https://doi.org/10.7840/kics.2023.48.3.359

멀티 에이전트 심층 강화학습 기반 CSMA/CA 프로토콜

조 영 제•, 황 경 호°

CSMA/CA Protocol Based on Multi Agent Deep Reinforcement

Yeong-je Jo•, Gyung-Ho Hwang°

요 약

본 논문은 CSMA/CA 프로토콜에 멀티 에이전트 강화학습을 적용하여 성능을 비교 분석한다. 기존 CSMA/CA 프로토콜은 랜덤 백오프 방식을 사용하여 백오프 값이 0인 단말들만 패킷 전송을 시도하여 채널에 접속 중인 단말의 수가 많을수록 패킷 충돌수가 증가하여 성능이 저하되는 문제가 있었다. 본 논문에서는 채널에 접속한 각각의 단말들을 하나의 에이전트로 설정하고 채널에 존재하는 모든 에이전트가 채널 상태를 관찰한 후 채널 상태에 맞춰 전송 성공률이 높은 Contention Window(CW)를 결정하여 성능을 개선 한다.

Key Words : CSMA/CA protocol, MAC protocol, Multi Agent, Deep Reinforcement Learning, MADDPG

ABSTRACT

In this letter, the performances are compared and analyzed by applying Multi Agent Deep Reinforcement Learning to the CSMA/CA protocol. In the case of the existing CSMA/CA protocol, a terminal with a backoff value of 0 using the random backoff method transmits a packet, so the higher the number of terminals connected to the channel, the higher the number of packet collisions, which degrades the performance. We propose that each terminal connected to a channel is regarded as one agent, and all agents belonging to the channel observe the channel state, and then determine the Contention Window (CW) with a high transmission success rate according to the channel state to improve performance.

Ⅰ. 서 론

최근 M2M(Machine to Machine), V2X (Vehicle to Everything) 등과 같이 다양한 서비스 환경에서 랜덤 액세스 방식을 사용하는 경우가 증가하는 추세이다.[1] IEEE 802.11 무선 전송 방식인 CSMA/CA 프로토콜은 채널에 접속한 단말 노드들이 0에서 CW값 사이의 랜덤 백오프 값을 선택하고 idle 상태마다 1만큼 감소하여 백오프 값이 0이 되었을 때 패킷을 전송한다. 2개 이상의 단말 노드들이 동시에 패킷을 전송하면 충돌이 발생하여 패킷 전송이 실패하게 되고 CW 값을 2배 증가시켜 재전송을 시도한다.[2] 다수의 단말 노드들이 동일한 AP에 접속하여 패킷 전송을 시도할 경우 충돌 발생 확률이 증가하여 지연 시간 증가, 무선자원 낭비, 처리율 감소 등과 같은 문제가 발생하여 해당 문제를 개선하기 위해 다양한 연구가 진행되고 있다.[3]

본 논문에서는 기존 CSMA/CA 프로토콜의 문제점을 개선하기 위하여 멀티 에이전트 심층 강화학습 알고리즘인 multi-agent deep deterministic policy gradient(MADDPG)를 적용하고 기존 IEEE 802.11a 프로토콜과 성능을 비교 분석한다.

Ⅱ. 관련 연구

2.1 싱글 에이전트 심층 강화학습

싱글 에이전트 강화학습은 단일 에이전트가 특정한 환경에서 행동을 취한 후 행동 가치에 따라 보상받는 과정으로 보상을 최대로 하는 행동을 학습하는 것을 목표로 한다. 행동의 가치를 평가하여 학습하는 Deep Q-Network(DQN) 알고리즘은 대표적인 가치기반 강화학습 알고리즘으로 Q-Learning에 인공신경망을 적

※ 본 논문은 2022년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다. (2021RIS-004)
• First Author : (ORCID:0000-0002-0663-0746)Dept. Computer Engineering, Hanbat National University, yeongjejo@edu.hanbat.ac.kr, 학생(석사), 학생회원
° Corresponding Author : ORCID:0000-0001-6795-8086)Dept. Computer Engineering, Hanbat National University, gabriel@hanbat.ac.kr, 정교수, 종신회원
논문번호 : 202212-296-B-LU, Received December 8, 2022; Revised December 22, 2022; Accepted December 22, 2022

용하여 다양한 상태 공간(state space) 환경에서도 행동 가치를 학습할 수 있다.[4] 행동을 결정하는 policy를 학습하는 정책기반 강화학습인 REINFORCE 알고리즘은 행동을 확률적으로 결정할 수 있어 다양한 행동 공간(action space)을 요구하는 환경에 적절한 알고리즘이다.

