A Q-Learning and Fuzzy Logic based Routing Protocol for UAV Networks

Sijin Huang*, Jie Tang*, Zihao Zhou*, Guangguang Yang‡, Maksim V. Davydov§ and Kai Kit Wong¶
*School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China
‡School of electronic information engineering, Foshan University, Foshan, China
§Department of Theory of Electrical Engineering, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus
¶Department of Electronic and Electrical Engineering, University College London, WC1E 6BT London, UK
Email: skinhuang@foxmail.com, eejtang@scut.edu.cn, eezihaozhou@gmail.com, guangguangyanguop@gmail.com, davydov-mv@bsuir.by, kai-kit.wong@ucl.ac.uk

***Abstract*—With the development of ad hoc network technology, unmanned aerial vehicle (UAV) swarm has demonstrated significant promise across civil and military domains. However, owing to unique attributions, such as high dynamic topology, 3-D mobility and low density, it's extremely challenging to establish a reliable and robust communication between flying nodes. In this paper, we proposed a Q-Learning and Fuzzy Logic based Routing Protocol (QFRP) for UAV networks, which adopted an efficient Q-value update mechanism based on HELLO and ACK. In this mechanism, we take neighbor set coherence and link lifetime into account. Since routing exploration has an important impact on routing performance, we proposed a fuzzy logic based mechanism for exploration and exploitation that considers Q-value, link quality and access delay to mitigate the blindness of random exploration. Simulation results demonstrate that QFRP can make efficient routing decisions within dynamic multi-hop UAV networks, and outperforms existing protocols regarding packet delivery ratio (PDR), end-to-end (E2E) delay, and routing overhead.**

***Index Terms*—UAV swarm, Q-Learning, fuzzy logic, routing protocol, wireless communication, FANETs.**

I. Introduction

Unmanned aerial vehicles (UAV) swarms have become widespread in various applications and services in both civil and military domains due to their flexible networking, easy deployment and high maneuverability [1]. Without preset network infrastructure and control center, UAV swarm establishes an interconnected flying ad hoc network (FANET) through cooperative and collaboration. Distinct from vehicular ad hoc networks (VANETs) and mobile ad hoc networks (MANETs), nodes within FANETs have rapid mobility resulting in rapid topology changes [2], making the designed of routing protocols quite challenging.

There have been many classical routing protocols which can be classified into three categories: proactive (e.g., DSDV [3]), reactive (e.g., GRAd [4]), and geographic (e.g., GPSR [5]) routing protocols. Proactive protocol creates routes before packets are forwarded, which requires it to maintain the routing table that results in high routing overhead. Reactive protocol creates routes when packets need to be sent, which brings higher delay due to routing discovery. Geographic protocol forward packets based on the locations of packets' destinations, which makes strong assumptions about the known destination positions. In addition, the aforementioned traditional routing protocols exhibit restricted limited adaptability and flexibility due to insufficient intelligent awareness of communication environments and are not suitable for FANETs.

In recent years, Reinforcement Learning (RL) techniques have been used in FANETs to address routing problems [6]. In [7], Liu *et al.* introduced a Q-Learning based routing protocol known as QMR, focusing on optimizing multiple objectives such as one-hop delay and remaining energy. Besides, adaptive Q-Learning factors and a new mechanism for the exploration and exploitation based on weighted Q-value were used in QMR. In [8], Xue *et al.* proposed a Q-Learning empowered highly dynamic and latency-aware routing algorithm (QEHLR), which considered link remaining time and deleted predicted failed links to avoid connection loss. Extending the local view of the network topology to two-hop neighbors range, Arafat *et al.* [9] proposed a Q-Learning based topology-aware routing protocol (QTAR). Further, fuzzy logic is used in combination with Q-Learning based protocols due to its flexible framework for dealing with complex and uncertain variables. In [10], He *et al.* proposed a fuzzy logic reinforcement learning based routing algorithm for FANET and the fuzzy output was used to calculate action reward. In [11], Wu *et al.* employed fuzzy logic to assess wireless links by considering various metrics, including signal strength, relative vehicle movement and available bandwidth. The fuzzy output affected the learning rate.

However, the Q-Learning based routing protocol using data packets for exploration and update Q-value by ACK is considered inefficient for several reasons. First, this will inevitably lead to some exploratory packets being forwarded to unexpected relays, resulting in a decrease in packet delivery ratio. Second, the broken links cannot be discovered in time until next exploratory. Last but not least, if the ACK packet can not successfully received due to collision or other reasons, UAVs will not be able to gain experience to update Q-value.

