논문 21-46-11-09 The Journal of Korean Institute of Communications and Information Sciences '21-11 Vol.46 No.11
https://doi.org/10.7840/kics.2021.46.11.1834

랜덤 이동 모델을 따르는 무선 노드의 Q-learning 기반 애드혹 네트워크 구축 전략

김 나 영◆, 권 민 혜°, 박 형 곤°°

Q-learning Based Ad-Hoc Network Formation Strategy for Wireless Nodes with Random Mobility Models

Nayoung Kim◆, Minhae Kwon°, Hyunggon Park°°

요 약

분산형 네트워크는 고정된 네트워크 기반 망의 도움 없이 다수의 노드가 자율적으로 네트워크를 형성하고 유지할 수 있으며, 기반 시설에 구애받지 않고 다양한 환경에 적용할 수 있다는 장점이 있다. 그러나 모바일 기기의 수가 늘어남에 따라 최적의 네트워크를 구축하기 위한 계산 복잡도가 높아지고, 모바일 기기의 이동성으로 인해 네트워크 처리율을 안정적으로 유지하기 어렵다는 한계점이 존재한다. 또한, 모바일 기기의 경우, 일반적으로 배터리 기반으로 동작하기 때문에 제한된 배터리 용량 내에서 이동을 위한 모터 제어 에너지 소모량과 데이터 송수신을 위한 에너지 소모량을 모두 해결해야 한다는 어려움이 있다. 본 논문에서는 Q-learning 기반의 강화학습을 사용한 모바일 애드혹 네트워크 구축 알고리즘을 제시한다. 제안하는 방안은 랜덤하게 이동하는 노드들로 구성된 네트워크에서 중앙 제어 없이 전력 소모를 최소화하는 동시에 네트워크 처리율을 최대로 유지하는 것을 목표로 노드를 학습시킨다. 시뮬레이션 결과를 통해 제안한 알고리즘이 네트워크 크기와 이동 모델과 관계없이 항상 높은 네트워크 처리율을 유지할 뿐만 아니라 에너지 소비를 최소화할 수 있음을 확인하였다.

Key Words : Q-learning, Mobile Ad-Hoc Network, Random Mobility Model

ABSTRACT

Since a decentralized network can be formed and maintained by multiple nodes without direct control from network infrastructure, it can be deployed in a wide range of applications. As more nodes need to be taken into account, however, computational complexity required for an optimal network significantly increases. It is more challenging to maintain higher network throughput because of node mobility. For a mobile ad-hoc network that consists of battery-powered mobile devices, in particular, it is even more important to consider power consumption required for node mobility and data transmission. In this paper, we propose an algorithm for a mobile ad-hoc network formation based on Q-learning. The proposed algorithm enables relay nodes with random movement to learn how to manage transmission power for minimizing total power consumption and maximizing network throughput without a central coordination. We confirm that the proposed solution can achieve high network throughput while minimizing power consumption regardless of network size and mobility patterns.

※ 본 연구는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2020R1A2B5B01002528, NRF-2020R1F1A1069182) 및 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2021-0-00739)을 받아 수행되었습니다.

◆ First Author : Ewha Womans University Department of Electronic and Electrical Engineering, Graduate Program in Smart Factory, nayoung.kim@ewhain.net, 학생회원

° Corresponding Author : Soongsil University School of Electronic Engineering, minhae@ssu.ac.kr, 정회원

°° Corresponding Author : Ewha Womans University Department of Electronic and Electrical Engineering, Graduate Program in Smart Factory, hyunggon.park@ewha.ac.kr, 종신회원

논문번호 : 202106-120-B-RN, Received June 3, 2021; Revised July 26, 2021; Accepted August 11, 2014

Ⅰ. 서 론

모바일 기기의 수는 계속해서 증가하는 추세이며, 최근 기기 간 초연결 사회로 진입하며 무선 네트워크의 중요성은 더욱 대두되고 있다. 중앙형 네트워크는 기반 망의 확장을 통해 인터넷 수요의 폭발적 증가를 해결해왔으나, 네트워크 기반 시설의 확충이 한계에 다다르고 사물 인터넷 패러다임이 제시되면서 D2D (Device to Device)와 M2M (Machine to Machine)과 같은 분산형 네트워크가 새로운 해결책으로 부상하고 있다[1]. 이는 차량 간 네트워크, 센서 네트워크 등으로 확장되고 있다.

분산형 네트워크는 상호 접속된 복수의 개별적인 노드에 의해 통신 및 제어 기능이 수행된다. 대표적으로는 모바일 애드혹 네트워크 (Mobile Ad-hoc Network, MANET)가 있다[2]. MANET은 이동성을 가진 노드가 자율적으로 애드혹 네트워크를 구성한다. 애드혹은 릴레이 노드들이 데이터를 수신하고 송신하며, 소스 노드부터 목적 노드까지 데이터를 전송하는 방식을 말한다. 이는 사람의 접근이 어려운 환경에서 무선 모바일 기기를 통해 필요한 데이터를 수집하고, 수집한 데이터를 상당히 멀리 떨어진 목적지까지 전송해야 할 때 유용하다. 따라서 MANET은 군사 분야, 센서 네트워크, 재난 상황에서의 비상 네트워크와 같이 통신 기반 망이 부재하며 노드가 이동성을 가져야 하는 상황에 주로 활용된다[3-4]. 최근에는 릴레이 노드로 드론을 활용한 연구가 활발히 진행되고 있다[5].

MANET은 다양한 환경에 적용될 수 있으나, 여러 한계점을 갖는다. MANET에서는 안정적인 네트워크 처리율 (throughput)의 유지가 어렵다. 위치 정보를 비롯한 릴레이 노드의 정보가 실시간으로 변화하므로 환경의 다이나믹스 (dynamics)가 커져 노드 간 연결성이 불안정해질 수 있기 때문이다. 또한, 일반적으로 노드들은 배터리와 같은 한정된 에너지원으로 대기 전력과 이동에 필요한 전력, 통신에 필요한 전력을 사용해야 하므로 효율적인 에너지 사용이 중요하다. 마지막으로 노드 수가 증가함에 따라 모든 노드 정보의 실시간 수집이 어려워지는데, 노드 정보를 수집하는 과정은 통신 전력을 소모하게 되므로 에너지 소모 문제를 심화시킬 수 있다. 따라서 최소한의 주변 정보로 노드를 제어하는 방안이 필요하다.

결과적으로 MANET은 릴레이 노드의 이동성이 예측 불가한 상황에서 네트워크의 규모가 커지더라도 연결성을 유지할 수 있어야 한다. 또한, 릴레이 노드의 원활한 작동을 위해 전체 네트워크의 에너지 소비도 최소화하여야 한다[6,7].

MANET은 구성 노드의 분산적 의사결정에 의해 구축될 수 있으며, 가장 대표적인 분산 의사결정 방법이 강화학습이다[8]. 강화학습은 학습 주체인 에이전트가 환경과 직접 상호작용하면서 환경에 맞는 최적의 행동을 학습할 수 있게 한다. 에이전트가 환경의 상태 정보에 따라 행동을 결정하면 그에 따른 보상을 얻게 되는데, 최적의 행동은 각 상태에서 최대의 보상을 얻는 행동이다. 강화학습은 환경에 대한 사전 지식이 없어도 학습 가능하므로 환경의 다이나믹스가 존재하는 경우에 적절하다. 또한, 다수의 에이전트가 공통 목표를 달성하기 위해 개별적으로 학습할 수 있으므로 네트워크 구축과 같이 한 개 이상의 에이전트가 협력해야 하는 상황에 적용할 수 있다.

