강화학습을 이용한 IEEE 802.11be 동기화 전송 성능 향상

김병창, 권 람, 한승주, 박은찬
동국대학교-서울 정보통신공학과

mbc991028@dongguk.edu, lamk@dongguk.edu, tmdwn0324@dgu.ac.kr, ecpark@dongguk.edu

Enhancing IEEE 802.11be Synchronous Transmission Performance Using Reinforcement Learning

Byeongchang Kim, Lam Kwon, Seungjoo Han, Eun-Chan Park

요 약

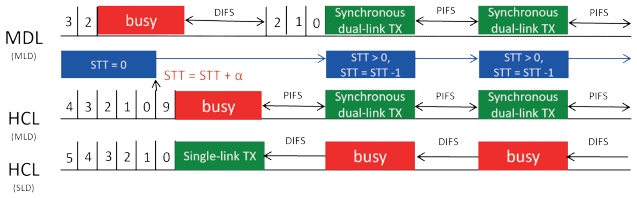
본 논문은 IEEE 802.11be 기반 Wi-Fi 7의 다중 링크 동작 환경에서 동시 송수신이 불가능한 Non Simultaneous Transmit and Receive (NSTR) 장치의 동기화 전송 성능 개선을 위한 연구이다. 본 연구의 선행연구로서 기존에 제안된 Contention-Less Synchronous Transmission (CLST) 기법은 전송 기회를 토큰화하여 다중 링크 전송을 여러 번 진행하면서 백오프 시간을 줄이고 다중 링크 단말과 단일 링크 단말의 처리율과 공정성을 증가시켰다. 그러나 이 기법의 성능은 제어 파라미터에 크게 영향을 받으며 다양한 환경에서 최적의 파라미터값을 도출하기 어려운 문제점을 지닌다. 이를 해결하기 위해 본 논문에서는 CLST 기법의 핵심 파라미터(α)를 강화학습을 통해 조절하는 방안을 제안한다. 제안하는 기법은 다중 링크 단말과 단일 링크 단말이 공존하는 혼잡한 링크의 처리율과 공정성 지표를 조합하여 강화학습의 보상으로 설정하고 ε-Greedy 정책을 이용하여 최적의 파라미터 값을 도출하는 것을 목표로 하였다. 모의실험을 통해 제안된 방법이 기존 기법보다 처리율과 공정성 성능을 개선한 것을 확인하였다.

Ⅰ. 서 론

IEEE 802.11be 기반의 Wi-Fi 7은 차세대 무선랜 표준으로, Medium Access Control (MAC) 계층에서 다중 링크 동작을 제안하였다. 다중 링크 동작 환경에서는 동시 송수신이 가능한 다중 링크 단말 (STR-MLD: Simultaneous Transmission and Reception - Multi Link Device), 동시 송수신이 불가능한 다중 링크 단말 (NSTR-MLD: Non-STR MLD), 단일 링크 단말 (SLD: Single-Link Device)의 세 가지 유형의 단말이 공존하는 상황을 고려해야 한다. 링크 간 또는 장치 내부의 간섭 문제로 인해 NSTR-MLD의 경우 다중 링크에서 비동기 전송을 하는 경우 전송 실패나 불필요한 전송 억제가 발생할 수 있어, 이러한 단말을 위한 다양한 동기 전송 방식이 제안되었다. 특히, 본 연구진의 선행연구로서 제안된 동기 전송 방식인 CLST는 NSTR-MLD와 SLD의 경쟁이 많은 링크 Heterogeneous Coexistence Link (HCL)와 경쟁이 상대적으로 적은 링크 MDL (MLD Dominant Link) 두 링크를 정의하고 HCL에서의 NSTR-MLD와 SLD간의 공정성 문제를 극복하면서 처리율을 개선하는 것을 목표로 하였다[1]. CSLT 기법에서는 동기화 전송을 허용하는 Synchronous Transmission Token (STT)을 조절함으로써 이러한 목표를 달성할 수 있는데, 성능이 STT를 조절하는 α값에 크게 영향을 받는 한계점을 지니고 있다.

