# Multi Agent Learning for Flying Base Stations

Ali Anser Jaffri
LUMS
Email: 26100201@lums.edu.pk

***Abstract*—This project presents a framework for optimizing the deployment and movement of Unmanned Aerial Vehicles (UAVs) in dynamic environments, formulated as a Potential Game to ensure system-wide utility maximization while preserving decentralized decision-making. At the core of this approach is the Marginal Contribution Utility (MCU), which quantifies the individual impact of each UAV on the system's global utility function, defined as the sum of satisfied users' utility. By aligning the change in UAV preferences with the increase in global utility, the system inherently satisfies the conditions of a potential game, ensuring convergence to stable and efficient configurations.**

**The framework utilizes log-linear learning, a probabilistic adaptation mechanism, where UAVs iteratively update their action preferences based on MCU-derived rewards. These preferences are evaluated through a softmax function, enabling UAVs to balance exploration and exploitation dynamically. The toolkit integrates simulation modules for UAV-user assignments, reward computation, and utility evaluation, alongside real-time visualization tools to analyze the impact of individual UAV deviations.**

**By employing this game-theoretic formulation, the framework achieves robust performance in environments with varying user demands and UAV constraints. Experimental results demonstrate the scalability and adaptability of the proposed method, highlighting its potential for applications in wireless communication, disaster response, and other scenarios requiring autonomous UAV networks.**

## I. Introduction

The rapid growth of Unmanned Aerial Vehicles (UAVs) as a versatile technology has made them indispensable in addressing diverse challenges, from enhancing wireless communication to supporting disaster response, agricultural monitoring, and urban planning. However, the dynamic nature of these environments introduces significant complexities in UAV deployment and user task assignment. UAVs must operate autonomously, adapt to changing demands, and maximize global performance under constraints such as limited communication range, energy capacity, and coverage areas. Traditional centralized optimization approaches often struggle with scalability and real-time adaptability, whereas decentralized strategies can lack coordination, leading to suboptimal outcomes. Thus, there is a pressing need for frameworks that balance individual UAV decision-making with global utility optimization in dynamic and distributed settings.

To address these challenges, this project proposes a game-theoretic framework for UAV deployment and user assignment, formulated as a *Potential Game*. The key innovation lies in leveraging the Marginal Contribution Utility (MCU), which quantifies the impact of each UAV on the overall system's performance. The global utility function, defined as the cumulative utility of satisfied users, serves as the potential function, ensuring alignment between individual actions and global optimization. Using log-linear learning, UAVs iteratively update their action preferences based on MCU-derived rewards, and actions are chosen using a softmax mechanism to balance exploration and exploitation. This decentralized approach ensures scalability, robustness, and convergence to stable and efficient equilibria. The integration of visualization tools allows for real-time analysis of UAV deviations and their impact on the system utility, enhancing the interpretability of results.

Game-theoretic frameworks have emerged as powerful tools for addressing the complex challenges of UAV-assisted networks, particularly in dynamic and resource-constrained environments. Carotenuto et al. [1] employ non-cooperative game theory to design adaptive user association strategies in post-disaster scenarios, achieving significant improvements in Quality of Service (QoS) by minimizing data loss rates and energy consumption. Extending this work, Zhang et al. [2] explore cooperative UAV-assisted non-orthogonal multiple access (NOMA) networks, demonstrating how game-theoretic resource allocation enhances spectral efficiency and optimizes overall network performance. Providing a broader perspective, Saad et al. [3] present a comprehensive survey on the application of game theory in UAV communications, emphasizing its utility in managing interference, user association, and energy efficiency. Li et al. [4] further highlight the combined potential of game theory and machine learning, showcasing their synergy in adaptively optimizing operations within UAV-assisted wireless networks. Liu et al. [5] take this a step further by formulating a joint optimization problem as a constrained potential game, focusing on adaptive uplink scheduling and UAV association to significantly enhance network throughput and fairness. Together, these studies underline the versatility and effectiveness of game-theoretic approaches in improving UAV network performance, forming a strong foundation for this work's exploration of scalable and adaptive methodologies for UAV deployment and resource optimization.