통신 주체이자 학습 주체가 되는 릴레이 노드를 강화학습을 이용해 학습시키면 이동성을 비롯한 네트워크의 변동성에도 불구하고 안정적인 네트워크를 구축할 수 있다. 본 논문은 강화학습의 일종인 Q-learning[9]을 기반으로 랜덤한 이동성을 가지는 환경에서 네트워크의 연결성을 유지하면서 에너지 소모량은 최소화하는 알고리즘을 제안한다.

본 논문은 다음과 같이 구성되어 있다. Ⅱ장에서는 본 연구와 관련된 선행 연구를 소개한다. Ⅲ장과 Ⅳ장은 Q-learning 기반 네트워크 구성과 알고리즘을 제안한다. Ⅴ장에서는 시뮬레이션을 통해 제안한 알고리즘의 성능을 평가한다. Ⅵ장에서는 본 논문에서 제안한 네트워크 구축 알고리즘에 대한 결론을 맺는다.

Ⅱ. 선행 연구

초기에는 효율적인 라우팅 방법을 중점으로 무선 애드혹 네트워크 구축 전략에 관한 연구가 진행되었다. 라우팅 방법 개선을 위해 분산 해시 테이블 (distributed hash table)을 이용한 분산 네트워크 구성 방법과 소프트웨어 정의 네트워크 (Software Defined Network, SDN) 컨트롤러를 사용한 무선 멀티 홉 네트워크 형성 방안이 제시되었다[10,11]. SDN 컨트롤러를 적용한 방안은 SDN 컨트롤러가 노드의 위치와 에너지 레벨을 저장하고 있으며 구성 노드는 2개 홉까지의 이웃 노드 정보와 라우팅 목록을 가지고 있는 상황을 가정하고 있다. 또한, 모든 노드가 1개 홉 거리의 이웃 노드를 기준으로 최적 네트워크 토폴로지가 될 수 있는 후보를 선정하고 현재까지 축적된 에너지 소모량을 기준으로 네트워크를 구축하는 방식도 제안되었다[12]. 이러한 네트워크 구축 방안은 분산형 네트워크

의 구조를 따르지만, 중앙 시스템의 제어가 필요하다.

빠르게 모바일 환경이 보편화 되면서, 이동성을 포함한 애드혹 네트워크 구축 전략에 관해 연구되었다. 그중 차량 애드혹 네트워크 (vehicular ad hoc network)를 가중 방향성 그래프 (weighted-directed graph)로 모델링 한 뒤 데이터 전송 경로를 선택하는 연구가 있다[13]. 이 연구에서는 밀도 측면에서 노드 변화율에 평균을 취한 뒤, 그 값을 기반으로 경로를 선택한다. 이는 이동성을 고려하여 네트워크를 구축하고자 한 방식이나, 기존의 중앙 제어형 라우팅 알고리즘에서 크게 벗어나지 못했다는 한계를 갖는다. 고정 노드와 이동성을 가지는 노드로 구성된 애드혹 네트워크에서 고정 노드의 데이터 송신 에너지 소모량을 감소시키기 위해 노드의 이동성을 예측한 연구가 있다[14]. 이 연구에서는 노드의 이동성을 예측 및 조절함으로써 안정적이며 에너지 효율적인 네트워크를 구축하였으나, 모든 노드의 데이터 송신 에너지가 노드 간 거리에 따라 자율적으로 조절된다고 가정하였다는 한계가 있다. 이동 모델을 기반으로 노드의 분포를 분석하여 최적의 데이터 전송 경로를 찾는 방안도 제안되었다[15]. 이 연구에서는 랜덤 방향성 이동 모델 (random direction mobility model)로 노드의 이동성을 한정 지었다.

개별 노드가 네트워크를 구축하는 분산 해결책은 강화학습을 중심으로 연구되었다. 특히 강화학습 알고리즘은 대표적인 분산 의사결정 방법으로 네트워크 구축에 다수 적용되었다. 보상 기반의 라우팅 프로토콜 (reward-based routing protocol)은 홉의 개수, 대역폭, 배터리, 그리고 노드의 이동 속도 등을 가중치로 두고 최적의 경로를 찾는 방식이다[16]. 이 연구에서는 강화학습에 사용하는 보상 함수를 도입했으나, 많은 노드 정보를 요구하므로 현실성이 부족하다는 한계가 있다.

최소의 노드 정보만을 사용하기 위해, 무작위로 토폴로지를 형성한 뒤 그에 대한 학습 과정을 거치는 방안이 제안되었다[17]. 이 연구에서는 최대의 전송 전력으로 브로드캐스트하여 이웃 노드의 수를 파악한다. 이후, 강화학습을 도입하여 에너지를 절약하는 동시에 네트워크 연결을 유지하는 것을 목표로 학습시킨다. 그러나 각 노드가 이동할 경우 획득한 인접 노드의 정보가 변경되는 문제가 발생할 수 있다.

MANET의 한정된 전력량 문제를 해결하기 위해 에너지 하베스트 노드 (energy harvest node)를 에이전트로 가정한 모델 기반의 학습 방안이 연구되었다[18]. 이 논문은 이동성 그룹 모델 (reference point group mobility)로 노드의 이동성을 모델링하였으며, 인접 노드와의 거리와 인접 노드의 배터리 상태를 상태 정보로 정의한다. 에너지 하베스트 노드를 활용한 다른 연구에서는 기본 사용자 (primary user)와 보조 사용자 (secondary user)로 노드를 구분하였으며, 상태 정보로는 축적한 에너지, 현재 배터리 수준, 그리고 인접 노드 정보를 사용한다[19].

MANET을 구축하는 방법과 MANET의 한계를 해결하기 위한 다양한 선행 연구가 존재한다. 그러나 네트워크의 연결성을 유지하는 것에 중점을 둔 연구는 필요한 상태 정보의 양이 많아 에너지 소모량을 감소시키기 어려우며, 이를 해결하기 위해 에너지 하베스트 모델을 사용한 연구는 일반적인 상황을 표현할 수 없다는 한계가 존재한다. 따라서 본 논문에서는 전체 에너지 소모량을 절약할 수 있으면서 네트워크의 연결성을 유지하는 MANET 구축 알고리즘을 제시하고자 한다.

Ⅲ. 이동성 가지는 노드의 Q-learning 기반 모바일 애드혹 네트워크

3.1 MANET 구성

본 논문에서 데이터 전송을 위하여 고려하는 MANET은 시간 t에서 소스 노드부터 목적 노드까지의 방향성 그래프 G_t로 표현한다. 소스 노드, 목적 노드, 릴레이 노드를 모두 포함한 전체 노드의 수는 k이며, 소스 노드와 목적 노드는 직접 연결될 수 없을 만큼 멀리 위치한 상황을 고려한다. 따라서 소스 노드가 수집한 데이터를 목적 노드까지 전달하기 위해서는 그림 1과 같이 릴레이 노드를 거치는 애드혹 형태로 네트워크를 구성해야 한다.

