Wi-Fi 7의 Multi-Link Operation 기반 차량 네트워크에서 강화학습 기반의 적응형 MLO 모드 제어 기술

방수정, 이미정*
이화여자대학교
tnwjd7732@ewhain.net, *lmj@ewha.ac.kr

Adaptive MLO Mode Control using Deep Reinforcement Learning for Wi-Fi 7 Vehicular Networks

Soo Jeong Bang, Mee Jeong Lee*
Ewha Womans Univ.

요 약

최근 차량 애플리케이션의 발전으로 대용량 센서 스트림 및 AR/VR 기반 작업과 같은 저지연·고대역폭 처리 요구가 증가하며 이에 따라 작업을 외부 서버에게 전달하여 처리하는 오프로딩 연구가 활발하다. 그러나 단일 주파수 대역폭만으로는 증가하는 트래픽 요구를 모두 충족시키에 제약이 존재한다. 이러한 배경에서 IEEE 802.11be는 2.4 GHz와 5 GHz 대역을 병렬 활용하는 Multi-link operation (MLO)을 포함하며 STR 및 NSTR 모드를 제공하고, 두 모드는 배터리 소비과 전송 성능 측면에서 상호 보완적 특성을 가진다. 따라서 차량이 오프로딩 데이터의 특성, 차량 배터리 상태, 링크 품질을 종합하여 MLO 모드를 선택할 경우 차량의 서비스 만족도를 최대화할 수 있다. 본 논문에서는 차량 오프로딩 환경에서 전송 지연 및 에너지 소비를 동시에 최소화하는 차량 MLO 모드 결정 문제를 최적화 문제로 형성하였고 이를 Markov Decision Process(MDP)로 재정의하여 강화학습 모델 Proximal Policy Optimization(PPO)을 적용한 알고리즘을 제안한다.

Ⅰ. 서 론

최근 차량용 어플리케이션은 고해상도 센서 스트림 처리, AR/VR 기반 운전자 지원, 실시간 지도 업데이트같은 대용량·저지연 서비스를 요구한다.[1] 이러한 작업을 차량이 자체적으로 처리하기에는 연산 자원 부족 문제와 전력 제약이 크므로 가까운 도로변의 엣지 서버로 데이터를 오프로딩하여 연산 부담을 분산하는 방식이 보편화되고 있다.[2] 그러나 차량의 이동성으로 인하여 데이터를 외부 서버로 전송하는 과정 중 Radio Access Network (RAN)에서의 불안정성이 가중되어 안정적인 전송 지연을 보장하기 어렵다는 한계가 존재하며, 동시에 많은 사용자가 하나의 대역에 데이터를 송신하면 네트워크 혼잡 문제가 발생할 수 있다.

위의 문제를 해결하기 위한 기술로, 차세대 무선 표준 IEEE 802.11be (Wi-Fi 7)은 Multi-Link Operation (MLO)을 도입하여 2.4GHz 및 5GHz 대역을 동시에 활용한다.[3] MLO는 두 개 이상의 링크를 동시 송수신하는 Simultaneous Transmit-Receive(STR) 모드와 수신 시 반대편 링크의 송신을 중지하는 Non-Simultaneous Transmit-Receive (NSTR) 모드를 지원한다. STR은 대역폭 집성이 가능해 처리량이 크나, 자기 간섭 문제가 발생할 수 있고 에너지 소모가 크다. 반대로 NSTR은 전력 효율이 높고 간섭 회피에 유리하나 순간 처리량이 제한될 수 있다. 따라서 STR/NSTR 모드를 고정 사용하는 것보다 차량의 상태, 멀티 링크의 품질, 업로드 하고자 하는 트래픽의 특성을 반영하여 MLO 모드를 결정할 경우 오프로딩 성능과 만족도를 크게 향상시킬 수 있다.

본 논문은 차량 네트워크에서 Wi-Fi 7 MLO를 탑재한 차량이 작업 오프로딩을 위해 도로변 이중 대역 AP(2.4GHz, 5GHz)와 통신하며, 주어진 상태를 기반으로 차량이 통신 지연과 에너지 소모를 동시에 최소화하기 위한 STR/NSTR 모드 결정 문제를 최적화 문제 형태로 형성한다. 실시간 차량 네트워크에서 연산 복잡도가 큰 최적화 문제를 효율적으로 해결하기 위하여 Deep Reinforcement Learning (DRL) 기법을 활용하고, 이를 위하여 앞서 정의한 최적화 문제를 MDP 문제로 재정의한다. 이후 DRL 기반 MLO 모드 제어 알고리즘을 제안한다.

