

引用格式:丁雪乾,薛建彬. 边缘计算下基于 Lyapunov 优化的系统资源分配策略[J]. 微电子学与计算机, 2020, 37(2): 63-68.
[DING X Q, XUE J B. System resource allocation strategy based on lyapunov optimization in edge computing[J]. Microelectron-
ics & Computer, 2020, 37(2): 63-68.]

边缘计算下基于 Lyapunov 优化的系统资源分配策略

丁雪乾¹, 薛建彬^{1,2}

(1 兰州理工大学 计算机与通信学院, 甘肃 兰州 730050;

2 东南大学 移动通信国家重点实验室, 江苏 南京 210096)

摘 要: 针对密集的计算任务卸载使得移动边缘计算系统网络开销大以及应用与服务交互延迟高等问题, 提出一种基于 Lyapunov 优化的资源分配策略. 该策略首先在保证用户服务质量需求的情况下, 考虑系统数据队列状态和任务执行开销, 建立系统开销模型; 然后运用 Lyapunov 优化理论设计了一种基于用户设备和边缘计算服务提供商数据队列状态的实时分配方案, 并通过构造拉格朗日乘子函数进行迭代优化. 实验结果表明, 所提方案在保证系统队列稳定性的同时降低了系统平均总开销.

关键词: 移动边缘计算; 任务卸载; Lyapunov 优化; 资源分配; 稳定性

中图分类号: TN929. 5; TP393

文献标识码: A

文章编号: 1000—7180(2020)02—0063—06

System resource allocation strategy based on lyapunov optimization in edge computing

DING Xue-qian¹, XUE Jian-bin^{1,2}

(1 School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China;

2 National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

Abstract: In view of the problem about high network overhead and high interaction delay between applications and services due to intensive computation task offloading in mobile edge computing system, a resource allocation strategy based on Lyapunov optimization is proposed. Firstly, the overhead model of the system is established considering system's data queue and task execution overhead while ensuring the user's service quality requirements. Then, a real-time allocation scheme based on user's devices and edge computing service provider data queue is designed by using Lyapunov optimization theory, and it is iteratively optimized by constructing a Lagrangian multiplier function. Experimental results show that the proposed scheme reduces the average overhead of system while ensuring the the queue stability.

Key words: mobile egde computing; task offloading; Lyapunov optimization; resource allocation; stability

1 引言

随着 5G 的迅速推进, 移动智能设备及其多样化的任务需求急剧增长, 基于云服务中心控制的传统移动云计算 (Mobile Cloud Computing, MCC) 已

难以满足越来越高的通信需求^[1]. 因此, 移动边缘计算 (Mobile Edge Computing, MEC) 技术通过任务迁移, 将传统云服务中心的部分计算、存储能力“下沉”到网络边缘, 实现了在更加靠近用户侧的无线接入网内提供 IT 服务和云计算功能, 极大地缩短了

应用与核心服务器的链路传输,同时降低了系统时延^[2-3].然而,相对于云服务而言,边缘端的频谱、物理资源依然是有限的^[4-5],密集的任务卸载容易造成边缘服务资源短缺和网络拥塞等问题.因此,如何在保证卸载过程中应用需求的前提下,找出一种最优的 MEC 资源分配策略具有重要的研究价值.在文献^[6]中,Liu Y 等人提出了边缘云多资源分配策略,以应对资源密集型和延迟敏感型移动应用的任务卸载.文献^[7]提出了一种激励-拍卖机制(Incentive-Compatible Auction Mechanism, ICAM)作为用户设备(买方)与服务提供商(卖方)之间的资源交易,通过定价策略,ICAM 可以有效地分配边缘云资源,以满足移动设备的处理需求.

尽管上述关于边缘计算下的系统资源配置研究有独特解决方案,但仍存在一定的局限性.如在文献^[6-8]中,就要求接边缘控制中心能够准确地获取卸载链路中的信道状态信息,而文献^[9]却未考虑无线资源的分配问题.为此,本文结合计算任务卸载和边缘端服务提出一种边缘系统资源分配策略,通过应用 Lyapunov 优化定理将时间变量约束目标转化为队列稳定性问题,可以离线统计获取应用任务的数据信息,避免了实时获取信息的额外开销.

