Federated Reinforcement Learning UAV Trajectory Design for Fast Localization of Ground Users

Arzhang Shahbazi*, Igor Donevski†, Jimmy Jessen Nielsen†, Marco Di Renzo*
*Laboratoire des Signaux et Systèmes, University of Paris-Saclay, CNRS, CentraleSupélec, Gif-Sur-Yvette, France
†Department of Electronic Systems, Aalborg University, Aalborg, Denmark
arzhang.shahbazi@centralesupelec.fr

***Abstract*—In this paper, we study the localization of ground users by utilizing unmanned aerial vehicles (UAVs) as aerial anchors. Specifically, we introduce a novel localization framework based on Federated Learning (FL) and Reinforcement Learning (RL). In contrast to the existing literature, our scenario includes multiple UAVs learning the trajectory in different environment settings which results in faster convergence of RL model for minimum localization error. Furthermore, to evaluate the learned trajectory from the aggregated model, we test the trained RL agent in an alternative environment which shows the improvement over the localization error and convergence speed. Simulation results show that our proposed framework outperforms a model trained with transfer learning by $\%30$.**

***Index Terms*—Unmanned aerial vehicle (UAV), localization, received signal strength (RSS), Reinforcement learning (RL), Federated Learning (FL).**

I. Introduction

In recent years, location-aware services have been recognized as a crucial component for broad applications in wireless communication. Generally, information regarding the location of objects can be exploited in different layers, from communication aided purposes to the application level where location information is desired to interpret the collected data [1]. For this purpose, the global positioning system (GPS) grants a suitable performance for outdoor applications. However, GPS is known of its expensive cost and vulnerability to jamming. Thus, alternative localization approaches have become more attractive for research focus over the past decade. In the literature, there are several ground anchor based localization techniques that have been broadly studied [2]. Specifically, the Received Signal Strength (RSS) technique is favorable because of its inherent simplicity and low complexity. This simplicity is due to the fact that RSS can be used without any modification to current systems, so it is the easiest way forward. Moreover, RSS based localization can achieve satisfactory performance in emergency situations [3]. Nonetheless, the variation around the mean signal power due to shadowing significantly impacts the reliability of this technique. This is especially important in urban and high urban environments where the shadowing effect is more severe and hence the localization accuracy drops significantly. To address this issue, unmanned aerial vehicles (UAVs) deployed as aerial anchors is an emerging solution in order to localize ground devices. The aerial anchors potentially are capable of resolving the main drawback of ground node localization when using RSS technique. In fact, UAV anchors can combine the benefits of satellites with a higher link probability of LoS and the advantages of ground anchors with a short link length and hence higher RSS resolution. Furthermore, UAVs are typically battery-limited which introduces an important challenge towards their deployment as aerial anchors. This fact restricts UAVs operational lifetime and hence reduces the number of measurements that can be collected during their mission, which can negatively affects the accuracy of localization. In fact, depending on the hovering duration, speed of the UAV, and length of the path, the energy consumption of the UAV varies.

The noteworthy success of Machine Learning (ML) is mainly associated to two key components – highly powerful computing and extremely efficient data analytic. However, such a impressive success in ML essentially relies on whether or not there are enough data to support ML algorithms so as to make them work convincingly, in which it becomes a crucial issue in many ML applications. Because of the proliferation of UAVs, collecting data through them becomes much practical and convenient such that a UAV anchor has gradually been a vast live database abounding with real-time information, which can be utilized by ML to optimize network operations and organization. It has become an important issue to appropriately and effectively use ML techniques based on data distributed over a massive mobile network. Specifically, when transporting raw data from all UAVs to a server in a huge network due to the many issues, such as network congestion, energy consumption, privacy, security, etc. To avoid transporting a huge amount of distributed data to a server for conducting centralized ML and to preserve the privacy of users, a distributed learning methodology without raw data transportation, such as federated learning (FL) [4], becomes a viable solution.

