

밀리미터파 통합 액세스 백홀 네트워크에서 심층 강화학습 기반의 대역 할당 기법

박정훈*, 진희태*, 정수민**, 김석찬^o

Spectrum Allocation Based on Deep Reinforcement Learning in mmWave Integrated Access and Backhaul Network

Jeonghun Park*, Heetae Jin*, Sumin Jeong**, Suk Chan Kim^o

요약

5G 이동 통신 네트워크에서 데이터 전송 속도를 높이기 위해 밀리미터파가 사용됨에 따라 네트워크 밀집화가 이루어지고 있다. 하지만 네트워크 밀집화는 기기 설치 및 운영 등에서 기존의 네트워크 대비 추가적인 비용 부담을 야기한다. 이에 따라 비용 효율성이 높은 통합 액세스 백홀 네트워크에 대한 관심이 높아지고 있다. 통합 액세스 백홀 네트워크는 기존의 유선 백홀 링크를 무선으로 대체하여 비용적인 부담을 완화한다. 하지만 액세스 링크와 백홀 링크가 동일한 주파수 대역을 공유함에 따라 링크 간 간섭이 심화되는 문제가 생긴다. 이에 효율적인 네트워크 운영을 위한 기지국의 대역 할당 최적화가 요구된다. 본 논문에서는 액세스 대역이 사용자에게 균일하게 할당되는 통합 액세스 백홀 네트워크에서 네트워크 용량을 최대화하는 백홀 대역 할당 문제를 공식화 한다. 이를 해결하기 위한 심층강화학습 기반의 백홀 대역 할당 기법을 제안한다.

키워드 : 밀리미터파, 통합 액세스 백홀, 강화학습, 이중 심층 Q-망, 대역 할당

Key Words : mmWave, Integrated Access and Backhaul, Reinforcement Learning, Double Deep Q-Network, Spectrum Allocation

ABSTRACT

As mmWave has been utilized on fifth-generation mobile communication networks for a high data rate, network densification is in progress. Under the circumstances, interest in an integrated access and backhaul network is increasing. The integrated access and backhaul network replaces a traditional wired backhaul link with a wireless backhaul link reducing financial burden. However, inter-link interference gets harsher as access and backhaul link shares the same bandwidth. Consequently, spectrum allocation optimization is required for network efficiency. In this paper, we formulate a problem maximizing network capacity through backhaul spectrum allocation when the access spectrum is allocated uniformly to users. Then, propose deep reinforcement learning-based backhaul spectrum strategy which can solve the problem.

* 이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

• First Author : Pusan National University, jhpark2021@pusan.ac.kr, 학생회원

^o Corresponding Author : Pusan National University, sckim@pusan.ac.kr, 중신회원

* Pusan National University, htjin@pusan.ac.kr, 학생회원

** Paris-Saclay University, sumin.jeong@centralesupelec.fr

논문번호 : 202211-286-A-RN, Received November 29, 2022; Revised January 31, 2023; Accepted February 9, 2023

1. 서론

최근 이동 통신 디바이스 기반의 다양한 서비스에 대한 수요가 기하급수적으로 증가하고 있다. 이에 따라 급증한 사용자의 수요를 감하기 위해 높은 데이터 전송 속도가 요구된다. 이러한 상황에서 광대역을 제공하는 6GHz 이상의 밀리미터파 기술은 fifth-generation(5G) New-Radio(NR)의 핵심 기술로 자리 잡고 있다. 하지만 밀리미터파는 높은 전파 손실로 인한 전파 가능 거리 제한 및 강한 직진성으로 인한 신호 음영 지역을 발생과 같은 한계를 가진다. 이러한 문제를 해결하기 위해 기지국들이 밀집하게 배치되어야 한다. 기존의 네트워크의 백홀 링크에서는 유선 광섬유 통신이 사용되고, 이러한 백홀 링크 설치 비용은 기지국 설치 및 운영 비용의 85%를 차지한다^[1].

