

# 移动边缘云计算网络的高效计算资源共享

张永民<sup>1</sup>，电气与电子工程师学会会员；兰晓龙<sup>2</sup>；任菊<sup>3</sup>，电气与电子工程师学会会员；蔡林<sup>4</sup>，电气与电子工程师学会会士。

**摘要-** 边缘和云都可以为移动设备提供计算服务，以提高其性能。边缘可以通过提供本地计算服务减少传输延迟，而云则可以支持巨大的计算需求。它们之间的合作可以提高计算资源的利用率并确保服务质量，因此对边缘-云计算商业模式至关重要。本文提出了一种高效的移动边缘-云计算网络框架，使边缘和云能够以批发和回购的形式共享计算资源。为了优化计算资源共享过程，我们提出了计算资源管理问题，让边缘服务器管理其批发和回购计划，让云确定批发价格及其本地计算资源。然后，我们从两个角度来解决这些问题：i) 社会福利最大化；ii) 边缘和云的利润最大化。对于 i)，我们证明了社会福利的凹性，并提出了最优云计算资源管理以实现社会福利最大化。对于 ii)，由于难以直接证明基元问题的凸性，我们首先证明了批发计算资源相对于批发价格的凹性，并设计了最优定价和云计算资源管理，以实现其利润最大化。数值评估表明，通过社会福利最大化可以实现总利润最大化，而通过最优定价和云计算资源管理可以实现各自利润最大化。

**索引词**--边缘、云、计算资源共享、批发与回购、批发价格。

## I. 引言

移动应用改变了我们的生活，在我们的日常生活中变得越来越重要，有许多新的应用层出不穷，如虚拟现实、增强现实、智能识别、自动驾驶、互动游戏和电子健康[1]-[4]。快速的

手稿于 2019 年 3 月 31 日收到；2019 年 9 月 11 日修订，并 2020 年 1 月 16 日；2020 年 2 月 28 日接受；IEEE/ACM TRANSACTIONS ON NETWORKING 编辑 J. Huang 批准。发表日期：2020 年 3 月 25 日；当前版本日期：2020 年 6 月 18 日。本研究部分得到国家重点研发计划（2019YFA0706403）、"111 计划"（B18059）、国家自然科学基金（61702450、61629302、61702562 和 U19A2067）、加拿大科学与工程研究理事会（NSERC）和加拿大计算组织（Compute Canada）的资助。（通讯作者：鞠仁）。

张永敏和任菊分别就职于中南大学计算机科学与工程学院（长沙 410083）和清华大学计算机科学与技术系（北京 100084）（电子邮箱：

zhangyongmin@csu.edu.cn; renju@csu.edu.cn）。

兰小龙，四川大学网络安全学院，中国成都 610065（电子邮箱：xiaolonglan1112@gmail.com）。

Lin Cai 现供职于加拿大维多利亚大学电气与计算机工程系（University of Victoria, Victoria, BC V8P 5C2）（电子邮箱：cai@uvic.ca）。

数字对象标识符 10.1109/TNET.2020.2979807

移动应用对计算的要求越来越高,这给硬件能力有限的移动设备设计带来了新的挑战。除了硬件技术的发展,边缘和云都可以为移动设备提供计算服务。一般来说,云计算拥有巨大的计算资源,因此可以同时处理大量的计算任务,而边缘则可以在更靠近数据源的地方执行合格的计算,并通过减少用户与云之间的传输开销最终提高服务质量(QoS)。边缘和云的合作可以提高计算资源的利用率,确保服务质量。因此,在未来,特别是在移动设备的发展赶不上应用需求的增长时,建立一个具有适当商业模式的高效边缘-云合作网络是非常必要的。

近年来,人们对移动边缘计算(MEC)和云网络进行了广泛的研究[5],包括系统架构[6]-[9]、能源管理[10]-[14]、数据传输[15]-[17]、计算资源优化[18]-[22]和运行效率[23]-[26]。现有研究大多集中在设计兼容的 MEC、可靠的云网络和高效的计算任务处理协议,以提高计算系统的效率。然而,对于商业公司来说非常重要的 MEC

与云的合作以及相应的商业模式却缺乏应有的关注,可能会延缓其发展,尤其是在其起步阶段,对计算的要求时变且较低,投资较高。

通常,由于广泛部署的特点,MEC的平均建设和运营成本较高。此外,MEC的盈利能力较低,因为MEC的计算需求通常是随时间变化且有限的,这导致计算资源利用率较低。由于云的集中建设特点和规模经济效应,其平均建设和运营成本相对较低。然而,云计算需求的快速增长不仅对服务质量提出了挑战,也极大地增加了运营成本。在这种情况下,云希望以较低的成本获得计算资源,而MEC则希望获得更多的利润,并能为突发的计算任务提供有保障的服务。借助核心网络,MEC和云可以通过有线连接,以较低的通信延迟共享计算资源[27],从而实现优势互补,进一步提高盈利能力和服务质量。

我们之前的工作[28]为移动边缘云计算网络中的MEC设计了一种高效的批发和回购方案(EWBS)来管理批发和回购过程。给定批发和回购价格后,MEC

1063-6692 © 2020 IEEE。允许个人使用,但再版/再发行须经 IEEE 许可。  
更多信息,请参见 <https://www.ieee.org/publications/rights/index.html>。

服务器可以在批发收入和回购成本之间做出很好的权衡，在满足服务质量要求的同时实现利润最大化。然而，云的运营成本和计算要求对批发价格的影响尚未得到考虑。事实上，批发价格通常由云根据其运营成本和计算要求来指定。因此，应联合设计 MEC 的批发和回购方案以及云的定价方案。

在批发和回购的形式下，边缘服务器根据云发布的批发价格和自身的计算需求确定批发和回购的计算资源。云根据运营成本和 QoS 惩罚调整批发价格并管理其本地计算资源。考虑到边缘和云属于同一实体，我们证明了社会福利是云计算资源的凹函数，选择最优的云计算资源可以使社会福利最大化。考虑到边缘和云属于不同实体的另一种情况，由于批发难以直接证明整个问题的凸性，我们首先证明了批发计算资源相对于批发价格的凹性，然后设计了最优定价和云计算资源管理，以实现 MEC 和云的利润最大化。这样，通过所提出的高效计算资源共享方案，就能提高边缘和云的盈利能力。我们的工作贡献可概括为以下几点：

- 我们为移动边缘-云计算网络提出了一个高效的框架，在这个框架中，移动计算中心和云可以以批发和回购的形式相互共享计算资源。
  - 我们将 MEC 和云计算资源管理表述为利润最大化问题，并通过批发和回购流程将其耦合。
  - 我们从两个角度解决计算资源管理问题：i) 通过优化云计算资源管理实现社会福利最大化；ii) MEC 和云计算的利润最大化。
- 优化定价和云计算资源管理。
- 仿真结果表明，所提出的计算资源共享方案可以通过社会福利最大化实现总利润最大化，并通过最优定价和云计算资源管理实现 MEC 和云的利润最大化。

本文的其余部分安排如下：第二节介绍了相关工作，第三节提出了移动边缘-云计算网络的高效框架，并提出了 MEC 和云的计算资源管理问题。第四部分考虑了无利润转移的同一实体的计算资源管理问题，并设计了最优计算资源管理以实现社会福利最大化。第五部分考虑了有利润转移的不同实体的计算资源管理问题，并提出了优化定价和云计算资源管理以实现各自利润最大化的方案。第六节通过模拟演示了所提算法的效率。最后，第七部分总结了

我们的工作并介绍了我们未来的工作。

## II. 相关作品

我们从两个方面介绍相关工作：

### i) 系统设计和 ii) 系统优化。

#### 系统设计

文献[6]全面介绍了移动电子设备系统，包括其架构和技术手段。在网络接入技术方面，[8]提出了启发式链路路径公式，可以在合理的时间内设计出移动接入网；[24]提出了一种基于 MEC 的物体检测架构，通过无线通信实现实时监控应用。考虑到能源稀缺的问题，[10]提出了一种基于微波功率传输的 MEC 解决方案，以实现无源低复杂度设备的计算。从建设成本的角度出发，[25]提出了一种启发式解决方案，以在服务水平目标约束下最大限度地降低数据中心的部署成本。考虑到移动设备的移动性，文献[7]提出了一种新型空地一体化移动边缘网络（AGMEN），利用无人机辅助边缘网络的通信、缓存和计算；文献[17]提出了一种用于自动驾驶服务的两级边缘计算架构。然而，很少有研究考虑到 MEC 与云之间计算资源的结合，而这为大幅提高系统性能提供了机会。

#### 系统优化

考虑到能量限制，[20]提出了一种 MEC 和无线功率传输的统一设计，以增强移动设备的计算能力和能量供应。参考文献[22]提出了一种高能效计算卸载管理方案，以尽量减少移动设备的能耗。

从能耗的角度来看，文献[11]提出了一种最优工作负载分配方案，在功耗和传输延迟之间做出权衡；文献[12]提出了一种能量感知卸载方案，联合优化通信和计算资源的分配；文献[16]提出了一种卸载方案，在延迟约束下最小化整体用户的能耗。参考文献[19]利用单变量搜索技术提出了一种局部最优算法，以最小化能耗和执行延迟。文献[14]和[13]分别提出了同步和异步 MECO 系统的节能资源分配方案，以在计算延迟约束下最大限度地减少移动能耗。

为了改善用户体验，[15]设计了一种无线电和计算资源联合分配方案，以优化系统性能并提高用户满意度。

#### 参考文献

参考文献[18]提出了一种启发式算法，以共同优化卸载决策、通信资源和计算资源。参考文献[21]为虚拟机分配设计了真实的多项式时间拍卖，以实现社会福利最大化和/或提供商利润最大化。参考文献[23]提出了一种基于博弈的多用户计算卸载方案，以实现多用户各种利益的最大化。参考文献[26]设计了一系列分布式动态云网络控制算法，可在不了解服务的情况下为流量处理和传输联合调度计算和通信资源。

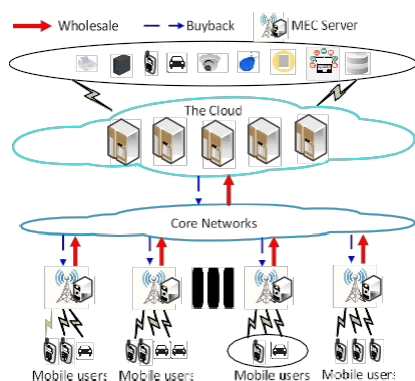


图 1. 移动边缘云计算网络架构。

需求, 确保云网络的稳定性并最大限度地降低其成本。然而, 如何共同优化 MEC 和云的计算资源管理, 进一步提高计算资源的利用率和收益率, 仍是一个有待解决的问题。

### III. 系统模型和问题表述

考虑一个移动边缘云计算网络，其中有  $N_1$  个 MEC 服务器和  $N_2$  云服务器。通常，MEC 服务器与蜂窝基站共建，为本地移动用户提供计算服务，而云服务器位于世界各地，通常通过核心网络为远程用户提供计算服务。一般来说，云端的计算任务量要比每个 MEC 服务器的计算任务量大得多，其延迟敏感性通常也较低。为了保证服务质量，MEC 服务器需要在紧迫的期限内完成接收到的计算任务，而云则需要尽快完成接收到的计算任务。

MEC 服务器和云通过有线核心网络连接, 可根据各自的计算需求共享计算资源。具体来说, 当 MEC 服务器的计算资源充裕时, 它们可以将部分计算资源批发给云, 以提高其盈利能力。相应地, 云可以低价从 MEC 服务器购买计算资源, 减少云计算资源, 降低高昂的运营成本, 提高服务质量。当 MEC 服务器的计算需求较高时, 它们可以以较高的回购价格从云中回购部分计算资源, 以确保其 QoS。<sup>1</sup>相应地, 云需要满足 MEC 服务器的回购请求, 以提高盈利能力。这样, 云可通过低批发价的批发方案从 MEC 服务器获取计算资源, 而 MEC 服务器则可通过高回购价的回购方案保证本地突发流量和紧急情况的延迟性能。通过共享计算资源, MEC 服务器和云都能从中受益。

