도심 항공 모빌리티 네트워크에서의 강화학습 기반 무선 자원 할당 최적화

김준식, 이호원
아주대학교
{kjs17766, howon}@ajou.ac.kr

A Study on the DRL-Based Resource Allocation in UAM Networks

Kim Jun Sik, Howon Lee
Ajou Univ.

요 약

본 논문에서는 도심 항공 모빌리티(Urban Air Mobility, UAM) 환경에서 무선 자원 활용 효율을 높이기 위해 Proximal Policy Optimization(PPO) 기반의 무선 자원 할당 방식을 제안한다. 서브 채널 수와 전송 전력을 연속적으로 조절하는 구조를 통해 에너지 효율성과 지연 시간 측면에서 성능을 살펴본다. 시뮬레이션 환경에는 ITU-R 기반의 Air-to-Ground(A2G) 채널 모델과 로그 정규 새도잉을 고려하고, 기존 정책 기반 강화학습 방식 대비 우수한 성능을 가짐을 보인다.

Ⅰ. 서 론

도심 항공 모빌리티(Urban Air Mobility, UAM)는 차세대 교통수단으로 주목받고 있으며, 비행 중에도 안정적인 무선 통신을 유지하기 위한 지상과 비지상 네트워크 인프라 구축이 주목받고 있다. 특히 UAM 내부의 탑승자들이 이용하는 영상 스트리밍, 온라인 회의와 같은 고용량 서비스는 높은 품질의 통신을 요구하므로, 제한된 무선 자원을 효과적으로 사용하는 방법이 필요하다 [1]. 기존 자원 할당 방식은 변화하는 사용자 요구와 네트워크 환경에 유연하게 대응하기 어렵다. 이에 따라 본 논문에서는 UAM 기반 네트워크에서 사용자 요청의 특성과 시스템 상태 정보를 바탕으로, 서브 채널 수와 전송 전력을 동적으로 조절하는 PPO(Proximal Policy Optimization) 알고리즘 기반 무선 자원 할당 기법을 활용하여 에너지 효율성과 지연 시간을 최적화하고자 한다.

Ⅱ. 본론

본 논문에서는 단일 UAM이 고도 300m에서 비행하며, 고정된 기지국(Base Station, BS)과 통신하고 N명의 탑승자가 존재하는 환경을 가정한다. 통신은 OFDMA(Orthogonal Frequency Division Multiplexing Access) 기반으로, 코드 심볼 유지 시간(Δt) 동안 K개의 서브 채널을 사용한다. 또한 각 사용자 요청은 포아송 분포를 따라 일정한 요청을 생성하고, 생성된 요청은 큐에 저장되며, 강화 학습의 정책에 따라 서브 채널의 개수와 전송 전력을 할당한다.

Ⅱ-Ⅰ. 시스템 모델

UAM과 BS 간의 A2G(Air-to-Ground) 통신 링크 채널은 ITU-R (ITU Radiocommunication Sector)에서 제안한 도심 환경 채널 모델을 적용하며, LoS(Line-of-Sight) 및 NLoS(Non-Line-of-Sight) 확률에 따라 평균 경로 손실을 계산한다. LoS 확률은 수식 (1)로 계산한다.

$$P^{LoS} = \frac{1}{1+\alpha \cdot exp(-\beta(\theta-\alpha))} \tag{1}$$

여기서 α, β는 도심 환경 모델의 특징을 나타내는 파라미터를 의미하고 θ는 UAM과 기지국이 이루는 각도이다. LoS 채널에서의 경로 손실은 수식 (2)와 같이 계산한다.

$$L^{LoS} = 20\log\left(\frac{4\pi f_{a2g} d_{a2g}}{c}\right) + \xi^{LoS} \tag{2}$$

d_{a2g}는 기지국과 UAM이 이루는 3차원 거리를 의미하며 f_{a2g}, ξ^{LoS}는 각각 A2G 채널의 중심 주파수, LoS 채널의 추가적인 경로 손실을 의미한다. A2G 채널에서의 최종적인 경로 손실은 수식 (3)과 같이 LoS 및 NLoS 경로 손실의 가중합으로 계산된다 [2].

