



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0041509
(43) 공개일자 2025년03월25일

(51) 국제특허분류(Int. Cl.)

H04W 72/02 (2009.01) G06N 3/092 (2023.01)
G06N 3/098 (2023.01) H04W 4/40 (2018.01)
H04W 72/25 (2023.01) H04W 92/18 (2009.01)

(52) CPC특허분류

H04W 72/02 (2023.01)
G06N 3/092 (2023.01)

(21) 출원번호 10-2023-0124321

(22) 출원일자 2023년09월18일

심사청구일자 2023년09월18일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대학교)

(72) 발명자

김덕경

서울특별시 서초구 바우피로7길 51, 101동 503호
(우면동, 대림아파트)

이인성

인천광역시 미추홀구 인하로 95, 401호(용현동)

(74) 대리인

양성보

전체 청구항 수 : 총 8 항

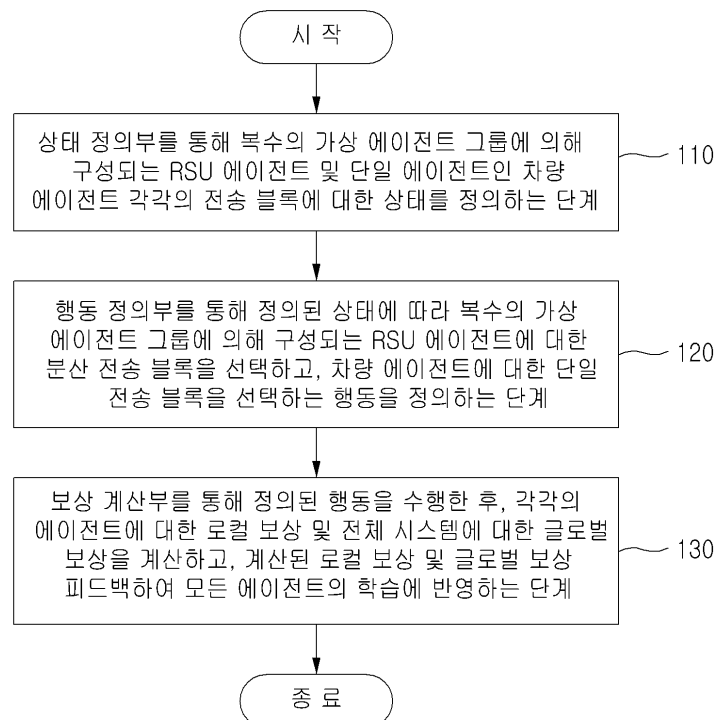
(54) 발명의 명칭 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법 및 시스템

(57) 요약

V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법 및 시스템이 제시된다. 본 발명에서 제안하는 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템은 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록

(뒷면에 계속)

대표도 - 도1



록에 대한 상태를 정의하는 상태 정의부 -상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 됨-, 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 행동 정의부 및 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 보상 계산부를 포함한다.

(52) CPC특허분류

G06N 3/098 (2023.01)

H04W 4/40 (2020.05)

H04W 72/25 (2023.01)

H04W 92/18 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711193423

과제번호 2022-0-00521-002

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원

연구사업명 방송통신산업기술개발사업

연구과제명 [Ezbaro] 레벨4 이상 자율주행 서비스 지원을 위한 5G-NR-V2X 통신 기술개발(2차년도)

기 여 율 1/1

과제수행기관명 대보정보통신

연구기간 2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 상태 정의부 -상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 됨-;

상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 행동 정의부; 및

상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 보상 계산부

를 포함하는 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템.

청구항 2

제1항에 있어서,

상기 상태 정의부는,

상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트의 각각의 에이전트는 단일 전송 블록을 선택하고,

상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우 간섭이 발생하므로, 전송 블록 재할당을 위해 동일한 전송 블록을 사용하는지 여부를 판단하여 상태를 정의하는

분산형 다중 에이전트 강화학습 기반 자원 할당 시스템.

청구항 3

제1항에 있어서,

상기 행동 정의부는,

상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우, 간섭을 제거하기 위해 해당 차량 에이전트의 전송 블록을 재할당하도록 행동을 정의하고, 각각의 차량 에이전트는 자율적인 의사 결정을 위해 개별 강화학습을 사용하며,

자원 소비 및 교환 지연을 최소화하기 위해 분산형 강화학습 모델을 사용하여 학습함으로써, 각각의 차량 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내리는

분산형 다중 에이전트 강화학습 기반 자원 할당 시스템.

청구항 4

제1항에 있어서,

상기 보상 계산부는,

제한된 공유 자원에 대한 자원 할당을 위해 MARL(Multi-Agent Reinforcement Learning) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트에 대한 개별 보상을 계산하고, 이후 우선순위 가중치를 적용하여 가중치가 부여된 글로벌 보상을 계산하며, 각각의 차량 에이전트가 협력적으로 행동하여 개별 성능을 최적화하며, 상기 강화학습 모델 학습을 위한 피드백으로 제공되는

분산형 다중 에이전트 강화학습 기반 자원 할당 시스템.

청구항 5

상태 정의부를 통해 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 단계 -상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 됨-;

행동 정의부를 통해 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 단계; 및

보상 계산부를 통해 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 단계

를 포함하는 분산형 다중 에이전트 강화학습 기반 자원 할당 방법.

청구항 6

제5항에 있어서,

상기 상태 정의부를 통해 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 단계는,

상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트의 각각의 에이전트가 단일 전송 블록을 선택하고,

상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우 간섭이 발생하므로, 전송 블록 재할당을 위해 동일한 전송 블록을 사용하는지 여부를 판단하여 상태를 정의하는

분산형 다중 에이전트 강화학습 기반 자원 할당 방법.

청구항 7

제5항에 있어서,

상기 행동 정의부를 통해 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 단계는,

상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우, 간섭을 제거하기 위해 해당 차량 에이전트의 전송 블록을 재할당하도록 행동을 정의하고, 각각의 차량 에이전트는 자율적인 의사 결정을 위해 개별 강화학습을 사용하며,

자원 소비 및 교환 지연을 최소화하기 위해 분산형 강화학습 모델을 사용하여 학습함으로써, 각각의 차량 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내리는

분산형 다중 에이전트 강화학습 기반 자원 할당 방법.

청구항 8

제5항에 있어서,

상기 보상 계산부를 통해 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 단계는,

제한된 공유 자원에 대한 자원 할당을 위해 MARL(Multi-Agent Reinforcement Learning) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트에 대한 개별 보상을 계산하고, 이후 우선순위 가중치를 적용하여 가중치가 부여된 글로벌 보상을 계산하며, 각각의 차량 에이전트가 협력적으로 행동하여 개별 성능을 최적화하며, 상기 강화학습 모델 학습을 위한 피드백으로 제공되는

분산형 다중 에이전트 강화학습 기반 자원 할당 방법.

발명의 설명

기술 분야

[0001] 본 발명은 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 자율 주행은 도로 안전, 교통 효율성 및 환경 지속 가능성을 향상시키는 해결책으로 인식되고 있다. 현대의 자율 주행 차량에는 카메라, 레이더/LiDAR, ADAS(Advanced Driver Assistance Systems) 등의 첨단 센서가 탑재되어 있으며, 이러한 센서는 자율 주행 기능을 가능하게 하고 도로 안전을 보장하는 데 중추적인 역할을 하지만 악천후, 시인성 저하, 시야 방해 등의 영향을 받을 수 있다. 따라서, EU(European Commission)은 도로 사용자와 교통 관리자 간의 정보 교환 및 조정을 촉진하기 위해 다양한 활용 사례를 배치하고 교통 시스템을 개선하기 위한 고급 메커니즘을 도입하는 데 핵심적인 역할을 해 온 C-ITS(Cooperative Intelligent Transport Systems)를 도입했다. 이러한 발전 중에서 V2V(Vehicle-to-Vehicle), V2I(Vehicle-to-Infrastructure), V2P(Vehicle-to-Pedestrian), V2N(Vehicle-to-Network) 통신 등을 포함하는 첨단 기술이 부상하고 있다. 특히, 3GPP(3rd Generation Partnership Project) Rel. 16에서 소개된 5G NR(New Radio) 기반 V2X는 이전 모델에 비해 더 넓은 커버리지와 향상된 서비스 품질(QoS)을 제공한다.

