논문 21-46-11-13 The Journal of Korean Institute of Communications and Information Sciences '21-11 Vol.46 No.11
https://doi.org/10.7840/kics.2021.46.11.1873

무인 항공 멀티홉 네트워크를 위한 Double Q-Learning 기반 라우팅 프로토콜

임 재 원◆, 고 영 배°

A Double Q-Learning Based Routing Protocol for Unmanned Aerial Multi-Hop Networks

Jae-won Lim◆, Young-Bae Ko°

요 약

무인 항공 멀티홉 네트워크 또는 공중 애드혹 네트워크는 감시정찰, 센싱 데이터 수집 등의 다양한 임무를 수행하기 위하여 다수의 무인 항공기들(UAVs: Unmanned Aerial Vehicles)로 구성된 특별한 형태의 이동 애드혹 망(MANETs: Mobile Ad-Hoc Networks)이다. 일반적으로, 무인 항공기는 동적 이동성과 제한된 자원으로 인한 열악한 연결성 및 낮은 네트워크 성능 문제가 있다. 그러므로 경량화되고 적응적으로 운용할 수 있는 라우팅 프로토콜설계가 매우 중요하다. 본 논문에서는 목적지까지의 최소 홉과 링크 품질을 고려하는, Double Q-learning 기반 라우팅 프로토콜을 제안한다. 제안 기법에서는 제어 메시지 부하를 낮추기 위하여 네트워크 상황에 따라 노드 탐색 메시지의 주기를 조정한다. OPNET 시뮬레이터를 통해 제안 기법의 성능 검증을 수행하였으며 기존 기법인 QMR(Q-learning based Multi-objective optimization Routing)이나 단순 큐 러닝 기반 라우팅 프로토콜과 비교하여 패킷 전송률이 향상되고 데이터 전송의 지연 시간이 줄어든다는 것을 확인하였다.

Key Words : Flying Ad-Hoc Networks, Double Q-learning based routing, Unmanned aerial vehicle

ABSTRACT

The Unmanned Aerial Multi-hop Network or Flying Ad-Hoc Network is a special type of mobile ad hoc networks, consisting of multiple UAVs(Unmanned Aerial Vehicles) to perform a variety of missions such as ISR(Intelligent Surveillance and Reconnaissance), sensing data collection, etc. In general, UAVs have problems with poor connectivity and low network performance due to their dynamic mobility and limited resources. Therefore, it is very important to design a routing protocol that operates in light-weight and adaptive manners. In this paper, we propose a double Q-learning based routing protocol that takes into account the minimum number of hops and link quality towards a destination. The proposed scheme adjusts an interval of node discovery messages according to network conditions to reduce the control message overhead. Via the OPNET simulator, we have performed a validation study of the proposed scheme and found out the fact that its packet delivery ratio becomes higher but the delay of data transmission is lower, compared to the existing QMR(Q-learning based Multi-objective optimization Routing) scheme as well as the simplest Q-learning based routing protocol.

◆ First Author : Department of AI Convergence Network Graduate School of Ajou Univ., gift21cna@ajou.ac.kr, 정회원
° Corresponding Author : Department of Software and Computer Engineering, Ajou Univ., youngko@ajou.ac.kr, 종신회원
논문번호 : 202106-135-B-RN, Received June 21, 2021; Revised August 1, 2021; Accepted August 3, 2021

Ⅰ. 서 론

'Multi-UAV Network' 또는 'Flying Ad-hoc Network(FANET)'는 감시정찰, 센싱 데이터 수집 등의 다양한 임무를 수행하기 위하여 다수의 무인 항공기로 구성된 이동 애드혹 망이다. 이러한 네트워크는 이동성이 높은 다수의 무인 항공기를 통하여 기반 시설 없이 통신 네트워크를 신속하게 구성할 수 있다. 하지만 UAV의 빠른 이동성으로 인한 잦은 토폴로지 변화로 인해 통신 품질이 떨어지는 문제가 존재한다. 예를 들어 UAV 통신 중계 어플리케이션은 원격 장치가 통신할 수 있도록 다수의 UAV가 기반 시설 없이 통신 영역을 확장하는 어플리케이션이다. 통신 중계 어플리케이션에서 UAV는 멀티 홉을 통해 원격 장치에서 데이터를 중계해야 하지만 UAV 노드의 이동성이 높기 때문에 연결이 끊기는 경우가 많아 패킷 손실 문제가 발생한다. 이러한 문제는 네트워크 성능을 저하시킬 뿐만 아니라 UAV의 에너지 소모를 증가시킨다[1]. 이러한 문제를 해결하기 위해 다양한 라우팅 프로토콜 연구가 진행되고 있다. Multi-UAV 네트워크의 라우팅 프로토콜 연구는 Multi-UAV 네트워크의 특성을 파악하여 동적인 네트워크 환경에서 최적의 라우팅 경로를 선택하기 위해 라우팅 메트릭을 설정하는 데 중점을 두었다. 네트워크 상태를 나타낼 수 있는 적절한 라우팅 메트릭을 사용하면 최적의 경로 선택을 통해 네트워크 성능을 향상시킬 수 있다. 또한 라우팅 프로토콜은 reactive 프로토콜과 proactive 프로토콜로 분류된다. Proactive 라우팅 프로토콜은 사전에 라우팅 정보를 저장하고 토폴로지가 변경될 때마다 라우팅 테이블을 업데이트한다. 이러한 종류의 라우팅 프로토콜은 패킷 전송 지연을 줄이지만 Multi-UAV 네트워크에서는 노드의 높은 이동성으로 인한 잦은 토폴로지 변경으로 큰 오버헤드가 발생한다. 특히 멀티 홉 네트워크에서는 사전에 라우팅 경로를 생성하기 위한 네트워크 오버헤드로 인해 통신이 실패할 수 있다. 반면에 reactive 라우팅 프로토콜은 데이터 전송이 필요할 때 라우팅 경로를 설정한다. 이러한 종류의 네트워크에서는 경로 설정에 대한 오버헤드는 감소하지만 경로 탐색 시간으로 인해 지연 시간이 늘어나는 단점이 존재한다[2].