MANET은 릴레이 노드 사이의 여러 동적 상호 작용의 결과에 따라 구축되고 유지되므로, 본 논문에서

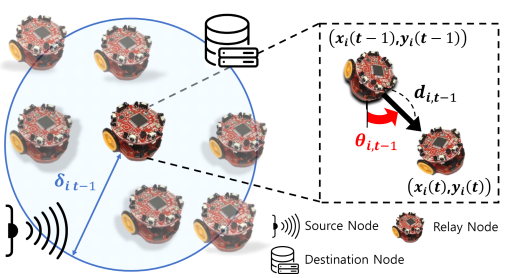


그림 1. 제안하는 모바일 애드혹 네트워크 구성 및 릴레이 노드의 위치
Fig. 1. Proposed mobile ad-hoc network formation and locations of relay nodes

는 릴레이 노드가 결정 주체인 마코프 결정 과정 (Markov Decision Process, MDP)으로 설정한다. i번째 릴레이 노드를 n_i라 할 때, 소스 노드와 목적 노드를 제외한 릴레이 노드 집합은 $N(G_t)=\{n_i \mid i=0,1,\cdots k-3\}$이다. 전체 릴레이 노드의 수는 $|N(G_t)|$이며, $|N(G_t)|$개의 릴레이 노드는 중앙 제어가 부재한 상황에서 랜덤하게 이동한다.

본 논문에서 릴레이 노드의 이동성은 랜덤 이동 모델 (random mobility model)을 따른다. i번째 릴레이 노드 n_i의 시간 t에서의 이동 거리는 $d_i(t)$이며, 이동 각도는 $\theta_i(t)$이다. 그림 1과 같이, 릴레이 노드의 시간 $t-1$에서의 위치를 $(x_i(t-1), y_i(t-1))$라 하면, 이동 후의 위치는 $(x_i(t), y_i(t))$로 나타낼 수 있다.

릴레이 노드 n_i의 시간 t에서의 전송 범위 반지름은 $\delta_{i,t}$라하고, 그림 1과 같이 전송 면적 $\pi\delta_{i,t}^2$내에 존재하는 모든 노드에 데이터를 전송한다. 애드혹 네트워크는 릴레이 노드 간의 전송 면적이 중첩되어 최소 2개 홉으로 데이터 전송이 가능할 경우 형성된다[20]. 모든 릴레이 노드는 이중 통신 (full duplex)을 가정하여 데이터를 동시에 송수신한다. n_i의 전송 에너지는 $\delta_{i,t}$에 비례한다. 네트워크 처리율은 단위 시간당 전송되는 데이터의 양으로, 소스 노드와 목적 노드 사이의 홉의 개수에 반비례한다[21]. 릴레이 노드가 전송 에너지를 증가시키면 네트워크 처리율이 향상되지만, 소비하는 에너지도 같이 증가한다. 이때, 릴레이 노드의 최대 전송 전력은 제한되어 있으나, 모든 노드의 최대 전력량은 충분하다고 가정한다.

3.2 MANET 구성

본 논문에서는 다양한 릴레이 노드의 이동 형태를 반영하기 위하여 랜덤 워크 (random walk), 랜덤 웨이포인트 (random waypoint), 그리고 레비 워크 (Lévy walk) 모델을 사용한다[22,23]. 각각의 릴레이 노드는 할당된 영역 내에서 움직인다고 가정하여 이동성에 따른 충돌문제는 발생하지 않는다.

랜덤 워크 모델은 브라운 운동 (Brownian motion)이라고도 하며, 분자의 확산 모델과 같이 일반적인 랜덤 이동 상황에 적용할 수 있다. 시간 t에서의 릴레이 노드 n_i의 각도 $\theta_i(t)$와 이동 거리 $d_i(t)$를 균등 확률 분포 (uniform random distribution, $\mathrm{U}(a,b)$)에 따라 결정하므로, $\theta_i(t)\sim\mathrm{U}(0,2\pi)$, $d_i(t)\sim\mathrm{U}(0,D_{\max})$를 만족한다. 이때 $D_{\max}$는 시간 t에서 최대로 움직일 수 있는 거리이다. 식 (1)은 랜덤 워크 모델을 따르는 릴레이 노드 n_i의 시간 t에서의 위치를 나타낸다.

$$\begin{aligned}&(x_i(t), y_i(t))\\&\quad=(x_i(t-1)+d_i(t)\cos\theta_i(t),\\&\qquad y_i(t-1)+d_i(t)\sin\theta_i(t))\end{aligned} \tag{1}$$

랜덤 웨이포인트 모델은 랜덤 워크 모델과 유사하지만, 각도와 이동 거리가 아닌 다음 시간 t에서의 x좌표와 y좌표를 균등 확률 분포에 따라 결정한다. 랜덤 웨이포인트 모델은 모바일 기기 사용자가 이동할 때, 사용자의 위치와 속도를 표현하기 위해 주로 사용된다. 따라서 다음 시간 t에서의 릴레이 노드 n_i의 위치는 식 (2)에 따라 결정된다.

$$x_i(t)\sim\mathrm{U}(0,D_{\max}),\ y_i(t)\sim\mathrm{U}(0,D_{\max}) \tag{2}$$

레비 워크 모델은 시간 t에서의 릴레이 노드 n_i의 이동 거리 $d_i(t)$를 레비 분포 (Lévy distribution, $\mathrm{Lévy}(\mu,c)$)에 따라 결정하는 모델이다. 레비 워크 모델은 인간의 움직임을 가장 잘 표현하는 것으로 알려져 있다. 레비 분포의 질량 함수에서 $\mu\in(-\infty,\infty)$는 분포의 평균을 결정하고, $c\in(0,\infty)$는 분산을 결정한다. 식 (3)은 레비 분포의 확률 밀도 함수이다[24].

$$x\sim\mathrm{Lévy}(\mu,c)=\sqrt{\frac{c}{2\pi}}\,\frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}} \tag{3}$$

레비 분포는 지수함수보다 느리게 감쇄되는 두꺼운 꼬리 분포 (heavy tail feature)를 가진다는 특징이 있다. 이로 인해 평균에서 크게 벗어난 값이 선택될 확률이 높다. 레비 워크 모델에서 각도 $\theta_i(t)$는 균등 확률 분포를 따르므로, $d_i(t)\sim\mathrm{Lévy}\left(\frac{1}{2},1\right)$이고 $\theta_i(t)\sim\mathrm{U}(0,2\pi)$이다. 레비 워크 모델을 따르는 릴레이 노드 n_i의 시간 t에서의 위치는 식 (1)을 통해 계산한다.

Ⅳ. 랜덤 이동 모델에 따른 이동성을 가지는 무선 노드의 Q-learning 기반 네트워크 구축

4.1 MDP 설정

MDP는 $<S,A,R,\gamma>$의 형태로 표현되며, S는 유한한 상태 공간 (state space), A는 유한한 행동 공

간 (action space), R은 행동에 따른 보상 (reward), γ는 감가율 (discount factor)이다. 에이전트 (agent)는 MDP에 의해 환경과 상호작용하며 누적 보상을 최대화하는 것을 목표로 동작한다.

4.1.1 상태(state)

본 논문에서 상태 정보 $s_{i,t} \in S$는 시간 t에서 릴레이 노드 n_i가 데이터를 전송했을 때, 수신 가능한 노드들의 수로 정의한다. 즉, 릴레이 노드 n_i의 상태는 시간 t에서의 전송 범위 반지름 $\delta_{i,t}$에 의해 결정된 전송 면적 $\pi\delta_{i,t}^2$ 안에 존재하는 주변 노드들의 수로 식 (4)와 같이 나타낼 수 있다.

$$s_{i,t} = \left|\{n_j \mid d(n_i, n_j) \le \delta_{i,t}, j \ne i\}\right| \tag{4}$$

이때 $0 \le s_{i,t} \le k-1$이다. 노드 n_i의 전송 범위 반지름 $\delta_{i,t}$이 주변 노드와의 유클리디안 거리 (Euclidian distance)보다 크다면 전송 범위 안에 주변 노드가 존재한다. 주변 노드와의 거리가 전송 범위 반지름보다 크다면 전송 범위 안에 위치하지 않는다.

