**Minimum Throughput Maximization for Multi-UAV Enabled WPCN: A Deep Reinforcement Learning Method**

[**https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8950047**](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8950047)

**0. SYSTEM MODEL**

|  |
| --- |
| 모든 UAV에는 **하나의 안테나가 있고 서로 같은 frequency band를 공유**한다.  **:** 특정 영역에 있는 IoT 장치의 집합으로, **은** 를 나타낸다.   * 이때 이 성립한다. * 이때 어떤 클러스터 에 대해서도 **HTT 프로토콜이 적용된 UAV-enabled TDMA 시스템을 고려**한다.   **:** UAV의 특정 flight period   * 이때 **Device** 과 UAV 은 다음과 같이 나타내어진다. * 여기서 는 **UAV 의 고도**를 나타낸다. * 분석을 쉽게 하기 위해서 **flight period T는 N+1개의 time slot으로 분할**된다고 하자.   가 **UAV의 최대 속도**일 때 **UAV의 위치**는 다음을 만족시켜야 한다.  이때 UAV와 IoT device들의 channel condition은 **air-to-ground channel**로 간주되며, 이때 **LoS와 non-line-of-sight(NLoS)가 랜덤하게 발생**한다. **LoS의 발생 확률**은 다음과 같다. |

**0. SYSTEM MODEL**

|  |
| --- |
| **UAV 과 wireless device** 에 대하여, **LoS와 NLoS link에서의 path loss 모델**은 다음과 같다.  이때 **UAV 과 wireless device**  간의 **channel power gain**은 다음과 같다. |

**1. TDMA and HTT transmission protocol**

|  |
| --- |
| N+1개의 time slot은 다음과 같이 할당된다.   * **0번째 time slot: downlink WPT** * **n번째 time slot (단, ): uplink WIT**   이때 **은 0 또는 1**이고, 이것은 **IoT device 이 UAV 과 통신하는지의 여부**를 나타낸다. 또한 TDMA 프로토콜이 적용되었으므로 **time resource allocation**에 대한 다음과 같은 **제약 조건**이 고려되어야 한다.  각 flight period의 **0번째 slot**에서 **UAV는 transmit power 로 downlink energy signal을 전송**하므로, **각 IoT device 의 period T에서의 collected energy**는 다음과 같다.  **Device 에서의 n-th time slot에서의** **이용 가능한 에너지**는 다음과 같다. |

**1. TDMA and HTT transmission protocol**

|  |
| --- |
| 따라서 **IoT deivce 의 uplink power의 upper bound**는 다음을 만족시킨다.  따라서, **IoT device** 와 연결된 **UAV 에 대한 time slot n에서의 received SINR**은 다음과 같다.  이때 **IoT device** 의 **instantaneous throughput**은 다음과 같다.  따라서 **flight cycle이 T**인 **IoT device** 의 **평균 throughput** 은 다음과 같다. |

**2. Problem Formulation**

|  |
| --- |
| **, ,** 이라고 하면 이 논문의 목표는 **multi-UAV enabled WPCN의 minimum average throughput을 IoT device의 association {}, uplink power {}, 3D trajectory {}을 joint하게 최적화하여 최대화**하는 것이다. 따라서 **throughput optimization problem**은 수학적으로 다음과 같이 정의된다. |

**3. Problem-solving using Deep Q Learning - basis**

|  |
| --- |
| Reinforcement Learning (RL)은 **최적의 해** 를 찾아서 **누적된 reward of expection을 최대화**시키는 것을 목표로 한다.  이때 **UAV는 3D area에서 비행하므로 모델은 크고 연속적인 state space**를 갖고 따라서 Q-table이 너무 커진다. 따라서 convergence rate가 느려지므로 **deep neural network (DNN)**를 사용한다.  Deep Q learning (DQL)에서는 **Q 값의 분포가 DNN에 의하여 추정**되고, 이 **DNN은 다음과 같은 loss function의 평균으로 학습**된다. |

**4. Problem-solving using Deep Q Learning – state space and action state**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **State space**는 각 agent에 대해 다음 3부분으로 구성된다.   |  |  | | --- | --- | |  | **의 위치** | |  | 각 device가 **과 통신하는 횟수** | |  | 에서의 **각 device의 average throughput** |   **Action space**는 27개의 element를 가지며 다음과 같다.   * **varying from**  |  |  | | --- | --- | |  | UAV가 좌회전한다. | |  | UAV가 우회전한다. | |  | UAV가 뒤로 돌아간다. | |  | UAV가 앞으로 이동한다. | |  | UAV가 아래로 이동한다. | |  | UAV가 위로 떠오른다. | |  | UAV가 그 위치에 계속 있다. |  * UAV가 다음 위치로 이동할 때, **energy flow를 broadcast**하거나 cluster 내에서 **best channel condition에 해당하는 device를 선택**한다. |

**5. Problem-solving using Deep Q Learning – reward**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 보상 함수는 다음과 같다.   |  |  |  | | --- | --- | --- | | 분류 | 조건 | 보상 | | Action | UAV가 행동 수행 후 **border 아래쪽에서 비행** | Penalty of **-1**  UAV는 **boundary로 이동** | | Step | 각 step에서, **UAV i와 UAV j의 trajectory에 교차**되는 부분이 있음 | Penalty of **-1**  UAV는 **이전 위치로 이동** | | Throu-ghput | 각 epoch에서 **device의 throughput이 증가하지 않음**, 즉 device가 UAV와 너무 많은 횟수로 통신하여, **energy가 고갈되어 UAV는 오직 interference만 수신** | Penalty of **-1** | | 각 epoch에서 **device의 throughput이 증가** | Reward of **+1** | | M.A.T. | 각 epoch 이후에 **어떤 cluster**의 **device의 minimum average throughput이 0**, 즉 해당 epoch에서 **어떤 device들은 UAV와 통신하지 않음** | Penalty of **-2** | | 각 epoch 이후에 **모든 device**에 대해 **minimum average throughput이 증가하지 않음** | **모든 UAV**에 대해 Penalty of **-1** | | 각 epoch 이후에 **minimum average throughput이 증가** | **모든 UAV**에 대해 Reward of **+1** |   이것을 알고리즘으로 나타내면 다음과 같다. |

**6. Simulation Result**

|  |  |
| --- | --- |
|  |  |
| Parameters | Uplink minimum throughput (# of iters) |
|  |  |
| Trajectories of UAVs optimized by Algorithm 1, for UAV=3 | Trajectories of UAVs optimized by Algorithm 1, for UAV=5 |
|  |  |
| Maximum minimum throughput (# of UAVs) |