

다중 에이전트 계층적 강화 학습을 활용한 군집 UAV 기반 ISAC 시스템 자원할당 최적화 기법

최재혁, 이호원

아주대학교 전자공학과

jpeace321@ajou.ac.kr, howon@ajou.ac.kr

Resource Allocation Optimization in UAV Swarm-Based ISAC Systems Using Multi-Agent Hierarchical Reinforcement Learning

Jae Hyuk Choi, Howon Lee

Department of Electrical and Computer Engineering, Ajou University

요약

본 논문에서는 군집 unmanned aerial vehicle(UAV) 기반 통합 센싱 및 통신(integrated sensing and communication : ISAC) 시스템의 자원할당 최적화 문제를 multi-agent hierarchical reinforcement learning(MA-HRL)을 활용하여 해결하고자 한다. 군집 UAV 시스템에서 타겟 탐지와 UAV 간 통신은 필수적인 기능이다. ISAC은 같은 하드웨어와 자원 속에서 센싱과 통신을 통합함으로써 군집 UAV 시스템의 효율적 자원 활용과 성능 극대화의 해답이 될 수 있다. 군집 UAV 기반 ISAC 시스템에서의 자원할당은 센싱과 통신 성능 간 최적화와 각 UAV 사이의 최적화가 이루어져야 하므로 매우 복잡한 문제를 형성한다. 본 논문에서는 MA-HRL을 이용하여 communication sum rate(CR)와 radar quality(SINR), 그리고 energy efficiency(EE)를 최대화하고자 한다.

I. 서론

Unmanned aerial vehicle(UAV)를 활용한 통신 네트워크는 6G의 다양한 확장성을 보여준다[1]. 통합 센싱 및 통신(integrated sensing and communication : ISAC)은 한정된 자원에서 센싱과 통신의 상호보완적 기능으로 효율적인 시스템을 제공한다. ISAC 시스템에서 군집 UAV는 커버리지 확장 및 line of sight(LoS) 링크 확보가 가능하여 ISAC 시스템에 풍부한 확장성을 가져온다[2]. 하지만, 군집 UAV 기반 ISAC 시스템은 여러 대의 UAV 간 신호 간섭 제어, 자원할당과 같은 복잡한 문제들이 존재한다[3]. 본 논문은 Reinforcement Learning(RL)을 이용한 자원할당 최적화를 통해 communication sum rate(CR), radar quality(SINR), 그리고 energy efficiency(EE)를 최대화하고자 한다. 이때, Q-learning(QL)으로 연산할 수 없는 방대한 action set을 고려하기 위해 hierarchical RL(HRL)을 활용하여 그 문제를 해결하고자 한다.

II. 시스템 모델

본 논문에서는 Fig. 1과 같은 군집 UAV 기반 ISAC 시스템을 고려한다. 공통된 목적지를 갖고 비행하는 군용 군집 UAV는 지속적인 실시간 UAV 탐지와 원활한 정보 통신이 필수적이다. UAV들은 mmWave 빔포밍으로 서로에게 통신하고 target을 탐지한다. 이때, UAV들을 개별 agent로 정의하며 각 agent들은 beam, channel, power를 어떻게 활용해야 통신과 탐지 성능을 동시에 향상하고 에너지 효율을 극대화할 수 있는지 행동과 보상의 관계를 학습한다. 이때, beam, channel, power의 자원할당은 경우의 수가 너무 많다. 따라서, Fig. 1과 같이 HRL을 활용하여 방대한 행동 집합에 대한 문제를 해결하고자 한다.

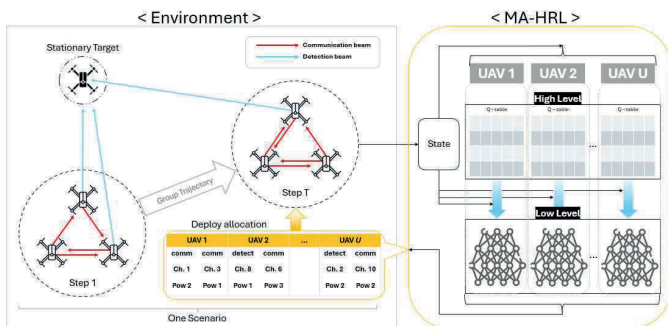


Fig. 1 제안하는 MA-HRL 프레임워크

A. 통신 모델

UAV 간 통신은 free space와 rician fading으로 정의되며 다음과 같다.

$$h_c = PL_c \times RF(K_c) \quad (3)$$

$$PL_c = \frac{\sqrt{G_T G_R} \lambda}{4\pi d} \quad (4)$$

$$RF(K_c) = \sqrt{\frac{K_c}{1+K_c}} LoS + \sqrt{\frac{1}{2(K_c+1)}} NLoS \quad (5)$$

