(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0044818
(43) 공개일자 2024년04월05일

(51) 국제특허분류(Int. Cl.)
H04B 7/185 (2006.01) *B64G 1/10* (2024.01)
B64G 1/66 (2024.01) *B64G 3/00* (2024.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
H04B 7/18519 (2013.01)
B64G 1/1007 (2024.01)
(21) 출원번호 10-2022-0124287
(22) 출원일자 2022년09월29일
심사청구일자 2022년09월29일

(71) 출원인
국방과학연구소
대전광역시 유성구 북유성대로488번길 160 (수남동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
한명훈
대전광역시 유성구 북유성대로 488번길 160
백상헌
서울특별시 성북구 안암로 145
(뒷면에 계속)
(74) 대리인
특허법인 광장리앤고

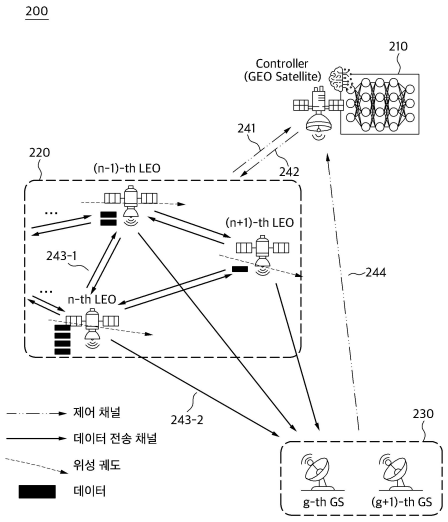
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **복수의 위성을 제어하는 장치 및 그의 동작 방법**

(57) 요 약

본 발명은 전자 장치가 복수의 위성을 제어하는 방법으로서, 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득하는 단계; 상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키는 단계; 및 상기 모델을 사용하여 상기 복수의 위성을 제어하는 단계를 포함하는 복수의 위성을 제어하는 방법에 관한 것이다.

대 표 도 - 도2



(52) CPC특허분류
B64G 1/66 (2024.01)
B64G 3/00 (2024.01)
G06N 3/08 (2023.01)
H04B 7/18515 (2013.01)
(72) 발명자
함재현
대전광역시 유성구 북유성대로 488번길 160
최홍록
서울특별시 성북구 안암로 145

명 세 서

청구범위

청구항 1

전자 장치가 복수의 위성을 제어하는 방법에 있어서,

상기 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국(ground station)에 대한 제2 정보를 획득하는 단계;

상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키는 단계; 및

상기 모델을 사용하여 상기 복수의 위성을 제어하는 단계를 포함하는, 복수의 위성을 제어하는 방법.

청구항 2

제1항에 있어서,

상기 제1 정보는 상기 복수의 위성이 이동하는 궤도 및 상기 복수의 위성이 보유하는 데이터에 대한 정보를 포함하고,

상기 제2 정보는 상기 하나 이상의 지상국의 위치 및 상기 하나 이상의 지상국이 상기 복수의 위성으로부터 수신한 데이터에 대한 정보를 포함하는, 복수의 위성을 제어하는 방법.

청구항 3

제1항에 있어서,

상기 데이터 전송과 연관된 행동은 상기 복수의 위성 간의 데이터 오프로드(offload) 동작 및 상기 복수의 위성에서 상기 하나 이상의 지상국으로의 데이터 다운로드(download) 동작을 포함하는, 복수의 위성을 제어하는 방법.

청구항 4

제1항에 있어서,

상기 모델은 심층 강화 학습(deep reinforcement learning)을 통해 훈련되고,

상기 심층 강화 학습과 연관된 보상(reward)은 상기 제2 정보를 기반으로 확인된 상기 복수의 위성에서 상기 하나 이상의 지상국으로 다운로드되는 데이터의 양으로 정의되는, 복수의 위성을 제어하는 방법.

청구항 5

제4항에 있어서,

상기 심층 강화 학습은 상기 보상과 엔트로피(entropy)의 가중치 합계(weighted sum)로 설정되는 목적 함수를 최대화하는 SAC(Soft Actor-Critic) 알고리즘을 기반으로 하는, 복수의 위성을 제어하는 방법.

청구항 6

제4항에 있어서,

상기 심층 강화 학습과 연관된 상태(state)는:

상기 복수의 위성 간의 통신 가능 여부 및 상기 복수의 위성과 상기 하나 이상의 지상국 간의 통신 가능 여부에 대한 접촉 정보;

상기 복수의 위성이 보유하는 데이터의 양에 대한 데이터 정보; 및

상기 복수의 위성과 상기 하나 이상의 지상국의 접촉 시간에 대한 접촉 시간 정보를 기반으로 정의되고,

상기 접촉 정보, 상기 데이터 정보, 및 상기 접촉 시간 정보는 상기 제1 정보 및 상기 제2 정보 중 적어도 하나를 기반으로 확인되는, 복수의 위성을 제어하는 방법.

청구항 7

제6항에 있어서,

상기 접촉 시간 정보는 상기 복수의 위성 중 제1 위성이 상기 하나 이상의 지상국 중 제1 지상국의 통신 범위에 진입하는 제1 시점 및 상기 제1 위성이 상기 통신 범위를 벗어나는 제2 시점에 대한 정보를 포함하는, 복수의 위성을 제어하는 방법.

청구항 8

제1항에 있어서,

상기 전자 장치는 상기 복수의 위성에 제어 메시지를 전송하는 정지궤도(geostationary earth orbit) 위성에 포함되고,

상기 복수의 위성은 복수의 저궤도(low earth orbit) 위성을 포함하는, 복수의 위성을 제어하는 방법.

청구항 9

제1항에 있어서,

상기 모델을 사용하여 복수의 위성을 제어하는 단계는

상기 복수의 위성에 대한 제3 정보 및 상기 하나 이상의 지상국에 대한 제4 정보를 획득하는 단계;

상기 모델을 사용하여 상기 제3 정보 및 상기 제4 정보에 대응하는 상기 복수의 위성의 행동을 결정하는 단계; 및

상기 결정된 행동을 기반으로 상기 복수의 위성을 제어하는 단계를 포함하는, 복수의 위성을 제어하는 방법.

청구항 10

비일시적 컴퓨터 판독 가능 저장 매체로서,

컴퓨터 판독 가능 명령어들을 저장하도록 구성되는 매체를 포함하고,

상기 컴퓨터 판독 가능 명령어들은 프로세서에 의해 실행되는 경우 상기 프로세서가:

복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득하는 단계;

상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키는 단계; 및

상기 모델을 사용하여 상기 복수의 위성을 제어하는 단계를 포함하는 상기 복수의 위성을 제어하는 방법을 수행하도록 하는, 비일시적 컴퓨터 판독 가능 저장 매체.

청구항 11

복수의 위성을 제어하는 전자 장치로서,

적어도 하나의 명령어를 저장하는 메모리; 및

상기 적어도 하나의 명령어를 실행함으로써, 상기 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득하고, 상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키고, 상기 모델을 사용하여 상기 복수의 위성을 제어하는 프로세서를 포함하는, 복수의 위성을 제어하는 전자 장치.

발명의 설명

기 술 분 야

[0001] 본 명세서의 실시 예는 복수의 위성을 제어하는 장치 및 그의 동작 방법에 관한 것이다.

배 경 기 술

[0002] 최근 인공위성의 기술적인 발전과 생산 비용의 감소 등을 통해 우주 통신 기술 개발과 B5G/6G에서 우주 네트워크를 활용한 연구가 주목을 받고 있다. 특히, 각각 양 38,000km와 약 5,000~20,000km 고도에 위치하는 정지궤도(geostationary earth orbit, GEO) 위성과 중궤도(medium earth orbit, MEO) 위성에 비해, 약 500~1,200km의 낮은 고도에 위치하는 저궤도(low earth orbit, LEO) 위성이 큰 관심을 받았다. LEO 위성은 GEO 위성과 MEO 위성보다 더 낮은 고도에 위치하기 때문에 지상과의 통신에서 더 좋은 신호 품질과 더 낮은 전파 지연을 보장할 수 있게 되어 글로벌 서비스 제공에 있어 더 유리할 것으로 기대된다. 또한 Starlink와 OneWeb 등 인공위성을 통해 글로벌 인터넷 서비스를 제공하는 기업에서는 2027년까지 수천 개에 이르는 위성을 발사할 계획을 갖고 있다. LEO 위성을 기반으로 위성을 통한 목표 지역 감시, 기상 측정뿐만 아니라 전세계를 커버하는 다중 액세스 에지 컴퓨팅(multi-access edge computing, MEC)과 유비쿼터스 지능(ubiquitous intelligence)와 같은 차세대 주요 핵심 기술이 실현 가능하게 할 것으로 기대된다.

[0003] 한편, 지상 노드의 관점에서 정지해 있는 GEO 위성과는 달리, LEO 위성은 약 27,000km/h의 속도로 궤도를 따라 지구 주변을 돌고 있기 때문에, LEO 위성이 이동하는 궤도에 따라 지상국(ground station, GS)과의 연결이 수분에 불과하는 정도의 짧은 접촉 시간을 가질 수 있다. 이와 더불어, 각각의 위성은 정해진 궤도를 따라 이동하며 데이터를 수집 및 운반하게 되는데, 지나온 환경에 따라 위성이 수집하는 데이터의 양에 차이가 발생할 수 있다. 이때, 만약 위성이 수집한 데이터의 양에 비해 하나의 GS와의 접촉 시간이 짧아서 위성이 주어진 접촉 시간 내에 보유하는 데이터를 모두 전송하지 못하는 경우, 해당 위성이 보유하는 데이터를 모두 다운로드하기 위해서는 이후 다른 GS와의 추가적인 접촉을 필요로 하게 되며, 이로 인해 데이터 전송이 완료되는 데에 더 많은 시간과 스케줄링 비용이 소요된다.

[0004] 이러한 문제를 해결하기 위한 종래 기술(X. Jia, T. Lv, F. He, and H. Huang, "Collaborative data downloading by using inter-satellite links in LEO satellite networks," IEEE Transactions on Wireless Communications, vol. 16, no. 3, pp. 1523-1532, March 2017)에서는 위성들의 협력적 다운로드 기법이 연구되었다. 해당 연구에서는 짧은 접촉 시간 내에 위성 네트워크에서 GS로의 데이터 다운로드 처리량(throughput)을 최대화하기 위해 주변에 수집 및 운반하는 데이터의 양이 적은 위성에 데이터를 오프로드(offload)하는 기법을 제안하였다. 그러나, 해당 연구에서는 위성 네트워크 환경을 비교적 정적인 것으로 고려하였다. 위성의 이동에 따른 위성과 위성 간의 무선 링크 연결 상태 변화와 위성과 GS 간의 무선 링크 연결 상태 변화는 고려하였지만, 위성이 특정 지역을 가로지르는 동안 발생하는 통신 및 연결 상태 변화에 대한 정보를 미리 예측 가능하며 변하지 않는 것으로 가정하였다. 그러나, 실제 LEO 위성 네트워크 환경은 위성의 높은 이동성으로 인해 위성 간의 무선 통신 환경과 위성과 GS 간의 무선 통신 환경이 실시간으로 변하는 매우 동적인 환경일 수 있다. 따라서, 최적의 해를 찾은 시점과 환경이 달라지더라도 실시간으로 변화하는 환경에 대응할 수 있는 기법이 요구된다.

발명의 내용

해결하려는 과제

[0005] 본 명세서의 실시 예는 상술한 문제점을 해결하기 위하여 제안된 것으로, 위성 및 지상국의 정보를 기반으로 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키고, 모델을 사용하여 위성을 제어하기 위한 방법을 제안한다.

[0006] 보다 구체적으로, 본 명세서의 일 실시 예는 동적인 네트워크 환경에 대해 학습하고 온라인 방식으로 상황의 변화에 따라 적절한 동작을 수행하기 위한 심층 강화 학습(Deep Reinforcement Learning, DRL) 기반의 협력적 다운로드 프레임워크(Cooperative Downloading Framework, CDF)에 관한 것이다. 실시 예에서, 네트워크 컨트롤러 역할을 수행하는 GEO 위성이 에이전트(agent)로 동작하면서 관할 지역을 지나는 복수의 LEO 위성과 GS 간의 접촉 정보 및 각 LEO 위성이 보유하는 데이터의 양을 수신하고, 시간에 따른 각 LEO 위성들의 연결 상태와 데이터의 양의 변화에 대한 관계를 학습하고, 이를 통해 주어진 상황에서 데이터 오프로드 또는 데이터 다운로드 동작 수행 여부를 결정함으로써 처리량을 최대화할 수 있는 정책(policy)을 학습하는 것을 목적으로 한다.

과제의 해결 수단

[0007] 상술한 과제를 달성하기 위하여, 본 명세서의 일 실시 예에 따르는 전자 장치가 복수의 위성을 제어하는 방법은, 상기 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국(ground station)에 대한 제2 정보를 획득하는 단계; 상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키는 단계; 및 상기 모델을 사용하여 상기 복수의 위성을 제어하는 단계를 포함하는 것을 특징으로 한다.

[0008] 일 실시 예에 따르면, 상기 제1 정보는 상기 복수의 위성이 이동하는 궤도 및 상기 복수의 위성이 보유하는 데이터에 대한 정보를 포함하고, 상기 제2 정보는 상기 하나 이상의 지상국의 위치 및 상기 하나 이상의 지상국이 상기 복수의 위성으로부터 수신한 데이터에 대한 정보를 포함하는 것을 특징으로 한다.

