Poster Abstract: A QoS-aware, Energy-efficient Trajectory Optimization for UAV Base Stations using Q-Learning

Shavbo Salehi
Electrical and Computer Engineering
Department, Urmia University
Urmia, Iran
shavbo.salehi@urmia.ac.ir

Jahan Hassan
School of Engineering and
Technology, The Central Queensland
University
Sydney, Australia
j.hassan@cqu.edu.au

Ayub Bokani
School of Engineering and
Technology, The Central Queensland
University
Sydney, Australia
a.bokani@cqu.edu.au

Sayed Amir Hoseini
School of Engineering and
Technology, The Central Queensland
University
Sydney, Australia
s.hoseini@cqu.edu.au

Salil S. Kanhere
School of Computer Science and
Engineering, The University of New
South Wales
Sydney, Australia
salil.kanhere@unsw.edu.au

ABSTRACT

Next generation mobile networks have proposed the integration of Unmanned Aerial Vehicles (UAVs) as aerial base stations (UAV-BS) to serve ground nodes with potentially varying QoS requirements. However, the dependence on the on-board, limited-capacity battery of the UAV-BS limits their service continuity. While conserving energy is important, meeting the QoS requirements of the ground nodes is equally important. We present an energy-efficient trajectory optimization for the UAV-BS while satisfying QoS requirements. We model the trajectory optimization as an MDP problem and solve it using Q-Learning. Simulation results reveal that our proposed algorithm decreases the average energy consumption by nearly 55% compared to a randomly-served algorithm.

1 INTRODUCTION

Wireless communication systems , e.g., 5G and beyond, are considering UAVs as aerial base stations (UAV-BS) to provide terrestrial communications services. Their dependence on the on-board, limited-capacity battery hinders their service continuity.

Energy-efficient methods for conserving UAV energy due to mechanical activities (e.g., flying) as well as electronic activities (e.g., wireless communication) have been proposed [1] [2]. Optimizing UAV trajectory for energy-efficiency saves flying energy. Zeng et al., [3] proposed an energy-efficient trajectory optimization for UAVs which also considered the communication throughput. Our focus is to design a UAV trajectory model that minimizes flying costs while maintaining the QoS by considering all effective parameters that can be considered as different state components. Therefore, an intelligent model that makes best decisions at different states is required. The problem can be formulated as selecting an action from a finite set of choices based on the repetitive observations. We choose to employ Markov Decision Process (MDP) to formulate the trajectory optimization problem which we solve using Q-Learning. Q-Learning not only learns in which order the nodes should be served after some experiences, but also can dynamically update the decision policy if the environment or nodes behaviour change.

2 SYSTEM MODEL

We consider a UAV-BS providing communication services by flying over ground nodes which may have varying QoS requirements that are typically specified in terms of data rate and delay tolerance. In our prior work [1], the UAV-BS trajectory design was optimized using the Travelling Salesman Problem (TSP) method for energy-efficiency. Due to the complexity and lack of scalability of TSP, we employ Q-Learning in this paper to optimize the trajectory to serve nodes based on their QoS requirements. We categorize services in four classes based on the Universal Mobile Telecommunications System (UMTS) QoS, and assign a service priority (SP) to each class. The highest priority level is Conversational class and the Streaming class, Interactive class, and Background class in descending order. We consider $SP = 4$ for highest priority to $SP = 1$ for the lowest priority. These numbers are used in the Q-Learning revenue function in Section 2.1.

For modeling the energy consumption of the UAV-BS, we consider both the energy expended in flying towards the destination and also for transmitting data. The UAV-BS flying energy consumption is calculated as [2]:

$$E_f = P_0\left(\frac{1}{V} + \frac{3V}{U_{tip}^2}\right) + P_i\left(\sqrt{V^{-4} + \frac{1}{4v_0^4}} - \frac{1}{2v_0^2}\right)^{\frac{1}{2}} + \frac{1}{2}d_0 DsAV \quad (1)$$

where P_0 and P_i are the blade profile power and induced power in hovering status respectively. Also, during the UAV-BS data transmission, it has hovered above the node. $E_{DT} = T_R * P_i$ which refers to the energy expended for data transmission where T_R is the time of data transmission.

