

# UAV 기반 무선전력전송 네트워크에서 강화학습을 활용한 IoT 기기 활성화 알고리즘

김대솔<sup>1</sup>, 김소영<sup>1</sup>, 손민정<sup>1</sup>, 하서영<sup>1</sup>, Sengly Muy<sup>1</sup>, 이정륜\*  
<sup>1</sup>\*중앙대학교 전자전기공학부

<sup>1</sup>bbpsqkds@cau.ac.kr, <sup>1</sup>soyoung1455@cau.ac.kr, <sup>1</sup>mjson01@cau.ac.kr,  
<sup>1</sup>hsy.wannabe@gmail.com, <sup>1</sup>muysengly@cau.ac.kr, \*jrlee@cau.ac.kr

## IoT device activation algorithm using reinforcement learning in UAV-based wireless power transmission network

Daesol Kim<sup>1</sup>, Soyoung Kim<sup>1</sup>, Minjung Son<sup>1</sup>, Seoyoung Ha<sup>1</sup>, Sengly Muy<sup>1</sup>, Jungryun Lee\*  
<sup>1</sup>\*Chungang University School of Electrical and Electronics Engineering

### 요약

본 논문은 IoT Device 이 전력을 유선으로 공급받지 않을 경우, 그 lifetime 이 취약하다는 점을 보완하기 위하여 Energy Harvesting-UAV 를 최적화할 수 있는 1) Initial Searching 2) Clustering 알고리즘 3) 강화학습 Q-learning 알고리즘을 탑재한 것을 전제로 시뮬레이션을 진행하였고, 그 결과 평균 사용 전력 25% 감소 및 평균 비행 소요시간 55%가 감소하는 것을 관찰하였다.

### I. 서론

에너지 하베스팅은 일상생활에서 소모되거나 버려지는 에너지를 수집하여 활용 가능한 전기 에너지를 생성 및 저장하는 기술로, 단순히 에너지를 재활용하는 것을 넘어 Internet of Things(IoT) 기기 등의 소형 전기 장치를 배터리 교체 없이 수 년 간 동작시킬 수 있는 기술로 활용될 수 있기에 주목받고 있다. [1]

IoT 기술은 네트워크를 활용하여 일상을 데이터화, 자동화해준다는 장점이 있으나 배터리 용량이 작아 기기의 수명이 짧다는 한계점을 갖고 있다. 이에 UAV를 통한 에너지 하베스팅 및 무선 충전 기술을 적용하여 지형에 구애받지 않고 기기를 탐색, 충전할 수 있다.

본 연구에서는 SWIPT와 같은 RF 통신 기술을 통한 에너지 하베스팅 기술을 전제로 UAV가 에너지를 수집, 이를 전기 에너지로 바꾸어[2] 일정 지역 내에 분포되어 있는 소형 IoT 기기에 전력을 공급하는 과정에서 UAV의 이동 경로의 최적화를 통해 UAV의 충전 효율을 최대화하는 저전력 시스템을 개발하고자 한다.

### II. 본론

미확인 IoT 기기들이 임의로 분포되어 있는 환경에서 UAV가 멈춰 충전하는 지점의 개수와 위치, UAV가 이동하는 경로, UAV의 충전 영역의 최적화를 통하여 저전력 시스템을 구축하고 충전 속도를 최적화한다.

Initial Search 알고리즘을 통하여 일정 영역 내의 Ground User (GU)의 위치 정보를 파악, 이후 Find Minimum Route (FMR) 알고리즘을 적용하여 충전 UAV가 이동할 최적 경로를 구하여 저전력 시스템을 구현한다.

#### A. Initial Search 알고리즘

Initial Search 알고리즘은 UAV가 임의의 영역을 맵 돌며 방전 및 연결 해제 등의 사유로 네트워크에 등록되지 않은 IoT 기기들을 충전, 탐색하여 해당 IoT 기기들의(이하 Ground User, GU) 위치를 파악한다. 이 때, UAV의 충전 범위가 최대한 겹치지 않으면서 모든 영역의 탐색을 시행 가능하게끔 UAV의 경로를 설정한다.

$$d_{charge\ beam} = 2 \times h_{uav} \times \tan\left(60^\circ \times \frac{\pi}{180}\right)$$
$$W = 1 \times 10^{\frac{32 - 20 \times \log_{10}\left(\frac{4\pi \times d_{uav to gu} \times 10^9}{2.9 \times 10^8}\right)}{10}} \times 1000$$

#### B. K-means Clustering

Initial Search 알고리즘을 통하여 파악된 GU들의 위치 정보를 활용하여 UAV의 충전 영역을 확정하는데 있어 K-means Clustering 알고리즘을 활용한다. UAV의 전력 공급 방식은 1 대 1이 아니라 UAV의 충전 영역 내의 모든 기기들에게 전력 공급이 이루어지는 형태를 띠기 때문에 UAV가 머물면서 충전을 하는 cluster의(이하 클러스터) 위치 및 크기, 개수를 최적화함으로써 충전 시스템의 효율을 조정할 수 있다.

$$k = \arg \min_{j=1, \dots, k} \|x_n - \mu_j^t\|$$

$$\mu_k^{t+1} = \frac{\sum_{i=1}^n r_{ik} x_i}{\sum_{i=1}^n r_{ik}}$$

#### C. Re-Clustering

클러스터의 개수를 최적화하기 위하여 클러스터의 영역 내에 1개 이상의 GU가 완전히 충전되었을 경우 한번 더 clustering 과정을 거쳐 UAV의 클러스터를 재설정,

충전이 필요 없는 기기를 충전함으로써 일어나는 이동 거리 증가를 최소화하였다.