[0009] 일 실시 예에 따르면, 상기 데이터 전송과 연관된 행동은 상기 복수의 위성 간의 데이터 오프로드(offload) 동작 및 상기 복수의 위성에서 상기 하나 이상의 지상국으로의 데이터 다운로드(download) 동작을 포함하는 것을 특징으로 한다.

[0010] 일 실시 예에 따르면, 상기 모델은 심층 강화 학습(deep reinforcement learning)을 통해 훈련되고, 상기 심층 강화 학습과 연관된 보상(reward)은 상기 제2 정보를 기반으로 확인된 상기 복수의 위성에서 상기 하나 이상의 지상국으로 다운로드되는 데이터의 양으로 정의되는 것을 특징으로 한다.

[0011] 일 실시 예에 따르면, 상기 심층 강화 학습은 상기 보상과 엔트로피(entropy)의 가중치 합계(weighted sum)로 설정되는 목적 함수를 최대화하는 SAC(Soft Actor-Critic) 알고리즘을 기반으로 하는 것을 특징으로 한다.

[0012] 일 실시 예에 따르면, 상기 심층 강화 학습과 연관된 상태(state)는: 상기 복수의 위성 간의 통신 가능 여부 및 상기 복수의 위성과 상기 하나 이상의 지상국 간의 통신 가능 여부에 대한 접촉 정보; 상기 복수의 위성이 보유하는 데이터의 양에 대한 데이터 정보; 및 상기 복수의 위성과 상기 하나 이상의 지상국의 접촉 시간에 대한 접촉 시간 정보를 기반으로 정의되고, 상기 접촉 정보, 상기 데이터 정보, 및 상기 접촉 시간 정보는 상기 제1 정보 및 상기 제2 정보 중 적어도 하나를 기반으로 확인되는 것을 특징으로 한다.

[0013] 일 실시 예에 따르면, 상기 접촉 시간 정보는 상기 복수의 위성 중 제1 위성이 상기 하나 이상의 지상국 중 제1 지상국의 통신 범위에 진입하는 제1 시점 및 상기 제1 위성이 상기 통신 범위를 벗어나는 제2 시점에 대한 정보를 포함하는 것을 특징으로 한다.

[0014] 일 실시 예에 따르면, 상기 전자 장치는 상기 복수의 위성에 제어 메시지를 전송하는 정지궤도(geostationary earth orbit) 위성에 포함되고, 상기 복수의 위성은 복수의 저궤도(low earth orbit) 위성을 포함하는 것을 특징으로 한다.

[0015] 일 실시 예에 따르면, 상기 모델을 사용하여 복수의 위성을 제어하는 단계는 상기 복수의 위성에 대한 제3 정보 및 상기 하나 이상의 지상국에 대한 제4 정보를 획득하는 단계; 상기 모델을 사용하여 상기 제3 정보 및 상기 제4 정보에 대응하는 상기 복수의 위성의 행동을 결정하는 단계; 및 상기 결정된 행동을 기반으로 상기 복수의 위성을 제어하는 단계를 포함하는 것을 특징으로 한다.

[0016] 본 명세서의 일 실시 예에 따르는 비일시적 컴퓨터 판독 가능 저장 매체는, 컴퓨터 판독 가능 명령어들을 저장

하도록 구성되는 매체를 포함하고, 상기 컴퓨터 판독 가능 명령어들은 프로세서에 의해 실행되는 경우 상기 프로세서가: 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득하는 단계; 상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키는 단계; 및 상기 모델을 사용하여 상기 복수의 위성을 제어하는 단계를 포함하는 상기 복수의 위성을 제어하는 방법을 수행하도록 하는 것을 특징으로 한다.

[0017] 본 명세서의 일 실시 예에 따르는 복수의 위성을 제어하는 전자 장치는, 적어도 하나의 명령어를 저장하는 메모리; 및 상기 적어도 하나의 명령어를 실행함으로써, 상기 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득하고, 상기 제1 정보 및 상기 제2 정보를 기반으로 상기 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키고, 상기 모델을 사용하여 상기 복수의 위성을 제어하는 프로세서를 포함하는 것을 특징으로 한다.

발명의 효과

[0018] 본 명세서의 실시 예에 따르면, 네트워크 컨트롤러 역할을 수행하는 정지궤도 위성이 에이전트로 동작하면서 특정 지역을 지나는 복수의 저궤도 위성과 하나 이상의 지상국 간의 접촉 정보 및 각 저궤도 위성이 보유하는 데이터의 양을 수신하고, 시간에 따른 각 저궤도 위성들의 연결 상태와 데이터의 양의 변화에 대한 관계를 학습하여, 주어진 상황에서 데이터 오프로드 또는 데이터 다운로드 동작 수행여부를 결정함으로써 데이터 처리량을 최대화할 수 있다.

도면의 간단한 설명

[0019] 도 1은 본 명세서의 일 실시 예에 따른 전자 장치의 각 구성을 개략적으로 도시하는 예시적인 도면이다.

도 2는 본 명세서의 일 실시 예에 따른 전자 장치의 복수의 위성을 제어하는 방법이 구현되는 예시적인 시스템 모델을 나타내는 도면이다.

도 3은 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반의 협력적 다운로드 프레임워크를 나타내는 예시적인 도면이다.

도 4는 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반 협력적 다운로드 프레임워크를 구성하는 신경망 설계를 나타내는 예시적인 도면이다.

도 5는 본 명세서의 일 실시 예에 따른 전자 장치가 복수의 위성을 제어하는 방법의 흐름을 나타내는 순서도이다.

도 6은 본 명세서의 일 실시 예에 따른 SAC 알고리즘의 동작을 나타내는 도면이다.

도 7은 본 명세서의 일 실시 예에 따른 성능 평가를 위해 설정한 실험 환경을 도시한 도면이다.

도 8은 본 명세서의 일 실시 예에 따른 에이전트의 액터 네트워크의 학습에 따른 손실의 수렴 양상을 도시한 도면이다.

도 9는 본 명세서의 일 실시 예에 따른 각 기법의 초기 데이터 분포에 따른 위성의 평균 접촉 시간 이용률을 나타내는 도면이다.

도 10은 본 명세서의 일 실시 예에 따른 각 기법의 저궤도 위성의 개수에 따른 성능의 변화를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0021] 실시 예를 설명함에 있어서 본 발명이 속하는 기술 분야에 익히 알려져 있고 본 발명과 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다. 이는 불필요한 설명을 생략함으로써 본 발명의 요지를 흐리지 않고 더욱 명확히 전달하기 위함이다.

[0022] 마찬가지 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시되었다. 또한, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다. 각 도면에서 동일한 또는 대응하는 구성요소에는 동일한 참조 번호를 부여하였다.

[0023] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시 예들은 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. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[0024] 이 때, 처리 흐름도 도면들의 각 블록과 흐름도 도면들의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수 있음을 이해할 수 있을 것이다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 흐름도 블록(들)에서 설명된 기능들을 수행하는 수단을 생성하게 된다. 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 메모리에 저장되는 것도 가능하므로, 그 컴퓨터 이용가능 또는 컴퓨터 판독 가능 메모리에 저장된 인스트럭션들은 흐름도 블록(들)에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다. 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 흐름도 블록(들)에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

[0025] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실행 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

[0026] 이 때, 본 실시 예에서 사용되는 '~부'라는 용어는 소프트웨어 또는 FPGA또는 ASIC과 같은 하드웨어 구성요소를 의미하며, '~부'는 어떤 역할들을 수행한다. 그렇지만 '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '~부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들, 및 변수들을 포함한다. 구성요소들과 '~부'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성요소들과 '~부'들로 더 분리될 수 있다. 뿐만 아니라, 구성요소들 및 '~부'들은 디바이스 또는 보안 멀티미디어카드 내의 하나 또는 그 이상의 CPU들을 재생시키도록 구현될 수도 있다.

[0027] 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. 또한, 명세서에 기재된 “~부”, “~모듈” 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0028] 명세서 전체에서 기재된 “a, b, 및 c 중 적어도 하나”의 표현은, ‘a 단독’, ‘b 단독’, ‘c 단독’, ‘a 및 b’, ‘a 및 c’, ‘b 및 c’, 또는 ‘a,b,c 모두’를 포괄할 수 있다.

[0029] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다.

[0031] 도 1은 본 명세서의 일 실시 예에 따른 전자 장치의 각 구성을 개략적으로 도시하는 예시적인 도면이다.

[0032] 도 1을 참조하면, 전자 장치(100)는 프로세서(110) 및 메모리(120)를 포함할 수 있고, 복수의 위성을 제어하는 방법을 수행할 수 있다. 도 1에 도시된 전자 장치(100)에는 본 실시 예들과 관련된 구성요소들만이 도시되어 있

다. 따라서, 전자 장치(100)에는 도 1에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있음은 당해 기술분야의 통상의 기술자에게 자명하다.

[0033] 프로세서(110)는 전자 장치(100)에서의 복수의 위성의 제어를 위한 전반적인 기능들을 제어하는 역할을 한다. 예를 들어, 프로세서(110)는 전자 장치(100) 내의 메모리(120)에 저장된 프로그램들을 실행함으로써, 전자 장치(100)를 전반적으로 제어한다. 프로세서(110)는 전자 장치(100) 내에 구비된 CPU(central processing unit), GPU(graphics processing unit), AP(application processor) 등으로 구현될 수 있으나, 이에 제한되지 않는다.

[0034] 메모리(120)는 전자 장치(100) 내에서 처리되는 각종 데이터들을 저장하는 하드웨어로서, 메모리(120)는 전자 장치(100)에서 처리된 데이터들 및 처리될 데이터들을 저장할 수 있다. 또한, 메모리(120)는 전자 장치(100)에 의해 구동될 애플리케이션들, 드라이버들 등을 저장할 수 있다. 메모리(120)는 DRAM(dynamic random access memory), SRAM(static random access memory) 등과 같은 RAM(random access memory), ROM(read-only memory), EEPROM(electrically erasable programmable read-only memory), CD-ROM, 블루레이 또는 다른 광학 디스크 스토리지, HDD(hard disk drive), SSD(solid state drive), 또는 플래시 메모리를 포함할 수 있다.

[0035] 일 실시 예에서, 전자 장치(100)는 복수의 위성의 동작을 제어할 수 있다. 전자 장치(100)는 복수의 위성에 제어 메시지를 전송하는 정지궤도 위성에 포함될 수 있고, 복수의 위성은 복수의 저궤도 위성을 포함할 수 있다. 정지궤도 위성이 복수의 저궤도 위성의 데이터 전송과 연관된 동작을 제어하는 방법을 설명하기 위해, 정지궤도 위성, 복수의 저궤도 위성, 및 저궤도 위성으로부터 데이터를 다운로드하는 지상국을 포함하는 시스템 모델을 이하 도 2를 통해 설명한다.

[0036] 도 2는 본 명세서의 일 실시 예에 따른 전자 장치의 복수의 위성을 제어하는 방법이 구현되는 예시적인 시스템 모델을 나타내는 도면이다. 도 2를 참조하면, 정지궤도(GEO) 위성(210), 복수의 저궤도(LEO) 위성(220) 및 하나 이상의 지상국(GS)(230)을 포함하는 위성 제어 시스템(200)이 도시된다.

[0037] 정지궤도 위성(210)은 특정 지역에 정지한 상태에서 복수의 저궤도 위성(220)과 하나 이상의 지상국(230)으로 이루어지는 네트워크 동작을 제어하는 네트워크 제어기로서 동작할 수 있다.

[0038] 정지궤도 위성(210)은 각자의 궤도로 특정 지역을 지나는 복수의 저궤도 위성(220)을 제어하기 위한 독립적인 제어 채널(241, 242)을 가질 수 있으며, 제어 채널(241, 242)은 정지궤도 위성(210)이 복수의 저궤도 위성(220)의 궤도 정보와 보유한 데이터의 양과 같은 정보를 획득하기 위한 채널(241) 및 정지궤도 위성(210)이 복수의 저궤도 위성(220)의 데이터 전송 및 전달을 위한 제어 메시지를 보내기 위한 채널(242)을 포함할 수 있다.

[0039] 정지궤도 위성(210)은 마찬가지로 하나 이상의 지상국(230)을 제어하기 위한 독립적인 제어 채널(244)을 가질 수 있으며, 제어 채널(244)을 통해 복수의 저궤도 위성(220)의 하나 이상의 지상국(230)에 대한 데이터 전송 결과 정보를 획득할 수 있다.

[0040] 일 실시 예에서, 복수의 저궤도 위성(220)은 특정 지역에 진입하기 전에 지나온 환경 및 다른 저궤도 위성 또는 지상국과 접촉한 경험에 따라 상이한 데이터 양을 보유한 채로 특정 지역에 진입할 수 있다. 복수의 저궤도 위성(220)은 특정 지역에 진입한 이후, 다른 위성으로부터 데이터를 오프로드받는 경우를 제외하고 추가적인 데이터를 수집하지 않는 것으로 가정될 수 있다.

[0041] 복수의 저궤도 위성(220) 간에는 위성 간 링크(inter-satellite link, ISL)(243-1)를 통해 데이터를 오프로드할 수 있고, 복수의 저궤도 위성(220)과 지상국(230) 간에는 지상-위성 링크(satellite-ground link, SGL)(243-2)를 통해 복수의 저궤도 위성(220)에서 지상국(230)으로 데이터를 다운로드할 수 있다.