여기서 G_T , G_R 은 각각 송/수신안테나 gain이다. LoS가 지배적인 A2A 통신이기 때문에 $RF(K_c)$ 와 같이 rician fading 모델을 이용한다. 이에 따라서 통신 신호의 SINR은 아래와 같이 계산할 수 있다.

$$SINR_i^c = \frac{|h_c| p_{i,j}}{I + \alpha \hat{T}(BW)} \quad (6)$$

여기서 I 는 같은 채널을 사용하는 빔들에 의한 간섭 신호의 세기, α 는 볼츠만 상수, \hat{T} 는 잡음 온도, BW 은 한 채널의 대역폭이다. 위의 SINR로 achievable communication rate를 계산하면 아래와 같다.

$$CR_i^c(SINR_i^c) = BW \log_2(1 + SINR_i^c) \quad (7)$$

B. 센싱 모델

탐지 성능을 계산하기 위한 센싱 모델은 다음과 같다.

$$h_s = PL_s \times RF(K_s) \quad (8)$$

$$PL_s = \sqrt{\frac{G_T G_R \eta \lambda^2}{(4\pi)^3 d^4}} \quad (9)$$

UAV와 target 사이의 탐지 전파 경로손실은 (9)와 같다. 여기서 η 는 radar cross section이다. 센싱 모델도 rician fading 모델을 이용한다. 이에 따라서 탐지 신호의 SINR은 아래와 같이 정의된다.

$$SINR_i^s = \frac{|h_s| p_{i,j}}{I + \alpha \hat{T}(BW)} \quad (10)$$

C. Problem Formulation

위의 모델 정의에 따라 문제를 아래와 같이 정의한다.

$$P1 : \max_{b, c, p} \sum_{i=1}^N R_u(B_i)^{[i]} \quad \forall t \in T \quad (11)$$

$$s.t. \quad n(A_u[B_i[b_i = detect]])^{[i]} \geq 1 \quad \forall t \in T \quad (12)$$

$$A_u[n(B_i[b_i = comm])] \geq 1 \quad \forall u \in U \quad (13)$$

$$A_u[c^{[t]}] \neq A_u[c^{[t+1]}] \quad \forall t \in T \quad (14)$$

t 는 state step을 의미하고 T 는 한 시나리오의 step 집합이다. (12)은 각 step에서 탐지를 선택한 beam이 하나 이상 있어야 함을 의미한다. (13)는 각 agent가 최소 1회 통신을 수행해야 함을 의미한다. (14)는 각 agent는 이전 step에서 사용한 채널을 다음 스텝에서 사용할 수 없음을 나타낸다. P1을 해결하기 위해 multi-agent HRL (MA-HRL)을 이용한 기법을 탐구하고 multi-agent QL(MA-QL)과 그 성능을 비교한다. MDP는 아래와 같다.

1) *State space* : 시스템의 상태는 UAV들의 위치/속도, target의 위치, state step 번호를 갖는다.

$$s^{[t]} = [u^{[t]}, v^{[t]}, e^{[t]}, s^{\#}] \quad (1 \leq \# \leq T) \quad (15)$$

2) *Action space* : 각 agent들의 행동 A_u 는 beam들이 사용되는 task, channel, power의 조합 $B_i[b_i, c_{i,k}, p_{i,j}]$ 를 선택하는 것이다.

$$A_u[B_1[b_1, c_{1,k}, p_{1,j}], \dots, B_i[b_i, c_{i,k}, p_{i,j}]] \quad (16)$$

$$(u = 1 \sim U, i = 1 \sim N, b = \begin{cases} comm \\ detect \end{cases}, k = 1 \sim M, j = 1 \sim P)$$

A_u 는 u 번째 UAV agent의 행동 집합이고 U 는 agent의 개수다. B 는 agent가 자원을 할당할 beam을 의미하고 b 는 beam이 어떤 task(통신 또는 탐지)에 활용될지를 나타낸다. 각 agent는 N 개의 beam, M 개의 channel, P 개의 power level 중에서 자원을 선택한다.

3) *Reward space* : 보상은 아래와 같이 상위 보상(R^{High})과 하위 보상(R^{Low})로 구분된다.

$$R_u = \sum_{i=1}^N (R_i^{High} + R_i^{Low}) \quad (17)$$

$$R_i^{High} = CR_i^c + SINR_i^s - \text{penalty}(\text{Bad Count}), \quad R_i^{Low} = EE_i$$

상위 보상으로는 CR과 SINR의 합과 패널티를 부여하고, EE는 하위 보상으로 이용한다. EE는 다음과 같이 계산한다.

$$EE_i = CR_i (SINR_i^b) \quad b = \begin{cases} comm \\ detect \end{cases} \quad (18)$$

패널티의 경우 (12), (13)의 제약을 위반하는 횟수를 *Bad count*로 세며 상위 보상을 그 횟수만큼 감소시킨다.

