Adversarial-Aware Multi-Agent Reinforcement Learning for Secure Resource Allocation in Heterogeneous Aerial Access IoT Networks

Umar Sa'ad[a], Woongsoo Na[b], Nhu-Ngoc Dao[c,*], Sungrae Cho[a,*]

[a]*School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea*
[b]*Department of Software, Kongju National University, Cheonan 31080, South Korea*
[c]*Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea*

Abstract

This paper proposes an Adversarial-Aware Multi-Agent Deep Deterministic Policy Gradient (AA-MADDPG) framework integrating adversarial risk analysis with deep reinforcement learning for robust resource allocation and offloading in heterogeneous aerial access networks. We model adversarial behavior using three rationality paradigms, including Nash equilibrium, level-k thinking, and prospect maximizing, employing Bayesian model averaging for robust adversarial action prediction. The framework enables collaborative decision-making among aerial access tiers and IoT devices while maintaining attack resilience. Simulations demonstrate that AA-MADDPG reduces task drop rates by 27% and energy consumption by 19% compared to baselines.

Keywords: Computation offloading, adversarial risk analysis, multi-agent reinforcement learning, aerial access networks.

1. Introduction

The proliferation of IoT devices has created unprecedented demand for distributed computing infrastructure with stringent quality-of-service (QoS) requirements [1]. Heterogeneous Aerial Access IoT (AAIoT) networks, integrating high-altitude platforms (HAPs) and unmanned aerial vehicles (UAVs) as aerial computing servers (ACSs), have emerged as a transformative solution for IoT service delivery in remote and disaster-affected areas [2], leveraging flexible deployment, line-of-sight communication, and on-demand computing for efficient task offloading and resource allocation [3, 4].

However, the open wireless nature of AAIoT networks introduces critical vulnerabilities. Unlike opportunistic attacks, intelligent adversaries can observe network behavior, learn optimal attack strategies, and dynamically adapt through jamming, spoofing, and resource exhaustion [5, 6], posing fundamental challenges to existing approaches that assume benign environments or rely on reactive defenses. Centralized optimization is further intractable for dynamic AAIoT networks with numerous agents making interdependent decisions under uncertainty. Recent surveys [7, 8] highlight the urgent need for adversarial-aware, proactive learning algorithms capable of anticipating and mitigating intelligent attacks while maintaining optimal network performance.

1.1. Related Work

Recent advances in intelligent offloading for heterogeneous aerial IoT networks employ multi-agent deep reinforcement learning (MADRL) frameworks, particularly multi-agent deep deterministic policy gradient (MADDPG) and multi-agent proximal policy optimization (MAPPO), demonstrating superior performance in handling nonconvex, dynamic environments [3, 9, 4, 10, 6]. Hierarchical UAV-HAP architectures manage delay and energy constraints through collaborative processing [10, 11], while game-theoretic approaches achieve Nash equilibria with reduced latency and energy consumption [12, 13, 14, 2], though typically assuming rational adversaries without behavioral biases. Space-air-ground integrated networks extend coverage through multi-tier offloading [15, 16], but security remains largely unaddressed. Despite MADDPG's effectiveness for cooperative decentralized learning [17], existing implementations lack adversarial awareness, and limited physical-layer security work [18, 19] does not provide comprehensive strategic adversary modeling.

1.2. Motivations

No existing work algorithmically integrates adversarial risk analysis (ARA) [20] into the value estimation architecture of MARL for aerial AAIoT networks. Current methods either ignore adversarial behavior, employ reactive defenses, or assume single-paradigm rationality. Robust MARL methods such as minimax-MADDPG [21] optimize against worst-case opponents, yielding overly conservative policies that sacrifice benign-condition performance. Bayesian model-averaged multi-paradigm opponent modeling capturing perfect rationality, bounded rationality, and behavioral biases remains unexplored for aerial IoT resource allocation, despite real-world adversaries rarely conforming to a single rationality model (mismatched assumptions cause 35–50% performance degradation, as our ablation demonstrates).

*Corresponding authors
Email addresses: umar@uclab.re.kr (Umar Sa'ad), wsna@kongju.ac.kr (Woongsoo Na), nndao@sejong.ac.kr (Nhu-Ngoc Dao), srcho@cau.ac.kr (Sungrae Cho)

1.3. Contributions

- We propose *AA-MADDPG*, introducing an ARA-conditioned centralized critic that embeds adversarial belief-state features directly into Q-value estimation, enabling risk-aware policy optimization beyond post-hoc robustness.
- We develop a multi-paradigm opponent model combining Nash equilibrium, level-k thinking, and prospect theory via Bayesian model averaging, enabling adaptive learning under adversary-type uncertainty while preserving near-optimal benign-condition performance, in contrast to worst-case robust RL.
- Simulations demonstrate 27% higher task completion, 19% lower energy consumption, and >80% resilience under 50% jamming versus state-of-the-art baselines, including adversarial MARL methods.

