DRHRP: A Deep Reinforcement Learning–Based Hybrid Routing Protocol for UAV–Enabled Wireless Networks

Mahdi Zareei(✉) and Jesús Arturo Pérez Díaz

Tecnologico de Monterrey, School of Engineering and Sciences, Monterrey 64849, Mexico
{m.zareei,jesus.arturo.perez}@tec.mx

Abstract. In this paper, we propose DRHRP, a novel deep reinforcement learning (DRL)–based hybrid routing protocol for unmanned aerial vehicle (UAV)–enabled wireless networks. DRHRP integrates concepts from geocast routing, energy-efficient and connectivity maintenance strategies, and multi-agent DRL to dynamically adapt to the challenges of highly mobile and resource-constrained networks. By modeling the routing decision as a Markov Decision Process (MDP) and employing a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework, our protocol learns optimal forwarding policies using only local observations. Extensive simulations demonstrate improvements in packet delivery ratio, delay reduction, and energy consumption compared to traditional schemes.

Keywords: UAV networks · wireless ad hoc networks · deep reinforcement learning · hybrid routing · energy efficiency · connectivity maintenance

1 Introduction

The rapid proliferation of unmanned aerial vehicles (UAVs) in applications ranging from environmental monitoring and surveillance to disaster management and emergency communications has led to a growing demand for robust and efficient routing protocols tailored for Flying Ad Hoc Networks (FANETs) [1]. Unlike traditional mobile ad hoc networks (MANETs), FANETs operate in a three-dimensional space where the network topology changes rapidly due to the high mobility of UAV nodes and the dynamic nature of their operational environments. In such scenarios, conventional routing protocols, which often rely on static or slowly changing network assumptions, tend to perform suboptimally, suffering from high delays, frequent packet losses, and elevated energy consumption.

In response to these challenges, recent research has explored a variety of routing techniques. Early work on geocast routing demonstrated that geographic information could be exploited to confine message propagation within a targeted

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15649, pp. 401–409, 2025.
https://doi.org/10.1007/978-3-031-96997-3_25

area, thereby reducing unnecessary overhead [2]. Parallel research efforts focused on designing energy-efficient protocols that extend the lifetime of resource-constrained networks [3] and on maintaining network connectivity, especially in critical applications such as post-disaster transportation where continuous communication is essential [4]. Hybrid routing protocols, which blend proactive and reactive strategies, have emerged as a promising solution to balance the trade-offs between overhead and responsiveness [5].

The advent of deep reinforcement learning (DRL) has opened new avenues for network optimization. By framing routing decisions as a sequential decision-making problem, DRL-based approaches can dynamically adapt to changing network conditions, learning optimal forwarding strategies in real time [10–12]. In particular, multi-agent DRL methods such as MADDPG have shown great promise in cooperative environments where each agent (or node) makes decisions based solely on local observations, yet contributes to a global objective [6]. Motivated by these advances, this manuscript introduces DRHRP, a DRL-based hybrid routing protocol that leverages both proactive neighbor discovery and reactive forwarding to improve packet delivery, reduce delay, and conserve energy in FANETs.

This work not only proposes a novel routing framework but also provides a comprehensive mathematical analysis, detailed algorithm design, and extensive experimental evaluation. In addition to addressing the dynamic topology and limited energy resources, DRHRP is designed to operate effectively even in scenarios with varying node densities, mobility speeds, and beaconing intervals. The remainder of this paper is organized as follows. Section 2 offers an extended literature review. Section 3 describes the system model and problem formulation. Section 4 details the proposed DRHRP protocol and its DRL-based algorithm. Section 5 provides mathematical analysis of routing cost and convergence properties. Section 6 outlines the simulation setup and parameters. Section 7 presents and discusses the experimental results, and Section 8 concludes the paper with suggestions for future work.

2 Literature Review

Routing protocols in UAV networks have evolved significantly over the past decades. One of the earliest approaches was geocast routing, where the objective was to deliver messages to nodes within a specific geographic region. Maihöfer [2] provided one of the first comprehensive surveys of geocast routing protocols, highlighting methods such as flooding, directed flooding, and techniques to confine message propagation to a desired area. These early studies laid the groundwork for understanding how geographic information could be used to improve routing efficiency in dynamic networks.

Subsequent research shifted focus towards energy efficiency, as UAVs and sensor nodes typically operate on limited battery power. Tanoli and Khan [3] introduced routing protocols that not only aimed to minimize energy consumption but also sought to balance the load among nodes to prevent early node

failures. This work was critical in demonstrating that routing decisions must account for both immediate and long-term energy costs, a principle that later influenced many energy-aware routing strategies in wireless networks.

