



(11) 공개번호 10-2025-0107553
(43) 공개일자 2025년07월14일

(51) 국제특허분류(Int. Cl.)
H04B 7/185 (2006.01) *G06N 3/092* (2023.01)
H02J 50/00 (2021.01) *H02J 50/80* (2016.01)
H04B 7/04 (2017.01) *H04B 7/0413* (2017.01)

(52) CPC특허분류
H04B 7/18502 (2013.01)
G06N 3/092 (2023.01)

(21) 출원번호 10-2024-0002274

(22) 출원일자 2024년01월05일
 심사청구일자 2024년01월05일

(71) 출원인
세종대학교산학협력단
 서울특별시 광진구 능동로 209 (군자동, 세종대학교)
 (72) 발명자
이병무
 서울특별시 송파구 송파대로 567 잠실주공아파트
 502-602
푸스피타사리 아니사 양건
 서울특별시 광진구 군자로 108
 (74) 대리인
김현승

전체 청구항 수 : 총 15 항

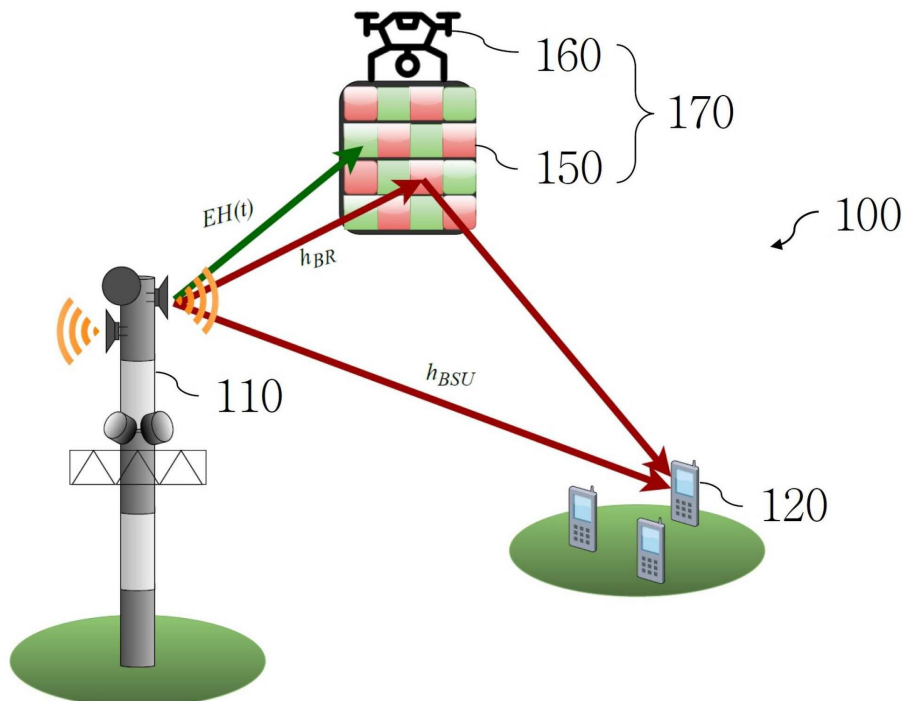
(54) 발명의 명칭 무인 비행체(UAV)-재구성가능 지능형 표면(RIS) 보조식 통신 시스템을 위한 쌍 지연형 심층 결정론적 정책 그래디언트(TD3) 알고리즘 기반 동시 무선 정보 및 전력 전송(SWIPT)

(57) 요약

통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 방법이 제공되는데, 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)을 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, RIS는 복수의 반

(뒷면에 계속)

대표도



사 요소를 포함한다. 제공된 방법은, TD3 알고리즘을 사용하여 에이전트를, BS로부터 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 시간 슬롯에 대해 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면서 시간 슬롯에 대해 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 판정하도록 훈련시키는 단계와, BS로 하여금 통신 시스템을 위한 SWIPT 동작을 수행하게 하기 위해 훈련된 에이전트를 가동하는 단계를 포함하되, 시간 슬롯의 EH 국면 동안에, 복수의 반사 요소 각각은 EH를 위해 할당되고, 시간 슬롯의 정보 전송 국면 동안에, 복수의 반사 요소 중 일부는 UT를 향한 주어진 신호의 반사를 위해 할당되되 복수의 반사 요소 중 나머지는 EH를 위해 할당된다.

(52) CPC특허분류

H02J 50/001 (2023.08)

H02J 50/80 (2023.08)

H04B 7/04013 (2023.05)

H04B 7/0413 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711188561
과제번호	2023R1A2C1002656
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	초고밀도 Massive IoT 연결성 확보를 위한 지능형 거대 다중 안테나 핵심 기술 연구
과제수행기관명	세종대학교
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179335
과제번호	00156345
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	ICT혁신인재4.0(세종대학교)
과제수행기관명	세종대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

통신 시스템을 위한 쌍 지연형 심층 결정론적 정책 그래디언트(Twin Delayed Deep Deterministic policy gradient: TD3) 알고리즘 기반 동시 무선 정보 및 전력 전송(Simultaneous Wireless Information and Power Transfer: SWIPT)을 위한 방법으로서, 상기 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)을 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 상기 RIS는 복수의 반사 요소를 포함하고, 상기 방법은,

TD3 알고리즘을 사용하여 에이전트(agent)를, 상기 BS로부터 상기 UT로의 시간 슬롯(time slot)에서의 송신을 위한 주어진 신호를 포함하는 상태(state)에 기반하여, 상기 시간 슬롯에 대해 상기 UT에서의 적어도 사전결정된 최소값의 용량(capacity)을 달성하면서 상기 시간 슬롯에 대해 상기 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션(action)을 판정하도록 훈련시키는 단계와,

상기 BS로 하여금 상기 통신 시스템을 위한 SWIPT 동작을 수행하게 하기 위해 상기 훈련된 에이전트를 가동하는 단계를 포함하되,

상기 시간 슬롯의 EH 국면 동안에, 상기 복수의 반사 요소 각각은 상기 EH를 위해 할당되고, 상기 시간 슬롯의 정보 전송 국면 동안에, 상기 복수의 반사 요소 중 일부는 상기 UT를 향한 상기 주어진 신호의 반사를 위해 할당되고 상기 복수의 반사 요소 중 나머지는 상기 EH를 위해 할당되는,

방법.

청구항 2

제1항에 있어서,

상기 통신 시스템은 상기 BS 및 상기 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원하는,

방법.

청구항 3

제1항에 있어서,

상기 판정된 액션은 상기 EH 국면의 지속기간(duration)과, 상기 BS의 송신 전력과, 상기 복수의 반사 요소 각각의 위상 천이(phase shift)와, 상기 정보 전송 국면에 걸친 상기 각각의 반사 요소의 상기 할당을 나타내도록 구성된 액션 공간의 원소인,

방법.

청구항 4

제3항에 있어서,

상기 상태는 상기 BS로부터 상기 UT로 송신되는 신호와, 상기 BS 및 상기 RIS 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 채널의 채널 행렬과, 상기 BS 및 상기 UT 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소인,

방법.

청구항 5

제4항에 있어서,

상기 TD3 알고리즘은 행위자 망과, 2개의 비평가 망과, 타겟 행위자 망과, 2개의 타겟 비평가 망을 사용하고, 상기 TD3 알고리즘을 사용하여 상기 에이전트를 훈련시키는 단계는,

상기 행위자 망 및 상기 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 상기 액션 공간 내의 선택된 액션을 반복적으로(iteratively) 판정하는 것 - 상기 선택된 액션의 판정은 상기 상태 공간 내의 새로운 상태가 상기 현재의 상태로 설정되는 것으로써 반복되고, 상기 새로운 상태는 상기 선택된 액션이 취해짐에 따라 관측됨 - 과,

상기 선택된 액션이 판정됨에 응답하여, 경험 재생 버퍼(experience replay buffer) 내에, 상기 선택된 액션, 상기 현재의 상태, 상기 새로운 상태, 그리고 상기 선택된 액션이 적어도 상기 사전결정된 최소값을 갖는 상기 용량을 초과하는 경우에 상기 효율에 기반하여 양의 값으로 주어지는 보상(reward)을 포함한 전이 정보(transition information)를 저장하는 것과,

상기 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치(mini-batch)가 상기 경험 재생 버퍼로부터 이용 가능함과 더불어 상기 전이 정보가 상기 경험 재생 버퍼에 저장됨에 응답하여, 상기 타겟 행위자 망, 상기 2개의 타겟 비평가 망 및 상기 미니배치를 사용하여 상기 2개의 비평가 망을 업데이트하는 것과,

상기 2개의 비평가 망의 상기 업데이트보다 덜 빈번하게, 상기 업데이트된 2개의 비평가 망 중 하나를 사용하여 상기 행위자 망, 상기 타겟 행위자 망 및 상기 2개의 타겟 비평가 망을 업데이트하는 것에 의해 상기 에이전트를 훈련시키는 단계를 포함하는,

방법.

청구항 6

컴퓨터 실행가능 명령어를 포함하는 컴퓨터 판독가능 저장 매체로서, 상기 컴퓨터 실행가능 명령어는 프로세서에 의해 실행되는 경우에 상기 프로세서로 하여금 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 가능하게 하는 동작을 수행하게 하되, 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)를 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 상기 RIS는 복수의 반사 요소를 포함하고, 상기 동작은,

TD3 알고리즘을 사용하여 에이전트를, 상기 BS로부터 상기 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 상기 시간 슬롯에 대해 상기 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면서 상기 시간 슬롯에 대해 상기 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 판정하도록 훈련시키는 것을 포함하되,

상기 시간 슬롯의 EH 국면 동안에, 상기 복수의 반사 요소 각각은 상기 EH를 위해 할당되고, 상기 시간 슬롯의 정보 전송 국면 동안에, 상기 복수의 반사 요소 중 일부는 상기 UT를 향한 상기 주어진 신호의 반사를 위해 할당되고 상기 복수의 반사 요소 중 나머지는 상기 EH를 위해 할당되는,

컴퓨터 판독가능 저장 매체.

청구항 7

제6항에 있어서,

상기 통신 시스템은 상기 BT 및 상기 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원하는,

컴퓨터 판독가능 저장 매체.

청구항 8

제7항에 있어서,

상기 판정된 액션은 상기 EH 국면의 지속기간과, 상기 BS의 송신 전력과, 상기 복수의 반사 요소 각각의 위상 천이와, 상기 정보 전송 국면에 걸친 상기 각각의 반사 요소의 상기 할당을 나타내도록 구성된 액션 공간의 원소인,

컴퓨터 판독가능 저장 매체.