최근 복잡한 무선통신 시스템의 성능을 향상시키기 위해 강화학습이 널리 이용되고 있으며[2]. 에너지 관리, 전송 파워, 변조 방식, 핸드 오버 제어 등 다양한 통신 제어 문제에 심층 강화 학습을 적용하는 연구가 이루어지고 있다. 본 논문에서는 기존 CLST 기법의 STT를 조절하는 핵심 파라미터인 α값을 강화학습을 이용하여 최적화하는 방법을 제안한다. CLST 기법의 메커니즘은 HCL 링크에서 백오프 카운터가 0이 되었으나 MDL 링크가 점유 중인 상태로 인해 즉각적인 전송이 불가능할 경우, STT를 저장하고 STT 값을 α만큼 증가시킨 후, MDL이 유휴해지는 경우 이를 소진하여 동기 전송을 수행한다. 하지만 최적의 성능을 얻을 수 있는 α값을 도출하기 어려운 문제점을 가진다. 본 연구에서는 최적의 α값을 얻기 위해 Multi-Armed Bandit 문제로 정의하고 ε-Greedy 정책을 통해 학습하는 방식으로 α를 탐색함으로써 처리율과 공정성 향상을 목표로 한다. CLST 기법에서 사용되는 파라미터 α와 ε-Greedy 학습을 통해 얻은 α를 구분하기 위해, 전자를 α_{clst}, 후자를 α_{greedy}라고 표기한다. 보상 함수는 비교적 경쟁이 많은 링크 HCL 에서의 평균 처리율과 디바이스 단위의 공정성 지수를 결합한 복합 지표를 활용하였다. NSTR-MLD와 SLD 비율을 다양하게 설정하고, 각 비율 및 처리율-공정성 가중치에 따라 α_{greedy} 값을 학습하였다. 그 후 모의실험을 통해 기존 CLST 기법과 비교하여 처리율과 장치들의 공정성 향상을 확인하였다.

Ⅱ. 제안 기법



[그림 1] CLST 기법 동작 방식 예시

Wi-Fi 7 환경에서 NSTR-MLD는 한 링크가 전송 중일 때 에너지 유출이 생기기 때문에 다른 링크가 차단되는 문제가 발생한다. 이를 해결하기 위해 CLST 기법에서는 STT를 제안하였다. [그림 1]은 CLST 기법의 예시를 나타낸다. HCL에는 MLD와 SLD가 공존한다. HCL에서 MLD의 백오프 카운터가 0이 되어 전송 기회를 얻었으나, MDL에서 점유 중인 상태로 인해 동기화 전송이 불가능한 경우 STT 값을 α_{clst}만큼 증가시킨다. 이후 MDL이 백오프 종료 후 유휴해지면 저장된 STT를 하나 소진하여 MDL과 HCL에서 동기화 전송을 수행하고, 이어 설정된 추가 보상 전송 횟수만큼 연속 전송하여 다중 링크 전송 효율을 높인다. STT의 증가율을 결정하는 파라미터인 α_{clst} 값이 커질수록, HCL과 MDL 양쪽 링크가 백오프 과정 없이 연속적으로 동기 전송하는 횟수가 증가하여 전송 효율을 향상시킬 수 있지만, 다중 링크 단말과 단일 링크 단말 간의 채널 점유 공정성을 저해시키는 문제가 발생한다. CLST 기법에서의 α_{clst} 값은 다중 링크 단말과 단일 링크 단말 수 또는 그 비율을 고려하여 설정해야 한다. 하지만 다양한 환경에서 최적의 성능을 얻을 수 있는 α_{clst}값을 시스템 모델을 통해 이론적으로 도출하는 것은 매우 어려운 문제이다. 따라서 STT 스택 증가 값 α_{clst} 값만을 제어 변수로 선정하여 Multi-Armed Bandit 기반의 학습 방법으로 α_{greedy} 학습을 진행하였다. 이를 위해 α_{greedy}의 후보값을 0.1에서 5.0 사이에 균일 간격으로 50개 정의하였고, ε-Greedy 정책을 이용해 각 에피소드에서 α_{greedy} 값을 확률 ε로 무작위로 선택하며 학습을 진행하였다.

