https://doi.org/10.7840/kics.2025.50.3.432

대한민국 트래픽 분포를 고려한 강화학습 기반 저궤도 위성 빔 호핑 알고리즘

문 태 한◆, 이 재 열*, 김 태 윤**, 이 영 포***, 김 동 욱***, 류 탁 기***, 김 재 현°

Reinforcement Learning-Based Low Earth Orbit Satellite Beam Hopping Algorithm Considering Traffic Distribution in South Korea

Taehan Moon◆, Jaeyeol Lee*, Tae-Yoon Kim**, Youngpo Lee***,
Dongwook Kim***, Takki Yu***, Jae-Hyun Kim°

요 약

저궤도 (low Earth orbit, LEO) 위성은 지상 네트워크 달리 공간의 제약을 받지 않고 여러 지역에 통신 서비스를 제공할 수 있는 장점으로 많은 연구가 진행되고 있다. 그중 다중 빔 위성 통신 시스템에서는 위성의 스펙트럼, 전력 및 용량 자원이 제한적이기 때문에, 효율적으로 자원을 활용할 수 있는 빔 호핑 (beam hopping, BH) 기술이 주목받고 있다. 본 논문에서는 대한민국의 다중 빔 저궤도 위성 서비스 시나리오에서의 심층 Q 네트워크 (deep Q-network, DQN) 기반 지구 고정 (Earth-fixed) 빔 호핑 알고리즘을 제안한다. 제안하는 알고리즘은 위성의 제한된 용량, 채널 상태와 대한민국 셀의 랜덤한 트래픽 분포를 고려하여 효율적인 빔 호핑을 수행한다. 모의실험 결과, 제안하는 알고리즘은 기존의 휴리스틱 알고리즘에 비해 위성 자원의 효율성을 향상시키고, 대한민국 내 셀 트래픽을 최대한 처리하는 데 우수한 성능을 보인다.

키워드 : 비지상 네트워크, 저궤도 위성, 빔 호핑, DQN 알고리즘
Key Words : Non-Terrestrial Network, Low Earth Orbit Satellite, Beam Hopping, Deep Q-network Algorithm

ABSTRACT

Low Earth orbit (LEO) satellites, unlike terrestrial networks, are not constrained by geographical limitations and have the advantage of providing data services to multiple regions. In multi-beam satellite communication systems, efficient resource management of spectrum, power, and capacity is essential, highlighting the importance of beam hopping (BH) technology. This paper proposes an Earth-fixed beam hopping algorithm based on a deep Q-network (DQN) for a multi-beam LEO satellite scenario over the South Korea. The

※ 본 연구는 SK 텔레콤 산학협력과제 (DTS서비스 필요지역/용량 분석 통한 위성망 설계기술 개발) 지원으로 연구되었음.
◆ First Author : Ajou University Department of Artificial Intelligence Convergence Network, ansxogks3@ajou.ac.kr, 학생회원
° Corresponding Author : Department of Electronic Engineering, Ajou University, jkim@ajou.ac.kr, 종신회원
* 아주대학교 우주전자정보공학과 (Department of Space Survey Information Technology, Ajou University), jaeyel98@ajou.ac.kr, 학생회원
** 아주대학교 AI융합네트워크학과 (Department of Artificial Intelligence Convergence Network, Ajou University), xodbsxogjs@ajou.ac.kr, 학생회원
*** SK telecom, Korea, {youngpo.lee, kimdw, takki.yu}@sk.com
논문번호 : 202411-285-B-RU, Received November 15, 2024; Revised November 19, 2024; Accepted November 19, 2024

proposed algorithm is designed to optimize beam hopping by efficiently managing the satellite's limited capacity, accounting for channel conditions, and accommodating the random traffic distribution of ground cells. Simulation results confirmed that the proposed algorithm improves the efficiency of satellite resources compared to existing heuristic algorithms, offering enhanced performance in maximizing the handling of cell traffic within South Korea.

Ⅰ. 서 론

최근 통신 기술이 6G로 발전함에 따라, 상대적으로 낮은 고도인 약 300 ~ 2,000 km에서 운용되는 저궤도 (low Earth orbit, LEO) 위성은 초공간 및 지속적인 연결성이라는 강점을 바탕으로 활발한 연구가 진행되고 있다[1]. 이러한 저궤도 위성들은 짧은 지연 시간과 광범위한 커버리지를 제공함으로써 차세대 글로벌 서비스의 핵심 기술로 주목받고 있다.

저궤도 위성은 광범위한 서비스를 제공하는 동시에 위성의 자원 관리 효율성을 높이고, 지상 셀의 요구 데이터 전송률을 충족해야 한다. 또한, 위성의 스펙트럼, 전력, 용량 자원이 제한된 상황에서 효율적인 자원 할당은 필수적이다. 특히 다중 빔을 방사하는 위성 시스템에서 빔 호핑 (beam hopping, BH) 기술은 제한된 자원의 최적화를 위한 효과적인 방법으로 주목받고 있다[2]. 빔 호핑 기술은 위성 자원을 시간 및 공간에 따라 동적으로 조정하여, 트래픽이 집중된 지역에 우선적으로 할당함으로써 불균형한 트래픽 분포에 효과적으로 대응하는 방안을 제시한다.

