트랜스포머 기반 강화학습을 활용한 ISM 대역 채널 할당 및 전력 제어 기법

문찬호, 오민택, 최진석
한국과학기술원

{ghcksans, ohmin, jinseok}@kaist.ac.kr

Channel Allocation and Power Control in the ISM Band Using Transformer-Based Reinforcement Learning

Chanho Moon, Mintaek Oh, Jinseok Choi
Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 2.4GHz ISM (Industrial, Scientific, and Medical) 대역에서의 효율적인 채널 할당 및 전력 제어를 위한 트랜스포머 기반 강화학습 기법을 제안한다. 기존의 주파수 도약 기법인 AFH (Adaptive Frequency Hopping)와 FHSS (Frequency Hopping Spread Spectrum)는 동적 채널 환경에서 효과적으로 대응하지 못하는 한계가 있다. 제안 모델은 트랜스포머 구조와 PPO(Proximal Policy Optimization) 알고리즘을 결합하여 채널 이득과 SINR 정보를 활용한 최적 자원 할당을 수행한다. 실험 결과, 제안 기법은 AFH 대비 평균 전송률이 20.8%, 공정성 지수는 15.3% 향상되었으며, FHSS 대비 각각 35.9%와 21.3%의 성능 향상을 보였다. 특히 관측 간격이 증가해도 성능 저하가 적어 불완전한 관측 환경에서도 효과적으로 동작함을 확인하였다.

Ⅰ. 서 론

ISM 대역은 비면허 주파수 대역으로 다양한 무선 장치들이 공존하면서 발생하는 간섭 문제가 통신 성능에 큰 영향을 미친다[1]. 기존의 AFH (Adaptive Frequency Hopping)나 FHSS (Frequency Hopping Spread Spectrum)와 같은 간섭 회피 및 보안 기법들은 동적 채널 환경에서의 적응력이 제한적이며, 최근 등장한 강화학습 기반 접근법 역시 시간적 종속성이 있는 채널 상태를 효과적으로 모델링하는 데 한계가 있다[2]. 본 논문에서는 시계열 데이터 처리에 효과적인 트랜스포머 구조[4]를 PPO 알고리즘[3]과 결합하여, ISM 대역 환경에서 채널 할당 및 전력 제어를 위한 새로운 접근법을 제안한다. 제안 모델은 채널 이득과 SINR 정보를 활용하여 다수의 사용자 단말(UT)에 대한 최적 자원 할당을 수행하며, 미래 채널 상태 예측을 통해 동적 환경에서의 적응력을 향상시키는 장점이 있으며, 공정성[5]-[7] 까지 고려하여 최적화가 진행된다.

Ⅱ. 본론

본 연구에서는 다음과 같은 ISM 대역 환경을 고려한다. 기지국(BS)은 C 개의 채널을 통해 최대 N 개의 사용자 단말(UT)과 통신하며, 각 UT 는 시간에 따라 변화하는 채널 상태를 가진다. 우리의 목표는 다음과 같은 최적화 문제를 해결하는 것이다.

$$\max_{\{c_i,p_i\}} \sum_{i=1}^{N} R_i + \alpha \cdot J$$

$$subject\ to \sum_{i=1}^{N} p_i \le P_{toal},$$

$$\mathrm{c_i} \in \{1,2,\dots,C\}, \quad \forall i \in \{1,2,\dots N\}.$$

여기서 R_i는 UT i의 전송률, J는 Jain 의 공정성 지수로 모든 UT 간 전송률의 균등함을 측정하는 지표, α 는 공정성과 전송률 간의 균형을 조절하는 가중치($\alpha = \max(1.0, \bar{R} \cdot 0.25)$), p_i는 UT i에 할당된 전력, P_{total}은 총 가용 전력, c_i는 UT i 에 할당된 채널이다. 이 최적화 문제는 전체 전송률과 시스템의 공정성을 동시에 고려하여 네트워크 성능을 최대화하는 것을 목표로 한다. t 시점에서 UT i의 채널 j에 대한 SINR 은

$$\mathrm{SINR}_{i,j}^{(t)} = \frac{p_i \cdot g_{i,j}^{(t)}}{\sum_{\substack{k=1 \\ k\neq i}}^{N} p_k \cdot g_{k,j}^{(t)} \cdot \mathbf{1}(c_k = j) + N_0 + I_{ext,j}^{(t)}}.$$

