Reinforcement Learning Based Non-Cooperative Radio Localization Assisted by UAV Swarm

Haoyu Chen*

*Institute of Artificial Intelligence, Xiamen University.

***Abstract*—The accuracy of non-cooperative radio localization aided by unmanned aerial vehicles (UAVs) depends on the swarm size and the spatial distribution. In this paper, we propose a non-cooperative radio localization scheme that chooses the cooperative UAVs based on the received signal strength (RSS) of the radio device, previous shared RSS from neighboring UAVs, and the previous localization error to increase the accuracy with reduced energy consumption. Based on the weighted least squares approach, the position of the radio device is estimated to evaluate the localization error to formulate the cooperative UAVs selection distribution with increased accuracy. Simulations based on the ten UAVs show that the proposed scheme decreases the root-mean-square error and energy consumption over the benchmark.**

***Index Terms*—Radio localization, unmanned aerial vehicles, received signal strength.**

I. Introduction

The unmanned aerial vehicle (UAV) swarm enables the localization of the non-cooperative radio device based on received signal strength (RSS) or angle of arrival (AOA) without any direct collaboration of feedback signaling from the radio device to support the safety applications, such as jammer and unauthorized transmitter localization [1]–[5]. The localization accuracy increases with the number of deployed UAVs at the expense of the communication costs.

The RSS or AOA measurements estimated by UAVs are fused to compute the position of the radio device using range-based and hypothesis-driven estimators, such as weighted least squares and maximum a posteriori algorithms [2], [9], followed by search strategies, such as convex relaxation and greedy method, to determine cooperative UAVs selection with increased accuracy. For example, the localization scheme named MAPL chooses the set of anchors based on the RSS of the radio device according to the maximum a posteriori based greedy algorithm to improve the accuracy. However, the variations of RSS from the moving radio device caused by signal decay under large-scale UAV networks and thus degrade the location accuracy.

The shared RSS among the cooperative UAVs indicates the features of the signal attenuation in the different spatial regions caused by the propagation effects such as path-loss and small-scale fading, which can be exploited to predict the possible area of the radio devices [1]. For example, the UAV near the moving radio device generally measures higher RSS compared with those situated at greater distances due to the shadowing effect and the radio signal decays.

In this paper, we propose a reinforcement learning (RL) based non-cooperative radio device localization scheme to optimize the cooperative UAVs selection, which can be regarded as a partially observable Markov decision process. Based on the state including the current RSS measurement from the radio device, previous shared RSS from the cooperative UAVs, and localization error, the proposed swarm-aided radio localization scheme chooses the number and subset of cooperative UAVs to balance accuracy with communication energy consumption.

The RSS measurements fused from selected agents are jointly applied for radio device localization based on the weight least squares method to improve the accuracy. The weights of selected UAVs are introduced to emphasize the reliability of the range observations since short-range measurements tend to result in higher localization accuracy than long-range ones. The localization error is utilized in risk level evaluation to avoid the selection of cooperative UAVs that compromise the localization accuracy. Simulation results with 10 UAVs show that our proposed localization scheme can strategically choose cooperative agents to reduce 41.67% root-mean-square error compared with the benchmark MAPL in [2].

The rest of this paper is organized as follows. First, the related work is reviewed in Section II and the system model presented in Section III. The proposed RL-based radio device localization scheme is proposed in Section IV, followed by the performance analysis in section V. The simulation results are presented and analyzed in Section VI and the conclusion is drawn in Section VII.

II. Related Work

The range-based radio localization scheme usually estimates the distance between the radio device and the UAVs based on the RSS, AOA, time of arrival, and time difference of arrival. These techniques involve inherent trade-offs between localization accuracy and implementation complexity, where the RSS-based approaches offer cost-effective and easy-implementation solutions [1], [2], [10]. For example, the scheme in [10] used RSS values and radio map estimations to achieve about 5 meters of accuracy in mean absolute

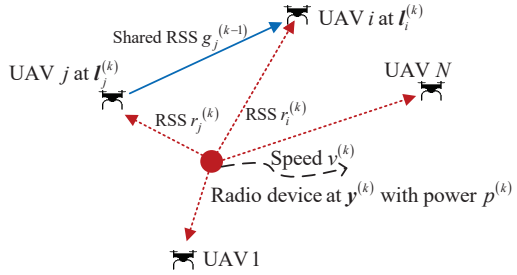


Fig. 1. Illustration of the non-cooperative radio device localization, in which UAV i measures the RSS $r_i^{(k)}$ from the radio device and obtain shared RSS from UAV j.

error since distance estimations were obtained from a path loss radio map rather than a statistical channel model.