The efficient deployment and resource allocation of UAVs have been extensively studied, with a focus on optimizing 3D placement, user association, and energy efficiency in UAV-enabled networks. Mkiramweni et al. [6] provide a comprehensive survey on game-theoretic approaches in UAV communications, emphasizing their role in optimizing resource utilization and enhancing network performance. Mozaffari et al. [7] address the deployment of multiple UAVs for maximizing wireless coverage, presenting an algorithmic framework that balances service quality and resource constraints. Similarly, Shakhatreh et al. [8] utilize particle swarm optimization for efficient 3D UAV placement, achieving improved network efficiency in complex environments. Alzenad et al. [9] focus on energy-efficient UAV placement for maximal coverage, proposing a 3D placement strategy that minimizes energy consumption while meeting coverage requirements. El Hammouti et al. [10] extend this to multi-UAV networks by introducing a distributed algorithm for joint 3D placement and user association, enabling UAVs to dynamically adapt to user demands in real-time. Liu et al. [11] explore resource allocation and 3D placement in UAV-enabled IoT communications, highlighting energy-efficient strategies tailored for IoT use cases. Zou et al. [12] further enhance network performance by jointly optimizing 3D placement and partially overlapped channel assignments to maximize throughput. Pan et al. [13] investigate UAV placement and user association in software-defined cellular networks, presenting solutions that integrate UAV capabilities with modern network infrastructure. Additionally, El Hammouti et al. [14] propose both optimal and greedy drone association and positioning schemes for the Internet of UAVs, offering a trade-off between computational efficiency and global utility maximization. Collectively, these works underscore the importance of integrating advanced optimization techniques into UAV deployment strategies, aligning with the objectives of this work to achieve scalable and energy-efficient UAV-enabled communication networks.

Reinforcement learning has become a cornerstone methodology for addressing the complex optimization challenges in UAV-assisted communication networks, enabling adaptive and scalable solutions. Chen et al. [15] propose a deep reinforcement learning framework to jointly optimize multi-UAV deployment and user association, focusing on long-term communication coverage. Their approach ensures sustained connectivity by dynamically adapting UAV positions and user assignments to changing environmental conditions. Building on this, Li et al. [4] explore multi-agent deep reinforcement learning for decoupled user association and resource allocation, highlighting its potential to improve network efficiency through cooperative decision-making among UAVs. Zhao et al. [16] further extend these ideas by applying multi-agent reinforcement learning to maximize user connectivity in UAV-based networks, investigating the role of inter-agent information exchange in enhancing network performance. Together, these studies demonstrate the versatility and effectiveness of reinforcement learning techniques in optimizing UAV operations, providing key insights into how decentralized frameworks can improve communication coverage, resource allocation, and user association in dynamic environments.

The theoretical foundations of potential games and their applications provide critical insights into designing efficient, decentralized decision-making frameworks for UAV networks. Monderer and Shapley [17] introduced potential games as a powerful class of games where individual agent actions align with a global potential function, ensuring convergence to Nash equilibria through simple learning dynamics. Building on this foundation, Rosenthal [18] developed the concept of congestion games, a subclass of potential games, to analyze equilibrium properties in systems where agents compete for shared resources. These ideas have been extended to multi-agent systems, including UAV networks, where agents (UAVs) must optimize their actions under shared constraints. Marden and Shamma [19] revisited log-linear learning as a practical, payoff-based implementation for achieving equilibria in potential games, demonstrating its robustness under asynchronous updates and incomplete information. These foundational works form the backbone of game-theoretic optimization, enabling the development of scalable and adaptive frameworks like the one proposed in this project. By leveraging the properties of potential games and incorporating payoff-based learning mechanisms, this work aligns individual UAV actions with system-wide objectives, ensuring efficient and robust operation in dynamic and resource-constrained environments

## II. Methodology

### A. Problem Statement

The deployment of UAVs for user association and communication coverage is modeled as an optimization problem that balances global utility and decentralized decision-making. The problem is formulated as a *Potential Game*, where each UAV acts as a player that selects strategies to maximize a global objective represented by a potential function.

