https://doi.org/10.7840/kics.2023.48.8.942

3차원 셀룰러 네트워크기법에서 분산 심층강화학습 기반 에너지 효율 최대화

이 승 민◆, 반 태 원*, 이 호 원°

Distributed Deep Reinforcement Learning-Based Energy Efficiency Maximization in 3D Cellular Networks

Seungmin Lee◆, Tae-Won Ban*, Howon Lee°

요 약

본 논문에서는 이동성을 지닌 지상 사용자에게 안정적인 공중-지상 통신 커버리지를 제공하기 위한 다중 unmanned aerial vehicle-base station(UBS) 기반 3차원 셀룰러 네트워크를 고려한다. 특히, UBS 네트워크의 매우 짧은 네트워크 라이프타임 문제를 해결하기 위해서, 네트워크 전체 에너지 효율을 극대화할 수 있도록 UBS의 이동성 및 전송전력을 제어하고자 한다. 하지만, 지상 사용자가 움직이는 동적 환경 문제를 기존 반복 및 최적화 기법으로 풀어내기에 그 어려움이 존재한다. 따라서, 본 논문에서는 분산 deep Q-network(DQN) 기반 UBS 제어 방안을 제안한다. 그리고 분산 학습의 강점을 보이기 위해, 두 가지 중앙집중형 학습 방안을 소개하고, 이 기법들과 다중-에이전트 분산 큐-러닝(multi-agent distributed Q-learning, MD-QL) 그리고 탐욕적 행동(greedy action, GA)을 비교방안으로 고려한다. 결과적으로, 제안 방안이 UBS의 수와 사용자 이동속도에 따라 기존 알고리즘보다 그 성능이 강건하고 우수함을 보인다.

키워드 : 무인항공기, 심층 큐-네트워크, UAV 제어, 공중-지상 채널, 에너지 효율 극대화
Key Words : UAV, Deep Q-Network, UAV Control, Air-to-Ground Channel, Energy Efficiency Maximization

ABSTRACT

In this paper, we consider the multiple unmanned aerial vehicle-base station(UBS)-based 3D cellular networks to provide air-to-ground(A2G) communication coverage to moving ground users. Especially, to alleviate the short network lifetime problem of the UBS networks, we aim to control the movement and the transmission power of UBS so that maximizing the network-wide energy efficiency. However, considering the dynamic environment in which ground users move, deriving the optimal solution to the problem is significantly difficult with existing iterative methods or optimization methods. Therefore, in this paper, we propose a distributed deep Q-network(DQN)-based UBS control method. Also, to show the advantages of the distributed learning, we introduce two centralized learning methods, and then we consider the two centralized learning method, multi-agent distributed Q-learning(MD-QL) and greedy action(GA) methods as benchmarks. Conclusionally, we verify that the performance of the proposed method outperforms the conventional methods according to the movement speed of the ground user and the number of UBSs.

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2022-0-00704, 초고속 이동체 지원을 위한 3D-NET 핵심 기술 개발, 50%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1010602, 50%).
◆ First Author : School of Electronic and Electrical Eng., Hankyong National Univ., julsin1@hknu.ac.kr, 학생회원
° Corresponding Author : School of Electronic and Electrical Eng., Hankyong National Univ., hwlee@hknu.ac.kr, 종신회원
* Dept. of Intelligent Communication Eng., Gyeongsang National Univ., twban35@gnu.ac.kr, 종신회원
논문번호 : 202302-025-A-RN, Received February 11, 2023; Revised April 23, 2023; Accepted May 3, 2023