2. System and Threat Model

2.1. Network and Task Model

We consider a heterogeneous aerial IoT system comprising a HAP, N UAV edge servers $\mathcal{N} = \{1, \dots, N\}$, and M IoTDs $\mathcal{M} = \{1, \dots, M\}$. At each epoch t, IoTD m generates task $\phi_m[t] = (w_m[t], v_m[t], \tau_m^{\max}[t])$ representing input size (bits), CPU cycles per bit, and latency deadline (s). Each task is partially offloaded to server $j \in \{h\} \cup \mathcal{N}$ via association $x_{mj}[t] \in \{0, 1\}$ with $\sum_j x_{mj}[t] = 1$ and offloading ratio $\beta_{mj}[t] \in (0, 1]$; fraction $1 - \beta_{mj}[t]$ executes locally.

2.2. Communication and Computation Models

With bandwidth b_j, share $s_{mj}[t] \in [0, 1]$ ($\sum_m s_{mj}[t] \le 1$), transmit power p_m, and channel gain $g_{mj}[t]$, the uplink rate is

$$r_{mj}[t] = s_{mj}[t]\, b_j \log_2\Big(1 + \tfrac{p_m g_{mj}[t]}{N_0 b_j + I_{mj}[t]}\Big), \tag{1}$$

where N_0 is noise spectral density and $I_{mj}[t]$ is attack-induced interference, giving transmission time $t_{mj}^{\text{off}}[t] = \beta_{mj}[t] w_m[t] / r_{mj}[t]$. With local and server CPU frequencies $f_m[t]$, $f_{mj}[t]$ ($\sum_m f_{mj}[t] \le f_j^{\max}$), execution times are

$$t_{mj}^{\text{exec}}[t] = \tfrac{\beta_{mj}[t] w_m[t] v_m[t]}{f_{mj}[t]}, \quad t_m^{\text{loc}}[t] = \tfrac{(1-\beta_{mj}[t]) w_m[t] v_m[t]}{f_m[t]}, \tag{2}$$

and since local and remote execution are parallel, total latency is

$$T_m[t] = \max\{t_{mj}^{\text{off}}[t] + t_{mj}^{\text{exec}}[t],\ t_m^{\text{loc}}[t]\} \le \tau_m^{\max}[t]. \tag{3}$$

2.3. Energy Model

Under dynamic voltage and frequency scaling, local, transmission, and server-side energies are

$$E_m^{\text{loc}}[t] = \kappa_m f_m^2[t](1 - \beta_{mj}[t]) w_m[t] v_m[t], \quad E_{mj}^{\text{tx}}[t] = p_m t_{mj}^{\text{off}}[t], \tag{4}$$

$$E_{mj}^{\text{exec}}[t] = \kappa_j f_{mj}^2[t] \beta_{mj}[t] w_m[t] v_m[t], \tag{5}$$

giving total energy $E_m[t] = E_m^{\text{loc}}[t] + E_{mj}^{\text{tx}}[t] + E_{mj}^{\text{exec}}[t]$.

2.4. Adversary Model and ARA Framework

An intelligent adversary launches one of three attacks per epoch: **jamming** (interference $I_{mj}[t] = \alpha[t] P_J \chi_{mj}[t]$, where P_J is jammer power, $\alpha[t] \in [0, 1]$ intensity, and $\chi_{mj}[t] \in \{0, 1\}$ link indicator); **spoofing** (falsified channel gains $\tilde{g}_{mj}[t] = g_{mj}[t](1 + \epsilon_{mj}[t])$, $|\epsilon_{mj}[t]| \le \epsilon^{\max}$, with injection probability π_s); or **hybrid** (simultaneous jamming and spoofing). The attacker's policy is modeled as an ARA mixture:

$$P(a^{\text{adv}}|O_t) = \sum_{k \in \{\text{GT}, \ell\text{-k}, \text{PT}\}} \omega_k P_k(a^{\text{adv}}|O_t), \quad \sum_k \omega_k = 1, \tag{6}$$

where the three components represent: **GT** (Nash/best-response based on utilities $U^{\text{adv}}, U^{\text{def}}$); **level-$k$** (best-responds to level-$(k-1)$ heuristic defender, $k \in \{1, 2\}$); and **PT** (subjective value $v(u) = u^\alpha$ for gains, $v(u) = -\lambda(-u)^\beta$ for losses, with Prelec probability weighting $\pi(p) = \exp\{-\eta(-\ln p)^\gamma\}$). Mixture weights ω_k are updated via Bayesian inference on observed attack traces (Figure 1).

