



西安交通大学学报
Journal of Xi'an Jiaotong University
ISSN 0253-987X, CN 61-1069/T

《西安交通大学学报》网络首发论文

题目: 移动边缘计算中分布式的设备发射功率优化算法
作者: 周文晨, 方维维, 李阳阳, 薛峰, 王子岳
收稿日期: 2018-04-07
网络首发日期: 2018-10-25
引用格式: 周文晨, 方维维, 李阳阳, 薛峰, 王子岳. 移动边缘计算中分布式的设备发射功率优化算法[J/OL]. 西安交通大学学报.
<http://kns.cnki.net/kcms/detail/61.1069.T.20181023.1652.018.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:

移动边缘计算中分布式的设备发射功率优化算法

周文晨¹, 方维维¹, 李阳阳², 薛峰¹, 王子岳¹

(1. 北京交通大学计算机与信息技术学院, 100044, 北京; 2. 中国电子科学研究院创新中心, 100041, 北京)

摘要: 针对大规模移动边缘计算网络架构中的用户设备卸载计算任务时所需的通信和计算资源难以协同优化的问题, 提出了一种基于马尔可夫近似的分布式发射功率优化算法。基于香农定理和链路传输特性将用户功耗最小化策略建模成组合优化模型, 通过 Log-Sum-Exp 原则将目标模型转化为最小权重配置的近似问题; 针对该近似问题, 提出了马尔可夫状态跳转的规则和分布式的设备自调节机制以实现高效求解。实验结果表明: 与随机优化算法相比, 该算法的系统用户设备发射总功耗优化效果提升了78.5%, 在给定场景下, 穷举搜索最优解的计算复杂度可达 4^{10} , 而该算法仅需要迭代优化130次后即可逼近最优解, 能够有效减少通信和计算时延, 确保发射功率的调整结果快速向最优目标收敛。

关键词: 移动边缘计算; 计算卸载; 马尔可夫近似; 分布式算法

中图分类号: TN915.6 **文章编号:** 0253-987X(2018)12-0000-00

A Distributed Transmit Power Optimization Algorithm for Mobile Edge Computing System

ZHOU Wenchen¹, FANG Weiwei¹, LI Yangyang², XUE Feng¹, WANG Ziyue¹

(1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;

2. Innovation Center, China Academy of Electronics and Information Technology, Beijing 100041, China)

Abstract: A novel distributed transmit power optimization algorithm based on Markov approximation framework is proposed to solve the problem that it is difficult to collaborative optimize communication and computing resources for computing task's offloading in large-scale mobile edge computing's network architecture. Strategy for minimizing power consumption of user devices is modeled as a combined optimization model based on Shannon's theorem and link transmission characteristics. According to the Log-Sum-Exp approximation rule, the original problem is transformed into a relevant approximated version with the aim to find a minimum weighted configuration. For the designed approximated version, the Markov state transition rule and the distributed self-adjustment mechanism are proposed to achieve efficient solution. The experimental results show that the total power consumption of the devices for the proposed algorithm is 78.5% better than that of random method. In a given scenario, the complexity of the exhaustive search's optimal solution is up to 4^{10} , while the proposed algorithm approaches the optimal solution after only 130 times Iterations. The proposed algorithm can effectively constrain the delay and ensure that the adjustment result of the transmission power converges to the optimal target quickly.

Keywords: mobile edge computing; computation offloading; Markov approximation; distributed algorithm

近年来, 移动互联网快速发展, 移动用户设备数量呈指数爆炸式增长^[1], 移动计算需求不断升级。根据思科最新预测报告显示, 2021年全球移动数据流量将比2016年增长7倍, 全球移动设备数

量将增长到116亿^[2], 但目前以长距离数据传输和集中式大数据处理为特点的移动云计算不仅占用大量网络带宽并且传输时延较大, 已无法满足时延敏感型业务需求, 例如无人驾驶汽车、医疗大数据

收稿日期: 2018-04-07。 作者简介: 周文晨(1993—), 女, 硕士生; 方维维(通讯作者), 男, 副教授。 基金项目: 国家自然科学基金资助项目(61501022); 中央高校基本科研业务费资助项目(2017JBM021)。

网络出版时间:

网络出版地址:

和智慧城市等^[3]。

欧洲电信标准化研究所 (ETSI) 在 2014 年首次提出移动边缘计算 (MEC), 移动设备可将高复杂度、高能耗计算任务卸载到 MEC 系统的接入网边缘节点, 例如基站、无线接入点等, 从而获得低时延、近距离的本地化云服务^[4]。MEC 系统结合了移动通信^[5]和云计算^[6]两种技术, 通过协同优化通信和计算资源实现低能耗、低时延的计算卸载服务。目前, 最新的 MEC 系统研究将关注点放在单服务器多用户条件下的通信、计算资源分配控制, 以实现总体能耗最优。Mao 等基于随机优化模型的优化方法, 联合本地计算能耗、最优发射功率和无线带宽资源的决策进行控制优化移动设备能耗, 但是无法有效约束时延, 且计算复杂性较大, 没有考虑某个区域内 MEC 系统的整体部署优化^[7]; Liu 等通过使用马尔可夫决策过程理论对任务队列的排队属性进行分析, 提出了发射功率约束条件下的最小化处理时延的问题, 但忽略了设备的能耗^[8]; You 等针对 TDMA、OFDMA 两种网络工作模式下, 通过确定卸载量和时间槽, 解决卸载通信和本地计算总体能耗最小化的凸优化问题, 但是所提算法需要依赖于本机计算能量和信道增益的先验知识做出资源分配策略, 降低了实际场景的可用性^[9]。

综上所述, 本文以 MEC 系统的多服务器、多用户的大规模场景下的任务计算卸载的时延和能耗的折衷平衡为研究目标, 基于边缘计算的通信和计算资源的耦合约束, 设计一种基于马尔可夫近似的分布式发射功率优化算法 (TPO): 基于马尔可夫状态概率跳转规则设计, 移动用户设备自主决策计算卸载的服务器对象, 有效降低系统级用户功耗。理论证明分布式 TPO 算法具有稳态概率分布的马尔可夫链, 所设计的转移速率满足马尔可夫链的状态相互转换条件; 实验仿真结果表明, 所提算法显著优于基准算法, 并在有效时间内逼近最优解决方案。

1 MEC 系统建模

本文研究的 MEC 系统场景示例如图 1 所示。根据已有相关研究, 本文基于以下 5 种假设: ①每个边缘服务器时间和频率完全同步; ②每个边缘服务器上的每个载波的信道增益在特定时间内保持恒定; ③由于每个边缘服务器的载波频率不同, 假设 MEC 系统中边缘服务器间不存在信号干扰; ④设备同一时刻只能与一个边缘服务器进行连通。⑤

卸载所需要执行的指令都可以利用边缘服务器进行处理^[10-13]。

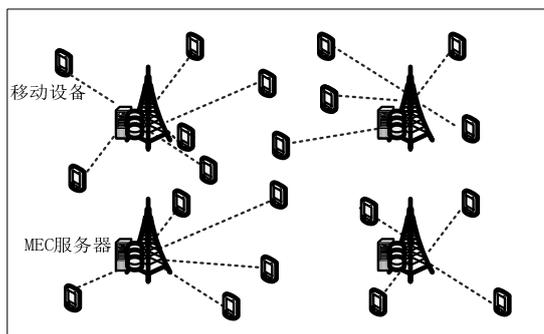


图 1 多 MEC 服务器多用户设备场景示例

假设当前区域内有 M 个边缘服务器 $X = \{1, 2, \dots, M\}$ 和 N 个移动用户设备 $Y = \{1, 2, \dots, N\}$ 。每个移动用户设备 n 的计算任务卸载到边缘服务器 m , 设备 n 必须传送所有信息到服务器 m 。我们设定执行程序的参数: ①每秒输入比特数; ②需要被 MEC 服务器执行计算的指令集; ③每秒输出比特数。对于移动用户设备 n 所需要卸载到服务器 m 的计算任务的输入数据为 C_{mn} , 执行的指令数为 $D_{mn} = D_{mn}(C_{mn})$ 。设备 n 是否卸载计算任务到服务器 m 主要取决于如下因素: 需要被处理的指令数、设备 n 与对应服务器 m 之间的无线信道状况和服务器 m 的并行多个进程的能力大小, 同时每个进程为了用户的满意体验 (QoE) 都规定了不应超过的最大时延^[14], 因此移动用户设备 n 决定是否卸载计算任务到服务器 m 之前必须考虑时延约束。在没有解码错误的前提下, 用户设备 n 的时延应该包含将输入信息传输到对应服务器 m 的时间、服务器 m 执行指令所需的时间和将结果返回给设备 n 的时间。

根据香农公式, 在加性白高斯噪声 (AWGN), 信道带宽为 B 的信道上传送 C_{mn} 比特的数据所需最小时间为

$$t_{mn} = C_{mn} / B \ln[1 + (P_n |H_{mn}|^2) / (G_{mn} d_{mn}^\gamma N_o)] \quad (1)$$

式中: P_n 是设备 n 计算任务卸载的发射功率; $|H_{mn}|^2$ 是信道衰落系数 (归一化距离); G_{mn} 是信噪比裕度; d_{mn} 是设备 n 和相连 MEC 服务器 m 的物理距离; γ 是路径损耗指数; N_o 表示噪声功率。其中 P_n 是可调功率, $P_n \in S$, S 是有限离散集合。