2 系统模型与方案规划

2.1 系统模型

假设无线接入网内存在个 N 用户设备,即 $i \in \{1, 2, \dots, N\}$.边缘计算服务提供商(Cloud Service Provider, CSP)在靠近用户的基站侧部署网络资源,即提供边缘化计算服务,从而提升单核用户设备对于延迟敏感或计算密集型任务的处理能力.并且, CSP 通过单跳的网络传输就可以与基站建立通信,形成“设备-基站-边缘侧服务”的 MEC 网络场景,如图 1 所示.

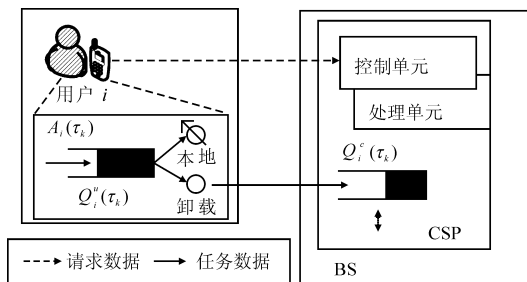


图 1 系统场景

本文研究 MEC 场景下任务卸载过程中的无线、计算资源分配问题,旨在最小化边缘网络系统的

平均成本.将移动设备及边缘端缓冲区的数据任务表示为不同的数据队列,基于队列模型,系统的执行时间看作由一系列连续的时间片 τ_k 构成, $k \in \{1, 2, \dots, K\}$.在每个时间片起始,移动设备通过接入无线链路向 CSP 发送任务卸载请求信息(包括任务数据大小、用户最大容忍时延等特性),作为边缘端的服务中心,当 CSP 控制单元接收到请求信息后进行分析,并将执行决策信息通过链路反馈到用户设备控制器.本文假设在一个执行时间片内用户 i 的任务是连续不可分的,用 α_i 来表示用户设备收到的决策变量,当 $\alpha_i = 1$ 时,表示用户 i 将任务卸载到边缘服务器上执行,反之,当 $\alpha_i = 0$ 时,表示由设备本地 CPU 处理.

2.1.1 通信模型

相比于卸载到边缘端的任务数据,设备发送的请求信息数据量非常小,因此,通常 MEC 系统中不再讨论发送与接收此信息所产生的能耗与延迟.而在任务卸载时, CSP 需要从网络运营商处租赁频谱资源从而为每个用户分配一定的卸载传输带宽,令 $x_i(\tau_k) \in [0, 1]$ 为用户 i 获得的带宽比例,则该用户进行计算卸载时的瞬时传输速率可记作为:

$$r_i(\tau_k) = x_i(\tau_k)B \log_2(1 + S_i) \quad (1)$$

式中, B 表示系统带宽; S_i 表示设备 i 在无线链路间的信干噪比.且分配的带宽量不能大于系统总量:

$$\sum_{i \in N} \alpha_i(\tau_k) x_i(\tau_k) \leq 1 \quad (2)$$

为确保用户应用时延需求,定义其卸载传输所需要满足的最低速率为:

$$r_i(\tau_k) \geq r_n, \forall i \in N \quad (3)$$

2.1.2 计算模型

在执行时间片内,将用户设备产生的数据任务记作 $A_i(\tau_k)$,即为时间片 τ_k 内设备 i 的任务到达比特数,且到达过程服从泊松分布.当任务在用户本地执行时,定义单个设备的 CPU 周期频率为 f_u ,并将处理器执行单位比特任务量所需的 CPU 运行周期定义为任务处理密度 ρ .用户设备实时产生的计算任务以数据队列的形式进入设备缓冲区,以便进行数据处理或卸载到边缘服务器上执行.用 $Q_i^L(\tau_k)$ 代表时间片 τ_k 内设备端缓冲区的任务队列量,则在下一个时间片队列的更新过程可表示为:

$$Q_i^L(\tau_{k+1}) = \left[Q_i^L(\tau_k) - (1 - \alpha_i(\tau_k)) \frac{f_u}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) + A_i(\tau_k) \right]^+ \quad (4)$$

式中, $[x]^+ = \max\{0, x\}$; τ_{k+1} 表示时间片的更新

状态.

