $a_M = \begin{cases} \text{Random action} \\ \arg\ \max_{a_{M,i} \in A_{M,i}} Q(s_{M,i}, a_{M,i}) \end{cases}$
    for $s = 1 : S_N$ **do** ▷ S_N is Number of SBSs
      SBS action is determined according to the ϵ-greedy policy.
      $a_S = \begin{cases} \text{Random action} \\ \arg\ \max_{a_{S,i} \in A_{S,i}} Q(s_{S,i}, a_{S,i}) \end{cases}$
    end for
    Calculates SINR between all SBSs and all users.
    SBS allocates channels to users with higher SINR
    MBS assigns channels to users with higher SINR.
    Calculate network-wide energy efficiency.
    Update SBS's and MBS's Q-tables.
  end for
end for

frequency channels, respectively. Also, $P_{tx,M}$ is the number of possible transmit power candidates of MBS. It can be seen that the total number of states of MBS is significantly related to the increase in the number of frequency channels. According to the PRP method, since the number of SBSs which can be assigned to each channel is obtained by the maximum common divisor of $|\mathbb{S}|$ and $|\mathbb{C}|$ (MPT). Accordingly, the number of channels that can be assigned to each SBS is partitioned, resulting in a reduction of $|\mathbb{S}|$.

In addition, channel allocation coefficient of SBS (SPT) is obtained by the least common divisor of $|\mathbb{C}|$ and $|\mathbb{S}|$ except 1. SBS can use the channel with the same remainder of division when its unique number and channel unique number are divided by SPT. As a result, the number of states of SBS applying the PRP method can be expressed as $(\frac{|\mathbb{C}|}{SPT}) \times P_{tx,S} \times \frac{|\mathbb{C}|}{MPT}$. Here, $P_{tx,S}$ is the number of possible transmit power candidates of SBS. The detailed operation of the PRP method is described in Fig. 3, and Algorithm 1 describes the overall procedure of the proposed MADQ-PRP algorithm.

TABLE 3. Simulation parameters.

Parameter	Value
Number of episodes	1000
Number of iterations	10000
Transmit power of MBS	2∼10 [W]
Transmit power of SBS	0∼1 [W]
MBS Pathloss	34+40log(d)
SBS Pathloss	37+30log(d)
Frequency Bandwidth	10 [MHz]
Nosie power	-174dBm/Hz + 10log(W)+10dB
Small scale fading	Rayleigh Fading

Table 2 shows the computational complexity of the benchmark and proposed algorithms. Based on Big-O notation, the computational complexity of the centralized QL algorithm can be given as $O(|\mathbb{S}_\mathbb{M}||\mathbb{S}_\mathbb{S}|^N|\mathbb{A}_\mathbb{M}||\mathbb{A}_\mathbb{S}|^N)$. Here, the size of MBS's state set ($\mathbb{S}_M$) can be obtained as $\mathbb{C}_M \times \mathbb{P}_M$ where $\mathbb{C}_M$ is the number of available frequency channels of the MBS and $\mathbb{P}_M$ is the number of transmit power adjustment levels of the MBS. Similarly, the size of SBS's state set ($\mathbb{S}_S$) can be represented as $\mathbb{C}_S \times \mathbb{P}_S$ where $\mathbb{C}_S$ is the number of available frequency channels of the SBS and $\mathbb{P}_S$ is the number of transmit power adjustment levels of the SBS. Also, $|\mathbb{A}_\mathbb{M}|$ and $|\mathbb{A}_\mathbb{S}|$ represent the sizes of the action sets of the MBS and the SBS, respectively. Since the centralized Q-learning considers a set of states for all agents, the computational complexity exponentially increases as the number of agents increases. However, in the distributed Q-learning, since agents only consider their own state information, the computational complexity becomes $O(|\mathbb{S}_\mathbb{M}||\mathbb{A}_\mathbb{M}| + N|\mathbb{S}_\mathbb{S}||\mathbb{A}_\mathbb{S}|)$. Furthermore, the proposed MADQ-PRP algorithm considering the pre-resource partitioning, can additionally reduce the computational complexity compared to the distributed Q-learning algorithm. The computational complexity of the proposed MADQ-PRP can be obtained as $O(\frac{|\mathbb{S}_\mathbb{M}||\mathbb{A}_\mathbb{M}|}{\psi_m} + N\frac{|\mathbb{S}_\mathbb{S}||\mathbb{A}_\mathbb{S}|}{\psi_s})$ where ψ_M and ψ_S are PRP factors of MBS and SBS, respectively.