In [1], Zhou *et al.* proposed a bidirectional Q-Learning routing protocol (BQLAODV) to improve Q-value iteration speed. In [12], utilizing node average Q-values, Wei *et al.* proposed link pre-learning and multi-Q learning to speed

up Q-Learning process. In [13], Sliwa *et al.* presented a Predictive Ad-hoc Routing fueled by Reinforcement learning and Trajectory knowledge, periodically generating and flood forwarding chirp message to update Q-value.

This inspires the Q-Learning and Fuzzy Logic based Routing protocol (QFRP) proposed in this paper. We take multiple metrics into account on routing including node access delay, link quality, link lifetime, and neighbor set coherence. A new Q-value update mechanism both based on HELLO and ACK packets is proposed. Moreover, we achieve a reasonable compromise between exploration and exploitation under network condition based on a flexible fuzzy system. The simulation results illustrate that QFRP accomplishes significant performance improvement in packet delivery ratio(PDR), end-to-end (E2E) delay and routing overhead compared with different routing protocols.

The remainder of this paper is organized as follows. In next section, we presents the system model. In Section III, the routing protocol proposed is described. Section IV includes the simulation results and discussions. Finally, Section V concludes this paper.

II. SYSTEM MODEL

In this paper, QFRP is designed for UAV network where consists a series of UAV nodes $\mathcal{U}$, denoted as $\mathcal{U} = [\mathcal{U}_0, \mathcal{U}_1, ..., \mathcal{U}_n]$. Each UAV in the networks can serve as a source, target or relay node. And the same omni-directional antenna with maximum communication range D_{max} is attached to each UAV. If the Euclidean distance $d_{\mathcal{U}_i\mathcal{U}_j}$ between UAV $\mathcal{U}_i$ and $\mathcal{U}_j$ (where $\mathcal{U}_i, \mathcal{U}_j \in \mathcal{U}$) is within the limit of D_{max}, a wireless transmission link can be established. Moreover, the links are symmetrical: if $\mathcal{U}_i$ can receive messages from $\mathcal{U}_j$, then $\mathcal{U}_j$ can receive messages from $\mathcal{U}_i$ as well. Each drone is equipped with Global Navigation Satellite Systems (GNSS) to determine its position, speed, and direction of motion. And each UAV periodically exchange the HELLO packets to sense the presence of neighbor UAVs $\mathcal{N}$. The neighbors of UAV $\mathcal{U}_i$ can be defined as $\mathcal{N}_i$.

III. PROPOSED ROUTING PROTOCOL: QFRP

A. HELLO message format

In QFRP, HELLO messages and ACK messages play a crucial role in updating Q-values. In addition, HELLO messages are used to update neighbor table, which is closely related to routing decisions. The detail fields of HELLO messages are shown in Fig.1. ACK messages contain *Originator ID*, *Position*, *Velocity*, *Coherence* and *V* fields at least.

Originator ID	Position	Velocity	HRR	AD	Coherence	V	Creation Time

Fig. 1: Format of HELLO message

- *Originator ID:* The identity of the drone originating HELLO message. Each drone has a unique identity.
- *Position and Velocity:* The current position and velocity of the drone originating this message obtained. It's used to calculate the link lifetime prediction (see Sec. III-C).
- *HRR:* The received ratio (HRR) of HELLO messages from neighbors. Upon origination, the HRR is calculated (see Sec. III-D) and filled into this field.
- *AD:* The access delay (AD) includes contention delay and queuing delay (see Sec. III-D).
- *Coherence:* A factor for measuring the neighbor set coherence (see Sec. III-C).
- *V:* The reverse path score to every destination (see Sec. III-C).
- *Creation Time:* The message creation time.

B. Neighbor Table

Each UAV maintains a dynamic Neighbor table, which plays an important role in routing decision. Each table entry holds state information about a neighbor drone, as shown in Fig.2.

Neighbor ID	AD	HRR	Time

Fig. 2: Format of neighbor table entry

1) Neighbor Table format:

- *Neighbor ID:* The id of a neighbor drone to which this neighbor table entry refers.
- *AD:* The access delay of the neighbor drone. When a HELLO message arrived, this field is set to the AD field of the HELLO packet.
- *HRR:* HELLO packets received ratio.
- *Time:* Creation time of the latest received HELLO packet.

2) Neighbor Table Maintenance: When a HELLO message is received at a drone, if no matching entry is found, a new neighbor table entry is created.Otherwise, the corresponding entry will be overwritten. The neighbor table will be kept 'fresh' by removing timeout entries whose creation time to the current time are greater than the maximum entry lifetime. Based on the fresh neighbor table, routing decision is made (see Sec. III-D).

