

支持 MEC 的天地一体化网络下任务卸载和资源分配

李晓青¹ 贺占权¹ 周卫彤²

¹(航天恒星科技有限公司 北京 100095)

²(北京航空航天大学电子信息工程学院 北京 100191)

摘要 针对天地一体化网络(Integrated Satellite-Terrestrial Network, ISTN)高动态、时延敏感的特性,用户服务质量(Quality of Service, QoS)保障成为新的研究重点。为此,在天地一体化网络场景下使用移动边缘计算(Mobile Edge Computing, MEC)技术,在低轨道(LEO)卫星上部署 MEC 服务器,通过排队理论分析场景的时延成本,提出任务卸载和资源分配的联合优化方案。为最小化用户与 LEO 卫星的传输时延,基于凸优化理论给每个用户分配最优的带宽资源。基于整数规划在多项式时间内给出接近全局最优的任务卸载决策。仿真结果表明提出算法与低复杂度的启发式算法相比,减少了 20% 的总时延成本,验证了 ISTN(Integrated Satellite-Terrestrial Network)下通过 MEC 技术保障地面用户 QoS 的可行性和有效性。

关键词 天地一体化网络 移动边缘计算 凸优化理论 整数规划

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2023.02.021

TASK OFFLOADING AND RESOURCE ALLOCATION FOR THE MEC ENABLED INTEGRATED SATELLITE-TERRESTRIAL NETWORK

Li Xiaoqing¹ He Zhanquan¹ Zhou Weitong²

¹(Space Star Technology Co., Ltd., Beijing 100095, China)

²(School of Electronic and Information Engineering, Beihang University, Beijing 100191, China)

Abstract Aimed at the highly dynamic and delay-sensitive characteristics of the integrated satellite-terrestrial network (ISTN), the quality of service (QoS) guarantee of users become the new research focus. Therefore, mobile-edge computing (MEC) is used in the ISTN scenario and MEC server is deployed on the low-orbit (LEO) satellite. By analyzing the delay cost of the scenario through queuing theory, we propose a joint task offloading and resource allocation algorithm. In order to minimize the transmission delay between users and LEO satellite, each user was assigned the optimal bandwidth resource based on the convex optimization theory. Based on integer programming, an approximate global optimal task unloading decision was given in polynomial time. The simulation results show that the proposed algorithm reduces the total delay cost by 20% compared with the low-complexity heuristic algorithm. It also verifies the feasibility and effectiveness of using MEC technology to guarantee the QoS of ground users in ISTN.

Keywords Integrated satellite-terrestrial network Mobile-edge computing Convex optimization theory Integer programming

0 引言

移动通信技术已经发展到第五代(5G),5G 网络

的主要工作频段为 3 000 ~ 5 000 MHz,远远高于 4G 网络,这导致 5G 信号在信道中传播的衰减要更高。5G 基站的覆盖半径仅为 100 ~ 300 m 左右,这为实现偏远地区用户以及海空中用户的组网覆盖提出了严重挑

战^[1]。近年来以小型化、低成本、低延迟和高吞吐量为特点的 LEO 卫星极大地促进了天地一体化网络 (ISTN) 的发展^[2], ISTN 已成为 5G 时代移动通信系统中一个必不可少的替代方案^[3]。

考虑到 5G 时代偏远地区及海、空用户节点的组网需求,无法直接通过布置地面基站来完成全域覆盖,这使得 ISTN 朝着卫星可直连用户的方向发展;此外,用户设备 (User Equipment, UE) 侧流量急剧增长,而由于自身资源有限,UE 在处理计算密集型和时间敏感型业务时能力不足的情况。为此,移动边缘计算 (MEC) 技术应运而生,相比于传统移动云计算 (Mobile Cloud Computing, MCC) 技术,MEC 技术可以在更靠近 UE 侧配置,显著减少云端处理任务的时延。相比于传统蜂窝网络架构,ISTN 架构的传播时延显著增加,天地链路变化频繁。随着服务的增多导致大量数据通过链路进行传输,大大增加了链路负载,给 ISTN 的网络架构设计带来严重挑战。使用 MEC 技术^[4],将计算资源从云端分配至边缘服务器,能够极大地改善对用户的 QoS,并且可以有效减少整个 ISTN 的流量。

在基于 MEC 的网络架构中,资源优化分配策略是研究热点。文献[5]考虑将多个独立用户的任务卸载到一个基站的蜂窝网络场景下,抽象为以时延和能量消耗最小化为目标的混合整数非线性优化 (Mixed-Integer Nonlinear Programming, MINLP) 问题,并设计了次优的算法来获得最优资源分配方案。文献[6]将拥有多个基站蜂窝网络划分为不同子区域,通过将多个独立计算任务卸载到 MEC 服务器或云端,对基站服务缓存和任务卸载进行联合决策以优化时延和能量消耗,基于 Lyapunov 优化和 Gibbs 采样对优化问题进行求解,提出一种次优的服务存储策略和任务卸载方案。文献[7]探讨 ISTN 场景下应用 MEC 技术来改善用户 QoS 的可行性,提出在近 UE 侧地面站和远端核心网地面网关布置 MEC 服务器,并提出一种协作计算卸载 (Cooperative Computation Offloading, CCO) 模型来实现天地一体化网络中多 MEC 服务器并行计算。文献[8]提出了一种在 LEO 卫星与近 UE 侧地面站布置 MEC 服务器的双边缘天地一体化网络,并提出基于双边缘 ISTN 中的协作分流方案,通过分析得出分流效率和能耗的性能。