3. Problem Formulation

3.1. Decision Variables and Constraints

At epoch t, the defender chooses associations $\{x_{mj}\}$, bandwidth shares $\{s_{mj}\}$, offloading ratios $\{\beta_{mj}\}$, CPU allocations $\{f_{mj}, f_m\}$, and UAV positions $\{\mathbf{q}_n\}$, subject to:

$$\textstyle\sum_j x_{mj}[t] = 1,\ x_{mj} \in \{0, 1\}; \quad \sum_m s_{mj}[t] \le 1,\ s_{mj} \in [0, 1]; \tag{7}$$

$$\textstyle\sum_m f_{mj}[t] \le f_j^{\max},\ f_{mj} \ge 0,\ f_m \in [0, f_m^{\max}]; \tag{8}$$

$$\beta_{mj} \in [0, 1],\ T_m[t] \le \tau_m^{\max}[t],\ \|\mathbf{q}_n[t+1] - \mathbf{q}_n[t]\| \le v_n^{\max} \Delta t. \tag{9}$$

3.2. Risk-Aware Objective

Define success rate $\text{Succ}[t] = M^{-1} \sum_m \mathbf{1}\{T_m[t] \le \tau_m^{\max}[t]\}$, normalized energy $\bar{E}[t] = M^{-1} \sum_m E_m[t] / E_{\text{ref}}$, and normalized delay $\bar{D}[t] = M^{-1} \sum_m T_m[t] / \tau_m^{\max}[t]$. The per-epoch utility is

$$U_t = \omega_c \text{Succ}[t] - \omega_e \bar{E}[t] - \omega_d \bar{D}[t] - \omega_a \mathbb{E}_{a^{\text{adv}} \sim P(\cdot|O_t)}[\text{Loss}(a^{\text{adv}}, s_t, a_t)], \tag{10}$$

where $\text{Loss}(\cdot)$ quantifies attack-induced degradation, and the objective maximizes $\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t U_t]$.

3.3. Multi-Agent RL Formulation

We formulate control as a partially-observed stochastic game solved via MADDPG with centralized training and decentralized execution. The global state s_t includes task parameters, channel gains, server resources, UAV positions, and ARA parameters θ_t^{ARA}; each agent observes local o_t^i, executes a_t^i, and receives reward $R_t = U_t$ from (10). The ARA-conditioned critic

$$Q_\psi(s_t,\ a_t,\ \text{feat}(\theta_t^{\text{ARA}})), \tag{11}$$

incorporates mixture weights and component parameters via $\text{feat}(\cdot)$, directly embedding risk-awareness into value estimation.

