

신체 영역 네트워크에서 의료 정보 수집 시스템을 위한 저지연 및 고신뢰성을 갖는 Q-Learning 기반의 라우팅 프로토콜

최 유진*, 권정훈*, 나호정*, 유상조*

Q-Learning Based Dynamic Routing Protocol with Low Latency and High Reliability for Medical Data Collection System Using Body Area Networks

Yoo-Jin Choi*, Jeong-Hoon Kwon*, Ho-Jeong Na*, Sang-Jo Yoo*

요 약

신체 영역 네트워크(BAN : Body Area Network)는 신체데이터를 지속적으로 수집하여 사용자의 건강에 대한 실시간 모니터링이 가능하다. 하지만 센서는 배터리 등 자원의 한계가 있기 때문에 에너지 효율성이 요구되며, 실시간 데이터 수집을 위해 지연 시간을 최소화 해야 한다. 또한, 사용자의 건강관리 및 문제 발생 시 빠른 대처를 위해 수집한 신체데이터를 건물 내부의 데이터 스테이션에서 관리하는 것도 중요하다. 따라서 본 논문에서는 신체 영역 네트워크에서 수집한 신체 데이터를 건물 내부의 데이터 스테이션으로 전달하기 위한 동적 라우팅 프로토콜을 제안한다. 우선 신체데이터 수집을 위해 신체 영역 네트워크에서 지연시간, 네트워크 수명, 에너지 효율성을 고려한 최적의 경로를 탐색한다. 그리고 수집된 데이터를 의료 관련 건물 내부의 데이터 스테이션으로 전송하기 위해 건물에 별도의 무선네트워크를 구성하고, 에너지 효율성을 고려한 최적의 경로를 탐색한다. 동적인 환경에서 적응적으로 동작하기 위해 Q-Learning을 기반으로 각 목적에 맞는 라우팅프로토콜을 설계하였다. 모의실험을 통해 제안된 방법이 최적 경로를 도출하는 것을 확인하였다.

Key Words : Wireless Sensor Network, Reinforcement Learning, Q-Learning, Body Area Network, Routing Protocol

ABSTRACT

A Body Area Network(BAN) continuously collects body data to enable real-time monitoring of the user's health. Because sensors are limited in resource such as battery, there is a demand for energy efficiency, also the latency should be minimized for real-time data collection. In addition it is important to manage the collected body data in the data station inside the building for the user's health care and quick response in case of problems. In this paper, we propose dynamic data routing protocol for a medical information collection system in

※ 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2017R1A2B4003512)

♦ First Author : Inha University Department of Information and Communication Engineering, gch04750@naver.com, 학생회원

◦ Corresponding Author : Inha University Department of Information and Communication Engineering, sjyoo@inha.ac.kr, 종신회원

* Inha University Department of Information and Communication Engineering, topkjh11@hanmail.net; world9304@naver.com, 학생회원

논문번호 : KICS2019-08-169, Received August 23, 2019; Revised September 19, 2019; Accepted September 19, 2019

a healthcare facility. First of all, the BAN considers the requirements for latency, network lifetime and energy efficiency. In order to transmit the collected data to the base station inside the medical-related building, a separate wireless network is constructed in the building and the optimal path considering energy efficiency is searched. To consider the dynamic environment, routing protocols for each purpose were designed based on Q-Learning. The simulation results show that the proposed method derives the optimal path.

I. 서 론

기술이 발전함에 따라 생활이 여유로워지고 건강에 대한 관심이 높아지고 있다. 병을 초기에 알아내 치료하는 것뿐만 아니라 발병 이전에 예측하고 예방하는 것 까지 발전할 것으로 기대하고 있다. 이에 따라 신체 정보를 확인할 수 있는 웨어러블 밴드의 출시 등 다양한 헬스케어 서비스가 생성되어 이용자는 자신의 건강정보를 쉽게 접할 수 있다. 또한, 다양한 신체 정보를 획득할 수 있는 신체 삽입형 센서에 대한 연구들도 활발히 이루어진다^[1]. 신체 삽입형 센서들은 무선 통신을 통해 싱크노드로 데이터를 전송한다. 여기서, 싱크노드는 주로 스마트워치, 스마트폰, 웨어러블 밴드와 같이 에너지자원이 크고 송신범위가 큰 제품이다. 이러한 센서 노드들과 싱크노드로 이루어진 무선센서네트워크를 신체 영역 네트워크(BAN)라고 한다. 신체 영역 네트워크의 발전으로 환자나 고령자들을 위한 의료 관련 서비스가 더욱 확대될 것이라 예상된다. 따라서 이를 위한 연구의 필요성이 부각되고 있다^[2].

신체 영역 네트워크를 구성하는 센서 노드들은 무선네트워크의 통신범위가 좁고 센서의 크기가 작으므로 배터리의 용량이 제한적이기 때문에 자원을 효율적으로 사용해야 한다. 또한 신체 의료 정보를 다루는 만큼 데이터의 신뢰성과 저지연성이 요구된다. 이 때문에 신체 영역 네트워크에 기존의 라우팅프로토콜을 그대로 적용시키면 크게 두 가지의 문제점이 발생한다. 첫 번째, 노드의 잔여에너지를 고려하지 않기 때문에 특정 노드에 에너지의 소비가 집중된다. 이로 인해 네트워크의 수명이 줄어든다. 두 번째, 신체의 움직임에 따라 변하는 네트워크 환경이 반영되지 않기 때문에 연결 상태가 좋지 않은 경로를 통해 데이터가 전송된다. 이는 데이터의 재전송을 발생시키기 때문에 에너지 소모율이 커지며 전송 지연이 발생한다. 따라서 본 논문에서는 신체 영역 네트워크 내부에서 센서의 에너지 잔량과 함께 동적인 신체의 움직임을 고려한 라우팅 프로토콜과 신체 영역에서 수집된 데이터를 데이터스테이션으로 전송하는 라우팅프로토콜 제안하여 의료시설내부 의료 정보 수집 시스템을 위한

라우팅 프로토콜을 완성하고자 한다.

신체 영역 네트워크의 라우팅프로토콜의 효율성을 높이기 위해 신체의 움직임에 고려한다. 또한 최소 홉수와 에너지 효율성까지 고려하며, 동적인 환경에서 원하는 목적에 맞게 동작하는 강화학습의 일종인 Q-Learning을 기반으로 라우팅프로토콜을 모델링한다. Q-Learning의 상태를 조건에 따라 여러 개로 구분지어 각 상태에 따라 다른 보상함수가 적용되도록 하여 신체 영역 네트워크에서 최적경로를 도출한다.

또한 병원과 같은 의료시설 내부에서 이용자의 신체 의료 정보가 실시간으로 전송된다면 이용자의 위급상황 시 의료시설에서 빠른 대처를 할 수 있을 것이며 수집된 신체 의료 정보를 기반으로 질병의 진행 정도를 판단하거나 예측할 수 있을 것이다. 이를 위해 신체 영역 네트워크에서 획득한 신체데이터의 처리가 중요하다. 효율적인 데이터의 처리를 위해 본 논문에서는 건물 내부에 별개의 무선 네트워크를 구성하여 사용자의 신체 영역 네트워크의 싱크노드부터 건물의 데이터스테이션까지의 최적경로를 설정한다. 결론적으로, 신체 영역 네트워크에서 신체의 움직임에 따라 동적으로 반응하는 최적의 경로를 도출하는 라우팅 프로토콜과 데이터를 효율적으로 처리하기 위해 건물 내부에서 이루어지는 라우팅프로토콜을 제안한다.