实际情况下, 因为传输错误需要重传错误的数据包, 所以服务器 m 接收到正确的输入数据所需要的时间通常大于 t_{mn} 。设定第 n 个链路上传输数据包的错误率为 P_{err}^n , 若每个传输数据包的错误是独

立的,那么平均重传次数为 $1/(1-P_{\text{err}}^n)$ 。设备 n 的计算任务正确卸载到 MEC 服务器 m 端应包含传输和重传过程,因此服务器 m 端正确接收 C_{mn} 比特数据的平均时延为

$$T_{nm}^{\text{trans}} = \frac{t_{mn}}{1-P_{\text{err}}^n} \quad (2)$$

服务器 m 执行指令所需的时间与其资源分配方式有直接关联。系统内用户设备与服务器的连接方式表示为

$$x_{mn} = \begin{cases} 1, & \text{服务器 } m \text{ 为 } n \text{ 提供服务} \\ 0, & \text{否则} \end{cases} \quad (3)$$

从式 (3) 可知,服务器 m 服务的用户设备数为 $K_m = \sum_{n=1}^N x_{mn}, m \in X, n \in Y$ 。设定服务器 m 每秒可执行 E_m 个指令,对其服务的用户设备 n 采取资源均分策略,故服务器 m 为设备 n 提供的计算能力是 E_m/K_m , m 运行 D_{mn} 条指令必要时间为

$$T_{mn}^{\text{comp}} = \frac{D_{mn}K_m}{E_m} \quad (4)$$

忽略服务器 m 把计算结果返回给设备 n 的极短时间^[11],则设备 n 的平均计算卸载的时延为

$$T_n = (T_{nm}^{\text{trans}} + T_{nm}^{\text{comp}})x_{mn} \quad (5)$$

T_n 不应超过每个用户设备的 QoE 所规定的最大时延约束 T_n^{max} , 以满足移动用户设备 n 的时延需求。

系统用户的总功率开销表示为 $\sum_{n=1}^N P_n$, 则我们的优化问题模型可以表示为

$$\begin{aligned} & \min \sum_{n=1}^N P_n \\ & \text{s.t.} \sum_{n=1}^N x_{mn} = 1, m \in X, n \in Y \\ & K_m = \sum_{n=1}^N x_{mn}, m \in X, n \in Y \\ & T_n \leq T_n^{\text{max}}, n \in Y \\ & P_n \in S, n \in Y \\ & x_{mn} \in \{0,1\}, m \in X, n \in Y \end{aligned} \quad (6)$$

基于香农定理和链路传输特性将用户功耗最小化策略建模成问题 (6), 作为典型的组合网络优化问题,其求解复杂度是 NP 难的,一般只能通过集中式的穷举搜索获得最优解^[15],可行解空间随着网络规模的增大而呈指数级增加,而随机法又不能获得

稳定而较优的解,因此寻求一种有效和稳定的求解方法是有必要的。

2 TPO 算法设计

针对问题 (6), 我们设计一种分布式并发的 TPO 算法,利用 Log-Sum-Exp 函数^[15]将问题转化为最小权重配置的近似问题,然后使用马尔可夫近似^[15]求解问题,即通过求解转化后的问题从而趋近问题 (6) 的最优解,并证明算法可进行有效求解。

2.1 Log-Sum-Exp 原则

我们设定配置 $f = \{x_{mn}, m \in X, n \in Y\}$, 即问题 (6) 所表示的系统中移动用户设备 n 和边缘服务器 m 相连的可行配置状态,所有可行的状态 f 组成 F , 即 $f \in F$ 。

则问题 (6) 的等效问题模型是

$$\begin{aligned} & \min_{\rho \geq 0} \sum_{f \in F} \rho_f \sum_{n=1}^N P_n(f) \\ & \text{s.t.} \sum_{f \in F} \rho_f = 1 \end{aligned} \quad (7)$$

式中: ρ_f 是配置 f 使用的时间比率。其中 $\sum_{n=1}^N P_n(f)$ 可看成配置 f 的权重,因此问题转化为寻找马尔可夫链最小权重的配置。

问题 (7) 是 NP 难的组合优化问题,对实际系统来说可行集合 F 非常大,因此通过 Log-Sum-Exp 函数,将问题 (7) 近似转化为

$$\begin{aligned} & \min_{\rho \geq 0} \sum_{f \in F} \rho_f \sum_{n=1}^N P_n(f) + \frac{1}{\beta} \sum_{f \in F} \rho_f \ln \rho_f \\ & \text{s.t.} \sum_{f \in F} \rho_f = 1 \end{aligned} \quad (8)$$

