A Reinforcement Learning-based Energy Efficient Trajectory and Communication Optimization for UAV-assisted 5G Network

Sarder Fakhrul Abedin, Md. Shirajum Munir, Ashis Talukder, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, South Korea
Email: {saab0015, munir, ashis, cshong}@khu.ac.kr

Abstract

In this paper, we consider an Unmanned Aerial Vehicle (UAV) based 5G network system supported by the millimeter wave (mmWave) back-haul communication. We also formulate a trajectory and communication optimization problem for the UAVs with the air-to-ground channel communication and mobility energy constraints. To solve the trajectory optimization problem, we propose a Q-learning based solution that optimizes the multi UAV trajectory path while maximizing the network utility. The performance analysis shows the convergence of the proposed approach under different learning configurations for the UAV trajectory optimization problem.

1. Introduction

In recent years, a broad spectrum of fifth generation (5G) Internet-of-Things applications [1] are supported by the Unmanned Aerial Vehicles (UAVs) as these can provide higher mobility and coverage than the traditional ground networks. More specifically, the UAVs are deployed as mobile small cells which is an alternative solution for the future communication networks [2] [3]. The UAVs are also promising for different remote data collection [4], disaster response [5], and Flash Crowds [6] solutions in order to meet the demand of seamless wireless communication by enabling effective network load-balancing [7]. Moreover, the energy-efficient trajectory for serving different IoT applications while providing efficient wireless connectivity is crucial for the battery-powered UAVs to sustain a green communication network[8].

In [9], the UAV trajectory path selection is studied for providing crowd sourcing service during the natural disaster situation. In [10], a Q-learning based IoT task offloading service is provided by the UAVs along with network resource allocation solution considering the UAV path selection constraint. In [11], the UAV-assisted cellular networks are considered to meet the demand of data rate due to flash crowds and a data-driven 3D placement of the UAVs are proposed. However, the core challenge of multiple UAV trajectory optimization with energy efficiency constraint is yet not well-investigated. Therefore the main contributions of this paper is summarized as below,

- First, a system model with the high frequency millimeter wave (mmWave)back-haul communication technology is proposed for UAV-aided wireless network.
- Second, we formulate a trajectory optimization problem for the UAVs with the air-to-ground channel modeling and mobility energy constraint. The objective of the trajectory optimization is to maximize the utilities of the UAVs while ensuring the network energy efficiency,
- Third, we model a Q-learning based solution to solve the trajectory optimization problem for the UAVs. The proposed approach effectively optimizes the multi UAV trajectory path and ensures effective data collection service for the remote IoT data collection.
- Finally, we perform the convergence analysis for the proposed approach under different learning configuration which illustrates the efficacy of the proposed approach to solve the UAV trajectory optimization problem.

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2019-2015-0-00742) supervised by the IITP(Institute for Information & communications Technology Promotion) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea goverment(MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing). *Dr. CS Hong is the corresponding author

2. System Model

In Fig. 1, we consider a set of UAV $\mathcal{U} = \{1, 2, \cdots, U\}$ as aerial user equipments and acts as service gateway for IoT devices $i \in \mathcal{I}$, a edge computing enabled ground base station [12] b. The set of UAV trajectory points are denoted as, $\mathcal{P} = \{1, 2, \cdots, P\}$ including the point of the ground base station. The links between the trajectory points are denoted as, $\mathcal{L} \subseteq \mathcal{P} \times \mathcal{P}$. Each UAV covers J number of trajectory points $\mathcal{P}_u = \{j_0, \cdots, j_J\}$ under given start and end point over a discrete time slot τ. We assume the end point of the UAVs is always the ground base station where they offload the data. We also assume, all the network equipments are equipped with directional antennas and the IoT devices transmit (i.e., uplink) the LTE-A/4G signals to the UAVs and 5G millimeter wave (mmWave) communication is considered between UAV and ground base station.

A. Air-to-Ground Communication Model

The air-to-ground LoS and NLoS path loss [11] probabilities for the the UAV $u \in \mathcal{U}$ serving IoT device $d \in \mathcal{D}$ is calculated as ,

$$\Omega_{i,u}^{p,LoS} = \frac{1}{1 + a\exp(-b(\frac{180}{\pi}\Theta_u - a))}, \Omega_{i,u}^{p,NLoS} = 1 - \Omega_{i,u}^{p,LoS} \quad (1)$$

In (1), a and b are the environment variables, $\Theta_u = \frac{h_u}{d_{u,i}}$ is the elevation angel of the UAV $u \in \mathcal{U}$ with UAV height h_u and euclidean distance from the IoT device $d_{u,i}$. The path

