

Cellular-V2X Mode 4를 위한 멀티에이전트 심층강화학습 기반 범위 적응형 분산 전력제어 알고리즘

양우열, 조한신

한밭대학교

dnduf1103@gmail.com, hsj0@hanbat.ac.kr

Multi-Agent Deep Reinforcement Learning based Range-Adaptive Distributed Power Control for Cellular-V2X Mode 4

Wooyeol Yang, Han-Shin Jo,

Hanbat National University

요약

Cellular Vehicle to everything (C-V2X)의 목적은 주변 차량에게 Cooperative Awareness Message (CAM)를 주기적으로 통신함에 있다. 그러나 C-V2X 통신은 차량 밀도가 높은 환경에서 통신성능이 저하되며, 해당 문제의 해결책으로 분산 혼잡 제어 알고리즘이 소개되었다. 기존에 소개된 알고리즘들은 CAM의 커버리지 요구 사항을 고려하지 않은 채 동작한다. 본 논문에서는 CAM의 커버리지 요구 사항을 만족시키기 위해 멀티에이전트 심층강화학습 기반의 분산 전력제어를 제안한다. 제안하는 범위 적응형 분산 전력제어 (Range-Adaptive Distributed Power Control, Ra-DPC) 알고리즘은 설정된 목표 메시지 전송 거리에 따라 각 차량이 적응적으로 전력을 선택할 수 있다. 시뮬레이션 결과를 통해 Ra-DPC 알고리즘이 기존 전력제어 알고리즘보다 에너지 효율과 Packet Delivery Ratio 측면에서 우수한 것을 확인할 수 있다.

I. 서론

Cellular Vehicle to everything (C-V2X)는 Long Term Evolution (LTE) 기반 V2X 통신을 지칭하며, 이 기술에는 차량 간 통신 (V2V), 차량과 인프라 및 네트워크 통신 (V2I/N), 차량과 단말 간 통신 (V2P)이 포함된다. C-V2X의 Mode 4 V2V 통신은 Cooperative Awareness Messages (CAM)를 주변 차량과 주기적으로 통신하는 것이 목적인 기술이며, 요구되는 통신 범위는 300-500 m이다[1]. 이러한 V2V 통신은 차량 밀도가 높은 혼잡한 상황에서 성능이 더욱 불리하게 나타나며, 이를 극복하기 위한 알고리즘으로 분산 혼잡 제어(distributed congestion control, DCC)가 소개되었다. 미국의 Society of Automotive Engineers (SAE)는 J2945/1 문서에서 Dedicated Short Range Communication (DSRC)의 전력제어와 전송 주기제어를 활용한 DCC 알고리즘을 소개했다[2]. 유럽의 European Telecommunications Standards Institute (ETSI)는 DCC 프레임워크를 구축했다.

그러나 이러한 DCC 알고리즘은 CAM의 통신 범위를 고려하지 않은 채 동작한다. CAM의 커버리지 요구사항이 300-500 m인 점을 고려한다면, DCC 알고리즘은 해당 범위 내 성능을 최적화시키도록 설계되어야 한다. 본 논문에서는 목표 메시지 전송 거리 (Target Range, TR) 내 성능을 최적화시키는 멀티에이전트 심층강화학습 기반 범위 적응형 분산 전력제어 (Range-Adaptive Distributed Power Control, Ra-DPC) 알고리즘을 제안한다.

심층강화학습이란 딥러닝과 강화학습이 결합된 학습 알고리즘을 일컫는다. 차량 네트워크에 심층강화학습이 적용된 사례가 있으며, 해당 연구들은 심층강화학습 중 Deep Q-Networks (DQN) 알고리즘을 이용하였다[3],[4]. DQN 알고리즘은 학습의 주체인 에이전트의 행동 수가 적을 때, 좋은 학습 결과를 보인다. 그러나 V2V 전력제어에서 학습은 정해진 전력 범위 내 매우 많은 수의 전력 값을 선택해야 하므로, DQN의 적용은 적절하지 않다. 따라서, Ra-DPC 알고리즘은 에이전트의 행동 수가 많아도 좋은 학습 결과를 보이는 Policy Gradient (PG)

알고리즘을 이용한다. 본 연구의 기여는 다음과 같다. 1) C-V2X DCC 전력제어에 처음으로 심층강화학습을 적용한다. 2) Ra-DPC는 기존 DCC 알고리즘에 없는 TR의 개념을 추가하고, TR에 따라 적응적으로 통신성능을 최적화시킨다.