청구항 9

제8항에 있어서,

상기 상태는 상기 BS로부터 상기 UT로 송신되는 신호와, 상기 BS 및 상기 RIS 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 채널의 채널 행렬과, 상기 BS 및 상기 UT 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소인,

컴퓨터 판독가능 저장 매체.

청구항 10

제9항에 있어서,

상기 TD3 알고리즘은 행위자 망과, 2개의 비평자 망과, 타겟 행위자 망과, 2개의 타겟 비평자 망을 사용하고, 상기 TD3 알고리즘을 사용하여 상기 에이전트를 훈련시키는 것은,

상기 행위자 망 및 상기 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 상기 액션 공간 내의 선택된 액션을 반복적으로 판정하는 것 - 상기 선택된 액션의 판정은 상기 상태 공간 내의 새로운 상태가 상기 현재의 상태로 설정되는 것으로써 반복되고, 상기 새로운 상태는 상기 선택된 액션이 취해짐에 따라 관측됨 - 과,

상기 선택된 액션이 판정됨에 응답하여, 경험 재생 버퍼 내에, 상기 선택된 액션, 상기 현재의 상태, 상기 새로운 상태, 그리고 상기 선택된 액션이 적어도 상기 사전결정된 최소값을 갖는 상기 용량을 초래하는 경우에 상기 효율에 기반하여 양의 값으로 주어지는 보상을 포함한 전이 정보를 저장하는 것과,

상기 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치가 상기 경험 재생 버퍼로부터 이용가능함과 더불어 상기 전이 정보가 상기 경험 재생 버퍼에 저장됨에 응답하여, 상기 타겟 행위자 망, 상기 2개의 타겟 비평자 망 및 상기 미니배치를 사용하여 상기 2개의 비평자 망을 업데이트하는 것과,

상기 2개의 비평자 망의 상기 업데이트보다 덜 빈번하게, 상기 업데이트된 2개의 비평자 망 중 하나를 사용하여 상기 행위자 망, 상기 타겟 행위자 망 및 상기 2개의 타겟 비평자 망을 업데이트하는 것에 의해 상기 에이전트를 훈련시키는 것을 포함하는,

컴퓨터 판독가능 저장 매체.

청구항 11

통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 장치로서, 상기 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)을 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 상기 RIS는 복수의 반사 요소를 포함하고, 상기 장치는,

프로세서와,

상기 BS로 하여금 상기 통신 시스템을 위한 SWIPT 동작을 수행하게 하도록 TD3 에이전트를 가동하기 위해 상기 프로세서에 의해 실행가능한 명령어의 세트으로써 인코딩된 메모리를 포함하되, 상기 TD3 에이전트는, 상기 BS로부터 상기 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 상기 시간 슬롯에 대해 상기 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면서 상기 시간 슬롯에 대해 상기 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 판정하도록 훈련되되,

상기 시간 슬롯의 EH 국면 동안에, 상기 복수의 반사 요소 각각은 상기 EH를 위해 할당되고, 상기 시간 슬롯의 정보 전송 국면 동안에, 상기 복수의 반사 요소 중 일부는 상기 UT를 향한 상기 주어진 신호의 반사를 위해 할당되되 상기 복수의 반사 요소 중 나머지는 상기 EH를 위해 할당되는,

장치.

청구항 12

제11항에 있어서,

상기 통신 시스템은 상기 BT 및 상기 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원하는,

장치.

청구항 13

제11항에 있어서,

상기 판정된 액션은 상기 EH 국면의 지속기간과, 상기 BS의 송신 전력과, 상기 복수의 반사 요소 각각의 위상 천이와, 상기 정보 전송 국면에 걸친 상기 각각의 반사 요소의 상기 할당을 나타내도록 구성된 액션 공간의 원소인,

장치.

청구항 14

제13항에 있어서,

상기 상태는 상기 BS로부터 상기 UT로 송신되는 신호와, 상기 BS 및 상기 RIS 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 채널의 채널 행렬과, 상기 BS 및 상기 UT 간의 채널의 채널 행렬과, 상기 RIS 및 상기 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소인,

장치.

청구항 15

제14항에 있어서,

상기 TD3 알고리즘은 행위자 망과, 2개의 비평자 망과, 타겟 행위자 망과, 2개의 타겟 비평자 망을 사용하고, 상기 TD3 에이전트는,

상기 행위자 망 및 상기 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 상기 액션 공간 내의 선택된 액션을 반복적으로 판정하는 것 - 상기 선택된 액션의 판정은 상기 상태 공간 내의 새로운 상태가 상기 현재의 상태로 설정되는 것으로써 반복되고, 상기 새로운 상태는 상기 선택된 액션이 취해짐에 따라 관측됨 - 과,

상기 선택된 액션이 판정됨에 응답하여, 경험 재생 버퍼 내에, 상기 선택된 액션, 상기 현재의 상태, 상기 새로운 상태, 그리고 상기 선택된 액션이 적어도 상기 사전결정된 최소값을 갖는 상기 용량을 초과하는 경우에 상기 효율에 기반하여 양의 값으로 주어지는 보상을 포함한 전이 정보를 저장하는 것과,

상기 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치가 상기 경험 재생 버퍼로부터 이용가능함과 더불어 상기 전이 정보가 상기 경험 재생 버퍼에 저장됨에 응답하여, 상기 타겟 행위자 망, 상기 2개의 타겟 비평자 망 및 상기 미니배치를 사용하여 상기 2개의 비평자 망을 업데이트하는 것과,

상기 2개의 비평자 망의 상기 업데이트보다 덜 빈번하게, 상기 업데이트된 2개의 비평자 망 중 하나를 사용하여 상기 행위자 망, 상기 타겟 행위자 망 및 상기 2개의 타겟 비평자 망을 업데이트하는 것에 의해 훈련된,

장치.

발명의 설명

기술 분야

[0001] 본 개시는 무인 비행체(Unmanned Aerial Vehicle: UAV)-재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS) 보조가 있는 통신 시스템을 위한 동시 무선 정보 및 전력 전송(Simultaneous Wireless Information and Power Transfer: SWIPT)에 관련되고, 더욱 구체적으로, UAV-RIS 보조식(assisted) 통신 시스템을 위한 쌍 지연형 심층 결정론적 정책 그라디언트(Twin Delayed Deep Deterministic policy gradient: TD3) 알고리즘 기반 SWIPT에 관련된다.

배경 기술

- [0002] 더 빠르고 더 양질의 데이터 송신에 대한 늘어나는 필요를 충족시키기 위해 무선 통신 기술이 꾸준히 발전되었다. 유망한 기술 중에서, 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO)은 스펙트럼 및 에너지 효율을 향상시키고 저지연시간(low latency)으로 다량의 데이터 트래픽(traffic)을 다룰 수 있다는 두드러진 능력을 갖는다(B. M. Lee, "Exploring the impact of power control strategies for enhanced iot connectivity in massive mimo," IEEE Internet of Things Journal, pp. 1-1, 2023).
- [0003] RIS는 종래의 릴레이(relay) 기술로부터 나온 것인데, 의도된 수신기에 전자기파를 반사하기 위해 인공적인 평탄 표면을 사용한다. RIS 보조식 무선 통신 방안을 이용하여 다양한 성능 메트릭(metric), 예를 들어 달성가능한 합산 레이트(sum rate), 용량(capacity) 내지 서비스 품질(Quality of Service: QoS)을 최적화하는 데에 연구 노력이 집중되었다(C. Xu, J. An, T. Bai, S. Sugiura, R. G. Maunder, Z. Wang, L.-L. Yang, and L. Hanzo, "Channel estimation for reconfigurable intelligent surface assisted high-mobility wireless systems," IEEE Transactions on Vehicular Technology, vol. 72, no. 1, pp. 718-734, 2023). 또한, RIS는 MIMO 통신에서 스펙트럼 및 에너지 효율을 향상시키는 역량을 보여주었다(N. T. Nguyen, Q.-D. Vu, K. Lee, and M. Juntti, "Hybrid relay-reflecting intelligent surface-assisted wireless communications," IEEE Transactions on Vehicular Technology, vol. 71, no. 6, pp. 6228-6244, 2022). UAV는 제어 특징 및 적응가능성(adaptability)으로 인해, 이전에 액세스할 수 없던 지역이나 비가시선(Non-Line-Of-Sight: NLOS) 조건에 영향을 받는 지역에서 네트워크의 제공을 원활하게 한다(T. V. Tung, T. T. An, and B. M. Lee, "Joint resource and trajectory optimization for energy efficiency maximization in uav-based networks," Mathematics, vol. 10, no. 20, 2022). 그렇지만, 이러한 기술은 UAV의 온보드(onboard) 배터리 용량이 한정된다는 새로운 난관에 놓이기 일쑤다.
- [0004] 나아가, 지상의 센서를 위한 무선 전력 전송 시스템에서 에너지 최소화 문제를 다루기 위해 RIS가 UAV와 통합된 바 있다(H. Ren, Z. Zhang, Z. Peng, L. Li, and C. Pan, "Energy minimization in ris-assisted uav-enabled wireless power transfer systems," IEEE Internet of Things Journal, vol. 10, no. 7, pp. 5794-5809, 2023). SWIPT에 대해 RIS-UAV 시스템에서의 에너지 효율을 살펴본 추가의 연구가 있다(X. Zhang, H. Zhang, W. Du, K. Long, and G. K. Karagiannidis, "Joint resource allocation and reflecting design in irs-uav communication networks with swipt," IEEE Transactions on Wireless Communications, pp. 1-1, 2023). 다만, 에너지 효율을 달성하면서 동시에 쓰루풋(throughput)을 개선하는 것은 비볼록 최적화(non-convex optimization)의 높은 복잡도로 인해 만만찮은 난제로 남아 있다.

발명의 내용

해결하려는 과제

- [0005] UAV-RIS 보조식 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT가 본 문서에 개시된다.