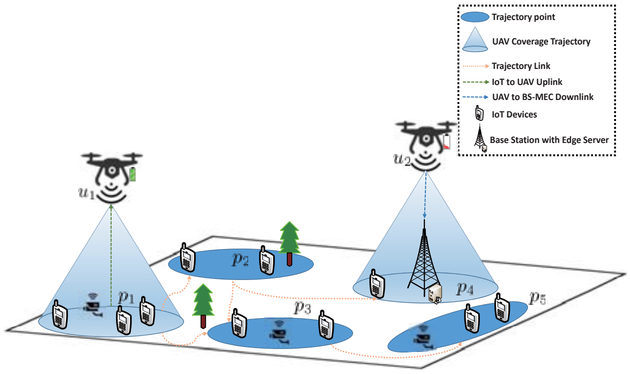


Figure 1: System model of UAV-aided 5G network.

loss in decibel (dB) for LoS and NLoS links between UAV $u \in \mathcal{U}$ and IoT device $i \in \mathcal{I}$ is calculated as,

$$\Upsilon^p_{i,u} = \begin{cases} 20\log(\frac{4\pi f_c d^p_{i,u}}{c}) + \nu, \text{LoS link}, \\ 20\log(\frac{4\pi f_c d^p_{i,u}}{c}) + \bar{\nu}, \text{NLoS link} \end{cases} \tag{2}$$

In (2), ν and $\bar{\nu}$ are the mean additional losses, f_c is the front-haul carrier frequency. The capacity of the communication link between UAV $u \in \mathcal{U}$ and IoT device $i \in \mathcal{I}$ is calculated as,

$$R^p_{i,u} = \begin{cases} \frac{B}{|\mathcal{I}|} \cdot \log\left(1 + \omega^p_{i,u}\right), \text{if } \omega^p_{i,u} > \omega_{th} \\ 0, \quad \text{otherwise.} \end{cases} \tag{3}$$

In (3), B is the front-haul subchannel bandwidth equally distributed to the IoT devices $\mathcal{I}$, $\omega^p_{i,u}$ is the signal-to-interference-plus-noise ratio (SINR) at the UAV $u \in \mathcal{U}$ which captures co-channel interference from other links, fixed transmission power and path-loss probability from (1). The data transmission is possible between the link is the $\omega^p_{i,u}$ is greater than a threshold SINR ω_{th}. The total front-haul capacity between the UAV $u \in \mathcal{U}$ and the IoT devices $i \in \mathcal{I}$ at trajectory points $p \in \mathcal{P}$, is calculated as,

$$R_{u,p} = \sum_{i \in \mathcal{I}} R^p_{i,u} \tag{4}$$

For the back-haul communication between the UAV $u \in \mathcal{U}$ and ground base station b the simple propagation in free space is adopted and the back-haul capacity is calculated as,

$$\hat{R}_{u,p} = \hat{B}^{mmWave} \cdot \log\left(1 + SNR\right) \tag{5}$$

In (5), $\hat{B}^{mmWave}$ is the fixed back-haul bandwidth.

B. UAV Mobility Energy Model

The mobility cost of UAV $u \in \mathcal{U}$ for covering different trajectory points in terms of energy consumption is calculated as,

$$\mathcal{E}_u = d_u \times [f_1||v||^3 + \frac{f_2}{||v||}(1 + \frac{||a||^2}{g^2})] \tag{6}$$

In (6), d_u covered by the UAV $u \in \mathcal{U}$ which is multiplied with the propulsion energy consumption with f_1 and f_2 depends on the UAV design and $g = 9.8 \ m/s^2$ is the gravitational acceleration.

C. Utility Function Design

The utility function of the UAV $u \in \mathcal{U}$ is defined as,

$$\mathcal{F}(\mathcal{P}_u) = \frac{\hat{R}_{u,p} + R_{u,p}}{\mathcal{E}_u} \tag{7}$$

In (7), the utility function captures the ratio between the benefit that the UAV $u \in \mathcal{U}$ provides in terms providing front-haul and back-haul capacity whereas the cost is defined as the mobility energy consumption of serving at different trajectory points $\mathcal{P}_u \subset \mathcal{P}$.