이외에 강화 학습(Reinforcement learning) 기법을 활용하는 라우팅 프로토콜에 대한 연구도 진행되고 있다[3]. 강화 학습은 모델 없이 주어진 환경을 학습하면서 자기 최적화를 통해 의사결정을 내리는 기계 학습 기법이므로 오버헤드가 적다는 장점이 있다. 따라서 배터리와 노드의 계산 능력에 제약을 갖는 Multi-UAV 네트워크 환경에 적용될 수 있다. 강화 학습의 한 가지 종류로 큐 러닝(Q-learning)[4]이 존재하지만 큐 러닝은 보상 값이 과대평가 되면서 잘못된 방향으로 학습이 진행될 수 있다는 단점을 가진다[5]. 따라서 본 논문에서는 목적지에 대한 최소 홉 수, 링크 품질 및 이동성을 고려하는 강화 학습 기반의 라우팅 프로토콜인 더블 큐 러닝(Double Q-learning) 기반 라우팅 프로토콜을 제안한다. Multi-UAV 네트워크를 구성하기 위해 링크의 안정성과 목적지에 대한 최소 홉 수를 반영하기 위해 단계적 인접 노드 탐색이 먼저 수행되며 단계적 노드 탐색으로 얻은 정보를 기반으로 더블 큐 러닝을 수행한다. 더블 큐 러닝을 적용함으로써 큐 러닝의 과대평가 문제를 감소시켰고 이에 대한 보상 함수와 홉 수와 함께 링크 안정성 계산 방법을 설계하였다. 이러한 라우팅 방법은 데이터 전송률 및 전송 속도 측면에서 QoS를 유지하는 데 유리하다. 또한 강화 학습 기반의 라우팅 프로토콜은 노드 간 패킷 교환을 통하여 보상 값에 대한 계산을 진행하는데 이러한 패킷의 주기 또한 Multi-UAV 네트워크의 QoS에 큰 영향을 미친다. Multi-UAV 네트워크에서는 UAV의 높은 이동성으로 패킷 교환이 잦은 문제가 발생한다. 따라서 본 논문에서는 노드의 속도와 링크의 품질을 고려한 패킷 교환 주기를 설계하여 네트워크 상태에 따라 패킷 교환 주기를 조절할 수 있도록 하였다. 본 논문은 2장에서 관련 연구로써 강화 학습 기반 라우팅 프로토콜과 더블 큐 러닝에 대하여 서술한다. 3장에서 목적지까지의 최소 홉과 링크 안정성을 고려하는 더블 큐 러닝 기반의 라우팅 프로토콜을 제안하고 4장에서 제안 프로토콜의 성능을 제시하며 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 관련 연구

강화 학습은 주어진 환경에서 현재 상태에 따라 최대한의 보상을 받는 방법에 관한 기계 학습 분야이다. 강화 학습 방법 중 하나인 큐 러닝은 결정을 내리기 위해 단순히 Q-value를 관찰하기 때문에 주어진 환경에 대해 학습하는 데 많은 계산량을 필요로 하지 않는다. 따라서 큐 러닝은 적은 계산량을 요구하기 때문에 제한된 성능을 갖는 환경인 UAV 네트워크에 적용하는 데 적합하다[3].

2.1 강화 학습 기반의 라우팅 프로토콜

큐 러닝 기반의 지리적 애드 혹 라우팅 프로토콜

(Q-Geo : Q-learning based geographic ad hoc routing protocol)[6]은 지리적 라우팅 체계와 패킷 전송 시간을 고려하기 위해 큐 러닝 알고리즘을 사용하였다. Q-Geo 라우팅 프로토콜은 수집한 이동성 정보로부터 노드의 위치를 추정하고 큐 러닝을 통해 라우팅 결정을 내리는 두 단계로 설계되었다. 큐 러닝의 보상을 계산하기 위한 정보로써 패킷 지연과 상대 노드까지의 거리를 활용하며 해당 정보는 노드의 속도에 따른 주기를 가진 hello 메세지로 전달한다.

[7]은 강화 학습 기반의 라우팅 프로토콜(RLSRP : Self-learning routing protocol)과 위치 예측 기반 MAC(PPMAC : Position prediction based directional MAC)을 사용하는 강화학습 기반의 하이브리드 통신 프로토콜을 제안하였다. [7]은 PPMAC을 사용하여 다른 노드의 위치를 예측하고 통신과 데이터 전송을 제어하였다. 또한 PPMAC은 지향성 안테나를 사용하여 방향 정보를 활용한 MAC을 제안하였고 RLSRP는 강화 학습을 사용하여 종단 간 지연이 가장 적은 라우팅 경로를 제공한다. RLSRP는 네트워크 토폴로지의 변화를 추적하기 위해 이웃 테이블을 유지하고 경로 선택에 있어 강화학습을 사용하지만 RLSRP에서 사용하는 강화 학습 알고리즘은 할인율, 학습율과 같은 강화 학습의 변수를 고정하여 사용한다.

