

# 다중 빔 저궤도 위성에서 다중 에이전트 심층 강화 학습을 활용한 핸드오버 최소화 기법

이 충녕<sup>\*</sup>, 김태훈<sup>\*</sup>, 방인규<sup>\*\*</sup>, 채승호<sup>○</sup>

## Handover Minimization Scheme Using Multi-Agent Deep Reinforcement Learning in Multi-Beam Low Earth Orbit Satellites

Chungnyeong Lee<sup>\*</sup>, Taehoon Kim<sup>\*</sup>, Inkyu Bang<sup>\*\*</sup>, Seong Ho Chae<sup>○</sup>

### 요약

본 논문에서는 중앙 집중 학습 및 분산 실행(CTDE, Centralized Training Decentralized Execution) 방식의 다중 에이전트 심층 강화 학습(MADRL, Multi-Agent Deep Reinforcement Learning) 알고리즘<sup>※</sup> Multi-Agent Proximal Policy Optimization,(MAPPO) 기반 다중 빔 저궤도 위성 핸드오버 전략을 제안한다. 제안된 방법은 빔 간 및 위성 간 핸드오버 발생 시의 비용 차이, 사용자 서비스 품질 제약 조건, 그리고 부하 균형을 고려하여 핸드오버 횟수를 최소화하고 처리율을 최대화하는 것을 목표로 한다. 각 사용자는 국소 정보(커버리지 내 부하 수 및 채널 정보)를 바탕으로 독립적으로 핸드오버를 결정하며, 이는 복잡하고 동적으로 변하는 다중 빔 저궤도 위성 환경에 즉각적인 대응을 가능하게 한다. 시뮬레이션 결과, 제안된 알고리즘은 기존 핸드오버 알고리즘에 비해 핸드오버 횟수를 39.1%에서 75.53%까지 감소시키고, 처리율을 14.6%에서 157.7%까지 향상시켰음을 보임으로써, 제안 알고리즘의 우수한 성능을 입증하였다.

**키워드 :** 저궤도 위성, 핸드오버 전략, 다중 빔, 다중 에이전트 심층 강화학습

**Key Words :** Low earth orbit satellite, handover strategy, multi-beam, multi-agent deep reinforcement learning

### ABSTRACT

In this paper, we propose a Multi-Agent Proximal Policy Optimization (MAPPO)-based handover strategy for multi-beam Low Earth Orbit (LEO) satellite networks, employing the Centralized Training and Decentralized Execution (CTDE) approach of Multi-Agent Deep Reinforcement Learning (MADRL). The proposed strategy aims to minimize the number of handovers and maximize throughput by considering the cost differences between inter-beam and inter-satellite handovers, user quality of service (QoS) constraints, and load balancing. Each user independently makes handover decisions based on local information (e.g., load and channel conditions

\* 이 논문은 2022년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원(KRIT-CT-22-047, 우주계층 지능통신망 특화연구실)과 2025년 국립한밭대학교 학술문화연구재단 연구비 지원을 받아 수행된 연구입니다.

\*\* 이 논문의 일부는 한국통신학회 2024년도 추계종합학술발표회('24.11.20.~'24.11.22.)에서 발표되었습니다.

◆ First Author : Tech University of Korea Department of IT Semiconductor Convergence Engineering, lc9902130509@tukorea.ac.kr, 학생회원

○ Corresponding Author : Tech University of Korea Department of Electronics Engineering, shchae@tukorea.ac.kr, 종신회원

\* Hanbat National University Department of Intelligence Media Engineering, ikbang@hanbat.ac.kr, 종신회원

\*\* Hanbat National University Department of Computer Engineering, thkim@hanbat.ac.kr, 정회원

논문번호 : 202502-034-B-RE, Received February 14, 2025; Revised March 6, 2025; Accepted March 11, 2025

within the coverage area), allowing for prompt immediate adaptation to the dynamic and complex environment of multi-beam LEO satellite networks. Simulation results indicate that the proposed algorithm reduces the number of handovers by 39.1% to 75.53% and improves throughput by 14.6% to 157.7% compared to benchmark handover algorithms, thereby objectively demonstrating the superior performance of the proposed approach.

## I. 서 론

최근 저궤도 위성 네트워크는 전 세계적인 커버리지(Coverage)와 끊김 없는 연결을 제공하는 역할로 많은 관심을 받고 있다<sup>[1,2]</sup>. 저궤도 위성은 낮은 전파 지연 시간과 낮은 에너지 소비와 같은 여러 이점을 제공할 수 있지만, 높은 공전 속도는 지상 사용자가 서비스 받는 동안 빈번한 핸드오버를 불가피하게 발생한다<sup>[3]</sup>. 특히, 다중 빔 저궤도 위성 환경에서는 사용자에게 여러 빔이 빠르게 지나감에 따라 매우 빈번한 빔 간 또는 위성 간 핸드오버가 발생한다<sup>[4]</sup>. 따라서, 빔 간 또는 위성 간 핸드오버 발생 비용은 서로 다르므로, 핸드오버 빈도뿐만 아니라 그 비용을 최적화하는 것이 중요하다<sup>[5]</sup>. 핸드오버 발생 횟수의 증가는 네트워크 오버헤드(Overhead)가 증가시켜 서비스 품질이 저하되는 문제를 야기할 수 있다. 따라서, 이러한 비용을 최소화하는 다중 빔 저궤도 위성을 위한 핸드오버 전략의 설계가 필요하다.

최근, 저궤도 위성 네트워크에 대해, 다양한 기준과 제약 조건을 가진 핸드오버 기법들이 제안되었다<sup>[6-8]</sup>. 먼저, 네트워크 흐름 그래프(Network-Flows Graph) 기반의 핸드오버 전략이 저궤도 위성 네트워크를 위해 제안되었다<sup>[6]</sup>. 여기서, 위성과 사용자 단말 간의 매칭은 제한된 용량과 최대 흐름 조건 하에서 그래프의 비용과 흐름의 곱을 최소화함으로써 결정되었다. Wu 등은 저궤도 위성 핸드오버를 위한 이분 그래프(Bipartite Graph) 프레임워크와 잠재적 게임(Potential Game) 이론 기반 핸드오버 알고리즘을 제안하여, 저궤도 위성 네트워크에서 모바일 단말기의 이득을 극대화하였다<sup>[7]</sup>. 대규모 사용자 단말을 위한 그룹 핸드오버 전략이 저궤도 위성 네트워크에서 제안되었다<sup>[8]</sup>. 그러나, 제안된 핸드오버 전략<sup>[6]-[8]</sup>들은 중앙화된(Centralized) 방식으로 중앙 제어 장치와 글로벌(Global) 정보를 필요로 한다. 이는, 고밀도 저궤도 위성 네트워크에서 더욱 큰 신호 오버헤드(Overhead)를 초래한다.