In this paper, we introduce a novel framework for ground users (GUs) localization in urban environments using UAVs. Our proposed framework incorporate reinforcement learning with federated learning which enables us to explore the optimal trajectory of the UAVs for maximum localization accuracy for different types of propagation environments. First, by for-

This work was supported by the European Commission through the H2020 PAINLESS Project under Grant 812991 and the H2020 ARIADNE Project under Grant 871464.

mulating the problem we investigate the paths that UAVs take for for minimum localization error for three environments with different parameters which impact the path loss and accuracy of localization. By utilizing federated learning technique we aggregate these models and finally we test the trained model in fourth environment. Our results show that the localization error achieved with same number of training episodes is %30 lower with trained FL model from three environment as compared to the model transferred sequentially from first environment to fourth environment.

The rest of this paper is organized as follows. In Section II, we introduce the system model and the path loss model for localization based on RSS. Then, the machine learning framework for UAVs is introduced in Section III. In Section IV the simulation results are presented. Finally, the work is concluded in Section V.

II. SYSTEM MODEL

In this paper, we assume multiple UAVs flying over an urban area at a fixed altitude h, operating as an aerial anchors to localize multiple terrestrial users. These devices are equipped with a wireless communication device which periodically broadcast a probe request. We resort to utilizing the following log-normal shadowing pathloss model as it is capable of modeling wireless environments with acceptable precision [5]. We formulate the path loss as:

$$L = 20\log(d) + 20\log(\frac{4\pi f}{c}) + A_\tau(\theta) \tag{1}$$

where d is the distance between the UAV and ground user, f and c are respectively the system frequency and speed of light, and $A_\tau(\theta)$ is a log-normal distributed random variable with mean μ_τ and variance $\sigma_\tau^2(\theta)$, i.e.,

$$A_\tau(\theta) \sim \mathcal{N}(\mu_\tau, \sigma_\tau^2(\theta)) \tag{2}$$

where the variance can be defined as:

$$\sigma^2(\theta) = \mathbb{P}_{LoS}^2(\theta)\sigma_\tau^2(\theta) + [1 - \mathbb{P}_{LoS}^2(\theta)]\sigma_\tau^2(\theta) \tag{3}$$

where $\sigma_\tau(\theta)$ corresponds to the shadowing effect of LoS and NLoS links between the UAV and the ground user, where $\tau = \{0, 1\}$ is an indicator that can have value 1 for LoS link and 0 for NLoS link, and they are expressed as:

$$\sigma_\tau(\theta) = d_\tau \exp(-c_\tau\theta\frac{180}{\pi}) \tag{4}$$

and $\mathbb{P}_{LoS}(\theta)$ is the probability of having LoS link, which is written as:

$$\mathbb{P}_{LoS}(\theta) = \frac{1}{1 + a\exp(-b(\theta\frac{180}{\pi} - a))} \tag{5}$$

where a, b, c_τ, d_τ and μ_τ are environment dependent parameters. Thus, the distance between the UAV and the device can be estimated as follows:

$$d = 10^\zeta \tag{6}$$

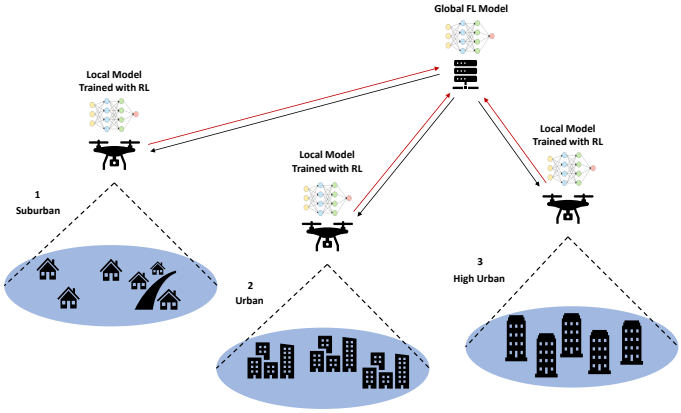


Fig. 1: Federated learning architecture.

$$\zeta = \frac{P_t - P_r - 20\log(\frac{4\pi f}{c}) - A_\tau(\theta)}{20} \tag{7}$$

where P_r and P_t denote the received and the transmitted power, respectively.

