

# 大规模移动边缘计算网络： 空间建模及计算吞吐量优化

Large-Scale Mobile Edge Computing Network:  
Spatial Modeling and Computation Throughput Optimization



韩凯峰/HAN Kaifeng, 胡昌军/HU Changjun, 刘铁志/LIU Tiezhi

(中国信息通信研究院, 中国 北京 100191)

(China Academy of Information and Communications Technology, Beijing 100191, China)

**摘要:**提出并定义了大规模移动边缘计算(MEC)网络中的空间计算吞吐量这一性能指标,通过运用随机几何、凸优化等理论,对这一性能指标进行了分析和最优化设计。利用随机几何理论为大规模 MEC 网络建立空间模型,该模型涵盖边缘云及用户的空间随机分布、无线接入、计算任务卸载、边缘端并行计算等重要的网络特征。基于网络模型,首次对 MEC 网络空间计算吞吐量进行定义和分析,并通过优化设计 MEC 服务范围半径( $r_0$ )以及用户计算卸载比例( $\rho$ )这两个指标,来实现 MEC 网络空间吞吐量的最大化。所提供的严谨的理论分析、富有物理内涵的优化结果将为部署大规模 MEC 网络提供了极为重要的设计指南。

**关键词:**移动边缘计算;移动计算卸载;无线网络建模;随机几何;凸优化

**Abstract:** The attempt on defining, analyzing, and optimizing the spatial computation throughput in a large-scale wireless mobile edge computing (MEC) network is made. The analysis involves the interplay of theories of stochastic geometry and convex optimization. Specifically, the large-scale MEC network features of wireless access and edge-computing are first modeled, such as random nodes distribution, computation tasks offloading, parallel computing at edges, by using stochastic geometry. Based on the proposed model, the spatial computation throughput of the MEC network is defined, studied and maximized in terms of the radius of MEC service range (denoted by  $r_0$ ) as well as offloading ratio of active mobile users (denoted by  $\rho$ ) under the constraints of latency and energy. The optimal solutions of  $r_0$  and  $\rho$  can be easily calculated via solving simple equations, and their closed-form results could be obtained in the extreme case. The tractable analysis and insightful results give useful design guidelines for MEC network planning and provisioning in a large-scale space.

**Keywords:** mobile edge computing; mobile computation offloading; wireless network modeling; stochastic geometry; convex optimization

DOI: 10.12142/ZTETJ.202004003

网络出版地址: <https://kns.cnki.net/KCMS/detail/34.1228.TN.20200715.1709.002.html>

网络出版日期: 2020-07-16

收稿日期: 2020-05-21

作为 5G 系统中的关键技术之一,移动边缘计算(MEC)可以利用部署在网络边缘的服务器为移动用户提供泛在、低时延的高质量计算服务<sup>[1-2]</sup>,例如多媒体云游戏、增强现实等。相比中心化的云计算,去中心化的 MEC 能够显著降低计算时延、移动

用户能耗以及网络传输复杂度<sup>[3-4]</sup>,适用于未来边缘智能网络<sup>[5]</sup>、大规模物联网<sup>[6]</sup>以及点对点网络<sup>[7]</sup>等应用场景。

在 MEC 中,一个研究热点领域是设计高效能、低时延的移动计算卸载方式,即移动用户可将其计算任务卸载到 MEC 服务器中进行计算。文献

[8]提出了一个卸载计算策略,在给定计算时限要求下,通过联合设计频谱和计算分配资源来最小化移动用户的能耗。对于类似的优化目标,文献[9]给出了基于 Lyapunov 优化理论的动态优化结果。文献[10]则将移动计算卸载及资源分配设计问题拓展至

车联网场景。

虽然上述研究工作仅考虑包含一个或几个边缘云和用户的小规模 MEC 网络,就能设计出复杂但有效的计算卸载方案或策略,但是研究并设计并优化大规模 MEC 网络(包含无限多个边缘云和用户的 MEC 网络)中的无线通信和边缘计算的性能指标也同样重要。文献[11]首次提出了基于随机几何理论的大规模 MEC 网络模型,并对 MEC 网络的传输时延和计算时延进行了理论分析以及优化设计,为大规模 MEC 网络部署提供了重要设计指南;但该文献并未研究如何定义并最优化地设计 MEC 网络空间吞吐量的问题。

为此,我们提出并定义了大规模 MEC 网络中的空间计算吞吐量,并通过优化设计 MEC 服务范围半径以及用户计算卸载比例,来实现 MEC 网络空间吞吐量的最大化,以期部署大规模 MEC 网络提供参考。

### 1 系统模型及性能指标

考虑一个包含无限多个 MEC 服务器和移动用户的大规模 MEC 网络(如图1所示)。

#### 1.1 MEC 网络空间分布

在二维空间内,假设 MEC 服务器和用户的位置都服从泊松点过程(PPP)分布,其中 MEC 服务器位置  $X \in \mathbb{R}^2$  服从密度为  $\lambda_s$  的 PPP 分布  $\Omega = \{X\}$ ,用户位置  $Y \in \mathbb{R}^2$  服从密度为  $\lambda_u$  的 PPP 分布  $\Phi = \{Y\}$ 。

##### 1) 多用户接入模型

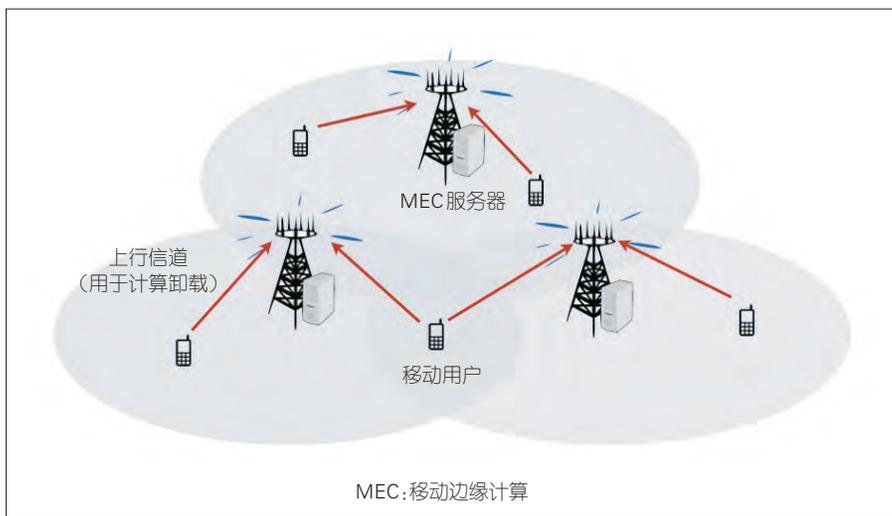
将总带宽资源分为  $M$  个正交的子信道  $\{1,2,\dots,M\}$ ,每个用户可以随机选择一个子信道将计算任务上传到 MEC 服务器进行计算(即计算卸载),选择同一子信道进行传输的多个用户将产生干扰。令时隙间隔为