式中: β 是正的常数。因为问题 (8) 的目标函数对于所有的 ρ_f 是二次可微、单调递增和严格的凹函数,其所有的约束都是线性的,所以 Karush-Kuhn-Tucker (KKT) 条件对于最优解是必要和充分的。故问题 (8) 的最优解是

$$\rho_f^*(x) = \frac{\exp(-\beta \sum_{n=1}^N P_n(f))}{\sum_{f' \in F} \exp(-\beta \sum_{n=1}^N P_n(f'))}, f \in F \quad (9)$$

与原问题 (6) 相比,近似问题 (8) 存在一个额外的熵项,即为优化差距,通过对比两问题的表达式可知优化差距的上限是

$$r^* = \max \left(\frac{1}{\beta} \sum_{f \in F} \rho_f \ln \rho_f \right) \quad (10)$$

从式 (10) 可知优化差距上限的大小取决于 β , 当 β 增加时, 优化差距上限会减少, 这意味着近似求解结果更加准确, 优化差距上限 $r^* = \lfloor F \rfloor / \beta$, 此时 $\rho_f = 1/\lfloor F \rfloor, f \in F$, 同时从文献[16]可知当 β 增加时, 收敛时间将变长。

综上所述, 当 β 增加时, 近似求解获得的系统移动用户设备发射总功耗更加准确, 因此我们选择相对较大的 β 来获得稳态分布。通过马尔可夫近似获得的近似系统功耗可表示为

$$\sum_{f \in F} \rho_f^* \sum_{n=1}^N P_n(f) \quad (11)$$

2.2 TPO 算法设计

如文献[15]的引理 1 所述, 存在至少一个具有稳态分布 ρ_f^* 的连续时间可逆遍历马尔可夫链。状态空间 F 必须满足两个条件: ①具有遍历性质, 该状态空间中的任何两个状态可以通过至少一条路径相互连通。②具有时间可逆性质, 满足细致平衡等式, 即 $\rho_f q_{f,f'} = \rho_{f'} q_{f',f}$ 。

根据[15]可知有很多设计选择, 我们设计状态转移速率为

$$q_{f,f'} = \frac{1}{\zeta(M-1)} \frac{\exp(-\beta \sum_{n=1}^N P_n(f'))}{\max\{\exp(-\beta \sum_{n=1}^N P_n(f')), \exp(-\beta \sum_{n=1}^N P_n(f))\}} \quad (12)$$

马尔可夫链通常具有分布式的结构, 因此我们可以设计分布式算法以实现符合马尔可夫链的移动用户设备的决策算法^[17]。算法设计的关键是建立不同配置 f 之间的链接, 实现最小化转换配置的系统级功耗。通过每次只执行一个用户设备切换服务器的选择来连接系统两个配置的链接。将算法命名为发射功率优化 (TPO) 算法, TPO 算法遵循以下 4 个步骤。

步骤 1 每个移动用户设备随机选择 MEC 服务器, 初始化计算资源和设备时延约束等信息。

步骤 2 当前状态定义为配置 f , 每个用户设备计算自己的发射功率并广播, 同时生成均值为 ζ 的指数分布随机数, 根据随机数进行倒计时, 最先到期的用户设备 n 将随机切换到新 MEC 服务器, 并通知其他所有用户设备终止倒计时。

步骤 3 进入新配置 f' , 每个用户设备重新计算自己的发射功率并广播。用户设备 n 根据配置 f 和 f' 的系统移动用户设备发射总功耗的改变做出决策,

以概率 $\rho_{f,f'}$ 停留在新的 MEC 服务器或者以概率 $(1-\rho_{f,f'})$ 切换到原先的 MEC 服务器, 其中

$$\rho_{f,f'} = \frac{\exp(-\beta \sum_{n=1}^N P_n(f'))}{\max\{\exp(-\beta \sum_{n=1}^N P_n(f')), \exp(-\beta \sum_{n=1}^N P_n(f))\}} \quad (13)$$

步骤 4 重复步骤 2~步骤 4, 直至收敛。

根据算法可知设备随机切换到新的 MEC 服务器并以概率 $\rho_{f,f'}$ 继续与新服务器相连, 当 $\exp(-\beta \sum_{n=1}^N P_n(f')) \leq \exp(-\beta \sum_{n=1}^N P_n(f))$ 时, 即设备 n 切换到新服务器, 系统能得到更优表现, 此时设备 n 将以概率 1 停留在新服务器。当 $\exp(-\beta \sum_{n=1}^N P_n(f')) \geq \exp(-\beta \sum_{n=1}^N P_n(f))$ 时, 即设备 n 与原先的服务器连接, 系统能得到更优表现, 此时设备 n 会以概率 $1-\rho_{f,f'}$ 切换回原先服务器进行连接获得计算任务卸载服务, 也即设备 n 切换回原先服务器的概率会随着新旧配置系统设备发射总功耗差距 $[\exp(-\beta \sum_{n=1}^N P_n(f')) - \exp(-\beta \sum_{n=1}^N P_n(f))]$ 增加而增大。

