저궤도 위성 시스템에서 Queue delay를 고려한 심층 강화학습 기반 최적 빔 호핑 기법

김유빈, 이호원

아주대학교

{youbin1323, howon}@ajou.ac.kr

Deep Reinforcement Learning-Based Beam Hopping Optimization in LEO Satellite Systems with Consideration of Queue Delay

Youbin Kim, Howon Lee

Ajou Univ.

요 약

본 논문은 저궤도 위성 통신 환경에서 지상 트래픽 수요의 지리적 불균형성과 시변적인 특성을 반영한 무선 자원 할당 기법을 제안한다. 특히, 빔 호핑(Beam hopping, BH) 기법을 활용하여 큐 지연(queue delay)을 고려한 빔 할당 및 대역폭 분배 방식을 설계하고, 이를 심층 강화학습(Deep reinforcement learning, DRL) 기반으로 최적화한다. 또한, 트래픽 예측값을 활용함으로써 보다 빠르고 안정적인 학습 수렴을 달성하였다.

Ⅰ. 서 론

최근 위성 통신 시스템은 넓은 커버리지와 높은 전송 용량을 기반으로, 지상 통신을 보완할 수 있는 차세대 인프라로 주목받고 있다 [1]. 그러나 지상 트래픽 수요는 지리적으로 불균형하고 시간에 따라 변동성이 크기 때문에, 제한된 위성 자원을 효율적으로 활용하는 방안이 중요한 과제로 대두되고 있다 [2]. 이에 따라, 다중 빔 위성 (Multi-beam satellite, MBS)에서는 위상 배열 안테나의 발전을 바탕으로 빔 호핑 (Beam hopping, BH) 기술이 도입되어 시간 및 공간 자원의 활용 효율을 높이고 있다 [3]. BH 시스템에서는 매 타임 슬롯마다 활성화할 빔 조합, 즉 BH 패턴의 설계가 시스템 성능에 큰 영향을 미친다. 하지만 기존의 휴리스틱 방식은 시변적이고 불규칙한 트래픽에 유연하게 대응하기 어려우며, 패턴 탐색의 복잡도 또한 높다는 한계가 있다 [4]. 이에 본 논문에서는 사용자 지연 요구를 반영하기 위해 큐 지연(queue delay)을 고려하고, 이를 기반으로 심층 강화학습 (Deep reinforcement learning, DRL)을 활용한 최적 BH 기법을 제안한다.

Ⅱ. 시스템 모델 및 제안방안

본 논문에서는 저궤도 위성이 그림 1과 같이 M개의 빔을 이용하여 K개의 지상 셀에 통신 서비스를 제공하는 시나리오를 고려한다. 각 빔은 동일한 대역폭을 사용하며, 전체 대역폭은 N개의 주파수 채널로 분할되어 셀별로 할당된다. 지상 트래픽 수요는 Telecom Italia Big Data Challenge 2014에서 수집된 데이터를 기반으로 구성된다 [5]. 위성은 각 셀별로 처리되지 않은 데이터를 저장할 수 있는 개별 큐(queue)를 보유하고 있다. 제안하는 프레임워크는 LSTM (Long short-term memory)을 활용하여 예측된 트래픽 수요를 기반으로, 각 타임스텝마다 적절한 빔 전환을 수행함으로써 최적의 통신 서비스를 제공한다.

이러한 문제를 해결하기 위해, 본 논문에서는 위성 빔에 두 개의 DRL 에이전트를 적용하였다. 제안하는 두 에이전트의 MDP (Markov decision process)는 다음과 같다. 타임스텝 t에서 두 에이전트의 상태정보는 글로벌 상태정보로 동일하게 정의되며, 각 큐에 저장된 트래픽 양과 향후 L 타임스텝까지의 트래픽 예측값으로 구성된다. 각 에이전트의 행동은 다음과 같다: 첫 번째 에이전트는 다음 타임스텝 t+1에서 서비스할 셀을 선택하고, 두 번째 에이전트는 해당 빔의 대역폭 분할 방안을 결정한다. 두 에이전트 간의 협력적 학습을 위해 설계된 통합 보상 함수는 식 (1)과 같다.