[0003] V2X 통신은 유니캐스트, 그룹캐스트, 방송 등 다양한 커뮤니케이션 캐스트 유형을 포함하고 있고, 이들 유형은 특정한 커뮤니케이션 목표와 설정에 부합한다. 유니캐스트와 그룹캐스트는 특정한 시나리오의 도로 안전성을 높이기 위해 지역적으로 활용되고 있으며, 이에 반해 방송은 종합적인 도로 안전을 위해 광범위한 지역에 서비스를 제공하고 있다. V2X 방송 통신을 통해 도로 안전성을 강화하는 데 중점을 두고 있다. 다양한 시나리오에 부합하는 것은 목적에 부합하는 특정한 형식과 크기의 메시지를 전송하는 것이다. 이중 트래픽 모델을 통합하는 것은 진정한 V2X 통신을 달성하기 위해 매우 중요하다.

[0004] 3GPP는 두 가지 자원 할당 모드, 즉 모드 1 및 모드 2를 도입한다. 모드 1에서, 기지국(Base Station; BS)은 Uu 인터페이스를 통해 사이드링크(Sidelink; SL)를 위한 사용자 장치(User Equipment; UE)에 무선 자원을 직접 할당한다. UE는 PUCCH(Physical Uplink Control Channel)를 이용하여 BS에 스케줄링 요청(Scheduling Request; SR)을 보내고, BS는 PDCCH(Physical Downlink Control Channel)을 통해 자원을 특정한다. 반대로, 모드 2에서, UE는 감지된 수신 신호 세기 지시자(Received Signal Strength Indicators; RSSI) 및 다른 UE의 제1 사이드링크 채널 정보(Sidelink Channel Information; SCI) 정보를 이용하여 SL 자원을 자율적으로 선택한다. UE는 사용되지 않은 자원과 예약되지 않은 자원을 정렬하고, 필요한 양의 자원을 랜덤하게 선택하여 반-영구적 방식으로 정의된 사이클 수 동안 반복적으로 재사용한다. 이러한 장점에도 불구하고, 두 모드는 모두 고유한 한계를 가지고 있다. 모드 1은 BS의 커버리지에 의해 지리적으로 제한되어 그 범위를 제한한다. 모드 2에서, 성능은 임의의 자원 선택, 불확실성 및 잠재적인 간섭을 겪는다. 따라서, 적절한 자원 할당 방안은 V2X 환경에 필수적이다.

[0005] RSU(Road Side Unit)는 중요한 인프라 역할을 하며, 확장된 커버리지, 사각지대 감지 및 멀티 홉 릴레이 기능을 통해 통신 성능을 증폭시킨다. RSU는 도로 정보를 주변 차량에 직접 방송함으로써 특히 도로 안전을 향상시킨다. RSU는 멀티 홉 릴레이 기능도 수행하여 V2X 시스템에서 통신 개선, 커버리지 확장 및 성능 향상의 기회를 유도한다. LiDAR 및 카메라와 같은 다양한 센서를 장착한 Smart RSU의 개념은 포괄적인 도로 데이터를 수집하기 위해 적극적으로 탐구되고 있다. 또한 Smart RSU는 작업을 오프로드하고 MEC(Multi-Access Edge Computing) 기술을 활용하여 사용자 계산 부하를 완화하고 전반적인 경험을 향상시킬 수 있다.

[0006] 강화 학습(Reinforcement Learning; RL)은 기존의 수학적 모델이 어려워하는 복잡한 작업과 요구 사항을 해결하는 데 핵심적인 이점을 제공한다. RL 모델은 다양한 요소를 설명하고 그에 따라 작업을 최적화할 수 있으며, 동적이고 불확실한 설정에서 유연함을 입증한다. 이 적응성은 일반적인 모델이 흔들리는 빠르게 변화하는 V2X 환경에서 중요한 다중 목표에 대한 문제에 적합하다. 특히 분산형 다중 에이전트(Multi-Agent; MA) RL에서 각 에이전트(예를 들어, 차량 또는 RSU)는 V2X 설정에서 실제 자원 할당 동작을 미러링하여 로컬 관찰을 기반으로 독립적으로 동작한다.

선행기술문헌

비특허문헌

- [0007] (비특허문헌 0001) L. Busoniu, R. Babuska, and S. B. De, "Multi-agent reinforcement learning: An overview." Chapter 7 in Innovations in Multi-Agent Systems and Applications-1 (D. Srinivasan and L.C. Jain, eds.), vol. 310 of Studies in Computational Intelligence, Berlin, Germany: Springer, pp. 183-221, 2010.

발명의 내용

해결하려는 과제

- [0008] 본 발명이 이루고자 하는 기술적 과제는 RSU와 차량이 모두 공유 자원 풀을 사용하여 다양한 크기의 패킷을 전송하고, 분산형 강화학습 모델을 사용하며, 각 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정하여 자원 소비 및 교환 지연을 최소화하기 위한 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법 및 시스템을 제공하는데 있다.

과제의 해결 수단

- [0009] 일 측면에 있어서, 본 발명에서 제안하는 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템은 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 상태 정의부 -상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 됨-, 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 행동 정의부 및 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 보상 계산부를 포함한다.
- [0010] 상기 상태 정의부는 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트의 각각의 에이전트는 단일 전송 블록을 선택하고, 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우 간섭이 발생하므로, 전송 블록 재할당을 위해 동일한 전송 블록을 사용하는지 여부를 판단하여 상태를 정의한다.
- [0011] 상기 행동 정의부는 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우, 간섭을 제거하기 위해 해당 차량 에이전트의 전송 블록을 재할당하도록 행동을 정의하고, 각각의 차량 에이전트는 자율적인 의사 결정을 위해 개별 강화학습을 사용하며, 자원 소비 및 교환 지연을 최소화하기 위해 분산형 강화학습 모델을 사용하여 학습함으로써, 각각의 차량 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내린다.
- [0012] 상기 보상 계산부는 제한된 공유 자원에 대한 자원 할당을 위해 MARL(Multi-Agent Reinforcement Learning) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트에 대한 개별 보상을 계산하고, 이후 우선순위 가중치를 적용하여 가중치가 부여된 글로벌 보상을 계산하며, 각각의 차량 에이전트가 협력적으로 행동하여 개별 성능을 최적화하며, 상기 강화학습 모델 학습을 위한 피드백으로 제공된다.
- [0013] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법은 상태 정의부를 통해 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 단계 -상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 됨-, 행동 정의부를 통해 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 단계 및 보상 계산부를 통해 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 단계를 포함한다.

발명의 효과

- [0014] 본 발명의 실시예들에 따르면 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당

방법 및 시스템을 통해 RSU와 차량이 모두 공유 자원 풀을 사용하여 다양한 크기의 패킷을 전송하고, 분산형 강화학습 모델을 사용하며, 각 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정하여 자원 소비 및 교환 지연을 최소화할 수 있다. 또한, 우선 순위 가중치에 의해 알려지는 공유 글로벌 보상을 위해 MARL 모델 훈련을 제안하고, 이를 통해 에이전트가 협력적으로 행동하여 개별 성능을 최적화할 뿐만 아니라 다양한 트래픽을 효율적으로 지원할 수 있다.