The transmit power of the radio device and path-loss exponent are crucial parameters that are assumed to be known by the anchors to achieve acceptable performance for the RSS-based scheme, and their dynamic fluctuations can lead to imprecise distance estimations [11], [12]. For example, a random variable that represents the non-line-of-sight (NLOS) path loss is introduced to the robust weighted least squares algorithm to estimate the position and transmit power of the radio device and the path-loss exponent in the mixed LOS/NLOS environments [11]. Moreover, the scheme in [12] localizes the device with anchor position uncertainties and unknown transmit power, where the semidefinite programming method is applied to relax the maximum likelihood estimator.

Compared to the RSS based methods, the AOA based schemes use azimuth and elevation angles from transmit signal from spatially distributed anchors to achieve higher accuracy at the expense of hardware complexity [4], [8], [13], [14]. For example, the stationary sensors equipped with antenna arrays measure the AOA to localize the moving object with constant velocity, in which bias reduced constrained weighted least squares problem is solved by semidefinite programming method [13].

III. System Model

A. Network Model

The swarm exchanges the spectrum information such as the RSS estimated by each UAV to support applications such as spectrum awareness in space-time. As shown in Fig. 1, at time slot k, UAV $i \in \mathcal{N} = \{1, 2, ..., N\}$ at position $\boldsymbol{l}_i^{(k)}$ is cruising over an urban area to localize moving radio device at position $\boldsymbol{y}^{(k)}$ and speed $v^{(k)} \in [0, V_m]$. Each UAV equipped with a sensor to measure the RSS denoted by $\hat{r}_i^{(k)}$ that depends on the path loss model from the radio device with power $p^{(k)}$ on communication channel $f_i^{(k)} \in \{1, 2, ..., B\}$, given by

$$\hat{r}_i^{(k)} = p^{(k)} - 10\gamma \log_{10} d_i{}^{(k)} + \psi + n_i^{(k)}, \tag{1}$$

where $d_i^{(k)}$ denotes the distance between the radio device and UAV i, ψ is the constant that related to the communication channel $f_i^{(k)}$ and speed of light, γ is the path loss exponent, and $n_i^{(k)} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ denotes the lognormal shadowing term. According to [16], the variance of $n_i^{(k)}$ is given by

$$\sigma_i^2 = \varpi^2 \sigma_1^2 + (1 - \varpi)^2 \sigma_2^2, \tag{2}$$

where ϖ represents the probability of LoS link, and $\left[\sigma_m^2\right]_{m \in \{1,2\}}$ denotes the shadowing effect of NLoS and LoS links between the radio device and the UAV, respectively. As a non-cooperative node, the transmit power $p^{(k)}$ of radio device is not known in advance in (1).

B. Communication Model

The swarm works in a collaborative manner for accurate and efficient radio localization. More specifically, each UAV broadcasts the localization requests that contain its position and destinations to the neighbors with power $p_u^{(k)} \in \left[\underline{P}, \overline{P}\right]$. Once receiving the localization requests, UAV j detects the signal strength of each radio channel for collision avoidance and broadcasts the request to send (RTS) message that contains its current position $\boldsymbol{l}_j^{(k)}$ and the intended destinations. UAV j will wait for clear to send (CTS) messages from the neighbors and starts to send the data packet including the estimated RSS $\hat{r}_j^{(k)}$ whose initial part is regarded as the confirmation for the CTS messages. To prevent interference from hidden terminals, UAV j transmits a busy tone and waits for ACK from its neighbors within a given time.

Upon detecting the transmission of the message and receiving a valid RTS, each neighbor of the UAV j decodes the information and determines its priority as the relay. More specifically, each neighbor obtains its priority according to the relative distance to the intended destination. Finally, the selected relay activates the busy tone on the control channel for a duration T_R, sends a CTS after the RTS, and decodes the shared information from the UAV j. UAV i waits for a valid RTS message from one of the neighbors and sends the CTS message. Finally, RSS $\hat{r}_j^{(k)}$ from UAV j is decoded and applied for localization. For convenience, the time index k in the superscript is omitted if no ambiguity occurs.

IV. RL-Based Localization Approach

We propose an RL-based non-cooperative radio device localization scheme named RLDL that optimizes the cooperative UAV selection with reduced localization error and energy consumption. According to the state consisting of the RSS, previous shared RSS from the cooperative agents, and previous localization error, the localization policy is selected to enhance the utility as the weighted sum of the localization error, energy consumption, and localization variance. The

state formulation of shared RSS comes from the fact that the UAVs generally receive a higher power value than the one that is away from the radio device, which confines possible area of the moving radio device. The risk value of localization error is utilized to formulate the cooperative UAVs selection to avoid choosing the UAVs with large distance estimation or on the same co-planar.

