

eMBB与URLLC业务共存的MEC场景下基于拍卖算法的计算卸载策略

陈锡宇¹, 窦海娥¹, 姚继明^{1,2}, 王磊¹, 夏志杰¹

(1. 南京邮电大学通信与信息工程学院, 江苏南京 210003; 2. 国网智能电网研究院有限公司, 北京 102211)

摘要: 根据5G通信的发展规划, 当前超可靠低时延通信(URLLC)与增强移动宽带(eMBB)业务共存是极为常见的应用场景。由于通信系统资源有限, 故大量用户请求的业务会产生激烈竞争。同时, 移动边缘计算指将移动用户计算任务卸载到资源更加充足的小区边缘服务器上计算, 具有低能耗和低时延的优点。针对两种业务共存的多用户多边缘服务器场景, 以用户计算能耗和传输时延以及边缘服务器收益为目标, 基于拍卖理论建立多对多拍卖模型, 提出一种基于多回合双向拍卖算法的计算卸载策略, 在每回合拍卖结束后根据资源剩余量动态调整用户出价和服务器报价, 以提高拍卖成功率, 并确保边缘服务器总收益。仿真结果表明, 该计算卸载算法能够合理分配边缘服务器的有限资源, 并提高边缘服务器总收益。

关键词: eMBB; URLLC; 移动边缘计算; 计算卸载策略; 多回合双向拍卖

DOI: 10.11907/tjdc.241302

开放科学(资源服务)标识码(OSID):

中图分类号: TP393

文献标识码: A

文章编号: 1672-7800(2025)005-0137-09



Computation Offloading Strategy Based on Auction Algorithm in MEC Scenario Where eMBB and URLLC Services Coexist

CHEN Xiyu¹, DOU Hai'e¹, YAO Jiming^{1,2}, WANG Lei¹, XIA Zhijie¹

(1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 2. State Grid Smart Grid Research Institute Co., Ltd., Beijing 102211, China)

Abstract: According to the development plan of 5G communication, the coexistence of URLLC and eMBB services is a very common application scenario at present. Due to the limited resources of the communication system, the service with a large number of user requests will produce fierce competition. At the same time, mobile edge computing (MEC) refers to the offloading of mobile user computing tasks to the edge server of the cell with more resources, which has the advantages of low energy consumption and low latency. Aiming at the multi-user multi-edge server scenario where two kinds of services coexist, this paper takes user calculation of energy consumption, transmission delay and edge server revenue as targets, establishes a many-to-many auction model based on auction theory, and proposes a computational offloading strategy based on multi-round bidirectional auction algorithm. After each round of auction, the user bid and server bid are dynamically adjusted according to the amount of resources remaining, to improve the auction success rate and ensure the total revenue of the edge server. Finally, the simulation results show that the algorithm proposed in this paper can reasonably allocate the limited resources of edge server and improve the total revenue of edge server.

Key Words: eMBB; URLLC; mobile edge computing; computation offloading strategy; multi-round two-way auction

收稿日期: 2024-04-26

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(62071255); 江苏省重点研发计划项目(BE2023087); 江苏省高等学校自然科学研究重大项目(20KJA510009); 南京邮电大学宽带无线通信与传感网技术教育部重点实验室开放研究基金项目(JZNY202312)

作者简介: 陈锡宇(1998-), 男, 南京邮电大学通信与信息工程学院硕士研究生, 研究方向为5G网络移动边缘计算; 窦海娥(1980-), 女, 南京邮电大学通信与信息工程学院博士研究生、高级实验师, 研究方向为多媒体无线通信; 姚继明(1987-), 男, 南京邮电大学通信与信息工程学院博士研究生, 国网智能电网研究院有限公司高级工程师, 研究方向为电力无线通信技术; 王磊(1977-), 男, 博士, 南京邮电大学通信与信息工程学院教授、博士生导师, 研究方向为多媒体无线通信; 夏志杰(1964-), 男, 博士, 南京邮电大学通信与信息工程学院教授、博士生导师, 研究方向为工业互联网。本文通讯作者: 窦海娥。

0 引言

国际电信联盟在ITU-RWP5D第22次会议明确指出5G网络主要覆盖三大应用场景:增强移动宽带(Enhanced Mobile Broadband, eMBB)、超可靠低时延通信(Ultra-reliable Low-latency Communication, URLLC)和大规模机器通信(Large-scale Machine Communication, mMTC)^[1]。其中, eMBB业务目的是提高用户间的通信体验,要求用户体验速率达到1 Gbps,特定场景下满足20 Gbps的峰值速率,涵盖了3D、超高清视频、AR/VR、云游戏等超大流量移动带宽和时延中等的业务;URLLC业务作为全新的领域,要求毫秒级别的端到端时延和高达99.999%的传输可靠性,涵盖了诸如智能电网、工业自动化、自动驾驶、移动医疗、触觉互联网等要求超低时延和超高可靠性的业务。随着通信设备数量的飞速增长,计算密集型应用正在成为人们日常生活中不可分割的一部分。然而,由于移动设备的计算能力和设备寿命有限,在本地设备上处理数据成为一个需要考虑的重要问题。针对该问题,移动云计算(Mobile Cloud Computing, MCC)是一种解决方案。在MCC中,通过蜂窝网络将计算密集型任务卸载到云服务器,可以减少资源受限设备的能源消耗。接着,云服务器执行卸载任务,并将输出返回给设备。但是,云服务器距离移动设备很远,会产生较大的传输时延,无法满足一些低时延的业务需求^[2]。于是,无线通信行业和学术研究界推出了一种名为移动边缘计算(Mobile Edge Computing, MEC)的新技术。根据当前发展策略,从5G通信发展需求的角度出发,在URLLC与eMBB两种业务共存的场景中,移动边缘计算便是其中之一^[3]。

MEC技术将网络控制、存储和移动计算等业务转移到网络边缘进行,以便资源有限的用户端移动设备可以进行计算复杂和低延迟的业务,将显著减少延迟和移动设备能耗,可以降低时延、设备节能、环境感知和增强隐私/安全性。