4.1.2 행동(action)

행동 $a_{i,t} \in A$은 식 (5)와 같이 이전 시간 $t-1$과 비교했을 때 시간 t에서 변화한 전송 범위 반지름을 나타낸다.

$$a_{i,t} = \delta_{i,t} - \delta_{i,t-1} \tag{5}$$

$a_{i,t} > 0$이면, $|\delta_{i,t} - \delta_{i,t-1}|$만큼 전송 범위를 확장하고, $a_{i,t} < 0$라면 전송 범위를 축소한다. $a_{i,t} = 0$인 경우에는 이전 시간에서의 전송 범위를 그대로 유지한다. 전체 행동 공간의 크기는 $|A|$이다.

4.1.3 보상(reward)

보상 $r_{i,t} \in R$은 학습 목표 달성에 의한 보상과 처벌로 구성된다. 보상 항은 네트워크 처리율의 변화량 $\phi(s_{i,t}) - \phi(s_{i,t-1})$이며, 처벌 항은 전송 범위 반지름 변경에 따른 송신 에너지 소모량 함수 $F(a_{i,t})$로 정의한다.

행동에 따른 송신 에너지 소모량 $F(a_{i,t})$은 각 릴레이 노드의 송신 에너지 함수 $G(\delta_{i,t-1}) = \frac{P_r \lambda^2}{A_t A_r}\delta_{i,t-1}^2$로 식 (6)과 같이 표현할 수 있다[25]. 이때, P_r은 수신 에너지, A_t는 송신 안테나의 유효 범위, A_r은 수신 안테나의 유효 범위, λ는 파장을 의미한다.

$$\begin{aligned} F(a_{i,t}) &= G(\delta_{i,t-1} + a_{i,t}) - G(\delta_{i,t-1}) \\ &= \frac{P_r \lambda^2}{A_t A_r}\left[(\delta_{i,t-1} + a_{i,t})^2 - (\delta_{i,t-1})^2\right] \\ &= \frac{P_r \lambda^2}{A_t A_r}\left[2\delta_{i,t-1} a_{i,t} + a_{i,t}^2\right] \end{aligned} \tag{6}$$

따라서 본 논문에서 보상 함수는 식 (7)과 같이 보상 항과 처벌 항의 선형 결합으로 나타낸다.

$$\begin{aligned} r_{i,t} = {} & \omega \times (\phi_\tau(s_{i,t}) - \phi_{t-1}(s_{i,t-1})) \\ & -(1-\omega) \times F(a_{i,t}) \end{aligned} \tag{7}$$

선형 결합 비율은 균형 계수 (balance coefficient) $0 \le \omega \le 1$에 의해 이루어진다. $\omega = 1$이면 에이전트는 네트워크 처리율의 최대화만을 목표로 행동을 결정하고, $\omega = 0$인 경우에는 에너지 소모의 최소화만이 목적이 된다. 이때, 네트워크 처리율이 $\phi(s_{i,t}) > 0$이면 소스 노드-목적 노드 간 네트워크 연결이 이루어졌음을 의미한다. 에너지 소모량이 노드별로 행동에 따라 결정되는 것과는 달리, 네트워크 처리율의 변화량은 모든 노드가 같은 값을 갖기 때문에 일반적으로 릴레이 노드의 송신 에너지의 총합이 커질수록 네트워크 처리율이 증가한다.

각 에이전트의 최종 목표는 미래의 누적 보상을 최대로 만드는 정책 π^*을 결정하는 것이다. 정책은 관찰된 상태를 기반으로 최적의 행동을 결정하는 방식이다. 본 논문에서는 제시한 MDP 문제를 풀기 위해 강화학습 알고리즘 중 Q-learning을 사용한다. Q-learning에서 최적의 행동이란 각 상태에서 최대의 Q-함숫값을 가지는 행동으로 정의된다. Q-함숫값은 Q-함수에 의해 도출되는데, Q-함수는 매시간 t마다 정책 π에 따라 감가율 $0 \le \gamma < 1$만큼 감소한 누적 보상 값에 평균을 취한다. 감가율 γ은 미래 보상과 현재 보상 사이의 균형을 맞춰주는 계수로, 1에 가까울수록 미래 보상의 가치를 높게 평가한다. 식 (8)과 식 (9)은 학습을 시간 t_{end}까지 지속할 때, 시간 t에서 i번째 릴레이 노드 n_i의 Q-함수와 최적의 행동 a^*을 나타낸 수식이다.

$$Q_i^{\pi}(s,a) = E_{\pi}\left[\sum_{\tau=t+1}^{t_{end}} \gamma^{\tau-t-1} r_{i,\tau} \mid s_{i,t}=s, a_{i,t}=a\right] \quad (8)$$

$$a^* = \pi^*(s) = \arg\max_a Q_i^{\pi^*}(s,a) \quad (9)$$

4.2 Q-learning 기반 네트워크 구축 알고리즘

본 논문에서는 tabular-Q 방법을 적용하여 에이전트를 학습시킨다[26]. 전체 학습 방식의 의사 코드는 알고리즘 1에 나타내었다. tabular-Q는 $Q(s,a)$로 구성된 Q-table을 에이전트마다 생성하는 방식으로, 정책 $\pi^*(s)$은 식 (9)과 같다. 학습은 모든 에이전트가 행동에 따른 상태를 관찰하고 $Q(s,a)$를 계산하여 Q-table을 완성하는 과정을 의미한다. 이는 정해진 에피소드(episode) 횟수 T_{end}가 끝날 때까지 반복되며, 새로운 에피소드가 시작될 때마다 상태와 행동, 네트워크 처리율, 그리고 에이전트의 위치를 초기화한다.

하나의 에피소드는 다수의 스텝 (step)으로 구성되며, 매 스텝 t마다 모든 에이전트는 개별적으로 현재 상태를 인식하고 그에 따라 분산적으로 행동 공간에서 행동을 결정한다. 각 에이전트는 결정된 행동에 따라 전송 범위 반지름 $\delta_{i,t}$을 변화시킨다. 이때, 최대 전송 범위를 초과하거나 최소 전송 범위인 0보다 작게 전송 범위 반지름을 변화시키는 행동이 선택된 경우, 다시 행동을 선택한다.