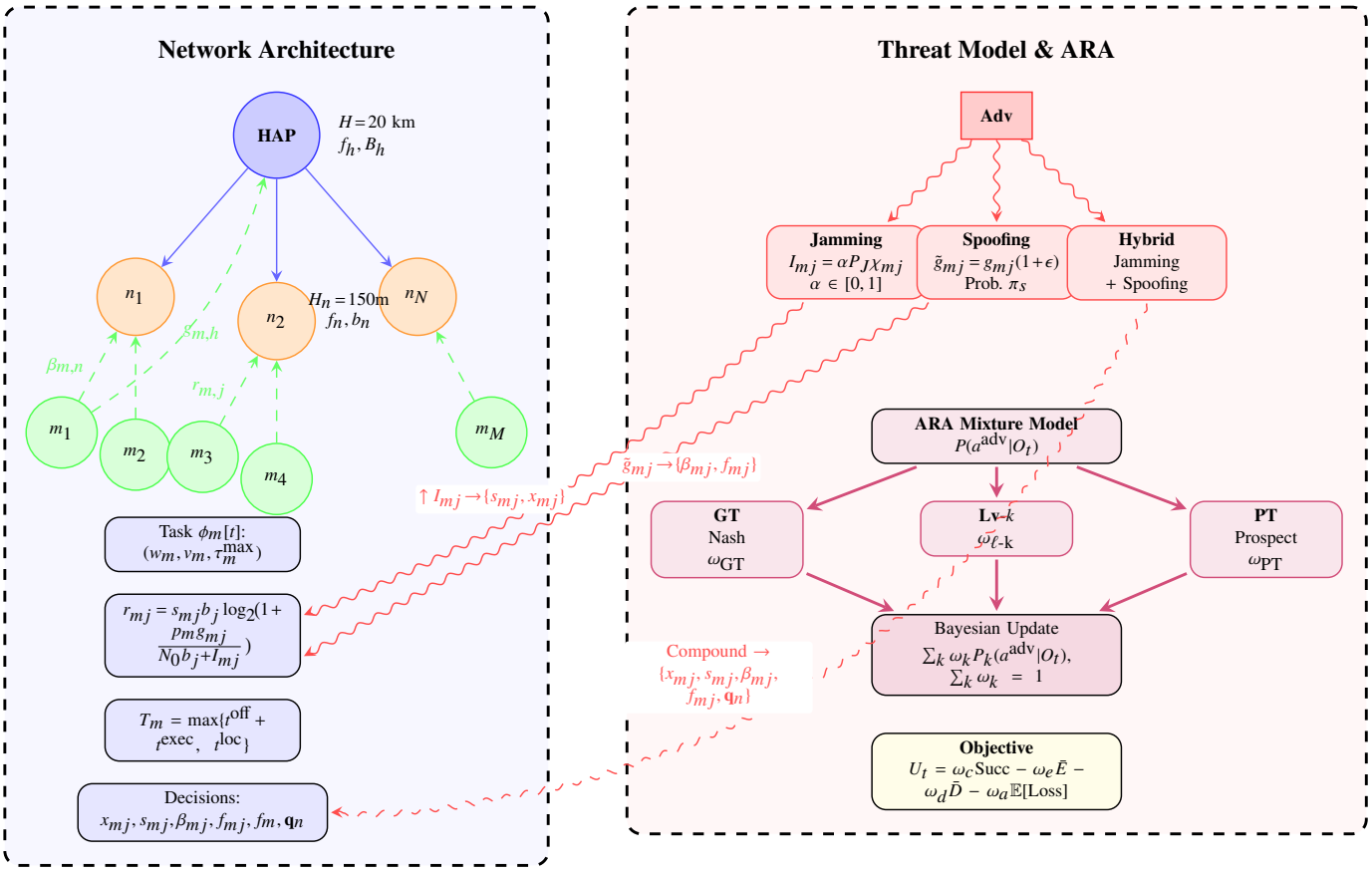

Figure 1: System and threat model. **Left:** HAP (20 km), UAV edge servers (150 m), and IoTDs with offloading ratio $\beta_{mj}[t]$. **Right:** Adversary launches jamming, spoofing, or hybrid attacks; annotated arrows show causal impact on decision variables. ARA combines three rationality paradigms via Bayesian averaging to predict $P(a^{\text{adv}}|O_t)$. Solid blue: association; dashed green: offloading; red: attack paths.

4. Simulation Results and Discussion

4.1. Simulation Setup

We compare AA-MADDPG against seven baseline approaches:

1. Standard MADDPG without adversarial awareness,
2. Minimax-MADDPG (M3DDPG) [21], which trains against worst-case opponent perturbations,
3. Robust Adversarial Reinforcement Learning (RARL) [22], which co-trains a destabilizing adversary alongside the policy agent,
4. Deep Q-Network (DQN)-based offloading,
5. Greedy resource allocation,
6. Random offloading, and
7. A game-theoretic Nash equilibrium approach.

Three adversarial attack scenarios are evaluated following realistic threat models reported in prior literature [23, 18, 19]: (i) jamming attacks with 30–70% interference power, (ii) spoofing attacks that inject false information into 10–40% of transmitted packets, and (iii) hybrid attacks combining both strategies.

4.2. Performance Under Normal Conditions

Figure 2 demonstrates that AA-MADDPG achieves 96.3% average task completion rate with 30 IoTDs, outperforming standard MADDPG (91.2%), DQN (87.5%), greedy allocation (83.1%), and random offloading (71.4%). The superior performance stems from AA-MADDPG's ability to learn coordinated offloading strategies that balance load across ACSs.

Among the adversarial-aware baselines, M3DDPG achieves 89.4% task completion under benign conditions but incurs a 12% energy overhead due to its worst-case optimization strategy, while RARL achieves 90.1% task completion with a more balanced energy profile. AA-MADDPG outperforms both approaches because its Bayesian-averaged threat model mitigates the conservatism inherent in minimax formulations while preserving robustness through multi-paradigm adversary prediction.