移动边缘-云计算网络的架构如图 1 所示。

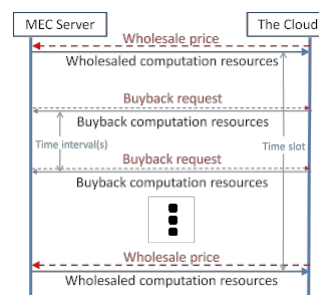


图 2. 移动边缘云计算网络的运行方案。

<sup>1</sup> 支付更高的价格可以保证回购计算资源的高优先级 [29]。



### A. 移动边缘云计算网络的运行

为了管理 MEC 服务器和云之间的计算资源共享过程，我们设计了一个高效的移动边缘-云计算网络框架，如图 2 所示。在这个系统中，有两种不同的时间尺度：第一种是时隙，另一种是时间间隔。通常，一个时隙可以是几十分钟，而一个时间间隔可以是几百毫秒。在每个时隙，MEC 服务器可将其部分计算资源批发给云，但在时隙内不能调整。在每个时间间隔内，MEC 服务器可从云端购回部分计算资源，这些资源应立即得到满足。这样，MEC 服务器需要保证批发给云的计算资源，并可根据自己的计算需求调整回购的计算资源。云可以在时隙内使用 MEC 批发的计算资源，并需要保证 MEC 服务器在每个时间间隔内的回购计算资源。

移动边缘云计算网络的运行方案如下：在前一个时隙结束时，云需要向 MEC 服务器发布批发价格。然后，每个 MEC 服务器确定批发的计算资源并估算回购的计算资源。在该时段内，MEC 服务器的批发计算资源由云管理。如果某个 MEC 服务器的预留计算资源无法满足其计算需求，该 MEC 服务器将通过向云发送回购请求来调整未来时间间隔内的回购计算资源。然后，云会相应地将回购的计算资源分配给相应的 MEC 服务器。

让  $t$  表示第  $t$  个时隙。将一个时隙划分为  $K$  个时间段，让  $k$  表示一个时隙中的第  $k$  个时间段。由于批发和回购方案的时间粒度不同，MEC 服务器需要分别确定批发和回购的计算资源[28]。云需要在每个时隙开始时确定批发和回购价格以及云计算资源。为简单起见，本文假设回购价格是给定的，而批发价格是可调的。

### B. MEC 服务器的运行模式

每个 MEC 服务器的计算资源可分为两部分。第一部分由 MEC 服务器保留，作为计算服务器处理接收到的计算任务。第二部分批发给云，作为灵活的计算服务器，处理接收到的计算任务。

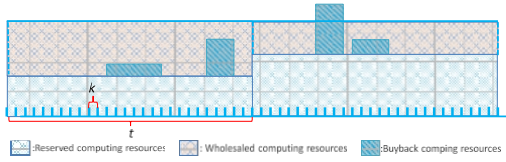


图 3. MEC 服务器的计算资源模型。

计算任务。鉴于计算任务的随机性，预留的计算资源可能不足以及时完成所有计算任务，因此 MEC 服务器需要从云上买回一些计算资源。因此，MEC 服务器的可用计算资源由其批发和回购计划决定。每个 MEC 服务器的可用计算资源如图 3 所示。

备注请注意，当批发/回购事件

发生时，云/MEC 服务器上的计算任务将为

卸载到相应的 MEC 服务器或/和云端进行处理。每个 MEC 服务器回购的计算资源可以来自自身、其他 MEC 服务器或/和云服务器。如果回购的计算资源来自自身，则相应数量的批发计算资源将被释放给 MEC 服务器，这样所有的 MEC 服务器都能获得相应的计算资源。

MEC 服务器上的计算任务将被卸载到其他 MEC 服务器或/和云端进行处理。通常情况下，云会将批发的计算资源释放回 MEC 服务器，以减少通信过载。这意味着，当 MEC 服务器发送回购信息时

请求，云将释放相应的  
将批发的计算资源数量返还给 MEC

如果本地计算资源足够，则会将其他多路复用器服务器/云的部分计算资源分配给多路复用器服务器；反之，则会将其他多路复用器服务器/云的部分计算资源分配给多路复用器服务器。

一般来说，MEC 服务器和云之间的通信延迟可以忽略不计，原因如下：i) 在 MEC 服务器本地处理的计算任务没有通信延迟；ii) 由于 MEC 服务器和云之间采用有线连接，因此它们之间的通信延迟相对于使用无线通信的通信延迟要低[30], [31]；iii) 卸载的计算任务可能会进入目标服务器的队列，而目标服务器正忙于为其他计算任务提供服务，因此通信延迟可以被队列延迟吸收[32]。当通信延迟较大且无法省略时，MEC 服务器可根据通信延迟提出回购请求，以确保计算和通信的整体延迟满足 QoS 要求。

让  $C_{e,t}$  表示可用计算资源总量， $C'$

其中  $C_{e,k}^I = C_{e,t}^I$  因为保留的计算资源在整个时段内都可以使用。

在不失一般性的前提下，我们假设 MEC 服务器  $e$  的计算任务到达遵循泊松分布，在时段  $t$  内的期望值为  $\lambda_{e,t}$ ，每个计算任务的计算工作量遵循指数分布，期望值为  $R_i$ 。MEC 服务器根据先到先服务 (FCFS) 策略 [33] 处理计算任务。在 FCFS 政策下，一个计算任务的计算延迟取决于其到达时间、计算工作量、之前未处理的计算工作量以及 MEC 服务器的可用计算资源。

让  $W_{e,k}$  表示到达的计算工作量  
表示 MEC 服务器  $e$  在时间间隔  $k$  内的累计未处理计算工作量， $Q_k$  表示 MEC 服务器  $e$  在时间间隔  $k$  内的累计未处理计算工作量。  
分别是时间间隔  $k$  结束时的  $e$ 。因此，我们有

$$Q_{e,k} = \max(0, Q_{e,k-1} + W_{e,k} - C_{e,k}), \quad (3)$$

在这里，计算工作量被转换为对计算资源的需求，即本文中的 GHz。

在 FCFS 策略下，如果  $\sum_{k'=k+1}^{\Sigma k+m-1} C_{e,k'}^I < \hat{C}_{e,k}^I$ ，其中  $\hat{C}_{e,k}^I$  表示总的  $C_{e,k}^I$ ，在接下的时间间隔  $[k+1, k+m]$  内，MEC 服务器  $e$  的可用计算资源。让  $D_{e,k}^I$  表示  $W_{e,k}$  中计算任务的最大计算延迟（包括排队延迟）。为简单起见，计算延迟  $D_{e,k}^I$  定义为

$$D_{e,k}^I = \begin{cases} m, & \text{如果 } \sum_{k'=k+1}^{\Sigma k+m-1} C_{e,k'}^I < \hat{C}_{e,k}^I \leq \sum_{k'=k+1}^{\Sigma k+m} C_{e,k'}^I \\ 1, & \text{if } Q_{e,k} \leq C_{e,k+1}^I \end{cases} \quad (4)$$

其中， $k'$  表示当前时间间隔  $k$  的下一个时间间隔。取决于时间间隔的时间尺度。为了保证 MEC 服务器  $e$  的服务质量，计算延迟  $D_{e,k}^I$  存在一个上限，用  $D_{e,t}^I$  表示。因此，应满足以下约束条件：

$$D_{e,t}^I \geq D_{e,k}^I. \quad (5)$$

由于可用计算资源  $\{C_{e,k'}^I, k' \in [k+1, k+m]\}$  决定了计算延迟  $D_{e,k}^I$ ，因此 MEC 服务器可以通过管理批发和回购计算资源。

### C. 云计算的运行模式

云的可用计算资源包括两部分：i) 云服务器的计算资源和 ii)

保留的计算资源, 以及  $C_{e,t}^I$  和  $C_{e,t}^C$  分别是 MEC 服务器  $e$  在时隙  $t$  期间的批发计算资源。我们有

$$C_{e,t} = C_{e,t}^I + C_{e,t}^C \quad (1)$$

让  $C_{e,k}^I$  和  $C_{e,k}^B$  表示可用计算总量

资源和 MEC 服务器的回购计算资源  $e$  分别在时间间隔  $k$  内。因此, 我们有

$$C_{e,k}^I = I_{e,k} + C_{e,k}^B, \quad (2)$$

计算资源。一般来说

第一部分取决于云服务器上的云计算资源, 这些资源可由云进行管理。第二部分取决于 MEC 的批发和回购计划。

服务器, 这受到批发价格的影响。

让  $C_{c,t}$  表示下列地点的可用计算资源总量

表示云服务器在时隙  $t$  内的总运行次数,  $C_{e,t}$  表示云服务器在时隙  $t$  内的总运行次数。

分别为  $C_{c,k}$  和  $C_{e,k}$  可由以下公式得出

$$C_k = C_{c,k} + \sum_e (C_{e,k} - C_{e,k}^B), \quad (6)$$



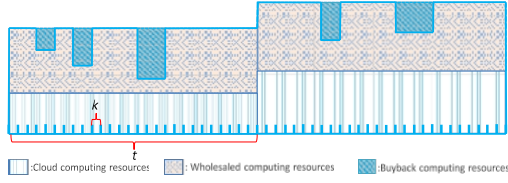


图4 云计算资源模型 云计算资源模型。

其中  $C_{c,t}^k = C_{c,t}$  和  $C_{e,t}^k = C_{e,t}$  由于计算

云服务器上的资源和来自 MEC 服务器的批发计算资源在时隙  $t$  期间始终可用。如图 4 所示。一般来说，在

云计算资源  $C_{c,t}$ ，用  $C_{c,t}^-$  表示。因此，我们有

$$0 \leq C_{c,t} \leq C_{c,t}^- \quad (7)$$

在本文中，计算任务到达云的次数遵循泊松分布，在时隙  $t$  内的期望值为  $\lambda_{c,t}$ ，每个计算任务的计算工作量为 0.5%。

计算任务遵循指数分布，其中  $R_t$  的预期值。考虑到云上异构应用的性能目标各不相同，优先级分配

是提高计算资源效率的常用方法之一 [34]。如何分配计算任务的优先级不在本文讨论范围之内。本文在本文中，我们将重点关注平均计算量。

在云端延迟。

以平均计算延迟作为云的服务质量要求，平均计算延迟存在一个上限，用  $D_{c,t}^-$  表示。一般情况下， $D_{c,t}^- > D_{c,t}^+$  始终成立。让  $D_{c,t}^+$  表示平均计算延迟在云端。我们有

$$D_{c,t}^+ \leq D_{c,t}^- \quad (8)$$

在这里，计算延迟  $D_{c,t}^+$  取决于计算工作量的到达和可用的计算资源

$C_{c,t}$ 。随着计算工作量的增加，越来越多的计算

资源  $C_{c,t}$  可降低计算延迟  $D_{c,t}^+$ ，这意味着更好的服务质量。

#### D. 多媒体中心和云的盈利模式

对于每台 MEC 服务器，其利润包括四个部分：i) 运行成本，ii) 处理计算任务的收入，iii) 向云批发计算资源的收入；以及 iv) 从云购回计算资源的成本。MEC 服务器和云的运营成本有几个重要因素。我们将这些因素（如能源消耗、日常维护和日常开支）视

让  $u_{e,t}$  表示批发一个的批发价单位计算资源， $U_{e,t}^S$  表示总的时段内批发计算资源的收入  $t$  分别为  $U_{e,t}^S$  的值为

$$U_{e,t}^S = a K C_{2,t} C_{e,t} \quad (10)$$

让  $g(C_{e,k}^{C^B})$  表示 MEC 服务器  $e$  在以下情况下的回购成本从云端买回  $C_{e,k}^{C^B}$  单位的计算资源