上述文献普遍基于在拥有存储和计算能力的固定基站中配置 MEC 服务器,为 UE 缓存热点服务,并将用户的计算任务卸载至 MEC 服务器,将场景抽象成以服务缓存和任务卸载为决策,以最小化时延和能量消耗为目标的优化问题,并提出最优或次优的服务存储策略和任务分流方案。

本文考虑一种在 LEO 卫星上配置 MEC 服务器,直接与地面用户建立天地链路的 ISTN 架构,为地面基站无法覆盖到的区域提供 MEC 服务,大幅度提高偏远地区用户的 QoS。基于串联排队理论得出 ISTN 场景的总时延成本,通过拉格朗日对偶理论和梯度下降法给出了 ISTN 网络的总传输时延成本的极小值,并基于模拟退火算法,提出任务卸载和资源分配联合调度方案,在多项式时间内给出近似全局最优的 ISTN 的时延成本。仿真结果表明本文算法相比于低复杂度启发式算法减少了 20% 的总时延成本,验证了 ISTN 下通过 MEC 技术保障地面用户 QoS 的可行性和有效性。

1 天地一体化网络模型

典型的 ISTN 包含地面网络与空间网络,具有规模庞大、支持的业务种类多、拓扑结构呈现立体多层次化和高动态变化的特点。

1.1 网络模型

本文考虑如图 1 所示的 LEO 卫星-UE 双层天地一体化网络模型,将 MEC 服务器部署在 LEO 卫星上,为 UE 提供任务卸载服务。每个时隙内由 MEC 服务器决策是否为 UE 提供 MEC 服务以及 LEO 卫星为 UE 分配的带宽资源。

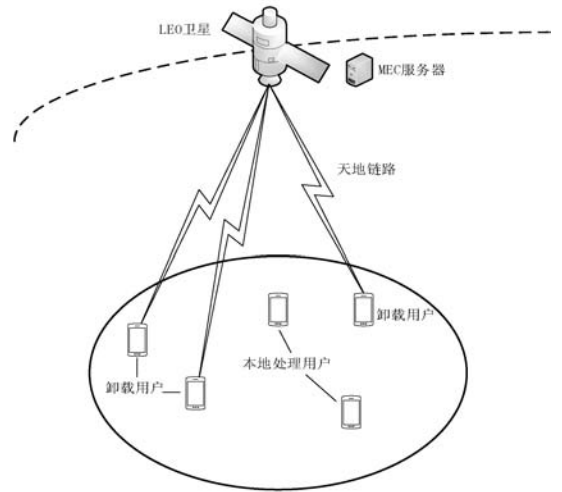


图 1 ISTN 架构

定义 LEO 卫星提供的通信能力为卫星提供的链路总带宽 W (单位为 Hz); 提供的计算能力为 MEC 服务器中 CPU 的主频 F (单位为 Hz), MEC 服务器的 CPU 可以处理用户卸载至 LEO 卫星的任务。

对于 UE 侧,定义 UE 集合为 $U = \{u_1, u_2, \dots, u_i\}$, UE u_i 配置主频为 f_i 的 CPU 以本地处理计算任务; u_i 产生的计算任务请求的 CPU 指令周期个数服从均值为 c_i (单位为 CPU cycles) 的负指数分布; 计算任务的平均

大小为 d_i (单位为 bits)。在时隙内 u_i 上计算任务的产生是一个速率为 λ_i 的泊松过程,在实际应用中,可以通过基于自回归模型等成熟的需求预测模型来估计时隙开始时的瞬时需求^[10],使 MEC 服务器可以根据 UE 端设备的需求,为 UE 动态提供 MEC 服务。

传统的 ISTN 网络架构下进行任务卸载时,卫星通过用户链路、馈电链路为 UE 提供与远端核心网的链接,时延较长。本文设计的 ISTN 网络架构下,星上 MEC 服务器直接通过用户链路为用户提供 MEC 服务,无须通过馈电链路。在此架构上初步探究 ISTN 网络针对偏远地区用户及海空中用户的计算密集型任务的服务能力。

1.2 任务卸载和资源分配模型

假设每个 UE 的等效全向辐射功率 (Equivalent Isotropically Radiated Power, EIRP) 相同,LEO 卫星采用正交频分复用 (Frequency Division Multiple Access, FDMA) 的接入方式,当不同用户共享频谱资源时不存在相互干扰。

MEC 服务器进行任务卸载的决策有本地执行、完全卸载和部分卸载三种方案。本地执行即整个计算任务在 UE 本地完成;完全卸载即整个计算任务由 MEC 卸载和处理;为了简化分析,我们不考虑部分卸载的情况。则可定义一个二元决策变量 $a_i = \{0, 1\}$ 表示 LEO 卫星上的 MEC 服务器是否为 UE u_i 提供 MEC 服务。其中 $a_i = 1$ 表示 MEC 服务器为 u_i 提供 MEC 服务; $a_i = 0$ 表示 u_i 在本地处理计算任务,则时隙内任务卸载决策是 $A = \{a_1, a_2, \dots, a_i\}$ 。

