

## 移动边缘计算中的分层资源部署与共享策略

张厚浩, 李晗琳, 高林

(哈尔滨工业大学(深圳)电子与信息工程学院, 广东 深圳 518055)

**摘要:** 随着 5G 时代的到来, 万物互联逐渐成为现实, 然而, 物联网中的很多终端因计算资源不足导致功能受限。移动边缘计算是一种新型网络计算模式, 其终端设备可以将资源敏感型应用卸载到附近计算资源充足的边缘节点, 从而降低运行成本/时延, 提高服务质量 (QoS, quality of service)。针对移动边缘计算资源部署问题进行研究, 考虑网络服务提供商 (ISP, Internet service provider) 支付部署资源的费用, 同时向终端设备出租资源以获得收入的场景。由于物联网终端设备的资源需求是快速随机变化的, 而边缘计算资源的部署是一个长期缓慢变化的过程, 因此, 合理的资源部署策略至关重要。提出了一个基于 5G 网络模型的分层网络结构, 其中, 计算资源可以部署在基站端、汇聚端以及云端等不同的网络层级。基于该分层网络架构, 最优资源部署问题被构建为一个以 ISP 的利润最大化为优化目标的混合整数规划问题, 并基于 MATLAB 中的 CVX 工具箱对该问题进行求解。仿真结果表明, 所提出的移动边缘计算资源分层部署方案在 ISP 收入和部署成本方面的性能均优于平坦型资源部署方案。

**关键词:** 物联网; 移动边缘计算; 分层资源部署

**中图分类号:** TN929.5

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-3750.2021.00200

## Hierarchical resource deployment and sharing strategy in mobile edge computing

ZHANG Houhao, LI Hanlin, GAO Lin

School of Electronic and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

**Abstract:** With the coming of 5G, Internet of things (IoT) has gradually become a reality. However, the function of many terminals is limited to insufficient computing resources in IoT. Mobile edge computing (MEC) is a new network paradigm, where terminals can offload the resource sensitive applications to nearby edge services with rich resources, and thus reduced the operational cost/delay and increased the quality of service (QoS). The resource deployment of MEC was studied, a scenario was considered that the Internet service provider (ISP) paid for deploying the MEC resources, and meanwhile gained revenue from leasing resources to terminals. Note that the deployment of MEC resource is a long-term strategy, while the demands from terminals are time varying. Therefore, it is critical to deploy the MEC resource properly. A hierarchical architecture built upon the 5G radio access network for MEC resource deployment and sharing was proposed, where the MEC resource could be deployed at different network levels. Based on the hierarchical network architecture, the optimal MEC resource deployment problem was formulated and solved as a mixed integer programming problem, aiming at maximizing the ISP's revenue. And the CVX toolbox in MATLAB was used to solve the problem. Simulation results demonstrate that the proposed solution outperforms the flat resource deployment solution in terms of both the ISP's revenue and the deployment cost.

**Key words:** Internet of things, mobile edge computing, hierarchical resource deployment

收稿日期: 2020-06-29; 修回日期: 2020-09-26

基金项目: 国家自然科学基金资助项目 (No.61972113)

**Foundation Item:** The National Natural Science Foundation of China (No.61972113)

## 1 引言

### 1.1 背景和研究意义

近年来,移动边缘计算逐渐成为 5G 和物联网时代的重要技术。移动边缘计算将具有一定计算能力的边缘服务器节点放置在边缘物联网节点附近,作为一种新型的网络模式被用于解决物联网节点计算和通信资源不足等问题<sup>[1]</sup>。在移动边缘计算中,物联网终端可以通过将任务卸载至附近的边缘服务器处理,从而以较小时延完成较多的任务。

显然,移动边缘计算技术的应用前提是网络中的边缘计算节点具备一定的计算资源用于服务距离较近的终端用户任务,从而减少任务流在网络中的传输转发数。然而,由于硬件计算资源在节点上的部署是一次性的(不随时间变化),而终端请求的任务量大小是随时间变化的。因此,在节点上部署资源过多可能会导致资源过剩,造成闲置;而部署的资源过少可能会导致资源不足,增大任务传输时延。由此可见,边缘计算网络中的资源部署问题是移动边缘计算技术应用的基础和重点,同时也是本文的研究重点。

本文重点研究由 ISP 部署边缘资源的场景。其中,ISP 在其网络节点上部署一定的计算资源,并将其转换为具有处理任务能力的边缘节点。由此,ISP 承担资源部署的成本,同时,部署的计算资源用于处理用户卸载至系统的任务,可以为 ISP 带来收益。本文重点研究如何在部署成本最低的情况下获取最大收益,使得 ISP 的利润最大化。

网络的拓扑结构对边缘计算的系统性能和资源部署策略影响很大。本文研究 5G 网络拓扑下的资源部署场景,5G 网络支持核心网的功能部分下沉至距离终端更近的网络节点,构成具有处理能力的边缘云<sup>[2]</sup>,并根据 5G 的网络拓扑结构设计一种分层的边缘计算网络结构。