3rd Generation Partnership Project (3GPP) 에서 개발한 5G New Radio (NR)에서는 이러한 비용적인 한계를 극복하기 위해 통합 액세스 백홀(Integrated Access and Backhaul, IAB) 네트워크의 사용을 고려한다^{[2],[3]}. [그림 1] IAB 네트워크는 5G 네트워크 기지국인 gNB와 다수의 IAB-node들로 구성된다. 기존 네트워크처럼 gNB는 핵심망에 유선 백홀 링크로 연결된다. 반면에 IAB-node(자식 node)는 Core network(핵심망) 대신 gNB 또는 다른 IAB-node(부모 node)에 의해 무선 백홀링 된다(다중 홉 구조). IAB-node는 기존 기지국과 같이 네트워크 사용자들에게 액세스 링크를 제공할 수 있다. IAB 네트워크와 기존 무선 백홀 네트워크 사이의 차이점은 공유된 주파수 대역 내에서 액세스, 백홀 전송이 이루어진다는 점이다(In-band 백홀). 이에 하프 듀플렉스 제약이 존재한다^[4]. IAB 네트워크는 낮은 운용비용, 높은 주파수 대역폭 사용률 등의 장점을

가진다. 하지만 무선 링크 간 간섭이 심화된다는 단점도 가진다^[5].

무선 링크 간 간섭이 심한 IAB 네트워크의 효율적인 운용을 위해서는 동적 대역 할당 알고리즘이 필요하다. IAB 네트워크에서 대역 할당 문제는 액세스 대역 할당 문제와 백홀 대역 할당 문제로 구분 될 수 있다. 액세스 대역 할당 문제는 기지국이 셀 안에 분포하는 유저에게 할당할 액세스 대역의 크기를 정하는 문제이다. 액세스 대역 할당 문제의 경우 사용 가능한 전체 대역 혹은 전체 대역 중 일부분을 사용자에게 동등 배분하는 정적인 기법이 존재한다. 백홀 대역 할당 문제는 부모 node가 자식 node에게 백홀 링크를 제공하기 위해 사용할 대역의 크기를 정하는 문제이다. 백홀 대역 할당 문제를 해결하는 방법에는 각각의 자식 node들이 서비스하는 사용자 수에 비례하게 백홀 자원을 할당하는 정적인 기법이 존재한다. 하지만 위와 같은 정적인 백홀 대역 할당 기법은 밀리미터파를 사용함에 따라 발생하는 신호 차폐에 대한 민감도, 높은 전파 손실 및 시변적인 변수를 고려하지 못하며, 이는 네트워크 성능 저하로 이어질 수 있다. 따라서 네트워크 전반적인 링크 상태를 고려한 동적인 백홀 대역 할당 기법이 필요하다. 대역 할당 문제는 주로 정적인 상황에서의 최적화 문제를 푸는 것으로 해결되어 왔지만, 환경의 시변성 및 계산 복잡도를 고려하였을 때 IAB 네트워크에 적용하기는 힘들다. 따라서 저자는 강화 학습 기반의 백홀 대역 할당 기법을 제안한다.

본 논문에서는 각 기지국이 사용 가능한 전체 액세스 대역 내에서 사용자에게 대역을 동등 배분하는 액세스 대역 할당 방식을 사용할 때, 하프 듀플렉스 제약 아래 네트워크 용량을 최대화하는 심층 신경망 기반의 백홀 자원 할당 기법을 제안한다. 심층 신경망을 학습시키기 위하여 강화학습을 사용했으며, 이중 심층 Q 신경망(Double Deep Q Network, DDQN) 구조를 사용했다. 시뮬레이션을 통해 제안한 백홀 대역 할당 기법을 사용함으로써 정적 백홀 대역 할당 기법 대비 더 높은 네트워크 용량을 얻을 수 있음을 볼 수 있다.

본 논문의 이후 내용은 다음과 같이 구성된다. 먼저 II에서 IAB 네트워크 시스템 모델을 설명하고, 백홀 대역 할당 문제를 공식화한다. 그리고 강화 학습 기반 접근에 대해 설명한다. III에서는 시뮬레이션을 통해 제안한 대역 할당 기법과 정적 기법의 성능을 비교 및 분석한다. 마지막으로 IV에서 논문을 마무리 한다.