유럽 우주국 (European Space Agency, ESA)의 연구에 따르면, Eutelsat Quantum과 SES-17과 같은 최신 위성들은 소프트웨어 기반 기술과 디지털 투명 프로세서를 통해 실시간으로 자원 할당을 가능하게 한다[3]. 정지궤도 위성에서는 다중 에이전트 심층 강화학습 (multi-agent deep reinforcement learning, MADRL) 기반의 빔 호핑 스케줄링 기법에 대한 연구가 진행되고 있다[4]. 균일하지 않고 동적인 트래픽 수요에 대응하기 위해 정지궤도 위성 다중 빔 호핑 기술을 활용하여 효율적인 자원 할당 및 간섭 문제를 해결한다. 또한, 샤논 채널 용량 공식을 기반으로 각 셀에 할당된 자원의 전송 용량을 정량화하고 최적의 빔 스케줄링을 설계한다[5]. 이를 통해 위성 시스템의 처리량을 향상시키고, 대기 지연을 최소화한다. 이처럼 빔 호핑 기술은 위성 네트워크의 유연성을 크게 향상시키며, 다양한 사용자 요구에 신속하게 대응할 수 있는 환경을 제공한다. 하지만 다른 궤도의 위성과 달리, 저궤도 위성은 빠른 속도로 인해 지속적으로 변화하는 위성의 위치와 트래픽 분포를 고려해야 한다. 또한, 저궤도 위성과 지상 셀 간의 채널 상태를 고려하여, 신호 세기에 따라 적절한 변조 방식을 동적으로 할당함으로써 데이터 전송 효율을 최적화할 필요가 있다.

지상의 셀 트래픽은 랜덤한 분포를 따르며, 저궤도 위성은 제한된 용량을 갖기 때문에, 최적의 셀을 선택하여 서비스를 제공하기 위한 전략적 자원 할당이 필요하다. 더불어, 위성과 지상 셀 간의 최소 고도각 및 간섭으로 인해 신호 세기가 낮아져 통신 서비스 제공이 불가능할 가능성이 존재한다. 따라서, 본 논문에서는 이러한 문제를 해결하기 위해 강화학습 기반의 지구 고정 (Earth-fixed) 빔 호핑 알고리즘을 제안한다. 심층 Q 네트워크 (deep Q-network, DQN) 강화학습 기법을 활용하여 저궤도 위성 네트워크의 자원 할당을 실시간으로 최적화한다. 제안하는 알고리즘은 지상 셀의 랜덤한 트래픽 수요를 충족시키고, 신호 대 간섭 잡음 비 (signal to interference plus noise ratio, SINR)에 따라 동적으로 변조 방식을 할당하는 빔 호핑 방식을 설계한다.

Ⅱ. 시스템 모델 및 빔 호핑 문제 구성

2.1 시스템 모델

본 논문에서는 대한민국 전역에 다중 빔을 방사하는 저궤도 위성 시스템을 고려한다. 그림 1은 시간에 따라 저궤도 위성이 이동하며, 빔 호핑을 통해 대한민국에 통신 서비스를 제공하는 과정을 나타낸다. 해당 시스템은 $\boldsymbol{N}=\{n|n=1,2,..,N\}$ 개의 저궤도 위성으로 구성된 군집 위성 시스템으로, 각 위성은 B_n의 제한된 용량을 보유한다. 대한민국은 $\boldsymbol{M}=\{m|m=1,2,...,M\}$ 개의 고정된 셀로 구성된다. 대한민국 내 셀에는 랜덤하게 분포된 트래픽 $Z=\{\xi_m|\xi_m=\xi_1,\xi_2,...,\xi_M\}$ 이 발생한다. 저궤도 위성은 대한민국에 $\boldsymbol{T}=\{t|t=1,2,..,T\}$ 시간 동안 통신 서비스를 제공하며, 시간 t에서 저궤도 위성 n의 위치는 지구 중심 고정 좌표계 (Earth-centered Earth-fixed, ECEF)로 표현한다. 이를 통해 위성의 위치는 $v_n(t)=\{x_n(t),y_n(t),z_n(t)\}$ 으로 나타낸다. 저궤도 위성이 시간에 따라 이동한 거리는 궤도 상의 위치

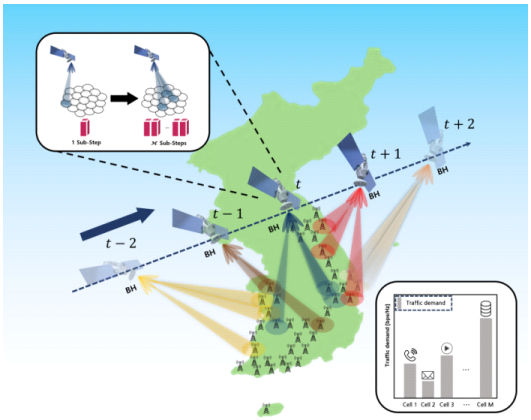


그림 1. 대한민국 내의 저궤도 위성 다중 빔 호핑 시스템
Fig. 1. LEO satellite multi-beam hopping system in South Korea

변화로 표현할 수 있다. 시간 t에 따른 위성의 위치 변화는 $\boldsymbol{P_T}=\{p_t|p_t=p_1,p_2,...,p_T\}$으로 나타내고, 다음과 같이 계산한다.

$$p_t = v_n(t) - v_n(t-1). \quad (1)$$

저궤도 위성은 궤도를 따라 이동하며, 최소 고도각을 만족하는 셀에게 통신 서비스를 제공한다. 시간 t에서 저궤도 위성 n이 대한민국 셀 m에게 하향 링크를 형성하는 상태는 $l_m^n(t)$으로 나타낸다. $l_m^n(t)=1$은 시간 t에서 저궤도 위성 n이 셀 m에 서비스를 제공함을 의미하고, $l_m^n(t)=0$은 제공하지 않음을 의미한다.

