(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0020537
(43) 공개일자 2024년02월15일

(51) 국제특허분류(Int. Cl.)
H04W 4/40 (2018.01) *G06N 3/04* (2023.01)
H04W 28/02 (2009.01)
(52) CPC특허분류
H04W 4/40 (2020.05)
G06N 3/04 (2023.01)
(21) 출원번호 10-2022-0098709
(22) 출원일자 2022년08월08일
심사청구일자 2022년08월08일

(71) 출원인
국립한밭대학교 산학협력단
대전광역시 유성구 동서대로 125 (덕명동)
(72) 발명자
조한신
대전광역시 서구 둔산로 155 크로바아파트 109동 506호
양우열
대전광역시 동구 계족로489번길 73 한숲아파트 101동 1103호
(74) 대리인
이은철, 이우영

전체 청구항 수 : 총 11 항

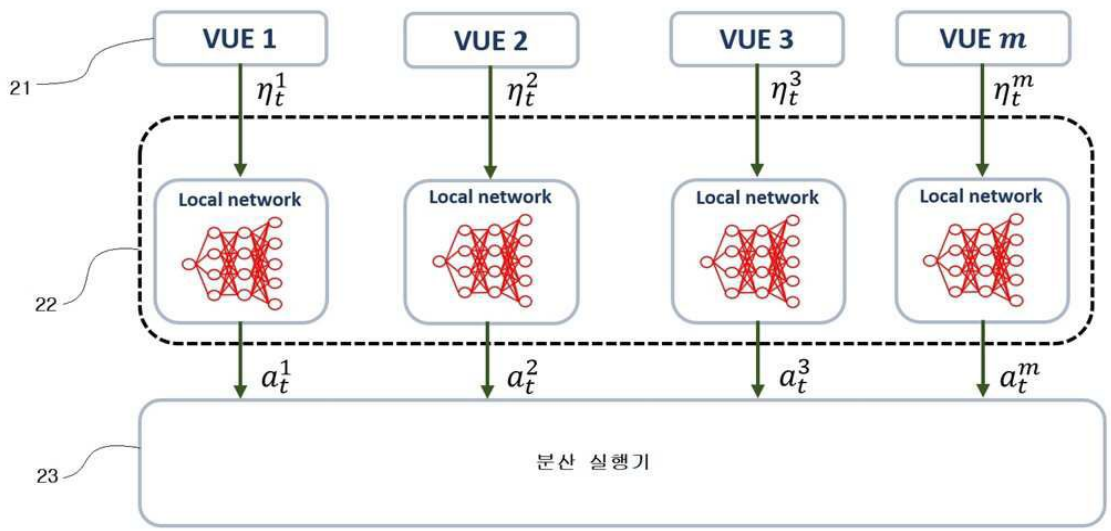
(54) 발명의 명칭 **심층 강화 학습 기반의 차량 통신 시스템 및 이의 방법**

(57) 요 약

본 기술은 심층 강화 학습 기반의 차량 통신 시스템 및 이의 방법이 개시되어 있다. 본 기술의 구체적인 구현 예에 따르면, 심층 강화 학습을 수행하여 각 차량 밀도에 따라 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값을 타겟 통신 범위 내에 존재하는 적어도 하나의 차량 단말로 공유하고, 타겟 통신 범위 내에 존재하는 차량의 노드 수로 입력된

(뒷면에 계속)

대 표 도 - 도5



현재 시점 상태에 대해 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값의 행동으로 도출된 행동으로 패킷 생성 시간의 상한치를 도출하며 도출된 패킷 생성 시간의 상한치로 통신 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출할 수 있고, 또한 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값의 행동으로 도출된 행동으로 패킷 생성 시간의 상한치를 도출하므로 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 연산 복잡도를 감소할 수 있고, 이에 경량 디바이스에 범용적으로 적용할 수 있으며, 이동통신망의 타겟 신경망의 타겟값으로 타겟 PDR을 제어하여 패킷 생성 시간의 상한치를 조정함에 따라 조정된 패킷 생성 시간에 의거 C-V2X 기반의 차량 통신 시스템의 서비스 품질이 급격히 저하되는 현상을 방지할 수 있다.

(52) CPC특허분류
H04W 28/0231 (2023.01)
H04W 28/0289 (2023.05)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711160483
과제번호 2022-0-01053-001
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 자율주행기술개발혁신사업(R&D)
연구과제명 다중 통신기술 네트워크 로드밸런싱 기술개발
기 여 율 1/2
과제수행기관명 대구경북과학기술원
연구기간 2022.04.01 ~ 2024.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711174097
과제번호 00156212
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성(R&D)
연구과제명 ICT혁신인재4.0 (한밭대학교)
기 여 율 1/2
과제수행기관명 한밭대학교산학협력단
연구기간 2022.07.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

이동 통신망과 차량 통신망을 이용하여 다수 차량과 사물 간의 통신을 수행하는 C-V2X(cellular vehicle to everying) 기반의 차량 통신 시스템에 있어서,

이동 통신망을 통해 다수의 차량과의 통신을 수행하는 중앙 제어기를 포함하고,

상기 중앙 제어기는,

V2V(Vehicle to Vehicle) 기반의 차량 통신 시스템의 분산 혼잡 제어(DCC: Distributed Congestion Control) 기반으로 학습 모델을 구축하고

구축된 학습 모델에 대해 적어도 하나의 차량 밀도 각각에 대해 심층 강화 학습으로 훈련하여 최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 도출하며,

도출된 최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 차량 통신망의 각 차량 단말로 공유하도록 구비되는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 2

제1항에 있어서, 상기 중앙 제어기의 로컬 신경망의 $Q(s,a;\theta)$ 값은,

심층 강화 학습을 수행하기 위해 현재 시점 t의 상태 s_t^{cc} , 행동 a_t^{cc} , 및 가중치 Θ를 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 3

제1항에 있어서, 상기 중앙 제어기는,

현재 시점의 상태의 입력으로 기 설정된 확률로 랜덤하게 행동을 선택하는 로컬 신경망;

상기 로컬 신경망의 행동을 입력으로 리워드 및 다음 시점의 상태를 도출하고, 도출된 현재 시점의 상태, 행동, 리워드 및 다음 시점의 상태를 포함하는 트랜지션을 학습 모델로 생성하며, 생성된 트랜지션을 재생 메모리에 저장하는 모델 구축부;

상기 트랜지션의 다음 시점의 상태로 손실함수의 타겟값을 설정하는 타겟 신경망; 및

상기 로컬 신경망의 $Q(s,a;\theta)$ 값과 타겟 신경망의 타겟값을 토대로 손실함수를 도출하고 도출된 손실함수로 상기 로컬 신경망 및 타겟 신경망의 각 가중치를 업데이트하여 상기 로컬 신경망 및 타겟 신경망을 훈련하는 손실함수 연산부;를 포함하고,

상기 로컬 신경망은

상기 트랜지션의 현재 시점의 상태를 입력으로 모델 구축부의 $Q(s,a;\theta)$ 값을 출력하도록 구비되는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 4

제3항에 있어서, 상기 로컬 신경망은

현재 시점의 상태의 입력으로 기 설정된 확률이 아닌 경우(1-설정된 확률인 경우) 학습 모델의 $Q(s,a;\theta)$ 값

중 최대 $Q(s,a;\theta)$ 값의 행동을 선택하도록 구비되는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 5

제1항에 있어서, 상기 중앙 제어기는,

상기 재생 메모리의 트랜지션이 기 정해진 미니 배치 사이즈 보다 큰 경우 상기 트랜지션을 미니 배치 사이즈로 샘플링하여 현지 시점의 상태를 상기 로컬 신경망에 전달하고, 다음 시점의 상태를 상기 타겟 신경망에 전달하는 샘플링부를 더 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 6

제1항에 있어서, 상기 차량 통신망의 각 차량 단말은,

기 정해진 단위 거리의 차량 통신망의 차량의 수를 현재 시점의 상태로 획득하는 상태 획득부;

상기 상태 획득부의 현재 시점의 상태 입력에 대해 훈련된 로컬 신경망으로 훈련하여 최대 리워드를 가지는 각 차량의 행동을 도출하는 로컬 신경망 구동부; 및

각 차량의 행동으로 각 차량의 패킷 생성 시간을 조정하고 조정된 패킷 생성 시간으로 통신 수행하여 각 차량의 패킷 전송비를 도출하는 분산 실행기를 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템.

청구항 7

C-V2X 기반의 차량 통신 시스템의 이동 통신망의 중앙 제어기에서 수행되는 심층 강화 학습 기반의 차량 통신 방법에 있어서,

로컬 신경망에서 입력된 현재 시점의 상태로 랜덤하게 또는 최대 $Q(s,a;\theta)$ 값의 행동을 선택하는 로컬 신경망 학습 단계;

상기 로컬 신경망의 행동을 입력으로 리워드 및 다음 시점의 상태를 도출하고, 도출된 현재 시점의 상태, 행동, 리워드 및 다음 시점의 상태를 포함하는 트랜지션을 학습 모델로 생성하며, 생성된 트랜지션을 재생 메모리에 저장하는 모델 구축단계;

타겟 신경망에서 상기 트랜지션의 다음 시점의 상태로 행동으로 훈련하고 손실함수의 타겟값을 설정하는 타겟 신경망 학습 단계;

입력된 현재 시점의 상태에 대해 상기 트랜지션의 최대 리워드를 가지는 행동을 선택하여 $Q(s,a;\theta)$ 값을 도출하고, 도출된 로컬 신경망의 최대 리워드를 가지는 $Q(s,a;\theta)$ 값과 타겟 신경망의 타겟값을 토대로 손실함수를 도출하며, 도출된 손실함수로 상기 로컬 신경망 및 타겟 신경망의 각 가중치를 업데이트하여 상기 로컬 신경망 및 타겟 신경망을 훈련하는 로컬 신경망 훈련단계; 및

최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 출력하고 출력된 로컬 신경망의 $Q(s,a;\theta)$ 값을 타겟 통신 범위 내에 존재하는 차량 단말에 공유하는 전송단계를 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템을 이용한 차량 통신 방법.

