Three-dimensional aerial base station location for sudden traffic with deep reinforcement learning in 5G mmWave networks

Peng Yu1,2, Jianli Guo1, Yonghua Huo1, Xiujuan Shi1, Jiahui Wu2 and Yahui Ding2

Abstract
Data volume demand has increased dramatically due to huge user device increase along with the development of cellular networks. And macrocell in 5G networks may encounter sudden traffic due to dense users caused by sports or celebration activities. To resolve such temporal hotspot, additional network access point has become a new solution for it, and unmanned aerial vehicle equipped with base stations is taken as an effective solution for coverage and capacity improvement. How to plan the best three-dimensional location of the aerial base station according to the users’ business needs and service scenarios is a key issue to be solved. In this article, first, aiming at maximizing the spectral efficiency and considering the effects of line-of-sight and non-line-of-sight path loss for 5G mmWave networks, a mathematical optimization model for the location planning of the aerial base station is proposed. For this model, the model definition and training process of deep Q-learning are constructed, and through the large-scale pre-learning experience of different user layouts in the training process to gain experience, finally improve the timeliness of the training process. Through the simulation results, it points out that the optimization model can achieve more than 90% of the theoretical maximum spectral efficiency with acceptable service quality.

Keywords
Sudden traffic, 5G mmWave networks, aerial base station, location planning, deep Q-learning

Date received: 20 December 2019; accepted: 30 March 2020
Handling Editor: Bo Rong

Introduction
Along with varieties of services and the Internet-of-Things (IoT) devices data communication requirements for different scenarios in 5G networks, traffic generations take on drastic spatial and temporal variations, which have caused tremendous pressure on the basic macro base stations (BSs). 5G networks integrate different network architectures, such as cloud radio access networks, ultra-dense networks with heterogeneous cells to improve network density and coverage, thus achieving higher transmission rate and network capacity. However, in some hotspots, such as sports or celebration activities will cause periodic bursts of traffic on networks, and the deployed BSs may be not to
accommodate all the sudden traffic. And it is difficult to reconstruct existing communication infrastructure immediately to resolve this problem. So additional capacity enhancement schemes are required. According to previous literature, unmanned aerial vehicle (UAV) equipped with BSs, called as aerial base station (aerial-BS), is an effective complementary method.1

The mmWave, also called mmW, is the frequency band used in various high-speed applications. The mmWave includes wireless frequencies of 30–300 GHz and the wavelength range of the radio wave in this frequency band is from 1 to 10 mm, so it is called millimeter wave. This frequency has obvious advantages, because it supports higher bandwidth, so it is very suitable for wireless infrastructure applications. Millimeter wave combines height-oriented antenna with beamforming and beam tracking to provide a very safe and reliable link. In addition to the advantages of large bandwidth and high-speed rate, millimeter wave has narrow beam, good directivity, and high spatial resolution, which improves the transmission efficiency, so mmWave transmission suitable for direct communications can be a feasible method here. When aerial-BSs can be used as air access points or disconnect relays and enhanced connections between networks, using aerial-BSs as air support for existing cellular networks can handle sudden traffic situation more economically and enhance network capacity better.

Aerial-BSs for emergency communication or hotspot have attracted much attentions from academic and industry field recently. As aerial-BSs can provide rapid deployment and cost reduction, making communication failure or burst traffic to be absorbed soon.2 And a few literatures have explored for aerial-BSs compensation for network damage or performance degradation due to abnormal traffic.3,4 However, these methods consider less about multiple BSs and the placement algorithms are complex, making them hard for dynamic variations in 5G mmWave networks.

In this article, for optimizing the three-dimensional (3D) deployment of aerial-BSs for 5G mmWave networks, a classic deep reinforcement learning (DRL) network which named deep Q-network (DQN) algorithm is adopted. Compared with traditional algorithm, it can solve high complexity and big state space and action space, so we choose DQN algorithm to solve our problem. DQN uses Q-network to fit Q-table, which solves the problem of dimension disaster well. First, the model is learned and saved with maximal spectrum efficiency, and then the optimal 3D deployment location is quickly found by applying the model in a simulation scenario.

With above analysis, the main contributions are shown as follows:

1. Modeling the aerial-BSs’ location as a maximal spectral efficiency (SE) problem with quality of service (QoS) constraints considering line-of-sight (LOS)/non-line-of-sight (NLOS) path loss under mmWave networks. Moreover, a simple decompose mechanism to reduce its complexity is proposed as well.

2. Applying DQN algorithm to the optimization of 3D deployment of aerial-BSs and sudden traffic accommodation. And the 3D location is used here for different aerial-BS action space. Considering the dynamic changes of the network topology environment, the users’ distribution is trained as part of the state, so that the model can be well adapted to different user distribution.

The remaining content is organized as follows: System model is analyzed in section “System model.” And DQN procedures for aerial-BSs’ 3D deployment are shown in section “DQN-based aerial-BS location optimization framework.” And simulations are taken in section “Simulation results.” Finally, conclusions and recommendations are given in section “Conclusion and future work.”