当前,5G系统中如何合理规划用户计算卸载选择,实现网络资源的充分利用是5G无线资源管理中亟待解决的问题^[4]。其中,MEC系统中的资源分配便是多用户MEC系统中非常值得研究的关键问题之一,由于不同业务对于服务质量(Quality of Service, QoS)的要求不同,在MEC系统中支持不同业务是一大挑战。当很多移动用户同时将计算任务卸载到MEC服务器上时,会出现严重干扰问题。随着系统中移动用户数量的不断增加,MEC服务器存在较多计算任务时会导致速率下降,易产生严重干扰,故进行卸载任务的移动用户数量并不会随之增加,反而会有所减少。因此,在保证服务质量的前提下,如何制定计算卸载策略显得尤为重要。

针对以上问题,本文研究了eMBB与URLLC业务共存

的多用户多服务器场景下的计算卸载策略。一方面,要保证用户时延和能耗在约束条件之内;另一方面,要考虑边缘服务器的收益最大化。因此,本文建立了多对多拍卖模型,设用户设备为买方,边缘服务器为卖方,基站为拍卖人,提出一种基于多回合双向拍卖的资源分配方法,随着拍卖轮次的推进动态调整用户出价与服务器报价以提高拍卖成功率。仿真结果表明,相比于现有的资源分配和计算卸载策略,本文方法能够合理分配资源,有效提高边缘服务器的收益。

本文建立eMBB与URLLC共存场景下的MEC网络模型,将计算卸载与资源分配决策问题建模为多对多拍卖模型,并对多个MEC服务器的情况进行研究,建立最优化问题。接着,基于多回合拍卖方法,在综合考虑传输和计算延迟以及能量消耗的同时,兼顾边缘服务器的收益,使得有限资源得到高效分配,并以此为目标函数解出最优解。对本文方法进行模拟仿真,并与现有其他方法进行比较与分析。最后,对未来可深入研究的方面进行了展望。

1 相关工作

1.1 eMBB与URLLC共存的MEC场景下资源调度现状

对于MEC场景下的资源调度策略,学者们进行了大量研究。文献[5]研究了移动边缘计算和人工智能的6G物联网覆盖漏洞(Coverage Holes, CHs)恢复问题,基于圆盘模型和置信信息模型的融合模型,移动边缘节点使用该实用程序引导其移动,并通过基于Q-learning的重复博弈以恢复CHs。在此基础上,提出了一种改进的可信信息覆盖漏洞恢复算法(Improved Confident Information Coverage Hole Recovery Algorithm, I-CICRA),提高了恢复性能和修复效率,降低了时间复杂度。与基于分布式收益的Q-learning算法的一系列仿真比较表明,所提出的算法在总覆盖、迭代次数、能耗和能效方面具有更好的性能。文献[6]为了提高MEC服务器上的资源管理效率,设计一种改进的基于广义第二价格的计算效率较高的算法,该算法通过考虑动态卸载请求的到达和服务器的计算工作量以决定定价和资源分配。在此基础上,提出了满足对称纳什均衡和个体理性的资源卖方自适应最优竞价策略。最后,通过大量数值结果证明了所提出的资源分配机制的有效性。文献[7]针对MEC预留资源会造成延迟和资源浪费等问题,提出了一个共享平台,预订计划用户可以将未使用的资源配额转售给按需用户。建立了一个三阶段非合作Stackelberg博弈模型,并研究了买卖双方的最优策略,求出了移动运营商对按需用户的最优价格。结果表明,当按需池的供给较低时,以及当按需池的佣金较低时,同时拥有两个池使得运营商获得最优收益。

关于eMBB与URLLC共存场景下存在的问题,许多学者对该领域进行了深入研究。文献[8]将eMBB和URLLC

用户的节能任务卸载和调度作为一个混合整数非线性问题,并分解为多个凸形式子问题,然后使用块坐标下降算法交替求解,结果表明该方法能有效降低能耗。文献[9]研究了同时支持URLLC和eMBB两种业务的网络模式,提出一种新的功率控制和指示方法,在保证URLLC可靠性要求的同时,为用户设备动态提供不同的功率控制参数,并将对eMBB的影响降到最低。文献[10]研究了如何在具有URLLC和eMBB业务的MEC系统中分析延迟并优化短数据包卸载方案,在MEC系统中部署Processor-Sharing(PS)服务器,将总计算能力平均分配给服务器中的所有报文,推导了短数据包所经历的延迟互补累积分布函数的封闭表达式,表明PS服务器在短数据包和长数据包的场景下都优于两种先到先服务的服务器。

目前,针对MEC系统中URLLC业务与eMBB业务共存的资源分配研究依然存在两大问题:①优化目标单一,当前一些研究仅以能耗或时延为优化目标,在MEC系统最新的研究方向中,时延和能耗应尽量共同考虑;②目前多数研究只着眼于移动终端的时延和能耗最优化,却忽略了边缘服务器端在计算卸载中的计算成本或收益。增加MEC服务器的数量并不一定能缩短整个系统的时延,因为部署过多的MEC服务器,将导致无线资源紧张,造成传输时延大于计算时延的情况,同时还可能导致能耗增加。因此,选择适当的MEC服务器集群对整个系统的性能表现起着关键作用。

1.2 拍卖算法

在经济学中,拍卖理论是一种重要理论,在许多其他领域中也得到了广泛的传播与运用。拍卖的本质是一种不完全信息博弈。其中,卖方的目标是利益最大化,买方的目标则是付出尽可能少的价格。通过这种双方竞价博弈的方式,可以发掘拍卖物品的最大价值,提高整体收益。按照拍卖方式可分为单向拍卖和双向拍卖两种不同类型。

单向拍卖指拍卖中一个买方对多个卖方或者一个卖方对多个买方,单向拍卖进行过程中会有“资源优势方”和“信息优势方”。若是一个买方对多个卖方,则卖方为“资源优势方”,买方称为“信息优势方”;相反,若是一个卖方对多个买方,则买方为“资源优势方”,卖方称为“信息优势方”。

双向拍卖指拍卖存在多个买方及多个卖方,相比于单向拍卖,双向拍卖中没有“资源优势方”和“信息优势方”,双方追求一种更加平等的供求关系,且在资源不够充足的情况下,双方能够通过博弈迅速取得平衡点,提高总收益和拍卖效率,在当前资源分配研究领域也是一种常用方法。

双向拍卖主要流程如下:①买卖双方分别提交各自的出价和要价,同时提交买方所需的拍卖品数量与卖方参加交易的拍卖品数量;②根据拍卖市场中的报价要求,判断双方出价或要价是否符合规范,若合理则进行下一步操