IV. SIMULATION RESULTS

A. SIMULATION ENVIRONMENTS

In this section, we consider a two-tier IAB-enabled small-cell network with non-uniform traffic distribution, and the

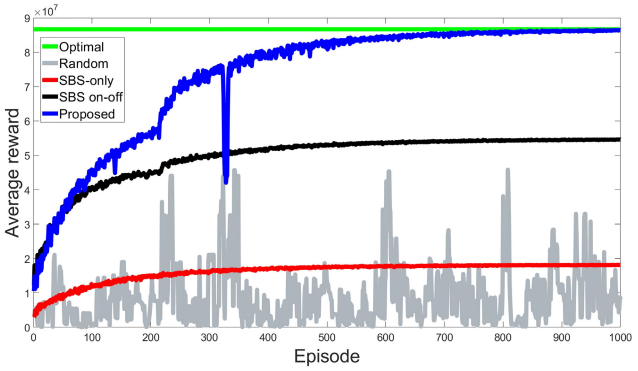


FIGURE 4. Accumulated average reward vs. episode when $|\mathbb{M}| = 1$, $|\mathbb{S}| = 2$, $|\mathbb{C}| = 6$, and $|\mathbb{U}| = 6$.

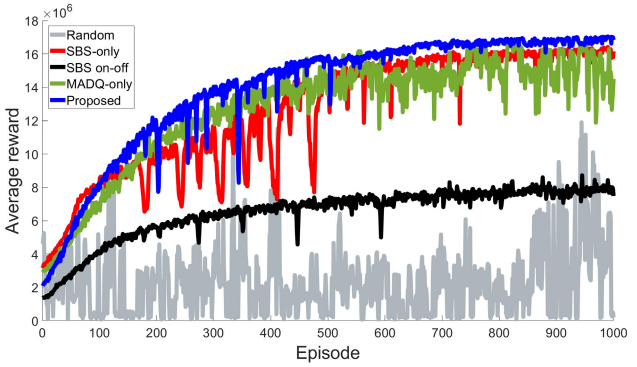


FIGURE 5. Accumulated average reward vs. episode when $|\mathbb{M}| = 1$, $|\mathbb{S}| = 6$, $|\mathbb{C}| = 6$, and $|\mathbb{U}| = 14$.

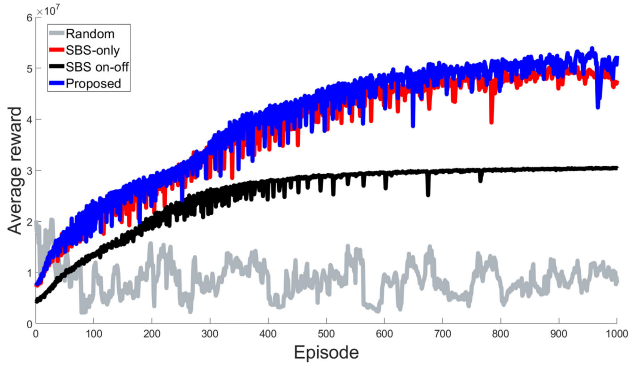


FIGURE 6. Accumulated average reward vs. episode when $|\mathbb{M}| = 1$, $|\mathbb{S}| = 12$, $|\mathbb{C}| = 12$, and $|\mathbb{U}| = 39$.