Related work

In recent years, experts pay more attentions to the usage of aerial-BSs in cellular networks as they can achieve rapid deployment solutions to meet the needs of wireless networks. In Yu et al.,1 a scheme of introducing air BS to enhance the capacity of the data traffic burst area is proposed. However, the location deployment of the aerial-BSs has become one of the key challenges. Unlike the traditional fixed-position BS, the location deployment of the aerial-BSs is flexible and can move flexibly in the air, and ultimately determine by its height and angle. Therefore, the aerial-BSs’ location planning is a 3D deployment problem. In Gomez et al.,2 a heuristic algorithm is proposed to serve multiple users using the least number of aerial-BSs, and to obtain 3D location of aerial-BSs. However, in the dynamic environment where the network topology changes, it is necessary to re-initialize the heuristic algorithm and with new topology. And this process will bring many computational complexities ultimately. In Deaton,3 the vertical and horizontal dimensions of the aerial-BSs are separated, and a deployment scheme of the aerial-BSs with the minimum transmit power serving the maximum number of users is proposed. In Zong,4 aiming at maximizing the total logarithm of users, an algorithm to deploy the aerial-BSs in 3D location is proposed, besides considering the user BS association and wireless return bandwidth allocation. In the previous literature, user mobility is not considered. A 3D deployment algorithm of aerial-BSs based on
Q-learning is proposed in the research work, which considers user mobility. But when the dimension of state space increases, the Q-table will occupy a lot of memory and bring a lot of time overhead.

Still, much research has designed algorithms for aerial-BSs’ placement according to different scenarios. A polynomial-time algorithm which aims at optimizing aerial-BSs’ placement is adopted in Lyu et al. And more unmanned aerial vehicle base station (UAV-BS) evaluation framework for user coverage with minimum transmit power is shown in Alzenad. And aerial-BSs’ locations considered with height and user locations from different optimization problems are studied as well. Several intelligent algorithms are taken for aerial-BS deployment, for instance, grid search algorithm for 3D UAV-BS placing and minimal aerial-BS number. However, the above literature assumes that all users take on the same QoS constraints. In addition, the aerial-BS locations are modeled as an optimal QoS requirement problem. However, the algorithm is still hard to obtain the optimal results due to the heuristic algorithms.

In recent years, both industries and experts pay lots of attentions to the DRL which was derived from DeepMind. As the name suggests, DRL is the combination of deep learning and reinforcement learning, so DRL makes up for the shortcomings of DL and RL. First, the RL is a study of the mapping of environment state to action space. It is based on the Markov decision process (MDP), that is, the current state is only related to the previous state, regardless of the cumulative influence before the previous state. MDP is usually defined as a quad (S, A, R, P); in addition to quads, RL has two important functions, namely, value function and Q function, so is a state-of-the-art method to produce control policies using the action set. With the development of DL, DQN algorithm appears, which is a promising tool to address multi-agent optimization problems such as the UAV navigation. In Mnih et al., the DQN algorithm which can receive appropriate strategies directly from high-dimensional perceptron inputs was proposed. For further applying DRL in continuous action control and mass discrete action control, a set of control tasks to measure procedures in continuous action control were proposed by Duan et al. In Duan et al., as an emerging deep machine learning method, DRL has been widely used in UAV-enabled wireless communication control. In Wu et al., in view of DQN, a 3D aerial-BS location planning algorithm is adopted in capacity enhancement, but just one aerial-BS is considered. Consider of environment learning, a two-step algorithm is applied in the UAVs’ intelligent arrangement in Luo et al. Next, in Liu et al., on account of UAV control in coverage and connectivity, a DRL-driven energy saving algorithm which outperforms baseline methods is introduced. For joint virtual reality (VR) content caching and transmission problem with cellular-connected UAVs, a distributed DL algorithm which integrates liquid state machine and echo state networks is introduced in Chen et al. Moreover, a DRL approach which aims at obtaining the optimum trace of UAV is analyzed in Saxena et al. Above DRL algorithms are all aiming at resolving a specific optimization problem, which show their efficiency. Wang et al. have formalized the UAV movement problem as a constrained Markov decision process (CMDP) problem and employed Q-learning to solve the UAV movement problem. But this study has not considered the different requirements of users and the convergence is decided by the number of UAVs. So it may not be suitable for sudden traffic compensation as well. Huang et al. have presented a DRL-based scheme for UAV navigation through massive multiple-input and multiple-output (MIMO). But the optimal location at the UAVs is obtained based on the received signal strengths without requiring global information, so it may fall into local optimum rather than global optimum.

In our previous work, user QoS requirements for different services with DQN are proposed, but just one aerial-BS is considered. Moreover, the reference signal received power (RSRP) and signal to interference and noise ratio (SINR) which are important indicators related to QoS are not considered as well. In this article, we will extend it to multiple BSs scenario with sudden traffic compensation, and the detailed description is shown below.