作,若不合理则需要重新提交出价或要价;③按照交易规范,由拍卖者匹配买卖双方的出价与要价,确认其交易的物品种类与数量,匹配成功后买方和卖方进行交易,分别获得拍卖品和资金;④根据信息公开规则,需要向市场公开交易成功的信息,该信息需要包含成交价格与成交拍卖品的数量;⑤若交易满足了结束条件,则关闭双向拍卖市场,完成交易结束流程;若不满足,则返回步骤①,重新开始新的交易。

除经济领域外,拍卖理论在其他领域也得到了广泛应用。文献[11]为了解决跨数据中心的资源分配问题,提出了一种基于优先级组合双拍卖(Priority Combinatorial Double Auction, PCDA)的资源分配方案。拍卖过程包括3个阶段:云用户代理报价、云资源提供商竞价、资源管理机构竞价协商。通过用户优先级和任务紧急性,估算拍卖过程中每个任务的违规情况,并计算云提供商的利润。仿真结果表明,所提出的PCDA能够保证投标成功率。与传统的双向拍卖相比,PCDA在拍卖期间可减少35.00%的能源消耗,并使云提供商的利润增加约38.84%。文献[12]提出了一种基于功率感知和可信组合双拍卖(Power-aware and Credible Combinatorial Double Auction, PACCDA)的云资源分配方法。该方法综合考虑了供应商和用户的信用、能源成本等几个因素,以抵制拍卖中的不诚实和恶意行为,在资源交易中,更加注重参与者得到的服务质量,减小过度的能源消耗。仿真结果表明,PACCDA在拍卖成功数量、服务提供商利润和消费者利润方面具有更好的性能。文献[13]为了激励服务器所有者提供计算资源,并提高服务购买者的端到端服务质量,通过计算服务拍卖(Computing Service Auction, COMSA)解决这些问题,与现有边缘计算双拍卖方案主要关注计算资源交易不同,COMSA明确考虑了频谱分配和数据路由,解决了双拍卖机制设计和网络资源分配的联合问题,为边缘计算服务提供了端到端服务质量保障。大量的仿真结果证明,COMSA具有重要的经济属性优势,即真实性、预算平衡性和个体合理性。

综上,拍卖理论在通信研究领域的应用十分广泛。本文所建立的模型为多边缘服务器与多用户共存的场景,与双向拍卖模型特征相似,故将该模型与本文场景相结合,以达到本文研究目标,即提高边缘服务器总收益。

2 系统模型

2.1 网络场景

本文考虑的场景为一个多用户、多MEC边缘服务器所在的小区,如图1所示。该模型中有 K 个基站, M 个eMBB用户和 N 个URLLC用户随机分布在一定范围内,每个用户都有一个计算任务待执行。其中,由于URLLC业务具有突发性和零散性的特点,几乎不会发生大量URLLC用户同时请求的情况,故 $M \gg N^{[14]}$ 。每个基站配备一个边缘服

服务器,用户可以选择在本地设备上计算,或者卸载至边缘服务器上加以处理。由于边缘服务器资源有限,将资源分配问题建模为多回合双向拍卖模型,通过设计竞价策略,使资源在买卖双方的竞争中得到高效分配,在减小时延和能耗的同时让边缘服务器获得更大收益。

假设本文场景中所有边缘服务器均可并行处理,每个边缘服务器中存有 L 种不同资源,例如通信资源、存储资源和计算资源等。可将边缘服务器 k 持有的资源状态用一组向量表示为 $resource_k = (c_{k1}, c_{k2}, \dots, c_{kL})$ 。

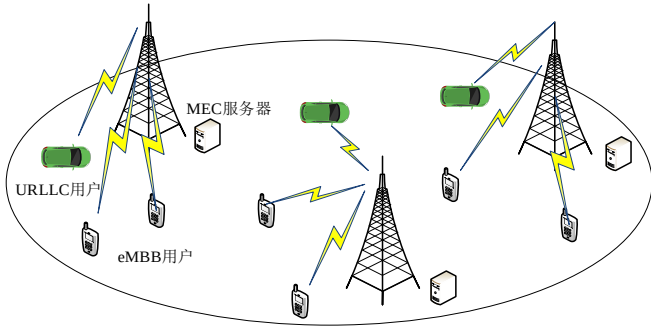


Fig. 1 Cell model diagram of multi-user and multi-MEC edge server coexistence

图1 多用户多MEC边缘服务器共存的小区模型

设eMBB业务的可达速率为 R_m ,可表示为式(1)^[15]。

$$R_m = N_m W \log_2 \left(1 + \frac{g_m G_m P_m}{N_0} \right) \quad (1)$$

其中, N_m 为配给eMBB用户 m 的子载波数目, W 为频带资源子载波带宽, P_m 为eMBB用户 m 的发射功率, N_0 为加性高斯白噪声功率, g_m 为eMBB用户的小尺度瑞利衰落, G_m 为eMBB用户 m 的路径损耗。根据文献[16],路径损耗模型可表示为式(2)。

$$G_m = 128.1 + 37.6 \lg(d_m) \quad (2)$$

其中, d_m 为eMBB用户 m 到基站的距离。

设eMBB用户 m 的本地计算时延为 T_m^l ,可表示为式(3)。

$$T_m^l = c_m / f_m^l \quad (3)$$

其中, c_m 为处理业务所需CPU周期数, f_m^l 是本地计算能力。

eMBB用户 m 本地计算能耗 $E_{m,l}$ 可表示为式(4)。

$$E_{m,l} = \sigma (f_m^l)^2 c_m \quad (4)$$

其中, σ 为能耗因子,与eMBB用户设备的芯片性能相关。

计算卸载时延一般分为3个部分:计算任务上传、服务器执行和计算结果返回。计算任务上传指在作出计算卸载决策后,移动终端将相应的计算任务通过无线链路,如移动蜂窝网络、Wi-Fi网络等,上传至边缘服务器协助执行的过程。服务器执行是指当任务被卸载到边缘服务器后,所在服务器会根据制定的卸载策略为其提供相应的计算资源,协助移动终端完成任务计算。计算结果返回指在

边缘服务器完成卸载部分的计算工作后,所得的计算结果将通过无线链路回传给移动终端。通常情况下,相比于输入数据量,计算结果的数据量非常小,其造成的传输时延可忽略不计^[17]。