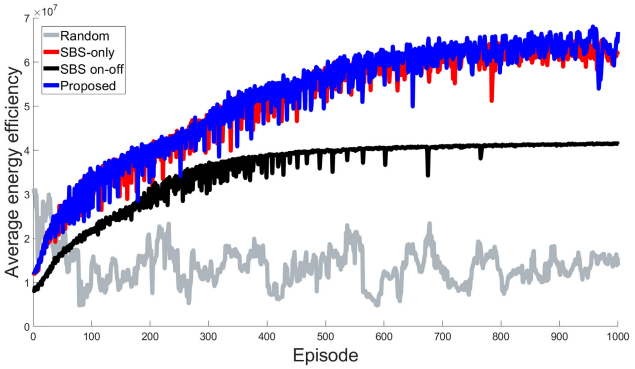
maximum speed of user is 0.1 m per episode. The simulation is conducted on a computer equipped with i5-12600 CPU 3.30 GHz and 32.0 GB RAM memory. Other simulation parameters used in this paper are summarized in Table 3. Also, we consider four benchmark algorithms such as 'Optimal', 'Random action', 'SBS-only', 'SBS on-off' and 'MADQ-only'. The detailed description for these benchmark algorithms is as follows.

- **(B1) Optimal:** In this paper, we obtain an optimal solution by using the exhaustive search algorithm considering all possible cases of both MBSs and SBSs. By comparing to this benchmark algorithm, we can confirm that the proposed MADQ-PRP algorithm can achieve the optimal solution.
- **(B2) Random action:** In the random action algorithm, each agent always acts randomly without learning about channel allocation and transmit power control of MBSs and SBSs.
- **(B3) SBS on-off:** This algorithm performs transmit power control, but the agent has only two options such as 0 W (off) and 1 W (on). This algorithm does not consider the detailed power control of SBSs.
- **(B4) SBS-only:** In this algorithm, SBSs only supports users. That is, MBS only provides backhaul communication links to SBSs.
- **(B5) MADQ-only:** This algorithm simply means MADQ without PRP. As mentioned before, MADQ-only has vey high computational complexity compared to MADQ-PRP.

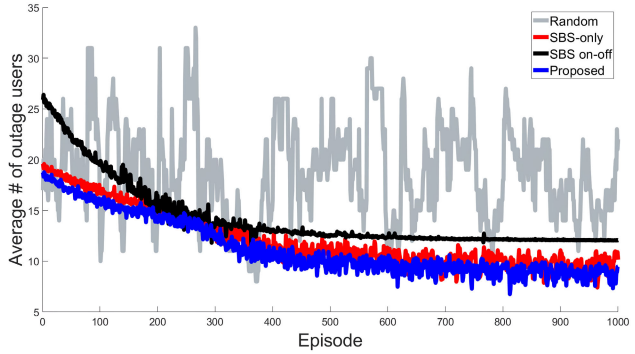
B. RESULTS AND DISCUSSION

First, we consider a simple network environment consisting of $|\mathbb{M}| = 1$, $|\mathbb{S}| = 2$, $|\mathbb{C}| = 6$, and $|\mathbb{U}| = 6$ to compare the performance of the proposed MADQ-PRP algorithm with the optimal solution obtained by the exhaustive search algorithm, as shown in Fig. 4. Furthermore, small-scale fading and user movement are not considered in this simple network model. As a result, because the proposed MADQ-PRP allocates the channel with the highest SINR to users, it can minimize the interference between MBS and SBSs as well as the transmit power consumption. Also, we can show that MADQ-PRP can converge to the optimal solution. In contrast, due to the number of outage users, it can be seen that the SBS-only algorithm has severe performance degradation.