当用户将数据任务卸载到边缘端时, CSP 接收任务数据同样以队列的形式置入边缘服务缓冲池中. 相比于用户设备, CSP 提供的边缘计算服务拥有较大执行能力, 表示为 f_c , 并在执行过程中为不同用户进行分配, 以参量 $y_i(\tau_k)$ 来表示在时间片 τ_k 下 CSP 分配给设备 i 的计算资源, 即 $y_i(\tau_k) \in [0, 1]$. 将时间片 τ_k 内 CSP 端缓冲区的任务队列量用 $Q_i^c(\tau_k)$ 表示, 则在下个时间片队列的更新表达式可表示为:

$$Q_i^c(\tau_{k+1}) = \left[Q_i^c(\tau_k) - y_i(\tau_k) \frac{f_c}{\rho} + \alpha_i(\tau_k) r_i(\tau_k) \right]^+ \quad (5)$$

同样地, 分配的计算资源量不能大于边缘端系统总量:

$$\sum_{i \in N} \alpha_i(\tau_k) y_i(\tau_k) \leq 1 \quad (6)$$

根据队列稳定性定理, 队列表达式(4)、(5)需满足如下有界性约束:

$$\limsup_{k \rightarrow \infty} \frac{1}{K} \sum_{k=0}^K E[Q_i^u(\tau_k) + Q_i^c(\tau_k)] < \infty \quad (7)$$

由此, 根据利特尔定律可知, 系统中任务执行中的时间约束得以保证.

2.2 方案规划

不同于云计算环境下的资源调度, MEC 场景中的资源分配更多地考虑边缘服务开销与服务质量的结合. CSP 在租赁频谱资源的同时将物理资源和无线资源抽象化为虚拟服务资源块为用户进行分配. 对于差异化的数据任务特性和用户服务尺度, 以 ϵ 表示频谱资源支出的单位费用, ω 表示单位计算成本费用. 因此, 可将 CSP 在用户 i 下的无线资源和计算资源支出分别表示为 $\epsilon x_i(\tau_k) B$ 和 $\omega y_i(\tau_k) f_c$. 因此, CSP 在时间片 τ_k 下的瞬时开销可表示为:

$$U_{\text{CSP}} = \sum_{i \in N} [u(\alpha_i(\tau_k) \epsilon x_i(\tau_k) B) + u(\alpha_i(\tau_k) \omega y_i(\tau_k) f_c)] \quad (8)$$

式中, 使用递增凸函数 $u(x) = \log(x)$, $x > 0$ 作为系统开销函数. 由于本文研究一种动态服务资源分配策略, 在保证系统任务队列稳定性和用户服务质量(Quality of Service, QoS)性能要求下, 最小化边缘网络的平均成本开销. 因此, 资源分配方案可表示为如下的数学模型:

$$\begin{aligned} \min_{x, y, u} \bar{U}_{\text{CSP}} &= \limsup_{k \rightarrow \infty} \frac{1}{K} \sum_{i \in N} [u(\alpha_i(\tau_k) \epsilon x_i(\tau_k) B) + \\ &\quad u(\alpha_i(\tau_k) \omega y_i(\tau_k) f_c)] \\ \text{s. t. } &(2), (3), (6), (7) \end{aligned} \quad (9)$$

式(9)为时间平均意义下的随机优化问题, 且执行开销与当前时间片内的任务产生与缓冲量有关, 并且在无法获取信道状态变化下的数据信息时, 可以利用 Lyapunov 优化方法来进行进一步分析该问题.