여기서 $g_{i,j}{}^{(t)}$는 채널 이득, N_0는 열 잡음, $I_{ext,j}^{(t)}$는 외부 간섭(주로 Wi-Fi)을 나타낸다. 각 UT 의 전송률은

$$R_i^{(t)} = B \cdot \log_2\left(1 + \mathrm{SINR}_{i,c_i}^{\mathrm{t}}\right).$$

여기서 B는 채널 대역폭이다. 시스템의 공정성은 Jain 의 공정성 지수로 평가된다.

$$J = \frac{\left(\sum_{i=1}^{N} R_i\right)^2}{N \cdot \sum_{i=1}^{N} R_i^2}.$$

이 지수는 0 과 1 사이의 값을 가지며, 1 에 가까울수록 모든 UT 가 동등한 전송률을 얻는 공정한 상태를 의미한다. 강화학습 에이전트의 보상(reward) 함수는

$$Reward_{total} = \sum_{i=1}^{N} R_i + \alpha \cdot J.$$

본 논문에서는 트랜스포머 구조[4]를 특징 추출기로 활용하는 PPO 기반 강화학습 모델을 제안한다. 제안 모델은 4 차원 텐서(UT, 채널, 특성 종류, 시간 히스토리)를 관측 상태로 사용하며, 특성은 채널 이득과 SINR 으로 구성된다. 행동 공간은 채널 선택 확률 벡터($\mathbb{R}^C$)와 전력 할당 비율 벡터($\mathbb{R}^N$)로 구성되어 채널 간섭을 최소화하고 시스템 성능을 향상시킨다. 트랜스포머의 Self-Attention 메커니즘은 시계열 데이터의 전역적 의존성을 효과적으로 포착하며, 보조 네트워크는 미래 채널 상태 예측을 통해 선제적 자원 할당을 가능하게 한다. 본 연구는 PPO 알고리즘[3]의 기본 손실 함수에 미래 채널 상태 예측 손실을 통합하였다.

$$L_{\text{total}} = L_{\text{PPO}} + \lambda \cdot L_{\text{prediction}}.$$

여기서 L_{PPO}는 [3]에서 제안된 표준 PPO 손실 함수이고, $L_{\text{prediction}}$ 은 미래 채널 상태 예측을 위한 MSE 기반 손실로, 시간적 예측 능력을 강화한다.

성능 비교를 위해 두 가지 기준 방식을 구현하였다. 첫째, AFH 방식은 각 UT 가 일정 주기로 채널 SINR 을 측정하여 상위 채널을 선택하는 방식이다. AFH 에서는 전력 할당을 위해 다음과 같은 waterfilling 알고리즘을 사용하였다.

$$p_i = \max\left(\mu - \frac{1}{SINR_i}, 0\right).$$

여기서 μ 는 총 전력 제약을 충족하기 위한 수위 매개변수이다. 둘째, FHSS 방식은 매 시점 완전 랜덤한 채널 선택을 수행하고 각 UT 에 균등한 전력을 할당한다. 반면, 제안하는 RL 방식은 채널 선택과 전력 할당 정책을 모두 학습하여 더 효율적인 자원 분배가 가능하다.

실험은 $10 \times 10 \times 10 m^3$ 3 차원 공간 내에 다양한 수의 UT 로 구성된 환경에서 수행되었다. 주요 파라미터는 다음과 같다: 채널 수 10 개, 최대 UT 수 10 개, 히스토리 길이 10, 에피소드 길이 1000 타임스텝이다.

실험 결과, 제안된 트랜스포머 RL 기법은 4.297 Mbps 의 평균 전송률과 0.370 의 공정성 지수를 달성하여 AFH(3.558 Mbps, 0.321)와 FHSS(3.163 Mbps, 0.305) 대비 우수한 성능을 보였다(그림 1). 또한 관측 간격(0~100 타임스텝)에 따른 성능 평가에서, AFH 는 관측 간격 증가 시 채널 정보 부족으로 성능이 급격히 저하되었고, FHSS 는 랜덤 선택 방식으로 일정한 성능을 유지했다. 반면 제안된 트랜스포머 RL 은 시간적 패턴 인식과 미래 채널 예측 능력으로 관측 간격이 증가해도 성능이 안정적으로 유지되어, 불완전한 관측 환경에서 더욱 효과적임을 입증하였다(그림 2).