边缘计算系统通过任务调度将用户的任务分发至不同的节点上进行处理,因此,任务调度方法对边缘计算的系统性能和资源部署策略有很大影响。软件定义网络(SDN, software defined network)通过将网络的控制层和数据层分离来提升任务流调度的灵活性<sup>[3]</sup>。然而,任务在节点之间的传输会产生一定的传输代价,如时延、带宽占用等,从而降低 QoS。QoS 的下降会造成收益的降低,由此可见,在每个节点上部署合理的资源以降低传输时延

对于 ISP 非常重要。本文将在资源部署策略中综合考虑上述因素。

### 1.2 现有研究工作总结

目前,大多数现有研究主要关注如何通过任务卸载<sup>[1]</sup>、资源分配<sup>[4]</sup>以及服务放置<sup>[5]</sup>以减少传输时延或节省终端能量。它们主要关注在边缘计算系统资源已经部署完成的情况下,如何提高边缘计算系统的使用效率。然而,在现实中,系统资源的部署是前提工作,只有少部分工作针对边缘计算中的资源部署进行了研究。文献[6-11]将边缘计算系统部署为平坦型网络结构,它们认为所有节点组成单层的网络,都以最小化传输代价获得最佳系统性能为系统目标。其中,文献[7]和文献[10]实现目标的方法为使系统各节点的负载均衡。在文献[12-14]中,将资源部署在高层节点实现分层的网络结构,但是没有考虑同层节点的资源共享机制。文献[12]认为分层的边缘计算系统在具有峰值负载时可以提升系统资源的利用率。

然而,现有的工作一般认为系统必须将终端请求的任务全部处理完<sup>[6,11,13]</sup>,但是这些目标都有其局限性,即系统必须将终端卸载的任务处理完而不考虑 ISP 的实际利润,但是 ISP 可以根据实际的利润情况拒绝部分请求。同时,以往的研究工作也忽略了终端请求任务量的随机性对系统利润的影响。在实际应用中,不同的终端会有不同的应用,且在不同的时间内有不同的处理需求,各个节点接收的任务量是随时间变化的。

但是,由于部署在节点上的硬件资源难以改变,每个节点的处理能力不会轻易变动。这样很可能会导致一些节点超载(资源不足),而另一些节点闲置(资源过剩)的情况出现。即使多个边缘节点可以通过任务调度的方式协同工作,但是调度产生的传输代价会降低系统收益。显然,系统性能会随着部署资源的增多而变得更好,但同时也增大了部署成本。因此,一个有效的部署策略对 ISP 有重要意义。

### 1.3 本文主要贡献

针对上述问题,本文根据 5G 网络的结构设计了一种有资源共享机制的分层型网络结构边缘计算系统。在靠近用户的低层节点部署资源可以降低传输代价,提高服务质量;在远离用户的高层节点部署资源可以用一定的传输代价来降低资源部署成本,同时,资源共享机制可以使各节点的资源利

用率最大化。

综上所述，任务调度和资源部署问题是耦合在一起的，即部署策略需要任务调度的结果反馈，而任务调度需要运行在已经部署完成的边缘计算系统上。本文研究的主要挑战在于如何通过一次部署使得系统可以应对随机变化的任务请求量，为了解决该问题提出了一种联合资源部署和共享 (JRDS, joint resource deployment and sharing) 方案，通过优化资源的共享方案实现合理的任务调度。通过 JRDS 方案，可以得出最优的资源部署和任务调度策略。

本文的主要贡献总结如下。

1) 本文基于 5G 网络结构提出了一种新型的边缘计算部署模型，该模型具有分层结构和资源共享机制，可以涵盖大多数现有的边缘计算模型。

2) 本文提出了 JRDS 方案解决资源部署和任务调度的耦合性，实现了面向 ISP 利润最大化的最优部署策略。

3) 通过求解优化问题，并将本方案与平坦型网络结构、其他分层型网络结构的边缘计算系统对比，表明本文所提方案可以使 ISP 获得最大利润，并且当任务需求量随机变化时也能保持稳定的系统性能。

## 2 分层型网络结构边缘计算系统

### 2.1 网络结构模型

在 5G 接入网中，网络模块被划分为多个单元：有源天线单元 (AAU, active antenna unit)、分布单元 (DU, distribute unit) 和集中单元 (CU, centralized unit)，这些单元可以灵活地进行结合和分离<sup>[15]</sup>。对应 5G 接入网的单元划分，本文提出的边缘计算系统由基站 (BS, base station)、汇聚节点 (AN, ag-

gregation node) 和核心节点 (CN, core node) 3 类节点组成。ISP 可以在这 3 类节点上部署资源，共享机制的分层型网络结构边缘计算系统如图 1 所示。

