

# Efficient 3D Aerial Base Station Placement Considering Users Mobility by Reinforcement Learning

Rozhina Ghanavi\*, Elham Kalantari†, Maryam Sabbaghian\*, Halim Yanikomeroglu‡, and Abbas Yongacoglu†

\*School of Electrical and Computer Engineering

University of Tehran, Tehran, Iran, Email: {rghanavi, msabbaghian}@ut.ac.ir

†School of Electrical Engineering and Computer Science

University of Ottawa, Ottawa, ON, Canada, Email: {ekala011, yongac}@uottawa.ca

‡Department of Systems and Computer Engineering

Carleton University, Ottawa, ON, Canada, Email: halim@sce.carleton.ca

**Abstract**—This paper considers an aerial base station (aerial-BS) assisted terrestrial network where user mobility is taken into account. User movement changes the network dynamically which may result in performance loss. To avoid this loss, guarantee a minimum quality-of-service (QoS) and possibly increase the QoS, we add an aerial-BS to the network. For fair comparison between the conventional terrestrial network and the aerial-BS assisted one, we keep the total number of BSs identical in both networks. Obtaining the best performance in such networks highly depends on the optimal placement of the aerial-BS. To this end, an algorithm which can rely on general and realistic assumptions and can decide where to go based on the past experiences is required. The proposed approach for this goal is based on a discounted reward reinforcement learning which is known as Q-learning. Simulation results show this method provides an effective placement strategy which increases the QoS of wireless networks when it is needed and promises to find the optimum position of the aerial-BS in discrete environments.

## I. INTRODUCTION

Ubiquitous connectivity, reliable quality-of-service (QoS), extremely high number of connected devices, and supporting diverse services are essentials of the fifth generation (5G) and beyond-5G (5G+) wireless networks. To obtain the distinctive attributes of 5G, it is imperative to exploit ingenious approaches. One of the emerging technologies which increases the agility of the network while enhancing the QoS, is utilizing aerial base stations (aerial-BSs), especially when the existing terrestrial infrastructure is insufficient to address the demand. Aerial-BSs can have several use cases, such as assisting the legacy terrestrial network to relieve congestion in case of an unusual excessive demand (such as a festival or a sport match). They can also provide temporal coverage when the ground-BSs are not accessible (such as a natural disaster which might damage the ground-BSs).

### A. Related Work

Use of aerial-BSs in wireless networks has gained attention in recent years. They may offer the most suitable way to

enhance the QoS of the next generation wireless networks. In [1], [2], the positioning of aerial relays under assumption of fixed altitude is discussed. In [3], a novel approach for serving a number of users while using minimum number of aerial-BSs along with finding an efficient 3D placement is proposed. In [4], a backhaul aware 3D placement of an aerial-BS for various network design parameters is presented. In [5], an aerial-BS is positioned with the use of numerical methods to maximize the number of served user. In [6], a free space optical link is proposed as a method to provide backhaul/fronthaul access for flying network platforms. In [7], the optimal placement of an aerial-BS to maximize the number of users served with minimum transmit power is discussed. In [8], considering delay-tolerant and delay-sensitive users, an algorithm is proposed to find efficient 3D locations of aerial-BSs. The algorithm also investigates the user-BS associations and wireless backhaul bandwidth allocations to maximize the sum logarithmic rate of the users in a heterogeneous network including a macro-BS and several drone-BSs. In [9], the downlink coverage performance of aerial-BSs is investigated. In [10], the optimization problem of placing aerial-BSs and ground-BSs to minimize the average network delay is investigated. In [11], [12], aerial-BSs are considered as backhaul/fronthaul links and the association problem of small cells with aerial-BSs is investigated. The effect of aerial relays between ground nodes in a wireless network is discussed in [13].

### B. Our Contribution

One of the drawbacks of the previous studies of aerial-BS systems is that none of them has considered users' movements. In other words, the optimization problem has been solved for a snapshot of the network. In reality, however, it is very likely that the position of the users changes gradually. Correspondingly, there is an indisputable link between the QoS of users and their movements. Another disadvantage of the previous studies is that the used optimization methods such as



Fig. 1. Aerial-BS can assist network through hard to predict situations.

heuristic algorithms may need long processing time and this makes these methods impractical in real situations.