3 算法描述

基于系统模型的分析, 需要考虑: 如何在保证用户延迟性能下, 对于这种相互耦合的队列, 以及状态依赖的动态过程进行稳态分析, 进而解决式(9)中的目标问题. 据此, 本文应用 Lyapunov 优化方法分析用户设备和 CSP 端的数据积压队列, 在保证用户 QoS 约束下, 通过 Lyapunov 漂移加罚(Lyapunov drift-plus-penalty)函数推导出含有目标问题下的上界值, 然后, 利用拉格朗日乘子迭代对资源的分配方案进行控制协调, 从而得到最佳的分配策略.

3.1 基于 Lyapunov 优化的理论分析

基于 Lyapunov 优化的理论分析方法是一种分析和解决系统稳定性、多目标优化的经典方法, 其思想主要是通过控制系统稳定性来实现性能的优化. 令系统整体的队列状态向量为 $Q_i(\tau_k) = (Q_i^u(\tau_k), Q_i^c(\tau_k))$, 针对数据积压队列的稳定性, 及队列更新表达式(4)、(5), 定义如下二次型 Lyapunov 函数:

$$L(Q_i(\tau_k)) = \frac{1}{2} \left[\sum_{i \in N} (Q_i^u(\tau_k))^2 + \sum_{i \in N} (Q_i^c(\tau_k))^2 \right] \quad (10)$$

上述函数通过对用户设备和边缘服务端队列的平方求和, 表示出在 τ_k 时刻系统中各队列的残余量, 该函数值越大, 则说明系统中积压了较多待执行的数据, 在一定程度上代表了当前系统执行的时延性能. 则在时间片 τ_k 下的单步 Lyapunov 漂移(Lyapunov drift) $\Delta(Q_i(\tau_k))$ 可表示为:

$$\Delta(Q_i(\tau_k)) = E[L(Q_i(\tau_{k+1})) - L(Q_i(\tau_k)) | Q_i(\tau_k)] \quad (11)$$

以 $A_i^{\max}(\tau_k)$ 代表任务最大到达量, $r_i^{\max}(\tau_k)$ 代表最大数据传输速率, 与 f_u, f_c, ρ 均为不含控制变量的常量, 令 $C = \frac{1}{2} \sum_{i \in N} \left[\left(\frac{f_u}{\rho} \right)^2 + 2 (r_i^{\max}(\tau_k))^2 + (A_i^{\max}(\tau_k))^2 + \left(\frac{f_c}{\rho} \right)^2 \right]$, 根据类比不等式 $([x]^+)^2 \leq x^2$, 可以计算得到(11)式中 $L(Q_i(\tau_{k+1})) - L(Q_i(\tau_k))$ 满足条件:

$$\begin{aligned} L(Q_i(\tau_{k+1})) - L(Q_i(\tau_k)) &\leq C - \\ &\sum_{i \in N} \left[\left((1 - \alpha_i(\tau_k)) \frac{f_u}{\rho} + \alpha_i(\tau_k) r_i(\tau_k) - A_i(\tau_k) \right) Q_i^u(\tau_k) \right] - \end{aligned}$$

$$\sum_{i \in N} \left[\left(y_i(\tau_k) \frac{f_c}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) \right) Q_i^c(\tau_k) \right] \quad (12)$$

根据最小化目标函数及 Lyapunov 优化的设计准则,可以将原目标问题转化为最小化每个时间片 τ_k 内的 Lyapunov 漂移加罚(drift-plus-penalty) 函数的问题.联合(12) 式可得该函数:

$$\begin{aligned} & \Delta(Q_i(\tau_k)) + VE[U_i(\tau_k) | Q_i(\tau_k)] \leq C - \\ & E \left\{ \sum_{i \in N} \left[\left((1 - a_i(\tau_k)) \frac{f_u}{\rho} + a_i(\tau_k) r_i(\tau_k) - A_i(\tau_k) \right) Q_i^u(\tau_k) \right] \middle| Q_i(\tau_k) \right\} - \\ & E \left\{ \sum_{i \in N} \left[\left(y_i(\tau_k) \frac{f_c}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) \right) Q_i^c(\tau_k) \right] \middle| Q_i(\tau_k) \right\} + \\ & VE[U_i(\tau_k) | Q_i(\tau_k)] \end{aligned} \quad (13)$$