[8]은 다목적 라우팅 프로토콜을 위한 큐 러닝 기반 fuzzy logic을 제안하였다. 소스 노드는 전송 속도, 배터리 잔량, 배터리 소모율, hop 수 및 패킷 전달 시간을 고려하면서 제안하는 알고리즘을 사용하여 라우팅 경로를 결정한다. Fuzzy 시스템은 신뢰할 수 있는 링크를 식별하기 위해 사용되며, 큐 러닝은 경로에 보상을 제공함으로써 fuzzy 시스템을 지원한다. 또한 전체 경로에 대한 큐 러닝 알고리즘을 제안하여 전체 경로에 대한 성능도 고려한다. 단일 링크와 전체 경로에 대한 각각의 Q-value를 얻은 후 fuzzy logic이 Q-value를 평가하여 라우팅을 위한 최적의 경로를 결정한다.

[9]는 무선 네트워크에서 비디오 스트리밍을 라우팅하기 위한 라우팅 프로토콜이다. 강화 학습 기반 기회주의적 라우팅(RLOR: Reinforcement learning-based opportunistic routing)은 비디오 스트리밍의 주요 요구사항인 전송 지연을 고려하여 전송 지연 추정치에 기초한 보상 함수를 제안하였다. 또한 RLOR은 노드의 링크 품질과 혼잡 수준의 변화를 고려하여 각 경로에 대해 예상되는 지연을 예측하는 강화 학습 기반의 알고리즘을 제안하고 이를 기반으로 하는 라우팅 알고리즘을 제안하였다.

[10]은 UAV를 포함하는 5G 네트워크 환경에서 배터리 잔량과 링크의 안정성을 고려하는 강화 학습 기반 라우팅 프로토콜을 제안하였다. [10]은 Multi-UAV 네트워크의 높은 이동성에 의한 잦은 토폴로지 변화와 제한된 배터리 용량을 고려하여 끊어지는 링크의 수를 줄이고 네트워크를 장기간 유지할 수 있도록 강화 학습을 통해 최적의 경로를 선택한다. 강화 학습의 보상 요소로써 UAV의 배터리 잔량과 이동성의 수준을 고려한다.

[11]은 UAV 네트워크를 위한 큐 러닝 기반의 라우팅 프로토콜을 제안하였다. [11]은 종단 간 지연과 배터리를 동시에 최적화하는 큐 러닝 기반의 다목적 라우팅 프로토콜(QMR : Q-learning based Multi-objective optimization routing protocol)을 제안하였다. 또한 학습률, 할인율 등의 큐 러닝에서 사용되는 변수들을 동적으로 변경하였다. QMR은 인접 노드 탐색, 큐 러닝 알고리즘, 라우팅 결정 및 페널티 프로세스로 구성되며 GPS 데이터를 사용하여 이웃 노드의 위치를 수집하고 hello 메세지를 전송하여 경로 탐색 프로세스를 시작한다. 또한 노드의 배터리 잔량과 링크 지연을 Q-value 계산에 포함하여 충분한 배터리 잔량을 가지면서 적은 링크 지연을 갖는 이웃 노드를 경로로 선택하여 데이터를 전송한다. 하지만 네트워크 정보 수집을 수행하는 hello 메세지의 주기가 적응적으로 조절되지 않으며 적은 지연을 갖는 이웃 노드를 경로로 선택하기 때문에 데이터 전송에 사용되는 홉 수가 크다는 단점이 존재한다.

2.2 Double Q-learning

큐 러닝은 현재 상태(State)에 대한 보상(Reward)을 선택하는 방식으로 환경에 대한 학습을 수행하는 러닝 방식이다. 다음은 큐 러닝의 작동 방식에 대한 수식이다.

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

s_t는 현재의 상태를 의미하고 a_t는 시간 t에서 선택된 행동(Action)을 의미한다. 또한 α는 현재의 Q-value를 얼마나 반영할지에 대한 학습률, γ는 할인율을 의미한다. 따라서 이 식의 반복은 Q-value가 미래에 얻게 될 보상을 표현한다. 하지만 큐 러닝은 현재의 상태에서 취할 수 있는 행동 중 어떤 행동의 보상이 가장 클 것인지 선택하는 과정에서 single estimator에 대해 Maximum Evaluation을 진행한다.

따라서 이러한 estimation에서 실제 기대 값 보다 커지게 되는 과대평가가 발생하게 된다. 과대평가 문제는 최적이 아닌 행동을 선택하는 결과를 초래하여 수렴 속도를 늦출 수 있다. 더블 큐 러닝은 큐 러닝의 과대평가 문제를 감소시키기 위해 고안된 강화 학습 기법으로 두 개의 estimator를 적용하여 큐 러닝의 과대평가 문제를 감소시켰으며 큐 러닝 대비 안정적인 학습과 빠른 수렴을 보장한다. 더블 큐 러닝의 알고리즘은 다음과 같다.

더블 큐 러닝은 Q^A, Q^B의 두 개의 Q-value를 유지하며 Q^A를 업데이트하기 위해 $Q^A(s', a^*) = \max_a Q^A(s', a)$를 사용하는 것이 아닌 $Q^B(s', a^*)$를 사용한다. Q^B는 같은 문제를 해결하지만 다른 상태에서 업데이트되기 때문에 표본이 다르고 따라서 이 행동의 보상은 편향되지 않는다고 할 수 있으며 과대평가를 줄이는 결과를 갖는다. 또한 더블 큐 러닝은 조건에 따라 하나의 Q-value만 업데이트 하므로 기본적으로 큐 러닝과의 계산 복잡도는 같게 된다. 따라서 더블 큐 러닝은 큐 러닝과 계산 복잡도는 같지만 과대평가를 감소시키는 특성을 갖는다.