C. Q-Learning Based Route Discovery and Route Maintenance

1) Q-Leaning model: Q-learning is a value-based reinforcement learning algorithm that does not rely on a model, and is utilized to address control issues within Markov decision processes, commonly seen in ad hoc networks routing protocols. In QFRP, we model the routing process, also known as packets transmission, as a Markoc process, which can be defined as a 4-tuple $< S, A, R, P >$. In this tuple, $S \in \mathcal{U}$ denotes the current state, indicating the location of the packet. Action $A \in \mathcal{N}_S$ corresponds to the selection of the next hop node. Reward *R* signifies the reward following the execution of an action, and *P* is the state transition probability. Q-Learning aims to learn the state-action values iteratively to obtain the optimal policy. The modified update rule in QFRP is as follows:

$$\begin{aligned} Q_{\mathcal{U}_i}^{t+1}(\mathcal{U}_d, \mathcal{U}_a) =& Q_{\mathcal{U}_i}^{t}(\mathcal{U}_d, \mathcal{U}_a) + \\ & \alpha \left[\gamma V(\mathcal{U}_d, \mathcal{U}_a) - Q_{\mathcal{U}_i}^{t}(\mathcal{U}_d, \mathcal{U}_a)\right] \end{aligned} \tag{1}$$

where $Q_{\mathcal{U}_i}^{t+1}(\mathcal{U}_d,\mathcal{U}_a)$ is the Q-value of $\mathcal{U}_i$ selecting $\mathcal{U}_a$ to relay the packet whose destination is $\mathcal{U}_d$ at iteration $t+1$. α is the learning rate, and γ is the discounted factor. $V_{\mathcal{U}_a}$ is the reverse path score to $\mathcal{U}_d$ via forwarder $\mathcal{U}_a$ which is contained in the HELLO or ACK message and is defined as:

$$V(\mathcal{U}_d,\mathcal{U}_a) = \begin{cases} 1, & \mathcal{U}_\mathrm{a} \text{ is destination} \\ \underset{\mathcal{U}_b \in \mathcal{N}_a}{max}\ Q_{\mathcal{U}_a}^t(\mathcal{U}_d,\mathcal{U}_b), & \text{otherwise} \end{cases} \tag{2}$$

The adaptive discount factor γ serves as a multidimensional routing metric, defined as:

$$\gamma = \gamma_0 \times \gamma_{LLT(\mathcal{U}_i,\mathcal{U}_j)} \times \gamma_{Coh(\mathcal{U}_j)} \tag{3}$$

where γ_0 operates as a time constant to ensure the avoidance of loops in routing by implementing a guaranteed metric degradation per hop.

$\gamma_{LLT(\mathcal{U}_i,\mathcal{U}_j)}$ is a variable factor based on the Link lifetime prediction (LLT), and is computed as:

$$\gamma_{LLT(\mathcal{U}_i,\mathcal{U}_j)} = \begin{cases} \sqrt{\dfrac{LLT(\mathcal{U}_i,\mathcal{U}_j)}{\tau_0}}, & LLT(\mathcal{U}_i,\mathcal{U}_j) < \tau_0 \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where τ_0 is a given prediction horizon, set to 2.5 second. And LLT t can be calculated by solving a quadratic equation $(\Delta P_x+\Delta V_x t)^2+(\Delta P_y+\Delta V_y t)^2+(\Delta P_z+\Delta V_z t)^2 = D_{max}^2$, where $\Delta P = P_{\mathcal{U}_j} - P_{\mathcal{U}_i}$ is the relative position, $\Delta V = V_{\mathcal{U}_j} - V_{\mathcal{U}_i}$ is the relative velocity, and D_{max} is the maximum communication range.

$\gamma_{Coh(\mathcal{U}_j)}$ is factor to gauges the neighbor set coherence of node $\mathcal{U}_j$ based on the disparity amid the neighboring sets $\mathcal{N}_j(t)$ and $\mathcal{N}_j(t-\Delta t)$ in $\Delta t = 2.5s$. According to [14], it is derived as:

$$\gamma_{Coh(\mathcal{U}_j)} = \begin{cases} 0, & \mathcal{N}_j(t) \cup \mathcal{N}_j(t-\Delta t) = \emptyset \\ \sqrt{\dfrac{\mathcal{N}_j(t) \cap \mathcal{N}_j(t-\Delta t)}{\mathcal{N}_j(t) \cup \mathcal{N}_j(t-\Delta t)}}, & \text{otherwise} \end{cases} \tag{5}$$

After defining the reverse path score and the adaptive discount factor, we can maintain Q-table by computing the corresponding Q-value according to (1).