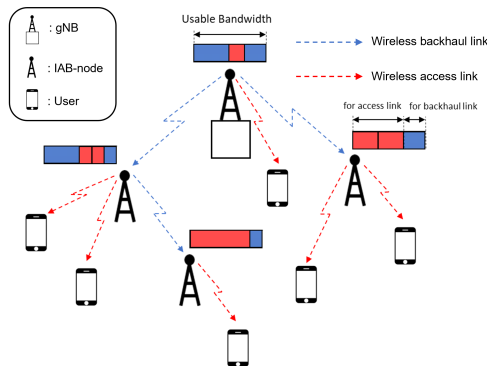


그림 1. 다중 홉 IAB 네트워크 구조 및 백홀 자원할당
Fig. 1. Multi-hop IAB network architecture and backhaul spectrum allocation

II. 본 론

2.1 IAB 네트워크 시스템 모델 및 문제 공식화

본 논문에서는 2계층(단일 홉) IAB 네트워크의 하향 링크 통신을 고려한다. 백홀 링크와 액세스 링크는 주파수 분할 듀플렉스(FDD)된다. 네트워크는 하나의 gNB, L 개의 IAB-node, K 명의 사용자로 구성된다. 본 논문에서는 gNB를 b_0 , l ($l > 0$) 번째 IAB-node를 b_l 으로 표기한다. $l \in \{1, \dots, L\} = N$. 그리고 k 번째 사용자를 u_k 로 표기한다. $k \in \{1, \dots, K\} = U$. 각 기지국이 사용할 수 있는 대역은 M 개의 직교하는 부-채널로 나누어져 있다고 가정한다. 모든 부-채널의 집합을 $D = \{1, 2, 3, \dots, M\}$ 라 한다. l 번째 IAB-node를 위한 백홀 링크에 사용되는 부-채널 집합을 $D_{b,l}$, k 번째 사용자를 위한 액세스 링크에 사용되는 부-채널 집합을 $D_{a,k}$ 라 표기한다.

2.1.1 위치 설정 및 사용자 연결

gNB는 네트워크 정중앙에 위치하며, IAB-node들과 사용자들은 커버리지 영역 내에서 무작위로 위치한다. 사용자는 평균 수신 신호 세기를 최대로 제공하는 기지국에 연결된다. 즉 b_l 이 u_k 를 서비스한다면 (1)을 만족한다. l_k 는 u_k 을 서비스하는 기지국의 인덱스다. P_i 는 기지국 b_i 의 송신 전력, $G_{i,k}$, $L_{i,k}$ 은 각각 b_i 과 u_k 사이의 유효 안테나 이득, 경로 손실이다. (1)에 따라 b_l 에 연결된 사용자들의 인덱스 집합을 U_l 이라 한다.

$$l_k = l = \arg \max_{i \in \{0, 1, 2, \dots, L\}} P_i G_{i,k} L_{i,k} \quad (1)$$

2.1.2 신호 대 간섭 및 잡음비 모델

기지국의 송신 전력은 각 부-채널에 균일하게 사용된다고 가정한다. 간섭은 한 링크에 사용하는 부-채널을 동시에 다른 링크도 사용할 경우 발생한다. $D_{b,l}$ 을 사용하는 백홀 링크와 $D_{a,k}$ 를 사용하는 액세스 링크가 겹치는 간섭 신호의 세기는 각각 (2), (3)와 같다. 링크가 m 번째 부-채널을 사용할 때, h_{m,l,l_k} 는 b_l 과 b_{l_k} 사이의, h_{m,k,l_k} 는 u_k 와 b_{l_k} 사이의 레일리 페이딩을 반영한다. $h \sim CN(0, 1)$.

$$I_{b,l} = \sum_{k \in U_l} \sum_{m \in D_{a,k} \cap D_{b,l}} \frac{P_{l_k}}{M} |h_{m,l,l_k}|^2 G_{l,l_k} L_{l,l_k} \quad (2)$$

$$I_{a,k} = \sum_{l \in N/l_k} \sum_{m \in D_{b,l} \cap D_{a,k}} \frac{P_0}{M} |h_{m,k,0}|^2 G_{k,0} L_{k,0} + \sum_{k' \in U/k} \sum_{m \in D_{a,k'} \cap D_{a,k}} \frac{P_{k'}}{M} |h_{m,k,l_k}|^2 G_{k,l_k} L_{k,l_k} \quad (3)$$