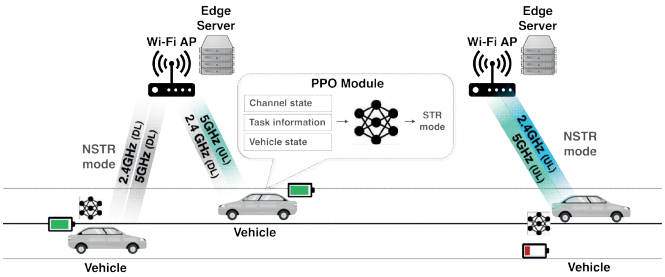

그림 1 제안하는 시스템 아키텍처

Ⅱ. 시스템 모델과 최적화 문제

본 연구에서 고려하는 네트워크는 도로변 Wi-Fi 7 엣지 AP와 이동 중인 차량 단말(OBU)로 구성된다. 각 AP는 2.4 GHz(20MHz)·5 GHz(80MHz) 두 대역을 제공하며, OBU는 Multi-Link Device(MLD) 기능으로 두 링크를 동시에 활성화할 수 있다. STR 모드는 한 링크를 송신(Tx), 다른 링크를 수신(Rx) 용으로 동시 사용하고, NSTR 모드는 모든 링크가 동일 시점에 같은 방향으로 동작한다. 차량이 오프로딩하는 작업 i는 데이터 크기 $D_i[bit]$와 최대 허용 지연 $T_i^{\max}[\mathrm{sec}]$으로 표현한다.

STR 모드는 링크 1개를 송신에 사용하므로 업로드 처리량은 단일 링크 용량 중 더 나은 품질의 링크로 제한된다. STR 모드의 전송 지연 T_{trans}^{STR} 은 수식 (1)과 같다.

$$T_{trans}^{STR} = D/\max_{k \in L}[B_k \log_2(1+SNR_k)] \quad (1)$$

여기서 B_k는 링크 k의 채널 대역폭(Hz), SNR_k는 차량-AP 링크 k의 신호 대 잡음 비이다.

NSTR 모드에서는 두 링크를 동시에 업로드로 사용하지만 현재 프레임이 수신이라면 송신 가능 상태로 전환되기까지 평균 대기시간 T_{wait}이 추가되며 이 수치는 통계적 값을 따른다. 따라서 NSTR 모드의 전송 지연 T_{trans}^{NSTR}는 수식 (2)과 같다.

$$T_{trans}^{NSTR} = T_{wait} + D/\sum_{k \in L} B_k \log_2(1+SNR_k) \quad (2)$$

차량 단말이 작업 데이터를 업로드할 때 소모되는 총 에너지 E_{tx}는 송수신 체인의 활성 구성이 STR/NSTR 모드에 따라 달라진다는 점을 반영하여 다음 수식 (3)로 모델링한다.

$$E_{tx} = \begin{cases} (P_{TX,k} + P_{RX,k'} + \triangle P_{SI})\, T_{trans}^{STR}, & (STR) \\ \sum_{k \in L} P_{TX,k}\, T_{trans}^{NSTR}, & (NSTR) \end{cases} \quad (3)$$

여기서 $P_{TX,k}$와 $P_{RX,k'}$ 는 각각 링크의 송수신 체인 전력이고 k는 업로드에 선택된 링크, k'은 수신에 사용되는 링크이며 $\triangle P_{SI}$는 STR 모드에서 자기 간섭 소거 회로가 추가로 소비하는 전력이다.

차량이 시각 t에 MLO 모드 $a_i(t) \in \{0=STR, 1=NSTR\}$를 결정하고 송신하는 과정의 비용 함수 U_i는 수식 (4)로 계산된다.

$$U_i = \alpha E_{tx}(a_t) + \beta T_{trans}(a_t) \quad (4)$$

따라서 MLO 모드 결정 최적화 문제는 (P1)으로 정의한다.

$$(P1)\ \min_{a_i(t)} U_i$$
$$C1: a_i(t) \in 0, 1$$
$$C2: T_{trans} \le T_i^{\max}$$

(P1)에서는 작업 i에 대한 비용 함수 U_i를 최소화하는 MLO 모드 $a_i(t)$를 결정하며 $C1$은 이진 결정 변수 제약이고 $C2$는 작업의 최대 허용 지연 제약이다. 하지만 (P1)은 이진 결정 변수 $a_i(t)$를 포함하고, 목적 함수U_i가 전송 지연과 에너지 소모를 포함한 비선형 함수로 구성되어 있어 비볼록 최적화 문제로 분류된다. 이러한 문제는 전통적인 수치 최적화 기법으로 해결할 경우, 연산 복잡도가 높고 반복 계산이 필요하므로 실시간 처리가 요구되는 차량 네트워크 환경에서 적용이 어려운 한계가 있다. 이에 따라 본 연구는 DRL 기법을 적용하여 채널 상태와 트래픽 부하가 빠르게 변화하는 환경에서도 적응적으로 MLO 모드를 결정할 수 있는 학습 기반 프레임워크를 제안한다. 이를 위해 (P1)을 MDP로 재정의하였으며 MDP의 구성 요소는 다음과 같이 설정된다.