整个 LEO 卫星可提供的总带宽为 W , 单位为 bit/s, 定义时隙内 LEO 卫星带宽资源分配决策 $B = \{b_1, b_2, \dots, b_i\}$, 其中 $b_i = [0, 1]$, 表示 MEC 服务器为 u_i 分配的带宽占总带宽比例。因为本地计算的 UE 不占用频带资源,若 $a_i = 0$, 则 $b_i = 0$, 则用户链路通信速率 R_i 有:

$$R_i = b_i W \quad (1)$$

由于采用 FDMA 多址接入技术,所有 UE 分配的频谱资源不超过卫星的总通量,则 ISTN 的频带资源约束表示为:

$$\sum_i b_i = 1 \quad (2)$$

1.3 时延模型

计算卸载的性能通常以时间延迟和能量消耗作为衡量指标。在 ISTN 场景下,我们主要关注 MEC 服务器在减小业务时延上的能力。ISTN 下 UE 产生的计算任务可以在本地执行或完全卸载,在本地执行时,时延

成本是指在 UE 处执行本地计算所花费的时间,在完全卸载时,时延包括传输时延、传播时延、排队时延和计算时延。

本地执行时,计算任务在本地执行过程可建模为一个 M/M/1/FCFS 型队列。队列服务时间即本地计算时延服从均值为 c_i/f_i 的指数分布,我们假设本地计算队列稳定,即 $f_i > \lambda_i c_i$, 则本地执行计算任务的时延成本 T_i^l 等于队列的总停留时间期望,即:

$$T_i^l = \frac{\lambda_i c_i}{f_i - \lambda_i c_i} \quad (3)$$

为了分析 ISTN 网络下计算任务完全卸载的时延成本,根据分组交换网络思想,可将整个网络建模成如图 2 所示的串联排队系统。

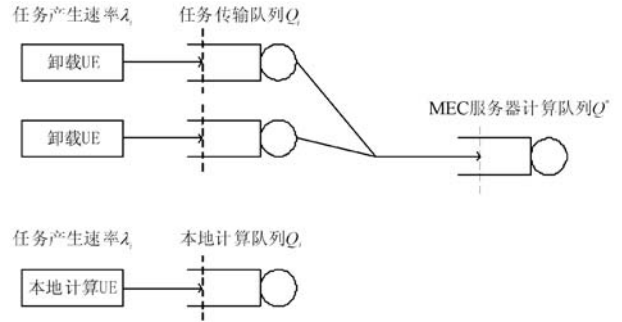


图2 串联排队系统

图2中队列集 $Q = \{Q_i | \forall a_i \neq 0\}$ 是独立的 M/D/1/FCFS 队列, Q_i 表示卸载 UE 产生计算任务并上传 LEO 卫星的过程, Q 是一个 M/G/1/FCFS 队列, 表示 MEC 服务器处理计算任务的过程。队列 Q_i 的服务时间 t_i 为用户 u_i 产生的计算任务的传输时延:

$$t_i = \frac{d_i}{R_i} \quad \forall b_i \neq 0 \quad (4)$$

根据排队论的 Pollaczek-Khinchin 公式和 Little 定理任务有队列 Q_i 的平均停留时间期望 T_i 为:

$$T_i = t_i + \frac{\lambda_i t_i^2}{2(1 - \lambda_i t_i)} \quad \forall b_i \neq 0 \quad (5)$$

此时队列遵循稳态约束条件:

$$\lambda_i t_i = \frac{\lambda_i d_i}{b_i W} \leq 1 \iff b_i \geq \frac{\lambda_i d_i}{W} \quad \forall b_i \neq 0 \quad (6)$$

由于 u_i 将计算任务上传至 MEC 服务器上缓存以待处理的过程是一个速率为 $\tilde{\lambda}_i = a_i \lambda_i$ 的泊松过程,所有任务在 MEC 服务器上缓存以待执行的过程是一个速率为 $\tilde{\lambda} = \sum_i \tilde{\lambda}_i$ 的泊松过程。对于队列 Q^* , MEC 服务器对计算任务的处理时间为 c_i/F 的概率为 $\tilde{\lambda}_i/\tilde{\lambda}$, 假设 s 为队列中服务时间随机变量, s 的一阶原点矩和二阶原点矩分别为:

$$E(s) = \frac{1}{F} \sum_i c_i \tilde{\lambda}_i / \tilde{\lambda} \quad (7)$$

$$E(s^2) = \frac{1}{F^2} \sum_i c_i^2 \tilde{\lambda}_i / \tilde{\lambda} \quad (8)$$

根据 Pollaczek-Khinchin 公式和 Little 定理有队列 Q^* 的平均停留时间期望 T_s 为:

$$T_s(A) = E(s) + \frac{\tilde{\lambda} E(s^2)}{2[1 - \tilde{\lambda} E(s)]} \quad (9)$$

其中队列遵循稳态约束条件:

$$\tilde{\lambda} E(s) = \frac{1}{F} \sum_i c_i \tilde{\lambda}_i \leq 1 \stackrel{\Delta}{=} \sum_i a_i c_i \lambda_i \leq F \quad (10)$$

实际上, Pollaczek-Khinchin 公式的稳态约束条件可理解为计算资源约束。由此,我们定义计算任务卸载至 MEC 服务器处理的时延成本为:

$$T_i^M(A, B) = T_{s_i}(b_i) + T_s(A) + 2t_c \quad \forall b_i \neq 0 \quad (11)$$

式中: $T_{s_i}(b_i)$ 是 u_i 与 LEO 卫星的链路传输时延; $T_s(A)$ 是计算在 LEO 卫星上的平均停留时间; t_c 表示 UE 与 LEO 卫星之间的链路传播时延, 由于我们忽略不同 UE 与 LEO 的距离变化, 因此对于每个 UE, t_c 是常数。