Deep Deterministic Policy Gradient (DDPG) 알고리즘은 가치기반 방식과 정책기반 방식을 접목한 Actor Critic 강화학습 알고리즘이다.[5] DDPG 알고리즘은 DQN 알고리즘이 연속 행동(continuous action)을 결정할 수 없었던 문제를 해결하기 위하여 행동을 결정하는 Actor 네트워크와 행동의 가치를 평가하는 Critic 네트워크로 구성하여 연속 행동을 결정할 수 있게 하였다.

2.2 멀티 에이전트 심층 강화학습

다수의 에이전트가 존재하는 환경에서 에이전트 하나의 행동으로 인하여 다른 에이전트에 영향을 주어 성능이 저하되는 싱글 에이전트 강화학습의 문제를 해결하기 위해 멀티 에이전트 강화학습 관련 연구가 진행되고 있다. 멀티 에이전트 강화학습 같은 경우 다수의 에이전트가 협업하여 보상을 최대화할 수 있는 정책을 학습하게 된다.

멀티 에이전트 심층 강화학습 알고리즘 중 하나인 MADDPG는 DDPG 알고리즘을 기반으로 자신의 관찰 정보를 바탕으로 행동을 취하고 정책을 학습할 때 다른 에이전트의 정보를 공유하는 Centralized Training and Decentralized Execution(CTDE) 방식을 사용한다.[6] 에이전트는 Actor 네트워크를 통해 관찰한 정보를 바탕으로 행동을 결정하고 행동에 관한 결과를 리플레이 메모리에 저장하고 모든 에이전트의 정보를 모아 Critic 네트워크를 통해 Q값을 구하게 된다. Actor 네트워크와 Critic 네트워크를 업데이트하는 식은 식(1), (2)와 같다. 식(1)은 모든 에이전트의 관찰 정보 x상태에서 에이전트 i는 다른 에이전트들의 policy를 예측하여 Actor 네트워크를 업데이트하고, 식(2)는 에이전트들의 행동에 대한 예측 보상 값인 Q-value와 실제로 받게 된 보상 값 y에 대한 loss를 구하여 Critic 네트워크를 업데이트한다.

$$\nabla_{\theta_i} J(\theta_i) = E_{s\ p^{\mu}, a_i\ \pi_i}\left[\nabla_{\theta_i} \log \pi_i(a_i \mid o_i)\, Q_i^{\pi}(x, a_1, ..., a_N)\right] \quad (1)$$

$$L(\theta_i) = E_{x,a,r,x'}\left[\left(Q_i^{\mu}(x, a_1, ..., a_N) - y\right)^2\right] \quad (2)$$

Ⅲ. MADDPG 기반 CSMA/CA 프로토콜

본 논문에서는 CSMA/CA 프로토콜의 성능을 개선하기 위해 심층 멀티에이전트 강화학습 알고리즘인 MADDPG를 적용 방법을 제안한다. 패킷을 전송하기 위해 채널에 접속한 단말 노드들이 1초의 time step동안 채널 상태를 관찰하고 패킷 전송 성공률이 높은 CW값을 선택하여 패킷을 전송한다.

3.1 관찰(Observation)

채널에 접속한 단말 노드 i는 이전 time step 동안의 CW값인 CW_{t-1}^{i}, 전송 성공횟수 S_{t-1}^{i}, 충돌로 인한 전송 실패 횟수 C_{t-1}^{i} 그리고 현재 전송해야 하는 패킷의 수 P_t^i를 관찰한다.

3.2 행동(Action)

행동은 802.11a 프로토콜과 동일하게 time step 동안 {15, 31, 63, 127, 255, 511, 1023} 총 7개의 행동 공간 중 하나의 CW값을 선택한다.