### B. Game Formulation

The game is defined as $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \{u_i\}_{i \in \mathcal{N}})$, where:

- $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of UAVs (players).
- $\mathcal{A} = \prod_{i \in \mathcal{N}} \mathcal{A}_i$ is the joint action space, where $\mathcal{A}_i$ is the set of feasible actions for UAV $i$, such as selecting a position or resource allocation.
- $u_i : \mathcal{A} \to \mathbb{R}$ is the utility function of UAV $i$, defined to align with the global objective.

### C. Global Utility Function

The global utility function $\Phi : \mathcal{A} \to \mathbb{R}$ is defined as the total number of users being served:

$$\Phi(a) = \sum_{j \in \mathcal{U}} \mathbb{I}_j(a),$$

where:

- $\mathcal{U}$ is the set of all users.
- $\mathbb{I}_j(a)$ is the indicator function:

$$\mathbb{I}_j(a) = \begin{cases} 1, & \text{if user } j \text{ is served by at least one UAV under } a, \\ 0, & \text{otherwise.} \end{cases}$$

### D. UAV Utility Function

The utility function for each UAV $i$ is defined as the *Marginal Contribution Utility (MCU)*:

$$u_i(a) = \Phi(a) - \Phi(a_{-i}),$$

where:

- $\Phi(a)$ is the global utility with all UAVs taking actions $a$.
- $\Phi(a_{-i})$ is the global utility when UAV $i$ does not participate, i.e., the actions of all UAVs except $i$.

### *E. Potential Game Properties*

The game $\mathcal{G}$ satisfies the properties of a *Potential Game*:

1) **Potential Function:** $\Phi(a)$ is a potential function such that:
$$u_i(a) = \Phi(a) - \Phi(a_{-i}),$$
ensuring alignment between individual utilities and the global objective.
2) **Convergence:** The game converges to a Nash equilibrium, where no UAV has an incentive to unilaterally deviate, via iterative best-response dynamics or learning algorithms.
3) **Decentralization:** UAVs optimize their actions independently, based only on local utility evaluations, enabling scalable implementations.

### *F. Learning Algorithm*

To find a Nash equilibrium, a *Log-Linear Learning (LLL)* algorithm is employed, which iteratively updates UAV actions using a probabilistic approach:

1) **Initialization:** Each UAV initializes its action $a_i \in \mathcal{A}_i$ randomly.
2) **Utility Evaluation:** For each action $a_i \in \mathcal{A}_i$, UAV $i$ evaluates its utility:
$$u_i(a) = \Phi(a) - \Phi(a_{-i}).$$
3) **Action Preference Update:** UAV $i$ updates its preference for each action $a_i$ using a Boltzmann distribution:
$$P(a_i) = \frac{\exp(\beta u_i(a))}{\sum_{a_i' \in \mathcal{A}_i} \exp(\beta u_i(a_i', a_{-i}))},$$
where $\beta > 0$ is the inverse temperature parameter controlling the balance between exploration and exploitation.
4) **Action Selection:** UAV $i$ selects an action $a_i$ with probability $P(a_i)$.

### *G. Algorithm Workflow*

The proposed framework operates iteratively as follows:

1) **Input:** Initialize UAV positions, user demands, and constraints (e.g. distance limits).
2) **Iterative Updates:**
   - Each UAV selects an action based on its updated preferences.
   - Utilities are evaluated, and the potential function $\Phi(a)$ is updated.
   - UAVs update their preferences using Log-Linear Learning.
3) **Convergence:** The algorithm continues until the UAVs reach a Nash equilibrium.
4) **Output:** The final UAV positions and user associations optimize the global utility.

### *H. Advantages of the Proposed Methodology*

- **Optimality:** The alignment of UAV utilities with the global potential function ensures system-wide optimization.
- **Scalability:** Decentralized decision-making allows the framework to scale with increasing numbers of UAVs and users.
- **Robustness:** The framework adapts to changes in user demands and UAV constraints dynamically.
- **Fairness:** The marginal contribution utility ensures equitable resource allocation across users.

