A Distributed 3D UAV Placement Algorithm for Integrated Ground-Air Cellular Networks

Xinwei Wang, Ye Liu, Shan Chang*
Computer Science and Technology, Donghua University, Shanghai, China
Emails: 2191959@mail.dhu.edu.cn, 1199131@mail.dhu.edu.cn, changshan@dhu.edu.cn

2022 8th International Conference on Big Data Computing and Communications (BigCom) | 978-1-6654-7384-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIGCOM57025.2022.00058

***Abstract*—The problem of deploying unmanned aerial vehicles (UAVs) to form a wireless cellular network has gained much attention. Most of the recent works either need to know the locations of users or to make decision in a centralized way. In this paper, we design a distributed learning-based 3D UAV placement approach, named SDQ-H. Given a set of UAVs, SDQ-H builds an integrated ground-air cellular network to maximize the number of ground users (GUs) covered, without requiring knowledge of geographic locations of GUs. SDQ-H decouples the 3D UAV placement problem into two sub-optimal optimization problems. First, we utilize a Q-learning algorithm to optimize UAV positions in 2D space, i.e., in longitude and latitude. UAVs are modeled as agents, each of which takes action independently according to its position and reward. Then, we finetune the position of UAVs in the third dimension, i.e., height. The above two optimization processes are carried out alternately until convergence. We conduct simulations on several different user distributions. Simulation results demonstrate the superiority of SDQ-H.**

***Index Terms*—Unmanned aerial vehicles, Reinforcement learning, 3D UAV deployment.**

I. Introduction

In recent years, the rapid development of wireless technology has provided new solutions for solving some specific problems in communication scenarios. In order to satisfy the growing number of requirements of users, UAVs equipped with base stations have gained attention in the field of wireless communication systems due to the characteristics of easy deployment, mobility, high altitude, and less cost. As air base stations, UAVs can be deployed in relatively high positions, providing ground users with a communication link with a higher probability of line-of-sight (LoS) [1], moreover, they can improve wireless network performance in scenarios such as natural disasters, field surveys, and traffic congestion. When the terrestrial network is damaged by an accident, UAVs can quickly set up a temporary reliable communication network for users on the ground. Unlike fixed ground base stations, which are expensive to deploy and cannot move, UAV can help the ground cellular network to extend the coverage to remote areas, and release the burden of overloaded ground base station traffic in large-scale activity scenarios.

The deployment of UAV base stations has become the most important issue in the construction of aerial networks, and many researchers are studying the optimization problem of deploying UAVs base stations in cellular networks [2–5]. These works are mainly focused on the energy efficiency [5, 11], time delays [19, 23], network throughput [14, 15, 22, 24, 25] and so on, however, there are some shortages in most of existing works. At first, in actual deploying scenarios such as remote areas and destroyed terrestrial network caused by disasters, the information of ground users are hard to acquire. Users' location could be set as global knowledge before deploying [10, 12, 23], or synthesized to follow a cluster [17] and homogeneous Poisson point process (HPPP) [18] distribution. Unfortunately, there works are not possessed of authenticity without realistic scenarios. Then, some work [9, 19] deploying UAVs with centralized method, which cost a lot on computation. E. Arribas et al. [35] design a heuristic algorithm for UAVs 3D deployment in polynomial time, but they ignore the cooperative among UAVs. Both Liu and Chen [11, 27] propose distributed method based on reinforcement learning, but they only consider target areas without the concreate link between UAVs and ground users. Moreover, in the scenario of UAVs deployment without gBSs, it is feasible that UAVs deploy as relays, in order to assist ground users to transfer information and access the internet by gBSs. Nevertheless, A. Shamsoshoara et al. think of a UAV-assisted network model which adopt autonomous method [36] , in other words, each UAV establishes only one direct connection with, which restrict the flexibility of UAVs, because UAVs fly above their own autonomous area without cooperative among other UAVs and global information. In our system, we deploy a set of UAVs as relays to help users access the network, which share local information when connected with each other when flying. A fully connected network will be constructed after deployment, which means that every user can connect to the gBS directly or through UAVs.

In this paper, we aim to navigate a set of UAVs to find optimal deployment locations to form a fully connected network with gBSs in unfamiliar environment, where as many GUs as possible could access the cellular gBSs in singlehop or multi-hops. In our first 2D optimization sub-problem, due to the partial information of each UAVs, we model each UAV as an agent, which collects GUs' locations and shares with other UAVs when connecting with each other. We finetune the height given 2D sub-optimal positions of UAVs and iteratively find the final 3D deployment positions. We conduct our experiments on simulated and realistic scenarios, the results shows that UAVs final positions could serve more of GUs than initial positions, and guarantee the fully-connected network at the

* is the corresponding author.

DOI 10.1109/BigCom57025.2022.00058

TABLE I: DIFFERENCE BETWEEN SDQ-H AND EXISTING WORKS

Condition \ Reference	[10]	[17]	[18]	[11]	[20]	[27]	[35]	[36]	Ours
Unknown users' distribution					✓				✓
Distributed deployment			✓	✓		✓		✓	✓
With fixed gBSs	✓	✓	✓				✓		✓
Cooperative									✓

same time. In some distributions, SDQ-H even achieves full coverage. Compared with OnDrone [35] method, SDQ-H w/o IEM, random deployment with and without fully-connected, SDQ-H shows a better performance than them.

The rest of the paper is organized as follows. We introduce the related research in Section II , In Section III, and IV we elaborate on the system model and SDQ-H details, respect-tively. Simulation results and analysis is discussed in Section V. Finally, this article is summarized in Section VI.

II. Related Work

For different purposes, researchers have investigated the use of mobile aerial base stations to improve cellular network performance through various methods. An improved genetic algorithm [12] that can maximize the number of users covered according to different QoS requirements. Lai et al. [13] propose a density-aware deployment algorithm that maximizes the number of users covered under the premise of meeting the minimum data rate requirements of each user. In addition to optimizing the number of users on the ground, the throughput [9, 15] of the network is also an extensively studied content. E. Arribas et al. [14] optimize the network throughput with thought of the fairness of resource allocation. H. El Hammouti et al. [18] design the LAYF algorithm, considering the allocation of bandwidth resources and the QoS of users. However, most of them ignore the connectivity or three-dimensional deployment.