청구항 8

제7항에 있어서, 상기 차량 통신 방법은

상기 트랜지션이 기 정해진 미니 배치 사이즈 보다 큰 경우 상기 트랜지션을 미니 배치 사이즈로 샘플링하여 현지 시점의 상태를 상기 로컬 신경망에 전달하고, 다음 시점의 상태를 상기 타겟 신경망에 전달하는 샘플링 단계를 더 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템을 이용한 차량 통신 방법.

청구항 9

제7항에 있어서, 상기 각 차량 단말에 의거 수행하는 차량 통신 방법은,

기 정해진 단위 거리의 통신 커버리지 내의 차량 단말의 노드 수로 획득된 현재 시점의 상태를 입력하여 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값으로 각 차량 단말의 행동을 도출하는 단계; 및

도출된 차량의 행동으로 각 차량 별 패킷 생성 시간의 상한치를 조정하고 조정된 패킷 생성 시간의 상한치로 차량 간의 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 단계를 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템을 이용한 차량 통신 방법.

청구항 10

C-V2X 기반의 차량 통신 시스템의 차량 통신망의 각 차량 단말에 의거 수행하는 차량 통신 방법에 있어서,

기 정해진 타겟 통신 범위 내의 차량 단말의 노드 수로 입력된 현재 시점의 상태에 대해, 이동 통신망의 중앙 제어기에서 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값을 토대로 각 차량의 행동을 도출하는 단계; 및

도출된 차량의 행동으로 각 차량 별 패킷 생성 시간의 상한치를 조정하고 조정된 패킷 생성 시간의 상한치로 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 단계를 포함하는 것을 특징으로 하는 심층 강화 학습 기반의 차량 통신 시스템을 이용한 차량 통신 방법.

청구항 11

제7항 내지 제10항 중 한 항의 심층 강화 학습 기반의 차량 통신 시스템을 이용한 차량 통신 방법을 실행하기 위한 프로그램이 기록되고 컴퓨터에서 판단 가능한 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 심층 강화 학습 기반의 차량 간의 통신 시스템 및 이의 방법에 관한 것으로서, 보다 상세하게는, C-V2X 기반의 차량 통신 시스템에 심층 강화 학습을 수행하여 패킷 생성 시간을 조정하고 조정된 패킷 생성 시간으로 통신하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 향상시킬 수 있도록 한 기술에 관한 것이다.

배 경 기 술

[0002] 일반적으로 차량통신은, 차량과 차량 간의 통신인 V2V(Vehicle to Vehicle)와, 차량과 기지국 간의 통신인 V2I(Vehicle to Infrastructure)로 구분될 수 있다.

[0003] 일정한 이격거리 내에서 각 차량에 장착된 차량 단말을 통해 데이터 통신하여 V2V 서비스를 이용할 수 있다. 여기서, 차량 단말은, 도로를 주행하는 차량에 각각 장착되어, 각 차량으로부터 요청되는 차량통신 서비스(V2I 서비스, V2V 서비스)를 제공하는 V2X 단말에 의해 구현될 수 있다.

[0004] V2X 단말은, 대용량의 차량통신을 수행하는 차량통신 모듈(WAVE+)(231, 241)과, 이동통신 모듈(LTE)(232, 242)이 결합된 구조로, 최적의 차량통신 네트워크 접속을 수행할 수 있다.

[0005] V2X 단말은 상대적으로 대용량의 데이터 신호를 고속으로 전송이 가능한 차량통신 모듈을 우선적으로 이용하되, 전국에 설치되어 있는 다수의 이동통신 기지국을 기반으로 끊김 없는 데이터 신호의 전송이 가능한 이동통신 모듈(LTE)을 상호 보완해 이용하여 V2I 서비스를 제공할 수 있다.

[0006] 다른 차량과 주변 사물과의 통신을 통해 교통사고를 예방하기 위한 C-V2X(Cellular Vehicle to Everything)는 이동통신 모듈의 Uu 인터페이스와 차량통신 모듈의 PC5 인터페이스를 사용하는 통신모드로 나눌 수 있으며, Uu 인터페이스는 셀룰러 네트워크와 차량 간의 통신을 지원하고 셀룰러 네트워크는 데이터 전송을 위해 각 차량의 하위 채널 및 무선 자원을 예약한다. PC5 인터페이스는 차량 대 차량 통신을 지원하며 상기 셀룰러 네트워크 통신과 무관하게 동작한다.

[0007] 그러나 차량과 차량 간의 통신은 차량의 고밀도 환경에서 높은 간섭으로 인해 서비스 품질(QoS : Quality of Service)을 만족하기 어렵다.

[0008] 이에 최근에는 근거리 통신망을 위한 메시지 전송 전력 및 데이터 전송률제어를 수행하는 분산형 혼합 제어

(Distributed Congestion Control, 이하 DCC로 약칭함)에 대한 연구가 활발하게 이루어지고 있다.

[0009] 이러한 DCC 기법은 C-V2X의 고밀도 환경에서 통신 성능을 향상시킬 수 있으나, 서비스 품질(QoS) 관련 매개변수가 없이 작동되므로 C-V2X 환경의 서비스 품질(QoS)을 충족할 수 없는 문제점이 있었다.

[0010] 이에 본 출원인은 다양한 차량 밀도의 V2V(차량 대 차량)의 통신 범위 및 타겟 서비스 품질(QoS)을 토대로 데이터 전송률을 조정하여 C-V2X의 타겟 서비스 품질 QoS를 만족하는 심층 강화 학습 기반의 적응적 DCC 방안을 제안하고자 한다.

선행기술문헌

비특허문헌

[0011] (비특허문헌 0001) 선행문헌 1: I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, “Neu- ral combinatorial optimization with reinforcement learning,” Jan. 2017, arXiv:1611.09940. [Online]. Available: https://arxiv.org/abs/1611.09940

(비특허문헌 0002) 선행문헌 2: H. Qie, D. Shi, T. Shen, X. Xu, Y. Li, and L. Wang, “Joint optimization of multi-UAV target assignment and path planning based on multi-agent reinforcement learning,” IEEE Access, vol. 7, pp. 146264-146272, 2019.

발명의 내용

해결하려는 과제

[0012] 본 발명은, 패킷 생성 시간을 심층 강화 학습을 통해 최대 리워드를 가지는 로컬 신경망으로 조정하여 적응적 DCC 기반으로 C-V2X의 서비스 품질을 향상시키기 위한 것이다.

[0013] 이에 패킷 생성 시간을 조정하는 연산 복잡도를 감소할 수 있고, 경량의 디바이스에 범용적으로 적용할 수 있다.

[0014] 본 발명의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있으며, 본 발명의 실시 예에 의해 보다 분명하게 알게 될 것이다. 또한, 본 발명의 목적 및 장점들은 특허청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0015] 이러한 기술적 과제를 달성하기 위한 본 발명의 심층강화학습 기반의 차량 통신 시스템은,

[0016] 이동 통신망과 차량 통신망을 이용하여 다수 차량과 사물 간의 통신을 수행하는 C-V2X(cellular vehicle to everying) 기반의 차량 통신 시스템에 있어서,

[0017] 이동 통신망을 통해 다수의 차량과의 통신을 수행하는 중앙 제어기를 포함하고,

[0018] 상기 중앙 제어기는,

[0019] V2V(Vehicle to Vehicle) 기반의 차량 통신 시스템의 분산 혼잡 제어(DCC: Distributed Congestion Control) 기반으로 학습 모델을 구축하고

[0020] 구축된 학습 모델에 대해 적어도 하나의 차량 밀도 각각에 대해 심층 강화 학습으로 훈련하여 최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 도출하며,

[0021] 도출된 최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 차량 통신망의 각 차량 단말로 공유하도록 구비될 수 있다.

[0022] 바람직하게 상기 중앙 제어기의 로컬 신경망은,

[0023] 심층 강화 학습을 수행하기 위해 현재 시점의 상태, 행동, 및 가중치를 포함할 수 있다.

[0024] 바람직하게 상기 중앙 제어기는,

[0025] 현재 시점의 상태의 입력으로 기 설정된 확률로 랜덤하게 행동을 선택하는 로컬 신경망;

[0026] 상기 로컬 신경망의 행동을 입력으로 리워드 및 다음 시점의 상태를 도출하고, 도출된 현재 시점의 상태, 행동, 리워드 및 다음 시점의 상태를 포함하는 트랜지션을 학습 모델로 생성하며, 생성된 트랜지션을 재생 메모리에 저장하는 모델 구축부;

[0027] 상기 트랜지션의 다음 시점의 상태로 행동으로 훈련하고 손실함수의 타겟값을 설정하는 타겟 신경망; 및

[0028] 상기 로컬 신경망의 최대 리워드를 가지는 $Q(s,a;\theta)$ 값과 타겟 신경망의 타겟값을 토대로 손실함수를 도출하고 도출된 손실함수로 상기 로컬 신경망 및 타겟 신경망의 각 가중치를 업데이트하여 상기 로컬 신경망 및 타겟 신경망을 훈련하는 손실함수 연산부;를 포함하고,

[0029] 상기 로컬 신경망은

[0030] 상기 트랜지션의 현재 시점의 상태를 입력으로 모델 구축부의 최대 리워드를 가지는 행동을 선택하여 $Q(s,a;\theta)$ 값을 출력하도록 구비될 수 있다.

[0031] 바람직하게 상기 로컬 신경망은

[0032] 현재 시점의 상태의 입력으로 기 설정된 확률이 아닌 경우(1-설정된 확률인 경우) 학습 모델의 $Q(s,a;\theta)$값 중 최대 $Q(s,a;\theta)$ 값의 행동을 선택하도록 구비될 수 있다.

[0033] 바람직하게 상기 중앙 제어기는,

[0034] 상기 재생 메모리의 트랜지션이 기 정해진 미니 배치 사이즈 보다 큰 경우 상기 트랜지션을 미니 배치 사이즈로 샘플링하여 현지 시점의 상태를 상기 로컬 신경망에 전달하고, 다음 시점의 상태를 상기 타겟 신경망에 전달하는 샘플링부를 더 포함할 수 있다.

