(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0170322
(43) 공개일자 2023년12월19일

(51) 국제특허분류(Int. Cl.)
H04L 67/289 (2022.01) *H04L 47/52* (2022.01)
H04L 49/00 (2022.01) *H04L 67/1038* (2022.01)
H04W 28/08 (2023.01)
(52) CPC특허분류
H04L 67/289 (2022.05)
H04L 47/528 (2013.01)
(21) 출원번호 10-2022-0070682
(22) 출원일자 2022년06월10일
심사청구일자 2022년06월10일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
곽경섭
인천광역시 연수구 해돋이로84번길 10, 604동 802호(송도동, 송도풍림아이원6단지아파트)
(74) 대리인
양성보

전체 청구항 수 : 총 8 항

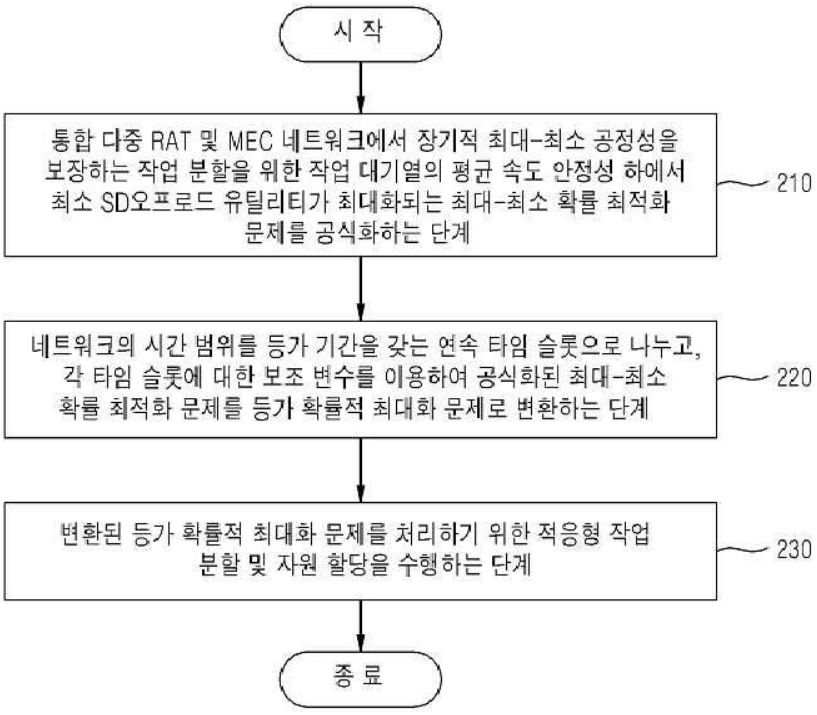
(54) 발명의 명칭 **다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 방법 및 시스템**

(57) 요 약

다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성 보장 방법 및 시스템이 제시된다. 본 발명에서 제안하는 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성 보장 방법은 최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing)

(뒷면에 계속)

대 표 도 - 도2



네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계, 최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계 및 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계를 포함한다.

(52) CPC특허분류
H04L 49/3045 (2013.01)
H04L 67/1038 (2022.05)
H04W 28/0861 (2023.05)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711166941
과제번호 2020R1A2B5B02002478
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 중견후속연구
연구과제명 [Ezbaro] 빅데이터 분석 기반 퍼지 온톨로지 및 기계학습을 활용한 IoT u-헬스케어 시스템 핵심기술 개발 및 설계
기 여 율 1/1
과제수행기관명 인하대학교
연구기간 2022.03.01 ~ 2023.02.28

명 세 서

청구범위

청구항 1

최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계;

최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계; 및

적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계

를 포함하는 모바일 엣지 컴퓨팅 통합 방법.

청구항 2

제1항에 있어서,

상기 최적화 계산부가 통합 다중 RAT 및 MEC 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계는,

상기 장기적 최대-최소 공정성을 조사하기 위해 최대 최소 확률 최적화 문제를 모든 작업 대기열의 평균 속도 안정성을 보장하는 조건(C1), 작업 분할 인과 관계 제약 조건(C2), SD 하위 작업의 오류 없는 전송을 보장하는 조건(C3), RAT(Radio Access Technology)에 대한 SD의 최대 허용 업링크 전송 전력을 제공하는 조건(C4), 각 MECS의 최대 허용 계산 주파수를 부여하는 조건(C5), 유일하게 하나의 SD에 할당할 수 있는 부반송파를 나타내는 조건(C6), 부반송파 할당 변수가 2진수임을 명시하는 조건(C7)을 만족하도록 공식화하는

모바일 엣지 컴퓨팅 통합 방법.

청구항 3

제1항에 있어서,

상기 최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계는,

상기 공식화된 최대-최소 확률 최적화 문제에 대하여 가상 대기열 기술을 이용하여 타임 슬롯에서 백로그로 가상 대기열 그룹을 정의하고, 각 SD에 대한 가상 대기열 역학을 통해 등가 확률적 최대화 문제로 변환하는

모바일 엣지 컴퓨팅 통합 방법.

청구항 4

제1항에 있어서,

상기 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계는,

상기 변환된 등가 확률적 최대화 문제를 적응형 작업 분할 및 자원 할당 알고리즘을 통해 현재 NSI(Network State Information)와 QSI(Queue State Information)에만 기초하여 각 타임 슬롯에서 적응형 작업 분할 및 자원 할당을 결정하고, 임의의 파라미터의 통계적 분포를 필요로 하지 않으며, 최적화를 위한 복수의 변수들이 목적과 제약 모두에서 서로 독립되어 있어 병렬로 처리 가능하도록 복수의 분리된 하위 문제로 분해하는

모바일 엣지 컴퓨팅 통합 방법.

청구항 5

통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하고, 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 최적화 계산부; 및

상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 적응형 작업 분할 및 자원 할당부

를 포함하는 모바일 엣지 컴퓨팅 통합 방법.

청구항 6

제5항에 있어서,

상기 최적화 계산부는,

상기 장기적 최대-최소 공정성을 조사하기 위해 최대 최소 확률 최적화 문제를 모든 작업 대기열의 평균 속도 안정성을 보장하는 조건(C1), 작업 분할 인과 관계 제약 조건(C2), SD 하위 작업의 오류 없는 전송을 보장하는 조건(C3), RAT(Radio Access Technology)에 대한 SD의 최대 허용 업링크 전송 전력을 제공하는 조건(C4), 각 MECS의 최대 허용 계산 주파수를 부여하는 조건(C5), 유일하게 하나의 SD에 할당할 수 있는 부반송파를 나타내는 조건(C6), 부반송파 할당 변수가 2진수임을 명시하는 조건(C7)을 만족하도록 공식화하는

모바일 엣지 컴퓨팅 통합 방법.

청구항 7

제5항에 있어서,

상기 최적화 계산부는,

상기 공식화된 최대-최소 확률 최적화 문제에 대하여 가상 대기열 기술을 이용하여 타임 슬롯에서 백로그로 가상 대기열 그룹을 정의하고, 각 SD에 대한 가상 대기열 역학을 통해 등가 확률적 최대화 문제로 변환하는

모바일 엣지 컴퓨팅 통합 방법.

청구항 8

제5항에 있어서,

상기 적응형 작업 분할 및 자원 할당부는,

상기 변환된 등가 확률적 최대화 문제를 적응형 작업 분할 및 자원 할당 알고리즘을 통해 현재 NSI(Network State Information)와 QSI(Queue State Information)에만 기초하여 각 타임 슬롯에서 적응형 작업 분할 및 자원 할당을 결정하고, 임의의 파라미터의 통계적 분포를 필요로 하지 않으며, 최적화를 위한 복수의 변수들이 목적과 제약 모두에서 서로 독립되어 있어 병렬로 처리 가능하도록 복수의 분리된 하위 문제로 분해하는

모바일 엣지 컴퓨팅 통합 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 최근 몇 년 동안 안면 인식, 증강 현실, 대화형 온라인 게임 등과 같은 계산 집약적이고 대기 시간에 민감한 모바일 애플리케이션의 급속한 성장은 스마트폰에서 다양한 웨어러블 장치에 이르는 광범위한 스마트 장치(Smart Devices; SD)에 대한 컴퓨팅 수요를 증가시켰다[1], [2]. SD는 일반적으로 자원 제약을 받기 때문에 로컬에서 이러한 애플리케이션을 직접 계산하는 데 시간이 너무 많이 걸릴 수 있으며, 따라서 엄격한 지연 시간 요구 사항을 충족하지 못한다. 더욱이, 이러한 계산 집약적인 애플리케이션을 위한 작업 컴퓨팅은 너무 많은 전기를 소비할 것이며, 이는 또한 로컬 컴퓨팅에 적합하지 않게 만든다[3]. 이 과제를 해결하기 위해 모바일 클라우드 컴퓨팅(Mobile Cloud Computing; MCC)이 제안되어 지난 10년간 업계에 적용되었다[4]. 그럼에도 불구하고, MCC는 일반적으로 원격 중앙 집중식 클라우드에 따라 애플리케이션/작업 오프로드 서비스를 제공한다. 작업 트래픽이 기하급수적으로 증가함에 따라, 그러한 전통적인 컴퓨팅 패러다임은 새로운 컴퓨팅 요구를 거의 충족할 수 없으며 심각한 핵심 네트워크 혼잡, 불만족스러운 종단 간 서비스 지연 시간 등과 같은 일부 고유한 결함을 노출한다[5].

[0003] 이러한 문제를 추가로 해결하기 위해, 유럽전기통신표준연구소(European Telecommunications Standard Institute; ETSI)에 의해 MCC에 추가로 모바일 엣지 컴퓨팅(Mobile Edge Computing MEC)이 제안되었다[6]. MCC와 비교하여, MEC는 분산된 소형 에지 서버에 의해 SD에 가까운 무선 접속 네트워크 엣지에서 작업 컴퓨팅 서비스를 제공하여 서비스 대기 시간을 단축한다. 또한, MEC는 SD와의 근접성으로부터 이익을 얻는 실시간 무선 컨텍스트에 대한 인식을 가지고 있다. 이러한 장점은 적응형 작업 오프로드와 공동 통신 및 계산 자원 할당을 MEC 네트워크에서 실행하게 한다[7]. 여러 종래기술에서 적응형 작업 오프로드 문제를 연구했지만, 대부분 SD 작업이 단일 특정 RAT(Radio Access Technology)에 의해 오프로드되고 SD 간의 공정성을 간과했다고 가정했다. 네트워크 엣지의 제한된 자원을 고려할 때, 공정한 작업 오프로드 및 자원 할당 메커니즘이 가장 중요하다.

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 스마트 장치가 여러 RAT 링크를 통해 병렬로 계산 작업을 오프로드할 수 있는 통합 멀티 RAT 및 MEC 네트워크 프레임워크를 제공하는데 있다. 제안하는 네트워크의 시간 변동 특성뿐만 아니라 자원 제한 특성을 고려하여 스마트 장치 간의 공정한 자원 공유/배분을 용이하게 하기 위한 장기 공정성 보장 메커니즘을 제안한다.

과제의 해결 수단

[0005] 일 측면에 있어서, 본 발명에서 제안하는 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성 보장 방법은 최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계, 최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계 및 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계를 포함한다.

[0006] 상기 최적화 계산부가 통합 다중 RAT 및 MEC 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계는 상기 장기적 최대-최소 공정성을 조사하기 위해 최대 최소 확률 최적화 문제를 모든 작업 대기열의 평균 속도 안정성을 보장하는 조건(C1), 작업 분할 인과 관계 제약 조건(C2), SD 하위 작업의 오류 없는 전송을 보장하는 조건(C3), RAT(Radio Access Technology)에 대한 SD의 최대 허용 업링크 전송 전력을 제공하는 조건(C4), 각 MECS의 최대 허용 계산 주파수를 부여하는 조건(C5), 유일하게 하나의 SD에 할당할 수 있는 부반송파를 나타내는 조건(C6), 부반송파 할당 변수가 2진수임을 명시하는 조건(C7)을 만족하도록 공식화한다.

[0007] 상기 최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계는 상기 공식화된 최대-최소 확률 최적화 문제에 대하여 가상 대기열 기술을 이용하여 타임 슬롯에서 백로그로 가상 대기열 그룹을 정의하고, 각 SD에 대한 가상 대기열 역학을 통해 등가 확률적 최대화 문제로 변

환한다.

[0008] 상기 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계는 상기 변환된 등가 확률적 최대화 문제를 적응형 작업 분할 및 자원 할당 알고리즘을 통해 현재 NSI(Network State Information)와 QSI(Queue State Information)에만 기초하여 각 타임 슬롯에서 적응형 작업 분할 및 자원 할당을 결정하고, 임의의 파라미터의 통계적 분포를 필요로 하지 않으며, 최적화를 위한 복수의 변수들이 목적과 제약 모두에서 서로 독립되어 있어 병렬로 처리 가능하도록 복수의 분리된 하위 문제로 분해한다.