所有 BS 构成了网络的第 1 层，所有 AN 构成网络的第 2 层，CN 构成网络的第 3 层。多个 BS 通过前传链路连接至 AN，每个 BS 只能连接一个 AN，每个 AN 可以连接多个 BS 形成集合，所有 AN 通过回传链路连接至 CN。同层节点的资源共享机制为：连接至同一个 AN 的 BS 之间可以通过 AN 作为中转节点传输任务，AN 之间可以通过 CN 传输任务，并且本文规定不是共享的任务流只能自下向上传输。

本文用  $\mathcal{K} = \{1, 2, \dots, K\}$  表示 AN 的集合， $\mathcal{N} = \{1, 2, \dots, N\}$  表示 BS 的集合， $\mathcal{N}_k = \{1, 2, \dots, N_k\}$  表示连接至 AN  $k$  的 BS 集合。其中， $N_k^i \in \mathcal{N}_k$  表示连接至 AN  $k$  的 BS  $i$ ，本文将包含 AN  $k$  和 BS 集  $\mathcal{N}_k$  的集合统称为区域  $k$ 。

### 2.2 部署模型

在节点上部署资源需要 ISP 支付费用，包括固定费用 (机房建造、空调制冷和维护等费用) 和动态费用 (由部署的资源量决定)。值得注意的是，低层节点的固定费用要高于高层节点的固定费用<sup>[13]</sup>。

本文用  $r_{BS}^{k,i}$ 、 $r_{AN}^k$  和  $r_{CN}$  分别表示资源部署在区域  $k$  内的 BS  $i$ 、AN  $k$  和 CN 上的固定费用。在动态部署费用方面，ISP 需要支付费用  $p$  以获取 1 单位计算资源，该计算资源有  $s$  单位的处理能力 (可处理任务量的缓冲量)。每个节点需要做两个部署决定：是否部署资源和部署的资源量大小。对于是否部署资源，本文引入 0-1 变量  $a_{BS}^{k,i}$ 、 $a_{AN}^k$  和  $a_{CN}$  分别表示是否在 BS、AN 和 CN 处部署资源 ( $a=0$  表示不部署资源， $a=1$  表示部署资源)。对于部署的

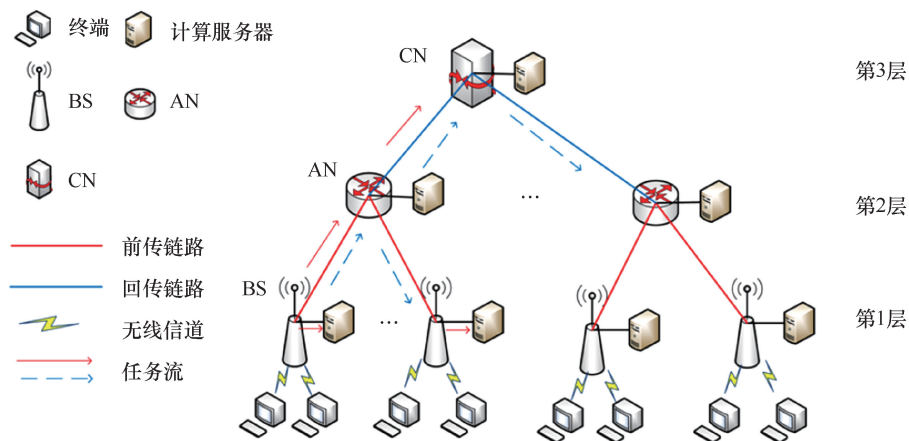


图 1 共享机制的分层型网络结构边缘计算系统

资源量, 本文引入非负整数变量  $b_{BS}^{k,i}$ 、 $b_{AN}^k$  和  $b_{CN}$  分别表示在 BS、AN 和 CN 上部署资源的数量。上述 6 个变量为 JRDS 方案中的部分决策变量。

### 2.3 任务模型

每个 BS 都会覆盖一片区域, 区域内有若干物联网终端设备, 这些区域相互之间没有重叠。根据前面的讨论, 每个区域内的任务请求是时变的, 为了刻画这种时变属性, 本文分时隙进行讨论, 时长为  $\tau$ , 下标为  $t$ , 且  $t \in \mathcal{T} = \{0, 1, \dots, T\}$ 。本文用  $\lambda_{UE}^{k,i}(t)$  表示区域  $k$  内第  $i$  个 BS 在  $t$  时刻接收的任务请求量。

本文用  $v$  表示 ISP 在处理完 1 单位用户任务后所获取的收益。被 ISP 接收的任务会被系统调度至合适的节点上进行处理, 并且每个任务可以被分割为任意部分并被不同的节点处理<sup>[8]</sup>。本文用  $\lambda_{BS}^{k,i}(t)$ 、 $\lambda_{AN}^k(t)$  和  $\lambda_{CN}(t)$  分别表示 BS、AN 和 CN 在  $t$  时刻的工作负载, 即需要处理的任务量。

### 2.4 调度模型

若当前节点没有足够的缓存量接收任务, 任务需要通过 SDN 的控制传输至其他节点处理。为了使方案更加通用, 本文认为任意两个可以路由到达的节点之间可以互相传输任务。由于划分的时隙较大, 认为可以完成在时隙内所有节点上的任务, 各个边缘节点的计算资源在时隙末被释放。本文重点关注每个时隙内的任务调度情况, 并且由于是线性问题, 所以通过 CVX 工具箱可以得到当前资源部署情况下关于任务调度的最优策略。