도면의 간단한 설명

- [0015] 도 1은 본 발명의 일 실시예에 따른 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법을 설명하기 위한 흐름도이다.
- 도 2는 본 발명의 일 실시예에 따른 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템의 구성을 나타내는 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 이기종 V2X 트래픽을 갖는 시스템 모델을 설명하기 위한 도면이다.
- 도 4는 본 발명의 일 실시예에 따른 글로벌 보상 계산 과정을 설명하기 위한 도면이다.
- 도 5는 본 발명의 일 실시예에 따른 MARL 모델의 학습 알고리즘을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0016] V2X(Vehicle-to-everything) 통신은 자율 주행과 ITS(Intelligent Transportation Systems)을 아우르는 첨단 주행을 위한 중추적인 기술이다. 직접 V2V(Vehicle-to-Vehicle) 통신을 넘어, RSU(Road Side Unit)를 통한 V2I(Vehicle-to-Infrastructure) 통신은 효율적인 교통 관리와 첨단 주행을 위해 중요한 역할을 할 수 있으며, 주변 차량에 적절한 도로 정보를 제공한다. 다양한 V2X 시나리오를 수용하기 위해서는 V2X 통신에 다양한 목적, 형식 및 크기를 가진 이중 트래픽을 지원해야 한다.
- [0017] 본 발명에서는 RSU 배치 V2X 통신에서 이중 트래픽에 대한 자원 할당 문제를 해결하고, 제한된 공유 자원을 가진 MARL(Multi-Agent Reinforcement Learning) 기반 자원 할당 체계를 제안한다. 모델 복잡성을 줄이기 위해 RSU는 단일 에이전트가 동시에 여러 자원을 선택하는 대신 작은 작업 공간을 가진 가상 에이전트 모음으로 모델링된다. 트래픽 이중성을 효율적으로 통합하기 위해 가중 글로벌 보상이 도입된다. 성능은 PRR(Packet Reception Ratio)과 통신 범위 측면에서 랜덤 및 최적 할당 체계와 평가 및 비교된다. 제안된 기법은 최적 기법의 성능과 거의 일치하며 부하 부족 및 과부하 상황에서 랜덤 할당 체계 성능을 크게 개선한다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0019] 본 발명에서는 혼합 RSU-차량 시나리오에 대한 RL 기반 자원 할당 전략을 제안한다. 제안하는 접근 방식에서, RSU와 차량은 모두 공유 자원 풀을 사용하여 다양한 크기의 패킷을 전송한다. 프레임워크는 분산형 강화학습, 예를 들어 DQN(Deep Q-Network) 모델을 사용하고, 각 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내리므로 자원 소비 및 교환 지연을 최소화한다. 또한, 우선 순위 가중치에 의해 알려지는 공유 글로벌 보상을 위해 MARL 모델 훈련을 제안한다. 이는 에이전트가 협력적으로 행동하여 개별 성능을 최적화할 뿐만 아니라 다양한 트래픽을 효율적으로 지원할 수 있다.
- [0020] 본 발명의 실시예에 따르면, 다양한 패킷 크기를 가진 다양한 V2X 트래픽을 처리함에 있어서, 초기에 V2X 엔티티에 대한 개별 보상을 계산한다. 이어서, 가중치가 부여된 글로벌 보상을 생성하기 위해 우선순위 가중치를 적용한다. 다양한 크기의 메시지에 대한 이러한 가중치의 변조를 통해, 작은 메시지에 비해 자원 집약적인 특성에도 불구하고 큰 메시지에 대한 PRR 성능을 향상시킨다.
- [0021] 본 발명의 실시예에 따르면, 공유 자원 풀 내의 과부하 및 과소부하 상황을 모두 수용한다. 다양한 채널 혼잡 수준에 걸쳐 자원 할당 체계의 효과를 보여준다. 이를 통해 적절한 통신 범위를 위한 효율적인 자원 활용을 보장한다. 변동하는 자원 수요와 가용성에 대한 적응력은 동적 자원 수요로 특징지어지는 실제 V2X 설정에 적합하게 만든다.
- [0022] 본 발명의 실시예에 따르면, RSU를 여러 가상 에이전트와 함께 배포하고 각 에이전트가 단일 자원을 선택하는 것을 포함한다. 이는 MA-DQN 모델에서 단일 에이전트가 여러 자원을 동시에 선택하는 것과 대조된다. 이러한 구현은 기존의 접근 방식에 비해 행동 공간 크기를 최소화하고 학습 효율성을 향상시킨다.

- [0023] V2X 통신을 위한 다양한 RL 기반 자원 할당 관련 종래기술은 다양한 V2X 설정에서 다양한 정보 가용성을 가진 다양한 목적을 추구하기 위해 별개의 RL 모델을 활용했다. 종래기술에서는 교차 환경에서 V2V 링크에 대한 최적의 자원 할당과 이산 전력 레벨 할당을 위해 MADQN 모델을 활용했다. 이러한 맥락에서, V2V 및 V2I 링크는 스펙트럼 자원을 공유하여 V2V 링크 지연 시간 및 신뢰성 제약을 충족하면서 V2I 링크 채널 용량을 최대화하는 것을 목표로 한다. 그러나 이 접근 방식은 모든 V2I 및 V2V 링크에 대한 채널 정보, 주변 차량의 할당 세부 정보, 자원에 대한 간섭 전력 및 V2V 링크에 대한 잔여 패킷 전송 시간 제약을 포함하는 광범위한 정보 집합이 필요했다.
- [0024] 또 다른 종래기술에서 에이전트 결정은 언더레이 상황에서 자원 및 전력 할당과 함께 모드 선택을 포괄했다. 예를 들어, 분산형 다중 에이전트 행위자-비평가(Actor-Critic; AC) 모델은 연속 값 채널 정보를 능숙하게 처리하고 전력을 전송하는 반면, 연합 심층 강화 학습(Deep Reinforcement Learning; DRL) 모델은 에이전트 간의 분산 학습을 용이하게 했다. 후자의 모델은 에이전트가 개별 관찰 및 목표를 고려하는 동안 집단적인 네트워크 지식을 활용할 수 있도록 했다. 또 다른 종래기술에서는 DQN과 DDPG(Deep Deterministic Policy Gradient)를 결합하여 RL 모델을 제작했다. DQN 요소는 차량 간 최적의 자원 할당을 결정하는 반면, DDPG는 에이전트가 연속적인 가변성을 가진 전력 레벨을 선택할 수 있도록 했다.
- [0025] V2X 통신에 대한 대부분의 이전 연구는 두 가지 일반적인 특성을 공유한다. 첫째, 분산형 다중 에이전트 모델을 채택했다. NR V2X 모드 2에서 각 차량이 자율적으로 자원을 선택한다는 것은 주목할 만하다. 이는 에이전트가 자신의 관찰과 학습을 기반으로 의사 결정을 내리는 분산형 다중 에이전트 프레임워크를 다른 RL 모델보다 V2X 컨텍스트를 포착하기에 더 적합하게 만든다. 둘째, RL 학습에 대한 글로벌 보상을 통합했다. MARL의 분산된 특성은 여러 에이전트의 행동으로 인해 비정규성을 도입하여 RL 훈련 성과를 손상시킬 수 있다. 그럼에도 불구하고 모든 에이전트에게 동일한 글로벌 보상 피드백을 제공하면 이러한 비정규성이 완화된다. 또한 글로벌 보상은 경쟁적 에이전트 기반 자원 할당에서 전체 네트워크 성능 향상을 목표로 하는 완전한 협력 노력으로 초점을 옮긴다. 이 접근 방식은 V2X 네트워크의 전반적인 효과를 향상시키는 것이다. 또 다른 종래기술에서는 V2X 통신 성능, 특히 처리량을 극대화하기 위해 자원 할당에 대한 글로벌 보상을 특징으로 하는 분산형 MA-DQN 전략을 소개했다. 하지만, 본 발명에서는 RSU를 생략하고 CAM 메시지만을 중심으로 하는 동질적인 트래픽 환경을 고려했다.
- [0026] 본 발명에서는 ITS(Intelligent Transportation Systems)를 위한 전용 주파수 대역에 초점을 맞추어 V2I 링크를 통한 자원 공유를 방지하고자 한다. 따라서 모드 선택의 필요성이 줄어들고 전력 제어의 역할이 두드러지지 않게 된다. 이는 자원 할당의 핵심적인 강조점을 도로 주체 간의 효율적인 분배로 이동시킨다. 또한 두 가지 트래픽 부하 상황을 설명한다. 가용 자원이 필요한 할당을 초과하는 부하량의 경우, 자원 할당의 초점은 충돌을 방지하는 것이다. 반면에 자원이 부족한 과부하 시나리오에서는 충돌에 따른 간섭을 최소화하는 것으로 목표가 변한다.
- [0027] 종래기술들은 V2X 통신에서 RSU(Roadside Units)의 중추적인 역할을 강조했다. 특히 인구 밀집 및 혼잡 지역에서 도시 교통 문제를 해결하기 위한 중요한 해결책으로서 RSU 배치를 강조했다. LiDAR 및 카메라와 같은 센서를 장착한 RSU는 도로 데이터를 수집하고 차량과 보행자에게 배포할 수 있으므로 교통 혼잡을 개선하고 안전을 강화하며 긴급 대응을 가속화할 수 있다. 또 다른 종래기술에서 자율 주행에 대한 개요는 도로 효율성을 극대화하고 고급 주행 성능을 높이기 위해 RSU 배치의 중요성을 거듭 강조했다. 또 다른 종래기술은 SDSM(Sensor Data Sharing Messages)의 사용 사례와 구조를 소개하며 도로 센서 데이터 공유에서 RSU의 중요한 역할을 강조했다.
- [0028] 여러 SDO(Standard Development Organizations)는 ITS(Intelligent Transportation Systems) 구현을 형성하는데 적극적으로 참여해 왔다. 특히, SAE(Society of Automotive Engineers)는 초기에 IEEE 1609.3 및 IEEE 802.11p에 기반한 WAVE(Wireless Access in Vehicular Environments) 통신과 조화를 이루며 V2X 통신 표준을 제정했다. 이 표준은 이후 3GPP의 산물인 C-V2X 통신을 수용하도록 채택되었다. ETSI(European Telecommunications Standards Institute)는 WiFi 기술에 기반한 ETSI-ITS-G5를 활용하여 V2X 통신을 위한 표준에 기여했다. 그 후, ETSI는 3GPP의 C-V2X를 수용하고 C-V2X도 포괄하도록 표준을 개정했다. 통신과 차량 산업 주체들을 연합하는 국제 컨소시엄인 5GAA(5G Automotive Association)도 이러한 환경에서 중추적인 역할을 한다.
- [0029] 다양한 V2X 환경에서 목적, 형식 및 크기가 뚜렷한 메시지가 정의된다. SAE와 ETSI가 구축한 주요 V2X 메시지에 대한 포괄적인 개요는 표 1에 제시되어 있다.