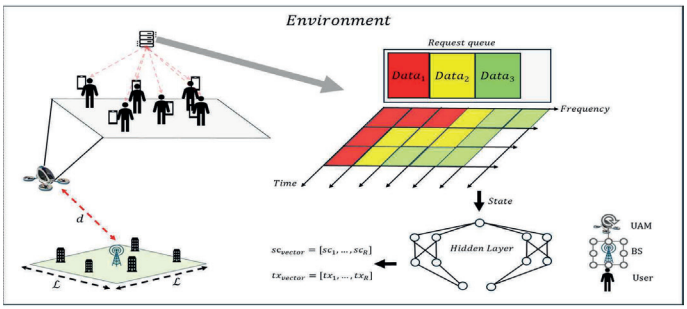

그림 1. UAM 환경에서의 자원 할당 시스템 모델

$$L_{a2g}^{avg} = P^{LoS} \cdot L^{LoS} + (1-P^{LoS}) \cdot L^{NLoS} \tag{3}$$

반면, UAM 내부 사용자와의 통신 링크는 자유 공간 손실(Free Space Path Loss, FSPL) 모델에 로그 정규 새도잉을 적용하여 구성한다. 그에 대한 수식은 (4)와 같다.

$$L_{infra} = 20\log_{10}\left(\frac{4\pi f_{infra} d_{infra}}{c}\right) + \psi_{dB} \tag{4}$$

여기서 f_{infra}는 UAM-사용자 채널에서의 중심 주파수, d_{infra}는 사용자의 거리를 의미한다. 이 채널에서 새도잉 모델은 $\psi_{dB} \sim N(0,8)$의 가우시안(Gaussian) 확률 분포를 따른다. 경로 손실을 통해 수신 전력, SINR(Signal-to-Interference Plus Noise Ratio)를 수식 (5)를 사용하여 계산하고, 이 값을 통해 데이터 전송률과 요청 i에 대한 전송 에너지, 그리고 에너지 효율성은 수식 (6)을 통해 확인할 수 있다. 지연 시간은 수식 (8)과 같이 모델링하였다.

$$\gamma_i = \frac{P_{r_i}}{I+N_0}, \quad P_{r_i} = P_{t_i} - L_i, \quad I = 0.1 \cdot \sum_{i=1, i\neq n}^{N} (P_{t_i} - L_i) \tag{5}$$

$$EE_i = \frac{C_i}{E_i}, \quad C_i = \frac{B}{K}\log_2(1+\gamma_i), \quad E_i = P_{t_i} \cdot \Delta t \tag{6}$$

$$\tau_{total} = \tau_{uam-bs} + \tau_{bs-uam} + \tau_{uam-user} \tag{7}$$

$$\tau_{uam-bs} = \tau_{bs-uam} = \Delta t, \quad \tau_{uam-user} = \frac{D_i}{C_i} \tag{8}$$

Ⅱ-Ⅱ. 강화학습 프레임워크

본 논문에서는 무선 자원 할당 문제를 마르코프 결정 과정(Markov Decision Process, MDP)으로 모델링하고, 이를 기반으로 PPO 알고리즘을 통해 최적의 자원 할당 정책을 학습한다. MDP는 상태, 행동, 보상 함수로 설계한다 [3].

1) State(상태) : $s_i = [D_i, t_i, t_{elapsed,i}, sc_{ratio}, E_{remain}, d_{n,x}, d_{n,y}]$ 으로 정의한다. UAM 내부 사용자 n의 요청 i에 대한 요청 데이터 크기(D_i), 마감 기한(t_i), 요청 도착 후 경과 시간($t_{elapsed,i}$), 전체 서브 채널 수 대비 사용 서브 채널 수(sc_{ratio}), UAM의 남은 배터리 에너지(E_{remain}) 그리고 사용자 n의 위치($d_{n,x}, d_{n,y}$)로 정의한다. 타임 슬롯마다 요청 큐에 존재하는 R개의 요청에 대해 매번 7차원 상태 벡터를 관측한다. 따라서 전체 상태는 $R \times |S|$ 행렬로 구성된다.

2) Action(행동) : 행동은 요청 큐에 존재하는 R개의 요청 각각에 대해 상태를 확인하고 서브 채널 비율 (α_i) 및 전송 전력 비율(β_i)을 연속적인 값으로 결정한다. 전체 행동은 다음과 같은 형태의 행렬로 출력된다.

$$A = \begin{bmatrix} \alpha_1, \beta_1 \\ \alpha_2, \beta_2 \\ \vdots \quad \vdots \\ \alpha_R, \beta_R \end{bmatrix}, \qquad (\alpha_i, \beta_i \in [0,1]) \tag{9}$$

이를 다시 정수 형태로 변환하여 다음의 자원 벡터를 생성한다.

$$sc_{vector} = [sc_1, \cdots, sc_R], \ \sum_{r=1}^{R} sc_r < K \tag{10}$$

$$tx_{vector} = [p_1, \cdots, p_R], \ P_{\min} \le p_r \le P_{\max} \tag{11}$$

sc_i는 요청 i에 할당되는 서브 채널의 수, p_i는 요청 i의 서브 채널에 할당되는 전송 전력을 나타낸다.