## III. Results

### *A. Theoretical Analysis*

The proposed framework is grounded in potential game theory, ensuring convergence to a Nash equilibrium through log-linear learning. By aligning UAV utilities with the global utility function, the framework guarantees that individual UAV actions positively contribute to the overall system objective. Theoretical properties include:

- **Convergence:** Log-linear learning guarantees convergence within a finite number of iterations.
- **Scalability:** The decentralized nature of the framework ensures efficient operation with increasing numbers of UAVs and users.
- **Fairness:** Marginal contribution-based utilities ensure equitable resource allocation.

### *B. Simulation Setup*

Two simulation scenarios were conducted to evaluate the framework under varying grid sizes and configurations.

*1) Scenario 1: Medium-Scale Environment:*

- **Grid Size:** $200 \times 200$
- **Number of UAVs:** $N = 20$
- **Number of Users:** $M = 60$
- **Learning Parameters:**
  - Number of Iterations: 300
  - Learning Rate: $\alpha = 0.2$
  - Inverse Temperature: $\beta = 7.5$
- **Coverage Constraints:**
  - Coverage Radius: 15.0
  - SNR Threshold: 0.1
- **Collision Avoidance:**
  - Collision Distance: 5.0
  - Collision Penalty: $\gamma_c = -0.5$

*2) Scenario 2: Small-Scale Environment:*

- **Grid Size:** $100 \times 100$
- **Number of UAVs:** $N = 10$
- **Number of Users:** $M = 30$
- **Learning Parameters:**
  - Number of Iterations: 100
  - Learning Rate: $\alpha = 0.2$
  - Inverse Temperature: $\beta = 7.5$
- **Coverage Constraints:**

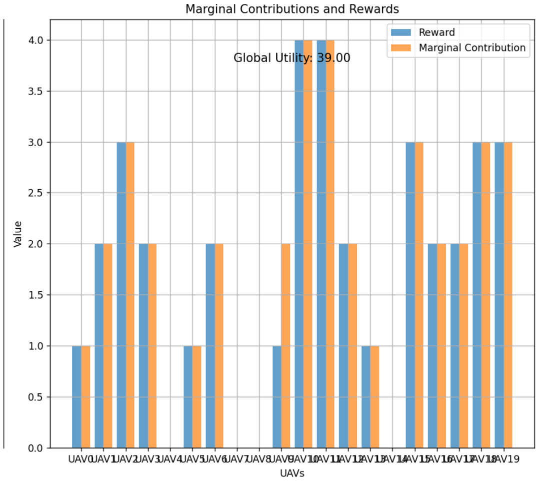
  - Coverage Radius: 15.0
  - SNR Threshold: 0.1
- **Collision Avoidance:**
  - Collision Distance: 5.0
  - Collision Penalty: $\gamma_c = -0.5$

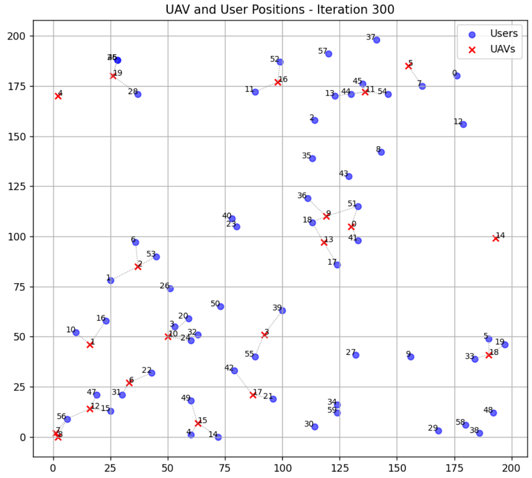
## C. Simulation Results

The results for both scenarios are summarized below:

### 1) Scenario 1: Medium-Scale Environment:

- **Users Served:** 40 out of 60 users.
- **Convergence:** Achieved within 300 iterations.
- **Global Utility:** 40, representing the number of users served.