Energy consumption analysis in Fig. 3 shows that AA-MADDPG reduces average energy consumption to 1.34 J per IoTD, representing 19% savings compared to standard MADDPG (1.65 J) and 31% savings compared to greedy allocation (1.94 J). Training convergence (Fig. 4) reveals that AA-MADDPG achieves stable performance after approximately 800 episodes, 30% faster than standard MADDPG (1,150 episodes).

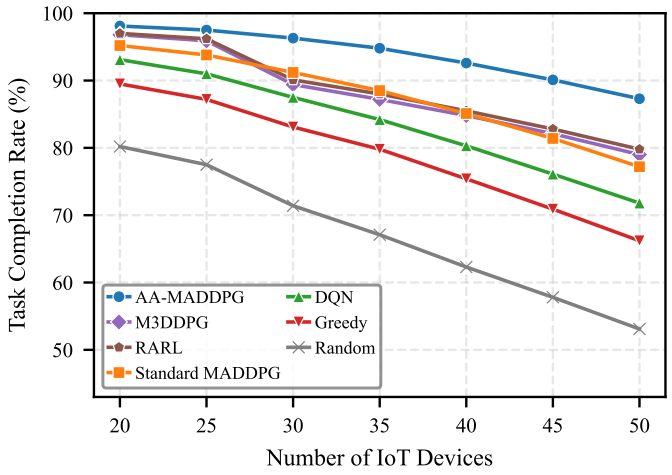

Figure 2: Task completion rate vs. number of IoT devices. AA-MADDPG maintains high completion rates even as network size increases.

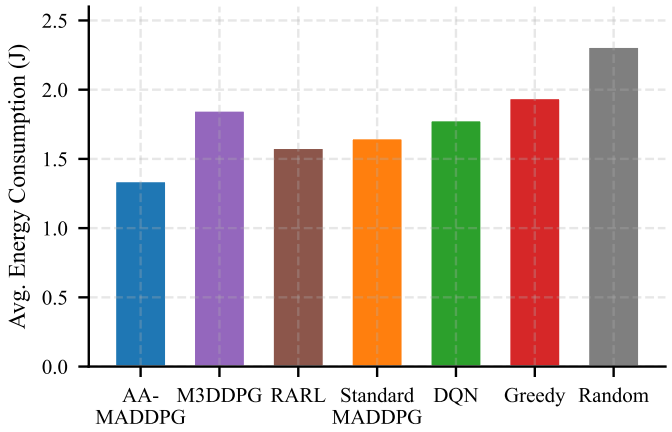

Figure 3: Average energy consumption comparison. AA-MADDPG achieves 19% energy savings compared to standard MADDPG.

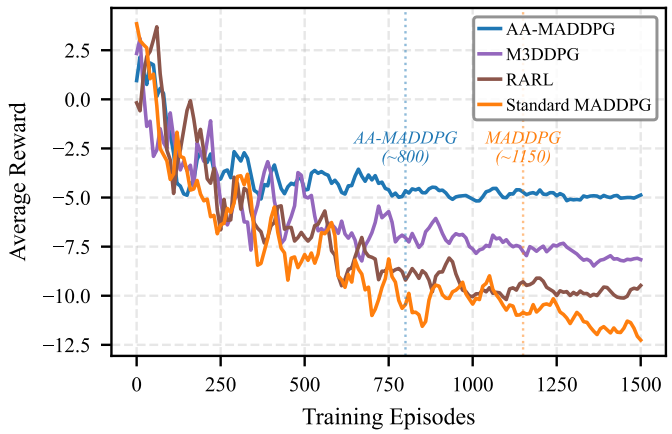

Figure 4: Training convergence comparison. AA-MADDPG converges 30% faster than standard MADDPG.

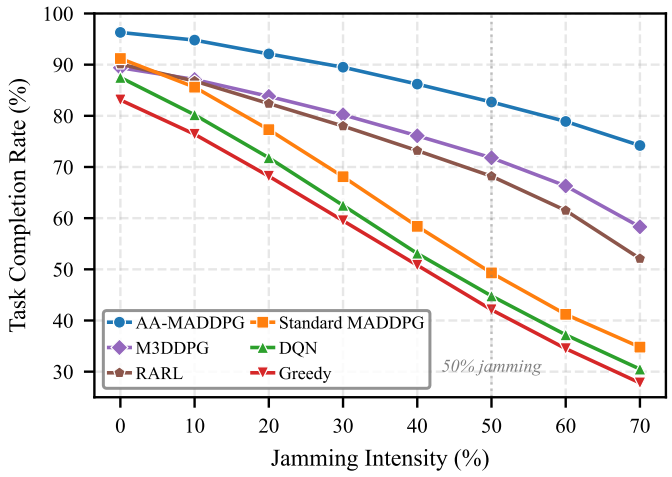

Figure 5: Task completion rate under various jamming intensities. AA-MADDPG maintains over 80% performance under 50% jamming.