과제의 해결 수단

- [0006] 예에서, 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 방법(통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)를 포함하고, UAV에 탑재된 재구성가능 RIS에 의해 보조되고, RIS는 복수의 반사 요소를 포함함)은, TD3 알고리즘을 사용하여 에이전트를, BS로부터 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 시간 슬롯에 대해 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면서 시간 슬롯에 대해 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 판정하도록 훈련시키는 단계와, BS로 하여금 통신 시스템을 위한 SWIPT 동작을 수행하게 하기 위해 훈련된 에이전트를 가동하는 단계를 포함하되, 시간 슬롯의 EH 국면 동안에, 복수의 반사 요소 각각은 EH를 위해 할당되고, 시간 슬롯의 정보 전송 국면 동안에, 복수의 반사 요소 중 일부는 UT를 향한 주어진 신호의 반사를 위해 할당되되 복수의 반사 요소 중 나머지는 EH를 위해 할당된다.
- [0007] 전술된 개요는 상세한 설명에서 추가로 후술되는 몇몇 양상을 단순화된 형태로 소개하기 위해 제공된다. 이 개요는 청구된 주제(subject matter)의 중요 특징 또는 필수적 특징을 식별하도록 의도되지 않고, 청구된 주제의 범위를 정하는 데 사용되도록 의도되지도 않는다. 나아가, 청구된 주제는 본 명세서에서 논의되는 임의의 또는 모든 이점을 제공하는 구현에 한정되지 않는다.

발명의 효과

[0008] 본 개시에 따르면, UAV-RIS 보조식 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT에서 UAV-RIS의 에너지 수확(Energy Harvesting: EH) 효율을 최대화하면서 사용자 단말의 용량을 보장할 수 있다.

[0009] 본 개시에 따르면, 과대조정 바이어스의 우려가 있는 DDPG 알고리즘에 비해, 그리고 RIS에서의 EH 및 정보 전송 간의 시간 스위칭(time switching)만을 이용하는 방안에 비해, 성능 향상을 달성할 수 있다.

도면의 간단한 설명

[0010] 도 1은 예시적인 SWIPT가 적용가능한 UAV-RIS 보조식 MIMO 통신 시스템의 예를 보여주는 개략도이다.

도 2는 도 1의 MIMO 통신 시스템을 위한 예시적인 에너지 수확(Energy Harvesting: EH) 방안을 설명하기 위한 도면이다.

도 3은 도 1의 MIMO 통신 시스템에서 적용가능한 TD3 알고리즘의 예를 의사코드(pseudocode)의 형태로 제시한다.

도 4 내지 도 6은 도 3과 관련하여 설명된 예시적인 TD 알고리즘의 시뮬레이션의 결과를 보여주는 도면이다.

도 7은 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 프로세스의 예를 보여준다.

발명을 실시하기 위한 구체적인 내용

[0011] 본 개시에서 사용되는 다양한 용어는 본 문서에서의 기능을 고려하여 상용 용어의 용어범으로부터 선택되는데, 이는 당업자의 의도, 준례, 또는 새로운 기술의 출현에 따라서 달리 인식될 수 있다. 특정한 사례에서, 몇몇 용어에는 상세한 설명에서 개진된 바와 같이 의미가 주어질 수 있다. 따라서, 본 문서에서 사용되는 용어는, 단순히 그 명칭에 의해서가 아니라, 본 개시의 맥락에서 그 용어가 갖는 의미와 일관되게 정의되어야 한다.

[0012] 본 문서에서 용어 "포함하다", "가지다" 등은 이후에 열거된 요소, 예컨대, 어떤 특징, 숫자, 단계, 동작, 구성 요소, 정보 또는 이들의 조합의 존재를 명시하는 경우에 사용된다. 달리 표시되지 않는 한, 이런 용어 및 이의 변형은 다른 요소의 존재 또는 추가를 배제하도록 의도되지 않는다.

[0013] 본 문서에서 사용되는 바와 같이, 용어 "제1", "제2" 등은 몇 개의 서로 닮은 요소를 식별하도록 의도된다. 달리 기재되지 않는 한, 그러한 용어는 이들 요소의 또는 이들의 사용의 특정한 순서와 같은 한정을 부과하도록 의도된 것이 아니라, 단지 여러 요소를 따로따로 지칭하기 위해 사용된다. 예를 들면, 어떤 요소가 일례에서 용어 "제1"로써 참조될 수 있는 한편 동일한 요소가 다른 예에서 "제2" 또는 "제3"과 같은 상이한 서수로써 참조될 수 있다. 그러한 예에서, 이들 용어는 본 개시의 범위를 한정하지 않는 것이다. 또한, 여러 요소의 리스트에서 용어 "및/또는"을 사용하는 것은 열거된 항목 중 임의의 하나 또는 복수 개를 비롯하여 이들 항목의 모든 가능한 조합을 포함한다. 나아가, 단수 형태의 표현은 명확하게 달리 사용되지 않는 한, 복수 형태의 의미를 포함한다.

[0014] 첨부된 도면을 참조하여 본 개시의 소정의 예가 이제 상세히 기술될 것이다. 다만, 본 개시는 많은 상이한 형태로 구현될 수 있으며, 본 문서에 개진된 예에 한정되는 것으로 해석되어서는 안 된다. 오히려, 이들 예는 본 개시의 범위의 더 나은 이해를 제공하기 위해서 주어지는 것이다.

[0015] 도면에 묘사된 개별 특징 중 몇몇은 서로 독립적으로 구현되거나 조합되어 구현될 수 있음이 고려된다. 다양한 예의 이들 특징 및 다른 특징은 첨부된 도면과 함께 이하의 상세한 설명으로부터 명백하게 될 것이다.

시스템 모델

[0017] 도 1은 예시적인 SWIPT가 적용가능한 UAV-RIS 보조식 MIMO 통신 시스템(100)의 예를 보여주는 개략도이다.

[0018] 도시된 예에서, MIMO 통신 시스템(100)은 K개의 안테나(K는 1보다 큰 정수임)를 갖는 베이스 스테이션(Base Station: BS)(110) 및 L개의 안테나(L은 1보다 큰 정수임)를 갖는 사용자 단말(User Terminal: UT)(120)을 포함하고, RIS(150)을 더 포함하며 이에 의해 보조된다. RIS(150)은 UAV(160) 상에 탑재되는데(carried), 이들 양자 모두는 집합적으로 UAV-RIS(170)로 지칭될 수 있다. 도 1에 묘사된 바와 같이, BS(110)는 UT(120) 및 UAV-RIS(170)와 통신할 수 있다. 몇몇 예시적인 환경에서, 어떤 장애물이 BS(110) 및 UT(120) 간의 가시선(Line Of Sight: LOS) 링크를 차단하도록 놓일 수 있다.

[0019] 도 1의 예에서, RIS(150)은 복수의 반사 요소를 갖는다. 이하에서, 논의의 편의를 위해, RIS(150)의 반사 요소의 총 개수는 $M \times N$ 로서, 이들 반사 요소는 M개의 행 및 N개의 열에 배열된다고 가정된다. RIS(150)의 반사 요소

중에서 제i 행 및 제j 열에 배치된 것은 $R_{i,j}$ 로 표기될 수 있다. 또한, 예로서, RIS(150)에 마이크로제어기(microcontroller)가 더 구비될 수 있는데, 이는 채널 상태 정보(Channel State Information: CSI)와 같은 정보를 생성하여 (예컨대, BS(110)에) 발신하고/거나 (예컨대, BS(110)로부터) 수신하여 처리할 수 있다.

[0020] 도 1의 예에서, UAV-RIS(170)에 충전가능 배터리가 구비되는데, 이는 아래에서 더욱 상세히 논의는 바와 같이 RIS(150)에 의해 수확된(harvested) 에너지를 보유할 수 있다. 예로서, 그러한 배터리는 UAV(160)에 온보드로 장착될 수 있고, UAV(160)에, 그리고 가능하게는 RIS(150)에도, 전력을 공급할 수 있다.

[0021] 도 2는 MIMO 통신 시스템(100)을 위한 예시적인 에너지 수확(Energy Harvesting: EH) 방안을 설명하기 위한 도면이다. 도 2에 도시된 바와 같이, 신호 송신을 위한 전체 시간 기간은 T개의 동일 길이(equal-length) 시간 슬롯으로 구분될 수 있다. 각각의 시간 슬롯 $t \in \{1, 2, \dots, T\}$ 는 EH 국면(EH phase) 및 정보 전송 국면(information transfer phase)으로 구분될 수 있는데, 이들 양자의 시간 길이 비율은 $\tau(t) : 1 - \tau(t)$ 이다(다시 말해, 시간 슬롯 t의 지속기간(duration)이 하나의 단위 시간이라면, EH 국면의 지속기간은 $\tau(t)$ 이고, 정보 전송 국면의 지속기간은 $1 - \tau(t)$ 임). 그러한 시간 스위칭(time-switching) 모델에서, 시간 슬롯 t의 EH 국면 동안에 RIS(150)의 복수의 반사 요소 각각은 EH를 위해 할당되는(allocated) 한편, 시간 슬롯 t의 정보 전송 국면 동안에 RIS(150)의 복수의 반사 요소의 할당은 다음과 같이 행해질 수 있다: RIS(150)의 복수의 반사 요소 전부를 UT(120)로의 신호 반사를 위해 할당하는 것(이는 "시간 도메인" 방안으로 칭해질 수 있음) 대신에, RIS(150)의 복수의 반사 요소 중 일부분은 UT(120)로의 신호 반사를 위해 할당하고 RIS(150)의 복수의 반사 요소 중 나머지는 EH를 위해 할당하는 것(이는 "시간-공간 도메인" 방안으로 칭해질 수 있음).

[0022] 구체적으로, MIMO 통신 시스템(100)에서 프리코딩(precoding) 프로세스가 구현된 예에서, BS(110)의 제k 안테나($k \in \{1, 2, \dots, K\}$)로부터 송신되는 신호 X_k 는 다음과 같이 표현될 수 있다.

수학식 1

$$X_k = V_k X$$

[0023]

[0024] 여기서 X는 UT(120)를 위해 의도된 데이터 심볼을 신도록 인코딩된 신호를 나타내고, V_k 는 BS(110)의 제k 안테나 상에서의 송신을 위한 프리코딩을 나타낸다. 위 식은 다음과 같이 고쳐 쓸 수 있다.

수학식 2

$$\bar{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix} = \begin{bmatrix} V_1 \\ \vdots \\ V_K \end{bmatrix} X = \bar{V} X$$

[0025]

[0026] 예를 들어, 신호 X의 각각의 성분이 평균이 0이고 분산이 1인 순환 대칭 복소 가우시안 신호(circularly symmetric complex Gaussian signal)(즉, $X \sim \mathcal{CN}(0,1)$)이도록 BS(110)가 가우시안 코드북(Gaussian codebook)과 함께 프리코딩을 이용한다면, BS(110)에서의 총 송신 전력에 대한 다음의 제약이 고려될 수 있다.

수학식 3

$$\mathbb{E}(\bar{X}^H \bar{X}) = \sum \|V_k\|^2 \leq P_{max}$$

[0027]

[0028] 여기서 $\sum \|V_k\|^2$ 는 $\bar{X}$ 의 공분산 $\bar{V} \bar{V}^H$ 의 트레이스(trace)로서 UT(120)를 위한 송신 전력을 나타내고, P_{max} 는 BS(110)에서 허용되는 총 전송 전력을 나타낸다.