강화 학습 기반의 라우팅 알고리즘은 앞서 조사한 것과 같이 패킷 교환을 통하여 네트워크의 환경에 대한 데이터를 수집하는데 Multi-UAV 네트워크는 노드의 이동성이 높아 토폴로지의 변화가 잦다. 따라서 패킷 교환 주기를 적응적으로 조정하지 않으면 제어 오버헤드의 증가를 야기할 수 있다. 또한 기존 큐 러닝 기반의 라우팅 프로토콜은 Maximum Estimation에 의한 과대평가 문제가 발생한다. 이와 같은 문제를 해결하기 위해 본 연구에서는 노드의 이동 속도와 방향을 고려하여 패킷 교환 주기를 적응적으로 조정하고 경로 결정에 있어 목적지까지의 최소 홉 수와 링크의 안정성을 고려하는 더블 큐 러닝 기반의 라우팅 알고리즘을 제안한다.

Algorithm 1. Double Q-learning

1: ***Initialize*** Q^A, Q^B, s
2: ***repeat***
3: Choose a, based on Q^A, Q^B, observe r, s'
4: Choose (e.g. random) UPDATE(A) or UPDATE(B)
5: ***if***(UPDATE(A))
6: ***Define*** $a^* = \arg max_a Q^A(s', a)$
7: $newQ^A(s,a) = Q^A(s,a) + \alpha(s,a)*(r + \gamma Q^B(s', a^*) - Q^A(s,a))$
8: ***else if***(UPDATE(B))
9: ***Define*** $b^* = \arg max_a Q^B(s', a)$
10: $newQ^B(s,a) = Q^B(s,a) + \alpha(s,a)(r + \gamma Q^A(s', b^*) - Q^B(s,a))$
11: s <- s'

Ⅲ. Multi-UAV 네트워크를 위한 더블 큐 러닝 기반의 라우팅 프로토콜

큐 러닝 기반의 라우팅 프로토콜은 수집된 정보로 네트워크 상태를 학습한다. 또한 Q-value를 사용하여 노드의 주변 환경에 따라 적절한 경로를 결정하며 Q-value는 데이터 전송 후 보상을 얻음으로써 업데이트된다. 이에 따라 각 노드는 주변 환경을 학습하여 적절한 경로를 선택하여 데이터를 전송할 수 있다. 하지만 큐 러닝은 Maximum Estimation에 의하여 잘못된 환경을 과대평가하여 잘못된 방향으로 학습되는 경우가 존재한다. 이와 같은 경우 Q-value를 라우팅 메트릭으로 사용하는 환경에서 QoS에 악영향을 미칠 수 있다. 따라서 본 연구에서는 큐 러닝의 과대평가를 방지하기 위해 더블 큐 러닝을 사용한다. 그리고 UAV는 배터리의 수용력에 대한 제한이 존재하며 또한 각 UAV이 맡은 임무에 따라서 단위 시간 당 배터리 소모량이 크게 차이가 날 수 있다. 예를 들어 배송 어플리케이션에서 배송 임무를 맡은 UAV는 운반품의 무게에 의해 중계임무를 맡은 UAV보다 배터리 소모가 빠르다. QMR과 같은 기존 연구에서는 배터리 잔량으로 에너지를 관리하였는데 배터리 소모율을 반영하지 않았기에 이동에 쓰이는 에너지가 통신에 쓰이는 에너지 보다 더 큰 UAV 환경에서는 부족하다. 따라서 본 연구에서는 통신에서 배터리 소모율을 라우팅 메트릭의 한 가지 요소로 고려하여 이와 같은 환경에 대응하도록 하였다.

- 더블 큐 러닝 모델 적용 : 큐 러닝은 확률적인 환경에서 Maximum Estimation에 의하여 행동을 과대평가하기 때문에 종종 성능이 좋지 않은 결과를 보여준다. 따라서 본 연구에서는 더블 큐 러닝을 적용하여 과대평가 문제를 감소시켰다.
- 적응적으로 조절하는 노드 탐색 주기 : 네트워크 내 UAV를 그룹으로 분할하여 기존 연구에서의 제어 오버헤드를 줄이고 UAV 노드의 속도와 방향으로 그룹을 이탈하는 경우를 예측하여 탐색을 진행하여 네트워크의 제어 오버헤드를 감소시켰다.

- 적응적인 탐험(Exploration)과 활용(Exploitation) : 일반적으로 강화학습에 적용되는 정적인 의사 결정이 아닌 네트워크의 환경에 대해 적응적으로 탐험과 활용을 결정하는 방안을 제시하였다.

본 연구에서는 최적의 경로를 선택하기 위해 세 가지 메트릭을 사용한다. 먼저 제안한 프로토콜과 같이 사용한 목적지로부터의 최소 홉과 링크 안정성을 사용하여 안정적이면서도 가급적 최단의 경로를 선택하도록 고려하였다. 하지만 UAV 네트워크에서는 UAV 노드가 제한된 배터리 용량을 가지기 때문에 UAV 노드의 배터리 또한 고려해야 한다. 따라서 본 연구에서는 UAV 노드의 배터리 잔량과 배터리 소모율을 추가로 고려하여 라우팅 메트릭의 요소로 사용하였다.