2.1 Q-Learning Formulation

In our considered scenarios, we assume static ground nodes randomly positioned in different locations and request for services with different priority levels. With no information of these variables in such stochastic environment, Q-Learning can be the best option to train an agent through numerous state-action experiences. Therefore, we model the service area on the ground as a grid and consider our UAV-BS as the Q-Learning agent. The state space is

Table 1: Simulation Parameters

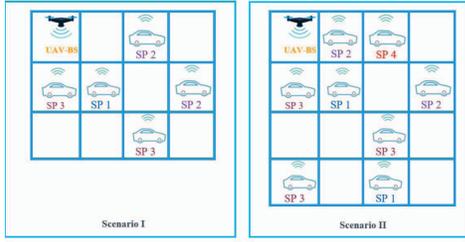| Q-Learning Component	Value
Cell side	40 (m)
Learning rate	0.5
Discount factor	0.7
Revenue adjusting factors (w_1, w_2, w_3)	100, -10, -15
Air density (D)	1.225 (kg/m^3)
Rotor radius (R)	0.5
Rotor disc area (A)	0.79
Tip speed of the rotor blade (U_{tip})	200
Rotor solidity (s)	0.05
Aircraft forward speed (V)	20
Fuselage drag ratio (d_0)	0.3
Mean rotor induced velocity in hover (v_0)	7.2

Figure 1: Simulation scenarios

created by the UAV-BS's observation of its location, nodes' locations, and service priority of each node. The Q-Learning agent must take action on the next node which it should serve. Therefore, the number of possible actions equals to the number of nodes. We use *epsilon greedy* scheme in our simulation where actions are taken randomly at the beginning of learning process when the agent is fully in exploration mode. As time goes, it decreases the exploration rate and increases the exploitation rate where the action with the highest Q-Value is taken. The Q-values are updated after serving each node using the below Q-Learning equation:

$$Q^{new}(s, a) = (1 - \alpha)Q(s, a) + \alpha(R(s, a) + \gamma Q(s', a^*)) \quad (2)$$

where s' is the next state after taking action a at state s and a^* is the maximum Q-value of all state-action pairs on state s':

$$a^* = argmax_a Q(s', a) \quad (3)$$

where α and γ are learning rate and discount factor respectively.

For our Q-Learning revenue function $R(s, a)$, we consider one positive reward when the UAV-BS serves high priority services and we apply penalty for the data transmission energy consumption and flying energy consumption. Hence, The revenue function $R(s, a)$ is calculated by equation below:

$$R = w_1 \times \overline{SP} + w_2 \times \overline{E_{DT}} + w_3 \times \overline{E_F} \quad (4)$$

where w_1, w_2 and w_3 are the tuning parameters of the normalized service priority $\overline{SP}$, the normalized data transition cost $\overline{E_{DT}}$ and the normalized flying cost $\overline{E_F}$ respectively.

3 PERFORMANCE EVALUATION

To evaluate the proposed method, we simulated two scenarios as illustrated in Fig. 1 which are service areas of a 3 x 4 and a 4 x
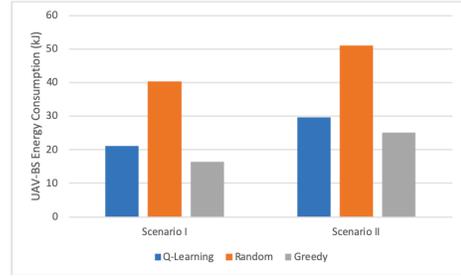

Figure 2: Energy consumption of the UAV-BS for Q-Learning-based and random serving schemes.

4 grids with five and eight randomly distributed static ground, respectively. Each node has data for transmission with a random service priority. The UAV-BS is randomly located in a cell at the start of the simulation and makes the movement decision based on the Q-Learning output to serve nodes. Performance comparison is made against two other methods wherein each state the next serving node is chosen by (i) Random: the UAV-BS chooses a random action and (ii) Greedy: the UA-BS chooses the nearest node to its current location. Simulation parameters are shown in Table 1.

As can be seen in Fig. 2, our Q-Learning based method consumes nearly 55% less energy compared to that of the Random method for both scenarios. We note that the UAV-BS consumes the least amount of energy when the greedy node-serving method is used, however it ignores application QoS requirements while only focusing on reducing energy consumption.

4 CONCLUSION AND FUTURE WORK

We presented our initial implementation of Q-Learning based, energy-efficient trajectory optimization for the UAV-BS. The optimized trajectory reduces flying costs by generating an optimal path to serve the nodes, while improving the nodes' QoS requirements by serving high-priority requests first. The results illustrate that our optimized trajectory decreases energy consumption by more than nearly 55% compared to the random nodes-serving method. In our future work, more complex scenarios including node mobility and various battery levels will be considered.

5 ACKNOWLEDGMENTS

REFERENCES

[1] Shabvo Salehi, Ayub Bokani, Jahan Hassan, and Salil S. Kanhere. 2019. AETD: An Application Aware, Energy Efficient Trajectory Design for Flying Base Stations. In *2019 IEEE 14th Malaysia International Conference on Communication (MICC)*.
[2] Yong Zeng, Jie Xu, and Rui Zhang. 2019. Energy minimization for wireless communication with rotary-wing UAV. *IEEE Transactions on Wireless Communications* 18, 4 (2019), 2329–2345.
[3] Y. Zeng and R. Zhang. 2017. Energy-Efficient UAV Communication With Trajectory Optimization. *IEEE Transactions on Wireless Communications* 16, 6 (June 2017), 3747–3760. https://doi.org/10.1109/TWC.2017.2688328