2.3 TPO 算法可行性证明

定理 1 TPO 算法实现了一个时间可逆的马尔可夫链, 其稳态分布如式 (9) 所示。

证明 通过算法步骤 2 的移动用户设备的转换条件可知, 所有配置可以在有限转换次数内相互可达, 因此构造的马尔可夫链是不可约的。此外, 它是具有唯一稳态分布的有限状态遍历马尔可夫链。现在证明稳态分布是式 (9)。

基于式 (12) 设定的系统状态转移速率可得

$$\begin{aligned} \rho_f q_{f,f'} &= \exp(-\beta \sum_{n=1}^N P_n(f)) q_{f,f'} \\ &= \exp(-\beta \sum_{n=1}^N P_n(f')) q_{f',f} = \rho_{f'} q_{f',f} \end{aligned} \quad (14)$$

它满足细致平衡等式, 证毕。

定理 2 通过上述完备的马尔可夫链设计, 对于任意两个配置 $f, f' \in F$ 满足相互转换条件, 转移速率为式 (12)。

证明 从系统的角度来看, 状态 f 转移到状态 f' 是由于配置 f 下的设备 n 切换服务器所致, 设备 n 从已连接的 MEC 服务器断开, 随机切换到新的服务器, 可切换的服务器有 $M-1$ 个。故对于设备 n 离开状态 f 转移概率为

$$\frac{1}{(M-1)} \frac{\exp(-\beta \sum_{n=1}^N P_n(f'))}{\max\{\exp(-\beta \sum_{n=1}^N P_n(f')), \exp(-\beta \sum_{n=1}^N P_n(f))\}}$$

(15)

其中配置 $f \in F$ 是马尔可夫链的状态之一, 因为设备是根据倒计时机制激活切换服务器步骤, 倒计时时间是均值为 ζ 的指数分布随机数, 因此状态 f 转移到状态 f' 的转移速率为

$$\frac{1}{\zeta(M-1) \max\{\exp(-\beta \sum_{n=1}^N P_n(f')), \exp(-\beta \sum_{n=1}^N P_n(f))\}} \exp(-\beta \sum_{n=1}^N P_n(f')) \quad (16)$$

因此转移速率与式 (12) 相等, 证毕。因此本文算法满足近似最优解, 是可行设计。

3 仿真与结果分析

3.1 实验设置

为了验证 TPO 算法的有效性和稳定性, 采用 C++ 编程进行仿真实验, 与随机法和穷举法的实验结果进行对比分析, 同时分析关键参数对系统的影响。参数设置见表 1^[11,18]。移动用户设备 n 的输入数 C_{mn} 设定为 $(0.2+0.02n)$ MB, 为简化起见, 设定 MEC 服务器执行的指令数 D_{mn} 与输入数 C_{mn} 成线性相关, 即 $D_{mn} = \sigma C_{mn}$ 。集合 $S = \{0\text{w}, 10^{-2}\text{w}, 2 \times 10^{-2}\text{w}, \dots, 10\text{w}\}$ 。

表 1 TPO 算法仿真参数

参数	数值
M	4
N	100
B/MHz	6.4×10^2
$P_{\text{err}}^n / \%$	50
T^{max}/s	2.3×10^{-2}
E_m / s	$[1, 1.3, 1.6, 1.8] \times 10^9$
$ H_{mn} ^2$	1
σ	40
G_{mn}	0.65
N_o / W	10^{-3}
β	20
ζ	30

3.2 实验结果分析

因采用穷举搜索获得系统 M 个移动用户设备与 N 个 MEC 服务器的连接方式的最优解共需计算 N^M 次, 复杂度过高, 因此我们先将 TPO 算法与随

机法实验结果进行对比。图 2 分别给出了 TPO 算法与随机法得到的系统用户设备发射总功率。图中红色标注为随机方法实验多轮获得的系统用户设备发射总功率的平均值, 蓝色标注为 TPO 算法多次实验获得的迭代过程的平均值曲线。从图中可知虽然 TPO 算法实现的系统用户设备发射总功率平均值在算法执行初期比随机方法平均值大, 但是在 50 次迭代后小于随机方法平均值。TPO 算法实现的总功率均值随着算法迭代次数逐步下降, 可收敛至 4.6W, 而多次随机方法的均值为 21.4W, 由此可推算出 TPO 算法性能比随机法的效果提升了 78.5%, 因此 TPO 算法不仅可行, 并且得到的收敛解比随机方法结果更趋近最优解, 更加有效。