为了避免大量的回购计算资源， $g(C_{e,k}^{C^B})$  通常是  $C_{e,k}^{C^B}$  的递增凸函数 [35]。具体来说，我们设置  $g(C_{e,k}^{C^B}) = c_1 C_{e,k}^{C^B} + c_2 (C_{e,k}^{C^B})^2$  在本文中。让  $U_{e,t}^B$  表示总的

时隙  $t$  期间 MEC 服务器  $e$  的回购成本。

$$U_{e,t}^B = \sum_{k=1}^K g(C_{e,k}^{C^B}) \quad (11)$$

让  $U_{e,t}$  表示 MEC 服务器  $e$  在时隙  $t$  期间的总利润。根据 MEC 的利润模型，我们有

$$U_{e,t} = U_{e,t}^I + U_{e,t}^S - U_{e,t}^B \quad (12)$$

可以发现，总利润  $U_{e,t}$  取决于 MEC 服务器的批发和回购计划  $e$ 。

云计算的利润包括四个部分：i) 在云上处理计算任务的利润；ii) 本地运营的利润；iii) 云计算的利润。云服务器的成本，iv) 交易计算的云服务器的成本，v) 云服务器的成本，vi) 云服务器的成本，vii) 云服务器的成本。资源，其计算公式为  $U_{e,t}^S$

$U_{e,t}^B$ ；以及 iv) 计算延迟造成的服务质量损失。

假设云的服务费与 MEC 相同。因此，利润表示为  $U_{e,t}^I$ 、

是  $U_{e,t}^I = \sum_{k=1}^K a_{e,k} C_{e,k}^{C^B}$ 。

让  $g(C_{e,t})$  表示云运行期间的本地运行成本。

通常情况下， $g(C_{e,t})$  被假定为  $C_{e,t}$  的递增凸函数。具体

而言，我们设置  $g(C_{e,t}) =$

$c_3 C_{e,t} + c_4 (C_{e,t})^2$  在本文中。云可以确定其

通过管理云服务器上的云计算资源  $C_{c,t}$ ，实现本地运行成本  $g(C_{c,t})$ 。

根据 MEC 服务器的运行模式，与 MEC 服务器交易计算资源的成本为

$U_{e,t}^S - U_{e,t}^B$  所获得的计算资源

时间间隔  $k$  内的云量为  $\sum_{e,k} (C_{e,k}^C - C_{e,k}^{C^B})$ 。

为运营成本的一部分。为简单起见，我们将每个 MEC 服务器和云的运营成本都视为运营成本的一部分。服务器作为一个常数，取决于计算资源。

在云中，计算延迟是 QoS 的一个重要参数。因此，本文定义了云计算延迟的 QoS 惩罚，用  $U(D_{c,t}^{\wedge})$  表示。通常情况下， $U(D_{c,t}^{\wedge})$  是云计算延迟的一个递增函数。

因此，本文仅将后三部分视为 MEC 服务器的利润。让  $U_{e,t}^I$  表示在时间段  $t$  内处理一个单位计算工作量的服务费用。让  $U_{e,t}^I$  表示在时间段  $t$  内 MEC 服务器  $e$  处理计算任务的收入。

$$U_{e,t}^I = \sum_k a_{1,t} W_{e,k} \quad (9)$$

因为所有计算任务都应及时完成

增凸函数。

平均计算延迟  $\bar{D}_{c,t}$ 。

因此，云的总利润（用  $U_{c,t}$  表示）可由以下公式得出

$$U_{c,t} = U_{c,t}^I - g(C_{c,t}) - U(D_{c,t}^{\wedge}) - \sum_e (U_{e,t}^S - U_{e,t}^{B_e}) \quad (13)$$

云需要在运行成本和服务质量损失之间做出权衡。

## E. 问题的提出

在移动边缘云计算网络中，MEC 服务器和云都有各自的运行模式和目标。

在保证服务质量的前提下与云计算交易计算资源。云计算的目标是通过在运营成本和 QoS 惩罚之间进行良好权衡，实现总利润最大化。由于运营模式不同，它们的计算资源管理问题也不尽相同。

对于 MEC 服务器  $e$ ，由于计算工作量的不确定性，MEC 服务器需要在批发收入和回购成本之间做出权衡。MEC 服务器  $e$  的计算资源管理问题可表述为

$$\mathbf{P\_1:} \quad \max_{C_{e,t}^C, C_{e,k}^{CB}} \sum_t U_{e,t} \quad (14)$$

$$\text{s.t. } C_{e,t} = C_{e,t}^C + C_{e,k}^{CB}, \forall t, \quad (15)$$

$$D_{e,t}^- \geq D_{e,k}^-, \forall k, t, \quad (16)$$

$$C_{e,k}^{CB} \geq 0, \forall k. \quad (17)$$

目标是使 MEC 服务器的总利润最大化

变量是批发和回购计算

资源。第一个约束条件定义了批发计算资源的可用范围。

第二个约束

确保完成所有计算任务

在时间上。第三个约束条件给出了每个时间间隔内回购计算资源的可用范围。由于批发和回购方案的时间粒度不同，MEC 服务器需要分别设计批发方案和回购方案。

对于云计算，MEC 服务器的计算资源将受到批发价格的影响，而云计算服务器的计算资源则决定了本地运行成本。所有可用计算资源决定了 QoS 惩罚。

为了实现利润最大化，云计算需要做出权衡

云计算资源、批发价格和 QoS 惩罚之间的关系。云计算资源管理问题可表述为

$$\mathbf{P\_2:} \quad \max_{a_{2,t}, C_{c,t}} \sum_t U_{c,t} \quad (18)$$

$$\text{s.t. } 0 \leq C_{c,t} \leq \bar{C}_{c,t}, \forall t, \quad (19)$$

$$D_{c,t}^- \leq D_{c,t}^-, \forall t, \quad (20)$$

$$a_{2,t} \geq 0, \forall t. \quad (21)$$

云计算的目标是利润最大化，可控变量是批发价格和云计算资源。第一个约束条件定义了云服务器上云计算资源  $C_{c,t}$  的可用范围。第二个约束给出了服务质量要求。第三个约束条件显示了批发价格的范围。一般来说，可用计算资源取决于批发价格  $a_{2,t}$  和云计算资源  $C_{c,t}$ 。因此，云需要共同确定批发价格  $a_{2,t}$  和云计算资源  $C_{c,t}$ 。

根据问题 **P 1** 和 **P 2**，可以发现多用途电子设备服务器的计算资源管理，以及

2) 计算资源共享只发生在利润转移时，因此 MEC 和云需要根据彼此的决策来设计各自的计算资源管理。下面我们将逐一介绍这些解决方案。

#### IV. 无批发和回购计划

##### 利润转账

如果所有 MEC 服务器和云都属于同一个实体，则利润转移的总利润。因此，计算资源  $U_{e,t}^{e,t} - U_{e,t}^S$ ，不会影响

将问题 **P 1** 和 **P 2** 中的目标和约束条件结合在一起，就可以提出管理问题，可改写为

$$\mathbf{P1:} \quad \min_{C_{e,t}^C, C_{e,k}^{CB}, C_{c,t}} \sum_t (D_{c,t}^- + U_{c,t}) + g(C_{c,t}) \quad (22)$$

与(15) - (17)、(19) - (20)一致、

因为  $\{U_{e,t}^I, \forall e, t\}$  和  $\{U_{c,t}^I, \forall t\}$  是常数，而总利润转移  $\{\sum_e (U_{e,t}^S - U_{e,t}^B), \forall t\}$  为零。实体的

问题 **P1** 的目标是最小化云的本地运行成本和服务质量惩罚。此外，由于 MEC 服务器和云之间不存在利润转移，因此可以省略约束 (21)，例如  $a_{2,t} \geq 0$ 。要解决这个问题，我们需要模拟变量与云计算任务的平均计算延迟  $D_{c,t}^-$ 。

对于问题 **P1**，我们有如下的 Lemma：

**推论 1：** 最优批发计算资源

$C_{e,t}^C$  问题 **P1** 的  $C_{e,t}^C = C_{e,t}$ 。

这是因为考虑到复用增益，单队列多服务器模型比多队列多服务器模型更好 [36]。MEC 服务器和云统一管理计算资源。

由于所有 MEC 服务器都可以通过  $C_{e,t} = C_{c,t}$  可以改善系统性能。

这意味着 MEC 服务器和云服务器的计算资源都可以由云管理，这样所有计算任务都将在云的管理下进行处理。我们首先建立一个  $M/M/N$  队列系统模型，其中  $N = N_1 + N_2$ ，计算任务的到达遵循泊松过程，在时隙  $t$  内， $\lambda = \lambda_{c,t} + e \lambda_{e,t}$ ，平均服务速率为  $\mu = C_{c,t} + e C_{e,t} / NR$ ，为了确保队列系统的稳定性，我们有如下的定理：

**定理 2：** 云计算资源  $C_{c,t}$  应满足  $C_{c,t} > (\lambda_{c,t} + e \lambda_{e,t}) R t - e C_{e,t}$ 。

**证明** 根据队列理论，必要条件为

排队系统稳定的条件是  $\rho = \lambda / \mu < 1$ 。

否则，队列长度和计算延迟都将无穷大。因此  $\sum_e \bar{C}_{c,t} + e$

$C_{e,t} > (\lambda_{c,t} + \lambda_{e,t})R_t$   
云计算与批发和回购相关联

通过批发价格来实现。实现利润最大化、

MEC 服务器和云计算应共同优化其计算资源管理。本文

从两个方面解决上述优化问题：1) 有

MEC 服务器和云之间没有利润转移、

这样

应该满足。那么，我们有  $C_{e,t} > (\lambda_{c,t} + \lambda_{e,t})R_t$  -  
 $C_{e,t}$ ，用于云计算资源  $C_{c,t}$ 。 ■

考虑到计算任务的到达时间  $\lambda$  和服务速率  $\mu$ ，平均计算  
延迟  $D_{c,t}$  可由以下公式求得

$$D_{c,t} = \left[ \frac{(\lambda/\mu)^N \mu}{(N-1)!} \right] P_0 + \frac{1}{\mu}, \quad (23)$$

其中

$$P_0 = \sum_{n=0}^N \frac{(N\lambda)^n}{n!} + \frac{(N\lambda)^N}{N!} \frac{\mu}{\mu - \lambda} - 1. \quad (24)$$

这里,  $D_{c,t}$  是计算器的预期计算延迟。

在 MEC 服务器和云服务器上执行任务。

请注意, MEC 服务器和云中的计算任务具有不同的

QoS 要求。对于计算

MEC 服务器上的所有任务都应完成

在截止日期前完成。对于云上的计算任务, 应尽快完成。

鉴于

根据计算任务集及其计算工作量分布, MEC 服务器和云服务器的平均计算延迟由  $D_{c,t}$  给出。但是, 很难直接获得云服务器计算任务的平均计算延迟  $\hat{D}_g$ 。因此, 我们将下面是任务调度和近似值:

### 任务调度

当 MEC 服务器上所有计算任务的计算延迟 (包括队列延迟) 都在其最大可容忍延迟范围内时, 所有计算任务都将按照 FCFS 策略进行处理。否则, 将在其他 MEC 服务器/云上以高优先级处理, 以确保 MEC 服务器上的所有计算任务都能及时完成。假设云有足够的资源处理所有请求,

没有计算任务会被阻塞。因此, MEC 服务器  $e$  上计算任务的平均运行延迟可以估算。根据利特尔定律 [37] 中计算任务的长期平均数量。

静止系统的长期平均有效

到达率  $\lambda_{c,t} + \lambda_{e,t}$  乘以计算任务在系统中花费的平均时间  $D_{c,t}$ 。由于队列系统中不存在阻塞任务, 因此到达云端的计算任务的平均计算延迟  $\hat{D}^c_{c,t}$  可通过以下方法估算

$$\hat{D}^c_{c,t} \approx \frac{D_{c,t}(\lambda_{c,t} + \lambda_{e,t}) - \sum_{e \in \mathcal{E}} \lambda_{e,t} \min\{D_{c,t}, D_e^-\}}{\lambda_{c,t} + \lambda_{e,t}}. \quad (25)$$