D. UAV Trajectory Optimization Problem Formulation

The objective of the UAV Trajectory Optimization Problem is to maximize the utilities of the UAVs subject to communication and mobility constraints where the decision variable is the trajectory and communication variable of the UAVs. The problem is defined as,

$$\max_{\Pi_{u,p}, \Psi^p_{i,u}} \sum_{u \in \mathcal{U}} \Pi_{u,p} \Psi_{i,u} \mathcal{F}(\mathcal{P}_u), \tag{8}$$

subject to

$$\Pi_{u,p} \hat{R}_{u,p} \le R^b_{max}, \forall u \in \mathcal{U}, \forall b \in \mathcal{P}, \tag{9}$$

$$\Pi_{u,p} \Psi^p_{i,u} R_{u,p} \ge R_i^{min}, \forall i \in \mathcal{I}, \forall p \in \mathcal{P}, \forall u \in \mathcal{U}, \tag{10}$$

$$\Pi_{u,p} \mathcal{E}_u \ge \mathcal{E}^{th}_u, \forall u \in \mathcal{U}, \forall p \in \mathcal{P}, \tag{11}$$

$$\sum_{i \in \mathcal{I}} \Psi^p_{i,u} \le K_u, \forall u \in \mathcal{U}, \tag{12}$$

$$\sum_{u \in \mathcal{U}} \Psi^p_{i,u} \le 1, \forall i \in \mathcal{I}, \tag{13}$$

$$\sum_{p \in \mathcal{P}} \Pi_{u,p} \le 1, \forall u \in \mathcal{U}, \tag{14}$$

$$\sum_{u \in \mathcal{U}} \Pi_{u,p} \le 1, \forall p \in \mathcal{P} \backslash \{b\}. \tag{15}$$

where two indicator functions $\Pi_{u,p}$ and $\Psi_{i,u}$ are defined as,

$$\Pi_{u,p} = \begin{cases} 1, & \text{if } u \text{ covers the trajectory points in } \mathcal{P} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

In (16), $\Pi_{u,p} = 1$, if the UAV $u \in \mathcal{U}$ covers the trajectory point $p \in \mathcal{P}$ in order to collect information update and $\Pi_{u,p} = 0$ otherwise.

$$\Psi_{i,u} = \begin{cases} 1, & \text{if } i \text{ is served by } u \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

In (17), $\Psi_{i,u} = 1$, if the UAV $u \in \mathcal{U}$ assigns communication sub-channel to the IoT $i \in \mathcal{I}$ and $\Psi_{i,u} = 0$ otherwise.

In problem (18), the constraints (9)-(13) are the back-haul and front-haul communication constraints whereas constraints (14)-(15) are the trajectory constraints.

3. Proposed Q-Learning based UAV Trajectory Optimization

In this paper, we the propose a Q-Learning [13] based UAV trajectory and communication optimization where we model the state space over discrete time slots $T = \{1, 2, \cdots, t\}$ is defined as, $\mathcal{S} = \{s_1, s_2, \cdots, s_t\}$ where the each state tuple is $<$ start trajectory, end trajectory, $\mathcal{E}, R, \hat{R} >$. The action space is the $\mathcal{A} = \{a_1, a_2, \cdots, a_t\}$ is the joint actions of visiting the trajectory state with the UAV in $\mathcal{U}$. Each valid action means the UAVs in $\mathcal{U}$ can cover the trajectory points in $\mathcal{P}$ in the given state $\mathcal{S}$ and the reward is calculated. Otherwise, the UAV agents receive no reward. The reward function for the actions is defined as,

$$\zeta = \begin{cases} \mathcal{F}(s,a), & \text{if conditions (9)-(15) are fulfilled for,} \\ & 1, 2, \cdots, T \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

At time t, the UAV agents take actions a_u in the given state s_t and the Q-value is defined as,

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha(\zeta + \beta \min(Q(s_{t+1}, a_{t+1}))) \tag{19}$$

In (19), learning rate is $\alpha = [0, 1]$, discount factor is β and $Q(s_{t+1}, a_{t+1}))$ is the next state-action pair.

4. Performance Evaluation

The simulation parameters are depicted in Table I. We consider the episodic training for the proposed approach where the maximum number of training episodes are set to 700.

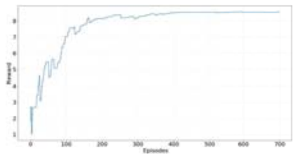
Table I: Simulation Parameters

Parameters	Values
$\lvert\mathcal{U}\rvert$, $\lvert\mathcal{P}\rvert$, $\lvert\mathcal{I}\rvert$	2, 7, 70
B, $\hat{B}^{mmWave}$	20 MHz, 100 MHz
f_c, f_c^{mmWave}	2 GHz, 28 GHz
UAV and IoT Transmission Power, ω_{th}	20 dBm, 10 dBm, 5 dBm
h_u	200 m
α	[0.5, 0.8]