[0029] 그러면, 시간-공간 도메인 방안에 따르면, 시간 슬롯 t에서 UAV-RIS(170)(특히, RIS(150))에 의해 수확된 에너지

지는 다음과 같이 표현될 수 있다.

수학식 4

$$EH(t) = \tau(t) \sum_{i=1}^M \sum_{j=1}^N \eta \|h_{BR_{i,j}}^H \bar{X}\|^2 + (1 - \tau(t)) \sum_{i=1}^M \sum_{j=1}^N \omega_{i,j} \eta \|h_{BR_{i,j}}^H \bar{X}\|^2$$

여기서 $h_{BR_{i,j}}$ 는 BS(110)로부터 RIS(150)의 반사 요소 $R_{i,j}$ 로의 채널, 예컨대, 소규모 페이딩 채널(small-scale fading channel)을 나타내는 벡터로서 $h_{BR_{i,j}}^H = [h_{R_{i,j}}^1, \dots, h_{R_{i,j}}^K]$ 로 표현될 수 있는데, $h_{R_{i,j}}^k$ 는 BS(110)의 제k 안테나로부터 반사 요소 $R_{i,j}$ 로의 채널의 이득이고, η 는 에너지 변환 효율 인자(energy conversion efficiency factor)로서 $0 \leq \eta \leq 1$ 이며, $\omega_{i,j}$ 는 $R_{i,j}$ 가 신호 반사를 위해 채택된 경우에는 0으로, 그리고 $R_{i,j}$ 가 에너지 수확을 위해 채택된 경우에는 1로 설정되는 계수이다(이하에서 $\{\omega_{1,1}, \dots, \omega_{M,N}\}$ 는 ω 로 표기될 수 있음).

또한, 시간 슬롯 t에서 UAV-RIS(170))(특히, RIS(150))에 의해 수신된 총 에너지는 다음과 같이 표현될 수 있다.

수학식 5

$$P_r(t) = \sum_{i=1}^M \sum_{j=1}^N \|h_{BR_{i,j}}^H \bar{X}\|^2$$

그러면, 시간 슬롯 t에 대한 EH 효율 $\varepsilon(t)$ 은 다음과 같이 표현될 수 있다.

수학식 6

$$\varepsilon(t) = \frac{EH(t)}{P_r(t)}$$

나아가, 시간 슬롯 t의 정보 전송 국면에서, BS(110)로부터 RIS(150)로의 채널을 나타내는 채널 행렬을 h_{BR} , 그리고 RIS(150)로부터 UT(120)로의 채널을 나타내는 채널 행렬을 h_{RU} 라고 할 때($h_{BR}^H \in \mathbb{C}^{(M \times N) \times K}$ 이고 $h_{RU}^H \in \mathbb{C}^{1 \times (M \times N)}$ 임), BS(110)로부터 UAV-RIS(170)를 거쳐 UT(120)에 의해 수신된 신호 y_{BRU} 는 다음과 같이 표현될 수 있다.

수학식 7

$$y_{BRU} = h_{RU}^H \Phi^H h_{BR}^H \bar{X} + \omega_{AWGN}$$

여기서 Φ 는 RIS(150)에서의 위상 천이(phase shift) 행렬로서 $\Phi = \text{diag}(\phi_{1,1}, \dots, \phi_{M,N}) \in \mathbb{C}^{(M \times N) \times (M \times N)}$ 이

고, $\phi_{m,n} = e^{j\varphi_{m,n}}$ 이며(해당 반사 요소에 대해 연속적인 위상 천이 $\varphi_{m,n} \in [0, 2\pi]$ 가 고려됨에 유의하시오), ω_{AWGN} 는 평균이 0이고 분산이 σ_l^2 인 가산성 백색 가우시안 잡음(Additive White Gaussian Noise: AWGN)이다.

[0039] 나아가, 전송된 신호 y_{BRU} 를 BS(110)로부터 직접적으로 UT(120)에 의해 수신된 신호 y_{BSU} 와 조합함으로써 합산 레이트가 향상될 수 있다. 이 신호 y_{BSU} 는 다음과 같이 표현될 수 있다.

수학식 8

[0040]
$$y_{BSU} = h_{BSU}^H \bar{X} + \omega_{AWGN}$$

[0041] 여기서 h_{BSU} 는 BS(110)로부터 UT(120)로의 채널을 나타내는 채널 행렬이다($h_{BSU}^H \in \mathbb{C}^{1 \times K}$ 임).

[0042] 따라서, UT(120)에서 수신된 신호는 다음과 같이 표현될 수 있다.

수학식 9

[0043]
$$y = y_{BRU} + y_{BSU}$$

[0044] 그러면, 이러한 코히어런트(coherent) 신호에 대한 잡음의 비율 SNR은 다음과 같이 주어진다.

수학식 10

[0045]
$$SNR = \frac{|h_{RU}^H \Phi^H h_{BR}^H \bar{X}|^2}{\sigma_l^2} + \frac{|h_{BSU}^H \bar{X}|^2}{\sigma_l^2}$$

[0046] 그러면, 섀넌(Shannon)의 용량 공식에 따라, 시간 슬롯 t 동안에 대역폭 B에 걸친 합산 레이트는 다음과 같이 표현될 수 있다.

수학식 11

[0047]
$$C(t) = (1 - \tau(t))B \log_2(1 + SNR)$$

[0048] 신호 품질을 유지하기 위해, UT(120)에서의 용량 C(t)은 주어진 최소 용량 값 C_{min} 이상이어야 한다(즉, $C(t) \geq C_{min}, t \in \{1, 2, \dots, T\}$).

[0049] 최적화 문제 규정

[0050] 이제, 시간 슬롯 t에 대해 신호 품질을 개선하기 위한 최소 요구사항 이상의 용량을 유지하면서 UAV-RIS(170)에서의 EH의 효율을 향상시키기 위해, 다음과 같이 몇 개의 제약을 조건으로 하는 최적화 문제가 정의될 수 있다.

수학식 12

$$\begin{aligned} & \max_{\tau(t), \Phi, P_t, \omega} \sum_{t=1}^T \varepsilon(t) \\ \text{s.t.} \quad & \text{C1:} \quad 0 \leq \tau(t) \leq 1, \forall t \in \{1, 2, \dots, T\} \\ & \text{C2:} \quad C(t) \geq C_{\min}, \quad t \in \{1, 2, \dots, T\} \\ & \text{C3:} \quad 0 \leq P_t = \sum \|V_k\|^2 \leq P_{\max} \end{aligned}$$

[0051]

[0052] 여기서 C1은 시간 슬롯 t 의 EH 국면의 지속기간에 대한 제약이고, C2는 시간 슬롯 동안 UT에서의 용량에 대한 제약이며, C3은 시간 슬롯 동안 UT(120)를 위한 BS(110)의 송신 전력에 대한 제약이다. 이러한 비볼록 최적화 문제는 종래의 수학적 최적화 기법으로 풀기가 쉽지 않다.

[0053] TD3 알고리즘 기반 프레임워크

[0054] 도 3은 MIMO 통신 시스템(100)에서 적용가능한 TD3 알고리즘(300)의 예를 의사코드(pseudocode)의 형태로 제시한다.

[0055] UAV-RIS(170)의 보조와 함께 BS(110)로부터 UT(120)로의 통신이 행해지는 예에서, 위 식의 비볼록 상황을 처리해 가면서 모든 제약을 만족시키는 최적의 해를 찾는 TD3 기반 프레임워크가 이용될 수 있다. TD3는 정책 기반의 심층 강화 학습(Deep Reinforcement Learning: DRL)이며, Q 학습(Q-learning)에서 발생하는 과대추정 바이어스(overestimation bias)를 줄이기 위해 행위자 망(actor network)와 함께 2개의 비평가 망(critic network)을 활용한다. 도 3에 제시된 바와 같이, TD3 알고리즘 기반 프레임워크는 에이전트(agent)로 하여금 그것이 받는 보상(reward)을 최대화하도록 다차원의 연속적 상태(state)-액션(action) 공간으로부터 학습할 수 있게 한다.

[0056] 전술된 예에서, 그러한 TD3 에이전트의 훈련(training)을 위한 각각의 시간 스텝(time step) t 에서 주어지는 상태 s_t 는 시간 스텝 $t-1$ 에서의 액션에 기반하여, 그리고 또한 시간 스텝 t 에서의 특정한 특징, 예컨대, BS(110)로부터 UT(120)에 송신되는 신호, BS(110) 및 UAV-RIS(170) 간의 채널, UAV-RIS(170) 및 UT(120) 간의 채널, BS(110) 및 UT(120) 간의 채널, 그리고 UAV-RIS(170)(특히, $R_{i,j}$) 및 UT(120) 간의 거리에 기반하여 정해질 수 있다. 예로서, 상태 s_t 는 이들 특징을 나타내는 변수로 구축된 상태 공간(state space)의 원소일 수 있다. TD3 알고리즘(300)은 그러한 상태 공간의 각 성분으로서 $\bar{X}$, h_{BR} , h_{RU} , h_{BSU} 및 $d_{i,j}$ 을 각각 입력 받도록 구성된다(문장 부분(310)).

[0057] 또한, 시간 스텝 t 에 대해 취해지는 액션 a_t 는 다음을 나타내는 변수로 구축된 액션 공간의 원소일 수 있다: EH 국면의 지속기간, BS(110)의 송신 전력, RIS(150)의 반사 요소의 위상 천이, 그리고 RIS(150)의 반사 요소의 할당. 도 3의 문장 부분(330)에 선언된 바와 같이, TD3 알고리즘(300)은 그러한 액션 공간의 이들 성분 $\tau(t)$, Φ , P_t 및 ω 를 출력하도록 구성된다.

[0058] 나아가, 액션 a_t 이 취해지는 것에 응답하여 관측되는 보상 r_t 은 다음과 같이 누적되는 $\rho(t)$ 로서 표현될 수 있다.

수학식 13

$$\bar{r} = \sum_t \rho(t)$$

[0059]

[0060] 여기서 $\rho(t)$ 는 다음과 같이 주어진다.

수학식 14

$$\rho(t) = \begin{cases} \varepsilon(t), & \text{if } C(t) \geq C_{min} \\ 0, & \text{if } C(t) < C_{min} \end{cases}$$

[0061]

[0062] 이에 따라 보상 r_t 은 EH 효율의 측면에서 양의(positive) 보상을 달성할 것이다. 즉, 취해진 액션 a_t 이 최소 용량 값 C_{min} 이상의 용량 $C(t)$ 을 초래하는 경우에 보상 r_t 은 EH 효율 $\varepsilon(t)$ 에 기반하여 양의 값으로 설정되고, 누적적 보상 $\bar{r}$ 은 증가된다.