[0009] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성 보장시스템은 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하고, 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 최적화 계산부 및 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 적응형 작업 분할 및 자원 할당부를 포함한다.

발명의 효과

[0010] 본 발명의 실시예들에 따르면 최소 장기적 시간 평균 오프로드 유틸리티를 최대화하는 것을 목표로 최대-최소 확률 최적화 문제를 공식화하고, 스마트 장치 작업 분할 및 업링크 전송 전력 할당, RAT 부반송파 할당 및 MEC 계산 주파수 할당을 공동으로 최적화하는 Lyapunov 최적화 기술을 기반으로 적응형 작업 분할 및 자원 할당을 수행할 수 있다. 제안하는 적응형 작업 분할 및 자원 할당 알고리즘은 분포 정보를 요구하지 않고 시간 변동 네트워크 역학을 수용할 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 발명의 일 실시예에 따른 통합 다중 RAT 및 MEC 네트워크를 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 시스템의 구성을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 적응형 작업 분할 및 자원 할당 알고리즘을 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 분산 공동 작업 분할 및 통신 자원 할당 알고리즘을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0012] 최근 모바일 엣지 컴퓨팅(Mobile Edge Computing; MEC)은 날로 증가하는 컴퓨팅 수요를 충족시키기 위한 새로운 패러다임으로 인식되고 있다. 또한 네트워크 처리량 및 서비스 신뢰성을 향상시키기 위해 다중 무선 접속 기술(Multiple Radio Access Technology; multi-RAT)이 제안되었다. 그러나 두 기술은 독립적으로 발전해 왔다.

[0013] 본 발명에서는 스마트 장치(Smart Devices; SD)가 여러 RAT 링크를 통해 병렬로 계산 작업을 오프로드할 수 있는 통합 멀티 RAT 및 MEC 네트워크 프레임워크를 제안한다. 제안하는 네트워크의 시간 변동 특성뿐만 아니라 자원 제한 특성을 고려하여 SD 간의 공정한 자원 공유/배분을 용이하게 하기 위한 장기 공정성 보장 메커니즘을 제안하는 데 중점을 둔다. 특히, 최소 장기적 시간 평균 오프로드 유틸리티를 최대화하는 것을 목표로 최대-최소 확률 최적화 문제를 공식화한다. SD 작업 분할 및 업링크 전송 전력 할당, RAT 부반송파 할당 및 MEC 계산 주파수 할당을 공동으로 최적화하는 Lyapunov 최적화 기술을 기반으로 적응형 작업 분할 및 자원 할당 알고리즘을 제안한다. 적응 알고리즘은 분포 정보를 요구하지 않고 시간 변동 네트워크 역학을 수용할 수 있다. 또한, 제안하는 알고리즘은 다항식 계산 복잡성을 가지며 엄격한 분석에 의해 점근적으로 최적의 것으로 나타낼 수 있다. 시뮬레이션 결과는 이론적 분석을 검증하고 제안된 알고리즘의 Jain 공정성 지수가 타임 슬롯 수가 증가함에 따라 1로 수렴되어 벤치마크 알고리즘보다 성능이 우수하다는 것을 보여준다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0014] 본 발명에서는 작업 오프로드 효율성을 높이기 위해 다중 RAT(multi-RAT) 다양성을 활용하여 SD 작업을 병렬로 오프로드할 것을 제안한다. 이를 기반으로, 고려된 네트워크에 대한 공정한 작업 오프로드 및 자원 할당을 조사하는 데 중점을 둔다.

[0015] 다중 RAT 수렴은 5G와 네트워크를 넘어 네트워크 처리량 및 서비스 신뢰성을 향상시키기 위한 핵심 기술 중 하나로 간주되어 왔다[8]. 다중 RAT 네트워크에서 SD는 다중 RAT 기능(예를 들어, WiFi, LTE, 5NR 등)을 갖추고 이러한 이기종 RAT를 서로 다른 프로토콜 계층에 통합할 수 있다. 이렇게 함으로써, SD는 여러 RAT 링크를 동시에 유지하고 작업/트래픽 흐름을 여러 하위 흐름으로 분할하여 다른 RAT 링크를 통해 이러한 하위 흐름을 전송할 수 있으므로, 이러한 병렬 전송에 의해 네트워크 처리량과 서비스 신뢰성이 향상될 수 있다[9]. 실제로, 다중 RAT의 상호 운용성을 지원하기 위한 새로운 네트워크 프레임워크는 문헌 및 표준화 기구에서 논의되었다. 예를 들어, 이중 연결은 SD를 다른 유형의 기지국(예를 들어, 매크로 및 피코 기지국)에 연결하여 이동성-강력 통신을 지원하기 위해 LTE-A Release 12에서 제안되었다[10]. 또한, LTE와 WiFi의 공존을 촉진하기 위해, LTE 지원 접속(LTE Assisted Access; LAA)[11]와 LTE-WiFi 링크 애그리게이션(LTE-WiFi link aggregation; LWA)[12]이 3세대 파트너십 프로젝트(3GPP)에 의해 지정되었다. 그럼에도 불구하고, 다중 RAT는 MEC 컨텍스트와 독립적으로 발전해 왔다. 두 첨단 기술을 함께 병합하기 위해 새로운 통합 네트워크가 필요하다.

[0016] 최근에는 MEC 네트워크에 대한 작업 오프로드 및 자원 할당은 크게 정적 오프로드와 적응형/동적 오프로드의 두 가지 범주로 나눌 수 있다.

[0017] 정적 오프로드의 경우, 네트워크 무작위성이 무시되므로 결정론적 최적화 모델이 일반적으로 채택된다. 또한 에너지 소비 또는 지연 시간 최소화가 최적화 목표로 널리 활용된다. 예를 들어, 종래기술[15]에서 볼록 최적화 방법을 사용하여 단일 사용자 작업 오프로드에 대한 에너지 소비와 작업 계산 지연 시간을 각각 최소화하려고 시도했다. 종래기술[16]에서, 주어진 채널 조건에서 에너지 효율을 개선하기 위해 지연 시간 제약 작업 오프로딩을 위한 사용자 협력 접근법을 제안했다. 유사한 최적화 목표를 가진 문제도 다중 MECS 및 다중 사용자 시나리오로 확장되었다. 예를 들어, 종래기술[17]에서 매크로 셀 지원 이기종 MEC 네트워크에서 부분 작업 오프로드에 대한 저복잡도 연속 볼록 근사 기반 알고리즘이 제안되었다. 종래기술[18]에서, 다중 MECS 및 다중 사용자 작업 오프로드에 대한 에너지 소비 및 작업 계산 지연 시간 트레이드오프 문제를 공식화하고 중앙 집중식 휴리스틱 알고리즘으로 해결했다.

[0018] 적응형/동적 오프로드의 경우, 작업 오프로드 및 자원 할당 전략은 시시각각 변하는 네트워크 역학을 수용해야 한다. 이러한 컨텍스트에서 몇 가지 작업이 수행되었다. 종래기술[19]에서 다중 사용자 작업 오프로드 프로세스는 사전 네트워크 정보가 없는 비협조적인 정확한 잠재적 게임으로 구성되었으며, 이는 최상의 응답 역학을 통해 내쉬 균형으로 수렴되는 것으로 입증되었다. 고차원성의 저주를 풀기 위해 종래기술[20]과 [21]에서 각각 심층 강화 학습 기반 알고리즘을 제안하여 최적의 정지 작업 오프로드 정책을 학습하였다. 종래기술[22]에서, Lyapunov 최적화 기술은 다중 사용자 MEC 시스템을 위한 적응형 작업 오프로드 알고리즘을 설계하는 데 사용되었다. 또한, 종래기술[23]에서 초신뢰성 및 저지연 다중 MECS 네트워크를 위해 2시간 규모의 Lyapunov 최적화 기반 작업 오프로드 알고리즘이 제안되었다.

[0019] 앞서 언급한 종래기술은 정적 및 적응/동적 작업 오프로드에 대해 광범위하게 조사했지만, 대부분은 사용자 작업이 단일 RAT에 의해서만 오프로드될 수 있다고 가정하고 사용자 간의 공정성을 간과했다. 공정한 커뮤니케이션 자원 할당을 보장하기 위해, 종래기술[24]에서 솔루션의 존재와 고유성이 분석된 공정한 내시 협상 게임으로 문제를 공식화했다. 종래기술[25]에서 사용자 간의 최대 가중 에너지 소비와 작업 계산 지연 시간을 최소화하여 최소-최대 공정성을 연구했다. 그러나 위의 작업은 네트워크 무작위성을 무시했다. 네트워크 무작위성을 협력하기 위해, 종래기술[26]은 단일 RAT MEC 네트워크에 대한 부분 네트워크 정보를 사용하여 비례 공정 및 적응 알고리즘을 도출했다. 종래기술[27]에서, 다중 모드 사용자 장치에 대한 애플리케이션 처리량 공정성과 에너지 효율을 연구하였고, 여기서 사용자 장치는 채널 조건에 따라 최적의 네트워크 인터페이스를 선택할 수 있다. 작업 오프로드 효율성을 높이기 위해 이전 회의 작업[13]은 통합 다중 RAT 및 MEC 네트워크 프레임워크를 제안하고 첫 번째 시도에 대한 장기적 최대-최소 공정성을 조사했다. 그러나 종래기술은 제안된 네트워크 프레임워크에 대한 자세한 설명을 제공하지 않았으며 제한된 공간 때문에 필요한 일부 기술적 세부 사항을 생략했다.

[0020] 본 발명에서는 통합된 다중 RAT 및 MEC 네트워크의 프로토콜 설계를 제안한다. 연속성 완화와 Lagrange 이중 분해를 사용하여 다항식 계산 복잡성을 가진 적응형 작업 분할 및 자원 할당 알고리즘을 제안한다.

[0022] 도 1은 본 발명의 일 실시예에 따른 통합 다중 RAT 및 MEC 네트워크를 설명하기 위한 도면이다.

[0023] 도 1을 참고하면, 이 통합 네트워크에서 각 SD의 사용자 평면 프로토콜 스택은 하위 계층 RAT 프로토콜의 여러 그룹에 의해 공유되는 공통 패킷 데이터 수렴 프로토콜(Packet Data Convergence Protocol; PDCP)로 구성된다. RAT 프로토콜의 각 그룹은 무선 링크 제어(Radio Link Control; RLC), 매체 접속 제어(Medium Access Control; MAC), 물리적(physical; PHY) 프로토콜로 구성된다. 이러한 프로토콜 설계를 통해 SD는 PDCP 계층에서 작업 분할 운영을 실행하고 각 RAT에 해당하는 하위 프로토콜 계층에 하위 작업 흐름을 매핑할 수 있다. 이와 같이 SD 작업은 여러 RAT 기지국으로 오프로드되고 연결된 MEC 서버(MEC servers; MECS)에 의해 동시에 계산될 수 있으므로 작업 오프로드 및 컴퓨팅 효율성을 향상시킬 수 있다.

[0024] 그러나 무선 RAT 링크는 일반적으로 대규모 및 소규모 채널 페이딩으로 인해 공간적으로 다양하다. SD 작업도 무작위로 도착한다. 따라서 통합 다중 RAT 및 MEC 네트워크에 대한 작업 분할은 확률적 네트워크 역학에 적응해야 한다. 또한 자원이 제한된 동적 네트워크 환경에서 SD 공정성을 보장하기 위해 공정성 보장 메커니즘이 주의 깊게 연구되어야 한다. 이를 위해, 모든 SD 중에서 최소/최악의 SD 오프로드 유틸리티를 최대화할 것을 제안한다. 여기서 SD 오프로드 유틸리티는 장기적 최대-최소 공정성이라고 하는 장기적 시간 평균 오프로드 작업 양으로 정의된다. 또한, 적응형 작업 분할 및 자원 할당 알고리즘 설계를 통해 SD 간 오프로딩 유틸리티 측면에서 차별되지 않은 서비스 품질을 달성할 수 있는 통합 다중 RAT 및 MEC 네트워크에 대한 장기적 최대-최소 공정성 보장 메커니즘을 설계하는 것을 목표로 한다.

[0025] 통합 다중 RAT 및 MEC 네트워크에 대한 장기적인 최대-최소 공정성을 보장하기 위해, 본 발명에서는 통신 및 계산 리소스 할당은 물론 작업 분할을 공동으로 최적화하여 MECS에서 작업 대기열의 평균 속도 안정성 하에서 최소 SD 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화한다.