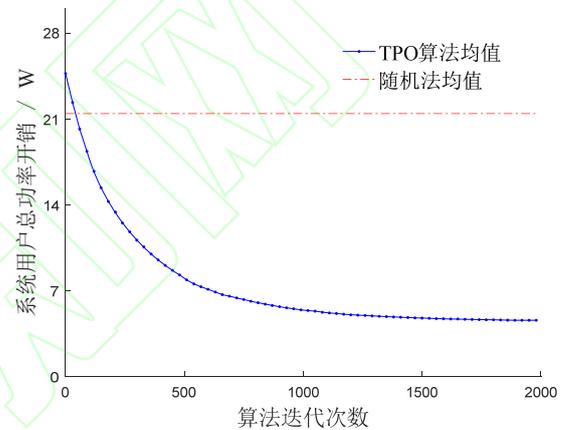


图 2 TPO 算法与随机算法对比

为了进一步验证 TPO 算法的准确性, 我们建立 4 个 MEC 服务器服务 10 个用户设备的小型 MEC 系统场景, 在此场景下对比分析 TPO 算法多次实

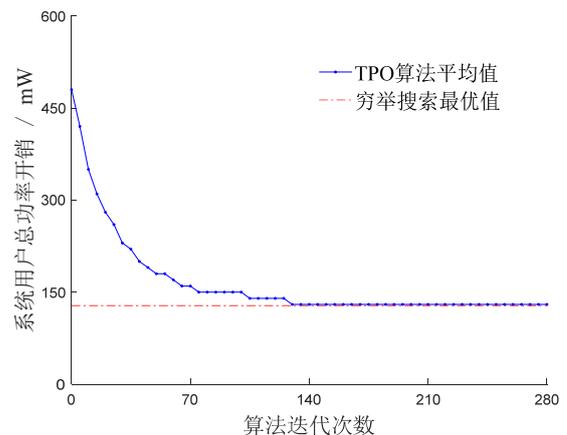


图 3 TPO 算法与穷举搜索结果对比

验平均值和穷举搜索最优结果。图 3 给出了 TPO 算法与穷举搜索实现的系统用户设备发射总功率, 图 3 中点划线是穷举搜索实现的最优解, 实线为

TPO 算法的系统用户设备发射总功率均值迭代结果。实验数据显示 TPO 算法值仅在第 130 次迭代后就逼近最优解, 远小于理论优化差距, 与穷举法的 4^{10} 计算次数相比, TPO 算法复杂度大大降低, 更具有实际应用价值。

以上实验结果验证了 TPO 算法的有效性和实用性, 下面将分析系统关键参数对 TPO 算法结果的影响。图 4 显示了 TPO 算法的参数 β 对算法收敛时间和系统用户设备发射总功耗的影响。从图 4 中可知, 当 $\beta=20$ 时, 系统总功率接近最优解, 收敛时间较长; 当 $\beta=5$ 时, 系统总功率远离最优解, 收敛时间较短。因此伴随 β 的增加, TPO 算法获得的近似系统总功耗越接近最优值, 同时收敛时间变长, 符合式 (10) 理论结果。

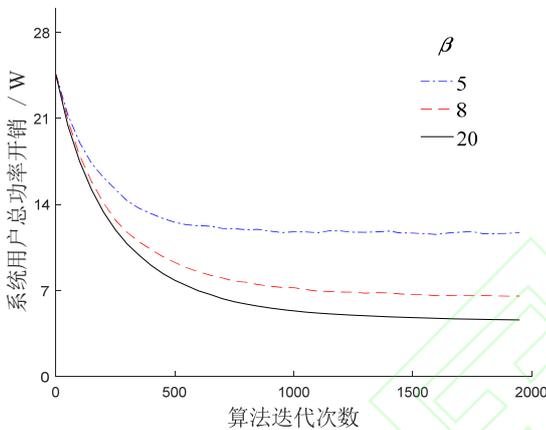


图 4 系统用户设备发射总功耗与 β 的关系

图5是将所有移动用户设备的输入数据大小 C_{mn} 设定为相同值并统一调整, 观察系统用户设备总功耗与设备所需计算卸载的输入数据量大小 C_{mn} 的关系。从图5中可以看出, 当 C_{mn} 越大, TPO 算法实现的系统用户设备的总发射功率会增大, 这是因为设备需要更多的能耗传输更多的数据, 但是总发射功率的收敛时间并没有伴随 C_{mn} 增加发生显著变化, 说明当设备的输入比特量增加时, TPO 算法的收敛速度会更快, 这对实际系统中用户设备计算任务卸载高峰期的处理具有很大的优势。