$T_s$  秒,假设用户位置及信道状态在不同时隙相互独立,并要求每个用户需要在  $T_s$  秒内完成计算任务。

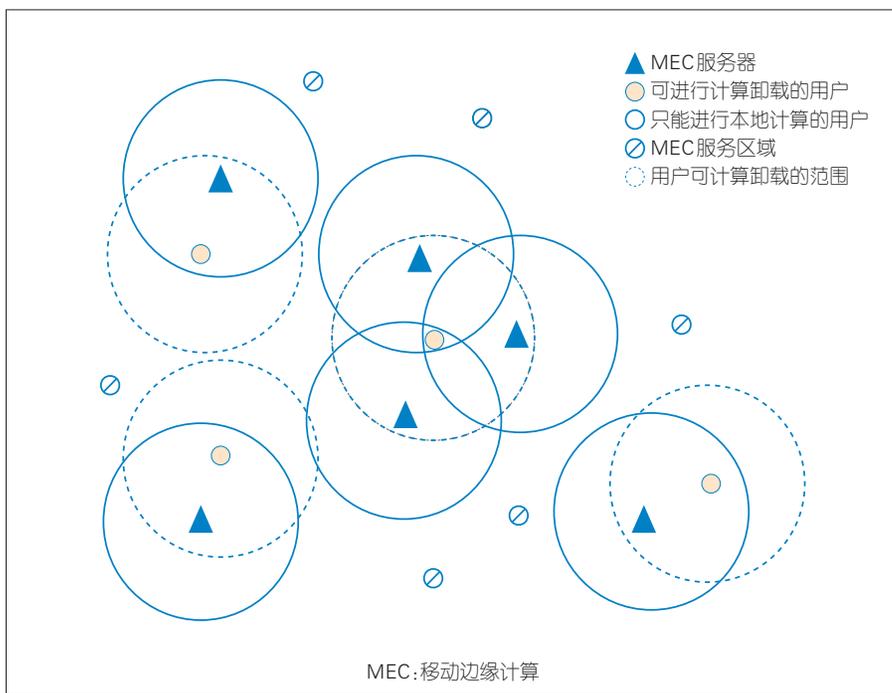
##### 2) MEC 服务区域

图2展示了大规模 MEC 网络空间分布图。每个 MEC 服务器  $X$  均可形成一个半径为  $r_0$  的圆形 MEC 服务区域,表示为  $O = (X,r_0)$ ,其中  $r_0$  为常量。假设只有被 MEC 服务区域覆盖的用户可以选择计算卸载,其他未被

覆盖的用户只能进行本地计算。任意一个 MEC 服务区域内的用户数量  $N$  是均值为  $\pi r_0^2 \lambda_u$  的泊松随机变量。对于任意一个用户,可供其进行计算卸载的服务器数量  $W$  是均值为  $\pi r_0^2 \lambda_s$  的泊松随机变量。当同一用户被多个 MEC 服务器覆盖时,它将把计算任务同时卸载到多个服务器进行计算以提高计算成功概率(其中有一个服务器完成计算,即代表计算成功)。



▲图1 MEC系统模型



▲图2 大规模 MEC 网络空间分布

### 1.2 边缘计算卸载模型

#### 1) 计算卸载比例

所有在 MEC 服务区域内的用户,均可决定是进行计算卸载还是进行本地计算。将网络中的选择计算卸载的用户所占比例称作用户计算卸载比例,表示为  $\rho \in (0,1)$ 。若  $\rho = 1$ ,则表示全部被服务区域覆盖的用户均选择计算卸载。因此,网络中选择计算卸载的用户密度表示为  $(1 - e^{-\lambda_u \pi r_0^2}) \rho \lambda_u$ 。

#### 2) 上行信道模型

考虑用户计算卸载时的上传信道,忽略功率控制影响,将 MEC 服务器端接收到的用户信号功率表示为  $qg|Y - X|^{-\alpha}$ 。其中,  $q$  为发射功率,  $g$  为小尺度锐利衰落系数,  $\alpha$  为大尺度路径损耗系数,  $|Y - X|$  为用户  $Y$  与服务器  $X$  之间的欧氏距离。假设在任意一子信道  $m \in \{1,2,\dots,M\}$  内,进行计算卸载的用户数服从均值为  $\frac{(1 - e^{-\lambda_u \pi r_0^2})}{M} \rho \lambda_u$  的 PPP 分布  $\Phi_u^{(m)} = \{Y_m\} \subseteq \Phi_u$ , 对于选择第  $m$  个子信道进行计算卸载的用户所受到的干扰信号  $I$  可以表示为  $I = \sum_{Y_{x,m} \in \Phi_u^{(m)} \setminus \{Y_{x,m}^*\}} qg_m |Y_{x,m}|^{-\alpha}$ , 其中  $g_m$  服从独立恒等分布。我们对任意选取的一个典型用户进行分析,在 MEC 服务器端,该典型用户信号的信号干扰比 SIR 可以表示为  $SIR = \frac{g_0 |Y_{x,m}^*|^{-\alpha}}{\sum_{Y_m \in \Phi_u^{(m)} \setminus \{Y_{x,m}^*\}} g_m |Y_m|^{-\alpha}}$ , 其中,当 SIR 不低于固定阈值  $\theta$  时,则为卸载传输成功。