관련 연구로써 무선센서 네트워크에서 경로설정을 위해 머신러닝을 사용한 연구들이 있다. Qin Yang은 무선 Ad-hoc 네트워크에서 강화학습을 적용하여 패킷전송과 에너지 소비의 신뢰성이 고려되는 라우팅경로 선택 알고리즘을 제안하였다^[3]. Jinhyuk Yim은 모바일 Ad-hoc 네트워크에서 이동성이 높은 모바일노드 간 라우팅의 어려운 점을 해결하고자 Q-Learning을 적용한 위치기반 라우팅 프로토콜을 제안하였다^[4]. Jeong-Soo Han은 강화학습에 정책 기울기 값을 적용한 라우팅 기법을 제안하였다. 이 논문은 기대 보상값의 기울기 값을 정책에 반영하여 빠른 네트워크 환경을 학습하는 것을 보였다^[5]. Alam은 인체영역에 Q-Learning을 적용한 라우팅 프로토콜을 사용할 수 있음을 보였다^[6]. 이 논문에서 제안한 모델은 에너지 효율과 최소의 홉 수를 고려하였지만 신체의 동적인

움직임에 따른 네트워크의 환경 변화가 고려되지 않았다.

본 논문의 구성은 다음과 같다. 2장에서는 제안된 의료수집시스템의 모델을 신체 영역 네트워크와 건물 내부네트워크로 나누고 각 네트워크의 시스템모델과 고려사항을 설명한다. 3장에서는 Q-Learning을 이용하여 제안된 시스템에서 경로를 설정하는 방법에 대해 설명한다. 4장은 실험결과로, 앞서 소개한 라우팅 프로토콜을 별도로 가정한 실험 환경 내에서 시뮬레이션을 수행한 결과에 대해서 설명한다. 마지막 5장에서는 본 논문에 대한 결론을 설명한다.

II. 시스템 모델 및 고려사항

본 논문에서는 신체 의료 정보 수집 시스템에서 저지연성과 신뢰성 높은 데이터 수집을 위해 신체 영역 네트워크 내외부에서 데이터 전송을 위한 라우팅프로토콜을 제안한다. 이때, 신체 의료 정보 수집 시스템을 구성하는 네트워크의 모든 센서노드가 헬로 패킷을 주기적으로 브로드캐스팅하여 자신의 이웃노드의 정보를 알 수 있는 것으로 가정한다.

신체 의료 정보 수집 시스템은 병원같은 의료시설 내부에서 이용자들의 신체데이터를 수집하기 위한 시스템이다. 각 사용자의 신체에 부착되거나 삽입된 센서를 통해 데이터가 측정되고, 이 데이터는 신체 영역에 존재하는 싱크노드로 전달된다. 이후 데이터는 싱크노드로부터 해당 건물의 데이터스테이션으로 전달되어 최종적으로 수집된다.

따라서, 그림 1과 같이 신체 의료 정보 수집 시스템은 2가지의 네트워크로 구성이 된다. 첫 번째는 그림 1(a)와 같이 센서노드에서 획득한 신체데이터를 주기

적으로 싱크노드로 전송하는 신체 영역 네트워크이다. 두 번째는 그림1(b)처럼 사용자들의 신체 영역 네트워크에서 수집된 정보를 의료 관련 시설의 데이터스테이션으로 전달하기 위해 구성되는 건물 내부 네트워크이다.

2.1 신체 영역 네트워크

본 논문에서는 신체 영역 네트워크의 싱크 노드가 왼쪽 손목에 존재하며, 센서 노드는 머리를 제외한 신체 영역에 총 12개가 존재한다고 가정한다.

신체 영역 네트워크는 주로 400MHz ~ 60GHz 주파수 대역을 사용하는 무선센서 네트워크로 의료 정보를 다룬다⁷⁾. 또한, 의료 정보를 다루는 특성상 저지연성이 요구되며, 제한된 배터리활용을 위해 에너지 효율성과 네트워크 수명에 대한 요구를 만족시키며 기존의 모델에서 신체의 움직임에 고려하지 않아서 생기는 비효율성을 해결하기 위한 새로운 라우팅 프로토콜을 제안한다. 신체의 움직임에 따라 변하는 네트워크 환경을 반영하기 위해 강화학습 기법의 일종인 Q-Learning을 사용하며 최적의 라우팅경로를 도출하기 위해 최단경로, 네트워크의 수명, 신체의 활동을 고려한다.

2.1.1 최단경로

신체 영역 네트워크는 신체의 움직임에 따라 센서노드들의 위치가 동적으로 변화한다. 따라서 데이터를 싱크노드까지 전송하기 위한 경로가 수시로 바뀌게 된다.

신체의 자세가 그림2(a)에서 그림2(b)로 변할 때, 센서노드의 배치가 변화하기 때문에 싱크노드로 데이터를 전송하는 최단 경로가 바뀐다. 이처럼 신체의 움직임에 따라 변하는 네트워크 환경을 반영하기 위해

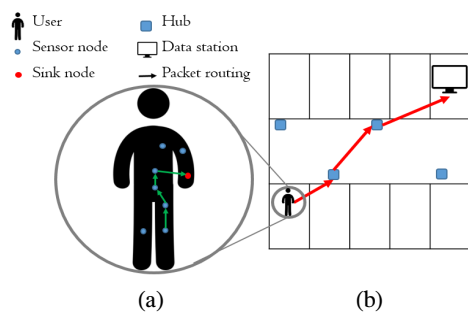


그림 1. 신체 의료 정보 수집 시스템, (a) 신체 영역 네트워크에서 데이터 수집, (b) 건물내부네트워크에서 데이터 수집
Fig. 1. Body medical data collection system, (a) Data collection in BAN, (b) Data collection in in-building network.

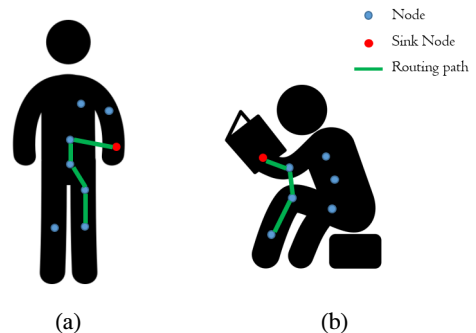


그림 2. 신체의 움직임에 따른 데이터 전송 경로 변경, (a) 서있는 상태, (b) 앉아있는 상태
Fig. 2. Change of data transmission path according to body movement, (a) Standing state, (b) Sitting state

강화학습 기법의 일종인 Q-Learning을 사용한다. 각 노드는 헬로패킷의 주기적인 송수신을 통해 이웃노드를 알 수 있으므로 주어진 환경 네트워크에 대해 이웃노드로 패킷의 송수신이라는 행동을 무수히 많이 취하며 싱크노드로 향하는 최단경로를 학습할 수 있다.

2.1.2 네트워크의 수명

신체 영역 네트워크에서 에너지의 사용이 특정한 센서 노드에 집중되지 않고 네트워크의 센서 노드들이 균등한 에너지 소비 분포를 가지도록 하여 네트워크의 수명을 연장한다.