In parallel, the need to maintain network connectivity in highly dynamic environments spurred research into connectivity maintenance algorithms. Kurt et al. [4] proposed a distributed heuristic for maintaining connectivity in a swarm of drones, particularly under post-disaster conditions where conventional communication infrastructure might be unavailable. This line of work emphasized the importance of ensuring that even as individual nodes move or fail, the overall network remains connected–a challenge that is particularly acute in three-dimensional UAV networks.

Hybrid routing protocols emerged as a solution to the conflicting requirements of low latency and low overhead. Ravilla et al. [5] explored the combination of proactive and reactive routing approaches, arguing that maintaining local routing tables while discovering routes on-demand could offer the best of both worlds. These protocols were particularly effective in scenarios where the network topology changed frequently but localized information could be maintained with modest overhead.

More recently, the field of deep reinforcement learning (DRL) has begun to influence the design of routing protocols. The foundational work by Sutton and Barto [10] established the theoretical underpinnings of reinforcement learning, while subsequent studies by Lillicrap et al. [11] and Lowe et al. [12] demonstrated the practical applicability of DRL in continuous control and multi-agent environments. The application of DRL to networking problems has been explored in various contexts, including congestion control, power management, and, more recently, routing optimization. Qin et al. [6] extended multi-agent DRL frameworks to solve joint optimization problems in content ranking, showing that coordinated decision-making among agents can significantly outperform independent approaches.

Additional contributions to the literature include studies on traditional routing algorithms such as Dynamic Source Routing (DSR) [7] and Ad Hoc On-Demand Distance Vector (AODV) routing [8]. These protocols, while effective in certain scenarios, often suffer from scalability issues in highly dynamic networks like FANETs. Furthermore, research in wireless sensor networks by Akyildiz et al. [9] has underscored the importance of balancing energy consumption with network performance–a balance that is critical for UAV networks as well.

Despite these advances, a comprehensive framework that integrates the adaptability of DRL with the robustness of hybrid routing protocols for FANETs remains underexplored. Many studies have addressed specific aspects such as energy efficiency or connectivity maintenance, but few have combined these objectives with a DRL-based decision mechanism. The literature indicates a gap in designing protocols that can adapt in real time to rapidly changing network conditions while optimizing multiple performance metrics simultaneously. This motivates the development of DRHRP, which seeks to build on the strengths of

existing approaches while addressing their limitations through a unified DRL-based framework.

3 System Model and Problem Formulation

We consider a FANET represented as a dynamic graph $G(V, E)$ where V is the set of UAV nodes and E is the set of wireless links. Each UAV is equipped with a GPS receiver and periodically transmits beacon messages that include its location, residual energy, and queue status. This exchange of information enables nodes to maintain an up-to-date view of their local neighborhood even as the network topology evolves.

For each link $(i, j) \in E$, the quality metric q_{ij} is determined by factors such as the signal-to-noise ratio (SNR), frame error rate (FER), and the Euclidean distance d_{ij}. The energy consumption for transmitting a packet over link (i, j) is modeled as

$$E_{ij} = E_0 + E_d \, d_{ij}^{\eta} \tag{1}$$

where E_0 is a fixed energy cost, E_d is a distance-dependent factor, and η is the path-loss exponent.

Routing is formulated as a Markov Decision Process (MDP) with the tuple $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$. Here, $\mathcal{S}$ represents the state space (local neighborhood information, including node locations, residual energy, and queue lengths), and $\mathcal{A}$ represents the set of available actions (selecting a neighboring node for packet forwarding). The transition probabilities $P(s'|s, a)$ capture the dynamics of the network, and the reward function is defined as

$$r(s, a) = \alpha \cdot \text{PDR} - \beta \cdot \text{Delay} - \gamma \cdot \text{Energy} \tag{2}$$

with weighting factors α, β, and γ chosen to balance the trade-offs among packet delivery ratio (PDR), delay, and energy consumption. The goal is to maximize the cumulative reward over time:

$$R = \sum_{t=0}^{T} \gamma^t r_t \tag{3}$$

4 Proposed DRHRP Protocol

DRHRP integrates proactive neighborhood discovery with a DRL-based forwarding decision mechanism. Each UAV periodically broadcasts beacon messages containing its state information, allowing neighboring nodes to update their local views. A lightweight DRL agent, implemented using the MADDPG framework, observes the current state s_t and selects the next-hop neighbor a_t based on its policy $\pi(s_t)$. The agent's objective is to maximize the long-term cumulative reward by learning optimal forwarding strategies.