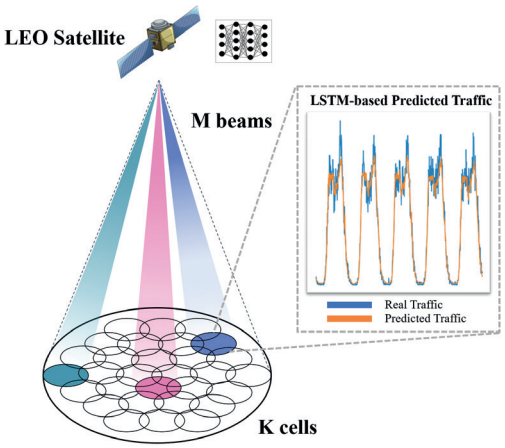

그림 1. System Model

$$r_t = \alpha \frac{\sum_{k=1}^{K}(x_t^k \cdot |B_t^k| \cdot B_{ch}\log(1+\gamma_t^k))}{\sum_{k=1}^{K} D_t^k} \times \beta e^{\frac{\sum_{k=1}^{K} s_t^k}{\vartheta_s}} \times \mu e^{\frac{\max\{\tau_t^k\}-\min\{\tau_t^k\}}{\vartheta_F}} \quad (1)$$

여기서 $x_t^k, |B_t^k|, B_{ch}, \gamma_t^k$는 각각 타임스텝 t에서 k번째 빔이 활성화되었는지를 나타내는 변수, 해당 빔이 사용하는 채널의 개수, 채널 대역폭 및 SINR (Signal to interference plus noise ratio)을 의미한다. $\sum_{k=1}^{K} D_t^k$는 큐에 저장되어 있는 트래픽의 양을 의미하며, $\sum_{k=1}^{K} S_t^k$는 타임스텝 t에서 빔 전환 횟수를 의미한다. ϑ_s, ϑ_F는 보상의 크기를 조정하는 변수이며 α, β, μ는 보상 별 가중치이다. τ_t^k는 타임슬롯 t에서 k 셀의 평균 큐 지연을 의미하며 식 (2)와 같이 계산된다.

$$\tau_t^k = \frac{\sum_{t=1}^{T} U \cdot \phi_{t,u}^k}{\sum_{t=1}^{T} \phi_{t,u}^k} \quad (2)$$

여기서 $\phi_{t,u}^k$은 큐 k에서 u 타임스텝만큼 기다린 데이터 패킷 수를 의미한다.

Ⅲ. 시뮬레이션 결과

A. 시뮬레이션 환경

시뮬레이션의 전체적인 파라미터는 표 1과 같다.

표 1. Simulation Parameters

Parameter	Value	Parameter	Value
# of Cells (K)	16	Noise power spectral density	-204 [dbm/Hz]
# of Beams (M)	6	Total bandwidth	200 [MHz]
# of Channels (N)	4	# of episodes	10000
Satellite altitude	1200 [km]	# of iterations	200
Carrier frequency	20 [GHz]	Learning rate	10^{-3}
Total transmit power	200 [Watt]	Discount factor	0.95

B. 시뮬레이션 결과

본 논문에서는 제안방안의 성능을 검증하기 위해 벤치마크 알고리즘들과 비교 실험을 수행하였다. 벤치마크 알고리즘은 다음과 같다.

- Random Action: 각 에이전트가 모든 에피소드에서 무작위로 행동을 선택하는 방식
- DQN-Max-Delay-based Beam Hopping (DQN-MD-BH): 평균 큐 지연이 가장 큰 셀을 우선적으로 선택하여 서비스하는 방식
- DQN-Max-Traffic-based Beam Hopping (DQN-MT-BH): 트래픽 수요가 가장 높은 셀을 우선적으로 선택하여 서비스하는 방식
- DQN-Full-Bandwidth per Beam (DQN-FBB): 각 빔이 전체 채널을 모두 사용하여 서비스를 제공하는 방안
- DQN-Single-Channel per Beam (DQN-SCB): 각 빔이 하나의 채널만을 사용하여 서비스를 제공하는 방식
- Proposed w/o Traffic Prediction (Proposed w/o TP): 제안방안에서 트래픽 예측값 없이 학습을 수행하는 방식

그림 2는 트래픽 예측값을 활용하는 방안과 활용하지 않는 방안 간의 성능을 비교한 시뮬레이션 결과를 나타낸다. 시뮬레이션 결과를 통해, 트래픽 예측값을 활용한 제안방안이 보다 빠른 수렴 속도를 보임을 확인할 수 있으며, 이는 학습 단계에서 불필요한 탐색을 줄이고 효율적인 정책 학습을 유도함을 보여준다.