2) Update Mechanism: In this paper, we propose a new Q-value update mechanism, instead of flooding update messages or ACK based method. In the considered network, every drone can serve as a message destination. HELLO packets carrying $V(\mathcal{U}_i,\mathcal{U}_i) = 1$ are periodically broadcast by every node $\mathcal{U}_i$ to realize routing discovery and provide alternate routing options. When a HELLO packet is received, the Q-value is updated. When a node transmits a data packet and receive the ACK, the Q-value is also updated rather than waiting until next hello packet reception, in order to discover link breaks and adapt to network changes. To improve the accuracy of Q-table, if the ack timeout retransmission occurs for packets sent to next hop node $\mathcal{U}_i$, the Q-values through node $\mathcal{U}_i$ will be punished by a punishment factor of 0.8 decay.

It is worth noting that unlike the basic Q-learning algorithm, in QFRP a single learning iteration (packet reception) updates the Q-values for all directions (to all destinations), achieving a faster Q-value iteration. It's more efficient than ACK mechanism and less overhead than the flood forwarding mechanism.

D. Fuzzy Logic Based Routing Decision

A new mechanism for the exploration and exploitation is proposed in QFRP. It balances exploration and exploitation by considering link information, including node access delay, link quality and Q-value. Instead of selecting the next hop node with the highest Q-value directly for routing decisions, the next hop nodes' scores are evaluated by fuzzy logic system, and the node with the highest score is chosen.

1) Fuzzy Logic: In FANET, owing to the rapid movement of UAV and the fading feature of wireless channel, the wireless communication link is highly dynamic. The selection of the optimal next hop is heavily influenced by the communication environment, which makes the mathematical model of the routing problem complicated or inflexible. Fuzzy logic provides a solution for handling imprecise and uncertain data. By employing a fuzzy system, we address this problem without mathematical model derivation. The inputs of this fuzzy system to be considered for routing are Q-value, Link Quality (LQ) and Access Delay (AD).

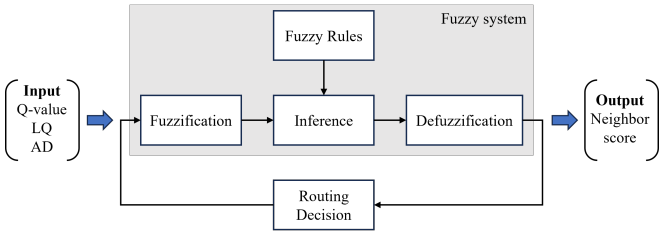

Fig. 3: Proposed fuzzy system structure

2) Structure: The fuzzy system architecture of the proposed mechanism is illustrated in Fig.3, including fuzzification, fuzzy inference and defuzzification. The multiple input factors of the system need to be calculate in advance. The following is the procedure for the fuzzy logic based routing decision.

3) Calculation of input factors: Before origination of a HELLO message, the link quality factor and access delay factor should be calculated. And the updated Q-value in the previous section is taken without any modification as an input to the fuzzy system, ranging between 0 and 1.

Link Quality (LQ) Factor: In FANETs, the transmission success rate between two nodes can represents the link quality. In this paper, we use HELLO received ratio (HRR) to evaluate link quality. For precise estimation in dynamic environments where topology changes frequently and packet collisions may occur, a sliding window size equivalent to ten HELLO intervals is used. A UAV node keeps a record of every neighbors' HELLO packet reception within the sliding window size to estimate HRR. The HELLO packets are sent with a defined time interval, so that HRR can be computed by each drone. Upon origination of a HELLO packet, the HRR is calculated by the record of HELLO packet reception within ten HELLO intervals, defined as:

$$HRR_{\mathcal{U}_d}^{\mathcal{U}_s}(t) = \begin{cases} \frac{count_{\mathcal{U}_d}^{\mathcal{U}_s}}{\lceil t/\tau \rceil} \times (1 - 0.5^{count_{\mathcal{U}_d}^{\mathcal{U}_s}}), & t < w \\ \sum_{i}^{count_{\mathcal{U}_d}^{\mathcal{U}_s}} \omega_i r_i, & t \geq w \end{cases} \tag{6}$$