[0030] <표 1>

Standard Organization	Messages	Continuity [†]	Casting ^{††}	Message size ^{†††} [bytes]	Link type	Abbreviation / Objective
SAE	BSM	C	Broadcast	320~350	V2V, V2I, V2P	Basic Safety Message Vehicles inform their speed, direction, and position to their surroundings.
	MAP	C	Broadcast	-	I2V	Map Data Provide geographical information of intersections to vehicles
	SPAT	C	Broadcast	-	I2V	Signal Phase and Timing Message Provide status of intersection and signal information such as priority in signal processing
	TIM	C	Broadcast	-	I2V, I2P	Traveler Information Message Transmit various information to V2X devices.
	PVD	E	All	-	I2V, V2I	Probe Vehicle Data Exchange of vehicle information with surrounding V2X devices (mainly RSUs).
ETSI	CAM	C	Broadcast	Around 350 (200 ~ 800)	V2V, V2I	Cooperative Awareness Message Vehicles inform their speed, direction, and position to their surroundings
	DENM	E	All	Up to 1200	V2V, I2V, V2P, I2P	Decentralized Environmental Notification Message Notify the information of accidents, personal issues (engine malfunctions, brake failures, etc.) or hazardous situations on the road to their surroundings
	CPM	E	All	121 for header, 35 per included information	V2V, V2I, V2P, I2V, I2P	Collective Perception Message Alert the surroundings about uncooperative road users and obstacles on the road
	MAPEM	C	Broadcast	-	I2V	Map Message Provide geographical information of intersections to vehicles
	SPATEM	C	Broadcast	-	I2V	Signal Phase and Timing Message Provide status of intersection and signal information such as priority in signal processing
	IVIM	C	Broadcast	-	I2V	In-Vehicle Information Message Provide road information (speed limits, road conditions, etc.) to vehicles

[†] C: Continuous, E: Event-triggered.
^{††} All: Unicast, groupcast, and broadcast.
^{†††} Hyphen: Not specified clearly.

[0031]

[0033]

연속성, 캐스팅 유형, 메시지 크기, V2X 링크 유형 및 목적을 아우르는 메시지 특성에 대한 통찰력을 제공한다. 5GAA 분류에 따르면, V2X 메시지는 연속 및 이벤트 트리거 두 가지 범주로 분류된다. BSM 및 CAM과 같은 연속 메시지는 주기적인 방송을 통해 도로 사용자 및 인프라 엔티티 간의 지속적인 정보 공유를 용이하게 한다. 이 메시지는 헤딩, 속도, 가속 및 위치와 같은 세부 정보를 공유한다. 반대로, 이벤트 트리거 메시지는 사고, 도로 장애물, 엔진 고장 또는 연료 부족과 같은 차량 관련 문제와 같은 특정 이벤트에 대해 주변에 경고하는 역할을 한다. 이러한 메시지는 정보 요청 또는 교환의 역할도 한다. 이벤트 트리거 메시지는 연속 메시지와 달리 사용 사례에 따라 유니캐스트, 그룹캐스트 또는 브로드캐스트 모드를 채택할 수 있다. 유니캐스트는 특정 도로 사용자에게 경고하는 Emergency Stops에 선택될 수 있고, 그룹캐스트는 군집 주행을 위한 CACC(Cooperative Adaptive Cruise Control)를 활성화할 수 있으며, 브로드캐스트는 위험한 도로 상황을 알리는 EVW(Emergency Vehicle Warning)에 적합함을 입증한다.

[0034]

메시지 크기는 전달하는 정보의 양에 따라 달라진다. 그 대표적인 예는 교차로 기반 지리 데이터를 브로드캐스트하는 MAP 메시지이다. 그것의 크기는 교차로 수와 세부 레벨과 같은 요소에 따라 달라진다. CAM과 같이 광범

위하게 연구되는 메시지의 경우, 평균 크기는 약 350 바이트이다. 그러나 차량 범주 및 보안 콘텐츠와 같은 선택적 데이터로 인해 200-800 바이트 범위에 걸쳐 있다. 많은 메시지가 I2V, I2P 및 V2I 링크를 통해 인프라구조와 엔티티 간에 교환된다. 이는 인프라구조의 더 넓은 시야와 센서 배열 덕분이며, 표준 도로 사용자가 얻기 어려운 과제가 될 수 있는 정보의 전송을 가능하게 한다.

[0035] 앞서 언급한 메시지의 배열은 하나의 활용 사례에 공동으로 기여할 수 있다. 자율 주행 차량이 교차로를 협상하는 교차로 자동화 시나리오를 고려해보자. 여기서 차량은 다양한 메시지를 이용하여 주행 계획을 수립한다. 인프라구조로부터 SPAT 및 MAP 메시지를 획득하여 신호등 단계, 타이밍 및 지리적 특성에 대한 통찰력을 얻는다. 동시에 동료 차량의 CAM 또는 BSM 메시지에 의존하여 주행 의도를 관측한다. 이러한 메시지 병합은 자율 주행 차량이 자신 있게 교차로를 통과하고 횡단 기동을 안전하게 수행할 수 있도록 지원한다.

[0036] 표 1에 자세히 설명된 V2X 메시지 외에도 몇 가지 보충 메시지도 설정된다. SAE는 다양한 사용 사례에 적합한 새로운 메시지를 적극적으로 작성하고 있다. 예를 들어, 2022년 11월에 발표된 SAE J2735 문서의 가장 최신 버전은 새로운 메시지를 소개한다. 여기에는 날씨 데이터의 교환을 용이하게 하는 RWM(Road Weather Message), 취약한 도로 사용자의 안전을 강화하기 위해 설계된 PSM(Pedestrian Safety Message), 감지를 위해 개별 센서를 활용하여 차량 및 인프라에서 파생된 도로 위험 정보의 공유를 위한 SDSM(Sensor Data Sharing Message)가 포함된다.

[0038] 도 1은 본 발명의 일 실시예에 따른 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법을 설명하기 위한 흐름도이다.

[0039] 제안하는 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 방법은

[0040] 상태 정의부를 통해 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의하는 단계(110), 행동 정의부를 통해 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택하는 행동을 정의하는 단계(120) 및 보상 계산부를 통해 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영하는 단계(130)를 포함한다.

[0041] 단계(110)에서, 상태 정의부를 통해 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의한다. 본 발명의 실시예에 따르면, 상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 된다.

[0042] 본 발명의 실시예에 따르면, 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트의 각각의 에이전트는 단일 전송 블록을 선택한다.

[0043] 본 발명의 실시예에 따르면, 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우 간섭이 발생하므로, 전송 블록 재할당을 위해 동일한 전송 블록을 사용하는지 여부를 판단하여 상태를 정의한다.

[0044] 단계(120)에서, 행동 정의부를 통해 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택한다.

[0045] 본 발명의 실시예에 따르면, 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우, 간섭을 제거하기 위해 해당 차량 에이전트의 전송 블록을 재할당하도록 행동을 정의한다.

[0046] 본 발명의 실시예에 따르면, 각각의 차량 에이전트는 자율적인 의사 결정을 위해 개별 강화학습, 예를 들어 DQN을 사용하며, 자원 소비 및 교환 지연을 최소화하기 위해 분산형 DQN(Deep Q-Network) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내리도록 한다. 본 발명의 실시예에서는 DQN을 예시로서 설명하지만, 이는 일 실시예일뿐 이외에도 다양한 강화학습을 사용할 수 있다.

[0047] 단계(130)에서, 보상 계산부를 통해 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영한다.