图 1 展示了在 SDN 调度下的任务流流向。物联网终端通过无线链路卸载任务至 BS, BS 接收部分任务至网络中处理。对于 BS, 共有 4 种任务流方向: 在当前 BS 处理、传输至 AN 处理、传输至同区域内的 BS 处理和以上 3 个方向的组合处理。AN 的任务流流向同 BS 类似, 本文引入 4 个非负连续型决策变量来表示系统中任务流调度过程。本文用  $x_k^i(t)$ 、 $y_k^{i_1, i_2}(t)$ 、 $z_k(t)$  和  $w_{k_1, k_2}(t)$  分别表示任务流从 BS  $i$  至 AN  $k$ 、从区域  $k$  内的 BS  $i_1$  至 BS  $i_2$ 、从 AN  $k$  至 CN 和从 AN  $k_1$  至 AN  $k_2$  的任务流大小。其中,  $i_1 = i_2$  表示 BS  $i_1$  本地处理的任务流,  $k_1 = k_2$  表示 AN  $k_1$  本地处理的任务流。

### 2.5 系统传输代价模型

在实际应用中, 边缘节点的计算能力远大于终端节点的计算能力, 每个节点只需要很短的时间即可处理完成终端的任务请求。因此相较于计算代

价, 本文更关注由节点间任务传输产生的传输代价, 忽略节点上的计算代价。这里有两种数据传输形式: 终端与节点间的无线传输、节点间的有线传输。其中, 由于移动边缘计算系统不受无线传输的影响, 因此本文只关注由节点间有线传输产生的代价。传输代价主要受两个维度的影响, 包括传输数据大小以及传输链路的长度、损耗。

为了量化任务流和链路的关系, 本文定义  $\beta_k^i$ 、 $\alpha_k^{i_1, i_2}$  分别为从 BS  $i$  向 AN  $k$  以及从 BS  $i_1$  至 BS  $i_2$  传输单位任务时的代价, 其中  $\alpha_k^{i_1, i_2} = \beta_k^{i_1} + \beta_k^{i_2}$ 。  $\delta_k$ 、 $\gamma_{k_1, k_2}$  分别表示从 AN  $k$  向 CN 以及从 AN  $k_1$  至 AN  $k_2$  传输单位任务的代价, 其中  $\gamma_{k_1, k_2} = \delta_{k_1} + \delta_{k_2}$ 。同时, 任务在链路传输过程中也会占用链路的带宽, 而链路容量是有限的。本文用  $h_{BS}^{k,i}$  和  $h_{AN}^k$  分别表示从 BS  $i$  至 AN  $k$  和从 AN  $k$  至 CN 的链路容量。

## 3 建立 JRDS 优化问题

系统的传输时延会因部署资源的增多而减小, 但同时部署资源的增多又会增加部署成本。降低部署成本和降低传输代价之间存在矛盾, 并且这种矛盾会因终端需求的随机性加剧。

在大多数现有的工作中, 需要系统必须将终端卸载的任务处理完, 不考虑 ISP 的实际利润。在实际应用中, ISP 不仅要承担部署成本, 还要对终端的 QoS 负责, 并且仅在完成一定任务后才能获取收益。由于 ISP 是理性的, 所以系统会拒绝收益小于代价的任务。JRDS 方案的目标是在长期条件下获取最大利润, 本文需要找出最优的资源部署策略实现这个目标。

### 3.1 限制条件

#### 3.1.1 链路容量限制

链路中的任务流总和不能超过链路容量。本文存在两种链路: 前传链路和回传链路。对于每个节点而言, 无论是来自其他节点的输入量还是传输至其他节点的输出量都会占用相应链路的容量。

对于区域  $k$  内的 BS  $i$  与 AN  $k$  之间的前传链路, 其任务传输量应满足式(1)。

$$\sum_{i \in \mathcal{N}_k^i/i} y_k^{i,i}(t) + x_k^i(t) + \sum_{i_1 \in \mathcal{N}_k^i/i} y_k^{i_1,i}(t) \leq h_{BS}^{k,i}, \forall i \in \mathcal{N}_k, \forall k \in \mathcal{K} \quad (1)$$

其中,  $y_k^{i,i}(t)$  为传输至同区域内其他 BS 的任务量,

$x_k^i(t)$  为传输至 AN 的任务量,  $y_k^{i,j}(t)$  为接收的来自同区域内其他 BS 的任务量,  $h_{BS}^k$  为链路容量。

同式(1)类似, 对于 AN  $k$  与 CN 之间的回传链路, 其任务传输量应满足式(2)。

$$\sum_{k_1 \in \mathcal{K}/k} w_{k,k_1}(t) + z_k(t) + \sum_{k_1 \in \mathcal{K}/k} w_{k_1,k}(t) \leq h_{CN}^k, \forall k \in \mathcal{K} \quad (2)$$

