



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0164519  
(43) 공개일자 2023년12월04일

- |  |  |
|--|--|
| (51) 국제특허분류(Int. Cl.)<br><i>H04W 28/08</i> (2023.01) <i>G06N 20/00</i> (2019.01)<br>(52) CPC특허분류<br><i>H04W 28/0942</i> (2020.05)<br><i>G06N 20/00</i> (2021.08)<br>(21) 출원번호 10-2022-0064364<br>(22) 출원일자 2022년05월25일<br>심사청구일자 2022년05월25일 | (71) 출원인<br>포항공과대학교 산학협력단<br>경상북도 포항시 남구 청암로 77 (지곡동)<br>(72) 발명자<br>양현중<br>경상북도 포항시 남구 청암로 77<br>장종규<br>경상북도 포항시 남구 청암로 77<br>김영준<br>울산광역시 울주군 언양읍 유니스트길 50<br>(74) 대리인<br>특허법인(유한)아이시스 |
|--|--|

전체 청구항 수 : 총 14 항

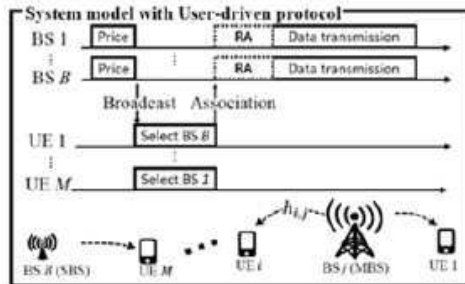
(54) 발명의 명칭 강화학습을 이용한 채널 정보가 제한된 네트워크에서의 기지국 제어 방법 및 장치

(57) 요약

본 개시는 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법을 제공한다. 상기 방법은, 복수의 기지국들 중 제 1 기지국의 제 1 제어장치가 상기 제 1 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 수신하는 단계; 상기 제 1 제어장치가 상기 제 1 기지국을 제외한 상기 복수의 기지국들의 제어장치들로부터 정보를 수신하는 단계; 상기 제 1 제어장치가 상기 상태정보를 강화학습 모델에 입력하여 상기 제 1 기지국의 과금 정보를 예측하는 단계; 및 상기 제 1 제어장치가 상기 과금 정보를 예측한 결과를 이용하여 상기 제 1 기지국 및 상기 복수의 단말들 간의 연결 상태를 제어하는 단계;를 포함한다.

대표도 - 도1

100



(52) CPC특허분류

*H04W 28/0835* (2020.05)

*H04W 28/095* (2020.05)

---

**명세서**

**청구범위**

**청구항 1**

복수의 기지국들 중 제 1 기지국의 제 1 제어장치가 상기 제 1 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 수신하는 단계;

상기 제 1 제어장치가 상기 제 1 기지국을 제외한 상기 복수의 기지국들의 제어장치들로부터 정보를 수신하는 단계;

상기 제 1 제어장치가 상기 상태정보를 강화학습 모델에 입력하여 상기 제 1 기지국의 과금 정보를 예측하는 단계; 및

상기 제 1 제어장치가 상기 과금 정보를 예측한 결과를 이용하여 상기 제 1 기지국 및 상기 복수의 단말들 간의 연결 상태를 제어하는 단계;를 포함하는 방법.

**청구항 2**

제 1 항에 있어서,

상기 상태정보는 상기 제 1 기지국 및 상기 복수의 단말들 간의 채널 이득 및 상기 제 1 기지국의 이전 상태의 행동 정보를 포함함을 특징으로 하는 방법.

**청구항 3**

제 2 항에 있어서,

상기 행동 정보는 상기 제 1 기지국 및 상기 복수의 단말들 간 연결 프로토콜에 필요한 과금 정보를 포함함을 특징으로 하는 방법.

**청구항 4**

제 1 항에 있어서,

상기 정보는 상기 제 1 기지국을 제외한 상기 복수의 기지국들에 대한 상태정보 중 그래디언트(Gradient) 계산에 필요한 정보를 포함함을 특징으로 하는 방법.

**청구항 5**

제 2 항에 있어서,

상기 상태정보는 이하의 수학식에 따라 표현됨을 특징으로 하는 방법,

$$s_j^{(t)} = [\tilde{x}_{1j}^{(t)}, \dots, \tilde{x}_{Mj}^{(t)}, H_{1j}^{(t)}, \dots, H_{Mj}^{(t)}, a_j^{(t-1)}]$$

여기에서  $s_j^{(t)}$  는 시간 슬롯 t에서 j번째 기지국의 상태 정보이고,  $\tilde{x}_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 i 번째 단말의 최소 수신율을 최대화하는 변수이고,  $H_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국 및 상기 i번째 단말 간의 채널 이득을 나타내고,  $a_j^{(t-1)}$  는 시간 슬롯 t-1에서 상기 j번째 기지국의 행동 정보를 나타냄.

