AI-Enabled Swarm Clustering for UAVs-Aided IoT Data Acquisition

1st Thanh Phung Truong
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
tptruong@uclab.re.kr

2nd Tung Son Do
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
tsdo@uclab.re.kr

3rd Junsuk Oh
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
jsoh@uclab.re.kr

4th Dongwook Won
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
dwwon@uclab.re.kr

5th The Vi Nguyen
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
tvnguyen@uclab.re.kr

6th Chunghyun Lee
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
chlee@uclab.re.kr

7th Sungrae Cho
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
srcho@cau.ac.kr

***Abstract*—This paper presents a novel artificial intelligence (AI)-enabled framework for optimizing data acquisition in Unmanned Aerial Vehicle (UAV)-aided Internet of Things (IoT) networks through intelligent swarm clustering. The proposed framework addresses the challenge of efficient data collection from distributed IoT devices (IoTDs) by employing a two-stage optimization approach: a K-Means-based clustering algorithm for IoTD association and a data-weighted centroid optimization strategy for UAV positioning. We introduce a dynamic clustering mechanism that considers spatial distribution and data volume heterogeneity of IoTDs to determine optimal UAV hovering locations. Extensive simulations demonstrate the superior performance of our proposed framework compared to conventional approaches. Results show that our method achieves up to 41.5% reduction in data collection time compared to random clustering with greedy UAV positioning (RC-GU) and maintains consistent performance advantages across various transmission power levels. The framework exhibits particular efficiency in low-power scenarios, making it especially suitable for energy-constrained IoT environments. This work advances intelligent UAV swarm management and provides a promising solution for large-scale IoT data acquisition scenarios.**

***Index Terms*—artificial intelligence, internet of things, unmanned aerial vehicle, data acquisition**

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2024-00436887) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

I. Introduction

The Internet of Things (IoT) has revolutionized modern industries through its capability to enable real-time data collection and processing, fundamentally transforming how we monitor and control our environments [1]. As IoT deployments continue to expand, organizations increasingly rely on vast networks of interconnected devices to optimize their operations and enhance decision-making processes [2], [3]. However, this exponential growth in IoT implementations presents significant challenges in data acquisition and management, particularly in scenarios where traditional communication infrastructure proves inadequate or impractical [4], [5].

In response to these challenges, Unmanned Aerial Vehicles (UAVs) have emerged as a groundbreaking solution, offering unprecedented flexibility and mobility in IoT data acquisition [6]. These aerial platforms excel in scenarios requiring dynamic data collection from distributed IoT devices, especially in remote or infrastructure-limited environments. UAVs' ability to navigate large areas while adaptively adjusting their flight paths has demonstrated remarkable potential in enhancing network coverage and reducing data collection latency [7]–[9]. However, as IoT networks scale up, incorporating multiple UAVs and numerous devices, the complexity of coordinating

efficient data acquisition becomes increasingly challenging [10]. The concept of swarm intelligence, derived from natural biological systems, provides an innovative framework for orchestrating multiple UAVs in data collection missions. This approach enables decentralized coordination among UAV groups, potentially optimizing route planning, task allocation, and energy efficiency [11]. Nevertheless, the dynamic nature of IoT environments, characterized by spatial and temporal variations in data generation patterns, demands sophisticated adaptive algorithms capable of real-time optimization.

Fortunately, artificial intelligence (AI) is a transformative force enabling effective swarm-based clustering for UAV-aided IoT data acquisition. By incorporating AI algorithms, UAV swarms can autonomously identify and group IoT devices based on multiple criteria, including spatial distribution, data priority, and communication requirements [12]–[14]. This intelligent clustering approach significantly reduces system overhead while enhancing data aggregation efficiency. The integration of advanced machine learning techniques enables UAV swarms to make informed decisions about data collection strategies, resulting in more resilient and adaptive IoT systems [15].

Motivated by these observations, this paper introduces a novel AI-enabled swarm clustering framework specifically designed for UAV-assisted IoT data acquisition. Our approach integrates AI algorithms with swarm intelligence principles to optimize UAV coordination in dynamic IoT environments. The proposed framework addresses several critical challenges:

- Efficient clustering of IoT devices considering both spatial and demand factors.
- Dynamic optimization of UAV positioning based on data volume priority.
- Real-time coordination of multiple UAVs to ensure comprehensive network coverage.

