Analysis of Reinforcement Learning Schemes for Trajectory Optimization of an Aerial Radio Unit

Hossein Mohammadi[†], Vuk Marojevic[†], and Bodong Shang[⋆]
[†]Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA
[⋆]Eastern Institute for Advanced Study, Ningbo, China
{hm1125|vuk.marojevic}@msstate.edu, bdshang@hotmail.com

ICC 2023 - IEEE International Conference on Communications | 978-1-5386-7462-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICC45041.2023.10279738

***Abstract*—This paper introduces the deployment of unmanned aerial vehicles (UAVs) as lightweight wireless access points that leverage the fixed infrastructure in the context of the emerging open radio access network (O-RAN). More precisely, we introduce the aerial radio unit (A-RU) that dynamically serves an underserved area and connects to the distributed unit (O-DU) via a wireless fronthaul between the UAV and the closest fixed network infrastructure tower. In this paper we employ artificial intelligence (AI) for determining the UAV trajectory for serving User Equipment (UEs) while maintaining the fronthaul connectivity to the O-DU at the same time in a multiple-input multiple-output (MIMO) fading channel. We first formulate the trajectory time and throughput rate; however, owing to the nonconvexity of the problem of maximizing the network throughput based on UAV location, we put our effort to achieve these goals by RL approach. Three different approaches have been presented. We first divide the area into a grid and let the UAV explore the environment by flying from point A to point B using both the offline Q-learning and the online SARSA algorithm and the pathloss as the reward. With the intention of maximizing the average payoff, the trajectory in the first scenario is described as a Markov decision process (MDP). According to simulations, MDP produces better results in a smaller area and in less time. In contrast, SARSA performs better in larger environments at the expense of a longer flight duration.**

***Index Terms*—Reinforcement learning, Markov decision process, open radio access network, unmanned aerial vehicles, trajectory.**

I. Introduction

Drones, also referred to as unmanned aerial vehicles (UAVs), have gained significant interest in recent years for variety cases, including improving network coverage and delivering freight [1]. Moreover, according to shorter-term forecasts, the recreational and commercial UAV fleets are expected to reach up to 3 million by 2023, and the drone services market is expected to be worth more than 63.6 billion dollars by 2025 [2].

Due to interference issues and high handover rates, as well as other simulation-based research, it has been demonstrated that present LTE networks do not provide adequate coverage above building heights [3]. The authors of [4] provide an overview of the many cutting-edge methods for managing resources and interference in orthogonal frequency division multiple access (OFDMA)-based femtocell networks and offer a qualitative comparison of the various methods. Interference management is highlighted as the unresolved issue. In terms of coverage and handover rates, a number of measurement studies produced satisfactory results. However, they were carried out in rural areas where the base station density is significantly lower [5].

Although there are opportunities using scalable learning techniques for data-driven wireless networks, these techniques have their own problems. First, in order to reduce the load on the central nodes and decrease the latency, the computations will likely be shifted to edge devices in the future. Therefore, in order to work within the limitations of on-device computation, storage, and battery capacities, low-complexity learning models must be developed. One proposed solution is that of a self-organizing network to enable the automatic deployment of cellular networks. The difficulties include, among others, data imbalance, data insufficiency, cost insensitivity, and non-real time reaction [6]. However, by increasing the global network and the volume of data transferred among devices, these approaches become impractical, in particular for Beyond 5G (B5G) and future 6G networks.

UAV network entities must decide locally how to optimize the network performance in the face of a network environment that is unknown. When the state and action spaces are small, reinforcement learning (RL) has been effectively employed to enable network entities to achieve the best policy for decisions or actions given their states. RL however may not be able to establish the ideal policy in a reasonable amount of time in complex large-scale networks where the state and action spaces are large. In order to address this, deep RL (DRL), is proposed [7].

Data transmission through UAVs has been studied in [8], which provides a data delivery scheme to address the pricing issue for UAVs and a data delivery strategy for users to increase the effectiveness of data transmission with the aim of maximizing the functionality of both UAVs and users. This paper introduces the deployment of UAVs as lightweight wireless access points that leverage the fixed infrastructure in the context of the emerging open radio access network (O-RAN) [9]. More precisely, we introduce the aerial radio unit (A-RU) that dynamically serves an underserved area and connects to the O-RAN distributed unit (O-DU) via the wireless fronthaul between the UAV and the closest fixed network infrastructure tower [10].