**청구항 6**

제 5 항에 있어서,

상기 행동 정보는 이하의 수학적식에 따라 표현됨을 특징으로 하는 방법,

$$a_j^{(t)} = \lambda_j^{(t)}$$

여기에서  $a_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국의 행동 정보이고,  $\lambda_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국의 라그랑주 이중 변수인 과금 정보를 나타냄.

### 청구항 7

제 6 항에 있어서,

상기 강화학습 모델은 이하의 수학적식에 따라 보상이 최대값이 되도록 학습됨을 특징으로 하는 방법,

$$r_j^{(t)} = 2 \sqrt{\sum_{j'=1}^B \sum_{i=1}^M \tilde{P}_{ij'}^{(t)} + \sum_{i=1}^M P_{ij}^{(t)} - W \sum_{j'=1}^B a_{j'}^{(t-1)}}$$

여기에서  $r_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 강화학습 모델의 보상이고,  $P_{ij}^{(t)}$  는  $\tilde{x}_{ij}^{(t-1)} a_{ij}^{(t-1)} R_{ij}^{-1}$  로 정의되고,  $R_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 상기 i번째 단말 및 상기 j번째 기지국의 주파수 효율이고,  $\hat{P}_{ij}^{(t)}$  는  $P_{ij}^{(t)}$  의 양자화된 정보로 상기 정보를 나타내고, W는 상기 j번째 기지국의 총 주파수 대역을 나타냄.

### 청구항 8

제 1 기지국에 포함된 제어장치에 있어서, 상기 제어장치는,

복수의 기지국들 중 상기 제 1 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 수신하고, 상기 제 1 기지국을 제외한 상기 복수의 기지국들의 제어장치들로부터 정보를 수신하는 통신장치;

강화학습 모델을 저장하는 저장장치; 및

상기 상태정보를 강화학습 모델에 입력하여 상기 제 1 기지국의 과금 정보를 예측하고, 상기 과금 정보를 예측한 결과를 이용하여 상기 제 1 기지국 및 상기 복수의 단말들 간의 연결 상태를 제어하는 연산장치;를 포함하는 제어장치.

### 청구항 9

제 8 항에 있어서,

상기 상태정보는 상기 제 1 기지국 및 상기 복수의 단말들 간의 채널 이득 및 상기 제 1 기지국의 이전 상태의 행동 정보를 포함함을 특징으로 하는 제어장치.

### 청구항 10

제 9 항에 있어서,

상기 행동 정보는 상기 제 1 기지국 및 상기 복수의 단말들 간 연결 프로토콜에 필요한 과금 정보를 포함함을 특징으로 하는 제어장치.

### 청구항 11

제 8 항에 있어서,

상기 정보는 상기 제 1 기지국을 제외한 상기 복수의 기지국들에 대한 상태정보 중 그래디언트(Gradient) 계산에 필요한 정보를 포함함을 특징으로 하는 제어장치.

### 청구항 12

제 9 항에 있어서,

상기 상태정보는 이하의 수학적식에 따라 표현됨을 특징으로 하는 제어장치,

$$s_j^{(t)} = [\tilde{x}_{1j}^{(t)}, \dots, \tilde{x}_{Mj}^{(t)}, H_{1j}^{(t)}, \dots, H_{Mj}^{(t)}, a_j^{(t-1)}]$$

여기에서  $s_j^{(t)}$  는 시간 슬롯 t에서 j번째 기지국의 상태 정보이고,  $\tilde{x}_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 i 번째 단말의 최소 수신율을 최대화하는 변수이고,  $H_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국 및 상기 i번째 단말 간의 채널 이득을 나타내고,  $a_j^{(t-1)}$  는 시간 슬롯 t-1에서 상기 j번째 기지국의 행동 정보를 나타냄.

### 청구항 13

제 12 항에 있어서,

상기 행동 정보는 이하의 수학적식에 따라 표현됨을 특징으로 하는 제어장치,

$$a_j^{(t)} = \lambda_j^{(t)}$$

여기에서  $a_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국의 행동 정보이고,  $\lambda_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 j번째 기지국의 라그랑주 이중 변수인 과금 정보를 나타냄.