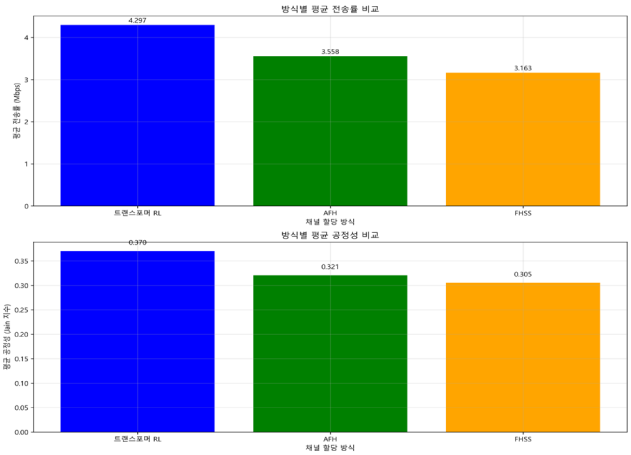

그림 1. 채널 할당 방식별 성능 비교

Ⅲ. 결론

본 논문에서는 ISM 대역에서의 효율적인 채널 할당 및 전력 제어를 위한 트랜스포머 기반 강화학습 기법을 제안하였다. 이 기법은 트랜스포머 구조와 PPO 알고리즘을 결합하여 채널 이득과 SINR 정보를 통합적으로 활용하고, 미래 채널 상태 예측을 통해 동적 환경에서의 적응력을 향상시킨다. 전송률과 공정성을 동시에 최적화하는 보상 함수를 설계하였으며, 커리큘럼 학습 전략을 도입하여 복잡한 다중 UT 환경에서도 효과적인 학습이 가능하게 하였다. 실험 결과, 제안 기법은 AFH 및 FHSS 대비 각각 20.8%/35.9%의 전송률 향상과 15.3%/21.3%의 공정성 개선을 달성하였으며, 관측 간격이 증가해도 성능 저하가 적어 실제 ISM 대역 환경에서 효율적인 스펙트럼 활용과 간섭 관리를 위한 실용적 솔루션으로 활용될 수 있을 것으로 기대된다.

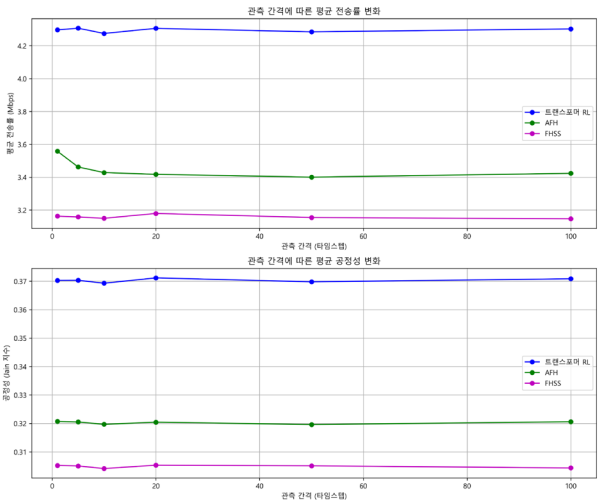

그림 2. 관측 간격에 따른 성능 변화

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원 (KRIT-CT-22-078) 및 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2024-00395824, (총괄 1-세부 2) Upper-mid Band 를 지원하는 Cloud virtualized RAN (vRAN) 시스템 기술 개발)

참 고 문 헌

[1] Zhao, Q., & Sadler, B. M. (2007). A survey of dynamic spectrum access. *IEEE signal processing magazine,* 24(3), 79-89.

[2] Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y. C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. IEEE communications surveys & tutorials, 21(4), 3133-3174..

[3] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[4] Ashish, V. (2017). Attention is all you need. Advances in neural information processing systems, 30, I.

[5] D. Kim, J. Choi, J. Park, and D. K. Kim "Max-Min fairness beamforming with rate-splitting multiple access: Optimization without a toolbox." *IEEE Wireless Communications Letters,* 12.2 (2022): 232-236.

[6] S. Lee, E. Choi, and J. Choi, “Max-Min Fairness Precoding for Physical Layer Security with Partial Channel Knowledge,” *IEEE Wireless Communications Letters*, vol. 12, no. 9, pp. 1637-1641, Sep. 2023

[7] S. Kim, E. Choi, M. Oh, and J. Choi, “Artificial Noise-aided Max-Min Fairness Secrecy Precoding with Partial Wiretap Channel Knowledge,” *IEEE Transactions on Vehicular Technology*, Early Access