In this paper, we consider a traditional wireless network. User movement inevitably entails considerable changes in the QoS, raising the probability of the desired QoS not being delivered. We cope with this issue by exploiting an aerial-BS (Fig. 1). For a fair comparison between the aerial-BS assisted network and the terrestrial one, we keep the number of BSs in both networks identical. As previously mentioned, the optimal placement of the aerial-BS can significantly influence the system performance. This issue has been the subject of several papers some of which are mentioned in Section I.A. However, most of the articles considering this problem does not provide a closed form solution but present heuristic algorithms [3], [8]. In dynamic environments where the network topology changes, the heuristic algorithm needs to be reinitialized and run for the new topology. This process can be time consuming and imposes more computational complexity to the system. In our case where the network topology changes gradually due to users' movement, reinforcement learning is a promising candidate to solve the problem. Once the training phase is completed, this method continuously adapts the solution for small changes. This prevents the repeated running of the heuristic algorithm for continual topological changes. To the best of our knowledge, 3D positioning of aerial-BSs has not been solved using reinforcement learning. We propose an algorithm which significantly increase the QoS of the network and also it has lower complexity compare to other discussed method in this area.

The rest of this paper is organized as follows. The system model is given in Section II. Section III gives the problem formulation and our novel approach to solve it. Section IV presents the simulation results and finally the study is concluded in Section V.

## II. SYSTEM MODEL

We consider a downlink of a wireless cellular network including several macro-BSs and an aerial-BS. The parameters

$I$  and  $J$  denote the set of users and macro-BSs. We use  $i = I_1, I_2, \dots, I_k$ , and  $j = J_1, J_2, \dots, J_p$ , to index users and BSs, respectively. We assume that the users are served if the network QoS is higher than a threshold QoS which is denoted by  $QoS_{th}$  throughout the paper. This parameter presents the QoS required by the network below which the quality is not acceptable. We assume that at the initial time of  $t_{st}$ , all users in the system fulfill this QoS requirement. However, as time passes, due to gradual movements of the users the QoS changes and it might fall below  $QoS_{th}$ . We propose a framework where by the use of an aerial-BS, we reduce the probability of losing the QoS of the users. Since the aerial-BS movement is costly, we cannot constantly change the position of the aerial-BS. In fact, our goal is to support at least the  $QoS_{th}$  while the position of the aerial-BS is fixed at least for the minimum time interval of  $t_{min}$ . While users are moving in the area users' association to BSs change continuously. Our key solution is to include the prediction of users positions and calculate the users' association for each configuration in optimizing the location of the aerial-BS. This way, the QoS meets the required level without the cost being prohibitive.

### A. Air-to-ground Path Loss Model

The air-to-ground path loss depends on the height of the aerial-BS and its elevation angle with the user. Several studies have investigated the air-to-ground path loss model. In this paper, we adopt the path loss model proposed in [14] and [15]. This model derives the air-to-ground path loss equation by considering two propagation classes, the first one is having a line-of-sight (LoS) links and the second one is the group of non line-of-sight (NLoS) links. The probability of having a LoS link is formulated as

$$P(\text{LoS}) = \frac{1}{1 + \kappa \exp(-\zeta(\frac{180}{\pi})\theta - \kappa)}, \quad (1)$$

where the elevation angle of  $\theta$  equals  $\arctan(\frac{h}{l})$ . The parameters  $h$  and  $l$  denote the altitude and the horizontal distance between the aerial-BS and the user, respectively. Constants  $\kappa$  and  $\zeta$  depend on the environment. The average path loss is defined as

$$PL(\text{dB}) = 20 \log(\frac{4\pi f_c d}{c}) + P(\text{LoS})\eta_{\text{LoS}} + P(\text{NLoS})\eta_{\text{NLoS}}. \quad (2)$$

The first term of (2) presents the free space path loss which depends on the carrier frequency,  $f_c$ , the speed of light,  $c$  and the distance between aerial-BS and user denoted by  $d$ . The constants  $\eta_{\text{LoS}}$  and  $\eta_{\text{NLoS}}$  present the additional loss due to free space propagation. We recall that  $P(\text{NLoS})=1-P(\text{LoS})$ .