Ⅰ. 서 론

6G 시대에는, 매우 다양한 종류 디바이스들이 나타나며 방대한 양의 모바일 트래픽이 발생할 것으로 전망된다[1-4]. 이를 위해, 6G에서는 저비용 배치, 신속한 이동성 및 유연한 토폴로지 구축 등의 장점을 기반으로 unmanned aerial vehicle(UAV)의 활용이, 기존 네트워크에서 발생하는 문제점들을 해결하기 위한 핵심 솔루션들 중 하나로 중요하게 고려되고 있다. 하지만, 아직 무선 통신 네트워크에서 UAV의 활용을 위해 해결해야 할 많은 문제가 존재한다. 특히, 네트워크 유지시간과 직결되는 UAV의 짧은 배터리 라이프 타임을 완화하기 위해서, 에너지 효율 성능 향상을 위한 UAV의 최적 제어에 대한 연구가 매우 활발히 진행되고 있다[5-11]. 구체적으로, UAV swarm-to-ground control station (GCS) 패킷 전송 시나리오에서, UAV의 불필요한 재전송으로 인한 배터리 절약을 위해, UAV의 잔여 에너지에 따라 패킷 전송 확률을 제어하는 residual energy-aware online random access (RE-ORA) 연구가 수행되었다[7]. 또한, 지상 사용자들에게 하향링크 커버리지를 제공해 주기 위해서, 최대 커버리지 범위를 보장하는 UAV-BS(UBS)의 최적 고도를 수학적 분석을 통해 도출하는 연구가 수행되었다[8]. 비록, 기존 연구[7][8]에서는 최적 솔루션을 도출함으로써 그 성능을 입증했지만, 사용자가 정적인 환경만을 고려하였고, 반복 알고리즘을 제안하였다. 이러한 방식은 적은 계산 복잡도로 최적 솔루션 도출이 가능하지만, 그 솔루션은 결국 사용자 이동성이 고려된 동적인 환경에서 최적이 될 수 없다. 동적인 환경은 현실성을 위해서 반드시 고려되어야 할 요소이지만, 기존의 선형 알고리즘으로 접근하기에는 그 어려움이 존재한다. 최근, 딥러닝 분야의 호황과 더불어 강화학습 분야의 성공 사례가 많아지면서, 이러한 동적 환경에 대응하기 위해 심층 강화학습을 이용하고 있다. 구체적으로, 공중-지상 통신 네트워크에서 정지해있는 지상 사용자들에게 하향링크 커버리지 제공을 위한 강화학습 기반 다중 UBS의 최적 제어 연구가 수행되었다[9]. 그리고, 지상 사용자가 움직이는 동적 환경에서 사용자 서비스 품질(quality-of-service; QoS)를 극대화하기 위해, 분산 큐 러닝(Q-Learning, QL) 기반 UBSs의 3차원 최적 배치 및 최적 제어 연구가 수행되었으며[10], 최적 주파수 자원 사용 문제를 고려하면서 이동성을 지닌 UAV 터미널 노드의 배터리 제약 문제를 완화하기 위해, 계층적 다중-에이전트 큐-러닝 기반 ground control station(GCS) 제어 연구가 수행되었다[11]. 비록, 기존 연구[9]에서는 강화학습 기법을 이용하여 에너지 효율을 개선했지만, 사용자가 정적인 학습 환경을 가정하였다. 또한, 기존 연구[10,11]에서는 사용자가 지속적으로 움직이는 동적 환경을 고려하였지만, 사용자 이동속도에 따른 제안 방안의 성능 분석을 제공하지 않았다. 그리고, 동적 환경에 대응하기 어려운 테이블 기반 학습 방식인 큐-러닝 기법을 이용하였다. 본 연구에서는 지상 사용자가 지속적으로 움직이는 동적 환경을 고려하고, 중앙집중형 deep Q-Network(DQN) 알고리즘 기반 UBSs 제어 방안에서 나타날 수 있는 신호처리를 위한 상당한 오버헤드 문제나, 개인정보 및 보안 측면에서 초래될 수 있는 문제 등을 완화하기 위해서 분산형 DQN 기반 UBS 이동성 및 전송전력 제어 방안을 제안한다. 이에 따라, 본 논문에서의 주요 제안 사항은 다음과 같다.

- 본 논문에서는 지상 사용자가 지속적으로 움직이는 다중 UBS 기반 3차원 셀룰러 네트워크에서, 네트워크 전체 에너지 효율 극대화를 위해, 심층 강화학습 기법 중 하나인 분산 DQN 기반 UBS 제어 방안을 제안한다.
- 에이전트 수에 따른 제안 방안의 성능 분석을 위해, 중앙집중형 순차적 제어 DQN(round‐robin DQN, RR-DQN)과 중앙집중형 선택적-K DQN(selective‐K DQN, SK-DQN) 그리고 사용자 이동속도에 따른 성능 분석을 위해, multi-agent distributed Q-learning(MD-QL)[9]와 탐욕적 행동(greedy action, GA) 방안을 비교 방안으로서 고려한다.
- 시뮬레이션을 통해, 에이전트의 수와 사용자 이동속도에 따른 제안 방안의 그 성능의 강건함(robustness)을 보이고, 높은 에너지 효율을 달성함을 보인다.