[0026] 또한, 문제를 쉽게 해결하기 위해 먼저 각 타임 슬롯에 대한 보조 변수를 도입하여 공식화된 최대-최소 문제를 등가 최대 버전으로 변환한다. 그런 다음 Lyapunov 최적화 기술[14]을 사용하여 변환된 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당 알고리즘을 제안한다. 여기서 사전 네트워크 정보를 요구하지 않고 각 타임 슬롯에서 여러 분리된 하위 문제가 해결된다.

[0027] 제안하는 알고리즘은 다항식 계산 복잡성을 갖는 것으로 분석된다. 또한, 최소 SD 오프로드 유틸리티 최적성과 작업 대기열 백로그 간에 [O(1/V), O(V)] 트레이드오프를 달성한다. 시뮬레이션 결과는 제안된 알고리즘의 Jain 공정성 지수가 점근적으로 1로 수렴한다는 것을 보여주며, 이는 일부 다른 벤치마크 알고리즘을 크게 능가한다. 따라서 장기적 최대-최소 공정성을 보장할 수 있다.

[0028] 도 1과 같이, 본 발명에서는 서비스 영역 내에 서로 다른 RAT를 갖춘 SD의 집합 N과 기지국의 집합 M이 분포하는 통합 다중 RAT 및 MEC 네트워크를 고려한다. 각 RAT 기지국에는 SD에서 오프로드된 컴퓨팅 작업을 위한 MECS가 연결되어 있다. 앞서 설명된 바와 같이 SD에는 여러 RAT 기지국과의 연결을 동시에 유지하고 연결된 MECS에 작업을 동시에 오프로드할 수 있도록 다중 RAT 기능이 부여된다. 또한, 본 발명에서는 시간 범위를 등가 기간이 τ인 연속 타임 슬롯으로 나누는 동기화된 시간 제한 시스템 프로토콜 모델을 고려한다. 타임 슬롯 지수는 $t \in T = \{0, 1, 2쪋\}$로 표시된다. 더 잘 이해하기 위해, 본 발명에서는 표 I에 주요 주석을 요약한다.

[0029] <표 1>

Notation	Description
$\mathcal{N}$	Set of SDs
$\mathcal{M}$	Set of RATs/MECSs
$\mathcal{K}_m$	Set of subcarriers occupied by RAT m
$\mathcal{T}$	Set of time-slots
τ	Duration of a time-slot
$A_n(t)$	Tasks generated by SD n at time-slot t
$A_n^{\max}$	Maximum size of tasks generated by SD n per time-slot
β_n	CPU cycles needed for computing a-bit task of SD n
$a_{n,m}(t)$	Tasks of SD n offloaded by RAT m at time-slot t
$b_{n,m}(t)$	Tasks of SD n computed by MECS m at time-slot t
N_0	Single-sided spectral density of AWGN
B_k	Spectrum bandwidth of subcarrier k
$x_{n,m}^k(t)$	Indicator of whether subcarrier k of RAT m is allocated to SD n at time-slot t
$p_{n,m}^k(t)$	Transmit power of SD n allocated to RAT m on subcarrier k at time-slot t
$g_{n,m}^k(t)$	Channel power gain from SD n to RAT m on subcarrier k at time-slot t
$r_{n,m}^k(t)$	Achievable transmit rate from SD n to RAT m on subcarrier k at time-slot t
$f_{n,m}(t)$	CPU computation frequency allocated by MECS m to SD n at time-slot t
$f_m^{\max}$	Maximum allowed computation frequency of MECS m
$p_{n,m}^{\max}$	Maximum allowed transmit power of SD n allocated to RAT m
$Q_{n,m}(t)$	Task queue backlog of SD n at MECS m at time-slot t
$Z_n(t)$	Virtual queue backlog of SD n at time-slot t
$U_n^o(t)$	Offloading utility of SD n at time-slot t

[0030]

[0032] 도 2는 본 발명의 일 실시예에 따른 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 방법을 설명하기 위한 흐름도이다.

[0033] 제안하는 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 방법은 최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계(210), 최적화 계산부가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계(220) 및 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계(230)를 포함한다.

[0034] 단계(210)에서, 최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화한다.

[0035] 이때, 상기 장기적 최대-최소 공정성을 조사하기 위해 최대 최소 확률 최적화 문제를 모든 작업 대기열의 평균 속도 안정성을 보장하는 조건(C1), 작업 분할 인과 관계 제약 조건(C2), SD 하위 작업의 오류 없는 전송을 보장하는 조건(C3), RAT(Radio Access Technology)에 대한 SD의 최대 허용 업링크 전송 전력을 제공하는 조건(C4), 각 MECS의 최대 허용 계산 주파수를 부여하는 조건(C5), 유일하게 하나의 SD에 할당할 수 있는 부반송파를 나타내는 조건(C6), 부반송파 할당 변수가 2진수임을 명시하는 조건(C7)을 만족하도록 공식화한다.

[0036] 단계(210)에서, 최적화 계산부가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화한다.

[0037] 상기 공식화된 최대-최소 확률 최적화 문제에 대하여 가상 대기열 기술을 이용하여 타임 슬롯에서 백로그로 가상 대기열 그룹을 정의하고, 각 SD에 대한 가상 대기열 역학을 통해 등가 확률적 최대화 문제로 변환한다.

[0038] 단계(230)에서, 적응형 작업 분할 및 자원 할당부가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행한다.

[0039] 상기 변환된 등가 확률적 최대화 문제를 적응형 작업 분할 및 자원 할당 알고리즘을 통해 현재 NSI(Network State Information)와 QSI(Queue State Information)에만 기초하여 각 타임 슬롯에서 적응형 작업 분할 및 자원 할당을 결정한다. 이때, 임의의 파라미터의 통계적 분포를 필요로 하지 않으며, 최적화를 위한 복수의 변수들이 목적과 제약 모두에서 서로 독립되어 있어 병렬로 처리 가능하도록 복수의 분리된 하위 문제로 분해한다.

[0041] 도 3은 본 발명의 일 실시예에 따른 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 시스템의 구성을 나타내는 도면이다.

[0042] 제안하는 다중무선접속기술과 모바일 엣지컴퓨팅 통합 네트워크를 위한 장기적 최대-최소 공정성을 보장하는 시스템(200)은 최적화 계산부(310) 및 적응형 작업 분할 및 자원 할당부(320)를 포함한다.

[0043] 본 발명의 실시예에 따른 최적화 계산부(310)는 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하고, 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환한다.

[0044] 본 발명의 실시예에 따른 적응형 작업 분할 및 자원 할당부(320)는 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행한다.

[0045] 최적화 계산부(310) 및 적응형 작업 분할 및 자원 할당부(320)는 도 2의 단계들(210~230)을 수행하기 위해 구성될 수 있다.

[0046] 단계(210)에서, 최적화 계산부(310)가 통합 다중 RAT(Radio Access Technology) 및 MEC(Mobile Edge Computing) 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD(Smart Devices) 오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화한다.

[0047] 본 발명의 실시예에 따르면, 각 시간대가 시작될 때 각 SDn의 응용 계층에서 많은 양의 $A_n(t) \in [0, A_n^{\max}]$ 작업이 생성된다고 가정한다. 여기서 $A_n^{\max}$ 은 $A_n(t), \forall t \in \mathcal{T}$ 의 최대값이다. An(t)가 독립적으로 동일하게 분포된 프로세스이며 이 확률적 과정을 실험적으로 유효하게 채택했음에 유의한다. 예를 들어, 타임 슬롯이 고려된 프로토콜 모델은 균일하기 때문에 타임 슬롯당 포아송 작업 도착 수는 독립적으로 동일하게 분포된다. 본 발명에서는 벡터 $\mathbf{A}(t) = [A_n(t)]_{n \in \mathcal{N}}$ 에 의해 모든 An(t)을 수집하고 A^{av} = E{A(t)}로 기댓값을 나타낸다. 본 발명의 실시예에 따르면, 작업이 세분화되고 데이터 분할되어 임의로 여러 비율/세그먼트로 분할되고 독립적으로 병렬로 계산될 수 있다고 가정한다. 통합 다중 RAT 및 MEC 네트워크의 이점을 활용하기 위해 각 SD는 PDCP 계층에서 도착한 작업을 분할한 다음 다른 RAT 링크를 통해 결과 하위 작업을 병렬 컴퓨팅을 위한 분산 MECS로 전달한다. $a_{n,m}(t)$ 가 작업 분할 연산을 기반으로 시간 단위로 RATm을 통해 전달되는 것으로 결정된 SDN의 하위 작업을 나타내도록 하고 행렬 $\mathbf{a}(t) = [a_{n,m}(t)]_{m \in \mathcal{M}, n \in \mathcal{N}}$ 이 모든 SD에 대한 하위 작업을 수집한다. 그렇다면, 다음과 같은 인과관계 제약이 충족되어야 한다.

[0048] $$\sum_{m \in \mathcal{M}} a_{n,m}(t) \leq A_n(t), \forall n \in \mathcal{N}. \qquad (1)$$

[0049] 인과 관계 제약 조건(1)은 SDN이 타임 슬롯에서 모든 RAT를 통해 오프로드한 총 작업 양이 현재 작업 도착을 초과하지 않아야 함을 의미한다.

[0050] 효율적인 전송 견고성 때문에, 직교 주파수 분할 다중 접속(Orthogonal Frequency Division Multiple Access; OFDMA) 기술은 LTE, WiMax, WiFi[28]를 포함한 다양한 무선 표준에서 성공적으로 채택되었고, 5Gera[29]에 사용하기로 결정되었다. 따라서, 본 발명에서는 OFDMA 기술이 일반성의 손실 없이 고려된 네트워크의 모든 RAT에 채택되었다고 가정한다. 사용 가능한 시스템 스펙트럼은 부반송파들의 집합 $\mathcal{K} = [\mathcal{K}_m]_{m \in \mathcal{M}}$ 으로 나뉘며, 여기서

하위 집합 $\mathcal{K}_m, \forall m \in \mathcal{M}$ 은 RATm이 점유하는 부반송파를 포함한다. 다른 RAT를 통한 데이터 신호는 일반적으로 다른 주파수 스펙트럼에서 변조되므로, $\mathcal{K}_m, \forall m \in \mathcal{M}$ [9] 사이에 중복이 없다고 가정한다. $\mathbf{x}(t) = [x^k_{n,m}(t)]_{m\in\mathcal{M},n\in\mathcal{N},k\in\mathcal{K}_m}$ 는 부반송파 할당 행렬을 나타낸다. 여기서 $x^k_{n,m}(t)$ 은 이진 지시자이고, RATm의 부반송파 k가 SDn에 시간 단위로 할당될 경우 $x^k_{n,m}(t) = 1$ 이다. 그렇지 않으면, $x^k_{n,m}(t) = 0$ 이다. 어떤 부반송파라도 OFDMA 시스템에서 오직 하나의 SD에만 할당될 수 있기 때문에, 본 발명에서는 $\sum_{n\in\mathcal{N}} x^k_{n,m}(t) \le 1, \forall m \in \mathcal{M}, k \in \mathcal{K}_m$ 을 가진다. 업링크 RAT 링크에 대한 플랫 블록 페이딩 채널 모델, 즉 SDn에서 모든 부반송파 k의 RAT 기지국 m으로 가는 채널 전력 게인(Channel Power Gain; CPG) $g^k_{n,m}(t)$ 는 한 타임 슬롯 동안 변경되지 않고 독립적이고 동일한 분포 방식으로 다음 타임 슬롯에서 변경될 수 있는 가능성을 고려한다. 마르코프 채널 모델처럼 더 일반적인 채널 모델도 마찬가지이다[30].

[0051] 이때, 모든 CPG를 수집하기 위해 행렬 $\mathbf{g}(t) = [g^k_{n,m}(t)]_{m\in\mathcal{M},n\in\mathcal{N},k\in\mathcal{K}_m}$ 을 사용한다. $p^k_{n,m}(t)$ 를 부반송파 k에서 SDn이 RATm에 할당한 업링크 송신 전력으로 하고 $\mathbf{p}(t) = [p^k_{n,m}(t)]_{m\in\mathcal{M},n\in\mathcal{N},k\in\mathcal{K}_m}$ 을 각각 업링크 송신 전력 행렬로 한다. 그러면, 타임 슬롯에서 부반송파 k에서 SDn으로부터 RATm으로 도달할 수 있는 업링크 전송 속도는 다음과 같이 기록될 수 있다.

[0052]
$$r^k_{n,m}(t) = x^k_{n,m}(t) B_k \log_2\left(1 + \frac{p^k_{n,m}(t) g^k_{n,m}(t)}{N_0 B_k}\right), \quad (2)$$

[0053] 여기서 Bk는 부반송파 $k \in \mathcal{K}_m$ 의 스펙트럼 대역폭이고 N0은 가산 백색 가우스 노이즈(Additive White Gaussian Noise; AWGN)의 단측 스펙트럼 밀도이다. 모든 부반송파 $k \in \mathcal{K}_m$ 의 전송 속도를 집계함으로써 SDn에서 RATm까지의 총 전송 속도를 얻는다.