II. Related Work

This section's literature review assesses the feasibility of UAV data transmission and discusses the benefits and draw-

978-1-5386-7462-8/23/$31.00 ©2023 IEEE

backs of artificial intelligence (AI) for UAV trajectory optimization.

The authors of [11] maximize the lowest throughput over all ground users in the downlink by concurrently optimizing scheduling and user association with the UAV trajectory and power control in order to achieve equitable performance among users. The formulated problem is a mixed integer nonconvex optimization problem. As a result, the authors suggest an effective iterative solution for solving it using consecutive convex optimization and block coordinate descent.

In the context of full-duplex multi-UAV networks, reference [12] investigates the decoupling of the uplink-downlink association and trajectory design challenge. The goal of the joint optimization task is maximizing the sum-rate of the UEs in both uplink and downlink. The paper suggests employing a multi-agent DRL (MADRL) approach that allows each UAV to choose its policy in a distributed fashion.

Several studies have considered the user association problem for UAV-aided communications. UAV-UE edge computing system was studied in [13]. Reference [14] investigates how to link a terminal to a UAV in the downlink and a ground station in the uplink for a mmWave UAV network. Reference [15] explores how to operate multi-UAV networks where UAVs can associate with terrestrial UEs and periodically update their locations with the aim of maximizing the accumulated downlink rate.

Another approach is leveraging beamforming or directional antennas, such as in [16] which assumes that each base station has directional antennas with a two-dimensional radiation pattern, whereas the UEs feature omnidirectional antennas. The authors present a load balancing multi-objective RL (MORL) framework. They provide a method based on meta-RL and develop a general policy that can rapidly adjust to new tradeoffs.

Reference [17] suggests designs for transmit beamformers and receive beamformers for a UAV network with a target, a UAV joint communications and radar (JCR) base station, and numerous user devices. The UAV broadcasts orthogonal frequency division multiplexing (OFDM) signals with transmit and receive beamforming for simultaneous MIMO radar and multi-user (MU)-MIMO communication tasks.

A. AI Models

Machine learning and deep learning, two subcategories of AI, have advanced to the point where they will support B5G and 6G wireless networks In Particular in smart cities [18]. Additionally, a new era of wireless communications with full AI support is anticipated between 2027 and 2030 in order to meet the growing need for wireless connectivity. AI can help improve conventional schedulers and congestion control systems, reduce packet losses in vehicle-to-everything networks, and mitigate receiver non-linearity effects , among others [19].

A variety of algorithms and approaches are used in deep learning to provide high-level representations of data. Deep learning's primary objective is to eliminate the need for manual data structure definition through automatic data-driven learning. Its name is an allusion to the fact that any neural network with two or more hidden layers is commonly referred to as a deep neural network. Artificial neural networks (ANN) are the basis for the majority of deep learning models. The ANN is a type of computational nonlinear model inspired by the neural organization of the brain that can be trained to carry out tasks including classification and prediction [20].

Over the past 20 years, the development of AI has been significantly impacted by RL [21]. An agent can periodically make decisions, evaluate the outcomes, and autonomously modify its approach to obtain the best possible policy through its interactions with the environment. The agent first observes its existing condition, then acts, reaping both its immediate reward and its new state. The agent's policy is adjusted based on the observed information and this process is repeated until the agent's policy approaches the optimal policy. DRL considerably speeds up the learning, especially for problems with large state and action spaces. As a result, DRL enables network controllers or Internet of Things (IoT) gateways to dynamically govern user association, spectrum access, and transmit power for a sizable number of IoT devices and mobile users in large-scale networks with thousands of devices [7].

III. System Model

The time variant data rate and the signal-to-noise plus interference ratio are defined as

$$R_t = log_2(1 + SINR_t) \tag{1}$$

and

$$SINR_t = \frac{P_{receivedpower,t}}{\sigma^2 + \sum_k I_k}, \tag{2}$$

where σ is the noise power and I_k is the co-channel interference between the k different sub-regions.
For the received power we need to estimate the path loss:

$$\begin{aligned} PL(dB) = 20log(\frac{4\pi f_c d}{c}) + Prob(LoS)\eta_{LoS} + \\ Prob(NLoS)\eta_{NLoS}. \end{aligned} \tag{3}$$

The probability of line of sight (LoS) can be formulated as

$$Prob(LoS) = \frac{1}{1 + \mu_{1,2} exp(-\psi_{1,2}(\frac{180}{\pi})\theta_{i,j} - \mu_{1,2})}, \tag{4}$$

where $\mu_{1,2}$ and $\psi_{1,2}$ are the environment factors for the uplink and downlink channels and

$$\theta = arctan(\frac{h}{l}) \tag{5}$$

is the angle with h being the UAV altitude and l the horizontal distance to the O-CU/O-DU. Moreover, η_{LoS} and η_{NLoS} are additional losses due to the free space propagation [22].