- [0048] 본 발명의 실시예에 따르면, 제한된 공유 자원에 대한 자원 할당을 위해 MARL(Multi-Agent Reinforcement Learning) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트에 대한 개별 보상을 계산하고, 이후 우선순위 가중치를 적용하여 가중치가 부여된 글로벌 보상을 계산한다.
- [0049] 본 발명의 실시예에 따르면, 각각의 차량 에이전트가 협력적으로 행동하여 개별 성능을 최적화하며, 상기 DQN 모델 학습을 위한 피드백으로 제공된다.
- [0051] 도 2는 본 발명의 일 실시예에 따른 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템의 구성을 나타내는 도면이다.
- [0052] 본 실시예에 따른 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)은 프로세서(210), 버스(220), 네트워크 인터페이스(230), 메모리(240) 및 데이터베이스(250)를 포함할 수 있다. 메모리(240)는 운영체제(241) 및 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 루틴(242)을 포함할 수 있다. 프로세서(210)는 상태 정의부(211), 행동 정의부(212) 및 보상 계산부(213)를 포함할 수 있다. 다른 실시예들에서 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)은 도 2의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)은 디스플레이나 트랜시버(transceiver)와 같은 다른 구성요소들을 포함할 수도 있다.
- [0053] 메모리(240)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(240)에는 운영체제(241)와 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당 루틴(242)을 위한 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 드라이브 메커니즘(drive mechanism, 미도시)을 이용하여 메모리(240)와는 별도의 컴퓨터에서 판독 가능한 기록 매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록 매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록 매체(미도시)를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록 매체가 아닌 네트워크 인터페이스(230)를 통해 메모리(240)에 로딩될 수도 있다.
- [0054] 버스(220)는 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)의 구성요소들간의 통신 및 데이터 전송을 가능하게 할 수 있다. 버스(220)는 고속 시리얼 버스(high-speed serial bus), 병렬 버스(parallel bus), SAN(Storage Area Network) 및/또는 다른 적절한 통신 기술을 이용하여 구성될 수 있다.
- [0055] 네트워크 인터페이스(230)는 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)을 컴퓨터 네트워크에 연결하기 위한 컴퓨터 하드웨어 구성요소일 수 있다. 네트워크 인터페이스(230)는 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)을 무선 또는 유선 커넥션을 통해 컴퓨터 네트워크에 연결시킬 수 있다.
- [0056] 데이터베이스(250)는 V2X 통신에서 이중 트래픽을 위한 분산형 다중 에이전트 강화학습 기반 자원 할당을 위해 필요한 모든 정보를 저장 및 유지하는 역할을 할 수 있다. 도 2에서는 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)의 내부에 데이터베이스(250)를 구축하여 포함하는 것으로 도시하고 있으나, 이에 한정되는 것은 아니며 시스템 구현 방식이나 환경 등에 따라 생략될 수 있고 혹은 전체 또는 일부의 데이터베이스가 별개의 다른 시스템 상에 구축된 외부 데이터베이스로서 존재하는 것 또한 가능하다.
- [0057] 프로세서(210)는 기본적인 산술, 로직 및 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)의 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(240) 또는 네트워크 인터페이스(230)에 의해, 그리고 버스(220)를 통해 프로세서(210)로 제공될 수 있다. 프로세서(210)는 상태 정의부(211), 행동 정의부(212) 및 보상 계산부(213)를 위한 프로그램 코드를 실행하도록 구성될 수 있다. 이러한 프로그램 코드는 메모리(240)와 같은 기록 장치에 저장될 수 있다.
- [0058] 상태 정의부(211), 행동 정의부(212) 및 보상 계산부(213)는 도 1의 단계들(110~130)을 수행하기 위해 구성될 수 있다.
- [0059] 분산형 다중 에이전트 강화학습 기반 자원 할당 시스템(200)은 상태 정의부(211), 행동 정의부(212) 및 보상 계산부(213)를 포함할 수 있다.
- [0060] 본 발명의 실시예에 따른 상태 정의부(211)는 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트 및 단

일 에이전트인 차량 에이전트 각각의 전송 블록에 대한 상태를 정의한다. 본 발명의 실시예에 따르면, 상기 RSU 에이전트는 분산 전송 블록이 할당되고, 상기 차량 에이전트는 단일 전송 블록이 할당 된다.

- [0061] 본 발명의 실시예에 따르면, 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트의 각각의 에이전트는 단일 전송 블록을 선택한다.
- [0062] 본 발명의 실시예에 따른 상태 정의부(211)는 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우 간섭이 발생하므로, 전송 블록 재할당을 위해 동일한 전송 블록을 사용하는지 여부를 판단하여 상태를 정의한다.
- [0063] 본 발명의 실시예에 따른 행동 정의부(212)는 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트에 대한 분산 전송 블록을 선택하고, 상기 차량 에이전트에 대한 단일 전송 블록을 선택한다.
- [0064] 본 발명의 실시예에 따른 행동 정의부(212)는 상기 RSU 에이전트 및 상기 차량 에이전트 중 동일한 전송 블록을 사용하는 경우, 간섭을 제거하기 위해 해당 차량 에이전트의 전송 블록을 재할당하도록 행동을 정의한다.
- [0065] 본 발명의 실시예에 따르면, 각각의 차량 에이전트는 자율적인 의사 결정을 위해 개별 강화학습, 예를 들어 DQN을 사용하며, 자원 소비 및 교환 지연을 최소화하기 위해 분산형 DQN(Deep Q-Network) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트는 RSSI 값을 공유하지 않고 로컬로 관찰된 환경 상태만을 기반으로 자율적으로 결정을 내리도록 한다. 본 발명의 실시예에서는 DQN을 예시로서 설명하지만, 이는 일 실시예일뿐 이외에도 다양한 강화학습을 사용할 수 있다.
- [0066] 본 발명의 실시예에 따른 보상 계산부(213)는 상기 정의된 행동을 수행한 후, 각각의 에이전트에 대한 로컬 보상 및 전체 시스템에 대한 글로벌 보상을 계산하고, 계산된 로컬 보상 및 글로벌 보상을 피드백하여 모든 에이전트의 학습에 반영한다.
- [0067] 본 발명의 실시예에 따른 보상 계산부(213)는 제한된 공유 자원에 대한 자원 할당을 위해 MARL(Multi-Agent Reinforcement Learning) 모델을 사용하여 학습함으로써, 각각의 차량 에이전트에 대한 개별 보상을 계산하고, 이후 우선순위 가중치를 적용하여 가중치가 부여된 글로벌 보상을 계산한다.
- [0068] 본 발명의 실시예에 따르면, 각각의 차량 에이전트가 협력적으로 행동하여 개별 성능을 최적화하며, 상기 DQN 모델 학습을 위한 피드백으로 제공된다.
- [0070] 도 3은 본 발명의 일 실시예에 따른 이기종 V2X 트래픽을 갖는 시스템 모델을 설명하기 위한 도면이다.
- [0071] 도 3은 RSU가 있는 고속도로에 배치된 V2X 네트워크의 시스템 모델을 보여준다. 예를 들어, RSU(353)는 도로의 왼쪽 옆, 상부 차선에 근접하게 위치한다. 고속도로는 N대의 차량을 수용한다. 이 시나리오에서는 RSU가 큰 크기로 메시지를 전송하는 반면, K-1 차량은 작은 크기로 메시지를 전송한다고 가정한다. 단순화를 위해 DENM(Decentralized Environmental Notification Message)이 일반적으로 크기가 CAM(Cooperative Awareness Message)을 능가한다는 점을 고려하여 크기가 큰 메시지와 작은 메시지에 대해 각각 DENM과 CAM을 지정한다. 두 메시지 모두 100 msec의 기본 전송 간격을 유지한다.
- [0072] NR 프레임 구조에서 가장 작은 자원 블록은 7개의 OFDM 심볼과 12개의 부반송파를 포함하는 PRB(Physical Resource Block)이다. 본 발명에서는 CAM이 여러 개의 PRB로 구성된 단일 전송 블록(Transmission Block; TB)이 필요하다고 가정한다. 또한, DENM은 D TB가 필요하다고 가정한다. TB 전송을 위한 필요한 PRB의 수는 메시지 크기, MCS(Modulation and Coding Scheme) 및 수비학(numerology)에 달려 있다. 본 발명의 실시예에 따른 시나리오에서는 RSU를 포함하여 도로 위의 모든 엔티티가 M TB를 포함하는 공유 자원 풀에 액세스한다. 이 중 차량 7대는 수신 차량(Rx UE)(352)을 의미하고 차량 3대는 자원 할당이 필요한 송신 차량(Tx UE)(351)을 의미한다.
- [0073] 도 3의 왼쪽 부분(310)에서, 복수의 TB(341, 352, 353, 344, 345)를 포함하는 자원 블록에서, UE 1과 UE 2는 모두 동일한 TB(341)를 공유하여 그들 사이에 간섭을 야기한다. 유사하게, UE 3과 RSU는 동일한 TB(344)를 사용하므로 간섭 또한 발생한다. 결과적으로, V2X 네트워크의 전체 성능이 저하된다. 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트(353) 및 단일 에이전트인 차량 에이전트(351, 352) 각각의 전송 블록에 대한 상태를 정의한다(311).
- [0074] 상기 정의된 상태에 따라 상기 복수의 가상 에이전트 그룹에 의해 구성되는 RSU 에이전트(353)에 대한 분산 전

송 블록을 선택하고, 상기 차량 에이전트(351, 352)에 대한 단일 전송 블록을 선택하는 행동을 정의한다(320).