II. 멀티에이전트 심층강화학습 기반 범위 적응형 분산 전력제어 (Ra-DPC)

심층강화학습은 신경망의 가중치 (θ)를 최적화하여, 에이전트의 정책 (μ_θ)을 최적화시키는 알고리즘이다. 여기서 μ_θ 는 입력으로 시점 t 의 관측치 (S_t)를 입력으로 받아 행동 (A_t)의 확률을 출력하는 함수이다. 최적화된 μ_θ 는 보상 (R_t)을 최대화한다. 학습을 위해 PG 알고리즘의 목적함수는 다음과 같이 정의된다.

$$J(\theta) = E\left[\sum_{t=0}^{T-1} R_{t+1} | \mu_\theta\right] \quad (1)$$

여기서 T 는 에피소드의 종료 시점이며, 목적함수를 최대화하는 것이 심층강화학습의 목표이다. Ra-DPC 알고리즘에 적용된 PG 알고리즘은 Monte Carlo Policy Gradient (MCPG)이며, T 시점에서 θ 를 업데이트한다. 이때, MCPG는 Gradient descent가 아닌, Gradient ascent를 사용한다.

$$\nabla_\theta J(\theta) = \sum_{t=0}^{T-1} G_t \nabla_{\theta_\mu} \log(\mu(S_t | \theta_\mu)) \quad (2)$$

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta) \quad (3)$$

식(2)에서 G_t 는 보상의 기댓값을 의미하며 다음의 식으로 정의된다.

$$G_t = \sum_{k=t}^T \gamma^{k-t} R_k \quad (4)$$

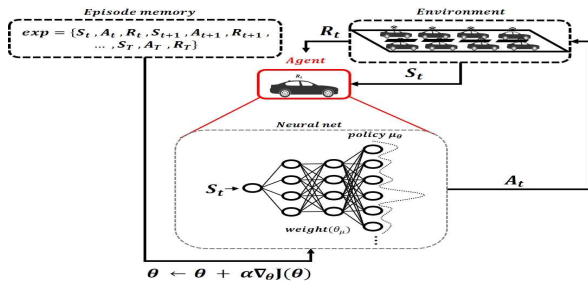
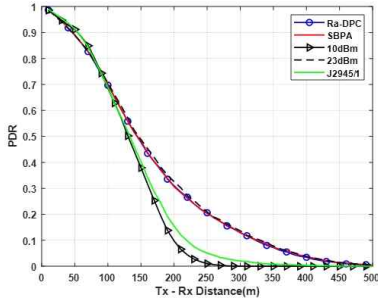


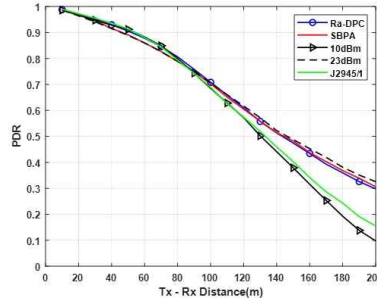
그림 1. Ra-DPC 시스템 구조

표 1. 시뮬레이션 파라미터

parameter	Value
Density (ρ)	300 veh/km
Message size	300 byte
Channel model	Winner+B1
Bandwidth	10 MHz
Antenna gain (G_t, G_r)	3 dB
Noise figure	9 dB
Target range (TR)	[200,500] m
MCS index	3



(a) TR = 500 m



(b) TR = 200 m

그림 2. 송 수신기 거리별 PDR

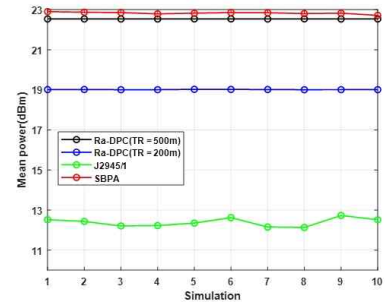


그림 3. 시뮬레이션 평균 전력

식 (3),(4)에서 α , γ 는 학습률과 감가율로 0과 1사이 값이다.