### 1.3 计算模型

在计算过程中,主要考虑两个约

束条件:一是时延约束  $T_s$ ,即每个用户的计算任务需要在  $T_s$  秒内计算完毕;二是能量约束  $\xi$ ,即每个用户在每个时隙用于计算的能量不得超过  $\xi$ 。为便于分析,令  $\xi = qT_s$  以保证每个用户均有足够的能量用于计算卸载。

#### 1) 边缘计算

假设每个 MEC 服务器的计算能力有限,每当其接收到一个用户卸载的计算任务(包含  $\ell$  比特数据),便启动一个虚拟机进行独立的边缘计算。对于进行计算卸载用户,其每个计算任务的时延包括 3 部分:卸载传输时延  $T_t$ 、边缘计算时延  $T_c$ 、计算结果下载时延  $T_d$ 。由于计算结果数据量很小,因此可忽略  $T_d$ 。为满足时延约束条件,要求  $T_t + T_c \leq T_s$ 。对于卸载传输时延  $T_t$ ,令用户数据传输速率为  $\eta = B \cdot \log_2(1 + \theta)$ ,其中  $B$  为子信道带宽,  $T_t$  可表示为  $T_t = \ell/\eta$ ,期间能量消耗为  $qT_t$ 。对于边缘计算时延  $T_c$ ,根据文献[13]中的计算时延模型,  $T_c$  可表示为  $T_c = T_0(1 + d)^{i-1}$ ,其中,  $i$  为虚拟机数量,  $d \geq 0$  为多个虚拟机的复用退化因子,  $T_0 = \ell/\mu_{ec}$  为单个虚拟机计算每个任务的时延( $\mu_{ec}$  是虚拟机计算能力,单位为比特/秒)。

#### 2) 本地计算

令  $\mu_{lc}$  (单位为比特/秒) 为用户终端的本地计算算力。为满足时延约束条件,需  $\ell/\mu_{lc} \leq T_s$ ,等价于  $\ell \leq T_s \mu_{lc}$ ,即  $T_s \mu_{lc}$  是最大本地计算数据量。令  $\tau_{lc}$  为本地计算每比特数据所消耗的能量,为满足能量约束条件,最大本地计算数据量为  $qT_s/\tau_{lc}$ 。综上所述,最大本地计算数据量可表示为  $\ell_{lc}^{(max)} = \min\{T_s \mu_{lc}, qT_s/\tau_{lc}\}$ 。

### 1.4 性能指标

#### 1) 卸载传输成功概率

为定量刻画上传信道(即卸载传

输)的可靠性,定义卸载传输成功概率  $p_{\ell,s}$ ,数学表达式为  $p_{\ell,s} = \Pr(SIR \geq \theta)$ 。

#### 2) MEC 成功概率

首先,为刻画用户的计算任务,可以在规定时间  $T_s$  内计算完毕的概率  $p_c$ ,数学表达式为  $p_c = \Pr(T_t + T_c \leq T_s)$ 。考虑到每个用户可以将计算任务卸载至其附近的  $W$  个 MEC 服务器,当其中任意一个 MEC 服务器能够在规定时间内完成计算任务,就意味着该用户的卸载计算成功。基于此,为衡量 MEC 服务成功概率,定义 MEC 成功概率  $p_{mec}(W)$  为  $p_{mec}(W) = 1 - (1 - p_c p_{\ell,s})^W$ 。

#### 3) MEC 网络空间吞吐量

为刻画大规模网络中成功完成计算的用户空间密度,定义 MEC 网络空间吞吐量  $C = C_{ec} + C_{lc}$ ,其中  $C_{ec}$  表示利用边缘计算完成的吞吐量,  $C_{lc}$  表示利用本地计算完成的吞吐量,其数学表达式分别为:

$$C_{ec} = \mathbf{E} \left[ \rho \lambda_u (1 - e^{-\lambda_u \pi r_0^2}) \cdot p_{mec}(W) \right], \quad (1)$$

$$C_{lc} = \lambda_u \left[ (1 - \rho) (1 - e^{-\lambda_u \pi r_0^2}) + e^{-\lambda_u \pi r_0^2} \right] \cdot I(\ell_{lc}^{(max)} \geq \ell), \quad (2)$$

其中,  $I(A)$  为指示函数,即当事件  $A$  发生时,  $I(A)$  为 1,否则为 0。

结合公式(1)和(2),可以得到  $C$  的表达式:

$$C = \lambda_u \left[ (1 - \rho) (1 - e^{-\lambda_u \pi r_0^2}) + e^{-\lambda_u \pi r_0^2} \right] \cdot I(\ell_{lc}^{(max)} \geq \ell) + \mathbf{E} \left[ \rho \lambda_u (1 - e^{-\lambda_u \pi r_0^2}) p_{mec}(W) \right]. \quad (3)$$

## 2 对 MEC 网络空间计算吞吐量的理论分析

在本节中,我们将对 MEC 成功

概率以及 MEC 网络空间计算吞吐量等关键性能指标进行分析,为后续对网络进行优化设计提供理论基础。

### 2.1 卸载传输成功概率分析

鉴于直接得到  $p_{\ell,s}$  的表达式存在难度,故先推导出  $p_{\ell,s}$  的下界来分析最差情况下的卸载传输成功概率,这同样对网络优化设计具备参考意义。首先,假设典型用户与 MEC 服务器间的距离为  $r_0$ ,可得到 SIR 的下界  $\text{SIR} \geq \frac{g_0 r_0^{-\alpha}}{\sum_{Y_m \in \Phi_u^{(m)} \setminus \{Y_m^*\}} g_m |Y_m|^{-\alpha}} = \text{SIR}^{(\text{low})}$ , 由此可得到  $p_{\ell,s}$  的下界:

$$p_{\ell,s} \geq \exp\left[-\frac{2\pi}{\alpha}\left(1 - e^{-\lambda_u \pi r_0^2}\right)\left(\theta r_0^2\right)^{\frac{2}{\alpha}}\left(\frac{\rho \lambda_u}{M}\right)B\left(\frac{2}{\alpha}, 1 - \frac{2}{\alpha}\right)\right] = p_{\ell,s}^{(\text{low})}, \quad (4)$$