[0035] 바람직하게 상기 차량 통신망의 각 차량 단말은,

[0036] 기 정해진 단위 거리의 차량 통신망의 차량의 수를 현재 시점의 상태로 획득하는 상태 획득부;

[0037] 상기 상태 획득부의 현재 시점의 상태 입력에 대해 훈련된 로컬 신경망으로 훈련하여 최대 리워드를 가지는 각 차량의 행동을 도출하는 로컬 신경망 구동부; 및

[0038] 각 차량의 행동으로 각 차량의 패킷 생성 시간을 조정하고 조정된 패킷 생성 시간으로 통신 수행하여 각 차량의 패킷 전송비를 도출하는 분산 실행기를 포함할 수 있다.

[0039] 본 발명의 다른 실시예에 의거한, 심층 강화 학습 기반의 차량 통신 시스템에 의거 수행되는 차량 통신 방법은,

[0040] C-V2X 기반의 차량 통신 시스템의 이동 통신망의 중앙 제어기에서 수행되는 심층 강화 학습 기반의 차량 통신 방법에 있어서,

[0041] 로컬 신경망에서 입력된 현재 시점의 상태로 랜덤하게 또는 최대 $Q(s,a;\theta)$ 값의 행동을 선택하는 로컬 신경망 학습 단계;

[0042] 상기 로컬 신경망의 행동을 입력으로 리워드 및 다음 시점의 상태를 도출하고, 도출된 현재 시점의 상태, 행동, 리워드 및 다음 시점의 상태를 포함하는 트랜지션을 학습 모델로 생성하며, 생성된 트랜지션을 재생 메모리에 저장하는 모델 구축단계;

[0043] 타겟 신경망에서 상기 트랜지션의 다음 시점의 상태로 행동으로 훈련하고 손실함수의 타겟값을 설정하는 타겟 신경망 학습 단계;

[0044] 입력된 현재 시점의 상태에 대해 상기 트랜지션의 최대 리워드를 가지는 행동을 선택하여 $Q(s,a;\theta)$값을 도출

하고, 도출된 로컬 신경망의 최대 리워드를 가지는 $Q(s,a;\theta)$ 값과 타겟 신경망의 타겟값을 토대로 손실함수를 도출하며, 도출된 손실함수로 상기 로컬 신경망 및 타겟 신경망의 각 가중치를 업데이트하여 상기 로컬 신경망 및 타겟 신경망을 훈련하는 로컬 신경망 훈련단계; 및

[0045] 최대 리워드를 가지는 로컬 신경망의 $Q(s,a;\theta)$ 값을 출력하고 출력된 로컬 신경망의 $Q(s,a;\theta)$값을 타겟 통신 범위 내에 존재하는 차량 단말에 공유하는 전송단계를 포함하는 것을 일 특징으로 한다.

[0046] 바람직하게 상기 차량 통신 방법은

[0047] 상기 트랜지션이 기 정해진 미니 배치 사이즈 보다 큰 경우 상기 트랜지션을 미니 배치 사이즈로 샘플링하여 현지 시점의 상태를 상기 로컬 신경망에 전달하고, 다음 시점의 상태를 상기 타겟 신경망에 전달하는 샘플링 단계를 더 포함할 수 있다.

[0048] 바람직하게 상기 각 차량 단말에 의거 수행하는 차량 통신 방법은,

[0049] 기 정해진 단위 거리의 통신 커버리지 내의 차량 단말의 노드 수로 획득된 현재 시점의 상태를 입력하여 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값으로 각 차량 단말의 행동을 도출하는 단계; 및

[0050] 도출된 차량의 행동으로 각 차량 별 패킷 생성 시간의 상한치를 조정하고 조정된 패킷 생성 시간의 상한치로 차량 간의 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 단계를 포함할 수 있다.

[0051] 본 발명의 또 다른 실시예에 의거한 C-V2X 기반의 차량 통신 시스템의 차량 통신망의 각 차량 단말에 의거 수행하는 차량 통신 방법에 있어서,

[0052] 기 정해진 타겟 통신 범위 내의 차량 단말의 노드 수로 입력된 현재 시점의 상태에 대해, 이동 통신망의 중앙 제어기에서 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값을 토대로 각 차량의 행동을 도출하는 단계; 및

[0053] 도출된 차량의 행동으로 각 차량 별 패킷 생성 시간의 상한치를 조정하고 조정된 패킷 생성 시간의 상한치로 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 단계를 포함하는 것을 일 특징으로 한다.

발명의 효과

[0054] 상기와 같은 본 발명에 따르면 심층 강화 학습을 수행하여 각 차량 밀도에 따라 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값을 타겟 통신 범위 내에 존재하는 적어도 하나의 차량 단말로 공유하고, 타겟 통신 범위 내에 존재하는 차량의 노드 수로 입력된 현재 시점 상태에 대해 훈련된 로컬 신경망의 $Q(s,a;\theta)$값의 행동으로 도출된 행동으로 패킷 생성 시간의 상한치를 도출하며 도출된 패킷 생성 시간의 상한치로 통신 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출할 수 있다.

[0055] 또한 훈련된 로컬 신경망의 $Q(s,a;\theta)$ 값의 행동으로 도출된 행동으로 패킷 생성 시간의 상한치를 도출하므로 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하는 연산 복잡도를 감소할 수 있고, 이에 경량 디바이스에 범용적으로 적용할 수 있다.

[0056] 또한 본 발명에 의거, 이동통신망의 타겟 신경망의 타겟값으로 타겟 PDR을 제어하여 패킷 생성 시간의 상한치를 조정함에 따라 조정된 패킷 생성 시간에 의거 C-V2X 기반의 차량 통신 시스템의 서비스 품질이 급격히 저하되는 현상을 방지할 수 있다.

도면의 간단한 설명

[0057] 본 명세서에서 첨부되는 다음의 도면들은 본 발명의 바람직한 실시 예를 예시하는 것이며, 후술하는 발명의 상세한 설명과 함께 본 발명의 기술사상을 더욱 이해시키는 역할을 하는 것이므로, 본 발명은 그러한 도면에 기재된 사항에만 한정되어 해석되어서는 아니된다.

도 1은 일 실시예에 적용되는 C-V2X 기반의 차량 통신 시스템의 구성도이다.

도 2는 도 1의 이동 통신망의 중앙 제어기의 세부 구성도이다.

도 3은 도 1의 차량 단말의 스케줄링을 보인 예시도이다.

도 4는 도 2의 중앙 제어기의 처리 알고리즘을 보인 전체 흐름도이다.

도 5는 도 1의 차량 통신망의 차량 단말의 세부 구성도이다.

도 6은 도 4의 차량 단말의 처리 알고리즘을 보인 흐름도이다.

도 7은 일 실시예의 DDQN과 DQN 과의 누적 리워드를 보인 예시도이다.

도 8은 도 8은 일 실시예의 DDQN 및 DQN의 평균 Q을 보인 예시도이다.

도 9는 일 실시예에 의거한 평균 패킷 전송비 PDR을 보인 예시도이다.

도 10은 일 실시예의 IPG를 보인 예시도들이다.

도 11은 일 실시예에 의거 DDQN과 DQN 각각의 학습 결과를 보인 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0058] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

[0059] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 발명에 대해 구체적으로 설명하기로 한다.

[0060] 본 발명에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0061] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.

[0062] 따라서, 구성요소들과 "부"들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 "부"들로 결합되거나 추가적인 구성요소들과 "부"들로 더 분리될 수 있다.

[0063] 아래에서는 첨부한 도면을 참고하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략한다.

[0064] 일 실시예가 적용되는 각각의 구성요소에 대해 임의 개를 임의의 적절한 구성으로 포함할 수도 있다. 일반적으로, 컴퓨팅 및 통신 시스템들은 광범위한 구성들로 나타나며, 도면은 본 개시의 범위를 어떤 특정 구성으로 한정하지 않는다. 도면은 본 특허 문서에서 개시된 다양한 특성들이 사용될 수 있는 하나의 동작 환경을 도시하고 있지만, 그러한 특성들은 어떤 다른 적절한 시스템에서 사용될 수도 있다.

[0065] 명세서 상의 설명에 있어서, 동작을 수행하는 주체는 차량의 간의 통신 시스템을 제어하는 프로세서일 수 있으며, 다른 한 예로써, 측정 및 제어과정을 컴퓨터에 의해 수행하는 프로그램이 기록된 기록매체 또는 이를 포함하는 장치일 수 있다.

[0066] 일 실시예는 C-V2X 기반의 의거 차량과 사물과의 통신하는 차량 통신 시스템에 있어, V2V 기반의 차량 통신 시스템의 분산 혼잡 제어(DCC: Distributed Congestion Control)의 패킷 전송 시간을 심층 강화 학습(Deep Reinforcement Learning: 이하 DRL로 약칭함)에 의거 조정하고 조정된 패킷 전송 시간으로 차량 간의 차량의 통신을 수행함에 따라, C-V2X 기반의 차량 통신 시스템의 서비스 품질을 향상시킬 수 있고, 조정된 패킷 전송 시간에 따른 통신 지연을 방지할 수 있다.

[0067] 도 1은 일 실시예의 C-V2X 기반의 차량 통신 시스템의 구성도로서, 도 1을 참조하면, 일 실시예는 차량과 중앙 제어기 간의 통신하는 이동통신망(S1)과 차량과 차량 간의 통신을 제어하는 차량 통신망(S2)을 포함한다.

[0068] 여기서, 이동통신망(S1)은 기 정해진 단위 도로 길이의 노드의 수를 현재 시점상태로 입력한 후 입력된 현재시점 상태에 대해 로컬 신경망으로 행동을 선택한 다음 심층 강화 학습으로 로컬 신경망을 훈련하는 중앙 제어기(1)를 포함할 수 있다.

[0069] 도 2는 도 1의 중앙 제어기(1)의 세부 구성도로서, 도 2를 참조하면, 중앙 제어기(1)는 현재 지점의 상태 을 입력으로 현재시점의 상태 s_t^{cc}, 행동 a_t^{cc}, 리워드 r_t^{cc} 및 다음시점의 상태 s_{t+1}^{cc}를 포함하는 트랜지션(transtion) exp을 재생 메모리(13)에 저장한다음 저장된 트렌지션을 이용해 로컬 신경망 및 타겟 신경망을 훈련하여 최대 리워드를 가지는 로컬 신경망을 출력하는 구성을 갖추며, 이에 중앙 제어기(1)는 로컬 신경망(11), 모델 구축부(12), 재생 메모리(13), 샘플링부(14), 타겟 신경망(15), 손실함수 연산부(16) 중 적어도 하나를 포함한다.