[0063] 도 3의 문장 부분(320)에 기술된 바와 같이, TD3 알고리즘(300)은 입력 후에 다음을 초기화하도록 구성된다: 경험 재생 메모리(experience replay memory)(이는 경험 재생 버퍼로도 지칭될 수 있음) $\mathcal{M}$; 크기 $\mathcal{N}_D$ 의(즉, $\mathcal{N}_D$ 개의 전이 튜플(transition tuple)(이는 경험 튜플로 지칭될 수도 있고, 간단히 전이 또는 경험으로 지칭될 수도 있음) (s_j, a_j, r_j, s_{j+1}) 이 실장될(populated) 수 있는) 미니배치(mini-batch) 데이터 구조 $\mathcal{D}$; 행위자 망(더욱 구체적으로는 그것의 정책 함수 근사화)을 나타내는 $a = \pi(s|\vartheta_\pi)$; 타겟(target) 행위자 망을 나타내는 $a' = \pi(s|\vartheta'_\pi)$; 한 쌍의 비평가 망(더욱 구체적으로는 각자의 가치 함수 근사화)을 나타내는 $Q(s, a|\vartheta_{q1})$ 및 $Q(s, a|\vartheta_{q2})$; 그리고 한 쌍의 타겟 비평가 망을 나타내는 $Q'(s, a|\vartheta'_{q1})$ 및 $Q'(s, a|\vartheta'_{q2})$.

[0064] TD3 알고리즘(300)은 각각의 에피소드에 대해 초기 상태를 얻은 후(문장 부분(340)), 매 훈련 시간 스텝 t 마다, 탐색 잡음(exploration noise) $\mathcal{R}$ (이는, 가령, 클리핑됨(clipped))이 더해진, 액션 공간 내의 선택된 액션 a_t 을 행위자 망 및 현재의 상태에 기반하여 판정하도록 구성된다(문장 부분(350)). 계속해서 문장 부분(350)을 참조하면, TD3 알고리즘(300)은 선택된 액션 a_t 을 실행하여 달성되는 보상 r_t 및 다음 상태 s_{t+1} 를 관측하고, 전이 (s_t, a_t, r_t, s_{t+1}) 를 경험 재생 메모리에 저장하도록 구성된다.

[0065] 이후에, 도 3의 문장 부분(360)에 기술된 바와 같이, TD3 알고리즘(300)은 경험 재생 메모리로부터 $\mathcal{N}_D$ 개의 전이의 미니배치를 랜덤하게 샘플링하도록 구성된다. 계속해서 문장 부분(360)을 참조하면, 경험 재생 메모리에 전이가 저장되고 그러한 미니배치가 이용가능한 시간 스텝마다 샘플링된 미니배치 $\mathcal{D} = \{(s_j, a_j, r_j, s_{j+1})\}$ 는 이른바 타겟 정책 평활화(target policy smoothing)에 이용되는데, TD3 알고리즘(300)은 타겟 행위자 망 및 다음 상태 s_{j+1} 에 기반하여, 그리고 클리핑된 잡음 ζ 을 가산함으로써, 타겟 액션 $\bar{a}$ 을 계산하고, 두 타겟 비평가 망으로부터 각각 제공된, 이 타겟 액션의 가치 중 최소값 $\min_{i=1,2} Q'(s_{j+1}, \bar{a}|\vartheta'_{qi})$ 에 할인 인자(discount factor) γ 를 곱한 것 및 보상 r_j 의 합인 가치 함수 타겟 y 을 계산하도록 구성된다. 이어서, TD3 알고리즘(300)은 도 3의 문장 부분(360)에서와 같이 벨만(Bellman) 손실 함수 $L(\vartheta_{qi})$ 에 따라(특히, 샘플링된 전이 모두에 걸쳐 이 손실을 최소화하는 방식으로) 각각의 비평가 망의 파라미터 ϑ_{qi} 를 업데이트함으로써 해당 비평가 망을 업데이트하도록 구성된다.

[0066] 한편, TD3 알고리즘(300)에 따르면, 행위자 망에 이른바 지연된 업데이트가 적용되며, 타겟 망 역시 마찬가지로에 유의한다. 도 3의 문장 부분(370)에 기술된 바와 같이, TD3 알고리즘(300)은 비평가 망이 업데이트된 때 시간 스텝이 아니라 그러한 때 $delay$ 번째 시간 스텝마다(예를 들어, 그러한 시간 스텝 하나 건너, 즉

$delay = 2$) 두 비평자 망 중 하나(예컨대, 도 3에서와 같이 $Q(s, a|\vartheta_{q1})$, 또는 다른 예로서, 두 비평자 망 중 다른 것에 비해 더 작은 가치를 출력하는 것)에 기반하여 획득되는 정책 그래디언트 $\nabla_{\vartheta_{\pi}} J(\vartheta_{\pi})$ 를 사용하여 행위자 망의 파라미터 ϑ_{π} 를 업데이트함으로써 행위자 망을 업데이트하도록 구성되는데, 정책 그래디언트는 샘플링된 전이 모두에 걸쳐서, 행위자 망에 의해 계산된 행위에 대해서 구해진 비평자 망의 출력의 그래디언트 $\nabla_a Q(s_j, a_j|\vartheta_{q1})$ 및 행위자 망의 파라미터에 대해서 구해진 행위자 망의 출력의 그래디언트 $\nabla_{\vartheta_{\pi}}(a)$ 의 곱을 평균함으로써 획득된다. 계속해서 문장 부분(370)을 참조하면, TD3 알고리즘(300)은 비평자 망의 파라미터 ϑ_{qi} 와 함께 평활화 인자 ψ 를 사용하여 각각의 타겟 비평자 망의 파라미터 ϑ'_{qi} 를 업데이트함으로써 해당 타겟 비평자 망을 업데이트하고, 마찬가지로 행위자 망의 파라미터 ϑ_{π} 와 함께 평활화 인자 ψ 를 사용하여 타겟 행위자 망의 파라미터 ϑ'_{π} 를 업데이트함으로써 타겟 행위자 망을 업데이트하도록 구성된다.

[0067] 시뮬레이션 결과

[0068] TD3 알고리즘(300)의 성능 평가를 위해, 다음의 파라미터가 설정되었다: BS(110) 및 RIS(150)의 반사 요소 R_{ij} 간의 채널 전력 이득을 계산하는 데에 사용되는 $P_{i,j,LOS} + \alpha P_{i,j,NLOS}$ 에서 LOS 링크 확률이 $P_{i,j,LOS} = \frac{1}{1 + \mathcal{A} \exp(-B(\theta_{i,j} - 1))}$ (여기서 $\theta_{i,j}$ 는 BS(110) 및 R_{ij} 간의 양각(elevation angle)임)로 주어질 때 $\mathcal{A} = 11.95$ 및 $B = 0.136$; 채널 전력 이득의 계산에서 NLOS 링크 확률 $P_{i,j,NLOS}$ 에 곱해진 감쇠 인자 $\alpha = 20\text{dB}$; 에너지 변환 효율 인자 $\eta = 0.7$; BS(110)의 안테나 개수 $K = 8$; 최대 송신 전력 $P_{max} = 500\text{W}$; 및 최소 용량 $C_{min} = 70 \text{ Mbps}$. 또한, RIS(150)의 반사 요소의 개수는 100개로 설정되었다. 채널은 라이시안 인자(Rician factor)가 5인 페이딩 분포를 갖는다고 가정되었다. 나아가, UT(120)가 단일 사용자인 시나리오로 시뮬레이션되었다.

[0069] 도 4는 행위자 망 및 비평자 망을 위한 학습률(learning rate)이 TD3 알고리즘(300)의 성능에 어떤 영향을 미치는지를 묘사한다. 도 4에서 볼 수 있는 바와 같이, 0.001의 학습률이 0.0001 및 0.00001에 비해 더 높은 보상을 제공한다. 0.00001의 학습률이 시간-도메인 방안에서 더 많은 안정성을 주지만, 평균 보상은 0.001의 학습률로 최선의 시스템 성능이 달성된다는 것을 나타내는바, 0.00001 및 0.0001에 비해 각각 3.5% 및 3.76% 개선을 보여준다. 더욱이, 시간-도메인 방안에서, 0.001의 학습률은 0.0001 및 0.00001에 비해 각각 최대 13.17% 및 24.7% 효율을 향상시킨다. 평균 보상을 고려하면, 0.001의 학습률로 최적의 시스템 성능이 달성된다고 볼 수 있다.

[0070] 도 5는 시간-공간 도메인 방안에서 적용된 TD3 알고리즘(300)(이하 "TD3 시간-공간 도메인"으로 표기될 수 있음)으로써 시뮬레이션된 시간 스텝당 EH 효율을 시간 도메인 방안에서 적용된 TD3 알고리즘(이하 "TD3 시간 도메인"으로 표기될 수 있음), 시간-공간 도메인 방안에서 적용된 통상의 심층 결정론적 정책 그래디언트(Deep Deterministic Policy Gradient: DDPG) 알고리즘(이하 "DDPG 시간-공간 도메인"으로 표기될 수 있음) 및 시간 도메인 방안에서 적용된 DDPG 알고리즘(이하 "DDPG 시간 도메인"으로 표기될 수 있음)과 비교하여 도시한다. 이 비교에서, 학습률은 공히 0.001로 설정되었다.

[0071] 도 5에 묘사된 바와 같이, TD3 시간-공간 도메인은 TD3 시간 도메인에 비해 거의 두 배의 효율을 보여준다. TD3 시간 도메인 및 DDPG 시간 도메인은 비슷한 효율을 내나, TD3 시간-공간 도메인은 DDPG 시간-공간 도메인에 비해 40% 개선으로 귀결된다.

[0072] 도 6은 TD3 시간-공간 도메인으로써 시뮬레이션된 시간 스텝당 용량을 TD3 시간 도메인, DDPG 시간-공간 도메인 및 DDPG 시간 도메인과 비교하여 도시한다. 이 비교에서, 학습률은 공히 0.001로 설정되었다.

[0073] 도 6에 묘사된 바와 같이, TD3 시간-공간 도메인은 DDPG 시간-공간 도메인 및 DDPG 시간 도메인에 비해 각각 41.83% 및 80.93% 용량이 증가된다. 한편, TD3 시간 도메인은 DDPG 시간 도메인에 필적하는 평균 용량을 초래한다.