### 청구항 14

제 13 항에 있어서,

상기 강화학습 모델은 이하의 수학적식에 따라 보상이 최대값이 되도록 학습됨을 특징으로 하는 제어장치,

$$r_j^{(t)} = 2 \sqrt{\sum_{j' \neq j}^B \sum_{i=1}^M \tilde{P}_{ij'}^{(t)} + \sum_{i=1}^M P_{ij}^{(t)} - W \sum_{j'=1}^B a_{j'}^{(t-1)}}$$

여기에서  $r_j^{(t)}$  는 상기 시간 슬롯 t에서 상기 강화학습 모델의 보상이고,  $P_{ij}^{(t)}$  는  $\tilde{x}_{ij}^{(t-1)} a_{ij}^{(t-1)} R_{ij}^{-1}$  로 정의되고,  $R_{ij}^{(t)}$  는 상기 시간 슬롯 t에서 상기 i번째 단말 및 상기 j번째 기지국의 주파수 효율이고,  $\hat{P}_{ij}^{(t)}$  는  $P_{ij}^{(t)}$  의 양자화된 정보로 상기 정보를 나타내고, W는 상기 j번째 기지국의 총 주파수 대역을 나타냄.

## 발명의 설명

### 기술 분야

[0001] 본 개시는 강화학습을 이용하여 채널 정보가 제한된 네트워크에서 기지국을 제어하는 방법 및 장치에 관한 것이다.

### 배경 기술

[0002] 다중 셀 이종 네트워크(Heterogeneous Network)에서 다수의 사용자들이 서비스를 요청할 때, 일반적으로 신호 대 잡음비(Signal to Interference plus Noise Ratio, SINR)가 가장 높은 기지국에 단말이 연결된다. 그리고 기지국의 전송 전력은 송신할 수 있는 최대전력을 이용한다. 즉, 이러한 이동통신 기지국 시스템은 기본적으로 단말이 신호 대 잡음비가 가장 높은 기지국에 연결되기 때문에 특정 기지국에 과도하게 부하가 걸리는 문제가 발생한다.

[0003] 최적화 이론으로 본 문제를 해결할 수 있지만, 최적화 이론 기반의 알고리즘은 셀 내 모든 기지국의 채널 정보를 요구하여 네트워크 오버헤드가 증가하며, 기지국들 중 하나의 대표 기지국이 모든 계산을 하고 다시 계산 결

과를 다른 기지국에 전달하여 지연시간 측면에서도 큰 손해가 있다.

**발명의 내용**

**해결하려는 과제**

[0004] 본 개시는 강화학습을 이용하여 채널 정보가 제한된 네트워크에서 기지국을 제어하는 방법 및 장치를 제공하는 데 있다.

**과제의 해결 수단**

[0005] 본 개시의 일 실시예에 따르면, 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법을 제공한다. 상기 방법은, 복수의 기지국들 중 제 1 기지국의 제 1 제어장치가 상기 제 1 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 수신하는 단계; 상기 제 1 제어장치가 상기 제 1 기지국을 제외한 상기 복수의 기지국들의 제어장치들로부터 정보를 수신하는 단계; 상기 제 1 제어장치가 상기 상태정보를 강화학습 모델에 입력하여 상기 제 1 기지국의 과금 정보를 예측하는 단계; 및 상기 제 1 제어장치가 상기 과금 정보를 예측한 결과를 이용하여 상기 제 1 기지국 및 상기 복수의 단말들 간의 연결 상태를 제어하는 단계;를 포함한다.

[0006] 본 개시의 일 실시예에 따르면, 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국의 제어장치가 제공된다. 상기 제어장치는, 제 1 기지국에 포함된 제어장치에 있어서, 상기 제어장치는, 복수의 기지국들 중 상기 제 1 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 수신하고, 상기 제 1 기지국을 제외한 상기 복수의 기지국들의 제어장치들로부터 정보를 수신하는 통신장치; 강화학습 모델을 저장하는 저장장치; 및 상기 상태정보를 강화학습 모델에 입력하여 상기 제 1 기지국의 과금 정보를 예측하고, 상기 과금 정보를 예측한 결과를 이용하여 상기 제 1 기지국 및 상기 복수의 단말들 간의 연결 상태를 제어하는 연산장치;를 포함한다.

**발명의 효과**

[0007] 본 개시의 실시예들은 다음의 장점들을 포함하는 효과를 가질 수 있다. 다만, 본 개시의 실시 예들이 이를 전부 포함하여야 한다는 의미는 아니므로, 본 개시의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0008] 본 개시의 일 실시예에 따르면 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법 및 장치는 특정 기지국에 집중되는 부하를 조절하는 효과가 있다.

[0009] 또한, 본 개시의 일 실시예에 따르면, 이동 통신 네트워크 내의 사용자 연결을 재설정하여 네트워크 전역의 성능을 향상시키는 효과가 있다.