In the emergency scenario, the deployment delay and path planning of the UAV base station are also a concern, X. Zhang et al. [19] propose a low-complexity optimization algorithm to decrease the delay time of the multi-UAV base station and solve the min-max and min-sum problems in the specific target coverage area. J. Liu et al. [20] propose a learning-based algorithm on geographic location information, which enables the UAV base station to cope with different user distribution scenarios, they reduce computing costs, and maximize the data rate in the network downlink. In order to maximize the total data rate of the uplink with limit time, F. Jiang et al. [22] discuss a centralized algorithm based on linear search to optimize the trajectory of the drone.

The learning-based approach can also solve the deployment optimization problem, however, in many realistic scenarios, the true distribution and geographical location of the user are unknown, which brings great challenges to the effective deployment of the UAV base station. Some work [24, 25] use the Q-learning method to optimize the two-dimensional position of the UAVs to maximize the user's data rate, C. H. Liu et al.

TABLE II: KEY NOTATIONS USERED IN THIS PAPER

Notation	Definition
$P_{LOS}^{i,k}$	The probability of LoS
$g_{i,k}$	Average path loss between BS i and GU k
$\Gamma_{i,k}$	SINR
σ	Power of Gaussian white noise
$p_{tx}^{i,k}$	The transmit power of gBSs/UAVs
U, G, A	The set of GUs, gBSs, UAVs
Π_{μ},Π^{g},Π^{a}	The positions of GUs, gBSs, UAVs
c^{B}	BS-BS link matrix
b^{G}	gBS-GU link matrix
b^{A}	UAVs-GU link matrix
f_0	gBSs/UAVs carrier frequency
R_c	Communication distance between BSs

[11] propose a deep reinforcement learning(DRL) method to achieve fully distributed autonomous deployment of multiple drones at a given height in advance, in order to maximize the average coverage score and fairness of the coverage area under the condition of meeting limited power consumption. Abeywickrama et al. [26] use DRL to design the UAV's path to improve the coverage performance of the wireless network and reduce interference and collisions between the drones. D. Chen [27] combines reinforcement learning with mean field game theory (MFG) to maximize communication efficiency while ensuring network connectivity, but for certain coverage areas, without taking into account communication connections with real users. All of the above work is to discuss the two-dimensional deployment of UAV base stations, some authors [28–31] mainly optimize the 3D deployment location of UAVs.But these works ignore the backhaul links of UAVs and the robustness of cellular networks jointly, Table I shows the difference between exsisting works and our system model as summary.

In the context of UAV-assisted cellular network connectivity and 3D deployment, this paper uses a learning-based approach designed to enable UAVs to deploy independently and maximize network coverage efficiency.

III. System Model

A. Network Model

The list of notations is demonstrated in Table II. We consider an UAV-assisted terrestrial system, where multi-UAV are served as aerial based stations for ground users (GUs), which are irregularly distributed on the target area, as illustrated in

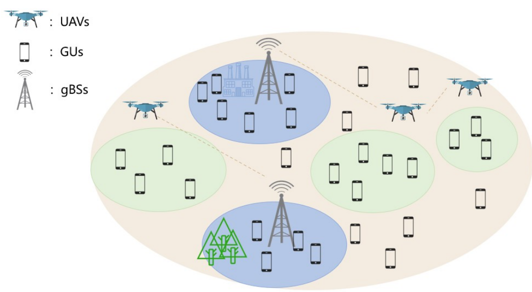


Fig. 1: UAV-assisted cellular network

Fig. 1. Each UAV is equipped with a mobile relay in order to transfer information from GUs to other base stations. We consider that UAVs communicate with a ground base station (gBS) in a single-hop or multi-hop manner to form a full-connected network. Except GUs covered by the gBSs, UAVs need to provide wireless access services for uncovered GUs on an orthogonal downlink bandwidth with respect to the gBSs.

In our 3D deployment problem, we constraint that a GU can only be served by one base station (UAV or gBS) for reducing redundant. Furthermore, without the knowledgement of GUs' distribution, all UAVs can share the information of covered GUs and local state.

B. Channel Model

UAV's height can affect the signal to interference plus noise ratio (SINR) which GU receives. In our scenario, there are two main channel models involved, i.e., the base-to-base (B2B) channel model (including UAV-to-UAV, gBS-to-UAV channels) and the base-to-ground (B2G) channel model (including UAV-to-GU, gBS-to-GU channels). Considering the spatiality of the UAVs, the communication link of B2B is mainly depended by high probability of Line-of-Sight (LoS) [32].

The B2G channel can be modeled as a probability of LoS link and non-LoS link[33], which depends on the height H of a UAV and the 2D projection distance $r_{i,k}$ between UAV i and GU k. The probability of LoS can be expressed as:

$$P_{LOS}^{i,k}(H, r_{i,k}) = \frac{1}{1+\xi . e^{-\varpi \frac{180}{\pi} tan^{-1} H/r_{i,k} - \xi}}, \tag{1}$$

where ξ ,ϖ is related to the environment of the target coverage area, and the probability of non-LoS can be expressed as $P_{LOS}^{i,k} = 1 - P_{LOS}^{i,k}$[6].

The SINR received by GU k from BS i can measure the quality of service (QoS) of the B2G channel. Given the transmit power $p_{tx}^{i,k}$, we assume that when GU receives a SINR exceeding a certain threshold Γ_{th}, the GU can be served by the base station (BS) and its QoS can be guaranteed. SINR is expressed [34] as:

$$\Gamma_{i,k} = \frac{p_{tx}^{i,k} g_{i,k}}{\sum_{j \neq i} p_{tx}^{j,k} g_{j,k} + \sigma} \geq \Gamma_{th}, \tag{2}$$

where $g_{i,k}$ is average path loss between base station i and GU k, σ is power of Gaussian white noise.