In equation (6), $HRR_{\mathcal{U}_d}^{\mathcal{U}_s}(t)$ denotes the $\mathcal{U}_s$ HELLO packet received ratio of $\mathcal{U}_d$, $count_{\mathcal{U}_d}^{\mathcal{U}_s}$ is the amount of reception record within sliding window time w, τ is the HELLO interval and $\lceil t/\tau \rceil$ is the expected amount of reception. For a more precise estimation of HRR, we think that recently successfully received HELLO packets carry higher weights ω_i, $\sum_i^{count_{\mathcal{U}_d}^{\mathcal{U}_s}} \omega_i = 1$, r_i denotes the i-th reception record. Therefore, before routing decision, we have:

$$LQ_d^s(t) = \begin{cases} HRR_{\mathcal{U}_d}^{\mathcal{U}_s}(t), & \mathcal{U}_s \in \mathcal{N}_d \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Access delay (AD) Factor: The end-to-end delay is a crucial metric in UAV networks routing protocols. To enhance end-to-end latency, we consider not only link quality and Q-values in routing decisions but also the access delay of nodes. In QFRP, when a UAV node transfer a packet, it calculates its own access delay (AD). Upon HELLO packet origination, AD will be filled into the message for informing neighbor drones. Assuming that node $\mathcal{U}_i$ sends a packet, node $\mathcal{U}_i$ calculates the access delay using the channel contention delay and the queuing delay. Access delay can be expressed as follows:

$$delay_{\mathcal{U}_i} = delay_{con} + delay_{que} \tag{8}$$

where $delay_{con}$ represents the contention delay within the channel, which is defined as the duration required by the medium access protocol to send the packet. $delay_{que}$ indicates the queuing delay, described as the period it takes for the packet to reach the front of transmission queue. The fuzzy input AD is normalized as:

$$AD = \frac{delay_{\mathcal{U}_i} - D_{min}}{D_{max} - D_{min}} \tag{9}$$

4) Fuzzification: Fuzzification is the procedure of converting numerical inputs into fuzzy values via Fuzzy Membership function (FM), which is shown in Fig.4. The ordinate indicates the degree of membership, ranging from 0 to 1.

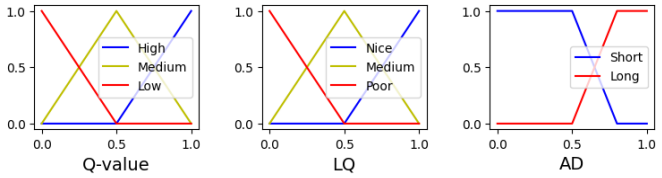

Fig. 4: Input fuzzy membership function

5) Fuzzy rules and fuzzy inference: In a fuzzy rule, the premises and conclusions correspond to the fuzzy input and output sets respectively. The three fuzzy values can be mapped by IF/THEN rules, and neighbor scores are mapped into six levels: Perfect, Good, Acceptable, Unpreferable, Bad, Very bad, as shown in Table I.

TABLE I: Fuzzy Rules

Rule	AD	LQ	Q-value	Rank
1	Short	Nice	High	Perfect
2	Short	Medium	High	Good
3	Short	Poor	High	Acceptable
4	Short	Nice	Medium	Acceptable
5	Short	Medium	Medium	Unpreferable
6	Short	Poor	Medium	Bad
7	Short	Nice	Low	Unpreferable
8	Short	Medium	Low	Bad
9	Short	Poor	Low	Very bad
...	...	...	...	...
18	Long	Poor	Low	Very bad

6) Defuzzification: Defuzzification serve as the last step in fuzzy logic. The neighbor scores obtained from IF/THEN rules will be transferred into numerical values based on the output membership function defined as in Fig.5. In this paper, we use Center of Gravity (COG) method for defuzzification. By applying this fuzzy system, the score of each neighbor node is computed, and the node with the highest score is selected as the next hop.

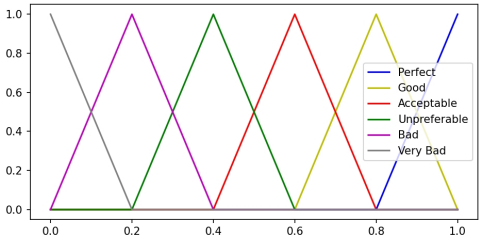

Fig. 5: Output fuzzy membership function

IV. SIMULATION RESULTS

A. Simulation Scenario

In this section, we construct several experiments to show the superiority of our proposed routing protocol by comparing it with GRAd [4] and PARRoT [13]. We use a FANET simulation platform based on Python3 to implement QFRP and make the comparison.