[0010] 또한, 본 개시의 일 실시예에 따르면, 이동 통신 네트워크 내 사용자의 위치에 관계없이 일관적인 전송률을 제공할 수 있는 효과가 있다.

[0011] 또한, 본 개시의 일 실시예에 따르면, 필요한 정보 교환량 및 네트워크 프로토콜의 복잡도를 감소시키는 효과가 있다.

**도면의 간단한 설명**

[0012] 도 1은 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국의 제어 프로토콜을 나타낸 도면이다.

도 2는 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법에 대한 순서도를 나타낸 도면이다.

도 3은 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어장치에 대한 블록도를 나타낸 도면이다.

도 4는 기지국의 제어장치가 교환하는 정보의 총량 별 단말들의 보장된 QoS를 나타낸 도면이다.

도 5는 본 개시의 일 실시예에 따른 정해진 QoS 요구치를 만족하지 않는 단말들의 비율을 나타낸 도면이다.

**발명을 실시하기 위한 구체적인 내용**

- [0013] 이하에서는 첨부한 도면을 참고하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 지닌 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나, 이는 본 개시를 특정한 실시 형태에 대해 한정하는 것이 아니며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 그리고 도면에서 본 개시를 설명하기 위해 설명과 관련없는 부분은 생략하였으며, 명세서 전체를 통해 유사한 부분에 대해서는 유사한 도면 부호를 붙인다.
- [0014] 제 1, 제 2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0015] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 한다. 그리고 "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.
- [0016] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다.
- [0017] 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다. 따라서, 본 명세서를 통해 설명되는 각 구성부들의 존재 여부는 기능적으로 해석되어야 할 것이다.
- [0018] 명세서 전체에서, 단말 (terminal)은 사용자 장비 (user equipment, UE), 이동 단말 (mobile terminal, MT), 이동국 (mobile station, MS)을 지칭하거나 포함한다. 또한, 기지국(base station, BS)은 노드 B (node B), 고도화 노드 B (evolved node B, eNB), 차세대 노드 B(next generation node B, gNB), 소형 기지국 (small cell BS)을 지칭하거나 포함할 수 있다.
- [0019] 본 개시의 실시예들에 따른 단말과 기지국 간의 연결 시스템은 다양한 전송 전력을 가지는 기지국이 배치된 이종 기지국 네트워크를 포함한다. 긴 시간간격으로 사용자 연결을 제어하는 것을 가정하여, 주파수 측면으로 변하는 선택적 페이딩 효과를 상쇄하여 흡사 주파수 비 선택적 페이딩환경과 같이 나타낼 수 있으며, 주파수 비 선택적 환경에서의 사용자 연결 또한 포함한다. 또한, 모든 기지국들이 동일한 주파수 자원을 사용하기 때문에 기지국들 간 신호 간섭이 발생한다. 각 기지국의 제어 장치에서는 로컬 채널 상태 정보 (local channel state information)을 얻을 수 있다고 가정한다.
- [0020] 도 1은 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국의 제어 프로토콜을 나타낸 도면이다.
- [0021] 도 1을 참조하면, 복수의 기지국들(BS1, BS2, ... BS B)에는 각각 하나의 제어장치가 탑재되어 있으며 각 기지국들은 과금 정보를 계산하여 네트워크 내의 복수의 단말들(UE1, UE2, ... UEM)에게 예를 들어, 브로드캐스트 (Broadcast) 방식으로 전달한다. 각 기지국들의 과금 정보를 수신한 복수의 단말들은 각 기지국마다의 주파수 효율을 연산한다. 복수의 단말들은 각 기지국들의 과금 정보 및 연산된 각 기지국의 주파수 효율을 기반으로 최적의 기지국에 연결을 수행한다. 예를 들어 복수의 단말들은 랜덤 액세스(Random Access) 절차를 통해 최적의 기지국에 연결을 수행할 수 있다. 또한, 각 기지국은 연결된 사용자들에게 주파수 자원을 할당하고 필요한 정보 및/또는 데이터를 전송한다. 예를 들어, 각 기지국은 랜덤 액세스(Random Access) 절차를 통해 연결 요청을 한 단말들에 주파수 자원을 할당할 수 있다. 본 개시는 이와 같이 복수의 기지국들 및 복수의 단말들 간의 연결에 필요한 최적의 과금 정보를 복수의 기지국들 및 복수의 단말들이 포함된 이동통신 네트워크의 상태 정보에 기반한 강화학습 모델을 통해 구하는 기법을 제안한다.
- [0022] 한편, 네트워크의 상태정보는 각 기지국 및 복수의 단말들 간의 무선통신 네트워크에 대한 상태정보를 의미한다. 예컨대, 각 기지국 및 복수의 단말들 간 채널 이득 및 이전 상태의 각 기지국의 행동 정보를 포함한

다. 이 때, 해당 정보는 각 기지국과 단말 간 연결 프로토콜에 필요한 과금 정보를 포함한다.