(2)와 같이 본 논문에서 고려된 2계층 IAB 네트워크에서 백홀 링크는 동일 계층(첫 번째 계층) 간섭에 자유롭다. 즉 두 번째 계층 IAB-node의 액세스 링크로부터 발생하는 간섭에만 영향을 받는다. 백홀, 액세스 링크의 신호 대 간섭 및 잡음비는 각각 (4), (5)와 같다. σ^2 는 잡음 신호 세기다.

$$\Gamma_{b,l} = \frac{|D_{b,l}| \frac{P_0}{M} |h_{l,0}|^2 G_{l,0} L_{l,0}}{\sigma^2 + I_{b,l}} \quad (4)$$

$$\Gamma_{a,k} = \frac{|D_{a,k}| \frac{P_k}{M} |h_{k,l_k}|^2 G_{k,l_k} L_{k,l_k}}{\sigma^2 + I_{a,k}} \quad (5)$$

2.1.3 IAB 네트워크 사용 합 용량

gNB와 핵심망을 잇는 유선 백홀 링크의 용량이 무한대라고 가정할 때, IAB 네트워크의 용량 C 은 (6)과 같다. BW_k, BW_l 은 액세스, 백홀 링크에 할당된 대역이다.

$$C = C_0 + \sum_{l=1}^L C_l \quad (6)$$

$$C_0 = \sum_{k \in U_0} BW_k \times \log_2(1 + \Gamma_{a,k}) \quad (7)$$

$$C_l = \min \left\{ BW_l \times \log_2(1 + \Gamma_{b,l}), \sum_{k \in U_l} BW_k \times \log_2(1 + \Gamma_{a,k}) \right\} (l > 0) \quad (8)$$

(7)은 gNB가 서비스하는 사용자들의 합 용량이다. l 번째 IAB-node는 백홀 링크를 통해 수신한 데이터를 U_l 에게 릴레이하기 때문에, (8)와 같이 백홀 링크의 용량과 U_l 의 합 용량 중에 작은 값으로 결정된다. 하프 듀플렉스 제약 아래에서 각 기지국이 사용가능한 액세스 대역폭 전체를 사용자에게 동등 배분할 때, 네트워크 용량을 최대화 하는 백홀 대역 할당을 찾는 문제는 (P1)으로 공식화 된다.

$$(P1) \max_{D_{b,1}, \dots, D_{b,L}} C(D_{b,1}, \dots, D_{b,L})$$

$$D_{b,l} \cup \bigcup_{k \in U_l} D_{a,k} = D$$

for $l \in \{0, 1, 2, \dots, L\}$; (9)

$$\bigcup_{k \in U_0} D_{a,k} \cup \bigcup_{l=1}^L D_{b,l} = D;$$
 (10)

$$D_{b,l} \cap D_{a,k} = \emptyset$$

for $l \in N, k \in U_l$; (11)

$$D_{b,l} \cap D_{a,k} = \emptyset$$

for $l \in N, k \in U_0$; (12)

(9), (10)은 각 기지국이 사용가능한 액세스 대역폭 전체를 활용한다는 조건을 반영한다. 그리고 (11), (12)은 하프 듀플렉스 제약조건 및 주파수 대역에서 백홀 링크와 액세스 링크가 구분됨을 나타낸다.