3) Reward(보상) : 각 요청 i에 대해 전송이 이루어진 후, 수식 (12)를 통해 보상이 계산된다.

$$r_i = \omega \cdot EE_i \times \exp(-0.5 \cdot \frac{\tau_{total}}{t_i}) \tag{12}$$

가중치를 통해 에너지 효율성과 지연 시간의 스케일을 맞추었고 지연 시간이 마감 기간보다 더 커지게 되면 보상이 줄어든다.

표 1. 시뮬레이션 / 강화학습 파라미터

Parameter	Value	Parameter	Value
UAM 내부 사용자 (N)	6	ω	10^{-6}
서브 채널 수(K)	15	ψ	-0.5
L	1000m	Episode	1000
전체 시간(T)	100s	Batch size	64
타임 슬롯(Δt)	0.1s	할인율	0.99
$P_{\min}$, $P_{\max}$	25, 30(dBm)	학습률	5×10^{-6}
f_{a2g}, f_{infra}	2.5, 5(GHz)	엔트로피 계수	0.005
B	80MHz	가치함수 계수	0.3
α, β	9.61, 0.16	클리핑 계수	0.3

Ⅲ. 시뮬레이션 결과 및 분석

그림 2는 제안하는 강화학습 프레임워크를 반복 학습하여 얻은 누적 보상에 대한 그래프이다. 초기에는 5.0에서 낮게 출발하였으나 에피소드가 200을 넘어가면서 약 17.0에 수렴한다. 또한 그림 3은 학습을 통해 진행한 에너지 효율성과 지연 시간에 대한 그래프이다. 각 지표는 에피소드 동안의 평균값을 기록한 것으로 점차 수렴하는 과정을 보여준다. 또한 서브 채널과 전력을 균등하게 할당하는 정책 기반의 환경보다 에너지 효율성과 지연 시간이 더 좋은 성능을 보인다.

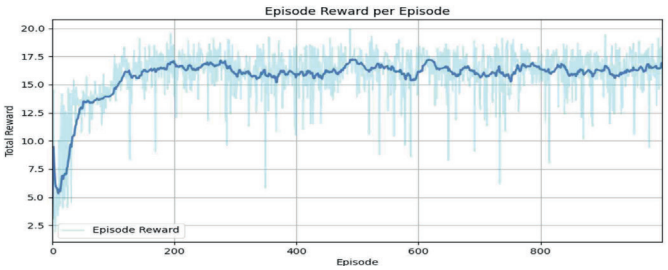

그림 2. 에피소드 당 평균 보상

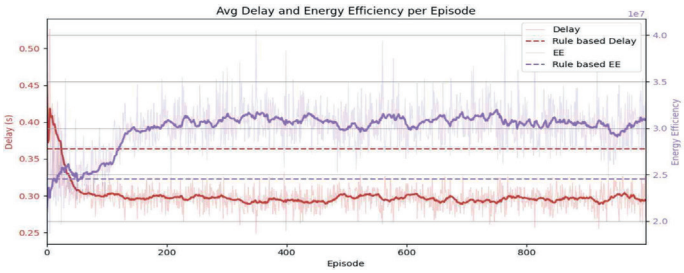

그림 3. 에피소드 당 평균 에너지 효율 및 지연 시간

Ⅳ. 결론

본 논문에서는 PPO 알고리즘을 사용하여 에너지 효율성과 지연 시간을 줄이고자 동적으로 자원을 할당하는 기법을 사용하여 기존 정책 기반의 방식보다 더 좋은 성능을 보이는 것을 확인하였다. 향후 연구에서는 QoS를 구현하여 보다 현실적인 네트워크 환경을 구성하고 기지국 간 핸드오버, 네트워크 간섭 등 여러 요소를 고려하여 구현하고자 한다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. RS-2024-00396992, 저궤도 위성통신 핵심 기술 기반 큐브위성 개발)과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. 2022-0-00704, 초공간 이동체 지원을 위한 3D-NET 핵심 기술 개발)과 2025년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2025-00563401, 3차원 공간에서 에너지 효율적 멀티레벨 AI-RAN 구현을 위한 AI-for/and-RAN 핵심 원천 기술 연구)을 받아 수행된 연구임.

참 고 문 헌

[1] 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, June 2023.

[2] S. Lee, H. Yu and H. Lee, "Multiagent Q-Learning-Based Multi-UAV Wireless Networks for Maximizing Energy Efficiency: Deployment and Power Control Strategy Design," in IEEE Internet of Things Journal, vol. 9, no. 9, pp. 6434-6442, 1 May1, 2022.

[3] X. Li, W. Zhou, H. Zhang, J. Zhao, D. Zhao and Z. Dong, "Joint Subcarrier and Power Allocation in Mobile Scenario of the OFDM Systems Based on Deep Reinforcement Learning," 2023 8th International Conference on Computer and Communication Systems (ICCCS), Guangzhou, China, 2023, pp. 209-214.