[0042] 도 2에 도시된 실시 예에서, 복수의 저궤도 위성(220)은 각각의 할당된 궤도를 이동하면서 데이터를 수집하고 특정 지역에서 하나 이상의 지상국(230)과 접촉할 수 있다. 명세서 전반에서, 저궤도 위성이 지상국과 "접촉(contact)"하는 것, 또는 하나의 저궤도 위성이 다른 저궤도 위성과 "접촉"하는 것은 접촉하는 두 개체가 서로 통신 가능한 범위에 위치하여 서로 데이터를 송수신할 수 있다는 것을 지칭할 수 있다. 복수의 저궤도 위성(220)이 하나 이상의 지상국(230)과 접촉하는 시간 동안, 복수의 저궤도 위성(220)은 SGL(243-2)를 통해 하나 이상의 지상국(230)으로 데이터를 다운로드할 수 있다. 이때 복수의 저궤도 위성(220)은 그들의 경로에 따라 각각 상이한 양의 데이터와 접촉 시간을 가질 수 있으므로, 복수의 저궤도 위성(220) 중 접촉 시간에 비해 과도한 양의 데이터를 보유한 저궤도 위성 A와, 상대적으로 적은 양의 데이터를 보유하고 지상국과의 여유있는 접촉 시간을 갖는 저궤도 위성 B가 존재할 수 있다. 이 경우, 더 많은 양의 데이터를 다운로드할 수 있도록 ISL(243-1)이 이용될 수 있다. 예를 들어, 과도한 양의 데이터를 보유한 저궤도 위성 A는 ISL(243-1)을 통해 상대적으로 적은 양의 데이터를 보유하는 저궤도 위성 B에 데이터를 오프로드할 수 있다. 이에 따라, 저궤도 위성 B는 그의

지상국과의 접촉 시간에 오프로드받은 데이터를 지상국으로 다운로드할 수 있다. 이와 같은 복수의 저궤도 위성(220)의 다운로드 기법은 협력적 다운로드 기법으로 지칭될 수 있다.

[0043] 이와 같이 지상국(230)으로 다운로드되는 데이터의 양을 최대화하려면, 모든 저궤도 위성(220)은 그의 지상국(230)과의 접촉 시간을 최대한 활용하여야 한다. 그러나, 복수의 저궤도 위성(220)은 그들의 경로에 따라 각각 상이한 양의 데이터와 접촉 시간을 가지며, ISL(243-1)이 복수의 위성의 접촉 시간 및 보유 데이터 간의 불일치를 해소하는 데 활용될 수 있다고 하더라도, 저궤도 위성의 높은 이동성과 이로 인한 ISL(243-1) 및 SGL(243-2)의 동적 변경은 불일치의 극복을 더욱 어렵게 한다. 이를 해결하기 위해, 일 실시 예에서, 전자 장치가 복수의 저궤도의 데이터 전송과 연관된 최적의 행동을 결정하기 위한 학습이 수행될 수 있고, 학습은 최적의 정책을 획득하기 위한 MDP(Markov Decision Process, 마르코프 결정 프로세스)와 같은 확률 모델로 표현되는 심층 강화 학습(deep reinforcement learning)을 포함할 수 있다. 이는 전자 장치로 하여금 복수의 저궤도 위성이 현재 관측된 네트워크 상태 정보에 따라 데이터를 오프로드 또는 다운로드할지에 대한 최적의 결정을 내리도록 할 수 있으며, 궁극적으로 다운로드되는 데이터의 양을 최대화할 수 있다.

[0044] 한편, 복수의 저궤도 위성(220)을 포함하는 저궤도 위성 네트워크 환경은 비슷한 시점에 여러 위성이 동시에 접촉하는 경우가 발생할 수 있으며, 위성의 이동은 주기적인 특성을 가지므로 여러 위성이 동시에 접촉하는 경우가 반복적으로 발생할 수 있다. 즉, 동시에 접촉하는 위성의 수가 많으므로 데이터 처리 방식에 대한 선택지가 많아져 최적의 데이터 전송 수단을 찾는 것이 어려워 학습이 지체될 수 있다. 이러한 상황에서 문제를 해결할 수 있는 최적의 정책을 학습하기 위해서는 적절한 탐험 메커니즘을 통해 여러 선택지에 대한 탐험이 수행되어야 하며 최적에 가까운 행동을 수행할 수 있어야 한다.

[0045] 이와 같은 목적을 달성하기 위해, 일 실시 예에서, 심층 강화 학습은 보상과 엔트로피(entropy)의 가중치 합계(weighted sum)로 설정되는 목적 함수를 최대화하는 SAC(Soft Actor-Critic) 알고리즘을 기반으로 할 수 있다. 심층 강화 학습에서 행동의 주체인 에이전트의 행동은 저궤도 위성이 데이터를 어떤 저궤도 위성 또는 지상국으로 전송해야 하는지를 나타내는 인덱스로 표현되는 이산 행동 공간(discrete action space)에 포함되기 때문에, 에이전트의 행동의 학습에 적합한 SAC 알고리즘을 기반으로 하는 협력적 다운로드 기법을 고려할 수 있다. SAC 알고리즘은 정책 그래디언트(policy gradient) 알고리즘의 하나로, 학습 알고리즘의 목적 함수가 보상(reward)와 엔트로피의 가중치 합에 대한 평균으로 설정되며, 목적 함수를 최대화하는 것으로 정의된다는 특징을 갖는다. 이와 같은 목적 함수의 설정은 학습 과정의 탐험적 성능을 향상시켜주며, 이는 최적에 가까운 행동을 수행할 수 있는 방향으로 학습을 진행시키므로 강건한(robust) 정책 학습을 가능하게 할 수 있다.

[0046] 심층 강화 학습을 기반으로 하는 협력적 다운로드 프레임워크가 이하 도 3을 통해 상세히 설명된다.

[0047] 도 3은 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반의 협력적 다운로드 프레임워크를 나타내는 예시적인 도면이다.

[0048] 도 3을 참조하면, 저궤도 위성 네트워크 환경(310)과 SAC 에이전트(320)로 구성된 심층 강화 학습 기반 협력적 다운로드 프레임워크(300)가 도시된다. SAC 에이전트(320)는 최적의 협력적 다운로드 정책을 근사하기 위한 액터(actor) 네트워크(321)와 정책을 평가하기 위한 크리틱(critic) 네트워크(322)를 가질 수 있다. 일 실시 예에서, 크리틱 네트워크(322)는 2개의 메인 크리틱 네트워크와 2개의 타겟 크리틱 네트워크를 포함할 수 있다.

[0049] 매 타임 슬롯의 시작 시, SAC 에이전트는 네트워크 환경으로부터 현재 관측된 네트워크 상태 정보를 갖는 상태(330)를 그래프로 정의할 수 있다. 일 실시 예에서, 상태(330)는 복수의 저궤도 위성 간의 연결 또는 복수의 저궤도 위성과 하나 이상의 지상국 간의 연결 상태, 복수의 저궤도 위성이 보유하는 데이터의 양, 복수의 저궤도 위성과 하나 이상의 지상국의 접촉 시작 시간 및 접촉 종료 시간에 대한 정보를 기반으로 설정될 수 있다.

[0050] 그 다음, SAC 에이전트는 액터 네트워크(321)를 사용하여 상태(330)를 행동(340)으로 매핑하고, 행동(340)으로 환경을 제어할 수 있다. 이에 대응하는 결과는 보상(350) 및 상태의 전이(즉, 다음 상태)로 관측될 수 있다. SAC 에이전트는 이 모든 정보(즉, 상태(330), 행동(340), 보상(350) 및 다음 상태의 튜플)을 경험(360)으로서 경험 버퍼(361)에 저장할 수 있다. 매 타임 슬롯의 종료 시, SAC 에이전트는 경험 버퍼(361)에서 일부 경험의 미니-배치(mini-batch)를 샘플링하고, 최적의 정책을 훈련시키기 위한 손실(loss)을 계산할 수 있다.

[0051] 도 4는 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반 협력적 다운로드 프레임워크를 구성하는 신경망 설계를 나타내는 예시적인 도면이다.

[0052] 도 4를 참조하면, 도 3에 도시된 액터 네트워크(321) 및 크리틱 네트워크(322)가 따르는 신경망(400)이 도시된다. 신경망(400)은 GAT(Graph Attention network)로 이루어진 입력층(410), 완전연결(Fully Connected, FC)층

으로 이루어진 은닉층(420), 및 병렬 FC 층으로 이루어진 출력층(430)으로 구성될 수 있다.

[0053] 입력층(410)에서는, 그래프 컨벌루션 네트워크(graph Convolutional networks) 중 하나인 GAT가 그래프 기반의 네트워크 상태를 효율적으로 처리하기 위해 사용될 수 있다. GAT는 각 노드의 피처 벡터를 인접 노드로 전파하고, 수신한 벡터들을 어텐션(attention) 메커니즘을 기반으로 결합할 수 있다. 어텐션 메커니즘은 피처 벡터로서 추출된 피처들 중 어떤 부분이 중요한 부분인지를 가중치 점수로서 나타내고, 피처 벡터에 상기 가중치를 적용함으로써 학습의 성능을 개선할 수 있는 메커니즘이다. 그 후, 그래프 기반 상태들의 잠재(latent) 피처들이 노드 간 관계를 유지하면서 추출될 수 있다.

[0054] 은닉층(420)의 다음 FC 층은 각 노드로부터 획득된 추출된 잠재 피처를 병합하고 전체 그래프에 대해 은닉 피처를 추출할 수 있다.

[0055] 마지막으로 출력층(430)에서, |L|개의 병렬 FC 층(여기서, L은 저궤도 위성의 집합 $\{l_1, l_2, \ldots, l_L\}$임) 각각은 은닉 피처를 각 저궤도 위성에 대한 로짓(logit)으로 매핑할 수 있다. 크리틱 네트워크의 경우, 로짓은 소프트-Q 값으로 활용될 수 있다. 한편 액터 네트워크의 경우, 로짓은 소프트맥스(softmax) 함수를 적용하여 행동 확률로 매핑될 수 있다.

[0057] 이하에서는, 전술한 프레임워크를 기반으로 도 1의 전자 장치(100)가 수행하는 복수의 위성을 제어하는 방법을 도 5를 통해 설명하기로 한다.

[0058] 도 5는 본 명세서의 일 실시 예에 따른 전자 장치가 복수의 위성을 제어하는 방법의 흐름을 나타내는 순서도이다. 도 5에 도시된 각 단계의 수행 주체는 도 1에 도시된 전자 장치(100)를 포함할 수 있다.

[0059] 일 실시 예에서, 전자 장치는 복수의 위성에 제어 메시지를 전송하는 정지궤도 위성에 포함될 수 있고, 복수의 위성은 복수의 저궤도 위성을 포함할 수 있다.

[0060] 단계 S510에서, 전자 장치는 복수의 위성에 대한 제1 정보 및 하나 이상의 지상국에 대한 제2 정보를 획득할 수 있다. 일 실시 예에서, 제1 정보 및 제2 정보는 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련시키기 위한 훈련 데이터로 사용될 수 있다.

[0061] 일 실시 예에서, 제1 정보는 복수의 위성이 이동하는 궤도 및 복수의 위성이 보유하는 데이터에 대한 정보를 포함할 수 있다. 제1 정보는 도 2에 도시된 채널(241)을 통해 획득될 수 있다. 전자 장치는 제1 정보로부터 전자 장치가 통신할 수 있는 범위 내에 진입하는 복수의 위성 각각의 이동 궤도와, 복수의 위성이 지상국으로 다운로드하기 위해 보유하는 데이터의 양에 대한 정보를 획득할 수 있다.

[0062] 일 실시 예에서, 제2 정보는 하나 이상의 지상국의 위치 및 하나 이상의 지상국이 복수의 위성으로부터 수신한 데이터에 대한 정보를 포함할 수 있다. 제2 정보는 도 2에 도시된 채널(244)을 통해 획득될 수 있다. 전자 장치는 제2 정보로부터 하나 이상의 지상국의 위치와, 하나 이상의 지상국 각각에 대한 복수의 위성의 데이터 전송 결과에 대한 정보를 획득할 수 있다. 일 실시 예에서, 하나 이상의 지상국에 대한 복수의 위성의 데이터 전송 결과는 심층 강화 학습과 연관된 보상으로 정의될 수 있다.

[0063] 단계 S520에서, 전자 장치는 제1 정보 및 제2 정보를 기반으로 복수의 위성의 데이터 전송과 연관된 행동을 결정하는 모델을 훈련할 수 있다. 일 실시 예에서, 데이터 전송과 연관된 행동은 복수의 위성 간의 데이터 오프로드 동작 및 복수의 위성에서 하나 이상의 지상국으로의 데이터 다운로드 동작을 포함할 수 있다.

[0064] 일 실시 예에서, 모델은 심층 강화 학습을 통해 훈련될 수 있다. 심층 강화 학습이란 주어진 상태에서 보상을 최대화할 수 있는 행동을 학습하는 강화 학습에 딥러닝을 결합함으로써 복잡한 상태 공간과 행동 공간을 가지는 문제를 해결하기 위해 정책 함수를 근사하고 학습하는 기술이다.

[0065] 일 실시 예에서, 저궤도 위성 네트워크에서 각 저궤도 위성의 데이터 전송과 연관된 행동을 결정하는 모델은 심층 강화 학습을 통해 훈련될 수 있고, 심층 강화 학습과 연관된 상태(state), 행동(action), 및 보상(reward)을 정의하는 두 가지 실시 예(제1 실시 예 및 제2 실시 예)를 이하에서 설명한다.

[0066] 제1 실시 예에서는 심층 강화 학습과 연관된 상태, 행동 및 보상이 아래의 1), 2) 및 3)과 같이 정의될 수 있다. 제1 실시 예에서는 첫번째 층이 완전연결층으로 이루어진 신경망이 사용된다.