저궤도 위성은 궤도를 따라 고속으로 이동하므로, 시간에 따라 통신 서비스를 제공하는 위치가 지속적으로 변화한다. 이에 따라 대한민국에서 최소 고도각을 만족하는 셀이 시간 경과에 따라 달라진다. ECEF 좌표계는 지구의 곡률을 반영하지 않고, 각 셀의 위치마다 수평면 정의가 달라 저궤도 위성과 셀 간의 고도각을 계산하기 위해서는 동북상 좌표계 (east-north-up, ENU)으로 변환이 필요하다. ENU 좌표계는 특정 위치의 수평면을 기준으로 한 지역 좌표를 제공하여, 위성과 셀 간의 상대적 위치와 고도각을 효과적으로 계산할 수 있다. 이를 활용하여 시간 t에서 저궤도 위성 n과 셀 m의 고도각 $\theta_m^n(t)$은 다음과 같이 계산된다.

$$\theta_m^n(t) = \sin^{-1}\left(\frac{u}{\sqrt{e^2+n^2+u^2}}\right)\times\frac{180}{\pi}, \quad (2)$$

e, n, u는 ENU 좌표계의 위치를 나타낸다.

ECEF 좌표계를 ENU 좌표계로 변환하는 과정은 다음과 같이 표현한다.

$$\begin{bmatrix} e \\ n \\ u \end{bmatrix} = R \cdot \nabla v, \quad (3)$$

∇v는 위성과 셀 간의 위치 관계를 나타내는 벡터이다. R은 셀 위도, 경도를 활용한 좌표계 변환 행렬로, 식 (4)와 같이 계산된다.

$$R=\begin{bmatrix} -\sin(\lambda) & \cos(\lambda) & 0 \\ -\sin(\phi)\cos(\lambda) & -\sin(\phi)\sin(\lambda) & \cos(\phi) \\ \cos(\phi)\cos(\lambda) & \cos(\phi)\sin(\lambda) & \sin(\phi) \end{bmatrix}, \quad (4)$$

λ과 ϕ는 각각 셀의 경도, 위도를 나타낸다.

2.2 채널 모델

본 논문에서 저궤도 위성과 한반도 셀 간의 채널 모델로 자유 공간 경로 손실 (free space path loss, FSPL)을 가정하고, 다음과 같이 표현한다.

$$FSPL_m^n = 32.45 + 20\log_{10}(f_c) + 20\log_{10}(d_m^n), \quad (5)$$

f_c는 중심 주파수를 의미한다. d_m^n은 위성 n과 셀 m간의 경사 길이를 나타내고, 다음과 같이 표현한다.

$$d_m^n = \sqrt{R_E^2\sin^2\theta_m^n + h_0^2 + 2h_0R_E} - R_E\sin\theta_m^n, \quad (6)$$

R_E는 지구 반지름, h_0는 위성의 천저점 (nadir point)에서 저궤도 위성까지의 높이를 의미한다.

$FSPL_m^n$을 활용하여 시간 t에서 위성 n과 셀 m간의 신호 대 잡음 비 (signal to noise ratio, SNR)을 다음과 같이 표현한다.

$$SNR_m^n(t) = \frac{P_n(t)\cdot G_m(t)}{k\cdot T_\sigma \cdot BW_m^n \cdot FSPL_m^n(t)}, \quad (7)$$

$P_n(t)$는 저궤도 위성 n의 송신 전력, $G_m(t)$는 시간 t에서 셀 m의 수신 안테나 이득, k는 볼츠만 상수, T_σ는 시스템 잡음 온도를 의미한다. BW_m^n는 위성 n이 셀 m에게 방사하는 빔의 대역폭을 의미한다.

다중 빔 시나리오에서는 저궤도 위성이 여러 셀에게 서비스를 제공할 경우, 셀 간 간섭이 증가하여 하향 링

크 신호 세기가 감쇄된다. 따라서 저궤도 위성은 채널 상태를 고려하여 적절한 셀에게 빔 호핑을 해야 한다. 본 논문에서 저궤도 위성과 서비스 받는 셀 간의 신호 세기를 SINR로 측정하며, 시간 t에서 저궤도 위성 n과 셀 m간의 SINR은 다음과 같이 표현한다.

$$SINR_m^n(t) = \frac{SNR_m^n(t)}{\left(\sum_{j=1, j\neq m}^{M} l_j^n(t) I_j^n(t)\right)/\sigma^2 + 1}, \quad (8)$$

I_j^n는 동일 시간대에 저궤도 위성 n으로부터 선택된 셀에 의해 발생하는 간섭을 나타내고, σ^2는 잡음 전력을 표현한다. $SINR_m^n$ 값에 따라 변조 코딩 구성 (modulation and coding scheme, MCS) index가 결정되며, 다음과 같이 표현한다.

$$\gamma_{index} = 2^{SE_{index} \times \frac{CR_{index}}{1024}} - 1, \quad (9)$$

γ_{index}는 MCS index를 기반으로 정의된 SINR 값이다. SE_{index}와 CR_{index}는 각각 MCS index에 해당하는 주파수 효율성 (spectral efficiency, SE)와 부호화율 (code rate, CR)이다.