그림 3은 제안방안과 다양한 벤치마크 알고리즘 간의 성능 비교를 나타낸다. DQN-MD-BH 및 DQN-MT-BH는 각각 대기열 지연이 가장 크거나 트래픽 수요가 가장 높은 셀을 우선적으로 선택하지만, 빔 전환에 따른 오버헤드를 고려하지 않기 때문에 제안방안에 비해 낮은 보상으로 수렴하는 경향을 보인다. DQN-SCB는 각 빔에 하나의 채널만 할당하는 구조로 인해 처리량이 매우 낮아, 전체 알고리즘 중 가장 낮은 보상으로 수렴하였다. 한편, DQN-FBB는 모든 빔에 전체 대역폭을 할당하여 비교적 높은 보상을 달성하였으며, 이는 지상 트래픽 수요가 높은 환경에서는 전체 대역폭을 사용하는 방식이 유리할 수 있음을 시사한다. 하지만 트래픽 수요를 고려하지 않고 모든 빔에서 전체 대역폭을 동시에 사용하므로, 수요가 낮은 셀에서는 자원이 낭비되어 전체 자원 활용 효율이 저하된다.

Ⅳ. 결론

본 논문에서는 지상 트래픽의 지리적 불균형성과 시변적 특성 그리고 큐 지연을 고려한 DRL 기반의 최적 빔 호핑 기법을 제안하였다.

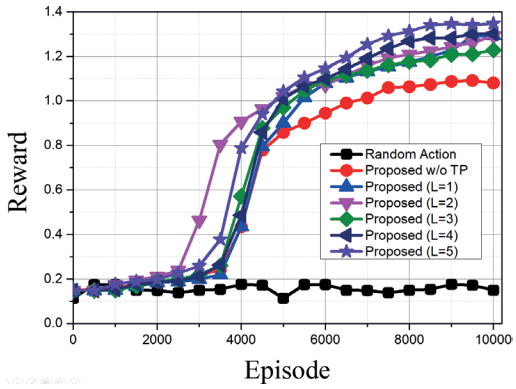

그림 2 Reward vs. episode ($\alpha=0.6, \beta=0.2, \mu=0.2$)

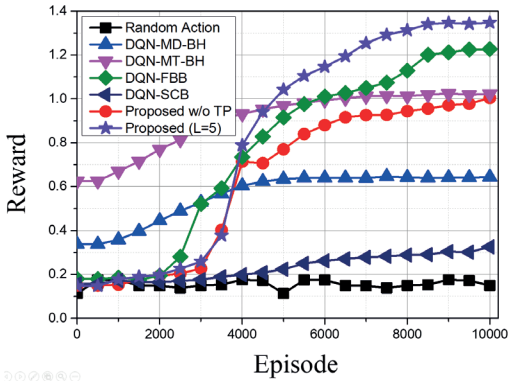

그림 3 Reward vs. episode ($\alpha=0.6, \beta=0.2, \mu=0.2$)

제안방안은 LSTM을 활용한 트래픽 예측을 통해 학습의 수렴 속도와 안정성을 향상시켰으며, 빔 할당과 대역폭 분할을 담당하는 두 개의 에이전트 구조를 통해 자원의 운용 효율 또한 향상시켰다. 시뮬레이션 결과를 통해 제안방안이 벤치마크 알고리즘에 비해 셀 간 지연을 완화하고 전체 네트워크 성능을 효과적으로 향상시킴을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. RS-2024-00396992, 저궤도 위성통신 핵심 기술 기반 큐브 위성 개발)과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. 2022-0-00704, 초고속 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 원천기술 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, Jun. 2023.

[2] J. P. Choi et al., “Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks,” in IEEE Transactions on Wireless Communications, vol. 4, no. 6, pp. 2983-2993, Nov. 2005.

[3] J. Lei et al., “Multibeam satellite frequency/time duality study and capacity optimization,” in Journal of Communications and Networks, vol. 13, no. 5, pp. 472-480, Oct. 2011.

[4] Z. Lin et al., "Dynamic Beam Pattern and Bandwidth Allocation Based on Multi-Agent Deep Reinforcement Learning for Beam Hopping Satellite Systems," in IEEE Transactions on Vehicular Technology, vol. 71, no. 4, pp. 3917-3930, Apr. 2022.

[5] Barlacchi et al., “A multi-source dataset of urban life in the city of Milan and the Province of Trentino,” Scientific data 2 (2015).