#### D. Find Minimum Cluster Number

Clustering 과정에서 cluster 개수를 몇 개로 정하느냐에 따라 그 결과 역시 달라진다. 최소한의 비행 소요시간을 도출해내는 최적의 cluster 개수를 찾기 위해 본 알고리즘이 사용된다.

#### E. 강화학습 Q-learning 알고리즘

강화학습 Q-learning 알고리즘을 이용하여 설정된 클러스터들 사이의 경로를 생성한다. Q-learning은 마르코프 결정 과정에서 Agent가 현재 state에서 어떠한 action을 취함으로써, 그 action으로부터 reward를 받아 최적의 policy를 찾아가는 알고리즘으로 머신러닝의 일종이다. 이러한 state, action, Q-value등을 표로 나타낸 것이 Q-table이다.

$$Q(s, a) = Q(s, a) + \alpha(r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s, a))$$

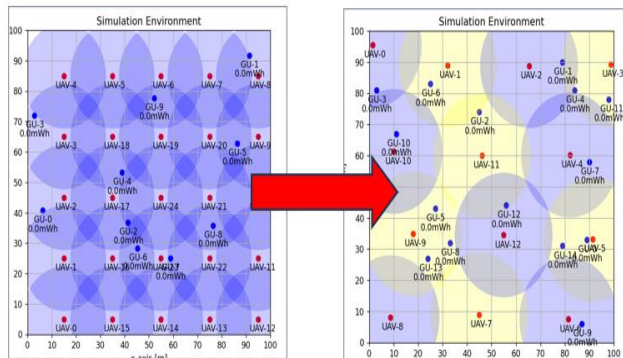
Q-value(Q(s,a))는 Q-table을 형성하고 어떤 것이 최적의 policy인지 비교할 수 있게 수치화할 수 있다. 본 연구에서 사용된 Q-value의 식은 위와 같고 Q-value에 식에서 사용되는 parameter들의 값은 아래의 표와 같다.

Learning rate( $\alpha$ )	0.8
Discount rate( $\gamma$ )	0.95
reward	소요거리에 반비례한 값
Q(s,a)	old state의 Q-value
$\max [Q(s_{t+1}, a)]$	Current state의 max Q-value

[표1. Q-value parameter table]

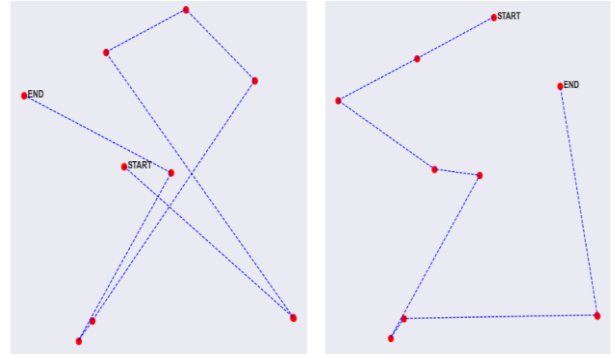
위의 parameter table에서 중요한 요소중 하나가 바로 Q(s,a), 즉 q-value이다. Q-value의 값을 선택하기 위해 선 state와 action이 필요한데, state는 현재 initial searching 및 clustering으로부터 형성된 map에서 UAV의 위치(node)를 의미하고, action은 그 UAV가 다음에 갈 수 있는 위치(node)를 의미하며, 이동 방향은 직선이고 한 episode당 한 번 다녀간 node는 더 이상 못가는 것을 전제로 하여 action과 reward를 정의하고 q-value 식을 만들었다.

#### F. 시뮬레이션 결과



[그림1. Initial Search 단계에서 Optimization 사용결과]

Initial Search 단계에서 UAV의 충전 영역이 겹치지 않게끔 hovering point를 optimization을 통해 최적화하여 전력의 낭비를 최소화하였고 수치상 평균 1200 mW에서 평균 850mW로 줄어드는 등 사용 전력이 25% 감소하였다.



[그림 2. 강화학습 전, 그림 3. 강화학습 후]



[그림 4. Rewards over training]

그림 4는 rewards over training의 graph이고 1000개의 episode를 거쳐 기계가 학습하면서 받은 reward를 뜻한다. Reward가 점점 직선과 가까운 형태를 띄우면서 최적의 policy를 발견함을 알 수 있고 그림 4와 같은 결과를 만들어냈다.

강화학습 시뮬레이션 결과 평균 비행 소요시간이 약 460초에서 210초 수준으로 감소하는 등 평균적으로 55% 감소하는 추세를 보였고, 이는 모든 경우의 수를 경험시켜 best output을 추려내는 brute forcing 기법과 비교했을 때 같은 결과를 냈음을 고려하면 강화학습이 잘 설계되었음을 확인할 수 있다.

### III. 결론

본 논문에서는 UAV의 이동 경로를 최적화하는 알고리즘을 통하여 일정 범위 내에서 미등록 IoT 기기들을 탐색 후 mapping하였고 충전 경로를 설정하여 시뮬레이션 결과 평균 사용 전력이 25% 감소, 평균 비행 소요시간을 55% 감소하는 것을 보여주며 UAV의 충전 효율을 극대화하였다.

### 참 고 문 헌

- [1] Garg, N., & Garg, R. (2017). Energy harvesting in IOT devices: A survey. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*.
- [2] Muy, S., Ron, D., & Lee, J. (2022). Energy efficiency optimization for SWIPT-based D2D-underlaid cellular networks using Multiagent Deep Reinforcement Learning. *IEEE Systems Journal*, 16(2), 3130– 3138.