本文eMBB用户 m 的卸载计算时延由两部分组成:任务卸载时延和服务器计算时延,结合无线传输速率模型,任务卸载时延可以写成式(5)。

$$t_m^{off} = \frac{l_m}{R_m} \quad (5)$$

其中, l_m 表示eMBB用户 m 的任务比特量(bits),该用户卸载计算所需时间可以写成式(6)。

$$t_m^c = \frac{c_m}{F_m^c} \quad (6)$$

其中, F_m^c 表示MEC服务器处理eMBB用户 m 的任务时的计算能力。

考虑到上行传输所需时间要远大于下行传输,故本文仅考虑数据从用户传输到边缘服务器,将传输时延与任务计算时延相加,故该用户的卸载计算总时延可以写成式(7)。

$$T_m^s = t_m^{off} + t_m^c = \frac{l_m}{R_m} + \frac{c_m}{F_m^c} \quad (7)$$

如式(8)所示,设 x_m^k 为二进制参数为分配参数,用来决定eMBB用户 m 的计算任务是否卸载到服务器 k 上进行处理。

$$x_m^k = \begin{cases} 1 & \text{当eMBB用户 } m \text{ 任务卸载到服务器 } k \text{ 进行计算} \\ 0 & \text{当eMBB用户 } m \text{ 任务不在服务器 } k \text{ 进行计算} \end{cases} \quad (8)$$

eMBB用户 m 的数据卸载在边缘服务器 k 上进行计算所需要的能耗如式(9)所示。

$$E_{m,off} = x_m^k P_m \frac{l_m}{R_m} \quad (9)$$

对于不同用户和任务,其时间延迟和能耗要求也不同。在某些低时延的场景中,用户可能会以牺牲一部分能耗为代价,换取更短的时延;反之,在对时延要求不高的场景下,用户也可以选择牺牲时延,以降低能耗,减小成本。故式(7)、式(9)可改写为式(10)、式(11)。

$$T_m^s = \zeta_m \left(\frac{l_m}{R_m} + \frac{c_m}{F_m^c} \right) \quad (10)$$

$$E_{m,off} = \frac{1}{\xi_m} x_m^k P_m \frac{l_m}{R_m} \quad (11)$$

其中, ζ_m 为用户 m 的代价因子,根据任务的具体要求可表现用户期望实现的卸载代价。

2.2 多回合双向拍卖模型

在进行资源分配时除考虑到用户的时延与能耗外,还要兼顾服务器的收益,故本文采用一种基于拍卖理论的资源分配方法,通过用户间的竞争和相互博弈以提高分配效率。假设边缘服务器扮演卖方和拍卖人的角色,用户为买方,如图2所示。边缘服务器为用户提供计算资源,以完成用户的卸载计算任务。同时,用户也要向边缘服务器支

付一定报酬, 支付规则由拍卖人决定。为了保证用户端效用, 用户最终支付不能大于自身对资源需求的估值^[18]。

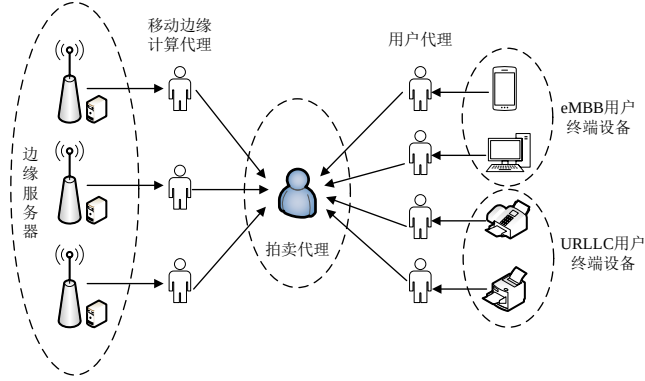


Fig. 2 Two-way auction model

图2 双向拍卖模型

在本文所建立的 MEC 模型中, 计算卸载包括两个部分: 一是用户设备将计算任务发送至服务器的传输时间; 二是边缘服务器计算过程所需时间, 下行传输时间相对较短, 可忽略不计。为了完成计算任务, 用户终端需要向服务器请求无线资源和计算资源。当边缘服务器持有的资源量能够满足移动终端需求时, 则双方成功完成拍卖。

(1) 成本。对于移动边缘计算而言, 成本与诸多因素有关, 例如存储资源成本、通信资源成本和计算资源成本等。边缘服务器的存储空间越大, 通信资源越多, 计算速度越快, 则性能越好, 但该服务器部署建设和维护成本也越高。上文假设了边缘服务器持有 L 种不同资源, 考虑到各项因素可能影响程度不同, 故边缘服务器 k 的成本可表示为式(12)。

$$C_k = \sum_{i=1}^L \lambda_{ki} c_{ki} \quad (12)$$

其中, c_{ki} 为边缘服务器 k 中的一种资源总量, λ_{ki} 为该种资源对应的成本影响权重。

(2) 估值。估值指用户愿意接收的最高价格, 估值与自身业务需求有关, 业务优先级越高, 用户愿意支付的价格越高, 即估值越高^[19]。对于本文建立的模型而言, 主要考虑了 eMBB 用户的业务时延和能耗, 可表示为式(13)。

$$v_m = \beta_m \delta_{mT} T_{e,m}^s + (1 - \beta_m) \delta_{mE} E_{m,off} \quad (13)$$

其中, β_m 和 $1 - \beta_m$ 分别为用户 m 业务时延和能耗的估值权重参数, 考虑到时延与能耗难以同时兼顾, 故令二者之和为 1。 δ_{mT} 和 δ_{mE} 分别为时延代价因子和能耗代价因子, 用户根据实际业务需求进行调整, 因子越大, 估值越小, 表示该任务优先级越小, 在拍卖中的优势也就越小。

(3) 出价策略。对于用户而言, 首先要处理优先级高的任务, 否则会导致截止时间长而失去竞价优势^[20]。因此, 本文基于拍卖补偿策略, 设计了面向不同优先级的用户出价策略。

首先, 由于 URLLC 业务具有高可靠和低延迟的特性, 故 URLLC 用户优先级普遍高于 eMBB 用户, 当接收到 URLLC 用户的请求时, 服务器应优先进行处理。此外, 考

虑到 URLLC 业务具有突发性和零散性, 几乎不会出现多个 URLLC 业务同时发送请求从而产生竞争的情况^[21]。因此, 下文就 eMBB 用户间资源竞争进行出价策略设计。

对于 eMBB 用户而言, 对资源的需求较大, 故不同优先级用户需要有不同的出价策略以进行资源分配, 用户出价如式(14)所示。

$$b_m(t+1) = \frac{\varepsilon_b \text{Sum}}{\text{Sum} - t} \text{Pro}_m b_m(t) \quad (14)$$

