计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(10):99−107 [doi: 10.15888/j.cnki.csa.008754]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

移动边缘计算网络中的资源分配与定价①

吕晓东[1], 邢焕来[2], 宋富洪[2], 王心汉[2]

[1](西南交通大学 信息科学与技术学院, 成都 610031)
[2](西南交通大学 计算机与人工智能学院, 成都 610031)
通信作者: 邢焕来, E-mail: hxx@home.swjtu.edu.cn

摘　要: 移动边缘计算 (mobile edge computing, MEC) 使移动设备 (mobile device, MD) 能够将任务或应用程序卸载到 MEC 服务器上进行处理. 由于 MEC 服务器在处理外部任务时消耗本地资源, 因此建立一个向 MD 收费以奖励 MEC 服务器的多资源定价机制非常重要. 现有的定价机制依赖于中介机构的静态定价, 任务的高度动态特性使得实现边缘云计算资源的有效利用极为困难. 为了解决这个问题, 我们提出了一个基于 Stackelberg 博弈的框架, 其中 MEC 服务器和一个聚合平台 (aggregation platform, AP) 充当跟随者和领导者. 我们将多重资源分配和定价问题分解为一组子问题, 其中每个子问题只考虑一种资源类型. 首先, 通过 MEC 服务器宣布的单价, AP 通过解决一个凸优化问题来计算 MD 从 MEC 服务器购买的资源数量. 然后, MEC 服务器计算其交易记录, 并根据多智能体近端策略优化 (multi-agent proximal policy optimization, MAPPO) 算法迭代调整其定价策略. 仿真结果表明, MAPPO 在收益和福利方面优于许多先进的深度强化学习算法.

关键词: 移动边缘计算 (MEC); 资源定价; 博弈论; 深度强化学习; 资源分配

引用格式: 吕晓东,邢焕来,宋富洪,王心汉.移动边缘计算网络中的资源分配与定价.计算机系统应用,2022,31(10):99–107. http://www.c-s-a.org.cn/1003-3254/8754.html

Resource Allocation and Pricing in Mobile Edge Computing Networks

LYU Xiao-Dong[1], XING Huan-Lai[2], SONG Fu-Hong[2], WANG Xin-Han[2]

[1](School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)
[2](School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Mobile edge computing (MEC) enables mobile devices (MDs) to offload tasks or applications to MEC servers for processing. As a MEC server consumes local resources when processing external tasks, it is important to build a multi-resource pricing mechanism that charges MDs to reward MEC servers. Existing pricing mechanisms rely on the static pricing of intermediaries. The highly dynamic nature of tasks makes it extremely difficult to effectively utilize edge-cloud computing resources. To address this problem, we propose a Stackelberg game-based framework in which MEC servers and an aggregation platform (AP) act as followers and the leader, respectively. We decompose the multi-resource allocation and pricing problem into a set of subproblems, with each subproblem only considering a single resource type. First, with the unit prices announced by MEC servers, the AP calculates the quantity of resources for each MD to purchase from each MEC server by solving a convex optimization problem. Then, each MEC server calculates its trading records and iteratively adjusts its pricing strategy with a multi-agent proximal policy optimization (MAPPO) algorithm. The simulation results show that MAPPO outperforms a number of state-of-the-art deep reinforcement learning algorithms in terms of payoff and welfare.

Key words: mobile edge computing (MEC); resource pricing; game theory; deep reinforcement learning; resource allocation

① 收稿时间: 2022-01-22; 修改时间: 2022-02-22; 采用时间: 2022-03-10; csa 在线出版时间: 2022-06-28

1 引言

随着移动应用和物联网的快速发展, 功率和计算资源有限的移动设备 (MDs) 不再满足资源密集型应用的严格要求, 如低延迟、高可靠性、用户体验连续性的要求[1]. 在移动边缘计算 (MEC) 中, 网络运营商和服务提供商进行合作, 在网络边缘提供优秀的通信和计算资源, 以增强 MD 能力[2].

在 MEC 中, 计算资源的管理在提高资源利用率和优化系统资源效益方面起着至关重要的作用[3]. 由 MEC 服务器处理外部任务会消耗本地计算资源. 因此, 每个 MD 根据某种资源定价机制支付一定的服务费, 以激励边缘云提供足够的计算资源.