其中  $B(x, y) = \int_0^1 \kappa^{x-1} (1-\kappa)^{y-1} d\kappa$ 。

从公式(4)中可观察到,当网络中卸载用户密度增大时(即  $\lambda_u$  或  $\rho$  增大),将引起更为严重的用户间干扰,致使  $p_{\ell,s}^{(\text{low})}$  变小,扩大 MEC 服务范围  $r_0$  有类似效果。

### 2.2 MEC 成功概率分析

首先计算分析  $p_c$ 。根据 MEC 时延约束,在 MEC 服务器中每个时隙可产生的虚拟机最大数目  $i^{(\text{max})}$  为  $1 \leq i \leq$

$$\left\lfloor \frac{\ln\left(\frac{T_s \eta - \ell}{T_0 \eta}\right)}{\ln(1+d)} \right\rfloor + 1 = i^{(\text{max})}, \text{ 其中, } x \text{ 为下取整函数。}$$

在每个时隙中,每个 MEC 服务范围内有平均  $\rho \tilde{N} = \rho N p_{\ell,s}$  个用户成功地将计算任务卸载到服务器端,为保证

每个任务都得到计算,需要  $\rho \tilde{N} \leq i^{(\text{max})}$ ,因此  $p_c$  可改写为:

$$p_c = \Pr(\rho \tilde{N} \leq i^{(\text{max})}). \quad (5)$$

由于  $\rho \tilde{N}$  是均值为  $\rho \lambda_u \pi r_0^2 p_{\ell,s}$  的泊松随机变量,故  $\Pr(\rho \tilde{N} = k) = \frac{(\rho \lambda_u \pi r_0^2 p_{\ell,s})^k}{k!} \cdot \exp(-\rho \lambda_u \pi r_0^2 p_{\ell,s})$ , 因此  $p_c$  可表示为:

$$p_c = \sum_{k=0}^{i^{(\text{max})}} \frac{(\rho \lambda_u \pi r_0^2 p_{\ell,s})^k}{k!} \exp(-\rho \lambda_u \pi r_0^2 p_{\ell,s}) = \frac{\Gamma(i^{(\text{max})} + 1 - \rho \lambda_u \pi r_0^2 p_{\ell,s})}{\Gamma(i^{(\text{max})})}, \quad (6)$$

基于公式(6),  $p_c$  下界可表示为:

$$p_c \geq \frac{1}{\Gamma(i^{(\text{max})})} (\rho \lambda_u \pi r_0^2 p_{\ell,s}^{(\text{low})})^{i^{(\text{max})} - 1} \exp(-\rho \lambda_u \pi r_0^2 p_{\ell,s}^{(\text{low})}) = p_c^{(\text{low})}, \quad (7)$$

基于公式(4)和(7),可得到 MEC 成功概率的下界:

$$p_{\text{mec}}(W) \geq 1 - \left(1 - p_c^{(\text{low})} \cdot p_{\ell,s}^{(\text{low})}\right)^W = 1 - \left[1 - \frac{1}{\Gamma(i^{(\text{max})})} \left(\rho \lambda_u \pi r_0^2 e^{-\frac{2\pi \rho \lambda_u (\theta r_0^2)^{\frac{2}{\alpha}} \beta \left(\frac{2}{\alpha} - \frac{2}{\alpha}\right)}{\alpha M} \left(1 - e^{-\lambda_u \pi r_0^2}\right)}\right)^{i^{(\text{max})} - 1} \times \exp\left(\rho \lambda_u \pi r_0^2 e^{-\frac{2\pi \rho \lambda_u (\theta r_0^2)^{\frac{2}{\alpha}} \beta \left(\frac{2}{\alpha} - \frac{2}{\alpha}\right)}{\alpha M} \left(1 - e^{-\lambda_u \pi r_0^2}\right)}\right)\right]^W, \quad (8)$$

观察公式(8),假设  $p_c^{(\text{low})} \cdot p_{\ell,s}^{(\text{low})} \rightarrow 0$ , 将有  $p_{\text{mec}}(W) \approx W \cdot (p_c^{(\text{low})} \cdot p_{\ell,s}^{(\text{low})})$ 。从中可见,用户的 MEC 成功概率会随着周围的服务器数量的增多而增长,符合直观预期。

根据公式(3)和公式(8),MEC 网络空间计算吞吐量的下界则可以表示为:

$$C \geq \rho \lambda_u \left(1 - e^{-\lambda_u \pi r_0^2}\right) \left(1 + e^{-\lambda_u \pi r_0^2} - e^{-\lambda_u \pi r_0^2} p_c^{(\text{low})} p_{\ell,s}^{(\text{low})}\right) + \lambda_u \left[ \left(1 - \rho\right) \left(1 - e^{-\lambda_u \pi r_0^2}\right) + e^{-\lambda_u \pi r_0^2} \right] \cdot I(\ell_{lc}^{(\text{max})} \geq \ell). \quad (9)$$

讨论1:观察公式(9),若当用户具备足够的本地能量和计算能力,即  $I(\ell_{lc}^{(\text{max})} \geq \ell) = 1$ ,使用计算卸载将不会带来对网络计算吞吐量的增益,  $C$  的下界将随着  $r_0$  和  $\rho$  的增加而线性减小。当用户本地能量和计算能力不足而选择计算卸载时,即  $I(\ell_{lc}^{(\text{max})} \geq \ell) = 0$ ,可以通过设计最优的  $r_0$  和  $\rho$  来最大化网络计算吞吐量,最优设计见第3节。

## 3 对 MEC 网络空间计算吞吐量的优化设计

当用户选择计算卸载时,即  $I(\ell_{lc}^{(\text{max})} \geq \ell) = 0$ ,通过设计最优的 MEC 服务范围  $r_0$  以及计算卸载比例  $\rho$  可以使 MEC 网络空间吞吐量  $C$  最大化。