결과적으로, 저궤도 위성 n이 시간 t에서 셀 m에게 요구되는 랜덤한 트래픽 ζ_m에 따라 제공하는 처리 용량은 다음과 같이 계산된다.

$$C_m^n(t) = \zeta_m SE_{index}(t), \quad (10)$$

$C_m^n(t)$는 저궤도 위성 n과 셀 m 간의 하향 링크 처리 용량을 의미한다.

2.3 저궤도 위성 다중 빔 호핑 문제 구성

본 논문에서는 저궤도 위성이 제한된 채널 용량 내에서 최대한 많은 대한민국 전역의 셀 트래픽을 처리하는 것을 목표로 한다. 그러나 저궤도 위성이 다수의 셀에 동시에 서비스를 제공할 경우, 간섭으로 인한 신호 세기가 감쇄될 수 있다. 또한, 셀의 트래픽이 랜덤하게 발생하는 특성으로 고려하여 동적 자원 할당 및 간섭 관리 전략이 필요하다. 이러한 문제는 아래와 같은 수식으로 정의할 수 있다.

$$\max_{l_m^n} \sum_{m=1}^{M} \sum_{t=1}^{T} (C_m^n(t)), \quad (11)$$

$$\text{s.t. } B_n \geq \sum_{m=1}^{M} \zeta_m l_m^n(t), \quad (12)$$

$$\sum_{n=1}^{N} l_m^n(t) = 1, \quad (13)$$

$$\theta_m^n(t) \geq \theta_{\min}, \ \forall n, \forall m, \forall t \quad (14)$$

$$SINR_m^n(t) \geq \delta_{th}, \ \forall n, \forall m, \forall t \quad (15)$$

식 (11)은 다중 빔 시나리오에서 각 셀에 제공되는 용량을 최대화하는 것을 정의한다. 식 (12)는 제한된 위성 용량 내에서 지상 셀의 트래픽을 처리해야 함을 의미한다. 식 (13)은 특정 시간 t에서 한반도 셀 m은 오직 한 개의 저궤도 위성으로부터 서비스를 받는 것을 정의한다. 식 (14)는 최소 고도각 $\theta_{\min}$을 만족할 경우, 저궤도 위성이 해당 셀에 서비스를 제공할 수 있음을 의미한다. 마지막으로, 식 (15)는 저궤도 위성과 연결된 셀 간의 신호 세기가 SINR 임계 값 δ_{th}를 초과해야 함을 의미한다.

Ⅲ. DQN 기반 Earth-fixed 빔 호핑 알고리즘 설계

3.1 Markov Decision Process 설계

본 논문에서는 앞서 정의한 저궤도 위성 빔 호핑 문제를 마르코프 결정 과정 (Markov decision process, MDP)으로 공식화한다. MDP는 일반적으로 상태 (S), 행동 (A), 보상 (R), 천이 확률 (P), 감가율 (Υ)로 구성된다. 시간 t에서 S는 다음과 같이 표현한다.

$$s(t) = (v_n(t), \boldsymbol{Z}(t), B_n(t), \boldsymbol{L_n}(t)), \ \ s \in S, \quad (16)$$

$v_n(t)$는 시간 t에서 저궤도 위성 n의 ECEF 위치 좌표, $\boldsymbol{Z}(t)$는 시간 t에서 대한민국 내 모든 셀이 요구하는 랜덤한 트래픽을 의미하고, $B_n(t)$는 시간 t에서 저궤도 위성 n의 잔여 용량을 표현한다. $\boldsymbol{L_n}(t)$는 시간 t에서 저궤도 위성 n과 지상의 모든 셀 간 링크 형성 여부를 정의하며, 아래와 같이 표현한다.

$$\boldsymbol{L_n}(t) = \{l_1^n(t), l_2^n(t), ..., l_m^n(t)\}. \quad (17)$$

저궤도 위성은 최적의 빔 호핑을 하기 위해서는, 단계마다 적절한 A를 선택해야 하며, 시간 t에서 A는

다음과 같이 표현한다.

$$a(t) = (p_{t+1}, \bigcup_{m=1}^{M} (l_m^n(t))), \quad a \in A, \tag{18}$$

p_{t+1}는 저궤도 위성의 이동, $\bigcup_{m=1}^{M} (l_m^n(t))$는 시간 t에서 저궤도 위성과 대한민국 셀 간의 하향 링크 연결성 여부를 표현한다.