Ⅱ. 시스템 및 채널 모델

본 논문에서는 사용자가 지속적으로 움직이고 있는 다중 UBS 기반 3차원 셀룰러 네트워크를 고려하며, 이는 그림 1에 요약되어 있다. 구체적으로, 3차원 셀룰러 네트워크에서 G 지상 사용자들에게 하향링크 커버리지를 제공해 주기 위해 U UBSs가 목표 영역(A^{area}) 내에 배치된다. 이때, UBSs와 지상 사용자들의 위치 표현을 위해, 데카르트 좌표계 시스템을 고려한다. 지상 사용자($j \in \{1,2,\cdots,G\}=\mathbb{G}$)는 목표 영역 내에 무작위로 분포되며, 모든 사용자의 위치는 UBSs에게 알려진 것으로 가정한다. 한편, 강화학습에서 에이전트의 초기 위치는 수렴 속도에 충분한 영향을 줄 수 있으므로[10], UBS($i \in \{1,2,\cdots,U\}=\mathbb{U}$)의 초기 위치는 지상 사용자

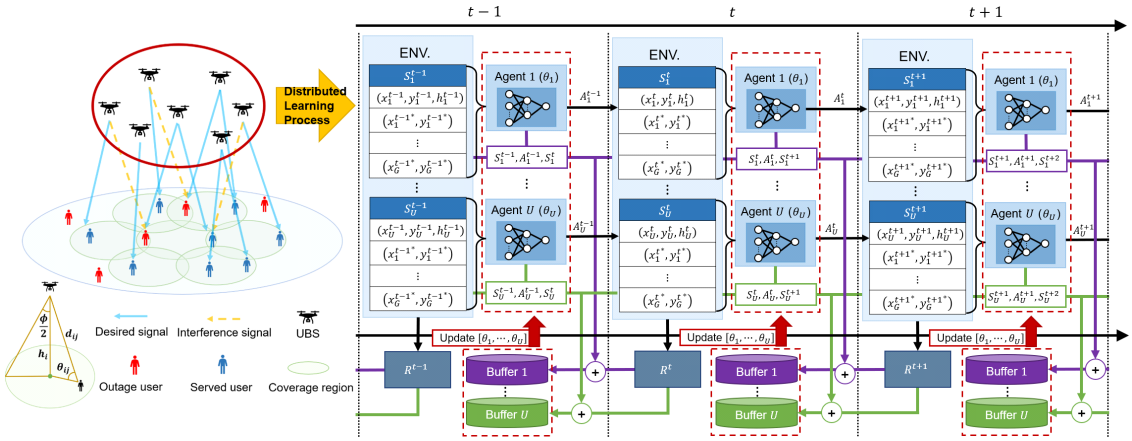


그림 1. UBS 기반 3차원 셀룰러 네트워크를 위한 분산형 DQN 프레임워크
Fig. 1. Distributed DQN framework for UBS-aided 3D cellular networks

분포에 대해 k-means 알고리즘을 적용하여 도출된 중심점을 이용하여 결정한다. UBS는 특정 출사각(ϕ)을 가진 지향성 안테나가 탑재된 것으로 고려하며, UBS의 커버리지 영역은 UBS의 ϕ에 의해 결정된다. 또한, 지상 사용자들은 random way point mobility(RWPM) 모델[12]에 따라서 시뮬레이션 동안 지속적으로 움직이는 것으로 고려한다.

본 논문에서는 UBS의 3차원 이동성에 따른 공중-지상 전파 특성을 반영하기 위해, 고도 각도에 의존하는 경로 손실 모델[13]을 이용한다. 이 모델은 소규모 페이딩(small-scale fading)의 영향이 고려되지 않지만, 도심 속 건물들의 분포나 밀도에 따른 공중-지상 전파 특성이 반영되었기 때문에, suburban, urban, dense urban 그리고 highrise urban 4가지 도심 환경 시나리오에 적용할 수 있다. 이 모델에 따라서, UBS i의 송신 신호(P_{ij}^{TX})로부터 지상 사용자 j가 수신한 전력(P_{ij}^{RX})은 다음과 같이 계산된다[13].

$$P_{ij}^{RX}(d_{ij},\theta_{ij}) = P_{ij}^{TX} - \left[L_{ij}^{LoS}(d_{ij})P_{ij}^{LoS}(\theta_{ij}) + L_{ij}^{NLoS}(d_{ij})P_{ij}^{NLoS}(\theta_{ij})\right] \quad (1)$$

수식 (1)에서, d_{ij}와 $\theta_{ij} = sin^{-1}(h_i/d_{ij})$는 각각 UBS i와 지상 사용자 j사이의 직선 거리와 고도 각도이다. 또한, P_{ij}^{LoS}와 P_{ij}^{NLoS}은 각각 UBS i와 지상 사용자 j사이의 링크가 가시선 (line-of-sight; LoS)과 비가시선(non-LoS; NLoS)일 확률이며, L_{ij}^{LoS}와 L_{ij}^{NLoS}은 각각 UBS i와 지상 사용자 j사이의 링크가 LoS와 NLoS일 때 고려되는 경로 손실을 나타낸다. 이 파라미터들의 상세한 내용은 기존 연구[13]에 기술되어 있다.