The cooperative UAVs are selected based on the RSS $r_i^{(k)}$ from the radio device, previous shared RSS $\boldsymbol{g}_i^{(k-1)}$, previous localization error $\rho_i^{(k)}$, the state of UAV i is formulated as

$$\boldsymbol{s}^{(k)} = \left[r_i, \boldsymbol{g}_i, \rho_i\right]. \tag{3}$$

The two-level hierarchical RL decomposes the localization policy into two sub-policies to reduce the state and action spaces that increase exponentially with the selected number of UAVs for fast learning. Specifically, according to the current state $\boldsymbol{s}^{(k)}$, the long-term reward and risk level, i.e., Q value and the E value, are used to formulate the modified Boltzmann distribution π_1 given by (4) for each state-action pair to choose the number of selected UAVs $x^{(k)} \in \{M, M+1, ..., N\}$. With the localization state formulated by $\hat{\boldsymbol{s}}^{(k)} = \left[\boldsymbol{s}^{(k)}, x\right]$, the cooperative agents $\boldsymbol{a} \subseteq \{1, 2, \ldots, N\}, |\boldsymbol{a}| = x$ are selected based on π_2 to get a more reliable position estimation of the radio device, which is given by

$$\pi_j = \frac{\exp\left(c_j Q\left(\boldsymbol{z}^{(k)}, \boldsymbol{x}\right) - E\left(\boldsymbol{z}^{(k)}, \boldsymbol{x}\right)\right)}{\sum\limits_{\boldsymbol{x}' \in \mathcal{A}_j} \exp\left(c_j Q\left(\boldsymbol{z}^{(k)}, \boldsymbol{x}'\right) - E\left(\boldsymbol{z}^{(k)}, \boldsymbol{x}'\right)\right)}, \tag{4}$$

where $\boldsymbol{z}^{(k)} \in \left\{\boldsymbol{s}^{(k)}, \hat{\boldsymbol{s}}^{(k)}\right\}$, $\boldsymbol{x} \in \{x, \boldsymbol{a}\}$, $[c_j]_{1 \le j \le 2}$ denotes the weights for number and subset selection, respectively. The sub-policy space $\mathcal{A}_1$ in the first level of hierarchical RL includes $(N - M + 1)$ feasible actions, the sub-policy space $\mathcal{A}_2$ in the second level consists of C_N^x available actions.

Similarly to [3], upon receiving the RSS from the selected UAVs $\boldsymbol{a}, |\boldsymbol{a}| = x$, the unknown parameters including position and the transmit power of the radio device denoted by $\boldsymbol{\theta} = [\boldsymbol{y}, 10^{p/5\gamma}]^T$ are computed by the weighted least squares estimator, which is given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{A}^T \boldsymbol{w}^T \boldsymbol{w} \mathbf{A}\right)^{-1} \mathbf{A}^T \boldsymbol{w}^T \boldsymbol{w} \mathbf{b}, \tag{5}$$

where

$$\mathbf{A} = \begin{bmatrix} 2\boldsymbol{l}_1 & -1 & 10^{-\hat{r}_1/5\gamma} \\ \vdots & \vdots & \vdots \\ 2\boldsymbol{l}_m & -1 & 10^{-\hat{r}_m/5\gamma} \\ \vdots & \vdots & \vdots \\ 2\boldsymbol{l}_x & -1 & 10^{-\hat{r}_x/5\gamma} \end{bmatrix} \tag{6}$$

$$\boldsymbol{w} = \left[1 - \frac{\hat{r}_m}{\sum_{m' \in \boldsymbol{a}} \hat{r}_{m'}}\right]_{1 \le m \le x} \tag{7}$$

and $\mathbf{b} = [\boldsymbol{l}_m \boldsymbol{l}_m^T]_{1 \le m \le x}^T$. The weights $\boldsymbol{w} = [w_m]_{1 \le m \le x}$ are introduced because short-range RSS measurements are inherently more reliable than long-range ones.