그림 3과 같이 에너지의 분포를 고려하여 최적 라우팅 경로를 도출한다. 현재노드가 node p 일 때 이웃노드 node u 와 node v 중 어느 노드로 패킷을 송신할지 결정할 때, 수식(1)과 같이 이웃노드들의 상대적인 잔여 에너지를 나타낼 수 있다.

$$E_p = \frac{E_r(x_j)}{E_{mean}} \quad (1)$$

여기서, $x_j \in X$ 이며 X 는 현재노드의 이웃노드들의 집합이다. $E_r(x_i)$ 는 센서노드 x_i 의 잔여 에너지이며, E_{mean} 는 이웃노드의 평균 잔여 에너지로 수식(2)와 같이 계산된다.

$$E_{mean} = \frac{\sum_{x \in X} E_r(x)}{n(X)} \quad (2)$$

여기서, $n(X)$ 는 집합 X 의 원소의 개수이다. E_p 는 현재노드의 이웃노드들의 잔여에너지 평균에 대해 이웃노드 x_j 의 상대적인 잔여에너지의 값으로 각 이웃노드들의 E_p 를 비교하여 이웃노드들 간의 상대적인

E_r : remain energy

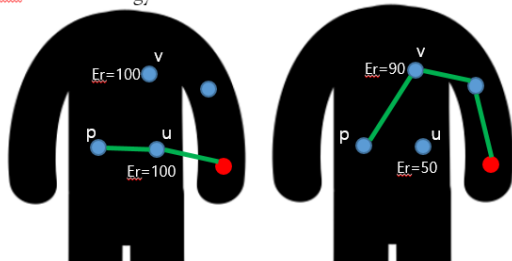


그림 3. 에너지 분포에 따른 경로 변화
Fig. 3. Change of routing path according to energy distribution.

잔여에너지를 고려한다. 따라서, 절대적인 잔여에너지 양을 비교하는 것보다 잔여에너지의 차이를 부각시킨다. 그림3(a)와 같이 node p 에서 싱크노드까지의 경로는 node u 로 데이터를 전송하는 것이 최소 홉 수만을 고려하였을 때 최적인 경로이다. 하지만 그림3(b)에서 node p 의 이웃노드들의 상대적 잔여에너지를 비교하면 node v 가 node u 에 비해 잔여에너지가 많으므로 node v 를 거치는 경로를 택한다. 즉, 수식(1)을 이용하여 적절한 잔여에너지의 분배를 추구하는 최적경로를 도출함으로써 네트워크의 수명을 증진시킨다.

2.1.3 신체의 활동을 고려한 패킷라우팅 경로

신체 영역 네트워크에서 신체의 움직임에 따라 센서노드들의 위치가 항상 변하기 때문에 데이터를 전송하기 위한 최적 경로의 조건이 수시로 바뀐다. 특히, 신체의 움직임이 빠른 상황에는 네트워크의 상황, 연결여부가 빠르게 변하기 때문에 기존의 방법으로는 최적경로를 도출하지 못한다. 달라진 네트워크 상황이 학습되기 이전에 네트워크 상황이 다시 변동되기 때문이다. 이를 위해 본 논문에서는 센서노드 간 링크의 연결성이라는 요소를 추가하여 신체의 활동까지 고려된 최적경로를 도출하고자 한다.

그림4(a)는 센서노드 간에 링크가 연결된 경우 링크가 계속 안정적인 상태로 유지된다. 하지만 그림 4(b)의 경우처럼 신체의 움직임이 활발할 때는 센서노드 간의 링크가 이어짐과 끊어짐이 반복된다.

이렇게 링크의 연결이 변동이 심한 경우 링크의 연결성이 낮다고 정의한다. 연결성이 낮은 경로로 패킷을 전송하게 되면 패킷이 손실될 가능성이 높아지기 때문에 에너지낭비와 지연이 생기게 된다. 따라서 연

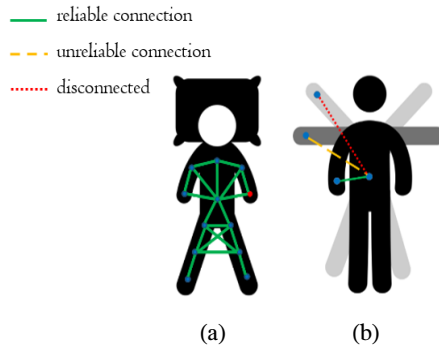


그림 4. 신체의 활동성, (a) 비활동적인 신체의 상태, (b) 활동적인 신체의 상태
Fig. 4. Activity of the body, (a) Inactive body, (b) Active body

결성을 계산하여 경로 선택 시 연결성을 고려한다.

연결성은 각 헬로패킷의 응답에 대해서 계산되며 링크의 연결성은 과거 N 개의 헬로패킷의 연결성의 평균으로 결정된다. 연결된 링크의 안정성은 수신세기에 비례하므로 수신세기를 이용하여 연결성을 계산한다. 경로손실은 자유공간에서 발생하며, 페이딩효과가 없다고 가정한다. 그래서 수신 세기는 수식 (3)과 같이 계산된다.

$$P_r = P_t(G_t G_r) \left(\frac{\lambda}{4\pi d} \right)^2 \quad (3)$$

여기서, P_t 는 송신 세기이며, G_t, G_r 는 각각 송신, 수신 안테나 이득이다. 또한, λ 는 동작 주파수, d 는 송신부와 수신부 사이의 거리를 나타낸다. 하지만 이 값은 크기가 작아서 비교에 적합하지 않으므로 [0,100] 사이의 값이 되도록 정규화 시킨 값을 연결성으로 정의한다. 정규화를 위해 수식(4)의 sigmoid 함수를 이용한다.

$$s(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

입력을 0과 1사이의 값으로 만들어주는 sigmoid함수는 입력을 0과 1사이의 값으로 만들어주는 함수이며, 이를 기반으로 연결성 CON 을 수식 (5)와 같이 계산한다.

$$CON = \frac{100}{1 + e^{-C_d(P_r - C_m)}} \quad (5)$$

여기서, C_d 와 C_m 은 상수로 모의실험에서 각 10^5 과 4×10^{-5} 를 이용하였다.

수식(5)를 이용하여 각 헬로 패킷마다 수신세기를 계산한다. 계산된 연결성 CON 을 이용하여 x_i 노드와 x_j 노드 사이의 링크의 연결성 $L_{con}(x_i, x_j)$ 를 수식 (6)

과 같이 계산한다.

$$L_{con}(x_i, x_j) = \frac{\sum^N CON(x_i, x_j)}{N} \quad (6)$$

여기서 N 은 최근 송수신한 패킷의 개수를 나타내는 상수로 만약 N 이 4라면 최근 수신한 4개의 헬로패킷으로 연결성을 계산한다. 따라서 링크의 연결성 $L_{con}(x_i, x_j)$ 은 과거 N 개의 헬로패킷의 연결성의 평균으로 구한다. 이와 같이 연결성을 계산하여 연결성이 좋지 않은 링크로의 패킷라우팅은 피하도록 설계한다. 연결성을 계산하는 알고리즘은 그림 5와 같다.