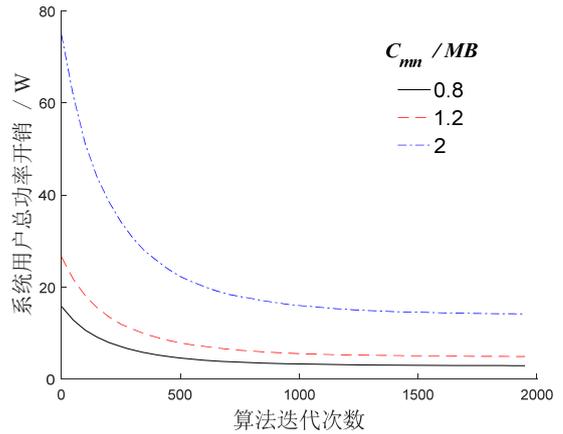


图 5 系统用户设备发射总功耗与 C_{mn} 的关系

为了显示 TPO 算法对用户设备资源分配的调节效果, 我们建立了用户设备高度聚集的特殊场景, 将 TPO 算法与其他常用算法的处理效果进行对比。图 6 中实线为 MEC 服务器, 虚线为用户, 点划线表示用户和相应的 MEC 服务器连接情况。通过观察可知, 图 6b 中传统的依赖接收信号强度指示 (RSSI) 进行连接的方式会导致 MEC 服务器资源分配不合理, 进而无法满足用户需求, 图 6c 中随机法实现的系统设备发射总功率为 48.3 W, 图 6d 中 TPO 算法实现的系统设备发射总功率仅为 8.7W, 因此随机方法的实验结果不满足最小化目标, 本文提出的 TPO 算法能够合理有效的建立移动用户设备和 MEC 服务器的连接, 达到系统设备发射总功耗优化的最优效果。

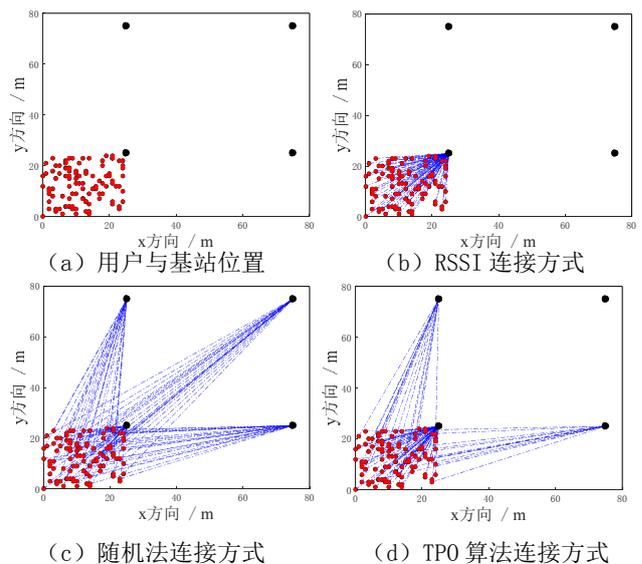


图 6 不同算法用户设备连接服务器信息

4 结论

传统的云计算模型已无法有效解决海量的边缘数据计算卸载问题,如何提升移动边缘计算的大规模网络架构的计算卸载的效率和用户体验受到越来越多的关注^[19]。本文在传输带宽和计算资源参数化的基础上,使用马尔可夫近似框架将系统用户设备的发射功率最小化问题转化,设计分布式的发射功率优化算法求得原目标的近似最优解。对比基准算法测试结果,TPO算法的系统用户设备发射总功耗优化效果比随机法提升了78.5%,同时执行效果更加稳定;相较于穷举搜索的指数阶 $O(2^n)$ 计算复杂度,该算法的复杂度为线性阶 $O(n)$,故该算法在有限迭代轮次后即可逼近穷举搜索的最优解;与传统的RSSI连接方式相比,TPO算法能够在用户量小范围高度集中时争抢计算资源的现象,减少网络拥塞和资源分配不均衡的情况发生,满足用户计算卸载需求。TPO算法能够在保证用户设备的时延约束下有效降低系统用户功耗成本,确保时延敏感型业务的时延约束下的用户体验。

参考文献:

- [1] FANG Weiwei, YAO Xuening, ZHAO Xiaojie, et al. A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms [J]. IEEE Transactions on Systems Man and Cybernetics Systems, 2016, (99):1-13.
- [2] Cisco. Cisco visual networking index: global mobile data traffic forecast update [EB/OL]. (2018-01-22)[2018-04-07]. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf>.
- [3] 何秀丽, 任智源, 史晨华, 等. 面向医疗大数据的云雾网络及其分布式计算方案[J]. 西安交通大学学报, 2016, 50(10): 71-77.
HE Xiuli, REN Zhiyuan, SHI Chenhua, et al. A cloud and fog network architecture for medical big data and its distributed computing scheme [J]. Journal of Xi'an Jiaotong University, 2016, 50(10):71-77.
- [4] 田辉, 范绍帅, 吕昕晨, 等. 面向5G需求的移动边缘计算[J]. 北京邮电大学学报, 2017, 40(2):1-10.
TIAN Hui, FAN Shaoshuai, LÜ Xinchun, et al. Mobile edge computing for 5G requirements [J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(2): 1-10.
- [5] FANG Weiwei, ZHAO Xiaojie, AN Yuan, et al. Optimal scheduling for energy harvesting mobile sensing devices[J]. Computer Communications, 2016, 75(C):62-70.
- [6] 李健, 黄庆佳, 刘一阳, 等. 云计算环境下的大规模图状数据处理任务调度算法[J]. 西安交通大学学报, 2012, 46(12): 116-122.
LI Jian, HUANG Qingjia, LIU Yiyang, et al. A task scheduling algorithm for large graph processing cloud in computing [J]. Journal of Xi'an Jiaotong University, 2012, 46(12): 116-122.
- [7] MAO Yuyi, ZHANG Jun, SONG S H, et al. Power-delay tradeoff in multi-user mobile-edge computing systems [J]. IEEE Transactions on Communications, 2015, 63(10): 3842-3855.
- [8] LIU Juan, MAO Yuyi, ZHANG Jun, et al. Delay-optimal computation task scheduling for mobile-edge computing systems [J]. IEEE Transactions on Mobile Computing, 2016: 1451-1455.
- [9] YOU Changsheng, HUANG Kaibin, HYUKJIN C, et al. Energy-efficient resource allocation for mobile-edge computation offloading [J]. IEEE Transactions on Wireless Communications, 2017, 16(3): 1397-1411.
- [10] CHEN Siwei, GAN Xiaoying, FENG Xinxin, et al. Markov approximation for multi-RAT selection [C]// Proceedings of the IEEE International Conference on Communications. Piscataway, NJ, USA: IEEE, 2015: 3045-3050.
- [11] BARBAROSSA S, SARDELLITTI S, LORENZO P D. Joint allocation of computation and communication resources in multiuser mobile cloud computing [C]// Proceedings of the Signal Processing Advances in Wireless Communications. Piscataway, NJ, USA: IEEE, 2013: 26-30.
- [12] HAJIESMAILI M H, MAK L T, WANG Z, et al. Cost-effective low-delay cloud video conferencing [C]// Proceedings of the International Conference on Distributed Computing Systems. Piscataway, NJ, USA: IEEE, 2015: 103-112.
- [13] LIEW S C, KAI C H, LEUNG H C, et al. Back-of-the-envelope computation of throughput distributions in CSMA wireless networks [J]. IEEE Transactions on Mobile Computing, 2007, 9(9): 1319-1331.

- [14] 王大鸣, 陈松, 崔维嘉, 等. 多用户 MIMO-OFDM 系统基于 QoE 效用函数的跨层资源分配 [J]. 通信学报, 2014, 35(9): 175-183.
WANG Daming, CHEN Song, CUI Weijia, et al. QoE utility function-based cross-layer resource allocation in multi-user MIMO-OFDM systems [J]. Journal on Communications, 2014, 35(9): 175-183.
- [15] CHEN Minghua, LIEW S C, SHAO Ziyu, et al. Markov approximation for combinatorial network optimization [C]// Proceedings of the Conference on Information Communications. Piscataway, NJ, USA: IEEE, 2010: 1783-1791.
- [16] SHAO Ziyu, ZHANG Hao, CHEN Minghua, et al. Reverse-engineering bittorrent: a Markov approximation perspective [C]// Proceedings of the 2012 IEEE INFOCOM. Piscataway, NJ, USA: IEEE, 2013: 2996-3000.
- [17] FANG Hui, LIN Xia, LOK T M. Power allocation for multiuser cooperative communication networks under relay-selection degree bounds [J]. IEEE Transactions on Vehicular Technology, 2012, 61(7): 2991-3001.
- [18] JIANG Libin, WALRAND J. A distributed CSMA algorithm for throughput and utility maximization in wireless networks [C]// Proceedings of the Conference on Communication, Control, and Computing. Piscataway, NJ, USA: IEEE, 2010: 960-972.
- [19] 施巍松, 孙辉, 曹杰, 等. 边缘计算:万物互联时代新型计算模型 [J]. 计算机研究与发展, 2017, 54(5): 907-924.
SHI Weisong, SUN Hui, CAO Jie, et al. Edge computing-an emerging computing model for the internet of everything era [J]. Journal of Computer Research and Development, 2017, 54(5): 907-924.

(编辑 武红江)