수식 (1)을 이용하여 UBS i와 지상 사용자 j사이의 signal-to-interference-to-noise ratio(SINR, γ_{ij})을 다음과 같이 계산한다.

$$\gamma_{ij} = \frac{P_{ij}^{RX}}{\sigma^2 + \sum_{n=1,n\neq i}^{U}\left(P_{nj}^{RX} \times I_j(n)\right)} \quad (2)$$

수식 (2)에서, σ^2는 열 잡음 전력을 나타내며, $I_j(i)$는 지상 사용자 j가 UBS i의 커버리지 영역 내에 존재하는지 여부를 나타내는 지시 함수이다. 지상 사용자 j는 다수의 UBSs로부터 수신한 신호의 $\gamma_{\mathbb{U}j} = [\gamma_{1j},\cdots,\gamma_{Uj}] \geq \gamma^{th}$로부터 가장 높은 γ_{ij}를 제공한 UBS i에게 연결되며, γ^{th}는 SINR 임계값을 나타낸다. $\gamma_{\mathbb{U}j} = \varnothing$일 때, 지상 사용자 j는 아웃티지 사용자가 된다. 한편, $\gamma_{\mathbb{U}j} \neq \varnothing$일 때, UBS i와 지상 사용자 j사이의 데이터율(R_{ij})은 다음과 같이 계산된다.

$$R_{ij} = B_{ij} \times \log_2(1+\gamma_{ij}) \quad (3)$$

수식 (3)에서 $B_{ij} = B_i / C_i$는 UBS i로부터 지상 사용자 j가 할당받은 대역폭 자원을 나타내며, B_i^{tot}는 UBS i의 전체 대역폭 자원을 나타낸다. 또한, C_i는 UBS i에게 커버되고 있는 지상 사용자의 수를 나타낸다. 수식 (2)와 수식 (3)을 이용하여, 본 논문에서는 UBS i의 에너지 효율(η_i)을 다음과 같이 정의한다.

$$\eta_i = \frac{\sum_{m=1}^{C_i}(R_{im})}{P^C + P_{ij}^{TX}} = \frac{\sum_{m=1}^{C_i}\left(B_{im}\times \log_2(1+\gamma_{im})\right)}{P^C + P_{ij}^{TX}} \quad (4)$$

수식 (4)에서, P^C는 UBS의 고정 소비 전력을 나타낸다. 본 논문에서는 네트워크 에너지 효율 극대화를 달성하도록 에이전트에게 학습 방향을 제시해주기 위해 η_i를 이용한다.

Ⅲ. 분산 DQN 기반 UBS 제어 방안

DQN은 행동을 선택할 때, 행동 가치함수인 큐-함수($Q(S_i^t, A_i^t;\theta_i)$)를 이용한다. $Q(S_i^t, A_i^t;\theta_i)$의 잠재적인 의미는 시간 t에 UBS i가 위치한 상태 S_i^t에서, 메인 네트워크(θ_i)로부터 행동 A_i^t를 선택했을 때, 받을 것이라고 기대하는 보상($\mathbb{E}\left[R^{t+1}+\kappa R^{t+2}+\cdots|S_i^t, A_i^t;\theta_i\right]$)이다. 여기에서, κ는 감가율이다. 따라서 DQN에서는 최고의 보상을 보장하도록, $Q(S_i^t, A_i^t;\theta_i)$를 최적의 $Q(S_i^t, A_i^t;\theta_i)^*$로 발전시키는 것이 궁극적인 목표이다. 하지만, DQN은 모든 행동에 대한 $Q(S_i^t, A_i^t;\theta_i)$를 발전시켜야 할 필요가 있기 때문에, 행동 크기의 증가가 학습을 어렵게 만든다. 본 논문에서는 행동 크기가 에이전트의 수에 의존하지 않는 분산 DQN 방안을 제안한다. 먼저, 본 논문에서 다루고 있는 네트워크 전체 에너지 효율 극대화를 위한 다중 UBS 제어 문제를, 강화학습으로 풀어내기 위해 마르코브 결정 과정(markov decision process, MDP)으로 다음과 같이 문제를 정의한다.

