

## 강화학습을 이용한 주파수 공유 알고리즘 연구

김 지 수\*, 김 건 우\*, 박 종 인\*, 최 계 원°

## Study on the Spectrum Sharing Algorithm Using Reinforcement Learning

Ji Su Kim\*, Gon Woo Kim\*, Jong In Park\*, Kae Won Choi°

요 약

밀집된 네트워크 환경에서는 기존의 매체 접근 제어(MAC, Medium Access Control) 방식으로 인해 소모되는 채널 자원이 늘어난다. 본 논문에서는 소모되는 채널 자원을 최소화하기 위해 강화학습을 이용하여 통신 환경에 따라 채널 사용량을 최적화하는 매체 접근 방식을 학습하는 주파수 공유 알고리즘 개발에 대해 다룬다. 주파수 공유 알고리즘을 이용하면 802.11ax의 다운 링크(DL, Down Link) 데이터 전송 시 밀집된 환경에서 늘어나는 Backoff 알고리즘의 오버헤드로 인한 채널 사용량 감소와 CB(Channel Bonding) 시 하나의 채널만을 통해 매체에 접근할 수 있다는 한계점을 개선할 수 있다. 본 논문은 802.11ax의 다운 링크 환경 무선 통신 시스템의 구현방법과 시스템 내에서 주파수 공유 알고리즘으로 사용될 DQN(Deep Q-Network)의 구조와 매개변수들에 대해 다룬다. 무선 통신 환경 시스템에서 학습된 강화학습 모델의 통신 성능을 확인하고, 파라미터를 변경하며 학습한 결과를 비교하여 변경된 파라미터가 주파수 공유 알고리즘에서 가지는 의미를 확인하였다.

**Key Words** : 802.11ax, Down Link, MAC, Deep Q Network, Channel Bonding

## ABSTRACT

In dense network environments, channel resources consumed by traditional media access control (MAC) methods increase. In this paper to reduce consumed channel resources, frequency sharing algorithm using reinforcement learning is developed that optimize channel usage according to current wireless communication environment. The use of proposed frequency sharing algorithm can address the reduction in channel usage due to the increasing overhead of the Backoff algorithm in a dense environment while down-link (DL, Down Link) data transfer at 802.11ax and It can also improve the threshold that media can be accessed through only one channel in CB(Channel Bonding).

This paper deals with the implementation of 802.11ax down link environment wireless communication system and DQN(Deep Q-Network) structure and parameters, which will be used as frequency sharing algorithm within the system. Performance of trained reinforcement learning model in wireless communication system is verified and by changing the parameters, we compared the training results and identified the meaning of the changed parameters in frequency sharing algorithm.

※ 본 연구 논문은 과학기술정보통신부(MSIT) 및 정보통신기획평가원(IITP)의 출연금 등으로 수행하고 있는 한국전자통신연구원의 스펙트럼 쉐어링을 통한 기존 무선국 보호 및 주파수 공유기술 개발(2019-0-00964)의 위탁연구과제의 연구결과입니다.

• First Author : Sungkyunkwan University, Department of Electrical and Computer Engineering, kimjoker0@skku.edu, 정희원

° Corresponding Author : Sungkyunkwan University, Department of Electrical and Computer Engineering, kaewonchoi@skku.edu, 종신회원

\* Sungkyunkwan University, Department of Electrical and Computer Engineering, 학생회원

논문번호 : 202011-291-A-RU, Received November 22, 2020; Revised December 17, 2020; Accepted December 22, 2020

### 1. 서 론

최근 Wi-Fi 6E의 상용화에 따라 사람들은 더 방대하고 다양한 정보를 실시간으로 주고받을 수 있는 기술을 기대하고 있다. IoT 기기, 스마트폰과 노트북, 태블릿, 스마트 워치 등 통신 기기들의 종류는 더 다양해지고 증강현실, 가상현실 및 4K 스트리밍 서비스 등 전송되는 정보량 또한 크게 증가하고 있다. 빠르게 늘어나는 통신량으로 인해 생기는 가장 큰 문제점은 고정된 주파수 대역폭, 즉 한정된 채널 자원이다.

각 국가에서는 통신기술에 따라 사용할 수 있는 주파수 대역폭을 지정한다. 국내의 경우 Wi-Fi 6E는 6GHz 대역에서 1200MHz의 대역폭을 할당받아 기존 Wi-Fi 대역폭인 250MHz에 비해 약 5배 증가하였다. 이는 이전보다 통신속도를 증가시키고 데이터 전송 지연을 감소시키는 효과를 가져올 것이다. 하지만 위에서 말했듯 가용 대역폭의 증가량보다 통신량의 증가 폭이 더 높아 대역폭은 빠르게 포화상태가 될 것으로 예상하며, 한정된 채널 자원을 얼마나 효율적으로 사용할 수 있는지가 통신 성능 향상의 관건이 되었다.