2.2 DDQN 기반 강화학습 프레임 워크

강화학습에서 에이전트는 동적인 환경과 상호작용하며, 시행착오를 통해 좋은 정책을 찾는다. 에이전트와 환경의 상호작용은 다음과 같이 이루어진다. 이산시간 t 에서 에이전트는 환경이 제공한 상태 s_t 를 바탕으로 정책 $\pi(a|s)$ 에 따라 행동 a_t 을 취한다. 이때 정책 $\pi(a|s)$ 는 제공된 상태가 s 일 때, 행동 a 를 취할 조건부 확률이다. 환경은 기존 상태 s_t 와 에이전트가 취한 행동 a_t 에 따라 다음 상태 s_{t+1} 와 보상 r_{t+1} 을 제공한다. 일련의 과정은 마르코프 결정과정을 따른다. 에이전트의 목표는 누적 보상 G_t 의 기댓값을 최대화하는 정책

$\pi(a|s)$ 을 찾는 것이다. 여기서 G_t 는 아래와 같이 정의된다.

$$G_t = \sum_{k=t+1}^T \gamma^{k-t-1} r_k$$
 (13)

$\gamma \in [0, 1]$ 는 미래 보상을 얼마나 중요시 여길지를 제어하는 감쇠 계수다. Q-함수로도 불리는 상태-행동 가치 함수 $Q_\pi(s_t, a_t) = E_\pi[G_t | s_t, a_t]$ 는 상태 s_t 에서 행동 a_t 을 취한 후 정책 π 을 따를 때, 누적 보상의 기댓값으로 정의된다. Q-학습에서 에이전트는 최적의 정책을 찾기 위하여 (14)와 같이 Q-값을 갱신한다. α 는 학습률이다.

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left(r + \gamma \times \max_{a'} Q(s', a') - Q(s, a) \right)$$
 (14)

에이전트의 행동 선택을 위한 정책으로는 일반적으로 ϵ -탐욕 정책이 활용된다. ϵ -탐욕 정책을 사용할 시 에이전트는 ϵ 의 확률로 무작위 행동을 선택하며, $1 - \epsilon$ 의 확률로 가장 높은 Q-값을 가지는 행동을 선택한다. 무작위 행동 선택 확률은 에이전트에게 환경 탐색의 기회를 제공하는 의미를 가진다.

심층 Q-학습은 심층 신경망을 통해 모방된 Q-함수를 사용한다. $\tilde{Q}(s, a; w) \approx Q(s, a)$. w 는 심층 신경망의 가중치이다. 심층 Q-신경망의 학습은 손실 $L(w)$ 을 최소화 하는 방향으로 w 을 조정함으로써 이루어지며, $L(w)$ 은 아래와 같다.

$$L(w) = (r + \gamma \times \max_{a'} \tilde{Q}(s', a'; w) - \tilde{Q}(s, a; w))^2$$
 (15)

Q-학습을 활용하는 에이전트는 실제 상태-행동 가치보다 더 높게 상태-행동 가치를 예측하는 최대화 편향 (Maximization bias)현상에 취약하다는 문제를 가지는데, 이중 Q-학습을 활용함으로써 이를 해결할 수 있다. 이중 Q-학습과 심층 Q-학습을 동시에 활용하는 DDQN은 학습 Q-신경망과 목표 Q-신경망으로 이루어져있다 [6]. 가중치 w 로 정의되는 학습 Q-신경망은 에이전트의 행동 선택에 사용되는 반면, 가중치 w^* 로 정의되는 목표 Q-신경망은 학습 Q-신경망을 학습시키기 위한 목표 값을 정하기 위해 사용된다. 이중 심층 Q-신경망을 사용할 시, 손실은 아래와 같다.

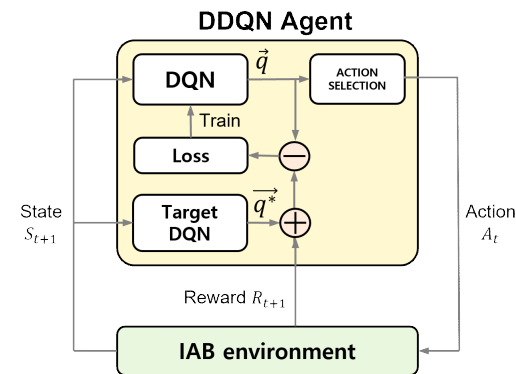


그림 2. 마르코프 결정과정 및 DDQN 에이전트
Fig. 2. Markov decision process and DDQN agent

$$L_{DDQN}(w) = (r + \gamma \times \underset{\mathbf{a}}{\tilde{Q}}(\mathbf{s}', \arg \max_{\mathbf{a}} \tilde{Q}(\mathbf{s}', \mathbf{a}; w); w^*) - \tilde{Q}(\mathbf{s}, \mathbf{a}; w))^2 \quad (16)$$