其中, t 表示当前拍卖轮数, Sum 表示拍卖总轮数, Pro_m 为用户任务的优先级, ε_b 为调节用户出价的参数, 用于调节不同优先级的任务间出价差异, 优先级越高的用户, 出价上涨速度也越快, 在下一轮拍卖中也就具有更大的竞争优势, 其计算任务更容易优先得到服务器处理^[22]。

对于服务器而言, 随着拍卖轮数的推进, 其剩余资源量和用户数量越来越少, 为了保持在拍卖中的竞争力, 需适当地降低价格以保持一定优势, 故边缘服务器要价如式(15)所示。

$$a_k(t+1) = \varepsilon_a \frac{\text{Sum} - t}{\text{Sum}} \mu_k(t) a_k(t) \quad (15)$$

其中, ε_a 为调节边缘服务器要价的参数, $\mu_k(t)$ 为边缘服务器剩余资源量对报价的影响参数, 具体如式(16)所示。

$$\mu_k(t) = \prod_{i=1}^L \frac{c_{ki,t}}{c_{ki}} \quad (16)$$

其中, $c_{ki,t}$ 为当前轮数 t 时边缘服务器 k 剩余的某种资源 i 的数量, 随着拍卖的推进, 资源量逐渐减少, $\mu_k(t)$ 减小, 边缘服务器要价也随之降低, 使得服务器能够保持一定优势, 以提高拍卖成功率。

3 问题公式化与求解

根据所建立的系统模型, 将问题公式化, 利用双向拍卖思想, 解决如何为用户分配最佳的边缘服务器以进行卸载计算, 并综合考虑服务器收益以及卸载计算时延和能耗最优化。

3.1 问题建立

根据式(8)可知, 本文考虑计算任务不可分割的情况, 用户只能选择将任务整体卸载到边缘服务器进行计算或是进行本地计算。因此, 用户 m 可以带给边缘服务器 k 的效益可表示为该用户支付报酬减去该服务器成本^[23], 如式(17)所示。

$$U_m^k = x_m^k (b_m - C_k), \quad (17)$$

边缘服务器的总收益可写成每个用户带给对应服务器的收益总和, 如式(18)所示。

$$U = \sum_{k=1}^K \sum_{m=1}^M U_m^k, \quad (18)$$

除考虑服务器收益外, 本文还需综合考虑计算卸载的时延与能耗, 故最优化问题如式(19)所示。

$$\begin{aligned}
\max \quad & U - \alpha_T T_m^s - \alpha_E E_{m, \text{off}} \\
\text{s.t.} \quad & C1: \sum_{m=1}^M x_m^k c_{m, ki} \leq C_{ki}, \forall k \in \{1, 2, \dots, K\}, \forall i \in \{1, 2, \dots, L\} \\
& C2: \sum_{k=1}^K x_m^k \leq 1, \forall m \in \{1, 2, \dots, M\} \\
& C3: x_m^k = \{0, 1\} \\
& C4: T_m^s \leq T_m, \forall m \in \{1, 2, \dots, M\} \\
& C5: \xi_m \geq 0
\end{aligned} \tag{19}$$

其中, C1 表示每个边缘服务器为对应用户提供的任意资源总和不能大于其自身该资源的持有量; C2 表示一个任务若选择卸载计算, 只能由一个边缘服务器完成; C3 表示用户的每个计算任务只能整体卸载或是本地计算, 不可分割; C4 表示 eMBB 用户 m 的时延不得超过业务需求; C5 表示时延和能耗必须均大于 0。

3.2 问题求解

对于本文考虑的 eMBB 与 URLLC 用户共存的场景, 由于 URLLC 业务具有低时延的要求, 且具有突发性和零散性的特点, 一般不会出现大量 URLLC 用户同时请求计算的情况。此外, URLLC 业务传输是以短间隙和迷你间隙为主, 其数据包长度和业务数据量远小于 eMBB 业务, 故可以此为依据判断 URLLC 用户的请求, 并对 URLLC 用户优先进行处理, 之后再考虑 eMBB 用户的资源分配。主要流程如下:

Step1: 对 eMBB 用户对于边缘服务器的偏好进行排行, 首先根据式 (19) 中的约束条件筛选出符合条件的服务器, 在不考虑障碍物等因素的情况下, 用户与候选服务器之间距离越小, 信道状态越好, 故根据候选服务器的位置信息将其按照升序进行排列。

Step2: 假设基站为拍卖人, eMBB 用户为买方, 需要向基站提供业务需求、估值和偏好排行, 边缘服务器为卖方, 向基站提供各类资源总量和要价。

Step3: 基站根据收集的信息对边缘服务器进行偏好度降序排序, 先对用户偏好度高的服务器的用户进行资源分配, 按照用户估值的排行情, 根据上述最优化问题中建立的目标函数, 优先对估值较高的用户进行资源分配。

Step4: 若满足拍卖条件, 则该用户将任务卸载到对应的边缘服务器上计算; 若不满足则使得用户向下一个偏好服务器发送请求。若该用户的所有偏好服务器均不能满足拍卖条件, 则本轮拍卖结束。