- 에이전트(agent): 에이전트는 학습의 주체이며 제안 방안에서는 각 UBS i가 에이전트로서 역할을 수행하게 된다.
- 행동(action): 각 에이전트 i의 3차원 이동성과 전송 전력 제어를 행동(A_i)으로 정의한다.

$$A_i^t \in \left\{\pm\Delta_x, \pm\Delta_y, \pm\Delta_h, \pm\Delta_P^{TX}, 0\right\} \quad (5)$$

수식 (5)에서, $\pm\Delta_x$, $\pm\Delta_y$, $\pm\Delta_h$, $\pm\Delta_P^{TX}$ 그리고 0은 각각, UBS의 "좌/우 이동", "상/하 이동", "고도상승/하강", "전송전력 증대/증감/유지"를 의미한다.

- 상태(state): 각 에이전트 i의 3차원 좌표(x_i, y_i, h_i)와 전송전력(P_{ij}^{TX}) 그리고 지상 사용자의 2차원 좌표(x_j^*, y_j^*)를 이용하여, 상태(S_i^t)를 정의한다.

$$S_i^t = \left[(x_i^t, y_i^t, h_i^t), (x_1^{t*}, y_2^{t*}), \cdots, (x_G^t*, y_G^t*)\right] \quad (6)$$

에이전트의 상태에 지상 사용자의 좌표 정보를 고려함으로써, 에이전트는 사용자의 위치에 따른 최적 행동을 학습할 수 있다.

- 보상(reward): 에이전트가 네트워크 전체 에너지 효율 극대화 하도록 학습 방향성을 제시해주기 위해, 수식 (4)를 이용하여, 공동 보상(R^t)을 다음과 같이 정의한다.

$$R^t = \sum_i^U \eta_i^t = \sum_i^U \left(\frac{\sum_{m=1}^{C_i^t}\left(B_{im}^t \times \log_2\left(1+\gamma_{im}^t\right)\right)}{P^C + P_{ij}^{TX,t}} \right) \quad (7)$$

에이전트는 MDP로 정의된 환경 속에서 그림 1의 절차에 따라, 최대 R^t를 보장하는 정책을 찾기 위해 $Q(S_i^t, A_i^t;\theta_i)$를 발전시켜나간다. 구체적으로, 그림 1의 $t-1$에서 각 에이전트 i는 각자 자신의 3차원 좌표 및 전송 전력과 모든 지상 사용자의 좌표로 구성된 상태 S_i^{t-1}을 이용하여 각자의 θ_i로부터 행동 A_i^{t-1}을 선택한다. 각 에이전트는 선택된 행동을 취함으로써, 상태 S_i^{t-1}에서 다음 상태 S_i^t로 전이하며, S_i^t에서 공동 보상 R^t을 받는다. 동시에, 각 에이전트는 학습 샘플($\left\{S_i^{t-1}, A_i^{t-1}, R^t, S_i^t\right\}$)을 자신의 경험 리플레이 버퍼($D_i^{buf}$)에 저장한다. 또한, D_i^{buf}로부터 일부 학습 샘플을 이용하여, 공동 보상 R^t을 극대화 하도록 각 에이전트의 θ_i를 업데이트한다. 이를 위해 손실 함수(L_i)가 정의된다.

$$L_i = \frac{1}{D_i}\sum_k^{D_i}\left(Q(S_i^k, A_i^k;\theta_i) - y_i^k\right)^2 \quad (8)$$

$$y_i^k = R^k + \kappa \times max_{A_i^k} Q\left(S'^k_i, A'^k_i;\theta_i^*\right) \quad (9)$$

수식 (8), (9)에서 D_i은 D_i^{buf}의 배치 크기를 의미하며, S_i^k, A_i^k, S'^k_i 그리고 A'^k_i는 각각 D_i^{buf}의 k번째 샘

플의 현재 상태, 현재 행동, 다음 상태 그리고 다음 행동을 나타낸다. 또한, θ_i^*은 목표 네트워크를 나타낸다. 학습률(l)에 따라, 손실함수를 최소화함으로써, $Q(S_i^t, A_i^t; \theta_i)$가 잠재적 보상인 y_i^k에 가까워지도록 θ_i가 업데이트된다.