에이전트가 최적의 정책을 학습하기 위해서는 관찰한 상태에서 다양한 행동을 시도해보는 탐색(exploration) 과정이 필요하다. 즉, 에이전트는 반복적으로 현재 상태에 대한 행동을 탐색하거나 정책에 따라 행동을 결정하면서 정책을 학습한다. 본 논문에서는 알고리즘 2의 ϵ-greedy 방식에 따라 행동을 선택한다. ϵ-greedy 방식은 ϵ의 확률로는 균등확률분포에 따라 전체 행동 공간에서 행동을 탐색하고, 1-ϵ의 확률로는 만들어진 정책에 따라 행동을 결정한다. ϵ은 식 (10)와 같이 $\epsilon_{\min}$에 도달할 때까지 매 에피소드 T마다 ϵ_{decay}만큼 선형적으로 감소하며, 이를 통해 탐색과 이용 (exploitation)이 적정하게 시행될 수 있다.

$$\epsilon = \min(\epsilon_{\min}, 1-\epsilon_{decay} \times T) \quad (10)$$

이때 $T=0,1,\cdots T_{end}$이다. 에이전트가 현재 상

알고리즘 1. 제안하는 Q-learning 기반의 랜덤 이동 모델을 따르는 MANET 구축 알고리즘
Algorithm 1. The proposed Q-learning based MANET formation algorithm with random mobility models

Algorithm 1 Q-learning Based Random Moving Node Activation

Require node set $N(G_t)$, action set A, Q-table Q, greedy rate ϵ

for *episode* = 0. 1, . . ., T_{end} **do**
  for all $n_{i,t} \in N(G_t)$ **do**
    Initialize $(x_{i,0}, y_{i,0})$
    $\delta_{i,0}=0, s_{i,1}=0, \phi_{i,0}=0$
  end for
  for *step* = 1, 2, . . ., t_{end} **do**
    for all $n_{i,t} \in N(G_t)$ do
      $a_{i,t} \leftarrow$ Get_Action($s_{i,t}, \epsilon, A, \delta_{i,t}$)
      $\delta_{i,t} \leftarrow \delta_{i,t+1} + a_{i,t}$
      Transmit data
      Update_Q_table($s_{i,t}, a_{i,t}, r_t, s_{i,t+1}$) using (11)
      $(x_{i,t+1}, y_{i,t+1})$ $\leftarrow$ Random_Mobility$(x_{i,t}, y_{i,t})$
    end for
  end for
end for

알고리즘 2. ϵ-greedy 방식의 행동 선택
Algorithm 2. Action selection based on ϵ-greedy method

Algorithm 2 ϵ-greedy Action Selection of Node $n_{i,t}$

function Get_Action($s_{i,t}, \epsilon, A, \delta_{i,t}$)
  Generate random variable $p \sim U(0,1)$
  if $p < \epsilon$ **then**
    Randomly select $a_{i,t} \in INA$
  else:
    $a_{i,t} \leftarrow \arg\max_a Q_i^{\pi}(s_{i,t}, a_i)$
return $a_{i,t}$

알고리즘 3. 랜덤 이동 모델을 따르는 노드의 다음 위치 결정
Algorithm 3. Selection of the next location based on random mobility models

Algorithm 3 Random Mobility of Node $n_{i,t}$

function Random_Mobility$(x_{i,t}, y_{i,t})$
  Generate random variables
  Evaluate next location $(x_{i,t+1}, y_{i,t+1})$ using (1) or (2)
  if next location is in pre-allocated region **then**
    return $(x_{i,t+1}, y_{i,t+1})$
  else:
    return Random_Mobility$(x_{i,t}, y_{i,t})$

태에 따른 행동을 확정하면, 네트워크 처리율이 계산되며 이에 대한 보상이 주어진다. 에이전트는 식 (11)에 나타난 바와 같이, 학습률 (learning rate) α에 의해 기존의 Q-값과 현재 계산된 Q-값을 선형 결합한 형태로 Q-table을 갱신한다. 이때, 학습률 α가 0에 가까울수록 Q-값은 보수적으로 갱신된다.

$$Q_i(s_{i,t-1}, a_{i,t-1}) \leftarrow Q_i(s_{i,t-1}, a_{i,t-1}) + \alpha(r_{i,t} + \gamma \max_a Q_i(s_{i,t}, a_i) - Q_i(s_{i,t-1}, a_{i,t-1})) \quad (11)$$

이후, 모든 에이전트는 알고리즘 3에 의해 할당된 영역 내에서 랜덤 이동 모델에 따라 이동한다. 최대 이동 거리는 할당된 영역의 크기에 비례하고, 하나의 에피소드에 할당된 스텝의 수 t_{end}에 반비례한다.

V. 네트워크 구축 시뮬레이션 결과

5.1 제안한 알고리즘의 성능 검증

본 논문에서 제안한 알고리즘의 성능을 검증하기 위해 랜덤 이동 모델 종류에 따른 실험과 네트워크 크기에 따른 실험을 진행하였다. 랜덤 이동 모델은 랜덤 워크, 랜덤 웨이포인트, 레비 워크 모델을 적용하고, 네트워크 크기는 3×3, 5×5, 10×10으로 설정한다. 모든 노드는 네트워크 내 1×1 영역마다 배치한다. 소스 노드와 목적 노드는 각각 1개로 고정되며, 에이전트인 릴레이 노드는 전체 네트워크 크기가 커짐에 따라 비례해서 증가한다. 즉, 전체 네트워크 크기가 3×3, 5×5, 10×10일 때, 에이전트의 수 $|N(G_\tau)|$는 각각 7개, 23개, 98개이다. 이때 에이전트의 최대 전송 전력 역시 네트워크 크기에 비례해 증가한다. 전체 네트워크 크기가 3×3, 5×5, 10×10일 때, 에이전트의 최대 전송 전력은 2, 3, 8로 설정한다. 균형 계수 ω는 0.65, 0.75, 0.85를 각각 사용하였다. 이는 소스 노드-목적 노드 사이의 릴레이 노드 수가 늘어나며 발생하는 변동성을 반영한 것이다.

에이전트 1개에 할당된 영역은 1×1로 일정하고 에피소드 1개는 100개의 스텝으로 구성된다. 따라서 이동 모델과 관계없이 1번의 스텝 t에서 최대 이동 거리는 0.01이다. 학습에 사용한 감가율은 $\gamma = 0.9$이며, 학습률은 $\alpha = 0.001$이다.

그림 2는 학습 시간에 따른 노드 당 평균 에너지 소모량과 네트워크 처리율을 나타낸 그래프이다. 탐색은 10000번째 에피소드까지 진행되었다. 그림 2를 통해, 탐색 기간에는 평균 에너지 소모량은 감소하고 네트워크 처리율은 증가함을 확인할 수 있다. 또한, 10000번째 에피소드 이후 탐색이 종료되었을 때 평균 에너지 소모량과 네트워크 처리율 모두 수렴한다.

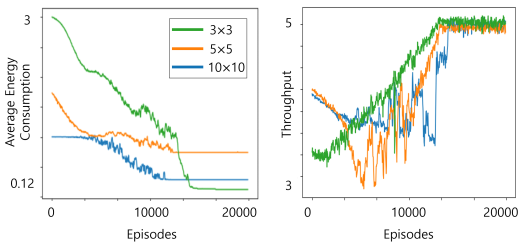


그림 2. 시간에 따른 3×3, 5×5, 10×10 네트워크에서 랜덤 워크 이동성을 가질 때, 평균 에너지 소모량과 네트워크 처리율
Fig. 2. Average energy consumption and throughput in 3×3, 5×5 network and 10×10 network with random walk mobility

그림 3, 4, 5, 6은 학습 시간에 따른 릴레이 노드별 평균 전송 에너지 변화를 나타낸다. 전송 에너지의 양은 우측에 서로 다른 색으로 나타내었다. 노란색에 가까울수록 높은 전송 에너지를 사용하였음을 의미하며 남색에 가까울수록 낮은 전송 에너지를 사용하였음을 의미한다.