[0074] 예시적인 TD3 알고리즘 기반 SWIPT 프로세스

- [0075] 이제, 도 7을 참조하여, 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 예시적인 프로세스(700)가 설명된다. 예를 들어, 프로세스(700)는 통신 시스템(가령, MIMO 통신 시스템(100))에 포함된 개체(가령, BS(110)) 또는 이에 포함된 컴퓨팅 장치에 의해 수행될 수 있다. 프로세스(700)의 다른 예시적인 흐름이 또한 고려된다. 예를 들어, 프로세스(700)는 도시되지 않은 추가적인 동작을 또한 포함할 수 있고/거나, 도 7과 관련하여 예시된 동작 중 일부를 포함하나 전부를 포함하지는 않을 수 있다.
- [0076] 기저의 통신 시스템은 BS(가령, BS(110)) 및 (가령, UT(120))를 포함하며, UAV(가령, UAV(160))에 탑재된 RIS(가령, RIS(150))의 보조를 받는다.
- [0077] 동작(710)에서, TD3 알고리즘(가령, TD3 알고리즘(300))을 사용하여 에이전트가, BS로부터 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 시간 슬롯에 대해 UT에서의 적어도 사전결정된 최소 값의 용량을 달성하면서 시간 슬롯에 대해 RIS에서의 EH의 효율을 최대화하기 위한 액션을 판정하도록 훈련된다. 구체적으로, 시간 슬롯의 EH 국면 동안에, RIS의 복수의 반사 요소 각각은 EH를 위해 할당되고, 시간 슬롯의 정보 전송 국면 동안에, RIS의 복수의 반사 요소 중 일부분은 UT를 향한 주어진 신호의 반사를 위해 할당되는 한편 RIS의 복수의 반사 요소 중 나머지는 여전히 EH를 위해 할당된다. 예를 들어, 판정된 액션은 시간 슬롯에 대해 판정된 다음 사항을 포함할 수 있다: EH 국면의 지속기간; BS의 송신 전력; 각각의 반사 요소의 위상 천이; 및 (특히, 정보 전송 국면에 걸친) 각각의 반사 요소의 할당. 또한, 예로서, 전송된 상태는 시간 슬롯에 대해 주어진 다음 사항을 더 포함할 수 있다: BS로부터 RIS로의 채널의 채널 행렬; RIS로부터 UT로의 채널의 채널 행렬; BS로부터 UT로의 채널의 채널 행렬; 및 RIS 및 UT 간의 거리.
- [0078] 동작(720)에서, 훈련된 에이전트는 BS로 하여금 통신 시스템을 위한 SWIPT 동작을 수행하게 하기 위해 가동된다.
- [0079] 추가 예
- [0080] 다음은 UAV-RIS 보조식 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT에 관련된 다양한 예이다.
- [0081] 예 1에서, 통신 시스템을 위한 쌍 지연형 심층 결정론적 정책 그라디언트(Twin Delayed Deep Deterministic policy gradient: TD3) 알고리즘 기반 동시 무선 정보 및 전력 전송(Simultaneous Wireless Information and Power Transfer: SWIPT)를 위한 방법이 제공되는데, 위 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)을 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 위 RIS는 복수의 반사 요소를 포함하고, 위 방법은, TD3 알고리즘을 사용하여 에이전트(agent)를, 위 BS로부터 위 UT로의 시간 슬롯(time slot)에서의 송신을 위한 주어진 신호를 포함하는 상태(state)에 대해, 위 시간 슬롯에 대해 위 UT에서의 적어도 사전결정된 최소값의 용량(capacity)을 달성하면서 위 시간 슬롯에 대해 위 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션(action)을 판정하도록 훈련시키는 단계와, 위 BS로 하여금 위 통신 시스템을 위한 SWIPT 동작을 수행하게 하기 위해 위 훈련된 에이전트를 가동하는 단계를 포함하되, 위 시간 슬롯의 EH 국면 동안에, 위 복수의 반사 요소 각각은 위 EH를 위해 할당되고, 위 시간 슬롯의 정보 전송 국면 동안에, 위 복수의 반사 요소 중 일부는 위 UT를 향한 위 주어진 신호의 반사를 위해 할당되되 위 복수의 반사 요소 중 나머지는 위 EH를 위해 할당된다.
- [0082] 예 2는 예 1의 주제를 포함하는데, 위 통신 시스템은 상기 BT 및 상기 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원한다.
- [0083] 예 3은 예 1 또는 예 2의 주제를 포함하는데, 상기 판정된 액션은 상기 EH 국면의 지속기간(duration)과, 위 BS의 송신 전력과, 위 복수의 반사 요소 각각의 위상 천이(phase shift)와, 위 정보 전송 국면에 걸친 위 각각의 반사 요소의 위 할당을 나타내도록 구성된 액션 공간의 원소이다.
- [0084] 예 4는 예 3의 주제를 포함하는데, 위 상태는 위 BS로부터 위 UT로 송신되는 신호와, 위 BS 및 위 RIS 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 채널의 채널 행렬과, 위 BS 및 위 UT 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소이다.
- [0085] 예 5는 예 4의 주제를 포함하는데, 위 TD3 알고리즘은 행위자 망과, 2개의 비평자 망과, 타겟 행위자 망과, 2개의 타겟 비평자 망을 사용하고, 위 TD3 알고리즘을 사용하여 위 에이전트를 훈련시키는 단계는, 위 행위자 망 및 위 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 위 액션 공간 내의 선택된 액션을 반복적으로(iteratively) 판정하는 것(위 선택된 액션의 판정은 위 상태 공간 내의 새로운 상태가 위 현재의 상태로

설정되는 것으로써 반복되고, 위 새로운 상태는 위 선택된 액션이 취해짐에 따라 관측됨)과, 위 선택된 액션이 관정됨에 응답하여, 경험 재생 버퍼(experience replay buffer) 내에, 위 선택된 액션, 위 현재의 상태, 위 새로운 상태, 그리고 위 선택된 액션이 적어도 위 사전결정된 최소값을 갖는 위 용량을 초래하는 경우에 위 효율에 기반하여 양의(positive) 값으로 주어지는 보상(reward)을 포함한 전이 정보(transition information)를 저장하는 것과, 위 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치(mini-batch)가 위 경험 재생 버퍼로부터 이용가능함과 더불어 위 전이 정보가 위 경험 재생 버퍼에 저장됨에 응답하여, 위 타겟 행위자 망, 위 2개의 타겟 비평자 망 및 위 미니배치를 사용하여 위 2개의 비평자 망을 업데이트하는 것과, 위 2개의 비평자 망의 위 업데이트보다 덜 빈번하게, 위 업데이트된 2개의 비평자 망 중 하나를 사용하여 위 행위자 망, 위 타겟 행위자 망 및 위 2개의 타겟 비평자 망을 업데이트하는 것에 의해 위 에이전트를 훈련시키는 단계를 포함한다.

[0086] 예 6에서, 컴퓨터 실행가능 명령어를 포함하는 컴퓨터 판독가능 저장 매체가 제공되는데, 위 컴퓨터 실행가능 명령어는 프로세서에 의해 실행되는 경우에 위 프로세서로 하여금 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 가능하게 하는 동작을 수행하게 하되, 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)를 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 위 RIS는 복수의 반사 요소를 포함하고, 위 동작은, TD3 알고리즘을 사용하여 에이전트를, 위 BS로부터 위 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 위 시간 슬롯에 대해 위 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면서 위 시간 슬롯에 대해 위 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 관정하도록 훈련시키는 것을 포함하되, 위 시간 슬롯의 EH 국면 동안에, 위 복수의 반사 요소 각각은 위 EH를 위해 할당되고, 위 시간 슬롯의 정보 전송 국면 동안에, 위 복수의 반사 요소 중 일부는 위 UT를 향한 위 주어진 신호의 반사를 위해 할당되되 위 복수의 반사 요소 중 나머지는 위 EH를 위해 할당된다.

[0087] 예 7은 예 6의 주제를 포함하는데, 위 통신 시스템은 위 BT 및 위 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원한다.

[0088] 예 8은 예 6 또는 예 7의 주제를 포함하는데, 위 관정된 액션은 위 EH 국면의 지속기간과, 위 BS의 송신 전력과, 위 복수의 반사 요소 각각의 위상 천이와, 위 정보 전송 국면에 걸친 위 각각의 반사 요소의 위 할당을 나타내도록 구성된 액션 공간의 원소이다.

[0089] 예 9는 예 8의 주제를 포함하는데, 위 상태는 위 BS로부터 위 UT로 송신되는 신호와, 위 BS 및 위 RIS 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 채널의 채널 행렬과, 위 BS 및 위 UT 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소이다.

[0090] 예 10은 예 9의 주제를 포함하는데, 위 TD3 알고리즘은 행위자 망과, 2개의 비평자 망과, 타겟 행위자 망과, 2개의 타겟 비평자 망을 사용하고, 위 TD3 알고리즘을 사용하여 위 에이전트를 훈련시키는 것은, 위 행위자 망 및 위 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 위 액션 공간 내의 선택된 액션을 반복적으로 관정하는 것(위 선택된 액션의 관정은 위 상태 공간 내의 새로운 상태가 위 현재의 상태로 설정되는 것으로써 반복되고, 위 새로운 상태는 위 선택된 액션이 취해짐에 따라 관측됨)과, 위 선택된 액션이 관정됨에 응답하여, 경험 재생 버퍼 내에, 위 선택된 액션, 위 현재의 상태, 위 새로운 상태, 그리고 위 선택된 액션이 적어도 위 사전결정된 최소값을 갖는 위 용량을 초래하는 경우에 위 효율에 기반하여 양의 값으로 주어지는 보상을 포함한 전이 정보를 저장하는 것과, 위 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치가 위 경험 재생 버퍼로부터 이용가능함과 더불어 위 전이 정보가 위 경험 재생 버퍼에 저장됨에 응답하여, 위 타겟 행위자 망, 위 2개의 타겟 비평자 망 및 위 미니배치를 사용하여 위 2개의 비평자 망을 업데이트하는 것과, 위 2개의 비평자 망의 위 업데이트보다 덜 빈번하게, 위 업데이트된 2개의 비평자 망 중 하나를 사용하여 위 행위자 망, 위 타겟 행위자 망 및 위 2개의 타겟 비평자 망을 업데이트하는 것에 의해 위 에이전트를 훈련시키는 것을 포함한다.

