Movable-GS-Mounted TUBS의 자원 할당 및 위치 제어를 위한 계층적 분산 다중 에이전트 심층 강화학습 기법

박은서, 이예린, 김형섭°, 정방철, 이호원
아주대학교, 한국전자통신연구원°

parkes0312@ajou.ac.kr, yerin1205@ajou.ac.kr, mobman@etri.re.kr, bcjung@ajou.ac.kr, howon@ajou.ac.kr

Hierarchical Distributed Multi-Agent DQN for Joint Resource Allocation and Position Control of Movable-GS-Mounted TUBSs

Eunseo Park, Yerin Lee, Hyungsub Kim°, Bang Chul Jung, Howon Lee
Ajou University, ETRI°

요 약

UAV를 활용한 공중기지국은 배터리 제약으로 인해 장기간의 안정적인 통신 서비스 제공에 어려움이 있다. 이를 해결하기 위해 지상 기지국과 테더로 연결되어 지속적으로 전력을 공급받는 tethered UAV(TUAV)가 제안되었으나, 한정된 테더 길이로 인해 전체적인 이동 범위가 제한되는 문제점을 가지고 있다. 본 논문에서는 공중 기지국을 이동차량에 탑재한 Movable-GS-Mounted TUBS(TUAV base station) 시스템을 고려하여, 네트워크 전체 전송률을 최대화하고 아웃티지 사용자 수를 최소화하기 위한 계층적 분산 다중 에이전트 심층 강화학습 기법을 제안하고 그 성능을 비교 분석한다.

Ⅰ. 서 론

UAV를 활용한 공중 기지국은 높은 이동성을 가지며, 지상 기지국 대비 우수한 가시선(Line-of-Sight, LoS) 환경을 제공할 수 있다 [1]. 그러나 배터리 제약으로 인해 장시간 안정적인 통신 서비스 제공에는 한계가 있다. 이를 극복하기 위해 지상 기지국과 테더로 연결되어 지속적으로 전력을 공급받는 tethered UAV(TUAV)가 제안되었다 [2]. 하지만 한정된 테더 길이에 따른 이동 범위 제약으로 인해, 반구형 서비스 영역(Hovering region) 내에서 TUAV 기반 공중 기지국(TUAV base station, TUBS)의 3차원 위치 최적화가 필요하다 [3]. 한편, 위치가 고정되어 있는 기존 지상 기지국의 경우에도 고정된 위치로 인해 네트워크 커버리지 유연성과 확장성에 그 한계를 가지고 있다. 본 논문에서는 이러한 한계를 극복하기 위해 공중 기지국을 이동 차량에 탑재한 Movable-GS-mounted-TUBS 시스템을 고려하고, 지상 이동성을 활용함으로써 서비스 범위 확장과 커버리지 유연성 향상을 목표로 한다. 더 나아가, 복잡하고 동적인 통신 환경에서의 자원 할당과 위치 제어 문제를 동시에 최적화하기 위해 TUBS와 Movable GS를 각각 독립적인 에이전트로 정의한 계층적 다중 에이전트 심층 강화학습 기법을 제안한다.

Ⅱ. 시스템 및 채널 모델

A. 시스템모델

본 논문에서는 그림 1과 같이 Movable GS와 TUBS가 테더로 연결된 네트워크를 고려한다. Movable GS 집합은 $G=\{G_1, G_2, \cdots, G_N\}$, TUBS 집합은 $T=\{T_1, T_2, \cdots, T_N\}$로 정의한다. 전체 사용자 수는 U로 나타내며, U_i는 i번째 TUBS에 연결되어 서비스를 제공받는 사용자 수를 의미한다. 전체 사용자 수는 $U=\sum_{i=1}^{N} U_i$로 표현된다. TUBS는 테더로 연결된 Movable GS를 기준으로 정의된 반구형 서비스 영역 내에서 3차원 이동이 가능하다. Movable GS는 지상 평면에서 이동하며, 이에 따라 TUBS의 서비스 영역 또한 동적으로 변화한다. 사용자 이동성을 고려하기 위해 사용자의 이동을 Random Walk 모델로 모델링한다 [4]. 사용자 k의 위치는 매 time step마다 무작위로 선택된 이동각 $w_k \in (0, 2\pi)$과 이동속도 $v_k \in [0, v_{\max}]$에 따라 갱신된다.

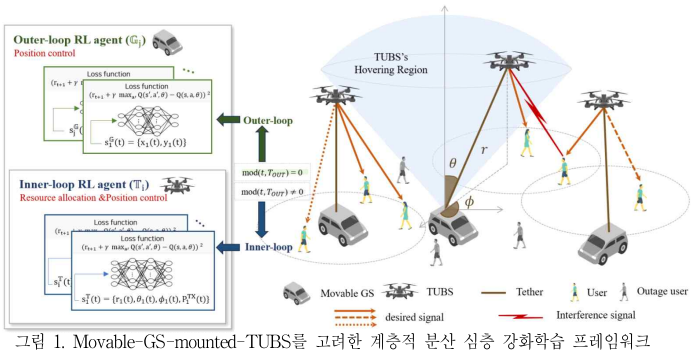

그림 1. Movable-GS-mounted-TUBS를 고려한 계층적 분산 심층 강화학습 프레임워크

B. 채널모델

TUBS와 사용자 간의 Air-to-Ground(A2G) 채널은 ITU-R에서 제안한 도심 환경 기반 확률적 채널 모델을 따른다 [5]. TUBS i와 사용자 k 간의 LoS 확률과 non-LoS(NLoS) 확률은 아래와 같이 계산한다.