The Q-value update in the MADDPG framework is given by

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right) \tag{4}$$

and the policy gradient is approximated as

$$\nabla_{\theta^\pi} J \approx \mathbb{E}_{s \sim \mathcal{D}} \left[\nabla_a Q(s, a) \big|_{a=\pi(s)} \nabla_{\theta^\pi} \pi(s) \right] \tag{5}$$

Algorithmically, the forwarding decision at each node involves (i) observing the current state (ii) selecting an action with added exploration noise, (iii) forwarding the packet to the chosen neighbor, and (iv) updating the DRL agent based on the received reward. This integrated approach enables DRHRP to adapt dynamically to varying network conditions.

5 Mathematical Analysis

We analyze the total energy consumption and delay of a multi-hop path. Let L denote the number of hops from the source to the destination. The total energy consumption is approximated by

$$E_{\text{total}} \approx L \cdot \left(E_0 + E_d \, \bar{d}^{\eta} \right) \tag{6}$$

where $\bar{d}$ is the average hop distance. Similarly, the end-to-end delay D is modeled as

$$D = L \cdot (t_{\text{proc}} + t_{\text{tx}}) \tag{7}$$

with t_{proc} and t_{tx} representing processing and transmission delays per hop, respectively. Under the standard assumptions of MADDPG [11,12], the DRL agent's policy converges to a locally optimal solution, as characterized by the Bellman optimality equation:

$$Q^*(s, a) = \mathbb{E}_{s'} \left[r + \gamma \max_{a'} Q^*(s', a') \right] \tag{8}$$

6 Simulation Setup and Experimental Results

Simulations were conducted using NS-3 in a FANET scenario with 50 UAV nodes randomly deployed within a $1\,\text{km}^3$ area. The nodes follow a Gauss–Markov mobility model with a maximum speed of 100 km/h. Each node broadcasts beacon messages every 0.5 s, and the transmission range is set to 250 m. Table 1 summarizes the simulation parameters.

The experimental evaluation focuses on several key performance metrics: packet delivery ratio (PDR), average delay, energy consumption, and routing overhead. In the following, we describe the experiments in detail.

In the first experiment, we vary the node density by simulating networks with 30, 50, and 100 UAV nodes. Table 2 presents the results. The table shows that

Table 1. Simulation Parameters

Parameter	Value
Number of UAVs	50
Simulation Area	1 km^3
Mobility Model	Gauss–Markov
Maximum Speed	100 km/h
Beacon Interval	0.5 s
Transmission Range	250 m
Path-loss Exponent (η)	2–4
Simulation Duration	500 s

Table 2. Node Density Variation Results

Nodes	Protocol	PDR (%)	Delay (ms)	Energy (J)	Overhead (packets)
30	HRP	84.0	40	1.10	180
	HEER	88.0	35	0.95	165
	DRHRP	92.0	28	0.80	150
50	HRP	87.2	35	1.20	210
	HEER	90.1	28.5	0.90	195
	DRHRP	95.3	22.4	0.70	170
100	HRP	80.0	50	1.50	250
	HEER	84.5	42	1.20	230
	DRHRP	90.0	35	1.00	210

DRHRP consistently achieves higher PDR and lower delay than the baseline protocols (HRP and HEER), even as the network size increases. The reduction in energy consumption and overhead is also notable, which is attributed to the adaptive routing decisions made by the DRL agent.

In the second experiment, the impact of mobility speed is evaluated for speeds of 30 km/h, 100 km/h, and 200 km/h (see Table 3). The results indicate that while all protocols suffer performance degradation at higher speeds, DRHRP still maintains a superior performance by leveraging its DRL agent, which adapts to the rapid topology changes.

The third experiment investigates the effect of varying the beacon interval (0.5 s, 1.0 s, and 2.0 s) on performance, as shown in Table 4. Shorter beacon intervals tend to improve the freshness of neighborhood information, resulting in higher PDR and lower delay. However, this comes at the cost of increased control overhead. DRHRP benefits significantly from more frequent updates, as the DRL agent makes more accurate forwarding decisions.