**System model**

It can be seen in Figure 1, a macro BS and a lower power node (LPN) are deployed in the ground. When the gymnasium holds concerts or other activities, it will

---

**Figure 1.** Scenario description for sudden traffic.
cause sudden traffic hotspots, which put forward higher requirements for the capacity of mobile data access and exceed the allowed capacity. To compensating additional data requirement temporarily, we can enhance the capacity of local networks by deploying mmWave aerial-BSs with high bandwidth, which can also provide services for users beyond the coverage of macro BSs. However, due to its signal attenuation and bad building penetration feature as shown in the figure, the path loss of mmWave link should be considered as well. The deployment location of aerial-BSs not only affects the number of users in its coverage area, but also affects the quality of the air-to-ground link. Because of the characteristics of aerial-BSs, the air-to-ground channel suitable for aerial-BSs should consider the 3D location effect; moreover, it has a higher chance of LOS connection.

In 5G mmWave networks, signal transmission will be affected by LOS and NLOS connections. Although aerial-BSs are suitable for direct beamforming, all the loss should be considered as well. To evaluate the received signal properly, in this article, an air-to-ground propagation model for aerial-BSs is constructed first. Next, the correlations from path loss to the maximum coverage radius are analyzed. Finally, average SE is calculated for optimizing the best aerial-BS location next.

**Air-to-ground path loss model**

As shown in our previous work, the air-to-ground channel should consider the impact on the occurrence of LOS. Based on recommendation from the International Telecommunication Union (ITU) for radio transmission, the important parameters for determining geometric probability of LOS transmission are given as follows:

1. \(\alpha\), which denotes the proportion of places that taken over by all buildings.
2. \(\beta\), which denotes the average building number per unit area (span/km\(^2\)).
3. \(\gamma\), which denotes the scale of building height distribution, and it is always taken as the Rayleigh probability density function. With these parameters, LOS propagation probability equation of link from user \(i\) to aerial-BS \(j\) can be shown as

\[
p_{\text{LOS}}(i,j) = \prod_{n=0}^{m} \left( 1 - \exp \left( -\frac{(h_j - (n + 1) \frac{h_j - h_i}{n + 1})^2}{2\gamma^2} \right) \right)
\]

(1)

with

\[
m = \left| (h_j - h_i) \tan \theta \sqrt{\alpha \beta} - 1 \right|
\]

(2)

where \(h_j\) and \(h_i\) are the height of transmitter (aerial-BS) and receiver (user), respectively. And can be found that the geometric LOS formula has no relation with the system frequency. Still, it can be replaced by sigmoid function as well.

Next, the probability of NLOS link from aerial-BS \(j\) to ground user \(i\) is given by

\[
p_{\text{NLOS}}(i,j) = 1 - p_{\text{LOS}}(i,j)
\]

(3)

Moreover, the path loss of the entire link from aerial-BS \(j\) to user \(i\) is given by

\[
\eta_{\text{path}}(i,j) = A_{\text{path}} + 10\delta_{\text{path}} \log d_{i,j}
\]

(4)

where \(A_{\text{path}}\) is the path loss, which is based on the reference distance and can be used both for LOS and NLOS, and \(\delta_{\text{path}}\) represents the path loss parameter. \(d_{i,j}\) is the distance from aerial-BS \(j\) to user \(i\).

For aerial-BS \(j\), its coverage radius \(r_j\) is correlated with its antenna height. With previous work in Guo et al. and related work in Moaffar et al., the relationship between aerial-BS location and its coverage is given by

\[
r_j^2 = h_j^2 + d_{i,j}^2 \leq 10^{\frac{A_{\text{path}} - \log(h_j d_j)}{10}}
\]

(5)

where \(h_j\) is the height of aerial-BS \(j\), \(A = \eta\text{LOS} - \eta\text{NLOS}\), and \(\eta\text{LOS}\) and \(\eta\text{NLOS}\) denote the average additional losses related to environment. Moreover, \(E = 20\log4\pi f_c/c + \eta\text{NLOS}\), where \(f_c\) denotes the carrier frequency. And \(\text{loss}(\cdot)\) denotes the path loss from aerial-BS \(j\) to user \(i\).

**Optimization problem**

Based on above channel model, we will analyze the aerial-BS requirement scenario for compensating sudden traffic when macro BS cannot satisfy all the data requirement. We assume that \(K\) different QoS requirements exist in the network under sudden traffic. Assuming that \(U\) is the user set and \(U_k \subseteq U\) is the users set with class \(k\) QoS type, so that \(\bigcup_k U_k = U\). We use \((x_i, y_i)\) to represent the position of user \(i\).