### 3.1 优化 MEC 服务范围半径

当  $I(\ell_{lc}^{(\text{max})} \geq \ell) = 0$ ,根据公式(9),  $C$  的下界可简化为:

$$C \geq \rho \lambda_u \left(1 - e^{-\lambda_u \pi r_0^2}\right) \left(1 + e^{-\lambda_u \pi r_0^2} - e^{-\lambda_u \pi r_0^2} p_c^{(\text{low})} p_{\ell,s}^{(\text{low})}\right) = C^{(\text{low})}. \quad (10)$$

观察公式(10)并考虑优化  $r_0$  的物理含义。一方面,当  $r_0$  变小时(即每个 MEC 服务范围变小),每个 MEC 服务器接收到的卸载计算任务数量会减小,由此将缩短对每个任务的计算时延从而提升 MEC 成功概率;另一方面,当  $r_0$  变大时,每个用户将会被更多的 MEC 服务器所覆盖并有更大概率实现 MEC,由此 MEC 成功概率也会提升。因此,我们可以设计最优

的 $r_0$ 使 $C$ 最大化。

由于直接根据公式(10)来优化 $r_0$ 存在一定难度,首先考虑两种特殊情况下的优化设计,情况1是假设上行信道十分可靠时,以致卸载传输成功概率为1(即 $p_{\ell_s} = 1$ ),情况2是当MEC服务器计算能力很强时,以致卸载任务总能在规定时间内完成计算(即 $T_0 \rightarrow 0, p_c = 1$ )。

首先考虑情况1,当 $p_{\ell_s} = 1, C^{(low)}$ 表示为:

$$C^{(low)} = \rho \lambda_u \left(1 - e^{-c_1 r_0^*}\right) \left(1 + e^{-c_1 r_0^*} - e^{-c_1 r_0^{*(low)}}\right), \quad (11)$$

其中 $c_1 = \lambda_b \pi$ 。由于 $C^{(low)}$ 对 $r_0$ 是可微的,通过优化设计 $r_0$ 来最大化 $C^{(low)}$ 的问题可以表示为P1:

$$P1: \max_{r_0 \in \mathbb{R}^+} \rho \lambda_u \left(1 - e^{-c_1 r_0^*}\right) \left(1 + e^{-c_1 r_0^*} - e^{-c_1 r_0^{*(low)}}\right). \quad (12)$$

P1为简单的凸优化问题,对P1的最优解 $r_0^*$ 可通过求解等式(13)得到:

$$\left(r_0^* (1 - p_c^{(low)})' - 2p_c^{(low)} - 2\right) e^{c_1 r_0^* (1 - p_c^{(low)})} = \left(r_0^* (1 - p_c^{(low)})' - 2p_c^{(low)}\right) e^{c_1 r_0^* (2 - p_c^{(low)})} - 4, \quad (13)$$

其中, $p_c^{(low)} = c_2 (r_0^*)^{2c_1^{(max)} - 2} \exp(-c_2 r_0^*)$ ,  $(1 - p_c^{(low)})' = 2c_2 (r_0^*)^{2c_1^{(max)} - 3} \exp(-c_2 r_0^*) (c_2 (r_0^*)^2 - i^{(max)} + 1)$ ,  $c_1$ 和 $c_2$ 为常量,分别为 $c_1 = \frac{(\rho \lambda_u \pi)^{i^{(max)} - 1}}{\Gamma(i^{(max)})}$ 。求解等式(13)可以利用MATLAB等常用软件工具。

当MEC服务器密度极高时,即 $\lambda_b \rightarrow \infty$ ,P1可简化为P2:

$$P2: \max_{r_0 \in \mathbb{R}^+} \rho \lambda_u \left(1 - e^{-c_1 r_0^{*(low)}}\right). \quad (14)$$

对P2的最优解 $r_0^*$ 可表示为如下

$$\text{闭式解 } r_0^* = \left(\frac{i^{(max)}}{\rho \lambda_u \pi}\right)^{\frac{1}{2}}.$$

讨论2:观察 $r_0^*$ 的闭式解,发现 $r_0^*$ 会随着最大可产生的虚拟机数量 $i^{(max)}$ 的增加而变大,这是因为 $i^{(max)}$ 越大意味着MEC服务器计算能力越强,从而能提供更大范围的服务。同时, $r_0^*$ 会随着卸载用户密度 $\rho \lambda_u$ 的增加而变小,是因为用户卸载的计算任务越多,将会给MEC服务器带来更大计算压力,故需要缩小MEC服务范围来减小计算压力,用以保证一定的MEC成功概率。

接下来,考虑情况2。当 $p_c = 1, C^{(low)}$ 可表示为:

$$C^{(low)} = \rho \lambda_u \left(1 - e^{-c_1 r_0^*}\right) \left(1 + e^{-c_1 r_0^*} - e^{-c_1 r_0^{*(low)}}\right). \quad (15)$$

优化设计 $r_0$ 的问题可表示为P3:

$$P3: \max_{r_0 \in \mathbb{R}^+} \rho \lambda_u \left(1 - e^{-c_1 r_0^*}\right) \left(1 + e^{-c_1 r_0^*} - e^{-c_1 r_0^{*(low)}}\right). \quad (16)$$

对P3的最优解 $r_0^*$ 可通过求解等式(17)得到:

$$\left(2p_{\ell_s}^{(low)} + r_0^* (p_{\ell_s}^{(low)})' + 2\right) e^{c_1 (r_0^*)^2 (1 - p_{\ell_s}^{(low)})} = \left(2p_{\ell_s}^{(low)} + r_0^* (p_{\ell_s}^{(low)})'\right) e^{c_1 (r_0^*)^2 (2 - p_{\ell_s}^{(low)})} - 4, \quad (17)$$

其中, $p_{\ell_s}^{(low)} = \exp\left(-c_3 (r_0^*)^2 \left(1 - e^{-c_1 (r_0^*)^2}\right)\right)$ ,  $(p_{\ell_s}^{(low)})' = 2c_3 \left(r_0^* \left(e^{-c_1 (r_0^*)^2} - 1\right) - c_1 (r_0^*)^3 e^{-c_1 (r_0^*)^2}\right) \exp\left(-c_3 (r_0^*)^2 \left(1 - e^{-c_1 (r_0^*)^2}\right)\right)$ , 其中, $c_1 = \lambda_b \pi$ ,  $c_3 = \frac{2\pi\rho\lambda_b\theta^\alpha}{\alpha M} B\left(\frac{2}{\alpha}, 1 - \frac{2}{\alpha}\right)$ 。