可以发现,  $\hat{D}^c_{c,t}$  可视为线性函数

的平均计算延迟  $D_{c,t}$ 。

让  $Q^k$  表示累积的未处理计算工作量。来自 M/M/N、MEC 和云的负载

队列系统。让  $k$  表示即将到来的时间间隔在当前时间间隔  $k$ , 我们可以找到一个满足以下条件的  $k'$

$$k' - 1 \leq C_{c,t} + C_{e,t} \leq Q^k \leq k' C_{c,t} + C_{e,t}. \quad (26)$$

为确保 MEC 的所有计算任务都能在截止日期前完成,

回购计算

资源  $C^B_{e,k}$  的设置方法如下

用于计算任务卸载。这意味着, 只有在 FCFS 条件下无法及时完成的计算任务, MEC 服务器上的计算任务才会被赋予更高的优先级。

因此, 最佳  $C^C_{e,t}$  和  $C^B_{e,k}$  已获得并

$$P1': \min_{C_{c,t}} \sum_t g^*(C_{c,t}) + U(\hat{D}^c_{c,t}) \quad (28)$$

$$s.t. 0 \leq C_{c,t} \leq C^+_{c,t}, \forall t, \quad (29)$$

$$\hat{D}^c_{c,t} \leq \bar{D}^c_{c,t}, \forall t, \quad (30)$$

目标是最小化云的本地运行成本和 QoS 惩罚, 而可控变量是云计算资源  $C_{c,t}$ 。由于  $g^*(C_{c,t})$  是  $C_{c,t}$  的递增凸函数, 我们需要分析  $U(\hat{D}^c_{c,t})$  相对于  $C_{c,t}$  的凸性。

首先, 我们分析平均计算量的凸性相对于云计算资源  $C$  的延迟  $D_{c,t}$ , 并得出以下 Lemma:

定理 3: 平均计算延迟  $D_{c,t}$  是云计算资源  $C_{c,t}$  的欺骗性

凸函数。证明根据二阶 C 公式的定义、

用  $C(N, \lambda/\mu)$  表示, 由于所有服务器都很忙, 到达的通信任务在队列中等待的概率可由以下公式给出

$$C(N, \lambda/\mu) = \frac{(\lambda/\mu)^N \mu}{(N-1)!(N\mu - \lambda)}, \quad (31)$$

证明是  $\lambda/\mu$  的递增凸函数 [38]-[40]。因此,  $\frac{\partial C(N, \lambda/\mu)}{\partial \lambda/\mu} > 0$  和  $\frac{\partial^2 C(N, \lambda/\mu)}{\partial (\lambda/\mu)^2} > 0$

成立。我们可以得出  $\frac{\partial^2 C(N, \lambda/\mu)}{\partial \mu} < 0$ ,  $\frac{\partial^2 C(N, \lambda/\mu)}{\partial \mu^2} > 0$ 。

根据 (23),  $D_{c,t}$  可改写为

$$D_{c,t} = \frac{C(N, \lambda/\mu)}{N\mu - \lambda} + \frac{1}{\mu}. \quad (32)$$

因此, 我们有

$$\frac{\partial D_{c,t}}{\partial \mu} = \frac{\frac{\partial C(N, \lambda/\mu)}{\partial \mu} (N\mu - \lambda) - NC(N, \lambda/\mu)}{(N\mu - \lambda)^2} - \frac{1}{\mu^2}. \quad (33)$$

和

$$\frac{\partial^2 D_{c,t}}{\partial \mu^2} = \frac{\frac{\partial^2 C(N, \lambda/\mu)}{\partial \mu^2}}{N\mu - \lambda} - \frac{2N}{(N\mu - \lambda)^2} \frac{\partial C(N, \lambda/\mu)}{\partial \mu} + \frac{NC(N, \lambda/\mu)}{(N\mu - \lambda)^3} + \frac{2}{\mu^3}.$$

可以发现,  $\frac{\partial D_{c,t}}{\partial \mu} < 0$ ,  $\frac{\partial^2 D_{c,t}}{\partial \mu^2} > 0$ 。因此,  $D_{c,t}$  是一个

的递减凸函数。由于  $\mu$  是一个递增的  $c,t$  的线性函数, 平均计算延迟  $D_{c,t}$  是云计算资源  $C_{c,t}$  的递减凸函数。 ■

根据 (25), 由于  $\hat{D}^c_{c,t}$  可视为  $D_{c,t}$  的线性函数, 因此  $\hat{D}^c_{c,t}$  是一个递减的凸函数

由于  $U(\hat{D}^c_{c,t})$  是一个递增的凸函数  $c,t$ 。



$$\begin{aligned} C_{e,k}^{C^B} &= W_{e,k}^{C^B}, \text{ 如果 } k' \geq_{e,t} D_{e,t}^-; \\ C_{e,k}^{C^B} &:= W_{e,k}^{C^B}, \text{ 否则。} \end{aligned} \quad (27)$$

当  $W_{e,k}^{C^B} \leq C_{e,k}^{C^B}$  时。否则，计算资源  $W_{e,k}^{C^B} - C_{e,k}^{C^B}$  期间，将从其他 MEC 服务器/云中购买。

时间间隔  $k + D_{e,t}^- - \delta$ ，其中  $\delta$  是通信延迟

$\hat{D}_{C,t}$  和  $\hat{U}_{D_{C,t}}$  是  $C_{C,t}$ 。由于目标函数和约束条件

关于  $C_{C,t}$  是凸的，问题 **P1'** 是一个凸优化问题，可用现有工具（如 Matlab 中的 fmincon）求解。

至此，最优  $C_{e,t}^{C^B}$  和  $C_{e,t}$ ：

$C_{e,t}^{C^B}$  是  $C_{e,t}^{C^B}$  由 (27) 求得，最优  $C_{e,t}$

**算法 1** 最佳云计算资源管理 (社会福利最大化)

---

**1 输入:** 用于计算任务的  $(\lambda_{e,t}, R_t, D_{e,t}^-, (C_{e,t}, K, T))$  每个 MEC 服务器, 以及  $(C_{e,t}^-, D_{e,t}^-, (C_{e,t}, K, T))$  的云;

**2 输出:**  $\{C_{e,t}^-, C_{e,t}^+ \forall t\}$  和  $\{C_{e,t}^-, D_{e,t}^-, (C_{e,t}, K, T)\}$

**3 对每个时段  $t$  执行**

4   1) 设置  $C_{e,t}^+ = C_{e,t}$

5   2) 通过求解问题 **P1'** 计算最佳  $C_{e,t}$  使用 **Fmincon**;

6   **对于每个时间间隔  $k$**

7   | 根据公式 (27) 设  $C_{e,t}^+$

**8 结束**

**9 结束**

---

分别由问题 **P1'** 求解得到。这样社会福利可以通过最优云最大化

计算管理, 如算法 1 所示。不过, 在这种情况下, 实体需要同时建立 MEC 和云计算的成本很高, 甚至不可行。因此, 我们研究批发和回购计划与利润转移。

## V. 利润批发和回购计划

一般来说, 移动多媒体中心和云属于不同的应用领域。

在本文中, MEC 打算最大化

通过向移动用户提供计算服务和将其丰富的计算资源批发给云, 而云旨在提供更好的计算服务

为客户端提供服务, 并通过购买云计算资源。从问题 **P1** 和 **P2** 可以发现, 多边环境协定服务器决定其批发的

而云计算决定批发价格和云计算资源。在批发和回购方案下, 这两个问题因利润转移而耦合在一起。此外, 由于利润转移, 还需要

很难证明这两个问题的凸性。根据移动边缘云的运行方案

在图 2 所示的计算网络中, 云需要发出一个批发价给多方共赢服务器, 然后多方共赢服务器确定其批发的计算资源。为了解决这些问题, 我们首先分析

批发价格和批发的计算资源来自

请注意, 在本节中, 所有计算任务的预期计算延迟都可以通过两种排队系统计算出来。第一种是计算任务由每个 MEC 服务器上的预留计算资源处理, 可模拟为 M/M/1 队列模式系统。另一个是由

在第四节中, 可将 MEC 服务器和到达云端的数据建模为 M/M/N 队列系统。

### A. 批发价格与批发计算资源之间的关系

鉴于计算任务的到来, 批发和回购的计算资源通过计算耦合在一起

延迟  $D_{e,t}^+$  给定预留计算资源  $C^l$  和截止时间  $D_{e,t}$ , 时隙  $t$  期间的预期回购计算资源 (用  $C^{-B}$  表示) 可由以下公式求得

$$C_{e,t}^{-B} = K \frac{\lambda_{e,t} R_t^2}{(C_{e,t} - \lambda_{e,t} R_t) D_{e,t}} e^{(\lambda_{e,t} - \frac{C_{e,t}}{R_t}) D_{e,t}}, \quad (34)$$

和最小预期回购成本  $U^B$ , 用  $U^{-B}$  表示、

可以用 (35) 求出, 如页面底部所示、[28]  $C^{-B}$  和  $U^B$  都是递减凸函数的预留计算资源  $C^l$  [28]。由于  $C = C_{e,t} - C_{e,t}^l$  的递增凸函数

根据  $U^B$  和  $C^l$  之间的关系, 我们分析了批发的计算资源之间的关系  $C_{e,t}^l$  和批发价  $a_{2,t}$ , 并有以下结果

**定理 1:** 给定  $C_{e,t}^l$  到达的到来 计算资源  $C^l$  是非递减型

的函数 **证明:** 根据利润模型,  $U_{e,t}$  包括三个部分:  $U^l$ 、 $U^s$  和  $U^B$ , 分别由 (9)-(11) 给出。

$U_{e,t}^l$  可视作常数, 因为计算的到达任务及其预期工作量。给定值  $U_{e,t}^s$ ,  $U_{e,t}^B$  是  $C^l$  的线性增函数, 因为  $U_{e,t} = a_{2,t} K C_{e,t} - U_{e,t}^B$  是  $C^l$  的递减凹函数。因此, 根据 [28] 的结果, 总利润  $U_{e,t}$  是  $C^l$  的凹函数, 最优  $C^l$  满足以下条件

$$\begin{cases} \frac{\partial U_{e,t}}{\partial C_{e,t}^l} = C_{e,t} - \lambda_{e,t} R_t, & \text{如果 } \frac{\partial U_{e,t}}{\partial C_{e,t}^l} \Big|_{C_{e,t}^l = \lambda_{e,t} R_t} > 0; \\ C_{e,t}^l = 0, & \text{如果 } \frac{\partial U_{e,t}}{\partial C_{e,t}^l} \Big|_{C_{e,t}^l = \lambda_{e,t} R_t} \leq 0; \\ \frac{\partial U_{e,t}}{\partial C_{e,t}^l} = 0, & \text{否则} \end{cases} \quad (36)$$

随着  $C_{e,t}^l$  的增加,  $\frac{\partial U_{e,t}}{\partial C_{e,t}^l}$  增加, 相应的

$C^l$  是不递减的, 原因如下: 对于

MEC 服务器。然后，我们为

最佳可用计算资源。最后，我们设计了  
优化定价和云计算资源的管理  
管理，以实现 MEC 和云计算的利润最大化

同时

在 (36) 中的第一种  $e, t$  是一个常数，因为它已达到  
情况下， $C^C$   
最大值；对于 (36) 中的第二种情况， $C^C$  将增加  
当  $\frac{\partial U_{e,t}}{\partial C^C} > 0$ ；对于 (36) 中的第三种情况 将  
增加，以确保  $\frac{\partial U_{e,t}}{\partial C^C} = 0$ 。

$$\bar{U}_{e,t}^B = Ke^{(\lambda e, t - \frac{C^I}{R_{e,t}})D} \frac{c1\lambda e, t R_t^2}{(C_{e,t}^I - \lambda e, t R_t)D_{e,t}} + \frac{2C_{e,t}^I \lambda R_t^3 I}{(C_{e,t}^I - \lambda e, t R_t)D_{e,t}^2} \quad (35)$$