C. Problem Formulation

We consider UAVs are hovering at an alterable altitude while users are static. It is assumed that the target coverage area is flat, let set $\mu \in U$ to represent a GU and its position can be denoted by $\Pi_{\mu} = (x, y, 0)$,GU's distribution is mainly collected from FousSquare, detailed in Section V. The position of a gBS $g \in G$ is denoted by $\Pi^{g} = (X_g, Y_g, H_g)$. The positions of multiple UAVs $a \in A$ are the mainly element of our optimization problem, they can be denoted by $\Pi^{a} = (X_a, Y_a, H_a)$. UAVs are deployed to assisted the fixed gBS to jointly cover as many GUs as possible. $b \in B = G \cup A$ represents a BS (including UAV, gBS).We use matrix $c^B = \left[c_{i,j}^b\right]_{B \times B}$ to represent the single-hop connection links between BS i and BS j. The fixed gBSs are fully connected by the fiber, so we ignore the connectivity between gBSs. Wherein, if $c_{i,j}^b = 1$, it means that BS i has a single-hop connection with BS j otherwise,$c_{i,j}^b = 0$. Note that we can determine the n-hop connection between BS i and BS j by calculating $((c^B)^n)_{i,j}$,if $((c^B)^n)_{i,j} \neq 0$, it means there exists a n-hop connection links between BS i and BS j. We use matrix $b^G = \left[b_{i,k}^g\right]_{g \times \mu}$, $b^A = \left[b_{j,k}^a\right]_{A \times U}$ to represent the connection between gBS-GU, and UAVs-GU, respectively.If $b_{i,k}^g = 1\,(i.e., \Gamma_{i,k} \geq \tau)$,it means that GU k is covered by gBS i, likewise, if $b_{j,k}^a = 1\,(i.e., \Gamma_{j,k} \geq \tau)$, it means that GU k is covered by UAV j, represented by $\sum_{k=1}^{U} b_{i,k}^g, \sum_{k=1}^{U} b_{j,k}^a$, respectively. We take the unidirectional network as our mainly constrains, in other words, when any UAV is connected to a gBS by single hop or n-hop links, the network is full-connected. Once the unidirectional network is determined, we can get the positions of UAVs, as well as a set of gBS which UAVs get to. This paper aims to find a three-dimensional deployment positions for UAVs that satisfies the unidirectional network and give access to as many GUs as possible. Thus, our optimization problem can be formulated as follows:

$$maximize_{(\{\Pi^a\}_{a \in A})} \frac{\sum_{k \in U}\left(\sum_{j \in A} b_{j,k}^a + \sum_{i \in G} b_{i,k}^g\right)}{|U|} \tag{3}$$

$$\begin{aligned}
s.t. \quad & \Gamma_{j,k} b_{j,k}^a \geq \Gamma_{th}, \Gamma_{i,k} b_{i,k}^g \geq \Gamma_{th}, \forall k \epsilon U, \forall j \epsilon A, \forall i \epsilon G && \text{(3a)}\\
& R_c c_{i,j}^b \geq d_{i,j}, \forall i,j \in B = A \cup G, i \neq j && \text{(3b)}\\
& \sum_{i=1}^{G} b_{i,k}^g + \sum_{j=1}^{A} b_{j,k}^a \leq 1, \forall k \in U && \text{(3c)}\\
& \exists n < |A|, ((c^B)^n)_{j,i} \neq 0, \forall j \in A, \forall i \in G && \text{(3d)}\\
& x^{min} \leq X_i \leq x^{max}, \forall i \in A && \text{(3e)}\\
& y^{min} \leq Y_i \leq y^{max}, \forall i \in A && \text{(3f)}\\
& H_i \in H, \forall i \in A && \text{(3g)}\\
& c_{i,j}^b, b_{i,k}^g, b_{j,k}^a \in \{0,1\}, \forall i \in G, \forall j \in A, k \in U && \text{(3h)}
\end{aligned}$$

Constraint (3a) means that the SINR is no less than the predefined threshold Γ_{th}. Constraint (3b) guarantees the communication links between BSs. Constraint (3c) ensures that a GU can be served by at most one BS. Constraint (3d) restricts that any UAV to be associated with one gBS by single-hop or multi-hop. Constraint (3e), (3f) and (3g) guarantee that the limited deployment space of UAVs. Constrains (3h) is the boolean value for deciding the single-hop link of BS-to-BS, gBS-to-GU, UAV-to-GU.

The optimization problem in the paper is hard to find the optimal solution. Specially, constraint (3d) means that the final 3D positions of UAVs are not unique. Brute force search costs a lot of time for finding global optimal, classical heuristics algorithm could not solve the problem well, especially in unknown environment. In order to achieve a suboptimal, we decouple the optimization problem into two sub-problems. First, given the height of UAVs, we optimize the 2D positions of them by a learning-based approach. The 2D optimization sub-problem can be expressed as:

$$maximize_{(X_a,Y_a)} \frac{\sum_{k\in U}\left(\sum_{j\in A} b_{j,k}^{a}+\sum_{i\in G} b_{i,k}^{g}\right)}{|U|} \quad (4)$$

$$\text{s.t} \quad \Gamma_{j,k} b_{j,k}^{a} \geq \Gamma_{th}, \Gamma_{i,k} b_{i,k}^{g} \geq \Gamma_{th}, \forall k \epsilon U, \forall j \epsilon A, \forall i \epsilon G \quad (4a)$$

$$R_c c_{i,j}^{b} \geq d_{i,j}, \forall i,j \in B = A \cup G, i \neq j \quad (4b)$$

$$\sum_{i=1}^{G} b_{i,k}^{g} + \sum_{j=1}^{A} b_{j,k}^{a} \leq 1, \forall k \in U \quad (4c)$$

$$\exists n < |A|, ((c^{B})^{n})_{j,i} \neq 0, \forall j \in A, \forall i \in G \quad (4d)$$

$$x^{min} \leq X_i \leq x^{max}, \forall i \in A \quad (4e)$$

$$y^{min} \leq Y_i \leq y^{max}, \forall i \in A \quad (4f)$$

$$c_{i,j}^{b}, b_{i,k}^{g}, b_{j,k}^{a} \in \{0,1\}, \forall i \in G, \forall j \in A, k \in U \quad (4g)$$