채널 자원은 데이터를 전송할 때뿐만 아니라 통신 간의 충돌을 예방할 때도 일어난다. 실제로 매체 접근 제어 방식 중 CSMA/CA(Carrier Sense Multiple Access with Collision Avoidance)에서는 충돌을 예방하기 위해서 Backoff 알고리즘을 사용하는데, 같은 채널에서 통신 간의 충돌을 방지하기 위해 0부터 초기에 설정된 CW(Contention Window) 값 사이에서 랜덤한 값을 골라 Backoff 카운트로 설정하고 해당 카운트만큼 채널이 빈 상태를 확인하고 송신하는 방식이다<sup>11</sup>. 충돌이 일어날 때마다 CW가 배로 증가하므로 한 채널에서, 많은 통신이 이루어질수록 통신 간 충돌과 충돌로 인해 증가하는 CW로 인해 채널이 비어있어도 전송하지 못하는 시간이 많아져, 이는 채널 자원의 낭비로 이어지게 된다. [2]에서는 이를 해결하기 위해 현재 통신 환경에 맞춰 초기 CW를 조정하는 적응형 Backoff 메커니즘을 제시하였다. 이는 고정된 초기 CW를 가지는 기존의 Backoff 알고리즘과 다르게 통신 환경의 변화에도 항상 일정 수준 이상의 처리량을 보여주었다<sup>2,3</sup>. 하지만 랜덤한 크기의 backoff count를 이용하여 충돌을 방지한다는 기본적인 개념은 다르지 않기 때문에, 무의미한 채널 자원 소모가 발생할 수 있다는 여지는 사라지지 않았다.

그림 1에서 볼 수 있듯이 802.11ax에서도 다운 링크 데이터 통신 시 Backoff 알고리즘을 사용하고 있

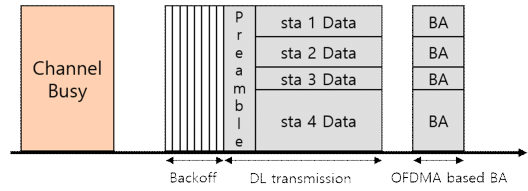


그림 1. 802.11ax의 다운 링크 데이터 전송 방식  
Fig. 1. Down Link data transfer in 802.11ax

다. 이는 곧 위에서 설명했듯이 같은 채널에서 통신하는 기기가 많아질수록 채널 사용량이 감소하는 한계를 가진다는 것을 의미한다.

802.11ax에서는 연속된 여러 채널을 동시에 이용하여 데이터를 전송하는 CB(Channel Bonding) 기술에 불연속적인 채널도 사용할 수 있게 만드는 PP(Preamble Puncturing) 기술이 적용되어 그림 2와 같이 중간에 사용되는 채널이 있어도 다른 나머지 채널은 사용할 수 있다<sup>4</sup>. 하지만 제1채널(Primary Channel)에서만 매체 접근이 이뤄지므로 제1채널이 바쁘다면 다른 채널이 비어있어도 데이터를 보내지 못한다는 한계가 있다.

802.11ax에서는 연속된 여러 채널을 동시에 이용하여 데이터를 전송하는 CB(Channel Bonding) 기술에 불연속적인 채널도 사용할 수 있게 만드는 PP(Preamble Puncturing) 기술이 적용되어 그림 2와 같이 중간에 사용되는 채널이 있어도 다른 나머지 채널은 사용할 수 있다<sup>4</sup>. 하지만 제1채널(Primary Channel)에서만 매체 접근이 이뤄지므로 제1채널이 바쁘다면 다른 채널이 비어있어도 데이터를 보내지 못한다는 한계가 있다.

기존 연구 중 강화학습과 순환 신경망(RNN, Recurrent Neural Network)을 이용하여 통신 환경을 파악하고 데이터 전송 가능 여부 확인과 적합한 채널 선택을 통해 높은 채널 사용률을 보인 경우가 있다<sup>5</sup>. 하지만, 동시에 최대 두 개의 채널만을 사용할 수 있도록 설정하여 실험을 진행하였기에 채널 사용량을 최대로 증가시켰다고 보긴 어렵다. 본 논문에서는 충

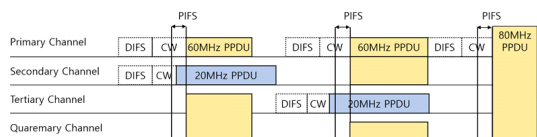


그림 2. 802.11ax의 preamble puncturing을 이용한 channel bonding  
Fig. 2. channel bonding using preamble puncturing in 802.11ax

돌 방지를 위해 소모되는 채널 자원을 최소화하고, CB 시 제1 채널만을 통한 매체 접근 제어의 한계를 해결하기 위해 AP의 다운 링크 통신 시 현재 채널 상황에 맞춰 동시에 다중 채널을 사용하여 데이터를 전송하는 주파수 공유 알고리즘을 제시한다.

시나리오 설정에 따라 강화학습 모델을 학습시키기 위해 BSS(Basic Service Set), 주파수 채널, AP와 통신 기기들의 위치 등을 조정할 수 있는 시뮬레이터를 구현하여 무선 통신 환경 시스템을 구성하였으며, 동일한 환경에서 강화학습 모델의 구조와 매개변수를 바꿔가며 학습을 진행하고, 학습된 모델이 채널에서 신호를 감지한 시간, 성공한 통신 시간, 실패한 통신 시간을 총 채널 사용량에 비교하여 확인하였다.

## II. 802.11ax 무선 통신 시스템

### 2.1 802.11ax(Wi-Fi 6E)

802.11ax(Wi-Fi 6E)는 차세대 IEEE의 표준 규격으로, 기존의 2.4GHz와 5GHz의 주파수 대역을 포함하고 있어 이전 규격과 함께 작동되도록 설계가 되어 있을 뿐만 아니라, 6GHz 주파수 대역에서 80MHz 14개, 160MHz 7개의 채널을 추가로 이용해 연결된 기기가 많은 밀집된 환경에서 안정적이고 빠른 통신이 가능하다. 또한, 802.11ax에서 새로 연구된 Preamble Puncturing이 기존 CB와 달리 제2채널(Secondary Channel)의 중간에 사용되는 채널이 있어도 해당 채널을 제외한 나머지 채널을 사용할 수 있게 하여 채널 사용량이 매우 증가하였던<sup>6)</sup>.