Ⅳ. 실험 결과 및 분석

본 논문에서는 UBSs를 순차적으로 제어하는 중앙집중형 RR-DQN 방안, K대의 UBSs를 제어하는 중앙집중형 SK-DQN 방안, MD-QL 방안[9] 그리고 GA 방안을 비교 방안으로 고려했다. 방안별 특징은 다음과 같다.

- RR-DQN: 다중 UBSs를 제어하기 위해 GCS가 에이전트 역할을 수행하며, 매 타임 슬롯마다 한 대의 UBS를 순차적으로 제어한다.
- SK-DQN: RR-DQN과 유사하지만, 네트워크 성능 향상에 기여할 수 있는 K대의 UBSs를 선택하고 제어한다.
- MD-QL: 각 UBS가 에이전트인 분산형 시스템으로서, DQN 방안들과 다르게 테이블 기반 학습 방법을 이용한다. DQN에 비해서 저 복잡도 알고리즘이지만 사용자가 움직이는 동적 환경을 학습하기에 그 어려움이 있다.
- GA: 매 타임 슬롯마다 가장 높은 보상 받을 수 있는 최적의 행동만을 선택한다. GA의 그 성능은 일반적으로 제안 방안 성능의 상계(upper bound)이며, GA와 제안 방안의 비교로서 제안 방안의 수렴성을 검증할 수 있다.

본 논문에서 시뮬레이션은 2-, 3-에이전트 환경에 대해 진행되었다. 각 방안은 다양한 사용자 위치 분포에 대해 학습하기 위해, 사용자 이동속도가 빠른 5 [m/s]인 환경에서 학습하였으며, 학습된 최적 모델을 이용하여 사용자 이동속도 0.05, 1, 3, 5 [m/s]인 각 환경에 따라 100,000번 스텝의 테스트를 진행하였다. 이 외의 공통 파라미터는 표 1에 요약되어 있다.

표 2에서는 각 에이전트 시나리오에서 네트워크 사이즈에 따른 제안 방안의 에너지 효율 비교 결과가 나타나며 그림 2, 3의 결과로부터 신경망을 이용한 학습 기법들은 지상 사용자 이동속도에 따라 에너지 효율 성능이 강건함을 보인다. 또한, RR-DQN은 매 타임 슬롯마다 한 대의 UBS만을 제어하기 때문에 에이전트의 수가 증가함에 따라서 그 성능이 크게 저하되는 것을 알 수 있다. 또한, SK-DQN ($k=1$)의 결과 또한 RR-DQN과 동일한 이유로 에이전트의 수의 증가에 따른 그 성능의 저하가 나타난다. 중앙집중형 DQN 방안에서 한 타임 슬롯 마다 여러 UBS를 제어하게 될 경우에, 행동의 크기는 지수 함수적으로 증가하게 된다. 이는 학습에 어려움을 부과하며, SK-DQN ($k=2$)와 ($k=3$) 결과로부

표 1. 시뮬레이션 파라미터
Table 1. Simulation parameter

파라미터	값
지상 사용자의 수(G)	11x U
초과 경로 손실(ξ^{LoS}, ξ^{NLoS})	1, 20 [dB]
환경 파라미터(α, β)	9.6117, 0.1581
안테나 출사각(ϕ)	60 °
SINR 임계값(γ^{th})	-2 [dB]
열 잡음 전력(σ^2)	-174 [dBm/Hz]
대역폭(B^{tot})	200 [kHz]
반송 주파수(f_c)	1 [GHz]
UBS 최대/최소 고도	150, 120 [m]
고정 소비 전력(P^C)	0.1 [W]
최대/최소 전송 전력	0.5, 0.1 [W]
상/하/좌/우 이동량($\pm\Delta_y$,$\pm\Delta_x$)	±5 [m]
고도 상승/하강 이동량($\pm\Delta_h$)	±5 [m]
전송 전력 증대/증감량($\pm\Delta_P^{TX}$)	±0.1 [W]
배치 샘플 크기(K)	64
목표 네트워크 학습 주기	200 [step]
옵티마이저 / 학습률	RMSProp / 0.0001
감가율(κ)	0.8
2-, 3-에이전트 환경 각각에서 제안 방안의 신경망 구성(relu)	300x300 450x400
에포크(epoch)	1000
2-, 3-에이전트 환경 각각의 에포크 당 반복수	5000/epoch, 7000/epoch
2-, 3-에이전트 환경 각각의 목표 영역(A^{area})	200*200, 280*280 [m^2]