其中,  $w_{k,k_1}(t)$  为传输至其他 AN 的任务量,  $z_k(t)$  为传输至 CN 的任务量,  $w_{k_1,k}(t)$  为接收的来自其他 AN 的任务量。

### 3.1.2 处理能力限制

由于节点上部署的资源有限, 每个节点的处理能力(可容纳的任务量)也是有限的, 节点上的负载不能超出其处理能力。BS 的处理能力限制如式(3)所示。

$$\sum_{j \in \mathcal{N}_k} y_k^{j,i}(t) \leq b_{BS}^{k,i} \cdot s, \forall i \in \mathcal{N}_k, \forall k \in \mathcal{K} \quad (3)$$

其中,  $y_k^{j,i}(t)$  为 BS  $i$  在  $t$  时刻的工作负载,  $b_{BS}^{k,i} \cdot s$  为其处理能力上限。

同理, AN 上的处理能力限制如式(4)所示。

$$\sum_{i \in \mathcal{N}_k} x_k^i(t) + \sum_{k_2 \in \mathcal{K}} w_{k_2,k}(t) \leq b_{AN}^k \cdot s, \forall k \in \mathcal{K} \quad (4)$$

其中,  $x_k^i(t)$  为所覆盖的 BS 传输至 AN  $k$  的任务量,  $w_{k_2,k}(t)$  为其他 AN 传输至 AN  $k$  以及 AN  $k$  自身的任务量总和,  $b_{AN}^k \cdot s$  为 AN 的处理能力。

CN 上的处理能力限制如式(5)所示。

$$\sum_{k \in \mathcal{K}} z_k(t) \leq b_{CN} \cdot s, \forall k \in \mathcal{K} \quad (5)$$

其中,  $z_k(t)$  为 CN 的工作负载,  $b_{CN} \cdot s$  为其处理能力上限。

### 3.1.3 调度限制

在移动边缘计算系统中, BS 作为终端需求的入网口, 其所接收的任务量在本地处理或传输至网络中的其他节点处理。因此, 对于 BS 而言, 其所接收的任务量不能超过用户请求的任务量, 如式(6)所示。

$$\sum_{i_1 \in \mathcal{N}_k} y_k^{i_1,i}(t) + x_k^i(t) \leq \lambda_{UE}^{k,i}(t), \forall i \in \mathcal{N}_k, \forall k \in \mathcal{K} \quad (6)$$

系统一旦接收了用户请求的任务, 就必须将这些任务处理完成。系统第 2 层和第 3 层的节点必须将第 1 层向上传输的任务处理完成, 即 AN 和 CN 的工作负载总和为所有 BS 接收的任务量, 如式(7)所示。

$$\sum_{k \in \mathcal{K}} \left[ \sum_{k_1 \in \mathcal{K}} w_{k,k_1}(t) \right] + \sum_{k \in \mathcal{K}} z_k(t) = \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}_k} x_k^i(t) \quad (7)$$

## 3.2 系统代价

本文主要考虑以下三部分系统代价/收益: 传输代价、部署代价和任务收益。

### 3.2.1 传输代价

任务传输产生传输代价, 其主要是链路长度和链路损耗带来的时延代价, 在本文中该项需要结合调度策略来计算, 在链路间传输的任务量同单位代价相乘即传输代价。系统第 1 层和第 2 层节点之间的传输代价由任务从 BS 传向同区域内 BS、从 BS 传向 AN 这两部分组成, 用  $T_{trans}^{BS}$  表示此部分传输代价, 如式(8)所示。

$$T_{trans}^{BS}(t) = \sum_{k \in \mathcal{K}} \left[ \sum_{i \in \mathcal{N}_k} \left( \sum_{i_1 \in \mathcal{N}_k/i} y_k^{i_1,i}(t) \cdot \alpha_k^{i,i_1} + x_k^i(t) \cdot \beta_k^i \right) \right] \quad (8)$$

与式(8)相似, 系统第 2 层和第 3 层节点之间的传输代价由任务自 AN 传至 AN、从 AN 传至 CN 组成, 如式(9)所示。

$$T_{trans}^{AN}(t) = \sum_{k \in \mathcal{K}} \left( \sum_{k_1 \in \mathcal{K}/k} w_{k,k_1}(t) \cdot \gamma_{k,k_1} + z_k(t) \cdot \delta_k \right) \quad (9)$$

因此, 总体传输代价如式(10)所示。

$$T_{trans}(t) = T_{trans}^{BS}(t) + T_{trans}^{AN}(t) \quad (10)$$

### 3.2.2 部署代价

部署代价由固定部署代价和动态部署代价两部分组成, 固定部署代价如式(11)所示。

$$C_f = \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}_k} r_{BS}^{k,i} \cdot a_{BS}^{k,i} + \sum_{k \in \mathcal{K}} r_{AN}^k \cdot a_{AN}^k + r_{CN} \cdot a_{CN} \quad (11)$$