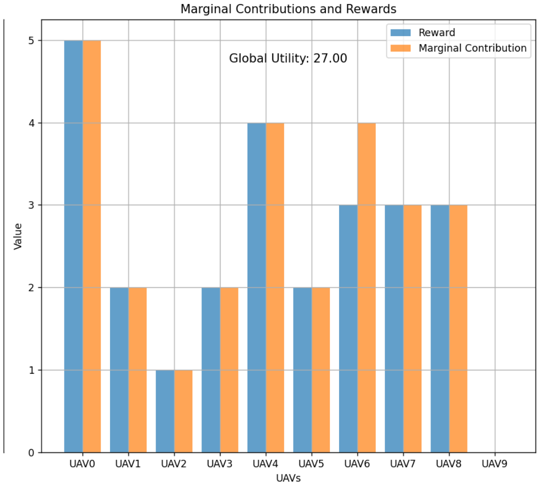




### 2) Scenario 2: Small-Scale Environment:

- **Users Served:** 27 out of 30 users.
- **Convergence:** Achieved within 100 iterations.
- **Global Utility:** 27, representing the number of users served.

### 3) Performance Metrics Comparison:



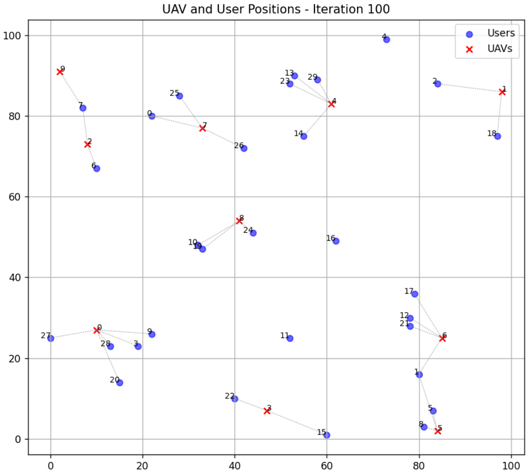


TABLE I
PERFORMANCE METRICS FOR BOTH SCENARIOS

| Metric | Scenario 1 | Scenario 2 |
|---|---|---|
| Users Served | 40 / 60 | 27 / 30 |
| Convergence Time | 300 iterations | 100 iterations |
| Average Utility per UAV | 2.0 | 2.7 |
| Maximum Coverage Radius | 15.0 | 15.0 |

## D. Discussion of Results

The simulation results validate the framework's adaptability across different grid sizes and user densities. In Scenario 1, serving 40 users within 300 iterations highlights the system's efficiency in a medium-scale environment. Scenario 2 demonstrated higher user coverage efficiency, serving 27 out of 30 users in a smaller grid, leading to an average utility of 2.7 per UAV. The collision avoidance mechanism effectively maintained safe operation in both scenarios. These results illustrate the framework's ability to optimize UAV deployment and user associations under varying environmental constraints.

## E. Impact of Parameters on Results

The performance of the framework is influenced by several key parameters. Their high-level significance and impact on the results are summarized below:

- **Learning Rate** ($\alpha$)**:** Controls the magnitude of updates to UAV action preferences. A higher $\alpha$ accelerates convergence but can lead to instability, while a smaller $\alpha$ ensures stability at the cost of slower convergence.
- **Inverse Temperature** ($\beta$)**:** Balances exploration and exploitation in the learning process. A higher $\beta$ favors exploitation for faster convergence but risks suboptimal solutions. A lower $\beta$ encourages exploration, potentially improving the global utility at the expense of delayed convergence.
- **Number of UAVs** ($N$) **and Users** ($M$)**:** Increasing $N$ relative to $M$ improves coverage and redundancy, leading to higher global utility. A lower $N$ can result in unserved users.
- **Grid Size:** Defines the spatial area of operation. A larger grid size spreads users across a wider area, reducing the global utility and requiring more UAVs to achieve similar coverage.
- **Number of Iterations:** Determines the time available for the system to converge. Sufficient iterations ensure UAVs stabilize at a Nash equilibrium, achieving optimal or near-optimal global utility. More iterations come at a computational cost
- **Coverage Radius:** Defines the maximum range within which UAVs can serve users. Increasing the radius improves coverage but may lead to overlapping services or interference. Reducing the radius restricts service areas, requiring more UAVs to maintain coverage.