[0075] 각 도로 엔티티는 자율적인 의사 결정을 위해 개별 DQN을 사용한다. 프로세스는 모든 TB에 걸쳐 RSSI 측정에 크게 의존한다. 예를 들어, UE 1의 RSSI를 분석하는 것은 RSU 및 UE 3으로부터 신호를 수신하는 TB(344)에 비해 인근 UE 2로부터의 TB(341)에 대한 더 강한 수신을 나타낸다. 또한, 먼 RSU에만 할당된 TB(345)는 약한 RSSI를 나타낸다. 이러한 RSSI 데이터로 무장하여, 각 도로 엔티티는 자원을 재할당한다(321).

[0076] 재할당 후, 수신 문제가 발생한 차량에 대한 평균 PRR 성능 및 채널 용량을 포괄하는 글로벌 보상이 계산된다(330). 이러한 글로벌 보상은 이후 DQN 학습을 위한 피드백으로 제공된다(331).

[0077] 도 3의 오른쪽(330)에서, 훈련된 MA-DQN을 사용하여 자원 재할당 후 시나리오가 펼쳐진다. UE 2와 UE 3은 미개 발된 TB(342) 및 TB(343)에 재할당된다. 이 목표 재할당은 모든 도로 엔티티에 걸쳐 간섭을 제거하여 전체 성능을 크게 향상시킨다.

[0078] 본 발명의 실시예에서는 대규모 페이딩 성분 α 를 고려하며, 이는 경로 손실 및 음영 효과를 통합한 것이다. 소 규모 페이딩 성분 $\tilde{g}$ 을 고려함으로써 복합 채널 이득 h 는 $h = \alpha\tilde{g}$ 로 표현된다. 매개 변수 $\tilde{g}$ 는 평균 및 단위 분산이 0인 Rayleigh 분포를 따른다. k번째 송신기와 j번째 수신기 사이의 채널 이득을 $h_{k,j}$ 라 하자. 그러면, k번째 송신기로부터 j번째 수신기에서 수신된 SINR(Signal-to-Interference-plus-Noise Rate)를 $\gamma_{k,j}$ 로 나타낼 수 있다.

$$\gamma_{k,j} = \frac{P_{Tx} \cdot h_{k,j}}{\sigma^2 + \sum_{m=1}^M \sum_{\substack{l=1 \\ l \neq k}}^K \rho_k[m] \cdot \rho_l[m] \cdot P_{Tx} \cdot h_{l,j}}, \quad (1)$$

[0080] 여기서 σ^2 는 노이즈 전력, P_{Tx} 는 송신 전력이다. 만약 k번째 송신기가 m번째 TB를 사용한다면 1이고 아니면 0인, 이진 자원 선택 지시자 $\rho_k[m] \in \{0,1\}$ 를 도입한다.

[0081] PRR을 측정하기 위해 SINR 임계값 γ' 을 활용한다. 즉, 수신한 SINR이 γ' 이상이면 메시지 수신이 성공한 것으로 간주한다. $\gamma_{k,j} \geq \gamma'$ 이면 1이고 아니면 0인 지시자 $\omega_{k,j} \in \{0,1\}$ 를 도입한다. 번째 송신기에서 수신하는 차량의 개수를 N_k 라고 하자. k번째 송신기의 PRR, P_k 는 다음과 같이 표현할 수 있다.

$$P_k = \frac{1}{N_k} \left(\sum_{j=1}^{N_k} \omega_{k,j} \right). \quad (2)$$

[0083] 따라서, 자원 할당 문제를 다음과 같이 공식화될 수 있다.

$$\arg \max_{\rho} P_k \quad (3a)$$

$$\text{s.t. } \sum_{m=1}^M \rho_k[m] = 1, (1 \leq k < K) \quad (3b)$$

$$\sum_{m=1}^M \rho_K[m] = D, \quad (3c)$$

[0087] 여기서 $\rho = \{\rho_1[1], \dots, \rho_k[m], \dots, \rho_K[M]\}$ 는 자원 선택 지시자 집합이다. 식(3b)는 CAM 전송을 위해 단일

TB만이 할당됨을 의미하며, 식(3c)는 마지막 송신기가 DENM을 브로드캐스트하기 위해 D TB를 필요로 하는 RSU임을 의미한다.

[0088] 본 발명의 실시예에 따르면, 먼저 IQL(Independent Q-Learning) 기반 MARL(Multi-Agent Reinforcement Learning) 모델을 고려한다. IQL은 다중 에이전트 RL에서 분산형 정책 기반 모델로, 각 에이전트는 다른 에이전트를 환경의 일부로 간주하여 자체 관찰을 통해 네트워크를 훈련한다. 그러나 IQL은 다른 에이전트도 학습 과정 중에 행동을 조정하기 때문에 모든 에이전트가 비정지 환경에서 학습한다는 단점이 있다. 이러한 단점은 학습 과정이 재생 메모리에 저장된 과거 경험에 의존하는 DQN 구조를 사용할 때 더욱 두드러진다. 비정지 환경에서는 현재 상태-행동 쌍과 과거 상태-행동 쌍 간의 상관관계가 약화되어 불안정한 학습을 초래할 수 있다.

[0090] 도 4는 본 발명의 일 실시예에 따른 글로벌 보상 계산 과정을 설명하기 위한 도면이다.

[0091] 도 4는 보상 공유를 사용하여 제안된 MA-DQN 모델의 구조를 보여주며, 전체 환경(400)에서 에이전트, 상태, 행동 및 보상은 다음과 같이 정의된다.

[0092] 본 발명의 실시예에 따르면, 각각의 송신 V2X 엔티티는 에이전트(421, 422)이며, 여기서 제 k번째 에이전트는 관찰된 상태인 $\mathbf{s}_k$ 를 기준으로 행동 a_k^s 을 결정한다. 메시지 전송을 위해 차량 에이전트는 1 TB를 필요로 하며, RSU 에이전트(410)는 D TB를 선택해야 한다. 이러한 이질성은 차량 에이전트와 RSU 에이전트의 서로 다른 행동 공간을 발생시킨다. 제안된 MA-DQN 모델에서 하나의 RSU 에이전트는 D개의 가상 에이전트 그룹(411, 412, 413)에 의해 구성되며, 각 에이전트는 차량 에이전트와 같이 단일 TB를 할당한다.

[0093] 본 발명의 실시예에 따르면, 모든 TB에 대한 이전 RSSI 값을 기반으로 상태를 정의한다. RSSI를 k번째 송신기에서 감지된 m번째 TB에 대한 수신 신호 세기를 나타내는 $q_k[m]$ 로 나타내면 다음과 같이 표현할 수 있다.

$$q_k[m] = \sum_{\substack{l=1 \\ l \neq k}}^K \rho_l[m] \cdot P_{Tx} \cdot h_{l,k}. \quad (4)$$

[0094] 그런 다음, k번째 에이전트의 상태를 다음과 같이 표현할 수 있다.

$$\mathbf{s}_k = \{q_k[1], \dots, q_k[M]\}. \quad (5)$$

[0097] 본 발명의 실시예에 따르면, 행동은 각 에이전트에 의한 TB 또는 TBs의 선택으로 정의된다. 행동 공간의 크기는 에이전트가 RSU인지 차량인지에 따라 달라진다. 차량의 경우 한 번에 하나의 TB만을 선택하면 되므로 차량의 행동 공간의 크기는 M과 같다. 반면 RSU는 사용 가능한 TB 중에서 D TB를 선택해야 한다. 따라서 RSU의 행동 공간의 크기는 $M C_D$ 가 되며, 이는 M TB로부터 D TB를 선택하는 조합의 수를 나타낸다. 이 경우 RSU의 행동 공간이 차량의 행동 공간보다 커져 RSU 에이전트 훈련에 방해가 된다. 이 문제를 완화하기 위해 단일 DQN 모델이 동시에 D TB를 선택하는 대신 도 4와 같이 각각 TB만을 선택하는 D개의 가상 에이전트로 RSU 에이전트를 구현한다. 이 접근 방식은 RSU 에이전트의 동작 공간을 보다 균형 있고 관리하기 쉽도록 보장하여 효과적인 학습을 가능하게 한다.

[0098] 본 발명의 실시예에 따르면, k번째 에이전트의 개별 보상인 R_k 는 실행 후 k번째 에이전트의 로컬 관찰을 바탕으로 계산되며, 이후 글로벌 보상인 R_G 를 계산하며, 이는 모든 에이전트의 행동과 연관되어 모든 에이전트의 학습에 공통적으로 적용된다.