Through extensive simulations and performance analysis, we demonstrate that our framework significantly outperforms conventional approaches in various IoT scenarios, particularly in terms of data collection time and energy efficiency. The results validate the effectiveness of combining AI techniques with swarm intelligence for optimizing UAV-aided IoT data acquisition systems.

II. Problem Statement

A. System Model

We consider a UAVs-aided IoT data acquisition system model illustrated in Fig. 1. In this system, a UAV swarm, composed of a set $\mathcal{U} \triangleq \{1, \ldots, U\}$ UAVs, plays the role of a data collection agent in a wide area and sends the data to the BS, which is distant and obstructed from the considered area. In this area, a total of I IoTDs, including sensors, cameras, and monitoring devices, transmit data to the BS for processing. Here, each UAV u can collect data from a set $\mathcal{I}_u \triangleq \{1, \ldots, I_u\}$ IoTDs and the number of IoTDs satisfies

$$\sum_{u \in \mathcal{U}} I_u = I. \tag{1}$$

The channel gain between IoTD i and UAV u is determined based on the communication distance. Following [16], we apply the free space path loss model, expressed as

$$g_{u,i} = \frac{\eta_0}{d_{u,i}^2}, \tag{2}$$

where η_0 denotes the channel gain at a reference distance of 1 meter, and $d_{u,i} \triangleq \| c_u - l_i \|_2$ is the communication distance, with c_u and l_i being the locations of UAV u and IoTD i, respectively. Accordingly, the transmission rate from IoTD i to UAV u can be calculated as

$$r_{u,i} = \alpha_{u,i} \frac{\beta_u B_I}{I_u} \log\left(1 + \frac{p_i g_{u,i}}{\sigma_i^2}\right), \tag{3}$$

where B_I denotes the bandwidth allocated for transmission between IoTDs and UAVs, $\beta_u \triangleq \frac{1}{U}$ denotes the fraction of bandwidth allocated for UAV u, p_i is the transmit power of IoTD i, σ_i^2 is the power of additive Gaussian white noise (AGWN) in the transmission, and $\alpha_{u,i}$ denotes the association variable, expressed as

$$\alpha_{u,i} = \begin{cases} 1, & \text{if } i \in \mathcal{I}_u, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Then, the time for transmitting data from IoTD i to the UAVs can be calculated as

$$t_i^U = \frac{D_i}{\sum_{u \in \mathcal{U}} r_{u,i}}, \tag{5}$$

where D_i denotes the data size of the IoTD i. In addition, the time for each UAV u to finish collecting data from its associated IoTDs can be calculated as

$$t_u = \max\{t_i^U | i \in \mathcal{I}_u\}. \tag{6}$$

After collecting data from IoTDs, UAVs send the data to BS for processing. By letting $h_u \triangleq \frac{\eta_0}{d_u^2}$ denote the channel gain between UAV u and BS, with $d_u^2 \triangleq \| c_u - l_B \|_2$ being the corresponding distance, the transmission rate from UAV u and BS can be calculated as

$$r_u = \frac{B_U}{U} \log\left(1 + \frac{p_u h_u}{\sigma_u^2}\right), \tag{7}$$

where B_U denotes the bandwidth allocated for transmission between UAVs and BS, p_u denotes the transmit power of UAV u, and σ_u^2 is the power of AGWN in the transmission. Accordingly, the time for transmitting data from UAV u to BS can be calculated as

$$t_u^B = \frac{\sum_{i \in \mathcal{I}_u} D_i}{r_u}. \tag{8}$$

Then, the time for BS to finish collecting data from UAVs can be determined as

$$t_B = \max\{t_u^B | u \in \mathcal{U}\}. \tag{9}$$

Consequently, the total time for data collection can be calculated as

$$t = t_B + \max\{t_u | u \in \mathcal{U}\}. \tag{10}$$

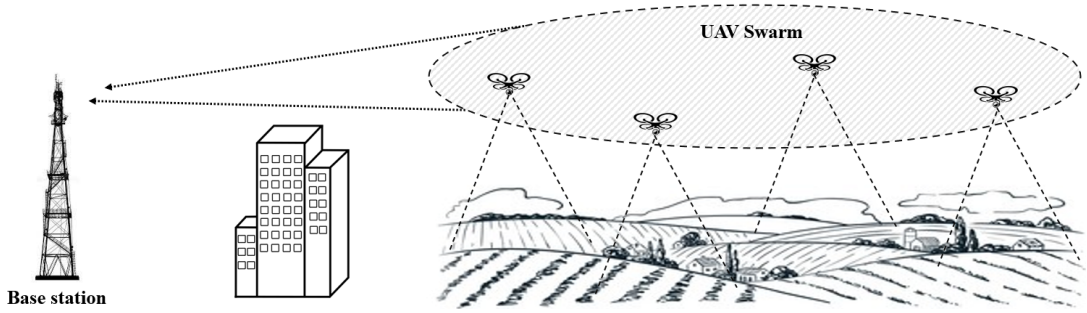

Fig. 1. UAVs-aided IoT data acquisition system model.