In the uplink from the UE to the UAV, the UAV receiver causes nonlinear signal distortion to the received signal, then retransmit it to the O-DU. The O-DU applies the model defined in [23] to eliminate the interference and phase shift in the received signal. The MIMO channel model between the UE and the UAV and between the UAV and the O-DU is obtained from [24]. Consequently, the signal at the UAV O-RU receiver and at the fixed node O-DU receiver can be modeled as

$$\begin{aligned} Y_{UAV} &= XH_{UL} + \frac{3}{2}\beta|X|^3, \\ Y_{DU} &= Y_{UAV}H_{FH} + \eta_{noise}. \end{aligned} \tag{6}$$

The second term in Y_{UAV} is the receiver nonlinearity applied to the received signal at the UAV [25]. Parameters H_{UL} and H_{FH} are the MIMO uplink and fronthaul (UAV to DU) channels, X is the transmitted signal from the UEs, β is the third order gain of the third order intermodulation product, and η_{noise} is the additive white Gaussian noise of zero mean and σ_{noise} variance.

IV. Problem Formulation

In this paper we assume that we have I users and J O-DUs with one UAV-based O-RU. In order to maximize the network throughput we need to maximize the following equation subject to certain criteria:

$$\begin{aligned} min\sum_{t} max \sum_{i,j} & R_{i,t} + V_{j,t}R_{j,t} \\ & s.t. \\ & h_{min,t} < h_t < h_{max,t} \\ & x_{min,t} < x_t < x_{max,t} \\ & y_{min,t} < y_t < y_{max,t} \\ & \sum_{j,t} V_{j,t}R_{j,t} > \sum_{i,t} R_{i,t} \\ & \sum_{j,t} V_{j,t} = 1; V_{j,t} \in \{0,1\}. \end{aligned} \tag{7}$$

$R_{i,t}$ is the data rate between the ith user and the UAV and $R_{j,t}$ is the data rate between the UAV and the O-DU at time t. Parameters h_t, x_t, and y_t capture the 3D location of the UAV, V_j determines whether the fronthaul connection between the UAV-based O-RU and the jth O-DU is active. The capacity of the fronthaul link should be larger than that of an access link. This assumption stems from the fact that several nodes might send data to the UAV. We do not consider the implications of the specific lower-layer functional split in this paper.

V. Proposed Approach

We examine the UAV trajectory under two distinct conditions which are illustrated in Figs. 1 and 2. In the first scenario, the UAV can choose from four possible actions, but passing through a previous state would result in a negative reward. In the second scenario, which is based on a Markov decision process (MDP) with two UAV action options, the UAV must skip passing through some states in order to reach the terminal state with the maximum reward. For both cases, the minimum number of steps to get to the terminal states is considered to minimize the UAV power consumption. We investigate the UAV trajectory for each scenario using two algorithms: Q-learning and State-Action-Reward-State-Action (SARSA).

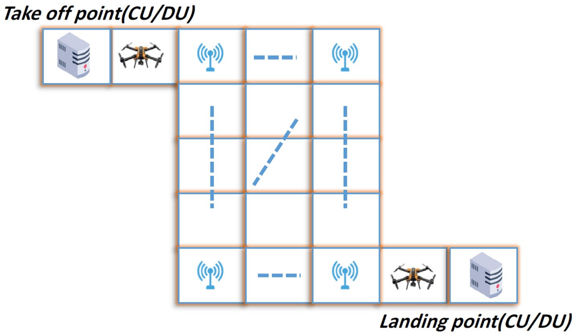

Fig. 1. The first scenario defines the start and end points for the UAV and lets it move around to service users based on SARSA and Q-Learning algorithm.

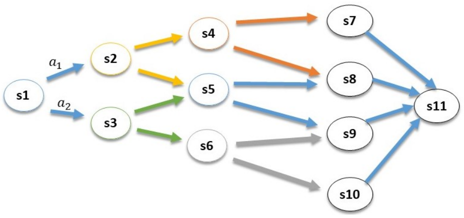

Fig. 2. In the second scenario the UAV movement follows a MDP.