2.2 건물내부네트워크

본 논문에서는 한 사용자의 신체 영역 네트워크에서의 데이터 라우팅뿐만 아니라, 병원과 같은 의료시설 내부에서 사용자의 싱크노드로부터 데이터를 수집하기 위해 실내 라우팅프로토콜을 제안한다.

그림 6과 같이 건물 내에 데이터 스테이션이 존재하며, 건물 내벽에 데이터 라우팅을 위한 허브(Hub)가 총 18개 있다고 가정한다. 또한, 네트워크의 유연한 망 구성을 위해 건물내부네트워크는 Ad-hoc과 같은 무선네트워크라고 가정한다. 이 때문에 네트워크를 구성하는 허브들은 제한된 에너지를 가지므로 잔여에너지를 고려한 라우팅프로토콜을 설계한다.

2.2.1 네트워크의 수명

건물 내 허브들은 앞서 말한바와 같이 제한된 에너지를 가지므로 건물 내부 네트워크의 수명 연장을 위해 잔여 에너지를 다음과 같이 고려한다.

$$E_p = \frac{E_r(x_j)}{E_s} \quad (7)$$

$$E_s = \sum_{x \in X} E_r(x) \quad (8)$$

여기서, E_s 는 이웃 노드들의 잔여에너지의 합을 나타내며, E_p 는 E_s 와 특정 이웃노드(x_j)의 잔여 에너지로 계산되며, 상대적인 잔여 에너지양을 나타낸다.

건물 내 부착된 허브들의 경우 신체 영역 네트워크의 센서노드에 비해 크기가 크기 때문에 에너지에 대한 제약이 비교적 적다. 따라서 수식(1)과 같이 잔여에너지의 평균을 사용하는 대신 그 합 E_s 을 사용하여 잔여 에너지양의 차에 대해 예민한 정도를 감소시켜

```

1: every Tseconds
2:   for all xj
3:     send hello packet
4:     if received respond packet
5:       calculate CON(xi, xj)
    
```

그림 5. 연결성 계산 알고리즘
Fig. 5. Connectivity calculation algorithm

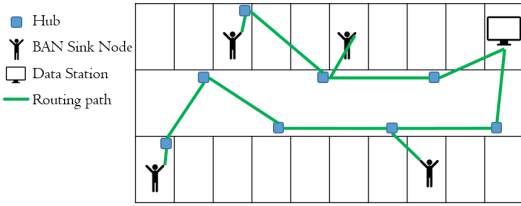


그림 6. 건물내부네트워크 구성도
Fig. 6. In-building network diagram

잔여 에너지를 고려한다.

III. Q-Learning 기반의 동적 라우팅 프로토콜

본 논문에서는 다양한 상황이 발생할 수 있는 동적인 환경을 고려하여 데이터를 전송하는 라우팅 프로토콜을 제안한다. 이를 위해 특정한 모델을 필요로 하지 않고 다양한 상황에 적용할 수 있다는 특징을 가진 Q-Learning을 사용한다.

Q-Learning은 기계학습 중 강화학습 기법의 일종이다. 강화학습은 그림 7과 같이 환경(environment)과 에이전트(agent), 에이전트가 환경에 행한 행동(action)과 행동의 결과에 의해 주어진 보상(reward)으로 이루어진다.

에이전트는 행동을 행하며 학습을 하는 주체이며 행동에 대해 보상이 보다 좋은 결과가 주어지는 방향으로 학습을 한다. 주어진 환경에 대해 무수히 많은 행동을 통해 학습하므로 동적인 환경에 대해서도 학습이 가능하다는 특징이 있다. Q-Table은 $Q(s_t, a_t)$ 에 현재 상태 s_t 에서 특정 행동 a_t 을 취했을 때 주어지는 보상 $R(s_t, a_t)$ 을 예측한 값을 저장한다.

정책은 이 보상을 극대화하도록 정의되며 정책에 따라 다음 행동을 선택하고 예측한 보상값에 따라 Q-Table을 갱신하며 학습이 진행되는 알고리즘이다^[8]. Q-Table의 Q값을 갱신하는 수식은 다음 수식(9)와 같다.

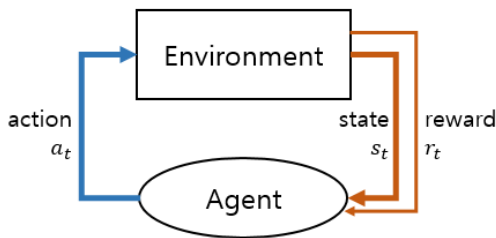


그림 7. Q-Learning 흐름도
Fig. 7. Q-Learning flow chart

$$Q(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha \{R(s_t, a_t) + \gamma^* Q_{\max}(s_{t+1}, a_{t+1})\} \quad (9)$$

여기서, $R(s_t, a_t)$ 는 행동 a_t 를 행함으로 얻어지는 보상이며 $Q_{\max}(s_{t+1}, a_{t+1})$ 는 상태 s_{t+1} 에서 미래의 행동에 대해 얻을 최대보상의 예측값이다. α ($\alpha \in (0, 1]$)는 학습이 진행되는 정도를 나타내는 변수인 학습률(learning rate)이다. α 를 변경하여 기존의 Q값과 미래의 Q값이 Q값을 갱신하는 데 미치는 영향력의 가중치를 조절하며 학습이 진행되는 속도를 조절할 수 있다. γ ($\gamma \in (0, 1]$)는 감소율(discount factor)로 미래 보상과의 거리가 멀수록 그 보상이 미치는 영향이 감소하도록 하여 미래 보상과의 거리가 짧은 최단 경로를 찾으려 한다^[8]. 본 논문에서는 미래의 Q값을 다른 요소에 의해서 조절하기 위해 Q-value 업데이트 식을 수식 (10)과 같이 설계하였다.

$$Q(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha \{R(s_t, a_t) + f_d(s_t, a_t) * \gamma^* Q_{\max}(s_{t+1}, a_{t+1})\} \quad (10)$$

수식(10)의 $f_d(s_t, a_t)$ 를 에너지 혹은 연결성과 같은 요소에 의해 조절되도록 설계하여 미래의 Q값을 조절할 수 있다.

3.1 신체 영역 네트워크에서 라우팅 프로토콜

신체 영역 네트워크에서 2.1절에서 설명했던 요소들을 고려하며 Q-Learning 모델을 모델링하여 라우팅 프로토콜을 설계한다. Q-Learning에서 에이전트는 학습을 하는 주체인 신체 영역 네트워크가 되며 상태 s_t 는 네트워크를 구성하는 노드, a_t 는 노드에서 노드로의 패킷전송으로 정의된다. 즉, a_t 는 s_t 에서 s_{t+1} 로 패킷을 전송하는 행위이며 $Q(s_t, a_t)$ 는 현재노드 s_t 에서 s_{t+1} 로 패킷을 전송했을 경우 예상되는 보상을 저장하게 된다. 보상이 가장 큰 경로가 최적경로이므로 $Q(s_t, a_t)$ 가 최대가 되는 a_t 를 선택하여 패킷을 전송함으로써 싱크노드까지의 최적경로를 도출한다. 따라서 2.1절에서 언급한 요소들이 Q값에 반영되도록 수식(10)의 $R(s_t, a_t)$ 와 $f_d(s_t, a_t)$ 를 설계한다.