当MEC服务器密度极高时(即

$\lambda_b \rightarrow \infty$ ),P3可简化为P4:

$$P4: \max_{r_0 \in \mathbb{R}^+} \lambda_u \left(1 - e^{-c_1 r_0^{*(low)}}\right). \quad (18)$$

对P4的最优解 $r_0^*$ 可表示为如下

$$\text{闭式解 } r_0^* = \left(\frac{2\pi\rho\lambda_b\theta^\alpha}{\alpha M} B\left(\frac{2}{\alpha}, 1 - \frac{2}{\alpha}\right)\right)^{-\frac{1}{2}}.$$

### 3.2 优化计算卸载比例

首先考虑通过优化 $\rho$ 来最大化MEC网络空间吞吐量 $C$ 的物理含义。太大或太小的 $\rho$ 都会降低 $C$ ,这是因为:太大的 $\rho$ 将会引起更严重的用户间干扰和卸载计算任务数量,导致MEC成功概率降低,从而降低 $C$ ;太小的 $\rho$ 会直接降低网络中卸载用户密度,导致 $C$ 的减小。因此,可以通过优化 $\rho$ 来最大化 $C$ 。

类似3.1节的步骤,首先考虑当上行信道十分可靠时(即 $p_{\ell_s} = 1$ ), $C^{(low)}$ 可表示为:

$$C^{(low)} = \rho \lambda_u \left(1 - e^{-c_4}\right) \left(1 + e^{-c_4} - e^{-c_4 p_c^{(low)}}\right), \quad (19)$$

其中 $c_4 = \lambda_b \pi r_0^2$ 。由于 $C^{(low)}$ 对 $\rho$ 是可微的,通过优化设计 $\rho$ 来最大化 $C^{(low)}$ 的问题可以表示为P5:

$$P5: \max_{\rho \in [0,1]} \rho \left(1 + e^{-c_4} - e^{-c_4 p_c^{(low)}}\right), \quad (20)$$

P5为凸优化问题,对P5的最优解 $\rho^*$ 可通过求解等式(21)得到:

$$\rho^* c_4 (p_c^{(low)})' e^{c_4 (1 - p_c^{(low)})} = e^{c_4 (1 - p_c^{(low)})} - 2, \quad (21)$$

其中, $p_c^{(low)} = c_6 (\rho^*)^{i^{(max)} - 1} e^{-c_5 \rho^*}$ ,  $(p_c^{(low)})' = c_6 (\rho^*)^{i^{(max)} - 2} e^{-c_5 \rho^*} (i^{(max)} - c_5 \rho^* - 1)$ ,  $c_4 = \lambda_b \pi r_0^2$ ,  $c_5 = \lambda_u \pi r_0^2$ , 以及 $c_6 = \frac{c_5^{i^{(max)} - 1}}{\Gamma(i^{(max)})}$ 。

为了得到物理含义更明显的结果,利用不等式 $1 - e^{-x} \leq x$ ,对公式(19)进行简化,可得到 $C^{(low)}$ 的近似结

果,  $C^{(low)} \approx \rho \lambda_u (1 - e^{-c_4}) (c_4 \rho_c^{(low)} + e^{-c_4})$ 。

由此,当 MEC 服务器密度极高时(即  $\lambda_b \rightarrow \infty$ ),最优化问题可设计为 P6:

$$P6: \max_{\rho \in [0,1]} c_4 \rho \rho_c^{(low)} \quad (22)$$

对 P6 的最优解  $\rho^*$  为如下闭式

$$\text{解 } \rho^* = \frac{i^{(\max)}}{\lambda_u \pi r_0^2}$$

讨论3:观察  $\rho^*$  闭式解,当  $i^{(\max)}$  增加时,表示 MEC 服务器计算能力强,便可提高  $\rho^*$  以加大卸载用户数目,来提高网络计算吞吐量;当用户密度  $\lambda_u$  增加时,则应减小  $\rho^*$  以减轻 MEC 服务器计算压力,从而保证一定的 MEC 成功概率。

接下来,考虑当 MEC 服务器计算能力很强(即  $\rho_c = 1$ ),  $C^{(low)}$  可表示为:

$$C^{(low)} = \rho \lambda_u (1 - e^{-c_4}) (1 + e^{-c_4} - e^{-c_4 \rho_{\ell,s}^{(low)}}) \quad (23)$$

基于公式(23),对于  $\rho^*$  的最优化问题可设计为 P7:

$$P7: \max_{\rho \in [0,1]} \rho (1 + e^{-c_4} - e^{-c_4 \rho_{\ell,s}^{(low)}}) \quad (24)$$

对 P7 的最优解  $\rho^*$  可通过求解等

式(25)得到:

$$\rho^* c_4 (p_{\ell,s}^{(low)})' e^{c_4(1-p_{\ell,s}^{(low)})} = e^{c_4(1-p_{\ell,s}^{(low)})} - 2, \quad (25)$$

其中,  $p_{\ell,s}^{(low)} = e^{-c_7 \rho^*}$ ,  $(p_{\ell,s}^{(low)})' = -c_7 e^{-c_7 \rho^*}$ ,

$$c_7 = \frac{2c_5(1 - e^{-c_4})\theta^\alpha}{\alpha M} B\left(\frac{2}{\alpha}, 1 - \frac{2}{\alpha}\right)$$

当 MEC 服务器密度极高时(即  $\lambda_b \rightarrow \infty$ ), P7 可简化为如下优化问题:

$$\max_{\rho \in [0,1]} c_4 \rho p_{\ell,s}^{(low)}, \quad (26)$$

其最优解  $\rho^*$  可表示为如下闭式

$$\text{解 } \rho^* = \frac{\alpha M}{2\lambda_u \pi r_0^2 \theta^\alpha B\left(\frac{2}{\alpha}, 1 - \frac{2}{\alpha}\right)}$$