[0070] 즉, 로컬 신경망(11)은 현재 지점 t의 상태 s_t^{cc} 을 입력으로 심층 강화 학습을 통해 훈련된 행동 a_t^{cc}을 출력하고, 출력된 행동 a_t^{cc} 는 모델 구축부(12)로 전달된다.

[0071] 모델 구축부(12)는 전달받은 행동 a_t^{cc} 을 토대로 리워드 r_t^{cc} 및 다음 시점 t+1 상태 s_{t+1}^{cc}를 도출하고, 도출된 현재 시점 t의 상태 s_t^{cc}, 행동 a_t^{cc}, 리워드 r_t^{cc} 및 다음 시점 t+1의 상태 s_{t+1}^{cc}는 재생 메모리(13)로 전달된다.

[0072] 재생 메모리(13)는 현재 시점 t의 상태 s_t^{cc}, 행동 a_t^{cc}, 리워드 r_t^{cc} 및 다음 시점 t+1의 상태 s_{t+1}^{cc}를 포함하는 트랜지션(transtion) exp을 소정 위치에 저장하여 학습 모델을 구축한다.

[0073] 또한, 중앙 제어기(1)는 구축된 학습 모델에 의거 로컬 신경망(11) 및 타겟 신경망(15)을 훈련하는 기능을 수행하되, 재생 메모리(13)에 저장된 트랜지션 exp의 사이즈가 기 정해진 미니 배치 사이즈 j보다 큰 경우 샘플링부(14)에 의거 트랜지션 exp을 미니 배치 사이즈로 샘플링한 다음 샘플링된 트랜지션 exp의 현재 시점 t의 상태를 로컬 신경망(11)에 입력되고 다음 시점 t+1의 상태 s_{t+1}^{cc}를 타겟 신경망(15)에 제공한다. 여기서, 다음 시점 t+1은 현재 시점을 기준으로 샘플링 주기 T를 경과한 시점을 말한다.

[0074] 이때 로컬 신경망(11)은 미니배치 사이즈로 샘플링된 트랜지션 exp s_j^{cc}, a_j^{cc} 를 입력으로 상태, 설정된 행동, 및 가중치를 포함하는 $Q(s_j, a_j; \theta)$ 을 도출한다.

[0075] 여기서, 로컬 신경망(11)은 최대 리워드를 가지는 행동을 훈련하기 위해 $Q(s_j, a_j; \theta)$값을 손실함수 연산부(16)로 전달한다.

[0076] 한편, 타겟 신경망(15)은 s_{j+1}^{cc}를 입력으로 미니배치 사이즈로 샘플링된 트랜지션 exp로 정해진 타겟값 y_i을 출력하고 출력된 타겟값 y_i은 손실함수 연산부(16)로 전달된다. 여기서 타겟값 y_i은 조정된 패킷 생성 시간으로 인한 통신 지연을 방지하도록 설정된다.

[0077] 이에 손실함수 연산부(16)는 로컬 신경망(11)의 $Q(s, a; \theta)$ 값을 토대로 손실함수 $L(\Theta)$를 도출하고 도출된 손실함수 $L(\Theta)$가 설정된 y_j에 도달하기 위한 가중치 Θ를 도출하고 도출된 가중치 Θ로 로컬 신경망(11)를 업데이트하고, Θ를 복사하여 타겟 신경망(15)의 가중치 Θ^- 를 업데이트한다.

[0078] 이때 심층 강화 학습을 통해 훈련된 로컬 신경망(11)은 중앙 제어기(1)의 통신 범위 내의 적어도 하나의 차량 단말 VUE에 각각 공유한다.

[0079] 중앙 제어기(1)에 의거 차량 밀도에 따라 V2V 기반의 차량 통신망의 DCC의 패킷 생성 시간을 심층 강화 학습을 통해 패킷 생성 시간 τ_m^{gen} 을 조정하여 리워드가 최대인 로컬 신경망(11)를 훈련하는 일련의 과정을 도 1 및 도

2를 참조하여 설명한다.

[0080] C-V2X 기반의 차량 통신 시스템은 싱글 캐리어 주파수 분할 멀티 엑세스(SC-FDMA: single-carrier frequency-division multiple access)를 기반으로 차량 간의 통신을 수행하되, 이때 채널의 대역폭 10 및 20MHz가 될 수 있고, 각 채널은 1ms의 서브 프레임으로 분할되며, 서브프레임은 0.5ms의 타임 도메인과 180kHz의 주파수 도메인구성된 리소스 블록 쌍(RB: Resource Block)으로 통신이 수행된다.

[0081] C-V2X 기반의 차량 통신 시스템의 이동 통신망(S1)은 할당 요구된 단위 주파수 소스 유닛인 리소스 블록(RB: Resource Block)과 1ms 서브 프레임 내의 적어도 하나의 서브 채널을 포함하는 패킷 세트(이하 패킷으로 약칭함)를 방송한다.

[0082] C-V2X 기반의 차량 통신 시스템의 차량 통신망(S2)의 각 차량 단말 VUE은 SB-SPS(Sensing Based Semi-Persistent Scheduling)를 사용하여 서브 채널을 스케줄링한다.

[0083] 도 3은 차량 단말의 SB-SPS를 사용하여 스케줄링된 서브 채널을 보인 예시도로서, 도 3을 참조하면, 차량 단말(VUE)은 주어진 샘플 주기 T_{CAM}에서 n번째 서브프레임의 주파수 영역에서 c번째 서브 채널을 통해 패킷(CAM)을 방송한다. 여기서, n은 N = {n|n = 1, 2, ..., N}에 의거 정의된 서브 프레임의 인덱스이고, C = {c |c = 1, 2, ..., C}은 n째 서브 프레임의 가용 서브 채널이다. 차량 통신망(S2)에서의 M의 차량 단말은 M = {m|m= 1, 2, ..., M}로 정의된다. 각 차량 단말(VUE)은 기 정

[0084] 해진 윈도우 T_{sense}에서 현재 통신 중인 하위 채널을 제외한 나머지 하위 채널과 인접된 링크 채널의 수신신호 세기(S-RSSI : sidelink received signal strength indication)를 감지할 수 있다. 감지된 수신신호 세기에 의해 차량 단말(VUE)는 확률 P_k로 현재 통신 중인 서브 채널을 유지할 수 있고, 확률 $1-P_k$로 서브채널을 재선택할 수 있다.

[0085] 여기서 차량 단말(VUE)는 측정된 인접 링크 수신신호 세기의 하위 20%에서 서브 채널의 재선택을 위한 서브 채널 후보를 생성하고 T1와 T2 사이의 윈도우에서 서브 채널 후보 중 새로운 서브 채널을 무작위로 선택한다.

[0086] 또한 패킷 수신 성공은 신호대 간섭 잡음 비(SINR: signal-to-interference-plus-noise ratio)로 결정된다.

[0087] 일 례로, C = 100, T_{CAM} =100 및 T_{Sense} =1000 이라고 가정하면, m번째 차량 단말(VUE)이 n번째 서브 프레임으로 c 번째 서브 채널을 사용하여 패킷을 전송하면, k번째 차량 단말(VUE)의 SINR은 하기 식 1로 도출된다.

[0088] [식 1]

[0089]
$$\gamma_{m,k}[n,c] = \frac{S_{m,k}[n,c]}{\sigma^2 + I_{m,k}[n,c]}$$

[0090] 여기서, σ^2 는 잡음 전력이고, $S_{m,k}[n,c]$은 m번째 차량 단말(VUE)에서 방송된 패킷이 k번째 차량 단말(VUE)에서 수신되는 전력이며, 전력 $S_{m,k}[n,c]$은 하기 식 2로 나타낼 수 있다.

[0091] [식 2]

[0092]
$$S_{m,k}[n,c] = (P_{tx}G_{tx}G_{rx}) \cdot \phi_m[n] \cdot \psi_{m,k}$$

[0094] 여기서, P_{tx}는 전송 전력이고, G_{tx}는 전송 안테나 이득, G_{rx}는 수신 안테나 이득, 및 $\Psi_{m,k}$는 음영 및 경로 손실을 고려한 스케일 페이딩이다. $\Phi_m[n]$는 패킷의 존재 또는 부재를 나타내는 바이어리 코드로서, 하기 식 3으로 나타낼 수 있다.

[0095] [식 3]

[0096]
$$\phi_m[n] = \begin{cases} 1 & \tau_m^{gen} \le \tau_n \\ 0 & \tau_m^{gen} > \tau_n \end{cases}$$

[0097] 여기서, τ_n은 n번째 서브 프레임의 시간, τ_m^{gen} 은 m번째 송신 차량 단말의 패킷 생성 시간이다. 식 3의 바이너리 코드가 1인 경우 패킷 버퍼 내에 패킷의 존재를 의미하고, 바이너리 코드가 0인 경우 패킷 버퍼 내의 패킷의 부재를 의미한다.

[0098] 이에 m번째 송신 차량 단말과 동일한 서브 채널을 사용하는 수신 차량 단말(VUE)의 누적 간섭 전력 $I_{m,k}[n,c]$은 하기 식 4로 도출될 수 있다.

[0099] [식 4]

[0100] $$I_{m,k}[n,c] = \sum_{\substack{v \in \mathcal{V} \\ v \neq m}} (P_t G_t G_r) \cdot \phi_v[n] \cdot \psi_{v,k}$$

[0101] 여기서, $^{\mathcal{V}}$ 는 m번째 송신 차량 단말과 동일한 n번째 서브 프레임의 c번째 서브 채널을 통해 통신하는 차량 단말들이다.

[0102] 측정된 SINR $\gamma_{m,k}[n,c]$가 MCS(Modulation Coding Scheme)에 의해 결정된 임계 SINR γ_{th}보다 높으면 k번째 수신 차량 단말은 패킷을 올바르게 수신하고, 그렇지 않으면 수신되지 아니한다.