[0067] 1) 상태

[0068] 일 실시 예에서, 심층 강화 학습과 연관된 상태는, 복수의 위성 간의 통신 가능 여부 및 복수의 위성과 하나 이상의 지상국 간의 통신 가능 여부에 대한 접촉 정보, 복수의 위성이 보유하는 데이터의 양에 대한 데이터 정보, 및 복수의 위성과 하나 이상의 지상국의 접촉 시간에 대한 접촉 시간 정보를 기반으로 정의될 수 있다. 일 실시 예에서, 접촉 정보, 데이터 정보, 및 접촉 시간 정보는 단계 S510에서 획득된 제1 정보 및 제2 정보 중 적어도 하나를 기반으로 확인될 수 있다. 일 실시 예에서, 접촉 시간 정보는 복수의 위성 중 제1 위성이 하나 이상의 지상국 중 제1 지상국의 통신 범위에 진입하는 제1 시점 및 제1 위성이 통신 범위를 벗어나는 제2 시점에 대한 정보를 포함할 수 있다.

[0069] 보다 구체적으로, 시점 t에서의 상태 s(t)는 다음과 같이 정의될 수 있다.

[0070] $$\boldsymbol{s}(t) = \{\boldsymbol{l}(t), \boldsymbol{d}(t), \boldsymbol{c}(t)\}$$.

[0071] 여기서, $\boldsymbol{l}(t) = \{l_{n,m}(t) \mid n \in \mathrm{N}, m \in \mathrm{M}\}$ 이며, N은 저궤도 위성에 대한 집합, G는 지상국의 집합, M은 집합 N과 집합 G의 합집합을 나타내고, 결국 $\boldsymbol{l}(t)$ 는 저궤도 위성과 주변 저궤도 위성 및 지상국과의 접촉 정보에 대한 집합을 나타낸다. 이때, $l_{n,m}(t) = f(t,n,m) \in \{0,1\}$ 로 정의되며, $f(t,n,m)$ 는 시점과 노드의 인덱스가 주어졌을 때 통신 가능 여부를 반환하는 함수로, 저궤도 위성 n과 다른 노드 m 간에 통신이 가능한 경우에는 1, 통신이 불가능한 경우에는 0을 반환한다.

[0072] $\boldsymbol{d}(t) = \{d_n(t) \mid n \in \mathrm{N}\}, d_n(t) = [0, D(t)]$ 이며, 이는 저궤도 위성이 시점 t에 보유하고 있는 데이터의 양을 나타내며, 이때 $D(t)$ 는 시점 t에 모든 저궤도 위성이 보유하고 있는 데이터의 합을 나타낸다.

[0073] $\boldsymbol{c}(t) = \left\{\left(c^s_{n,g}(t), c^e_{n,g}(t)\right) \mid n \in \mathrm{N}, g \in \mathrm{G}\right\}$ 이며, 이는 임의의 저궤도 위성 n과 임의의 지상국 g에 대해 시점 t에 남아있는 접촉 시간 정보에 대한 튜플의 집합을 나타낸다. 이때, 접촉 시간 정보 튜플의 $c^s_{n,g}(t)$ 는 시점 t에 지상국 g와 접촉을 시작하기까지 남은 시간에 대한 정보를 나타내며, $c^e_{n,g}(t)$ 는 시점 t에 지상국 g와 접촉이 끝나기까지 남은 시간에 대한 정보를 나타낸다. 각 원소는 t=0인 시점에 가장 큰 값을 가지고 시간이 지남에 따라 감소하며, 저궤도 위성이 지상국의 통신 반경에 들어가는 시점에 $c^s_{n,g}(t)$ 는 0이 되며, 저궤도 위성이 지상국의 통신 반경을 벗어나는 시점에 $c^e_{n,g}(t)$ 는 0이 된다.

[0074] 2) 행동

[0075] 시점 t에서의 행동 a(t)는 정지궤도 위성에 의해 결정되는 각각의 저궤도 위성이 수행할 동작의 집합을 나타내며, 다음과 같이 정의될 수 있다:

[0076] $$\boldsymbol{a}(t) = \{a_n(t) \mid n \in \mathrm{N}\}, a_n(t) \in \mathrm{M}$$.

[0077] 여기서, 저궤도 위성 n의 동작에 대해, $a_n(t) = n$ 인 경우 아무 동작도 수행하지 않고, $a_n(t) \in \mathrm{N} \backslash \{\mathrm{n}\}$ 인 경우 다른 위성으로 데이터를 오프로드하며, $a_n(t) \in \mathrm{G}$ 인 경우 지상국으로 데이터를 다운로드한다.

[0078] 3) 보상

[0079] 일 실시 예에서, 심층 강화 학습과 연관된 보상은 복수의 위성에서 하나 이상의 지상국으로 다운로드되는 데이터의 양으로 정의될 수 있다. 일 실시 예에서, 복수의 위성에서 하나 이상의 지상국으로 다운로드되는 데이터의 양은 S510에서 획득된 제2 정보를 기반으로 확인될 수 있다.

[0080] 보다 구체적으로, 시점 t에서의 보상 r(t)는 다음과 같이 정의될 수 있다:

[0081] $$r(t) = \sum_{n \in \mathrm{N}} r_n(t)$$.

$$r_n(t) = \begin{cases} 1, \text{if } a_n(t) \in \mathrm{G}, l_{n,m}(t) = 1, d_n(t) \neq 0 \\ 0, otherwise \end{cases}$$

[0082] 여기서, 으로 정의되며, 저궤도 위성이 현재 지상국과 연결 상태이면서 데이터를 보유하고 있을 때 지상국으로 보내는 행동을 선택하는 경우 보상으로 1을 받고, 그 외의 경우는 0을 받는다. 정의된 바와 같이 저궤도 위성이 정상적으로 통신이 가능한 상황에 지상국으로의 데이터 다운로드를 수행하게 되면 보상을 받으므로, 이러한 보상을 최대화하는 정책을 학습하는 것은 네트워크 처리량을 최대화하는 정책을 학습하는 것에 대응할 수 있다.

[0083] 일 실시 예에서, 모델을 단순화하기 위해 지구 표면은 평면인 것으로 가정할 수 있다. 또한, 저궤도 위성 간의 채널 및 저궤도 위성과 지상국 간의 채널은 가시선(line-of-sight) 채널이 주요하다는 특성을 고려하여 두 노드 사이의 통신 가능 여부는 거리에 의한 영향이 지배적인 것으로 가정할 수 있다. 또한, 모든 저궤도 위성의 고도 $h_n(t)$는 시간에 따라 변하지 않고 h_n으로 일정하다고 가정될 수 있다. 이에 따라 전술한 $f(t,n,m) = \delta\left[\| \hat{\boldsymbol{p}}_n(t) - \hat{\boldsymbol{p}}_m(t) \| \leq d_{th,n,m} \right]$으로 단순화될 수 있다. 여기서, 함수 $\delta[\cdot]$는 괄호 내부의 조건이 참이면 1을, 거짓이면 0을 반환하는 것으로 정의되며, $\hat{\boldsymbol{p}}_n(t)$과 $\hat{\boldsymbol{p}}_m(t)$은 각각 시점 t에서의 저궤도 위성 또는 지상국의 좌표를 나타내는 $\boldsymbol{p}_n(t) = \left(x_n(t), y_n(t), h_n\right)$와 $\boldsymbol{p}_m(t) = \left(x_m(t), y_m(t), h_m\right)$를 xy-평면으로 투영한 것을 나타내고, $d_{th,n,m}$은 통신 가능 거리를 나타낸다. $p_m(t)$가 저궤도 위성의 좌표를 나타낼 경우 $h_m=h_n$이고, $p_m(t)$가 지상국의 좌표를 나타낼 경우 $h_m=0$이다.

[0084] 한편, 제2 실시 예에서, 심층 강화 학습과 연관된 상태, 행동, 및 보상은 아래의 a), b) 및 c)와 같이 정의될 수 있다. 제2 실시 예에서는 첫번째 층이 GAT로 이루어진 신경망이 사용된다. 후술할 도 9 및 도 10의 실험 결과는 제2 실시 예를 기반으로 정의된 모델을 사용하여 실험한 결과를 나타낸다.

[0085] a) 상태

[0086] 위성 네트워크의 현재 상태는 위성의 데이터 보유량 위성들 간의 또는 위성 및 지상국 간의 연결에 대한 정보를 나타낼 수 있다. 먼저, 위성 네트워크의 위상 관계를 효과적으로 나타내기 위해, 위성 네트워크는 무향 그래프(undirected graph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$로 나타낼 수 있으며, 여기서 $\mathcal{V} = \mathrm{L} \cup \mathrm{K}$은 위성 네트워크의 전체 노드의 집합을 나타내고(본 실시 예에서, L은 저궤도 위상의 집합 $\{l_1, l_2, \ldots, l_L\}$, K는 지상국의 집합 $\{k_1, k_2, \ldots, k_K\}$을 나타냄), $\mathcal{E} = \{(v_i, v_j) | v_i, v_j \in \mathcal{V}\}$는 전체 링크의 집합을 나타낸다. 노드와 그래프 형태의 시간에 따라 변하는 특성은 피처 행렬 $F(t) \in \mathbb{R}^{|\mathcal{V}| \times (1+2|\mathcal{V}|)}$ 및 인접 행렬 $A(t) \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$로 각각 결정된다. 특히, 피처 행렬의 각 행은 각 노드의 피처 벡터를 나타낸다. 노드 $v \in \mathcal{V}$의 피처 벡터 $F_v(t)$는 $F_v(t) = \left[m_v(t) || c_v^0(t) || c_v^1(t) \right] \in \mathbb{R}^{(1+2|\mathcal{V}|)}$로 정의되며, 연산자 $\cdot||\cdot$는 접합(concatenation) 기호를 나타내고, $m_v(t) \in \mathbb{R}$는 보유 데이터의 양을 나타낸다.

$c_v^0(t) \in \mathbb{R}^{|\mathcal{V}|}$는 모든 다른 노드와의 접촉이 시작될 때까지 남은 시간을 나타내고, $c_v^1(t) \in \mathbb{R}^{|\mathcal{V}|}$는 모든 다른 노드와의 접촉이 종료될 때까지 남은 시간을 나타낸다. 인접 행렬 $A(t)$는 대칭 행렬이고, 그 원소 $A_{ij}(t) \in \{0,1\}$는 노드 v_i와 노드 v_j 간의 링크가 존재하는지 여부를 나타내며, 다음과 같이 정의된다:

[0087]
$$A_{ij}(t) = \begin{cases} 1, & \text{if } v_i, v_j \in \mathrm{L} \text{ and } d_{ij}(t) \le d_{\text{th}}^{\text{ISL}} \\ 1, & \text{if } v_i \in \mathrm{L}, v_j \in \mathrm{K}, \text{ and } d_{ij}(t) \le d_{\text{th}}^{\text{SGL}} \\ 0, & \text{otherwise.} \end{cases}$$

[0088] $d_{\text{th}}^{\text{ISL}}$와 $d_{\text{th}}^{\text{SGL}}$는 각각 ISL과 SGL의 최대 거리를 나타낸다.

[0089] 결국, 상태 s_t는 전술한 피처 행렬과 인접 행렬을 사용하여 $s_t = \{F(t), A(t)\}$와 같이 정의된다.

[0090] b) 행동

[0091] 행동 a_t는 타임 슬롯 t에서 저궤도 위성에 의해 수행되는 행동의 집합을 나타내며, $a_t = \{a_l(t) | l \in \mathrm{L}\}$로 정의된다.

[0092] 여기서 $a_l(t) \in \mathcal{V}$는 위성 l의 데이터의 목적지를 나타낸다. 예를 들어, $a_l(t) = l' \in \mathrm{L} \backslash \{l\}$인 경우, 위성 l은 ISL을 통해 다른 위성 l'으로 데이터를 송신하도록 시도한다. 반면, $a_l(t) = k \in \mathrm{K}$인 경우, 위성 l은 SGL을 통해 지상국 k로 데이터를 다운로드하도록 시도한다.

[0093] c) 보상

[0094] 현재 상태 및 선택된 행동의 집합에 따라, 에이전트는 저궤도 위성을 제어하고 즉각적인 보상을 받는다. 더 많은 보상을 획득하기 위해, 에피소드마다 다운로드되는 데이터의 전체 양을 최대화할 필요가 있다. $n_l^D(t)$가 타임 슬롯 t에서 위성 $l \in \mathrm{L}$의 다운로드 데이터의 양을 나타내고, 각 위성은 각 타임 슬롯에서 하나의 데이터 유닛을 송신할 수 있는 것으로 가정한다. 데이터를 지상국으로 송신하기 위해서는, 위성 l은 데이터를 보유해야 하고, 그의 선택된 행동이 인접한 지상국들 중 하나로 데이터를 다운로드하는 것을 나타내야 하므로, $n_l^D(t)$는 다음과 같이 표현될 수 있다:

[0095]
$$n_l^D(t) = \delta[a_l(t) = k, A_{lk}(t) = 1, \text{ and } m_l(t) \ge 1], k \in \mathrm{K}$$

[0096] 여기서 연산자 $\delta[\cdot]$는 괄호 안의 조건이 참이면 1을 반환하고, 거짓이면 0을 반환하는 델타 함수를 나타낸다.