计算资源  $C^C$  不会随着时间的推移而减少。  
批发价  $a_{2,t}$ 。

这意味着，当云计算提高批发价格时  
从 MEC 批发的计算资源  
服务器的数量不会减少。根据定理 1，我们可以得出

下面是一个

Lemma:

定理 4: 考虑到计算任务的到来，整个

当  $C^C \in (0, C_{e,t})$  时，出售的计算资源  $C^C$  是批发价格  
 $a_{2,t}$  的递增凹函数。

$\lambda R$   
在  $e, t$  处

证明: 根据 (36)， $\frac{\partial U_{e,t}}{\partial C_{e,t}^C} = 0$  是充分  
最佳条件  $C_{e,t}^C$  当  $C_{e,t}^C \in (0, C_{e,t} - \lambda e_{e,t} R)$  时。

对于利润总额的组成部分  $U_{e,t}$ ，我们有  $\frac{\partial U_{e,t}^I}{\partial C_{e,t}^C} = 0$ ，

$\frac{\partial U_{e,t}^S}{\partial C_{e,t}^C} = a_{2,t} - \frac{\partial C_{e,t}^B}{\partial C_{e,t}^C} > 0$  以及  $\frac{\partial^2 U_{e,t}^S}{\partial (C_{e,t}^C)^2} < 0$ 。对于最佳

$C_{e,t}^C$ ，我们有一个  $\frac{\partial^2 U_{e,t}^S}{\partial (C_{e,t}^C)^2} < 0$ ，因为  $\frac{\partial^2 U_{e,t}^S}{\partial (C_{e,t}^C)^2} < 0$ ， $\frac{\partial U_{e,t}^S}{\partial C_{e,t}^C} > 0$ ，

是  $C^C$  的递增函数。这意味着

$C^C$ 、 $\frac{\partial U_{e,t}^S}{\partial C_{e,t}^C}$  会高出很多。因此， $a_{2,t}$  是  $C^C$  的增函数。让  $C^C(a_{2,t})$  表示最优批发计算资源  
与批发价格  $a_{2,t}$  和  $a_{2,t} = 2a_{2,t} - a_{2,t}$

分别为  $C^C(a_{2,t})$  和  $C^C(a_{2,t})$ 。因此， $C^C(a_{2,t})$  是  $a_{2,t}$  的递增凹函数，当

$C_{e,t}^C \in (0, C_{e,t} - \lambda R)$  时。

考虑到批发价和回购价以及计算任务的到达时间，最优

的预留和批发价

计算资源  $C^C$  和  $C_{e,t}^C$  可通过 (36) -

和预期回购计算资源  $C_{e,t}^B$  可以由 (34) 得出。注意，对于 (36) 中的第三种情况，问题

P\_1 是一个凸优化问题，可以用现有的工具来解决，如  
[28] 中的 Bisection 方法和 [41] 中的次梯度方法。

## B. 云的最低运营成本

除了从

云可以通过自己的服务器生成计算资源。云上的总计算资源

时间间隔  $k$  内的云  $C_{e,k}^C$  由 (6) 给出，其中

由本地运行成本  $g^C(C)$  决定

虽然  $\sum_{e,k} (C_{e,k}^C - C_{e,k}^B)$  是由批发

价格为  $a_{2,t}$  的  $C^C$  (34) 中。

$\frac{\partial C_{e,t}^B}{\partial C_{e,t}^C} < 1$  始终成立。因此， $C^C - C_{e,t}^B$  是一个递增且

当  $C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  时， $C_{e,t}^C$  的凹函数。

给定批发价格  $a_{2,t}$ ，MEC 服务器的预期计算资源可由  
可用计算资源之间的关系  
(6) 给出。

$C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  和批发价  $a_{2,t}$ ，我们得到

定理 6: 可用计算资源  $C_{e,t}^C - C_{e,t}^B$  是批发价的一个凹增函数  
当  $C^C \in (0, C_{e,t} - \lambda R_{e,t})$  时， $a_{2,t}$

证明  $C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  时，

根据定理 4，当  $C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  时，批发计算资源  $C_{e,t}^C$  是一个  
批发价  $a_{2,t}$  的递增凹函数。

因此，我们有  $\frac{\partial C_{e,t}^C}{\partial a_{2,t}} > 0$  和  $\frac{\partial^2 C_{e,t}^C}{\partial a_{2,t}^2} < 0$ 。

根据定理 5，可用计算资源  $C^C -$

$C_{e,t}^B$  是一个递增的凹函数。

$C^C$ 。我们有  $\frac{\partial C_{e,t}^C}{\partial a_{2,t}} > 0$  和  $\frac{\partial^2 C_{e,t}^C}{\partial a_{2,t}^2} < 0$ 。

0 和  $\frac{\partial C_{e,t}^C}{\partial a_{2,t}} < 0$ 。那么，我们可以得出  
 $\frac{\partial C_{e,t}^C}{\partial a_{2,t}} > 0$  和  $\frac{\partial^2 C_{e,t}^C}{\partial a_{2,t}^2} < 0$ 。  
 $\frac{\partial C_{e,t}^C}{\partial a_{2,t}} > 0$  和  $\frac{\partial^2 C_{e,t}^C}{\partial a_{2,t}^2} < 0$ 。

计算资源  $C_{e,t}^C - C_{e,t}^B$  是一个递增且

当  $C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  时。

这意味着，随着批发价格的上涨，可用计算资源递减且呈凹形。  
 $\sum_{e,t} C_{e,t}^C \in (0, C_{e,t} - \lambda R_{e,t})$  是非

云计算获取计算资源的成本

$\sum_{e,t} C_{e,t}^C$  来自 MEC 服务器  $\sum_{e,t} a_{2,t} C_{e,t}^C - U_{e,t}^B$  和

云服务器生成  $C$  的本地运行成本  $g^C(C_{e,t})$ 。让  $p_{e,t}(a_{2,t})$  和  $p_{e,t}(C_{e,t})$  分别表示从 MEC

和云服务器获取更多计算资源的最小成本。则有

$$p_{e,t}(a_{2,t}) = \frac{\sigma}{e} \frac{(u_{e,t}^S - u_{e,t}^B) - \sigma (u_{e,t}^S - U_{e,t}^B)}{(C_{e,t}^C - C_{e,t}^B) - \sigma (C_{e,t}^C - C_{e,t}^B)} \quad (37)$$

$$p_{e,t}(C_{e,t}) = \frac{\partial g^C(C_{e,t})}{\partial C_{e,t}} \quad (38)$$

其中

$\{U_{e,t}^S, U_{e,t}^B, C_{e,t}^C, C_{e,t}^B\}$  是参数的  $\sum_{e,t} (C_{e,t}^C - C_{e,t}^B)$

批发价格  $a_{2,t}$  满足  $a_{2,t} > a$ ，并且

$$C_{e,t}^{-B}) - \sum_{k \in e,k} (C_{e,t}^C - C_{e,t}^{-B})$$

对于云计算, MEC 的可用计算资源服务器  $e$  在时隙  $t$  期间的流量为  $C_{e,t}^C - C_{e,t}^{-B}$ 。对于

$C_{e,t}^C - C_{e,t}^{-B}$ , 我们可以得到下面的 Lemma:

定理 5: 可用计算资源  $C_{e,t}^C$

$$(0, C_{e,t}^C - \lambda R)$$

证明 可以证明, 回购计算资源  $C_{e,t}^{-B}$  是  $C_{e,t}^C$  的递增凸函数。

比较 (34) 和 (35) 可以发现,  $\bar{U}_{e,t}^B$

$$\frac{C_{e,t}^C - \lambda R}{(C_{e,t}^C - \lambda R) D_{e,t}^B} \bar{C}_{e,t}^B$$

=

$$a_{2,t} \frac{\partial U_{e,t}}{\partial C_{e,t}^C}$$

$$C_{e,t}^C \in (0, C_{e,t}^C - \lambda R). \text{ 由于 } \frac{\partial U_{e,t}}{\partial C_{e,t}^C} > c_{e,t}^C \text{ 和 } c_{e,t}^C > a_{2,t},$$

由

$$C_{e,t}^C \in (0, C_{e,t}^C - \lambda R)$$

$$C_{e,t}^C = (C_1 +$$

$C_{e,t}^C$

当

是一个足够小的常数。

为了使总运营成本最小化, 我们有以下定理:

定理 2: 存在一个满足以下条件的均衡点

证明 定义  $\epsilon$  为足够小的常数。得到  $\epsilon$  单位计算资源, 批发

价格应从

购买  $\epsilon$  单位计算资源为  $p_{e,t}(a_{2,t})$ 。生成

$\epsilon$  单位计算资源由云服务器、本地

运营成本可视为  $p_{c,t}(C_{c,t})$ 。鉴于总

$$\text{可用计算资源} = \frac{C_{c,t}^C}{k}, \quad (a_{2,t}) = p_{c,t}(C_{c,t})$$

当  $C_{c,t} < C_{c,t}^C$  时总是成立, 原因如下: 1)

如果  $p_{c,t}(C_{c,t}) > p_{e,t}(a_{2,t})$ , 这意味着存在一个更高的

果

批发价  $a_{2,t}$

为云计算购买更多计算

从 MEC 服务器获取资源, 以减少总运行时间。

成本。2) 如果  $p_{c,t}(C_{c,t}) < p_{e,t}(a_{2,t})$ , 则表示云



可以产生更多的云计算资源,以降低总运营成本,同时降低批发价格  $a_{2,t}$ 。因此,存在一个均衡点,满足  $p_{e,t}(a_{2,t}) = 0$ 。

■  $p_{c,t}(C_{c,t})$  当  $C_{c,t} < C_{c,t}^-$  时。  
根据定理 2,随着总量的增加  
可用计算资源  $C_{c,t}$  成本  $p_{e,t}(a_{2,t})$  和  
 $p_{c,t}(C_{c,t})$  将同时增加,直到  $C_{c,t} = C_{c,t}^-$ 。

但是,当云计算资源  $C_{c,t}$  达到上限  $C_{c,t}^-$  时,云只能通过提高批发价格从 MEC 处获取计算资源。

$a_{2,t}$ 。这样,我们就有了下面的 Lemma:

定理 7: 操作解满足  $p_{e,t}(a_{2,t}) \geq 0$

$p_{c,t}(C_{c,t})$  当  $C_{c,t} = C_{c,t}^-$  时。

证明根据定理 2,为使总运行成本最小化,云将调

整  $a_{2,t}$  和  $C_{c,t}$ , 以

确保  $p_{e,t}(a_{2,t}) = p_{c,t}(C_{c,t})$  直到云计算资源  $C_{c,t}$  达到其上  
限  $C_{c,t}^-$ 。如果 QoS 惩罚  $U(D_{c,t}^{\wedge c})$  非常高或约束  
 $D_{c,t}^{\wedge c} \leq D_{c,t}^-$  不能  
云必须购买更多的计算资源  
通过提高批发价格  $a_{2,t}$   
由于  $p_{e,t}(a_{2,t})$  是  $a_{2,t}$  的递增函数,我们有

$$p_{e,t}(a_{2,t}) \geq p_{c,t}(C_{c,t}) \text{ 当 } C_{c,t} = C_{c,t}^-$$

C. 最优定价和云计算资源

管理层

给定批发价格  $a_{2,t}$ ,  $\sum$  值  $(C^C - C^B)$   
确定。因此,在给定任意批发价格  $a_{2,t}$  的情况下,我们可以通  
过解决以下问题 10 计算出最佳云计算资源  $C_{c,t}$  :

$$\begin{aligned} \text{P2\_1: 最小} & \quad g^*(C_{c,t}) + U(D_{c,t}^{\wedge c}) \\ \text{s.t.} & \quad 0 \leq C_{c,t} \leq C_{c,t}^-, \forall t \end{aligned}$$

$$D_{c,t}^{\wedge c} \leq D_{c,t}^-, \forall t$$

目标函数是最大限度地减少本地运行总量  
成本和 QoS 惩罚,而约束条件  
定义云计算资源的范围  $C$

和计算延迟  $D_{c,t}$

鉴于  $a_{2,t}$  和云计算的批发价格  
资源  $C_{c,t}$ , 通过设置  $\lambda = \lambda$ , 可根据 (23) 计算  $D_{c,t}^{\wedge c}$  可以  
出预期计算延迟。