To maximize the effectiveness of aerial-BSs, SE should also be considered. Assuming the total bandwidth of aerial-BS \(j\) is \(B_j\), and the bandwidth allocated to each user \(i\) with QoS type \(k\) is \(b_{i,k}\). And user \(i\)'s received signal power from aerial-BS \(j\) with QoS type \(k\) is expressed as \(S_{\text{path}}((i,j,k))\), which is given by

\[
S_{\text{path}}((i,j,k)) = \frac{b_{i,k}}{B_j} p_{TX} 10^{-\frac{\text{loss}(i,j)}{10}}
\]

(6)

\[
= \frac{b_{i,k}}{B_j} p_{TX} A_{\text{path}}^{-\delta_{\text{path}}}
\]
where $A_{\text{path}} = 10^{-A_{\text{path}}/10}$, and $P_j^{\text{TX}}$ is the transmit power of aerial-BS $j$. Moreover, user $i$’s total noise power for QoS type $k$ is given by Wu et al. 28

$$N_{i,k} = 10^{(\frac{A_{\text{path}}}{10} + \rho_i)} b_{i,k} 10^{-3}$$  \hspace{1cm} (7)$$

where $\rho_i$ denotes the user $i$’s devices noise figure. Consequently, user $i$’s SINR from aerial-BS $j$ be shown as

$$\text{SINR}(i,j,k) = \frac{S_{\text{path}}(i,j,k)}{N_{i,k}}$$  \hspace{1cm} (8)$$

Basic SE is shown as below (consider of Shannon’s theorem)

$$\Phi(i,j,k) = \log_2(1 + \text{SINR}(i,j,k))$$  \hspace{1cm} (9)$$

According to above analysis, the average SE with mmWave path is shown as

$$\Phi(i,j,k) = P^\text{LOS}(i,j) \log_2 \left( 1 + \frac{S_{\text{path}}^\text{LOS}(i,j,k)}{N_{i,k}} \right)$$

$$+ P^\text{NLOS}(i,j) \log_2 \left( 1 + \frac{S_{\text{path}}^\text{NLOS}(i,j,k)}{N_{i,k}} \right)$$  \hspace{1cm} (10)$$

And in our article, we want to find best locations for all the aerial-BSs with highest SE with all the users under sudden traffic. And the optimization problem is shown below

$$\max_{\{b_i\}, \{s_i\}, \{t_{i,j}\}} \mathcal{P} = \sum_{k=1}^{K} \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} \Phi(i,j,k) I_{i,j,k}$$

s.t. $d_{i,j} = \sqrt{(h_j - h_i)^2 + (x_j - x_i)^2 + (y_j - y_i)^2}$

$$\forall j, h_{\text{min}} \leq h_j \leq h_{\text{max}}$$

$$\forall j, x_{\text{min}} \leq x_j \leq x_{\text{max}}$$

$$\forall j, y_{\text{min}} \leq y_j \leq y_{\text{max}}$$

$$\forall i, j, I_{i,j,k} \in \{0, 1\}$$

$$\forall i, k, \sum_{j=1}^{N_j} I_{i,j,k} \leq 1$$

$$\forall i, j, k, \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} I_{i,j,k} \geq \zeta N_U$$

$$\forall i, j, k, \text{SINR}(i,j,k) \geq \sigma_{\text{min}}$$

$$\forall i, j, k, S_{\text{path}}(i,j,k) \geq \mu_{\text{min}}$$

when the $i$th user is connected to the aerial-BS $j$ with service $k$, $I_{i,j,k} = 1$. $[x_{\text{min}}, x_{\text{max}}], [y_{\text{min}}, y_{\text{max}}], [h_{\text{min}}, h_{\text{max}}]$ denote the 3D area ranges of the aerial-BSs. Here, $\zeta$ is the lowest ratio of users need to be served. And the received signal and SINR should above target value $\sigma_{\text{min}}$ and $\mu_{\text{min}}$, respectively. Moreover, each user can be connected to one aerial-BS at most.

### DQN-based aerial-BS location optimization framework

In this section, in order to maximize the total SE with suitable aerial-BS number and users’ QoS constraints and consider of DQN, we propose an framework for 3D location deployment of aerial-BSs. First, we use $K$-means to cluster the users according to their geographical location to determine connections and aerial-BS number, and use DQN to seek out the optimal positions of aerial-BSs next. And the framework can be seen in Figures 2–4.

To resolve the aerial-BS location problem, we need a network entity which is responsible for the data collecting and solution planning, and pushing corresponding control information to the aerial-BSs. From our previous work, self-organized network (SON) architecture is a reasonable scheme for it. In this article, distributed SON architecture will be adopted here. We assume that macro BS can be set as the relay for aerial-BSs’ backhaul in this scenario. And then, the SON agent deployed in the macro BS is the entity responsible for the collection, analysis, planning, and evaluation.

### K-means algorithm for partitioning region with aerial-BSs

K-means algorithm can find out the solutions of clustering problems and belongs to unsupervised learning.
algorithms. Based on a fixed prior, the given data set can be classified through some clusters. In which the important matter is to set $K$ centroids using $K$ cluster. These centers of mass should be put skillfully, which can lead to various results. Therefore, it is preferably to keep the centers of mass as far apart as possible. And then classify each point to the given data set and link it to the nearest centroid. If all points have been processed, complete the first step and the early grouping. At this point, we need to recalculate $K$ new centroids as the focus of clustering generated in the previous procedure. New bindings must be made between the same data set and the latest centroids after obtaining $K$ new centroids. So the algorithm can be executed in a loop. Consequently, it can be easily found that $K$ centroids gradually change their positions until they are no longer changed. In other words, the center of mass is no longer moving.