이와 같은 중앙 집중 방식의 한계를 완화하고자, 저궤도 위성 네트워크에서 분산화(Distributed)된 핸드오버 방식이 제안되었다<sup>[9,10]</sup>. 다중 에이전트(Agent) 강화

학습 기반 저궤도 위성 핸드오버 전략은 각 위성의 부하(Load) 제약 조건을 만족시키면서 평균 핸드오버 횟수를 최소화하는 방식으로 제안되었다<sup>[9]</sup>. 연속 심층 Q-학습( Successive Deep Q-Learning) 알고리즘 기반으로 한 분산형 저궤도 위성 핸드오버 방식이 제안되었는데<sup>[10]</sup>, 이는 각 사용자가 핸드오버가 발생할 때, 핸드오버 지역, 실패, 서비스 품질 및 위성 간 트래픽 균형을 고려하여 지역 정보에 기반한 핸드오버 결정을 수행할 수 있도록 한다.

앞서 제안된 중앙화 또는 분산화된 위성 핸드오버 전략<sup>[6-10]</sup>들은 위성 핸드오버만 고려하였으며, 빔 간 핸드오버는 고려하지 않았다. 그러나 실질적이고 정교한 저궤도 위성 네트워크에서의 핸드오버 설계를 위해서는 위성 간 또는 빔 간의 핸드오버 비용 차이를 고려해야 한다.

최근, 위성 간 또는 빔 간의 핸드오버를 고려한 연구에 대한 관심이 증가하고 있으나, 현재 연구 태동 단계에 머물러있다. 다층(Multi-Layer) 저궤도 위성 시스템에서 중앙 핸드오버 관리자가 예측된 채널 상태와 위성 위치 정보를 활용하여 핸드오버 시점을 결정하는 방식을 제시하였으나<sup>[9]</sup>, 통신 거리가 긴 위성 통신 환경에서는 글로벌 정보를 수집하는 중앙 관리자를 통한 핸드오버 기법이 급변하는 위성 환경에 실시간으로 대응하기 어렵다는 중대한 한계를 내포하고 있다. 또한, 해당 연구는 빔 간과 위성 간 핸드오버 기법을 통합적으로 다루지 않고 별도로 제시하였다. 반면, 위성 간 및 빔 간 핸드오버를 동시에 고려하는 핸드오버 기법이 제안되었으며<sup>[11]</sup>, 해당 방식은 분산화 및 다중 목적 최적화 기법을 활용하여 처리율, 핸드오버 비용, 부하 균형을 각각의 최적화 목표로 설정하였다. 그러나 해당 방식은 각 목표에 대해 개별적인 Q-네트워크를 구축해야 하고, 각 사용자가 세 개의 Q-네트워크를 유지하고 조절해야 하는 등 연산 복잡성이 크게 증가하고 최적화 성능 면에서 한계가 있다. 또한, 다중 빔 환경의 채널 특성을 정밀하게 반영되지 않았다.

따라서, 본 논문에서는 매우 복잡하고 역동적인 다중 빔 저궤도 위성 네트워크에서 발생하는 핸드오버 문제를 효과적으로 해결하기 위해, 연산 복잡성을 최소화하

면서도 탁월한 성능과 실시간 적응 능력을 갖춘 중앙 집중 학습 및 분산 실행 (*CTDE, Centralized Training Decentralized Execution*) 알고리즘인 MAPPO (Multi-Agent Proximal Policy Optimization) 기반 핸드오버 기법을 제안한다. 본 알고리즘은 학습 단계에서 중앙 집중식으로 전역 정보를 활용하여 안정적인 정책 업데이트를 구현하고, 중앙 비평가와 클리핑 기법을 활용하여 환경의 비정상(Non-Stationary) 문제에 효과적으로 대응한다. 또한, 실행 단계에서는 분산 방식으로 작동하여 신속한 처리 및 확장성을 달성한다. 이러한 접근은 기존 중앙화 방식이 제공하는 높은 최적화 성능과 분산 방식의 실시간 적응 능력을 동시에 확보함으로써, 기존 방법들이 가진 한계를 극복하는 데 기여한다.

본 연구의 주요 기여도는 다음과 같다.

- 제안된 핸드오버 알고리즘은 각 사용자가 독립적으로 지역 정보를 활용하여 즉각적으로 핸드오버 결정 한다. 이때, 우리는 범 간 또는 위성 간 핸드오버 비용 차이, 사용자 서비스 품질 제약 조건, 그리고 부하 균형을 고려하면서 핸드오버 횟수를 최소화하고, 동시에 처리율(Throughput)을 최대화하는 다중 목적 최적화 문제를 공식화한다.
- 우리는 이 최적화 문제를 다중 범 안테나 이득, 채널 페이딩(Fading) 등 다중 범 저궤도 위성의 통신 특성을 고려한 MAPPO 최적화 문제로 변환한다.
- 시뮬레이션 결과, 제안된 알고리즘은 분산 학습 및 분산 실행(*DTDE, Decentralized Training Decentralized Execution*) 방식의 알고리즘인 Multi-agent Deep Q-network (MADQN)과 기존 핸드오버 알고리즘에 비해 가장 적은 핸드오버 횟수와 가장 높은 처리율을 보이며 우수한 성능을 입증하였다.

## II. 시스템 모델 및 문제 정의

### 2.1 시스템 모델

우리는 다중 범 저궤도 위성 네트워크에서 연속된  $T$ 개의 시간 슬롯 동안의 핸드오버 의사결정 문제를 고려하며, 여기서 총  $N$ 개의 저궤도 위성이  $J$ 개의 다중 범을 가지고 동일한 궤도 주기로 미리 구성된 궤도면을 따라 계속해서 궤도 운동을 하며, 유한한 관심 지역  $A$  내에 균일하게 분포된  $K$ 명의 지상 사용자를 서비스한다. 저궤도 위성의 범이 충분히 넓은 커버리지 영역을 가지고 있으므로, 각 시간 슬롯 동안 각 유저는 적어도 하나의 범 커버리지 내에 있다고 가정한다. 그림 1은 고려된 시나리오를 나타낸다. 사용자의 집합은

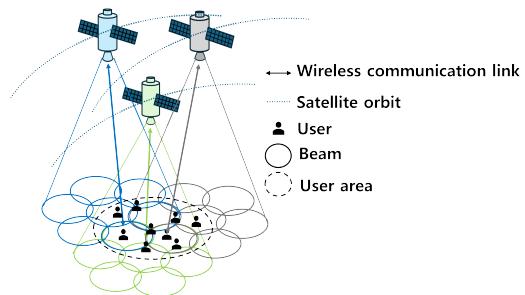


그림 1. 시스템 모델  
Fig. 1. System model

$\mathbf{K} = \{1, 2, \dots, k, \dots, K\}$ 로 표기한다. 위성의 집합은  $\mathbf{N} = \{1, 2, \dots, n, \dots, N\}$ 로 각 위성의 범 집합은  $\mathbf{J} = \{1, 2, \dots, j, \dots, J\}$ 로 정의한다. 편의를 위해 우리는 네트워크 내 모든 범에 관한 새로운 집합인  $\mathbf{V} = \mathbf{N} \times \mathbf{J}$ 을 정의하며, 위성  $n$ 의  $j$ 번째 범을 나타내는 이차원 벡터는  $v = (n, j) \in \mathbf{V}$ 로 표기한다. 각 위성 범의 커버리지 영역은 원으로 가정하며, 각 범의 중심은 커버리지 영역의 중심에 위치한다.