$\frac{C_{c,t} + (C^C - C^B)}{R}$ 。与 Lemma 3 类似,我们可以得出  
计算延迟  $D_{c,t}$  是一个递减的凸函数

云计算资源  $C_{c,t}$  的配置。由于  $g(C_{c,t})$  是

的递增凸函数

是一个关于云的凸优化问题

计算资源  $C_{c,t}$ 。因此,当  $C_{c,t} < C_{c,t}^-$  时,我们可以 33, 通

过二分法搜索最优批发价格  $a_{2,t}$  和最优 34 云计算资源  $C_{c,t}$  35  
与费用  $p_{e,t}(a_{2,t})$  和  $p_{c,t}(C_{c,t})$  相比,略 36

目标函数是使从 MEC 服务器购买计算资源的运营成本  
和云的 QoS 惩罚最小化,约束条件定义了  $a_{2,t}$  的批发价  
格范围和计算延迟。

■  $D_{c,t}^{\wedge c}$   
根据定理 1,可用计算资源

$\sum_e (C_{c,t}^e - C_{c,t}^{\wedge c})$  的非递减凹函数。

由于计算延迟  $D_{c,t}^{\wedge c}$  是总可用计算资源  $C_{c,t}^-$  的一个递  
减凸函数,因此  $U(D_{c,t}^{\wedge c})$  是  $a_{2,t}$  的一个非递增凸函数。因  
此,我们提出了一种基于二分法的启发式算法来找到最优  
批发价格  $a_{2,t}$ , 算法 2 中的步骤 19-35 略述了该算法  
。

## 算法 2 最佳定价和云计算资源管理

```

1 输入:  $(\lambda_{e,t}, R_t, D_{e,t}^-)$  表示计算任务,  $(C_{e,t}, K, T)$  表示每
   个 MEC 服务器,  $(C_{c,t}^-, D_{c,t}^-)$  表示云;
2 输出:  $\{a^*, C^*, \forall t\}$ ;
3 对每个时段  $t$  执行
4   1) 设  $a_{2,t} = p_{c,t}(C_{c,t}^-)$ 
   ;
   2) 计算最佳  $C_{c,t}^*$ 
   如果  $a_{2,t} < C_{c,t}^-$ , 设  $a_{2,t} = 0$  且  $a_{2,t} = a_{2,t}$ , 则
   如果  $a_{2,t} > C_{c,t}^-$ , 设  $a_{2,t} = C_{c,t}^-$ ;
   2) 通过解题 P2_1 更新最优  $C_{c,t}$ ;
   如果  $p_{e,t}(a_{2,t}) < p_{c,t}(C_{c,t}^*)$  那么
   | 设置  $a_{2,t} = a_{2,t}$ , 执行步骤 7;
   最后
   如果  $p_{e,t}(a_{2,t}) > p_{c,t}(C_{c,t}^*)$  那么
   | 设  $a_{2,t} = a_{2,t}$ , 执行步骤 7;
   结束
   如果  $p_{e,t}(a_{2,t}) = p_{c,t}(C_{c,t}^*)$  那么
   |  $a^* = a_{2,t}$ , 断开;
   结束
结束
如果  $a_{2,t} = C_{c,t}^-$ , 设  $a_{2,t} = a_{2,t}$ , 则
如果 1) 设  $a_{2,t} = (1 + G)a_{2,t}$ , 并更新成本  $U_{c,t}$ ; 当
 $U_{c,t}|_{a_{2,t}} < U_{c,t}|_{a_{2,t}}$  时, 做
| 设置  $a_{2,t} = a_{2,t}$ , 并通过步骤 20 更新  $U_{c,t}|_{a_{2,t}}$ ;
最后
2) 当  $U_{c,t}|_{a_{2,t}} \geq U_{c,t}|_{a_{2,t}}$  时, 设  $a_{2,t} = a_{2,t}$ ;
3) 设置  $a_{2,t} = \frac{a_{2,t} + a_{2,t}}{2}$  并更新成本  $U_{c,t}$ ;
如果  $U_{c,t}|_{a_{2,t}} > U_{c,t}|_{a_{2,t}}$  则
| 设置  $a_{2,t} = a_{2,t}$ , 执行步骤 27;
最后
如果  $U_{c,t}|_{a_{2,t}} < U_{c,t}|_{a_{2,t}}$  则
| 设置  $a_{2,t} = a_{2,t}$ , 执行步骤 27;
最后
如果  $U_{c,t}|_{a_{2,t}} = U_{c,t}|_{a_{2,t}}$  则
|  $a_{2,t}^* = a_{2,t}$ , Break;
结

```

## 束 结束

如算法 2 中的第 6-18 步。

但是，当云计算资源  $C_{c,t}$  达到其上限  $C_{c,t}^-$  时，云计算应选择一个最优的云计算资源  $C_{c,t}^*$ 。

$a_{2,t}$ ，以权衡运营成本和服务质量损失。因此，最佳批发价格  $a_{2,t}$  可以通过解决以下问题计算出来：

$$\begin{aligned} \text{P2\_2: } \min & \sum_{e,t} (U_{e,t}^S - U_{e,t}^B) + U(\hat{D}_{c,t}) \\ \text{s.t. } & a_{2,t}^e \geq 0, \forall t, \\ & \hat{D}_{c,t} \leq \bar{D}_{c,t}, \forall t. \end{aligned}$$

根据 [28] 中的 EWBS 计算最佳  $C^{C*}$  和  $\hat{C}^B$

$e,t$   $e,k$

37 结束

38 其中， $G$  是一个小步长。

## 趋同分析

当  $C_{c,t}^* < C_{c,t}^-$  时，基于二分法、批发价  $a_{2,t}$  的可用范围将按步骤 7-14 划分，直至  $p_{e,t}$  即最优批发价  $a_{2,t}^e$  已经得到。由于  $a_{2,t}^e = 0$  (见注 2)，这意味着  $a_{2,t}^e$  已经得到。由于总运行成本将逐级降低，步骤为

表 I

不同到达率的多路复用器服务器数量

$\lambda_{e,t}$	6	7	8	9	10	11	12
Low	20	10	5	2	1	1	1
Middle	4	5	6	8	6	5	6
High	1	1	2	2	4	10	20

6-18 将趋于统一解。当  $C^* = C_{c,t}^-$  时、

这意味着，通过提高批发价  $a_{2,t}$ ，可以降低总运营成本。因此，

采用步骤 20-24

以找到最佳批发价的范围  $a^*$ 。

$a_{2,t}$ ，总运营成本将达到无穷大。因此，存在一个统一的上限  $a_{2,t}^-$ ，使云的总利润最大化。同样，步骤 25-34 将缩小最优批发价的范围

$a_{2,t}^*$ 。由于每次运行的总成本都会减少

步，第 19-34 步将收敛到统一解。因此，建议的算法 2 将收敛到统一解。

## VI. 数值评估

我们评估了所提出的社会福利最大化以及移动边缘-云计算网络中的最优定价和云计算资源管理（模拟部分命名为最优定价方案），并在本节中展示了数值结果和分析。

评估设置如下：共有 40 台 MEC 服务器，每台服务器拥有

$C_{e,t} = 3.2\text{GHz}$  的计算资源，10 台云服务器位于

计算资源上限的云  $C_{c,t}^- =$

200 千兆赫。云计算资源的成本为  $c_3 =$

0.02487 美元/（GHz\*小时）， $c_4 = 0.002487$  美元/（

GHz\*小时）<sup>2</sup>。我们假设 MEC 服务器的计算任务到达率为每秒 [6, 12]，云的

计算任务到达率为每秒  $\lambda_{c,t} =$

500。为描述 MEC 服务器上随时间变化的计算工作量，

$\lambda_{e,t}$  在低、中、高三种状态下的具体分布见表 I。每个计

算任务的预期计算工作量为 100Kb，每个比特的处理需要约

2000 个周期的计算资源[42]。我们假设服务

例如，MEC 服务器和云的费用是一样的、

$a_{1,t} = 0.2764$  美元/Gb。买回计算资源的参数为  $c_1 =$

0.2736 美元/（千兆赫\*小时）和  $c_2 = 0.2736$  美元/（

千兆赫\*小时）<sup>2</sup>。MEC 服务器上计算任务的截止时间为  $D_{e,t}^- = 2$  秒，云上

计算任务的截止时间为  $D_{c,t}^- = 4$  秒。QoS 惩罚为  $U(D_{c,t}^+ \hat{c}_{c,t}) = 104 D_{c,t}^+ \hat{c}_{c,t}$

美元。为了评估所提算法的性能，我们比较了  $D_{c,t}^+$ 、 $D_{c,t}^-$  和  $D_{c,t}^0$

仿真结果有以下两种方案：1) MEC 服务器不能向云共

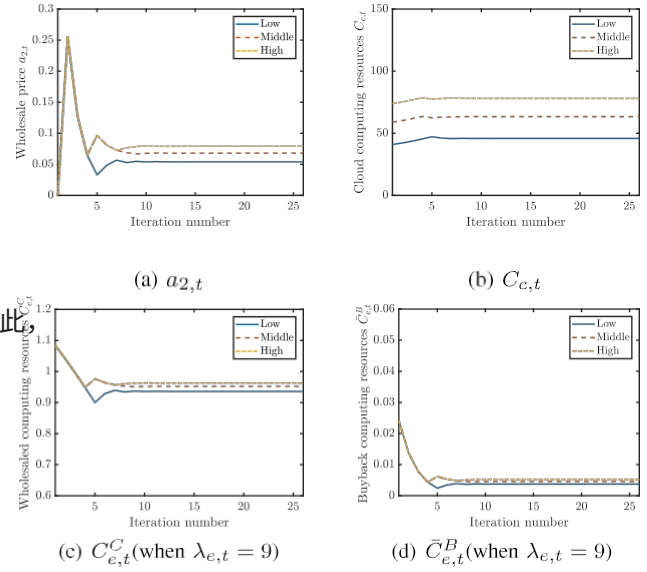


图 5. 变量轨迹：(a) 批发价格  $a_{2,t}$ ，(b)

(c)  $\lambda_{e,t} = 9$  的 MEC 服务器的批发计算资源  $C_{c,t}$ ，以及 (c) 批发计算资源  $C^C$  和回购计算资源  $C^B$ 。

随着 MEC 服务器上计算任务的增加，所有参数（包括批发价格  $a_{2,t}$ 、云计算资源  $C_{c,t}$  以及批发计算资源  $C^C$  和预期回购计算资源  $C^B$ （计算任务到达率给定））都将增加。原因可归纳如下：随着 MEC 计算任务的增加，

1) MEC

2) 对于云，由于 MEC 服务器批发的计算资源减少，因此需要提高批发价格和/或云计算资源，以便在运营成本和 QoS 损失之间进行良好权衡；

3) 对于  $\lambda_{e,t} = 9$  的 MEC 服务器来说，由于批发价格增加，它将向云批发更多计算资源以获得更多利润，同时回购的计算资源也将增加，以保证 QoS。