[0023] 한편, 각 기지국의 제어장치는 강화학습 모델을 포함한다. 강화학습 모델은 기지국의 과금 정보를 예측하는 모델이다. 할 수 있다. 예를 들어, 강화학습 모델은 Q 러닝(Q-Learning) 모델, DQN(Deep Q Network) 모델, 및/또는 액터-크리틱(Actor-Critic) 모델을 포함할 수 있다.

[0024] 강화학습 모델은 상태정보에 따라 각 기지국의 과금 정보를 예측할 수 있다. 강화학습 모델의 예측값은 각 기지국이 복수의 단말들 간의 연결상태를 유지할 것인지 해제할 것인지를 결정하는 값이다. 강화학습 모델은 사전에 과금 정보에 대한 보상이 최대가 되도록 학습하는 과정을 수행한다. 강화학습 모델의 예측 결과를 토대로 각 기지국은 복수의 단말들과의 연결을 효율적으로 제어할 수 있다.

[0025] 한편, 설명의 편의를 위해 단말과 기지국의 인덱스를 각각  $i \in \{1, \dots, M\}$  와  $j \in \{1, \dots, B\}$ 로 나타낸다. 또한, 각 사용자연결 및 전력제어의 최소 단위인 타임 슬롯은  $t \in \{1, \dots, T\}$ 로 표현된다. 기지국과 단말 간의 연결을 나타내는 변수는 아래 수학적 식 1과 같다.

### 수학적 식 1

[0026] 
$$x_{ij}^{(t)} = \begin{cases} 1, & \text{if user } i \text{ is served by BS } j \text{ at time slot } t \\ 0, & \text{otherwise} \end{cases}$$

[0027] 상기 수학적 식 1에서 변수  $x_{ij}^{(t)}$  는 0 또는 1의 값을 가지며, 시간 슬롯 t에서 단말 i가 기지국 j에 연결된 경우 해당 변수는 1의 값을 가지고 나머지 경우 0의 값을 가진다. 여기에서 기지국 및 단말 간의 연결은 1:1을 전제 조건으로 한다. 즉, 단말은 한번에 하나의 기지국에만 연결될 수 있다. 따라서, 기지국과 단말 간의 연결을 나타내는 변수  $x_{ij}^{(t)}$  에 대한 제약조건은 다음의 수학적 식 2로 정의된다.

### 수학적 식 2

[0028] 
$$\sum_{k=1}^B x_{ik}^{(t)} = 1, \text{ and } x_{ik}^{(t)} \in \{0,1\}$$

[0029] 한편, 기지국(들)이 단말에 할당하는 주파수 자원을 나타내기 위해 변수  $y_{ij}^{(t)}$  를 정의한다. 예를 들어, 총 주파수 대역폭이 20MHz일 때, 기지국 j가 단말 i에 10MHz만큼의 주파수 대역을 할당한 경우 변수  $y_{ij}^{(t)}$  는 0.5가 된다. 따라서, 변수  $y_{ij}^{(t)}$  에 대한 제약조건은 다음의 수학적 식 3으로 정의된다.

### 수학적 식 3

[0030] 
$$\sum_{k=1}^M \check{y}_{kj}^{(t)} = 1, \text{ and } \check{y}_{kj}^{(t)} \in [0,1]$$

[0031] 본 개시는 네트워크 내 단말들의 최소 데이터 수신율을 최대화하는 것으로 단말 i와 기지국 j 간 주파수 효율을  $R_{ij}^{(t)}$  로 정의하고, 기지국 j의 총 주파수 대역을 W라고 하였을 때, 단말 i의 데이터 수신율은  $\sum_{j=1}^B R_{ij}^{(t)} y_{ij}^{(t)} x_{ij}^{(t)}$  로



정의된다. 이 때, 단말들의 최소 수신율을 최대화하는 변수  $\check{y}_{ij}^{(t)}$  는 다음의 수학적 식 4로 정의된다.

**수학적 식 4**

$$\check{y}_{ij}^{(t)} = \frac{W}{R_{ij}^{(t)}} \left( \sum_{i=1}^M \left( \frac{\check{x}_{ij}^{(t)}}{R_{ij}^{(t)}} \right) \right)^{-1}$$

[0032]

단말들의 최소 수신율을 최대화하는 변수  $\check{x}_{ij}^{(t)}$  를 얻기 위한 라그랑주 이중 함수는 다음 수학적 식 5로 정의된다.