The position of a GU in 2D coordinates is described as (x_u, y_u). Given the projection of UAV on the ground (x, y), we can estimate the distances $r_i = \sqrt{(x - x_u)^2 + (y - y_u)^2}$ based on (7). Moreover, the multilateration technique can be utilized to estimate the user's position. In multilateration least squares are used to estimate the position of the user $(\hat{x}, \hat{y})$ according to the estimated distances. In a two-dimensional space, n_i distance measurements from n_i dissimilar positions are calculated to generate n_i circles centered at the position where the measurements are taken with radii equal to the respective measurements. If the distance measurements are accurate, the n_i circles intersect in one point that establish the position of the user. Now, given (x_i, y_i) the ground position of the UAV at sample point i, and $\hat{r}_i$ be the distance from sample point i to the middle of overlapping circles, then we can estimate the location $(\hat{x}, \hat{y})$ using N number of samples from the following minimization formula:

$$(\hat{x}, \hat{y}) = \min_{\hat{x},\hat{y}} \sum_{i=1}^{N} \sqrt{(x_{u_i} - \hat{x})^2 + (y_{u_i} - \hat{y})^2 - r_i} \tag{8}$$

III. PROPOSED METHOD

A. Q-Learning

To solve the problem explained in the previous section, we resort to a reinforcement learning framework based on double Q-learning. Compared to the existing reinforcement learning algorithms such as Q-learning that may leads to a suboptimal trajectory, the double Q-learning algorithm permit the UAV to find the optimal flying trajectory to minimize the localization error of all users. Furthermore, in comparison with the traditional Q-learning algorithm that generally uses one Q-table to record and update the values coming from different states and actions [6], the double Q-learning algorithm uses two Q-tables to separately select and evaluate the actions. Consequently, the double Q-learning algorithm prevent the

Algorithm 1: Federated averaging with DDQN.

Execution on Server:
Initialize w_0
for $j = 1$ to max_rounds do
  M = set of UAVs
  for Each UAV in parallel do
    $w_{t+1}^k = ClientUpdate(k, w_t)$
  end for
  $w_{t+1} = \sum \frac{1}{M} w_{t+1}^k$
end for
Return w_{t+1} to UAVs.

Execution on UAV:
Construct reward function R
Init: UAV position, s, $Q^i \quad i \in [A, B]$
Repeat
  if local_rounds $<$ max (local_rounds)
    Choose action:
    $a = argmax_a Q^i(s, a)$ from $Q^i(i \in [A, B])$
    Receive immediate reward
    Update table Q^i

overestimation of Q values. Next, we introduce the components of the double Q-learning algorithm. We utilize a RL framework modeled as a Markov Decision Process (MDP) to solve the localization problem. Each UAV independently make decisions with respect to standard MDP representation as described in the following:

1) **State Representation**: Each state considers the agent's location, represented by the UAV (x, y) coordinates in the trajectory taken, the localization error and estimated distances calculated by RSS signals explained in Section I.
2) **Action Space**: The action space is defined by movement directions on the sides of the hexagon as the only possible velocity vectors plus the action of remaining in the same place formatted into a 7-tuple.
3) **State Transition Model**: Considering a deterministic MDP, there is no randomness in the transitions that follow the agent's decisions. Thus, the next state is only affected by the action that the agent takes.
4) **Rewards**: The reward function is defined by the average localization error from the ground users at each step,

$$r[n] = \frac{L_s}{e[n]} \tag{9}$$

where L_s is desired localization error which is set to $10m$ and $e[n]$ is the evaluated localization error at time instant n.