다중 빔 저궤도 위성이 빔 호핑을 통해 지상 셀의 랜덤한 트래픽을 효율적으로 처리하기 위해, 시간 t에서 R은 다음과 같이 표현한다.

$$r(t) = \omega_1 \sum_{m=1}^{M} l_m^n(t) C_m^n(t) + \omega_2 \mu^n, \quad r \in R, \tag{19}$$

ω_1, ω_2은 구성 요소들의 정규화를 위한 파라미터이다. $\sum_{m=1}^{M} l_m^n(t) C_m^n(t)$는 시간 t에서 빔 호핑을 통해 처리되는 모든 용량의 합을 표현한다. μ^n은 저궤도 위성 n이 적절하지 않은 셀을 선택하거나, 선택할 수 있는 셀이 있음에도 불구하고 p_{t+1} 행동을 취할 때 발생하는 패널티이다. R은 다중 빔 저궤도 위성 시나리오에서 지상 셀의 랜덤한 트래픽 분포를 효율적으로 처리하고, 가능한 많은 셀에 서비스를 제공할 수 있도록 설계한다.

Υ는 0에서 1 사이의 값으로 미래에 받을 보상을 현재 시점에서 고려할 때 감가하는 비율을 의미한다.

3.2 강화학습 기반 빔 호핑 알고리즘 구조

본 논문에서는 DQN 모델을 기반한 지구 고정 빔 호핑 알고리즘을 제안한다. 다중 빔을 방사하는 저궤도 위성을 에이전트로 설정하고, DQN 알고리즘을 활용하여 대한민국 셀에 효율적으로 빔 호핑을 수행한다. 알고리즘 1은 DQN 기반 지구 고정 빔 호핑 알고리즘 전체적인 과정을 설명한다. DQN에서는 에이전트가 행동-가치 함수 Q-value를 통해 정책 π을 평가하고 개선할 수 있으며, π에 따라 행동을 선택한다. Q-value는 벨만 방정식을 기반으로 다음과 같이 계산한다.

$$Q(s_t, a_t) = E[r_t + \gamma \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1})], \tag{20}$$

신경망을 이용하여 Q를 근사하는 DQN을 활용하며, 이를 통해 $Q(s,a;\theta) \approx Q(s,a)$로 설정한다[6]. θ는 Q-network의 가중치를 나타낸다.

알고리즘 1. DQN 기반 LEO 위성 빔 호핑 알고리즘
Algorithm 1. DQN-based BH algorithm for LEO satellite

Algorithm 1 DQN 기반 저궤도 위성 빔 호핑 알고리즘

초기값: 행동-가치 함수 $\mathbf{Q}$의 랜덤 가중치 θ, 목표 행동-가치 함수 $\hat{\mathbf{Q}}$의 가중치 $\hat{\theta} = \theta$, 리플레이 버퍼 $\mathbf{B}$의 용량 $\mathbf{B_C}$, minibatch 크기 $\mathbf{F}$, 총 에피소드 수 $\mathbf{E}$
출력값: 위성의 위치 $\mathbf{V}$, 셀의 서비스 여부 $\mathbf{a}$
for Episode = 1 to $\mathbf{E}$ **do**
  저궤도 위성 시스템 환경 초기화
  for timestep $t = 1$ to T **do**
    Decaying ϵ-greedy 방법을 통한 행동 a_t 선택
    $= \begin{cases} \arg\max_a Q(s_t, a; \theta) & (\text{확률 } 1-\epsilon) \\ \text{임의의 행동} & (\text{확률 } \epsilon) \end{cases}$
    샘플 (s_t, a_t, r_t, s_{t+1})을 $\mathbf{B}$에 저장
    if $\mathbf{B} > \mathbf{B_C}$ **then**
      $\mathbf{B}$로 부터 랜덤한 minibatch (s_j, a_j, r_j, s_{j+1}) 샘플 선택
      for $j = 1$ to $\mathbf{M}$ **do**
        $y_j = r_j + \gamma \max_{a'} \hat{Q}(s_{j+1}, a'; \hat{\theta})$
      end for
      가중치 θ의 관점에서 경사 하강 단계 수행
      $(y_j - Q(s_j, a_j; \theta))^2$ 수행
      매 스텝마다 $\mathbf{Q}$의 파라미터를 $\hat{\mathbf{Q}}$로 복사
    end if
  end for
end for

DQN은 학습의 안정성을 위해 target network를 사용한다. Target network는 Q-network와 동일한 구조를 가지지만, 일정 주기마다 θ를 복사하여 업데이트함으로써, 학습 과정에서 변화하는 Q-value의 급격한 변동을 완화하는 역할을 한다. Target network에서 계산된 Q-value는 다음과 같이 표현한다.

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-), \tag{21}$$

y_t는 target Q-value, θ^-는 target network의 가중치를 나타낸다.

에피소드가 시작되면 저궤도 위성 환경은 초기화되고, 에피소드는 여러 단계로 구성된다. 에이전트는 각 단계에서 decaying epsilon-greedy 방법을 사용하여 행동을 선택한다. 이 방법은 p_{t+1}의 행동을 선택할 때까지 epsilon 값을 점진적으로 감소시켜, 에이전트가 다양한 행동을 선택할 수 있도록 충분한 탐색을 보장한다. 각 단계에서 행동을 수행하면 (s_t, a_t, r_t, s_{t+1})이 replay buffer에 저장된다. Replay buffer에 일정량 이상의 데이터가 축적되면, 에이전트는 학습을 시작하며 무작위로 mini batch를 선택한다. 이후, 에이전트는 선택한 데이터를 기반으로 Q-value와 target Q-value를 계산하고, 두 값의 차이를 최소화하기 위해 손실 함수를 활용한다. 손실 함수는 다음과 같이 정의된다.

$$L(t) = (r_t + \gamma \max_a Q(s', a'; \hat{\theta}))^2 - (Q(s, a; \theta))^2, \quad (22)$$

이 과정을 통해 에이전트는 Q-value 예측의 정확성을 점진적으로 개선하며, 최적의 정책 π^*을 학습한다.