For example, in Figure 2, if there are emergencies in the gymnasium. At this time, the users can be classified according to their geographical location because they have different environments and QoS. Like the users who are in the gym and the connection between the macro BS is NLOS, they may often have communication interruption to compete for communication resources owing to many people, so the demand of these users is to provide adequate communication resource to keep the communication process connected. Like the users who are out of the gym, they may be affected by the emergencies in the gymnasium although they are within the communication range of the BS but the load of the macro BS is over. So the demand of these users is to eliminate the interference of the gymnasium and ensure continuous communication. Like the users who are out of the gym and at the edge of the coverage area, so the demand of these users is to attain ample communication resource to finish the process of the communication.

Here, the $K$-means algorithm aims at minimizing the square error function given by

$$J = \sum_{j=1}^{K} \sum_{x \in S_j} ||x - c_j||^2$$

where $K$ is the cluster number, and $c_j$ is the average value in data cluster $S_j$. The steps for finding best cluster center can be found in our previous work in Mnih et al.\textsuperscript{15} as well.

As users' call admission at wireless link is complex,\textsuperscript{13,29} in this article, we can just assume that bandwidth for each aerial-BS $B_j$ is equal and denoted as $B$. And $R$ is defined as average user bandwidth requirement here. Next, based on the largest number of users which can be served by an aerial-BS at the same time, $N_U$ is given by

$$N_U = \left[ \frac{CBS}{R} \right]$$

$$CBS = B \times \phi$$

where $CBS$ is the capacity of one aerial-BS, $B$ is each aerial-BS’s total bandwidth, and $\phi$ denotes the average spectrum efficiency defined in Alzenad et al.\textsuperscript{14} And then the number of aerial-BSs required is given by

$$N_B = \left[ \frac{N_A}{N_U} \right]$$

where $N_A$ is the number of users with sudden traffic. Here, users are clustered by $K = N_B$. After the aerial-BS number and location is determined, then the location planning of aerial-BSs is carried out for each small area is shown next.

**DQN algorithm for location planning**

In this section, consider of DQN an algorithm of 3D deployment of aerial-BSs is proposed with definite
aerial-BS number for the sake of finding the maximum total SE shown in Alzenad et al.\textsuperscript{14} The DQN model used here can be found in Guo et al.\textsuperscript{24} as well. The DQN model used here integrates convolutional neural networks with RL model named Q-learning. Through the process of DQN, in view of the rewards received from the interaction procedure with the environment, the agent can learn continuously for the aim of getting the target status. In this algorithm, \(<A, S, R, P>\) is a classic quaternion for learning, where action \(A\) denotes the action set of the agent, state \(S\) means the state set of the agent, reward \(R\) is the value sets denoting reward or punishment, and finally, \(P\) is the probability of the agent in taking an action in the state space. The \(Q\) value will be trained with DL model, all the information required coming from the network, and agent can be taken as the SON entity shown in Yu et al.\textsuperscript{1} We will give detailed description for this algorithm as below.

In this algorithm, the agent is the candidate aerial-BS sets whose state space, action space, and rewards are defined by as follows:

The state space: \(S = (h, x, \varphi, y)\), where they denote the height, \(x\)-axis and \(y\)-axis coordinates shown in Alzenad et al.\textsuperscript{14} respectively.

The action space: \(A = \{0, 1, 2, 3, 4, 5, 6\}\) denotes moving directions of aerial-BSs, which are upward, downward, positive and negative directions of the \(x\)-axis, positive and negative directions of the \(y\)-axis, and maintaining current locations, respectively.

The reward: the system SE shown in Alzenad et al.\textsuperscript{14} and maintaining current locations, respectively.

The classic expression of the Q-learning is
\[
Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha (r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t))
\]
where \(Q(s_t, a_t)\) is the reward discount received from when the agent choosing action \(a_t\) under state \(s_t\), and \(\alpha\) denotes the learning rate. The greater the learning rate, the less previous learning outcomes are retained, \(\gamma\) denotes the discount factor, the larger the discount factor, the more the learning entity pays attention to the previous learning experience, and also the more it pays attention to the maximization of the reward value at hand. And this algorithm will choose the action based on the greedy strategy until the function gets optimal strategy as below
\[
\pi(s) = \arg \max_{a \in A} Q(s, a)
\]
This will find the best action for each state; however, the \(Q\) matrix has limited ability to store information, when the state is too much or discrete, this algorithm will naturally cause dimensional disasters, so Google’s DeepMind team will combine DL with RL, the DQN model is proposed, which uses the value function \(f\) trained from DL model to approximate the \(Q\) value
\[
Q(s, a) = f(s, a)
\]
From the above value function, the functional relationship refers to learning through a neural network to obtain the \(Q\) values and the functional mapping relationship between states and actions. The neural network uses two fully connected neural networks which have same structure but has difference in parameters: main network and target network. During the training process, the first cycle randomly generates different user distribution environments, and the second cycle iterates aim at finding the 3D positions of the aerial-BSs which have maximum SE. First, initialize a random state \(s_t\), and then use the \(\varepsilon\)-greedy strategy to select the action. That is, an action \(a_t \in A\) is randomly selected from the action set by the probability \(\varepsilon\), and the action \(a_t = \max_a Q(s_t, b)\) having the highest action value is selected with the probability \(1 - \varepsilon\). Get the new state \(s_{t+1}\) and reward \(r_t\), and update the current \(Q\) value, then update the target \(Q\) value every \(C\) steps, and reverse transmission with the square of the difference between the two as a loss function. The algorithm architecture is shown in Figure 5.
Algorithm 1: 3D locations of aerial-BSs with DQN algorithm

