# Learning Driven Mobility Control of Airborne Base Stations in Emergency Networks

Rui Li, Chaoyun Zhang, Paul Patras
School of Informatics
The University of Edinburgh
{rui.li, chaoyun.zhang, paul.patras}@ed.ac.uk

Razvan Stanica, Fabrice Valois
Univ Lyon, INSA Lyon, Inria, CITI,
F-69621 Villeurbanne, France
{razvan.stanica, fabrice.valois}@insa-lyon.fr

## ABSTRACT

Mobile base stations mounted on unmanned aerial vehicles (UAVs) provide viable wireless coverage solutions in challenging landscapes and conditions, where cellular/WiFi infrastructure is unavailable. Operating multiple such airborne base stations, to ensure reliable user connectivity, demands intelligent control of UAV movements, as poor signal strength and user outage can be catastrophic to mission critical scenarios. In this paper, we propose a deep reinforcement learning based solution to tackle the challenges of base stations mobility control. We design an Asynchronous Advantage Actor-Critic (A3C) algorithm that employs a custom reward function, which incorporates SINR and outage events information, and seeks to provide mobile user coverage with the highest possible signal quality. Preliminary results reveal that our solution converges after $4 \times 10^5$ steps of training, after which it outperforms a benchmark gradient-based alternative, as we attain 5dB higher median SINR during an entire test mission of 10,000 steps.

## Keywords

Emergency networks, mobility control, airborne base stations, deep reinforcement learning, AI in networks.

## 1. INTRODUCTION

Public safety and civilian operations intrinsically require stable wireless connectivity for both rescue services and post-disaster recovery. Contemporary military activities such as territorial search and emergency response also rely heavily on reliable data connections. Certain areas under extreme conditions, e.g. following earthquakes, floods, fire, and nuclear plant emergencies, are hardly accessible with legacy emergency cellular infrastructure carried on vans (i.e. cells-on-wheels). Meanwhile, following recent hardware/software advances, commercially available unmanned aerial vehicles (UAVs) are increasingly used for various applications, including aerial imaging and asset inspection. As a result, regulatory bodies, such as the Federal Aviation Administration (FAA), defined rules to enforce the safe operation of commercial UAVs [14]. The telecom industry also shows growing interest in deploying UAV-mounted base stations (BSs) for sporadic cellular services, with an emphasis on challenging use cases. For instance, following hurricane Marie's devastation of Puerto Rico, AT&T obtained FAA approval to fly UAVs for temporary cellular coverage [3]. Similarly, Altaeros rolled out SoftTower, an UAV carrying multi-sector LTE BS to provide connectivity to hard-to-reach areas [2].

Providing wireless connectivity in a large area to a sizeable group of users, including citizens and rescue teams (e.g. police force, medical personnel, or firefighters), often requires more than one airborne vehicle with networking capabilities. In contrast to cellular networks, where the BS deployment is carefully planed and conducted, BSs mounted on UAVs are mobile themselves. Coordination among flying BSs and movement control, to set up a core network, provide the needed coverage, and ensure sufficient and stable user data rates, is of paramount importance, whilst any signal failure can be catastrophic to critical missions. Stochastic wireless channels and user equipment (UE) movement uncertainty, however, render the BS mobility management task complex, involving an exponentially growing action space as the number of BSs increases. Traditional solutions, such as optimal control, require precise environment models, which are hardly obtainable in real-time and require strong assumptions that can compromise their usefulness. Heuristic alternatives only produce sub-optimal results.

In this paper, we tackle these challenges facing UAV mobility control with a deep reinforcement learning (DRL) approach. We devise a *domain-specific* reward function that encourages the UAV mobility control agent to provide high quality signal coverage to users, and we leverage an Asynchronous Advantage Actor-Critic (A3C) scheme to learn the optimal action policy via interaction with the wireless environment. Our design is motivated by the rapid convergence requirements specific to emergency settings. Simulation results demonstrate that our solution converges rapidly, and once trained, it makes accurate movement control decisions, outperforming a benchmark gradient-based scheme that has perfect knowledge of the stochastic channel. More precisely, we obtain a 5dB median SINR improvement, while only requiring current location and association information.

## 2. SYSTEM MODEL

We consider a fleet of $\mathcal{B}$ UAVs, each carrying one BS. The BSs run LTE protocol stacks with simplified functionality, e.g. disabled MME authentication, to simplify the overall architecture and prolong network lifetime. Each BS serves a number of UEs and is connected wirelessly (via satellite or $\mu$-/mm-wave links) to a central controller. The controller hosts a DRL agent that learns to make optimal decisions

*Workshop on AI in Networks (WAIN) 2018* Toulouse, France

about the BSs mobility control.