$$P_{ik}^{LoS} = a\left(\tan^{-1}\left(\frac{z_i}{\sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}}\right) - 15\right)^b. \quad (1)$$

$$P_{ik}^{NLoS} = 1 - P_{ik}^{LoS}. \quad (2)$$

여기서 a와 b는 ITU-R에서 제시하는 도심별 환경 파라미터이다.

TUBS i의 전송 전력을 P_i^{TX}, TUBS i와 사용자 k간의 평균 경로 손실을 η_{ik}라 하면 수신 신호 전력은 아래와 같이 계산한다.

$$P_{ik}^{RX} = P_i^{TX} - \eta_{ik}. \quad (3)$$

사용자는 가장 높은 수신 전력을 제공하는 하나의 TUBS와 연결되며, 신호 대 잡음비(signal-to-interference-plus-noise ratio, SINR)는 아래와 같이 계산한다.

$$\Gamma_{ik} = \frac{P_{ik}^{RX}}{\sum_{l=1, l\neq i}^{N} P_{lk}^{RX} + \sigma^2}. \quad (4)$$

여기서 σ^2은 Additive White Gaussian Noise(AWGN)의 분산을 의미한다. SINR이 임계값(ζ) 미만인 사용자는 아웃티지 사용자로 간주한다. (4)로부터, 평균 데이터 전송률은 아래와 같이 계산한다.

$$\gamma_{ik}^{avg} = \begin{cases} \frac{B_i}{U_i}\log_2(1+\Gamma_{ik}) & \text{if } \Gamma_{ik} > \zeta \\ 0 & otherwise \end{cases} \quad (5)$$

여기서 B_i는 i번째 TUBS에 할당된 총 대역폭을 의미한다.

Ⅲ. 계층적 분산 심층 강화학습 프레임워크

Outer-loop reinforcement Learning(RL)에서는 일정 주기마다 movable GS의 위치를 제어하고, inner-loop RL에서는 매 time step마다 TUBS의 자원 할당 및 3차원 위치를 제어한다. 이러한 계층적 설계는 전체 행동 공간을 축소함으로써 계산 복잡도를 줄이고, 안정적인 학습을 가능하게 한다. 본 최적화 문제는 Markov decision process(MDP)로 모델링된다.

A. Outer-loop RL

Outer-loop RL에서의 에이전트는 movable GS로, 상태는 j번째 movable GS의 데카르트 좌표계에서의 위치로 표현된다. 행동은 지상 평면 상에서의 위치 제어로 다음과 같이 정의된다.

$$s_{G_j}(t) = \{x_j(t), y_j(t)\}. \tag{6}$$

$$a_{G_j}(t) \in \{\pm \triangle_x, \pm \triangle_y, \triangle_{x,y} = 0\}. \tag{7}$$

B. Inner-loop RL

Inner-loop RL에서의 에이전트는 TUBS로, 상태는 연결된 movable GS를 기준으로 한 i번째 TUBS의 구면 좌표계 상에서의 위치와 전송 전력으로 구성된다. 행동은 위치 제어 $a_i^{pos}(t)$과 전송 전력 제어 $a_i^{pow}(t)$로 구성되며, 전체 행동 공간은 $A_i^{pos} \times A_i^{pow}$로 정의된다. 여기서 집합 A_i^{pos}과 A_i^{pow}는 $\{\pm \triangle_r, \pm \triangle_\theta, \pm \triangle_\phi, \triangle_{r,\theta,\phi} = 0\}$, $\{\pm \triangle_{P^{TX}}, \triangle_{P^{TX}} = 0\}$로 정의된다.

$$s_{T_i}(t) = \{r_i(t), \theta_i(t), \phi_i(t), P_i^{TX}(t)\}. \tag{8}$$

$$a_{T_i}(t) = \{a_i^{pos}(t), a_i^{pow}(t)\} \in A_i^{pos} \times A_i^{pow} \tag{9}$$

C. Shared Reward

모든 에이전트는 공통의 목표를 가진다. 에이전트 간의 협력적 학습을 유도하기 위해 공유 보상을 사용하고 시간 t에서의 공유 보상은 아래와 같이 정의된다.

$$r(t) = \left(\sum_{i=1}^{N}\sum_{k=1}^{U_i}\gamma_{ik}\right) \times e^{-\frac{O_{ue}(t)}{U}} \tag{10}$$

이때, $O_{ue}(t)$는 시간 t에서의 서비스 불가 사용자 수를 나타낸다. 안정적인 학습을 위해 보상 값은 가능한 최대값으로 나누어 [0,1]로 정규화된다.