[0054]
$$r_{n,m}(t) = \sum_{k\in\mathcal{K}_m} r^k_{n,m}(t). \quad (3)$$

[0055] 하위 작업의 오류 없는 전송을 보장하기 위해 작업 분할은 모든 타임 슬롯에서 RAT 링크 용량에 적응해야 한다. 다시 말해, 보증할 필요가 있다.

[0056]
$$\tau r_{n,m}(t) \ge a_{n,m}(t), \forall m \in \mathcal{M}, n \in \mathcal{N}. \quad (4)$$

[0057] 제약 조건(4)는 SDn의 모든 RATm을 통해 오프로드하기 위해 분할된 하위 작업의 양이 하위 작업 간의 순간 채널 용량보다 크지 않아야 함을 나타낸다. 이러한 제약은 오프로드 효율성을 개선하기 위해 작업 분할 및 통신 자원 할당의 최적화가 공동으로 수행되어야 함을 의미한다.

[0058] 본 발명의 실시예에 따르면, 각 MECS에 단일 코어 CPU 프로세서가 장착되어 있으며 계산 주파수는 동적 전압 스케일링 기술을 사용하여 적응적으로 조정할 수 있다고 가정한다[31]. 오프로드된 작업을 수신한 후 MECS는 작업 컴퓨팅을 위해 CPU 계산 주파수를 SD에 할당한다. $f_{n,m}(t)$ 는 시간 제한에서 MECSm이 SDn에 할당한 CPU 계산 주파수를 나타내며 $\mathbf{f}(t) = [f_{n,m}(t)]_{m\in\mathcal{M},n\in\mathcal{N}}$ 을 모든 MECS에 대한 계산 주파수 할당 행렬이라고 하자. 그러면, SDn에 대해 MECSm으로 계산할 수 있는 작업의 양은 다음과 같이 주어질 수 있다.

[0059]
$$b_{n,m}(t) = \frac{f_{n,m}(t)\tau}{\beta_n}, \quad (5)$$

[0060] 여기서 β_n 은 SDn의 작업에 대한 계산 작업량(비트당 CPU 사이클)으로, 오프라인 측정으로 얻을 수 있다 [32].

[0061] 각 MECS에서 서로 다른 SD의 처리되지 않은 작업은 작업 대기열에서 버퍼된다. 본 발명에서는 타임 슬롯 t에 MECS m에서 SDn에 대한 작업 대기열 백로그를 $Q_{n,m}(t)$ 로 나타내며, 행렬 $\mathbf{Q}(t) = [Q_{n,m}(t)]_{m\in\mathcal{M},n\in\mathcal{N}}$ 이 모든 SD 및 MECS에 대한 작업 대기열 백로그를 수집하도록 한다. MECSm에서 SDn에 대한 작업 도착으로서 $a_{n,m}(t)$, 작

업 출발로서 $b_{n,m}(t)$ 와 함께 Qn,m(t)의 역학은 다음과 같이 주어질 수 있다.

[0062] $$Q_{n,m}(t+1) = [Q_{n,m}(t) - b_{n,m}(t)]^+ + a_{n,m}(t), \quad (6)$$

[0063] 여기서 $[x]^+ = \max\{x, 0\}$ 이다.

[0064] 대기열 이론[33]에 따르면, 장기 시간 평균 출발은 대기열의 안정성을 보장하기 위해 장기 시간 평균 도착 이상이어야 한다. 따라서, 본 발명에서는 다음 식을 얻는다.

[0065] $$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\{b_{n,m}(t)\} \geq \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\{a_{n,m}(t)\}, \quad (7)$$

[0066] 여기서 기대 연산자 E{ · }는 NSI(Network State Information) s(t) = {A(t), g(t)} 및 작업 분할 및 자원 할당 정책 α(t) = {a(t), x(t), p(t), f (t)}와 관련이 있다. 다음으로, 본 발명에서는 정의 1에 의한 평균 속도 안정성이라고 불리는 더욱 엄격한 대기열 안정성의 개념을 제공한다.

[0067] 정의 1: 대기열은 다음과 같은 경우 안정적인 평균 속도로 정의된다.

[0068] $$\lim_{T\to\infty}\frac{\mathbb{E}\{|Q_{n,m}(T)|\}}{T} = 0. \quad (8)$$

[0069] 종래기술[14]로부터 식(7)의 대기열 안정성 조건은 실제로 평균 속도 안정성 식(8)의 필수 조건이다. 다시 말해, 본 발명의 실시예에 따른 네트워크에서 모든 작업 대기열의 평균 속도 안정성을 보장할 수 있다면, 식(7)은 자연스럽게 충족될 수 있다.

[0070] 본 발명의 실시예에 따르면, 통합 다중 RAT 및 MEC 네트워크에 대한 장기적 최대-최소 공정성을 조사하기 위해 최대 최소 확률 최적화 문제를 공식화하는 데 중점을 둔다. 최적화 목표는 작업 대기열의 평균 속도 안정성에 따라 모든 SD 중에서 최소/최악의 장기 시간 평균 오프로드 효율성을 최대화하는 것이다. 각 SDn의 장기적 시간 평균 오프로드 효율성은 $\bar{U}_n^o = \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\{U_n^o(t)\}$ 로 정의되며, 여기서 $U_n^o(t) = \sum_{m\in\mathcal{M}} a_{n,m}(t)$ 은 SDn의 순간 오프로드 효율성이다. 그 결과, 최대 확률 최적화 문제는 다음과 같이 공식화될 수 있다.

$$\mathcal{P}1: \max_{\mathbf{a}(t),\mathbf{x}(t),\mathbf{p}(t),\mathbf{f}(t)} \min_{n\in\mathcal{N}} \bar{U}_n^o$$

s.t. C1: All task queues are mean rate stable,

C2: $\sum_{m\in\mathcal{M}} a_{n,m}(t) \leq A_n(t), \forall n \in \mathcal{N},$

C3: $\tau r_{n,m}(t) \geq a_{n,m}(t), \forall m \in \mathcal{M}, n \in \mathcal{N},$

C4: $\sum_{k\in\mathcal{K}_m} x_{n,m}^k(t) p_{n,m}^k(t) \leq p_{n,m}^{\max}, \forall m \in \mathcal{M}, n \in \mathcal{N},$

C5: $\sum_{n\in\mathcal{N}} f_{n,m}(t) \leq f_m^{\max}, \forall m \in \mathcal{M},$

C6: $\sum_{n\in\mathcal{N}} x_{n,m}^k(t) \leq 1, \forall m \in \mathcal{M}, k \in \mathcal{K}_m,$

C7: $x_{n,m}^k(t) \in \{0,1\}, \forall m \in \mathcal{M}, n \in \mathcal{N}, k \in \mathcal{K}_m, \quad (9)$

[0071]

[0072] 여기서, 제약 조건 C1은 모든 작업 대기열의 평균 속도 안정성을 보장하고, C2는 작업 분할 인과 관계 제약 조건이며, C3는 SD 하위 작업의 오류 없는 전송을 보장하는 것을 목표로 하며, C4는 각 RAT에 대한 SD의 최대 허용 업링크 전송 전력을 제공하고, C5는 각 MECS의 최대 허용 계산 주파수를 부여하며, C6은 유일하게 한 SD에 할당할 수 있는 부반송파를 나타내고, C7은 부반송파 할당 변수가 2진수임을 명시한다.

[0073] 문제 P1 해결의 어려움은 주로 다음과 같은 측면에 있다:

[0074] (a) 문제 P1은 엄청난 양의 최적화 변수를 포함하는 장기적 시간 평균 확률적 최적화 문제이고 통계 NSI의 분포가 사전에 알려져 있지 않기 때문에 오프라인 알고리즘에 의한 문제 해결의 비효율성을 초래한다.

[0075] (b) 식(6)의 대기열 역학에 따라 작업 분할 및 자원 할당 결정은 시간대에 걸쳐 시간 결합된다. 따라서 현재의 의사 결정 조치는 미래에 영향을 미칠 것이다.

[0076] (c) 목표는 장기적 최대-최소 공정성을 달성하도록 설계되어 문제 해결의 어려움을 더욱 증가시킨다. 이를 해결하기 위해 먼저 각 시간대에 대한 보조 변수를 도입하여 최대-최소 문제를 최대 문제로 동등하게 변환한 다음 Lyapunov 최적화 기술을 활용하여 적응 알고리즘을 제안한다. 이는 현재 NSI 및 작업 대기열 백로그에만 응답하여 결정을 내린다.

[0077] 단계(220)에서, 최적화 계산부(310)가 상기 네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 상기 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환한다.

[0078] 문제 P1을 해결하기 위해, 본 발명에서는 각 시간대에 보조 변수 Φ(t)를 도입하여 확률적 최대화 문제로 동등하게 변환될 수 있음을 보여준다. 자세한 내용은 다음 요약에서 설명한다.

[0079] 보조정리 1: 문제 P1은 다음과 같이 문제 P2로 등가 변환될 수 있다.

$$\mathcal{P}2: \max_{\mathbf{a}(t),\mathbf{x}(t),\mathbf{p}(t),\mathbf{f}(t),\phi(t)} \bar{\phi}$$
$$\text{s.t. } C1-C7,$$
$$C8: U_n^o - \bar{\phi} \geq 0, \forall n \in \mathcal{N},$$
$$C9: 0 \leq \phi(t) \leq \phi^{max}, \qquad (10)$$

[0080]

[0081] 여기서 $\bar{\phi} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\{\phi(t)\}$ 이고 $\phi^{max} \geq \max_{n\in\mathcal{N}} A_n^{max}$ 이다. 일반성의 손실 없이, $\phi^{max} = \max_{n\in\mathcal{N}} A_n^{max}$ 를 취한다.

[0082] 문제 P2는 두 개의 추가 제약 조건 C8과 C9를 도입하더라도 P1보다 처리하기가 더 쉽다. 제약 조건 C9는 간단한 순간 제약 조건이다. 어려움은 제약 조건 C8에서 온다. 본 발명에서는 이 문제를 해결하기 위해 가상 대기열 기술을 사용한다. 특히, 타임 슬롯 t에서 백로그 Z(t) = $\mathbf{Z}(t) = [Z_n(t)]_{n\in\mathcal{N}}$ 으로 가상 대기열 그룹을 정의한다. 각 $n \in \mathcal{N}$ 에 대한 가상 대기열 역학은 다음과 같이 주어질 수 있다.

$$Z_n(t+1) = [Z_n(t) - U_n^o(t)]^+ + \phi(t). \qquad (11)$$

[0083]

[0084] 다시, 평균 속도 안정성[14]의 필요 조건을 사용함으로써, 다음과 같이 문제 P2를 문제 P3으로 더 변환할 수 있다.

$$\mathcal{P}3: \max_{\mathbf{a}(t),\mathbf{x}(t),\mathbf{p}(t),\mathbf{f}(t),\phi(t)} \bar{\phi}$$
$$\text{s.t. } C1-C7, C9,$$
$$C10: \text{All virtual queues are mean rate stable.} \qquad (12)$$

[0085]

[0086] 단계(230)에서, 적응형 작업 분할 및 자원 할당부(320)가 상기 변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행한다.

[0087] 도 4는 본 발명의 일 실시예에 따른 적응형 작업 분할 및 자원 할당 알고리즘을 나타내는 도면이다.

[0088] 본 발명의 실시예에 따르면, Lyapunov 최적화 기술을 기반으로 문제 P3를 해결하기 위한 적응형 작업 분할 및 자원 할당(Adaptive Task Splitting and Resource Allocation; ATSRA) 알고리즘을 제안한다.

[0089] 문제 P3의 장기적 시간 평균 목표를 직접 최적화하는 것을 우회하기 위해, Lyapunov 최적화 기술은 원-타임 슬롯(one-time slot) Lyapunov 드리프트의 상한을 최소화하고 목표 페널티 항을 추가할 것을 제안한다. Y(t) = [Q(t), Z(t)] 작업 대기열 및 가상 대기열의 대기열 상태 정보(Queue State Information; QSI)를 나타낸다. 2차 Lyapunov 함수는 다음과 같이 정의할 수 있다.

$$L(\mathbf{Y}(t)) = \frac{1}{2}\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}} Q_{n,m}^2(t) + \frac{1}{2}\sum_{n\in\mathcal{N}} Z_n^2(t). \qquad (13)$$

[0090]

[0091] 그러면, 원-타임 슬롯 조건부 Lyapunov 드리프트는 다음과 같이 주어진다.

[0092] $$\triangle(\mathbf{Y}(t)) \triangleq \mathbb{E}\left\{L(\mathbf{Y}(t+1)) - L(\mathbf{Y}(t))|\mathbf{Y}(t)\right\}. \quad (14)$$

[0093] 문제 P3의 예상 목적 함수를 식(14)에 추가하면 다음과 같이 조건부 Lyapunov 드리프트 플러스 페널티 항을 얻는다.