3.3 보상(Reward)

행동에 대한 보상 R_t^i은 식(3)과 같다. 에이전트들은 time step 동안의 패킷 전송 성공횟수S_t^i와 실패 횟수 C_t^i을 토대로 개별 보상을 받게 된다.

$$R_t^i = S_t^i - C_t^i \quad (3)$$

Ⅳ. 시뮬레이션 및 성능분석

본 논문에서 제안하는 MADDPG 기반 CSMA/CA 프로토콜의 성능을 분석하기 위해 IEEE 802.11a 프로토콜과 비교하여 성능을 분석한다. 시뮬레이션 사용한 환경 변수와 하이퍼 파라미터값은 표 1, 표 2에 나타냈으며, 패킷 생성은 식(4)와 같이 time step마다 0부터 7000까지 무작위로 생성되고 채널에 접속 중인 에이전트 수 N_t과 반비례하여 생성되게 설정하였다.

$$P_i^t = P_{t-1}^i - S_{t-1}^i + Random(7000)_t^i / N_t \quad (4)$$

그림 1은 MADDPG기반 CSMA/CA 방식과 IEEE 802.11a CSMA/CA 프로토콜의 1분간의 패킷 전송 성공 비율을 나타낸 그래프로 기존 IEEE 802.11a 방식은 채널에 접속한 노드의 수가 상대적으로 적을 경우 전송 성공 비율이 높지만 노드 수가 증가할수록 패

표 1. 환경 변수
Table 1. Environment variables

Parameter	Value
SIFS	16us
DIFS	34us
Slot time	9us
ACK frame size	14bytes
Data size	1,000bytes
Data transfer rate	54Mbps

표 2. MADDPG 하이퍼파라미터
Table 2. MADDPG hyperparameters

Parameter	Value
Hidden layer node	[256, 256]
Activation function	ReLu, Softmax
Batch size	100,000
Mini batch size	512
Actor γ	0.00025
Critic γ	0.0005

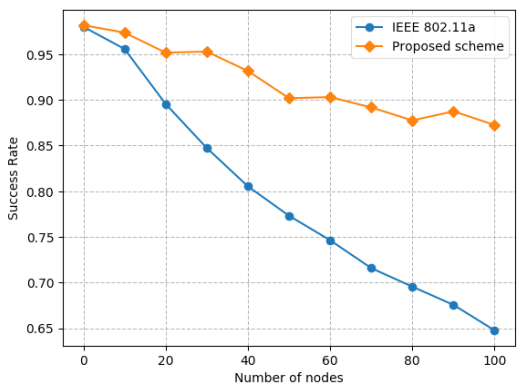

그림 1. CSMA/CA 프로토콜 성능 비교
Fig. 1. performance comparison of CSMA/CA protocol

킷 충돌이 비율이 높아져 전송 성공 비율이 감소하는 반면 제안한 MADDPG기반 CSMA/CA 프로토콜은 채널에 접속한 노드 수와 무관하게 IEEE 802.11a 프로토콜 대비 높은 전송 성공률이 보장되는 것을 확인하였다.

V. 결 론

본 논문에서는 MADDPG 기반 CSMA/CA 프로토콜을 제안하고 IEEE 802.11a 방식과 성능을 비교하여 시뮬레이션을 진행하였다. IEEE 802.11a CSMA/CA 프로토콜 같은 경우 초기 CW 최솟값인 15부터 충돌이 발생할 때마다 CW 최댓값인 1023까지 2배씩 증가시켜 채널에 접속한 단말 노드 수가 증가할수록 패킷 충돌률이 증가하여 성능이 낮아지지만 MADDPG 기반 CSMA/CA 프로토콜은 채널의 상태를 관찰하여 CW 값을 결정하고 패킷을 전송하여 IEEE 802.11a 방식 대비 성능이 개선된 것을 시뮬레이션을 통해 확인하였다.

References

[1] B. C. Jo, “Trend of V2X communication technology for connected & automated vehicles,” *AUTO JOURNAL : J. Korean Soc. Automotive Eng.*, vol. 42, no. 12, pp. 23-27, Dec. 2020.

[2] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE 802.11 Std., 2021. (https://doi.org/10.1109/ieeestd.2021.9363693)

[3] T.-W. Kim and G.-H. Hwang “Performance enhancement of CSMA/CA MAC protocol based on reinforcement learning,” *J. Inf. and Commun. Convergence Eng.*, vol. 19, no. 1, 2021. (https://doi.org/10.6109/JICCE.2021.19.1.1)

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” *NIPS Deep Learn. Wkshp.*, Dec. 2013. (https://doi.org/10.48550/arXiv.1312.5602)

[5] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *Mach. Learn.*, Sep. 2015. (https://doi.org/10.48550/arXiv.1509.02971)

[6] R. Lowe, et al., “Multi-agent actor-critic for mixed cooperative - competitive environments,” *NIPS 2017*, Jun. 2017. (https://doi.org/10.48550/arXiv.1706.02275)