B. Federated learning

In the UAV network proposed in Section II, our aim is to investigate the performance of FL over the UAV network that localize ground users via RSS reading, which lead to continuous FL between the edge server and the UAVs. Thus, we propose a FL model over the network in Fig. 1 as follows. Suppose there are 3 UAVs distributed in the network and their task is to jointly learn a global model with the edge server in T training rounds. To characterize the impact of different environment parameters on localization error, we assume each UAV is operating in a different environment setting i.e from sub urban to high urban.

Federated averaging (FedAvg) orchestrates training with a central server which hosts the shared global model w_t, where t is the communication round. The algorithm initialize by randomly setting the global model w_0. One communication round of FedAvg can be described in the following: At the beginning, the server distributes the current global model w_t to all UAVs. After updating their local models w_t^k to the shared model,$w_t^k \leftarrow w_t$, each UAV partitions its local data into batches and performs epochs of Stochastic Gradient Decent (SGD). Finally, UAVs upload their trained local models w_{t+1}^k to the server, which then generates the new global model w_{t+1} by computing a weighted sum of all received local models. Our approach for utilizing FedAvg reinforcement learning for localization is represented in Algorithm 1.

Table I: The path loss parameters for: Suburban (1), Urban (2), Dense Urban (3) and Highrise Urban (4) environments [7].

	a	b	μ_1	μ_0	d_1	d_0	c_1	c_0
env_1	4.88	0.43	0.1	21.0	11.25	32.17	0.06	0.03
env_2	9.61	0.16	1.0	20.0	10.39	29.6	0.05	0.03
env_3	12.08	0.11	1.6	23.0	8.96	35.97	0.04	0.04
env_4	14.32	0.08	2.3	34	7.37	37.08	0.03	0.03

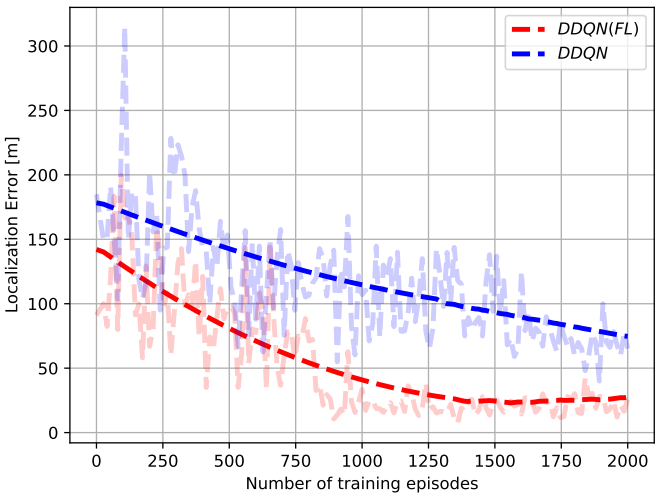
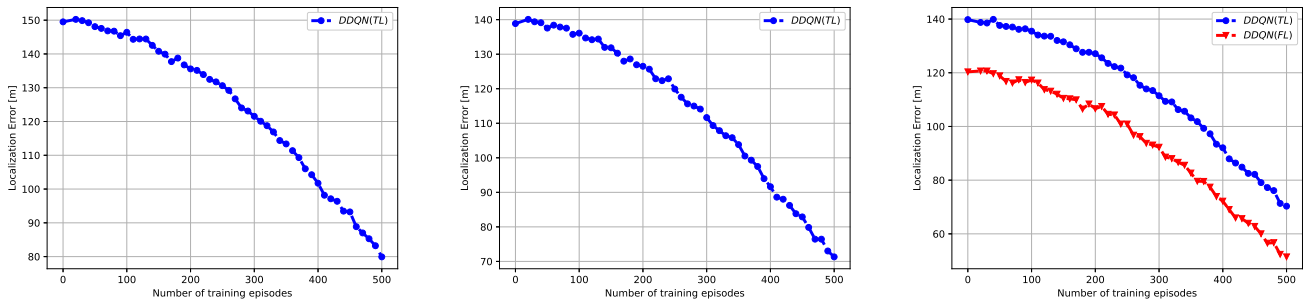


Fig. 2: Localization error versus training episodes in env_1. Comparison between FL model and baseline DDQN.