[0094] $$\triangle_V(\mathbf{Y}(t)) = \triangle(\mathbf{Y}(t)) - V\mathbb{E}\left\{\phi(t)|\mathbf{Y}(t)\right\}, \quad (15)$$

[0095] 여기서 V는 식(15)의 두 항 사이에서 균형을 이루는 양의 제어 파라미터이다.

[0096] 그런 다음 다음 요약에 의해 $\triangle_V(\mathbf{Y}(t))$의 상한을 부여한다.

[0097] 보조정리 2: 실행 가능한 결정 정책 $\alpha(t) = \{\mathbf{a}(t), \mathbf{x}(t), \mathbf{p}(t), \mathbf{f}(t)\}$보조 변수 $\phi(t)$, QSI $\mathbf{Y}(t)$ 및 제어 파라미터 V > 0에 대하여, 다음을 얻는다.

$$\begin{aligned}\triangle_V(\mathbf{Y}(t)) &\le B - V\mathbb{E}\left\{\phi(t)|\mathbf{Y}(t)\right\} \\ &+ \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}} Q_{n,m}(t)\mathbb{E}\left\{a_{n,m}(t) - b_{n,m}(t)|\mathbf{Y}(t)\right\} \\ &+ \sum_{n\in\mathcal{N}} Z_n(t)\mathbb{E}\left\{\phi(t) - U_n^o(t)|\mathbf{Y}(t)\right\}, \quad (16)\end{aligned}$$

[0098]

[0099] 여기서,

$$\begin{aligned}B \ge& \frac{1}{2}\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}}\mathbb{E}\left\{a_{n,m}^2(t) + b_{n,m}^2(t)|\mathbf{Y}(t)\right\} \\ &+ \frac{1}{2}\sum_{n\in\mathcal{N}}\mathbb{E}\left\{\phi^2(t) + (U_n^o(t))^2|\mathbf{Y}(t)\right\}. \quad (17)\end{aligned}$$

[0100] 이다.

[0101] 본 발명에서 제안하는 접근 방식의 주요 아이디어는 각 타임 슬롯에서 문제 P3의 모든 순간 제약 조건 하에서 식(16)의 오른손 추정치(estimate of the right hand)를 최소화하는 것으로, 이는 장기적으로 모든 작업 대기열을 점근적으로 안정화하면서 목표를 최대화하는 데 도움이 된다. 제안된 절차 ATSRA 알고리즘은 알고리즘 1에 요약되어 있다. ATSRA 알고리즘은 현재 NSI와 QSI에만 기반하여 각 타임 슬롯에서 결정을 내리고 임의의 파라미터의 통계적 분포를 요구하지 않는다는 점에 유의한다. 따라서 결정론적 문제 P4를 해결하기에 충분하다. 문제 P4의 최적화 변수 f(t), {a(t), x(t), p(t)} 및 Φ(t)는 목적과 제약 모두에서 서로 독립되어 있음을 알 수 있다.

[0102] 결과적으로, 문제 P4의 최적화 변수를 자연스럽게 병렬로 해결할 수 있는 몇 가지 분리된 하위 문제로 분해될 수 있다.

[0103] 세 가지 분리된 하위 문제, 즉 보조 변수 결정 하위 문제 SP4 - 1, 계산 빈도 할당 하위 문제 SP4 - 2, 공동 작업 분할 및 통신 자원 할당 하위 문제 SP4 - 3을 해결하여 문제 P4에 대한 최적의 솔루션을 제공한다.

[0104] 먼저, 보조 변수 결정 하위 문제는 P4에서 분리되어 다음과 같이 쓸 수 있다.

$$\begin{aligned}&\mathcal{SP}4-1: \min_{\phi(t)} \left(\sum_{n\in\mathcal{N}} Z_n(t) - V\right)\phi(t) \\ &\text{s.t.} \quad \text{C9:0} \le \phi(t) \le \phi^{\max}. \quad (19)\end{aligned}$$

[0105]

[0106] 분명히, 문제 SP4 - 1은 단일 변수 선형 프로그래밍 문제이며, 여기서 최적의 Φ(t)는 다음과 같이 구할 수 있다.

$$\phi(t) = \begin{cases}\phi^{\max}, \text{if } \sum_{n\in\mathcal{N}} Z_n(t) - V \le 0, \\ 0, \text{otherwise.}\end{cases} \quad (20)$$

[0107]

[0108] 또한, 최적의 계산 주파수 할당은 제약 조건 C5의 대상인 문제 P4의 목표에서 네 번째 항을 최소화함으로써 얻을 수 있다. 게다가, 이 문제는 분리할 수 있다. 각 m ∈ M에 대해 다음과 같이 쓸 수 있다.

$$\mathcal{SP}4-2: \max_{\mathbf{f}_m(t)} \sum_{n\in\mathcal{N}} \rho_{n,m}(t) f_{n,m}(t)$$

$$\text{s.t. C5:} \sum_{n\in\mathcal{N}} f_{n,m}(t) \le f_m^{\max},$$

$$\text{C11:} 0 \le f_{n,m}(t) \le \varpi_{n,m}(t), \forall n \in \mathcal{N}, \qquad (21)$$

[0109]

[0110] 여기서 $\mathbf{f}_m(t) = [f_{n,m}(t)]_{n\in\mathcal{N}}$, $\rho_{n,m}(t) = Q_{n,m}(t)\tau/\beta_n, \forall n \in \mathcal{N}$ 이고 $\varpi_{n,m}(t) = Q_{n,m}(t)\beta_n/\tau, \forall n \in \mathcal{N}$ 이다. 작업 대기열 백로그에 따라 허용된 계산 빈도 할당을 제한하기 위해 제약 조건 C11이 추가된다. 문제 SP4 - 2는 일반적으로 선형 프로그래밍 문제로 분류할 수 있으며, 이는 일반적인 심플렉스법으로 해결할 수 있다. 그러나 문제 SP4 - 2는 최적의 폐쇄형 해결책을 달성할 수 있도록 좋은 구조적 특성을 가지고 있다. 구체적으로, $\rho_{n,m}(t)$ 을 SDn의 단위 이익으로 해석하고 $f_m^{\max}$ 을 knapsack 용량으로 해석함으로써, 문제 SP4 - 2는 knapsack 문제의 선형 완화로 취급할 수 있다. 종래기술[34]에 따르면, 본 발명은 knapsack 용량 $f_m^{\max}$을 충족하기 위해 단위 이익이 더 높은 SD를 선택함으로써 최적의 계산 빈도 할당을 얻을 수 있다.

[0111] 표현을 위해, 본 발명에서는 SD를 단위 이익의 내림차순으로 정렬한다. 즉, i > j에 대해 $\rho_{i,m}(t) \ge \rho_{j,m}(t)$ 이다. 먼저 전체 계산 빈도가 할당되지 않은 SD로 구분 항목을 정의한다. breaking 항 w의 지수는 다음과 같이 식별할 수 있다.

$$w = \arg\min_i \sum_{j=1}^{i} \varpi_{j,m}(t) > f_m^{\max}. \qquad (22)$$

[0112]

[0113] MECS m에 대한 최적의 계산 주파수 할당은 다음과 같이 결정할 수 있다.

$$f_{n,m}(t) = \begin{cases} \varpi_{n,m}(t), & \text{if } n < w, \\ f_m^{\max} - \sum_{i=1}^{n-1} \varpi_{i,m}(t), & \text{if } n = w, \\ 0, & \text{otherwise.} \end{cases} \qquad (23)$$

[0114]

[0115] 작업 트래픽 정체를 완화하기 위해 MECS는 계산 빈도 할당을 결정할 때 작업 대기열 백로그가 크고 작업 계산 작업량이 작은 SD에 더 많은 우선순위를 부여한다.

[0116] 다음으로 공동 작업 분할 및 통신 리소스 할당에 있어서, 제약 조건 C2-C4, C6 및 C7에 따라 문제 P4의 목적에서 두 번째와 세 번째 항을 결합하면, 공동 작업 분할 및 통신 자원 할당 하위 문제는 다음과 같이 쓸 수 있다.

$$\mathcal{SP}4-3: \min_{\mathbf{a}(t),\mathbf{x}(t),\mathbf{p}(t)} \sum_{m\in\mathcal{M}} \sum_{n\in\mathcal{N}} \left(Q_{n,m}(t) - Z_n(t)\right) a_{n,m}(t)$$

$$\text{s.t. C2:} \sum_{m\in\mathcal{M}} a_{n,m}(t) \le A_n(t), \forall n \in \mathcal{N}, \qquad (24)$$

$$\text{C3:} \tau r_{n,m}(\mathbf{x},\mathbf{p})(t) \ge a_{n,m}(t), \forall m \in \mathcal{M}, n \in \mathcal{N}, \qquad (25)$$

$$\text{C4:} \sum_{k\in\mathcal{K}_m} x_{n,m}^k(t) p_{n,m}^k(t) \le p_{n,m}^{\max}, \forall m \in \mathcal{M}, n \in \mathcal{N}, \qquad (26)$$

$$\text{C6:} \sum_{n\in\mathcal{N}} x_{n,m}^k(t) \le 1, \forall m \in \mathcal{M}, k \in \mathcal{K}_m, \qquad (27)$$

$$\text{C7:} x_{n,m}^k(t) \in \{0,1\}, \forall m \in \mathcal{M}, n \in \mathcal{N}, k \in \mathcal{K}_m. \qquad (28)$$

[0117]

[0118] 문제 SP4 - 3은 해결하기 어려운 비볼록 혼합 정수 비선형 프로그래밍(Mixed Integer Nonlinear Programming;

MINLP) 문제이다. 그러나 $Q_{n,m}(t) - Z_n(t) \geq 0$ 인 경우, $a_{n,m}(t), x_{n,m}^k(t)$, $p_{n,m}^k(t)$에 대한 최적의 해결책이 0에서 달성될 수 있음을 발견했다. 이는 중복 변수를 제거하여 문제를 수정하도록 영감을 준다. 본 발명에서는 각 RAT m ∈ M에 대해 SD 집합 $\mathcal{N}_m(t) = \{n | Q_{n,m}(t) - Z_n(t) < 0, n \in \mathcal{N}\}$ 을 정의하고, 각 SD의 $n \in \mathcal{N}$에 대해 RAT 집합 Mn(t) = $\mathcal{M}_n(t) = \{m | Q_{n,m}(t) - Z_n(t) < 0, m \in \mathcal{M}\}$ 을 정의한다. 집합 $\mathcal{N}_m(t), \forall m \in \mathcal{M}$ 과 $\mathcal{M}_n(t), \forall n \in \mathcal{N}$ 은 작업 대기열 및 가상 대기열의 역학에 따라 시간이 변한다는 점에 유의한다. 그러면, 각 타임 슬롯 t격에서, 임의의 $n \in \mathcal{N},\ m \in \mathcal{M} - \mathcal{M}_n(t),\ k \in \mathcal{K}_m$ 에 대해, $a_{n,m}(t) = 0, x_{n,m}^k(t) = 0$ 및 $p_{n,m}^k(t) = 0$ 을 간단히 설정할 수 있다. 따라서 문제 SP4 - 3의 $n \in \mathcal{N}, m \in \mathcal{M}_n(t), k \in \mathcal{K}_m$ 에 대해 $a_{n,m}(t),\ x_{n,m}^k(t)$, $p_{n,m}^k(t)$ 만 최적화하면 된다. 본 발명의 실시예에 따르면, 각 타임 슬롯 t마다 각각 행렬 $\breve{\mathbf{a}}(t) = [a_{n,m}(t)]_{m \in \mathcal{M}, n \in \mathcal{N}_m(t)}$, $\breve{\mathbf{x}}(t) = [x_{n,m}^k(t)]_{m \in \mathcal{M}, n \in \mathcal{N}_m(t), k \in \mathcal{K}_m}$, $\breve{\mathbf{p}}(t) = [p_{n,m}^k(t)]_{m \in \mathcal{M}, n \in \mathcal{N}_m(t), k \in \mathcal{K}_m}$ 로 줄어든 작업 분할, 부반송 전력 할당 변수를 수집한다. 그러나, 그 결과로 생긴 문제는 여전히 볼록하지 않고 해결하기 어렵다. 이를 위해 연속성 완화와 Lagrange 이중 분해 방법[35], [36]을 활용하여 이 문제를 처리할 것이다. 표현의 단순성을 위해, 본 발명에서는 다음에서 타임 슬롯 지수 t를 생략한다.