Step5: 在下一轮拍卖中, 用户和边缘服务器分别根据式 (14) 和式 (15) 修改估值和要价, 返回步骤 (1), 再次进行拍卖。

Step6: 若所有用户任务均已分配到相应的服务器中进行计算, 或边缘服务器中剩余资源量过少, 无法完成计算任务, 则拍卖结束。

算法流程如图 3 所示。

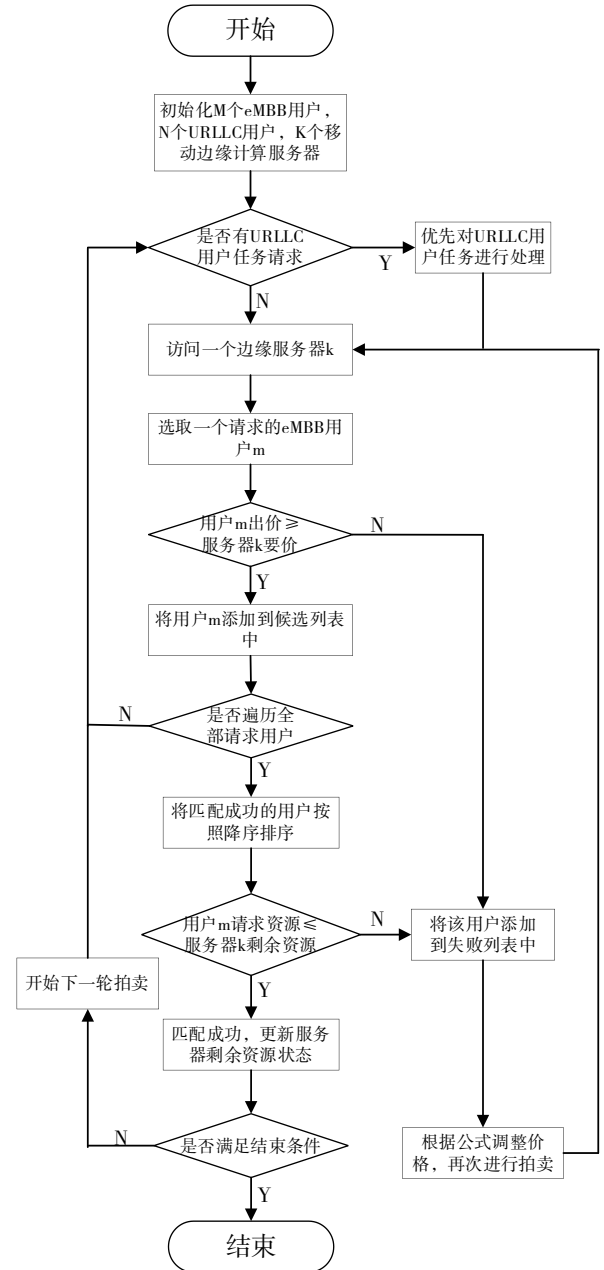


Fig. 3 Flow chart of resource allocation algorithm based on multi-round bidirectional auction

图3 基于多回合双向拍卖的资源分配算法流程

4 仿真结果与分析

本文利用仿真工具 MATLAB 对多回合双向拍卖算法进行仿真, 并对得出的数据加以分析。

4.1 仿真参数设置

假设本文场景为一个的半径 1 km 的圆形小区, 基站、eMBB 用户和 URLLC 用户随机分布在该小区内, 边缘服务器资源量、计算能力和用户任务数据量等参数均随机分布在一定区间内, 具体数据如表 1 所示^[24]。

Table 1 Simulation parameter

表 1 仿真参数

参数名称	值域
小区半径/km	1
eMBB 用户数量	[20, 80]
eMBB 用户数据量/MB	[1, 5]
URLLC 用户数量	[5, 10]
URLLC 用户数据量/MB	[0.5, 2.5]
eMBB 最大容忍时延/s	[20, 40]
边缘服务器通信资源/MHz	[1, 20]
边缘服务器计算资源/GB	[80, 400]
边缘服务器存储资源/GHz	[1, 10]
边缘服务器计算能力/GHz	1
移动终端计算能力/MHz	[0.5, 1]
噪声密度/dBm/Hz	-174

4.2 仿真结果分析

将本文提出的多回合双向拍卖算法与现有的卸载策略进行比较, 主要包括全部卸载、随机卸载和块坐标下降 (Block Coordinate Descent, BCD) 算法。其中, 对于最优化问题而言, BCD 方法具有易实现、全局搜索等优点, 是常用的复杂最优化问题求解方法之一, 故可作为常用传统方法与本文方法进行对比。文献[8]中将 eMBB 与 URLLC 用户的任务卸载和调度作为一个混合整数非线性问题, 再将问题分解为多个子问题, 并将其转化为凸形式, 然后使用 BCD 算法交替求解, 直到收敛到期望的解。

此外, 将本文算法与现有拍卖算法进行对比, 文献[24]在多用户多移动边缘服务器的场景下, 综合考虑智能设备性能和服务器资源, 提出了一种基于改进拍卖算法的计算卸载策略。根据带宽、计算任务量和服务器计算能力提出卸载决策, 并根据时间需求提出基于改进拍卖算法的任务调度模型。

不同方法下用户数量对边缘服务器总收益影响比较如图 4 所示。

可以看出, 相比于现有拍卖算法, 本文多回合双向拍卖算法能使服务器获得更高收益, 平均提高了约 9.02%。

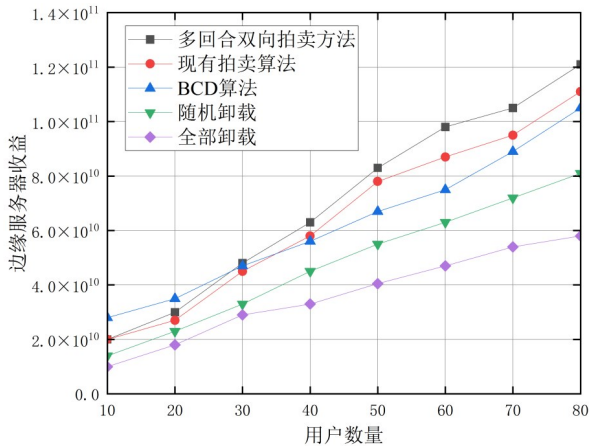


Fig. 4 Comparison of the effect of the number of users on the total revenue of edge servers under different methods