표 2. 2-, 3-에이전트 시나리오에서 네트워크 사이즈에 따른 제안 방안의 에너지 효율 비교
Table 2. Average energy efficiency comparison for network size in 2-agent and 3-agent scenarios

Network size [m^2]	Avg. energy efficiency [Gb/s/W]	
	2-Agent	3-Agent
200*200	45.51	66.38
280*280	45.23	66.18

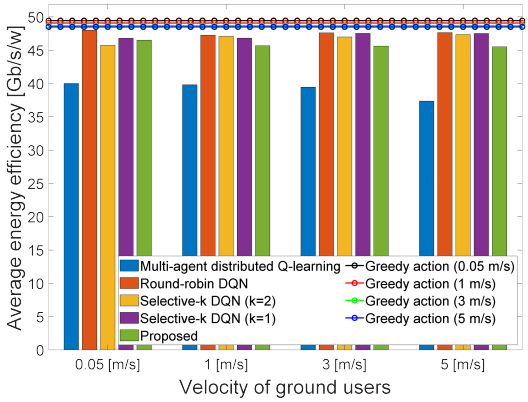


그림 2. 2-에이전트 시나리오에서 방안별 에너지 효율 비교
Fig. 2. Average energy efficiency comparison of each algorithm for 2-agent scenario.

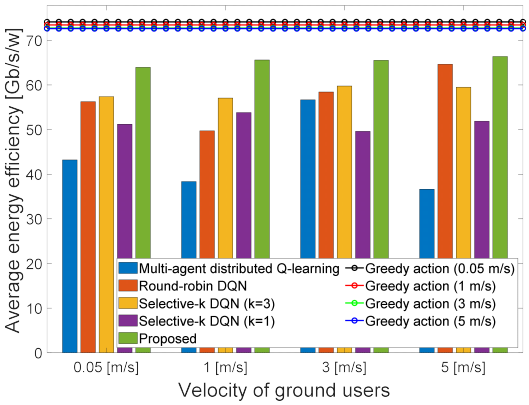


그림 3. 3-에이전트 시나리오에서 방안별 에너지 효율 비교
Fig. 3. Average energy efficiency comparison of each algorithm for 3-agent scenario.

터 그 사실을 알 수 있다. 여기에서 RR-DQN과 SK-DQN 방안은 GCS가 모든 UBSs의 상태 정보를 관찰하여 UBS를 제어하기 때문에, 2-agent 시나리오에서는 중앙집중형 방안의 에너지 효율 성능이 제안 방안보다 높게 나타나지만 그 차이는 미미하다. 하지만, GCS가 모든 UBSs의 상태 정보를 수집하고 행동을 브로드캐스트 하는 과정에서 상당한 신호처리를 위한 오버헤드가 발생하며, UBS의 상태 정보의 노출로 개인정보와 보안 측면에서 문제가 될 수 있다. 제안 방안에서 각 UBS는 자신의 상태 정보만을 관찰하여 행동을 결정하지만, 공동 보상을 이용하기 때문에 네트워크 전체 성능을 높이도록 학습할 수 있게 된다. 따라서, 제안 방안에서의 각 UBS는 자체적으로 행동을 결정하기 때문에, 에이전트 수가 증가함에 따라서, 그 성능이 중앙집중형 DQN 방안의 성능보다 강인함이 나타난다.

V. 결 론

본 논문에서는 지상 사용자가 지속적으로 움직이는 다중 UBS 기반 3차원 셀룰러 네트워크에서 네트워크 전체 에너지 효율 극대화를 위해 분산 DQN 기반 UBS 제어 기법을 제안하였다. 특히, 현실적인 이동속도와 에이전트 수에 따른 제안 방안의 성능을 분석하였고, 이를 위해 중앙집중형 RR-DQN, 중앙집중형 SK-DQN, MD-QL 그리고 GA 방안을 비교 방안으로 고려하였다. 시뮬레이션을 통해 사용자 이동속도와 에이전트 수에 따른 제안 방안의 그 성능이 비교 방안 대비 그 성능의 강건함과 우수함을 검증하였다.

References

[1] Samsung, *6G The Next Hyper-Connected Experience for ALL, Samsung's 6G White Paper*, 2020. (https://news.samsung.com/global/samsungs-6g-white-paper-lays-out-the-companys-vision-for-the-next-generation-of-communications-technology).