시간 슬롯(Time Slot)의 인덱스 집합을  $T = \{1, \dots, T\}$ 로 나타낸다. 각 시간 슬롯마다 채널은 준정적 페이딩(Quasi-static Fading) 겪지만, 시간 슬롯이 변경될 때마다 달라진다. 각 범의 시스템 대역폭은  $B_v$ 이다. 각 사용자는 GPS(Global Position System)를 사용해 자신의 위치를 파악하고, 저궤도 위성의 예측 가능한 움직임 덕분에 사용자는 서비스하는 위성의 범을 인식하게 된다<sup>[12]</sup>. 저궤도 위성의 높은 이동성 때문에 각 사용자가 사용할 수 있는 위성의 범이 각 시간 슬롯마다 동적으로 변한다. 따라서 사용자는 매 시간 슬롯마다 어떤 범과 통신할지 결정해야 한다.

시간 슬롯  $t$ 에서 각 위성의 범  $v = (n, j)$ 와 사용자  $k$  ( $\in \mathbf{K}$ ) 간의 커버리지 인디케이터(Indicator)는  $c_{k,v}^t$ 로 나타내며, 시간 슬롯  $t$ 에서 사용자  $k$ 가 범  $v$  커버리지 내에 있으면 1, 그렇지 않으면 0으로 정의된다. 따라서, 시간 슬롯  $t$ 에서 사용자  $k$ 를 커버하는 위성의 범 집합은  $C_k^t = \{v | c_{k,v}^t = 1, v \in \mathbf{V}\}$ 로 나타낸다.

시간 슬롯  $t$ 에서 사용자  $k$ 와 위성의 범  $v$  간의 연결은 인디케이터 변수  $x_{k,v}^t$ 로 나타내며, 이는 사용자  $k$ 가 위성의 범  $v$ 에 연결된 경우 1로 정의되며, 그렇지 않으면 0으로 정의된다. 시간 슬롯  $t$ 에서 각 위성의 범  $v$ 의

부하는  $l_v^t = \sum_{k=1}^K x_{k,v}^t$ 로 표현된다. 사용자  $k$ 는 자신을 커

버하는 위성 범 ( $\forall v \in C_k^t$ )에 대한 부하를 지역 정보로

얻을 수 있다. 시간 슬롯  $t$ 에서 사용자  $k$ 와 모든 빔 간의 연결을  $\mathbf{x}_k^t = \{x_{k,(1,1)}^t, \dots, x_{k,v}^t\}$ 로 나타낸다. 사용자  $k$ 는 각 시간 슬롯마다 하나의 빔에만 연결될 수 있으며, 핸드오버는  $X_k^t \neq X_k^{t+1}$  일 때 발생한다. 따라서 시간 슬롯  $t$ 에서 사용자  $k$ 의 핸드오버 비용은 다음과 같이 정의된다.

$$\varphi_k^t = \begin{cases} \xi, & \text{if } X_k^t \neq X_k^{t+1} \text{ and } E_1 \quad (\text{위성 간 핸드오버}) \\ \zeta, & \text{if } X_k^t \neq X_k^{t+1} \text{ and } E_2 \quad (\text{빔 간 핸드오버}) \end{cases} \quad (1)$$

여기서,  $E_1, E_2$ 는

$$\begin{aligned} E_1: \rho_k(t) &\neq \rho_k(t+1), \rho_k(t) = n \in N, \\ E_2: \rho_k(t) &= \rho_k(t+1) \text{ and } \delta_k(t) \\ &\neq \delta_k(t+1), \delta_k(t) = j \in J \end{aligned}$$

이고,  $\rho_k(t), \delta_k(t)$ 는 각각 시간 슬롯  $t$ 에서 사용자  $k$ 와 연결된 위성과 대응하는 빔을 나타낸다.

핸드오버는 위성 간 핸드오버와 빔 간 핸드오버 두 가지로 구분되며, 위성 간 핸드오버는 다른 위성의 빔으로 전환되는 경우를 의미한다. 특히, 두 핸드오버 방식에는 발생 비용에 차이가 있다. 빔 간 핸드오버인 경우 같은 위성 내에서 이뤄지기 때문에 리소스를 재할당이 상대적으로 수월하다. 그러나 위성 간 핸드오버인 경우 타겟(Target) 위성에게 핸드오버 요청을 보내고 핸드오버 승인을 받아야 하는 과정이 추가된다. 따라서 일반적으로 위성 간 핸드오버가 더 높은 신호 비용(Signalling Cost)과 신호 지연(Signalling Delay)<sup>[5]</sup>로 인해 더 높은 핸드오버 비용( $0 \ll \zeta < \xi$ )을 발생시킨다.

시간 슬롯  $t$ 에서 사용자  $k$ 와 위성 빔  $v$ 간의 안테나 이득을 포함한 채널 이득은 다음과 같이 나타낸다<sup>[10,13]</sup>.

$$|h_{k,v}^t|^2 = \frac{G^T G^R |\Psi|^2}{\left(4\pi \frac{d_{k,v}}{\lambda}\right)^2}. \quad (2)$$

여기서  $G^T, G^R$ 은 각 위성의 빔 송신, 사용자 수신 안테나 이득이며,  $\Psi$ 는 셰도우 라이시안 페이딩(Shadowed-Rician-fading) 나타내며,  $d_{k,v}$ 는 사용자와 빔 사이의 거리이며,  $\lambda$ 는 파장이다. 이때  $G^T$ 와  $G^R$ 은 위성과 사용자의 안테나 방사 패턴에 의해 결정된다는 점을 유의해야 한다. 구체적으로 ITU-R S.1528<sup>[14]</sup>를 참

조하면 위성의 안테나 방사 패턴은 다음과 같다.

$$G^T(\theta) = \begin{cases} G_{\max}^T & \theta \leq \theta_b \\ G_{\max}^T - 3(\theta/\theta_b)^2 & \theta_b < \theta \leq Y \\ G_{\max}^T + L_F - 25\log_{10}(\theta/Y) & Y < \theta \leq Z \\ L_F & Z < \theta \end{cases} \quad (3)$$

여기서  $\theta$ (°, 도 단위)는 오프 축 각도(Off-Axis Angle)이며  $\theta_b$ 는 3dB 빔폭  $\theta_{3dB}$ 의 절반이다.  $G_{\max}^T$ 는 송신 안테나의 최대 이득을 나타내며,

$$G_{\max}^T = 10\log_{10}\left[4.93\left(\frac{70}{\theta_{3dB}}\right)^2\right] \text{으로 계산한다.}$$

$Z = Y \times 10^{0.04(G_{\max}^T + L_s - L_F)}$ 에서  $L_s$ 는 주빔(Main Beam)과 피크(Peak) 이득 아래의 내부 측엽 마스크(Near-In Side-Lobe Mask) 교차점을 나타내며,  $L_F = 0 \text{ dBi}$ 는 외부 측엽(Far-Out Side-Lobe) 레벨(Level)이고,  $Y = 1.5\theta_b$ 다.

