

다중 UAV 무선 네트워크를 위한 계층적 강화학습 프레임워크 연구

이승민¹, 임수현¹, 채승호², 정방철³, 박찬이⁴, 이호원¹

한경대학교¹, 한국산업기술대학교², 충남대학교³, 국방과학연구소⁴

{julsin1, 03070226, hwlee}@hknu.ac.kr¹, shchae@kpu.ac.kr², bcjung@cnu.ac.kr³, chyipark@add.re.kr⁴

Hierarchical Reinforcement Learning Framework for Multi-UAV Wireless Networks

Seungmin Lee¹, Suhyeon Lim¹, Seong Ho Chae², Bang Chul Jung³, Chan Yi Park⁴, and Howon Lee¹
Hankyong National Univ.¹, Korea Polytechnic Univ.², Chungnam National Univ.³, Agency for Defense Development⁴

요약

본 논문에서는 다중 unmanned aerial vehicle (UAV)의 제한된 배터리 제약 문제를 완화하는 동시에 전송 효율 최대화를 위하여 최적 주파수 재사용 계수와 전송 전력을 찾기 위한 계층적 다중 에이전트 강화학습 프레임워크를 제안한다. 제안 방안은, 분산형 내부 강화학습 루프에서 UAV의 전송 전력을 최적으로 제어하고, 중앙집중형 외부 강화학습 루프에서 최적 주파수 재사용 계수를 찾도록 설계되었다. 시뮬레이션을 통해 제안하는 계층적 다중 에이전트 큐러닝 알고리즘이 다중 UAV 무선 네트워크에서 네트워크 전체 에너지 효율을 최대화하는 최적 주파수 재사용 계수와 전송 전력을 찾을 수 있으며, 기존 방안 대비 높은 성능을 가지는 것을 보인다.

I. Introduction

6세대 (6G) 무선 네트워크에서 3차원 공간 무선 연결성 제공을 위하여 unmanned aerial vehicle (UAV)은 매우 중요한 기술들 중 하나로 고려되고 있다 [1]. 그러나, UAV는 신뢰성 있는 무선 연결성을 제공하는데 있어서 아직까지 많은 문제를 가지고 있으며, 특히 UAV의 제한된 배터리 수명으로 인해 그 동작시간이 급격히 감소한다는 문제를 가지고 있다. 따라서, 이를 해결하기 위한 많은 연구들이 매우 활발히 진행되고 있으며 [2], [3], 본 논문에서도 이를 위해 계층적 다중 에이전트 강화학습 기반 주파수 재사용 계수 및 전송 전력 제어 기법을 제안한다. 본 논문에서는 계층적 다중 에이전트 강화학습 프레임워크를 기반으로 주파수 재사용 계수와 전송 전력을 최적으로 조정함으로써 다중 UAV 무선 네트워크 에너지 효율을 최대화하는 것에 그 목적을 둔다.

II. 제안방안: 계층적 다중 에이전트 큐러닝 프레임워크

본 논문에서는 air-to-ground (A2G) 상향링크 다중 UAV 무선 네트워크에서, N_{gcs} Ground Control Systems (GCSs)와 N_{uav} UAVs가 존재하는 3D 셀 아키텍처를 고려한다. GCSs의 다운링크 RSSI 신호를 기반으로 UAVs는 주 GCS를 결정하며, 한번 설정된 주 GCS는 변하지 않는 것으로 가정한다. 주 기지국 결정 이후 UAVs는 주 GCS에게 상향링크 전송을 시도하며, GCS는 수신된 SINR을 이용하여 UAVs의 전송 전력을 제어하거나 모든 GCSs의 수신된 SINR 정보를 알고 있는 중앙 컨트롤러가 네트워크 주파수 재사용 계수 κ 를 조정하여 네트워크 에너지 효율을 극대화한다. 이를 위해, 본 논문에서는 UAV 전송 전력 제어를 위한 분산형 내부 강화학습 루프와 네트워크 주파수 재사용 계수 κ 조정을 위한 중앙집중형 외부 강화학습 루프로 학습을 계층화하여 그 목적을 달성하고자 한다.