为了分析 ISTN 下用户 QoS 保障的问题, 定义 ISTN 的总时延成本为:

$$T(A, B) = \sum_i (1 - a_i) T_i^L + a_i T_i^M(A, B) \quad (12)$$

其表示 ISTN 网络场景下 MEC 服务器减小 UE 时延成本的能力。

1.4 优化问题模型

优化问题的目标是制定任务卸载决策 A 和带宽资源分配决策 B , 以最大限度地减少总时延成本 $T(A, B)$ 。优化问题 P1 的模型为:

$$\min_{A, B} T(A, B) \quad (13)$$

$$\text{s. t. } b_i > \frac{\lambda_i d_i}{\Phi}, \forall b_i \neq 0 \quad (14)$$

$$\sum_i b_i \leq 1 \quad (15)$$

$$\sum_i a_i c_i \lambda_i < F \quad (16)$$

式(14)是计算任务上传队列稳定约束条件, 代表 LEO 卫星在每个时隙内为 UE 分配的带宽应使计算任务上传队列稳定; 式(15)是 ISTN 的带宽资源约束条件; 式(16)是 MEC 服务器的 M/G/1 队列稳态约束, 也可将其理解为 MEC 服务器的计算资源上限。

可以很容易地注意到, P1 是一个 MINLP 问题, 且优化函数非凸, 这使得 P1 非常难以求解; 另一方面, 对于整数规划问题, 通过穷举法得出最优的任务卸载决策 A 的时间复杂度为 $O(2^N)$, 这是一个指数级复杂度, 并且对于每个卸载决策, 都需要确定最佳的带宽分配。由于算法的复杂性较高, 在实际应用中是不可行的。

2 ISTN 场景下任务卸载和资源配方案

本文基于模拟退火算法和梯度下降法提出一种天地一体化网络任务卸载和资源分配 (Satellite-Terrestrial Task Offloading and Resource Allocation, ST-TORA) 方案。对于如何求解 P1, 首先考虑将 A 和 B 解耦, 选定部分 UE 进行任务卸载, 确定卸载决策向量 A , 将原问题松弛并根据拉格朗日法求解, 得到带宽配置方案 B 和天地一体化网络总时延开销 $T(A, B)$ 。然后基于模拟退火算法迭代更新决策变量 A 直到找到 P1 的局部最优解, 将算法的时间复杂度降为多项式时间。

2.1 资源分配 (RA) 方案

当 MEC 服务器为多个 UE 提供 MEC 服务时, 如何有效地将 LEO 卫星有限的带宽资源分配给卸载 UE 是本节要解决的问题。取 $k_i = \lambda_i d_i / \Phi$, $K_i = d_i / 2\Phi$, 队列 Q_i 的平均服务时间为:

$$T_{s_i} = t_i + \frac{\lambda_i t_i^2}{2(1 - \lambda_i t_i)} = K_i \cdot \left(\frac{1}{b_i} + \frac{1}{b_i - k_i} \right), \forall b_i \neq 0 \quad (17)$$

式中: k_i 代表了传输队列 Q_i 稳定所需要分配最小带宽资源比例, 当 $b_i < k_i$ 时队列 Q_i 不稳定; 此时问题 P1 可松弛为问题 P2:

$$\min_B \sum_i T_{s_i}, \forall b_i \neq 0 \quad (18)$$

$$\text{s. t. } b_i > k_i, \forall b_i \neq 0 \quad (19)$$

$$\sum_i b_i \leq 1, \forall b_i \neq 0 \quad (20)$$

定理 问题 P2 在满足约束条件下是一个凸优化问题。

证明 式(19)的 Hessian 矩阵为:

$$\mathbf{H}_T = \text{diag} \left(\frac{\partial^2 T_{s_1}}{\partial b_1^2}, \frac{\partial^2 T_{s_2}}{\partial b_2^2}, \frac{\partial^2 T_{s_2}}{\partial b_2^2}, \frac{\partial^2 T_{s_3}}{\partial b_3^2}, \dots, \frac{\partial^2 T_{s_i}}{\partial b_i^2} \right) \quad (21)$$

其中:

$$\frac{\partial^2 T_{s_i}}{\partial b_i^2} = 2K_i \cdot \left[\frac{1}{b_i^3} + \frac{1}{(b_i - k_i)^3} \right] \quad (22)$$

由于约束条件 2 有 $\frac{\partial^2 T_{s_i}(b_i)}{\partial b_i^2} > 0$, \mathbf{H}_T 是半正定矩

阵, 目标函数是凸函数; 由于式(19)、式(20)是凸函数, 因此问题 P2 是一个凸优化问题, 其局部最优解即为全局最优解, 可采用拉格朗日对偶法和梯度下降法求解该问题。我们引入非负拉格朗日乘子 ε_i 和 μ 构建拉格朗日函数:

$$L(B, \varepsilon, \mu) = \sum_i 2K_i \cdot \left[\frac{1}{b_i^3} + \frac{1}{(b_i - k_i)^3} \right]^+ + \sum_i \varepsilon_i (b_i - k_i) + \mu \left(\sum_i b_i - 1 \right)^2 \quad (23)$$

根据次梯度法得到拉格朗日乘子的迭代公式为:

$$\varepsilon_i(n+1) = \left[\varepsilon_i(n) - m \frac{\partial L}{\partial \varepsilon_i} \right]^+ = \left[\varepsilon_i(n) - m(b_i - k_i) \right]^+ \quad (24)$$

$$\mu(n+1) = \left[\mu(n) - m \frac{\partial L}{\partial \mu} \right]^+ = \left[\mu(n) - m \left(\sum_i b_i - 1 \right)^2 \right]^+ \quad (25)$$