ITU-R S.465-6<sup>[15]</sup>을 참조하면 사용자 터미널(Terminal)의 안테나 방사 패턴은 다음과 같다.

$$G^R(\theta) = \begin{cases} G_{\max}^R & \theta \leq \theta_{\min} \\ 32 - 25\log_{10}(\theta) & \theta_{\min} < \theta \leq 48^\circ \\ -10 & 48^\circ < \theta \end{cases}. \quad (4)$$

여기서  $\theta_{\min} = \max(2^\circ, 144(D/\lambda)^{-1.09})$ 이고,

$G_{\max}^R = 10\log_{10}\left[\left(\frac{\pi D}{\lambda}\right)^2\right]$ 는 수신 안테나의 최대 이득이다. 이때  $D$ 는 사용자 안테나 개구면(Aperture)의 직경이다.

빔  $v$ 의 송신 전력은  $P_v$ 로 나타내며, 문제의 복잡성을 줄이고자 위성의 총 전력  $P_{SAT}$ 가 위성의 빔들에게 균등하게 분배된다고 가정한다(즉,  $P_v = P_{SAT}/J$ ). 간섭은 부하 수에 따라 대역폭  $B_W$ 을 공유하여 사용함에 따라 간섭은 존재하지 않음을 가정한다. 따라서 시간 슬롯  $t$ 에서 위성 빔  $v$ 에 의해 사용자  $k$ 의 신호 대 잡음 비율(SNR, Signal-to-Noise ratio)은 다음과 같이 표현된다.

$$SNR_{k,v}^t = \frac{P_v |h_{k,v}^t|^2}{\sigma^2}. \quad (5)$$

여기서  $\sigma^2 = \kappa T_{noise} B_W / l_v^t$ 는 잡음 전력을 나타내며,

$\kappa, T_{noise}$ 는 각각 볼츠만 상수(Boltzmann's Constant), 수신기 잡음 온도를 나타낸다. 더 나아가, 수식 (2)-(5)을 정리하면, 시간 슬롯  $t$ 에서 빔  $v$ 에 의해 서비스되는 사용자  $k$ 의 성취 가능한 처리율(Achievable Throughput)은 다음과 같이 표현할 수 있다.

$$R_{k,v}^t = \frac{B_W}{l_v^t} \log_2 (1 + SNR_{k,v}^t). \quad (6)$$

이때 우리는 사용자의 목표 서비스 품질을 만족시키고자, 사용자는 각자 최소 처리율( $R_{k,v}^t \geq R_k^{th}$ )을 만족한다고 가정한다.

## 2.2 최적화 문제 정의

본 논문에서는 다중 빔 저궤도 위성의 효율적인 핸드오버 관리를 위해,  $T$ 시간 슬롯 동안 위성 또는 빔 간 핸드오버의 비용 차이, 부하 균형, 그리고 서비스 품질 제약 조건을 고려하여 핸드오버 발생을 최소화하고 동시에 처리율을 최대화하는 것을 목표로 한다. 이에 따라 최적화 문제를 정의하면 다음과 같다.

$$\mathbf{P}: \min_{x_{k,v}^t} \sum_{t=1}^T \sum_{k=1}^K \varphi_k^t - \sum_{v \in \mathbf{V}} x_{k,v}^t (\omega_R R_{k,v}^t) \quad (7)$$

$$\text{s.t. } x_{k,v}^t \in \{0,1\}, \forall v \in \mathbf{V}, \quad (8)$$

$$\sum_{v \in \mathbf{V}} x_{k,v}^t = 1, \forall k \in \mathbf{K}, \quad (9)$$

$$\sum_{v \in \mathbf{V}} x_{k,v}^t R_{k,v}^t \geq R_k^{th}, \forall k \in \mathbf{K}. \quad (10)$$

여기서  $\omega_R$ 은 처리율 값을 스케일링(Scaling)하는 요소이다. 제약사항 (8)과 (9)는 각 사용자가 매 시간 슬롯마다 커버리지 내에 있는 오직 하나의 빔과 연결됨을 의미한다.

## III. 제안된 다중 빔 저궤도 위성 위한 다중 에이전트 심층 강화 학습 기반 핸드오버 전략

최적화 문제 (7)의 주요 목표는 사용자가 커버리지 내에 있는 빔들의 부하나 채널 정보 같은 현재 및 지역 정보를 기반으로 장기간(Long-term)적인 성능을 극대화 하는 것이다. MADRL은 복잡한 무선 통신 환경에서 고차원 상태 공간 및 행동 공간을 가진 다중 의사 결정

자(또는 행동자, Actor)가 도전적인 문제를 해결하기 위한 탁월한 접근법이다<sup>[16]</sup>. MADRL은 동적으로 변화하는 환경에서 에이전트들이 자신의 지역 정보를 기반으로 숨겨진 패턴(Hidden Pattern)을 학습할 수 있으므로 문제를 최적화할 수 있다. 따라서, 본 장에서는 최적화 문제 (7)를 MADRL 프레임워크(Framework)로 공식화한다.

### 3.1 다중 에이전트 심층 강화학습 공식화

최적화 문제 (7)를 MADRL 기반으로 해결하기 위해서는 최적화 문제를 분산된 부분 관측 마르코프 결정 과정(Dec-POMDP, Decentralized Partially Observable Markov Decision Process)<sup>[17]</sup> 프레임워크로 공식화해야 한다. 구체적으로 Dec-POMDP의 에이전트, 관측(Observation), 행동(Action) 그리고 보상(Reward)를 정의 내리면 다음과 같다.

1) 에이전트: 각 사용자  $k \in \mathbf{K}$ 가 하나의 에이전트로써 다른 에이전트 및 환경과 상호작용한다.

2) 관측: 각 시간 슬롯  $t$ 마다 사용자는 환경과 상호작용하면서 관측된 값을 업데이트한다. 사용자  $k$ 의 관측(상태) $s_k^t$ 은 다음과 같이 정의된다.

$$s_k^t = \langle h_{k,v}^t, l_v^t, a_k^{t-1} \rangle, \forall v \in C_k^t \quad (11)$$

여기서  $a_k^{t-1}$ 는 특징 정보(Fingerprint)로, 사용자  $k$ 의 이전 시간 슬롯에서의 행동 정보이다. 특징 정보 기반 MADRL은 리플레이 버퍼를 안정화를 통해 비정상성(Non-Stationary) 환경에서 보다 안정적인 학습을 가능하게 한다<sup>[18]</sup>. 이때 관측은  $2 \times NJ + 1$ 로 고정된 크기를 갖는 평탄화된 행벡터이고, 커버리지 내에 없는 빔의 정보는 -1로 설정한다. 채널과 부하 정보는 위성의 방송을 통해 얻을 수 있다.