现有的定价机制, 如基于拍卖的, 依赖于中间机构的静态定价[4–6]. 拍卖双方都需要向中介提供的服务支付费用. 总成本的增加使得双方都无法从资源交易中获得最佳的利益. 同时, 静态定价也不能适应 MD 不断变化的资源需求. 在这种情况下, MEC 服务器很难有效地利用其本地资源. 因此, 一个关键的问题是如何有效地将有限的 MEC 资源分配给具有不同需求和偏好的相互竞争的 MD. 为此, 我们将 MEC 环境中的多资源分配和定价问题表述为一个包含一个领导者和多个跟随者的多阶段 Stackelberg 博弈, 其中所有 MEC 服务器同时将它们的实时价格发送到一个聚合平台 (AP). 在第 1 阶段, 通过 MEC 服务器公布的价格, AP 通过解决一个效用最大化问题来找到最优的资源分配解决方案. 在第 2 阶段, 基于环境的反馈 (即资源如何分配给 MDs), 我们利用多代理近端策略优化 (MAPPO) 算法[7] 学习每个 MEC 服务器的最优定价策略, 这种策略不需要 MEC 服务器获取其他 MEC 服务器实现的定价策略就可以做出最优的决策.

我们的主要贡献如下: (1) 我们通过构建一个 Stackelberg 的领导者-跟随者博弈, 充分考虑了 AP 和 MEC 服务器之间的互动以及服务器之间的竞争. 纳什均衡给出了完全竞争的合理结果. (2) 该算法基于 MAPPO, 这是一种具有集中式训练和分布式执行的深度强化学习算法, 能够适应环境动力学和参数的不确定性. 每个 MEC 服务器作为一个智能体, 不断地与环境交互, 以生成一系列的培训经验. 智能体既不需要事先了解 MEC 资源的折扣成本, 也不需要知道其他智能体所采取的行动. 因此, 信号传递的开销就大大减少了. (3) 仿真结果表明, 该算法可以为每个 MEC 服务器学习一个优秀的策略, 以确定其资源的单价.

2 相关工作

边缘系统的最优资源分配和定价已经引起了越来越多的研究关注. 主要有两个研究方向: 定价导向因素和最优定价策略.

定价导向因素. Dong 等人[8] 提出了一种云计算资源定价算法, 分析历史资源使用情况并不断调整资源价格. 但该算法只考虑了资源利用率, 没有分析其他重要因素. Liu 等人[9] 提出了一种基于价格的分布式算法, 只强调了任务调度. Li 等人[10] 提出了云计算的静态资源定价方案. 定价操作简单, 但难以满足终端设备的动态需求. 他们没有考虑用户需求与资源价格之间的实时关系, 因此不能根据用户需求动态调整资源价格.

最优定价策略. 现有的最优定价策略大多是基于拍卖和博弈论的. Zhang 等人[11] 研究了通过基于拍卖的算法对系统效益和多维资源的联合优化. 解决方案是系统性能改进和单位效益的产物. 然而, 该算法以每一轮拍卖为优化目标, 难以接近全局最优. 因此, 执行成本非常高. Dong 等人[12] 采用了一种基于价格的双层 Stackelberg 博弈来模拟一个由单个 MEC 服务器和多个用户组成的 MEC 系统.

3 系统模型

3.1 网络模型

MEC 系统模型由多个利益相关者组成: (1) 希望出售免费资源的 MEC 服务器; (2) 希望购买资源以执行计算任务的 MD; (3)AP 作为第三方代理, 代表 MD 从 MEC 服务器购买资源. 不失一般性, 我们考虑一个通用的“多对多”场景, 即每个 MEC 都可以将资源出售给多个 MD. 同时, 每个 MD 可以购买多个 MEC 服务器出售的资源.