[0119] 본 발명의 실시예에 따른 연속성 완화 및 볼록성에 관하여, $x_{n,m}^k$ 을 연속 구간 [0,1]로 완화하고 새로운 변수 $s_{n,m}^k = x_{n,m}^k p_{n,m}^k$ 을 도입하여 $\breve{\mathbf{s}} = [s_{n,m}^k]_{m \in \mathcal{M}, n \in \mathcal{N}_m, k \in \mathcal{K}_m}$ 으로 함으로써 하위 문제 SP4-3의 완화 문제를 다음과 같이 나타낼 수 있다.

$$\mathcal{RSP}4-3: \min_{\breve{\mathbf{a}},\breve{\mathbf{x}},\breve{\mathbf{s}}} \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} (Q_{n,m} - Z_n)\, a_{n,m}$$

$$\text{s.t.}\quad \text{C12}: \sum_{m \in \mathcal{M}_n} a_{n,m} \leq A_n, \forall n \in \mathcal{N},$$

$$\text{C13}: \tau r_{n,m}(\breve{\mathbf{x}}, \breve{\mathbf{s}}) \geq a_{n,m}, \forall m \in \mathcal{M}, n \in \mathcal{N}_m,$$

$$\text{C14}: \sum_{k \in \mathcal{K}_m} s_{n,m}^k \leq p_{n,m}^{\max}, \forall m \in \mathcal{M}, n \in \mathcal{N}_m,$$

$$\text{C15}: \sum_{n \in \mathcal{N}_m} x_{n,m}^k \leq 1, \forall m \in \mathcal{M}, k \in \mathcal{K}_m,$$

$$\text{C16}: x_{n,m}^k \in [0,1], \forall m \in \mathcal{M}, n \in \mathcal{N}_m, k \in \mathcal{K}_m, \quad (29)$$

[0120]

[0121] 여기서,

$$r_{n,m}(\breve{\mathbf{x}}, \breve{\mathbf{s}}) = \sum_{k \in \mathcal{K}_m} x_{n,m}^k(t) B_k \log_2 \left(1 + \frac{s_{n,m}^k(t) g_{n,m}^k(t)}{x_{n,m}^k(t) N_0 B_k}\right). \quad (30)$$

[0122] 이다.

[0123] 완화된 속도 함수 $r_{n,m}(\breve{\mathbf{x}}, \breve{\mathbf{s}})$ 은 고전적인 원근 오목 함수의 형태를 가지고 있음을 알 수 있다. 따라서 제약 조건 C13은 볼록 제약 조건이다. 또한 목적 및 나머지 제약 조건 C12와 C14-C16은 affine이므로 문제 RSP4 - 3은 실질적으로 볼록한 문제이며 Slater의 조건을 만족시킬 수 있다. 따라서 문제 RSP4 - 3과 이중 문제[36] 사이에는 0의 이중 격차가 존재한다.

[0124] 본 발명의 실시예에 따른 Lagrange 이중 분해 방법에 관하여, 제약 조건 C13과 C14를 완화하기 위해 각각 $\breve{\boldsymbol{\omega}} = [\omega_{n,m}]_{m \in \mathcal{M}, n \in \mathcal{N}_m}$ 및 $\breve{\boldsymbol{\gamma}} = [\gamma_{n,m}]_{m \in \mathcal{M}, n \in \mathcal{N}_m}$ 의 두 그룹을 도입함으로써, 다음과 같이 문제 RSP4-3의 Lagrangian을 구할 수 있다.

$$\mathcal{L}(\breve{\mathbf{x}},\breve{\mathbf{s}},\breve{\mathbf{a}},\breve{\boldsymbol{\omega}},\breve{\boldsymbol{\gamma}}) = \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m}(Q_{n,m}-Z_n)\,a_{n,m}$$
$$+\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m}\omega_{n,m}\left(a_{n,m}-\sum_{k\in\mathcal{K}_m}\tau x_{n,m}^k B_k\right.$$
$$\left.\log_2\left(1+\frac{s_{n,m}^k g_{n,m}^k}{x_{n,m}^k N_0 B_k}\right)\right)$$
$$+\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m}\gamma_{n,m}\left(\sum_{k\in\mathcal{K}_m}s_{n,m}^k-p_{n,m}^{\max}\right). \quad (31)$$

[0125]

[0126] 그렇다면, 이중 함수는 다음과 같이 주어지고

$$g(\breve{\boldsymbol{\omega}},\breve{\boldsymbol{\gamma}}) = \min_{\breve{\mathbf{a}},\breve{\mathbf{x}},\breve{\mathbf{s}}}\mathcal{L}(\breve{\mathbf{x}},\breve{\mathbf{s}},\breve{\mathbf{a}},\breve{\boldsymbol{\omega}},\breve{\boldsymbol{\gamma}})$$
$$\text{s.t.}\quad \text{C12}: \sum_{m\in\mathcal{M}_n}a_{n,m}\le A_n,\forall n\in\mathcal{N},$$
$$\text{C15}: \sum_{n\in\mathcal{N}_m}x_{n,m}^k\le 1,\forall m\in\mathcal{M},k\in\mathcal{K}_m,$$
$$\text{C16}: x_{n,m}^k\in[0,1],\forall m\in\mathcal{M},n\in\mathcal{N}_m,k\in\mathcal{K}_m, \quad (32)$$

[0127]

[0128] RSP4-3의 이중 문제는 다음과 같다.

$$\max_{\breve{\boldsymbol{\omega}},\breve{\boldsymbol{\gamma}}\succeq\mathbf{0}} g(\breve{\boldsymbol{\omega}},\breve{\boldsymbol{\gamma}}). \quad (33)$$

[0129]

[0130] 원 문제의 최소값이 이중 문제의 최대값과 같다는 것은 잘 알려져 있다. 따라서 이중 도메인에서 문제 RSP4 - 3을 해결할 수 있다.

[0131] KKT(Karush-Kuhn-Tucker) 조건을 사용하면 다음과 같이 주어진 Lagrange 승수를 사용하여 최적의 $s_{n,m}^k$, $x_{n,m}^k$ 사이의 관계를 얻을 수 있다.

$$s_{n,m}^k = x_{n,m}^k\left[\frac{\omega_{n,m}\tau B_k}{\gamma_{n,m}\ln 2}-\frac{N_0 B_k}{g_{n,m}^k}\right]^+. \quad (34)$$

[0132]

[0133] $s_{n,m}^k$ 을 식(32)로 대체하면 이중 함수는 식(35)로 나타낼 수 있다.

[0134] $\Lambda_{n,m}=\omega_{n,m}+Q_{n,m}-Z_n$ 라 하고,

$$\Gamma_{n,m}^k$$
$$=\gamma_{n,m}\left[\frac{\omega_{n,m}\tau B_k}{\gamma_{n,m}\ln 2}-\frac{N_0 B_k}{g_{n,m}^k}\right]^+-\omega_{n,m}\tau B_k$$

[0135]

$$\times\log_2\left(1+\left[\frac{\omega_{n,m}\tau B_k}{\gamma_{n,m}\ln 2}-\frac{N_0 B_k}{g_{n,m}^k}\right]^+\frac{g_{n,m}^k}{N_0 B_k}\right),$$
$$\forall m\in\mathcal{M},n\in\mathcal{N}_m,k\in\mathcal{K}_m, \quad (36)$$

[0136]

[0137] 본 발명의 실시예에 따른 이중 함수를 다음과 같이 다시 나타낼 수 있다.

$$
\begin{aligned}
&g(\breve{\omega},\breve{\gamma}) \\
&= \min_{\breve{a},\breve{x}} \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m}\sum_{k\in\mathcal{K}_m} \Gamma_{n,m}^{k} x_{n,m}^{k} \\
&\quad + \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m} \Lambda_{n,m} a_{n,m} - \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m} \gamma_{n,m} p_{n,m}^{\max} \\
&\text{s.t.} \quad \text{C12, C15, C16.}
\end{aligned} \tag{37}
$$

[0138]

[0139] 식(37)은 $a_{n,m}$ 과 $x_{n,m}^{k}$ 에 대해 분리 가능한 고전적인 선형 프로그래밍 문제임을 알 수 있다. 식(37)과 제약 조건 C15와 C16의 목적에서 첫 번째 항을 결합하면, 최적의 부반송파 할당은 다음과 같이 결정될 수 있다.

$$
x_{n,m}^{k} = \begin{cases} 1, \text{if } \Gamma_{n,m}^{k} < 0 \text{ and} \\ \quad n = \arg\min\limits_{n\in\mathcal{N}_m} \left\{\Gamma_{n,m}^{k}, \forall m \in \mathcal{M}, k \in \mathcal{K}_m\right\}, \\ 0, \text{otherwise.} \end{cases} \tag{38}
$$

[0140]

[0141] $s_{n,m}^{k} = x_{n,m}^{k} p_{n,m}^{k}$ 이므로, 식(34)에 따라 최적의 $x_{n,m}^{k}$ 을 계산한 후 최적의 송신 전력 $p_{n,m}^{k}$을 복구할 수 있다.

[0142] 최적의 작업 분할은 다음과 같이 식(37)와 제약 조건 C12의 목적에서 두 번째 항을 결합함으로써 얻을 수 있다.

$$
a_{n,m} = \begin{cases} A_n, \text{if } \Lambda_{n,m} < 0 \text{ and} \\ \quad m = \arg\min\limits_{m\in\mathcal{M}} \left\{\Lambda_{n,m}, \forall n \in \mathcal{N}_m\right\}, \\ 0, \text{otherwise.} \end{cases} \tag{39}
$$

[0143]

[0144] 최적의 $a_{n,m}$, $x_{n,m}^{k}$, $p_{n,m}^{k}$ 이 주어지면, Lagrange 승수는 다음과 같이 주어지는 하위 그레이디언트 하강법[36]을 사용하여 업데이트될 수 있다.

$$
\begin{aligned}
g(\breve{\omega},\breve{\gamma}) = \min_{\breve{a},\breve{x}} & \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m} (Q_{n,m} - Z_n)\, a_{n,m} \\
& + \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m} \omega_{n,m} \left(a_{n,m} - \sum_{k\in\mathcal{K}_m} \tau x_{n,m}^{k} B_k \log_2 \left(1 + \left[\frac{\omega_{n,m}\tau B_k}{\gamma_{n,m}\ln 2} - \frac{N_0 B_k}{g_{n,m}^{k}} \right]^{+} \frac{g_{n,m}^{k}}{N_0 B_k} \right) \right) \\
& + \sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m} \gamma_{n,m} \left(\sum_{k\in\mathcal{K}_m} x_{n,m}^{k} \left[\frac{\omega_{n,m}\tau B_k}{\gamma_{n,m}\ln 2} - \frac{N_0 B_k}{g_{n,m}^{k}} \right]^{+} - p_{n,m}^{\max} \right) \\
& \text{s.t.} \quad \text{C12, C15, C16.}
\end{aligned}
$$

[0145]

$$
\begin{aligned}
\omega_{n,m}^{l+1} &= \left[\omega_{n,m}^{l} + \kappa \left(a_{n,m} - \tau \sum_{k\in\mathcal{K}_m} x_{n,m}^{k} B_k \right.\right. \\
&\qquad \left.\left. \log_2 \left(1 + \frac{p_{n,m}^{k} g_{n,m}^{k}}{N_0 B_k} \right) \right) \right]^{+}, \\
\gamma_{n,m}^{l+1} &= \left[\gamma_{n,m}^{l} + \kappa \left(\sum_{k\in\mathcal{K}_m} p_{n,m}^{k} - p_{n,m}^{\max} \right) \right]^{+},
\end{aligned} \tag{40}
$$

[0146]

[0147] 여기서 κ 는 충분히 작은 단계 크기이고 l은 반복 지수이다. 하위 그레디언트는 Lipchitz 연속성 조건을 만족하므로, 최적의 $\breve{\omega}^{*}$ 및 $\breve{\gamma}^{*}$ 를 향한 승수의 수렴이 보장된다.

[0149] 도 5는 본 발명의 일 실시예에 따른 분산 공동 작업 분할 및 통신 자원 할당 알고리즘을 나타내는 도면이다.

[0150] 본 발명의 실시예에 따른 분산 구현에 있어서, Lagrange 이중 분해 방법으로 인해 공동 작업 분할 및 통신 자원 할당을 분산 방식으로 구현할 수 있다. 분산 공동 작업 분할 및 통신 자원 할당(Distributed Joint task Splitting and Communication Resource Allocation; DJTSCRA) 알고리즘의 절차 세부 사항은 알고리즘 2에 도시되어 있다. 분산 DJTSCRA 알고리즘에서 SD는 업링크가 전력 할당 및 작업 분할 결정을 전송하도록 하고, RAT 기지국은 부반송파 할당 결정을 내리고 Lagrange 승수를 업데이트하는 역할을 한다. 두 파트너의 결정은 Lagrange 승수가 수렴 임계값을 충족하거나 반복 지수 수가 최대 Lmax에 도달할 때까지 각 반복에서 개별 결정 결과를 서로 교환함으로써 대체적으로 수행된다.