[0103] 이러한 차량 통신망의 서비스 품질은 패킷 전송비(PDR: Packet delivery ratio)로 나타낼 수 있다. 여기서 PDR은 수신 차량 단말로 방송된 총 패킷과 수신 성공한 패킷의 비율로 나타낸다.

[0104] m번째 송신 차량 단말이 n번째 서브 프레임의 c번째 서브 채널을 사용하여_패킷을 전송할 때 PDR은 하기 식 5로 도출될 수 있다.

[0105] [식 5]

[0106] $$\zeta_m[n,c,\beta] = \frac{Y_m[n,c,\beta]}{X_m[n,c,\beta]}$$

[0107] 여기서 $X_m[n,c,\beta]$은 n번째 서브 프레임의 c 번째 서브 채널로 β의 통신 범위의 수신 차량 단말로 전송되는 총 패킷 수이고, $Y_m[n,c,\beta]$은 n번째 서브 프레임의 c번째 서브 채널로 β의 통신 범위의 수신 차량 단말에서 수신 성공한 패킷의 수이다.

[0108] 이에 C-V2X 기반 차량 통신 시스템의 서비스 품질 QoS을 향상시키기 위한 적응적 분산 혼잡 제어 DCC의 문제 공식 P1은 하기 식 6에 의거 도출된다.

[0109] [식 6]

[0110] $$P1: \min_{\mathcal{U}} \sum_{\rho \in \mathcal{R}} |\zeta_{\rho,\beta}^{*} - \zeta_{\rho,\beta}^{avg}|,$$
$$\zeta_{\rho,\beta}^{avg} = \frac{\sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} \sum_{n \in \mathcal{N}} Y_m[n,c,\beta]}{\sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} \sum_{n \in \mathcal{N}} X_m[n,c,\beta]}$$
$$s.t. \quad C1: \mathcal{U} = \{u_1, \ldots, u_M\},$$
$$u_m \in \{\delta_1, \ldots, \delta_F\}, m \in \mathcal{M}$$

[0111] 여기서, $\zeta_{\rho,\beta}^{*}$는 차량 밀도 ρ인 차량 통신망(S2)의 타겟 통신 범위 β에서의 타겟 패킷 전송비(이하 PDR로 약칭함)이다. $\zeta_{\rho,\beta}^{avg}$는 차량 밀도 ρ인 차량 통신망(S2)의 타겟 통신 범위 β에서 평균 PDR이다.

[0112] 또한, 패킷 생성 시간 τ_m^{gen} 는 최소 전송 주기 100ms와 상한치 u_m 사이의 정수값이며, u는 M대차량 단말의 패킷 생성 시간 τ_m^{gen} 의 상한치 $\{u_1, \ldots, u_M\}$이고, 일 례로, u_m은 m번째 송신 차량 단말의 패킷 생성

시간 $\tau_m^{gen}(100 \leq \tau_m^{gen} \leq u_m)$의 상한치로, 집합 $\{\delta_1, \ldots, \delta_F\}$의 원소이다. δ번째 송신 차량 단말은 다수의 δ원소 중 하나를 u_m으로 선택하여 패킷 생성 시간 τ_m^{gen}를 조정한다. 일 례로, F=5인 경우 각 송신 차량 단말 VUE 1 ~ VUE m 각각의 패킷 생성 시간 τ_m^{gen}의 상한치 $u^1 \sim u^5$이고, 차량 단말 VUE m에 대한 상한치 u_m는 δ의 집합 $\{\delta_1, \ldots, \delta_F\}$={100, 200, 300, 400, 500}으로 중 하나로 도출된다.,

[0113] 식 6을 참조하면, 적응적 DCC의 문제 P1은 선택된 M개의 송신 차량 단말 M VUEs 중 하나인 m번째 송신 차량 단말은 VUE m의 패킷 생성 시간 τ_m^{gen}의 상한치 u_m로 통신하는 차량 통신 시스템의 타겟 PDR $\zeta_{\rho,\beta}^*$과 평균 PDR $\zeta_{\rho,\beta}^{avg}$간의 차의 최소화이다.

[0114] 이에 C-V2X 기반 차량 통신 시스템의 적응적 DCC의 문제 P1은 차량 밀도 ρ인 V2V 환경에서 M개의 각각의 송신 차량 단말이 최적의 패킷 생성 시간 τ_m^{gen}의 상한치를 선택하기 위한 문제이고 이러한 문제 P1의 해의 연산 복잡도는 패킷 생성 시간 τ_m^{gen}의 상한치의 범위와 M의 차량 단말의 수에 따라 증가될 수 있다. 이에 일 실시예는 적응적 DCC의 문제 P1의 해를 심층 강화 학습(DRL: Deep Reinforcement Learning) 알고리즘을 통해 훈련된 로컬 신경망(11)의 행동으로 패킷 생성 시간 τ_m^{gen}의 상한치를 조정함에 따라 도출하고, 이에 문제 P1의 해의 연산 복잡도를 감소할 수 있다.

[0115] 이에 심층 강화 학습(DRL)은 M개의 차량 단말의 차량 통신 환경에서 MDP(Markov Decision Processes)는 4개의 요소 <S,A,P,R>를 포함하고, 여기서, S는 상태 집합이고, A는 행동 집합이며, P는 상태 전이 확률로 $p(s,r|s,a') = \Pr(s_{t+1} = s'|s_t = s, a_t = a)$이고, R은 학습에 대한 리워드 집합이다. 현재 시점 t에서의 상태 $s_t \in S$에서 행동 $a_t \in A$을 수행하는 경우 MDP는 다음 시점 t+1에서의 상태 s_{t+1}와 다음 상태 s_{t+1}의 리워드 $r_t \in \mathrm{R}$이 제공될 수 있다. 이러한 MDP는 마지막 에피소드 T에 대한 상태 s_T에 도달할 때까지 계속된다.

[0116] 중앙 제어기(1)는 로컬 신경망(11) 및 모델 구축부(12)에 의거 상태-행동 쌍으로 구축된 학습 모델에 의거 입력된 상태 s에 대해 $Q(s,a) = \mathbb{E}_\pi[\sum_{i=t+1}^{T} \lambda^{i-t-1} r_t | s_t = s, a_t = a]$를 값을 출력한다.

[0117] 이에 로컬 신경망(11)은 확률 ϵ로 현재시점 t의 상태 s_t에 대한 랜덤한 행동 a_t를 선택할 수 있고, 확률 $1-\epsilon$로 $Q(s,a;\theta)$값이 최대인 행동 a_t를 선택할 수 있으며, 이에 행동 a_t는 하기 식 7로 나타낼 수 있다.

[0118] [식 7]

[0119] $$a_t = \arg\max_a Q(s_t, a; \theta)$$

[0120] 이 후 모델 구축부(12)는 로컬 신경망(11)의 행동 a_t으로 리워드 r_t와 다음시점 t+1의 상태 s_{t+1}를 선행문헌 1에 의거 도출하고, 도출된 $\{s_t, a_t, r_t, s_{t+1}\}$을 포함하는 트랜지션 *exp*를 재생 메모리 D(13)에 저장한다. 재생 메모리(13)에 저장된 트랜지션(Transition) *exp*는 미니 배치 사이즈 j보다 크면 샘플링부(14)는 미니 배치 사이즈로 트랜지션 *exp*를 샘플링한다.

[0121] 이러한 샘플링된 트랜지션 *exp*의 시점 j의 상태 s_j^{cc}는 로컬 신경망(11)에 전달되고, 다음 시점 j+1의 상태는 타겟 신경망(15)에 전달된다.

[0122] 그리고, 로컬 신경망(11)의 $Q(s,a;\theta)$ 값과 타겟 신경망(115)의 타겟값 y_i은 손실함수 연산부(16)로 전달된다.

[0123] 손실함수 연산부(16)의 로컬 신경망(11)의 훈련에 사용되는 가중치를 업데이트하기 위한 손실함수 $L(\theta)$는 하기 식 8로 나타낼 수 있다.

[0124] [식 8]

[0125] $$L(\theta) = \mathbb{E}[(y_j - Q(s_j, a_j; \theta))^2]$$

[0126] 여기서, 손실함수의 타겟값 y_i는 하기 식 9로 정의된다.

[0127] [식 9]

[0128] $$y_j = \begin{cases} r_j + \lambda \hat{Q}(s_{j+1}, \arg\max_{a'} Q(s_{j+1}, a'; \theta); \theta^-) & s_{j+1} \neq s_T \\ r_j & s_{j+1} = s_T \end{cases}$$

[0129] 여기서, λ는 감가율(discount factor)이고, 0과 1 사이의 값을 가진다. 로컬 신경망(11)의 $Q(s,a;\theta)$값의 가중치 Θ는 매 시간 t마다 업데이트되고, 타겟 신경망(15)의 $\hat{Q}(s,a;\theta^-)$ 값의 가중치 Θ^-는 매 주기 P마다 가중치 Θ를 복사하여 업데이트된다. 이에 로컬 신경망(11)은 가중치의 업데이트로 훈련하여 최대 리워드를 가지는 행동을 출력한다.

[0130] 일 실시예의 중앙 제어기(1)는 타겟 신경망(15)의 $\hat{Q}(s,a;\theta^-)$ 값을 사용하여 리워드가 급격하게 저하되는 현상을 방지하는 손실함수의 타겟값 y_i를 설정하고 설정된 타겟값 y_i에 도달하도록 로컬 신경망(11)의 가중치 Θ를 업데이트할 수 있다. 이러한 로컬 신경망(11)의 훈련으로 최대 리워드를 가지는 a_t을 출력한다.

[0131] 전술한 바와 같이 최대 리워드를 가지도록 훈련된 로컬 신경망(11)의 $Q(s,a;\theta)$값은 중앙 제어기 범위 내의 적어도 하나의 차량 단말에 공유한다.