II.1 채널 모델

본 논문에서는 확률적 Air-to-Ground (A2G) 채널모델을 기반으로 line-of-sight (LoS)와 non-LoS (NLoS) 신호를 개별적으로 고려한다 [4]. 해당 모델에서, UAV j 와 GCS i 사이의 SINR은 다음과 같이 표현된다.

$$\Gamma_{ij} = \sum_{k=1}^{N_{cf}} \left(\frac{P_{ij}^{Trans} \left(10^{\left(\frac{P_{ij}^{LoS} \times PL_{ij}^{LoS} + P_{ij}^{BLoS} \times PL_{ij}^{BLoS} \right)} \right)}{\sigma^2 + \sum_{m=1, m \neq j}^{N_{uav}} (P_{im}^{Rec} \times \sum_{n=1, n \neq i}^{N_{gcs}} I_t^{nm}(k))} \right) \times I_t^{ij}(k) \quad (4)$$

수식 (4)에서, $P_{ij}^{LoS(NLoS)}$ 는 UAV j 와 GCS i 사이의 LoS와 NLoS 확률이다. 또한, $PL_{ij}^{LoS(NLoS)}$ 는 LoS와 NLoS 링크에 대한 경로 손실을 나타내며, $P_{ij}^{Trans(Rec)}$ 은 UAV j 의 송신전력과 GCS i 의 수신 전력을 나타내고, N_{cf} 는 한 셀에서 이용가능한

주파수 자원의 수이다. σ^2 은 열 잡음 전력을 나타내며, $I_t^{ij}(k)$ 는 UAV j 가 GCS i 의 k 번째 주파수 자원에 할당 되었는지 아닌지를 나타내는 지시 함수를 의미한다.

II.2 분산형 내부 강화학습 루프 (Distributed inner-loop RL)

내부 강화학습 루프에서는 UAV 에너지 효율을 최대화하기 위한 최적 전송 전력을 찾기 위해, 각 GCS를 내부 강화학습 루프의 에이전트로 고려하였다.

■ 내부 강화학습 루프 상태

타임 스텝 t 에서, 주파수 재사용 계수가 κ 일 때, GCS i 의 내부 강화학습 루프의 상태 집합 ($S_t^{IN, (i, \kappa)}(k)$)은 각 주파수 자원의 SINR 수준과 $|S_t^{IN, (i, \kappa)}(k)| = N_{cf}$ 에 의해 정의되며, $S_t^{IN, (i, \kappa)}(k)$ 의 요소는 다음과 같이 표현된다.

$$S_t^{IN, (i, \kappa)}(k) = \begin{cases} \text{None}, & \text{assigned}, \\ 0, & \Gamma_{ij}(t) \leq \Gamma_1, \\ 1, & \Gamma_1 < \Gamma_{ij}(t) \leq \Gamma_2 \\ \vdots & \\ N_{IL-1}, & \Gamma_{N_{IL-1}} < \Gamma_{ij}(t) \leq \Gamma_{N_{IL}} \end{cases} \quad (5)$$

수식 (5)에서, “None”은 k 번째 주파수 자원이 이용되고 있지 않음을 의미하며, N_{IL} 은 내부 루프의 SINR 수준의 수를, $\Gamma_n (n \in \mathbb{N}_{IL})$ 은 내부 루프의 SINR 수준 n 의 임계값을 나타낸다.

■ 내부 강화학습 루프 행동

타임 스텝 t 에서, 주파수 재사용 계수가 κ 일 때, k 번째 주파수 자원을 사용하는 UAV j 의 전송 전력 제어를 위한 GCS i 의 행동은 다음과 같다.

$$A_t^{in, (i, \kappa)}(j, k) \in \{\Delta_P^{Trans}, -\Delta_P^{Trans}, 0\} \quad (6)$$

수식 (6)에서, $\Delta_P^{Trans}, -\Delta_P^{Trans}$, 그리고 “0”은 각각 “전송 전력 증가”, “전송 전력 감소”, “전송 전력 유지”를 나타낸다.