2.1.3절에서는 신체의 활동성에 대하여 언급하였다. 앞서 언급한대로 신체의 활동성에 따라 링크의 연결성을 보장할 수 없기 때문에 패킷 손실이 발생하며, 재전송을 통해 손실된 패킷을 확보해야 한다. 하지만 이 과정에서 센서 노드의 에너지 손실과 패킷 지연이

발생한다. 따라서 링크의 연결성을 고려하여 최적의 경로를 도출해야 한다. 본 논문에서는 신체의 활동성에 따라 노드의 상태를 2가지로 구분한다. 신체의 움직임이 거의 없고 센서들의 위치 변화가 거의 없는 경우를 상태1로 판단하고, 신체의 움직임이 활발하여 센서들의 위치변화가 빠른 경우를 상태2로 판단한다.

노드의 개수를 N 이라 할 때, 노드의 상태를 2개로 구분하므로 각 노드의 Q-table은 그림 8과 같이 $2 * N$ 의 크기를 가지며, 총 Q-table은 $(2N) * N$ 의 크기를 가진다. 상태 1인 경우 링크는 항상 안정적이므로 링크의 연결성은 고려하지 않는다. 하지만 상태 2인 경우, 링크의 상태는 항상 불안정하기 때문에 연결성을 우선적으로 고려한다. 따라서 상태에 따라 다른 보상 함수를 적용하기 위해 $R(s_t, a_t)$ 와 $f_d(s_t, a_t)$ 를 다르게 설계한다. 상태1인 경우 $R(s_t, a_t)$ 와 $f_d(s_t, a_t)$ 는 각각 수식(11)과 수식(12)와 같다.

$$f_d(s_t, a_t) = E_r(s_{t+1}) / E_{max} \quad (11)$$

여기서, E_{max} 는 센서노드의 에너지의 최대 값이다. 상태1의 경우 에너지를 고려해야 하므로 잔여 에너지를 이용해 미래의 Q값을 조정한다. 행동 a_t 를 통해 다음노드가 s_{t+1} 가 되므로 s_{t+1} 의 잔여에너지를 고려하여 잔여에너지의 비율에 따라 미래의 Q값을 감소시킨다. 이를 통해 잔여에너지가 적은 센서 노드의 경우 해당 경로에 대한 보상의 기댓값이 줄어든다. $R(s_t, a_t)$ 는 에너지에 의해 주어지는 보상과 링크에 의해 주어지는 보상의 합으로 주어진다.

$$R(s_t, a_t) = E_p * R_e(s_t) + R_c(s_t, a_t) \quad (12)$$

여기서, E_p 는 수식(1)과 같이 계산되며, $R_c(s_t, a_t)$

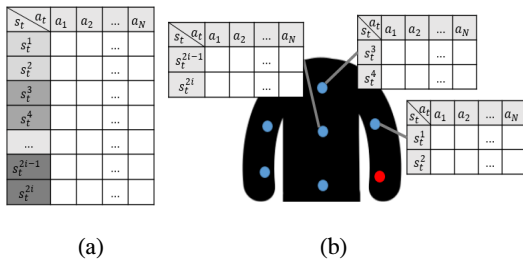


그림 8. 노드별로 상태가 2가지로 구분된 Q-table, (a) 전체 Q-table, (b) 각 노드별 Q-table
Fig. 8. Q-table that divide the state into two for each node, (a) Total Q-table, (b) Each node's Q-table

는 링크에 의해 주어지는 보상으로 정상적으로 패킷의 전달이 성공한 경우와 싱크노드에 도달한 경우에 대해서 미리 정해져있는 상수의 보상을 제공한다. 싱크노드에 도달한 경우의 보상은 패킷의 전달이 성공한 경우에 비해 월등히 크다. 또한, $E_p * R_e(s_t)$ 는 에너지에 의해 주어지는 보상으로 $R_e(s_t)$ 는 수식(13)과 같다.

$$R_e(s_t) = c_1 * N_c \quad (13)$$

여기서, N_c 는 연결된 링크의 개수이고, c_1 는 상수이다. 이는 연결된 이웃노드에 대해 보상 c_1 을 할당시켜 이 값을 주변 네트워크 환경에 대해 상대적인 값을 재분배한다. 따라서 s_{t+1} 가 다른 이웃노드들에 비해 큰 잔여에너지를 가진 경우 더 많은 보상을 분배받고 적은 경우 감춰된 양의 보상을 얻는다.

상태2의 경우 연결성이 에너지에 비해 우선적으로 고려되어야 한다. 따라서 $f_d(s_t, a_t)$ 는 연결성에 대한 수식(14)처럼 연결성에 의해 정해진다.

$$f_d(s_t, a_t) = \{CON(s_t, s_{t+1}) / C_{th}\}^2 \quad (14)$$

$f_d(s_t, a_t)$ 에 의해 연결성이 기준치 C_{th} 보다 적은 링크인 경우 미래의 Q값이 감소되며 높은 링크인 경우 그만큼 증가되어 영향을 미친다. 즉, 연결성이 높은 경로에 대한 보상의 기댓값이 증가하게 된다. 상태 2에서 $R(s_t, a_t)$ 는 에너지에 의한 보상, 연결성에 의한 보상, 링크에 의한 보상의 합으로 주어진다.

- 1: Current node = x_i
 $X = \{x_j | x_j = x_i$'s neighbor
 $A = \{a_t | a_t = \text{packet transfer to } x_j \in X$
- 2: Judge $s_t = \text{state of } x_i$
- 3: choose a_t in A (using e-greedy policy)
- 4: update Q-value of $Q(s_t, a_t)$
- 5: $Q(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha\{R(s_t, a_t) + f_d(s_t, a_t) * \gamma * Q_{max}(s_{t+1}, a_{t+1})\}$
- 6: $f_d(s_t, a_t) = \begin{cases} E_r(s_{t+1}) / E_{max} & (s_t \in \text{STATE1}) \\ \{CON(s_t, s_{t+1}) / C_{th}\}^2 & (s_t \in \text{STATE2}) \end{cases}$
- 7: $R(s_t, a_t) = \begin{cases} E_p * R_e(s_t) + R_c(s_t, a_t) & (s_t \in \text{STATE1}) \\ E_p * R_e(s_t) + \left\{ \frac{CON(s_t, a_t)}{C_s} \right\} * R_{con}(s_t) + R_c(s_t, a_t) & (s_t \in \text{STATE2}) \end{cases}$

그림 9. Q-table 갱신 알고리즘
Fig. 9. Q-table update algorithm

$$R(s_t, a_t) = 0.5 * E_p * R_c(s_t) + \{CON(s_t, a_t) / C_s\} * R_{con}(s_t) + R_c(s_t, a_t) \quad (15)$$

여기서, $\{CON(s_t, a_t) / C_s\} * R_{con}(s_t)$ 는 연결성에 의해 주어지는 보상이며, $R_c(s_t, a_t)$ 는 링크에 의한 보상이다. 또한, 에너지보다 연결성이 우선적으로 고려되어야하기 때문에 연결성보다 영향이 적도록 에너지에 의한 보상 $R_c(s_t)$ 에 0.5배를 하여 보상의 크기를 감소시킨다. $R_{con}(s_t)$ 은 수식 (16)과 같이 계산된다.

$$R_{con}(s_t) = c_2 * N_c \quad (16)$$

여기서, c_2 는 상수로 $R_{con}(s_t)$ 는 연결된 링크의 개수에 비례하여 주어지며 각 링크의 연결성에 비례하게 재분배된다.