학습 과정은 [그림 2]와 같다. 학습 Q-신경망을 나타내는 DQN은 주어진 상태에서 취할 수 있는 행동들의 가치(\vec{q}_t)를 예상한다. 에이전트는 예상된 행동-가치들을 기반으로 ϵ -탐욕 정책에 따라 행동을 취하며, 환경으로부터 보상과 다음 상태를 제공받는다. 보상과 목표-DQN이 예상한 행동들의 가치(\vec{q}_{t+1}^*)는 (16)과 같이, DQN을 학습시키기 위한 손실을 계산하기 위해 사용된다. DQN의 기중치는 손실을 최소화 하는 방향으로 학습된다. 위 학습 과정은 단위 시간마다 이루어지며, 효율적인 학습을 위해 주기적으로 목표 Q-신경망의 기중치를 학습 Q-신경망의 기중치와 동일하게 갱신한다.

DDQN을 활용하여 (PI)를 해결하기 위해 아래와 같이 상태, 행동, 보상을 정의하였다. e_l 은 b_l 을 위한 백홀 링크의 스펙트럼 효율이고, E_l 은 l 번째 기지국에 의해 서비스되는 유저들의 평균 스펙트럼 효율이다.

$$\text{상태 } \mathbf{s} \quad [e_1, \dots, e_b, |U_0|, \dots, |U_L|, E_0, \dots, E_L]^T \quad (17)$$

$$\text{행동 } \mathbf{a} \quad [|D_{b,1}|, \dots, |D_{b,L}|]^T \quad (18)$$

$$\text{보상 } r \quad C(D_{b,1}, \dots, D_{b,L}) \quad (19)$$

$$e_l = \log_2(1 + \Gamma_{b,l}) \quad (20)$$

$$E_l = \frac{\sum_{k \in U_l} \log_2(1 + \Gamma_{a,k})}{|U_l|} \quad (21)$$

III. 모의실험

시뮬레이션에 사용된 변수를 [표 1]에 나타내었다. 경로 손실 모델은 5GCM-UMA와 UMi-Street Canyon을 사용하였다⁷⁾. 두 경로 손실 모델을 통해 신호 전파 거리에 따른 신호 차폐 확률을 반영했다. 새로운 에피소드가 시작될 때 마다 IAB-node와 사용자의 위치는 200m×200m 평면 내에서 무작위로 배치된다. gNB, IAB-node, 사용자의 높이는 각각 25m, 10m, 2m이다. 본 논문에서는 단일 홉 IAB 네트워크를 다루므로, 모든 IAB-node는 gNB로부터 백홀링 된다. 반면에 사용자는

표 1. 시뮬레이션 변수
Table 1. Simulation parameters

| Parameter | Value |
|-----------------|-------------|
| L | 3 nodes |
| K | 10 users |
| M | 10 channels |
| f_c | 30GHz |
| Total bandwidth | 200MHz |
| P_0 | 40dBm |
| $P_{1,\dots,L}$ | 33dBm |
| $G_{l,k}$ | 8dB |
| $G_{l,l'}$ | 16dB |
| Spectral noise | -174dBm/Hz |

(1)에 따라 최대 평균 수신 세기를 제공하는 기지국에 연결된다.

DDQN에서 DQN과 목표 DQN은 두 개의 전결합(Fully-connected) 은닉층을 가지며, 첫 번째 은닉층은 256개, 두 번째 은닉층은 512개의 뉴런을 가진다. 모든 활성화 함수로 ReLU 함수를 사용하였다. 감쇠계수는 $\gamma = 0.8$ 이고, 학습 시간에 따라 최대값 $\epsilon_{\max} = 1$ 부터 최소값 $\epsilon_{\min} = 0.1$ 까지 ϵ 값이 감소하는 ϵ -탐욕 정책을 활용하였다. 학습률은 $\alpha = 10^{-5}$ 이다. DDQN의 학습은 총 10000개의 에피소드를 통해 진행되었으며, 한 에피소드는 1000개의 시간 단계 동안 진행된다. 시뮬레이션 되는 IAB 네트워크 환경에서 달성 가능한 사용자 합 용량은 IAB-node, 사용자의 위치 등의 변수에 따라 크게 다르기 때문에 많은 수의 에피소드를 통해 학습을 진행하였다.