我们考虑一个具有多个 MD 和具有多种类型资源的多个 MEC 服务器的 MEC 系统. 我们用$U=\{1,2,\cdots,U\}$和$M=\{1,2,\cdots,M\}$分别表示 MD 和 MEC 服务器的集合. $R=\{1,2,\cdots,R\}$表示资源种类的集合. 我们有$|U|=U$, $|M|=M,|R|=R$. MEC 服务器和 MD 之间的相互作用总结如下:

(1) MEC 服务器$i\in M$, 希望出售空闲资源$r\in R$, 于是向 AP 告知它的可用资源数量$Q_{i,r}$和其单位资源的期望价格$p_{i,r}$.

(2) 给定价格$p_{i,r}$和可用资源数量$Q_{i,r}$, MD$j \in U$决定从每个 MEC 服务器购买的资源数量

(3) MDj使用所购买的资源来处理其计算任务.

3.2 多阶段 Stackelberg 博弈

Stackelberg 领导者-跟随者博弈是一个策略游戏, 其中领导者承诺一个策略, 然后跟随者跟随[13]. 一般来说, 游戏中的所有玩家都是自私的, 因为他们每个人都考虑了他人的策略来最大化自己的利益. 具体来说, 考虑到跟随者可能采取的策略, 领导者选择了一种最大化其利益的策略. 基于观察领导者的策略, 每个跟随者都采用了使其利益最大化的策略. 然后, 我们解释了跟随者之间的竞争. 通过 MAPPO 算法, 得到了每个跟随者的最佳响应. 在这个算法中, 每个跟随者都与环境交互, 并学习一种策略来优化其长期奖励, 而不需要考虑他人采取的行动. Stackelberg 领导者-跟随者博弈的定义如下:

玩家: AP 和 MEC 服务器都是游戏玩家. AP 是领导者, 而所有的 MEC 服务器都是跟随者.

策略: 对于 MEC 服务器$i \in M$, 他的策略是确定资源$r \in R$的单价; 对于 AP, 策略是确定 MD $j \in U$从 MEC 服务器处购买的资源r的数量.

收益: MEC 服务器、MD 和 AP 的收益函数分别由式 (1)–式 (3) 给出.

令$x_{i,j,r}$表示 MD $j \in U$从 MEC 服务器$i \in M$处购买的资源$r \in R$的数量. MEC 服务器i的收益计算如下:

$$\psi_i(x_i, p_i) = \sum_{r \in R} p_{i,r} \sum_{j \in U} x_{i,j,r}, \forall i \in M \tag{1}$$

其中, $p_{i,r}$表示 MEC 服务器 i的资源 r的单价, $x_i = \{x_{i,j,r}\}_{j \in U, r \in R}$, $p_i = \{p_{i,r}\}_{j \in U}$.

MD j的收益定义如下:

$$\phi_j(x_i, p) = \sum_{r \in R} \log\left(\sum_{i \in M} \omega_{i,j,r} x_{i,j,r}\right) - \sum_{i \in M} p_{i,j,r} x_{i,j,r}, \forall j \in U \tag{2}$$

其中, $x_j = \{x_{i,j,r}\}_{i \in M, r \in R}$, $p = \{p_{i,r}\}_{i \in M, j \in U}$, $\omega_{i,j,r}$是 MEC 服务器i出售给 MD j的资源r的质量.

AP 的收益是所有 MEC 服务器和 MD 收益的总和 (即社会福利), 定义如下:

$$\begin{aligned}\phi(x, p) &= \sum_{i \in M} \psi_i(x_i, p_i) + \sum_{j \in U} \phi_j(x_i, p) \\ &= \sum_{j \in U} \sum_{r \in R} \log\left(\sum_{i \in M} \omega_{i,j,r} x_{i,j,r}\right)\end{aligned} \tag{3}$$

其中, $x = \{x_{i,j,r}\}_{i \in M, j \in U, r \in R}$.

由于所有资源都有独立的预算, 且彼此不受影响, 因此我们可以将多资源分配和定价问题分解为多个单资源分配和定价子问题. 因此, 我们将优化问题分解为 R 个独立的子问题, 每个子问题都与特定的资源类型相关联. 与整体处理原始优化问题相比, 该分解的主要优点是处理多个子问题显著降低了计算复杂度. 与$r \in R$相关的子问题表示为:

$$\psi_{i,r}(x_{i,r}, p_{i,r}) = p_{i,r} \sum_{j \in U} x_{i,j,r}, \forall i \in M \tag{4}$$

$$\phi_r(x_r, p_r) = \sum_{j \in U} \log\left(\sum_{i \in M} \omega_{i,j,r} x_{i,j,r}\right) \tag{5}$$

其中, $x_{i,r} = \{x_{i,j,r}\}_{j \in U}$, $x_r = \{x_{i,j,r}\}_{i \in M, j \in U}$, $p_r = \{p_{i,r}\}_{i \in M}$.