3) 행동: 시간 슬롯  $t$  때 사용자  $k$ 의 행동은 커버리지 내 빔 중 하나를 선택하는 것이다. 즉,  $a_k^t = v, \forall v \in C_k^t$ 이다.

4) 보상: 시간 슬롯  $t$  때 사용자  $k$ 의 보상  $r_k^t$ 은 다음과 같다.

$$r_k^t = \begin{cases} -\varsigma & \text{if } R_{k,a_k^t}^t < R_k^{th}, \\ -\varphi_k^t & \text{if } a_k^{t-1} \neq a_k^t \text{ and } R_{a_k^t}^t \geq R_k^{th}, \\ \min(R_{Ack}, \omega_R R_{k,v}^t) & \text{otherwise.} \end{cases} \quad (12)$$

여기서  $-\varsigma$ 는 사용자  $k$ 가 목표 사용자 서비스 품질  $R_k^{th}$  을 만족시키지 못한 통신 장애(Outage)에 대한 페널티(핸드오버 여부와 무관하게 적용됨)이다.  $-\varphi_k^t$ 는 서비스 품질을 만족하고 핸드오버가 일어났을 때 페널티(즉, 핸드오버 성공에 대한 페널티)이다.  $R_{Ack}$ 는 처리율의 편차가 크기 때문에 너무 큰 보상을 방지하는 특정 고정 값이다.

### 3.2 제안된 Multi-Agent Proximal Policy Optimization 알고리즘

본 절에서는 앞서 정의한 MADRL 프레임워크를 활용하여 최적화할 수 있는 MAPPO 알고리즘을 제안한다. 그림 2는 우리가 제안하는 MAPPO의 구조를 나타낸다. MAPPO는 CTDE 구조를 따른다. 학습 단계에서는 중앙 비평기(Critic) 네트워크가 모든 에이전트의 상태, 행동, 보상 정보를 통합하여, 에이전트 간의 상호작용을 고려하여 학습을 최적화(정책 업데이트)한다. 이 때 우리가 제안하는 MAPPO 구조는 모든 에이전트가 동일한 정책  $\pi(a|s)$ 을 공유하며, 각 에이전트(사용자)는 자신이 얻을 수 있는 지역 정보(즉, 관측 정보)를

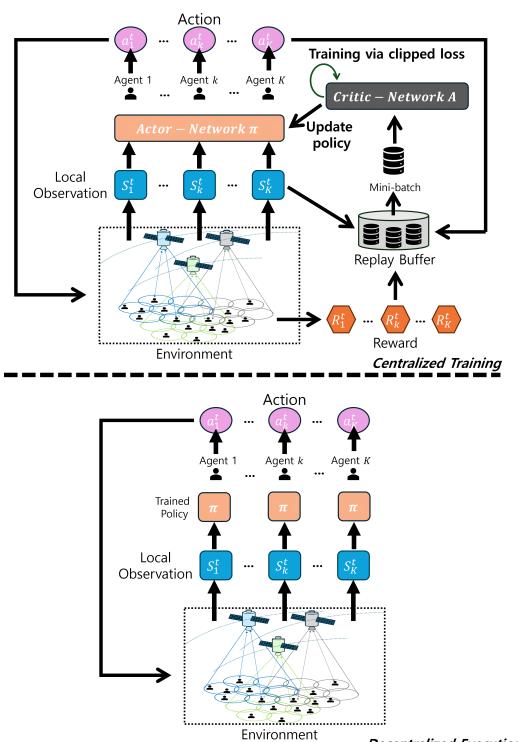


그림 2. 제안된 MAPPO 구조  
Fig. 2. Proposed architecture of MAPPO

바탕으로 독립적으로 행동을 결정한다. 학습이 완료된 후, 각 에이전트는 중앙 비평기 없이 학습된 정책을 바탕으로 독립적으로 실행된다. 이와 같은 방식은 DTDE에 비해 협력적인 환경에서 효율적이며, 복잡한 환경에서도 높은 성능을 발휘할 수 있다.

MAPPO의 핵심 개념은 행위자-비평기(Actor-Critic) 구조와 클리핑(Clipping) 기법이다. 행위자-비평기 구조는 에이전트가 상태(또는 관측)의 가치  $V(s)$ 를 학습하고 평가하는데, 이때 이점 함수(Advantage Function)  $A_k^t$ 를 활용한다. 이점 함수는 정책을 업데이트할 때, 행동의 상대적 가치를 평가하는데 사용되며, 상태  $s_k^t$ 에서 취한 행동  $a_k^t$ 가 얼마나 좋은지를 계산한다.

$$A_k^t = R^t - V(s_k^t) \quad (13)$$

여기서  $R^t$ 는 상태  $s_k^t$ 에서 얻은 실제 보상이다. 즉, 이점 함수는 비평기가 평가한 상태 가치와 실제로 얻은 보상 간의 차이를 계산하여, 행위자가 더 나은 행동을 선택할 수 있도록 정책 업데이트를 유도한다.

클리핑 비율  $r^t(\theta)$ 는 다음과 같이 정의된다.

$$r^t(\theta) = \frac{\pi_\theta(a_k^t | s_k^t)}{\pi_{old}(a_k^t | s_k^t)} \quad (14)$$

여기서  $\pi_\theta(a_k^t | s_k^t)$ 는 새로운 정책에서 상태  $s_k^t$ 에서 행동  $a_k^t$ 를 선택한 확률이고,  $\pi_{old}(a_k^t | s_k^t)$ 는 이전 정책에서 상태  $s_k^t$ 에서 행동  $a_k^t$ 을 선택할 확률이다. 이 비율은 새로운 정책이 이전 정책과 얼마나 다른지를 나타낸다. MAPPO는 클리핑된 정책 비율을 사용하여 손실 함수를 다음과 같이 정의한다.

$$L_{CLIP}(\theta) = \mathbb{E} \left[ \min \left( r^t(\theta) A_k^t, \text{clip}(r^t(\theta), 1 - \epsilon, 1 + \epsilon) A_k^t \right) \right] \quad (15)$$

여기서  $\epsilon$ 은 클리핑 파라미터(Parameter)로 정책이 지나치게 변하지 않도록 설정된 범위다. 즉, 이 손실 함수는 정책 업데이트 시 너무 큰 변화가 일어나지 않도록 클리핑하며, 학습 과정에서 정책의 안정적인 업데이트를 보장한다.