기존 강화학습 알고리즘에서는 에이전트가 다중 행동을 선택함에 따라 행동 공간의 크기가 급격히 증가한다. 예를 들어, 40개 셀 중에 10개에 빔 호핑을 수행하면, 행동 공간의 크기는 C_{40}^{10}이 된다. 이러한 거대한 행동 공간은 에이전트의 학습 복잡도를 지수적으로 상승시킨다. 이를 해결하기 위해 본 논문에서는 M개의 셀 중 1개의 셀을 반복 선택하여 행동 차원을 줄이는 방안을 활용한다[7].

Ⅳ. 저궤도 위성 빔 호핑 알고리즘 성능평가

4.1 시뮬레이션 환경 구성

본 논문에서는 SpaceX 문서를 기반으로 저궤도 위성 시스템을 설계하고[8], 안테나 패턴은 3rd generation partnership project (3GPP) 표준을 기반하여 설계한다[9]. 또한, 군집 저궤도 위성 중 동일 시간대에 한 개의 위성만 대한민국 내의 셀에 서비스를 제공한다고 가정한다. 대한민국 내 셀들은 handheld로 가정하며, 셀은 육각형 형태로, $25\sqrt{3}$ km 간격으로 배치한다. 표 1은 시뮬레이션에 사용된 파라미터를 나타낸다. MCS 표는 신호 세기에 따라 4-QAM 및 16-QAM 변조 방식을 동적으로 할당한다[10]. 또한, γ_{th}는 4-QAM MCS index에 해당하는 최저 SINR 값으로 정의한다.

표 1. 시뮬레이션 파라미터
Table 1. Simulation parameters

Parameters	Values
Altitude	530 km
Inclination	43°
M	40
$\theta_{\min}$	25°
B_n	570 MHz
δ_{th}	0.0192 dB
f_c	2 GHz
EIRP density	34 dBW/MHz
Satellite antenna aperture	2 m
Satellite antenna gain	30 dB
G_m/T_σ	-31.6 dB/T
k	-228.6 dBW/K/Hz
R_E	6,371 km
E	4,000
B_C	500,000
F	1,024
Step size	5 s

본 논문에서 사용한 neural network는 입력층, 3개의 은닉층, 출력층으로 구성된다. 은닉층은 각각 1,024개의 뉴런으로 구성되며, 각 층에서 정류 선형 유닛 (rectified linear unit, ReLU) 활성화 함수를 사용한다.

본 논문에서는 성능분석을 위해 두 가지 휴리스틱 알고리즘을 시뮬레이션에 적용한다. 첫 번째, Max selection 알고리즘은 저궤도 위성이 트래픽이 큰 셀부터 순차적으로 빔 호핑을 수행하는 방식이며, 두 번째, Min selection 알고리즘은 트래픽이 작은 셀부터 빔 호핑을 수행하는 방식이다. 비교 알고리즘은 신호 세기에 따라 동적으로 변조 방식을 할당하지 않으며, 4-QAM과 16-QAM 변조 방식을 각각 고정적으로 사용하는 경우를 고려한다.

4.2 시뮬레이션 결과

그림 2는 제안하는 DQN 기반 빔 호핑 알고리즘 과 비교 알고리즘의 에피소드에 따른 누적 보상을 나타낸다. 비교 알고리즘은 학습 과정을 거치지 않기 때문에, 일정한 보상 값을 유지한다. 변조 방식이 4-QAM일 때, Max selection과 Min selection의 보상 값은 각각 329.9와 208.3을 얻으며, 16-QAM일 때는 각각 598.5와 335.9의 보상 값을 얻는다. 제안한 알고리즘은 초기에 비교 알고리즘보다 낮은 보상을 얻지만, 1,550 에피소드를 넘어서면서 약 652의 높은 보상 값에 수렴하는 것을 확인할 수 있다. 이는 제안한 알고리즘이 최적의

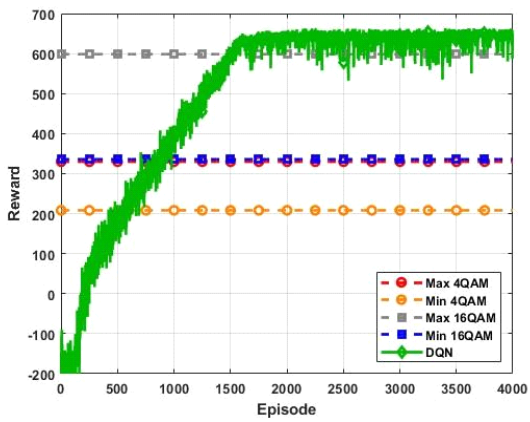


그림 2. 에피소드에 따른 누적된 reward
Fig. 2. Accumulated reward value per episode

빔 호핑 동작을 학습하여 더 나은 성능을 보이는 것을 의미한다.