4.3. Resilience Under Adversarial Conditions

Under jamming attacks with 50% interference intensity, M3DDPG maintains 71.8% task completion while RARL achieves 68.2%, both substantially below AA-MADDPG's 82.7%. At 70% jamming, the gap widens further: AA-MADDPG retains 74.2% of its baseline performance compared to 58.3% for M3DDPG and 52.1% for RARL.

This advantage arises from AA-MADDPG's ability to dynamically weight adversary models through Bayesian updating, whereas M3DDPG's fixed worst-case assumption leads to inefficient allocation of defensive resources when the actual attack deviates from the assumed worst-case profile.

4.4. Ablation Study

We evaluated the contribution of each rationality paradigm by testing AA-MADDPG variants using only Nash equilibrium, only level-k thinking, or only prospect theory. Results in Fig. 7 show that full AA-MADDPG with Bayesian averaging achieves 82.7% task completion under 50% jamming, compared to 76.4% (Nash only), 78.1% (level-k only), and 73.9% (prospect only). This 5–10% improvement validates the importance of combining multiple rationality paradigms.

The ARA module provides implicit attack detection by comparing predicted benign behavior with observed dynamics. At each epoch t, we compute $d(t) = D_{\mathrm{KL}}(P(a_{\mathrm{adv}} \mid O_t) \,\|\, P_0)$, where P_0 is the no-attack distribution, and flag an attack when $d(t)$ exceeds a threshold δ maintained as an exponential moving average over W=50 epochs. This yields an 87.3% detection rate with an 8.2% false positive rate under mixed attacks—an emergent property of ARA rather than a dedicated IDS; ROC analysis and comparison with anomaly-based methods are left for future work.

Parameter Sensitivity. Under 50% jamming, AA-MADDPG is robust to adversary model parameters. Varying the prospect theory loss-aversion coefficient $\lambda \in \{1.5, 2.25, 3.0\}$ yields task completion rates of 81.4%, 82.7%, and 81.9% (<1.6% spread); the probability weighting parameter $\gamma \in \{0.5, 0.69, 0.9\}$ produces 80.8%, 82.7%, and 82.1%. For level-k depth, k=2 achieves 82.7% versus 81.3% (k=1) and 82.4% (k=3), indicating diminishing returns consistent with behavioral game theory. This

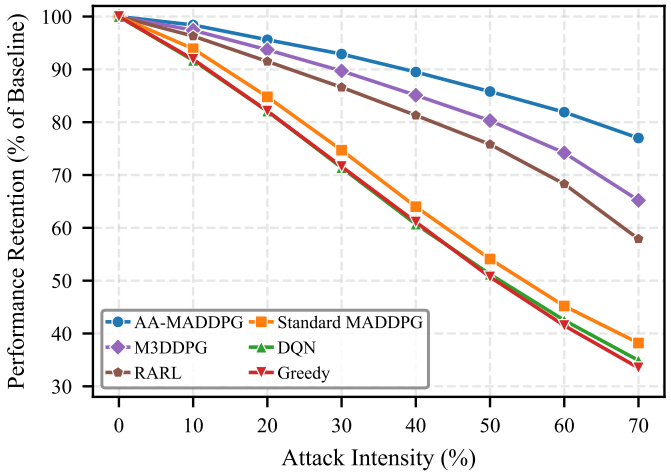

Figure 6: Performance retention under attacks. AA-MADDPG demonstrates superior resilience across all attack intensities.

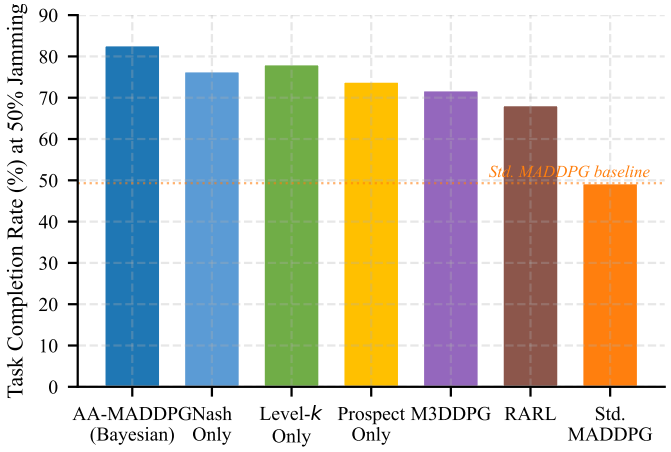

Figure 7: Ablation study showing the contribution of each rationality paradigm. Bayesian averaging provides superior robustness.

stability reflects the Bayesian averaging mechanism: adaptive mixture weights ω_k self-correct when individual paradigm parameters deviate from true adversary characteristics, confirming robustness across practically relevant parameterizations.