서론에서 설명한 다운 링크에서의 Backoff 알고리즘과 Preamble Puncturing의 한계를 해결한다면 더 높은 통신 성능을 얻을 수 있을 것이다. 이를 위해 본 논문에서 제안할 무선 통신 환경 시스템에서는 다운 링크 통신 상황을 가정하였다. 이에 더해 802.11ax의 CB와 Preamble Puncturing을 표현하기 위해 시스템에서 사용되는 하나의 주파수 채널을 20MHz의 채널로 정의하였고, 전체 채널은 CB로 연결되어있다 가정하였다.

### 2.2 Discrete Event Simulation(DES)

무선 통신 환경을 구현하기 위해서는 시스템 내의 모든 AP와 station이 독립적으로 동작하고 전체 시스템이 하나의 연속된 시간에서 동작하도록 해야 한다. 이를 위해 DES(Discrete Event simulation)를 사용하였으며 시뮬레이터의 모든 동작은 DES를 기반으로 실행된다.

시뮬레이터의 모든 동작은 이벤트를 통해 관리되고 이벤트 생성 시 해당 이벤트를 처리하기 위한 프로세스가 독립적으로 실행된다. 이벤트로 인해 생성된 모든 데이터는 로그를 생성하여 로그를 기반으로 시뮬레이션을 시각화하고 진행 상황을 확인할 수 있도록 구성하였다.

### 2.3 무선 통신 시스템 구조

무선 통신 시스템은 그림 3과 같은 형태로 구성되어 있다. DES를 기반으로 한 2차원 가상 환경에 그림 4와 같이 하나의 AP에 다수의 station이 모여 통신하는 구조인 BSS(Basic Service Set)를 추가하여 원하는 형태의 무선 통신 환경을 구성한다. BSS는 가상 환경에 추가 시 AP의 위치를 설정하고, AP로부터 일정 범위 내에 station들을 임의로 배치하게 된다. 이후 BSS들은 각자 설정된 가용 채널과 매체 접근 제어 방식을 통해 통신한다. 가용 채널이란 전체 채널 중 해당 BSS가 사용할 수 있도록 설정된 채널을 의미하며, 하나의 BSS는 모든 채널에서 같은 매체 접근 제어 방식을 사용한다.

각 BSS 내에서 station의 움직임은 mobility를 통해 설정할 수 있다. 본 시스템에서 사용된 mobility는 BSS의 범위 내에서 임의의 위치를 선택하여 일정한 속도로 이동 후 일정 시간 정지 후 다음 이동 위치로

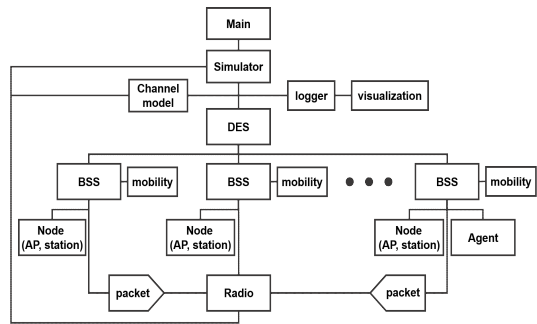


그림 3. 무선 통신 시스템 구조  
Fig. 3. wireless communication system

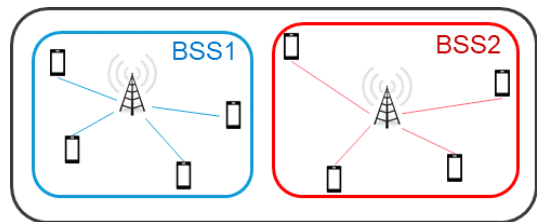


그림 4. BSS 구조  
Fig. 4. BSS structure

이동하는 형태이다.

BSS들에서 일어나는 모든 다운 링크 통신은 Radio로 전달되어 관리되며, 채널 별로 통신 간 간섭, 수신된 신호 세기 등이 계산된다. 그림 3의 오른쪽 끝에 있는 BSS의 경우 강화학습 모델인 에이전트를 연결하여 BSS 내의 AP가 얻은 채널 환경 정보로 현재 통신 환경에 맞는 매체 접근 제어 방식을 학습한다. 채널 환경 정보란 AP가 채널에서 얻을 수 있는 정보로 신호 감지를 했다면 수신된 신호 세기, 데이터 전송을 했다면 데이터 전송 성공 여부를 채널 환경 정보로 받는다.

그림 5는 시뮬레이터를 이용해 무선 통신 시스템으로부터 발생한 정보들을 토대로 시각화한 것이다. 원뿔 모양이 AP이며 구 모양은 station을 나타낸다. 노드 간에 연결된 흰 선은 데이터 전송이 이뤄지는 것을 의미한다.