[0099] 수신 장애가 발생한 차량의 채널 용량을 기반으로 보상을 계산할 때 패널티를 도입한다. 그러면 k번째 에이전트의 패널티인 D_k 는 다음과 같이 나타낼 수 있다.

$$D_k = \sum_{j=1}^{N_k^{(f)}} \left(\frac{1}{\max\{C_{min}, C_{k,j}^{(f)}\}} \right), \quad (6)$$

여기서 $N_k^{(f)}$ 는 해당 서비스 영역 내에서 k번째 송신기에 대해 수신 장애가 발생한 수신 차량의 총 개수이다. 그리고 $C_{k,j}^{(f)}$ 는 j번째 수신 장애 수신기의 채널 용량을 나타낸다. 특히 채널 상태가 매우 열악한 상황에서 페널티가 과도하게 커지지 않도록 보장하는 $C_{k,j}^{(f)}$ 에 대한 최소 임계값, C_{min} 을 도입하며, 이는 특히 채널 상태가 극도로 열악한 상황에서 페널티가 과도하게 커지지 않도록 보장한다. $C_{k,j}^{(f)}$ 는 다음과 같이 구할 수 있다.

$$C_{k,j}^{(f)} = W_k \cdot \log_2 (1 + \gamma_{k,j}^{(f)}), \quad (7)$$

여기서 W_k 는 대역폭이고, $\gamma_{k,j}^{(f)}$ 는 k번째 송신기로부터의 j번째 수신기에서의 SINR이다. 그러면, 개개의 보상 R_k 를 다음과 같이 나타낼 수 있다.

$$R_k = \lambda_P P_k - \lambda_D D_k, \quad (8)$$

여기서 λ_P 와 λ_D 는 각각 PRR과 페널티의 균형을 맞추기 위한 일정한 가중치이다. 메시지 크기가 다른 이중 트래픽의 자원 할당을 위해, 본 발명에서는 글로벌 보상의 계산을 위한 우선순위 가중치를 도입한다. 따라서 글로벌 보상 R_G 를 다음과 같이 나타낼 수 있다.

$$R_G = \lambda_V \left(\frac{1}{K-1} \sum_{k=1}^{K-1} R_k \right) + \lambda_R R_K. \quad (9)$$

이 사례에서, λ_R 및 λ_V 는 각각 RSU 및 차량에 대한 우선 순위 가중치로서, λ_R 값을 증가시킴으로써, $\lambda_R + \lambda_V = 1$ 을 충족시키면서 RSU가 글로벌 보상에서 차량보다 높은 우선 순위를 갖도록 제어될 수 있다.

각 DQN 모델은 ReLU 활성화 함수와 1차원 배치 정규화 과정을 가진 3개의 은닉 계층으로 구성되며, DQN은 기대 보상 G_k 를 최대화하는 최적의 정책 π^* 를 선택하도록 훈련된다. 여기서 보상의 감쇠 성분을 γ 로 고려하면 G_k 를 다음과 같이 나타낼 수 있다.

$$G_k = \sum_{t=0}^{T-1} \gamma^t R_k, \quad (0 \leq \gamma \leq 1) \quad (10)$$

여기서 R_k 는 각 시간 단계에서 획득한 k번째 에이전트의 개별 보상을 나타내고, T는 에피소드의 총 시간 단계 수이다. 할인율 γ 는 즉각적인 보상에 대한 미래 보상의 중요도를 결정한다. 기대 보상 G_k 를 최대화함으로써 에이전트는 시간이 지남에 따라 더 높은 누적 보상으로 이어지는 최적의 정책을 학습하는 것을 목표로 한다. 행동 a_k^s 는 다음과 같이 나타낼 수 있다.

$$a_k^s = \arg \max_{a_k^s \in A_k} [Q(s_k, a_k^s, \theta_k)], \quad \forall k \quad (11)$$

[0112] 여기서 $Q(s_k, a_k^s, \theta_k)$, θ_k 및 A_k 는 각각 k번째 에이전트의 출력 Q값, Q-네트워크의 가중치, 행동 공간을 나타낸다. 모든 에이전트가 행동을 취한 후, 각 에이전트의 글로벌 보상 R_G 및 다음 상태가 결정된다. 각 에이전트는 자신의 재생 메모리에 상태, 행동, 글로벌 보상, 다음 상태를 튜플로 저장한다. 이어서, k번째 에이전트는 자신의 재생 메모리에서 D 크기인 경험의 미니 배치를 랜덤으로 선택하여 손실 함수 $L_k^D(\theta_k)$ 를 계산한다. $L_k^D(\theta_k)$ 는 Q-네트워크의 Q값과 타겟 Q-네트워크의 차이를 기반으로 정의된다. k번째 에이전트의 타겟 Q-네트워크를 나타낼 때, $\tilde{Q}(s_k, a_k^s, \tilde{\theta}_k), L_k^D(\theta_k)$ 로 표현할 수 있다.

$$L_k^D(\theta_k) = \sum_{i=1}^D [R_G(i) + \gamma \max_{a_k^s(i+1)} \tilde{Q}(s_k(i+1), a_k^s(i+1), \tilde{\theta}_k) - Q(s_k(i+1), a_k^s(i+1), \theta_k)]^2. \quad (12)$$

[0113] 이후, 손실 함수를 최소화하기 위해 확률적 경사 강하를 이용하여 가중치 θ_k 를 갱신한다. 구체적으로, 손실 함수를 최소화하는 방향으로 가중치 갱신을 수행한다. 유사하게, k번째 에이전트에 대한 타겟 Q 네트워크의 가중치 $\tilde{\theta}_k$ 를 가중치 θ_k 에 일치시키도록 주기적으로 갱신한다. 이러한 동기화를 통해 타겟 Q 네트워크가 학습된 Q 네트워크를 근접하게 추종하고 학습 과정을 안정화하도록 보장한다.

[0115] 도 5는 본 발명의 일 실시예에 따른 MARL 모델의 학습 알고리즘을 나타내는 도면이다.

[0117] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0118] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0119] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체

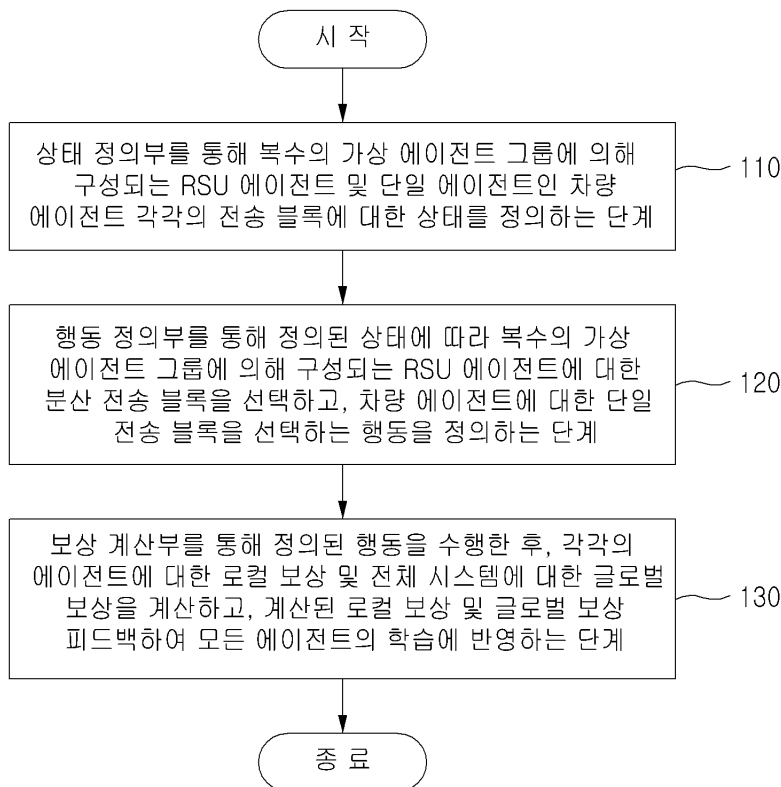
(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0120] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0121] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