[0132] 도 4는 중앙 제어기(1)의 처리 알고리즘을 보인 흐름도로서, 도 4를 참조하면, 단계 1 내지 단계3에서, 재생 메모리(13), 로컬 신경망(11)의 $Q(s,a;\theta)$ 값 및 타겟 신경망(15)의 $\hat{Q}(s,a;\theta^-)$을 초기화한 다음 단계 6 및 단계 7에서 중앙 제어기(1)는 기 설정된 확률 ϵ로 랜덤한 행동 a_t^{cc}을 선택할 수 있고, 확률 $1-\epsilon$로 가장 큰 $Q(s,a;\theta)$ 값을 가지는 행동 a_t^{cc}를 선택한다.

[0133] 이후 단계 8에서, 중앙 제어기(1)는 M대 차량 단말의 단위 거리(100m)내 차량수의 평균 값을 통해 행동 a_t^{cc}을 결정하고, 패킷 생성 시간 τ_m^{gen}을 조정한 다음 조정된 패킷 생성 시간으로 차량 간의 통신을 수행하고 이때 획득된 타겟 PDR과 평균 PDR의 차로 리워드 r_t^{cc}를 도출한다.

[0134] 단계 9에서 중앙 제어기(1)는 다음 시점 t+1의 상태 s_{t+1}를 연산하고 도출된 현재 시점 t의 상태 s_t^{cc}, 행동 a_t^{cc}, 리워드 r_t^{cc} 및 다음 시점 t+1의 상태 s_{t+1}를 포함하는 트랜지션 exp를 재생 메모리(13)에 저장한다.

[0135] 단계 10에서, 중앙 제어기(1)는 트랜지션 exp가 기 정해진 미니배치 사이즈 보다 크면 트랜지션 exp를 미니배치 사이즈로 샘플링을 수행한다.

[0136] 단계 11에서, 중앙 제어기(1)는 재생 메모리(13)에 저장된 트랜지션 exp또는 미니 배치 사이즈로 샘플링된 트랜지션 exp를 토대로 타겟값 y_i를 설정한다.

[0137] 그리고 단계 12에서, 중앙 제어기(1)는 로컬 신경망(11)의 Q값에 대한 경사 하강법을 수행하고 로컬 신경망(11)의 Q값의 가중치 Θ를 업데이트하고 가중치 Θ을 복사하여 타겟 신경망(15)의 가중치를 업데이트한다. 전술한

과정은 모든 훈련 에피소드 1~K의 각 스텝 1~T에 도달할 때 까지 반복 수행되며, 이에 로컬 신경망(11)은 심층 강화 학습 기반으로 훈련되며, 훈련된 로컬 신경망(11)은 각 차량 단말 VUE 1~VUE m에 공유된다.

[0138] 이하 로컬 신경망(11)의 공유하는 적어도 하나의 차량 단말은 로컬 신경망(11)으로 C-V2X 기반의 차량 통신 시스템의 서비스 품질 QoS를 도출하는 과정은 도 3을 참조하여 설명한다.

[0139] 도 5는 도 1에 도시된 차량 단말 VUE의 세부 구성도로서, 도 5를 참조하면, 차량 단말 VUE 1~ VUE m(2) 각각은 기 정해진 단위 거리(100m) 내의 차량의 수 즉, 노드의 수 $\eta_t^1 \sim \eta_t^m$ 인 현재 시점 t의 상태를 공유된 각 차량의 로컬 신경망(11)에 각각 입력한 후 훈련된 로컬 신경망(11)에서 행동 $a_t^1 \sim a_t^m$를 도출하고, 도출된 행동 $a_t^1 \sim a_t^m$ 토대로 패킷 생성 시간 τ_m^{gen} 을 조정한 다음 조정된 패킷 재생 시간 τ_m^{gen} 으로 차량 간의 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질 QoS를 도출하며, 이에 각 차량 단말 VUE1 ~ VUE m(2)은 상태 획득부(21), 로컬 신경망 구동부(22) 및 분산 실행기(23)를 포함할 수 있다.

[0140] 여기서, 상태 획득부(21)는 기 정해진 단위 거리(100m) 내의 차량의 수 즉, 노드의 수 $\eta_t^1 \sim \eta_t^m$를 각각 수집하고, 수집된 각 차량의 노드의 수 $\eta_t^1 \sim \eta_t^m$ 를 각 차량의 현재 시점 t의 상태로 공유된 로컬 신경망 구동부(22)에 입력한다.

[0141] 로컬 신경망 구동부(22)는 훈련된 로컬 신경망의 현재 시점의 상태를 입력으로 학습하여 행동 $a_t^1 \sim a_t^m$를 도출하며, 도출된 행동 $a_t^1 \sim a_t^m$ 는 분산 실행기(23)로 전달된다. 로컬 신경망(11)에서 행동 $a_t^1 \sim a_t^m$ 각각은 패킷 생성 시간 τ_m^{gen} 의 상한치로서 행동 a_t^m 과 100 ms내의 상수값으로 조정된다.

[0142] 분산 실행기(23)는 조정된 패킷 재생 시간 τ_m^{gen} 의 상한치로 차량 간의 통신을 수행하여 C-V2X 기반의 차량 통신 시스템의 서비스 품질 QoS를 도출할 수 있다.

[0143] 여기서 C-V2X 기반의 차량 통신 시스템의 서비스 품질 QoS은 조정된 패킷 생성 시간 τ_m^{gen} 의 상한치로 차량 간의 통신을 통해 획득된 송신 차량 단말에서 전송된 총 패킷의 수와 수신 차량 단말에서 수신된 패킷 수의 비로 패킷 전송비(PDR: Packet delivery ratio)로 도출될 수 있다.

[0144] 이에 일 실시예의 C-V2X 기반의 차량 통신 시스템의 각 차량 단말 VUE 1~VUE m(21)은 V2V 기반의 차량 통신 시스템의 분산 혼잡 제어(DCC)의 속도 제어를 통해 훈련된 로컬 신경망(11)의 패킷 생성 시간 τ_m^{gen} 으로 상호 통신하는 송신 차량 단말의 총 패킷 수과 수신 차량 단말의 수신 패킷 수의 비로 도출된 PDR로 차량 통신망의 서비스 품질 QoS를 도출할 수 있다.

[0145] 도 6은 도 5의 차량 단말(2)에서 수행되는 처리 알고리즘으로 도 6을 참조하면, 차량 단말에서 노드 수 정보를 수집하여 집합적으로 제어하는 MEC(Multi- Access Edge Computing)로 구비될 수 있다. 단계 19에서, 차량 단말(2)의 상태 획득부(21)는 해당 커버리지에서 차량 단말로부터 100m 이내의 노드 수를 수집하고, 수집된 노드 수를 평균하여 차량 통신망의 혼잡 수준을 추정하고 추정된 상태는 하기 식 10으로 나타낼 수 있다.

[0146] [식 10]

[0147]
$$s_t^{cc} = \frac{1}{M} \sum_{m \in \mathcal{M}} \eta_t^m$$

[0148] 여기서, η_t^m 는 현재 시점 t에서 m번째 차량 단말의 100m 이내의 차량 수 즉, 노드 수이고, 상태 s_t^{cc} 는 100m 이내의 평균 노드 수로 정해진다.

[0149] 이러한 획득된 현재시점의 상태 s_t^{cc} 는 로컬 신경망 구동부(22)로 전달된다.

[0150] 단계 20에서, 로컬 신경망 구동부(22)는 현재시점의 상태 s_t^{cc} 를 입력으로 공유된 각 차량의 로컬 신경망(11)에 의거 각 차량의 행동 a_t^{cc} 을 도출하고 도출된 각 차량의 행동 a_t^{cc} 은 분산 실행기(23)로 전달된다.

[0151] 여기서, 각 차량 단말의 행동 a_t^{cc} 은 하기 식 11로 나타낼 수 있다.

[0152] [식 11]

[0153] $$a_t^{cc} \in \{\delta_1, ..., \delta_f\}$$

[0154] 이때 분산 제어기(23)는 적응적 DCC의 문제 *P1*에서 타겟 PDR $\zeta_{\rho,\beta}^*$ 로 향상시키기 위해, 평균 PDR $\zeta_{\rho,\beta}^{avg}$ 동작을 조정하고, 이에 단계 21에서 분산 제어기(23)는 각 차량 단말 VUE의 패킷 생성 시간 τ_m^{gen} 을 상한치 δ_f 와 100ms 사이에서 패킷 생성 시간 τ_m^{gen} τ_m^{gen} 을 조정한다.

[0155] 단계 17 및 18에 의거 이러한 패킷 생성 시간에 대해 n=1, · · · ,N의 서브 프레임과 VUE=1, · · · ,M의 각 차량 단말에 대해 반복하여 수행한다.

[0156] 한편 분산 실행기(23)는 조정된 패킷 생성 시간 τ_m^{gen} 으로 차량 간의 통신을 수행하여 타겟 PDR $\zeta_{\rho,\beta}^*$과 평균 PDR $\zeta_{\rho,\beta}^{avg}$ 의 차의 최소화를 도출한다. 이때 분산 실행기(23)는 선행문헌 2에 의거 알고 있는 평균 PDR $\zeta_{\rho,\beta}^{avg}$으로 리워드를 도출하며 도출된 리워드는 하기 식 12로 나타낼 수 있다.

[0157] [식 12]

[0158] $$r_t^{cc} = -|\zeta_{\rho,\beta}^* - \zeta_{\rho,\beta}^{avg}|$$

[0159] 식 12에 의거, 분산 실행기(23)는 리워드값을 0으로 조정하여 타겟 PDR $\zeta_{\rho,\beta}^*$과 평균 PDR $\zeta_{\rho,\beta}^{avg}$ 차의 최소화를 통해 PDR을 도출한다.

[0160] 따라서 차량 단말은 중앙 제어기의 학습 결과로 도출된 로컬 정보만을 이용하므로 처리 시간을 단축할 수 있으며, 환경이 변경되어 타겟 PDR $\zeta_{\rho,\beta}^*$ 을 충족하지 아니한 경우 중앙 제어기는 변경된 차량 밀도에 따라 차량 단말의 패킷 생성 시간을 조정하고 조정된 패킷 생성 시간 을 할당할 수 있다.