Similar to [16], the localization error $\rho^{(k)}$ is computed as

$$\rho_i^{(k)} = \frac{1}{M} \sum_{m=1}^{M} w_m \left(\|\boldsymbol{l}_m - \hat{\boldsymbol{y}}_i\| - d_m\right)^2. \tag{8}$$

The variance of estimated positions $\hat{\boldsymbol{y}}_i^{(k)}, 1 \le i \le N$ denoted by $\lambda_i^{(k)}$ is used to investigate reliability of the proposed radio localization scheme, given by

$$\lambda_i^{(k)} = \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\boldsymbol{y}}_i^{(k)} - \mathbb{E}_i\left[\hat{\boldsymbol{y}}_i^{(k)}\right] \right\|^2. \tag{9}$$

The risk value is updated in the previous ν time slots via

$$E\left(\boldsymbol{s}^{(k)}, \boldsymbol{a}\right) = \sum_{l=k-\nu+1}^{k} \varphi^l \mathbf{I}\left(\rho_i^{(l)} > \omega\right), \tag{10}$$

where ω is the threshold of the risk value that depends on the localization error and φ is the risk discount factor.

To achieve accurate localization with reduced energy consumption, the utility u_i is formed by the weighted sum of the localization error ρ_i estimated by (8), transmission energy consumption e_i, and the variance λ_i calculated by (9), given by

$$u_i = -\rho_i - c_1 e_i - c_2 \lambda_i. \tag{11}$$

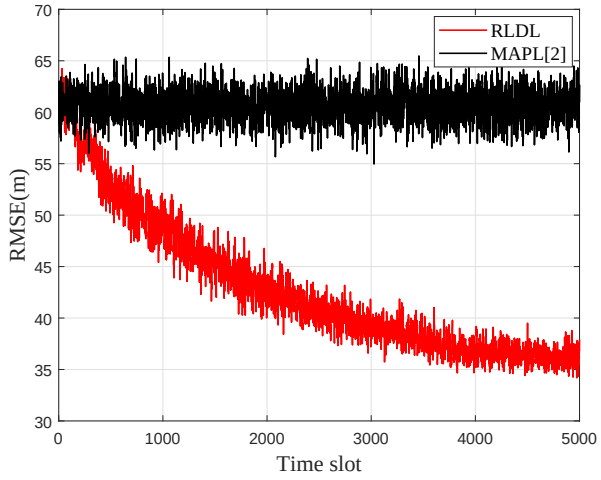
The Q-value is updated iteratively via the Bellman equation.

V. Simulation Results

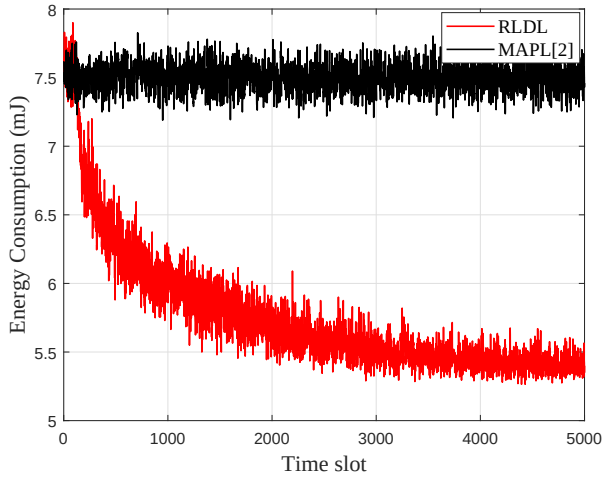
10 UAVs hover in an area of $200 \times 200 \times 10$ m^3 measure and exchange RSS from the radio device, which sends signal with power 100 mW and moves with speed 5 m/s. According to [16], the path-loss exponent is set as $\gamma = 2$, the standard deviation for the shadowing effect of LoS and NLoS links are $\sigma_1 = 1$ and $\sigma_2 = 20$, respectively. In the simulation, the learning rate $\alpha = 0.4$ and the discount factor $\beta = 0.3$. The RMSE is used to evaluate the bias of the localization between the estimated and the real position of the radio device.

As shown in Fig. 2, the proposed improve the accuracy with the reduced number of selected UAVs to localize the moving radio device. For example, the proposed RLDL decreases the RMSE from 65 to 36 with 26.67% reduced energy consumption and improve 36.70% utility after 4000 time slots. This is because the risk value is utilized in the policy selection to avoid choosing UAVs with an inappropriate geometric arrangement that reduces localization accuracy.