신체 영역 네트워크에서 라우팅 프로토콜을 위한 최종 보상함수와 Q-table 업데이트 식과 알고리즘은 그림 9와 같다.

3.2 건물내부네트워크에서 라우팅 프로토콜

건물내부네트워크에서 라우팅 프로토콜을 Q-Learning을 사용하여 설계한다. 2.2에서 언급하였듯이 허브들의 잔여에너지를 고려한다.

이때, Q-Learning의 에이전트는 건물 내부 네트워크이며 상태는 각 허브를 뜻하며 행동은 패킷의 전송이다. 허브의 잔여에너지가 고려된 라우팅을 위해 Q-Table 업데이트 식은 변형된 수식(10)을 사용한다. 또한 $f_a(s_t, a_t)$ 를 통해 다음 노드의 잔여 에너지양에 따라 미래의 Q값을 조절하므로 수식(11)을 그대로 사용한다. 하지만 신체 영역 네트워크에서와 달리 건물 내부 네트워크에서는 에너지에 대한 제약이 덜하기 때문에 잔여에너지양에 대해 예민하게 반응하지 않아도 된다. 따라서, 건물내부네트워크의 보상함수는 수식 (17)과 같다.

$$R(s_t, a_t) = E_p * R_c(s_t) + R_c(s_t, a_t) \quad (17)$$

여기서, E_p 는 수식(7)과 같이 결정되기 때문에 신체 영역 네트워크의 라우팅프로토콜보다 잔여에너지양에 덜 민감하다.

IV. 모의 실험 및 성능 평가

표 1에서 제시한 시뮬레이션 환경을 바탕으로 신체 영역 네트워크와 건물 내부 네트워크에서 모의 패킷 라우팅 실험을 진행하였다.

신체 영역 네트워크의 구성은 그림 10과 같이 관절 등과 같이 신체 주요 부위 곳곳에 싱크노드를 포함하여 총 13개의 센서가 부착되어있다고 가정하였다. 여기서, 파란색 원들은 노드들이며, 빨간색 원은 싱크노드를 의미한다. 또한 건물 내부 네트워크는 17개의 허브와 1개의 데이터스테이션으로 18개의 노드로 구성되어있으며, 네트워크를 구성하고 있는 센서들은 1초마다 헬로 패킷을 전송하여 응답 패킷으로 이웃노드와 이웃노드들에 관한 정보를 받는 것으로 가정하였다. 그 중 신체 영역 네트워크를 이루는 각 노드들은 하나의 응답패킷마다 연결성을 계산하여 1초마다 이전 8개의 응답패킷의 연결성의 평균으로 이웃노드와의 연결성을 결정한다.

표 1. 실험 파라미터
Table 1. Simulation parameter

Parameter		Value
BAN setting parameter	Node number	13
	Connection range	30[cm]
	N	8
	G_r	1
	G_t	1
	P_t	1mW
	f	400MHz
	C_u	10^5
	C_m	$4 * 10^{-5}$
	goal reward	1300
connect reward	50	
Hospital setting parameter	Node number	18
	Connection range	15[m]
	Iteration number	30000
	goal reward	500
	connect reward	50
Q-Learning parameter	c_1	50
	c_2	50
	α	0.9
	γ	0.4
	C_{th}	70

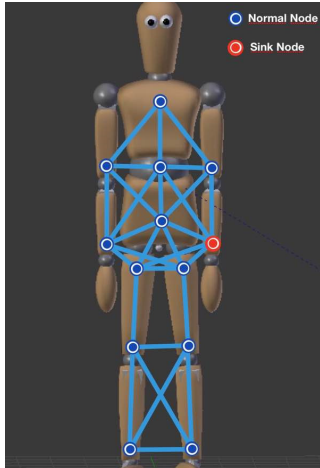


그림 10. 신체 영역 네트워크 구성
Fig. 10. Composition of BAN(Body Area Network)

실제 네트워크에서는 수신에너지를 통해 노드 간의 연결성을 계산할 수 있지만 시뮬레이션에서는 노드 간의 거리를 통해 계산한다. 본 실험은 신체를 3D 모델링하여 매트랩 환경에서 시뮬레이션을 진행하였다.

그림 11과 12는 신체 영역 네트워크의 각 상태에서 제한된 라우팅프로토콜에 의해 전송되는 패킷의 경로를 나타낸 것이다. 빨간색 마름모는 싱크노드, 분홍색 마름모는 출발노드이며 검은색 실선은 안정적인 연결, 점선은 불안정한 연결을 나타낸다. 빨간색 실선은 한 순간에 출발노드로부터 싱크노드까지의 패킷이동경로를 표시한다.

시뮬레이션 결과, 그림 11과 같이 상태 1일 때, 에너지를 고려하여 경로가 동적으로 바뀌며 매 순간의 최적경로를 찾는 것을 확인하였다. 그림 12와 같이 상태 2일 때는 각 노드간의 연결성을 고려하여 안정적인 연결이 선택되며 패킷이 전송되는 것을 확인하였다.