■ 내부 강화학습 루프 보상

제안된 계층적 다중 에이전트 큐러닝 알고리즘은 다중 UAV 무선 네트워크의 에너지 효율을 최대화하는 것을 목표로 한다. 따라서, 타임 스텝 t 에서, 주파수 재사용 계수가 κ 일 때, GCS i 의 에너지 효율을 이용하여 다음과 같이 보상을 정의한다.

$$R_t = \sum_i \frac{\sum_j \sum_k^{N_{gcs} N_{cf}} (BW_{\kappa} \log_2(1 + \Gamma_{ij}(t)) \times I_t^{ij}(k))}{\sum_j \sum_k^{N_{uav} N_{cf}} ((P_j^{CP} + P_{ij}^{Trans}) \times I_t^{ij}(k))} \quad (7)$$

수식 (7)에서, BW_{κ} 는 주파수 재사용 계수가 κ 일 때 각 주파수 자원의 대역폭 크기를 의미하며, P_j^{CP} 는 UAV j 의 고정 소비 전력을 의미한다.

II.3 중앙집중형 외부 강화학습 루프 (Centralized outer-loop RL)

외부 강화학습 루프에서는 최적의 주파수 재사용 계수 κ 를 찾기 위해, 모든 GCSs를 관리하는 중앙 컨트롤러를 외부 강화학습 루프의 에이전트로 고려하였다.

■ 외부 강화학습 루프 상태

타임 스텝 t 에서, 주파수 재사용 계수가 κ 일 때, GCS i 의 외부 강화학습 루프의 상태 집합 (S_t^{OUT})은 각 GCS에게 할당된 모든 UAVs의 평균 SINR 값과 $|S_t^{OUT}| = N_{gcs}$ 에 의해 정의되며, S_t^{OUT} 의 요소는 다음과 같이 표현된다.

$$S_t^{OUT}(i) = \begin{cases} 0, & \Gamma_i^{avg}(t) \leq \Gamma_1, \\ 1, & \Gamma_1 < \Gamma_i^{avg}(t) \leq \Gamma_2, \\ \vdots & \\ N_{OL}, & \Gamma_{N_{OL}-1} < \Gamma_i^{avg}(t) \leq \Gamma_{N_{OL}} \end{cases} \quad (9)$$

$$\Gamma_i^{avg} = \sum_j^{N_{nav}} (\Gamma_{ij}(t)) \div \sum_j^{N_{nav}} \sum_k^{N_f} (\Gamma_{ij}^k(k)) \quad (10)$$

수식 (9)에서, $\Gamma_i^{avg}(t)$ 는 GCS i 의 평균 SINR을 의미하며, N_{OL} 는 외부 루프의 SINR 수준의 수를, Γ_m ($m \in N_{OL}$)은 외부 루프의 SINR 수준 m 의 임계값을 나타낸다.

■ 외부 강화학습 루프 행동

외부 강화학습 루프에서, 주파수 재사용 계수 κ 를 조정하는 중앙 컨트롤러의 행동은 다음과 같이 나타낼 수 있다.

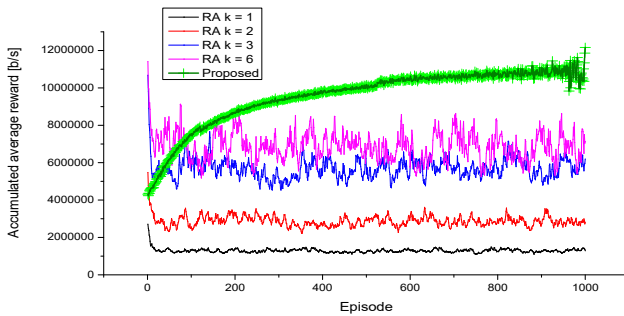
$$A_t^{out} \in \{\Delta_\kappa, -\Delta_\kappa, 0\} \quad (11)$$

수식 (11)에서, Δ_κ , $-\Delta_\kappa$, 그리고 "0"은 각각 " κ 증가", " κ 감소", " κ 유지"를 의미한다.

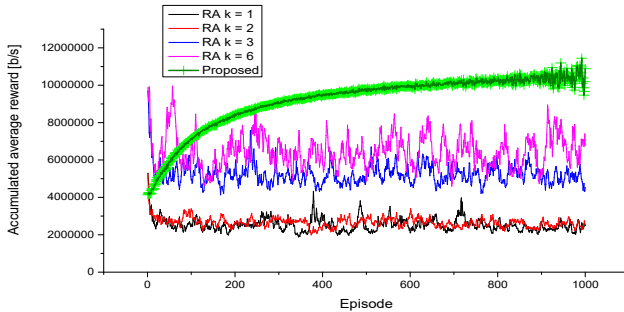
■ 외부 강화학습 루프 보상

외부 강화학습 루프의 보상은 내부 루프의 보상과 동일하게 설정하였다.