B. Problem Formulation

By optimizing the UAV distribution and the association of IoTDs to the UAVs, we aim to minimize the time needed to collect IoTD data. Accordingly, the problem can be formulated as

$$\text{(P1):} \quad \min_{\mathbf{C},\boldsymbol{\alpha}} \quad t \tag{11a}$$

$$\text{s.t.} \quad \alpha_{u,i} \in \{0,1\},, \tag{11b}$$

where $\mathbf{C} \triangleq \{c_u| \in \mathcal{U}\}$, and $\boldsymbol{\alpha} \triangleq \{\alpha_{u,i}|u \in \mathcal{U}, i \in \{1,\ldots,I\}\}$ are the optimization variables, and (11b) is constraint of variable value range.

III. Proposed Swarm Clustering-based Algorithm

Algorithm 1 K-Means-based IoTDs clustering algorithm.

Require: IoTDs' locations and number of UAV (N) (number of clusters).

Initialize N centroids randomly from the IoTDS' locations.
repeat
  for $i \in \{1,2,...,I\}$ **do**
    Compute $d_{u,i}, u \in \mathcal{U}$.
    Set $\alpha_{u,i} = 1$, where $u = \text{argmin}_{u\in\mathcal{U}}(d_{u,i})$.
  end for
  for $u \in \mathcal{U}$ **do**
    $c_u = mean(l_i, i \in \mathcal{I}_u)$.
  end for
until $c_u, u \in \mathcal{U}$, converge.
return $\alpha_{u,i}, c_u, u \in \mathcal{U}, i \in \{1,2,...,I\}$.

To solve the optimization problem $(P1)$, we propose a two-stage cluster-based algorithm: *(i)* IoTDs clustering optimization based on spatial distribution to determine association variables $\boldsymbol{\alpha}$, and *(ii)* UAV location optimization considering data-weighted centroids.

A. Cluster-based Association Optimization

Given the IoTDs' data volumes $\mathbf{D} \triangleq \{D_1,...,D_I\}$, the transmission time t_i is determined by the achievable rate $R_i \triangleq \sum_{u\in\mathcal{U}} r_{u,i}$ as per (5), which depends on the association variables $\boldsymbol{\alpha}$ and the channel gain matrix $\mathbf{H} \triangleq [h_{u,i}]$ between IoTDs and their associated UAVs (3). The channel gain is optimized by clustering proximate IoTDs and assigning dedicated UAVs to each cluster. We employ the K-Means algorithm with enhanced convergence properties, detailed in Algorithm 1.

Algorithm 2 K-Means-based IoTDs Clustering Algorithm with Dynamic Weight Optimization

Require: IoTDs' locations $\mathbf{L} = \{l_1,...,l_I\}$, number of UAVs N, convergence threshold ϵ

Initialize centroids $\mathcal{C} = \{c_1,...,c_N\}$ randomly from $\mathbf{L}$
Initialize iteration counter $t = 0$
repeat
  $t \leftarrow t+1$
  for $i \in \{1,2,...,I\}$ **do**
    Compute Euclidean distances:

$$d_{u,i} = \|l_i - c_u\|_2, \forall u \in \mathcal{U}$$

    Update association matrix:

$$\alpha_{u,i} = \begin{cases} 1, & \text{if } u = \text{argmin}_{u\in\mathcal{U}} d_{u,i} \\ 0, & \text{otherwise} \end{cases}$$

  end for
  Store previous centroids: $\mathcal{C}^{old} \leftarrow \mathcal{C}$
  for $u \in \mathcal{U}$ **do**
    Update cluster centroids: $c_u = \frac{\sum_{i\in\mathcal{I}_u} l_i}{|\mathcal{I}_u|}$
  end for
  Calculate centroid shift: $\Delta = \|\mathcal{C} - \mathcal{C}^{old}\|_F$
until $\Delta < \epsilon$ or $t > t_{max}$
return Association matrix $\boldsymbol{\alpha} = [\alpha_{u,i}]$, centroids $\mathcal{C}$