Input: Initial 3D positions aerial-BSs, locations of users with sudden traffic

Output: Optimal 3D locations of aerial-BSs and connected users

1 Initialize replay memory D to capacity N;
2 Initialize action-value function Q with random weights \( \theta \);
3 Initialize target action-value function \( \hat{Q} \) with weights \( \theta^n \leftarrow \theta \);

for episode = 1 to \( N \) do
5 For every user \( i \), get the location \((x_i, y_i)\) with sudden traffic;
6 Connect users to aerial-BSs under its coverage;
7 For every aerial-BS \( j \), get the 3D locations \((x_j, y_j, z_j)\);
8 Set sequence \( s_1 \leftarrow e_1 \), preprocess \( f_1 \leftarrow f(s_1) \);
9 for \( t = 1 \) to \( T \) do
10 Select a random action \( a_t \) for every aerial-BS \( j \) with probability \( \epsilon \) otherwise select \( a_t = \text{max}_aQ(f(s_t), a; \theta) \);
11 Execute action \( a_t \) in emulator and observe reward \( r_t \) and image \( x_{t+1} \);
12 If all the constraints in equation (11) are satisfied, Set \( s_{t+1} \leftarrow (s_t, a_t, x_{t+1}) \) and preprocess \( f_{t+1} \leftarrow f(s_{t+1}) \), otherwise go back to step 8;
13 Store transition \((d_{sj}, a_t, r_t, f_{t+1})\) in \( D \);
14 Sample random minibatch of transitions: \((d_{sj}, a_t, r_t, f_{t+1})\) from \( D \);
15 Set \( g_j = \left\{ \begin{array}{ll}
\gamma \max_aQ(f_{t+1}, a' ; \theta) & \text{steps at step } j+1 \\
\gamma \max_aQ(f_{t+1}, a' ; \theta) & \text{otherwise}
\end{array} \right. \)
16 Perform a gradient descent step on \((g_j - Q(f_j, a_j; \theta))<sup>2</sup>\) with respect to the network parameters \( \theta \);
17 Every \( \delta \) step do reset \( \hat{Q} \leftarrow Q \);
18 end
19 end

For high-dimensional state space, the DQN algorithm inputs state \( S \) and outputs a matrix, \([Q(s, a_1), Q(s, a_2), \ldots, Q(s, a_n)]\), which denotes the current reward and punishment values corresponding to all possible actions, by means of empirical learning, establish the mapping relationship between the state \( S \) and the matrix, and then select the optimal action from it. With the above analysis, the DQN procedures for 3D aerial-BS locations are shown in Algorithm 1.

In this algorithm, it has two key technologies, which are as follows:

1. Experience reply: first, all samples are placed in the sample pool, next in order to train network, choose a sample randomly from the sample pool. This process makes samples have no relation with each other.
2. Fixed Q-target network: as shown in Figure 5, the calculation of network target value requires the existing \( Q \) value. So the \( Q \) value is generated by a slower network. With the training steps, we can get the best locations for each aerial-BS.

Simulation results

Using an area of \( 3.0 \, \text{km} \times 3.0 \, \text{km} \) urban environment in simulation process. In this area, we assume that 1000 users distributed randomly under three subregions, which is shown in Figure 6. And a ground BS with same parameters is deployed in this region as well. Still, we assume that basic parameters and spectrum efficiency calculation method of all the BSs are the same. Moreover, users choose serving BSs with the higher spectrum efficiency BS (ground BS or aerial-BS). These assumptions make our method able to meet the sudden traffic demands in the region. This scenario can be extended to software-defined networking (SDN)/network function virtualization (NFV)-enabled 5G networks as well.\(^{30}\)

Next, Table 1 shows the parameters appeared in the algorithm. First, users are clustered by K-means algorithm and divided into small areas. And the result shows that three aerial-BSs are required here. Aim at making the scenario meets the real requirements, put the ground BS in the position of region center. The clustering result is shown in Figure 7 with different colors. As the users are distributed randomly under separate regions, so they can be easily be clustered. Next, we will find the best locations for different aerial-BSs with DQN algorithm.