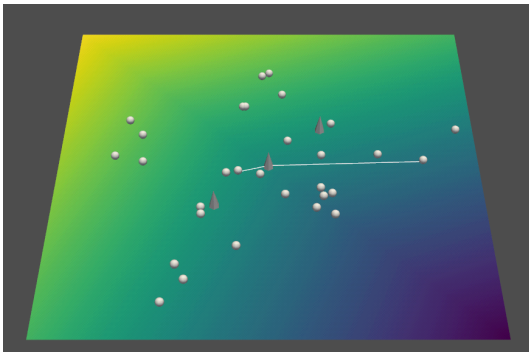


그림 5. 무선 통신 시스템 시각화  
Fig. 5. Wireless Communication System Visualization

### 2.4 주파수 공유 알고리즘

본 논문에서 제안하는 주파수 공유 알고리즘은 다중 채널 환경이라는 가정하에 AP가 얻을 수 있는 제한된 채널 정보를 일정한 시간만큼 쌓아 채널 맵을 만들어 채널 상황을 파악하고 데이터 전송이 가능한 채널들을 예측하여 다음 AP의 행동을 결정하는 방식이다. AP가 선택 가능한 행동으로는 크게 신호 감지와 데이터 송신이 있으며, 채널 정보는 AP가 실행한 행동에 따라 얻은 정보로 신호 감지의 경우 모든 채널에서 수신된 신호 세기를 확인하여 각 채널의 사용 여부를 얻을 수 있으며, 데이터 전송의 경우 데이터 전송 시 사용한 채널들의 데이터 전송 성공 여부를 얻을 수 있다. 위에서 행동을 통해 얻은 정보를 행동에 걸린 시간에 맞춰 기존 채널 맵에 업데이트하게 된다. 신호 감지를 하는 데 걸린 시간을 1로 설정하고 데이터 전

송에 걸린 시간은 신호 감지 시간과 데이터 전송 시간 간의 비율을 이용하여 정수 형태로 설정하였다. 채널 맵의 업데이트는 큐(Queue)의 방식으로 진행되어 앞에서 가장 오래된 채널 정보들을 최근 행동의 시간만큼 버리고 뒤에 새로운 채널 정보들을 해당 시간만큼 복사하여 추가하게 된다. 학습에 사용된 모델은 행동마다 얻은 결과를 통해 실시간으로 AP의 주파수 공유 알고리즘을 학습시키기 위해 DQN을 이용하였고, 연결된 데이터 셋인 채널 맵을 처리하기 위해 합성곱 신경망(CNN, Convolutional Neural Network)을 이용하였다.

## III. 강화학습 모델

### 3.1 알고리즘 및 신경망 구조

#### 3.1.1 학습 알고리즘

학습 알고리즘으로 사용된 DQN은 Q-Learning에 기반하여 심층 학습을 위한 학습 방법으로 Q-network 방식을 이용하며 깊어진 인공신경망 층과 Experience Replay 및 Target Network 알고리즘을 포함한다.

리플레이 메모리에 들어가는 상태 값은 2.4에서 설명한 채널 맵의 형태로 저장된다. 리플레이 메모리에 저장된 값들을 순서만 변경하여 한 번씩 학습시키기 위해 리플레이 메모리의 크기와 batch 크기를 같은 크기로 설정하였으며 리플레이 메모리 크기만큼 데이터를 쌓은 뒤, 리플레이 메모리에 쌓인 데이터를 임의의 순서대로 뽑아 Main Network를 업데이트한 이후 Target Network를 업데이트하는 과정을 에피소드라고 정의한다. 학습은 총 100회의 에피소드를 통해 진행되었다. 모델은 매 에피소드 이후 일정 횟수만큼 실행되어 신호 감지 횟수, 데이터 전송 성공 및 실패 횟수를 통해 전체 채널 사용량 중 AP의 신호 감지와 통신 성공, 실패가 차지하는 비중을 계산하여 학습 상태를 그래프로 나타내게 된다.

#### 3.1.2 인공신경망 구조

인공신경망에 입력값으로 들어가는 채널 맵은 학습 환경에서 사용된 총 채널의 개수를 높이로 설정하여 채널마다 얻은 채널 정보를 저장할 수 있도록 하였고, 신호 감지 시간 기준 쌓을 채널 정보의 길이를 채널 맵의 너비로 설정하였다. 또한, 학습 시 데이터 전송에 대한 채널 정보와 신호 감지에 대한 채널 정보의 혼동을 방지하기 위해 에이전트가 통신 환경에서 데이터 전송으로 얻은 채널 정보와 신호 감지로 얻은 채