The algorithm initializes with random UAV positions as centroids $\mathcal{C} \triangleq \{c_1,...,c_U\}$ and iteratively:

1) Updates the association matrix $\boldsymbol{\alpha}$ based on minimum Euclidean distance:

$$\alpha_{u,i} = \text{argmin}_{u\in\mathcal{U}} \|l_i - c_u\|_2, \forall i \in \mathcal{I} \tag{12}$$

2) Recalculates cluster centroids using the mean position:

$$c_u = \frac{1}{|\mathcal{I}_u|} \sum_{i\in\mathcal{I}_u} l_i, \forall u \in \mathcal{U} \tag{13}$$

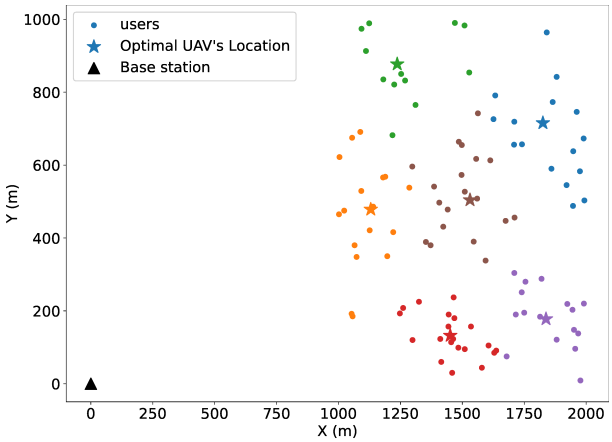

Fig. 2. Environment visualization.

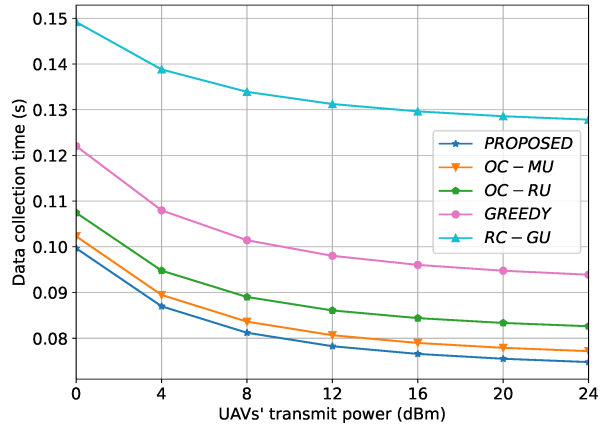

Fig. 4. Data collection time under different UAVs' transmit power.

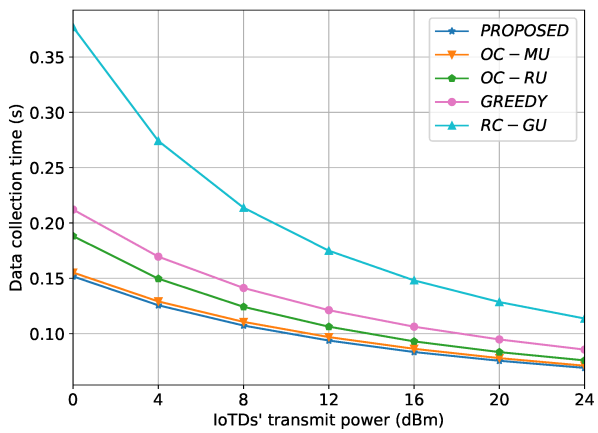

Fig. 3. Data collection time under different IoTDs' transmit power.

B. Data-Weighted Centroid Optimization

While Algorithm 1 provides spatial clustering, it doesn't account for heterogeneous data volumes D_i, which significantly impact transmission time as per (5). We propose a data-weighted centroid optimization where the UAV location c_u is determined by:

$$c_u = \frac{\sum_{i \in \mathcal{I}_u} D_i l_i}{\sum_{i \in \mathcal{I}_u} D_i}, \forall u \in \mathcal{U} \tag{14}$$

This weighted optimization ensures that:

1) IoTDs with larger data volumes ($D_i \gg 0$) have proportionally greater influence on UAV positioning
2) The channel gain $h_{u,i}$ is optimized for data-intensive nodes
3) The overall system achieves minimized data collection time through joint spatial-temporal optimization

The proposed algorithm is shown in Algorithm 2. The resulting UAV positions balance spatial clustering efficiency and data volume requirements, enhancing network performance.