III. 시뮬레이션 결과 및 결론



(a) GCS 간 거리 $D^g = 100(m)$, 셀 반경 $R^u = 50(m)$



(b) GCS 간 거리 $D^g = 100(m)$, 셀 반경 $R^u = 80(m)$

그림 1. D^g , R^u 변화에 따른 제안 방안 및 무작위 행동 방안의 누적 평균 보상 결과

표 1. 시뮬레이션 파라미터 요약

시뮬레이션 파라미터	값
주파수 재사용 계수 (κ)	{1,2,3,6}
GCS의 수 (N_{gcs})	12
UAV의 수 (N_{uav})	72
반송 주파수 (f_c)	1 (GHz)
대역폭 (BW_k)	200 (kHz)
초과 경로 손실 (PL^{LoS}, PL^{NLoS})	1, 20 (dB)
열 잡음 전력 (σ^2)	-120 (dBm)
내부 루프 SINR 수준 개수 (N_{IL})	3
외부 루프 SINR 수준 개수 (N_{OL})	2
UAV 최대, 최소 전송 전력 ($P_{min}^{Tran}, P_{max}^{Tran}$)	0.5, 2.0 (W)
UAV 고정 소비 전력 (P^{CP})	0.1 (W)
전송전력 증가량 (ΔP^{Tran})	0.5 (W)

본 연구에서 설계된 분산형 내부 강화학습 루프의 최적 솔루션에 대한

경우의 수는 $\left(\frac{P_{max}^{Tran} + P_{min}^{Tran}}{\Delta P^{Trans}} \right)^{N_{cf} \times N_{gcs}}$ 만큼 크기 때문에, exhaustive

search를 통한 최적 해를 찾는 것은 불가능하다. 따라서, 잠재적으로 최적 성능을 달성할 가능성을 가지고 있는 무작위 행동 방안을 이용하여 제안 방안과의 성능을 비교 분석해 본다. 본 논문에서, 무작위 행동 방안은 고정된 주파수 재사용 계수에 대하여 UAV의 전송 전력을 무작위로 제어하였다. 또한, 제안 방안의 성능 분석을 위해, GCS간 거리 D^g 와 셀 반경 R^u 의 변화에 따른 누적 평균 에너지 효율 결과를 도출하였다.

(D^g, R^u)가 (100,50)(m)와 (100,80)(m)으로 설정된 환경에서, 1000번의 에피소드와 에피소드당 200,000번의 반복으로 시뮬레이션이 진행되었다. 그림 1은 셀 반경 및 GCS 간 거리 변화에 대해서 에피소드 진행에 따른 각 방안의 누적 평균 보상을 나타낸다. 그림 1-(a)와 그림 1-(b)로부터, 제안 방안은 주파수 재사용 계수와 UAV의 전송 전력을 체계적으로 제어함으로써, 여러 무작위 행동 방안의 최대 성능에 수렴함을 확인할 수 있다. 또한, 두 그림으로부터 고정된 D^g 에서 R^u 가 증가할수록, 인접 기지국에 존재하는 UAV들의 간섭 신호의 세기가 감소하기 때문에, 네트워크 에너지 효율이 전반적으로 상승함을 알 수 있다.

ACKNOWLEDGMENT

이 연구는 국방과학연구소(ADD)의 지원을 받았습니다.

참고 문헌

- [1] G. Flagship, "Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence," 6G Research Visions 1, pp. 1-36, Sept. 2019.
- [2] S. Lee, H. Yu, and H. Lee, "Multi-Agent Q-Learning Based Multi-UAV Wireless Networks for Maximizing Energy Efficiency: Deployment and Power Control Strategy Design," IEEE Internet of Things Journal, pp. 1-1, 2021.
- [3] 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. 45504-45511, 2021.
- [4] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," IEEE Wireless Communications Letters, vol. 3, no. 6, pp. 569-572, Dec. 2014.