4.5. Computational overhead and scalability

Table 1 summarizes the computational cost of AA-MADDPG and the baseline approaches. The Bayesian update in the ARA module introduces an additional ≈ 3.2 ms per decision epoch (representing 6.4% of the 50 ms total inference latency), constituting a modest overhead relative to the resilience gains achieved.

Training requires approximately 4.8 hours on a single NVIDIA RTX 3090 GPU for the default scenario ($N = 4$ UAVs, $M = 30$ IoTDs), compared to 3.1 hours for standard MADDPG—a 55% increase attributable to adversary sampling and ARA feature computation during experience collection. The memory footprint scales as $O(K \times |S|)$ for the ARA module, where $K = 3$ denotes the number of rationality paradigms

Table 1: Computational cost comparison.

Method	Train Time (h)	Inference (ms)	Memory (MB)	Scalability (50 IoTDs)
AA-MADDPG	4.8	50	50.4	6.6 h (+38%)
Standard MADDPG	3.1	47	48.3	4.3 h (+39%)
M3DDPG	5.2	49	49.1	7.1 h (+37%)
DQN	1.8	12	22.7	2.4 h (+33%)
Greedy	—	< 1	—	—

and $|S|$ represents the state dimension, adding approximately 2.1 MB to the base MADDPG model (48.3 MB).

To assess scalability, we evaluated training time under increasing network sizes. Scaling the number of IoTDs from 30 to 50 increases training time by 38% (to 6.6 hours), while doubling the number of UAVs from 4 to 8 increases training time by 62% (to 7.8 hours), indicating manageable polynomial scaling for practical deployments.

4.6. Discussion, Limitations, and Future Work

AA-MADDPG demonstrates robust attack mitigation across three dimensions: under severe jamming (50% interference), task failure rates stay below 20% versus 40% for baselines; Bayesian-averaged multi-paradigm modeling maintains consistent effectiveness (±15% variation) across diverse attack types while single-model baselines suffer 35–50% drops under mismatched adversaries; and decentralized execution incurs only 50 ms per-epoch latency, enabling real-time UAV deployment. Two boundary conditions merit discussion. Non-stationary adversaries that mimic one rationality paradigm before switching may exploit the Bayesian updating lag; weakly informative Dirichlet priors ($\alpha = 1$) enable adaptation within $O(10)$ epochs, yet sufficiently deceptive adversaries can degrade performance during transition periods. Additionally, generalization across topologies remains unevaluated: the decentralized actor architecture should transfer across UAV/IoTD counts via local observation-conditioned policies, but the centralized critic requires retraining for substantially different configurations. Future work should address these through (i) meta-learning for rapid adaptation to novel attack strategies, (ii) transfer learning across network scales (e.g., N=4, M=30 to N=8, M=50), and (iii) differential privacy to prevent information leakage from defender behavior. Finally, our evaluation uses synthetic environments consistent with comparable MADRL frameworks [9, 10, 4]; trace-driven simulations using real-world aerial channel models (e.g., air-to-ground path loss models) and hardware-in-the-loop validation remain important future directions.

5. Conclusion

This paper presented AA-MADDPG, an adversarial-aware framework integrating ARA with multi-agent deep reinforcement learning for robust resource allocation in heterogeneous AAIoT networks. By modeling adversaries through three rationality paradigms combined via Bayesian model averaging, and embedding adversarial belief features into the critic architecture, the framework achieves 27% higher task completion and 19% lower energy consumption than baselines. Under 50% jamming, AA-MADDPG maintains over 80% task completion

while conventional and worst-case robust methods degrade significantly, establishing it as an effective and computationally tractable solution for secure aerial computing.

References

[1] H. Tran-Dang, D.-S. Kim, Digital twin-empowered intelligent computation offloading for edge computing in the era of 5g and beyond: A state-of-the-art survey, ICT Express (2025).

[2] J. Zhang, W. Xia, F. Yan, L. Shen, Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing, IEEE Access 6 (2018) 19324–19337.