[2] 5G Forum, *6G Technology Trends, in 6G Working Group White Paper*, 2021. (http://6gglobal.org/kr/sub/publication/publication_view.php?idx=2)

[3] 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," *IEEE J. Commun. and Netw.* (to appear) (https://doi.org/10.23919/JCN.2023.000006)

[4] H. Yu, H. Lee, and H. B. Jeon, "What is 5g? emerging 5g mobile services and network requirements," *Sustainability 2017*, vol. 9, no. 10, pp. 1-22, Oct. 2017. (https://doi.org/10.3390/su9101848).

[5] S. Lim, H. Yu, and H. Lee, "Optimal Tethered-UAV deployment in a2g communication networks: Multi-agent q-learning approach," *IEEE Internet of Things J.*, vol. 9, no. 19, pp. 18539-18549, Oct. 2022. (https://doi.org/10.1109/JIOT.2022.3161260).

[6] J. Lee, S. Lee, S. H. Chae, and H. Lee, "Performance analysis of cooperative dynamic-framed slotted ALOHA based on

random transmit power control in a2g communication networks," *IEEE Access*, vol. 10, pp. 106699-106707, Oct. 2022. (https://doi.org/10.1109/ACCESS.2022.3211943)

[7] S. Lim, S. H. Chae, and H. Lee, "RE-ORA: Residual energy-aware online random access for improving the lifetime of slotted ALOHA-Based swarming drone networks," *IEEE Access*, vol. 9, pp. 208-208, Mar. 2021. (https://doi.org/10.1109/ACCESS.2021.3066979).

[8] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," *2015 IEEE GLOBECOM*, pp. 1-6, San Diego, California, Dec. 2015. (https://doi.org/10.1109/GLOCOM.2015.7417609).

[9] S. Lee and H. Lee, "UAV-BS energy efficiency maximization based on multi-agent distributed q-learning," in *Proc. KICS Winter Conf.*, Pyeongchang, Korea, Feb. 2022. (https://doi.org/10.1109/JIOT.2021.3113128).

[10] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-uav networks: Deployment and movement design," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8036-8049, Aug. 2019. (https://doi.org/10.1109/TVT.2019.2922849).

[11] S. Lee, S. Lim, S. H. Chae, B. C. Jung, C. Y. Park, and H. Lee, "Optimal frequency reuse and power control in multi-uav wireless networks: Hierarchical multi-agent reinforcement learning perspective," *IEEE Access*, vol. 10, pp. 39555-39565, Apr. 2022. (https://doi.org/10.1109/ACCESS.2022.3166179).

[12] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 2, no. 3, pp. 257-269, Sep. 2003. (https://doi.org/10.1109/TMC.2003.1233531)

[13] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569-572, Dec. 2014. (https://doi.org/10.1109/LWC.2014.2342736).

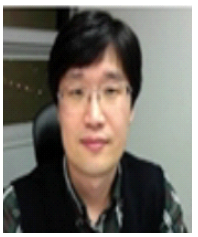
이 승 민 (Seungmin Lee)

2021년 2월 : 한경국립대학교 전기전자제어공학과 졸업

2021년 3월~현재 : 한경국립대학교 전자전기공학부 석사과정

<관심분야> B5G/6G 무선 통신, 고밀집 분산 네트워크, 강화학습기반 UAV 네트워킹

반 태 원 (Tae-Won Ban)



1998년 2월 : 경북대학교 전자공학과 학사

2000년 2월 : 경북대학교 전자공학과 석사

2010년 2월 : KAIST 전기 및 전자공학과 박사

2000년 2월~2010년 12월 : KT 모바일 R&D 연구원

2011년 1월-2012년 8월 : KT 네트워크그룹 프로젝트 매니저

2012년 9월-현재 : 경상국립대학교 지능형통신공학과 교수

<관심분야> 딥러닝기반 네트워크 알고리즘, 무선 자원관리 기술, OFDM/MIMO

이 호 원 (Howon Lee)

2003년 2월 : KAIST 전자전산학과 학사

2009년 8월 : KAIST 전기 및 전자공학과 박사 (석박사통합)

2009년 6월~2012년 2월 : KAIST ITC 연구조교수/팀장

2012년 3월~2021년 2월 : KAIST 겸직교수

2012년 3월~현재 : 한경국립대학교 전자전기공학부 전자공학전공 교수

<관심분야> 6G 모바일 네트워크, 무선자원관리, 드론/위성 통신, 머신러닝기반 통신 네트워크

[ORCID:0000-0001-5509-9202]