그림 3는 랜덤 워크 이동성을 갖는 5×5 네트워크에서 학습이 진행됨에 따라 0번, 3번, 4번, 5번, 11번, 19번 노드가 선택적으로 켜져 애드혹을 형성함을 보여준다. 특히 11번 노드는 최대 전송 전력을 계속 유지하며 2개 홉으로 소스 노드와 목적 노드를 연결한다. 이에 따라 네트워크 처리율에 필수적인 노드만 사

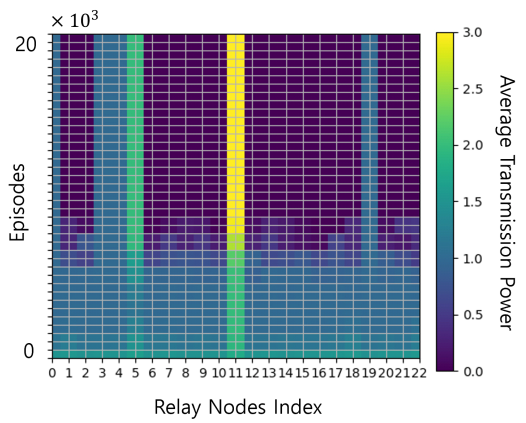


그림 3. 5×5 네트워크에서 랜덤 워크 이동성을 가지는 릴레이 노드들의 평균 전송 에너지
Fig. 3. Average transmission power of relay nodes in 5×5 network with random walk mobility

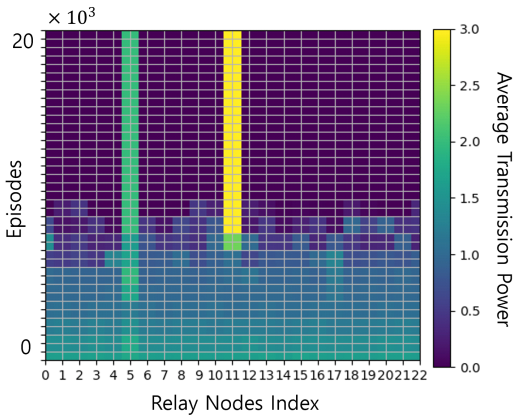


그림 4. 5×5 네트워크에서 랜덤 웨이포인트 이동성을 가지는 릴레이 노드들의 평균 전송 에너지
Fig. 4. Average transmission power of relay nodes in 5×5 network with random waypoint mobility

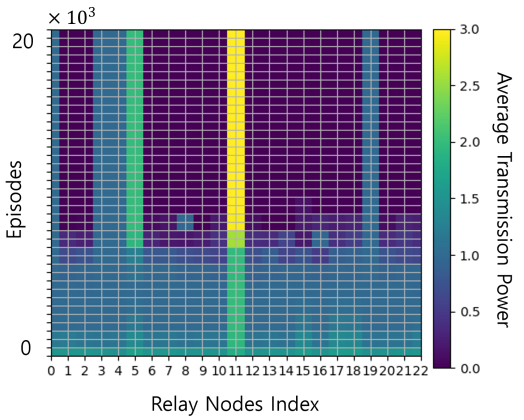


그림 5. 5×5 네트워크에서 레비 워크 이동성을 가지는 릴레이 노드들의 평균 전송 에너지
Fig. 5. Average transmission power of relay nodes in 5×5 network with Lévy walk mobility

용하고 다른 노드는 사용하지 않는 네트워크 토폴로지로 수렴함을 알 수 있다.

그림 4와 5를 통해 랜덤 이동 모델을 랜덤 웨이포인트 모델과 레비 워크 모델로 바꾸어도 동일함을 확인할 수 있다. 그림 5에서 5×5 네트워크에 랜덤 웨이포인트 모델을 적용하였을 때에는 5번과 11번 릴레이 노드가 사용됨을 확인할 수 있고, 그림 5에서 5×5 네트워크에 레비 워크 모델을 적용하였을 때 0번, 3번, 4번, 5번, 11번, 19번 릴레이 노드가 선택적으로 사용되는 것을 확인할 수 있다. 즉, 5×5 네트워크에서는 이동 모델과 무관하게 2개 홉으로 소스 노드와 목적 노드가 연결된다.

그림 6을 통해 네트워크 크기가 확장되어도 네트워크 처리율 향상에 중요한 역할을 하는 노드만 선택적으로 전송 에너지를 사용하는 것을 확인할 수 있다. 그림 6은 네트워크 크기가 확장된 랜덤 워크 이동성을 갖는 10×10 네트워크의 실험 결과를 나타낸다. 이때 10번 노드와 63번 노드는 큰 전송 에너지를 유지하고, 다른 노드는 전송 에너지를 줄이는 방향으로 학습이 이루어짐을 확인할 수 있다. 이는 3개 홉으로 소스 노드-목적 노드 연결이 이루어졌음을 의미한다.

릴레이 노드별 평균 전송 에너지 변화 그래프를 통해 랜덤 이동 모델의 종류 또는 네트워크의 크기와는 무관하게 네트워크 처리율 향상에 핵심적인 역할을 하는 릴레이 노드가 선택적으로 켜지는 것을 알 수 있다.

표 1은 랜덤 워크 이동성을 갖는 3×3 네트워크에서 전체 면적 대비 노드 밀도가 1인 경우와 노드 밀도가 $\frac{5}{9}$인 경우의 네트워크 처리율과 노드 당 송신 에너지 소모량을 비교한다. 표 1을 통해 네트워크 면적 내 노드 밀도가 1에서 절반 수준인 $\frac{5}{9}$로 감소하더라도 네트워크 처리율은 동일한 값인 4.8로 수렴함을 알 수 있다. 그러나 노드 당 송신 에너지 소모량은 노드 밀도가 1인 경우에 0.3으로 수렴하며, 노드 밀도가 $\frac{5}{9}$

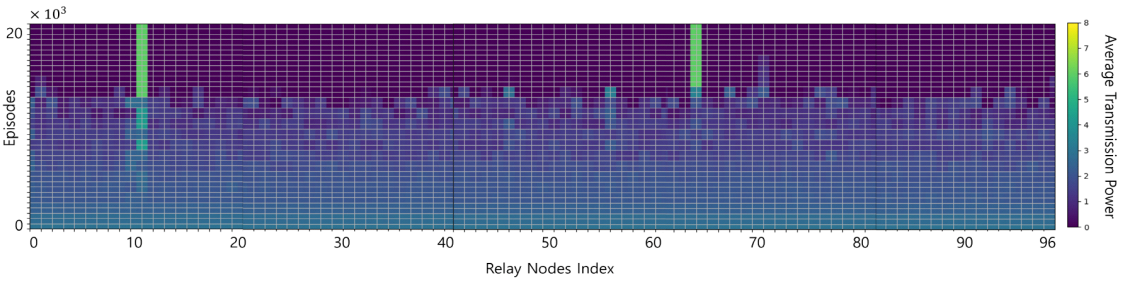


그림 6. 10×10 네트워크에서 랜덤 워크 이동성을 가지는 릴레이 노드들의 평균 전송 에너지
Fig. 6. Average transmission power of relay nodes in 10×10 network with random walk mobility

표 1. 랜덤 워크 이동성을 갖는 3×3 네트워크에서 전체 면적 대비 노드 밀도가 1인 경우와 노드 밀도가 인 경우의 네트워크 처리율과 노드 당 송신 에너지 소모량
Table 1. Throughput and average transmission power of relay nodes in 3×3 dense network and sparse network with random walk mobility

	Throughput (byte/sec)	Average transmission power per node (J)
Dense Network (density=1)	4.8	0.3
Sparse Network (density=5/9)	4.8	0.7

일 때에는 0.7의 값으로 수렴한다. 이는 노드 밀도가 상대적으로 높아 전송 면적 안에 다수의 이웃 노드가 존재할 경우, 낮은 송신 에너지를 사용하여 소스-목적 노드 간 네트워크 연결이 가능함을 의미한다. 즉, 노드 밀도가 1보다 낮아지면 노드 간 간격이 커져 소스-목적 노드 간 네트워크 연결을 유지하기 위해 노드 밀도가 1일 때보다 많은 송신 에너지를 사용해야 한다.