Then, based on the 2D positions, in order to get a height vector $\mathcal{H}$, we finetune each height of UAVs and update connection matrix, which is formulated as follows:

$$maximize_{(\mathcal{H})} \frac{\sum_{k\in U}\left(\sum_{j\in A} b_{j,k}^{a}+\sum_{i\in G} b_{i,k}^{g}\right)}{|U|} \quad (5)$$

$$\text{s.t.} \quad H_i \in H, \forall i \epsilon A,$$

IV. UAVs Positioning with SDQ-H

We aim to find the optimal 3D positions for UAVs constrained to the unidirectional connectivity which maximizes the total coverage of GUs. As mentioned above, we adopt a decomposition process to get a stable solution. Before explaining our approach, we firstly introduce the connection and communication strategy in BS-to-GU and BS-to-BS.

A. Initial Algorithm for Connection Stretagy

The target GUs in this paper conclude GUs covered by gBSs and those uncovered by gBSs. Firstly, gBSs can obtain the location information of GUs which are covered. To some degree, GUs are preferred to connect with gBSs than UAVs. Then, a GU will select the nearest UAVs when it is covered by multiple UAVs. UAVs fly above the area to find its' positions step by step, it cannot be guaranteed that each UAV connected with other BSs all the time. To make use of the instantaneous communication link of UAV-to-BSs, we propose an information exchange mechanism (IEM). Since it is impossible to obtain user information that is not covered by gBSs, a UAV i has GUs' information list to record locations of GUs which are covering or covered by UAV i. We define two kinds of communication states, namely, direct communication in Fig. 2.(a), and indirect communication (i.e., indirect communication through UAV or gBS, as shown in Fig. 2.(b) and Fig. 2.(c), respectively). When UAVs fly around, once UAV i and UAV j are in the communicative state, their own GUs' information list will be shared with each other.

Algorithm 1 Q-learning for 2D Deployment of UAVs

Input: Numbers of UAVs, initial 2D positions of UAVs
Output: final 2D position of UAVs, coverage rate

Initialize Q(s, a) arbitrarily, deploy UAVs
repeat
  for each step of episode:
    for each UAV i:
      Record covered GUs' information list
      Choose and execute action $a(i)_t$ according to policy
      if UAV i communicates with UAV j:
        Exchange and update each other's list
      If UAV i flies out of the border:
        Get a punish p_1
        Cancel the action
      If UAV i lose connection with gBS:
        Get a punish p_2
      else:
        Get a reward r_1
      Calculate the first reward $r_i^1(t)$
      Calculate the Second reward $r_i^2(t)$
      If UAV i communicates with any other BS:
        If current coverage is within expected coverage:
          Calculate the Third reward $r_i^3(t)$
      Update GUs' information list
      Update Q-table
until terminal

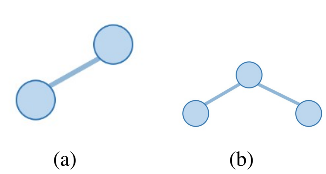


Fig. 2: Communication samples

Algorithm 2 Height Optimization for UAVs

repeat
  Given 2D positions from Algorithm 1:
  repeat
    for each UAV i:
      Choose the best height with other's height fixed
      $h_i^* = argmax_{h_i \in H} C_i \left(h_i, h_{-i}\right)$
      Update C_i
  until All UAVs achieve a NE
until Sum coverage of network is not significantly improved

B. 2D Position Optimization with Q-learning

Assuming that the locations of uncovered GUs are unknown, therefore, a centralized approach cannot solve our problem without global information. Moreover, the 2D deployment problem of UAVs in this paper involves decision-making and unkown environment, so we can use reinforcement learning (RL) methods [39] to model it as the Markov decision-making process(MDP) in order to find 2D positions of UAVs. Q-learning [40] is suitable for scenarios where agent learns a strategy to maximize utility in unknown environment. The process of finding 2D positions is a learning process, where UAVs are modeled as distributed RL agents when making decisions. To simplify the problem, we divide the target coverage area into equal-sized grids. The 2D position of a UAV is deployed on the center point of the grid, which represents the state of the drone at the current time step. UAV selects the appropriate action according to the instant reward. The strategy for choosing an action depends on the Q-table of each agent. The Q-table represents the value of selecting an action in a state at current time step, and an optimal strategy helps UAVs choose actions that can maximize the rewards obtained throughout the whole episode. The calculation of Q-value is:

$$\begin{aligned} &Q_i\left(S_t, A_t\right) \leftarrow Q_i\left(S_t, A_t\right) + \\ &\alpha\left(R\left(i\right)_{t+1} + \gamma \underset{a(i)}{max} Q_i(S(i)_{t+1}, a(i)) - Q_i(S_t, A_t)\right), \end{aligned} \tag{6}$$

The Q-value for our problem consists of the follow elements:

The agents, states representation, action space, and reward function design in the Q-learning model are defined as follows:

1) Agent: UAV $i, i \epsilon M = \{1, 2, ...M\}$.

2) States representation: For each UAV i at time step t, its 2D positions $\left\{x_t^i, y_t^i\right\}$ are modeled as states, which defined as:$S \triangleq \{s_t\} = \{x_i^t, y_i^t\}$

3) Action space: For each UAV i at time step t, its action can be represented as $a_i^t = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, which indicates 8 directions (forward, backward, left, right, left front, right front, back left, back right and hover). If UAV i is flying out of the boundary before performing action, we cancel the action.