학습을 위한 보상 함수(R)는 아래 식(1)과 같이 단말 단위 처리율을 이용하여 계산한 공정성 지수(F)와 HCL 링크의 처리율(TH)을 이론적 최대 처리율로 정규화한 값을 가중치(w)에 따라 결합한 형태로 구성하였다. 처리율만을 고려할 경우 공정성의 악화를 초래할 수 있으므로 공정성 지수를 반영하였다.

$$R = wF + (1-w)\frac{TH}{R_{\max}} \quad (1)$$

식(1)에서 가중치 w값을 변경하여 처리율과 공정성의 상대적 중요도를 조절할 수 있다.

가치 함수(Q-table)의 업데이트는 ε-Greedy 정책과 Sample-average 방식을 결합하여 이루어진다. 가치 함수는 식(2)과 같이 갱신된다.

$$Q_{t+1}(\alpha_{greedy}) = Q_t(\alpha_{greedy}) + \frac{1}{N(\alpha_{greedy})}[R - Q_t(\alpha_{greedy})] \quad (2)$$

식(2)에서 $Q(\alpha_{greedy})$는 행동 α_{greedy}에서의 가치 함수를, $N(\alpha_{greedy})$는 행동 α_{greedy}를 선택한 누적 횟수를 나타낸다. 매 에피소드마다 확률 ε로 탐험 (exploration)을, 확률 $(1-\varepsilon)$로 Q-table에서의 최적의 행동 (활용, exploitation)을 선택하며, 탐험 비율(ε)은 에피소드가 진행될수록 점진적으로 감소시키는 방식을 사용하여, 초반의 다양한 탐색으로부터 후반으로 갈수록 학습된 최적 α_{greedy} 값을 더 많이 활용하게 된다. ε의 감소 공식은 식(3)과 같다.

$$\varepsilon \leftarrow \max(\beta\varepsilon, \varepsilon_{\min}) \quad (3)$$

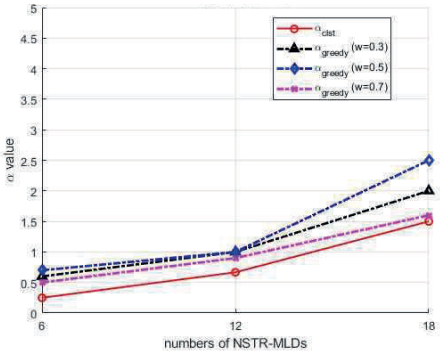
식(3)의 탐험 비율에 감쇠율(β)을 곱한 값과 탐험 비율 최솟값 중 최댓값으로 선택하여 총 70회의 에피소드를 반복 수행한 뒤, 학습된 Q-table에서 $Q(\alpha_{greedy})$가 가장 높은 가치 함수를 갖는 α_{greedy} 값을 최적 파라미터로 선정하였다.

Ⅲ. 모의실험

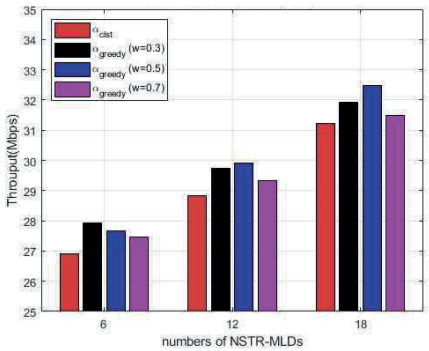
<표 1> 모의실험 환경

파라미터	값
모의실험 시간	1 sec
주파수 대역	2.4 GHz (HCL), 5 GHz (MDL)
다중 링크 수	2 개
단말 수 (HCL)	30 대
MCS	256 QAM (98 Mb/s)
MPDU	1000 bytes
대역폭	20 MHz
SIFS	18 μsec
에피소드 수	70 회
ε	1.0 → 0.01
β	0.98

<표 1>은 본 연구의 모의실험에 사용된 주요 파라미터를 나타낸다. 모의실험은 HCL에 연결된 전체 단말 (SLD와 NSTR-MLD) 수를 30으로 고정하고 SLD와 NSTR-MLD 수를 조정하면서 진행되었다. SLD 단말은 HCL에서만 동작한다고 가정하였다. 성능 평가 지표로는 HCL 링크의 평균 처리율과 단말 단위의 공정성 지수를 사용하였다.