그림 13은 건물 내부 네트워크에서 제한된 라우팅

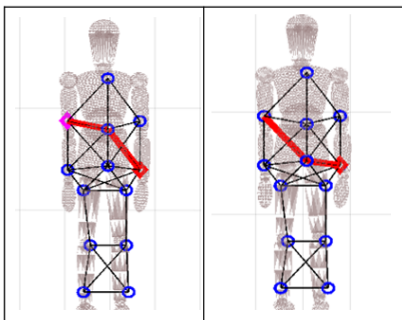


그림 11. 상태 1에서 패킷 라우팅 경로변화
Fig. 11. Packet routing path change on state 1

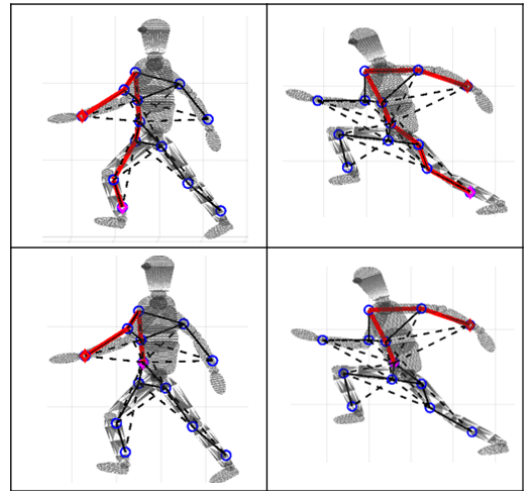


그림 12. 상태 2에서 패킷 라우팅 경로 변화
Fig. 12. Packet routing path change on state 2

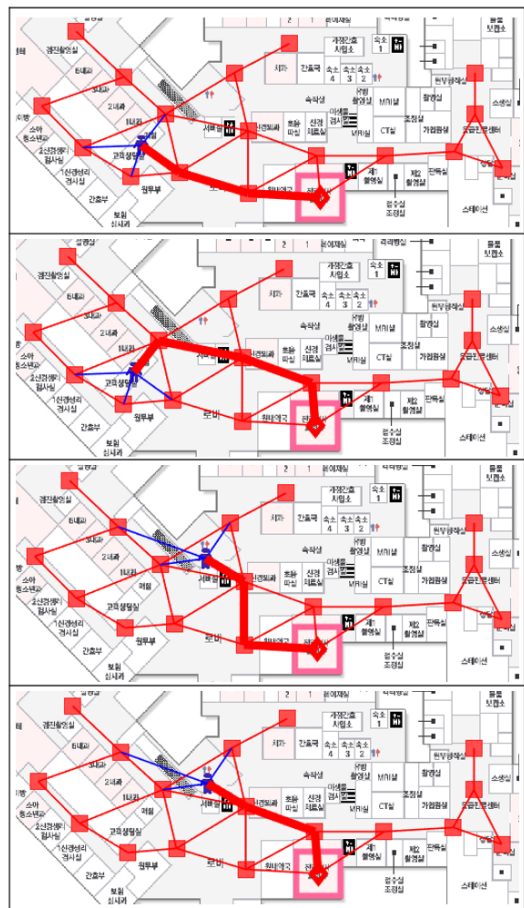


그림 13. 건물내부 패킷 라우팅 경로
Fig. 13. Indoor packet routing path

프로토콜에 의한 패킷전송경로를 보여준다. 빨간색마름모는 데이터스테이션이며 빨간색 네모는 부착되어 있는 허브이며 빨간색 선분은 허브 간의 연결을 표시한다. 파란색으로 사용자의 위치가 표시되며 파란색선분은 사용자의 싱크노드와 허브의 연결을 나타낸다. 두꺼운 빨간색 선분은 어느 한 순간에 사용자의 싱크노드로부터 데이터스테이션까지의 패킷이동경로를 나타낸다. 시뮬레이션 결과, 사용자의 위치에 따라 최적 경로를 찾아서 라우팅되며 허브의 잔여에너지가 고려되어 동적으로 경로가 변경되는 것을 확인하였다.

본 논문에서는 기존의 최단 홉 수만 고려된 라우팅 프로토콜^[9]을 적용한 네트워크 모델과 제안된 모델의 네트워크 수명향상을 비교하였다. 그림14는 비활동적인 상태에서 전송된 패킷의 수에 따른 에너지 소비를 비교하여 나타낸다. 센서노드의 위치와 잔여 에너지양을 동일하게 하여 진행하였으며, 네트워크를 구성한 노드 중 에너지가 모두 소모된 노드가 발생할 경우 네트워크의 수명이 완료된 것으로 한다.

여기서, 붉은색 선은 기존 모델의 결과이며, 파란색 선은 제안된 모델의 결과이다. 이 때, 에너지가 효율적으로 고려되는지 확인하기 위해 노드 하나의 잔여 에너지를 다른 노드의 잔여 에너지의 50%로 설정하였다.

실험 결과, 제안된 모델은 기존의 모델에 비해 최소 잔여에너지를 갖는 센서 노드의 에너지 소모율이 더 적으며, 에너지 소비가 균형적인 것을 확인하였다. 결과적으로 제안된 모델에서 네트워크의 수명이 더 늘어난 것을 확인하였다.

그림 15는 신체 활동이 상태1과 상태2가 반복될 때 에너지 소비를 비교한 그래프이다. 실험조건은 이전 실험과 같으며, 10000개의 패킷이 전송될 때 마다 신체의 상태가 바뀌도록 하였다.

그림 15와 같이 신체 활동이 계속적으로 변화는 환

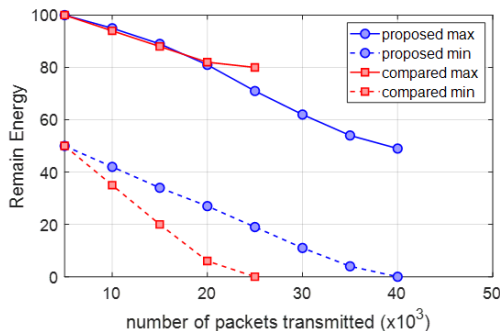


그림 14. 상태 1에서 성능 비교
Fig. 14. Performance comparison in state 1

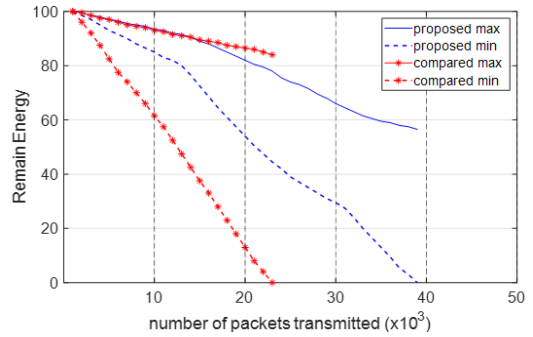


그림 15. 상태 변화에 따른 성능 비교
Fig. 15. Performance comparison with changing state

경에서도 네트워크 내부 최대 잔여 에너지를 가진 노드와 최소 잔여 에너지를 가진 노드의 에너지 차이가 제안된 모델에서 더 적었으며, 네트워크 수명이 더 연장된 것을 확인할 수 있다.

그림 16은 동적인 상태에서 패킷이 전달되는 비율을 비교하여 나타낸다. 패킷이 손실되면 해당 패킷에 대한 재전송이 발생하며, 이로 인해 네트워크에 추가적인 에너지 소비와 지연이 발생한다. 연결성이 클수록 패킷이 손실될 확률이 적기 때문에 본 실험에서는 패킷 전송 시 랜덤 값을 생성하며, 패킷을 송수신하는 두 노드간의 연결성이 생성된 랜덤 값보다 클 경우 패킷이 전달된 것으로 가정하였다.

그림 16과 같이 기존의 모델보다 제안된 모델에서의 패킷 전달률이 약 43% 증가하였다.

따라서, 제안된 모델이 기존의 모델보다 재전송이 발생하지 않는다.

그림 17과 18은 각각 state 1과 state 2 일 때, 학습에 따른 모든 노드의 평균 홉 카운트와 이상적인 최단 경로 라우팅의 평균 홉 카운트를 나타낸다. 학습 초기, 모든 노드는 충분한 에너지를 갖고 있으며, 잔여 에너지도 모두 비슷하기 때문에 홉 카운트가 감소하여 이

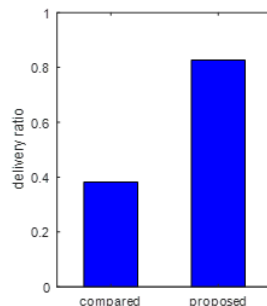


그림 16. 상태 2에서 패킷 전달률
Fig. 16. Delivery ratio in state 2

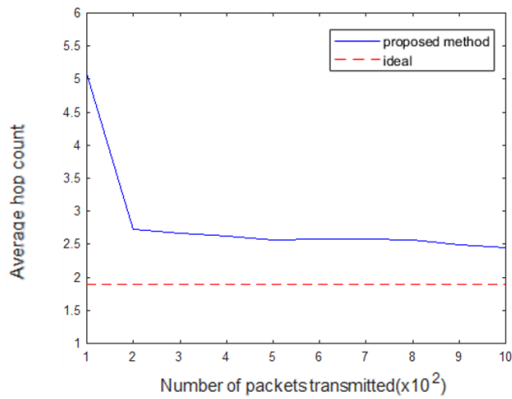


그림 17. 상태 1일 때, 학습이 진행됨에 따른 모든 노드의 평균 홉 카운트와 이상적인 홉 카운트
 Fig. 17. In state 1, the average hop count and the ideal hop count of all nodes as the learning progresses.