从而可得, Lyapunov 漂移加罚函数存在上界值:

$$\begin{aligned} & \max \{ \Delta(Q_i(\tau_k)) + VE[U_i(\tau_k) | Q_i(\tau_k)] \} = \\ & C + VE[U_i(\tau_k) | Q_i(\tau_k)] - \\ & E \left\{ \sum_{i \in N} \left[\left((1 - a_i(\tau_k)) \frac{f_u}{\rho} + a_i(\tau_k) r_i(\tau_k) - A_i(\tau_k) \right) Q_i^u(\tau_k) \right] - \right. \\ & \left. \left(\left(y_i(\tau_k) \frac{f_c}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) \right) Q_i^c(\tau_k) \right) \right] \middle| Q_i(\tau_k) \right\} \end{aligned} \quad (14)$$

式中, V 是单位转换系数,其作用在于权衡最优的目标函数和系统队列的稳定性,同时在实际中反映了决策者对于系统队列积压和成本支出的“着重程度”.由此,在任意的队列状态向量 $Q_i(\tau_k)$ 和权衡系数 V 下,最优的资源分配方案问题可表示为如下数学模型:

$$\begin{aligned} & \min_{X,Y,U} VU_i(\tau_k) - \\ & \sum_{i \in N} \left[\left((1 - a_i(\tau_k)) \frac{f_u}{\rho} + a_i(\tau_k) r_i(\tau_k) - A_i(\tau_k) \right) Q_i^u(\tau_k) \right] - \\ & \sum_{i \in N} \left[\left(y_i(\tau_k) \frac{f_c}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) \right) Q_i^c(\tau_k) \right] \\ & s. t. (2), (3), (6), (7) \end{aligned} \quad (15)$$

3.2 资源分配优化

由于方案(15)中包含除连续变量外的 0-1 卸载变量 $\alpha_i(\tau_k)$,属于 NP 难问题,本文根据松弛理论将该二元变量松弛至 $[0,1]$,即令 $\tilde{x}_i(\tau_k) = \alpha_i(\tau_k)x_i(\tau_k)$, $\tilde{y}_i(\tau_k) = \alpha_i(\tau_k)y_i(\tau_k)$. 则 $\tilde{x}_i(\tau_k)$ 、 $\tilde{y}_i(\tau_k)$ 分别表示 CSP 进行任务卸载时的无线、计算资源消耗比例.因此,该方案模型与约束条件都是线性的,是一个凸优化问题,可以应用拉格朗日乘子法求解.联合约束条件可得到对应的拉格朗日函数为:

$$\begin{aligned} & L(X,Y,\eta,\beta,\sigma) = VU_i(\tau_k) - \\ & \sum_{i \in N} \left[\left((1 - \alpha_i(\tau_k)) \frac{f_u}{\rho} + \tilde{r}_i(\tau_k) - A_i(\tau_k) \right) Q_i^u(\tau_k) \right] - \end{aligned}$$

$$\begin{aligned} & \sum_{i \in N} \left[\left(\tilde{y}_i(\tau_k) \frac{f_c}{\rho} - \tilde{r}_i(\tau_k) \right) Q_i^c(\tau_k) \right] + \\ & \eta \left(\sum_{i \in N} \tilde{x}_i(\tau_k) - 1 \right) + \beta \left(\sum_{i \in N} \tilde{y}_i(\tau_k) - 1 \right) + \sigma (r_n - \tilde{r}_i(\tau_k)) \end{aligned} \quad (16)$$

式中, η, β 和 σ 表示拉格朗日乘子.通过 KKT 条件,可以得到最优的资源分配为:

$$\tilde{x}_i(\tau_k)^* = \frac{V}{\ln 10} \cdot \frac{1}{[r_i'(Q_i^u(\tau_k) + Q_i^c(\tau_k) + \sigma^n) - \eta^n]} \quad (17)$$

$$\tilde{y}_i(\tau_k)^* = \frac{\rho V}{\ln 10 [Q_i^c(\tau_k) f_c - \rho \beta^n]} \quad (18)$$