最佳云计算资源  $C^*$

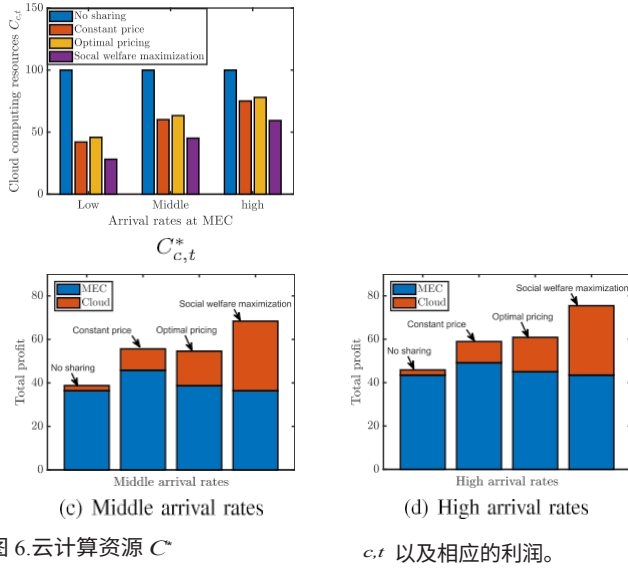
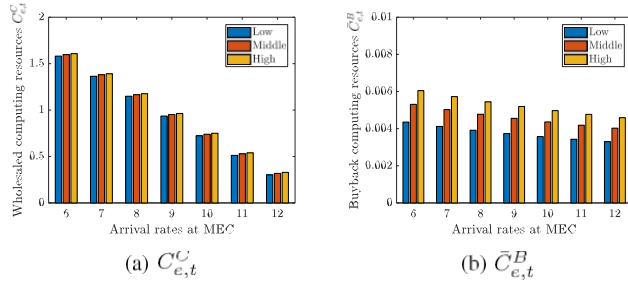
在不同的享其计算资源，可以向云购买计算资源以保证其服务质量，命名为 "不共享"；2) 批发价格是一个常数，即  $a_2 = 0.2487$  美元/（GHz\*小时），命名为 "不变价格"。需要注意的是，为了实现社会福利最大化，我们假设 MEC 和云的收入是它们提供计算服务的服务费，云的利润不会转移给 MEC。

## A. 模拟结果和性能分析

拟议算法中的变量轨迹如图 5 所示。可以发现, 所提出的算法可以在 15 次迭代内找到解。此外, 图 5 还显示了所提出算法的变量轨迹、

图 6 显示了各种方案以及 MEC 和云的相应利润。从图 6(a)中可以发现, "不共享"获得的云计算资源最高, 因为云上的所有计算任务都将由云计算资源处理; 而 "社会福利最大化"获得的云计算资源最低, 因为 MEC 服务器上的所有可用计算资源都将被利用, 以降低云的本地运营成本。此外, "恒定价格"的计算资源也低于最优定价方案。这是因为 "恒定价格"下的批发价格为  $_{2,t} = 0.02487$  美元/(千兆赫\*小时)、远高于最佳批发价格  $\alpha^* = \{0.05388 \text{ 美元}, 0.06790 \text{ 美元}, 0.07908 \text{ 美元}\}/(\text{千兆赫}^{\frac{2}{t}} \text{ 小时})$ , 因此 "Con-"是  $\{0.05388 \text{ 美元}, 0.06790 \text{ 美元}, 0.07908 \text{ 美元}\}/(\text{千兆赫}^{\frac{2}{t}} \text{ 小时})$ 。恒定价格"将以更高的运行成本从 MEC 获得更多的计算资源。

从图 6(b)-6(d) 可以发现, 随着 MEC 服务器到达率的增加, 所有方案的总利润都会增加。社会福利最大化获得的总利润最高, 云的利润也最高, 因为大部分计算资源来自 MEC, 没有任何利润转移。因此, 社会福利

图 6. 云计算资源  $C^*$ 图 7. 批发的计算资源  $C^C$ 

资源  $C^B$   $e,t$  为不同到达率的 MEC 服务器。

最大化 "获得的 MEC 服务器利润最低。显然, "不共享 "获得的总利润最低, 云的利润也最低, 因为云计算资源的运营成本很高。"不变价格 "的总利润高于最优定价方案。但是, "不变价格 "的云利润低于最优定价方案。因此, 所提出的最优定价方案可以提高云计算的利润率, 而这正是云计算所青睐的。此外, 与 "不共享 "和社会福利最大化相比, 最优定价方案可以提高 MEC 的盈利能力。

图 7 显示了 MEC 服务器的计算资源批发和回购情况。可以发现, 随着总计算工作量的增加, 到达率一定的 MEC 服务器会向云批发更多的计算资源, 以获取更多利润。为了满足计算需求, 到达率越高的 MEC 服务器会向云批发越少的计算资源。此外, 随着总计算工作量的增加, 到达率较低的 MEC 服务器将购买更多的计算资源。因为他们可以通过以下方式获得更多利润向云批发更多计算资源。请注意回购的计算资源与批发的计算资源相比非常少, 因为回购成本与批发价格相比非常高。

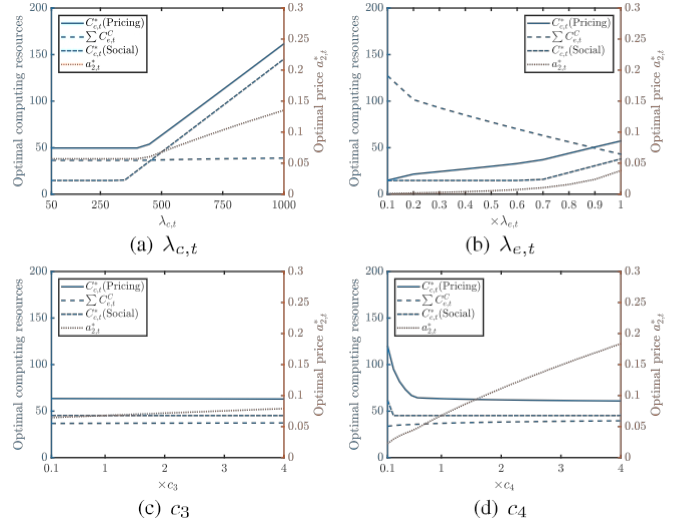


图 8. 系统参数对最优决策的影响。

批发和回购计算资源  $C^C$   $e,t$  和  $C^B_{e,t}$ 、受到多个系统参数的影响, 例如计算任务到达云  $\lambda_{c,t}$  的时间、计算任务到达 MEC 服务器  $\lambda_{e,t}$  的时间、在云上生成计算资源的成本参数等。本文分析了多个系统参数对最优决策的影响, 结果如图 8 所示。从图中可以看出, 云计算资源社会福利最大化下的  $C^*$  (社会)总是小于最优定价方案下的  $C^*$  (定价)  $c,t$ 。因为在社会福利最大化的前提下, 云计算可以利用 MEC 的更多计算资源。

#### 云计算工作量的影响 $\lambda_{c,t}$

如图 8(a) 所示, 随着云计算工作量  $\lambda_{c,t}$  的增加, 所有变量, 如社会福利最大化下的  $C^*$  和最优定价方案下的  $C^*$ 、 $a^*$  和  $C^C$  都不会递减。这就是  $c,t$   $2,t$   $e,t$  因为云中不断增加的计算工作量  $\lambda_{c,t}$  可能会增加对计算资源的需求, 从而在运营成本和 QoS 损失之间实现良好的权衡。可以发现,  $\lambda_{c,t}$  存在一个阈值。当计算工作量  $\lambda_{c,t}$  低于阈值时,  $C^*$   $c,t$  由于服务质量惩罚较低, 社会福利最大化和最优定价方案下的  $C^*$  以及最优定价方案下的批发价格  $a$  都将保持在最低水平。当计算工作量  $\lambda_{c,t}$  超过阈值时, 社会福利最大化和最优定价方案下的  $C^*$  和最优定价方案下的批发价格  $a^2$  几乎呈线性增长。请注意, 云计算资源  $C^C$   $c,t$  远大于批发价

$a_{2,t}^*$  和批发计算资源  $C^C$   $e,t$  由于高昂的回购成本和 MEC 的 QoS 保证。

#### 多媒体中心计算工作量的影响 $\lambda_{e,t}$

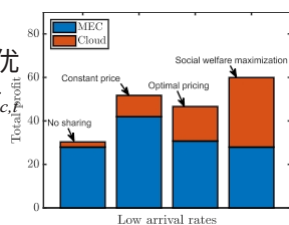
如图 8 (a) 所示, 随着 MEC 服务器计算工作量  $\lambda_{e,t}$  的增



加，批发的计算资源  
 B. 系统参数性能分析

一般来说，最优决策，如最优销售价格  $\alpha_{e,t}$ ，云计算资源  $C_{c,t}$  运行。

(a)



(b) Low arrival rates

$C_{e,t}$  将减少，而云计算资源  $C_{c,t}$  整体和批发价格  $\alpha_{e,t}$  增加。这是因为 MEC，服务器需要预留更多的计算资源，以保证云计算的正常

当计算要求较高时, QoS 就会降低。由于批发计算资源  $C^*$  的减少量、云必须增加云计算资源  $C^*$

$C_{e,t}$   
 $c_{e,t}$

和批发价  $a^*$  之间进行良好的权衡操作成本和 QoS 惩罚。因此, 随着  $\lambda_{e,t}$  的增加, 云的运行成本将增加, 而 MEC 的利润将增加, 如图 6 所示。

#### 成本参数 $c_3$ 和 $c$ 的影响

如图 8(c)-8(d)所示, 随着成本参数  $c_3$  和  $c_4$  的增大, 批发价  $a^*$  将增大, 而  $c_4$  的影响远大于  $c_3$ , 原因是  $(C)_{c,t}^2$  的大值。随着  $c_3$  的增加, 社会福利最大化和最优定价方案下的计算资源  $C^*$  和批发的计算资源  $C$  都会增加。

资源  $C^*$  基本保持不变, 因为云会稍稍提高批发价  $a^*$ , 以便在运营成本和服务质量损失之间进行良好权衡。随着  $c_4$  的增加, 批发价  $a^*$  也会迅速增加, 因为云计算公司会增加运营成本。

云计算资源  $C^*$  较大。需要注意的是, 当  $c_4$  的值非常小时, 采用最优定价方案的云计算资源  $C^*$  会更高, 因为云计算资源  $C$  的值较低。

成本参数。因此, 以较低的  $c_4$  构建云计算更容易产生利润。

## VII. 结论

在本文中, 我们提出了一个有效的框架, 让移动计算中心和云相互共享计算资源, 以提高它们的盈利能力。我们将 MEC 和云的计算资源管理表述为各自的利润最大化问题。然后, 我们考虑了两种情况来解决这些问题: i) 社会福利最大化, 即假设 MEC 和云属于同一实体, 计算资源共享不涉及利润转移; ii) 各自利润最大化, 即假设 MEC 和云属于不同实体, 计算资源共享只涉及利润转移。对于第一种情况, 我们证明了社会福利只取决于云计算资源和社会福利最大化问题的凹性。对于第二种情况, 我们分析了批发价格和 MEC 可用计算资源之间的关系, 然后设计了一种最优定价和云计算资源管理方法, 以实现总利润最大化。数值评估结果表明, 所提出的算法可以使社会福利最大化, 并分别使 MEC 和云计算的各自利润最大化。此

服务器同时运行。云可以向不同的实体发布不同的价格, 以提高其效益。此外, 我们还将考虑 MEC 服务器之间的通信延迟。

云/其他 MEC 服务器, 以及能源的影响。

在移动边缘云计算网络中管理计算资源共享时, 我们将考虑到计算资源的消耗。此外, 我们还将进一步讨论当计算任务的到达量超过 MEC 服务器和云的全部容量时的接纳控制设计。