IV. Performance Evaluation

We evaluate the proposed framework performance by conducting numerical simulation results. In this simulation framework, we present a comprehensive network topology consisting of a BS positioned at the coordinate origin with a height of 60 meters, serving as the central node for data aggregation and network management. The network infrastructure encompasses 100 IoTDs distributed stochastically across a square geographical region spanning 1000 m $\times$ 1000 m, situated at 1000 meters from the BS. The system employs 6 UAVs, which maintain a constant altitude of 100 meters above ground level and function as intermediate data collectors, establishing communication links between the ground-level IoTDs and the BS. The IoTDs' data sizes range from 0.1 to 5 Mbits to emulate different data types and applications. Other parameters are set as $B_I = B_U = 10$ MHz, $\eta_0 = 1.42 \times 10^{-4}$, and $\sigma_i^2 = \sigma_u^2 = -174$ dBm/Hz.

First, we visualize the result of IoTDs' clustering and UAVs' location determination in Fig. 2. The figure illustrates a comprehensive spatial distribution of network elements within a two-dimensional deployment area spanning 2000 meters horizontally and 1000 meters vertically. The topology encompasses multiple network components, with the BS strategically positioned at the coordinate origin $(0, 0)$, represented by a triangle marker. The devices, denoted by small dots, exhibit distinctive clustering patterns distributed far from the BS ($x > 1000$ m). These clusters demonstrate varying densities and spatial configurations, suggesting natural congregation points. The optimal UAV locations, indicated by star markers, are strategically positioned to maximize coverage efficiency, with each placement demonstrating a clear correlation with underlying cluster centroids. The visualization effectively captures the hierarchical nature of the network architecture, where UAVs serve as intermediate nodes between the base station and user clusters.

Then, to demonstrate the proposed framework, we compare the data collection performance of the proposed framework with other benchmark schemes. The comparison schemes in

this study are listed below.

- *PROPOSED*: This is the proposed framework, where we apply the K-MEans-based IoTDs clustering algorithm and the centroid-based UAVs location determination strategy to design the association variables and the UAVs' hovering location.
- *OC-MU (Optimal clustering - mean-based UAV location)*: This scheme maintains optimal clustering but determines UAV positions using the mean location of IoTDs within each cluster rather than our proposed centroid-based strategy, as shown in Algorithm 1. This comparison validates the superiority of our centroid-based positioning approach.
- *OC-RU (Optimal clustering - random UAV location)*: While preserving optimal clustering, this scheme randomly positions UAVs within their associated IoTD clusters, serving as a baseline for location optimization strategies.
- *GREEDY*: This approach employs uniform UAV distribution across the target area, where IoTDs associate with their nearest UAV based on proximity. This represents a distance-based greedy allocation strategy.
- *RC-GU (Random clustering - Greedy-based UAV locations)*: This scheme combines uniform UAV distribution with random IoTD-to-UAV associations, representing the most basic approach to network organization.

The performance evaluation illustrated in Fig. 3 demonstrates the comparative analysis of data collection time across varying IoTDs' transmit power levels (0-24 dBm) for different schemes. The PROPOSED framework, implementing K-Means clustering with centroid-based UAV positioning, achieves superior performance with a minimum collection time of 0.069 seconds at 24 dBm. This represents a significant 39.3% improvement over RC-GU (0.114 seconds), which suffers from inefficient random clustering and uniform UAV distribution. The OC-MU scheme, utilizing mean-based UAV positioning, shows comparable but slightly inferior performance (0.071 seconds at 24 dBm), validating the advantage of our centroid-based approach. The GREEDY algorithm's performance (0.086 seconds at 24 dBm) reveals the limitations of proximity-based association without proper clustering. The OC-RU scheme demonstrates intermediate performance despite optimal clustering, highlighting the impact of random UAV positioning on collection efficiency. All schemes exhibit performance improvements with increasing transmit power, with the most significant gains observed in the 0-8 dBm range, after which the improvements gradually occur.