노드의 에너지 소비량 균형을 맞추기 위해 다음 홉을 선택할 때 노드의 배터리 잔량과 소모율을 고려한다. 본 연구에서는 각 노드가 주어진 임무에 따라 에너지 소모에 있어 격차가 있다는 점을 고려하기 위해 노드의 배터리 잔량뿐만 아니라 배터리 소모율도 고려하여 네트워크의 지속 시간을 늘리고자 하였다. 본 연구에서는 에너지 메트릭을 다음과 같이 표현하였다.

$$E_j = \omega e^{B_{res}} + (1-\omega) e^{\frac{1}{B_{con}}} \tag{2}$$

ω는 가중치 계수이며 0 과 1 사이의 값을 갖는다. 또한 B_{res}는 배터리의 잔량, B_{con}은 배터리의 소모율을 나타낸다. 에너지 메트릭은 위 식과 같이 배터리의 잔량이 크고 단위 시간 당 배터리의 소모율이 적을수록 큰 값을 갖는다. 그리고 본 논문에서는 라우팅 메트릭의 요소로 CC(Control Center)로부터의 최소 홉과 링크 안정성을 사용한다. 링크 안정성은 노드 i, j가 멀어지는 속도와 링크의 품질을 고려한다. 링크의 품질을 측정하기 위해서 본 논문에서는 패킷 전송 시간과 패킷 전송률을 고려하였으며 이를 계산하기 위해서 Window Mean with Exponentially Weighted Moving Average (WMEWMA) 기법을 사용하였다 [12]. 링크 안정성을 계산하기위해 사용한 식은 다음과 같다.

$$link\_stability_{i,j}(t) = (1-\beta) e^{\frac{1}{v_{ij}}} + \beta \frac{\sum_{t-n}^{t-1} link\_quality_{i,j}}{n} \tag{3}$$

β(0<β<1)는 가중치 계수이며 n은 이웃 노드의 수를 나타낸다. $link_quality_{i,j}$는 노드 i에서 측정한 노드 j와의 링크의 품질을 나타낸다. 링크의 품질은 패킷이 전송되는 데 걸리는 시간과 패킷의 전송률로 계산된다. 또한 $v_{i,j}$는 노드 i와 j가 멀어지는 속도를 나타낸다. 따라서 링크 안정성은 노드 i와 j가 멀어지는 속도가 느릴수록, 링크 품질 값이 클수록 높은 값을 가지게 된다.

더블 큐 러닝은 큐 러닝의 Maximum Estimation에 의한 과대평가를 방지하기 위한 학습 방법이다. 큐 러닝은 확률적인 환경에서 종종 성능이 좋지 않은 결과를 보여주는데 이것은 Maximum Estimation에 의하여 행동을 과대평가하기 때문에 발생한다. 따라서 본 연구에서는 이러한 결과를 방지하기 위해 더블 큐 러닝을 적용하여 과대평가를 방지하고자 하였다.

본 연구에서는 더블 큐 러닝의 보상을 계산하기위해 목적지로부터의 홉 수와 링크의 안정성을 고려한다. 보상 함수(Reward Function)는 특정 상태에서 행동을 취할 때 활성화되며 에이전트(Agent)가 특정 행동을 수행할 때 보상 함수를 통해 보상을 받는다. 보상 함수는 목적지로부터의 최소 홉과 링크 안정성을 고려하여 링크 품질을 보장하고 네트워크 내의 데이터 전송 수를 줄이도록 설계하였고 보상 함수의 수식은 다음과 같다.

$$r = \begin{cases} r_{\max}, when\, s_{t+1} is\, destination \\ r_{\min}, when\, s_t is\, local\, minimum \\ \omega e^{\frac{1}{hop}} + (1-\omega) link\_stability, otherwise \end{cases} \tag{4}$$

다음 상태 중에 목적지가 있을 경우 데이터의 전송을 완료하기 위하여 최대치의 보상을 받는다. 또한 현재 노드와 목적지 간의 거리보다 더 큰 거리를 갖는 다음 상태는 최소치의 보상을 주어 가능한 목적지 노드에 가까운 상태를 선택하도록 하였다. 다른 경우에는 목적지까지의 최소 홉과 링크 안정성을 고려한 보상을 주어 가능한 적은 홉을 가지면서 안정성 있는 상태가 높은 Q-value를 가지도록 하였다.

강화 학습에서는 환경 변화를 업데이트하기 위해 탐험과 활용의 의사결정의 비율이 중요하다. 과도한 탐험의 비율은 네트워크의 높은 딜레이를 가져오며 반대로 과도한 활용 비율은 강화 학습 과정에 있어 환경에 대한 반영을 적게 하여 패킷 손실과 같은 결과를 초래할 수 있다. 강화 학습에서는 일반적으로 ϵ-greedy, 볼츠만 메커니즘 및 UCB (Upper-

Confidence-Bound)를 탐험과 활용의 의사결정으로 사용하지만 ϵ-greedy는 네트워크 환경에 맞게 적응적으로 비율 조정을 할 수 없으며 UCB와 볼츠만 메커니즘은 탐험과 활용의 비율을 바꿀 수 있지만 시간에 따라 비율을 결정하기 때문에 네트워크 환경의 반영이 어렵다는 단점이 존재한다. 따라서 Multi-UAV 네트워크와 같은 동적 환경의 경우 탐험과 활용 간의 균형은 단순히 시간에 의해서가 아니라 네트워크 조건에 의해서 조절되어야 한다. 본 연구에서는 노드의 이웃 테이블에 새로운 노드가 추가되었을 경우와 현재 링크의 데이터 전송 속도가 일정 이상 저하되었을 때 탐험을 수행하도록 하여 네트워크의 환경에 변화가 발생할 경우 반영할 수 있도록 하였다.