定义 1. 令 $x^*_{i,r} = \{x^*_{i,j,r}\}_{j \in U}$, $p^*_r = \{p^*_{i,r}\}_{i \in M}$, $x^*_r = \{x^*_{i,j,r}\}_{i \in M, j \in U}$, 解$(x^*_r, p^*_r)$是一个纳什均衡解的充分条件是:

$$\psi_{i,r}(x^*_{i,r}, p^*_{i,r}) \geqslant \psi_{i,r}(x^*_{i,r}, p_{i,r}), \forall p_{i,r} \neq p^*_{i,r} \tag{6}$$

$$\varphi_r(x^*_r, p^*_r) \geqslant \varphi_r(x_r, p^*_r), \forall x_r \neq x^*_r \tag{7}$$

3.3 AP 社会福利优化

针对与资源r相关的子问题, 给定所有 MEC 服务器对资源r的单价 (即p_r), AP 的目标是最大化它的收益.

问题 1:

$$\begin{cases}\max\limits_{x_r = \{x_{i,j,r}\}_{i \in M, j \in U}} \phi_r(x_r, p_r) \\ \text{s.t.} \sum\limits_{j \in U} x_{i,j,r} \leqslant Q_{i,r}, \forall i \in M \\ \sum\limits_{i \in M} p_{i,r} x_{i,j,r} \leqslant B_{j,r}, \forall j \in U\end{cases} \tag{8}$$

其中, $Q_{i,r}$是 MEC 服务器i中资源 r的可售卖数量, $B_{j,r}$是 MDj购买资源 r的预算.

定理 1. 问题 1 的最优解如下:

$$x^*_{i,j,r} = \begin{cases}\dfrac{B_{j,r}\omega_{i,j,r}}{p_{i,r}\sum_{i \in M}\omega_{i,j,r}}, & \text{if } I \geqslant 0 \\ \dfrac{Q_{i,r}\omega_{i,j,r}}{\sum_{j \in U}\omega_{i,j,r}}, & \text{otherwise}\end{cases} \tag{9}$$

其中,

$$I = \sum_{j \in U} \log\left(\sum_{i \in M} \frac{B_{j,r}\omega^2_{i,j,r}}{p_{i,r}\sum_{i \in M}\omega_{i,j,r}}\right) - \sum_{j \in U} \frac{Q_{i,r}\omega^2_{i,j,r}}{\sum_{j \in U}\omega_{i,j,r}} \tag{10}$$

证明见附录 A.

4 基于 MAPPO 的 AP 收益优化

本节将介绍每个 MEC 服务器如何选择其对 AP 所采用的策略的最佳响应.

4.1 基于深度强化学习的方法

我们使用多智能体强化学习 (multi-agent reinforcement learning, MARL) 来解决多重单一资源分配和定价子问题. 我们将每个子问题描述为一个马尔可夫决策过程 (Markov decision process, MDP), 以准确地反映资源分配和定价的决策过程. 然后, 我们将 MAPPO 应用于这些子问题. 由于其在全局优化方面的出色性能, MAPPO 可以在需要时快速为每个 MEC 服务器获得接近最优的单一资源分配和定价策略.

对于资源r, 给定来自环境的反馈 (即资源r如何分配给 MD), 每个 MEC 服务器需要确定资源r的单价以最大化它的收益.

$$\max_{p_{i,r}} \psi_{i,r}(x_{i,r}, p_{i,r}) \ \text{s.t.} \ p_{i,r} \geqslant 0 \tag{11}$$

MDP 的元素如下所示, 包括状态空间、动作空间和奖励函数.