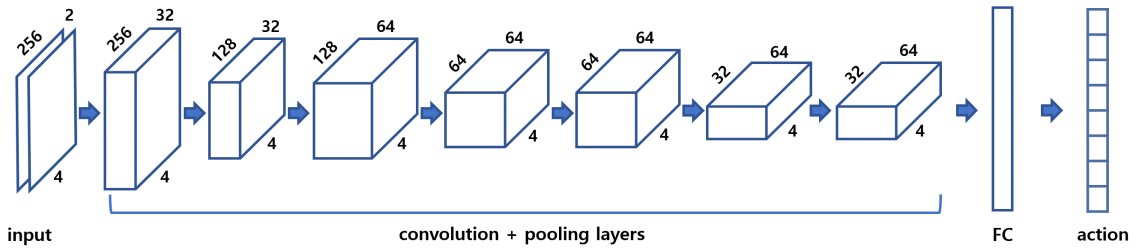


그림 6. 인공신경망 구조  
Fig. 6. Neural Network structure

널 정보를 따로 저장할 수 있도록 2개의 층으로 구성하였다. 본 논문의 학습 환경에서는 4개의 채널을 사용하며, 채널 맵에 들어갈 채널 정보의 길이를 256으로 설정하였기에 그림 6의 입력값과 같은 형태로 채널 맵이 구성된다. 학습에 사용될 합성곱 인공신경망은 그림 7과 같이 구성되어 있다. 모든 Convolution layer에서 시간 축에 대한 필터 크기는 4로, 채널 축의 필터 크기는 1로 설정되어 있기에 시간에 따른 채널 정보에 대해서만 Convolution이 이루어진다. 총 4개의 Convolution layer와 4개의 pooling layer로 이루어져 있으며, 이후 Fully-Connected Layer를 통해 정의된 행동들에 대한 Q-value 값을 계산하도록 하였다. 이 중 Q-value가 가장 높은 행동이 다음에 실행될 행동으로 AP에 전달되게 된다. 선택된 행동을 통해 무선 통신 환경으로부터 얻은 채널 환경 정보는 다시 시간 순서대로 쌓여 하나의 채널 맵으로 만들어진다.

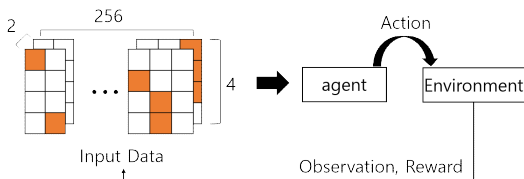


그림 7. 채널 맵을 이용한 DQN 구조  
Fig. 7. DQN using channel map

### 3.1.3 행동과 보상

본 시스템에서 에이전트가 선택할 수 있는 전체 행동의 수  $na$ (number of action)는 전체 채널의 개수를  $nc$ (number of channel), 한 번에 보낼 수 있는 최대 데이터 길이를  $mp$ (max packet)라고 했을 때 각 채널의 사용 여부와 보낼 데이터 길이의 경우의 수에 따라 다음과 같은 수식으로 정의된다.

$$na = (2^{nc} - 1) * mp \quad (1)$$

인공신경망을 통한 행동 선택은 모든 Q-value를 계산하고 이 중 Q-value가 최대인 색인을 찾는 것으로 이루어진다. 행동에 대한 색인은 표 1과 같이 일정한 규칙을 가지고 정의되어 있다. 색인 0에 대한 행동은 sensing(신호 감지)으로 모든 채널의 신호를 감지하며, 색인 1부터  $(2^{nc} - 1) * mp$ 까지는 transmit(데이터 전송)으로 데이터를 전송하는 채널의 수가 적은 행동부터 큰 행동 순으로, 전송할 데이터의 길이가 짧은 순서부터 긴 순서로 정의되어 있다.

행동에 따른 채널 환경 정보와 그에 따른 보상은 표 2와 표 3과 같이 sensing과 transmit의 결과에 따라 정의되어 있다. sensing의 경우 감지된 신호의 세기에 따라 나뉘며 두 경우 모두 데이터를 전송하지 않으므로 보상은 0이다, transmit의 경우 통신 성공 여부에

표 1. 행동에 대한 색인  
Table 1. action index

Index	Action
0	sensing
1	channel[0], 1 packet
2	channel[1], 1 packet
...	...
$(2^{nc} - 1) * mp$	channel[all], mp packet

표 2. Sensing에 대한 채널 환경 정보와 보상  
Table 2. Sensing Observation and Reward

Sensing	Sensed power = 0	Sensed power > 0
Reward	0	0

표 3. Transmit에 대한 채널 환경 정보와 보상  
Table 3. Transmit Observation & Reward

Transmit	Idle	Success	Failure
Reward	0	10*success channel	-40*failure channel



따라 나뉘며, 통신에 성공한 채널의 개수에 10을 곱한 값과 실패한 채널의 개수에 -40을 곱한 값을 합쳐 해당 채널 환경 정보의 보상으로 정한다.

### 3.2 학습 환경

학습을 위해 구성한 통신 환경은 두 개의 BSS와 강화학습 모델을 적용한 하나의 BSS로 구성되어 있으며, 전체 주파수 채널의 개수는 0번부터 3번까지 총 4개로 설정되어 있다. 두 BSS는 설정된 slot time 횟수만큼 대기한 후 일정한 패킷 전송 시간만큼 데이터를 전송하도록 동작한다. BSS 1의 경우 0번과 2번 채널을 사용하여 100회의 slot time만큼 대기하고 1회의 패킷 전송 시간만큼 데이터를 전송하며, BSS 2의 경우 1번과 3번 채널을 사용하여 200회의 slot time만큼 대기하고 2회의 패킷 전송 시간만큼 데이터를 전송한다.

각 AP의 MAC 프로토콜은 BSS에 따라 독립적으로 정의되는데, Agent의 경우에는 BSS와 TCP로 연결되어 해당 BSS의 AP로부터 환경 정보를 받아 학습하고 행동을 예측하여 AP로 전달한다. 이는 이후 다중 에이전트를 학습할 때 에이전트와 무선 통신 시스템 간에 정보 손실 없이 데이터를 주고받기 위함이다.

표 4. BSS 설정  
Table 4. BSS setup

BSS	slot time(9 us)	packet duration(217.6 us)
BSS 1	100	1
BSS 2	200	2

## IV. 실험 결과

### 4.1 학습 매개변수 설정

이 절에서는 학습에 사용된 매개변수에 관해 설명한다. 표 5는 사용된 매개변수들의 기본값이다. 학습을 진행하는 동안 특정 상태에서 새로운 행동의 Q value를 탐색하기 위해  $\epsilon$ -greedy 알고리즘을 이용하여 epsilon의 확률만큼 임의의 행동을 취하게 된다. 이때 모든 행동이 같은 확률을 가진다면 sensing을 할 확률은 전체 채널의 수와 최대 데이터 길이가 늘어남에 따라 기하급수적으로 낮아지므로 sensing probability라는 변수를 통해 임의의 행동 시 sensing을 선택할 확률을 50%로 설정하였다. 시간에 따라 주기적으로 변하는 환경을 가정했을 시 미래 상태에 대한 보상은 고려할 필요가 적어 미래에 얻을 보상에 대한 감가율(discount factor)을 0으로 하는 myopic evaluation을 기본값으로 하였다.