MAPPO는 위에서 언급한 대로 CTDE 구조로 동작하기 때문에, 복잡하고 협력적인 동적 환경에 대한 적응성이 뛰어나다. 또한, 안정적인 정책 업데이트를 위한

**Algorithm 1** 다중 빔 저궤도 위성 핸드오버를 위한 MAPPO 알고리즘 학습과정

- 1: 입력: 환경  $env$ , 에이전트 수  $K$ , 초기 정책  $\pi_0$ , 메모리  $memory$
- 2: 하이파라미터: 학습률  $\alpha$ , 할인 인자  $\gamma$ , 클리핑 파라미터  $\epsilon$ , 에폭 수  $I$ , 배치 크기  $B$ , 학습 에피소드 횟수  $EP$
- 3: 초기화 각 에이전트의 정책  $\pi_0$ 을 초기화
- 4: while 에피소드 횟수  $Ep = 0$  to  $EP$  do
- 5:     환경에서 초기 상태  $s^0$  얻기
- 6:     for 시간 슬롯  $t = 0$  to 에피소드 시간 길이 do
- 7:         for 각 에이전트  $k = 1$  to  $K$  do
- 8:             현재 상태  $s_k^t$ 에서 정책  $\pi_{old}$ 을 사용하여 행동  $a_k^t$  선택
- 9:             행동  $a_k^t$  실행, 다음 상태  $s_k^{t+1}$ , 보상  $r_k^t$ , 종료 플래그  $d_k^t$  얻기
- 10:             메모리에  $(s_k^t, a_k^t, r_k^t, s_k^{t+1}, d_k^t)$  저장
- 11:         end for
- 12:         if 메모리 크기  $> B$  then
- 13:             메모리로부터 무작위 배치 추출 및 보상 정규화
- 14:             for 정책 최적화 과정, 에폭  $i = 1$  to  $I$  do
- 15:                 각 행동  $a_k^t$  및 상태  $s_k^t$ 의 로그 확률과 상태 값을 계산
- 16:                 이점 함수  $A_k^t = R^t - V(s_k^t)$  계산
- 17:                 클리핑 비율 계산:
- 18:                 
$$r^t(\theta) = \frac{\pi_\theta(a_k^t | s_k^t)}{\pi_{old}(a_k^t | s_k^t)}$$
- 19:                 손실 함수 계산:
- 20:                 
$$L_{CLIP}(\theta) = \mathbb{E}^t [\min(r^t(\theta)A_k^t, \text{clip}(r^t(\theta), 1 - \epsilon, 1 + \epsilon)A_k^t)]$$
- 21:                 정책 업데이트  $\pi_0 \leftarrow \pi_\theta$
- 22:         end if
- 23:     end for
- 24: end while

그림 3. 제안된 MAPPO 학습과정에서의 의사코드  
Fig. 3. Proposed pseudocode of MAPPO

클리핑 기법과 이점 함수를 활용하여, 주어진 우리의 최적화 문제를 효과적으로 해결할 수 있다. 전반적인 MAPPO의 학습과정에서의 의사코드(Pseudocode)는 그림 3을 통해 기술했다.

#### IV. 시뮬레이션 결과 및 성능 분석

##### 4.1 시뮬레이션 파라미터

우리는 다중 빔 저궤도 위성이 정해진 궤도를 따라 움직이며, 반경 70km인 원 내부에 균일하게 분포된 지상 사용자들과 시간 슬롯 간격 1초로 총 40 슬롯 동안 통신하는 시뮬레이션을 진행하였다. 또한, 각 위성은 반경 약 100km의 빔 커버리지를 통해 서비스를 제공하며, 매 시간 슬롯마다 사용자가 1~5개의 빔 커버리지 영역 내에 위치하여 시뮬레이션 기간 동안 지속적으로 통신 서비스를 받는 상황을 고려하였다. 자세한 시스템 파라미터는 표 1에 정리했다. 제안된 MAPPO의 네트워크 및 학습 파라미터는 표2에 정리했다. 행위자 네트워크는 주어진 상태에서 적절한 행동을 선택하는 네트워크로, 입력 크기는 관측 정보의 크기이고, 출력 크기는 행동 공간의 크기이다. 비평가 네트워크는 주어진 상태에 대한 가치를 평가하는 네트워크로, 입력 크기는 관측 정보의 크기이고, 출력의 크기는 해당 상태의 가치이므로 1이다. 보상과 관련된 값  $-\zeta = -10$ ,  $-\xi = -5$ ,  $-\zeta = -0.1$ ,  $\omega_R = 10^{-7}$ ,  $R_{Ack} = 10$ 로 설정되었다.

표 1. 시스템 파라미터

Table 1. System parameters

파라미터	값
위성 수 $N$	3
각 위성의 빔 수 $J$	7
위성 고도	450 km
위성 공전 속도	7.6 km/s
빔의 전송 전력 $P_{n_j}$	9.8 W
빔의 시스템 대역폭 $B_W$	200 MHz
최소 처리율 $R_k^{th}$	10 Mbps
사용자 수 $K$	100
사용자 안테나 개구면 $D$	0.4 m
잡음 온도 $T_{noise}$	300 K
반송파 주파수	11.7 GHz(Ku-band)

표 2. MAPPO 네트워크 및 학습 파라미터

Table 2. Parameters of MAPPO networks and training

행위자 네트워크			
계층	입력 크기	출력 크기	활성화함수
입력 계층	$2 \times N + 1$	256	ReLU
완전연결계층	256	256	ReLU
출력 계층	256	$N \times J$ (행동)	-

##### 비평가 네트워크

계층	입력 크기	출력 크기	활성화함수
입력 계층	$2 \times N + 1$	256	ReLU
완전연결계층	256	256	ReLU
출력 계층	256	1(상태 가치)	-
학습 파라미터	값		
훈련 에피소드, $EP$	4700 에피소드		
리플레이 버퍼 사이즈	500 에피소드		
학습률 $\alpha$	0.0001		
클리핑 $\epsilon$	0.15		
배치 사이즈	128 에피소드		
에폭 수 $I$	15		

본 논문에서는 제안된 핸드오버 방식(MAPPO)을 다음과 같은 3가지 알고리즘과 비교한다.

1) 무작위(Random) 핸드오버 알고리즘<sup>[19]</sup>: 사용자가 매 시간 슬롯마다 커버리지 내에 있는 빔 중 하나를 무작위로 선택한다.

2) 최대 채널 이득(MAX-CG) 핸드오버 알고리즘<sup>[1]</sup>: 사용자가 매 시간 슬롯마다 커버리지 내에 있는 빔 중 채널 이득이 제일 큰 빔을 선택한다.

3) Multi-agent Deep Q-network (MADQN) 핸드오버 알고리즘<sup>[2]</sup>: DTDE 방식으로 사용자는 매 시간 슬롯마다 MADQN 통해 최적화 된 정책을 따라 빔을 선택한다.

#### 4.2 복잡도 및 수렴 분석

제안된 MAPPO 학습과정에서의 복잡도를 표 2를 참고하여 분석하면 다음과 같다.