**Wireless Channel:** We consider BSs share the same frequency band, i.e. reuse factor 1. We focus on the downlink communication, assuming the SINR is directly related to the quality of service received by UEs. The transmit power employed by BS $b$ to a user is $P_b$ and we denote $G_{b,u}$ the channel gain between BS $b$ and user $u$, which is a linear combination of the free-space path-loss $l_{b,u}$, shadow fading, and antenna gain $G_a$. The log-distance path-loss $l_{b,u}$ can be computed following the 3GPP model for urban cellular scenarios with standard coefficients $\alpha$ and $\beta$, i.e. $l_{b,u} = \alpha + \beta \log(D_{b,u})$, where $D_{b,u}$ is the Euclidean distance between $b$ and $u$. Given $\mathcal{I}_b \subset \mathrm{B} \backslash b$ the set of BSs that interfere with $b$ and $N_0$, the power of per-channel additive white noise, the SINR observed by UE $u$ is:

$$\mathrm{SINR}_{b,u} = \frac{P_b G_{b,u}}{N_0 + \sum_{b' \in \mathcal{I}_b} P_{b'} G_{b',u}} \tag{1}$$

**UE Mobility and LTE Handovers:** We assume a reference group mobility model, by which users cluster around group centres that move along random way points [6]. The motivation for employing this mobility model is that, in the envisioned emergency scenario, rescue and medical teams, or fire fighters, rush towards a target scene while the population may be moving away from that location. Further, as both UEs and BSs continuously change their position, we follow the standard LTE handover policy, i.e. employ hysteresis and time-to-trigger to avoid ping-pong effects. Specifically, when the received signal strength is below a certain threshold $SINR_{th}$ for a duration of $t_{trigger}$, the user can be handed off to adjacent cells, if a new BS provides an SINR higher by $\sigma$ than the value currently measured.

## 3. REINFORCEMENT LEARNING FOR AIRBORNE BS CONTROL

We address mobility control of airborne BSs in emergency settings, considering stochastic wireless channels and user mobility, as modelled in Sec. 2. While this is a challenging problem, deep reinforcement learning (DRL) has achieved promising results in similarly complex tasks, such as Atari game play [10] and adaptive video streaming [9]. This motivates us to take a DRL approach, formulating our task as Markov Decision Process (MDP), and designing an A3C based solution tailored to our target networking scenario.

**Markov Decision Process (MDP):** The BSs mobility control can be modelled as a 5-tuple MDP, $< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma >$, where $\mathcal{S}$ is the state of the environment including BSs locations, $\mathcal{L}_{\mathcal{B}}(t)$, UEs locations, $\mathcal{L}_{\mathcal{U}}(t)$, and their associations, $\alpha_{\mathcal{B},\mathcal{U}}(t)$. $\mathcal{A}$ is the action taken by each agent, i.e. the movement direction of each BS, $\mathcal{M}_b(t)$, and $\mathcal{P}$ is the state transition matrix. Precisely, the system moves from state $s$ to $s'$ following action $a$ according to $\mathcal{P}^a_{ss'} = \mathbb{P}[S_{t+1} = s' | S_t = s, A_t = a]$. $\mathcal{R}$ is the reward function, which quantifies the system performance following an action, i.e. $\mathcal{R}^a_s = \mathbb{E}[R_{t+1} | S_t = s, A_t = a]$. In our problem, this will depend on the SINR experienced by the UEs, as we detail next. $\gamma \in [0, 1]$ is the discount factor, which dictates the importance of future rewards.

A policy $\pi$ is the probability distribution of taking an action $a$ in a given state $s$, i.e. $\pi(a|s) = \mathbb{P}[A_t = a | S_t = s]$. To attain an optimal policy $\pi*$, we employ a DRL method and give an overview of the learning procedure in Fig. 1.

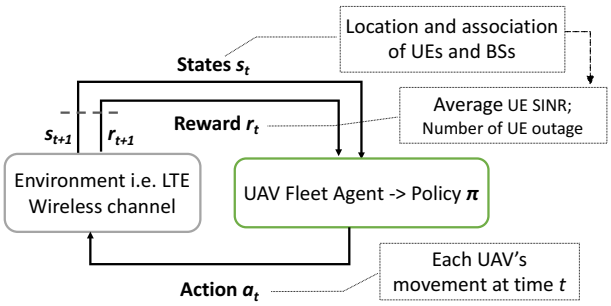


Figure 1: Overview of the learning loop for the proposed intelligent UAV mobility management agent.