Fig. 5 shows the accumulated average reward vs. episode when $|\mathbb{M}| = 1$, $|\mathbb{S}| = 6$, $|\mathbb{C}| = 6$, and $|\mathbb{U}| = 14$. Due to the computational complexity, we cannot obtain the optimal solution. Also, because of the coarse power control, we can show that the SBS on-off algorithm has a relatively lower reward compared to other algorithms. In the case of SBS-only algorithm, because the MBS does not provide access links to users, this algorithm might have more outage users compared to the proposed MADQ-PRP. It results in reward degradation of the SBS-only algorithm. Even though the MADQ-only algorithm considers all states and actions, because of their huge sizes, MADQ-only fails to converge and undergoes oscillation, as shown in Fig. 5. In contrast to the MADQ-only algorithm, the proposed MADQ-PRP algorithm converges and has the highest average reward compared to the benchmark algorithms. In addition, the CPU execution time of the MADQ-PRP, distributed RL, and centralized RL algorithms required for RL to converge was 2367.66s, 2605.05s, and 114334.89s, respectively. That is, it is confirmed that the proposed MADQ-PRP algorithm can reduce the CPU execution time by 9.1% and 97.9%



(a) Energy efficiency vs. episode



(b) Number of outage users vs. episode

FIGURE 7. Accumulated average energy efficiency and number of outage users vs. episode under $|\mathbb{M}| = 1$, $|\mathbb{S}| = 12$, $|\mathbb{C}| = 12$, and $|\mathbb{U}| = 39$.

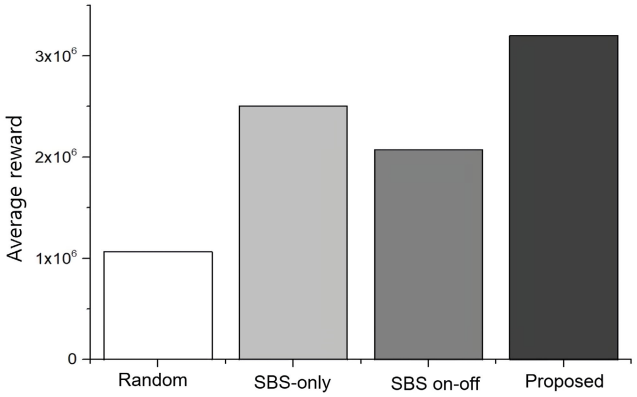


FIGURE 8. Test results for average reward of benchmark and proposed algorithms when $|\mathbb{M}| = 1$, $|\mathbb{S}| = 12$, $|\mathbb{C}| = 12$, and $|\mathbb{U}| = 39$.

compared to the distributed and centralized reinforcement learning algorithms, respectively.

Moreover, to consider more dense and complicated network environments, 12 SBSs, 12 channels, and 39 users were considered in Fig. 6. As shown in this figure, the proposed MADQ-PRP has the greatest reward compared to all benchmark algorithms. From equation (10), the reward of the proposed MADQ-PRP algorithm is defined as the combination of the network-wide energy efficiency and the number of outage users. Figs 7a and 7b show the original results of the accumulated average energy efficiency and the number of outage users vs. episode under $|\mathbb{M}| = 1$, $|\mathbb{S}| = 12$, $|\mathbb{C}| = 12$, and $|\mathbb{U}| = 39$, respectively. As shown in these figures, it is shown that the proposed MADQ-PRP algorithm outperforms the benchmark algorithms with respect to energy efficiency and the number of outage users because of the low-complexity multi-agent RL-based transmit power control and resource allocation. Furthermore, Fig. 8 shows average reward vs. algorithm through test experiments. Through this, we can demonstrate that the proposed method can find the highest reward compared to other benchmark algorithms, and the proposed MADQ-PRP algorithm may be flexibly applied to various IAB-enabled networks. Consequently, as shown in Figs. 4–8, even though an SBS has a relatively small network coverage compared to an MBS, it has an advantage with respect to capacity. On the other hand, although an MBS has strength in terms of coverage, it can be seen that it has a lower capacity than the SBS due to propagation loss proportional to the increase in distance. From this point of view, in practical heterogeneous networks, where SBSs and MBSs coexist, optimizing the network deployment, radio resource allocation, and transmit power control of MBSs and SBSs by considering topological dynamics and user mobility is very challenging. Therefore, the proposed algorithm based on the low-complexity RL and PRP could be one of the solutions to optimize the network performance from the perspective of the network operators in consideration of the coverage-capacity tradeoff.