1) Q-Learning: In an MDP, the goal is to identify the best course of action for the agent in order to maximize the anticipated long-term reward function of the system. We first develop a value function $\nu^{\pi} : S \to \mathbb{R}$. It quantifies the goodness of the policy and depicts the predicted value attained by adhering to state-specific legislation in which each state is in the set $s \in S$. With an unlimited horizon and discounted MDP, the value function can be defined as

$$\begin{aligned} \nu^{\pi}(s) &= E_{\pi}\left[\sum_{t=0}^{\infty} \gamma r_t(s_t, a_t)|s_0 = s\right] \\ &= E_{\pi}\left[r_t(s_t, a_t) + \gamma\nu^{\pi}(s_t + 1)|s_0 = s\right], \end{aligned} \tag{8}$$

where $r_t(s_t, a_t)$ is the achieved reward in iteration t for the agent being in state s_t and taking action a_t. Parameter $\gamma \in [0, 1]$ is the discount factor. In comparison to the current reward, the discount factor defines how important future rewards are. The key here is how much attention the UAV should give to potential rewards. Our suggested approach is to determine this value when, given the first observation of the environment, the assessment of the discounted long-term reward at the start of the episode is close to the average reward, which is defined as

$$\frac{|Q_0 - AR|}{|Q_0 + AR|} \leq \Delta, \tag{9}$$

where Δ is the predefined threshold to select the best discount factor.

The best action in each state can be discovered by using the best value function

$$\nu^{*}(s) = max_{a_t}\Big\{E_{\pi}\left[r_t(s_t, a_t) + \gamma\nu^{\pi}(s_{t+1})\right]\Big\}. \tag{10}$$

By defining

$$Q^*(s,a) \triangleq r_t(s_t,a_t) + \gamma E_\pi\left[\nu^\pi(s_t+1)\right], \quad (11)$$

if $\nu^* = max_a Q^*(s,a)$ is the best Q-function for all state-action pairs, then ν^* is the best value function. The objective is thus to determine the best Q-function $Q^*(s,a)$ values for all state-action pairs. This can be accomplished through an iterative procedure:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha_t\Big[r_t(s,a)+ \gamma max_{a'} Q_t(s,a') - Q_t(s,a)\Big]. \quad (12)$$

The learning rate $\alpha_t \in [0,1]$ is used in (12) to assess how new knowledge will affect the current Q value. The learning rate can be set to be constant or it can change dynamically as the learning advances. The Q-learning algorithm operates offline and can determine the agent's best course of action without knowing anything about the surrounding environment.

2) SARSA: The SARSA algorithm, as opposed to the Q-learning, is an online method that enables the agent to select the best course of action at each time step in real time without having to wait for the algorithm to converge. In fact, in contrast to Eq.12 which chooses the maximum value for different actions, in SARSA chooses the action based on the current policy. However, For the Q-learning process, the policy is changed using an off-policy strategy that considers the behaviors with the highest potential reward. As an on-policy method, the SARSA algorithm engages with the environment and modifies the policy immediately as a result of the taken actions.

3) MDP: The MDP offers a mathematical framework for simulating problem-solving situations where decisions are made by agents or decision makers and where results are somewhat influenced by randomness. Studying optimization issues that can be addressed using dynamic programming and RL methods is made possible by MDPs. An MDP cam be described by the tuple (S, A, P, R). S is the finite set of predefined states, A is the finite set of predefined actions the agent can choose in each state, P is the transition probability matrix for state transitions resulting from actions, and R is the set of immediate rewards. The mapping from a state to an action is determined by policy π. Finding the best policy to maximize the reward function is the purpose of an MDP. Its time horizon might be finite or infinite. An ideal strategy π^* that maximizes the anticipated total reward is defined for the MDP with a finite time horizon as

$$max_\pi E\left[\sum_{t=0}^{T} r_t(s_t, \pi(s_t))\right], \quad (13)$$

where $a_t = \pi(s_t)$. The goal for the infinite time horizon MDP can either be to maximize the average reward or the expected discounted total reward. The former is formulated as

$$max_\pi E\left[\sum_{t=0}^{T} \gamma r_t(s_t, \pi(s_t))\right] \quad (14)$$

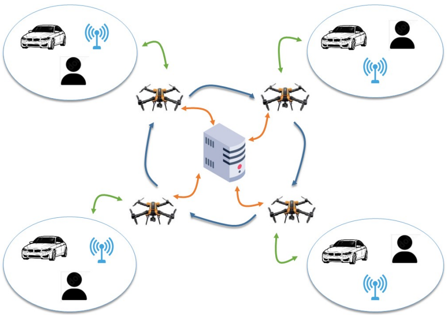

Fig. 3. One UAV is moving in a predefined area to provide network access to UEs.