[0091] 예 11에서, 통신 시스템을 위한 TD3 알고리즘 기반 SWIPT를 위한 장치가 제공되는데, 위 통신 시스템은 베이스 스테이션(Base Station: BS) 및 사용자 단말(User Terminal: UT)를 포함하고, 무인 비행체(Unmanned Aerial Vehicle: UAV)에 탑재된 재구성가능 지능형 표면(Reconfigurable Intelligent Surface: RIS)에 의해 보조되고, 위 RIS는 복수의 반사 요소를 포함하고, 위 장치는, 프로세서와, 위 BS로 하여금 위 통신 시스템을 위한 SWIPT 동작을 수행하게 하도록 TD3 에이전트를 가동하기 위해 위 프로세서에 의해 실행가능한 명령어의 세트으로써 인코딩된 메모리를 포함하되, 위 TD3 에이전트는, 위 BS로부터 위 UT로의 시간 슬롯에서의 송신을 위한 주어진 신호를 포함하는 상태에 기반하여, 위 시간 슬롯에 대해 위 UT에서의 적어도 사전결정된 최소값의 용량을 달성하면

서 위 시간 슬롯에 대해 위 RIS에서의 에너지 수확(Energy Harvesting: EH)의 효율을 최대화하기 위한 액션을 판정하도록 훈련되되, 위 시간 슬롯의 EH 국면 동안에, 위 복수의 반사 요소 각각은 위 EH를 위해 할당되고, 위 시간 슬롯의 정보 전송 국면 동안에, 위 복수의 반사 요소 중 일부는 위 UT를 향한 위 주어진 신호의 반사를 위해 할당되되 위 복수의 반사 요소 중 나머지는 위 EH를 위해 할당된다.

[0092] 예 12는 예 11의 주제를 포함하는데, 위 통신 시스템은 위 BT 및 위 UT 간의 다중 입력 다중 출력(Multiple-Input Multiple-Output: MIMO) 통신을 지원한다.

[0093] 예 13은 예 11 또는 예 12의 주제를 포함하는데, 위 판정된 액션은 위 EH 국면의 지속기간과, 위 BS의 송신 전력과, 위 복수의 반사 요소 각각의 위상 천이와, 위 정보 전송 국면에 걸친 위 각각의 반사 요소의 위 할당을 나타내도록 구성된 액션 공간의 원소이다.

[0094] 예 14는 예 13의 주제를 포함하는데, 위 상태는 위 BS로부터 위 UT로 송신되는 신호와, 위 BS 및 위 RIS 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 채널의 채널 행렬과, 위 BS 및 위 UT 간의 채널의 채널 행렬과, 위 RIS 및 위 UT 간의 거리를 나타내도록 구성된 상태 공간의 원소이다.

[0095] 예 15는 예 14의 주제를 포함하는데, 위 TD3 알고리즘은 행위자 망과, 2개의 비평자 망과, 타겟 행위자 망과, 2개의 타겟 비평자 망을 사용하고, 위 TD3 에이전트는, 위 행위자 망 및 위 상태 공간 내의 현재의 상태를 사용하여, 탐색 잡음이 가산된, 위 액션 공간 내의 선택된 액션을 반복적으로 판정하는 것(위 선택된 액션의 판정은 위 상태 공간 내의 새로운 상태가 위 현재의 상태로 설정되는 것으로써 반복되고, 위 새로운 상태는 위 선택된 액션이 취해짐에 따라 관측됨)과, 위 선택된 액션이 관측됨에 응답하여, 경험 재생 버퍼 내에, 위 선택된 액션, 위 현재의 상태, 위 새로운 상태, 그리고 위 선택된 액션이 적어도 위 사전결정된 최소값을 갖는 위 용량을 초과하는 경우에 위 효율에 기반하여 양의 값으로 주어지는 보상을 포함한 전이 정보를 저장하는 것과, 위 전이 정보의 랜덤하게 샘플링된, 고정된 크기의 미니배치가 위 경험 재생 버퍼로부터 이용가능함과 더불어 위 전이 정보가 위 경험 재생 버퍼에 저장됨에 응답하여, 위 타겟 행위자 망, 위 2개의 타겟 비평자 망 및 위 미니배치를 사용하여 위 2개의 비평자 망을 업데이트하는 것과, 위 2개의 비평자 망의 위 업데이트보다 덜 빈번하게, 위 업데이트된 2개의 비평자 망 중 하나를 사용하여 위 행위자 망, 위 타겟 행위자 망 및 위 2개의 타겟 비평자 망을 업데이트하는 것에 의해 훈련된다.

[0096] 특정한 예에서, 본 문서에서 언급된 장치, 디바이스, 시스템, 머신 등은 임의의 적합한 유형은 컴퓨팅 장치이거나, 이를 포함하거나, 이에 구현될 수 있다. 컴퓨팅 장치는 프로세서 및 프로세서에 의해 판독가능한 컴퓨터 판독가능 저장 매체를 포함할 수 있다. 프로세서는 컴퓨터 판독가능 저장 매체 내에 저장된 하나 이상의 명령어를 실행할 수 있다. 프로세서는 또한 컴퓨터 판독가능 저장 매체 내에 저장된 다른 정보를 판독할 수 있다. 추가로, 프로세서는 컴퓨터 판독가능 저장 매체 내에 새로운 정보를 저장할 수 있고 컴퓨터 판독가능 저장 매체 내에 저장된 어떤 정보를 갱신할 수 있다. 프로세서는, 예컨대, 중앙 처리 유닛(Central Processing Unit: CPU), 디지털 신호 프로세서(Digital Signal Processor: DSP), 그래픽 처리 유닛(Graphics Processing Unit: GPU), 프로세서 코어(processor core), 마이크로프로세서(microprocessor), 마이크로제어기(microcontroller), 필드 프로그램가능 게이트 어레이(Field-Programmable Gate Array: FPGA), 애플리케이션 특정 집적 회로(Application Specific Integrated Circuit: ASIC), 다른 하드웨어 및 로직 회로, 또는 이의 임의의 적합한 조합을 포함할 수 있다. 컴퓨터 판독가능 저장 매체는 다양한 정보, 예컨대, 프로세서에 의해 수행될 수 있는 프로세서 실행가능(processor executable) 명령어의 세트 및/또는 다른 정보로써 인코딩된다. 예를 들어, 컴퓨터 판독가능 저장 매체는 프로세서에 의해 실행되는 경우 컴퓨팅 장치(가령, 프로세서)로 하여금 본 문서에 개시된 몇몇 동작을 수행하게 하는 컴퓨터 프로그램 명령어 및/또는 그러한 동작에서 사용되는 정보, 데이터, 변수, 상수, 데이터 구조, 기타 등등이 내부에 저장될 수 있다. 컴퓨터 판독가능 저장 매체는, 예컨대, 판독 전용 메모리(Read-Only Memory: ROM), 랜덤 액세스 메모리(Random-Access Memory: RAM), 휘발성(volatile) 메모리, 비휘발성(non-volatile) 메모리, 착탈가능(removable) 메모리, 비착탈가능(non-removable) 메모리, 플래시(flash) 메모리, 솔리드 스테이트(solid-state) 메모리, 다른 타입의 메모리 디바이스, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광 기록 매체, 플롭티컬 디스크와 같은 자기-광 매체, 다른 타입의 저장 디바이스 및 저장 매체, 또는 이의 임의의 적합한 조합을 포함할 수 있다.

[0097] 특정한 예에서, 본 문서에 기술된 동작, 기법, 프로세스, 또는 이의 어떤 양상이나 부분은 컴퓨터 프로그램 제품 내에 구체화될 수 있다. 그러한 컴퓨터 프로그램은 컴퓨터에 의해 실행될 수 있는 어떤 유형의 (가령, 컴파일형(compiled) 또는 해석형(interpreted)) 프로그래밍 언어, 예컨대, 어셈블리(assembly), 기계어(machine language), 프로시저형(procedural) 언어, 객체지향(object-oriented) 언어 등등으로 구현될 수 있고, 하드웨

어 구현과 조합될 수 있다. 컴퓨터 프로그램 제품은 컴퓨터 판독가능 저장 매체의 형태로 배포될 수 있거나 온라인으로 배포될 수 있다. 온라인 배포를 위해, 컴퓨터 프로그램 제품의 일부 또는 전부가 서버(가령, 서버의 컴퓨터 판독가능 저장 매체) 내에 일시적으로 저장되거나 일시적으로 생성될 수 있다.

[0098] 이상의 설명은 상세하게 몇몇 예를 예시하고 기술하기 위해 제시되었다. 본 개시의 범주에서 벗어나지 않고서 위의 교시에 비추어 많은 수정 및 변형이 가능함을 당업자는 응당 이해할 것이다. 다양한 예에서, 전술된 기법이 상이한 순서로 수행되고/거나, 전술된 시스템, 아키텍처, 디바이스, 회로 및 유사한 것의 컴포넌트 중 일부가 상이한 방식으로 결합 또는 조합되거나, 다른 컴포넌트 또는 이의 균등물에 의해 대체 또는 치환되더라도 적절한 결과가 달성될 수 있다.

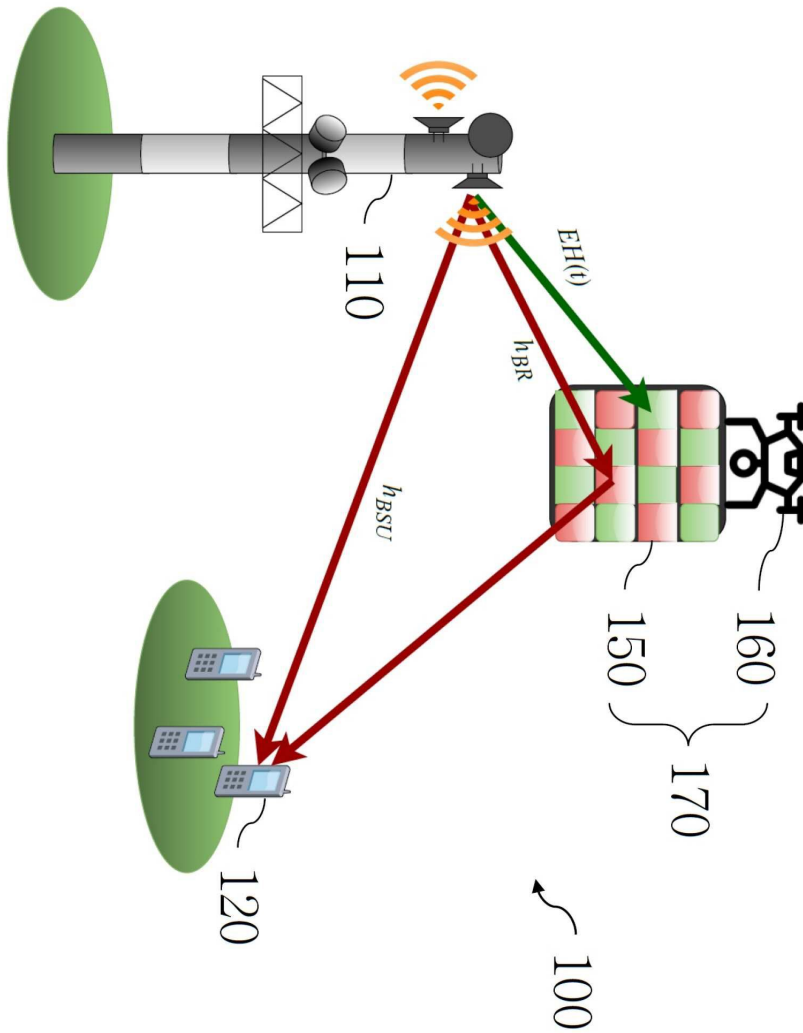
[0099] 그러므로, 본 개시의 범주는 개시된 그 형태에 한정되어서는 안 되며, 후술하는 청구항 및 이의 균등물에 의해 정해져야 한다.