[0097] 한편, 위성들 간의 데이터 교환도 보상에 영향을 미치므로, 위성들 간의 의미없는 데이터 교환을 방지할 필요가 있다. $n_l^O(t)$는 타임 슬롯 t에서 위성 $l \in \mathrm{L}$의 오프로드 데이터의 양을 나타낸다. 위성 l은 위성이 데이터를 보유하고, 그의 선택된 행동이 인접한 위성들 중 하나로 데이터를 오프로드하는 것을 나타낼 때 데이터를 오프로드할 수 있다. 따라서, $n_l^O(t)$는 다음과 같이 표현될 수 있다:

[0098] $$n_l^O(t) = \delta[a_l(t) = l', A_{ll'}(t) = 1, \text{ and } m_l(t) \geq 1], l' \in \mathrm{L}\backslash\{l\}$$

[0099] 결국, 타임 슬롯 t에서의 즉각적인 보상 $r_t := r(s_t, a_t)$ 은 다음과 같이 정의될 수 있고:

[0100] $$r_t = \sum_{l \in \mathrm{L}} \{n_l^D(t) + \lambda \cdot n_l^O(t)\},$$

[0101] $\lambda \leq 0$ 는 데이터 오프로드에 대한 가중치를 나타낸다.

[0102] 일 실시 예에서, 심층 강화 학습은 보상과 엔트로피의 가중치 합계로 설정되는 목적 함수를 최대화하는 SAC 알고리즘을 기반으로 할 수 있다. 이하에서는 도 6을 통해 SAC 알고리즘의 동작에 대해 설명한다.

[0103] 도 6은 본 명세서의 일 실시 예에 따른 SAC 알고리즘의 동작을 나타내는 도면이다.

[0104] 도 6을 참조하면 일 실시 예에 따른 이산화된(discretized) SAC 알고리즘(600)의 동작이 도시된다. 다운로드되는 데이터의 양을 최대화하는 최적의 협력적 다운로드 정책을 훈련시키기 위해서, SAC 기반 알고리즘(T. Haarnoja, et al., "Soft actor-critic algorithms and applications," arXiv preprint arXiv:1812.05905, 2018)이 활용될 수 있다. SAC 알고리즘은 정책을 평가하는 훈련 목적 함수에 엔트로피 항을 포함하며, 이를 통해 누적된 보상뿐만 아니라 엔트로피도 최대화하도록 훈련한다.

[0105] 파라미터 ϕ 를 갖는 액터 네트워크 $\pi_\phi(\cdot)$, 파라미터 θ_i 를 갖는 메인 크리틱 네트워크 Q_{θ_i}, 및 파라미터 $\hat{\theta}_i$ 를 갖는 타겟 크리틱 네트워크 $Q_{\hat{\theta}_i}$ 가 정의된다 ($i \in \{1,2\}$). SAC 알고리즘의 세부 단계는 도 6에 도시된다.

[0106] 먼저, 알고리즘은 경험 버퍼 D, 가중치 α, 및 파라미터(즉, ϕ, θ_i, $\hat{\theta}_i$ ($i \in \{1,2\}$))를 초기화한다 (알고리즘 1 1-5줄 참조). 각 에피소드마다, 알고리즘은 초기 상태 s_t를 관측한다 (알고리즘 1 7줄 참조). 또한, 각 타임 슬롯 t에 대해, 알고리즘은 액터 네트워크 $a_t = \pi_\phi(s_t)$ 를 사용하여 행동 a_t을 생성하고(알고리즘 1 9줄 참조), 행동 a_t를 실행한 후 다음 상태 s_{t+1} 및 보상 r_t 를 관측한다 (알고리즘 1 9-11줄 참조). 그 결과로서, 알고리즘은 경험 (s_t, a_t, r_t, s_{t+1})를 경험 버퍼 D에 저장한다 (알고리즘 1 12줄 참조).

[0107] 다음으로, 알고리즘은 파라미터 업데이트 단계로 진입한다. 파라미터 업데이트를 위해, 알고리즘은 버퍼 D로부터 일부 경험의 미니 배치 $\bar{D}$ 를 무작위로 샘플링한다 (알고리즘 1 13줄 참조). 이들 샘플링된 경험에 기초하여, 알고리즘은 타겟 상태 값 $V_{\hat{\theta}}(s')$ 를 다음과 같이 계산한다.

[0108] $$V_{\hat{\theta}}(s') = \mathbb{E}_{\tilde{a}' \sim \pi_\phi(s')} \left[Q_{\hat{\theta}}^{min}(s', \tilde{a}') - \alpha \log(\pi_\phi(s')) \right], \quad \ldots (1)$$

[0109] 여기서, s'는 미니 배치 샘플들 중 하나인 튜플 $(s, a, r, s') \in \bar{D}$ 의 다음 상태를 나타내고, $\tilde{a}'$ 는 s'와 액터 네트워크 $\pi_\phi(\cdot)$ 로부터 획득된 행동을 나타낸다. 또한, $Q_{\hat{\theta}}^{min}(s', \tilde{a})$ 은 $Q_{\hat{\theta}_1}(s', \tilde{a})$ 와 $Q_{\hat{\theta}_2}(s', \tilde{a})$ 중 최소 값을 나타낸다. 여기서 협력적 다운로드 정책에 의해 생성된 행동은 데이터를 수신하는 지상국 또는 인접한 저궤도 위성에 대한 인덱스가 되어야 한다. 즉, 정책은 이산 행동 공간에서 정의되어야 하므로, 이산 행동 공간을 고려한 SAC 파라미터 업데이트 알고리즘이 채택된다. 구체적으로, 신경망의 출력층은 |L| 개의 병렬 FC 층으로 배열되어 각 층이 각각의 위성에 대응되게 하고, 각 층의 출력 크기는 |V|로 설정되어 각 출력이 이산화된 행동 공간에 대응되게 한다 (L은 저궤도 위성에 대한 집합, V는 L과 지상국에 대한 집합의 합집합을 나타낸다).

이러한 네트워크 설계는 액터 네트워크 및 크리틱 네트워크가 상태를 각각 행동 확률과 Q 값으로 매핑할 수 있게 한다. 결과적으로, 식 (1)의 기대값 연산은 어떠한 행동 확률 밀도 함수를 도출하지 않고 액터 네트워크의 직접적인 출력으로 대체될 수 있다. 즉, 이산화된 행동 공간의 타겟 상태 값은 다음과 같이 계산될 수 있다 (알고리즘 1 14줄 참조).

[0110] $$V_{\hat{\theta}}(s') = \pi_{\phi}(s')^{\top}\left(Q_{\hat{\theta}}^{min}(s',\cdot) - \alpha \log\left(\pi_{\phi}(s')\right)\right) \quad \ldots (2)$$

[0111] 변환된 크리틱 네트워크 $Q_{\hat{\theta}}^{min}(s',\cdot)$ 는 행동 a'를 입력으로 요구하지 않는 것을 확인할 수 있다. 이는 크리틱 네트워크가 선택 가능한 행동들에 대해 직접 소프트 Q 값을 생성하기 때문이다. 타겟 상태 값에 기초하여, 알고리즘은 메인 크리틱 네트워크를 아래 목적 함수의 그래디언트로 업데이트한다 (알고리즘 1 15줄 참조).

[0112] $$\nabla_{\theta_i} J_Q(\theta_i) = \nabla_{\theta_i} \mathbb{E}_{(s,a,r,s')\sim D}\left[\frac{1}{2}\left(Q_{\theta_i}(s,a) - \left(r + \psi \mathbb{E}_{s'}\left[V_{\hat{\theta}}(s')\right]\right)\right)^2\right] \quad \ldots (3)$$

[0113] 여기서, ψ 는 할인율(discount rate)을 나타낸다.

[0114] 나아가, 알고리즘은 액터 네트워크를 아래 목적 함수의 그래디언트로 업데이트한다.

[0115] $$\nabla_{\phi} J_{\pi}(\phi) \approx \nabla_{\phi} \mathbb{E}_{s\sim D}[\mathbb{E}_{\tilde{a}\sim\pi_{\phi}(s)}[\alpha \log(\pi_{\phi}(s)) - Q_{\theta}^{min}(s,\tilde{a})]]. \quad \ldots (4)$$

[0116] 동일한 이산화에 의해, 식 (4)의 내부 기대값은 또한 직접 계산될 수 있으므로, 식 (4)는 아래와 같이 나타내어질 수 있다.

[0117] $$\nabla_{\phi} J_{\pi}(\phi) \approx \nabla_{\phi} \mathbb{E}_{s\sim D}[\pi_{\phi}(s)^{\top}(\alpha \log(\pi_{\phi}(s)) - Q_{\theta}(s,\cdot))] \quad \ldots (5)$$

[0118] 결국 알고리즘은 액터 네트워크 $\pi_{\phi}(\cdot)$ 를 식 (5)를 기반으로 업데이트할 수 있다 (알고리즘 1 16줄 참조).

[0119] 이어서, 가중치 α는 훈련이 진행됨에 따라 아래의 그래디언트를 통해 자동으로 조정된다 (알고리즘 1 17줄 참조).

[0120] $$\nabla_{\alpha} J(\alpha) = \nabla_{\alpha} \pi_{\phi}(s)^{\top}\left(-\alpha\left(\log\left(\pi_{\phi}(s)\right) + \hat{H}\right)\right) \quad \ldots (6)$$

[0121] 여기서, $\hat{H}$ 는 타겟 엔트로피를 나타낸다.

[0122] 마지막으로, 알고리즘은 매 B 스텝마다 지수적으로 변하는 평균을 사용하여 타겟 크리틱 네트워크를 메인 크리틱 네트워크로 업데이트한다 (즉, 소프트 업데이트) (알고리즘 1 18줄 참조).

[0123] $$\hat{\theta}_i = \tau \cdot \theta_i + (1-\tau) \cdot \hat{\theta}_i, i \in \{0,1\} \quad \ldots (7)$$

[0124] 여기서 τ 는 소프트 업데이트에 대한 계수를 나타낸다.

[0125] 다시 도 5를 참조하면, 단계 S530에서, 전자 장치는 모델을 사용하여 복수의 위성을 제어할 수 있다. 본 단계는 학습된 모델을 기반으로 실제 입력 데이터에 대응하는 출력 값을 추론하는 단계를 포함할 수 있다.

[0126] 일 실시 예에서, 전자 장치는 복수의 위성에 대한 제3 정보 및 하나 이상의 지상국에 대한 제4 정보를 획득하고, 모델을 사용하여 제3 정보 및 제4 정보에 대응하는 복수의 위성의 행동을 결정하고, 결정된 행동을 기반으로 복수의 위성을 제어할 수 있다. 제3 정보 및 제4 정보는 학습된 모델 기반의 추론을 위해 복수의 위성 및 하나 이상의 지상국에 대한 실제 동적 환경으로부터 획득한 실시간 데이터를 포함할 수 있다. 제3 정보 및 제4 정보가 입력되면, 전자 장치는 학습된 모델을 사용하여 제3 정보 및 제4 정보를 기반으로 보상을 최대화할 수 있는 복수의 위성의 행동을 결정할 수 있다.

[0128] 도 7은 본 명세서의 일 실시 예에 따른 성능 평가를 위해 설정한 실험 환경을 도시한 도면이다.

[0129] 도 7을 참조하면, 3대의 저궤도 위성(710a, 710b, 710c) 및 하나의 지상국(720)이 포함된 실험 환경(700)이 도시된다. 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반 협력적 다운로드 프레임워크의 성능을 평가하기 위해, 도 7과 같이 (7,800 x 4,500 x 1,150)(km^3)의 3차원 공간을 고려하였으며, 주어진 시간 동안 해당 공간을 지나는 서로 다른 궤도를 갖는 저궤도 위성의 수는 3, 그리고 지상국의 수는 1인 환경을 구성하였다. 저궤도 위성의 고도 h_n은 1,150km로 모두 동일하지만 위성 간의 충돌은 발생하지 않는다고 가정한다. 고도가 1,150km, 고도각(elevation angle)이 35° 인 저궤도 위성을 통해 지상으로 통신 가능한 반경의 반지름은 1,230km이므로, $m \in G$ 인 경우 SGL의 통신 가능 거리를 나타내는 d_{th}^{SGL} 는 1,230km로 설정된다. 즉, 도 7에 도시된 바와 같이 지상의 xy-평면으로 투영하였을 때 반지름 1,230km 의 원 내부에 지상국이 포함되면 지상국과의 통신이 가능한 것으로 간주한다.

[0130] 한편, 저궤도 위성 간의 통신은 이론상 7,733km까지 가능하지만, 보다 낮은 고도에 형성된 다른 궤도를 도는 저궤도 위성 또는 열권의 영향 등의 변수가 존재하므로, 일정 거리 내에 있는 위성과의 통신만을 고려한다. 이때 일정 거리는 인접 궤도면의 동일한 순번의 위성과의 거리를 기준으로 한다. 각각의 위성의 궤도 경사(inclination)가 53.0° , 인접 궤도와의 적경의 차(right ascension of ascending node, RAAN)가 8° , 각 위성의 근점 편각(argument of periapsis)의 차가 8° 인 경우, 인접 궤도 상의 같은 순번의 위성과의 최대 거리는 약 1,865km이기 때문에, ISL의 통신 가능 거리를 나타내는 d_{th}^{ISL} 는 1,865km로 설정된다. SGL 및 ISL의 대역폭은 5.1-5.2GHz의 C-밴드에서 33MHz로 설정되고, 6.2-6.5GHz의 L-밴드에서 16.5MHz로 설정된다.