其中,  $r_{BS}^{k,i} \cdot a_{BS}^{k,i}$  为系统第 1 层的固定部署代价,  $r_{AN}^k \cdot a_{AN}^k$  为系统第 2 层的固定部署代价,  $r_{CN} \cdot a_{CN}$  为系统第 3 层的固定部署代价。

动态部署代价取决于部署资源的数量, 分别由 BS、AN、CN 部署的资源量组成, 如式(12)所示。

$$C_d = \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}_k} p \cdot b_{BS}^{k,i} + \sum_{k \in \mathcal{K}} p \cdot b_{CU}^k + p \cdot b_{cloud} \quad (12)$$

### 3.2.3 任务收益

ISP 在完成所接收的任务后获得收益, 完成的任务量为系统第 1 层节点所接收的任务量总和, 式(13)为系统的总收益。

$$V(t) = \sum_{k \in \mathcal{K}} \left[ \sum_{i_1 \in \mathcal{N}_k} \left[ \sum_{i_2 \in \mathcal{N}_k} y_k^{i_1,i_2}(t) + x_k^{i_1}(t) \right] \right] \cdot v \quad (13)$$

### 3.3 优化问题建立

基于上述讨论，建立如下资源部署与共享联合优化问题即 JRDS 问题，该模型在任务调度、资源的限制条件下，以最小部署成本和传输时延完成尽可能多的任务量，详见式(14)。

$$\begin{aligned}
 & \max_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{a}, \mathbf{b}} && \frac{1}{T} \sum_{t \in \mathcal{T}} U(V(t), T_{\text{trans}}(t)) - (C_f + C_d) \\
 \text{s.t.} & && (1) \sim (7) \\
 & && x_k^i(t) \geq 0, \forall i \in N_k, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \\
 & && y_k^{i_1, i_2}(t) \geq 0, \forall i_1, i_2 \in N_k, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \\
 & && z_k(t) \geq 0, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \\
 & && w_{k_1, k_2}(t) \geq 0, \forall k_1, k_2 \in \mathcal{K}, \forall t \in \mathcal{T}
 \end{aligned} \tag{14}$$

其中， $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{a}, \mathbf{b}$  是决策变量的向量形式。 $U(\cdot)$  是与收益成正比、与传输代价成反比的效用函数， $\eta$  为收益和传输代价之间的系数。

$$U(V(t), T_{\text{trans}}(t)) = V(t) - \eta \cdot T_{\text{trans}}(t) \tag{15}$$

ISP 不仅要最小化部署成本和最大化收益，还要保证 QoS 尽可能小；部署策略是时不变的，而收益和传输代价在每一时刻随着终端输入的任务量变化而发生变化，即全局的部署策略要根据时变的任务调度情况做出调整，同时任务调度策略要依托已经部署完成的系统进行。这两个问题是耦合的，本文需要同时对这两者进行优化。在所给出的目标函数中， $U(\cdot)$  用于表征系统性能，其是一个长期的累积项，该项累积时间越长，最后得出的部署策略越稳定。本文最后对该项求时间平均以保证得出的部署策略对于任何时刻都是适用的。总体上，该优化问题是一个混合整数线性规划问题，是 NP-hard 问题，本文利用 CVX 工具箱求出最优部署策略。

## 4 仿真结果与分析

本节进行数值仿真，测试 JRDS 的系统性能。本文设置的网络拓扑结构包括 1 个 CN、2 个 AN、每个 AN 下连接 10 个 BS。本文进行 30 个时隙的仿真，即  $T = 30$ 。为了便于表示，所有参数采用归一化的相对量，本文重点关注相对量的改变对系统性能的影响情况。BS 所覆盖的区域内终端请求的任务量服从正态分布： $N(20, 10) \geq 0$ ，其中 10 为初始化方差，表示任务请求量的随机性。仿真参数设置如表 1 所示。

表 1 仿真参数设置

符号	定义	数值
$N$	BS 的数量	20
$K$	AN 的数量	2
$r_{\text{BS}}^{k,i}$	BS 的固定部署成本	$U(2,4)$
$r_{\text{AN}}^k$	AN 的固定部署成本	$U(1,3)$
$r_{\text{CN}}$	CN 的固定部署成本	$U(1,2)$
$p$	单位资源的成本	0.3
$s$	单位资源的处理能力	1
$h_{\text{BS}}^{k,i}$	BS $i$ 与 AN $k$ 之间的链路容量	$U(40,50)$
$h_{\text{AN}}^k$	AN $k$ 与 CN 之间的链路容量	$U(200,250)$
$\beta_k^i$	BS 与 AN 之间的任务传输单位代价	$U(0.075, 0.15)$
$\delta_k$	AN 与 CN 之间的任务传输单位代价	$U(0.25, 0.3)$
$\eta$	收益与传输代价之间的系数	0.7