5.2 제안한 알고리즘과 중앙 제어 방식의 성능 비교

본 논문에서 제안한 MANET 구축 알고리즘의 성능을 중앙 제어 방식과 비교하기 위해 실험을 진행하였다. 중앙 제어 방식은 전체 네트워크를 탐색 범위에 두고 최적의 네트워크 토폴로지를 찾는다. 중앙 제어 방식에서 각 릴레이 노드는 자기 자신을 포함한 다른 모든 릴레이 노드와 연결될 수 있다.

실험한 네트워크 규모는 3×3이며 소스 노드와 목적 노드는 각각 1개로 설정한다. 에이전트의 수 $|N(G_\tau)|$는 7개이며 랜덤 이동성 모델은 랜덤 워크 모델을 적용한다. 중앙 제어 방식은 3×3 네트워크에서 7개의 릴레이 노드가 선택 가능한 7^7가지의 모든 경우의 수를 고려한 후, 최대의 네트워크 처리율을 유지할 수 있는 행동 조합을 선택한다. 따라서 중앙 제어 방식의 네트워크 처리율은 제안한 알고리즘의 학습 결과 얻을 수 있는 네트워크 처리율의 상한선이 된다.

표 2는 제안한 알고리즘과 중앙 제어 방식의 네트워크 처리율과 평균 에너지 소모량을 비교한 결과이다. 중앙 제어 방식은 최대의 네트워크 처리율인 5의 값을 유지하며 릴레이 노드의 평균 에너지 소모량은 0.4의 값을 갖는다. 제안한 알고리즘에 의해 학습된 네트워크는 중앙 제어 방식보다 낮은 4.8의 네트워크 처리율로 수렴한다. 그러나 릴레이 노드 당 평균 에너지 소모량은 0.3으로 중앙 제어 방식보다 낮다. 이는

표 2. 랜덤 워크 이동성을 갖는 3×3 네트워크에서 중앙 제어 방식과 제안한 알고리즘의 네트워크 처리율과 노드 당 송신 에너지 소모량
Table 2. Throughput and average transmission power of relay nodes of the centralized algorithm and proposed algorithm in 3×3 network with random walk mobility

	Throughput (byte/sec)	Average transmission power per node (J)
Centralized Algorithm	5	0.4
Proposed Algorithm	4.8	0.3

두 학습 알고리즘에 의해 최종적으로 구축된 애드혹 네트워크의 형태가 다르기 때문이다. 제안한 강화학습 기반 알고리즘에 의해 학습된 네트워크는 네트워크 처리율 유지에 필수적인 1개의 릴레이 노드를 제외한 6개의 릴레이 노드가 모두 꺼진다. 그러나 중앙 제어 방식에 의해 구축된 네트워크는 7개의 릴레이 노드 중 2개의 릴레이 노드가 전송 에너지를 사용한다.

본 논문에서 제안하는 방안은 큰 규모의 네트워크에도 적용할 수 있다. 네트워크의 규모가 커지면 제안하는 방안이 중앙 제어 방식보다 계산 복잡도 면에서 효율적이다.

네트워크 구축 시의 계산 복잡도는 탐색 공간에 비례하고 탐색 공간은 가능한 노드 연결의 전체 조합 수로 정의한다. 전체 노드 수가 k이므로, 중앙 제어 방식을 적용하면 노드마다 k개의 연결이 가능하다. 즉, 연결 가능한 최대의 조합은 k^k개가 되고 계산 복잡도는 $O(k^k)$이다. 제안하는 알고리즘은 각 노드가 최대

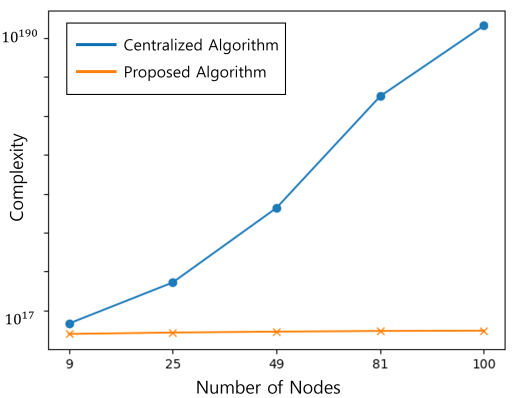


그림 7. 중앙 제어 방식과 제안한 알고리즘의 계산 복잡도 비교
Fig. 7. Required computational complexity in the centralized and proposed algorithm

$k-1$개의 이웃 노드를 가질 수 있으며 노드별 행동 공간의 크기는 $|A|$이므로 네트워크에서 가능한 최대 조합의 수는 $k(k-1)\times|A|$이다. 본 논문에서는 $|A|\ll k$인 다수의 노드로 구성되는 네트워크를 고려하므로 계산 복잡도는 $O(k^2)$이 된다. 따라서 노드의 수가 증가할수록 제안하는 알고리즘이 중앙 제어 방식과 비교하였을 때 상대적으로 낮은 계산 복잡도를 갖는다.

그림 7을 통해 네트워크 규모가 커졌을 때, 제안하는 방안의 계산 복잡도 면에서의 이득을 확인할 수 있다. 그림 7은 전체 노드 수의 증가에 따른 계산 복잡도를 나타낸다. 제안하는 알고리즘과 달리, 중앙 제어 방식은 노드 수가 증가함에 따라 지수적으로 계산 복잡도가 증가한다.

Ⅵ. 결 론

본 논문에서는 무선 모바일 기기들이 Q-learning을 기반으로 분산적으로 전송 에너지를 결정하여, 중앙 제어 없이 에너지 효율적인 MANET을 구성하는 방안을 제안하였다. 제안한 알고리즘에 의해 학습된 릴레이 노드는 랜덤 워크, 랜덤 웨이포인트, 레비 워크의 랜덤 이동 모델에 따라 이동함에도 안정적인 네트워크를 구축할 수 있다. 시뮬레이션 결과를 통해 학습된 릴레이 노드들이 구성하는 네트워크는 네트워크 처리율은 높게 유지하고 전체 에너지 소비량은 감소시킴을 확인하였다. 또한, 네트워크의 크기와 릴레이 노드의 수가 증가하여도 네트워크 처리율을 계속해서 안정적으로 유지함을 보였다. 이를 통해 본 논문에서는 네트워크의 크기와 노드의 랜덤한 이동성에도 불구하고 MANET을 분산적으로 구축할 수 있음을 확인하였다.

제안한 알고리즘에 의해 학습된 애드혹 형성 정책은 랜덤 이동 모델의 종류 또는 네트워크의 크기와는 무관하게 네트워크 처리율 향상에 핵심적인 릴레이 노드가 선택적으로 사용된다. 단, 개별 노드의 에너지 소모량이 불균등하게 분포할 수 있으므로 전력원이 한정된 경우 특정 노드의 수명이 짧아질 수 있는 문제가 있다. 이를 해결하고자 릴레이 노드의 최대 전력량과 에너지 레벨을 고려한 학습 기법에 관한 연구가 진행되어야 할 것이다.

References

[1] E. Baraneetharan, “Role of machine learning algorithms intrusion detection in WSNs: A survey,” *J. Info. Technol. and Digital World,* vol. 2, no. 3, pp. 161-173, 2020.