### B. Users' Mobility Model

As mentioned earlier, user movements can significantly affect the wireless network performance if the location of the aerial-BS is optimized only for a specific configuration. There are diverse methods to model the movement of ground users. The authors in [16] presented a comprehensive survey on such models. One of the most favored models is the random

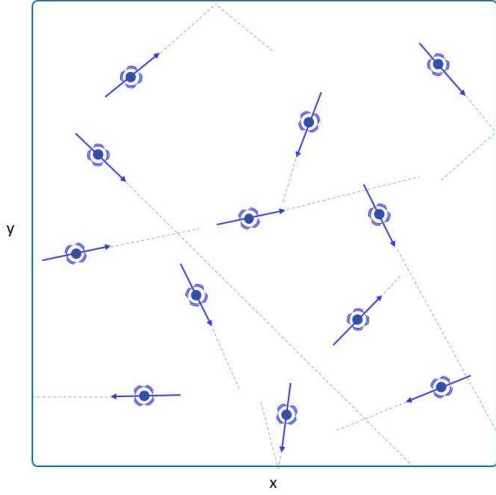


Fig. 2. Users' mobility scenario according to the random walk model.

walk model. In this model the direction of the movement is determined by an angle uniformly distributed between  $[0, 2\pi]$  and a user is assigned a random speed of a pedestrian between  $[0, c_{\max}]$ , where  $c_{\max}$  denotes the maximum speed of a pedestrian. A graphical example of the random walk model is depicted in Fig. 2.

### C. Optimization Problem

To obtain the optimum position of the aerial-BS added to the system, we solve the following optimization problem.

$$\max_{h_\chi, x_\chi, y_\chi} \sum_{j=1}^P \sum_{i=1}^k R_{ij} A_{ij}, \quad (3)$$

subject to

$$\sum_{j=1}^p A_{ij} = 1, \quad (4)$$

$$h_{\min} < h_\chi < h_{\max}, \quad (5)$$

$$x_{\min} < x_\chi < x_{\max}, \quad (6)$$

$$y_{\min} < y_\chi < y_{\max}, \quad (7)$$

$$A_{ij} \in \{0, 1\}, \quad (8)$$

where  $A_{ij}$  is equal to 1 if the  $i$ -th user is assigned to the  $j$ -th BS and it is zero otherwise. The parameter  $\chi$  is the assigned index to the aerial-BS,  $[x_{\min}, x_{\max}]$ ,  $[y_{\min}, y_{\max}]$ , and  $[h_{\min}, h_{\max}]$  denote the 3D area which the aerial-BS could be placed in. In fact, we obtain the optimal position of the aerial-BS such that the aggregate network throughput is maximized. The term  $R_{ij}$  denotes the throughput of the  $i$ -th user determined as

$$R_{ij} = \log_2(1 + \text{SINR}_{ij}), \quad (9)$$

where signal to interference plus noise ratio, SINR, is defined as

$$\text{SINR}_{ij} = \frac{P_{rij}}{\sigma^2 + \sum_{k=1, k \neq j}^p I_k}, \quad (10)$$

where  $P_r$  is the received power and  $\sigma^2$  is the noise power, the noise is a zero-mean white Gaussian noise with variance of  $\sigma^2$ . The term  $I_k$  illustrates the co-channel interference between the BSs. In this study, users choose the BSs based on the maximum SINR. In fact, as the users are moving along the area these assignments change and they decide which terrestrial or aerial-BS to connect based on the maximum SINR approach.

The problem of (3) is an NP-hard problem. Considering the fact that the environment is also changing and using heuristic approaches entails high computational complexity for repeated runnings, we solve this problem exploiting learning approaches. We utilize reinforcement learning specifically Q-learning as a novel approach to find the optimum position of the aerial-BS. It has been proved that under some constraints, this method can achieve the optimal solution in discrete environment if the learner has adequate time for learning [17]. Another advantage of this algorithm is that after sufficient learning time, the learning agent will learn the environment and will be able to find the optimum position in a very short time. This feature is essential to tackle the users' mobility problem, to the fact that this consideration makes the wireless network unsteady and it is necessary to have an algorithm which can guarantee a rapid solution to tune the wireless network with the new conditions.

## III. AERIAL-BS POSITIONING BASED ON Q-LEARNING

Section III.A presents an overview of reinforcement learning and Q-learning. Section III.B applies this algorithm to the problem.

### A. Q-learning Algorithm

Machine learning applications have gained attention in wireless networks in recent years. [18] has a survey on these works. Reinforcement learning has been widely applied in researches which need excessive sense of environments. It relies on the rewards and punishments received by a learning agent during the learning process. In the process of learning, the agent will learn taking the optimal procedure which in our case leads to maximizing the QoS. The learning agent, which is in the current state, takes an action and calculates the reward associated with taking that specific action. This continues until next action. Then, the state of the system changes and this process needs to be run on the new state. In each state transition, the agent generates a matrix including all information collected in state transitions. The information includes rewards and new states. This matrix is going to be used in later state transitions and helps the agent to improve the system performance further.