그림 3은 각 알고리즘에서 단계별로 할당된 대한민국 셀의 누적 용량을 나타낸다. 제안하는 알고리즘은 97.8503 Gbps의 누적 용량을 제공한다. 변조 방식이 4-QAM 일 때, Max selection과 Min selection은 각각 누적 용량 42.6209 Gbps, 22.2796 Gbps를 얻는다. 16-QAM의 경우, 비교 알고리즘은 각각 87.0451 Gbps, 42.8932 Gbps의 결과를 보인다. 결과적으로, 제안하는 알고리즘은 비교 알고리즘에 비해 각각 56.55%, 77.79%, 11.36%, 56.80% 향상된 성능을 보인다. Max selection 알고리즘은 트래픽이 많은 셀부터 우선적으로 처리하기 때문에 비교적 높은 누적 용량을 보이지만, Min selection 알고리즘은 트래픽 크기가 작은 셀을 주로 선택하기 때문에 간섭의 영향을 크게 받아 가장 낮은 누적 용량이 측정된다.

그림 4는 각 단계별로 위성의 잔여 용량을 표현한다. 변조 방식과 관계없이, 각 알고리즘의 위성 잔여 용량은 일관된 결과를 보인다. Max selection 알고리즘은 트래픽이 많은 셀부터 우선적으로 빔 스케줄링함으로써 Min selection 알고리즘에 비해 제한된 자원을 효율적으로 활용한다. 반면, Min selection은 간섭의 영향으로 낮은 SINR 값을 보이며, 많은 셀의 트래픽을 처리하지 못해 잔여 용량이 많이 남는 결과를 보인다. 제안한 알고리즘은 위성 용량을 최대한 활용하여 빔 호핑을 수행하며, 비교 알고리즘에 비해 최소한의 잔여 용량을 남긴다. 이는 자원을 효율적으로 관리하고 지상 셀의 랜덤한 트래픽을 효율적으로 처리함을 보여준다. 이처럼, 제안된 알고리즘은 최적의 빔 호핑 전략을 학습함으로써 자원을 효율적으로 사용하여 전체적인 네트워크 성능을 향상시키기 때문에 비교 알고리즘보다 뛰어난 결과를 보인다.

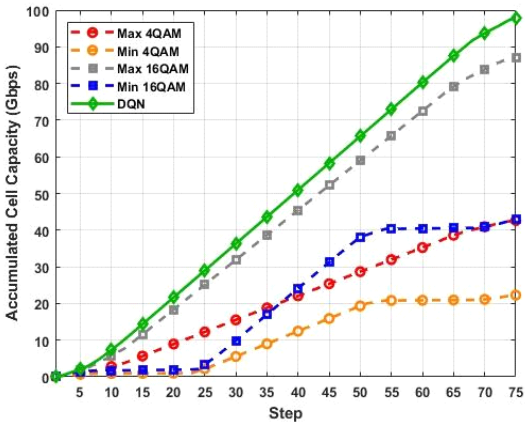


그림 3. 각 단계에서 셀에 제공되는 누적 채널 용량
Fig. 3. Accumulated channel capacity provided to cells by step

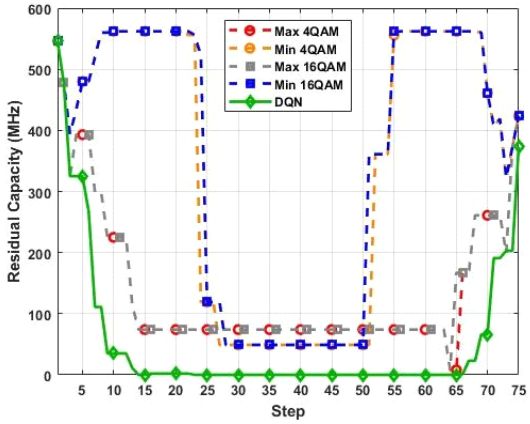


그림 4. 각 단계에서 저궤도 위성의 잔여 용량
Fig. 4. Residual LEO capacity per step

V. 결 론

본 논문에서는 저궤도 위성의 제한된 시간 및 용량 내에서 대한민국 셀 트래픽 분포를 고려한 DQN 기반 빔 호핑 알고리즘을 제안한다. 제안하는 알고리즘은 학습을 통해 트래픽 수요가 랜덤하게 분포된 상황에서 최적의 빔 호핑을 수행함으로써 위성의 자원을 최대한 효율적으로 활용할 수 있음을 보였다. 시뮬레이션 결과, 제안한 알고리즘은 기존 알고리즘에 비해 더 높은 누적 보상을 달성하며, 채널 용량 및 자원 활용 측면에서도 우수한 성능을 보인다. 결과적으로, 기존 휴리스틱 알고리즘에 비해 더 높은 효율성과 유연성을 갖춘 방안을 제시하였다.

References

[1] J. T. Lee, W. J. Lee, and J. H. Kim, "Performance evaluation of location-based conditional handover scheme using LEO satellites," in *Proc. ICTC 2023*, Jeju, Oct. 2023.

[2] Q. Zhao, Y. Hu, Z. Pang, and D. Ren, "Beam hopping for LEO satellite: challenges and opportunities," in *Proc. Int. Conf. Culture Oriented Sci. Technol.*, pp. 319-324, Aug. 2022.