Ⅳ. 시뮬레이션 결과 및 결론

본 논문에서는 suburban 환경 6-에이전트 시나리오에서 시뮬레이션을 진행하였다. 시뮬레이션 파라미터는 표 1에 요약되어있다.

표 1. 시뮬레이션 파라미터

Parameter	Value
Bandwidth (B)	200 [KHz]
Noise Power (σ^2)	-120 [dBm]
Minimum/Maximum transmit power ($P_{\min}^{TX}, P_{\max}^{TX}$)	27, 33 [dBm]
Maximum movement speed ($v_{\max}$)	1 [m/s]
TUBS's Learning rate (α_T)	0.00005
TUBS's Discount factor (β_T)	0.95
Movable GS's Learning rate (α_G)	0.0003
Movable GS's Discount factor (β_G)	0.7

비교 방안으로는 Random Action(RA), Multi-Agent Distributed DQN considering only TUBSs(MADDQN-OT), Multi-Agent Distributed DQN considering only Movable-GSs(MADDQN-OG), Fixed Position Control(FPC), Fixed Transmission Power Allocation(FTA), Distributed Q-Learning(D-QL)을 고려한다.

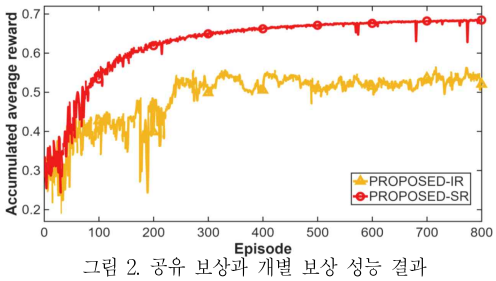

그림 2. 공유 보상과 개별 보상 성능 결과

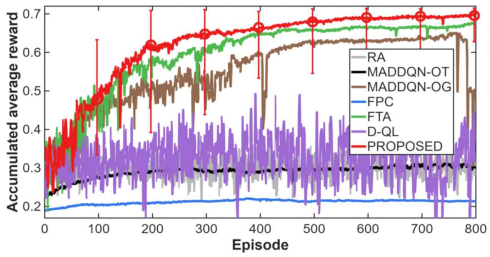

그림 3. 6-에이전트 환경에서의 학습 성능 결과

그림 2는 공유 보상(Shared Reward, SR)과 개별 보상(Individual Reward, IR)의 학습 성능을 비교한 결과로, 공유 보상이 더 빠른 수렴 속도와 높은 성능을 달성함을 확인할 수 있다. 이는 에이전트 간 협력적 학습이 네트워크 전체 성능 향상에 기여함을 의미한다. 그림 3은 제안 기법과 비교 방안들의 학습 성능을 비교한 결과를 나타낸다. 제안 기법은 학습 초기부터 빠르게 성능이 향상되며, 전 학습 구간에 걸쳐 안정적인 수렴성을 보인다. 이는 계층적 구조를 통해 행동 공간이 축소된 결과로 볼 수 있다. FTA 기법은 학습 초기에는 비교적 높은 보상을 달성하지만, 자원 할당을 고려하지 않는 한계로 인해 약 150 에피소드 이후부터는 제안 기법 대비 성능이 저하된다. FPC 기법은 전력 제어만을 고려한 방식으로, 위치 제어가 핵심적인 요소임을 보여준다. 또한, MADDQN-OT와 MADDQN-OG 기법의 비교 결과를 통해, Movable GS의 도입이 네트워크 커버리지 확장 및 전체 성능 향상에 유의미하게 기여함을 확인 할 수 있다. 한편, D-QL기법은 테이블 기반 학습의 한계로 보상 변동성이 크고 전반적으로 낮은 성능을 보이며, 복잡한 다중 에이전트 환경에서 적용하기에는 한계가 있음을 보여준다. 본 논문은 Movable GS와 TUBS를 결합한 네트워크 설계에 대한 새로운 가능성을 제시하며, 향후 테더의 무게에 따른 비행 제약을 고려한 현실적인 운용 환경으로 확장할 예정이다.

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단(No. RS-2025-02303435)과 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. RS-2024-00396992)과 2025년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2025-00563401)을 받아 수행된 연구임.

참 고 문 헌

[1] 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," Journal of Communications and Networks, vol. 25, no. 3, pp. 344 - 354, 2023.

[2] B. E. Y. Belmeguenai and M.-S. Alouini, "Unleashing the potential of networked tethered flying platforms: Prospects, challenges, and applications," IEEE Open Journal of Vehicular Technology, vol. 3, pp. 278 - 320, 2022.

[3] Y. Lee, H. Yu, H. Lee, and M.-S. Alouini, "D3QN-based IAB resource allocation and tethered UAV positioning for IoT networks," IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 5, pp. 6276 - 6287, 2025.

[4] Z. Gong and M. Haenggi, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," IEEE Trans. Mobile Comput., vol. 13, no. 2, pp. 337 - 349, Feb. 2014.

[5] "Propagation data and prediction methods for the design of terrestrial broadband millimetric radio access systems, radiowave propagation," Int. Telecommun. Union, Geneva, Switzerland, ITU-Recommendation P.1410-2, 2003.