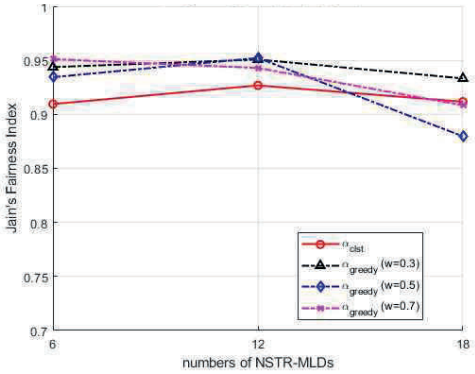
[그림 2] α_{clst}와 α_{greedy} 값 변화 그래프



[그림 3] NSTR-MLD 단말 수 변화에 따른 처리율

[그림 2]는 NSTR-MLD 단말 수 변화에 따른 α_{clst}와 α_{greedy}값의 변화를 나타낸다. α_{clst}값은 [1]에 주어진 것과 동일하게 NSTR-MLD 수를 SLD 수로 나눈 값으로 설정하였다. ε-Greedy 정책을 통해 학습된 α_{greedy}값은 대체로 α_{clst}값보다 큰 값을 가지며 가중치 w에 따라 달라지는데, NSTR-MLD 수가 6개인 상황에서 w가 커질수록 α_{clst}대비 높은 값으로 도출되었다.

[그림 3]은 NSTR-MLD의 단말 수 변화에 따른 처리율을 보여준다. NSTR-MLD의 단말 수가 6 이고 $\alpha_{greedy}(w=0.3)$일 때 α_{clst}대비 약 5% 높아진 처리율을 달성했다. 높은 α_{greedy}값은 HCL과 MDL 링크가 백오프 과정 없이 동시 전송을 횟수가 많아지는 것을 의미한다. 이는 학습된 α_{greedy}값이 적은 개수의 NSTR-MLD 환경에서 적극적인 연속 전송 전략을 선택하여 네트워크의 전반적 처리율을 증가시킨 결과이다.



[그림 4] Jain's Fairness Index 비교

[그림 4]는 NSTR-MLD의 단말 수 변화에 따른 공정성 지표를 보인다. HCL 처리율만을 고려할 경우 공정성의 악화를 초래할 수 있으므로 보상 함수에 공정성 지수를 반영하여 실험한 결과이다. 대체로 α_{clst}보다 높은 값을 보여주지만 $\alpha_{greedy}(w=0.5)$일 때 단말 수가 늘어남에 따라 공정성이 다소 감소하는 것을 확인할 수 있다.

Ⅳ. 결론

본 연구는 Wi-Fi 7 다중 링크 동작 환경에서 CLST 기법의 핵심 파라미터인 α_{clst}의 최적값을 얻기 위해 ε-Greedy 정책 기반의 Multi-Armed Bandit 학습 방법을 제안하였다. 모의실험을 수행한 결과, 학습된 α_{greedy} 값은 α_{clst} 방식 대비 HCL의 처리율과 공정성 지수를 모두 개선하였다. 본 연구의 시뮬레이션 결과는 에피소드 수와 시나리오 수의 제한으로 일부 한계가 있었으나, 추후 더욱 다양한 조건에서 추가적인 실험을 통해 더욱 개선된 결과를 얻을 수 있을 것으로 기대된다. 향후 연구에서는 ε-Greedy 외의 다양한 강화학습 알고리즘을 비교 및 평가하고, 실시간 네트워크 상태 변화에 따라 α_{greedy} 값을 동적으로 최적화하는 방법을 추가로 개발하여, 본 연구의 실질적인 적용성을 더욱 높이고자 한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-학석사연계ICT핵심인재양성 지원을 받아 수행된 연구임 (IITP-2024-00436744)

참 고 문 헌

[1] Kwon, Lam, and Eun-Chan Park. "Contention-Less Multi-Link Synchronous Transmission for Throughput Enhancement and Heterogeneous Fairness in Wi-Fi 7." Sensors 24.11 (2024): 3642.

[2] Luong, Nguyen Cong, et al. "Applications of deep reinforcement learning in communications and networking: A survey." IEEE communications surveys & tutorials 21.4 (2019): 3133-3174.