Q-learning based Stepwise Routing Protocol for Multi-UAV Networks

Jae Won Lim
Department of AI Convergence Network
Ajou University
Suwon, Republic of Korea
gift21cna@ajou.ac.kr

Young-Bae Ko
Department of AI Convergence Network
Ajou University
Suwon, Republic of Korea
youngko@ajou.ac.kr

2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 978-1-7281-7638-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIIC51459.2021.9415265

***Abstract*—A multi-UAV network is a wireless multi-hop network consisting of several Unmanned Aerial Vehicles (UAVs) that are supposed to communicate with a centralized control center. Due to high mobility, such a dynamic network faces frequent changes in network topology, resulting in poor wireless link quality and even frequent disconnection. UAVs' computational capability is also bound to a limited threshold. Therefore, it is important to design a routing protocol that works in a lightweight and adaptive manner. We propose an intelligent routing protocol for Multi-UAV Networks that ensure minimum hops to the destination and better link quality by employing the Q-learning technique. The performance of the proposed scheme was evaluated through the OPNET simulator. A preliminary result shows that it can improve the routing performance in terms of the end-to-end delivery and packet delivery ratio, compared to the de facto ad hoc routing protocol.**

Keywords—Multi-UAV Network, Q-learning based routing, Unmanned aerial vehicle

I. INTRODUCTION

Unmanned aerial vehicles(UAVs) have been extensively studied for the purpose of various real-life applications such as delivery services, disaster response, and industrial environment monitoring. Multiple UAVs can quickly construct a communication network without any pre-determined infrastructure in ad hoc fashion, so called 'Multi-UAV Networks' or 'Flying Ad Hoc Networks(FANET)' [1].

In those networks, there is a severe problem of poor communication quality due to frequent topology changes caused by the rapid mobility of UAVs. Note that the high mobility of nodes makes it difficult to predict the state of the link, so more frequent update for link quality is needed. How to make this update period adaptive to environments is challenging as UAVs have constraints on computational performance and battery capacity as well.

In order to solve this problem, we propose a Q-learning based routing protocol that considers the minimum hop count to the destination, link quality, and mobility. To construct Multi-UAV Networks, a stepwise neighbor node search is performed first to reflect the stability of the link and the minimum number of hops towards a destination. Then the Q-learning algorithm is performed based on information obtained by the stepwise node search. Thus, we designed the link stability factor along with the hop number, which is integrated with the reward function for the Q-learning mechanism. This procedure helps to maintain the QoS parameter in terms of data transmission ratio and transmission speed.

II. RELATED WORKS

A. Routing protocols in UAV networks

POLSR[2] is a speed-aware routing protocol that considers future location. Based on GPS information, the relative speed between two nodes can be obtained and used to evaluate communication link quality. So UAVs can make better routing decisions. Time-slotted AODV[3], which is proposed for UAV networks, is basically a time-slotted version of AODV. Time-slotted AODV uses dedicated time slots in which only one node can send data packets. Although it increases the use of network bandwidth, it mitigates packet collisions and ensures packet delivery. There are a good number of surveys done on the routing protocols for UAV network architecture [4], [5]. These surveys consolidate the published research on UAV networks' routing protocol by comprehending them in a brief manner. Besides, in these surveys, future research direction is given by pointing out open research issues from the existing literature. [4] covers the overall field of routing for UAV networks. It compares distinct kinds of routing protocols of the UAV network in terms of their major features, characteristics, and performance. In addition, [4] describes the design considerations of the UAV networks and the techniques used for the proposed routing protocol of UAV networks. [5] covers UAV classification, communication, and application architectures. It describes the strengths and the weaknesses of each routing protocol depending on different mobility modes. In addition, [5] proposes a taxonomy of the investigated routing protocols for UAV networks based on the routing mechanism which is used to establish a routing path between the UAVs and the control station.

B. Q-learning based routing for UAV networks

Reinforcement learning is an area of machine learning concerning how to get the maximum rewards in a given environment according to the current state. Q-learning, one of the reinforcement learning methods, requires low cost to learn about a given environment because it simply observes the Q-value to make a decision. This low-cost requirement of Q-learning makes it suitable to apply reinforcement learning to the routing protocol**.** Q-Geo[6] demonstrated the use of the Q-learning algorithm for considering the geographic routing scheme and packet travel time. The Q-Geo routing protocol is designed as two phases, which estimate node location from the collected mobility information and make routing decisions by Q-learning. QMR[7] is an another Q-learning based routing protocol for UAV networks. QMR aims to make a network long-lasting by taking into account the energy level of the nodes. QMR uses the energy level of a node and link delay as elements of reward function to balance power consumption and link delay.

III. Q-LEARNING BASED STEPWISE ROUTING PROTOCOL

We propose a Q-learning based routing protocol for multi-UAV networks that considers the minimum number of hops to the destination and link stability through a simple stepwise node discovery.

A. *Stepwise node discovery*

The proposed multi-UAV network structure assumes a network consisting of one control center(CC) and a number of UAVs. Algorithm 1 presents the pseudo code of the stepwise node discovery process. The process is initiated by the CC and performed step by step according to the period calculated by the mobility of the node. That is, the period of stepwise node discovery is adaptively calculated by the CC, who takes into account the mobility and location information of UAV nodes in the network.