As shown in the figure, the agent updates a policy $\pi$. By this, at time step $t$, given state $s_t$ including UEs and BSs locations and their associations, the agent takes an action $a_t$, according to which each drone $b \in \mathcal{B}$ moves from $(x_b, y_b, z_b)$ to $(x'_b, y'_b, z_b)$. The agent thereby receives a reward $r_t$, and the system enters a new state $s_{t+1}$. This process is repeated until the episodic return, i.e. the sum of future discounted rewards, converges. This indicates that an optimal policy was found.

**Proposed Solution:** There exist two main approaches to solving control problems through reinforcement learning (RL): policy-gradient RL methods learn by 'trial-and-error' a policy, i.e. the probability distribution of actions to take given the state of the environment; value-based methods learn to estimate the value for each action. Combining these two approaches, an actor-critic RL agent employs an 'actor' that performs actions to improve the policy, while a 'critic' makes judgements on the actor's performance and learns to estimate the action values.

To tackle the UAV mobility control problem, we propose a custom asynchronous advantage actor-critic (A3C) algorithm. A3C is a state-of-the-art actor-critic method that exploits multi-threading to create several learning agents, each exploring the state space in their own environment and updating periodically a global neural network with learned knowledge [11]. The key advantages of this approach are that it can be trained on a CPU, it de-correlates past experiences gained by each learning agent, and it converges rapidly. Such paralleled learning is a viable alternative to experience replay (i.e. Deep Q-Learning), as it removes large memory requirements.

In our A3C-based solution, we employ two neural networks that have the most simple deep learning architecture, i.e. the multi-layer perceptron (MLP). MLPs have fully connected layers and the neurons in the hidden layers implement non-linear activation functions of the output of the previous layer. The parameters (weights and biases) of these functions are obtained by training through backpropagation [7]. In our case, one of the MLPs acts as the actor and the other as the critic. Both of them employ 2 hidden layers, each consisting of 200 neurons.

We design a reward function that captures the specifics of mobile users served by airborne BSs, aiming to ensure best connectivity, i.e. highest SINR and lowest outage likelihood. Hence, the proposed reward function is

$$R = \theta * \overline{\mathrm{SINR}} - \frac{N_{out}}{N_{UE}}, \tag{2}$$

where $\overline{\text{SINR}}$ is the mean SINR computed across all the $N_{UE}$ users, $\theta$ denotes a normalising factor, and $N_{out}$ is the number of UEs whose SINR is below a minimum service requirement.

In what follows we discuss how we train the proposed algorithm and present the results of the performance evaluation campaign conducted.

# 4. EVALUATION

To evaluate the proposed A3C algorithm in UAV mobility control scenarios, we examine its convergence properties and compare its performance with that of an SINR gradient based benchmark.

## 4.1 Simulation Setup

We consider a 100×100 grid area, with a grid cell width of 5m. Within this area, 40 users move in groups of 10 UEs, following the group reference model [6], and are provided with connectivity by BSs mounted on UAVs. For each user we compute the SINR of the link to the serving BS, using the model described in Sec. 2. A sample SINR heatmap and UE locations are illustrated on the left in Fig. 2. The histogram on the right shows the distribution of the SINR experienced by users in this instance. The UAVs move at a fixed altitude, i.e. 10m from the ground level. The movement of a BS at each time step is chosen between 4 candidate directions (i.e. N, S, W, E) towards adjacent points on the grid, or idling (no move).

We train the DRL model with 10 random seeds, each time with 1,000 training episodes. A single episode lasts 2,000 time steps and the locations the UAVs are reset to the same coordinates at beginning of each episode. At each discrete time step, the mobility control agent performs actions based on the current learning policy, and chooses from $5^4 = 625$ possible actions to take (4 directions or movement plus idling, 4 UAVs). The simulation parameters used are summarised in Table 1.

Table 1: Simulation Parameters

| | Parameter | Value |
|---|---|---|
| Wireless Channel | BS Transmit power | 20 dBm |
| | Antenna gain | 2 dB |
| | Log normal shadowing | $\mathcal{N}(0, 2)$ |
| | Gaussian noise | -121 dBm |
| | Handover time-to-trigger | 3 |
| | Handover threshold | 1 dB |
| | Minimum SINR | -5 dBm |
| Learning | Learning rate | 0.0001 |
| | Discount factor | 0.9 |
| | Number of A3C workers | 4 |
| | Global update step | 10 |
| | Normalising factor $\theta$ | 0.05 |

The model is trained and tested on a 8-core desktop with Intel Xeon W-2125 CPU clocked at 4.00GHz, and we employ Python libraries TensorFlow to implement the neural networks [1].