**수학적 식 5**

$$2 \sqrt{\sum_{j=1}^B \sum_{i=1}^M \frac{\check{x}_{ij}^{(t)}}{R_{ij}^{(t)}}} - W \sum_{j=1}^B \lambda_j^{(t)}$$

[0034]

여기서, 변수  $\lambda_j^{(t)}$  는 시간 슬롯 t에서 기지국 j의 라그랑주 이중 변수인 과금 정보를 나타낸다. 또한, 일차 변수  $\check{x}_{ij}^{(t)}$  의 해는 다음 수학적 식 6으로 정의된다.

**수학적 식 6**

$$\check{x}_{ij}^{(t)} = \begin{cases} 1, & \text{if } j = \operatorname{argmin}_{j=1, \dots, B} \frac{\lambda_j^{(t)}}{R_{ij}^{(t)}} \\ 0, & \text{otherwise} \end{cases}$$

[0036]

한편, 본 개시의 실시예들에 따른 강화학습 모델은 아래의 과정에 따라 보상(Reward)이 최대값이 되도록 학습하는 과정을 거친다. 먼저 강화학습 모델의 상태 정보(State)는 다음 수학적 식 7로 정의된다.

**수학적 식 7**

$$s_j^{(t)} = [\check{x}_{1j}^{(t)}, \dots, \check{x}_{Mj}^{(t)}, H_{1j}^{(t)}, \dots, H_{Mj}^{(t)}, a_j^{(t-1)}]$$

[0038]

여기서,  $H_{ij}^{(t)}$  는 시간 슬롯 t에서 기지국 j와 단말 i 간 채널 이득을 나타내고,  $a_j^{(t-1)}$  는 이전 시간 슬롯 t-1에서 기지국 j의 행동(Action)을 나타낸다. 강화학습 모델에서 각 기지국의 제어장치는 과금 정보를 제어하므로 기지국 j의 행동  $a_j^{(t)}$  는 다음 수학적 식 8로 정의된다.

**수학적 식 8**

$$a_j^{(t)} = \lambda_j^{(t)}$$

[0040]

다음으로 강화학습 모델의 행동을 라그랑주 이중 변수인 과금 정보로 설정하였기 때문에, 기재된 기술의 보상은

[0041]

다음 수학적 식 9와 같이 라그랑주 이중 함수로 정의된다.

**수학적 식 9**

$$r_j^{(t)} = 2 \sqrt{\sum_{j'=1}^B \sum_{i=1}^M \hat{P}_{ij'}^{(t)} + \sum_{i=1}^M P_{ij}^{(t)} - W \sum_{j'=1}^B a_{j'}^{(t-1)}}$$

[0042]

[0043] 여기서,  $P_{ij}^{(t)}$  는  $\hat{x}_{ij}^{(t-1)} a_{ij}^{(t-1)} R_{ij}^{-1}$  로 정의되며  $\hat{P}_{ij}^{(t)}$  는  $P_{ij}^{(t)}$  의 양자화된 정보를 나타내며 각 기지국의 제어장치 간에 교환되는 정보를 나타낸다.

[0044] 수학적 식 7, 수학적 식 8, 수학적 식 9에서 정의된 강화학습 모델의 상태, 행동, 보상 정보는 DDPG(Deep Deterministic Policy Gradient) 방식을 통해 학습되며, 최적의 과금 정보를 얻을 수 있다. 따라서, 네트워크 내 기지국들은 제어장치의 강화학습 모델에서 계산된 결과를 각각 전송받아 단말과의 연결을 유지 또는 끊을 수 있고, 송신 전력을 제어할 수 있다.

[0046] 도 2는 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법에 대한 순서도를 나타낸 도면이다.

[0047] 도 2를 참조하면, 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법(200)은 210 내지 240 단계를 포함한다. 강화학습을 이용한 채널 정보가 제한된 기지국 제어 방법(200)은 기지국의 제어장치를 통해 수행될 수 있다.

[0048] 210 단계에서 각각의 제어장치는 해당 기지국과 복수의 단말들 간의 무선통신 네트워크에 대한 상태 정보를 수신한다. 상태 정보는 해당 기지국 및 복수의 단말들 간 채널 이득 정보 및 해당 기지국의 이전 상태의 행동 정보를 포함한다. 해당 기지국의 이전 상태의 행동 정보는 해당 기지국과 복수의 단말들 간의 연결 프로토콜에 필요한 과금 정보를 포함한다.

[0049] 220 단계에서 각각의 제어장치는 유·무선 통신을 통해 정보를 교환한다. 예컨대, 각각의 제어장치는 복수의 기지국들 간 설정된 백홀 시스템을 통해 해당 정보를 교환할 수 있다. 교환되는 정보는 수신한 상태 정보 중 그라디언트(Gradient) 계산에 필요한 정보를 포함한다.