##### 1) 행위자 네트워크:

$$O((2 \times N \times J + 1) \times 256 + 256 \times (N \times J)) = O(N \times J)$$

##### 2) 비평가 네트워크:

$$O((2 \times N \times J + 1) \times 256 + 256 \times 1) = O(N \times J)$$

따라서 MAPPO 학습과정에서의 총 복잡도는 다음과 같다.

$$O(N \times J) + O(N \times J) = O(N \times J)$$

MAPPO는 수렴성을 보장하기 위해 클리핑 기법, 이 점 합수 기반 업데이트, 그리고 중앙 비평가과 같은 여러 요소들을 활용한다. 그림 4는 제안된 알고리즘의 훈련 과정에서의 에피소드 보상 값의 변화를 보여준다. 이를 통해, 제안된 알고리즘은 훈련 에피소드가 증가할 수록 에피소드 보상 값이 점진적으로 증가하고, 약

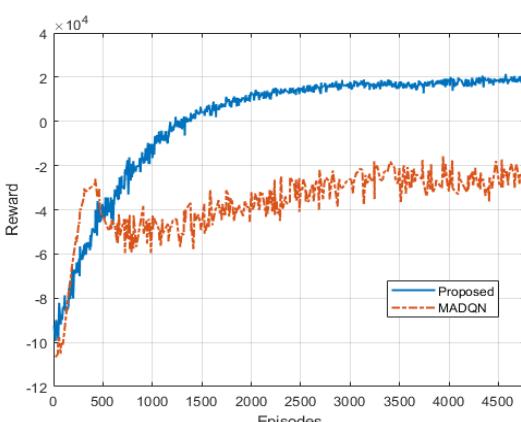


그림 4. 학습 과정에서의 에피소드 보상  
Fig. 4. An illustration of episode reward during training

3000 에피소드 이후부터 수렴하는 것을 확인할 수 있다. 이때, 제안된 방식이 DTDE 방식인 MADQN에 비해 수렴 값이 약 200% 더 좋은 것을 확인할 수 있다. 이는 우리가 제안한 CTDE 방식의 MAPPO 알고리즘이 보다 효과적으로 최적화 문제를 해결한 결과이다.

#### 4.3 성능 비교

표 3은 핸드오버 알고리즘에 따른 성능 비교표이다. 각 요소는 총 40 슬롯의 에피소드 중 첫 번째 슬롯을 제외한 39 슬롯 동안의 성능을 측정한 값이다. QoS'는 최소 처리율을 만족시키지 못한 횟수를 나타내며, 총 핸드오버, HO는 빔 간 핸드오버, BHO 또는 위성 간 핸드오버, SHO 발생 횟수를 합친 값이다. Ack는 서비스 품질을 만족시키면서 핸드오버가 발생하지 않은 횟수를 의미한다. 결과적으로, 제안된 알고리즘은 다른 알고리즘 대비 약 39.13% – 75.53% 적은 핸드오버를 발생시켜 성능이 우수함을 확인할 수 있다.

그림 5는 에피소드 동안의 사용자의 평균 처리율을 나타낸다. 해당 결과를 통해, 우리가 제안한 방식이 평균적으로 가장 우수한 처리율을 보인다는 것을 확인할 수 있다. 표 3과 그림 5의 결과를 종합해 보면, 제안된 알고리즘은 비교 알고리즘에 비해 가장 적은 핸드오버

표 3. 알고리즘에 따른 성능 비교표

Table 3. Performance comparison table by algorithm

기법	QoS'	Ack	BHO	SHO	HO
Random	12.81	6.28	3.89	16.02	19.91
MAX-CG	0.47	22.44	0.14	15.95	16.09
MADQN	1.82	29.18	0.65	7.35	8.00
Proposed	1.23	<b>32.9</b>	1.24	3.63	<b>4.87</b>

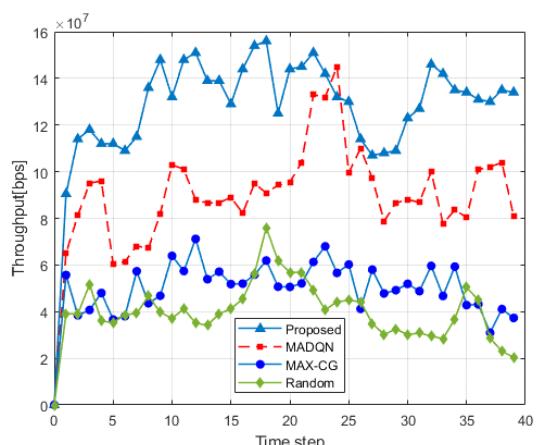


그림 5. 에피소드 동안의 사용자 평균 처리율  
Fig. 5. Average throughput of users during episode

횟수와 가장 높은 평균 처리율을 보였으며, 이는 최적화 문제 (7)를 매우 효과적으로 해결한 결과라고 할 수 있다.

## V. 결 론

본 논문에서는 CTDE 방식의 MADRL 알고리즘인 MAPPO 기반 다중 빔 저궤도 위성 핸드오버 전략을 제안하였다. 제안된 방식은 저궤도 위성 다중 빔 채널 특성을 반영하며, 빔 간 및 위성 간 핸드오버 비용 차이, 사용자 서비스 품질 제약 조건, 그리고 부하 균형을 고려하면서, 핸드오버 횟수 최소화하고 처리율 최대화한다. 제안된 알고리즘은 각 사용자가 독립적으로 지역 정보(커버리지 내 부하 수 및 채널 정보)를 바탕으로 핸드오버 결정을 하며 이는 복잡하고 동적인 다중 빔 저궤도 위성에 즉각적으로 대응할 수 있다. 시뮬레이션 결과, 제안된 알고리즘은 비교 알고리즘 대비 가장 적은 핸드오버 횟수와 가장 높은 처리율을 보이며 우수한 성능을 보였다.