[0151] 아래에서, 본 발명의 실시예에 따른 각 타임 슬롯에서 제안된 ATSRA 알고리즘의 계산 복잡성을 제공한다. 보조 변수 결정의 계산 복잡도는 폐쇄 형태로 해결되기 때문에 O(1)이다. 계산 빈도 할당의 계산 복잡도는 주로 $\{\rho_{n,m}(t)\}_{n\in\mathcal{N}}, \forall m \in \mathcal{M}$의 정렬에 의존하며, 이는 버블 정렬 알고리즘을 채택할 때 $O(|M||N|^2)$에 의해 주어질 수 있다. 공동 작업 분할 및 통신 리소스 할당 하위 문제는 DJTSCRA 알고리즘을 사용하여 해결된다. 종래기술 [37]에 따르면, δ-최적성 즉 $|g(\breve{\omega},\breve{\gamma}) - g(\breve{\omega}^*,\breve{\gamma}^*)| \le \delta$을 달성하기 위해, 하위 그레이디언트 방법은 변수 수에 의존하지 않는 $O(1\delta^2)$ 반복 비용을 사용하며, 이는 변수의 수에 의존하지 않는다. 각 SDn에 대해 송신 전력 $\breve{\mathbf{p}}_n = [p^k_{n,m}]_{m\in\mathcal{M}_n, k\in\mathcal{K}_m}$의 복잡도는 $O(\sum_{m\in\mathcal{M}_n}|\mathcal{K}_m|)$이고 각 반복에서 작업 분할 $\breve{\mathbf{a}}_n = [a_{n,m}]_{m\in\mathcal{M}_n}$ is $O(|\mathcal{M}_n|)$에 대해서는 O(|Mn|)이다. 결과적으로, SDn에 대한 총 계산 복잡도는 $O(1/\delta^2\sum_{m\in\mathcal{M}_n}(|\mathcal{K}_m|+1))$이다. 각 RAT 기지국 m에 대해, 부반송파 할당 $\breve{\mathbf{x}}_m = [x^k_{n,m}]_{n\in\mathcal{N}_m, k\in\mathcal{K}_m}$은 O(|Nm||Km|)이고, Lagrange 승수 $\breve{\omega}_m = [\omega_{n,m}]_{n\in\mathcal{N}_m}$ 및 $\breve{\gamma}_m = [\gamma_{n,m}]_{n\in\mathcal{N}_m}$을 업데이트하기 위한 복잡도는 $O(2|\mathcal{N}_m|)$이다. 결과적으로, RAT 기지국 m에 대한 총 계산 복잡도는 $O(1/\delta^2|\mathcal{N}_m|(|\mathcal{K}_m|+2))$ 이다. 따라서 DJTSCRA 알고리즘의 총 계산 복잡도는 $O(1/\delta^2[\sum_{m\in\mathcal{M}}|\mathcal{N}_m|(|\mathcal{K}_m|+2)+\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}_n}(|\mathcal{K}_m|+1)]) = O(1/\delta^2\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}_n}(|\mathcal{K}_m|+3))$ 이다. $\mathcal{M}_n \in , \mathcal{M}, \forall n \in \mathcal{N}$ 이므로, DJTSCRA 알고리즘의 계산 복잡도는 주어진 δ에서 $O(1/\delta^2\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}}(|\mathcal{K}_m|+3))$이다. 요약하면, ATSRA 알고리즘의 계산 복잡도는 각 타임 슬롯에 대해 $O(1/\delta^2\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}}(|\mathcal{K}_m|+3)+|\mathcal{M}||\mathcal{N}|^2)$ 이다.

[0152] 본 발명의 실시예에 따른 Slater 조건 가정에 따라 정리 1에 의해 제안된 ATSRA 알고리즘의 성능 결과를 제공한다.

[0153] 가정 1.(Slater 조건): 값 $\varepsilon > 0$ 및 $\phi(\varepsilon)$(여기서 $0 \le \phi(\varepsilon) \le \phi^{\max}$), 다음을 만족하는 s-전용 정책(s-only policies) $\hat{\alpha}(t) = \{\hat{\mathbf{a}}(t), \hat{\mathbf{x}}(t), \hat{\mathbf{p}}(t), \hat{\mathbf{f}}(t)\}$이 있다.

[0154]
$$\mathbb{E}\{\hat{\phi}(t)\} = \phi(\varepsilon),$$
$$\mathbb{E}\{\hat{\phi}(t)\} \le \mathbb{E}\{\hat{U}^o_n(t)\} - \varepsilon, \forall n \in \mathcal{N},$$
$$\mathbb{E}\{\hat{a}_{n,m}(t)\} \le \mathbb{E}\{\hat{b}_{n,m}(t)\} - \varepsilon, \forall m \in \mathcal{M}, n \in \mathcal{N}, \quad (41)$$

[0155] 여기서 s 전용 정책은 QSI Y(t)에 관계없이 관측된 NSI(t)의 랜덤 함수로서 독립적으로 제어 작용 $\hat{\alpha}(t)$를 선택하는 고정 및 무작위 정책의 특수 클래스로 정의된다.

[0156] 정리 1: SD $\bar{U}^o_{ws} = \min_{n\in\mathcal{N}}\bar{U}^o_n$ 가 SD 중 최소/최악의 오프로드 효율성, U^{opt}_{ws} 를 P1의 최적, ϕ^{opt}를 P3의 최적이라 하자. NSI s(t)가 타임 슬롯에 걸쳐 i.i.d. 이고 $\mathbb{E}\{L(\mathbf{Y}(0))\} < \infty$ 라 가정하면 가정 1이 충족된다. 그런 다음 모든 V >0에 대해 ATSRA 알고리즘은 다음과 같은 성능을 갖는다:

[0157] a) 최소/최악 오프로드 활용성의 성능 경계를 충족시킨다.

[0158] $$U_{ws}^{opt} - \bar{U}_{ws}^{o} \leq \phi^{opt} - \bar{\phi} \leq B/V. \qquad (42)$$

[0159] b) 모든 작업 대기열 및 가상 대기열은 평균 속도 안정적이므로 (7) 및 제약 조건 C8이 충족된다.

[0160] c) 시간 평균 작업 대기열 백로그에 성능 제한이 있다.

[0161] $$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}}\mathbb{E}\{Q_{n,m}\}\leq\frac{B+V(\phi^{max}-\phi(\varepsilon))}{\varepsilon}. \qquad (43)$$

[0162] 정리 1은 최소 오프로드 활용성 $\bar{U}_{ws}^{o}$가 최적 U_{ws}^{opt}의 O(1/V) 이내라 주장한다. 따라서, 충분히 큰 V 값을 사용하여 $\bar{U}_{ws}^{o}$를 최적에 근접하게 할 수 있다. 절충점은 O(1/V)가 최적에 근접하는 경우 O(V) 대기열 백로그 길이가 필요하다는 점이다.

[0163] 앞서 설명된 바와 같이 본 발명의 실시예들에 따르면 최소 장기적 시간 평균 오프로드 유틸리티를 최대화하는 것을 목표로 최대-최소 확률 최적화 문제를 공식화하고, 스마트 장치 작업 분할 및 업링크 전송 전력 할당, RAT 부반송파 할당 및 MEC 계산 주파수 할당을 공동으로 최적화하는 Lyapunov 최적화 기술을 기반으로 적응형 작업 분할 및 자원 할당을 수행할 수 있다. 제안하는 적응형 작업 분할 및 자원 할당 알고리즘은 분포 정보를 요구하지 않고 시간 변동 네트워크 역학을 수용할 수 있다.

[0165] 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.

[0166] 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.

[0167] 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.

[0168] 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.

[0169] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[0171] <참고문헌>

[0172] [1] M. Qin et al., "Service-oriented energy-latency tradeoff for IoT task partial offloading in MEC-enhanced multi-RAT networks," IEEE Internet Things J., vol. 8, no. 3, pp. 1896-1907, Feb. 2021.

[0173] [2] J. Ahn, J. Lee, S. Yoon, and J. K. Choi, "A novel resolution and power control scheme for energy-efficient mobile augmented reality applications in mobile edge computing," IEEE Wireless Commun. Lett., vol. 9, no. 6, pp. 750-754, Jun. 2020.

[0174] [3] E. El Haber, T. M. Nguyen, and C. Assi, "Joint optimization of computational cost and devices energy for task offloading in multi-tier edgeclouds," IEEE Trans. Commun., vol. 67, no. 5, pp. 3407-3421, May 2019.

[0175] [4] H. Wu, Y. Sun, and K. Wolter, "Energy-efficient decision making for mobile cloud offloading," IEEE Trans. Cloud Comput., vol. 8, no. 2, pp. 570-584, Jun. 2020.

[0176] [5] L. Lin, X. Liao, H. Jin, and P. Li, "Computation offloading toward edge computing," Proc. IEEE, vol. 107, no. 8, pp. 1584-1607, Aug. 2019.

[0177] [6] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," ETSI White Paper, vol. 11, pp. 1-16, 2015.

[0178] [7] Q. Zhang, L. Gui, F. Hou, J. Chen, S. Zhu, and F. Tian, "Dynamic task offloading and resource allocation for mobile-edge computing in dense cloud RAN," IEEE Internet Things J., vol. 7, no. 4, pp. 3282-3299, Apr. 2020.

[0179] [8] M. Qin et al., "Green-oriented dynamic resource-on-demand strategy for multi-RAT wireless networks powered by heterogeneous energy sources," IEEE Trans. Wireless Commun., vol. 19, no. 8, pp. 5547-5560, Aug. 2020.

[0180] [9] W. Wu, Q. Yang, P. Gong, and K. S. Kwak, "Energy-efficient traffic splitting for time-varying multi-RAT wireless networks," IEEE Trans. Veh. Technol., vol. 66, no. 7, pp. 6523-6535, Jul. 2017.

[0181] [10] 3GPP, "Study on small cell enhancements forE-UTRA and E-UTRAN: Higher Layer Aspects, Rel. 12," Tech. Rep. 36.842, Jul. 2014.

[0182] [11] 3GPP, "Feasibility study on licensed-assisted access to unlicensed spectrum, Rel. 13," Tech. Rep. 36.889, Jul. 2015.

[0183] [12] 3GPP, "Evolved universal terrestrial radio access network (E-UTRAN) and wireless LAN (WLAN); LTE-WLAN aggregation adaptation protocol, Rel. 13," Tech. Rep. 36.360, Mar. 2016.

[0184] [13] Z. Jing, Q. Yang, M. Qin, and K. Kwak, "Long term max-min fairness guarantee mechanism: Adaptive task splitting and resource allocation in MEC-enabled networks," in Proc. IEEE Int. Symp. Personal, Indoor Mobile Radio Commun., 2019, pp. 1-6.

[0185] [14] M. J. Neely, Stochastic Network Optimization With Application to Communication and Queueing Systems. San Rafael, CA, USA: Morgan and Claypool, 2010.

[0186] [15] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," IEEE Trans. Commun., vol. 64, no. 10, pp. 4268-4282, Oct. 2016.

[0187] [16] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," IEEE Internet Things J., vol. 6, no. 3, pp. 4188-4200, Jun. 2019.

[0188] [17] E. El Haber, T. M. Nguyen, C. Assi, and W. Ajib, "Macro-cell assisted task offloading in MEC-based heterogeneous networks with wireless backhaul," IEEE Trans. Netw. Service Manag., vol. 16, no. 4, pp. 1754-1767, Dec. 2019.

[0189] [18] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," IEEE Trans. Veh. Technol., vol. 68, no. 1, pp. 856-868, Jan. 2019.

[0190] [19] T. Q. Dinh, Q. D. La, T. Q. S. Quek, and H. Shin, "Learning for computation offloading in mobile edge computing," IEEE Trans. Commun., vol. 66, no. 12, pp. 6353-6367, Dec. 2018.

[0191] [20] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," IEEE Internet Things J., vol. 6, no. 3, pp. 4005-4018, Jun. 2019.

[0192] [21] Y. Liu, H. Yu, S. Xie, and Y. Zhang, "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," IEEE Trans. Veh. Technol., vol. 68, no. 11, pp. 11158-11168, Nov. 2019.

[0193] [22] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 5994-6009, Sep. 2017.

[0194] [23] C. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," IEEE Trans. Commun., vol. 67, no. 6, pp. 4132-4150, Jun. 2019.

[0195] [24] Z. Zhu et al., "Fair resource allocation for system throughput maximization in mobile edge computing," IEEE Access, vol. 6, pp. 5332-5340, 2018.

[0196] [25] X. Chen, Y. Cai, M. Zhao, and M. Zhao, "Joint computation offloading and resource allocation for min-max fairness in MEC systems," in Proc. IEEE Wireless Commun. Netw. Conf., Marrakesh, Morocco, 2019, pp. 1-6.