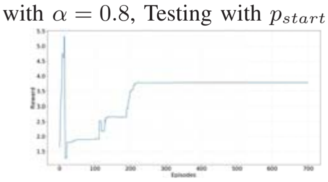
The Fig. 2 illustrates the convergence of the proposed Q-learning algorithm for trajectory optimization under two different learning rates. In Fig. 2(a), during the training, we set the learning rate at 0.8. and in the testing we set the initial trajectory state of the UAVs at $p_{start} = 1$ and the goal is set to $p_{end} = 7$. The training is converged around 500 episodes of training where the reward is maximum. On the contrary, in Fig. 2(b), the training phase is converged much faster at 250 episodes with learning rate 0.5 compare to higher learning rate. However, the reward is much lower when the learning rate is small which means with higher learning rate the training process can explore more states in order to enhance the performance gain than the lower learning rate. In this setting, for the testing, we get the optimal trajectory of given start and goal point is $[1, 2, 7]$.

5. Conclusion

In this paper, we solve the trajectory and communication optimization problem in the UAV-aided 5G wireless communication network where a Q-learning based solution optimizes the multi UAV trajectory path while maximizing



(a) Training with $\alpha = 0.8$, Testing with $p_{start} = 1, p_{end} = 7$



(b) Training with $\alpha = 0.5$, Testing with $p_{start} = 1, p_{end} = 7$

Figure 2: Convergence analysis over different learning rates

the network utility. The performance analysis shows the convergence of the proposed approach under different learning configurations and thus, enabling the UAVs to select the optimal trajectory path under communication and energy constraint.

References

[1] M. S. Munir, S. F. Abedin, M. G. R. Alam, D. H. Kim, and C. S. Hong, "Smart agent based dynamic data aggregation for delay sensitive smart city services," *Journal of KIISE*, vol. 45, no. 4, pp. 395–402, 2018.

[2] S. F. Abedin, A. K. Bairagi, M. S. Munir, and C. S. Hong, "A cooperative game based spatial correlation-aware resource allocation for narrowband internet of things," *Korea Computer Congress (KCC)*, pp. 392–394, 2018.

[3] N. I. Mowla, I. Doh, and K. Chae, "An efficient defense mechanism for spoofed ip attack in sdn based cdni," in *2015 International Conference on Information Networking (ICOIN)*. IEEE, 2015, pp. 92–97.

[4] S. F. Abedin, M. S. Munir, M. G. R. Alam, and C. S. Hong, "Energy efficient crowdsourcing for internet of things applications over nb-iot networks," *Korea Software Congress (KSC)*, pp. 434–436, 2017.

[5] S. F. Abedin, M. G. R. Alam, and C. S. Hong, "Multipolar natural disaster response management through uav and social media," *KIISE*, pp. 297–299, 2016.

[6] N. I. Mowla, I. Doh, and K. Chae, "Csdsm: Cognitive switch-based ddos sensing and mitigation in sdn-driven cdni word." *Comput. Sci. Inf. Syst.*, vol. 15, no. 1, pp. 163–185, 2018.

[7] S. F. Abedin, A. K. Bairagi, M. S. Munir, N. H. Tran, and C. S. Hong, "Fog load balancing for massive machine type communications: A game and transport theoretic approach," *IEEE Access*, vol. 7, pp. 4204–4218, 2018.

[8] S. F. Abedin, M. G. R. Alam, R. Haw, and C. S. Hong, "A system model for energy efficient green-iot network," in *2015 International Conference on Information Networking (ICOIN)*. IEEE, 2015, pp. 177–182.

[9] S. F. Abedin, M. G. R. Alam, A. K. Bairagi, A. Talukder, and C. S. Hong, "Uav-assisted intelligent crowdsourcing in natural calamity," in *The International Symposium on Perception, Action, and Cognitive Systems (PACS 2016)*, 2016.

[10] S. F. Abedin, M. G. R. Alam, and C. S. Hong, "Learning based intelligent iot task offloading and resource allocation in uav-assisted fog network," *Korea Computer Congress (KCC)*, pp. 1166–1168, 2017.

[11] C.-C. Lai, L.-C. Wang, and Z. Han, "Data-driven 3d placement of uav base stations for arbitrarily distributed crowds," *arXiv preprint arXiv:1909.11554*, 2019.

[12] S. F. Abedin, A. K. Bairagi, M. S. Munir, and C. S. Hong, "A transport theoretic approach for computational task migration in multi-access edge computing," *Korea Software Congress (KSC)*, pp. 328–330, 2018.

[13] M. S. Munir, S. F. Abedin, M. G. R. Alam, N. H. Tran, and C. S. Hong, "Intelligent service fulfillment for software defined networks in smart city," in *2018 International Conference on Information Networking (ICOIN)*. IEEE, 2018, pp. 516–521.