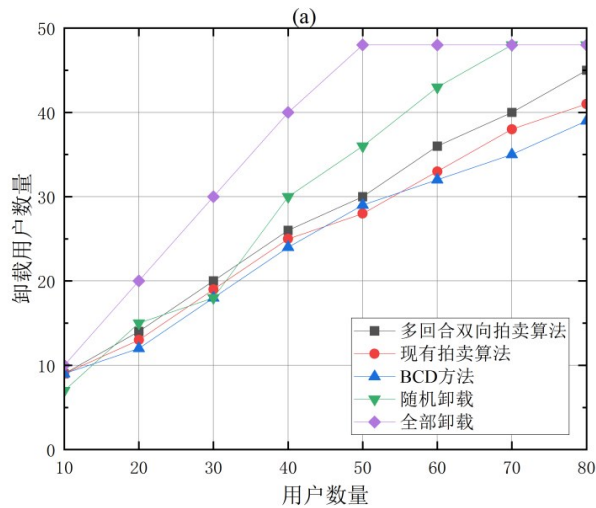
图 4 不同方法下用户数量对边缘服务器总收益影响比较

这是因为文献[24]中方法的主要目标为降低业务时延, 且未考虑不同业务共存场景, 因而缺少对于 URLLC 业务的优先处理机制, 故其总体收益小于本文方法。而与 BCD 算法相比, 当用户数量较少时, 本文方法不如 BCD 算法, 当用户数量为 10 时, 本文方法收益比 BCD 方法低大约 28.57%, 而当用户数量较大时, 本文方法要优于 BCD 方法, 收益比 BCD 方法提高约 12.5%~30.67%。这种现象的主要原因是, 当用户数量较小时, 资源相对充足, 能满足大部分用户的卸载需求, 故无需考虑用户任务的重要性以及能够带给服务器的收益, 而当用户数量增大时, 服务器资源无法满足所有需求, 故要考虑优先对收益大的用户进行卸载计算, 以获得更大收益。并且, 边缘服务器的维护和搭建需要成本, 当用户较少时, 边缘服务器可从用户处获得的收益减小, 但边缘服务器的成本不变, 因而总收益不如 BCD 方法。而随机卸载与全部卸载由于缺少合理的资源分配与卸载策略, 会造成较多的资源和能耗浪费, 故其性能远低于本文方法。综上所述, 在用户数量较大的场景下, 本文方法能够选择收益更大的用户进行卸载计算, 更加高效地进行资源分配, 显著提升边缘服务器总收益。

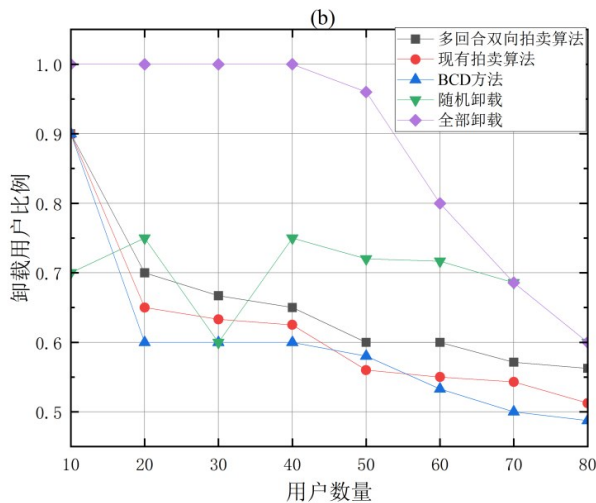
卸载用户数量和比例随用户数量变化曲线如图 5 所示。

从图 5(a)可以看出, 由于边缘服务器资源总量有限, 无法满足所有用户业务需求, 故当用户数量较多时, 全部卸载用户量无法超过最大承受数量。而本文方法中随着用户总数的增加, 选择卸载计算的用户数量也在增加, 即卸载计算成功的用户增多。当用户数量较多时, 本文方法卸载用户数量明显高于现有拍卖算法和 BCD 方法, 以用户数量 80 为例, 本文成功卸载数量相比于后二者提高约 9.76% 和 15.38%。这是因为当用户数量增大时, 边缘服务器资源无法满足所有用户的需求, 而本文方法提出的补偿策略能够动态地调整服务器报价和用户出价, 故相比于其他方法, 本文方法能够更加高效地进行资源分配, 提高卸载成功率。从图 5(b)可以看出, 随着用户数量增加, 用户卸载比例总体呈现下降趋势, 原因是边缘服务器资源有限, 当用户数量增大时, 剩余资源量减小, 无法满足所有用户的计算需求, 故对于一些计算任务量较大的用户, 无法达到拍卖要求, 因而不能进行卸载计算。

图 6 为 eMBB 用户数量与总用户数量不同比例下的边缘服务器收益情况。可以看出, 当用户总数相同时, URLLC 用户占比增大, 总收益也增大, 原因是 URLLC 业务的低时延特性且消耗的资源较少。因此, 在拍卖开始前要优先进行处理, 之后才对 eMBB 用户进行多回合双向拍卖的资源分配, 故当 URLLC 用户比例较大时, 分给 eMBB 用户的资源量减小。而本文仅考虑了对 eMBB 用户使用双向拍卖算法, 故本文方法优势也会缩小, 即总收益减小。但由于 URLLC 用户占比总体较小, 且消耗的资源量相对较少, 故对收益的整体影响不大。例如, 当用户数为 20 时,



(a) The impact of different user numbers on the number of unicast users
(a) 不同用户数量对卸载用户数量的影响



(b) The impact of different user numbers on the proportion of unicast users
(b) 不同用户数量对卸载用户比例的影响

Fig. 5 The impact of different user numbers on the number and proportion of unicast users

图5 不同用户数量对卸载用户数量与比例的影响

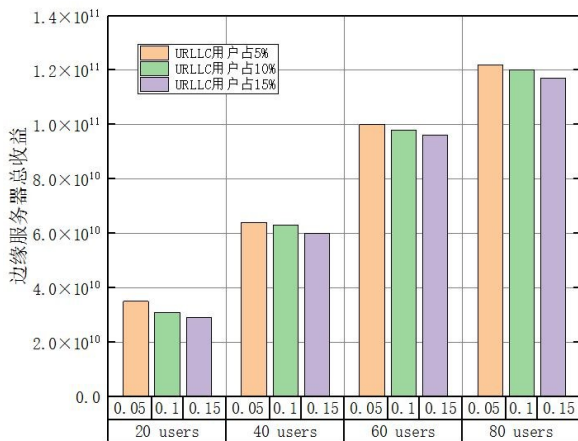


Fig. 6 The impact of different user ratios on revenue

图6 不同用户比例对收益的影响

URLLC用户占比为15%,比占比为5%时收益低17.14%,而当用户数量为80时,占比15%比占比5%收益低大约4.1%,当用户总数越大,其占比影响越小。

5 结语

本文研究了一种基于拍卖算法的eMBB与URLLC业务共存的MEC场景下资源分配算法。首先,考虑了eMBB与URLLC业务共存的MEC场景,建立了一个多用户多边缘服务器的小区模型。其次,根据该小区模型建立了一个由用户、基站和边缘服务器组成的拍卖模型,然后以用户传输速率和计算能耗以及边缘服务器收益加权和为目标函数建立最优化问题,并利用一种多回合双向拍卖的资源分配策略,即随着拍卖轮数的推进,根据业务特征和资源剩余量调整用户出价和边缘服务器报价,以提高卸载计算成功率。根据该策略进行多轮资源分配,直至满足结束条件。最后,仿真结果表明,相比于传统方法,本文所提出的计算卸载和资源分配方法能够使得边缘服务器收益最大化,且用户数量越高该优势越明显。