DDQN 에이전트는 에피소드 시작과 함께, (17)으로 설정된 IAB 네트워크 전반적인 링크 스펙트럼 효율에 관한 상태를 제공받고, (18)으로 설정된 백홀 스펙트럼 할당 행동을 취한다. 에이전트의 성능은 에피소드 내에서 얻은 보상 (19)의 평균으로 평가될 수 있다. DDQN 에이전트의 대조군으로 정적인 규칙 기반 에이전트를 사용하였다. 정적인 규칙 기반 에이전트는 gNB와 각 IAB-node가 서비스하는 사용자 수에 비례하게 대역을 할당한다.

[그림 3]은 학습초기 전형적인 에피소드 내에서 시간 단계 대 사용자 합 용량이다. 학습이 진행됨에 따라 사용자 합 용량이 증가하다가 최대에 수렴한다. 약 300 시간 단계 이후로 DDQN 에이전트의 성능이 규칙 기반 에이전트의 성능을 앞섬을 확인할 수 있다. 시간 단계에

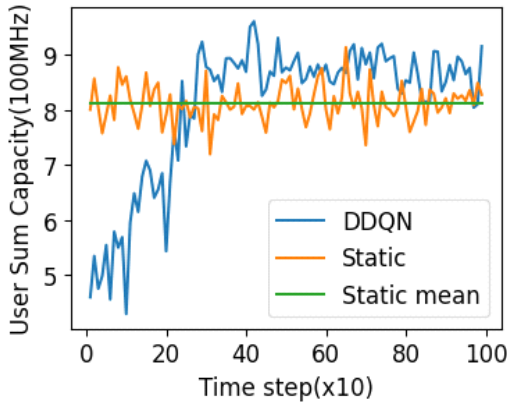


그림 3. 전형적 에피소드에서 시간 단계별 사용자 합 용량
Fig. 3. User sum capacity on a typical episode

다른 합 용량의 변동성은 ϵ -탐욕 정책 및 IAB 네트워크 환경의 시변성에 의한 것이다. 에피소드 별로 달성 가능한 사용자 합 용량이 다름을 고려하여, 다수의 에피소드를 진행하였을 때 에피소드 평균 사용자 합 용량 분포 비교를 통한 성능 평가도 진행하였다.

[그림 4]는 학습 단계를 마친 후 서로 다른 100개의 에피소드에서 DDQN 기반의 동적 백홀 대역 할당 기법과 사용자 수 기반 정적 백홀 대역 할당 기법을 사용했을 때, 각각의 에피소드 평균 사용자 합 용량의 분포를 나타내는 상자 그림이다. DDQN 기반의 동적인 대역 할당을 통해 정적으로 대역을 할당할 때 보다 평균 약 10% 더 높은 사용자 합 용량을 가질 수 있음을 확인할 수 있다. 상위 25%에 해당하는 사용자 합 용량과 하위 25%에 해당하는 사용자 합 용량의 경우, DDQN 에이전트를 사용하였을 때, 정적 백홀 대역 할당 기법 대비 각각 9.3%, 17.4% 상승하였다. 상위 25%의 사용자 합

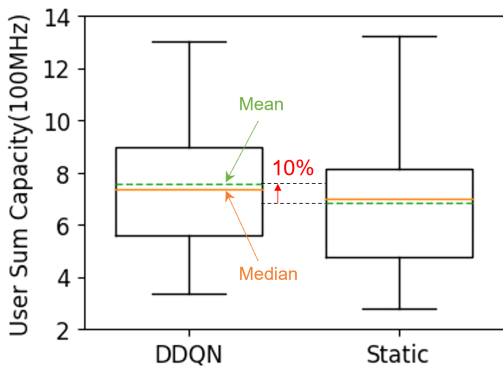


그림 4. 에피소드 평균 사용자 합 용량 분포
Fig. 4. Episode average user sum capacity distribution

용량 보다 하위 25%의 사용자 합 용량의 상승폭이 더 크다는 점에서, 네트워크 링크 스펙트럼 효율이 좋지 못한 상황에서 두 기법 간 성능 차이가 클 것으로 예상할 수 있다.