**Benchmark:** To assess the performance of the proposed DRL solution, we devise a benchmark SINR gradient based method, and test it with the same network settings as with the proposed DRL approach. At each time step, this benchmark computes the average SINR at the associated UEs along each of the possible directions of movement. It then moves that UAV in the direction of the lowest average SINR. By this approach, the aim is to avoid outage while maintaining good signal quality for all the UEs served.

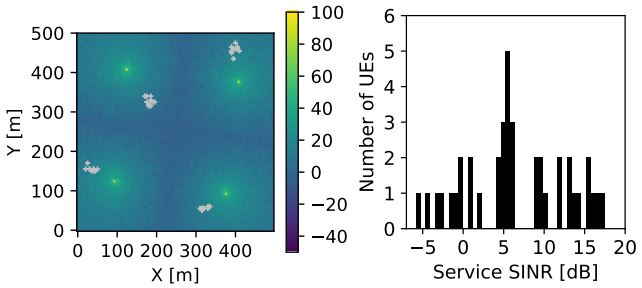


Figure 2: Left: SINR heatmap and UEs location (grey dots). Right: The corresponding distribution of UEs' experienced SINR.

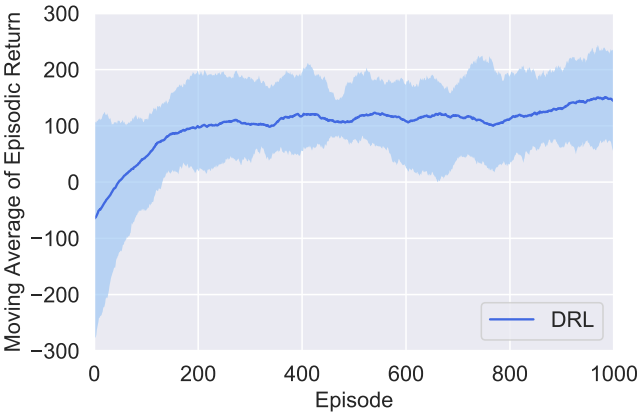


Figure 3: Moving average of the episodic return during 1,000 training episodes with different random seeds. Solid line represents the mean; shaded area represents the region between maximum and minimum return.

## 4.2 Simulation Results

**Training Convergence:** At the end of each episode, we compute the moving average of the episodic return $\mathcal{R}$ as:

$$\mathcal{R} = 0.01 * r_{ep} + 0.99\mathcal{R}, \quad (3)$$

where $r_{ep}$ is the total reward of the episode. To examine the learning convergence, in Fig. 3 we plot the evolution of $\mathcal{R}$. The training converges within 200 episodes (i.e. $4 * 10^5$ steps), corresponding to approximately 2.5 hrs of training in real world. Such training can be performed once during pre-deployment stage, after which the agent can be used for multiple missions, given the wireless channel characteristics remain largely similar. Observe that during this phase the proposed A3C solution improves the average episodic return from around -50 to 100, with the minimum value being improved from -280 to 80. Once trained, the agent can be used directly to make decisions about BSs movement, according to the current conditions.

**Performance:** Ultimately, we are interested in quantifying the performance gains our DRL approach can attain over other solutions, such as the SINR gradient-based benchmark considered. To this end, we test the trained neural network model over 10,000 steps, resetting the environment

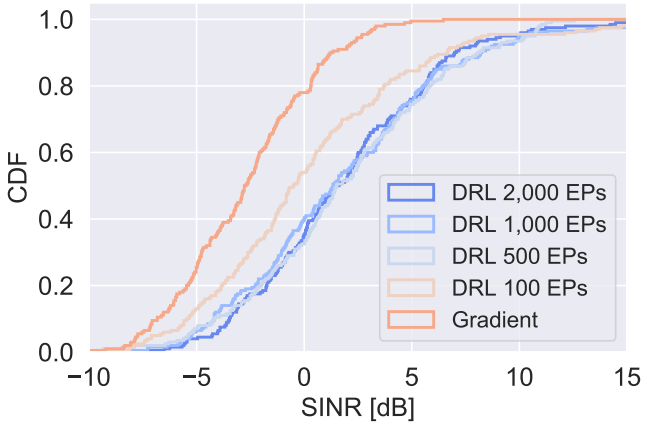


Figure 4: CDF of the SINR attained by all users with the proposed DRL over 10,000 testing steps, after 100, 500, 1,000, and 2,000 training episodes (EPs), and respectively with the benchmark gradient method.

every 2,000 steps, and examine the signal quality provided by both approaches. Specifically, in Fig. 4 we plot the cumulative distribution (CDF) of the SINR experienced by all users with the benchmark scheme and the proposed DRL algorithm, studying also the impact of the number of training episodes (i.e. 100, 500, 1,000, and 2,000 episodes) on the performance achieved.