본 연구에서는 3 단계를 수행하여 데이터를 전송한다. 첫 번째 단계는 노드 탐색 과정으로 CC에 의해 시작되어 네트워크 내의 노드에 대한 정보를 수집하는 프로세스이며 두 번째 단계는 더블 큐 러닝 기반 학습 단계로 더블 큐 러닝을 통해 Q-value를 업데이트하는 프로세스이다. 마지막 단계는 데이터 전송 단계로 경로는 Q-value와 에너지 메트릭을 사용하여 결정된다. 이를 통해 제안 기법은 Q-value와 에너지 메트릭을 사용하여 다음 경로를 결정하게 된다. 따라서 Q-value를 계산하는 데 사용된 목적지로부터의 최소 홉, 링크 안정성과 다음 노드의 배터리 환경을 고려하여 최선의 경로를 결정할 수 있다.

Ⅳ. 시뮬레이션을 통한 성능 검증

4.1 시뮬레이션 환경

본 연구의 제안 프로토콜은 OPNET 18.0.4를 통하여 구현하였으며 제안 프로토콜과 큐러닝 기반의 이전 연구[13] 및 QMR[11]과 비교를 진행하였다. 시뮬레이션은 1개의 CC와 29개의 UAV 노드로 구성된 총 30개의 노드로 구성하였으며 모든 노드는 500m*500m*100m 환경에 고르게 분포되어 있다. 또한 UAV 노드는 0m/s에서 30m/s로 이동하며 좌표 상 임의의 방향으로 이동하는 Random Way Point(RWP) 이동성 모델을 사용하고 CC는 고정된 위치에 존재한다. 트래픽 모델은 100byte의 데이터를 일정 주기로 송신하는 Constant Bit Rate(CBR) 트래픽 모델을 사용하였으며 이는 UAV의 센서 데이터를 가정한다. 표 1은 시뮬레이션의 설정 값을 나타낸다.

표 1. 시뮬레이터 변수
Table 1. Simulator Parameters

Parameters	Values
시뮬레이터	OPNET 18.0.4
네트워크 크기	500 m × 500 m × 100 m
시뮬레이션 시간	500 s
노드 수	30
UAV의 이동 속도	0 m/s - 30 m/s
이동성 모델	Random Way Point (RWP)
비교 프로토콜	QMR, 큐 러닝 기반의 라우팅 프로토콜, 더블 큐 러닝 기반의 라우팅 프로토콜
트래픽 종류	Barometer(100bytes, CBR)
MAC 레이어	802.11n
통신 거리	150 m

4.2 시뮬레이션 결과 분석

비교 프로토콜로 사용하는 QMR은 종단 간 지연과 노드의 배터리 잔량을 고려하는 큐 러닝 기반의 라우팅 프로토콜이다. QMR은 노드의 배터리 잔량을 고려하여 네트워크를 장기간 유지하는 데 장점을 가지지만 고정적인 Hello 패킷 교환을 가지고 다음 홉을 결정하는 데 목적지까지의 홉을 고려한 것이 아닌 보다 근거리의 노드를 선택하기 때문에 데이터를 전송하는 데 보다 많은 홉을 사용하는 경향을 보인다. 제안하는 프로토콜은 기본적으로 목적지까지의 홉 수, 링크 안정성을 고려하여 라우팅을 진행한다. 또한 큐 러닝 기반 프로토콜의 과대평가 문제를 감소시켰으며 배터리 잔량 및 배터리 소모율뿐만 아니라 목적지까지의 홉 수, 링크 안정성을 고려하여 다음 경로를 선정하여 최적의 경로를 선택하도록 하였다.

먼저 그림 1은 제안하는 더블 큐 러닝 기반의 프로토콜과 QMR이 데이터를 목적지까지 전송하는 데 사용한 평균 홉 수를 나타낸다. 제안 프로토콜은 에너지 요소를 고려하기 때문에 이를 고려하지 않은 이전 연구보다 근소하지만 더 많은 평균 홉 수를 보인다. 또한 QMR은 노드의 배터리 잔량을 고려하여 네트워크를 장기간 유지하는 데 장점을 가지지만 고정적인 Hello 패킷 교환을 가지고 다음 홉을 결정하는 데 목적지까지의 홉을 고려한 것이 아닌 보다 근거리의 노드를 선택하기 때문에 제안 기법보다 데이터를 전송하는 데 많은 홉을 사용하는 결과를 보인다. 제안 프로토콜은 목적지까지의 홉 수를 고려하기 때문에 보다 적은 홉을 사용하여 목적지까지 데이터를 전송하

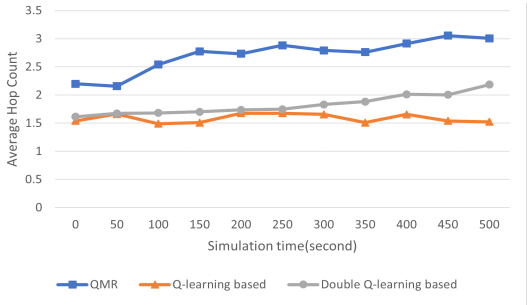

그림 1. 시뮬레이션 시간에 따른 라우팅 경로의 평균 홉 수
Fig. 1. Average Hop Count vs. Simulation Time

는 결과를 보여주었다.