[0161] 도 7은 일 실시예의 DDQN과 DQN 과의 누적 리워드를 보인 예시도로서, 도 7의 (a)를 참조하면, DQN와 DDQN의 타겟 PDR $\zeta_{\rho,\beta}^*$이 각각 90%와 85%로 DDQN의 리워드가 DQN의 리워드 보다 안정적이고 학습 결과가 향상됨을 알 수 있다. 또한 DQN은 훈련 과정 동안에 리워드가 급격하게 감소하는 반면, DDQN의 리워드는 급격하게 저하됨이 없이 안정적으로 훈련됨을 알 수 있다.

[0162] 또한 도 7의 (b)를 참조하면, 1000개의 훈련 에피소드 동안 동일한 타겟 PDR $\zeta_{\rho,\beta}^*$ 에 대해 DQN의 리워드는 DDQN의 리워드 보다 낮음을 알 수 있다.

[0163] 도 8은 일 실시예의 DDQN 및 DQN의 평균 Q을 보인 예시도로서, 도 8의 (a)를 참조하면, DQN의 평균 Q값은 같은 타겟 PDR $\zeta_{\rho,\beta}^*$에 대해 DDQN의 평균 Q값 보다 높음을 알 수 있고, 차량 밀도 ρ인 차량 네트워크의 타겟 통신 범위 β에서 평균 PDR $\zeta_{\rho,\beta}^{avg}$ = 85%의 DDQN의 평균 Q 값은 150 에피소드 인접에서 피크임을 알 수 있다. 또한 동일한 동작점에서 평균 PDR $\zeta_{\rho,\beta}^{avg}$ = 85%의 DQN의 누적 리워드는 도 3에 도시된 바와 같이 급격하게 떨어짐을 확인할 수 있다.

[0164] 즉, 누적 리워드의 급격한 감소 및 평균 Q 값의 급격한 상승은 DQN의 과추정을 나타낸다.

[0165] 도 8의 (b)를 참조하면, DQN은 DDQN에 비해 무선 자원 할당, 차량 위치, 차량 이동 등의 V2V 환경에서 노이즈로 인해 불안정한 Q값 추정을 나타내고 이에 리워드에 부정적임을 알 수 있다.

[0166] 도 9는 일 실시예에 의거한 평균 패킷 전송비 PDR을 보인 예시도로서, 도 9의 (a)는 (b)보다 더 좁은 범위의 V2V 통신을 평가하기 때문에 (b)보다 더 높은 평균 PDR 결과를 보여준다. 도 9의 (a)를 참조하면, ETSI DCC의 PDR은 100ms의 최소 전송 속도에 대해 하한의 PDR 보다 평균 15% 향상됨을 확인할 수 있다. 또한, DCC는 차량 밀도가 증가함에 따라 더 낮은 전력을 사용하므로, 차량 밀도 ρ = 200까지 하한의 PDR 보다 보다 낮음을 확인할 수 있으나 일 실시 예에 의거한 DDQN의 PDR은 DQN의 PDR 보다 12% 우수함을 알 수 있다. 일 실시예의 DDQN과 DQN 각각은 상한 임계치와 하한 임계치 사이의 타겟 서비스 품질 $\zeta_{\rho,\beta}^{*}$ 의 90%와 85%에 도달됨을 알 수 있다.

[0167] 또한 도 9의 (B)를 참조하면, 일 실시예의 DQN 및 DDQN 각각의 PDR은 일반적인 DCC보다 타겟 서비스 품질 $\zeta_{\rho,\beta}^{*}$ 에 근접됨을 알 수 있다.

[0168] 도 10은 일 실시예의 IPG를 보인 예시도들로서, 도 10을 참조하면, 각 차량의 밀도에서 50초 동안 평가된 IPG는 90% CDF를 가진 값임을 알 수 있다. 도 10의 (a) 및 도 10를 참조하면, 가장 좋은 PDR인 상한 임계치는 약 500ms의 가장 큰 IPG이고, PDR의 하한 임계치는 약 100ms에서 가장 최소 IPG 값이 나타남을 알 수 있습니다.

[0169] 또한 ETSI DCC는 저밀도(ρ = 120)에서 일 실시예의 DDQN 보다 더 큰 IPG가 출력되고, 고밀도에서 더 낮은 IPG가 도출됨을 알 수 있다.

[0170] 또한 도 10의 (b)를 참조하면, 차량 밀도 별로 QoS에 따른 전송률을 조절하여 타겟 PDR를 만족할 수 있다. (b)의 결과는 (a)보다 더 넓은 범위에 대해 평가되기 때문에 최대 IPG는 600ms로 증가하고, 기존 DCC 알고리즘은 하한보다 다소 큰 IPG를 나타내지만 타겟 PDR은 만족하지 못한다. 반면 일 실시예는 (a)보다 낮은 밀도에서 높은 IPG를 보여 더 넓은 커버리지에서 타겟 PDR가 만족됨을 알 수 있다.

[0171] 도 11은 일 실시예에 의거 DDQN에 의해 학습된 속도제어 정책과 DQN에 의해 학습된 속도 제어 정책을 보인 예시도로서, 도 11의 (a)를 참조하면, 속도 제어 정책은 행동 공간의 요소를 고르게 사용됨을 알 수 있다. 속도 제어 정책은 타겟 통신 범위 β 및 타겟 PDR $\zeta_{\rho,\beta}^{*}$ 을 충족시키기 위해 수행된다. 도 11의 (b)을 참조하면, 더 넓은 타겟 통신 범위 β의 범위와 더 높은 타겟 PDR $\zeta_{\rho,\beta}^{*}$ 이 필요할 때 낮은 상태 입력에 대해 패킷 생성 시간 τ_m^{gen} 은 높게 설정되어야 함을 알 수 있다.

[0172] 이에 일 실시예는, V2V 기반의 차량 통신 시스템의 이동통신망의 DDC에 대해 심층 강화 학습을 수행하여 훈련된 로컬 신경망을 다수의 차량 단말로 공유하고, 각 차량 밀도에 따라 훈련된 로컬 신경망으로 패킷 생성 시간을 조정함에 따라 C-V2X 기반의 차량 통신 시스템의 서비스 품질을 도출하기 위한 연산 복잡도를 감소할 수 있다.

[0173] 이상으로 본 발명의 기술적 사상을 예시하기 위한 바람직한 실시예와 관련하여 설명하고 도시하였지만, 본 발명은 이와 같이 도시되고 설명된 그대로의 구성 및 작용에만 국한되는 것이 아니며, 기술적 사상의 범주를 일탈함이 없이 본 발명에 대해 다수의 변경 및 수정이 가능함을 당업자들은 잘 이해할 수 있을 것이다. 따라서 그러한 모든 적절한 변경 및 수정과 균등물들도 본 발명의 범위에 속하는 것으로 간주되어야 할 것이다.