式中, $r_i' = B \log_2(1 + S_i)$, n 为步长.通过对拉格朗日乘子 η, β, σ 不断更新,可以得到 CSP 为卸载用户分配最优的无线与计算资源,具体资源分配优化算法步骤如下:

表 1 CSP 资源分配优化算法

步骤	内容
1	初始化: $k \leftarrow 0$, $A_i(\tau_k) \leftarrow 0$, $Q_i(\tau_k) \leftarrow 0$, $\forall i \in N$
2	初始化乘子变量: $\eta^0, \beta^0, \sigma^0$; 最大允许误差 $\Delta; n$.
3	for $k = 0, 1, \dots, K^{\max} - 1$, do
4	将当前的乘子变量 $\eta^0, \beta^0, \sigma^0$ 代入(17)、(18)求得局部最优分配策略 $\tilde{x}_i(\tau_k), \tilde{y}_i(\tau_k)$.
5	将 $\tilde{x}_i(\tau_k), \tilde{y}_i(\tau_k)$ 代入拉格朗日乘子更新方程,更新 $\eta^n, \beta^n, \sigma^n$
6	更新队列方程: $Q_i^u(\tau_{k+1}) = [Q_i^u(\tau_k) - (1 - \alpha_i(\tau_k)) \frac{f_u}{\rho} - \alpha_i(\tau_k) r_i(\tau_k) + A_i(\tau_k)]^+$ $Q_i^c(\tau_{k+1}) = [Q_i^c(\tau_k) - y_i(\tau_k) \frac{f_c}{\rho} + \alpha_i(\tau_k) r_i(\tau_k)]^+$
7	if $\ L^{n+1}(\tau_k) - L^n(\tau_k)\ \geq \Delta$ do
8	$n = n + 1$
9	返回步骤 3
10	else
11	输出 $Q_i(\tau_k), \tilde{x}_i(\tau_k)^*, \tilde{y}_i(\tau_k)^*$
12	end if
13	end for

3.3 性能分析

设用户设备产生的平均数据任务 $E[A_i(\tau_k)] = \lambda_i$,并且严格在 MEC 系统处理能力范围内: $\lambda_i + e \leq \Lambda$,其中 e 为非零常数, Λ 为处理区域.则通过本文算法“CSP 资源分配优化算法”获得的离线最优分配策略为 $\{\tilde{x}_i(\tau_k)^*, \tilde{y}_i(\tau_k)^*\}$,由于该策略独立于数据队列积压,基于随机优化理论,对于 Lyapunov 漂移加罚函数有:

$$\Delta L(Q_i(\tau_k)) + VE[U_i(\tau_k)] \leq C + VE[U_i^*(\tau_k)] - e \sum_{i \in N} [Q_i^c(\tau_k) + Q_i^u(\tau_k)] \quad (19)$$

对上式两边分别取期望值,同时对 $k = 0, 1, \dots, K-1$ 进行累加得到:

$$\begin{aligned} E[L(\tau_k) - L(\tau_0)] + V \sum_{k=0}^{K-1} \sum_{i \in N} E[U_i(\tau_k)] + \\ e \sum_{k=0}^{K-1} \sum_{i \in N} E[Q_i^c(\tau_k) + Q_i^u(\tau_k)] \\ \leq KC + V \sum_{k=0}^{K-1} \sum_{i \in N} E[U_i^*(\tau_k)] = \\ KC + V \sum_{k=0}^{K-1} \sum_{i \in N} E[U_i^*(\lambda + e)] \quad (20) \end{aligned}$$

对不等式两边分别除以 eK , 以及忽略其余非负项可以得到, 在本文算法控制下的任务队积压满足:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=0}^{K-1} E\left[\sum_{i \in N} (Q_i^c(\tau_k) + Q_i^u(\tau_k))\right] \leq \frac{C + VU_i^*(\lambda + e)}{e} \quad (21)$$