그림 1은 Ra-DPC의 시스템 구조를 나타낸다. 에이전트는 각 차량이며, 학습된 신경망을 통해 전력을 조정한다. 심층강화학습을 분산 전력제어에 적용하기 위해 S_t 는 점유 중인 채널의 누적 간섭으로 설정했으며, A_t 는 10-23 dBm까지의 14개의 전력 레벨로 설정했다. 학습을 통해 최적화하려는 것은 V2V 통신성능이기 때문에 R_t 는 TR 범위 내 packet delivery ratio (PDR)로 설정했다. 다음은 Ra-DPC알고리즘 슈도 코드이다.

Algorithm 1 Ra-DPC Algorithm

```

Initialize neural network with random weights  $\theta$ 
for Episode = 1, K do
  Initialize Episode memory
  for VUE = 1, M do
    for Step = 1, T do
      Select power level  $A_t$  through policy  $\mu_\theta(A_t|S_t)$ 
      broadcast a CAM with the power level  $A_t$ 
      Save the trajectory to the memory
    end for
  end for
   $\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$ 
return  $\theta$ 
end for

```

III. 시뮬레이션 성능 분석

성능 평가를 위해 표 1에 제시된 값을 적용하여 시뮬레이션을 수행하였으며, 기존 기술과의 비교분석을 위해 J2945/1의 전력제어[2] 및 기존 전력제어 Sensing Based Power Allocation (SBPA)알고리즘[4]의 성능 또한 평가한다. 그림 2은 TR에 따른 거리별 PDR을 보여준다. 결과에서와같이 Ra-DPC알고리즘은 J2945/1보다 모든 결과에서 우수한 PDR을 보여준다. 기존 전력제어 알고리즘인 SBPA와 비교했을 때, Ra-DPC는 (a)에서 더 우수한 결과를 보이며, 전반적으로 비슷한 성능을 보였다. 따라서, Ra-DPC알고리즘이 표준 및 기존 알고리즘과 비교하여 PDR측면에서 우수함을 확인 가능하다.

그림 3는 시뮬레이션의 평균 전력을 보여준다. Ra-DPC알고리즘은 설정된 TR에 따라 적응적으로 전력을 선택하며, TR의 범위가 넓을수록 전력이 높아지는 것을 확인 가능하다. 두 시뮬레이션 결과를 통해 제안하는 Ra-DPC알고리즘은 TR에 따라 전력을 효율적으로 이용하기 때문에 기존 알고리즘과 비교

하여 PDR과 에너지효율 측면에서 우수하다는 것을 알 수 있다.

IV. 결론

본 논문에서는 CAM의 커버리지 요구사항을 만족시키기 위해 심층강화학습 기반의 전력제어를 소개한다. 제안하는 알고리즘은 기존 알고리즘과 다르게 설정된 TR에 따라 적응적으로 전력을 선택했다. TR에 따라 적절한 전력을 선택하기 때문에 결과적으로 두 알고리즘보다 PDR과 에너지효율 측면에서 우수한 결과를 보였다. 향후 연구에서는 전력제어뿐만 아니라 전송 주기제어 등 다른 분산 혼잡 제어알고리즘에 심층강화학습을 적용할 계획이다.

ACKNOWLEDGMENT

본 성과물은 중소벤처기업부에서 지원하는 2020년도 맞춤형 기술파트너 지원사업(No. S2993642)의 연구수행으로 인한 결과물임을 밝힙니다.

참 고 문 헌

- [1] Study on LTE-based V2X services, 3GPP TR 36.885 v14.0.0, Jun. 2016.
- [2] On-Board System Requirements for V2V Safety Communications, SAE J2945/1, Mar. 2016.
- [3] Ye, H., Li, G. Y., Juang, B. H. F. "Deep reinforcement learning based resource allocation for V2V communications," IEEE Transactions on Vehicular Technology, 68.4, 3163-3173.
- [4] Zhang, H., Chong, S., Zhang, X., Lin, N.. "A deep reinforcement learning based D2D relay selection and power level allocation in mmWave vehicular networks," IEEE Wireless Communications 2019, 9.3: 416-419.
- [5] B. Kang S. Jung and S. Bahk, "Sensing-based power adaptation for cellular V2X mode 4," in 2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). IEEE, 2018. p. 1-4.