[0197] [26] X. Lyu et al., "Optimal schedule of mobile edge computing for internet of things using partial information," IEEE J. Sel. Areas Commun., vol. 35, no. 11, pp. 2606-2615, Nov. 2017.

[0198] [27] Y. Kim, H. Lee, and S. Chong, "Mobile computation offloading for application throughput fairness and energy efficiency," IEEE Trans. Wireless Commun., vol. 18, no. 1, pp. 3-19, Jan. 2019.

[0199] [28] L. Diez, A. Garcia-Saavedra, V. Valls, X. Li, X. Costa-Perez, and R. Agero, "LaSR: A supple multi-connectivity scheduler for multi-RAT OFDMA systems," IEEE Trans. Mobile Comput., vol. 19, no. 3, pp. 624-639, Mar. 2020.

[0200] [29] 3GPP, "Release Description; Summary of Rel-15 Work Items, Rel. 15," Tech. Rep. 21.915, Sep. 2019.

[0201] [30] L. Huang and M. J. Neely, "Max-weight achieves the exact [O(1/V), O(V)] utility-delay tradeoff under markov dynamics," Aug. 2010, arXiv:1008.0200.

[0202] [31] S. Kiamehr, M. Ebrahimi, M. S. Golanbari, and M. B. Tahoori, "Temperature-aware dynamic voltage scaling to improve energy efficiency of near-threshold computing," IEEE Trans. Very Large Scale Integr. Syst., vol. 25, no. 7, pp. 2017-2026, Jul. 2017.

[0203] [32] A. P. Miettinen and J. K. Nurminen, "Energy efficiency of mobile clients in cloud computing," in Proc. USENIX Conf. Hot Top. Cloud Comput., Boston, MA, USA, Jun. 2010, pp. 1-7.

[0204] [33] A. O. Allen, Probability, Statistics, and Queueing Theory. San Diego, USA: Academic Press, INC., 2014.

[0205] [34] D. Pisinger, "Algorithms for Knapsack Problems," Ph.D. dissertation, Dept. Comput. Sci., Copenhagen, Denmark: Univ. Copenhagen, 1995.

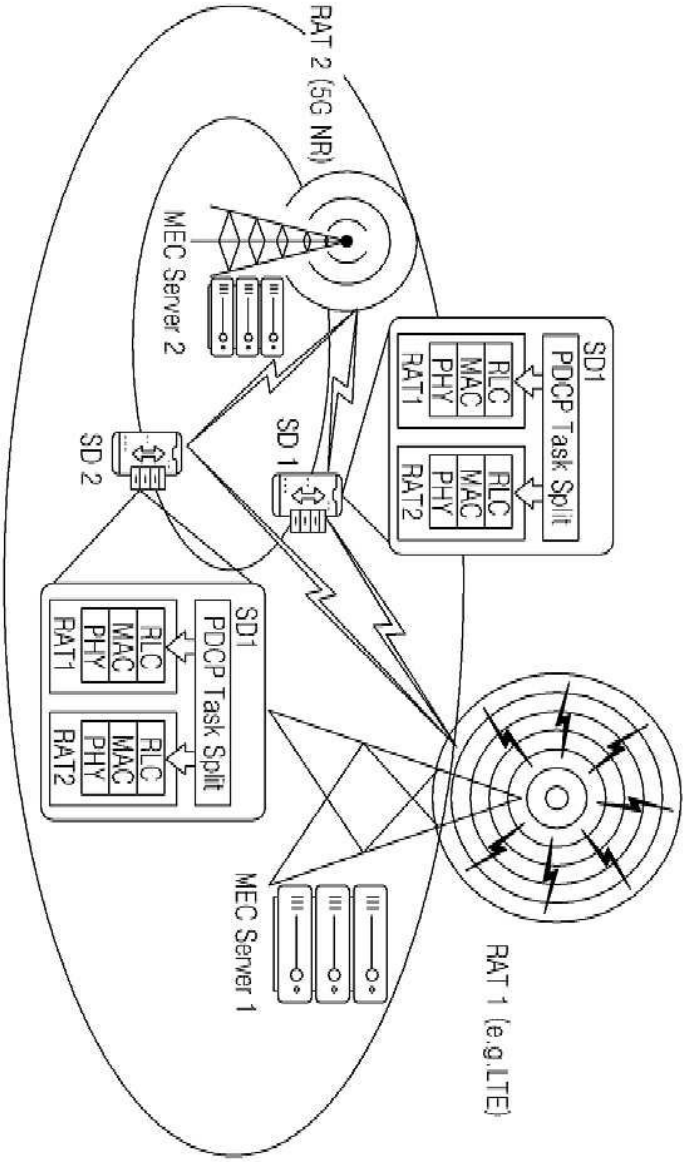
[0206] [35] M. Qin et al., "Learning-aided multiple time-scale SON function coordination in ultra-dense small-cell networks," IEEE Trans. Wireless Commun., vol. 18, no. 4, pp. 2080-2092, Apr. 2019.

[0207] [36] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

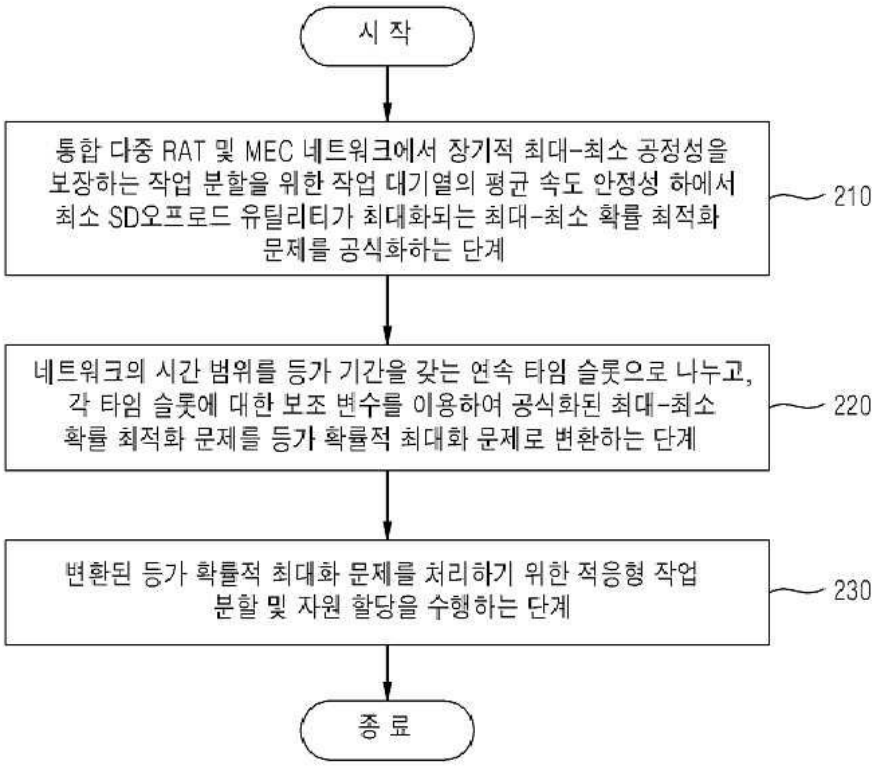
[0208] [37] N. Mokari, M. R. Javan, and K. Navaie, "Cross-layer resource allocation in OFDMA systems for heterogeneous traffic with imperfect CSI," IEEE Trans. Veh. Technol., vol. 59, no. 2, pp. 1011-1017, Feb. 2010.

도면

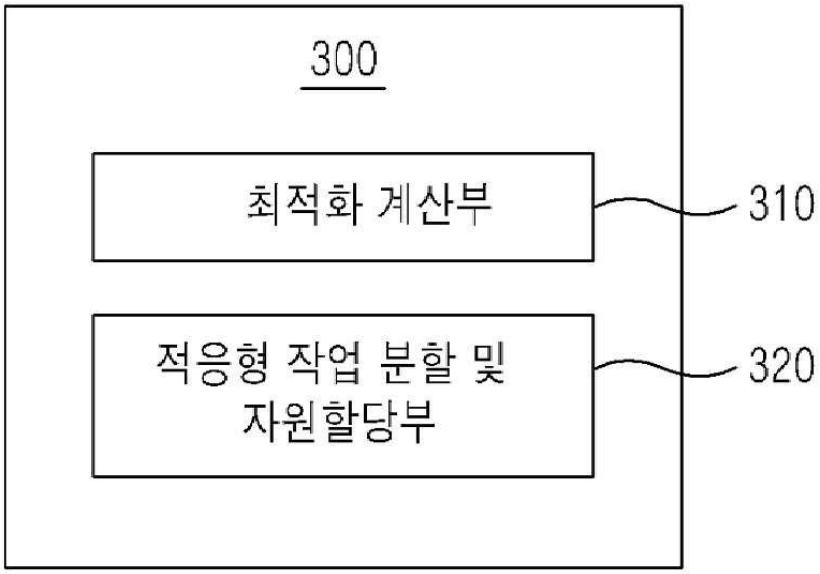
도면1


RAT 2 (5G NR)
MEC Server 2
SD 2
SD 1
SD1
PDCP Task Split
RLC
MAC
PHY
RAT1
RAT2
MEC Server 1
RAT 1 (e.g.LTE)

도면2


시 작
통합 다중 RAT 및 MEC 네트워크에서 장기적 최대-최소 공정성을 보장하는 작업 분할을 위한 작업 대기열의 평균 속도 안정성 하에서 최소 SD오프로드 유틸리티가 최대화되는 최대-최소 확률 최적화 문제를 공식화하는 단계
210
네트워크의 시간 범위를 등가 기간을 갖는 연속 타임 슬롯으로 나누고, 각 타임 슬롯에 대한 보조 변수를 이용하여 공식화된 최대-최소 확률 최적화 문제를 등가 확률적 최대화 문제로 변환하는 단계
220
변환된 등가 확률적 최대화 문제를 처리하기 위한 적응형 작업 분할 및 자원 할당을 수행하는 단계
230
종 료

도면3


300
최적화 계산부
310
적응형 작업 분할 및 자원할당부
320

도면4

Algorithm 1: The ATSRA Algorithm.

1: At the beginning of each time-slot $t \in \{0, 1, 2, \cdots\}$, observe the current NSI $s(t)$ and QSI $Y(t)$.

2: Determine the task splitting and resource allocation decision $\alpha(t) = \{a(t), x(t), p(t), f(t)\}$ and the auxiliary variable $\phi(t)$ by solving the following problem:

P4 :

$$\min_{a(t),x(t),p(t),f(t),\phi(t)} \sum_{n\in N} Zn(t)\,(\phi(t) - U_n^o(t))$$

$$+ \sum_{m\in M}\sum_{n\in N} Q_{n-m}(t)\,(a_{n-m}(t) - b_{n-m}(t)) - V\,\phi(t)$$

$$\text{s.t. } C2 - C7, C9.$$

3: Update task queues $Q(t)$ and virtual queues $Z(t)$ according to (6) and (11), respectively.

도면5

Algorithm 2: The DJTSCRA Algorithm.

Input: The iteration number $l = 1$ and its maximum value L^{max}, the convergence factor ϵ, the step size κ, the current QSI $Y(t)$ and NSI $s(t)$, the AWGN spectral density $N0$ and the CPG $g(t)$;

Output: The optimal $\check{a}^*$, $\check{x}^*$ and $\check{p}^*$;

1: for all RAT base stations $m \in M$ do
2: Initialize its Lagrange multipliers $\check{\omega}_m^1 = [\omega_{n,m}^1]_{n \in Nm}$ and $\check{\gamma}_m^1 = [\gamma_{n,m}^1]_{n \in Nm}$, then broadcast them to all SDs;
3: end for
4: repeat
5: for all SDs $n \in N$ do
6: Compute the uplink transmit power $\check{p}_n = [p_{n,m}^k]_{m \in M_n, k \in K_m}$ and the task splitting $\check{a}_n = [a_{n,m}]_{m \in M_n}$ according to (34) and (39), respectively, then broadcast them to RAT base stations $m \in M_n$;
7: end for
8: for all RAT base stations $m \in M$ do
9: Compute the subcarrier allocation $\check{x}_m = [x_{n,m}^k]_{n \in N_m, k \in K_m}$ according to (38) and update the Lagrange multipliers $\check{\omega}_m = [\omega_{n,m}]_{n \in N_m}$ and $\check{\gamma}_m^1 = [\gamma_{n,m}]_{n \in N_m}$ according to (40), then broadcast them to SDs $n \in N_m$;
10: end for
11: until $(\|\check{\omega}^{l+1} - \check{\omega}\| + \|\check{\gamma}_m^{l+1} - \check{\gamma}_m^{l+1}\|)/2 \leq \epsilon$ or