[0131] 지상국(720)의 위치는 (3900, 2250, 0)으로 설정되며, t=0인 시점에 각 위성(710a, 710b, 710c)의 초기 좌표는 xz-평면 상의 점 $p_n(0) = (x_n(0), 0, 1150)$에 위치하도록 설정되고, 이때 임의의 n에 대해 $x_n(0)$은 0과 7800 사이의 임의의 값을 갖도록 설정된다. 각 저궤도 위성(710a, 710b, 710c)은 초기 좌표에서 출발하여 지상국(720)과 통신 가능한 반경 내의 임의의 점을 지나는 직선 형태의 궤도를 따라 450km/m의 속도로 이동하는 것으로 가정한다. 각각의 위성(710a, 710b, 710c)은 무작위의 다운로드해야 할 데이터를 보유하고 있으며, 보유하는 데이터의 총합은 전체 저궤도 위성들(710a, 710b, 710c)이 지상국(720)과 접촉하는 시간의 합에 비례하는 것으로 설정된다. 그리고, 모든 위성(710a, 710b, 710c)이 지상국(720)과의 통신 가능 반경을 벗어나는 시점에 하나의 에피소드가 끝나는 것으로 가정하며, 매 에피소드에서 위와 같이 임의의 점에서 임의의 데이터를 보유한 채로 임의의 궤도를 따라 출발하는 것을 반복하도록 실험이 설정된다.

[0132] 에이전트는 하나의 액터 네트워크, 2개의 메인 크리틱 네트워크, 및 2개의 타겟 크리틱 네트워크를 가질 수 있다. 각각의 네트워크는 모두 1개의 GAT 층, 4개의 FC 은닉층과 위성 수만큼의 병렬 완전연결 출력층을 갖도록 설계될 수 있다. 다른 시뮬레이션 파라미터는 아래 표 1에 기재된 바와 같이 설계될 수 있다.

표 1

[0133]

파라미터	값
각 은닉층의 뉴런	512
각 출력층의 뉴런	위성 개수 + 지상국 개수
배치 크기	128
리플레이 버퍼 크기	1,000,000
학습률	3×10^{-4}
할인율	0.99
최적화기	Adam
타겟 엔트로피	0.98 x log(위성 개수)
오프로드 가중치(λ)	-0.3
소프트 업데이트 사이클(B)	2
소프트 업데이트 계수 (τ)	5×10^{-3}

[0134] [표 1: 시뮬레이션 파라미터]

[0135] 본 명세서의 일 실시 예에 따른 심층 강화 학습 기반 협력적 다운로드(DRL-CD) 프레임워크의 성능을 평가하기 위한 비교 기법으로는 저궤도 위성이 지상국과 접촉할 때까지 오프로드 결정을 랜덤으로 수행하는 기법(RANDOM)과 주변 위성과의 협력 없이 지상국으로 다운로드를 수행하는 기법(No-COOP)을 고려한다.

[0136] 도 8은 본 명세서의 일 실시 예에 따른 에이전트의 액터 네트워크의 학습에 따른 손실의 수렴 양상을 도시한 도면이다.

[0137] 도 8을 참조하면, 주어진 상태를 기반으로 다음 행동을 선택하는 에이전트의 액터 네트워크의 학습 스텝에 따른 손실의 양상을 나타내는 그래프(800)가 도시된다. 액터 손실(actor loss)은 훈련 시작 후 감소하다가 300k-번째 스텝 부근에서 값이 유지되는 것을 확인할 수 있다. 즉, 본 명세서의 일 실시 예에 따른 DRL-CD를 통해 위성 네트워크 환경에서 에이전트가 동적인 환경을 학습하고 있으므로, 액터의 손실이 발산하지 않고 수렴하는 것을 확인할 수 있다.

[0138] 도 9는 본 명세서의 일 실시 예에 따른 각 기법의 초기 데이터 분포에 따른 위성의 평균 접촉 시간 이용률을 나타내는 도면이다.

[0139] 도 9를 참조하면, 처음 주어진 데이터의 분포에 따라 비교 기법(DRL-CD, RANDOM, No-COOP)의 위성의 지상국과의 접촉 시간의 평균 이용률에 대한 성능 측정 결과를 나타내는 그래프(900)가 도시된다. 그래프(900)는 주어진 데이터 분포에서 각 기법에 대해 100회 시행 후 평균을 계산한 값을 나타낸다. 데이터의 분포는 통신사에서 대역폭 할당에 대한 공정성(fairness)을 측정할 때 사용되는 지표 중 하나인 G 공정성 인덱스(G Fairness Index)로 측정하며, 이는 $\mathcal{I}(\boldsymbol{m}) = \prod_{l \in \mathrm{L}} \sin\left(\frac{\pi \cdot m_l}{2 \cdot \max(\boldsymbol{m})}\right)$ 로 정의된다. $\boldsymbol{m} = \{m_l | l \in \mathrm{L}\}$ 은 저궤도 위성이 초기에 보유하는 데이터의 양의 집합을 나타낸다. 처음 주어진 데이터의 분포가 균일하면 $\mathcal{I}(\boldsymbol{m})$ 은 1에 가까운 값을 갖고, 데이터의 분포가 한쪽에 편향되어 있으면 $\mathcal{I}(\boldsymbol{m})$ 은 0에 가까운 값을 갖는다. 매 애피소드의 시작시, 초기 데이터의 총량은 각 경로의 최대 다운로드 가능한 데이터의 합으로 설정되며, $\mathcal{I}(\boldsymbol{m})$ 에 따라 각 위성에 데이터가 분배된다.

[0140] 비교 기법들과의 비교 결과, 처음 데이터의 분포에 상관없이 DRL-CD가 RANDOM 또는 No-COOP 기법보다 좋은 접촉 시간 이용률을 얻는 것이 확인된다. 특히, 균일한 데이터 분포(즉, $\mathcal{I}(\boldsymbol{m})$ =0.9)에서는 DRL-CD가 RANDOM 및 No-COOP 기법보다 각각 16.4% 및 8.2% 만큼 접촉 시간을 더 효율적으로 이용하는 것이 확인된다. 한편, 편향된 데이터 분포(즉, $\mathcal{I}(\boldsymbol{m})$ =0.1)에서는 DRL-CD는 RANDOM 및 No-COOP 기법도가 각각 15.2% 및 17.8%의 개선된 접촉 시간 이용률을 보여준다. 이는 DRL-CD가 모든 위성이 적절한 양의 데이터를 오프로드하도록 협력적 다운로드 정책을 성공적으로 훈련시키고, 초기 데이터 분포에 관계 없이 주어진 접촉 시간을 효율적으로 이용할 수 있게 하기 때문이다.

[0141] 또한, G 공정성 인덱스가 감소할수록(즉, 초기 데이터 분포가 편향될수록) 접촉 시간 이용률이 감소하는 것을 확인할 수 있는데, 이는 편향된 데이터 분포는 데이터를 과도하게 보유하는 위성이 적은 이용률을 갖는 위성들에게 데이터를 오프로드하는 데 더욱 많은 시간이 걸리게 하기 때문이다. 또한, 동적 위성 네트워크 환경에서 충분한 오프로드 시간을 보장하는 것이 어려울 수 있다.

[0142] 도 10은 본 명세서의 일 실시 예에 따른 각 기법의 저궤도 위성의 개수에 따른 성능의 변화를 나타내는 도면이다.

[0143] 도 10을 참조하면, 저궤도 위성의 개수가 저궤도 위성의 지상국과의 평균 접촉 시간 이용률에 미치는 효과를 나타내는 그래프(1000)가 도시된다. 100개의 에피소드가 무작위로 생성되고, G 공정성 인덱스($\mathcal{I}(\boldsymbol{m})$)는 여기서 0.2 또는 0.8로 설정된다. 도 10에서, 평균 접촉 시간 이용률은 위성의 수가 증가할수록 일반적으로 증가함을 확인할 수 있다. 예를 들어, 고도로 편향된 데이터 분포(즉, $\mathcal{I}(\boldsymbol{m}) = 0.2$)의 경우, DRL-CD, RANDOM 및

No-COOP의 접촉 시간 이용률은 저궤도 위성의 개수가 2에서 5로 증가하는 동안 각각 21.4%, 11.5%, 및 18.4%만큼 증가하였다. 이는 협력적 다운로드를 통해 다운로드되는 데이터의 양이 지상국과의 접촉 시간뿐만 아니라 다른 위성들 간의 접촉 시간에도 크게 의존하기 때문이다. 위성의 개수가 증가할수록, ISL 오프로드의 기회 제공 확률이 높아지고, 이는 과도한 양의 데이터를 보유하는 위성이 적은 양의 데이터를 보유하는 위성으로 데이터를 더 많이 오프로드할 수 있게 한다. 결과적으로, DRL-CD는 더 많은 저궤도 위성을 배치함에 있어서 더욱 효과적이고, 더욱 많은 저궤도 위성이 발사되는 미래에는 본 명세서에서 제안하는 심층 강화 학습 기반의 협력적 다운로드 기법이 데이터 전송 효율을 높이는 데에 상당한 기여를 할 것으로 기대된다.

[0145] 한편, 본 명세서와 도면에는 본 발명의 바람직한 실시 예에 대하여 개시하였으며, 비록 특정 용어들이 사용되었으나, 이는 단지 본 발명의 기술 내용을 쉽게 설명하고 발명의 이해를 돕기 위한 일반적인 의미에서 사용된 것이지, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예 외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다.