One of the most practical reinforcement learning techniques is Q-learning [19], which does not need the exact transition formulation. This makes the method more realistic as the agent faces the current and actual wireless network not the formerly designed one. There are number of studies related to the use of Q-learning in wireless networks. As an example, in [20] the authors discuss this method to provide an efficient QoS for multimedia communication. In a Q-learning algorithm, the

agent considers a class of states  $S = s_1, s_2, \dots, s_n$ , a class of actions  $A = a_1, a_2, \dots, a_m$ , and a knowledge matrix  $Q$ . In each state, the learning agent performs an action,  $a_t$  which triggers a state transition. Then, the agent calculates the reward in the new state. This procedure continues until the agent reaches to the desired goal and the algorithm converges [21].

### B. Proposed Algorithm

In this paper, in a cellular system where users are moving, the goal is to maximize the QoS which is the aggregate network throughput (3). We attain this goal by adding an aerial-BS to the existing wireless network while users are moving with the random walk model.

We use Q-learning to find the optimal 3D position of the aerial-BS to increase the QoS in  $t_{\min}$ . This goal is achieved by maximizing the reward  $r_t$  in the state  $s_t$  by taking the action  $a_t$ . In our learning process, we have events which are the movements of the users. If the QoS becomes lower than  $QoS_{th}$ , aerial-BS needs to change its position. In this situation learning agent must decide which action it should take to change the state. We define 6 actions through which the aerial-BS can explore the plane. The area is between  $[x_{\min}, x_{\max}]$ ,  $[y_{\min}, y_{\max}]$ , and  $[h_{\min}, h_{\max}]$  which are defined in Table II. When the mentioned event occurs and the learner decides to take the desired action,  $a_t$ , the current position of the aerial-BS changes and the action takes the aerial-BS to a new position. Here, the states are defined as the positions that the aerial-BS can be in, in the 3D coordinate. The system will receive deterministic rewards based on the action the learner takes in each state. The reward in the  $t$ -th time interval is defined as

$$r_t = QoS_t - QoS_{t-1}. \quad (11)$$

The  $Q$  knowledge matrix for this problem consists of the following elements

$$Q(s_t, a_t) = \alpha[r_{t+1} + \gamma \max Q(s_{t+1}, a) - Q(s_t, a_t)], \quad (12)$$

where  $s_t$  is the state of the agent at time  $t$  and  $a_t$  is the corresponding action the agent takes.  $0 < \alpha < 1$  is the learning rate, which decrease throughout the learning process for the convergence of the solution.  $0 < \gamma < 1$  is the discount factor which determines the speed of convergence and the accuracy of the process. In this study, we use  $\epsilon - greedy$  strategy for choosing exploration or exploitation action in the learning process. Exploration in learning procedure allows the agent to neglect the locally optimal results. While, exploitation occurs when the agent takes optimum actions based on its knowledge.

At initial time  $t = t_{st}$ , we calculate the throughput of the system with existing ground-BSs and users and consider it as  $QoS_{th}$ . Then users walk along the area with the random walk model. We assume that users continue walking in their decided direction for at least 10 s [22]. In this process users' association to BSs change and the new QoS is calculated. While users are moving in the area Q-learning algorithm is running in order to find the optimum position for the aerial-BS. This calculation would be run while the system is being used. As wireless network usage continues, the time required

in each iteration to obtain the optimal solution reduces. In each time when the system starts from  $t_{st}$ , we use the  $Q$  matrix of the previous iteration as the initial  $Q$  matrix. This increases the speed of convergence and can be considered as a tremendous motive to use Q-learning in the wireless network. Previous steps explained earlier are presented in Algorithm 1.