(https://doi.org/10.1109/CoST57098.2022.00072)

[3] L. Chen, et al., "The next generation of beam hopping satellite systems: Dynamic beam illumination with selective precoding," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2666-2682, Apr. 2023. (https://doi.org/10.1109/TWC.2022.3213418)

[4] Z. Lin, Z. Ni, L. Kuang, C. Jiang, and Z. Huang, "Dynamic beam pattern and bandwidth allocation based on multi-agent deep reinforcement learning for beam hopping satellite systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 3917-3930, Apr. 2022. (https://doi.org/10.1109/TVT.2022.3145848)

[5] J. Tang, D. Bian, G. Li, J. Hu, and J. Cheng, "Optimization method of dynamic beam position for LEO beam-hopping satellite communication systems," *IEEE Access*, vol. 9, pp. 57578-57588, Apr. 2021. (https://doi.org/10.1109/ACCESS.2021.3072104)

[6] V. Mnih, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529-533, Feb. 2015. (https://doi.org/10.1038/nature14236)

[7] X. Hu, et al., "Dynamic beam hopping method based on multi-objective deep reinforcement learning for next generation satellite broadband systems," *IEEE Trans. Broadcast.*, vol. 66, no. 3, pp. 630-646, Sep. 2020. (https://doi.org/10.1109/TBC.2019.2960940)

[8] *FCC Partially Grants SpaceX Gen2 Broadband Satellite Application*, Dec. 2020. (https://www.fcc.gov/document/fcc-partially-grants-spacex-gen2-broadband-satellite-application)

[9] 3GPP TR 38.811, "*Study on new radio (NR) to support non-terrestrial networks* (Release 16)," v.15.4.0, Sep. 2020.

[10] T. E. Humphreys, P. A. Iannucci, Z. M. Komodromos, and A. M. Graff, "Signal structure of the Starlink ku-band downlink," *IEEE Trans. Aerosp. Electron. Syst.*, Oct. 2023. (https://doi.org/10.1109/TAES.2023.3268610)

문 태 한 (Taehan Moon)

2024년 : 아주대학교 전자공학과 학사 졸업
2024년~현재 : 아주대학교 AI융합네트워크학과 석사과정
<관심분야> 비지상 네트워크, 자원할당

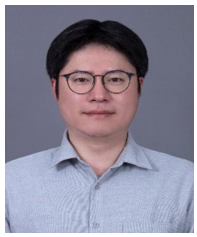
이 재 열 (Jaeyeol Lee)

2023년 : 순천향대학교 스마트자동차학과 학사 졸업
2023년~현재 : 아주대학교 우주전자정보공학과 석사과정
<관심분야> 다계층 네트워크, 강화학습

김 태 윤 (Tae-Yoon Kim)

2020년 : 아주대학교 전자공학과 학사 졸업
2020년~현재 : 아주대학교 AI융합네트워크학과 석박통합과정
<관심분야> 비지상 네트워크, 강화학습, 자원할당, 빔 호핑

이 영 포 (Youngpo Lee)



2008년 : 성균관대학교 전자공학과 학사 졸업
2010년 : 성균관대학교 전자공학과 석사 졸업
2014년 : 성균관대학교 전자공학과 박사 졸업
2014년~2023년 : 삼성전자 MX 사업부
2023년~현재 : SK텔레콤 New Connectivity 팀
<관심분야> 5G 셀룰러 통신, 저궤도 이동통신, 5G & 6G 표준 기술

김 동 욱 (Dongwook Kim)

2002년 : 아주대학교 전산학과 학사 졸업
2004년 : KAIST 전산학과 석사 졸업
2009년 : KAIST 전산학과 박사 졸업
2009~2017년 : 삼성전자 네트워크사업부 Air 시스템 연구실
2017년~현재 : SK텔레콤 New Connectivity
<관심분야> 5G 셀룰러 통신, UAM 통신, 저궤도 이동통신, 5G & 6G 표준 기술

류 탁 기 (Takki Yu)

1999년~2006년 : 연세대학교 전자전기공학과 학사 및 석박사 졸업
2009년~2010년 : Standford 대학교 박사 후 연구원
2019년~현재 : NGMN Alliance Board Director
2022년~현재 : 6G포럼 집행위원회 위원
2022년~현재 : SK텔레콤 Infra 기술 부사장
2023년~현재 : GSMA Technology Group 임원
2023년~현재 : 오프랜인더스티리얼라이언스(ORIA) 대표의장
<관심분야> 5G & 6G 셀룰러 통신, 이동통신 네트워크, 측위

김 재 현 (Jae-Hyun Kim)

1987년~1996년 : 한양대학교 전산과 학사 및 석박사 졸업
1997년~1998년 : 미국 UCLA 전기전자과 박사 후 연수
1998년~2003년 : Bell Labs, NJ, USA, 연구원
2003년~현재 : 아주대학교 전자공학과 교수
2018년~현재 : 6G포럼 디지털공간기술위원장
2021년~현재 : 위성통신포럼 대외협력위원장
2022년~현재 : 아주대학교 정보통신대학 학장
2024년~현재 : 과학기술정보통신부 전파연구센터 센터장
<관심분야> 5G/6G, 저궤도 위성 시스템, 국방 전술 네트워크, 무선 MAC 프로토콜