[0050] 230 단계에서 제어장치는 수신한 상태 정보 및 교환된 정보를 이용하여 강화학습 모델을 통해 해당 기지국의 과금 정보를 예측한다. 예컨대, 강화학습 모델에 수신한 상태 정보를 입력하여 과금 정보를 예측할 수 있다. 강화학습 모델의 예측값은 해당 기지국이 연결된 단말들과의 주파수 효율을 고려하여 연결을 유지할 것인지 연결을 종료할 것인지를 결정하기 위한 값이다. 강화학습 모델은 Q 러닝 모델, DQN 모델, 및/또는 액터-크리틱 모델을 포함한다.

[0051] 240 단계에서 제어장치는 강화학습 모델의 예측값을 통해 해당 기지국 및 복수의 단말들 간의 연결 상태를 제어한다. 예컨대, 강화학습 모델의 예측값을 기준값으로 하여 해당 기지국이 복수의 단말들과의 연결 상태를 유지할 것인지 연결을 종료할 것인지를 결정할 수 있다.

[0053] 도 3은 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어장치에 대한 블록도를 나타낸 도면이다.

[0054] 도 3을 참조하면, 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어장치(300)는 통신장치(310), 저장장치(320) 및 연산장치(330)를 포함한다. 제어 장치(300)는 기지국 내 탑재된 단말기 또는 서버일 수도 있고 기지국과 연결된 단말기 또는 서버일 수 있다.

[0055] 통신장치(310)는 해당 기지국과 복수의 단말들 간의 무선통신 네트워크에 대한 상태 정보를 수신한다. 또한, 통신장치(310)는 그라디언트 계산에 필요한 정보를 다른 제어장치들과 교환한다. 예컨대, 통신장치(310)는 네트워크

크 내 글로벌 채널을 통해 제어장치(300)이 탑재된 기지국 외 다른 기지국의 제어장치들로부터 그래디언트 계산에 필요한 정보를 수신할 수 있다.

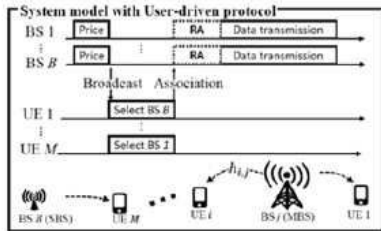
- [0056] 저장장치(320)는 강화학습 모델을 저장한다. 저장장치(320)는 제어장치(300)의 메모리로 구현될 수 있다. 저장장치(320)는 학습이 완료된 강화학습 모델을 저장할 수도 있고, 초기 상태의 모델을 저장하고 이후 학습 과정에 따라 저장된 모델을 불러올 수도 있다.
- [0057] 연산장치(330)는 상태정보를 강화학습 모델에 입력하여 자신이 탑재된 기지국의 과금 정보를 예측한다. 연산장치(330)는 제어장치(300)의 CPU 또는 AP로 구현될 수 있다. 또한, 연산장치(330)는 강화학습 모델의 예측 결과를 기반으로 제어장치(300)가 탑재된 기지국 및 상기 복수의 단말들 간의 연결상태를 제어한다.
- [0058] 한편, 상술한 바와 같은 강화학습을 이용한 채널 정보가 제한된 기지국 제어 장치(300)는 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 즉, 드론에 탑재되거나 기지국 내 설치된 컴퓨터 상에서 실행되는 프로그램일 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.
- [0059] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM (read-only memory), PROM (programmable read only memory), EPROM(Erasable PROM, EPROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.
- [0060] 일시적 판독 가능 매체는 스태틱 램(Static RAM, SRAM), 다이내믹 램(Dynamic RAM, DRAM), 싱크로너스 디램(Synchronous DRAM, SDRAM), 2배속 SDRAM(Double Data Rate SDRAM, DDR SDRAM), 증강형 SDRAM(Enhanced SDRAM, ESDRAM), 동기화 DRAM(Synclink DRAM, SLDRAM) 및 직접 램버스 램(Direct Rambus RAM, DRRAM) 과 같은 다양한 RAM을 의미한다.
- [0062] 도 4는 기지국의 제어장치가 교환하는 정보의 총량 별 단말들의 보장된 QoS를 나타낸 도면이다. 도 4의 결과를 얻기 위한 실험 환경을 10개의 기지국과 40개의 단말들이 배치된 네트워크를 구축하고, 네트워크 내에는 각 6개의 피코셀 기지국들, 3개의 마이크로 셀 기지국들, 및 1개의 매크로셀 기지국이 배치된 것으로 가정하였다. 또한, 피코셀 기지국, 마이크로 셀 기지국, 및 매크로셀 기지국의 송신 전력을 각각 24dBm, 30dBm, 및 46dBm으로 설정하였다. 도 4에서 x축은 각 기지국의 제어장치가 교환하는 정보의 총량을 나타내며, 예를 들어, 1비트의 로컬 보상을 10ms 마다 기지국 간에 교환한다면 9kbps만큼의 링크 용량이 기지국들 간에 필요하다.
- [0063] 도 4를 참조하면, 본 개시에서 제안된 기법(예: DRL-MMF UARA)은 기존의 기법들(예: max-SINR-MMF-UARA 또는 CSI-EX-MMF-UARA)대비 높은 QoS(quality of service)를 보장 가능하다. 즉, 본 개시에서 제안된 기법은 기존 기법들 대비 강화학습을 통해 적은 량의 정보 교환만으로 정확한 과금 정보를 예상할 수 있으며, 18kbps 만큼의 정보 교환만으로도 성능의 상한값에 가깝게 달성할 수 있다.
- [0065] 도 5는 본 개시의 일 실시예에 따른 정해진 QoS 요구치를 만족하지 않는 단말들의 비율을 나타낸 도면이다. 도 5의 실험환경은 도 4와 마찬가지로 10개의 기지국과 40개의 단말들이 배치된 네트워크를 구축하고, 네트워크 내에는 각 6개의 피코셀 기지국들, 3개의 마이크로 셀 기지국들, 및 1개의 매크로셀 기지국이 배치된 것으로 가정하였다. 또한, 피코셀 기지국, 마이크로 셀 기지국, 및 매크로셀 기지국의 송신 전력을 각각 24dBm, 30dBm, 및 46dBm으로 설정하였다.
- [0066] 도 5를 참조하면, QoS 요구치가 올라갈수록 QoS를 만족하지 않는 단말들의 수가 기하급수적으로 증가하는 것을 알 수 있다. 따라서, 본 개시에서 제안된 기법(예: DRL-MMF UARA)은 단말 사용자들의 최소 정보 수신율을 최대화하기 위해 사용자 연결과 주파수 자원 할당을 최적화하여 기존 기술들(예: Max-SINR-MMF UARA) 대비 더 많은 단말들의 QoS를 보장할 수 있다.
- [0068] 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국 제어 방법 및 장치는