IV. Simulation Results

We assume N GUs uniformly distributed in a circular area with a radius of 750m, centered at $(x, y) = (0, 0)$. The values for the path loss model considered in this paper are



(a) env_2 (b) env_3 (c) env_4

Fig. 3: Localization error versus training episodes (a) pre-trained model from env_1 transferred and retrained in env_2; (b) pre-trained model from env_1 and env_2 transferred and retrained in env_3; (c) test and comparison of model transferred from previous environments and FL architecture in env_4.

chosen as recommended in [7] for urban environments and are summarized in Table I. We assume all UAVs are flying at a fixed altitude and their position is known by GPS, and they can measure the RSSI from all users in their communication range. We resort to Python as a programming language to simulate the operation of our proposed method, and the numerical results are averaged over ten runs.

Fig. 2 shows the convergence of the proposed FL method. From Fig. 2 we observe that the FL algorithm required approximately 1300 episodes to reach convergence, which is much less than the number of episodes required for convergence of the DDQN. Fig. 2 also shows that the FL algorithm achieves the localization error of $25m$ after only 1200 episodes, which is about $\%75$ lower that the one reached by the DDQN baseline. This stems from the fact that the FL algorithm has already trained a set of weights from training in 3 environments and starts the training process with a pre-trained model.

In Fig. 3, we test the performance of the FL trained model from 3 urban environment on the 4th environment with a scenario when Transfer Learning (TL) is applied to transfer the model RL agent trained in one environment to next environment. Transfer learning aims at improving the process of learning new tasks using the experience gained by solving predecessor problems which are somewhat similar. Fig. 3 shows the results obtained in the scenario when considering different training options for the DDQN algorithm: in Fig. 3 (a) a training of $N_e = 500$ is done in the environment 2 on the basis of pre-trained model in environment 1; followed by a training of $N_e = 500$ in the environment 3 based on the transferred model from environment 2 Fig. 3 (b), and finally the agent is trained with $N_e = 500$ episodes in environment 4 based on the pre-trained model from previous environments and also $N_e = 500$ episodes is training with FL pre-trained model from environment $1-3$, Fig. 3 (c). As we can see the localization error achieved with 500 episodes of training in the 4th environment with the pre-trained model from transfer learning is approximately equal to $70m$, while with 500 episode, the FL pre-trained model reaches the localization error of $50m$. This result shows that our proposed framework is efficient in reducing convergence speed by $\%30$ and achieving better generalization performance in comparison with transfer learning approach.

V. CONCLUSION

The enhancement in localization accuracy of ground users when using UAV as base station and relying on RSS techniques has been studied. Specially, we utilized a FL framework to find an optimal trajectory through training an agent with RL algorithm which reached convergence faster. This paper validated the effectiveness of placing anchors at different position with respect to different environment setting in terms of both localization error and the required number of episodes for training an RL agent. Finally, the reported results motivate inspecting other localization methods, such as angle-of-arrival, and possibly integrate them with the proposed FL-based framework for further improvements.

REFERENCES

[1] A. Dammann, G. Agapiou, J. Bastos, L. Brunelk, M. Garcia, J. Guillet, Y. Ma, J. Ma, J. J. Nielsen, L. Ping, *et al.*, "Where2 location aided communications," in *European Wireless 2013; 19th European Wireless Conference*, pp. 1–8, VDE, 2013.

[2] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.

[3] D. Ebrahimi, S. Sharafeddine, P.-H. Ho, and C. Assi, "Autonomous uav trajectory for localizing ground objects: A reinforcement learning approach," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1312–1324, 2020.

[4] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.

[5] 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, 2014.

[6] U. Challita, A. Ferdowsi, M. Chen, and W. Saad, "Machine learning for wireless connectivity and security of cellular-connected uavs," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 28–35, 2019.

[7] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE global communications conference*, pp. 2898–2904, IEEE, 2014.