## References

- [1] C. Lee, T. Kim, I. Bang, and S. H. Chae, "Multi-agent deep Q-network based low earth orbit satellite handover algorithm," in *Proc. Symp. KICS*, pp. 1028-1029, Pyeongchang, Korea, Feb. 2024.
- [2] G. Kim, S. Lee, H. Lim, B. C. Jung, and S. H. Chae, "Coverage probability analysis of LEO satellite communication systems with directional beamforming," in *Proc. 14th Int. Conf. Ubiquitous and Future Netw. (ICUFN)*, pp. 243-247, Paris, France, Jul. 2023. (<https://doi.org/10.1109/ICUFN57995.2023.10200817>)
- [3] S. H. Chae, H. Lim, H. Lee, and B. C. Jung, "Performance analysis of dense low earth orbit satellite communication networks with stochastic geometry," *J. Commun. and Netw.*, vol. 25, no. 2, pp. 208-221, Apr. 2023. (<https://doi.org/10.23919/JCN.2023.000009>)
- [4] C. Lee, I. Bang, T. Kim, H. Lee, B. C. Jung, and S. H. Chae, "Multi-agent deep reinforcement learning based handover strategy for LEO satellite networks," *IEEE Commun. Lett.*, vol. 29, no. 5, pp. 1117-1121, May 2025. (<https://doi.org/10.1109/LCOMM.2025.3554818>)
- [5] Y. Li, W. Zhou, and S. Zhou, "Forecast based handover in an extensible multi-layer LEO mobile satellite system," *IEEE Access*, vol. 8, pp. 42768-42783, Mar. 2020. (<https://doi.org/10.1109/ACCESS.2020.2977831>)
- [6] S. Zhang, A. Liu, C. Han, X. Ding, and X. Liang, "A network-flows based satellite handover strategy for LEO satellite networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2669-2673, Dec. 2021. (<https://doi.org/10.1109/LWC.2021.3111680>)
- [7] Y. Wu, G. Hu, F. Jin, and J. Zu, "A satellite handover strategy based on the potential game in LEO satellite networks," *IEEE Access*, vol. 7, pp. 133641-133652, Sep. 2019. (<https://doi.org/10.1109/ACCESS.2019.2941217>)
- [8] L. Yang, X. Yang, and Z. Bu, "A group handover strategy for massive user terminals in LEO satellite networks," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC)*, pp. 1-6, London, United Kingdom, Sep. 2022. (<https://doi.org/10.1109/VTC2022-Fall57202.2022.10012912>)
- [9] S. He, T. Wang, and S. Wang, "Load-aware satellite handover strategy based on multi-agent reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1-6, Taipei, Taiwan, Dec. 2020. (<https://doi.org/10.1109/GLOBECOM42002.2020.9322449>)
- [10] H. Liu, Y. Wang, and Y. Wang, "A successive deep Q-learning based distributed handover scheme for large-scale LEO satellite networks," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC)*, pp. 1-6, Helsinki, Finland, Jun. 2022. (<https://doi.org/10.1109/VTC2022-Spring5431.2022.9860376>)
- [11] Y. Sun, Y. Zhai, W. Wu, P. Si, and F. R. Yu, "Handover for multi-beam LEO satellite networks: A multi-objective reinforcement learning method," *IEEE Commun. Lett.*, vol. 28, no. 12, pp. 2834-2838, Dec. 2024.

- (<https://doi.org/10.1109/LCOMM.2024.3470890>)
- [12] I. Ali, N. Al-Dahir, and J. E. Hershey, "Predicting the visibility of LEO satellites," *IEEE Trans. Aerospace and Electronic Syst.*, vol. 35, no. 4, pp. 1183-1190, Oct. 1999.  
(<https://doi.org/10.1109/7.805436>)
- [13] Z. Lin, C. Jiang, and Z. Huang, "Satellite-terrestrial coordinated multi-satellite beam hopping scheduling based on multi-agent deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 10091-10103, Aug. 2024.  
(<https://doi.org/10.1109/TWC.2024.3368689>)
- [14] *Satellite Antenna Radiation Patterns for Non-Geostationary Orbit Satellite Antennas Operating in the Fixed-Satellite Service Below 30 GHz*, document ITU-R S.1528, 2001.
- [15] *Reference Radiation Pattern for Earth Station antennas in the Fixed-Satellite Service For use in Coordination and Interference- assessment in the Frequency Range from 2 to 31 GHz*, document ITU-R S.465-6, 2010.
- [16] C. Lee, S. Lee, T. Kim, I. Bang, J. H. Lee, and S. H. Chae, "Multi-agent deep reinforcement learning-based multi-UAV path planning for wireless data collection and energy transfer," in *Proc. 15th Int. Conf. Ubiquitous and Future Netw. (ICUFN)*, pp. 500-504, Budapest, Hungary, Jul. 2024.  
(<https://doi.org/10.1109/ICUFN61752.2024.10625275>)
- [17] F. A. Oliehoek, C. Amato, et al., "A concise introduction to decentralized POMDPs," Springer, 2016.
- [18] J.-H. Lee, H. Seo, J. Park, M. Bennis, and Y.-C. Ko, "Learning emergent random access protocol for LEO satellite networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 257-269, Jan. 2023.  
(<https://doi.org/10.1109/TWC.2022.3192365>)
- [19] J.-H. Lee, C. Park, S. Park, and A. F. Molisch, "Handover protocol learning for LEO satellite networks: Access delay and collision minimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7624-7637, Jul. 2024.  
(<https://doi.org/10.1109/TWC.2023.3342975>)

이 총 녭 (Chungnyeong Lee)



2023년 2월 : 한국공학대학교 전  
자공학과 학사

2025년 2월 : 한국공학대학교 IT  
반도체융합공학과 석사

2025년 3월~현재 : 한국공학대  
학교 IT반도체융합공학과 박  
사과정

<관심분야> 이동통신, 위성통신, 강화학습

[ORCID:0009-0004-3522-6330]

김 태 훈 (Taehoon Kim)



2011년 2월 : 한양대학교 정보통  
신공학부 (공학사)

2013년 2월 : 한국과학기술원 전  
기및전자공학과 (공학석사)

2017년 8월 : 한국과학기술원 전  
기및전자공학부 (공학박사)

2017년 9월~2020년 2월 : 국방  
과학연구소 국방첨단기술연구원 선임연구원

2020년 3월~2022년 9월 : 국립한밭대학교 컴퓨터공학  
과 조교수

2022년 10월~현재 : 국립한밭대학교 컴퓨터공학과 부  
교수

<관심분야> Wireless Communications, Satellite  
Communications, Wireless Network Security

[ORCID:0000-0002-9353-118X]

방 인 규 (Inkyu Bang)



2010년 2월 : 연세대학교 전기전  
자공학부 (공학사)  
2012년 1월 : 한국과학기술원 전  
기및전자공학과 (공학석사)  
2017년 8월 : 한국과학기술원 전  
기및전자공학부 (공학박사)  
2017년 9월~2019년 2월 : 싱가

포르 국립대학 컴퓨터과학과 박사후연구원

2019년 3월~2019년 7월 : 국방과학연구소 지상기술연  
구원 선임연구원

2019년 8월~2022년 9월 : 국립한밭대학교 정보통신공  
학과 조교수

2022년 10월~현재 : 국립한밭대학교 지능미디어공학  
과 부교수

<관심분야> 6G/5G, 무선네트워크보안, 이동통신보안,  
물리계층보안, 위성통신, 인공지능 무선통신 응용

[ORCID:0000-0001-7109-1999]

채 승 호 (Seong Ho Chae)



2010년 2월 : 성균관대학교 정보  
통신공학부 (공학사)  
2012년 2월 : 한국과학기술원 전  
기 및 전자공학과 (공학석사)  
2016년 2월 : 한국과학기술원 전  
기 및 전자공학부 (공학박사)  
2016년 3월~2016년 9월 : 한국

과학기술원 정보전자연구소 연수연구원

2016년 10월~2018년 2월 : 국방과학연구소 선임연구원  
2018년 3월~2023년 8월 : 한국과학대학교 전자공학부  
조교수

2023년 9월~현재 : 한국과학대학교 전자공학부 부교수

<관심분야> 6G 통신 시스템, 저궤도 위성통신, 에지  
컴퓨팅, 인공지능 무선통신 응용

[ORCID:0000-0001-9092-5039]