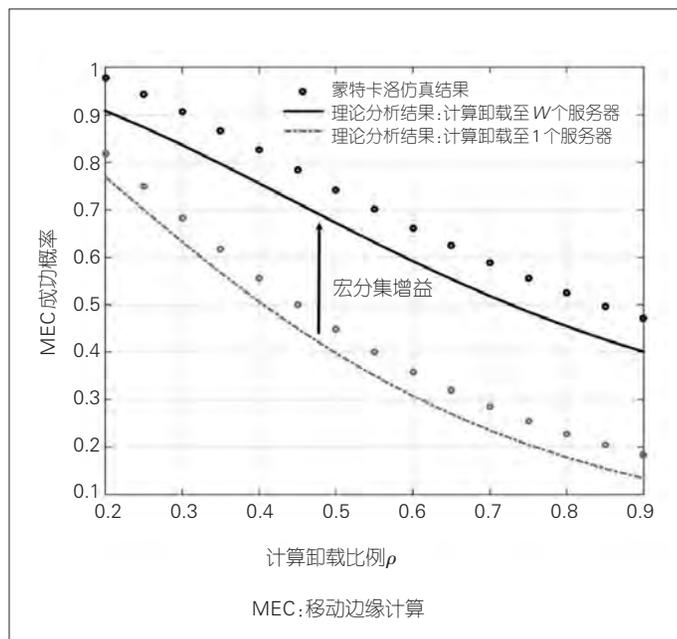
### 4 仿真结果

在本节中,我们利用 MATLAB 仿真对上文中得到的理论结果加以验证,主要仿真参数设置如下:  $\lambda_b = 0.01/\text{m}^2$ ,  $\lambda_u = 0.1/\text{m}^2$ ,  $r_0 = 8 \text{ m}$ ,  $\rho = 0.7$ ,  $\theta = 10 \text{ dB}$ ,  $B = 3 \text{ kHz}$ ,  $\alpha = 3$ ,  $T_s = 100 \text{ ms}$ ,  $\ell = 10^3 \text{ bits}$ ,  $T_0 = 1 \text{ ms}$ ,  $d = 0.3$ 。其中,

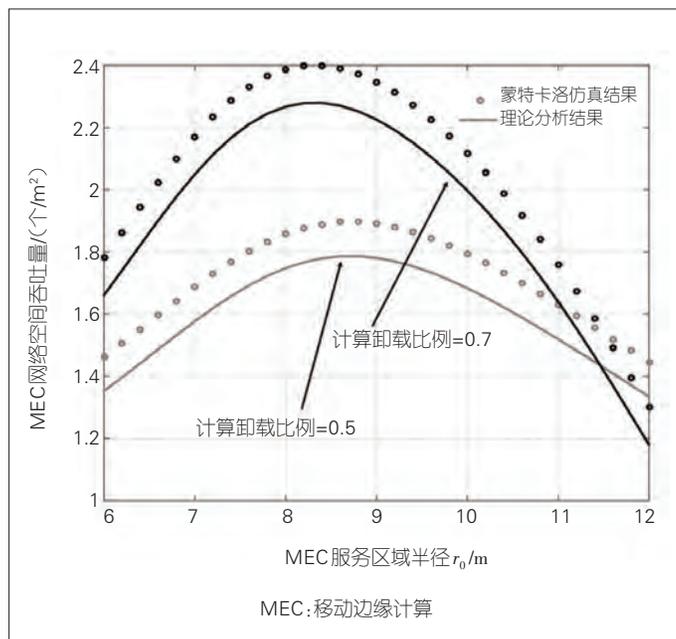
蒙特卡洛仿真结果由圆圈表示,理论分析结果由实/虚线表示。限于篇幅,这里只展示最重要的3个仿真结果。

图3展示了典型用户的 MEC 成功概率,其中,虚线表示用户将计算任务只卸载到任意一个服务器时 MEC 成功概率,实线则表示用户将计算任务卸载到附近的  $W$  个服务器时的 MEC 成功概率,即公式(8)。首先,文中得到的理论结果(下界)与仿真结果之间的差值较小,这证明理论结果比较准确;其次,可以观察到,相比选择一个 MEC 服务器进行计算卸载,当用户选择向  $W$  个服务器同时进行计算卸载时的 MEC 成功概率有明显提升,这得益于宏分集增益。

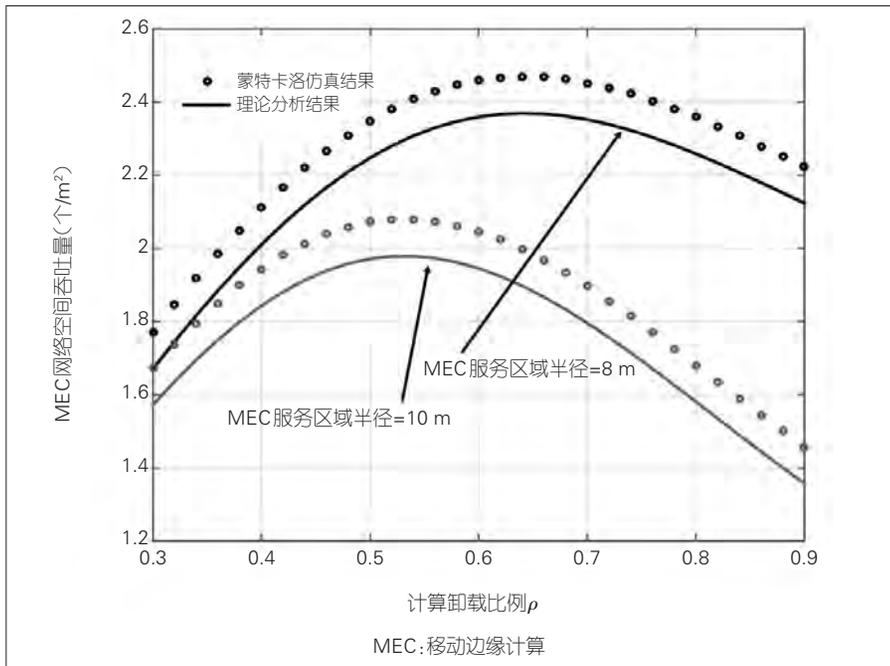
图4展示了通过优化  $r_0$  来最大化 MEC 网络吞吐量  $C$ 。首先,文中我们所求得的  $C^{(low)}$ (下界)相比  $C$  仅有少量差值,这表示理论下界较为准确;其次,当给定  $\rho$  的值,  $C$  及  $C^{(low)}$  是变量  $r_0$  的凹函数,因此可以通过设计最优的  $r_0^*$  来最大化  $C$ ,如当  $\rho$  在  $0.5 \sim 0.7$  之间时,  $r_0^*$  在  $8 \sim 9 \text{ m}$  之间。另外,当提高  $\rho$  的值,  $C$  的最大值会随之变大,这是因