[3] D. S. Lakew, A.-T. Tran, N.-N. Dao, S. Cho, Intelligent offloading and resource allocation in heterogeneous aerial access iot networks, IEEE Internet of Things Journal 10 (7) (2022) 5704–5718.

[4] T.-H. Nguyen, T. P. Truong, A.-T. Tran, N.-N. Dao, L. Park, S. Cho, Intelligent heterogeneous aerial edge computing for advanced 5g access, IEEE Transactions on Network Science and Engineering 11 (4) (2024) 3398–3411.

[5] J. Sharma, P. S. Mehra, Secure communication in iot-based uav networks: A systematic survey, Internet of Things 23 (2023) 100883.

[6] J. Rheey, H. Park, Robust hierarchical anomaly detection using feature impact in iot networks, ICT Express 11 (2) (2025) 358–363.

[7] A. Nabi, S. Moh, Offloading decision and resource allocation in aerial computing: A comprehensive survey, Computer Science Review 56 (2025) 100734.

[8] H. Guo, X. Zhou, J. Wang, J. Liu, A. Benslimane, Intelligent task offloading and resource allocation in digital twin based aerial computing networks, IEEE Journal on Selected Areas in Communications 41 (10) (2023) 3095–3110.

[9] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, W. Jiang, Multi-agent drl for task offloading and resource allocation in multi-uav enabled iot edge network, IEEE Transactions on Network and Service Management 18 (4) (2021) 4531–4547.

[10] H. Kang, X. Chang, J. Mišić, V. B. Mišić, J. Fan, Y. Liu, Cooperative uav resource allocation and task offloading in hierarchical aerial computing systems: A mappo-based approach, IEEE Internet of Things Journal 10 (12) (2023) 10497–10509.

[11] A. M. Seid, G. O. Boateng, S. Anokye, T. Kwantwi, G. Sun, G. Liu, Collaborative computation offloading and resource allocation in multi-uav-assisted iot networks: A deep reinforcement learning approach, IEEE Internet of Things Journal 8 (15) (2021) 12203–12218.

[12] P. Qin, Y. Fu, X. Zhao, K. Wu, J. Liu, M. Wang, Optimal task offloading and resource allocation for c-noma heterogeneous air-ground integrated power internet of things networks, IEEE Transactions on Wireless Communications 21 (11) (2022) 9276–9292.

[13] Y. Chen, K. Li, Y. Wu, J. Huang, L. Zhao, Energy efficient task offloading and resource allocation in air-ground integrated mec systems: A distributed online approach, IEEE Transactions on Mobile Computing 23 (8) (2023) 8129–8142.

[14] Y. Peng, X. Guang, X. Zhang, L. Liu, C. Wu, L. Huang, A cloud-edge collaborative computing framework using potential games for space-air-ground integrated iot, EURASIP Journal on Advances in Signal Processing 2024 (1) (2024) 54.

[15] C. Gao, X. Bian, B. Hu, S. Chen, H. Wang, Intelligent online offloading and resource allocation for hap drones and satellite collaborative networks, Drones 8 (6) (2024) 245.

[16] T. Lyu, Y. Xu, F. Liu, H. Xu, Z. Han, Task offloading and resource allocation for satellite-terrestrial integrated networks, IEEE Internet of Things Journal (2024).

[17] H. Peng, X. Shen, Multi-agent reinforcement learning based resource management in mec-and uav-assisted vehicular networks, IEEE Journal on Selected Areas in Communications 39 (1) (2020) 131–141.

[18] A. Garnaev, W. Trappe, An eavesdropping and jamming dilemma with sophisticated players, ICT Express 9 (4) (2023) 691–696.

[19] E. O. Frimpong, T. Kim, I. Bang, Deep learning approach for physical-layer security in gaussian multiple access wiretap channel, ICT Express 9 (4) (2023) 728–733.

[20] D. Banks, V. Gallego, R. Naveiro, D. Ríos Insua, Adversarial risk analysis: An overview, Wiley Interdisciplinary Reviews: Computational Statistics 14 (1) (2022) e1530.

[21] S. Li, Y. Wu, X. Cui, H. Dong, F. Fang, S. Russell, Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 4213–4220.

[22] L. Pinto, J. Davidson, R. Sukthankar, A. Gupta, Robust adversarial reinforcement learning, in: International conference on machine learning, PMLR, 2017, pp. 2817–2826.

[23] H. Pirayesh, H. Zeng, Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey, IEEE communications surveys & tutorials 24 (2) (2022) 767–809.