式中: $[x]^+ = \max\{0, x\}$; n 表示迭代次数; m 表示迭代步长, 应取足够小的正数。令:

$$\frac{\partial L}{\partial b_i} = -K_i \cdot \left[\frac{1}{b_i^2} + \frac{1}{(b_i - k_i)^2} \right]^+ + \varepsilon_i + 2\mu \left(\sum_i b_i - 1 \right)^2 + \mu \quad (26)$$

通过梯度下降法求解带宽分配的迭代公式为:

$$b_i(n+1) = \left[b_i(n) - m \frac{\partial f_i}{\partial b_i} \right]^+ \quad (27)$$

当迭代步长小于最小迭代步长 δ 时迭代停止。因此当确定任务卸载决策 A 时, 资源分配 (Resource Allocation, RA) 算法描述如算法 1 所示。

算法 1 RA 算法

输入: $A, k_i, K_i, I, W, F, t_c$ 。

输出: B, T 。

初始化 $m, n, \delta, \varepsilon_i(0), \mu(0), b_i(0)$;

while $n=0 \parallel \max_{\forall b_i \neq 0} |b_i(k+1) - b_i(k)| \geq \delta$

$$\varepsilon_i(n+1) = \left[\varepsilon_i(n) - m \frac{\partial L}{\partial \varepsilon_i} \right]^+, \forall b_i \neq 0;$$

$$\mu(n+1) = \left[\mu(n) - m \frac{\partial L}{\partial \mu} \right]^+, \forall b_i \neq 0;$$

$$b_i(n+1) = \left[b_i(n) - m \frac{\partial f_i}{\partial b_i} \right]^+, \forall b_i \neq 0;$$

$n = n + 1$;

end while

计算 $T(A, B)$;

return B, T

在任务卸载场景下, 不同的 UE 业务有着不同的计算任务大小, RA 算法可以为计算任务更大的 UE 分配更多的带宽资源, 最小化网络场景下的传输时延成本。

2.2 任务卸载决策

任务卸载决策 A 的求解是一个整数规划问题, 由穷举法寻找 A 的最优解的时间复杂度为 $O(2^N)$, 为指数级复杂度。可以预见, MEC 服务器应优先为业务数据量更大、请求计算资源更多、计算资源更少的 UE 提

供 MEC 服务, 这意味着最优解附近的解也相对较优, 适用于整数规划中的模拟退火算法。本文基于模拟退火算法, 在多项式时间内寻找问题 P1 的近似全局最优值。

首先确定一个满足约束条件的初始任务卸载决策 A , 并通过 RA 算法得到 B 和 T 。之后对任务卸载决策向量 A 进行 N^* 次如算法 2 所示的随机扰动得到 A_{new} , 通过 RA 算法得到 B_{new} 和 T_{new} 。其中 N^* 为邻域解空间大小。

算法 2 disturb 算法

输入: A, I, W, F 。

输出: A_{new} 。

随机选取扰动 UE i ;

$chosen = rand$;

if $chosen \leq 0.6$

$a_i = 1$;

else if $chosen \leq 0.85 \parallel A$ 中所有元素相等

随机选取与 UE i 卸载决策不同的 UE j ;

$a_i = 1 - a_i, a_j = 1 - a_j$;

else

$a_i = 0$;

return A_{new}

令 $\Delta = T_{\text{new}} - T$ 。若 $\Delta < 0$ 则接受该结果; 否则将有概率接受这个结果, 概率计算公式为 $P = \exp(\Delta/\tau)$, 其中 τ 为当前温度, 初始值设置为用户总数量。更新 $\tau_{\text{new}} = \alpha\tau$, 其中 α 为温度下降率。然后重新开始迭代, 直到温度 τ 到达温度下界 τ_{min} , 其中 τ_{min} 设为一个足够小的正数。基于此, 本文提出 ST-TORA 算法如算法 3 所示。

算法 3 ST-TORA 算法

输入: $c_i, d_i, \lambda_i, f_i, I, W, F, t_c$ 。

输出: A, B, T 。

初始化 A, k_i, K_i, T_i^t ;

$[B, T] = RA(A, k_i, K_i, I, W, F, t_c)$;

While $\tau > \tau_{\text{min}}$

for $i = 1 : N^*$;

$A_{\text{new}} = disturb(A, W, F)$;

$[B_{\text{new}}, T_{\text{new}}] = RA(A_{\text{new}}, k_i, K_i, I, W, F, t_c)$;

$\Delta = T_{\text{new}} - T, P = \exp(\Delta/\tau)$;

if $\Delta < 0 \parallel P > rand$

$A = A_{\text{new}}$;

$B = B_{\text{new}}$;

$T = T_{\text{new}}$;

end if

end for

$\tau = \alpha\tau$;

end while

return A, B, T

ST-TORA 算法的时间复杂度取决于邻域解空间个数大小以及温度下降率,对于 N 位向量 A ,取 $N^* = \tau = N$,ST-TORA 算法的时间复杂度为 $O(N \log N)$ ^[12]。相比于穷举法,ST-TORA 算法可在多项式时间内趋于全局最优。

3 仿 真

在本节中,通过仿真对本文提出的 ST-TORA 算法进行验证,仿真结果表明了应用于天地一体化网络场景,在 LEO 卫星上部署 MEC 服务器,为 UE 提供边缘计算服务,以保障网络 QoS 的可行性及有效性。