状态空间: MEC 服务器i在时隙t时刻的状态空间表示为o_i^t, 包括对前面的L个时隙的观察, 如式 (12) 所述:

$$o_i^t = [p_{i,r}^{t-1}, x_{i,r}^{t-1}, \cdots, p_{i,r}^{t-L}, x_{i,r}^{t-L}], \ \forall i \in M \tag{12}$$

全局状态: MAPPO 基于全局状态s而不是本地观察o_i学习策略π_θ和值函数$V_\varsigma(s)$. 我们使用所有局部观测结果的连接来作为 critic 网络的输入.

$$s^t = (o_1^t, \cdots, o_M^t) \tag{13}$$

动作空间: 在时隙t, MEC 服务器$i \in M$观察前L个时隙的资源分配情况, 决定在当前时隙的单价, 即$p_{i,r}^t$.

$$a_i^t = u_i(o_i^t|\theta^{u_i}) \tag{14}$$

奖励函数: 奖励函数定义如下:

$$r^t = \sum_{i=1}^{M} \log\left(1 + p_{i,r}^t \sum_{j \in U} x_{i,j,r}^t - c_i^t \sum_{j \in U} x_{i,j,r}^t\right) \tag{15}$$

其中, c_i^t是 MEC 服务器i的折扣成本. log 函数确保当所获收益 (即$p_{i,r}^t \sum_{j \in U} x_{i,j,r}^t$) 不足以抵扣成本 (即$c_i^t \sum_{j \in U} x_{i,j,r}^t$) 时, 奖励是负的.

4.2 基于 MAPPO 的资源分配和定价策略 (RAP-MAPPO)

MARL 算法可以分为两种框架: 集中式学习和分散式学习. 集中式方法假设合作博弈, 并通过学习单一策略直接扩展单智能体强化学习算法, 以同时产生所有智能体的联合动作. 在分散学习中, 每个智能体都优化自己的独立奖励; 这些方法可以解决非零和博弈, 但可能会受到非平稳转换的影响. 最近的工作已经开发出两条研究路线来弥合这两个框架之间的差距: 集中培训和分散执行 (centralized training and decentralized execution, CTDE) 和值分解 (value decomposition, VD). CTDE 通过采用 actor-critic 结构并学习集中的 critic 来减少方差, 从而改进了分散的强化学习. 代表性的 CTDE 方法是多智能体深度确定性策略梯度 (multi-agent deep deterministic policy gradient, MADDPG)[14]. VD 通常与集中式 Q 学习相结合, 将联合 Q 函数表示为每个代理的局部 Q 函数的函数[15], 这已被视为许多 MARL 基准测试的黄金标准. MAPPO 通过将单个近端策略优化算法 (proximal policy optimization algorithms, PPO)[16] 训练与全局值函数相结合, 属于 CTDE 类别. 不同的是 PPO 是一个 on-policy 算法, MADDPG 是基于 off-policy 的算法, 但是 MAPPO 经过简单的超参数调整就能获得比较好的成绩.

在 RAP-MAPPO 中, 每个 MEC 服务器都被视为一个智能体. 每个智能体都有一个参数为θ的 actor 网络和一个参数为ς的 critic 网络. RAP-MAPPO 为每个智能体训练这两个神经网络. 我们用V_ς表示 critic 网络, 用π_θ表示 actor 网络. 我们使用统一的经验重放缓冲区来存储历史数据点以进行训练. RAP-MAPPO 是一种基于集中式训练和分布式执行的方法. 在集中训练阶段, actor 网络只从自身获取观测信息, 而 critic 网络获取全局状态. 在分布式执行阶段, 每个代理只需要它的 actor 网络 (而不需要 critic 网络). 通过与环境的交互, 每个代理都可以做出合适的资源分配和定价策略.

Actor 网络被训练用来最大化:

$$L(\theta) = \left[\frac{1}{Bn}\sum_{i=1}^{B}\sum_{k=1}^{n} \min(r_{\theta,i}^k A_i^k, clip(r_{\theta,i}^k, 1-\varepsilon, 1+\varepsilon)\, A_i^k)\right] + \sigma \frac{1}{Bn}\sum_{i=1}^{B}\sum_{k=1}^{n} S[\pi_\theta(\mathrm{o}_i^k)] \tag{16}$$

其中, $r_{\theta,i}^k = \dfrac{\pi_\theta(a_i^k|o_i^k)}{\pi_{\theta_{old}}(a_i^k|o_i^k)}$, B和n分别是抽样次数和智能体数量, A_i^k基于广义优势估计 (generalized advantage estimation, GAE) 方法计算得出, S是策略熵, σ是超参数.