and the latter as

$$\lim inf_{T\to\infty} max_\pi E\left[\sum_{t=0}^{T} r_t(s_t, \pi(s_t))\right]. \quad (15)$$

In this research we are analyzing the UAV trajectory from point A, which is our takeoff point, to point B, which is the landing point according to Fig. 1. While flying over the different subregions in which the UAV is supposed to transmit the received data from the UEs to the network as illustrated in Fig. 3. In each subregion a set of IoT nodes transmit data to the UAV at the same time, causing interference at the UAV receiver.

We choose two reward functions,

$$R_{PL} = \frac{1}{1+e^{-x}} \quad (16)$$

and

$$R_{PL^{-1}} = \frac{1}{1+e^{-\frac{1}{x}}}, \quad (17)$$

where x corresponds to the path loss defined in the system model. For the first reward function, the UAV will follow the trajectory with the highest path loss since the reward will be higher and for the other, the UAV will follow a trajectory with the lowest path loss, which means being closer to the subregion with higher received power. Trajectories with higher path losses cause lower received power; hence the UAV should get closer to that subregion. Moreover, taking an action would lead to a reward of -1 unless the UAV moves to either a new state or closer to the terminal state. The path loss is applied to a sigmoid function which is bounded to $[0,1]$.

VI. Simulation Results and Analysis

For the simulations, the carrier frequency is $f_c = 6$ GHz and the area in which the UAV can move is $400 \times 400m^2$. Each subregion has an area of $80 \times 80m^2$. Therefore, if we assume that each sub-region center is the center of Cartesian plane, the UAV movement in each region has a uniform distribution between $[-40, 40]$; otherwise, handover happens. The UAV altitude is defined in the range of 100 to 400 m. The path loss is obtained from (3) based on the distance between the UAV and the subregion it is flying over. Table I summarizes the simulation parameters.

Flying at a consistent speed of $V = 20m/s$ rather than

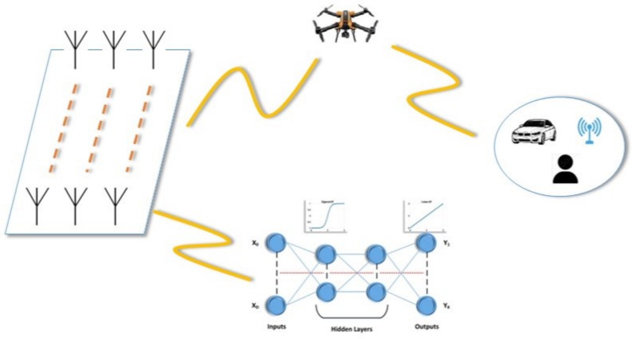

Fig. 4. Access link between UEs and UAV-based O-RU and wireless fronthaul link between the UAV and the nearest fixed infrastructure.

TABLE I
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
$\mu_{1,2}$	9.6	f_c	6GHz
$\psi_{1,2}$	0.15	α	0.9
η_{LoS}	1dB	γ	0.8
η_{NLoS}	20dB	x	[0..400]
Δ	0.1	y	[0..400]
ϵ_{greedy}	0.3	h	[100..200]
$n_{t_{UE}}$	2	$n_{r_{UAV}}$	2
$n_{t_{UAV}}$	2	$n_{r_{CU}}$	100

hovering produces the most power-efficient operation [26], [27]. Therefore, for the access channel the receiver and for the fronthaul channel the transmitter are moving at the same speed. We assume that there are 2 antennas in each subregion for transmitting the data to the UAV and 4 antennas at the UAV, 2 for receiving and 2 for transmitting. Moreover, a rectangular array with 25×4 antennas is employed at the O-DU site receiver. Indeed, each sub-region is supposed to collect the UEs data which is our IoT nodes and transmit them to the UAV. Further explanation of how ANN in rectangular array is going to help us in demodulation is shown in [23].