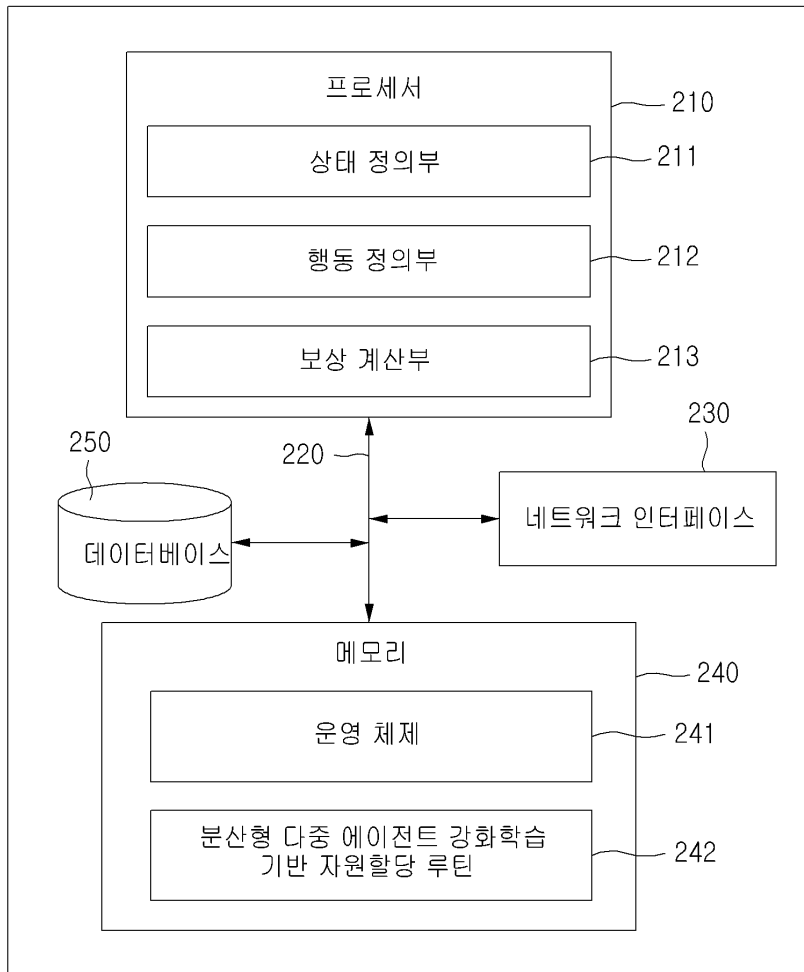
도면

도면1

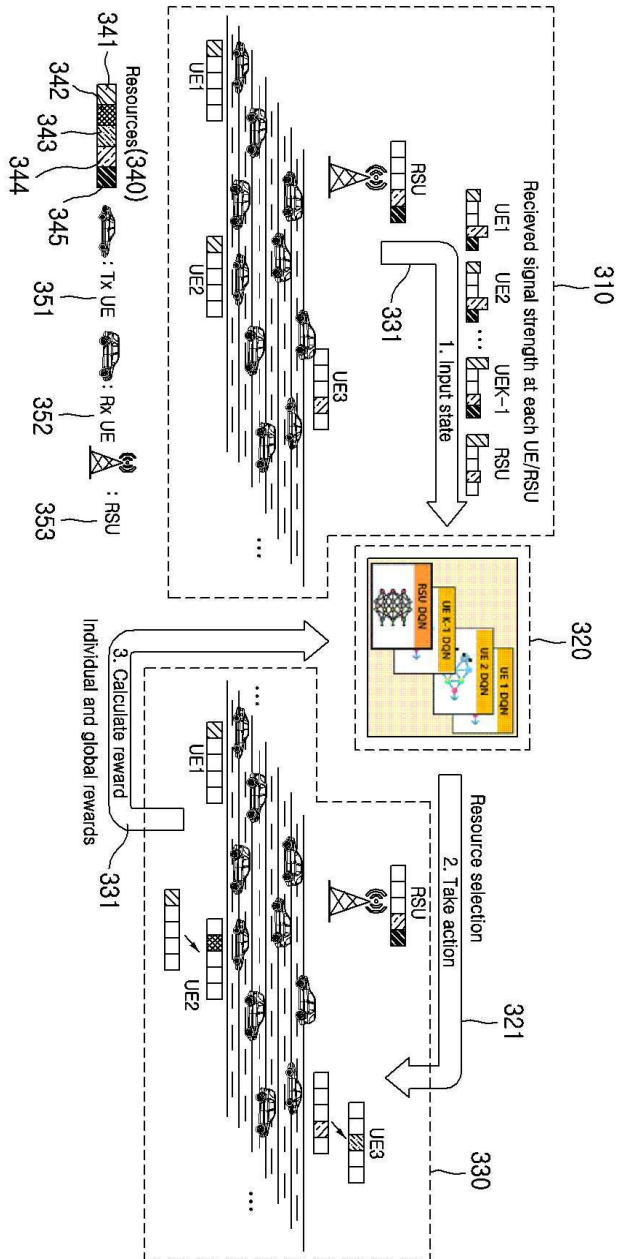


도면2

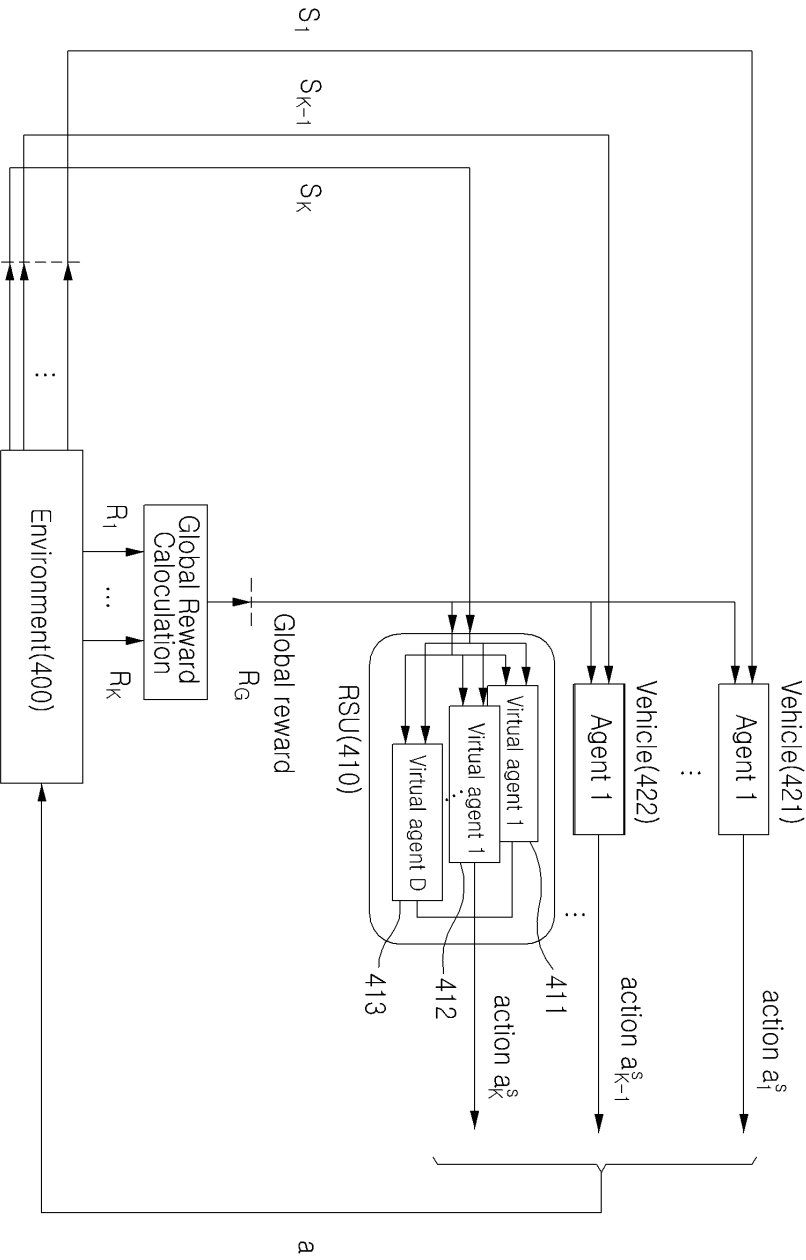
200



도면3



도면4



도면5

Algorithm 1: Learning process of the proposed MARL model

```

1: Generate  $N$  vehicles randomly and 1 RSU at the left
   side of the highway;
2: Initialize Q-networks for all agents
3: for each position do
4:   Update vehicle positions and large scale fading
   component
5:   for each step  $t$  do
6:     for each agent  $k$  do
7:       Observe  $s_k$ 
8:       Choose action  $a_k^S$  according to  $\epsilon$ -greedy
       policy
9:     end for
10:    Update small scale fading component
11:    All agents take actions and receive global
    reward  $R_G$ 
12:    for each agent  $k$  do
13:      Observe  $s_k$ 
14:      Store  $(s_k, a_k^S, R_G)$  in replay memory
15:    end for
16:  end for
17:  for each agent  $k$  do
18:    Randomly sample mini-batches from replay
    memory
19:    Optimize error between Q-network and target
    network using stochastic gradient descent
20:  end for
21: end for

```