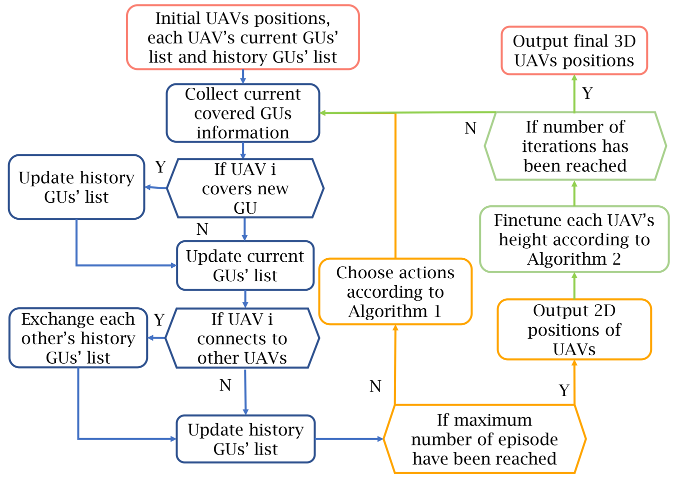


Fig. 3: The flow chart of SDQ-H

4) Reward function design: There are two goals in our optimization problem, the connectivity of UAV-assisted network and maximized coverage. Assuming that UAV cannot obtain GUs' distribution information of the target area, combined with IEM, we design the instant hierarchical rewards for UAVs in terms of both connectivity and coverage. Firstly, a UAV i will get a penalty p_1 when it flies out of the boundary, when it can communicate with any gBS, it will get a reward r_1, otherwise, it will get a penalty p_2, represented as:

$$r_i^1\left(t\right) = p_1 + p_2 + r_1, \tag{7}$$

these three quantities are constants, and the sum of them is considered as the first reward layer. The second layer's reward is:

$$r_i^2\left(t\right) = \Delta G_i\left(t\right), \tag{8}$$

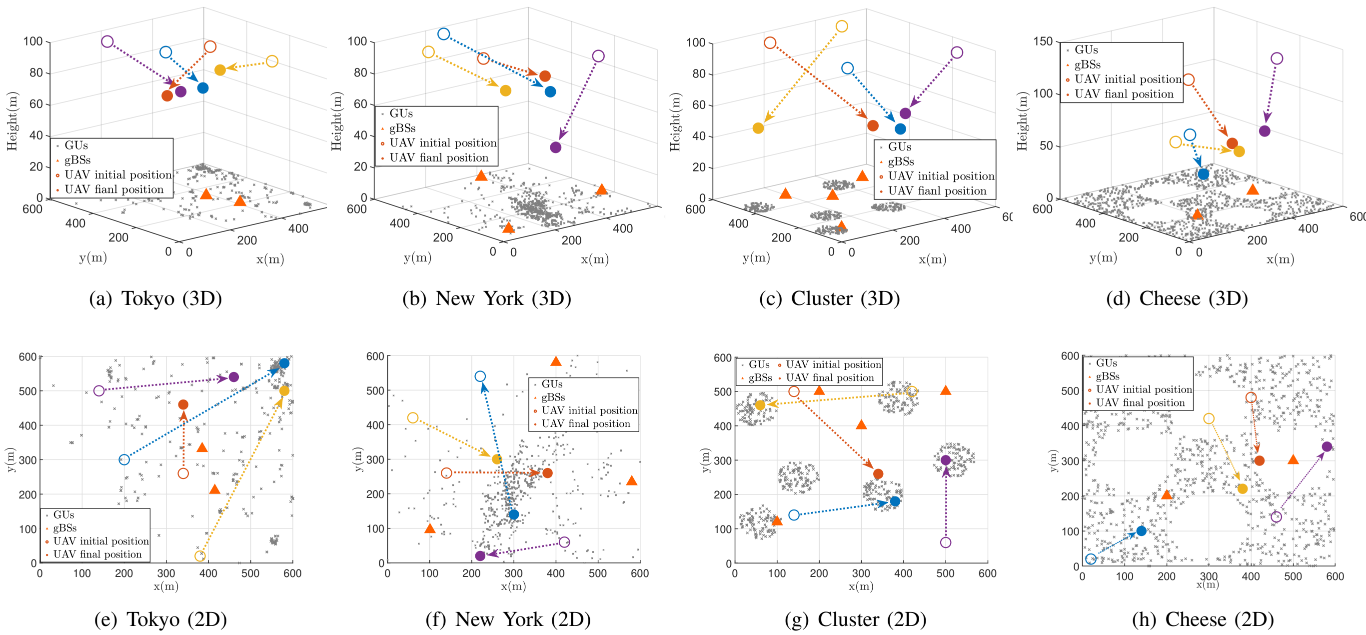
where $\Delta G_i\left(t\right) = G_i\left(t\right) - G_i\left(t-1\right)$ is coverage difference. Finally, when any UAVs can reach gBSs, we introduce additional reward depending on the expected coverage and the current coverage, and the expected coverage rate is taken between [0.88, 0.98]. Only if current coverage is within expected coverage, each UAV will be given an additional reward, represented as:

$$r_i^3\left(t\right) = w_1 \cdot expect_{cover} + w_2, \tag{9}$$

where w_2, w_2 are constants. At time step t, each UAV calculates its own total sum instant rewards and determines whether to exchange information. Pseudocode for 2D-position is presented in Algorithm 1.

C. Height Optimization

In this section, we optimize the height given the connected 2D positions of UAVs. In multi-UAV cooperative optimization issues, we combine Nash equilibrium [41] to finetune each UAV's height h_i. We denote C_i as local coverage of UAV i.



(a) Tokyo (3D) (b) New York (3D) (c) Cluster (3D) (d) Cheese (3D)

(e) Tokyo (2D) (f) New York (2D) (g) Cluster (2D) (h) Cheese (2D)

Fig. 4: The deployment (2D and 3D) results among four distributions.

Therefore, the conditions for Nash equilibrium are expressed as follows:

$$C_i\left(h_i^*, h_{-i}\right) \geq C_i\left(h_i, h_{-i}\right), \forall h_i \neq h_i^*, \tag{10}$$

h_i^* represents optimal height of UAV i, h_{-i} represents the UAVs height vector other than i.