本文将提出的 JRDS 策略与平坦型部署策略、两层型部署策略进行对比。平坦型部署策略是指资源仅部署在 BS 上，网络中不存在高层节点和资源共享机制。两层型部署策略是指由 BS 和 AN 构成两层网络，无 CN，AN 之间无法共享资源。

### 4.1 不同部署策略的系统性能

不同部署策略的系统性能如图 2 所示。显然，本文提出的 JRDS 策略可以在最低部署成本的情况下得到最大的利润，平坦型部署策略的总体利润最低。然而，JRDS 策略付出了一定的传输代价，一定程度上降低了 QoS。由于平坦型部署策略没有资源共享，因此不存在传输代价。

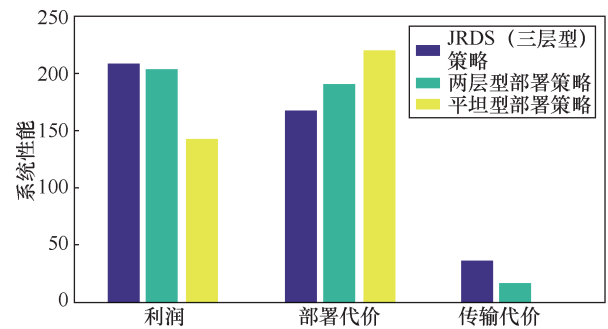


图 2 不同部署策略的系统性能

在固定部署成本下不同部署策略的表现如图 3 所示，从图 3 可以看出，在固定部署成本下，JRDS 策略可以达到最高利润。因为 JRDS 通过同层节点资源共享以及高低层节点协助实现资源最大化利用。

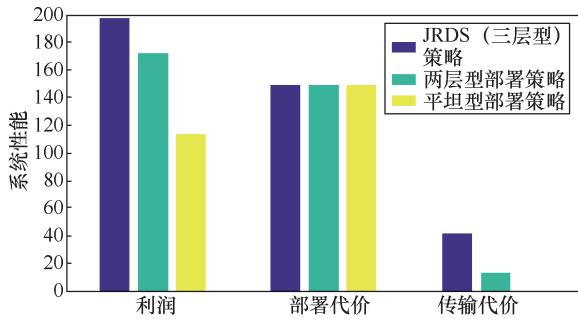
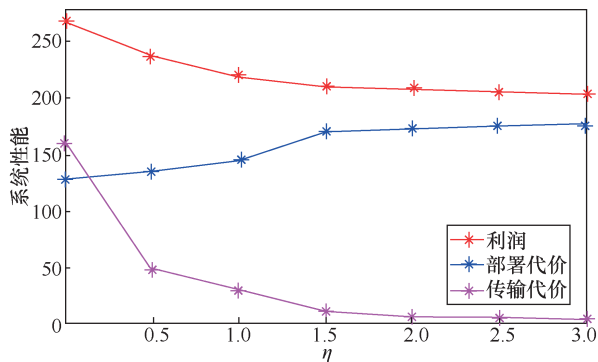


图 3 在固定部署成本下不同部署策略的表现

## 4.2 $\eta$ 对系统性能的影响

系统性能与  $\eta$  的关系如图 4 所示。当  $\eta$  增大时，系统利润逐渐减少，部署成本随之增大，传输代价随之减少。当  $\eta$  很大时，部署策略对于传输代价更敏感，JRDS 此时会在节点上部署更多资源以保证较低的传输代价，进而使得部署代价增加。由图 4 还可以看出，当  $\eta=0$  时，系统可以获得最大利润和最小部署成本。这是因为此时 ISP 仅需承担较少的部署成本，将资源部署在高层节点上并维持系统运行即可，而不用考虑由低层节点向高层节点调度任务时产生的传输代价。

图 4 系统性能与  $\eta$  的关系

## 4.3 任务量随机性对系统性能的影响

在改变终端请求任务量大小的随机性情况下，本文将 JRDS 分层策略同平坦型部署策略进行了对比。首先在终端请求任务量的随机性为 10 的情况下，分别得出 JRDS 分层型和平坦型网络结构的最优部署策略。然后控制该种部署方式不变，逐渐增大终端请求任务量的随机性，测试系统性能，改变任务量随机性后的系统表现如图 5 所示。

图 5 展示了两种策略在增大任务量随机性后的系统利润浮动，JRDS 的利润浮动方差为 19.450 4，远小于平坦型网络结构的利润浮动方差 150.234 2。并且平坦型部署策略的变化情况逐渐呈下降趋势，

且整体远小于 JRDS 策略。这显示出 JRDS 策略有较强的抵抗任务量随机变化的能力，因为 JRDS 策略可以灵活将过载节点上的任务传输至闲置节点，以平衡各节点之间的负载。