## 参考资料

- [1] N.N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: 调查", 《IEEE 物联网杂志》, 第 5 卷, 第 1 期, pp.450-465, Feb. 2018.
- [2] L.Cai, J. Pan, L. Zhao, and X. Shen, "Networked electric vehicles for green intelligent transportation," *IEEE Commun. 标准* 杂志, 第 1 卷, 第 2 期, 第 77-83 页, 2017 年 6 月。
- [3] Y.Li, K. Sun, and L. Cai, "Cooperative device-to-device communication with network coding for machine type communication devices," *IEEE Trans. 无线通信*, 第 17 卷, 第 1 期, 第 296-309 页, 2018 年 1 月。
- [4] J.Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi 和 S. He, "窄带物联网: 实施与应用", *IEEE Internet*

外, 我们还分析了系统参数对系统性能的影响, 结果表明成本参数  $c_4$  对批发价格的影响更大。

在本文中, 我们假设所有 MEC 服务器都具有在未来的工作中, 我们将考虑大规模移动边缘-云计算网络中的计算资源共享问题。在未来的工作中, 我们将考虑大规模移动边缘-云计算网络中的计算资源共享问题, 在这种网络中, 有多个隶属于不同实体、具有不同计算资源和 QoS 要求的 MEC 服务器, 同时有多个云网络竞争从 MEC 处批发的计算资源。

- Things J.*》, 第4卷, 第6期, 第2309-2314页, 2017年12月。
- [5] J.Ren, D. Zhang, S. He, Y. Zhang, and T. Li, "A survey on end edge-cloud orchestrated network computing paradigms: 透明计算、移动边缘计算、雾计算和小云", 《*ACM Comput.Surv.*》, vol. 52, no. 6, pp.
  - [6] H.Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. de Foy, and Y. Zhang, "Mobile edge cloud system: 架构、挑战和方法", 《*IEEE Syst.J.*》, vol. 12, no.3, pp.
  - [7] N.Cheng 等人, "空地一体化移动边缘网络: Architecture, challenges, and opportunities," *IEEE Commun. 杂志*》, 第56卷, 第8期, 第26-32页, 2018年8月。
  - [8] A.Ceselli, M. Premoli, and S. Secci, "Mobile edge cloud network design optimization," *IEEE/ACM Trans.网络*》, 第25卷, 第3期, 第1818-1831页, 2017年6月。3, 第1818-1831页, 2017年6月。
  - [9] J.Ren, H. Guo, C. Xu, and Y. Zhang, "Serving at the edge: 基于透明计算的可扩展物联网架构", 《*IEEE Netw.*》, 第31卷, 第5期, 第96-105页, 2017年9月/10月。5, 第96-105页, 2017年9月/10月。
  - [10] C.You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE J. Sel.Areas Commun.*》, 第34卷, 第5期, 第1757-1771页, 2016年5月。5, 第1757-1771页, 2016年5月。
  - [11] R.Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet Things J.*, vol. 3, no. 6, pp.
  - [12] J.Zhang *et al.*, "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet Things J.*, vol. 5, no.4, 第2633-2645页, 2018年8月。
  - [13] C.You, Y. Zeng, R. Zhang, and K. Huang, "Asynchronous mobile-edge computation offloading : Energy-efficient resource management," *IEEE Trans. 无线通信*》, 第17卷, 第11期, 第7590-7605页, 2018年11月。
  - [14] C.You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans.Kim*, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans.*16, no.3, pp.
  - [15] Y.Gu, Z. Chang, M. Pan, L. Song, and Z. Han, "Joint radio and computational resource allocation in *IoT* fog computing," *IEEE Trans.Veh. Technol. 技术*》, 第67卷, 第8期, 第7475-7484页, 2018年8月。
  - [16] S.Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans.Signal Inf.Process.Netw.*》, 第1卷, 第2期, 第89-103页, 2015年6月。
  - [17] Q.Yuan, H. Zhou, J. Li, Z. Liu, F. Yang, and X. S. Shen, "Toward efficient content delivery for automated driving services: 边缘计算解决方案", 《*IEEE Netw.*》, 第32卷, 第1期, 第80-86页, 2018年1月。
  - [18] X.Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Trans.Veh.Technol.*4, 第3435-3447页, 2017年4月。
  - [19] Y.Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: 使用动态电压缩放的部分计算卸载", 《*IEEE Trans. 通信*》, 第64卷, 第10期, 第4268-4282页, 2016年10月。

- [20] F.Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans.* 17, no.3, pp.
- [21] X.Zhang, Z. Huang, C. Wu, Z. Li, and F. C. M. Lau, "Online auctions in IaaS clouds: Welfare and profit maximization with server costs," *IEEE/ACM Trans.Netw.*, 第 25 卷, 第 2 期, 第 1034-1047 页, 2017 年 4 月。
- [22] F.Guo, H. Zhang, H. Ji, X. Li, and V. C. M. Leung, "An efficient computation offloading management scheme in the densely deployed small cell networks with mobile edge computing," *IEEE/ACM Trans.Netw.*, 第 26 卷, 第 6 期, 第 2651-2664 页, 2018 年 12 月。
- [23] X.Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans.Netw.* 5, 第 2795-2808 页, 2016 年 10 月。
- [24] J.Ren, Y. Guo, D. Zhang, Q. Liu, and Y. Zhang, "Distributed and efficient object detection in edge computing: 挑战与解决方案", 《*IEEE Netw.*》, 第 32 卷, 第 6 期, 第 137-143 页, 2018 年 11 月。
- [25] G. Liu and H. Shen, "Minimum-cost cloud storage service across multiple cloud providers," *IEEE/ACM Trans.Netw.* 4, pp.2498-2513, Aug. 2017.
- [26] H.Feng, J. Llorca, A. M. Tulino, and A. F. Molisch, "Optimal dynamic cloud network control," *IEEE/ACM Trans.Netw.* 5, pp.2118-2131页, 2018年10月。
- [27] Y.Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: 通信视角", 《*IEEE Commun.Surveys Tuts.*》, vol. 19, no.4, pp.
- [28] Y.Zhang, X. Lan, Y. Li, L. Cai, and J. Pan, "Efficient computation resource management in mobile edge-cloud computing," *IEEE Internet Things J.*, vol. 6, no. 2, pp.
- [29] H.Mendelson 和 S. Whang, "M/M/1 队列的最优激励兼容优先定价", 《*运营. Res.*》, vol. 38, no.5, 第 870-883 页, 1990 年 10 月。
- [30] WikiHow 工作人员。(2019年3月29日)。如何在 Microsoft Windows 中测试网络和互联网延迟(滞后)。[Online].Available: [https://www.wikihow.com/Test-Network-and-Internet-Latency-\(Lag\)-in-Microsoft-](https://www.wikihow.com/Test-Network-and-Internet-Latency-(Lag)-in-Microsoft-) [Online]. 视窗
- [31] Y.Hao, Y. Miao, L. Hu, M. S. Hossain, G. Muhammad, and S. U. Amin, "Smart-Edge-CoCaCo: AI-enabled smart edge with joint computation, caching, and communication in heterogeneous IoT," *IEEE Netw.*, vol. 33, no. 2, pp.
- [32] S.Sundar 和 B. Liang, "Offloading dependent tasks with communication delay and deadline constraint," in *Proc.Comput.Commun.*, 2018 年 4 月, 第 37-45 页。
- [33] L.L. F. Bittencourt, J. Diaz-Montes, R. Buyya, O. F. Rana, and M. Parashar, "Mobility-aware application scheduling in fog computing," *IEEE Cloud Comput.*, vol. 4, no. 2, pp.
- [34] N.Jain and J. Lakshmi, "PriDyn: Enabling differentiated I/O services in cloud using dynamic priorities," *IEEE Trans.服务计算*, 第 8 卷, 第 2 期, 第 212-224 页, 2015 年 3 月。
- [35] I.Menache, A. Ozdaglar, and N. Shimkin, "Socially optimal pricing of cloud computing resources," in *Proc.ICST Conf.Perform.Eval.方法工具*, 2011 年, 第 322-331 页。
- [36] V.V. S. Prasad、B. Vh 和 T. A. Koka, "单队列多服务器和多队列多服务器排队模型的数学分析: Comparison Study," *Global J. Math.Anal.* 3, pp.
- [37] J.J. D. C. Little and S. C. Graves, "Little's law," in *Building Intuition.美国马萨诸塞州波士顿: 施普林格出版社*, 2008 年, 第 81-100 页。
- [38] W.Grassmann, "The convexity of the mean queue size of the M/M/c queue with respect to the traffic intensity," *J. Appl.* 4, pp.
- [39] H.H. L. Lee and M. A. Cohen, "A note on the convexity of performance measures of M/M/c queueing systems," *J. Appl. Probab.* 4, pp.第 920-923 页, 1983 年 12 月。
- [40] Y.Zhang, P. You, and L. Cai, "Optimal charging scheduling by pricing for EV charging station with dual charging modes," *IEEE Trans.Intell.Transp.* 20卷, 第9期, 第3386-3396页, 2019年9月。
- [41] Y.Zhang, S. He, and J. Chen, "Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks," *IEEE/ACM Trans.Netw.* 3, pp.
- [42] X.Hu, K.-K. Wong, and K. Yang, "Wireless powered cooperation-assisted mobile edge computing," *IEEE Trans.无线通信*, 第 17 卷, 第 4 期, 第 2375-2388 页, 2018 年 4 月。4, pp.

**张勇民** (IEEE 会员) 于 2015 年获得中国杭州浙江大学控制科学与工程博士学位。2015 年至 2019 年, 他在加拿大不列颠哥伦比亚省维多利亚市维多利亚大学电气与计算机工程系任博士后研究员。现为中南大学计算机科学与工程学院教授。他的研究兴趣包括无线网络、智能电网和移动计算中的资源管理和优化。他的研究领域包括无线网络、智能电网和移动计算中的资源管理和优化。

荣获 IEEE PIMRC'12 最佳论文奖和 2018 年 IEEE 亚太地区优秀论文奖。

论文奖。她是电气和电子工程师学会维多利亚分会车辆技术和通信联合会分会的共同创始人和主席。她已当选为 IEEE Vehicular Technology Society 理事会成员 (2019-2021 年)。



**兰小龙** 分别于 2012 年和 2019 年获得成都理工大学数学与应用数学学士学位和西南交通大学信息与通信工程博士学位。2017 年至 2019 年, 他在加拿大不列颠哥伦比亚省维多利亚市维多利亚大学攻读访问博士生。他目前是中国成都四川大学网络安全学院的助理研究员。他目前的研究兴趣包括物理层安全、

缓冲辅助通信、能量收集无线通信以及移动边缘计算。



**鞠仁** (IEEE 会员) 分别于 2009 年、2012 年和 2016 年获得中国中南大学计算机科学学士、硕士和博士学位。2013 年至 2015 年, 他在加拿大滑铁卢大学电子与计算机工程系攻读访问博士生。现任清华大学计算机科学与技术系副教授, 中南大学计算机科学与工程学院兼职教授、

中国。他的研究兴趣包括物联网、无线通信、网络计算和云计算。他曾多次获得 IEEE 旗舰会议最佳论文奖, 包括 IEEE ICC'19 和 IEEE HPCC'19 等, 以及 IEEE TCSC 早期职业研究员奖 (2019)。



**Lin Cai** (电气和电子工程师学会会士) 分别于 2002 年和 2005 年获得加拿大滑铁卢大学电气和计算机工程硕士学位和博士学位 (研究生学习杰出成就奖)。自 2005 年以来, 她一直在维多利亚大学电气与计算机工程系工作, 现任教授。她的研究兴趣涉及通信和网络的多个领域, 重点是支持新兴多媒体的网络协议和架构设计。

交通和物联网。她是 NSERC E.W.R. Steacie 纪念奖获得者。研究员。她曾分别于 2010 年和 2015 年获得 NSERC 发现加速器补助金 (DAS), 并获得 2008 年 IEEE ICC 和 2011 年 IEEE WCNC 最佳