In game theory, when a player has an optimal return on its own strategy while the other players' strategies are fixed, the Nash equilibrium has been reached. As shown in Algorithm 2, we iteratively select the height that can reach the maximum coverage in the current positions for each UAV, and then use Algorithm 1 to update the 2D position, BS-to-BS connection matrix and UAVs-to-GUs connection matrix given the height, until the conditions of Nash equilibrium are reached. The flow chart of SDQ-H is shown as Fig. 3.

V. Experimental Simulations

We analyze the performance of SDQ-H in this section, given the fixed transmit power, and GUs are static on the area. We divide simulation area into 15×15 small grids equally, where each UAV has 225 states to deploy. To demonstrate the feasibility of SDQ-H, we conduct our experiment on simulated and real-world scenarios, where the GUs distribution are synthetic and realistic respectively. On the one hand, in simulated scenarios, the locations of GUs follow clustered and Cheese [35] distribution, we randomly deploy several gBSs within GUs. On the other hand, we select GUs' locations from Tokyo and New York in the FourSquare check-in datasets [36] as the simulation area. 429 GUs' first-check-in GPS locations are selected from Tokyo, where longitude is from 139.4733 to 139.9028 and latitude is from 35.5124 to 35.8638. 652 GU's locations are from New York, where longitude is from -72.2707 to -73.6999 and latitude is from 40.5508 to 40.9876. We also randomly selected several gBSs' locations corresponding 2 cities from Mozilla Location Service [37], an open website that collects ground base station information among world. The environment and simulation parameters are shown in Table III. To evaluate the coverage performance of SDQ-H, we compare a existing method and establish three benchmarking scenarios.

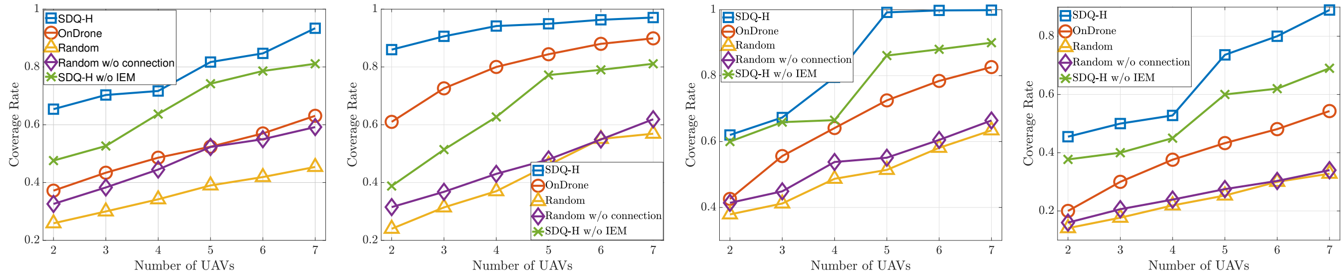
1) OnDrone: We arbitrarily deploy a set of UAVs that satisfies the constraint of unidirectional network, then find UAV i with the least number of coverage and redeploy it (after updating its position, all UAVs' positions still satisfy the unidirectional network). Repeat the steps above until the total coverage no longer changes.

2) Random scenario: In this scenario, we arbitrarily deploy a set of UAVs that satisfies the constraint of unidirectional network.

3) Random w/o connection scenario: In this scenario, we do not care about the positions of gBSs, then arbitrarily deploy a set of UAVs.

4) SDQ-H w/o IEM scenario: In this scenario, we do not consider IEM, and only use the reward function (7),(8).

The initial and final positions of UAVs for SDQ-H are depicted in Fig. 4. The 3D deployment results among four distributions are shown in Fig. 4(a)-(d),and Fig. 4(e)-(h) depict the 2D deployment results among these distributions. Take New York as an example, there are three gBSs deployed on the area in separate manner. We randomly initialize 4 UAVs' 3D positions in three-dimensional space. It is obvious that all UAVs are almost far away from gBSs and the area where GUs' distributions are dense. After several episodes, we determine



(a) Tokyo (b) New York (c) Cluster (d) Cheese

Fig. 5: The coverage rate versus the number of UAVs

the final 2D coordinates that achieve the suboptimal coverage in first sub-problem, as shown in Fig. 4(f). We can find that the distances between UAVs are smaller than before, and UAVs are deployed near the dense area. Correspondingly, each UAV could transfer the information from GUs to any gBS by single-hop or multi-hop. Fig. 4(b) shows 3D deployment results after finetuning the heights, each UAV's height is reduced more or less. It is worth mentioning that the final coverage is improved by comparing with initial coverage, and other GUs distributions also demonstrate the feasibility of SDQ-H.

Fig. 5 plots the average coverage rate of fully connected network versus the number of UAVs under different GUs distributions. Clearly, our proposed approach significantly improves the overall coverage rate as compared with 4 different scenarios.

Fig. 6 represents the jointly coverage rate of gBSs and UAVs at different thresholds. The numbers of UAVs are set as M = 2, 4, 6, we select GUs distribution of Tokyo as simulation area, and map the coordinates of the actual distribution into an area of 600×600 m. Fig. 6 shows that the coverage rate is decreased as the SINR threshold increases. We set all gBSs to the same height, each UAV has different height from 50 to 150 m. Given transmit power, received SINR of a GU has an impact on coverage of the network, when SINR threshold increases, there is a decreasing trend in coverage rate with different number of UAVs.