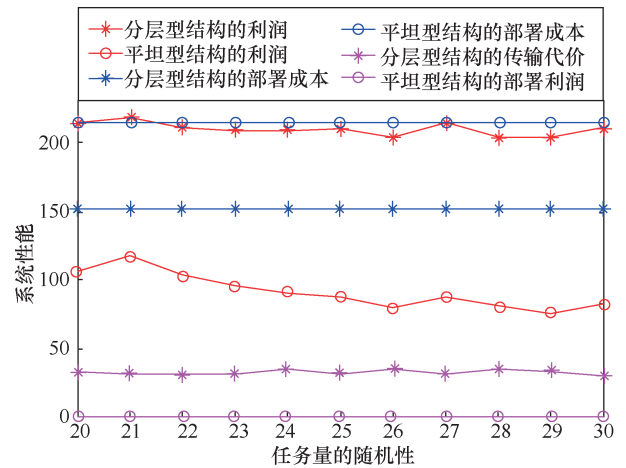


图 5 改变任务量随机性后的系统表现

## 5 结束语

在移动边缘计算的资源部署问题中，合理的资源部署策略可以提升网络中资源的利用效率，降低 ISP 的部署成本，提高利润。本文基于 5G 无线接入网结构，并结合 SDN 技术提出了一种资源共享机制的分层型资源部署策略——JRDS 策略，利用 CVX 工具箱对建立的混合整数线性规划问题进行求解。仿真结果表明，分层型网络结构通过牺牲部分传输代价能实现更优的系统性能，且网络层数越多，系统性能越好。本文提出的 JRDS 策略在对抗任务量随机性增大的情况时，由于资源共享机制，使得系统的表现很稳定。在未来的工作中，将考虑更精细化的网络结构和资源调度机制，在保证终端 QoS 的前提下实现以更低的部署成本得到更高的收益。

## 参考文献：

- [1] XU J, CHEN L X, ZHOU P. Joint service caching and task offloading for mobile edge computing in dense networks[C]//Proceedings of 2018 IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 207-215.
- [2] KEKKI S, FEATHERSTONE W, FANG Y, et al. MEC in 5G networks[S]. 2018.
- [3] KREUTZ D, RAMOS F M V, VERÍSSIMO P E, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1): 14-76.
- [4] GUO H Z, ZHANG J, LIU J J, et al. Energy-efficient task offloading and transmit power allocation for ultra-dense edge compu-

- ting[C]//Proceedings of 2018 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2018: 1-6.
- [5] WANG L, JIAO L, HE T, et al. Service entity placement for social virtual reality applications in edge computing[C]//Proceedings of 2018 IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 468-476.
- [6] FAN Q, ANSARI N. Cost aware cloudlet placement for big data processing at the edge[C]//Proceedings of 2017 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2017: 1-6.
- [7] XU Z, LIANG W, XU W, et al. Capacitated cloudlet placements in wireless metropolitan area networks[C]//Proceedings of 2015 IEEE 40th Conference on Local Computer Networks (LCN). Piscataway: IEEE Press, 2015: 570-578.
- [8] SANG Y, JI B, GUPTA G R, et al. Provably efficient algorithms for joint placement and allocation of virtual network functions[C]//Proceedings of 2017 IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.
- [9] DUAN P F, LI Q, JIANG Y, et al. Toward latency-aware dynamic middlebox scheduling[C]//Proceedings of 2015 24th International Conference on Computer Communication and Networks (ICCCN). [S.l.:s.n.], 2015: 1-8.
- [10] JIA M K, LIANG W F, XU Z C, et al. Cloudlet load balancing in wireless metropolitan area networks[C]//Proceedings of the 35th Annual IEEE International Conference on Computer Communications. Piscataway: IEEE Press, 2016: 1-9.
- [11] LI D W, DONG B X, WANG E, et al. A study on flat and hierarchical system deployment for edge computing[C]//Proceedings of 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). Piscataway: IEEE Press, 2019.
- [12] TONG L, LI Y, GAO W. A hierarchical edge cloud architecture for mobile computing[C]//Proceedings of the 35th Annual IEEE International Conference on Computer Communications. Piscataway: IEEE Press, 2016: 1-9.
- [13] LIU Z, ZHANG J W, JI Y F. Cost aware mobile edge computing hierarchical deployment in optical interconnection network[C]//Proceedings of 2018 Asia Communications and Photonics Conference (ACP). [S.l.:s.n.], 2018: 1-3.
- [14] CHEN Y, WU J, JI B. Virtual network function deployment in tree-structured networks[C]//Proceedings of 2018 IEEE 26th International Conference on Network Protocols (ICNP). Piscataway: IEEE Press, 2018: 132-142.
- [15] 3GPP. Study on new radio access technology: radio access architecture and interfaces[S]. 2017.

## [作者简介]



张厚浩（1996- ），男，哈尔滨工业大学（深圳）电子与信息工程学院硕士生，主要研究方向为移动边缘计算中的资源部署优化策略。



李晗琳（1996- ），女，哈尔滨工业大学（深圳）电子与信息工程学院硕士生，主要研究方向为移动边缘计算中的资源部署、任务卸载、经济模型分析等。



高林（1980- ），男，博士，哈尔滨工业大学（深圳）电子与信息工程学院副教授，主要研究方向为移动边缘计算、群智计算、群体智能、博弈论、强化学习等。