먼저 상태 s_t는 각 링크의 SNR과 busy ratio, 차량 배터리 State of Charge (Soc), 그리고 최근 N개 슬롯 동안 측정된 차량 트래픽 통계량로 구성된다. 행동 a_t는 $\{0=STR, 1=NSTR\}$ 중 하나의 MLO 모드를 선택하는 이진 결정이고, 마지막으로 보상은 $r_t = -U_i$으로 비용 함수와 동일한 구조로 설정하여 전송 지연과 에너지 소비를 동시에 최소화하는 원 문제의 목적 함수를 그대로 반영한다.

Ⅲ. 적응형 MLO 모드 제어 알고리즘

본 논문은 DRL을 활용하여 차량이 MLO 모드를 결정하는 정책을 학습한다. 정책은 MDP의 상태 s_t를 입력받아 최적의 MLO 모드를 출력한다. 단, 본 알고리즘은 각 작업 단위로 동작하지 않고 시스템 상태가 유의미하게 변화했을 때에만 DRL 모델이 호출된다. 이러한 long-time scale 의사결정 구조는 정책 실행 빈도를 줄여 계산 오버헤드를 최소화하면서도 주어진 환경에 최적의 모드를 선택할 수 있다.

PPO 알고리즘은 높은 성능을 보이는 DRL 모델로, 주어진 데이터로 현재의 정책을 최대한 향상시키지만 한번에 과도하게 정책이 업데이트되어 발산하지 않도록 한다.[4] PPO의 정책 네트워크는 환경에 대한 행동 확률 분포를 생성하고, 가치 함수 네트워크는 동일한 상태의 기대 누적 보상을 평가한다. PPO를 구성하는 두 네트워크를 훈련하기 위하여 각 차량은 자신이 경험한 s_t, a_t, r_t, s_{t+1}을 메모리에 저장하고, 주기적으로 버퍼의 데이터를 이용하여 정책 네트워크와 가치 함수 네트워크의 가중치를 업데이트한다. 저장된 데이터로부터 정책 네트워크는 클리핑 기반의 손실 함수를 최소화하는 방향으로 업데이트되고 가치 함수는 실제 누적 보상과 예측 값 간의 오차를 줄이도록 학습된다.

충분한 데이터로 학습된 모델은 각 차량에 배포되며, 실행 단계에서는 자신의 상태만을 기반으로 STR 또는 NSTR 중 하나를 선택하여 상황에 가장 적합한 MLO 모드를 결정한다.

Ⅵ. 결론

본 연구에서는 Wi-Fi 7의 MLO 기반 차량 네트워크에서 전송 지연과 에너지 소비를 동시에 최소화하기 위한 MLO 모드 제어 최적화 문제를 형성하였다. 동적으로 변화하는 환경에서 최적의 행동을 빠르게 결정할 수 있는 DRL을 활용하기 위하여 최적화 문제를 MDP로 재정의하였다. PPO 기반 MLO 모드 제어 알고리즘은 에너지 소모와 통신 시간을 동시에 최적화할 수 있는 정책을 학습하며, 최근 발생한 차량 내부 작업 통계, 차량의 배터리 상태, 링크 품질을 종합한 상태 표현을 바탕으로 NSTR 및 STR 모드 선택 정책을 학습하도록 설계되었다. 향후 연구에서는 알고리즘을 구현하고 다양한 이동성 및 트래픽 시나리오를 대상으로 성능을 종합적으로 평가함으로써 알고리즘의 실용성과 확장 가능성을 검증할 계획이다.

ACKNOWLEDGMENT

본 연구는 2024년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. RS-2024-00408566)

참 고 문 헌

[1] Ma, Yifang, et al. "Artificial intelligence applications in the development of autonomous vehicles: A survey." IEEE/CAA Journal of Automatica Sinica 7.2 (2020): 315-329.

[2] Tang, Ming, and Vincent WS Wong. "Deep reinforcement learning for task offloading in mobile edge computing systems." IEEE Transactions on Mobile Computing 21.6 (2020): 1985-1997.

[3] López-Raventós, Álvaro, and Boris Bellalta. "Multi-link operation in IEEE 802.11 be WLANs." IEEE Wireless Communications 29.4 (2022): 94-100.

[4] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).