Then, position planning is carried out for each subregion with corresponding aerial-BSs. Next, choose a 3D position randomly as the initial status of the aerial-BSs. Through training, it is shown in Figure 8 that as the learning process progresses, the aerial-BSs gradually move to the positions where the total SE of the system is the highest. As shown in the figure, the red, green, and blue dots at the bottom represent the users in the
three hotspot areas, the triangle of the red, green, and blue dots at the top represents the mobile process of the aerial-BSs related in each subregion, and the yellow pentagrams represent the optimized locations of the aerial-BSs.

After a certain scale of learning, the network structure parameters of DQN algorithm are obtained and saved as a model. When the model is directly applied in this scenario, the aerial-BSs will stay in the optimal position of system spectrum efficiency. Figure 9 shows the results of three different clusters comparing to the average SE. It can be seen that, the spectrum efficiency is converged along with training steps, and the lowest spectrum efficiency of the learned aerial-BS can reach 92.77% of the maximum spectrum efficiency of the system under ideal conditions. Moreover, the spectrum efficiency of different aerial-BSs is very near, which denotes the balance of our proposed algorithm. The advantage of using DQN algorithm is that after learning the model, it can be applied directly with near global optimal result, and the application time is very short, so the efficiency is very high.

Besides the spectrum efficiency, it is important to evaluate the users’ quality from signal strength and interference perspective. So we analyze the cumulative probability distribution function of RSRP and SINR for each aerial-BS here, which are shown in Figures 10 and 11 in the three regions, respectively. As shown in Figure 10, different aerial-BSs take on different effect. However, all the users’ RSRP values are higher than $-120$ dBm and lower than $-50$ dBm in our scenario, which means that the RSRP constraints are all satisfied. And 95% users’ RSRP is lower than $-70$ dBm.
Next, as shown in Figure 11, different aerial-BSs take on different variations as well, but all of them are higher than $-10$ dB, and their values range from $-10$ to 23 dB. As an important indicator, it means that all the users can be served perfectly. Although the SINR distributions for different aerial-BSs at the beginning have some discrepancies among each other, they will be very close to each other at last, and more than 10% users’ SINR values are higher than 20 dB, which still take on acceptable performance. From above analysis,
it is easy to see that our proposed mechanism can receive optimum spectrum efficiency with mmWave aerial-BSs, and make users’ quality above acceptable levels.

**Conclusion and future work**

Through the article, we have studied multiple aerial-BSs’ optimal locations to cope with sudden traffic with optimal spectrum efficiency. First, in view of probabilistic LOS/NLOS mmWave wireless connections, we can obtain the downlink coverage probability and path loss. And then, consider of DQN, we present an effective location method to satisfy the data requirements of users with different QoS. It can be seen from the simulation results that under certain constraints, if the learning entity has a sufficiently long learning time, after enough iterations, the learning entity will learn the environmental characteristics and save the learning model, and find the best deployment locations in a very short time when applying the model. The learned SE of the system can reach 70.76% of the maximum SE under ideal conditions.

As communication failure in sudden disaster or such emergency scenario has similar traffic problem, so we will extend our solutions for such applications in our future work. Moreover, we will try to find more efficient learning model for this problem, and explore its applications to new network such as WiMAX or use new technologies such as SDN/NFV as well. Still, how to model the interference and UAV formation with dense regions should be considered as well.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the open fund project of Science and Technology on Communication Networks Laboratory (grant no. SXX18641X024) and the National Natural Science Foundation of China (grant no. 61971053).

**ORCID iD**

Peng Yu  
https://orcid.org/0000-0002-0402-5390

**References**

1. Yu P, Li W, Zhou F, et al. Capacity enhancement for 5G networks using MmWave aerial base stations: self-organizing architecture and approach. *IEEE Wirel Commun* 2018; 25(4): 58–64.

2. Gomez K, Rasheed T, Reynaud L, et al. On the performance of aerial LTE base-stations for public safety and emergency recovery. In: *GLOBECOM workshops*, Atlanta, GA, 9–13 December 2013, pp.1391–1396. New York: IEEE.

3. Deaton J. High altitude platforms for disaster recovery: capabilities, strategies, and techniques for emergency telecommunications. *EURASIP J Wirel Commun Netw* 2008; 2008: 153469.

4. Zong R, Gao X, Wang X, et al. Deployment of high altitude platforms network: a game theoretic approach. In: *International conference on computing, networking and communications (ICNC)*, Maui, HI, 30 January–2 February 2012, pp.304–308. New York: IEEE.

5. Zhang T, Zhou F, Feng L, et al. Capacity enhancement for next generation mobile networks using MmWave aerial base station. In: *International conference on communication*, Paris, 21–25 May 2017, pp.1–6. New York: IEEE.

6. Kalantari E, Yanikomeroglu H and Yongacoglu A. On the number and 3D placement of drone base stations in wireless cellular networks. In: *IEEE 84th vehicular technology conference (VTC-Fall)*, Montréal, QC, Canada, 18–21 September 2016, pp.1–6. New York: IEEE.