Fig. 5 shows the average reward for different methods that the UAV can follow to get to the terminal state. Among Q-Learning, SARSA, and MDP, Q-Learning provides a smooth reward as the time episode increases for learning. It should be noted that this is for the case where PL and R has the direct relation, which means that the UAV goes through the direction with the lower received power. In contrast, the average reward deteriorates if the UAV wants to follow the path of lower path loss. For SARSA the result is a little different. Since SARSA is an online RL and updates itself in each iteration based on the environment and current state, we can get the maximum reward if the UAV has enough time to analyze the environment at the cost of longer flight time and, hence, higher power consumption. Finally, for MDP the average reward has an almost constant shape which is appropriate for an unknown environment and states. Fig. 6 provides insights for finding the best discount factor and how much the UAV should care about the future reward to get to the terminal point. As the figure shows, MDP reaches to this condition ($\Delta \leq 10\%$) sooner at the cost of a lower average reward. SARSA is the slowest but has a higher average reward, in fact this figure shows how optimistic different algorithms are at the beginning of the training and after how many episodes they reach to the true average reward. Fig. 7 shows the final UAV trajectory from the takeoff point to the landing point for SARSA. It illustrates that the UAV chooses the shortest path to get to the terminal state which equals to minimizing the time to get to the terminal state as shown in Eq.7. Further analysis of different modulation and MIMO fading channel is left for future research.

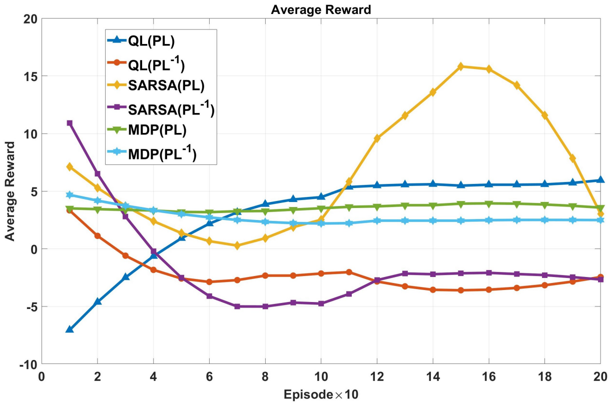

Fig. 5. Average rewards for Q-learning, SARSA and MDP based on (17)

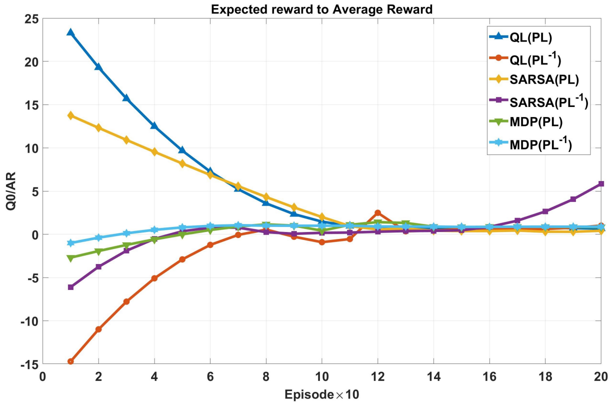

Fig. 6. Expected reward at the beginning of each episode to the average reward

VII. CONCLUSIONS

In this study, we presented an aerial radio unit that connects to the distributed unit through a wireless fronthaul between the UAV and the nearest tower for dynamically serving an underserved area with limited network infrastructure. We solved the UAV trajectory problem employing different RL solutions while the UAV needs to maintain simultaneous links to the UEs it serves and the network in a MIMO fading channel. In order to solve the problem of maximizing the

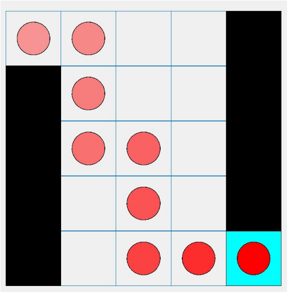

Fig. 7. UAV trajectory from initial point to the terminal state

network throughput based on UAV location, we have proposed two machine learning solutions. Assuming at first that the environment is a grid world, the UAV explores the area by flying from point A to point B using the offline Q-learning or the online SARSA algorithm, with the path loss determining the reward. The trajectory in the second scenario is characterized as a MDP with the goal of maximizing the average reward. Simulation results demonstrate that MDP achieves better results in a more constrained environment and with less effort. In contrast, SARSA requires a longer flight time but performs better in wider areas.