표 5. 학습 매개변수  
Table 5. Training parameter

Parameter	Value	Parameter	Value
num of channel	4	episode	100
max packet unit	3	replay memory	1000
history length	256	batch size	1000
sensing prob	0.5	initial epsilon	1
discount factor	0	epsilon decay	0.95
learning rate	0.001	min epsilon	0.1
epoch	1	test length	1000

### 4.2 학습 및 시뮬레이션 결과

학습된 에이전트의 성능 확인을 위해 같은 환경에서 기존 CSMA/CA 알고리즘과 결과를 비교하였다. 비교에 사용된 CSMA/CA는 한번에 2회의 데이터 패킷 전송 시간만큼 데이터를 전송하며, Backoff 알고리즘의 초기 CW 크기는 16으로 충돌 시 CW의 크기를 2배씩 늘리는 Exponential Backoff를 사용하여 최대 65536까지 증가하도록 설정하였다. 802.11ax와 같이 다중 채널을 위한 CB와 PP가 가능하다고 가정하여 제1채널을 가장 낮은 주파수 채널로 설정하였고 backoff count가 0일 때 제2채널들의 상태를 확인하여 가능한 모든 채널에 데이터를 전송하도록 하였다.

AP가 학습된 에이전트나 CSMA/CA를 이용하여 통신하는 모습을 무선 통신 환경에서 확인하기 위해 그림 8과 같이 채널 상태를 시각화하여 나타내었다. 붉은색이 확인할 BSS의 통신을 나타내며 파란색이 BSS 1의 통신, 노란색이 BSS 2의 통신을 의미한다.

그림 9, 10, 11은 동일한 환경에서 학습된 에이전트와 CSMA/CA의 성능 분석을 위한 전체 채널 사용량 대비 데이터 전송 성공 비율, 데이터 전송 실패 비율, 신호 감지 비율 그래프이다. 결과 그래프의 x축은 진행된 에피소드의 수를 나타내고 y축은 전체 채널 사용량에서 차지하는 비율을 나타낸다. 전반적으로 CSMA/CA의 데이터 전송 성공 비율이 학습된 에이전트보다 높지만, 실패 비율과 신호 감지의 비율 또한 매우 높다. 특히 에피소드 40과 50 사이에서 통신 성공 비율이 급격히 낮아지고 신호 감지 비율이 매우 높

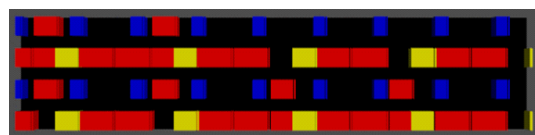


그림 8. 채널 상태 시각화  
Fig. 8. Channel Status Visualization

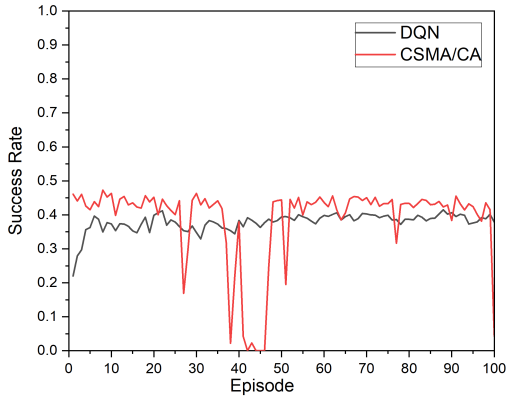


그림 9. 전체 채널 사용량 대비 전송 성공 비율  
Fig. 9. Transmission success ratio to total channel usage

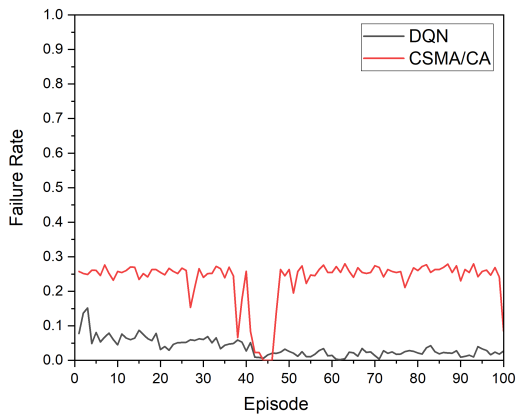


그림 10. 전체 채널 사용량 대비 전송 실패 비율  
Fig. 10. Transmission failure ratio to total channel usage