TABLE II: Simulation Parameters

Parameters	Values
Simulation Area	700m × 700m × 120m
Number of UAVs	10-30
Velocity of UAVs	10-50 m/s
Antenna	Omni-directional
UAV Transmit Power	1.0 W
SINR Threshold	2 dB
MAC	IEEE 802.11n
Mobility Model	Gauss Markov Mobility Model
HELLO Interval τ	0.5 s
Neighbor Entry Lifetime	1.0 s
Learning Rate α	0.5
Discounted Factor γ_0	0.8

For the considered scenario, UAVs are initially randomly distributed in a 3-D space of 700m × 700m × 120m. Then, UAV nodes move at a speed of 10-50 m/s based on Gauss Markov Mobility Model, reflecting the impact of mobility on routing performance. The number of UAV nodes varies

between 10 and 30, reflecting the impact of density on routing performance. Furthermore, a HELLO message is generated every 0.5 second and the life time for each neighbor entry is 1.0 second. The detailed parameters set in our simulation are summarizes in Table II. To evaluate QFRP, we consider three performance metrics: packet delivery ratio (PDR), average end-to-end (E2E) delay and normalized routing overhead [15].

B. Impact of UAV Velocity

We firstly assess the influence of various UAV velocities on network performance. We maintain a constant number of UAVs at 15, with a packet sending rate of about 2Hz. Fig.6 illustrates the PDR under different velocities of drones. Our proposed protocol, which considers link quality and neighbor set coherence, consistently achieves the highest PDR performance across different node velocities. As the velocity of drones increases gradually, for PARRoT, layer-by-layer forwarding chirp packets through flooding to update Q-value, the Q-value will be inaccurate due to congestion, collision and retransmission delay, which affects the next hop selection. For GRAd, the rapidly changing topology leads to frequent links disconnection, which greatly increases the probability that route reply packet cannot be delivered to the requester, and thus affects the PDR performance. In comparison, QFRP updates Q-value through HELLO and ACK packets instead of flooding to reduce the network load, and considers neighbor set coherence and link quality to ensure that packets can be successfully forwarded. In addition, QFRP updates the Q-value of each neighbor relay, so that it can quickly switch to an alternative path when the link is disconnected. Consequently, QFRP exhibits a higher PDR than PARRoT and GRAd.

Fig.7 shows the average E2E delay of the compared protocols for different drone velocities. When the velocity of drones increases gradually, these three protocols has a growth in different degree. GRAd shows the longest E2E delay due to the reactive routing form and frequent link breaks. For PARRoT, the inaccurate estimation of Q-value leads to the selection of invalid next hop, which in turn leads to a large number of packet retransmissions and worsens the E2E delay. In comparison, QFRP considers access delay and link conditions by fuzzy logic technology on routing, maintains the average E2E delay at a low level.

Fig.8 shows the normalized routing overhead under different velocities of drones. As the velocity of drones gradually increases, there is a tendency for the normalized routing overhead of the compared protocols to grow. PARRoT shows the largest routing overhead because its layer-by-layer chirps flooding. For GRAd, frequent link breaks cause a large number of routing requests and thus affects routing overhead. For QFRP, compared with GRAd, the routing overhead can be reduced by 73.13%, when the velocity of drones is 50 m/s.

C. Impact of UAV Node Density

Herein, we explore the influence of different node densities in UAV networks. The number of UAV nodes range from 10 and 30. The UAV velocity and packet sending rate are set to 20m/s and about 2 Hz respectively. Fig.9 illustrates how