7. Alzenad M, El-Keyi A, Lagum F, et al. 3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage. *IEEE Wirel Commun Lett* 2017; 6(4): 434–437.

8. Ghanavi R, Kalantari E and Sabbaghiyan M. Efficient 3D aerial base station placement considering users mobility by reinforcement learning. In: *IEEE wireless communications and networking conference, WCNC*, Barcelona, 16–18 April 2018, pp.1–6. New York: IEEE.

9. Lyu J, Zeng Y, Zhang R, et al. Placement optimization of UAV-mounted mobile base stations. *IEEE Commun Lett* 2017; 21(3): 604–607.

10. Al-Hourani A, Kandeepan S and Lardner S. Optimal LAP altitude for maximum coverage. *IEEE Wirel Commun Lett* 2014; 3(6): 569–572.

11. Bor-Yaliniz RI, El-Keyi A and Yanikomeroglu H. Efficient 3-D placement of an aerial base station in next generation cellular networks. In: *IEEE international conference on communications (ICC)*, Kuala Lumpur, Malaysia, 22–27 May 2016, pp.1–5. New York: IEEE.

12. Kalantari E, Shakir MZ1, Yanikomeroglu H, et al. Backhaul-aware robust 3D drone placement in 5G + wireless networks. In: *International conference on communication*, Paris, 21–25 May 2017, pp.109–114. New York: IEEE.

13. Plachy J, Becvar Z, Mach P, et al. Joint positioning of flying base stations and association of users: evolutionary-based approach. *IEEE Access* 2019; 2019: 11454–11463.

14. Alzenad M, El-Keyi A and Yanikomeroglu H. 3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements. *IEEE Wirel Commun Lett* 2017; 7(1): 38–41.

15. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature* 2015; 518(7540): 529–533.

16. Duan Y, Chen X, Houthooft R, et al. Benchmarking deep reinforcement learning for continuous control. In: *The 33rd international conference on machine learning*.
17. Wu J, Yu P, Feng L, et al. 3D aerial base station position planning based on deep Q-network for capacity enhancement. In: IFIP/IEEE symposium on integrated network and service management (IM), Washington, DC, 8–12 April 2019, pp.482–487. New York: IEEE.

18. Luo X, Zhang Y, He Z, et al. A two-step environment-learning-based method for optimal UAV deployment. *IEEE Access* 2019; 7: 149328–149340.

19. Liu CH, Chen Z, Tang J, et al. Energy-efficient UAV control for effective and fair communication coverage: a deep reinforcement learning approach. *IEEE J Select Area Commun* 2018; 36(9): 2059–2070.

20. Chen M, Saad W and Yin C. Echo-liquid state deep learning for 360° content transmission and caching in wireless VR networks with cellular-connected UAVs. *IEEE T Commun* 2019; 67(9): 6386–6400.

21. Saxena V, Jaldén J and Klessig H. Optimal UAV base station trajectories using flow-level models for reinforcement learning. *IEEE T Cognit Commun Netw* 2019; 5: 1101–1112.

22. Wang Q, Zhang W, Liu Y, et al. Multi-UAV dynamic networking with deep reinforcement learning. *IEEE Commun Lett* 2019; 23: 2243–2246.

23. Huang H, Yang Y, Wang H, et al. Reinforcement learning for UAV navigation through massive MIMO technique. *IEEE T Veh Technol* 2020; 69: 1117–1121.

24. Guo J, Huo Y and Shi X. 3D aerial vehicle base station (UAV-BS) position planning based on deep Q-learning for capacity enhancement of users with different QoS requirements. In: 15th international wireless communications & mobile computing conference (IWCMC), Tangier, 24–28 June 2019, pp.1508–1512. New York: IEEE.

25. Geneva, Rec. P.1410-2. *Propagation data and prediction methods for the design of terrestrial broadband millimetric radio access systems*. Geneva: ITU_R, 2003.

26. Mozaffari M, Saad W, Bennis M, et al. Unmanned aerial vehicle with underlaid device-to-device communications: performance and tradeoffs. *IEEE T Wirel Commun* 2016; 15(6): 3949–3963.

27. Mirahsan M, Schoenen R and Yanikomeroglu H. HetHetNets: heterogeneous traffic distribution in heterogeneous wireless cellular networks. *IEEE J Select Area Commun* 2015; 33(10): 2252–2265.

28. Wu Y, Rong B, Salehian K, et al. Cloud transmission: a new spectrum-reuse friendly digital terrestrial broadcasting transmission system. *IEEE Trans Broadcast* 2012; 58(3): 329–337.

29. Rong B, Qian Y, Lu K, et al. Call admission control optimization in WiMAX networks. *IEEE T Veh Technol* 2008; 57(4): 2509–2522.

30. Sun S, Kadoch M, Gong L, et al. Integrating network function virtualization with SDR and SDN for 4G/5G networks. *IEEE Netw* 2015; 29(3): 54–59.