3.1 仿真场景及参数设置

考虑由单个 LEO 卫星和 UE 构成的天地一体化网络场景,通过在 LEO 卫星部署 MEC 服务器为 UE 提供边缘计算服务。其中每个 UE 的计算任务数据平均大小 d_i ,所需 CPU 转数 c_i ,UE 端 CPU 计算能力 f_i 以及任务产生速率 λ_i 均随机生成,其中任务平均大小 d_i 与所需 CPU 转数 c_i 呈线性相关。其中仿真平台配置为: Intel(R) Core(TM) i5-4210H CPU @ 2.90 GHz; 8 GB RAM; 硬盘 1 TB; Windows 10 Education 64 位。在 MATLAB 2020a 环境下进行仿真,仿真参数如表 1 所示。

表 1 ISTN 场景仿真参数

仿真参数	设定值
LEO 卫星轨道高度/km	300
计算任务数据大小 d_i /MB	[5,15]
计算任务所需 CPU 转数 c_i /G	[0.05,0.2]
UE 计算能力 f_i /GHz	[1.5,3.3]
UE 任务产生速率 λ_i	[1,5]
迭代步长 m	0.000 1
最小迭代步长 δ	0.000 1
温度下降率 α	0.96

3.2 仿真结果及分析

3.2.1 资源分配(RA)算法

图 3 所示为相同任务卸载决策下,本文基于梯度下降法设计的 RA 方案与带宽资源平均分配(Equal Bandwidth,EB)方案下,LEO 卫星提供边缘计算的时延成本 T 与卸载用户数量的关系,其中 LEO 卫星总吞吐量 W 为 1 Gbit/s,MEC 服务器的 CPU 主频 F 为 10 GHz。

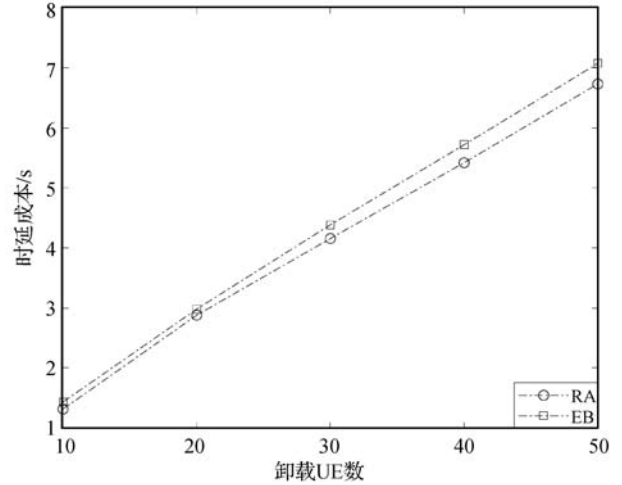


图 3 卸载时延与卸载 UE 数量关系

可以看出,随着卸载 UE 数量的增多,RA 算法按业务需求量分配带宽的优势更加明显。当任务卸载决策确定时,计算任务在 MEC 服务器的总停留时间的期望 $T_s(A)$ 确定,RA 算法可动态地为业务需求高的 UE 分配更高的带宽资源,避免带宽资源浪费并最小化总时延成本。

图 4 给出了 LEO 卫星总吞吐量 W 为 1 Gbit/s,MEC 服务器的 CPU 主频 F 为 10 GHz 的场景下,RA 算法的平均带宽利用率与卸载 UE 数量的关系。平均带宽利用率定义为五十次仿真下带宽分配向量 B 中各元素之和的平均值。

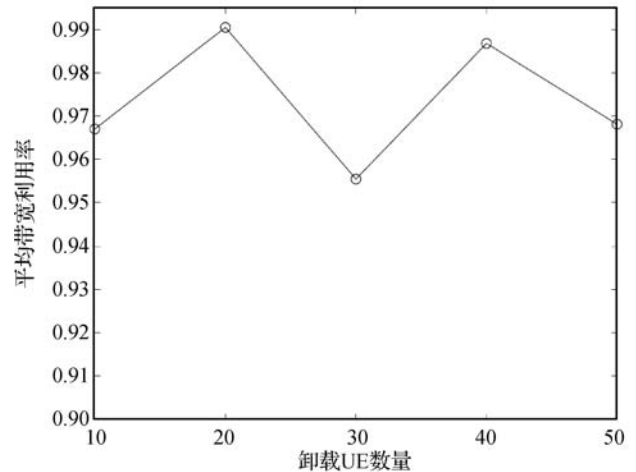


图 4 平均带宽利用率与卸载 UE 数量关系

RA 算法的性能由每次迭代的步长 m 以及拉格朗日乘子 ε_i 和 μ 的初值决定,若步长 m 过大则可能越过极值点, m 过小则算法收敛时间过慢;而拉格朗日乘子会随着迭代次数的增加而逐渐收敛。由图 4 可以看出,在带宽利用率保持在 0.95 以上的同时,RA 算法可以很好地逼近目标函数的极小值。