---

#### Algorithm 1 Q-learning algorithm for 3D placement of an aerial-BS

---

**Input:** Distribution of BSs which are in hexagonal pattern and users with Poisson point process.

**Output:** Optimum position of the aerial-BS

- 1:  $QoS_{th}$  is the calculated aggregate network throughput for the given ground wireless network.
  - 2: **while** system is running **do**
  - 3:   Let the users move for  $t_{\min}$  with the given users' mobility model.
  - 4:   Calculate the aggregate network throughput for the new system.
  - 5:   **if** ( QoS of the system with 19 ground-BSs <  $QoS_{th}$ ) **then**
  - 6:     Initialize  $Q(s, a)$  equal to the calculated  $Q(s, a)$  from the previous system running session.
  - 7:     **for** episode = 1 to max episode **do**
  - 8:       Initialize  $s$  as the previous state of the aerial-BS;
  - 9:       **for** step=1 to max step **do**
  - 10:          Decide which action to be taken, using  $\epsilon - greedy$ ;
  - 11:          Take action  $a_t$ , calculate the  $r_t$ , calculate the next  $s$  of the aerial-BS;
  - 12:           $Q(s_t, a_t) = \alpha[r_{t+1} + \gamma \max Q(s_{t+1}, a) - Q(s_t, a_t)]$ ,
  - 13:          Update the state  $s$  of the aerial-BS;
  - 14:       **end for**
  - 15:     **end for**
  - 16:   **else**
  - 17:     There is no need to use an aerial-BS.
  - 18:   **end if**
  - 19: **end while**
- 

## IV. SIMULATION RESULTS

We consider a cellular network with frequency reuse 1 in an urban area where 19 ground-BSs are placed in a hexagonal cell distribution. The number of users is between 150 and 800 and they are positioned based on Poisson point process. A random realization of users along with ground-BSs is presented in Fig. 3. The system parameters are presented in Table II.

Ground-BSs, new distribution of users and the 3D optimal position of the aerial-BS at  $t=100$  s is shown in Fig. 4. Fig. 5 presents the cumulative distribution function (CDF) for average SINR of users. This figure compares the CDF of average SINR for the traditional cellular system with 19 ground-BSs and our proposed system with 18 ground-BSs and supported by 1 aerial-BS. In the latter system, the position of the aerial-BS determined by Q-learning. This shows that

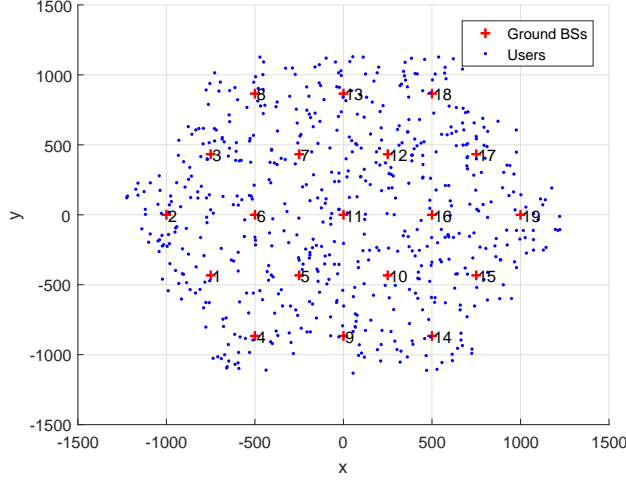


Fig. 3. 2D distribution of users and ground-BSs at the beginning of the simulation ( $t = t_{st}$ ).

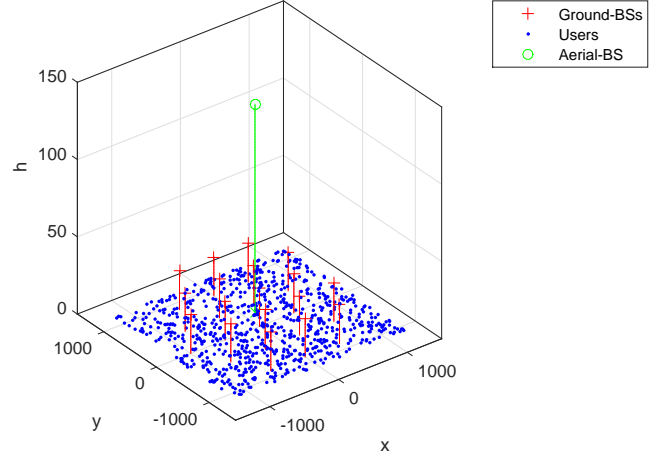


Fig. 4. 3D distribution of users, ground-BSs and optimum position of aerial-BS at  $t=100$  s.