[2] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, “Protocols for self-organization of a wireless sensor network,” *IEEE Pers. Commun.,* vol. 7, no. 5, pp. 16-27, Oct. 2000.

[3] M. Deruyck, J. Wyckmans, L. Martens, and W. Joseph, “Emergency ad-hoc networks by using drone mounted base stations for a disaster scenario,” *2016 IEEE 12th Int. Conf. WiMob*, pp. 1-7, 2016.

[4] T. Plesse, C. Adjih, P. Minet, A. Laouiti, A. Plakoo, M. Badel, P. Muhlethaler, P. Jacquet, and J. Lecomte, “OLSR performance measurement in a military mobile ad hoc network,” *Ad Hoc Netw.,* vol. 3, no. 5, pp. 575-588, 2005.

[5] İ. Bekmezci, O. K. Sahingoz, and Ş. Temel, “Flying ad-hoc networks (FANETs): a survey,” *Ad Hoc Netw.*, vol. 11, no. 3, pp. 1254-1270, 2013.

[6] M. Kwon, J. Lee, and H. Park, “Intelligent IoT connectivity: Deep reinforcement learning approach,” *IEEE Sensors J.,* vol. 20, no. 5, pp. 2782-2791, Mar. 2020.

[7] N. Kim, M. Kwon, and H. Park, “Q-learning based network formation strategy for wireless nodes with random-walk mobility,” *2021 13st JCCI,* pp. 257-258, Busan, Korea, Apr. 2021.

[8] R. S. Sutton and A. G. Barto, “*Introduction*,” *Reinforcement Learning: An Introduction,* 2nd Ed., pp. 1-22, USA: MIT Press, 2018.

[9] C. Watkins and P. Dayan, “Q-leaninrg,” *Mach. Learning*, vol. 8, pp. 279-292, 1992.

[10] A. C. Viana, M. D. de Amorim, S. Fdida, and J. F. de Rezende, “Self-organization in spontaneous networks : The approach of DHT-based routing protocol,” *Ad Hoc Netw.,* vol. 3, no. 5, pp. 589-606, 2005.

[11] J. Wang, Y. Miao, P. Zhou, M. S. Hossain, and Sk Md M. Rahman, “Software defined

network routing in wireless multihop network," *J. Netw. and Comput. Appl.,* vol. 85, no. 1, pp. 76-83, May 2017.

[12] K. Genda, "Topology control adopting optimal topology over update interval in mobile adhoc networks," *IEICE Commun. Express,* vol. 9, no. 3, pp. 83-88, 2020.

[13] K. Shafiee, and V. C. M. Leung, "Connectivity-aware minimum-delay geographic routing with vehicle tracking in VANETs," *Ad Hoc Netw.,* vol. 9, no. 2, pp. 131-141, Mar. 2011.

[14] A. Venkateswaran, V. Sarangan, T. F. La Porta, and R. Acharya, "A mobility-prediction-based relay deployment framework for conserving power in MANETs," *IEEE Trans. Mob. Comput.,* vol. 8, no. 6, pp. 750-765, Jun. 2009.

[15] G. Carofiglio, C. Chiasserini, M. Garetto, and E. Leonardi, "Route stability in MANETs under the random direction mobility model," *IEEE Trans. Mob. Comput.,* vol. 8, no. 9, pp. 1167-1179, Sep. 2009.

[16] S. Tabatabaei, M. Teshnehlab, and S. J. Mirabedini, "A new routing protocol to increase throughput in mobile ad hoc networks," *Wirel. Pers. Commun.,* vol. 83, pp. 1765-1778, Mar. 2015.

[17] T. T. T. Le and S. Moh, "Reinforcement-learning-based topology control for wireless sensor networks," *Advanced Sci. and Technol. Lett.,* vol. 142, pp. 4-27, 2016.

[18] M. Maleki, V. Hakami, and M. Dehghan, "A model-based reinforcement learning algorithm for routing in energy harvesting mobile ad-hoc networks," *Wirel. Pers. Commun.,* vol. 95, pp. 3119-3139, Feb. 2017.

[19] X. He, H. Jiang, Y. Song, C. He, and H. Xiao, "Routing selection with reinforcement learning for energy harvesting multi-hop cognitive radio network," *IEEE Access,* vol. 7, pp. 54435-54448, 2019.

[20] D. Yu and Hui Li, "On the definition of ad hoc network connectivity," *Int. Conf. Commun. Technol. Proc.,* vol. 2, pp. 990-994, 2003.

[21] F. Osterlind and A. Dunkels, "Approaching the maximum 802.15.4 multi-hop throughput," in *Proc. 5th Assoc. Computing Machinery Workshops Embedded Netw. Sensor (HotEmNets 2008)*, Jun. 2008.

[22] F. Bai and A. Helmy, "A survey of mobility models in wireless adhoc networks," *Wirel. Ad Hoc and Sensor Netw.*, pp. 1-29, Kluwer Academic Publishers, 2004.

[23] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, "Lévy dynamics of enhanced diffusion: application to turbulence," *Phys. Rev. Lett.*, vol. 58, pp. 1100-1103, Mar. 1987.

[24] P. Lévy, *Calcul des probabilités*, Paris: Gauthier-Villars, 1926.

[25] H. T. Friis, "A note on a simple transmission formula," in *Proc. IRE,* vol. 34, no. 5, pp. 254-256, 1946.

[26] R. S. Sutton and A. G. Barto, "Planning and leaning with tabular methods," *Reinforcement Learning: An Introduction,* 2nd Ed., pp. 182-216, USA: MIT Press, 2018.

김 나 영 (Nayoung Kim)

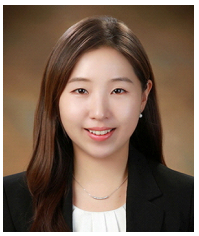
2021년 2월 : 이화여자대학교 전자공학과 학사

2021년 3월~현재 : 이화여자대학교 전자전기공학과 석사 과정

<관심분야> 멀티에이전트 네트워크 시스템, 인공지능, 강화학습

[ORCID:0000-0001-7235-3209]

권 민 혜 (Minhae Kwon)



2011년 8월 : 이화여자대학교 전자정보통신공학과 학사
2013년 8월 : 이화여자대학교 전자공학과 석사
2017년 8월 : 이화여자대학교 전자전기공학과 박사
2017년 9월~2018년 8월 : 이화여자대학교 전자전기공학과 박사 후 연구원
2018년 9월~2020년 2월 : 미국 Rice University, Electrical and Computer Engineering, Postdoctoral Researcher
2018년 9월~2020년 2월 : 미국 Baylor College of Medicine, Center for Neuroscience and Artificial Intelligence, Postdoctoral Researcher
2020년 3월~현재 : 숭실대학교 전자정보공학부 IT융합전공 조교수
<관심분야> 심층강화학습, 계산신경과학, 모바일네트워크
[ORCID:0000-0002-8807-3719]

박 형 곤 (Hyunggon Park)

2004년 2월 : 포항공과대학교 전자전기공학과 졸업
2006년 3월 : University of California, Los Angeles (UCLA) M.S.
2008년 12월 : University of California, Los Angeles (UCLA) Ph.D.
2010년~현재 : 이화여자대학교 전자전기공학과 교수
<관심분야> 멀티에이전트 시스템 최적화, 머신러닝, 인공지능, 게임이론
[ORCID:0000-0002-5079-1504]