V. CONCLUSION

In this paper, we proposed the MADQ-PRP algorithm to maximize network-wide energy efficiency as well as minimize the number of outage users in IAB-enabled small-cell networks. Through the simulation results, we demonstrated extensibility and flexibility of the proposed algorithm and performance excellency compared to the various benchmark algorithms such as 'Random action', 'SBS on-off', 'SBS-only', and 'MADQ-only'. In addition, we showed the convergence of the proposed MADQ-PRP algorithm to the optimal solution obtained by the exhaustive search algorithm. Furthermore, by proposing the PRP method, our proposed algorithm could reduce the computational complexity of the reinforcement learning algorithm significantly. Through the reward-sharing approach, our proposed algorithm consistently outperforms the benchmark algorithms in terms of both energy efficiency and the number of users experiencing outages. we hope that the proposed MADQ-PRP algorithm may be applied to many kinds of actual IAB-enabled networks in the forthcoming 6G mobile networks.

REFERENCES

[1] V. K. Quy, A. Chehri, N. M. Quy, N. D. Han, and N. T. Ban, "Innovative trends in the 6G era: A comprehensive survey of architecture, applications, technologies, and challenges," *IEEE Access*, vol. 11, pp. 39824–39844, 2023.

[2] D. Renga and M. Meo, "Can high altitude platform stations make 6G sustainable?" *IEEE Commun. Mag.*, vol. 60, no. 9, pp. 75–80, Sep. 2022.

[3] H. Tataria, M. Shafi, M. Dohler, and S. Sun, "Six critical challenges for 6G wireless systems: A summary and some solutions," *IEEE Veh. Technol. Mag.*, vol. 17, no. 1, pp. 16–26, Mar. 2022.

[4] H. R. Chi and A. Radwan, "Quality of things' experience for 6G artificial intelligent Internet of Things with IEEE P2668," *IEEE Commun. Mag.*, vol. 61, no. 6, pp. 58–64, Jun. 2023.

[5] H. Lee, B. Lee, H. Yang, J. Kim, S. Kim, W. Shin, B. Shim, and H. V. Poor, "Towards 6G hyper-connectivity: Vision, challenges, and key enabling technologies," *J. Commun. Netw.*, vol. 25, no. 3, pp. 344–354, Jun. 2023, doi: 10.23919/JCN.2023.000006.

[6] L. M. P. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "Toward greener 5G and beyond radio access networks—A survey," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 768–797, 2023.

[7] A. Jahid, K. H. Monju, S. Hossain, and F. Hossain, "Hybrid power supply solutions for off-grid green wireless networks," *Int. J. Green Energy*, vol. 16, no. 1, p. 12 3, Oct. 2018.

[8] E. Kim, H.-H. Choi, H. Kim, J. Na, and H. Lee, "Optimal resource allocation considering non-uniform spatial traffic distribution in ultra-dense networks: A multi-agent reinforcement learning approach," *IEEE Access*, vol. 10, pp. 20455–20464, 2022.

[9] *6G: The Next Hyper-Connected Experience for All*, Samsung Research, Suwon-si, South Korea, Jul. 2020.

[10] W. Chen, X. Lin, J. Lee, A. Toskala, S. Sun, C. F. Chiasserini, and L. Liu, "5G-advanced toward 6G: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1592–1619, Jun. 2023.

[11] *Study on Integrated Access and Backhaul: Release 16*, 3GPP, document TR 38.874, 2018.

[12] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 62–68, Mar. 2020.