IV. 결론

본 논문에서는 링크 간 간섭이 심한 IAB 네트워크에서 사용자 합 용량을 최대화하는 DDQN 기반 동적 백홀 대역 할당 알고리즘을 제안했다. 그리고 시간에 따라 변화하는 환경 내에서 동적인 백홀 대역 할당을 통해 평균 사용자 합 용량을 10% 가량 증가시킬 수 있음을 보였다. 하지만, 제안한 에이전트는 기지국이 고정된 액세스 스펙트럼 할당 방식을 사용한다는 가정 아래 백홀 스펙트럼 할당을 최적화 하는 한계를 가진다. 즉, 최적의 액세스 스펙트럼 할당 방식이 적용될 경우, 더 높은 사용자 합 용량을 얻을 수 있을 것으로 기대된다. 결합된 백홀, 액세스 스펙트럼 할당 최적화는 추후 연구할 주제이다.

References

- [1] H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *IEEE Spectrum*, vol. 38, no. 8, pp. 40-45, Aug. 2001. (<https://doi.org/10.1109/6.938713>)
- [2] 3GPP, "NR; Study on Integrated Access and Backhaul; Release 15," TR 38.874, Jan. 2018.
- [3] M. Polese, et al., "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 62-68, Mar. 2020. (<https://doi.org/10.1109/MCOM.001.1900346>)
- [4] S. K. Baek, J. S. Song, H. S. Jung, and I. K. Kim, "Standardization trends for integrated access and backhaul network," in *Proc. KICS*, pp. 964-965, Pyeongchang, Korea, Aug. 2020.
- [5] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmWaves," *2018 IEEE GLOBECOM*, pp. 1-7, Abu Dhabi, UAE, Dec. 2018. (<https://doi.org/10.1109/GLOCOM.2018.8647977>)
- [6] H. Van Hasselt, A. Guez, and D. Silver,

“Deep reinforcement learning with double q-learning,” in *Proc. AAAI Conf. Artificial Intelligence*, pp. 2094-2100, Phoenix, Arizona USA, Feb. 2016.

(<https://doi.org/10.48550/arXiv.1509.06461>)

- [7] 3GPP, “*Study on Channel Model for Frequencies from 0.5 to 100GHz; Release 16*,” TR 38.901, Jan. 2020.

정 수 민 (Sumin Jeong)



2017년 2월 : 부산대학교 전자공학과 졸업

2018년 2월 : 부산대학교 전자공학과 석사

2022년 12월 : 유니버시티 칼리지 더블린 전자공학과 박사

2022년 11월~현재 : 파리-사클레 대학교 박사후연구원

<관심분야> 전자공학, 무선통신

[ORCID: 0000-0002-8951-9508]

박 정 훈 (Jeonghun Park)



2021년 8월 : 부산대학교 전자공학과 졸업

2021년 9월~현재 : 부산대학교 전자공학과 석사과정

<관심분야> 전자공학, 강화학습

[ORCID:0009-0002-8432-7811]

김 석 찬 (Suk Chan Kim)



1993년 2월 : 부산대학교 전자공학과 졸업

1995년 2월 : 한국과학기술원 전자공학과 석사

2000년 2월 : 한국과학기술원 전자공학과 박사

<관심분야> 전자공학, 강화학습, 인공지능, 무선통신, 초음파 신호처리, 수중 신호처리, MIMO, RIS, V2X, 객체 탐지, 음원 추적

[ORCID:0000-0002-5699-8565]

진 희 태 (Heetae Jin)



2016년 2월 : 부산대학교 전자공학과 졸업

2016년 2월~현재 : 부산대학교 전자공학과 석박사통합과정

<관심분야> 전자공학, 강화학습

[ORCID:0000-0003-0627-8022]