그림 2는 제안 프로토콜과 QMR의 평균 패킷 전송률(PDR)을 나타낸다. PDR은 소스 노드에서 보내는 데이터 패킷 수와 목적지 노드에서 받는 데이터 패킷 수 사이의 비율을 의미한다. UAV 속도가 0m/s일 때, 네트워크는 고정되어 있으므로 안정성이 최대이다. 결과적으로 제안 기법과 QMR은 데이터 전송에 대해 가장 높은 PDR을 갖는다. UAV 속도가 증가함에 따라 네트워크 안정성이 저하되고 제안 기법 및 QMR의 PDR이 네트워크 상태에 영향을 받는다. 제안 프로토콜은 더블 큐 러닝을 적용함으로써 큐 러닝의 과대평가 문제를 감소시켰으며 앞서 그림 1에서 나타나듯 QMR은 데이터를 전송하는 데 제안 기법보다 많은 수의 홉을 사용한다. 따라서 패킷의 충돌 위험이 증가하고 제안 기법 보다 다소 낮은 PDR을 보여주었다. 이에 따라 제안하는 더블 큐 러닝 기반의 라우팅 프로토콜은 QMR보다 2.2% 높은 PDR을 보여주었다.

그림 3은 제안 프로토콜과 QMR이 데이터를 목적지까지 전송하는 데 소요되는 종단 간 지연을 나타낸다. 종단 간 지연은 소스 노드가 패킷을 전송하고 목적지 노드가 수신하는 패킷 간의 시간 차이를 의미한

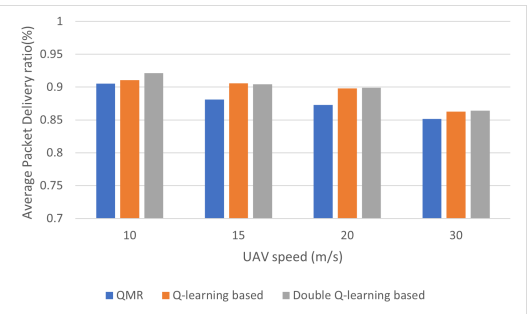

그림 2. UAV 이동 속도에 따른 평균 패킷 전송률
Fig. 2. Average Packet Delivery Ratio vs. UAV's Moving Speed

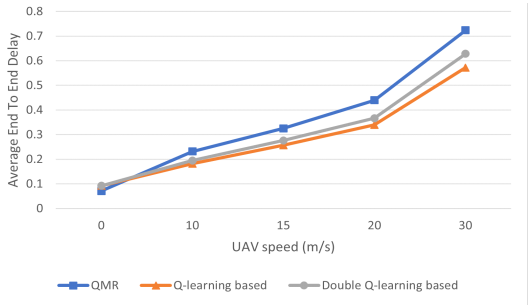

그림 3. UAV 이동 속도에 따른 평균 단대단 지연시간
Fig. 3. Average End To End Delay vs. UAV's Moving Speed

다. 제안 프로토콜은 Q-value를 계산할 때 목적지까지의 홉 수를 고려하기 때문에 근거리의 노드를 선택하는 QMR에 비교하여 데이터를 전송하는 데 있어 적은 수의 홉을 사용한다. 그 결과로 제안 프로토콜은 QMR과 비교하여 낮은 종단 간 지연을 나타내었다.

그림 4는 링크의 평균 에러율을 나타낸다. 링크의 에러율은 소스 노드가 경로로 선정한 다음 노드와의 패킷 전송 실패율을 나타내며 이는 급격한 네트워크 환경 변화와 다음 노드의 Q-value를 과대평가함으로써 발생할 수 있다. 본 연구에서는 큐 러닝의 과대평가 문제를 줄이기 위해 더블 큐 러닝을 적용하였으며 과대평가 문제를 줄임으로써 보다 안정적인 학습을 진행하고 두 개의 Q-value를 유지함으로써 행동에 대한 보상이 편향되지 않도록 하였다. 그 결과로 제안 프로토콜은 큐 러닝 기반인 이전 연구보다 약 10% 낮은 링크 에러율을 보였다.

그림 5는 상대적인 에너지 소모율을 나타낸다. 해당 값은 큐 러닝 기반의 이전 연구의 에너지 소모량을 1로 두었을 때 제안 기법과 QMR의 상대적인 에너지 소모량을 나타낸다. 시뮬레이션 시간이 지날수록 큰 차이를 보이며 제안 기법은 다음 홉 선정에 목적지까

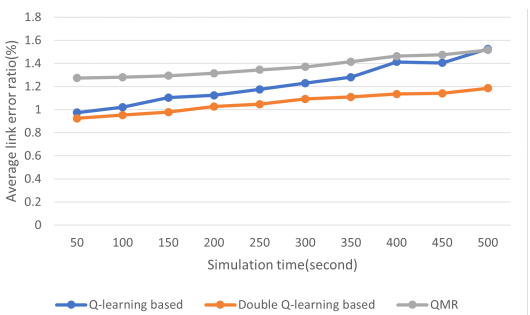

그림 4. 시뮬레이션 시간에 따른 평균 링크 에러율
Fig. 4. Average Link Error Ratio vs. Simulation Time