[13] Y. Zhang, H. Shan, M. Song, H. H. Yang, X. S. Shen, Q. Zhang, and X. He, "Packet-level throughput analysis and energy efficiency optimization for UAV-assisted IAB heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9511–9526, Jul. 2023.

[14] H. Alghafari, M. S. Haghighi, and A. Jolfaei, "High bandwidth green communication with vehicles by decentralized resource optimization in integrated access backhaul 5G networks," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1438–1447, Sep. 2022.

[15] Q. Cheng, Z. Wei, and J. Yuan, "Deep reinforcement learning-based spectrum allocation and power management for IAB networks," in *Proc. IEEE Int. Conf. Commun. Workshops*, Jun. 2021, pp. 1–6.

[16] W. Lei, Y. Ye, and M. Xiao, "Deep reinforcement learning-based spectrum allocation in integrated access and backhaul networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 3, pp. 970–979, Sep. 2020.

[17] H. Lee, E. Kim, H. Kim, J. Na, and H.-H. Choi, "Multi-agent Q-learning based cell breathing considering SBS collaboration for maximizing energy efficiency in B5G heterogeneous networks," *ICT Exp.*, vol. 8, no. 4, pp. 525–529, Dec. 2022.

[18] G. Yang, A. Esmailpour, N. Nasser, G. Chen, Q. Liu, and P. Bai, "A hierarchical clustering algorithm for interference management in ultra-dense small cell networks," *IEEE Access*, vol. 8, pp. 78726–78736, 2020.

[19] W. Lee, H. Lee, and H.-H. Choi, "Deep learning-based network-wide energy efficiency optimization in ultra-dense small cell networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8244–8249, Jan. 2023.

[20] A. Shahid, V. Maglogiannis, I. Ahmed, K. S. Kim, E. De Poorter, and I. Moerman, "Energy-efficient resource allocation for ultra-dense licensed and unlicensed dual-access small cell networks," *IEEE Trans. Mobile Comput.*, vol. 20, no. 3, pp. 983–1000, Mar. 2021.

[21] J. Jang and H. J. Yang, "Deep learning-aided user association and power control with renewable energy sources," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2387–2403, Apr. 2022.

[22] P. Frenger, C. Friberg, Y. Jading, M. Olsson, and O. Persson, "Radio network energy performance: Shifting focus from power to precision," Ericsson Rev., White Paper, Feb. 2014, p. 9. [Online]. Available: https://www.ericsson.com/4ac61e/assets/local/reports-papers/ericsson-technology-review/docs/2014/er-radio-network-energy-performance.pdf

[23] F. Bai and A. Helmy, "A survey of mobility models," in *Wireless Adhoc Networks*. Los Angeles, CA, USA: Univ. Southern California, 2004.

JUNSEUNG LEE (Student Member, IEEE) received the B.S. degree from the School of Electronic and Electrical Engineering, Hankyong National University, Anseong, South Korea, in 2022. His current research interests include B5G/6G wireless communications, ultra-dense distributed networks, reinforcement learning for UAV networks, unsupervised learning for wireless communication networks, and the Internet of Things.

HYUN-HO CHOI (Senior Member, IEEE) received the B.S. (cum laude), M.S., and Ph.D. degrees (summa cum laude) from the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2001, 2003, and 2007, respectively.

From February 2007 to February 2011, he was a Senior Engineer with the Communication Laboratory, Samsung Advanced Institute of Technology (SAIT), Suwon, South Korea. Since March 2011, he has been a Professor with Hankyong National University, Anseong, South Korea, where he is currently a Professor with the School of ICT, Robotics and Mechanical Engineering. He has also experienced as a Visiting Researcher with TeleCIS Wireless Inc., Mountain View, CA, USA, in 2006, and a Visiting Scholar with Stanford University, Stanford, CA, USA, in 2008, and the University of California at Irvine, Irvine, CA, USA, in 2017. He has published 77 international journal articles and 37 international conference papers. His current research interests include bio-inspired networking, distributed optimization, machine learning, wireless energy harvesting, mobile ad-hoc networks, and next-generation wireless communication.