Our proposed RL localization scheme also outperforms the MAPL in [2]. For example, RLDL improves the localization performance with 41.67% reduced error, 26.67% energy consumption, and 36.70% improved utility after 4000 time slots. The reason is that the shared RSS in the



(a) RMSE



(b) Energy consumption

Fig. 2. Performance of the 10 UAVs aided radio localization scheme, in which the radio device moves along a trajectory and sends radio signals with power 100 mW.

state formulation enables the proposed RLDL scheme to choose the cooperative UAVs with reduced error to localize the moving radio device. However, the accrracy of MAPL scheme in [2] degrades as it is unable to choose the UAVs to improve the localization accuracy according to the variations of RSS from the moving radio device caused by signal decay.

VI. Conclusion

In this paper, we have proposed a RL based non-cooperative radio localization scheme based on current RSS, previous shared information by the cooperative UAVs, and the previously localization error to choose the cooperative UAVs to enhance the localization accuracy with reduced energy consumption. The policy distribution is formulated with safety constraints to avoid selecting the cooperative agents with high localization error. The computational complexity increases with the swarm size and the minimal number of selected UAVs. Simulation results have shown that RLDL reduces 41.67% RMSE and 26.67% energy consumption compared with the benchmark scheme MAPL in [2].

References

[1] M. Khaledi, M. Khaledi, S. Sarkar, *et al.*, "Simultaneous power-based localization of transmitters for crowdsourced spectrum monitoring," in *Proc. ACM MobiCom*, pp. 235–247, Oct. 2017.

[2] A. Bhattacharya, C. Zhan, H. Gupta, *et al.*, "Selection of sensors for efficient transmitter localization," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, pp. 2410–2419, Jul. 2020.

[3] S. Tomic, M. Beko, and R. Dinis, "3-D target localization in wireless sensor networks using RSS and AoA measurements," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3197–3210, Apr. 2017.

[4] Y. Sun, K. C. Ho, L. Gao, *et al.*, "Three dimensional source localization using arrival angles from linear arrays: Analytical investigation and optimal solution," *IEEE Trans. Signal Process.*, vol. 70, pp. 1864–1879, Mar. 2022.

[5] N. H. Nguyen, K. Doğançay, and E. E. Kuruoğlu, "An iteratively reweighted instrumental-variable estimator for robust 3-D AOA localization in impulsive noise," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4795–4808, Sept. 2019.

[6] M. S. Oh, S. Hosseinalipour, T. Kim, *et al.*, "Dynamic and robust sensor selection strategies for wireless positioning with TOA/RSSmeasurement," *IEEE Trans. Veh. Technol.*, vol. 72, no. 11, pp. 14656–14672, Nov. 2023.

[7] L. Xiao, L. J. Greenstein, and N. B. Mandayam, "Sensor-assisted localization in cellular systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4244–4248, Dec. 2007.

[8] S. Monfared, E. I. P. Copa, P. De Doncker, *et al.*, "AoA-based iterative positioning of IoT sensors with anchor selection in NLOS environments," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6211–6216, Jun. 2021.

[9] B. Beck, S. Lanh, R. Baxley, *et al.*, "Uncooperative emitter localization using signal strength in uncalibrated mobile networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7488–7500, Nov. 2017.

[10] Ç. Yapar, R. Levie, G. Kutyniok, *et al.*, "Real-time outdoor localization using radio maps: A deep learning approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9703–9717, Dec. 2023.

[11] Y. Sun, S. Yang, G. Wang, *et al.*, "Robust RSS-based source localization with unknown model parameters in mixed LOS/NLOS environments," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3926–3931, Apr. 2021.

[12] Y. Zou and H. Liu, "RSS-based target localization with unknown model parameters and sensor position errors," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6969–6982, Jul. 2021.

[13] G. Wang, P. Xiang, and K. C. Ho, "Bias reduced semidefinite relaxation method for 3-D moving object localization using AOA," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7377–7392, Nov. 2023.

[14] Y. Zou, L. Wu, J. Fan, *et al.*, "A convergent iteration method for 3-D AOA localization," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8267–8271, Jun. 2023.

[15] X. Kang, D. Wang, Y. Shao, *et al.*, "An efficient hybrid multi-station TDOA and single-station AOA localization method," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5657–5670, Aug. 2023.

[16] D. Ebrahimi, S. Sharafeddine, P.-H. Ho, *et al.*, "Autonomous UAV trajectory for localizing ground objects: A reinforcement learning approach," *IEEE Trans. Mob. Comput.*, vol. 20, no. 4, pp. 1312–1324, Apr. 2021.

[17] J. Wang, J. Chen, and D. Cabric, "Cramer-Rao Bounds for joint RSS/DoA-based primary-user localization in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1363–1375, Mar. 2013.