## IV. Conclusion

This work presents a scalable and adaptive framework for UAV deployment and user association, leveraging the principles of potential games and log-linear learning. The proposed framework aligns individual UAV actions with a global utility function, ensuring decentralized yet efficient optimization of system performance. Key contributions include:

- Formulating the UAV deployment and user association problem as a potential game, guaranteeing convergence to a Nash equilibrium.
- Designing a global utility function based on the number of users served, ensuring alignment with individual UAV utilities.
- Implementing log-linear learning for UAVs to iteratively improve their actions, balancing exploration and exploitation for optimal coverage.
- Demonstrating the scalability and adaptability of the framework across different scenarios, including varying grid sizes, UAV counts, and user densities.
- Validating the framework's efficiency through simulations, achieving high global utility and rapid convergence within $100 - 300$ iterations in current tested scenarios.

The results highlight the framework's ability to optimize UAV deployment and user associations in resource-constrained and dynamic environments. Future work may explore enhancements such as integrating energy-efficient algorithms, dynamic user mobility models, and multi-UAV coordination in real-world settings.

## References

[1] N. Carotenuto *et al.*, "Game theoretic optimal user association in emergency networks," *Springer*, 2023.

[2] S. Zhang *et al.*, "Game theoretical approaches for cooperative uav noma networks," *IEEE Xplore*, 2023.

[3] W. Saad *et al.*, "A survey of game theory in unmanned aerial vehicles communications," *IEEE Xplore*, 2023.

[4] X. Li *et al.*, "Game theory and machine learning in uavs-assisted wireless communication networks: A survey," *Academia.edu*, 2023.

[5] Y. Liu *et al.*, "Adaptive uplink scheduling and uav association in uav-assisted cellular networks: A game-theoretic approach," *IEEE Xplore*, 2023.

[6] M. E. Mkiramweni, C. Yang, J. Li, and W. Zhang, "A survey of game theory in unmanned aerial vehicles communications," *IEEE Communications Surveys  Tutorials*, vol. 21, no. 4, pp. 3386–3416, 2019.

[7] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.

[8] H. Shakhatreh, A. Khreishah, A. Alsarhan, I. Khalil, A. Sawalmeh, and N. S. Othman, "Efficient 3d placement of a uav using particle swarm optimization," in *2017 8th International Conference on Information and Communication Systems (ICICS)*, 2017, pp. 258–263.

[9] 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 Communications Letters*, vol. 6, no. 4, pp. 434–437, 2017.

[10] H. El Hammouti, M. Benjillali, B. Shihada, and M.-S. Alouini, "Learn-as-you-fly: A distributed algorithm for joint 3d placement and user association in multi-uavs networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5831–5844, 2019.

[11] Y. Liu, K. Liu, J. Han, L. Zhu, Z. Xiao, and X. Xia, "Resource allocation and 3d placement for uav-enabled energy-efficient iot communications," *IEEE Internet of Things Journal*, 2020.

[12] C. Zou, X. Li, X. Liu, and M. Zhang, "3d placement of unmanned aerial vehicles and partially overlapped channel assignment for throughput maximization," *Elsevier Digital Communications and Networks*, 2020.

[13] C. Pan, C. Yin, N. C. Beaulieu, and J. Yu, "3d uav placement and user association in software-defined cellular networks," *Springer Wireless Networks*, vol. 25, no. 7, pp. 3883–3897, 2019.

[14] H. El Hammouti, D. Hamza, B. Shihada, and M.-S. Alouini, "The optimal and the greedy: Drone association and positioning schemes for internet of uavs," *IEEE Transactions on Wireless Communications*, 2023.

[15] X. Chen *et al.*, "Joint optimization of multi-uav deployment and user association via deep reinforcement learning for long-term communication coverage," *IEEE Xplore*, 2023.

[16] L. Zhao *et al.*, "Distributed user connectivity maximization in uav-based communication networks using multi-agent reinforcement learning," *IEEE Xplore*, 2023.

[17] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[18] R. W. Rosenthal, "A class of games possessing pure-strategy nash equilibria," *International Journal of Game Theory*, vol. 2, no. 1, pp. 65–67, 1973.

[19] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, 2012.