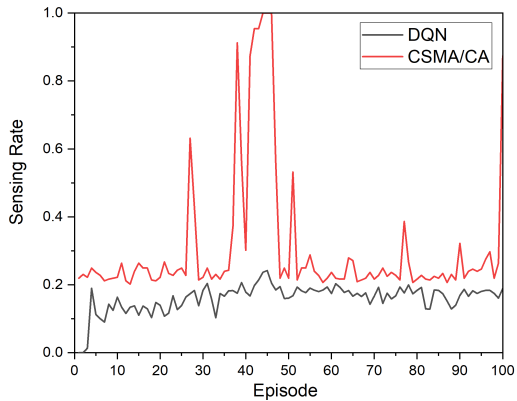


그림 11. 전체 채널 사용량 대비 신호 감지 비율  
Fig. 11. Sensing ratio to total channel usage

아지는 것을 볼 수 있다. 이는 해당 시간에 잇따른 통

신 실패로 backoff count가 크게 증가하여 데이터를 전송할 수 없었기 때문이다. 제1채널만을 이용하여 채널에 접근하고,

데이터 전송 직전에만 전체 채널의 사용 여부를 판단하므로 상대적으로 높은 실패 비율을 보이는 CSMA/CA에 비해 학습된 에이전트는 통신 환경의 채널 상황에 맞춰 데이터 전송 여부를 결정하므로 통신 성공 비율에 큰 폭 차가 없으며 데이터 전송 실패 비율이 매우 낮은 것을 확인할 수 있다. 그림 12와 13은 학습된 에이전트와 CSMA/CA의 채널 상황을 시각화한 것이다. 학습된 에이전트는 채널 상황에 맞춰 통신하는 반면 CSMA/CA의 경우 임의의 backoff count 동안 신호 감지 후 데이터를 보내고 있다. 학습 환경의 BSS는 고정된 시간 간격으로 데이터를 전송하므로 CSMA/CA는 잦은 데이터 전송 실패로 채널 사용량이 감소하는 것으로 확인된다.

매개변수에 따른 학습에 대한 영향을 확인하기 위해 표 6에 나와 있는 시나리오대로 총 6개의 매개변수를 변화하여 학습을 진행하였다.다음은 매개변수 변화에 따른 성능 비교로 그림 14, 15, 16으로 각각 데이터 전송 성공 비율, 데이터 전송 실패 비율, 신호 감지 비율로 나타내었다.

Discount factor를 증가시킨 경우 학습이 진행되는

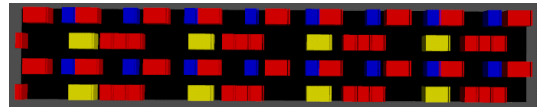


그림 12. 학습된 DQN 에이전트 시각화(default)  
Fig. 12. Trained DQN agent visualization(default)

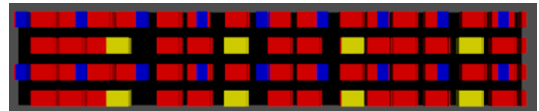


그림 13. CSMA/CA 시각화  
Fig. 13. CSMA/CA visualization

표 6. 시나리오별 매개변수 변화  
Table 6. Parameter change by scenario

Scenario	Parameter	Value
1	default	-
2	discount factor	0 → 0.99
3	kernel size	(4,1) → (4,4)
4	failure reward	-40 → -10
5	sensing probability	0.5 → 0.9
6	learning rate	0.001 → 0.0003