The results confirm that our DRL solution achieves a 5dB improvement of the median SINR, over the gradient-based benchmark. If considering -5dB as the signal outage level, the proposed DRL scheme only experiences approximately 5% user outage, which is one forth of that experienced by the benchmark gradient method, i.e. 20%. The results further confirm that after convergence, our learning algorithm performs stably. Specifically, the distribution of the SINR after 2,000 training episodes achieves a median value that is only marginally better than that of an algorithm trained over 500 episodes. We conclude that the proposed DRL scheme for mobility control of airborne BSs attains more than 3× higher median SINR and 4× lower outage rate compared to the gradient-based benchmark. Furthermore, our solution performs stably after training.

## 5. RELATED WORK

Fotouhi *et al.* propose a mobility model to improve spectral efficiency in drone-based BS scenarios, though consider fixed user group coverage without UE handover [4]. In [13], Oueis *et al.* provide technical overview of LTE operation for public safety. Orsino *et al.* then highlight that drones carrying radio transceivers improve network coverage and bring higher data rates to challenging locations [12]. Grossglauser and Tse study the per-user throughput in mobile ad-hoc networks and conclude that performance can be improved dramatically when BSs are mobile [5].

Employing deep learning to solve complex networking problems is becoming a hot research area [15]. For instance, Li *et al.* leveraged supervised learning to optimise utility in backhaul networks [8] and Mao *et al.* employed a reinforcement learning technique for adaptive video streaming [9]. Advances in deep reinforcement learning (DRL) such as the asynchronous actor-critic method (A3C) achieved remarkable performance in game-play applications, using only half the training time on a multi-core CPU [11]. To our knowledge, our work is the first to employ DRL for airborne BS mobility control.

## 6. CONCLUSION

In this paper, we introduced a deep reinforcement learning solution for the mobility control of a fleet of UAV-mounted base stations, with the goal of providing reliable service coverage in scenarios where wireless infrastructure is unavailable. We took a A3C approach and designed a reward function that captures the specifics of such scenarios. By means of simulation experiments of our study reveals that the proposed DRL algorithm converges fast, and achieves 5dB higher median SINR and 4× lower outage range, as compared to a gradient-based benchmark solution.

## References

[1] M. Abadi et al. 2016. TensorFlow: a system for large-scale machine learning. In *OSDI*, Vol. 16. 265–283.
[2] Altaeros. 2018. Next generation global telecom infrastructure.
[3] FAA. 2017. FAA Approves Drone to Restore Puerto Rico Cell Service. https://www.faa.gov/news/updates/?newsId=89185.
[4] A. Fotouhi et al. 2017. DroneCells: Improving 5G Spectral Efficiency using Drone-mounted Flying Base Stations. *CoRR* abs/1707.02041 (2017).
[5] M. Grossglauser et al. 2002. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. Networking* 10, 4 (2002), 477–486.
[6] X. Hong et al. 1999. A group mobility model for ad hoc wireless networks. In *ACM MSWiM*. 53–60.
[7] Y. LeCun et al. 2015. Deep learning. *nature* 521, 7553 (2015), 436.
[8] R. Li et al. 2018. DELMU: A Deep Learning Approach to Maximising the Utility of Virtualised Millimetre-Wave Backhauls. *arXiv preprint arXiv:1810.00356* (2018).
[9] H. Mao et al. 2017. Neural adaptive video streaming with pensieve. In *SIGCOMM*. ACM, 197–210.
[10] V. Mnih et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529.
[11] V. Mnih et al. 2016. Asynchronous methods for deep reinforcement learning. In *ICML*. 1928–1937.
[12] A. Orsino et al. 2017. Effects of Heterogeneous Mobility on D2D-and Drone-Assisted Mission-Critical MTC in 5G. *IEEE Comms. Mag.* 55, 2 (2017), 79–87.
[13] J. Oueis et al. 2017. Overview of LTE Isolated E-UTRAN Operation for Public Safety. *IEEE Comms. Standards Magazine* 1, 2 (2017).
[14] RCR Wireless. 2018. Drone test trials to shape FAA rules. https://www.rcrwireless.com/20180424/wireless/176063/drone_test_trials_to_shape_FAA_rules_tag41.
[15] C. Zhang et al. 2018. Deep Learning in Mobile and Wireless Networking: A Survey. *arXiv preprint arXiv:1803.04311* (2018).