본 논문에서는 Fig. 1과 같이 계층적 선택 구조로 나눠 MA-HRL을 구성한다. 상위 계층은 행동($b_i, c_{i,k}$)을 정하고(Q-table), 상위 행동은 하위 상태로 적용되며 하위 계층에는 $p_{i,j}$ 을 결정한다(Deep Q-Network). 세 자원 중 $p_{i,j}$ 를 하위로 분리함으로써 더 높은 빈도의 선택 경험으로 EE를 빠르게 최대화할 수 있고, DQN을 적용하여 다른 agent의 행동보다는 자신의 행동과 보상 간의 관계를 효과적으로 학습할 수 있다.

IV. 시뮬레이션 결과

Table. 1 시뮬레이션 파라미터

Parameter	Value
$UAV(U), Beam(N), Channel(M), Power(P), Step(T)$	3, 2, 10, 3, 5
$Carrier\ Freq.(f_c^1 \sim f_c^M), BW$	(28 ~ 29GHz), 100MHz
x, \hat{T}	1.38×10^{-23} , 290K
$\eta, K_c, K_s, K_{interference}$	$1m^2$, 30, 15, 5
$G_T, G_R, G_{sidelobe}$	10dBi, 10dBi, -30dBi
$power\ level(1, 2, 3)$	1W, 2W, 3W
$penalty$	100
$(episode, max\ step)$	(8000, 500)

MA-QL은 Fig. 2-(b)의 경우 행동공간이 455,000개가 되어 학습이 동작하지 않았다. 따라서, episode와 max step을 (8000, 10)으로 진행하였다.

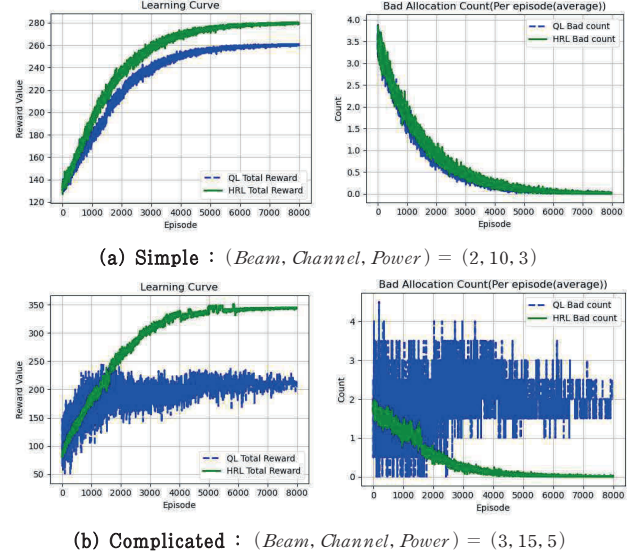


Fig. 2 시뮬레이션 결과

시뮬레이션 결과, 단순한 상태 Fig. 2-(a)에서는 MA-HRL이 MA-QL보다 최종 도달 보상이 우세하나 전반적으로 비슷한 성능을 보인다. 하지만, 복잡한 상태 Fig. 2-(b)에서는 MA-HRL이 MA-QL보다 보상 및 학습 안정성 측면에서 모두 우월함을 확인할 수 있다.

V. 결론

본 논문에서는 군집 UAV 기반 ISAC 시스템에서의 MA-HRL을 활용한 자원할당 최적화 기법을 탐구하였다. 시뮬레이션의 결과를 통해서 MA-HRL을 이용한 행동공간의 계층화는 증가하는 행동공간에 대한 대안이 될 수 있음을 확인했다. 하지만, 상위 계층에서 여전히 Q-table을 활용하고 있기에 더욱 복잡한 상황에서는 성능개선의 한계가 있음을 예상해 볼 수 있다. 따라서, MA-HRL의 한계와 문제점을 검토하고 개선하여 동적인 시스템에 대한 자원할당 최적화 문제를 해결할 수 있는 연구로 발전시킬 수 있을 것이다.

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 멀티테이블 AI-RAN 구현을 위한 AI-for/and-RAN 핵심 원천기술 연구)을 받아 수행된 연구임.

참고 문헌

- [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] K. Meng et al., "UAV-Enabled Integrated Sensing and Communication: Opportunities and Challenges," in IEEE Wireless Communications, vol. 31, no. 2, pp. 97-104, April 2024
- [3] Wang M, Chen P, Cao Z, Chen Y. "Reinforcement Learning-Based UAVs Resource Allocation for Integrated Sensing and Communication (ISAC) System." Electronics. 2022