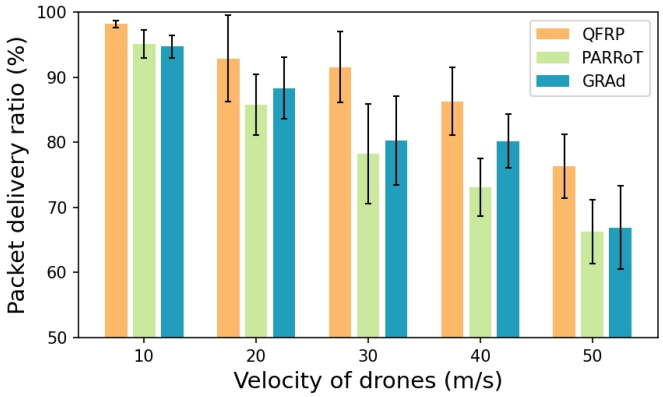

Fig. 6: PDR for varying UAV node velocities

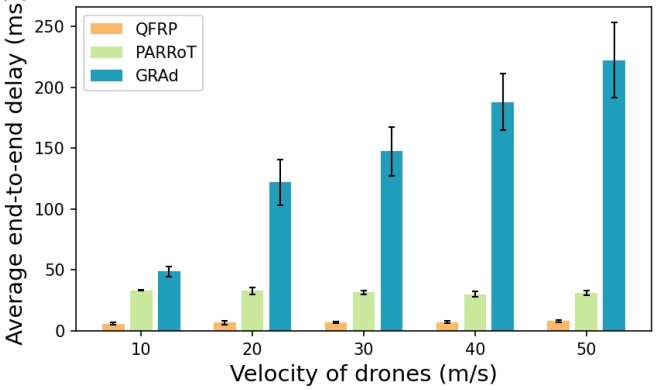

Fig. 7: Average E2E delay for varying UAV node velocities

network density affects PDR. It can be seen that with the increase of the network density, there are fewer isolated UAV nodes and the connectivity of UAV network is improved, PDR performance of QFRP and PARRoT improves. For GRAd, as the node density increases, the number of the communication hops may increase, which makes it difficult to maintain the link and ultimately shows a decrease in PDR.

Fig.10 shows the average E2E delay under different node densities. We take the logarithm of average E2E delay for ease of presentation. As the density increases, the E2E delay of the three protocols becomes gradually longer. For PARRoT, its network load is huge due to chirps flooding, result in more collision, retransmissions and delay. Obviously, QFRP has the best E2E delay performance in the compared routing protocols.

The performance of normalized routing overhead under different network densities is shown in Fig.11. For PARRoT, there is a tendency close to linear growth of its routing overhead because every chirp packet generated from each drone will be forwarded. For GRAd, with the increase of node density, more control packets are needed to maintain path and deal with broken links. In comparison, the routing overhead of QFRP is little affected by network density and maintained at a low level because the routing maintenance most depends on HELLO control packets sent periodically.