Algorithm 1. Stepwise Node Discovery

Phase 1 — **Initial Broadcast of Discovery packets periodically generated by Control Center**

```
if (self == CC):
  if (The time of periodic stepwise node discovery arrives):
    Generate Discovery_packet:
      Discovery_packet.Sender_ID = CC's ID;
      Discovery_packet.Hop_Count = 0
    Broadcast(Discovery_packet);
```

Phase 2 — **Stepwise Transfer of Discovery & ACK packets for the purpose of gathering neighbor UAV nodes information**

```
if (self != CC):
  if(Discovery_packet first received):
    Parent_node = Discovery_packet. Sender_ID
    Generate ACK:
      ACK.Target_ID =Discovery_packet. Sender_ID
      ACK.Sender_ID = self's ID
      ACK.Node_INFO = self's moving speed, location, and hop_count_from_CC information
    Unicast(ACK, Parent_node)
    Modify Discovery_packet:
      Discovery_packet.Sender_ID = self's ID
      Discovery_packet.Hop_Count += 1
    Broadcast(Discovery_packet)
    While (ACK_waiting_time != 0):
      if (receive(ACK)):
        Generate Neighbor_table:
          Neighbor_table.append(ACK.Sender_ID))
          Neighbor_table.append(ACK.Node_INFO))
    Unicast(Neighbor_table, Parent_node)
```

Phase 3 — **Discovery of the Multi-UAV nodes information in the whole network**

```
if (Neighbor_table packet received):
  if (self != CC):
    Generate New_Neighbor_Table by removing any neighbor nodes (duplicated with its own neighbors) from Neighbor_table
    Unicast (New_Neighbor_table, Parent_node)
  if (self == CC):
    Generate Whole_Nodes_Table by removing any neighbor nodes (duplicated with its own neighbors) from New_Neighbor_table
```

B. *Q-learning model*

Q-learning is a kind of reinforcement learning and does not require high computation in the learning process. The characteristics of Q-learning[8] are suitable for UAVs with constraints on computational capability. In Q-learning, the Q-value is updated as follows.

$$NewQ(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha[r_{(s_t, a_t)} + \gamma \max_a Q(s_{t+1}, a)] \quad (1)$$

where s_t means the current state and a_t means the action selected at time t, α is the learning rate and γ is denoted as the discount factor. The repetition of this equation represents the sum of rewards that will be gained in the future. In our work, a node to select a routing path is defined as 'state', and selecting a routing path is defined as 'action'.

The minimum hop from CC and link stability are used as elements of the routing metric. Link stability takes into account the speed at which nodes i and j move away, and the link quality. In order to measure the latter, we considered the packet transmission time and packet delivery ratio. To calculate this, the Window Mean with Exponentially Weighted Moving Average (WMEWMA) method[9] was used. The equation used to calculate the link stability is as follows.

$$link\ stability_{i,j}(t) = (1-\beta)e^{\frac{1}{v_{i,j}}} + \beta\frac{\sum_{t-n}^{t-1} link\ quality_{i,j}(k)}{n} \quad (2)$$

In this equation, $\beta(0 < \beta < 1)$ is a coefficient weight value and n is the number of neighboring nodes. In addition, $v_{i,j}$ represents the speed at which nodes i and j are moving away. $link\ quality_{i,j}(k)$ represent the link quality between the said nodes. Therefore, the link stability has a higher value when the speed at which nodes i and j move away from each other is slow and the value of link quality is high.

C. *Reward Function*

Reward function is activated when an action is taken in a specific state. When an agent takes an action, the agent gets a reward for the selected action through the Reward function. The Reward Function considers the minimum hop to the destination and link stability to ensure the quality of the link and to reduce the number of transmissions within the network. The equation of the reward function is as follows;

$$r = \begin{cases} r_{max}, & when\ s_{t+1}\ is\ destination \\ r_{min}, & when\ s_t\ is\ local\ minimum \\ \omega e^{\frac{1}{hop}} + (1-\omega)link\ stability, & otherwise \end{cases} \quad (3)$$

where r represents the obtained reward, r_{max} is the maximum reward, and r_{min} denotes the minimum reward. ω is the weighting factor that balances the bias between the link stability and the number of hops. If there is a destination among the following states, the r_{max} is received to complete the data transmission. In addition, among the states that can be selected, those with a greater distance than that between the current node and the destination are rewarded with a r_{min} so that states close to the possible destination node are selected. In other cases, reward is given in consideration of link stability and minimum hops to the destination, so that the stable state has a high Q-value while having as few hops as possible. The value of ω varies between 0 to 1.

IV. PERFORMANCE EVALUATION

A. Simulation Environment

We used OPNET 18.0 to evaluate the performance of our protocol and compared it with AODV. We deployed 16 nodes, which include 1 control center in a 3D network environment of 1000m*1000m*100m. All nodes use the 802.11n MAC layer. We set the mobility of UAVs to Random Way Point(RWP) at a speed of 0 m/s ~ 30 m/s. The direction and speed of UAVs were changed as the simulation goes by. Table I shows simulation parameters.