Fig. 7 shows the performance of SDQ-H according to various grid sizes and the density of grids could impact the position and cover rate of UAVs. We conduct the experiment

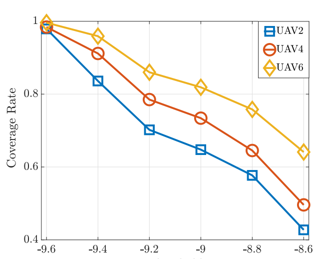


Fig. 6: SINR shreshold

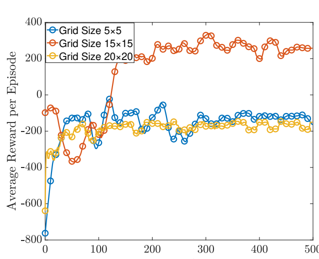


Fig. 7: The average reward per episode

TABLE III: PRIMARY SIMULATION PARAMETERS

Parameter	Value
p_{tx}	15 dBm
f_0	2 GHz
σ	-140 dBm
c	3×10^8 m/s
R_c	165 m
grid size	15×15
Γ_{th}	-9dB
H	[50-150]m
x^{max}, y^{max}	600m, 600m

on distribution of New York with 4 UAVs. As shown in this figure, we can see the similarity of the convergence rate for grid size 5×5 and 20×20, but there is a high converged values in grid size 15×15, therefore, we choose the 15×15 as our final grid size.

VI. Conclusion

In this paper, we have proposed a distributed 3D deployment approach of UAVs based on reinforcement learning and Nash equilibrium. Aiming to find optimal positions that maximize the sum coverage of all BSs and satisfy the fully connected network, we decouple our problems into 2 sub-problem. Firstly, under unknown environments, we design the information exchange mechanism and hierarchical reward function in order to find 2D postions of UAVs with Q-learning. Secondly, we finetune the height of each UAV given the suboptimal 2D positions. The simulation results show that our proposed approach is more effective in deploying UAVs comparing with 4 studied method. In the future work, we will consider the bandwidth limit of UAVs and throughoutput in network.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61972081), and the Natural Science Foundation of Shanghai (Grant No. 22ZR1400200).

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, Y. Nam and M. Debbah, "A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems," in IEEE Communications Surveys Tutorials, vol. 21, no. 3, pp. 2334-2360, thirdquarter 2019.

[2] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. M. Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani et al., "Evolution of non-terrestrial networks from 5G to 6G: A survey," arXiv preprint arXiv:2107.06881, 2021.

[3] A. H. Arani, M. Mahdi Azari, W. Melek and S. Safavi-Naeini, "Learning in the Sky: Towards Efficient 3D Placement of UAVs," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2020, pp. 1-7, doi: 10.1109/PIMRC48278.2020.9217381.

[4] J. Wang, Y. Liu, S. Niu, and H. Song, "Extensive throughput enhancement for 5G enabled UAV swarm networking," IEEE Journal on Miniaturization for Air and Space Systems, pp. 1–1, 2021

[5] M.Alzenad, A.ei-Key, F. Lagum, and H. Yanikomeroglu, "3D placement of an unmanned aerial vehicle base station (UAV-BS) for energy effificient maximal coverage," IEEE Wireless Commun. Lett., May 2017.

[6] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," IEEE Wireless Commun. Lett., vol. 3,no. 6, pp. 569–572, Dec. 2014.

[7] Mozaffari M, Saad W, Bennis M, et al. Drone small cells in the clouds: Design, deployment and performance analysis[C]//2015 IEEE global communications conference (GLOBECOM). IEEE, 2015: 1-6.

[8] Bor-Yaliniz R I, El-Keyi A, Yanikomeroglu H. Efficient 3-D placement of an aerial base station in next generation cellular networks[C]//2016 IEEE international conference on communications (ICC). IEEE, 2016: 1-5.

[9] Kalantari E, Yanikomeroglu H, Yongacoglu A. On the number and 3D placement of drone base stations in wireless cellular networks[C]//2016 IEEE 84th Vehicular Technology Conference (VTC-Fall). IEEE, 2016: 1-6.

[10] Zhao H, Wang H, Wu W, et al. Deployment algorithms for UAV airborne networks toward on-demand coverage[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(9): 2015-2031.

[11] C. H. Liu, X. Ma, X. Gao and J. Tang, "Distributed Energy-Efficient Multi-UAV Navigation for Long-Term Communication Coverage by Deep Reinforcement Learning," in IEEE Transactions on Mobile Computing, vol. 19, no. 6, pp. 1274-1285, 1 June 2020, doi: 10.1109/TMC.2019.2908171.

[12] Chen Y, Li N, Wang C, et al. A 3D placement of unmanned aerial vehicle base station based on multi-population genetic algorithm for maximizing users with different QoS requirements[C]//2018 IEEE 18th International Conference on Communication Technology (ICCT). IEEE, 2018: 967-972.

[13] Lai C C, Chen C T, Wang L C. On-demand density-aware UAV base station 3D placement for arbitrarily distributed users with guaranteed data rates[J]. IEEE Wireless Communications Letters, 2019, 8(3): 913-916.

[14] E. Arribas, V. Mancuso, and V. Cholvi, "Fair cellular throughput optimization with the aid of coordinated drones," in Proc. IEEE Int. Conf. Comput. Commun., pp. 1–6.

[15] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Wireless communication using unmanned aerial vehicles (UAVs): Optimal transport theory for hover time optimization," IEEE Trans. Wireless Commun., vol. 16, no. 12, pp. 8052–8066, Dec. 2017.

[16] S. A. W. Shah, T. Khattab, M. Z. Shakir, and M. O. Hasna, "A distributed approach for networked flflying platform association with small cells in 5G+networks," inProc. IEEE Global Commun. Conf. (GLOBECOM),Singapore, Dec. 2017, pp. 1–7.

[17] E. Kalantari, I. Bor-Yaliniz, A. Yongacoglu, and H. Yanikomeroglu,"User association and bandwidth allocation for terrestrial and aerial base stations with backhaul considerations," inProc. IEEE 28th Annu. Int.Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[18] H. El Hammouti, M. Benjillali, B. Shihada and M. Alouini, "Learn-As-You-Fly: A Distributed Algorithm for Joint 3D Placement and User Association in Multi-UAVs Networks," in IEEE Transactions on Wireless Communications, vol. 18, no. 12, pp. 5831-5844, Dec. 2019, doi: 10.1109/TWC.2019.2939315.