▲图3 MEC成功概率与计算卸载比例关系图



▲图4 MEC网络空间吞吐量与MEC服务区域半径关系图



▲图5 MEC成功概率与计算卸载比例关系图

为增大 $\rho$ 意味着更多的用户选择计算卸载,从而有效增大网络吞吐量。

图5展示了通过优化 $\rho$ 来最大化MEC网络吞吐量 $C$ 。给定 $r_0, C$ 及 $C^{(low)}$ 同样是变量 $\rho$ 的凹函数。当 $r_0$ 在8~10m之间时,最优值 $\rho^*$ 在0.5~0.7间。另外,当 $r_0$ 增大时,MEC服务范围将扩大,更多用户的计算任务可卸载至服务器;因此,网络吞吐量也会增加。

### 5 结束语

本文中,我们首次定义了大规模MEC网络中的空间计算吞吐量这一性能指标,并通过优化设计MEC服务范围半径 $r_0$ 以及用户计算卸载比例 $\rho$ 这两个指标,实现MEC网络空间吞吐量的最大化。所提供的理论分析与优化结果将为部署大规模MEC网络提供了极为重要的设计参考。

#### 参考文献

[1] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading [J]. IEEE communications surveys and tutorials, 2017, 19(3): 1628–1656. DOI: 10.1109/comst.2017.2682318

[2] WU D, WANG F, CAO X, et al. Joint communication and computation optimization for wireless powered mobile edge computing with D2D offloading [J]. Journal of communications and information networks, 2019, 4(4): 72–86

[3] 马洪源. 面向5G的边缘计算及部署思考 [J]. 中兴通讯技术, 2019, 25(3): 77–81. DOI: 10.12142/ZTETJ.201903011

[4] 丁春涛, 曹建农, 杨磊, 等. 边缘计算综述:应用、现状及挑战 [J]. 中兴通讯技术, 2019, 25(3): 2–7. DOI: 10.12142/ZTETJ.201903001

[5] YANG X, HUA S, SHI Y, et al. Sparse optimization for green edge AI inference [J]. Journal of communications and information networks, 2020, 5(1): 1–15

[6] QIN M, CHEN L, ZHAO N, et al. Power-constrained edge computing with maximum processing capacity for IoT networks [J]. IEEE Internet of Things journal, 2019, 6(3): 4330–4343. DOI:10.1109/jiot.2018.2875218

[7] QIN M, CHEN L, ZHAO N, et al. Computing and relaying: utilizing mobile edge computing for P2P communications [J]. IEEE transactions on vehicular technology, 2020, 69(2): 1582–1594. DOI:10.1109/tvt.2019.2956996

[8] YOU C, HUANG K, CHAE H, et al. Energy-efficient resource allocation for mobile-edge computation offloading [J]. IEEE transactions on wireless communications, 2017, 16(3): 1397–1411. DOI:10.1109/twc.2016.2633522

[9] MAO Y, ZHANG J, LETAIEF K. Dynamic computation offloading for mobile-edge computing with energy harvesting devices [J]. IEEE communications surveys and tutorials, 2016, 14(12): 3590–3605. DOI:10.1109/jsac.2016.2611964

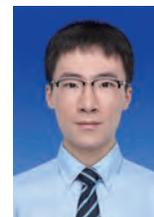
[10] ZHAO J, LI Q, GONG Y, et al. Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks [J]. IEEE transactions on vehicular technology, 2019, 68(8): 7944–7956. DOI: 10.1109/tvt.2019.2917890

[11] KO S, HAN K, HUANG K. Wireless networks for mobile edge computing: spatial modeling and latency analysis [J]. IEEE transactions on wireless communications, 2018, 17(8): 5225–5240. DOI:10.1109/twc.2018.2840120

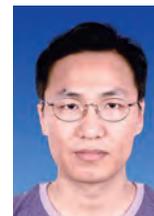
[12] HAENGGI M. Stochastic geometry for wireless networks [M]. Cambridge: Cambridge University Press, 2009. DOI: 10.1017/cbo9781139043816

[13] BRUNEO D. A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems [J]. IEEE transactions on parallel and distributed system, 2014, 25(3): 560–569. DOI:10.1109/tpds.2013.67

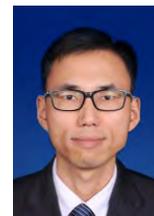
### 作者简介



**韩凯峰**, 中国信息通信研究院政策与经济研究所工程师, 工学博士; 主要研究领域为无线通信(5G/6G)、边缘计算、车联网、人工智能等前沿技术及战略; 已发表论文15篇。



**胡昌军**, 中国信息通信研究院政策与经济研究所战略研究部副主任、高级工程师, 工程硕士, 工信部《新一代人工智能产业创新重点任务揭榜工作方案》起草团队的主要参与者; 主要研究领域为人工智能、5G等ICT前沿领域技术及战略; 曾先后参加工信部电子发展基金等多项课题研究。



**刘铁志**, 中国信息通信研究院政策与经济研究所战略研究部主任、高级工程师, 经济学博士, 曾担任《国务院关于加快构建大众创业万众创新支撑平台的指导意见》、《互联网+人工智能三年行动实施方案》、工信部《促进新一代人工智能产业发展三年行动计划(2018–2020年)》《关于促进人工智能和实体经济深度融合的指导意见》等起草专家团队的主要负责人; 主要研究领域为信息通信产业政策和国际经济。