이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 개시의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

도면

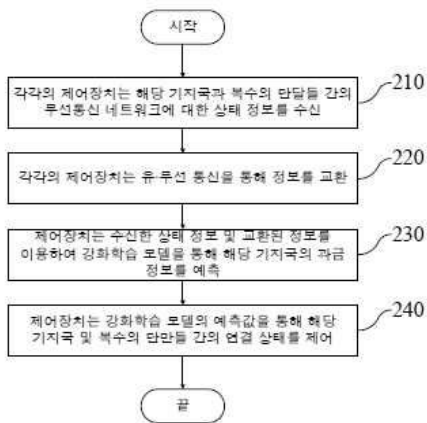
도면1

100

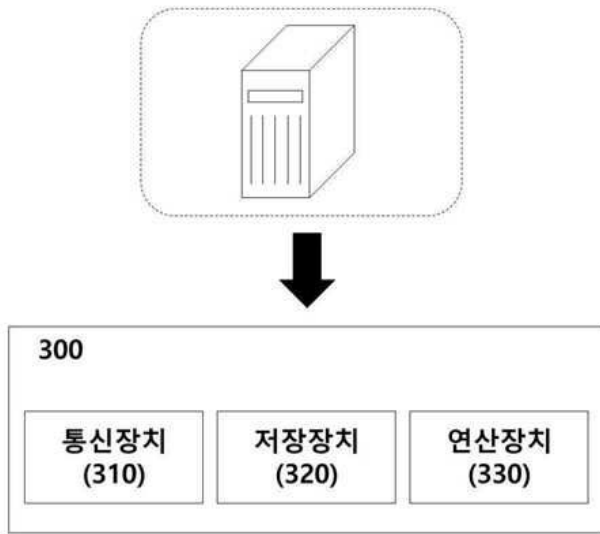


도면2

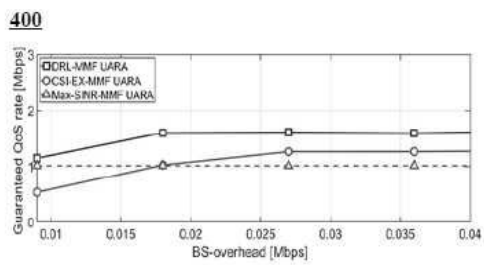
200



도면3



도면4



도면5