References

[1] B. Shang, V. Marojevic, Y. Yi, A. S. Abdalla, and L. Liu, "Spectrum sharing for uav communications: Spatial spectrum sensing and open issues," *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 104–112, 2020.

[2] A. Colpaert, M. Raes, E. Vinogradov, and S. Pollin, "Drone delivery: Reliable cellular UAV communication using multi-operator diversity," in *ICC 2022-IEEE International Conference on Communications*, pp. 1–6, IEEE, 2022.

[3] A. Colpaert, E. Vinogradov, and S. Pollin, "3D beamforming and handover analysis for UAV networks," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–6, IEEE, 2020.

[4] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference management in OFDMA femtocell networks: Issues and approaches," *IEEE Wireless Communications*, vol. 19, no. 3, pp. 86–95, 2012.

[5] M. Gharib, S. Nandadapu, and F. Afghah, "An exhaustive study of using commercial LTE network for UAV communication in rural areas," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2021.

[6] T. Zhang, K. Zhu, and E. Hossain, "Data-driven machine learning techniques for self-healing in cellular wireless networks: Challenges and solutions," *Intelligent Computing*, vol. 2022, 2022.

[7] N. C. Luong *et al.*, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[8] M. Dai *et al.*, "Joint channel allocation and data delivery for UAV-assisted cooperative transportation communications in post-disaster networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16676–16689, 2022.

[9] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What O-RAN can and cannot do!," *IEEE Network*, vol. 36, no. 6, pp. 206–213, 2022.

[10] M. Kouchaki and V. Marojevic, "Actor-critic network for o-ran resource allocation: xapp design, deployment, and analysis," in *2022 IEEE Globecom Workshops (GC Wkshps)*, pp. 968–973, 2022.

[11] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2109–2121, 2018.

[12] C. Dai, K. Zhu, and E. Hossain, "Multi-agent deep reinforcement learning for joint decoupled user association and trajectory design in full-duplex multi-UAV networks," *IEEE Transactions on Mobile Computing*, 2022.

[13] B. Shang and L. Liu, "Mobile-edge computing in the sky: Energy optimization for air–ground integrated networks," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7443–7456, 2020.

[14] C.-H. Liu, K.-H. Ho, and J.-Y. Wu, "MmWave UAV networks with multi-cell association: Performance limit and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2814–2831, 2019.

[15] H. El Hammouti, M. Benjillali, B. Shihada, and M.-S. Alouini, "A distributed mechanism for joint 3D placement and user association in uav-assisted networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2019.

[16] A. Feriani *et al.*, "Multiobjective load balancing for multiband downlink cellular networks: A meta-reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2614–2629, 2022.

[17] H. Noh, H. Lee, and H. J. Yang, "ICI-robust transceiver design for integration of MIMO-OFDM radar and MU-MIMO communication," *IEEE Transactions on Vehicular Technology*, 2022.

[18] M. Ilbeigi, A. Morteza, and R. Ehsani, "Emergency management in smart cities: Infrastructure-less communication systems," in *Construction Research Congress*, pp. 263–271, 2022.

[19] H. Mohammadi, W. AlQwider, T. F. Rahman, and V. Marojevic, "AI-driven demodulators for nonlinear receivers in shared spectrum with high-power blockers," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 644–649, IEEE, 2022.

[20] H. Mohammadi and V. Marojevic, "Artificial neuronal networks for empowering radio transceivers: Opportunities and challenges," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–5, IEEE, 2021.

[21] M. T. Ramezanlou, V. Azimirad, S. V. Sotubadi, and F. Janabi-Sharifi, "Spiking neural controller for autonomous robot navigation in dynamic environments," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 544–548, 2020.

[22] R. Ghanavi, E. Kalantari, M. Sabbaghian, H. Yanikomeroglu, and A. Yongacoglu, "Efficient 3D aerial base station placement considering users mobility by reinforcement learning," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2018.

[23] H. Mohammadi, M. Sabbaghian, and V. Marojevic, "Self interference management in in-band full-duplex systems," *arXiv preprint arXiv:2202.00764*, 2022.

[24] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.

[25] J. Dsouza, H. Mohammadi, A. V. Padaki, V. Marojevic, and J. H. Reed, "Symbol error rate with receiver nonlinearity," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–5, IEEE, 2020.

[26] M. Bliss and N. Michelusi, "Power-constrained trajectory optimization for wireless UAV relays with random requests," in *2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.

[27] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.