The comparative performance analysis in Fig. 4 evaluates data collection efficiency across varying UAVs' transmit power levels (0-24 dBm). The PROPOSED framework consistently demonstrates superior performance, achieving a minimum collection time of 0.0748 seconds at 24 dBm. This performance represents a substantial improvement over alternative approaches: 41.5% reduction compared to RC-GU (0.128 seconds), 20.4% reduction versus GREEDY (0.094 seconds), and 9.9% improvement over OC-RU (0.083 seconds). The OC-MU scheme, while implementing optimal clustering, shows marginally higher collection times (0.077 seconds at 24 dBm), confirming the effectiveness of our centroid-based positioning strategy over mean-based placement. The consistently poor performance of RC-GU across all power levels underscores the critical importance of proper clustering and strategic UAV positioning. Notably, the performance advantages of the PROPOSED scheme are most pronounced in the lower power range (0-8 dBm), where efficient positioning and clustering demonstrate their highest impact on system performance.

REFERENCES

[1] S. El Jaouhari and E. Bouvet, "Secure firmware over-the-air updates for IoT: Survey, challenges, and discussions," *Internet of Things*, vol. 18, p. 100508, 2022.

[2] Y. Wu, H.-N. Dai, H. Wang, Z. Xiong, and S. Guo, "A survey of intelligent network slicing management for industrial IoT: Integrated approaches for smart transportation, smart energy, and smart factory," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1175–1211, 2022.

[3] D. Kwon, J. Jeon, S. Park, J. Kim, and S. Cho, "Multiagent DDPG-based deep learning for smart ocean federated learning IoT networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895–9903, 2020.

[4] Z. Qadir, K. N. Le, N. Saeed, and H. S. Munawar, "Towards 6G internet of things: Recent advances, use cases, and open challenges," *ICT express*, vol. 9, no. 3, pp. 296–312, 2023.

[5] J. A. Fraire, O. Iova, and F. Valois, "Space-terrestrial integrated internet of things: Challenges and opportunities," *IEEE Communications Magazine*, vol. 60, no. 12, pp. 64–70, 2022.

[6] J. Won, D.-Y. Kim, Y.-I. Park, and J.-W. Lee, "A survey on UAV placement and trajectory optimization in communication networks: From the perspective of air-to-ground channel models," *ICT Express*, vol. 9, no. 3, pp. 385–397, 2023.

[7] A. S. Thabit, C. A. Kerrache, and C. T. Calafate, "A survey on monitoring and management techniques for road traffic congestion in vehicular networks," *ICT Express*, 2024.

[8] T.-H. Nguyen, T. P. Truong, A.-T. Tran, N.-N. Dao, L. Park, and S. Cho, "Intelligent heterogeneous aerial edge computing for advanced 5G access," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 4, pp. 3398–3411, 2024.

[9] W. J. Yun, S. Park, J. Kim, M. Shin, S. Jung, D. A. Mohaisen, and J.-H. Kim, "Cooperative multiagent deep reinforcement learning for reliable surveillance via autonomous multi-UAV control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086–7096, 2022.

[10] T. P. Truong, V. D. Tuong, N.-N. Dao, and S. Cho, "FlyReflect: Joint flying IRS trajectory and phase shift design using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4605–4620, 2023.

[11] U. Iqbal, Y. Lee, S. Cho, I.-H. Lee, and H. Jung, "Cooperative transmission of UAV swarm using orthogonal time–frequency space modulation," *ICT Express*, 2024.

[12] N.-N. Dao, Q.-V. Pham, N. H. Tu, T. T. Thanh, V. N. Q. Bao, D. S. Lakew, and S. Cho, "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1193–1225, 2021.

[13] S. Park, C. Park, and J. Kim, "Learning-based cooperative mobility control for autonomous drone-delivery," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870–4885, 2024.

[14] S. Park, J. P. Kim, C. Park, S. Jung, and J. Kim, "Quantum multi-agent reinforcement learning for autonomous mobility cooperation," *IEEE Communications Magazine*, vol. 62, no. 6, pp. 106–112, 2024.

[15] S. Javed, A. Hassan, R. Ahmad, W. Ahmed, R. Ahmed, A. Saadat, and M. Guizani, "State-of-the-art and future research challenges in UAV swarms," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 023–19 045, 2024.

[16] T. P. Truong, N.-N. Dao, and S. Cho, "HAMEC-RSMA: Enhanced aerial computing systems with rate splitting multiple access," *IEEE Access*, vol. 10, pp. 52 398–52 409, 2022.