TABLE I  
URBAN ENVIRONMENT PARAMETERS

Parameter	Value
$\kappa$	9.61
$\zeta$	0.16
$\eta_{LoS}$	1 dB
$\eta_{NLoS}$	20 dB

TABLE II  
SIMULATION PARAMETERS

Parameter	Value
$f_c$	2 GHz
$c_{max}$	1.3 m/s
$h_{min}$	25 m
$h_{max}$	525 m
Total 2D area	4 km <sup>2</sup>
$t_{min}$	10 s
$\gamma$	0.9
$\epsilon$	0.9

overall, the given method improves the SINR parameter of system. The average spectral efficiency of both systems are presented in Fig. 6. As can be seen, the system assisted by aerial-BS outperforms the traditional ground-BS system in terms of the spectral efficiency. Reward per iteration calculated from (11) for the presented wireless network with 18 ground-BSs and 1 aerial-BS is shown in Fig. 7. This figure shows the gradual improvement of the system as the reward reduces in each iteration, it also shows system approaches the optimum position after sufficient iterations and become stable as the reward becomes zero.

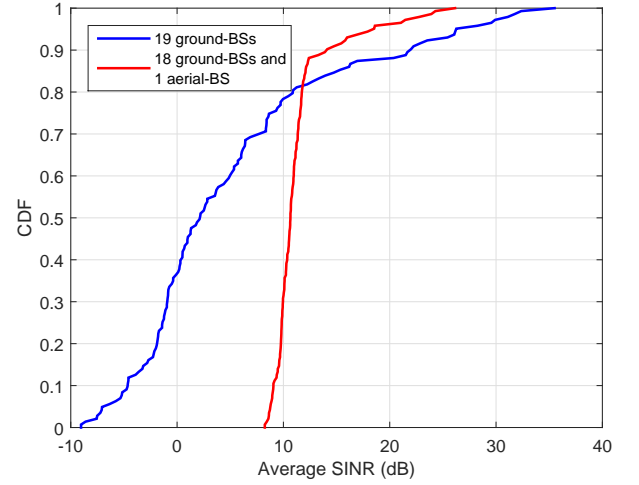


Fig. 5. CDF of average SINR of users for both the traditional ground-BSs system with 19 ground-BSs and our proposed system supported by 18 ground-BSs and 1 aerial-BS.

## V. CONCLUSION

In this paper, a traditional network of ground-BSs has been considered. To compensate the QoS loss due to user movements, the ground wireless network is assisted by an aerial-BS while one of the ground-BSs is switched off to keep the total number of BSs even and save energy. We apply Q-learning to the system as a novel approach to find the optimal position of the aerial-BS. Simulation results indicate that this method can bring much higher QoS to the network considering users' movements. By exploiting this approach, after giving the agent sufficient time to learn the environment, the processing time to find the optimum position of the aerial-BS becomes really low; therefore, it is a promising approach that can keep the agility and flexibility of future wireless networks.



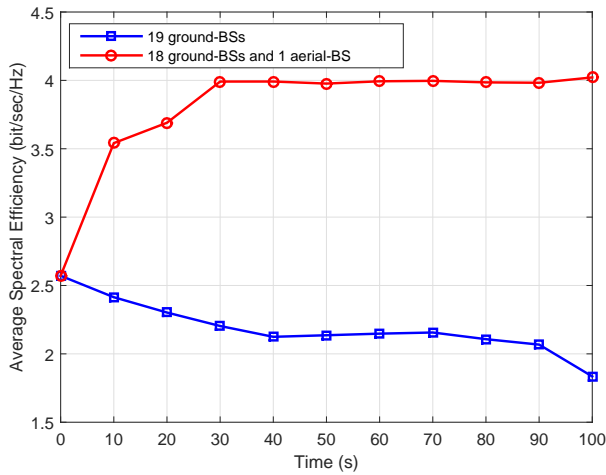


Fig. 6. The average spectral efficiency from both the traditional ground-BSs system with 19 ground-BSs and our proposed system supported by 18 ground-BSs and 1 aerial-BS.

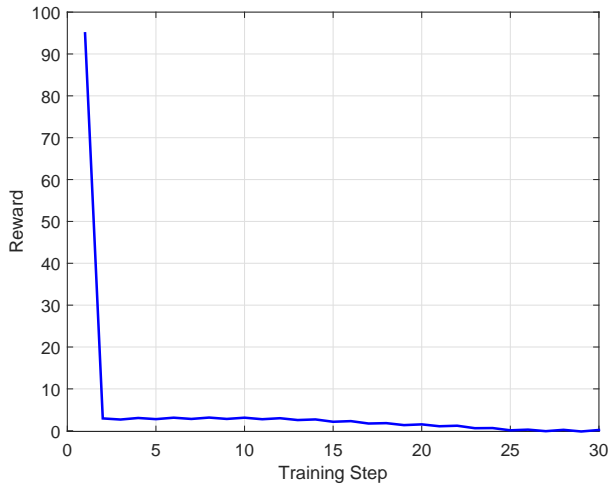


Fig. 7. Reward per iteration from solving (11), for finding the aerial-BS placement in a time slot.