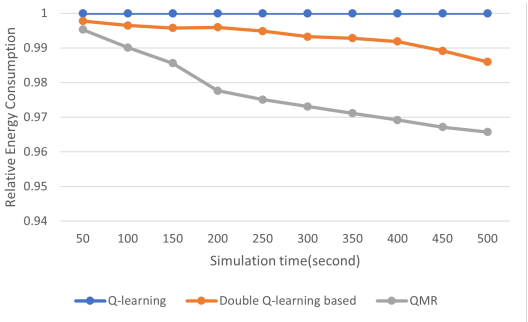

그림 5. 시뮬레이션 시간에 따른 에너지 소비율 비교
Fig. 5. Comparison to Energy Consumption Ratio with Simulation Time

지의 최소 홉 수를 고려하기 때문에 QMR보다 다소 많은 에너지를 사용하지만 QMR 보다 적은 종단 간 지연과 적은 에러율을 보인다. 또한 배터리 환경을 고려하기 때문에 이전 연구보다 적은 배터리 소모를 보인다.

V. 결 론

Multi-UAV 네트워크는 다양한 센서 데이터가 생성되며 노드의 이동성이 매우 높은 네트워크 환경을 가진다. 또한 이러한 네트워크의 환경 및 토폴로지가 빠르게 변화하는 특성은 기존 MANET과 VANET의 라우팅 프로토콜을 적용하기 어렵게 만들었다. 환경이 빠르게 변화하는 특성을 가진 Multi-UAV 네트워크에서는 강화 학습 기반의 적응적인 라우팅 프로토콜이 좋은 성능을 보여주고 있다. 본 논문에서는 Multi-UAV 네트워크의 특성을 적절하게 반영하고 큐 러닝의 문제점인 과대평가를 감소시키기 위해 더블 큐 러닝을 적용한 라우팅 프로토콜 제안하였다. 제안하는 라우팅 프로토콜은 노드 탐색을 통해 목적지까지의 홉 수를 사전에 인지하고 링크 안정성을 고려하여 우수한 성능을 보여주었다. 또한 큐 러닝 기반의 라우팅 프로토콜의 과대평가 문제를 감소시켰고 노드 탐색에 있어 그룹화를 진행하여 제어 오버헤드를 감소시켰다. 이에 따라 QMR과 비교하여 데이터를 전송하는 데 있어 적은 홉을 사용하는 결과를 보여주었고 이에 따라 적은 종단 간 지연을 보여주었다.

References

[1] A. Chriki, et al., "FANET: Communication, mobility models and security issues," *Comput. Netw.*, vol. 163, no. 106877, 2019.

[2] M. F. Khan, et al., "Routing schemes in FANETs: A survey," *Sensors*, vol. 20, no. 1, 2020.

[3] S. Rezwan and W. Choi, "A survey on applications of reinforcement learning in flying ad-hoc networks," *Electronics*, vol. 10, no. 4, 2021.

[4] C. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, pp. 279-292, 1992.

[5] H. Hasselt, "Double Q-learning," in *Advances in NIPS*, 2010.

[6] W.-S. Jung, J. Yim, and Y.-B. Ko, "QGeo: Q-learning based geographic ad hoc routing protocol for unmanned robotic networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2258-2261, 2017.

[7] Z. Zheng, et al., "Adaptive communication protocols in flying ad hoc network," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 136-142, 2018.

[8] Q. Yang, S. J. Jang, and S. J. Yoo, "Q-learning based fuzzy logic for multi-objective routing algorithm in flying ad hoc networks," *Wirel. Pers. Commun.*, vol. 113, pp. 115-138, 2020.

[9] K. Tang, et al., "Reinforcement learning-based opportunistic routing for live video streaming over multi-hop wireless networks," in *IEEE Int. Wkshps. MMSP*, 2017.

[10] M. F. Khan and K. L. Alvin Yau, "Route selection in 5G-based flying ad-hoc networks using reinforcement learning," in *10th IEEE ICCSCE*, Penang, Malaysia, 2020.

[11] J. Liu, et al., "QMR: Q-learning based multi-objective optimization routing protocol for flying ad hoc networks," *Elsevier Comput. Commun.*, vol. 150, pp. 304-316, 2020.

[12] A. Woo and D. Culler, "*Evaluation of efficient link reliability estimators for lowpower wireless networks*," Technique Report UCB// CSD-03-1270, U.C. Berkeley Computer Science Division, 2003.

[13] J. W. Lim and Y.-B. Ko, "Q-learning based

stepwise routing protocol for Multi-UAV networks," *2021 ICAIIC*, Apr. 2021.

임 재 원 (Jae-Won Lim)

2019년 2월 : 아주대학교 정보통신대학 소프트웨어학과 학사
2019년 2월~2021년 8월 : 아주대학교 일반대학원 AI융합네트워크학과 석사
<관심분야> 지능형 IoT 네트워크, 다중드론네트워크 등.
[ORCID:0000-0003-1495-2662]

고 영 배 (Young-Bae Ko)

1996년 9월~2000년 7월 : Texas A&M University (College Station) 컴퓨터공학 박사
2000년 8월~2002년 8월 : IBM T.J Watson Research Center (New York) 전임연구원
2002년 9월~현재 : 아주대학교 정보통신대학 소프트웨어학과/AI융합네트워크학과 교수
<관심분야> 차세대 초지능 통신네트워크(6G), 지능형 사물인터넷(AIoT), 고신뢰 저지연 네트워크(URLLC), 지능형 에지컴퓨팅(Edge Intelligence), 고정밀 실시간 측위 기술 및 서비스 등.
[ORCID:0000-0002-8799-1761]