3.2.2 ST-TORA 算法

图 5 给出 ST-TORA 算法下 3 种总通量的 LEO 卫星进行任务卸载的时延成本与 UE 总数的关系。

MEC 服务器的 CPU 主频 F 设置为 10 GHz。

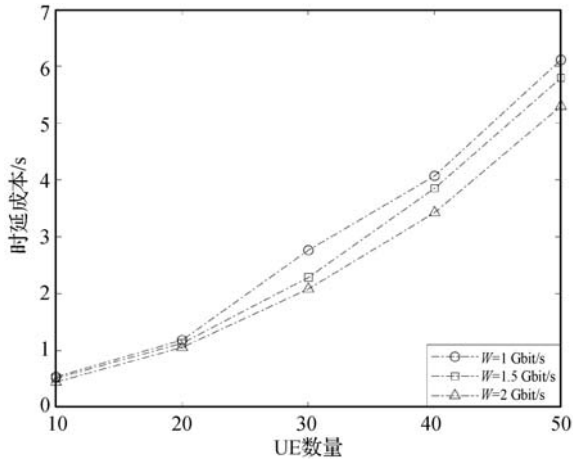


图5 不同总容量下时延成本与 UE 数量关系

可以看出天地一体化网络场景下时延成本与卫星的总容量直接相关,这是因为卫星总容量的增长直接减少计算任务的传输时延,大大保障 UE 端的 QoS。随着 UE 数量的增长,每 0.5 Gbit/s 的容量增长可为 UE 侧减少 100 ms 的传输时延。随着卫星通信技术的不断进步,超低轨道高通量卫星提供的总容量可高达 100 Gbit/s 级,传输时延可缩短至 μ s 级;而相同容量下,LEO 卫星可提供的覆盖性远高于地面基站。

图 6 给出了 ST-TORA 算法下 3 种计算能力的 MEC 服务器进行任务卸载的 UE 卸载比例与 UE 总数的关系。LEO 卫星提供的总容量 W 设置为 1 Gbit/s。

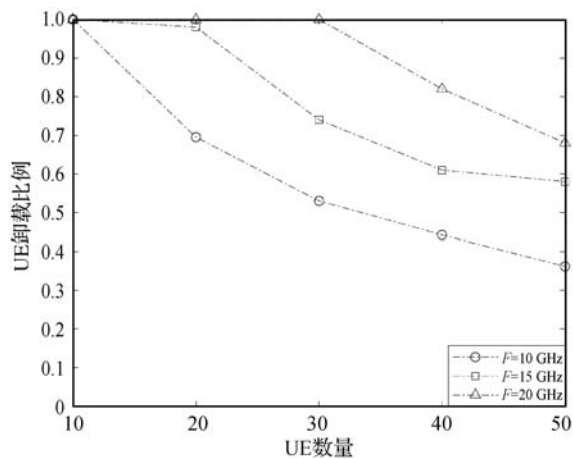


图6 不同计算能力下时延成本与 UE 数量关系

可以看出,卸载 UE 比例是 MEC 计算能力的一个递增函数。由于计算资源约束条件的存在,MEC 服务器的 CPU 主频越高,计算任务在 MEC 服务器的总停留时间越短,MEC 服务器可为更多的 UE 提供服务。当 UE 总数量较少时,天地一体化网络中 LEO 卫星可以提供足够的计算资源和带宽资源时,所有 UE 可以将计算任务上传至 MEC 服务器;随着 UE 数量的增多,由于 MEC 服务器的计算资源和 LEO 卫星提供的带宽资源有限,相同 UE 总数量下卸载 UE 个数会趋于

一个稳定值。

图 7 引入了一种基于 Add 和 Exchange 操作的低复杂度启发式(Heuristic)算法^[13]、本地成本优先式算法与本文设计的 ST-TORA 算法进行对比。本地成本优先式算法将每个 UE 的本地处理计算任务的时延成本 T_i^l 进行排序,然后在满足队列稳定约束和计算资源约束条件下,尽可能多地为 UE 提供 MEC 服务。MEC 服务器的主频 F 设置为 10 GHz,LEO 卫星提供的总容量 W 设置为 1 Gbit/s,不同算法均使用 RA 算法进行资源配置。表 2 给出了相同参数 UE 总数量为 50 的场景下,不同算法的收敛时间。

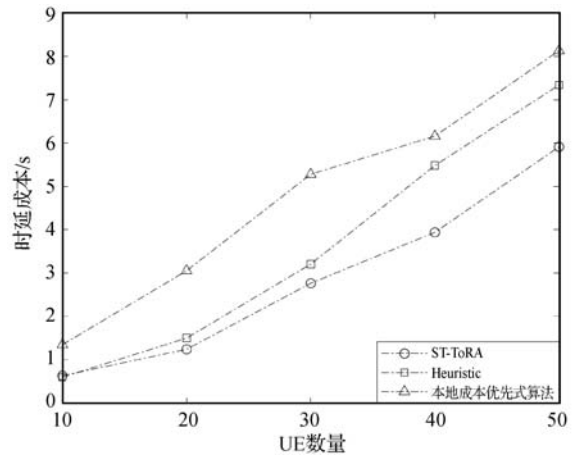


图7 不同算法下时延成本与 UE 数量关系

表 2 不同算法收敛时间对比

单位:s

算法	收敛时间
ST-TORA	397.05 ± 71.95
Heuristic	14.25 ± 0.36
本地成本优先式算法	2.54 ± 0.85

由图 7 可以看出,由于为过多的 UE 提供了 MEC 计算服务,LEO 卫星为每个用户分配的带宽资源只能勉强维持传输队列稳定,因此本地成本优先式算法的性能最差。随着 UE 数量的增多,相比于低复杂度的 Heuristic 算法的时延成本,本文所提 ST-TORA 算法的时延成本可降低约 20%。两种算法同样基于启发式算法,但基于模拟退火思想的 ST-TORA 算法可以跳出局部最优解达到全局最优解。但由表 2 可以看出 ST-TORA 算法的收敛时间要高于 Heuristic 算法,这是因为模拟退火算法的时间复杂度受温度下降率以及温度下界的直接影响,在本文仿真场景参数下,为了保证算法在减小时延成本方面的优越性,ST-TORA 算法的时间复杂度高于 Heuristic 算法。