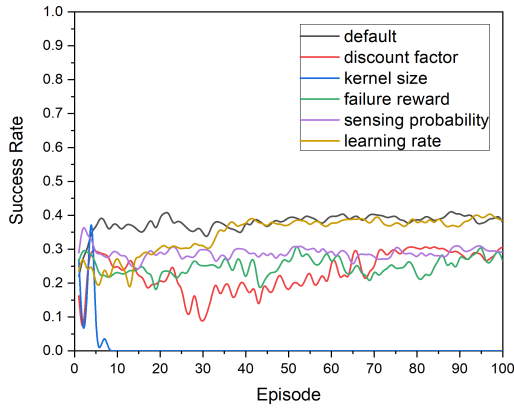


그림 14. 매개변수별 전송 성공 비율  
Fig. 14. Transmission success rate by parameter

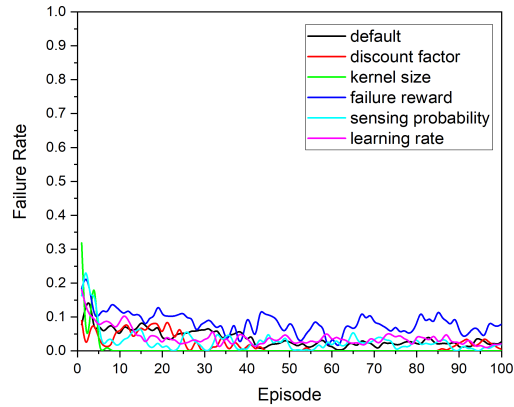


그림 15. 매개변수별 전송 실패 비율  
Fig. 15. Transmission failure rate by parameter

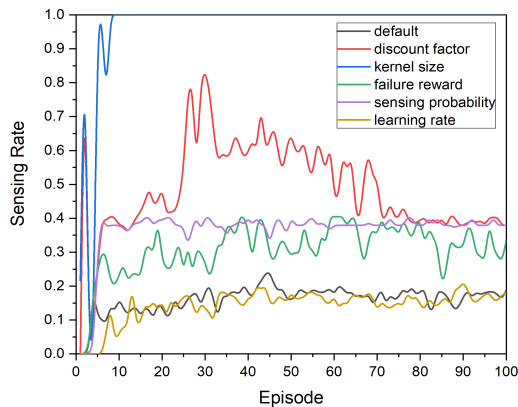


그림 16. 매개변수별 신호 감지 비율  
Fig. 16. Sensing rate by parameter

동안 신호 감지 비율이 다른 결과에 비해 매우 높게 나왔는데, 미래에 대한 보상이 포함되므로 전송 실패

로 얻는 음수의 보상값 누적되는 경우를 방지하기 위한 것으로 분석하였다. Kernel size의 경우 채널 축의 kernel size를 1에서 4로 늘려 채널 간 상관관계를 학습하도록 하였다. 하지만, 에이전트의 신호 감지 비율이 1로 수렴하며 학습에 실패하였고, 이를 통해 채널 간 상관관계가 적거나 없다는 것을 확인하였다. failure reward를 감소시킨 경우 다른 에이전트와 전송 성공 비율은 비슷하지만, 전송 실패 비율이 증가하였다. 이는 전송 실패 시 없는 보상 값이 줄어들어 기존보다 신호 감지 횟수가 줄어들고 데이터 전송 횟수가 늘어나서 발생한 현상으로 확인되었다.

나머지 매개변수의 경우 비슷한 형태로 최종값에 수렴하였다. 하지만 학습된 DQN 모델들의 수렴값이 크게 2개로 나뉘는 것을 확인할 수 있는데, 4개의 주파수 채널을 모두 사용하여 통신하거나 2개의 주파수 채널을 사용하여 통신하도록 학습된 결과의 차이이다. 이는 데이터 전송에 대한 행동의 개수가 많아 특정 상태에 따라 최적의 행동을 찾지 못하고 학습한 행동 중 가장 높은 보상을 얻을 수 있는 행동으로 수렴하여 발생한 차이로 분석하였다.

## V. 결론

본 논문에서는 강화학습을 이용하여 802.11ax의 다중 링크 통신 환경에 맞춰 매체 접근 제어 방식을 학습하는 주파수 공유 알고리즘을 제안하였다. 합성곱 인공신경망을 응용한 DQN을 통해 이전 채널 상황의 특징을 학습하여 행동을 결정하므로 채널 자원의 소모를 줄이며 통신할 수 있다. DES를 기반으로 한 무선 통신 시스템을 구현하여 동일한 환경에서 CSMA/CA와의 성능 비교를 통해 제안된 주파수 공유 알고리즘을 통한 채널 사용량 증가 가능성을 확인하였고, 에이전트의 매개변수를 변경해가며 학습을 진행하여 각 매개변수가 학습에 미치는 영향을 분석하였으며, 채널 개수에 따라 기하급수적으로 증가하는 행동의 개수가 학습 성능을 낮추고 있다는 것을 확인하였다.

실험을 통하여 통신하는 채널 간에 상관관계가 적은 것을 확인하였기에 추후 연구에서는 채널별로 데이터 전송 여부를 선택하여 총 행동의 수를 줄여 일관된 학습 결과가 나오는 것을 확인하고, 다음 연구에서는 새로운 BSS의 추가나 주변 BSS들의 통신 방식의 변경에 대한 학습을 진행할 것이다. 또한, 이전 채널 환경 정보들을 이용하여 다음 행동이 아닌 미래의 채널 상태를 예측하도록 학습하여 사용 가능한 채널을



찾아 통신하는 방식을 테스트해 볼 것이다.

### References

- [1] D. Coleman, "Dynamic frequency selection," *Wireless LAN Professionals*, Oct. 2019.
- [2] M. Park, "IEEE 802.11ac: Dynamic bandwidth channel access," *2011 IEEE Int. Conf. Commun.*, Kyoto, Japan, 2011.
- [3] M. S. Gast, *802.11ac A Survival Guide*, O'REILLY, 2013.
- [4] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Commun. Surv. & Tuts.*, vol. 21, no. 1, pp. 197-216, First Quarter 2019.
- [5] Y. Xu, J. Y. Yu, and R. M. Buehrer, "The application of deep reinforcement learning to distributed spectrum access in dynamic heterogeneous environments with partial observations," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 7, pp. 4494-4506, Jul. 2020.
- [6] M. Yang, B. Li, and Z. Yan, "MAC technology of IEEE 802.11ax: Progress and tutorial," *Mob. Netw. and Appl.*, Aug. 2020.
- [7] V. Minh, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller, "Playing atari with deep reinforcement learning," *ArXiv abs/1312.5602*, 2013.

김 지 수 (Ji Su Kim)



2019년 2월 : 성균관대 컴퓨터공학과 학사 졸업  
 2019년 3월~현재 : 성균관대학교 전기전자컴퓨터공학과 석사과정  
 <관심분야> 무선통신, 머신러닝  
 [ORCID:0000-0002-0182-6125]

김 건 우 (Gon Woo Kim)



2014년 3월~현재 : 성균관대학교 전기전자공학과 학사과정  
 <관심분야> 무선통신, 머신러닝  
 [ORCID:0000-0003-1536-431X]

박 종 인 (Jong In Park)



2020년 8월 : 한국외국어대학교 전자공학과 학사 졸업  
 2020년 9월~현재 : 성균관대학교 전기전자컴퓨터공학과 석사과정  
 <관심분야> 무선통신, 머신러닝  
 [ORCID:0000-0002-4035-6728]

최 계 원 (Kae Won Choi)



2007년 8월 : 서울대학교 전기컴퓨터공학부 박사  
 2010년 9월~2016년 8월 : 서울과학기술대학교 컴퓨터공학과 조교수  
 2016년 9월~현재 : 성균관대학교 전기전자컴퓨터공학과 부교수

<관심분야> 무선통신, 무선전력전송  
 [ORCID:0000-0002-3680-1403]