其次, 由于式(20)中

$$\begin{aligned} e \sum_{k=0}^{K-1} \sum_{i \in N} E[Q_i^c(\tau_k) + Q_i^u(\tau_k)] \geq 0, \text{ 同样有:} \\ V \sum_{k=0}^{K-1} \sum_{i \in N} E[U_i(\tau_k)] \leq KC + KV \sum_{i \in N} E[U_i^*(\lambda + e)] \quad (22) \end{aligned}$$

对上式两边同时除以 KV , 以及令 $K \rightarrow \infty$ 可以得到, 在本文算法控制下的 CSP 成本开销满足:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=0}^{K-1} E[U_i(\tau_k)] \leq \sum_{i \in N} U_i^*(\lambda + e) + \frac{C}{V} \quad (23)$$

4 实验结果

为了验证上述理论, 本文基于 Matlab R2016a 软件平台进行仿真实验, 仿真中其他参数设置如表 2 所示。本文考虑配置 CSP 的基站部署在中心, 多个卸载用户在半径为 200 m 的范围内随机分布, 仿真实验执行 1 000 个时间片, 每个时间片长度设置为 5 ms。

(1) 有效性评估

图 2 给出了时间平均成本开销及时间平均队列积压与权衡参数 V 的关系, 由图可见, 当 V 值的取值从 0 增大到 8 时, 即系统逐步倾向于以积压来换取更低地成本开销, 因此, 系统的时间平均成本开销随 V 值的增大而降低, 同时系统时间平均队列积压

随之增大, 这体现了成本开销 U 和系统队列积压 $Q_i(\tau_k)$ 之间存在的 $[C(1/V), C(V)]$ 平衡关系, 因而可以在实际的场景中根据需求灵活地选择 V 的取值。

表 2 仿真参数表

参数名	参数取值
系统带宽	15 MHz
高斯白噪声密度	-174 dBm/Hz
信噪比	16
用户设备发送功率	0.4 W
本地执行 CPU 频率	0.5~1.5 G cycles/s
边缘服务器执行频率	15 G cycles/s
任务处理密度	75 cycles/bit

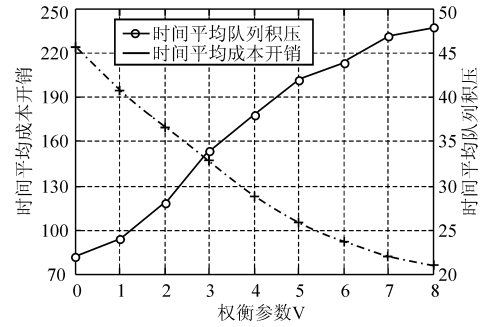


图 2 时间平均成本开销及时间平均队列积压

与权衡参数 V 的关系

当权衡参数 V 取值为 2、4 时, 对所提算法性能进行了有效性评估。如图 3(a) 所示, 随着仿真时间片的增加, 系统时间平均队列积压随之增大, 并最终稳定在值 27.64、38.14, 这说明了本文所提优化分配算法能够有效保证数据队列稳定性, 满足系统时延要求。同样地, 如图 3(b) 所示, 时间平均成本开销开始随仿真时间片的增加而迅速增大, 且最终趋于一稳定状态。二者均符合式(21)、(23)中的性能表述。