## REFERENCES

- [1] S. Rohde, M. Putzke, and C. Wietfeld, "Ad-hoc self-healing of OFDMA networks using UAV-based relays," *Ad Hoc Networks*, vol. 11, no. 7, pp. 1893-1906, September 2013.
- [2] X. Li, D. Guo, H. Yin, and G. Wei, "Drone-assisted public safety wireless broadband network," in *IEEE Wireless Communication and Network Conference Workshops (WCNC 2015)*, March 2015, pp. 323-328.
- [3] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, "On the number and 3D placement of drone base stations in wireless cellular networks," in *IEEE Vehicular Technology Conference (VTC 2016-Fall)*, September 2016, pp. 1-6.
- [4] E. Kalantari, M. Z. Shakir, H. Yanikomeroglu, and A. Yongacoglu, "Backhaul-aware robust 3D drone placement in 5G+ wireless networks," in *IEEE International Conference on Communications (ICC 2017)*, May 2017, pp. 1-6.
- [5] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *IEEE International Conference on Communications (ICC 2016)*, May 2016, pp. 1-5.
- [6] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M. Alouini, "FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 218-224, January 2018.
- [7] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage," *IEEE Wireless Communication Letter*, vol. 6, no. 4, pp. 434-437, August 2017.
- [8] E. Kalantari, I. Bor-Yaliniz, A. Yongacoglu, and H. Yanikomeroglu, "User association and bandwidth allocation for terrestrial and aerial base stations with backhaul considerations," in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2017)*, October 2017. [Online]. Available: <https://arxiv.org/abs/1709.07356>.
- [9] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *IEEE Global Communications Conference (GLOBECOM 2015)*, December 2015, pp. 1-6.
- [10] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for cell association in UAV-enabled cellular networks," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2053-2056, September 2017.
- [11] S. A. W. Shah, T. Khattab, M. Z. Shakir, and M. O. Hasna, "Association of networked flying platforms with small cells for network centric 5G+ C-RAN," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2017)*, October 2017. [Online]. Available: <https://arxiv.org/abs/1707.03510>.
- [12] S. A. W. Shah, T. Khattab, M. Z. Shakir, and M. O. Hasna, "A distributed approach for networked flying platform association with small cells in 5G+ networks," in *IEEE Global Communications Conference (GLOBECOM 2017)*. [Online]. Available: <https://arxiv.org/abs/1705.03304>.
- [13] P. Zhan, K. Yu, and A. Swindlehurst, "Wireless relay communications with unmanned aerial vehicles: Performance and optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 2068-2085, July 2011.
- [14] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569-572, December 2014.
- [15] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *IEEE Global Communications Conference (GLOBECOM 2014)*, December 2014, pp. 2898-2904.
- [16] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *IEEE Wireless Communication & Mobile Computing (WCMC)*, vol. 2, no. 5, pp. 483-502, 2002.
- [17] C.J.C.H. Watkins, and P. Dayan, "Q-learning," *Springer Journal Machine Learning*, vol. 8, no. 3, pp. 279-292, 1992.
- [18] A. Forster, "Machine learning techniques applied to wireless ad hoc networks: Guide and survey," in *IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2007)*, December 2007, pp. 365-370.
- [19] C. Watkins and P. Dayan, "Q-learning," *Journal Machine Learning*, vol. 8, no. 3, pp. 279-292, 1992.
- [20] F. Yu, V. W. S. Wong, and V. C. M. Leung, "Efficient QoS provisioning for adaptive multimedia in mobile communication networks by reinforcement learning," in *International Conference on Broadband Networks (ICBN 2004)*, October 2004, pp. 579-588.
- [21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press, 1998.
- [22] V. Petrov, D. Moltchanov, and Y. Koucheryavy, "Applicability assessment of Terahertz information showers for next-generation wireless networks," in *IEEE International Conference on Communications (ICC 2016)*, May 2016, pp. 1-7.