부호의 설명

[0174] S1 : 이동 통신망

S2 : 차량 통신망

1 : 중앙 제어기

11 : 로컬 신경망

12 : 모델 구축부

13 : 재생 메모리

14 : 샘플링부

15 : 타겟 신경망

16 : 손실함수 연산부

2 : 차량 단말

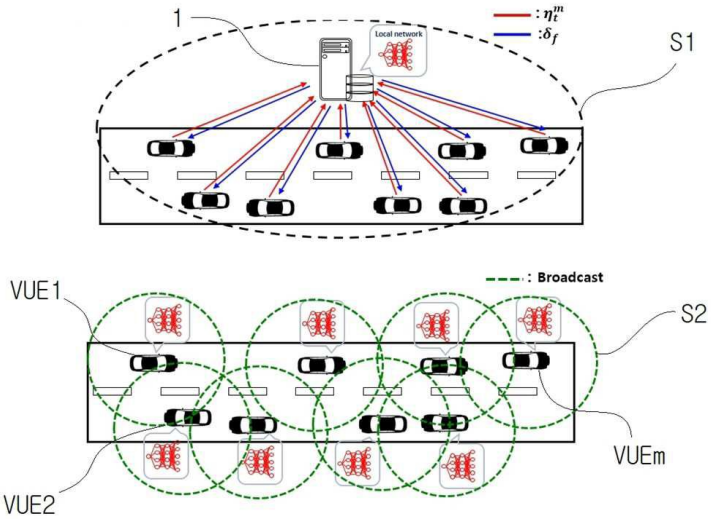
21 : 상태 획득부

22 : 로컬 신경망 구동부

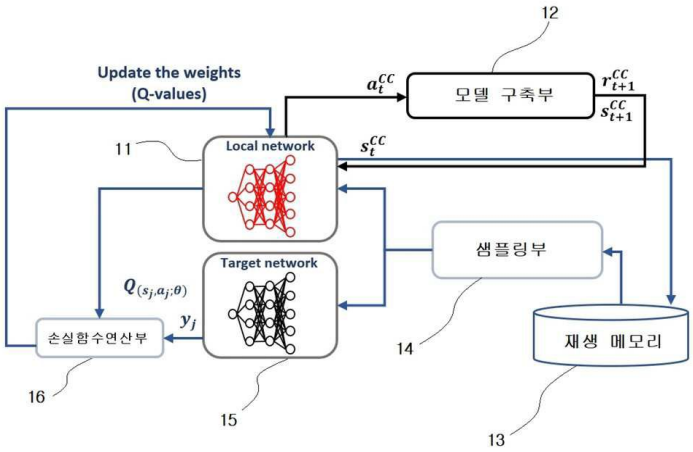
23 : 분산 실행기

도면

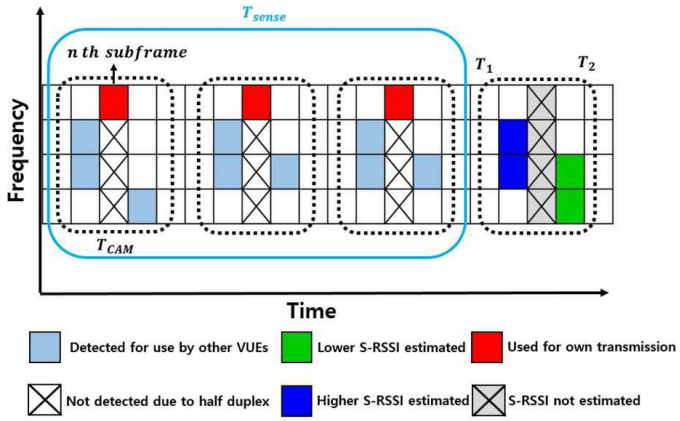
도면1



도면2



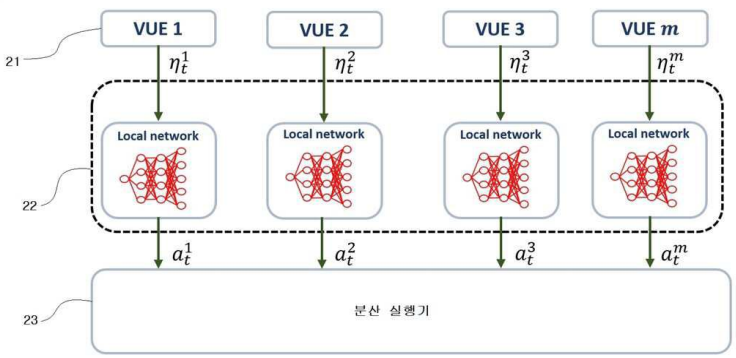
도면3



도면4

1: Initialize replay memory $\mathcal{D}$
2: Initialize local network $Q(s,a;\theta)$ with random weights θ
3: Initialize target network $\hat{Q}(s,a;\theta^{-})$ with weights $\theta^{-}=\theta$
4: **for** Episode =1,$\cdots$,K **do**
5: **for** Step=1,$\cdots$,T **do**
6: With probability select a random action a_t^{cc}
7: Otherwise select $a_t^{cc}=\arg\max_{a^{cc}} Q(s_t^{cc},a^{cc};\theta)$
8: Control packet generation time $\tau_{gen}^{m}=\{\tau_{gen}^{m}|100\leq \tau_{gen}^{m}\leq a_t^{cc}\}$ and observe reward $r_t^{cc}=-|\zeta_{\rho,\beta}^{*}-\zeta_{\rho,\beta}^{avg}|$
9: Calculate s_{t+1} and store experience $exp=\{s_t^{cc},a_t^{cc},r_t^{cc},s_{t+1}^{cc}\}$ in $\mathcal{D}$
10: Sample random mini-batch from $\mathcal{D}$
11: Set y_j via (9) or (10)
12: Perform gradient descent via (8) and update θ
13: Every P steps reset $\theta^{-}=\theta$
14: **end for**
15: **end for**

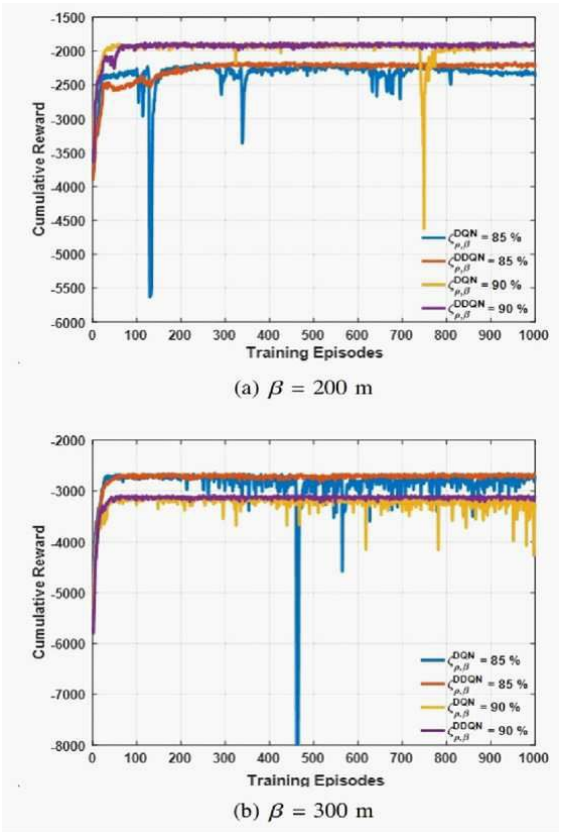
도면5



도면6

16: Share the optimal single-policy $Q(s, a; \theta)$ trained above to M VUEs
17: **for** Subframe $n = 1, \cdots, N$ **do**
18: **for** VUE $= 1, \cdots, M$ **do**
19: Set the number of nodes within 100 m η_t^m to state s_t^m
20: Select action $a_t^m = \arg\max_{a^m} Q(s_t^m, a^m; \theta)$
21: Adjust packet generation time $\tau_{gen}^m = \{\tau_{gen}^m | 100 \leq \tau_{gen}^m \leq a_t^m\}$
22: **end for**
23: **end for**

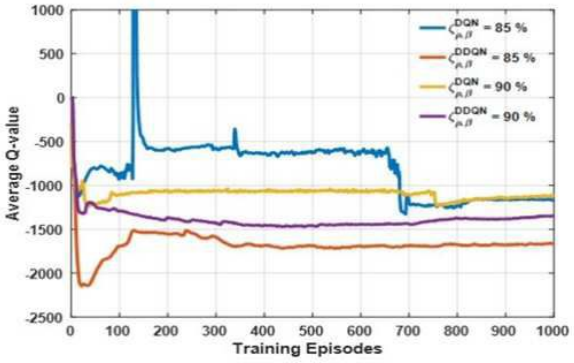
도면7



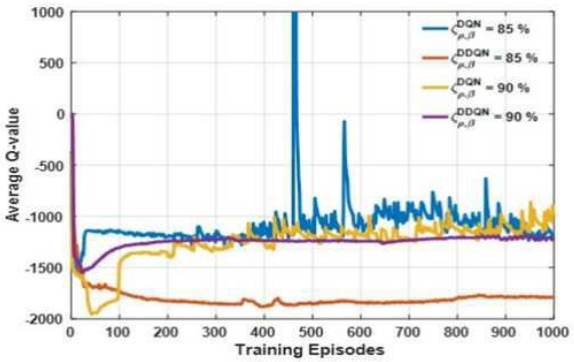
(a) β = 200 m

(b) β = 300 m

도면8

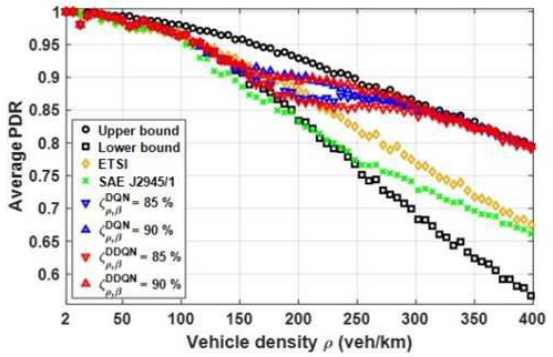


(a) β = 200 m

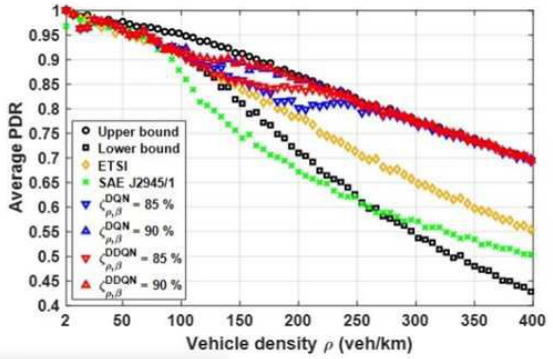


(b) β = 300 m

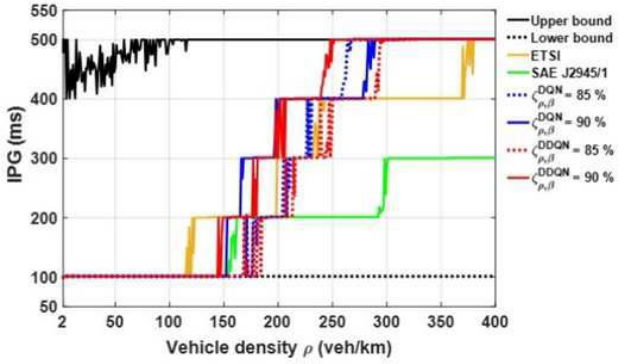
도면9



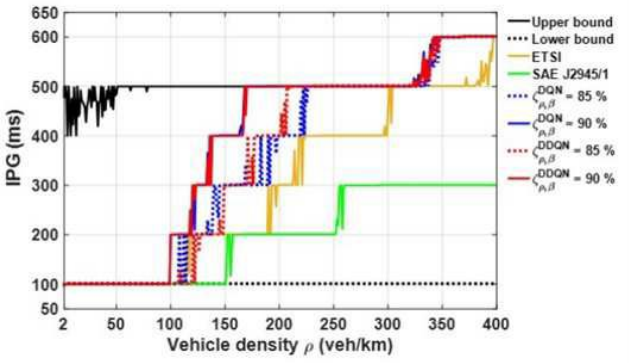
(a) β = 200 m



도면10

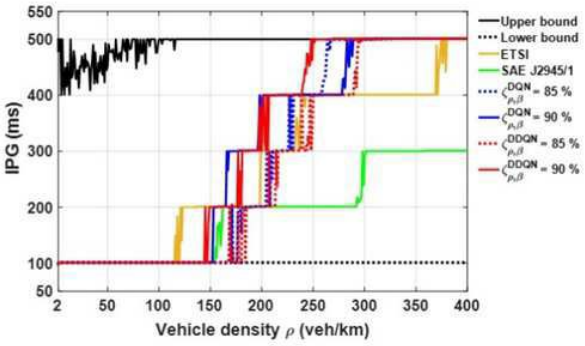


(a) β = 200 m

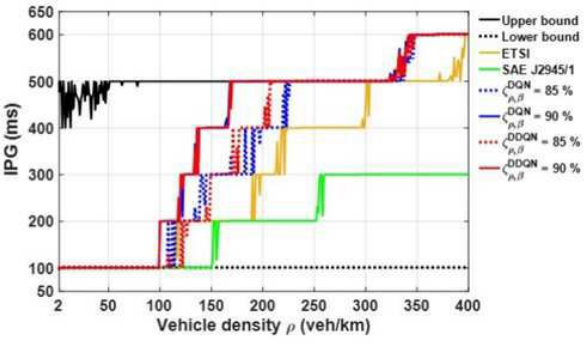


(b) β = 300 m

도면11



(a) β = 200 m



(b) β = 300 m