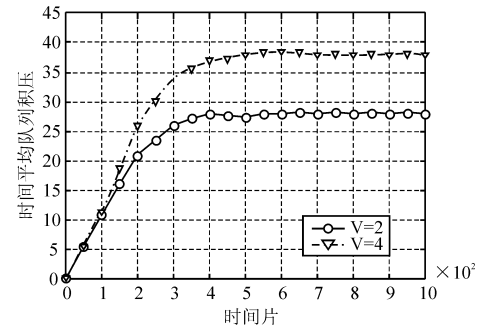


图 3(a) 时间平均队列积压与时间片的关系

(2) 性能对比分析

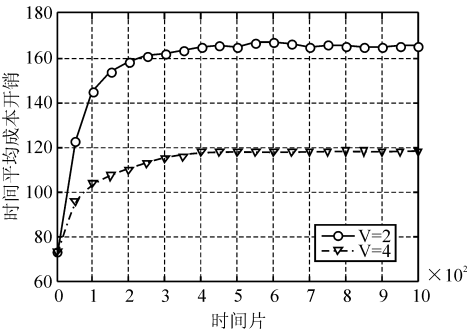


图 3(b)时间平均成本开销与时间片的关系

本文依据不同算法下的时间平均成本开销来进行优化算法的性能对比分析. 图 4 给出了本文算法在权衡参数值取 $V=4$ 时, CSP 中平均成本开销与用户数量的关系, 并与采用随机资源分配算法 (Random Resource Allocation Algorithm, RRA) 和文献[4]提出的联合启发式资源分配算法 (Joint Heuristic Resource Allocation Algorithm, JHRA) 进行了比较, 比较得到, 本文所提算法可以有效降低 CSP 中的成本开销, 且当用户数较多时, 这一趋势更加明显.

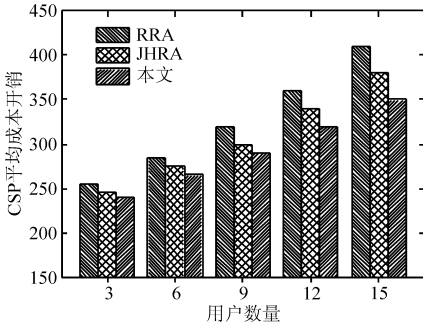


图 4 CSP 平均成本开销和用户数量的关系

5 结束语

基于 MEC 系统环境提出了一种最小化边缘服务成本开销的资源分配策略. 考虑到任务特性和服务尺度的差异, 该策略以无线与计算资源消耗单元为载体, 构建成本开销模型, 并联合分析动态任务数据状态变化, 应用 Lyapunov 优化设计一种实时动态分配方案. 仿真结果表明, 本文算法可以在满足系统数据队列稳定下显著降低资源成本开销, 从而边缘服务端可以通过本文算法实现资源开销与卸载时延之间的权衡控制.

参考文献:

[1] ZHANG K, MAO Y, LENG S, et al. Mobile-Edge Computing for Vehicular Networks: A Promising Network Paradigm with Predictive Off-Loading[J]. IEEE Vehicular Technology Magazine, 2017, 12(2): 36-44.

[2] TALEB T, SAMDANIS K, MADA B, et al. On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Architecture & Orchestration[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1657-1681.

[3] 李波, 黄鑫, 牛力, 等. 车载边缘计算环境中的任务卸载决策和优化[J]. 微电子学与计算机, 2019, 36(2): 78-82.

[4] KAN T Y, CHIANG Y, WEI H Y. Task offloading and resource allocation in mobile-edge computing system[C]//2018, 27th Wireless and Optical Communication Conference (WOCC). Hualien, Taiwan, IEEE, 2018: 1-4.

[5] TRAN T X, HAJISAMI A, PANDEY P, et al. Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges[J]. IEEE Communications Magazine, 2017, 55(4): 54-61.

[6] LIU Y, LEE M J, ZHENG Y. Adaptive Multi-Resource Allocation for Cloudlet-Based Mobile Cloud Computing System[J]. IEEE Transactions on Mobile Computing, 2016, 15(10): 2398-2410.

[7] JIN A L, SONG W, ZHUANG W. Auction-Based Resource Allocation for Sharing Cloudlets in Mobile Cloud Computing[J]. IEEE Transactions on Emerging Topics in Computing, 2018, 6(1): 45-57.

[8] YOU C, HUANG K, CHAE H, et al. Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading[J]. IEEE Transactions on Wireless Communications, 2016, 16(3): 1397-1411.

[9] 于博文, 蒲凌君, 谢玉婷, 等. 移动边缘计算任务卸载和基站关联协同决策问题研究[J]. 计算机研究与发展, 2018, 55(3): 537-550.

作者简介:

丁雪乾 男, (1993-), 硕士研究生. 研究方向为移动边缘计算. E-mail: dingxueqn@163.com

薛建彬 男, (1973-), 博士, 教授. 研究方向为物联网技术、无线通信理论与技术、移动边缘计算等.