Prof. Choi is currently a Life Member of KICS and KIICE. He received the Excellent Paper Award at ICUFN, in 2012; the Best Paper Award at ICN, in 2014; the Best Paper Award at Qshine, in 2016; the Excellent Paper Award in *Journal of Korean Institute of Communications and Information Sciences*, in 2018; and the Hankyong Academic Awards, in 2014, 2016, and 2017. He was a co-recipient of the SAIT Patent Award, in 2010, and the Paper Award at Samsung Conference, in 2010.

SEUNG-CHAN LIM (Member, IEEE) received the B.S. degree in electronic and electrical engineering from Hongik University, Seoul, South Korea, in 2011, the M.S. degree in electronic and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2013, and the Ph.D. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2019.

He was a Post-Doctoral Researcher with the Information and Electronics Research Institute, KAIST, in 2019. From 2019 to 2021, he was a Senior Researcher with the Agency for Defense Development (ADD), Daejeon. Since 2021, he has been with the School of ICT, Robotics and Mechanical Engineering, Hankyong National University (HKNU), Anseong, South Korea, where he is currently an Assistant Professor. His current research interests include wireless communication, signal processing, and machine learning.

HYUNGSUB KIM received the B.S. and M.S. degrees from the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2002 and 2004, respectively. Since 2004, he has been with the Intelligent Ultra-Dense Small Cell Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon. His current research interests include 4G LTE, 5G small cells, EPC, RRC, and interface protocols for mobile communication networks.

JEEHYEON NA (Member, IEEE) received the B.S. degree from the Department of Computational Statistics, Chonnam National University, in 1989, and the M.S. and Ph.D. degrees from the Department of Computer Science, Chungnam National University, in 2000 and 2008, respectively. She has been with the Electronics and Telecommunications Research Institute (ETRI), since 1989, where she is currently the Director of the Intelligent Ultra Dense Small Cell Research Section. Her current research interests include 4G and 5G small cells, self-organizing networks (SON), and location management and paging for mobile communication networks. She is a member of IEICE Communication Part.

HOWON LEE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2003, 2005, and 2009, respectively.

From 2009 to 2012, he was a Senior Research Staff/Team Leader of the Knowledge Convergence Team, KAIST Institute for Information Technology Convergence (KI-ITC). Since 2012, he has been with the School of Electronic and Electrical Engineering and the Institute for IT Convergence (IITC), Hankyong National University (HKNU), Anseong, South Korea. He has also experienced as a Visiting Scholar with the University of California at San Dieo (UCSD), La Jolla, CA, USA, in 2018. His current research interests include B5G/6G wireless communications, ultra-dense distributed networks, in-network computations for 3D images, cross-layer radio resource management, reinforcement learning for UAV networks, unsupervised learning for wireless communication networks, and the Internet of Things. He was a recipient of the 2006 Joint Conference on Communications and Information (JCCI) Best Paper Award and the Bronze Prize at the Intel Student Paper Contest, in 2006. He was a recipient of the Telecommunications Technology Association (TTA) Paper Contest Encouragement Award, in 2011; the Best Paper Award at the Korean Institute of Communications and Information Sciences (KICS) Summer Conference, in 2015; the Best Paper Award at the KICS Fall Conference, in 2015; the Honorable Achievement Award from 5G Forum Korea, in 2016; the Best Paper Award at the KICS Summer Conference, in 2017; the Best Paper Award at the KICS Winter Conference, in 2018; the Best Paper Award at the KICS Summer Conference, in 2018; the Best Paper Award at the KICS Winter Conference, in 2020; and the Best Paper Award at the KICS Winter Conference, in 2022. He received the Minister's Commendation by the Minister of Science and ICT, in 2017.