부호의 설명

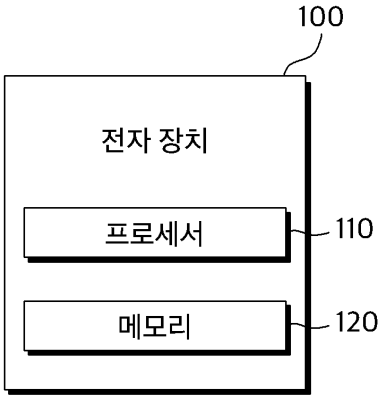
[0146] 100: 전자 장치

110: 프로세서

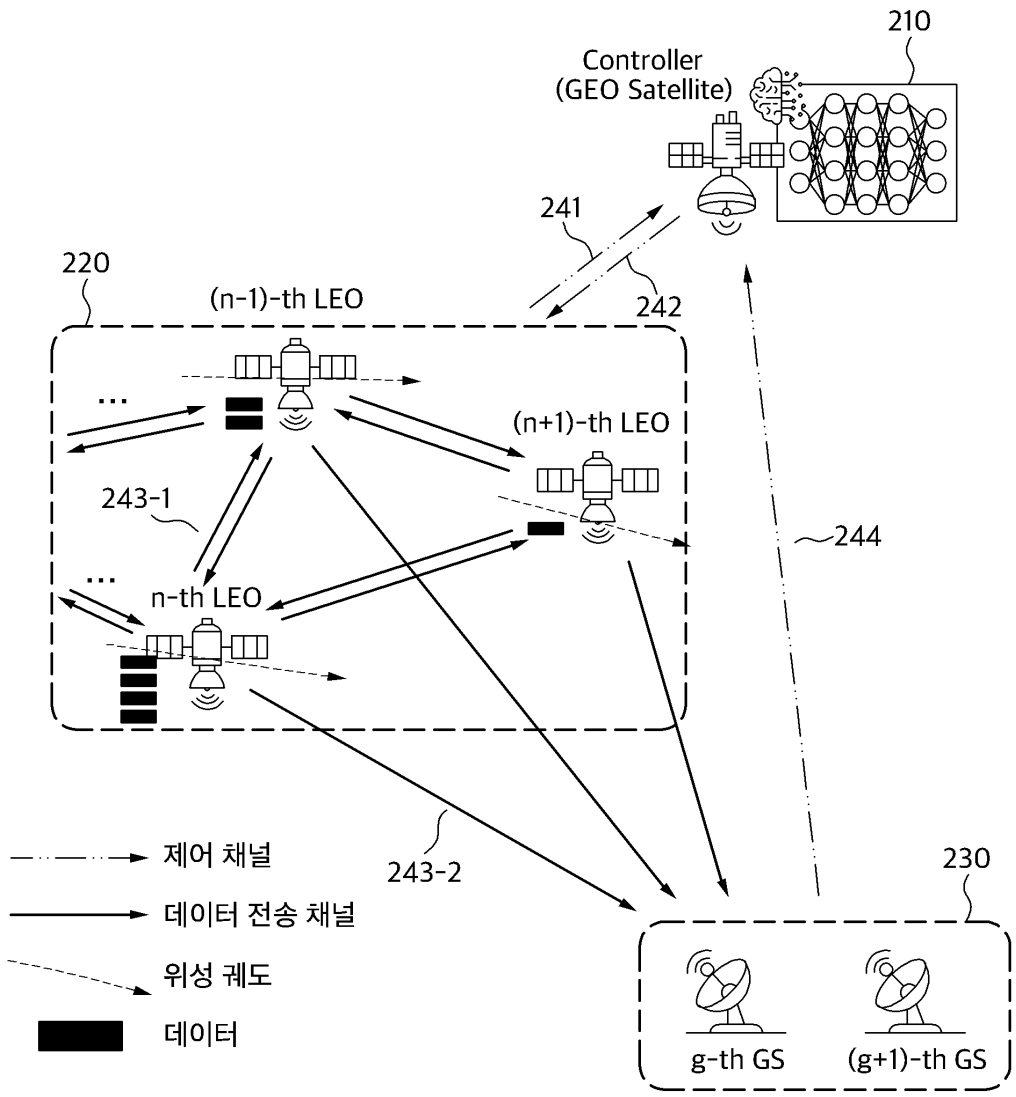
120: 메모리

도면

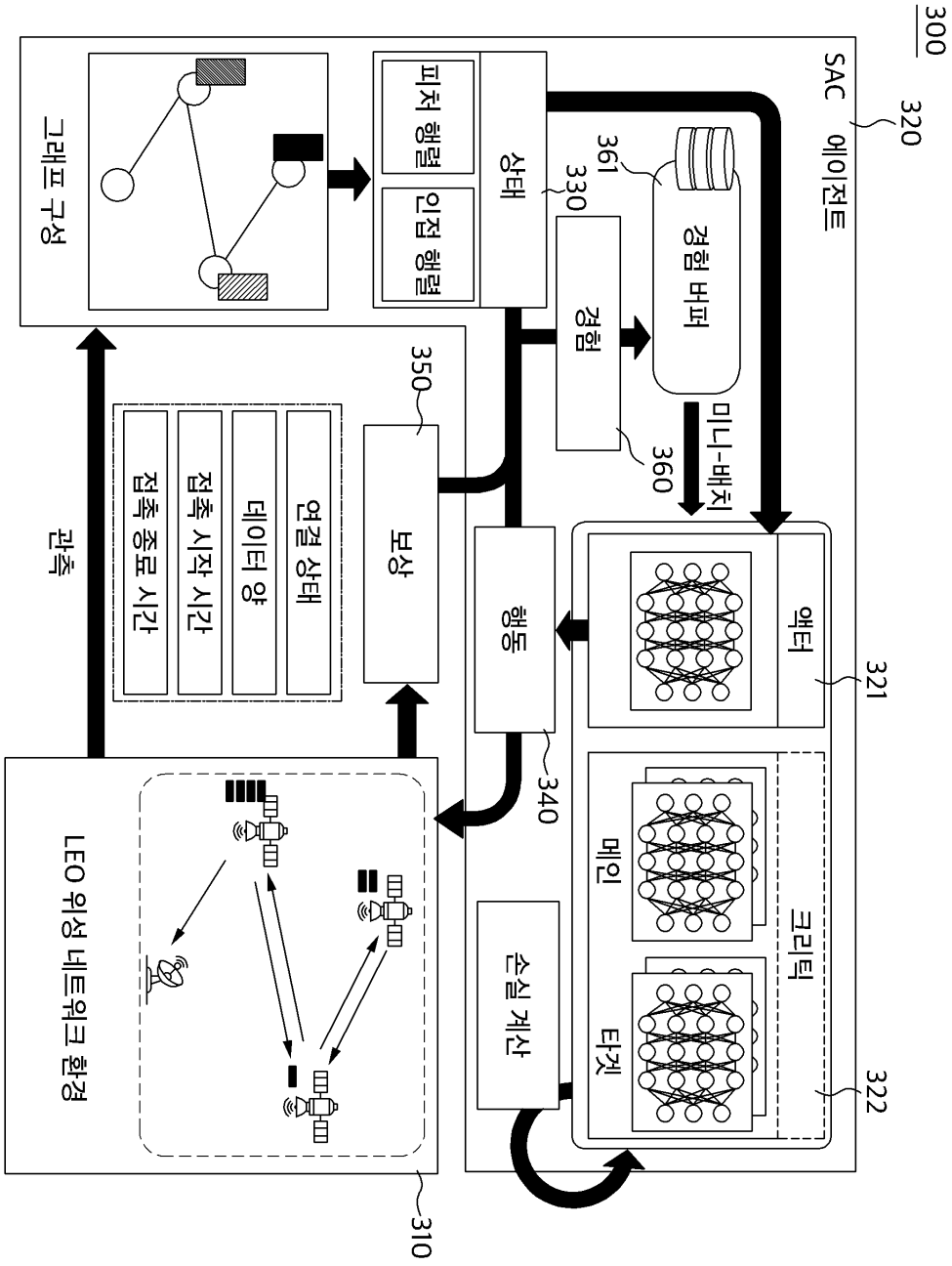
도면1



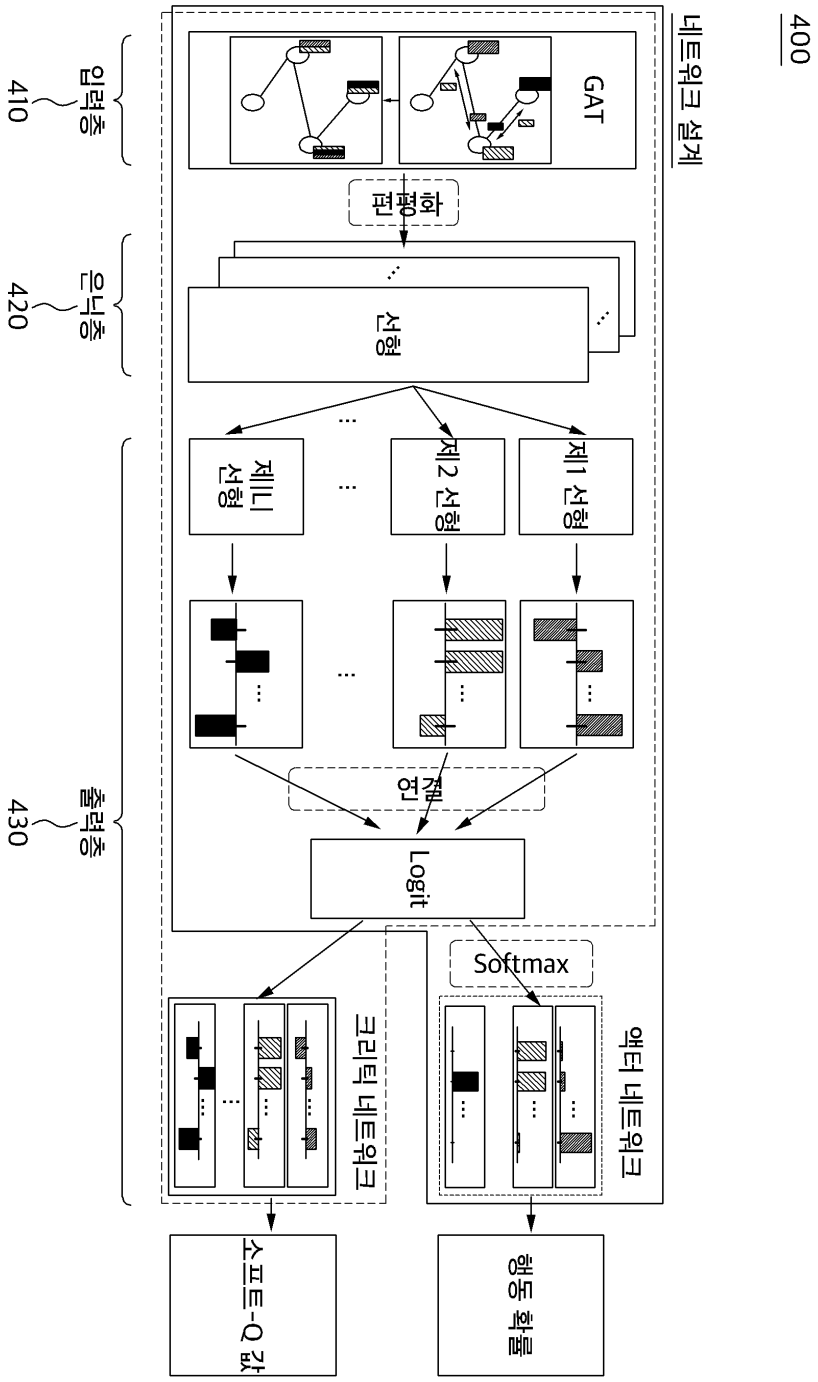
도면2


200
210
Controller
(GEO Satellite)
241
242
220
(n-1)-th LEO
(n+1)-th LEO
243-1
244
n-th LEO
243-2
230
g-th GS
(g+1)-th GS
제어 채널
데이터 전송 채널
위성 궤도
데이터

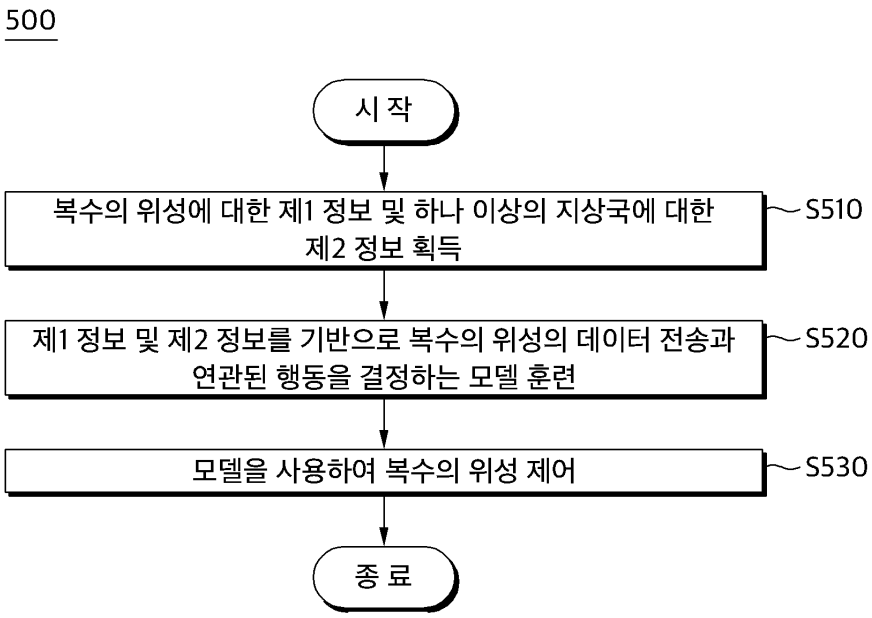
도면3


300
SAC 에이전트
320
그래프 구성
피처 행렬
인접 행렬
상태
330
361
경험 버퍼
경험
360
미니-배치
350
보상
행동
340
액터
321
메인
타겟
크리틱
322
손실 계산
연결 상태
데이터 양
접촉 시작 시간
접촉 종료 시간
관측
LEO 위성 네트워크 환경
310

도면4


400
네트워크 설계
입력층
410
은닉층
420
출력층
430
GAT
평탄화
선형
제1 선형
제2 선형
제N 선형
결합
Logit
Softmax
액터 네트워크
크리틱 네트워크
행동 확률
소프트-Q 값