TABLE I. SIMULATION PARAMETERS

Parameters	Values
Simulator	OPNET 18.0
Network size	1000m*1000m*100m
Number of Nodes	16
Mobility model	Random Way Point
MAC layer	802.11n
Traffic type	Barometer (100bytes, CBR)

B. Simulation Results

Simulation results are shown in terms of the end-to-end delivery and PDR of the protocol.

By definition, the end-to-end delay means the time difference between the packet received and sent from the source to the destination. The PDR means the ratio between the number of data packets sent by a source node and the number of data packets received by a destination node. The packet used in the simulation modeled barometer data of 100bytes.

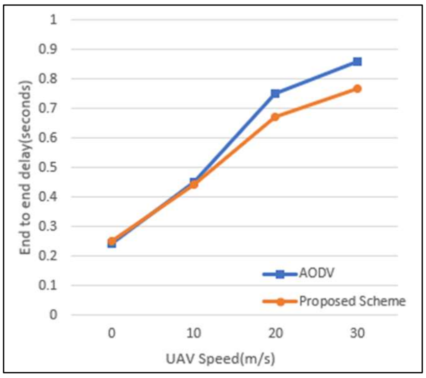

Fig. 1. Average end-to-end latency in terms of different UAV speed

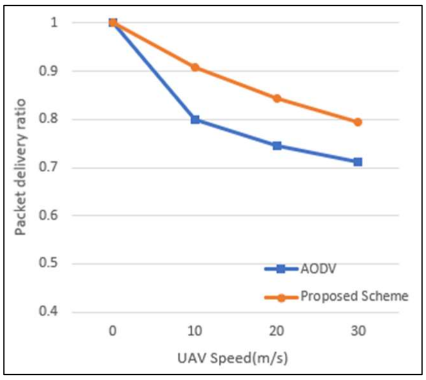

Fig. 2. Packet delivery ratio in terms of different UAV speed

Figure 1 shows the average end-to-end delay of AODV and the proposed scheme, compared in terms of the speed of the node. When the nodes do not move, the topology is fixed, so the delay is similar to that of AODV. On the other hand, the proposed routing protocol has a low delay when the node's moving speed is high. Since the proposed scheme adaptively determines the period of node search according to the node's moving speed, it shows shorter delay than the standard scheme. Figure 2 shows the performance outcome of the PDR comparison. The proposed scheme shows higher PDR than AODV because it considers not only the number of hops but also link stability.

V. CONCLUSIONS

In the multi-UAV networks, since the UAV needs to relay a lot of data, the network environment is heavily influenced by the control overhead and interference. We propose a Q-learning-based routing protocol using stepwise node search, that considers link quality and operates with the minimum hop to the destination. As a result, our work has the advantage of low delay while reducing the control overhead. For future work, we plan to study how to improve the performance through applying Q-learning in the grouping of Geo-Cast.

ACKNOWLEDGEMENT

This work was supported by the Civil-Military Technology Cooperation Program in Republic of Korea (19-CM-TN-07)

REFERENCES

[1] Amira Chriki, Haifa Touati, et al., “FANET: Communication, mobility models and security issues”, ELSEVIER Computer Networks, vol. 163, August.2019.

[2] Stefano Rosati, Karol Kru_zelecki, Gregoire Heitz, Dario Floreano, and Bixio Rimoldi, “Dynamic routing for flying ad hoc networks”, IEEE Transactions on Vehicular Technology, 65(3), pp.1690-1700, 2015

[3] H. Chen , F. Yu , H.CB Chan , et al. , “A novel multiple access scheme in wireless multimedia networks with multi-packet reception”, in: Proceedings of the 1st ACM Workshop on Wireless Multimedia Networking and Performance Mod- eling, ACM, 2005, pp. 24–31 .

[4] Muhammad Yeasir Arafat, Sangman Moh, “Routing Protocols for Unmanned Aerial Vehicle Networks: A Survey”. IEEE Access, vol. 7, pp.99694-99720, 2019.

[5] Demeke Shumeye Lakew, Umar Sa’ad, et al. , “Routing in Flying Ad Hoc Networks: A Comprehensive Survey”, IEEE Communications Surveys & Tutorials, vol. 22, 1071-1120, 2020

[6] Woo-Sung Jung, Jinhyuk Yim, and Young-Bae Ko, “QGeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks”. IEEE Communications Letters, 21(10): pp.2258-2261, 2017.

[7] Jianmin Liu, Qi Wang, et al., “QMR:Q-learning based Multi-objective optimization Routing protocol for Flying Ad Hoc Networks”, ELSEVIER Computer Communications, vol. 150, pp.304-316, 2020.

[8] Christopher JCH Watkins and Peter Dayan. “Q-learning”. Machine learning, 8(3-4) pp. 279-292, 1992.

[9] A. Woo and D. Culler, “Evaluation of Efficient Link Reliability Estimators for Low-Power Wireless Networks”, Technique Report UCB//CSD-03-1270, U.C. Berkeley Computer Science Division, September 2003.