[19] X. Zhang and L. Duan, "Fast Deployment of UAV Networks for Optimal Wireless Coverage," in IEEE Transactions on Mobile Computing, vol. 18, no. 3, pp. 588-601, 1 March 2019, doi: 10.1109/TMC.2018.2840143.

[20] J. Liu, Q. Wang, X. Li and W. Zhang, "A Fast Deployment Strategy for UAV Enabled Network Based on Deep Learning," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2020, pp. 1-6, doi: 10.1109/PIMRC48278.2020.9217205.

[21] Y. Li, G. Feng, M. Ghasemiahmadi and L. Cai, "Power Allocation and 3-D Placement for Floating Relay Supporting Indoor Communications," in IEEE Transactions on Mobile Computing, vol. 18, no. 3, pp. 618-631, 1 March 2019, doi: 10.1109/TMC.2018.2840989.

[22] F. Jiang and A. L. Swindlehurst, "Optimization of UAV heading forthe ground-to-air uplink,"IEEE J. Sel. Areas Commun., vol. 30, no. 5, pp. 993–1005, Jun. 2012.

[23] O. Esrafifilian, R. Gangula, and D. Gesbert, "Learning to communicate in UAV-aided wireless networks: Map-based approaches,"IEEE Internet Things J., vol. 6, no. 2, pp. 1791–1802, Apr. 2019.

[24] H. Bayerlein, P. De Kerret, and D. Gesbert, "Trajectory optimization for autonomous flying base station via reinforcement learning," in IEEE 19th International

Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2018, pp. 1–5.

[25] B. Khamidehi and E. S. Sousa, “Reinforcement learning-based trajectory design for the aerial base stations,” in IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2019, pp. 1–6.

[26] H. V. Abeywickrama, Y. He, E. Dutkiewicz, B. A. Jayawickrama and M. Mueck, ”A Reinforcement Learning Approach for Fair User Coverage Using UAV Mounted Base Stations Under Energy Constraints,” in IEEE Open Journal of Vehicular Technology, vol. 1, pp. 67-81, 2020, doi: 10.1109/OJVT.2020.2971594.

[27] D. Chen, Q. Qi, Z. Zhuang, J. Wang, J. Liao and Z. Han, ”Mean Field Deep Reinforcement Learning for Fair and Efficient UAV Control,” in IEEE Internet of Things Journal, vol. 8, no. 2, pp. 813-828, 15 Jan.15, 2021, doi: 10.1109/JIOT.2020.3008299.

[28] L. Zhang, A. Celik, S. Dang and B. Shihada, ”Energy-Efficient Trajectory Optimization for UAV-Assisted IoT Networks,” in IEEE Transactions on Mobile Computing, doi: 10.1109/TMC.2021.3075083.

[29] Y. Nie, J. Zhao, J. Liu, J. Jiang and R. Ding, ”Energy-efficient UAV trajectory design for backscatter communication: A deep reinforcement learning approach,” in China Communications, vol. 17, no. 10, pp. 129-141, Oct. 2020, doi: 10.23919/JCC.2020.10.009.

[30] U. Challita, W. Saad and C. Bettstetter, ”Interference Management for Cellular-Connected UAVs: A Deep Reinforcement Learning Approach,” in IEEE Transactions on Wireless Communications, vol. 18, no. 4, pp. 2125-2140, April 2019, doi: 10.1109/TWC.2019.2900035.

[31] X. Li, Q. Wang, J. Liu and W. Zhang, ”3D Deployment with Machine Learning and System Performance Analysis of UAV-Enabled Networks,” 2020 IEEE/CIC International Conference on Communications in China (ICCC), 2020, pp. 554-559, doi: 10.1109/ICCC49849.2020.9238795.

[32] Y. Zeng, R. Zhang and T. J. Lim, ”Wireless communications with unmanned aerial vehicles: opportunities and challenges,” in IEEE Communications Magazine, vol. 54, no. 5, pp. 36-42, May 2016, doi: 10.1109/MCOM.2016.7470933.

[33] A. Al-Hourani, S. Kandeepan and A. Jamalipour, ”Modeling air-to-ground path loss for low altitude platforms in urban environments,” 2014 IEEE Global Communications Conference, 2014, pp. 2898-2904, doi: 10.1109/GLOCOM.2014.7037248.

[34] L. Zhou, Z. Sheng, L. Wei, X. Hu, H. Zhao, J. Wei, and V. C. Leung, “Green cell planning and deployment for small cell networks in smart cities,” Ad Hoc Networks, vol. 43, pp. 30–42, 2016.

[35] E. Arribas, V. Mancuso and V. Cholvi, ”Coverage Optimization with a Dynamic Network of Drone Relays,” in IEEE Transactions on Mobile Computing, vol. 19, no. 10, pp. 2278-2298, 1 Oct. 2020, doi: 10.1109/TMC.2019.2927335.

[36] A. Shamsoshoara, F. Afghah, A. Razi, S. Mousavi, J. Ashdown and K. Turk, ”An Autonomous Spectrum Management Scheme for Unmanned Aerial Vehicle Networks in Disaster Relief Operations,” in IEEE Access, vol. 8, pp. 58064-58079, 2020, doi: 10.1109/ACCESS.2020.2982932.

[37] https://www.kaggle.com/chetanism/foursquare-nyc-and-tokyo-checkin-dataset.

[38] https://location.services.mozilla.com/downloads.

[39] R. S. Sutton and A. G. Barto, ”Reinforcement Learning: An Introduction,” in IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 1054-1054, Sept. 1998, doi: 10.1109/TNN.1998.712192.

[40] B. Jang, M. Kim, G. Harerimana and J. W. Kim, ”Q-Learning Algorithms: A Comprehensive Classification and Applications,” in IEEE Access, vol. 7, pp. 133653-133667, 2019, doi: 10.1109/ACCESS.2019.2941229.

[41] E. Maskin, “Nash equilibrium and welfare optimality,” Rev. Econ. Stud., vol. 66, no. 1, pp. 23–38, 1999.