在后续工作中,还需要从以下几个角度进行研究:

①本文方法目前仅考虑了业务不可分割的情况,在未来研究中可考虑将一些计算量较大的任务分割并卸载到多个边缘服务器上,再结合本文多回合双向拍卖算法进行资源分配,理论上可对资源进行更加充分的利用;②本文对URLLC用户在拍卖开始前单独考虑,不参与拍卖,未来若面对URLLC用户较多的场景,需要将URLLC和eMBB用户一起考虑,共同进行拍卖以完成资源分配。

参考文献:

- [1] ZHANG K, XU X, ZHANG J, et al. Dynamic multiconnectivity based joint scheduling of eMBB and uRLLC in 5G networks [J]. IEEE Systems Journal, 2021, 15(1): 1333–1343.
- [2] ZHANG Y, LI W, SEAH W K G. Admission control with latency considerations for 5G mobile edge computing [C]//Boston: 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2023.
- [3] YUN J, GOH Y, YOO W, et al. 5G multi-RAT URLLC and eMBB dynamic task offloading with MEC resource allocation using distributed deep reinforcement learning [J]. IEEE Internet of Things Journal, 2022, 9(20): 20733–20749.
- [4] ALSENWI M, TRAN N H, BENNIS M, et al. Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: a deep reinforcement learning based approach [J]. IEEE Transactions on Wireless Communications, 2021, 20(7): 4585–4600.
- [5] XIA Y Z, DENG X J, YI L Z, et al. AI-driven and MEC-empowered confident information coverage hole recovery in 6G-enabled IoT [J]. IEEE Transactions on Network Science and Engineering, 2022, 10(3): 1256–1269.
- [6] UMMY H, MAGHSUDI S, HOSSAIN E. A repeated auction model for load-aware dynamic resource allocation in multi-access edge computing [J]. IEEE Transactions on Mobile Computing, 2024, 23(7): 7801–7817.

- [7] MARIE S, SHARMA S, GUO K, et al. Towards effective resource procurement in MEC: a resource re-selling framework[J]. *IEEE Transactions on Services Computing*, 2023, 17(1): 82–97.
- [8] TUN Y K, KIM D H, ALSENWI M, et al. Energy efficient communication and computation resource slicing for eMBB and URLLC coexistence in 5G and beyond[J]. *IEEE Access*, 2020, 8: 136024–136035.
- [9] HAN X, XIAO K, LIU R, et al. Dynamic resource allocation schemes for eMBB and URLLC services in 5G wireless networks[J]. *Intelligent and Converged Networks*, 2022, 3(2): 145–160.
- [10] DUAN Y, SHE C, ZHAO G, et al. Delay analysis and computing offloading of URLLC in mobile edge computing systems[C]//Hangzhou: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), 2018.
- [11] MAO Y, XU X, WANG L, et al. Priority combinatorial double auction based resource allocation in the cloud[C]//Oxford: 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService), 2020.
- [12] DENG L, XU F, REN Y, et al. Resource allocation method based on combinatorial double auction mechanism in cloud computing[C]//Wuhan: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018.
- [13] CHEN X, ZHU G, DING H, et al. End-to-end service auction: a general double auction mechanism for edge computing services[J]. *IEEE/ACM Transactions on Networking*, 2022, 30(6): 2616–2629.
- [14] BUCCI U, CASSIOLI D, MAROTTA A. Performance of spatially diverse URLLC and eMBB traffic in cell free massive MIMO environments[J]. *IEEE Transactions on Network and Service Management*, 2024, 21(1): 161–173.
- [15] CHI X, XU X, HAN S, et al. Learning-based cooperative multiplexing mode selection and resource allocation for eMBB and uRLLC[C]//Austin: 2022 IEEE Wireless Communications and Networking Conference (WCNC), 2022.
- [16] INTERDONATO G, BUZZI S, ANDREA C D, et al. On the coexistence of eMBB and URLLC in multi-cell massive MIMO[J]. *IEEE Open Journal of the Communications Society*, 2023, 4: 1040–1059.
- [17] ALMEKHLAFI M, ARFAOUI M A, ELHATTAB M, et al. Joint scheduling of eMBB and URLLC services in RIS-aided downlink cellular networks[C]//Athens: 2021 International Conference on Computer Communications and Networks (ICCCN), 2021.
- [18] PAN J, MCELHANNON J. Future edge cloud and edge computing for internet of things applications[J]. *IEEE Internet of Things Journal*, 2018, 5(1): 439–449.
- [19] WANG C, LIANG C, YU F R, et al. Computation offloading and resource allocation in wireless cellular networks with mobile edge computing[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(8): 4924–4938.
- [20] HOU Y, WANG C, ZHU M, et al. Joint allocation of wireless resource and computing capability in MEC-enabled vehicular network[J]. *China Communications*, 2021, 18(6): 64–76.
- [21] SHENG J F, TENG X Y, LI W M, et al. Computational offloading strategy based on improved auction model in mobile edge computing[J]. *Application Research of Computers*, 2020, 37(6): 1688–1692.
盛津芳, 滕潇雨, 李伟民, 等. 移动边缘计算中基于改进拍卖模型的计算卸载策略[J]. *计算机应用研究*, 2020, 37(6): 1688–1692.
- [22] SHANG Q Q. Research on resource allocation algorithms for URLLC and eMBB services coexistence scenario[D]. Beijing: Beijing University of Posts and Telecommunications, 2020.
商晴庆. URLLC与eMBB业务共存场景下的资源分配算法研究[D]. 北京: 北京邮电大学, 2020.
- [23] SUN W S, ZHAO Y Y. Research on resource allocation algorithm based on the coexistence of eMBB and URLLC services[J]. *Journal of Hangzhou Dianzi University (Social Sciences)*, 2023, 43(1): 20–25.
孙文胜, 赵莹莹. eMBB和URLLC业务共存场景下的资源分配算法[J]. *杭州电子科技大学学报(自然科学版)*, 2023, 43(1): 20–25.
- [24] YANG Z. Research on joint resource allocation and computation offloading strategy for mobile edge computing network[D]. Chongqing: Chongqing University, 2021.
杨卓. 移动边缘计算网络中基于资源联合配置的计算卸载策略研究[D]. 重庆: 重庆大学, 2021.

(责任编辑: 孙娟)