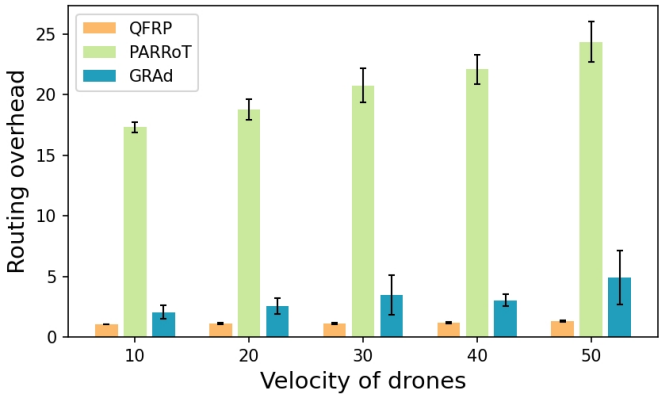

Fig. 8: Normalized routing overhead for varying UAV node velocities

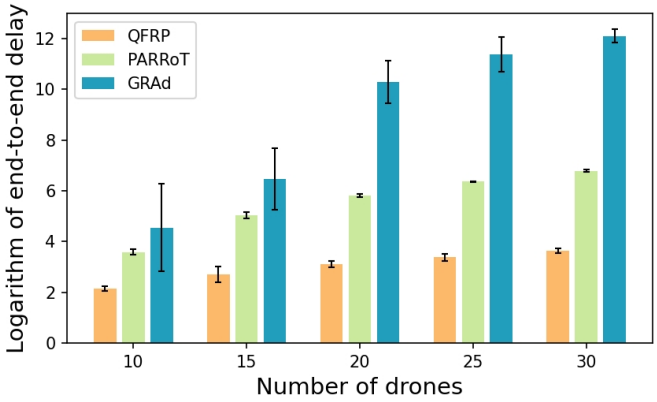

Fig. 10: Average E2E delay for varying UAV node velocities

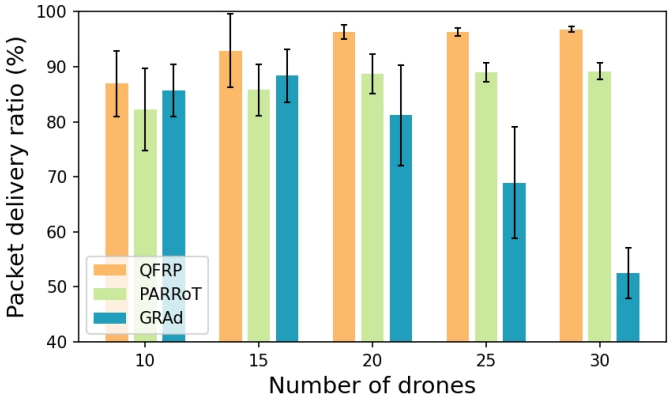

Fig. 9: PDR for varying UAV node velocities

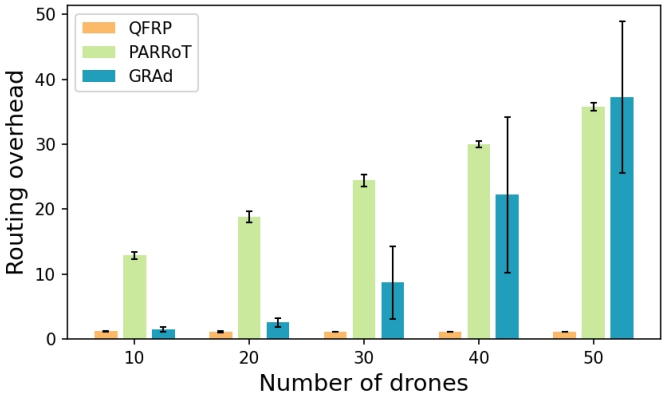

Fig. 11: Normalized routing overhead for varying UAV node velocities

V. Conclusion

In this paper, we propose QFRP based on Q-Learning and fuzzy logic, considering neighbor set coherence, link lifetime, link quality and access delay on routing. QFRP utilizes a fuzzy logic based exploration and exploitation mechanism to adapt to network topology changes. To mitigate the impact of blind exploration and speed up Q-value iteration, we propose a Q-value update mechanism based on HELLO and ACK packets which replaces flood-based method or inefficient ACK-based method. Simulation results illustrate that QFRP effectively improve PDR, reduce E2E delay and routing overhead in UAV swarm networks.

References

[1] J. Zhou, *et al.*, "A Bidirectional Q-learning Routing Protocol for UAV Networks," in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2021, pp. 1-5.

[2] Z. Zhou, *et al.*, "Optimized Routing Protocol Through Exploitation of Trajectory Knowledge for UAV Swarms," in *IEEE Transactions on Vehicular Technology*, 2024.

[3] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-sequenced Distance-vector Routing for Mobile Computers," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 4, pp. 234-244, 1994.

[4] R. Poor, "Gradient Routing in Ad Hoc Networks," Massachusetts Institute of Technology: Cambridge, MA, USA, 2000.

[5] B. Karp, "GPSR : Greedy Perimeter Stateless Routing for Wireless Networks," in *Proceedings of the 6th International Conference on Mobile Computing and Networking*, 2000, pp. 243-254.

[6] J. Lansky, *et al.*, "Reinforcement Learning-based Routing Protocols in Flying Ad Hoc Networks (FANET): A review," *Mathematics*, vol. 10, no. 16, pp. 3017–3017, 2022.

[7] J. Liu, *et al.*, "QMR: Q-learning based Multi-objective Optimization Routing Protocol for Flying Ad hoc Networks," *Computer Communications*, vol. 150, no. 15, pp. 304-316, 2020.

[8] Q. Xue *et al.*, "QEHLR: A Q-Learning Empowered Highly Dynamic and Latency-Aware Routing Algorithm for Flying Ad-Hoc Networks," *Drones*, vol. 7, no. 7, pp. 459, 2023.

[9] M.Y. Arafat *et al.* "A Q-learning-based Topology-aware Routing Protocol for Flying Ad Hoc Networks," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1985-2000, 2021.

[10] C. He *et al.* "A Fuzzy Logic Reinforcement Learning-based Routing Algorithm for Flying Ad Hoc Networks," in *2020 International Conference on Computing, Networking and Communications*, 2020, pp. 987-991.

[11] C. Wu *et al.* "Routing in Vanets: A Fuzzy Constraint Q-learning Approach," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 195-200.

[12] C, Wei *et al.* "QFAGR: A Q-learning-based Fast Adaptive Geographic Routing Protocol for Flying Ad hoc Networks," in *2023 IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 4613-4618.

[13] B.Sliwa, *et al.* "PARRoT: Predictive Ad-hoc Routing Fueled by Reinforcement Learning and Trajectory Knowledge," in *IEEE Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1-7.

[14] G. Oddi *et al.* "A Proactive Linkfailure Resilient Routing Protocol for MANETs based on Reinforcement Learning," in *2012 20th Mediterranean Conference on Control Automation (MED)*, 2012, pp. 1259–1264.

[15] A. Boukerche. "Performance Evaluation of Routing Protocols for Ad Hoc Wireless Networks," *Mobile Networks and Applications*, vol. 9, no. 4, pp. 333-342, 2004.