4 结 语

本文研究支持 MEC 的天地一体化网络场景下任

务卸载和资源分配联合优化方案。首先建立了 LEO 卫星(边缘网络)-UE 的双层 ISTN 模型;然后通过排队论与分组交换网络的思想对模型的时延成本进行考察,并最终将降低时延成本的过程抽象为一个 MINLP 问题,本文将其转为 2 个子问题进行求解:(1) 为最小化计算任务卸载至部署了 MEC 服务器的 LEO 卫星的传输时延成本,基于拉格朗日乘数法和梯度下降法,为每个 UE 分配最优的带宽资源。(2) 基于模拟退火算法,降低整数优化的时间复杂度,寻找到逼近最优的联合任务卸载和资源分配方案。仿真结果表明本文算法能够更好地满足天地一体化网络下的需求,有效降低其时延成本。下一步的工作考虑引入核心网-LEO-UE 三层模型的缓存网络场景^[14],进一步优化天地一体化网络下 UE 的 QoS。

参 考 文 献

- [1] Azzarelli T. Onweb Global Access//Satellite Symposium, 2016.
- [2] Gregory M, Heine F, Kämpfner H. et al. TESAT laser communication terminal performance results on 5.6 Gbit coherent inter satellite and satellite to ground links[C]//International Conference on Space Optics, 2017.
- [3] Cho J P, Joo C. Challenges for efficient and seamless space-terrestrial heterogeneous networks[J]. IEEE Communications Magazine, 2015, 53(5): 156 - 162.
- [4] Hu Y C, Patel M, Sabella D, et al. Mobile edge computing: A key technology towards 5G[R]. European Telecommunications Standards Institute, 2015.
- [5] Mao Y, Zhang J, Letaief K B. Joint task offloading scheduling and transmit power allocation for mobile edge computing systems[C]//2017 IEEE Wireless Communications and Networking Conference, 2017: 1 - 6.
- [6] Xu J, Chen L, Zhou P. Joint service caching and task offloading for mobile edge computing in dense networks[C]//IEEE Conference on Computer Communications, 2018: 207 - 215.
- [7] Neely M. Stochastic network optimization with application to communication and queueing systems[J]. Synthesis Lectures on Communication Networks, 2010, 3(1): 1 - 211.
- [8] Series M. Framework and overall objectives of the future development of IMT 2000 and systems beyond IMT 2000[R]. Recommendation ITU-R, M. 1645, 2003.
- [9] 杨增印,吴茜,李贺武,等.天地一体化信息网络域间路由协议 NTD-BGP[J].清华大学学报(自然科学版), 2019, 59(7): 512 - 522.
- [10] Chen T, Wang X, Giannakis G B. Energy and workload management for data centers in renewable-integrated power grid[C]//Signal & Information Processing, 2016.
- [11] Wang K, Yang K, Magurawalage C. Joint energy minimization and resource allocation in C-RAN with mobile cloud[J]. IEEE Transactions on Cloud Computing, 2018, 6(3): 760 - 770.
- [12] Cormen T H, Leiserson C E, Rivest R L, et al. Introduction to algorithms[M]. MIT Press, 2009.
- [13] Tran T X, Pompili D. Joint task offloading and resource allocation for Multi-Server Mobile-Edge computing networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(1): 856 - 868.
- [14] 黄晓舸,崔艺凡,张东宇,等.基于 MEC 的任务卸载和资源分配联合优化方案[J].系统工程与电子技术, 2020, 42(6): 1386 - 1394.

(上接第 123 页)

- [7] Sahai A, Waters B. Fuzzy identity-based encryption[C]//Proceedings of the 24th Annual International Conference on Theory and Applications of Cryptographic Techniques. Springer, 2005: 457 - 473.
- [8] Goyal V, Pandey O, Sahai A, et al. Attribute-based encryption for fine-grained access control of encrypted data[C]//Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006: 89 - 98.
- [9] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption[C]//IEEE Symposium on Security and Privacy(S&P). IEEE, 2007: 321 - 334.
- [10] Chase M. Multi-authority attribute based encryption[C]//Theory of Cryptography Conference. Springer, 2007: 515 - 534.
- [11] Lin H, Cao Z F, Liang X H, et al. Secure threshold multi authority attribute based encryption without a central authority[C]//International Conference on Cryptology in India. Springer, 2008: 426 - 436.
- [12] Gennaro R, Jarecki S, Krawczyk H, et al. Secure distributed key generation for discrete-log based cryptosystems[C]//Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques. ACM, 1999: 295 - 310.
- [13] 梁艳丽,凌捷.基于区块链的云存储加密数据共享方案[J].计算机工程与应用, 2020, 56(17): 41 - 47.
- [14] 李非非,韩笑,曾琦.具有隐私保护的固定密文长度分布属性基加密方案[J].计算机应用与软件, 2018, 35(5): 323 - 327.
- [15] 杨诗雨,李学俊.可密钥验证的多授权属性基加密方案[J].计算机应用与软件, 2017, 34(5): 311 - 316.