부호의 설명

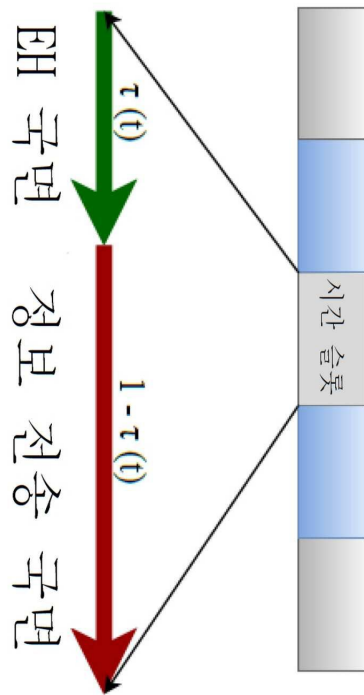
[0100] 100: MIMO 통신 시스템
110: BS
120: UT
150: RIS
160: UAV

도면

도면1



도면2

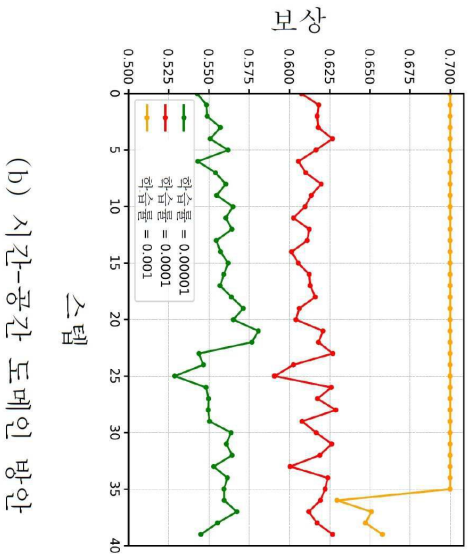
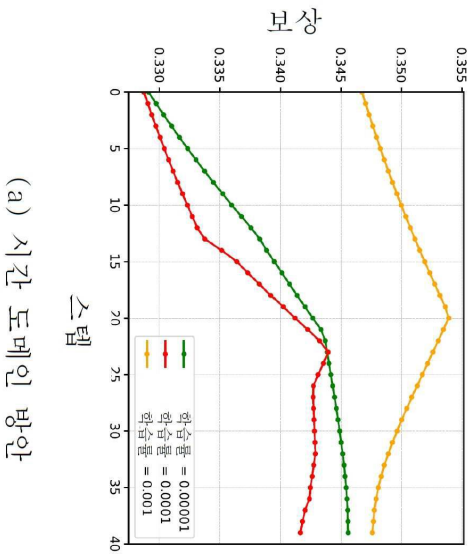


도면3

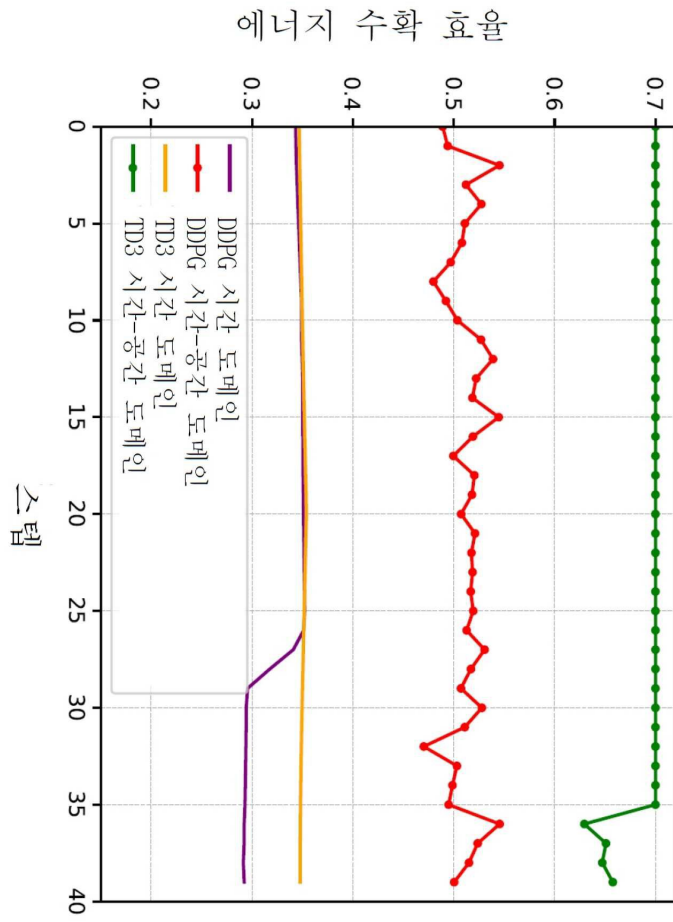
300
↙

Input: $\bar{X}, h_{BR}, h_{RU}, h_{BSU}, d_{i,j}$	} 310
Initialization: $\mathcal{M}, \mathcal{D}, a = \pi(s \vartheta_\pi), a' = \pi(s \vartheta'_\pi), Q(s, a \vartheta_{q1}), Q(s, a \vartheta_{q2}), Q'(s, a \vartheta'_{q1}), Q'(s, a \vartheta'_{q2})$	} 320
Output: $a_t = \{\tau(t), \Phi, P_t, \omega\}, \varepsilon(t), C(t)$	} 330
for episode $N_e = 0, 1, \dots, N - 1$ do Receive $\Phi, h_{BR}, h_{RU}, h_{BSU}$ to obtain s_0	} 340
for $t = 0, 1, \dots, T - 1$ do Select $a_t = \pi(s_t \vartheta_{\pi_t}) + \mathcal{R}$ Execute a_t to achieve r_t and s_{t+1} Store (s_t, a_t, r_t, s_{t+1}) in $\mathcal{M}$	} 350
Sample a mini-batch of $\mathcal{N}_D$ transitions $\mathcal{D} = \{(s_j, a_j, r_j, s_{j+1})\}$ from $\mathcal{M}$ Do target policy smoothing $\bar{a} = \pi(s_{j+1} \vartheta'_{\pi_j}) + \zeta$ Get $\mathcal{Y} = r_j + \gamma \min_{i=1,2} Q'(s_{j+1}, \bar{a} \vartheta'_{qi})$ Update the critic networks ϑ_{q1} and ϑ_{q2} using $L(\vartheta_{qi}) = \operatorname{argmin}_{\vartheta_{qi}} \frac{1}{\mathcal{N}_D} \sum_{j=1}^{\mathcal{N}_D} (\mathcal{Y} - Q(s_j, a_j \vartheta_{qi}))^2$	} 360
if $t \bmod \text{delay} = 0$ then Update the actor network using $\nabla_{\vartheta_\pi} J(\vartheta_\pi) = \frac{1}{\mathcal{N}_D} \sum_{j=1}^{\mathcal{N}_D} \nabla_a Q(s_j, a_j \vartheta_{q1}) \nabla_{\vartheta_\pi} (a)$ Update the target networks with $\vartheta'_\pi \leftarrow \psi \vartheta_\pi + (1 - \psi) \vartheta'_\pi$ $\vartheta'_{qi} \leftarrow \psi \vartheta_{qi} + (1 - \psi) \vartheta'_{qi}, \quad i = 1, 2$	} 370
end if end for end for	

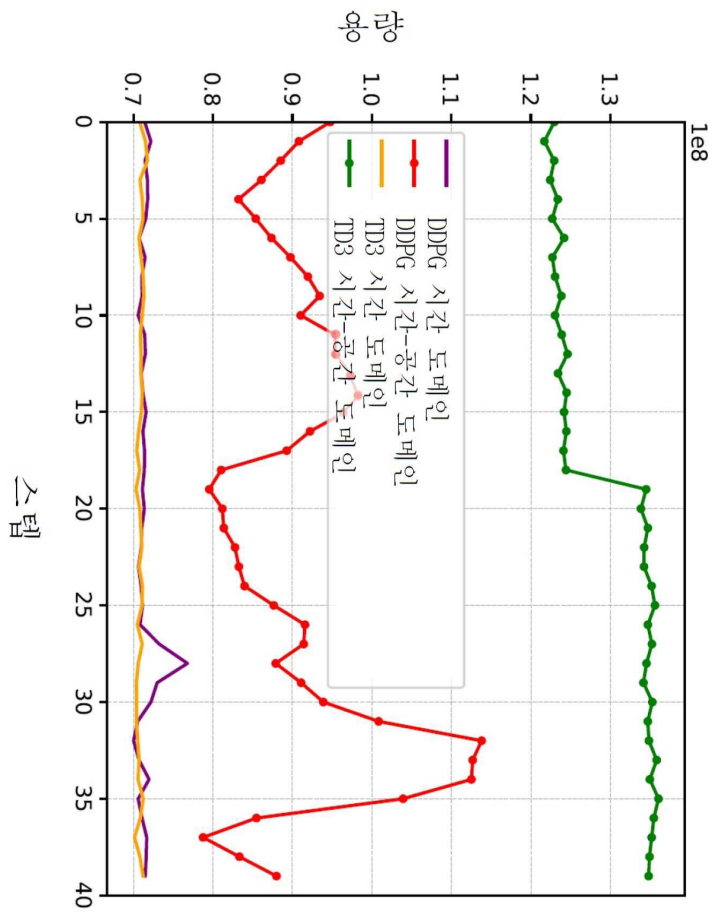
도면4



도면5



도면6



도면7