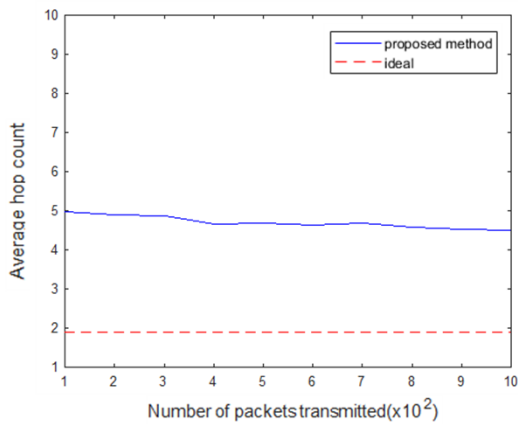


그림 18. 상태 2일 때, 학습이 진행됨에 따른 모든 노드의 평균 홉 카운트와 이상적인 홉 카운트
 Fig. 18. In state 2, the average hop count and the ideal hop count of all nodes as the learning progresses.

상적인 홉 카운트에 가까워지는 것을 볼 수 있다.

그림 16과 같이 재전송이 기존의 방법에 비해 발생하지 않으며, 그림 17, 18과 같이 제안된 방법은 학습이 진행됨에 따라 홉 카운트가 감소한다. 그러므로 제안된 방법에서 저지연성이 확보되는 것을 확인할 수 있다.

V. 결론

본 논문에서는 강화학습의 한 종류인 Q-Learning을 이용해 신체 의료 정보 수집 시스템에서 필요한 2가지 네트워크에서의 라우팅프로토콜을 제안하였다. 그 중 신체 영역 네트워크에서 정보 수집을 위해 제안

된 라우팅 프로토콜 모델은 신체의 움직임과 활동성, 노드의 잔여 에너지를 고려한 라우팅 경로를 통해 지연성과 에너지 낭비를 최소화하여 신속, 정확하게 싱크 노드로 데이터를 전송하는 것을 목표로 하였다. 학습 모델뿐만 아니라 신체 영역 네트워크의 데이터를 얻기 위해 신체를 3D 모델링하여 시뮬레이션을 진행하였다. 결과적으로 제안된 학습모델에서 네트워크의 수명이 연장된 것을 확인하였으며, 패킷의 전달률이 증가하고 학습에 따른 홉 카운트가 감소하므로 저지연성이 확보된 것을 확인하였다. 건물내부네트워크에서 제안된 라우팅프로토콜은 병원과 같은 건물 내에서 사용자들의 신체 의료 정보가 신속하게 수집되어 활용될 수 있도록 하였다. 건물 내 별도의 네트워크를 가정하여 사용자의 싱크노드에서 건물 내 데이터스테이션으로 패킷 라우팅프로토콜을 제안하였다. 건물 내 허브들을 통해 사용자에서 출발한 패킷이 데이터스테이션까지 이동하는 것과 허브의 에너지를 고려하여 라우팅 경로가 동적으로 변화하는 것을 시뮬레이션을 통해 확인하였다. 하지만 시뮬레이션에서 신체를 모델링하여 측정된 노드간의 연결성이나 움직임, 활동성은 실제 데이터가 아닌 시뮬레이션을 이용해 가정된 데이터이기 때문에 추후 실제 데이터를 이용하여 연구할 필요가 있다.

References

- [1] H. J. Yoon, "Design and development of body implantable powering device based on triboelectric effect," Ph. D. dissertation, Dept. of Advanced Materials Sci. & Eng., Sungkyunkwan university, Seoul, Aug. 2018.
- [2] W. S. Choi, L. H. Park, S. R. Cho, and I. H. Lee, "A study of routing design for wireless body area networks," in *Proc. KICS Symp.*, pp. 443-446, Korea, Jul. 2008.
- [3] Q. Yang and S. J. Yoo, "Dynamic routing path selection algorithm using reinforcement learning in wireless Ad-Hoc networks," *J. KICS*, pp. 1227-1235, Korea, Jul. 2018.
- [4] J. H. Yim, "Q-learning based geographic routing protocol for mobile ad-hoc networks," M.S. Thesis, Ajou university, Suwon, Korea, Dec. 2016.
- [5] J. S. Han, "A study of adaptive QoS routing scheme using policy-gradient reinforcement

learning,” *J. Korea Soc. Comput. and Inf.*, pp. 93-96, Korea, Nov. 2011.

- [6] Md. G. R. Alam, S. I. Moon, Rim Haw, and C. S. Hong, “A q-learning based routing protocol for body area networks,” *J. KISS: Inf. Netw.*, pp. 250-257, Oct. 2013.
- [7] L. Petrillo, T. Mavridis, J. Sarrazin, J. M. Dricot, and A. B. Delai, et al., “BAN working frequency: A trade-off between antenna efficiency and propagation losses,” *Conf. EuCAP 2014*, La Haye, Netherlands, Apr. 2014.
- [8] Y. A. Lee and T. C. Chung, “A function approximation method for Q - learning of reinforcement learning,” *J. KIISE*, pp. 1431-1438, 2004.
- [9] P. Khan, G. Konar, and N. Chakraborty, “Modification of floyd-warshall’s algorithm for shortest path routing in wireless sensor networks,” *Annu. IEEE INDICON*, Pune, India, 2014.

최 유 진 (Yoo-Jin Choi)



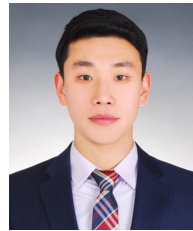
2016년 2월~현재 : 인하대학교
정보통신공학과 학사
<관심분야> 머신러닝, 네트워크
통신, 소프트웨어 개발, IoT,
빅데이터

권 정 훈 (Jeong-Hoon Kwon)



2014년 2월~현재 : 인하대학교
정보통신공학과 학사
<관심분야> 머신러닝, 네트워크
통신, IoT

나 호 정 (Ho-Jeong Na)



2018년 2월 : 인하대학교 정보통신공학과 학사 졸업
2018년 2월~현재 : 인하대학교
정보통신공학과 석사
<관심분야> 무선 통신 네트워크,
FANET, 머신러닝, IoT
[ORCID:0000-0003-2274-3107]

유 상 조 (Sang-Jo Yoo)



1990년 2월 : 한국과학기술원 전
기및전자공학과(공학석사)
2000년 8월 : 한국과학기술원 전
자전산학과(공학박사)
1990년 3월~2001년 2월 : KT 연
구 개발 본부
1990년 3월~2000년 11월 :
NIST(미국 표준기술연구원) 초빙연구원
2001년 3월~현재 : 인하대학교 정보통신공학과 교수
<관심분야> 무선 네트워크 프로토콜, Cognitive Radio
Network, 무선 센서네트워크, 지능형 사물인터넷
[ORCID:0000-0003-1533-0814]