도면5

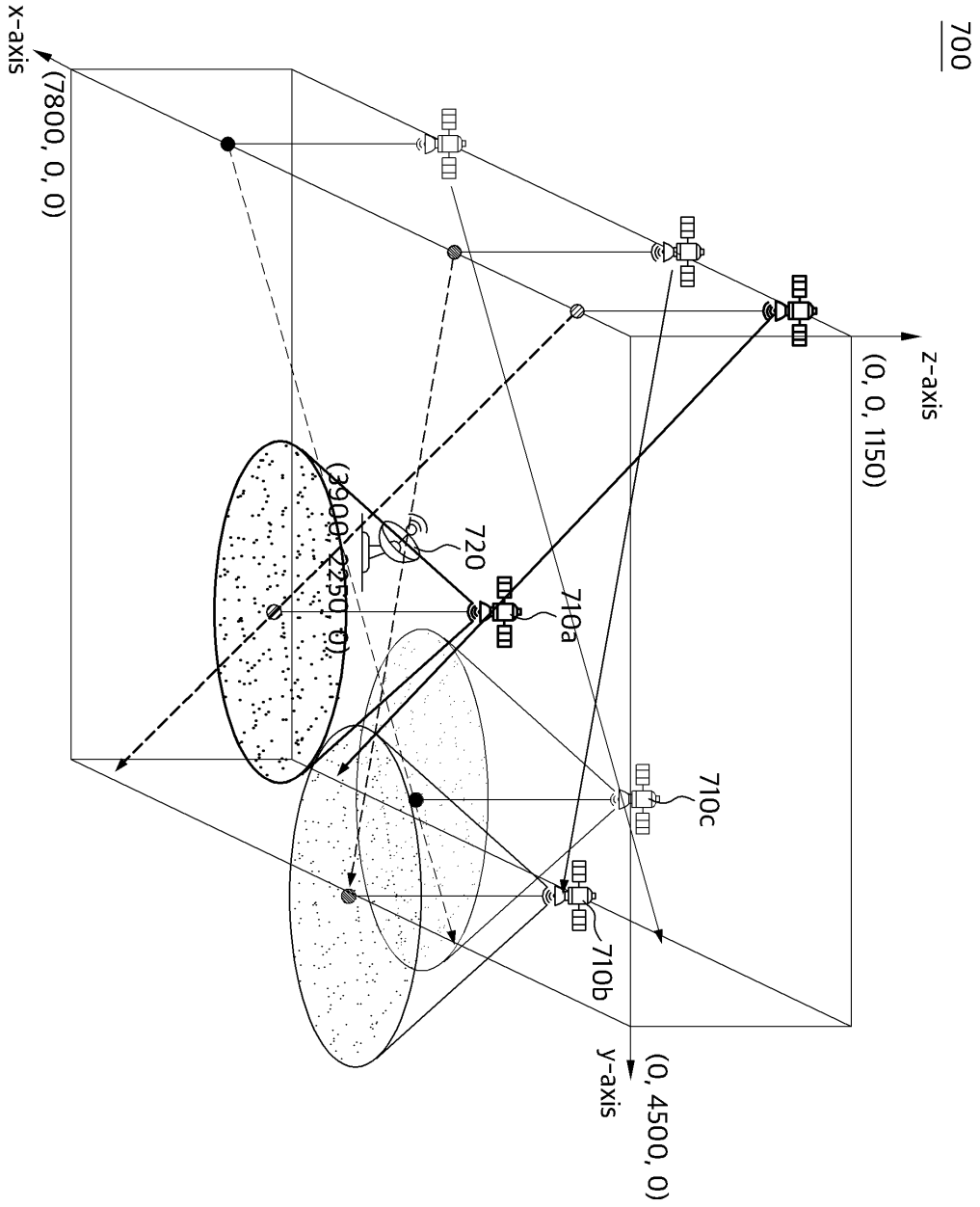


도면6

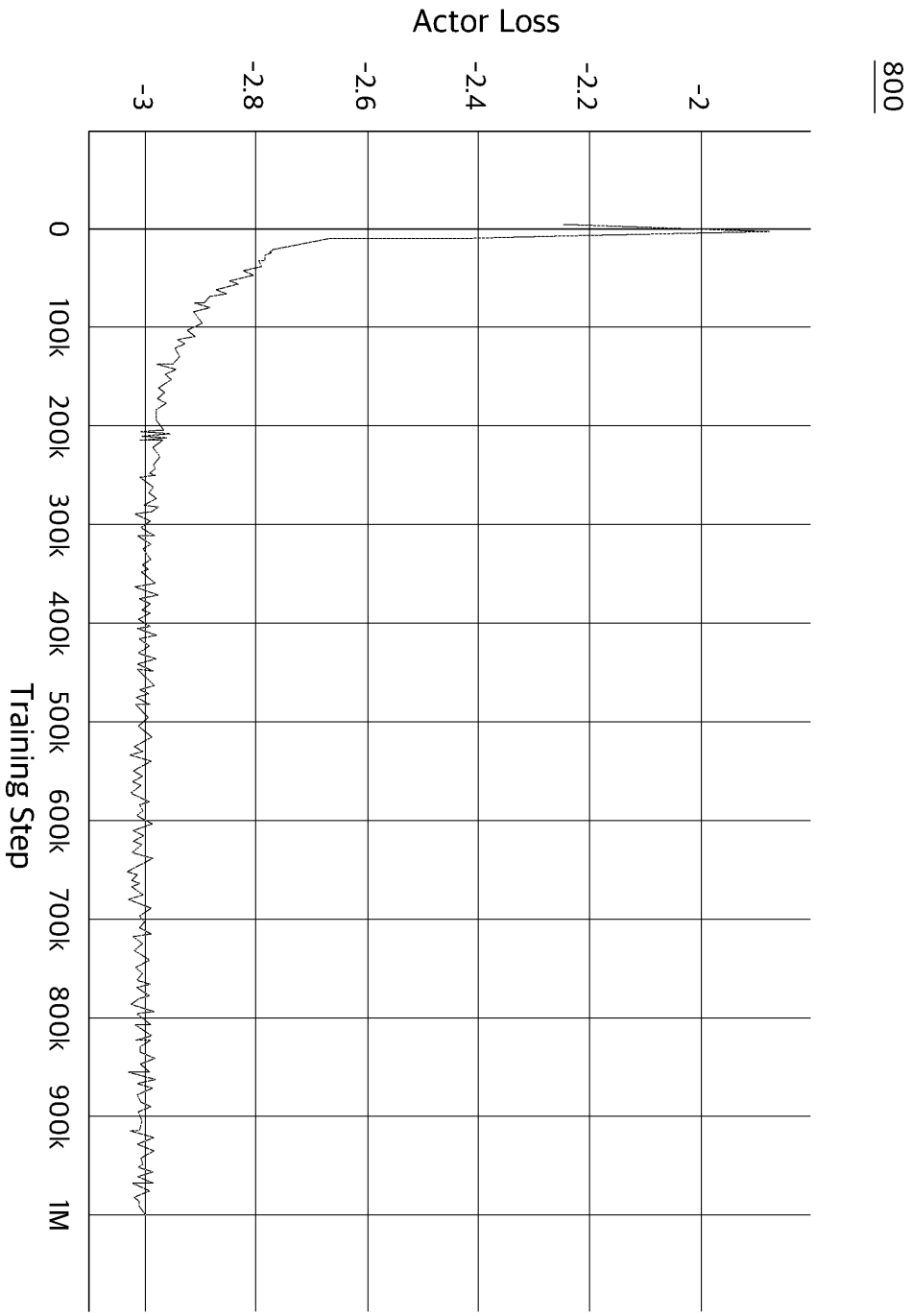
알고리즘 1 이산화된 SAC 알고리즘

Initialize the experience buffer D
Initialize the weight α with 1
Initialize the actor network π_ϕ with random parameter ϕ
Initialize main critic networks Q_{θ_i} with random parameters θ_i for $i \in \{1,2\}$
Initialize target critic networks $Q_{\hat{\theta}_i}$ with parameters $\hat{\theta}_i$ as main critic networks Q_{θ_i}
for each training episode **do**
  Observe initial state s_t
  for each step $t = 1,2,...,T$ **do**
    Generate the action $a_t = \pi_\phi(s_t)$
    Execute the action a_t
    Observe the next state s_{t+1} and the reward r_t
    Store the experience (s_t, a_t, r_t, s_{t+1}) at the experience buffer D
    Sample a mini-batch $\tilde{D}$ of a few experiences from the buffer D
    Calculate the target state value $V_{\hat{\theta}}(s')$ based on Equation (2)
    Update the main critic network Q_{θ_i} based on the gradient $\nabla_{\theta_i} J_Q(\theta_i)$ in Equation (3)
    Update the actor network π_ϕ based on the gradient $\nabla_\phi J_\pi(\phi)$ in Equation (5)
    Update the weight α based on the gradient $\nabla_\alpha J(\alpha)$ in Equation (6)
    For every B steps, use soft update for the target critic networks based on Equation (7)
  end for
end for

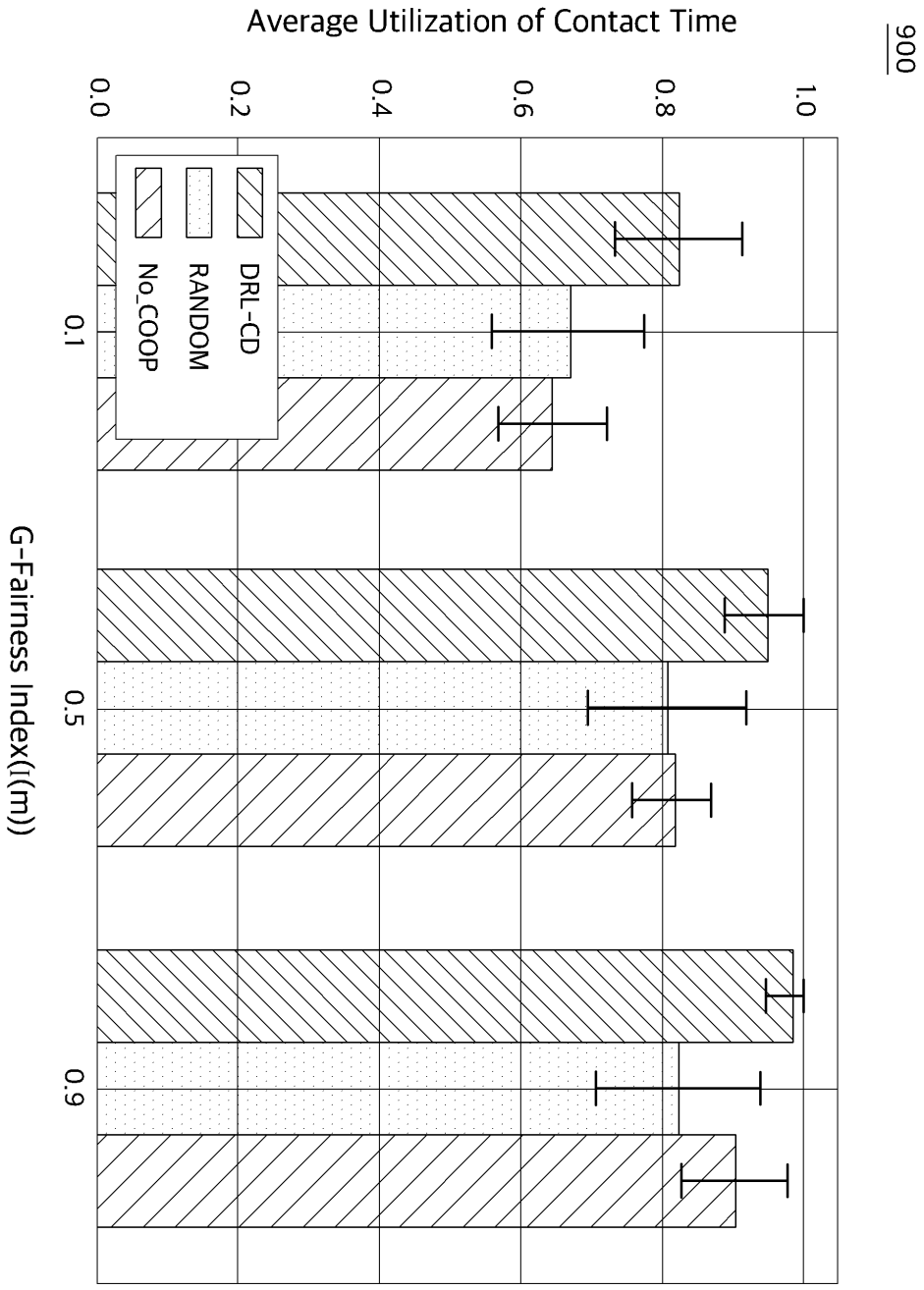
도면7


700
x-axis
(7800, 0, 0)
z-axis
(0, 0, 1150)
y-axis
(0, 4500, 0)
(3900, 2250, 0)
720
710a
710b
710c

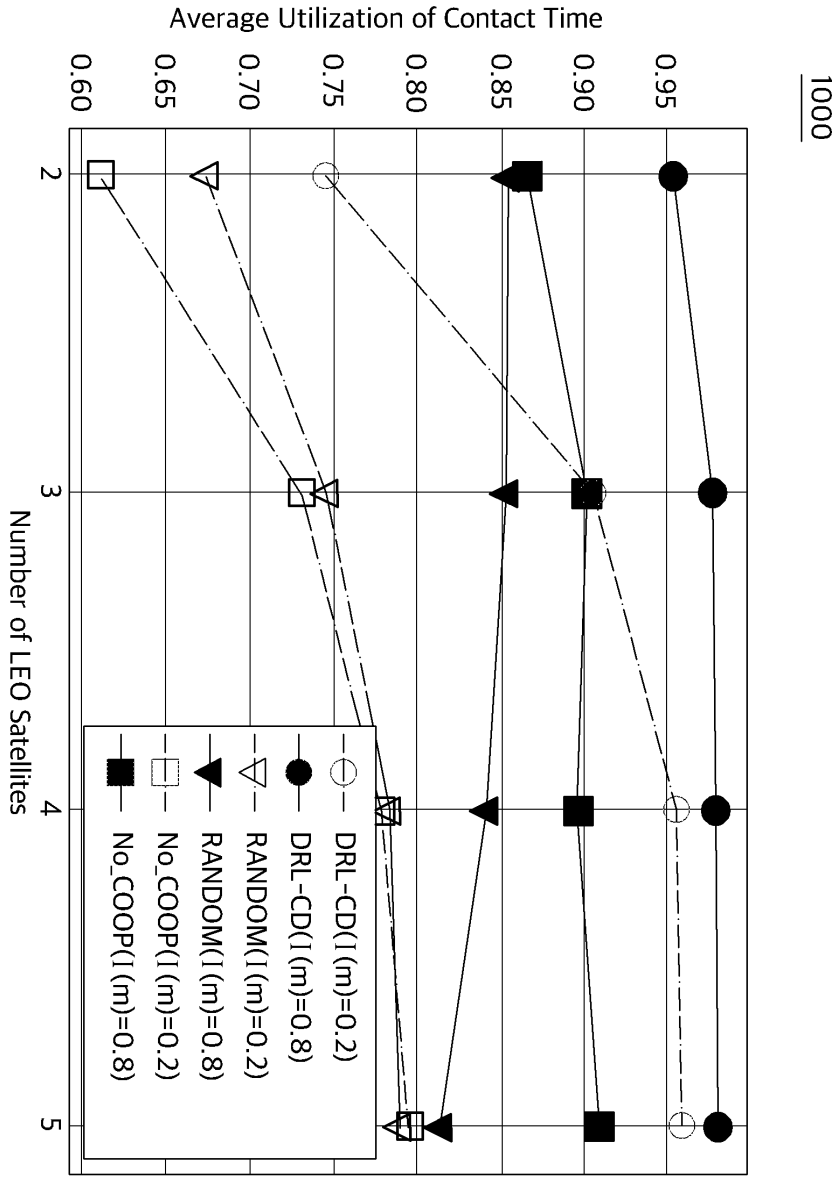
도면8


Actor Loss
-3
-2.8
-2.6
-2.4
-2.2
-2
800
0
100k
200k
300k
400k
500k
600k
700k
800k
900k
1M
Training Step

도면9


Average Utilization of Contact Time
0.0
0.2
0.4
0.6
0.8
1.0
DRL-CD
RANDOM
No_COOP
0.1
0.5
0.9
G-Fairness Index(I(m))
900

도면10


Average Utilization of Contact Time
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1000
Number of LEO Satellites
2
3
4
5
DRL-CD(I(m)=0.2)
DRL-CD(I(m)=0.8)
RANDOM(I(m)=0.2)
RANDOM(I(m)=0.8)
No_COOP(I(m)=0.2)
No_COOP(I(m)=0.8)