Finally, we conduct a sensitivity analysis on the reward function parameters for DRHRP. Table 5 shows the impact of three different weight sets on the

Table 3. Mobility Speed Variation Results

Speed (km/h)	Protocol	PDR (%)	Delay (ms)	Energy (J)	Overhead (packets)
30	HRP	90.0	25	1.00	190
	HEER	92.0	20	0.80	180
	DRHRP	97.0	15	0.60	160
100	HRP	87.2	35	1.20	210
	HEER	90.1	28.5	0.90	195
	DRHRP	95.3	22.4	0.70	170
200	HRP	75.0	60	1.60	230
	HEER	80.0	50	1.30	215
	DRHRP	85.0	45	1.10	200

Table 4. Beacon Interval Variation Results

Beacon Interval (s)	Protocol	PDR (%)	Delay (ms)	Energy (J)	Overhead (packets)
0.5	HRP	87.2	35	1.20	210
	HEER	90.1	28.5	0.90	195
	DRHRP	95.3	22.4	0.70	170
1.0	HRP	85.0	38	1.30	220
	HEER	88.0	32	1.00	205
	DRHRP	93.0	26	0.90	190
2.0	HRP	80.0	45	1.40	240
	HEER	83.0	40	1.10	225
	DRHRP	89.0	34	1.00	210

protocol's performance. The analysis demonstrates that increasing the weight on PDR (weight set B) improves performance metrics across the board, while a higher emphasis on delay (weight set C) can lead to a trade-off with overall packet delivery.

Table 5. Reward Weight Sensitivity Results (DRHRP)

Weight Set	α	β	γ	PDR (%)	Delay (ms)	Energy (J)
A	1	1	1	92.0	25	0.80
B	2	1	1	95.3	22	0.70
C	1	2	1	90.0	28	0.75

Overall, the experimental results confirm that DRHRP consistently outperforms traditional hybrid protocols such as HRP and HEER in diverse network

scenarios. The adaptive forwarding decisions made by the DRL agent contribute to a higher packet delivery ratio, lower delay, and reduced energy consumption, making DRHRP a strong candidate for real-world FANET deployments.

7 Additional Analysis and Future Work

The results presented indicate that DRHRP is robust under varying network conditions, including changes in node density, mobility, and beacon intervals. However, there remains significant potential to further optimize the protocol. Future work will focus on integrating real-world UAV mobility traces to refine the DRL model further and extending the framework to support multi-destination geocast. Moreover, advanced multi-agent cooperation strategies are under investigation to enhance scalability and robustness in larger networks.

8 Conclusion

In this manuscript, we introduced DRHRP, a deep reinforcement learning–based hybrid routing protocol tailored for UAV–enabled wireless networks. By combining proactive neighbor discovery with a MADDPG–driven reactive forwarding mechanism, DRHRP effectively addresses the challenges of dynamic topology and resource constraints inherent in FANETs. The extended mathematical analysis and comprehensive simulation results demonstrate the superiority of DRHRP over conventional routing schemes in terms of packet delivery, delay, energy efficiency, and overhead. Future research will include real-world implementation and further enhancements in multi-agent cooperation to tackle even more challenging network environments.

References

1. Bekmezci, I., Sahingoz, O.K., Temel, Ş: Flying ad hoc networks (FANETs): a survey. Ad Hoc Netw. **11**(3), 1254–1270 (2013)
2. Maihöfer, C.: A survey of geocast routing protocols. IEEE Commun. Surv. Tutorials **6**(1), 2–16 (2004)
3. Tanoli, T.S., Khan, M.K.: Energy efficient and QOS sensitive routing protocol for ad hoc networks. IOP Conf. Ser. Mater. Sci. Eng. **51**, 012026 (2013)
4. Kurt, A., Saputro, N., Akkaya, K., Uluagac, A.S.: Distributed connectivity maintenance in swarm of drones during post-disaster transportation applications. IEEE Trans. Intell. Transp. Syst. **22**(9), 6061–6074 (2021)
5. Ravilla, D., Sumalatha, V., Reddy Putta, C.S.: Hybrid routing protocols for ad hoc wireless networks. Int. J. Ad hoc, Sensor Ubiquitous Comput. **2**(4), 79–90 (2011)
6. Qin, Z., Yuan, K., Lahiri, P., Liu, W.: Cooperative multi-agent deep reinforcement learning in content ranking optimization. In: Proceedings of the eCom 2024, Washington, DC (2024)
7. Perkins, C., Royer, E.: Dynamic source routing in mobile ad hoc networks. In: Proceedings 2nd IEEE Workshop on Mobile Computing Systems and Applications (1999)

8. Johnson, D.B., Maltz, D.A.: Dynamic source routing in ad hoc wireless networks. In: Mobile Computing, pp. 153–181 (1996)
9. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**(4), 393–422 (2002)
10. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction, 2nd edn., MIT Press (2018)
11. Lillicrap, T.P., et al.: Continuous control with deep reinforcement learning. In: Proceedings of the ICLR (2015)
12. Lowe, R. et al.: Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings if the NeurIPS (2017)