Critic 网络被训练用来最大化.

$$L(\varsigma)=\frac{1}{Bn}\sum_{i=1}^{B}\sum_{k=1}^{n}\max[(V_{\varsigma}(s_i^k)-\widehat{R}_i)^2,(clip(V_{\varsigma}(s_i^k),V_{\varsigma_{\text{old}}}(s_i^k)-\varepsilon,V_{\varsigma_{\text{old}}}(s_i^k)+\varepsilon)-\widehat{R}_i)^2] \tag{17}$$

其中, $\widehat{R}_i$是折扣奖励. RAP-MAPPO 的训练步骤见算法 1.

算法 1. RAP-MAPPO 的训练过程

初始化: 初始化 actor 网络和 critic 网络的参数

1: 设置学习率α
2: while $step \leqslant stem_{\max}$
3: 令 data buffer D={}
4: for i=1 *to batch_size* do
5: τ=[]
6: for t=1 to T do
7: $p_t^a=\pi(o_a^t;\theta),u_a^t\sim p_a^t,v_a^t=V(s_a^t;\varsigma)$
8: 计算动作u^t, 得到r^t,o^{t+1},s^{t+1}
9: τ+=[$s^t,o^t,u^t,r^t,o^{t+1},s^{t+1}$]
10: 计算$\widehat{A},\widehat{R}$, 将τ分成长度为 L 的块
11: for 1=1, 2, ⋯, T //L do
12: $D=D\cup(\tau[l:l+L,\widehat{A}[l:l+L],\widehat{R}[l:l+L]])$
13: 从 D 中随机抽取 K 个样本
14: 通过$L(\theta)$更新θ, $L(\varsigma)$更新ς

5 仿真结果

5.1 参数设置

我们考虑一个由多个 MEC 服务器和多个 MD 组成的 MEC 系统. 收益最大化问题取决于可用的资源和预算. 为简单起见, 我们设置$\omega_{i,j,r}=1+0.1j+i/10$. 资源质量在整个实验过程中都是固定的. 我们将长期奖励的折扣系数设置为零, 因为自私的智能体的目标是最大化他们的即时收益. 为了加快训练过程, 我们对每个智能体都采用了一个相对较小的网络. Actor 网络和 critic 网络都是由 1 个输入层, 3 个隐藏层和 1 个输出层组成. 这 3 个隐藏层分别有 128、64 和 32 个神经元. 此外, actor 网络和 critic 网络都使用 ReLU 作为所有隐藏层的激活函数, 参与者网络采用 tanh 函数激活输出层进行策略生成. 其他模拟参数见表 1. 进行性能比较的算法如下:

质量比例最优定价 (quality proportional optimal pricing, QPOP): 我们假设 AP 对每个 MEC 服务器提供的资源质量有一个先验的知识. 单价设置与服务器的资源质量成正比, 同时消耗确定的资源和货币预算 (即 $I=0$). $I=0$在实际系统中是不可能的, 但由 QPOP 找到的解决方案可以作为一个最优的基准.

随机: 单价在 (0, 5) 区间内随机产生.

表 1 仿真参数

参数	值
MEC服务器的数量 (M)	4
MD的数量 (U)	8
资源种类 (R)	3
MDj在资源r上的预算	[20, 30]
MEC服务器i中资源r的数量	[30, 40]
MEC服务器i的折扣成本	0.05
MEC服务器i的资源质量	$1+0.1j+i/10$
Actor学习率	0.001
Critic学习率	0.001
Soft update	0.01
折扣系数	0

MADDPG: 每个 MEC 服务器都被视为一个智能体, 状态空间由前 L 个时隙的价格和资源分配组成, 动作是资源的单价, 奖励函数基于 MEC 服务器的资源收入和成本设计.

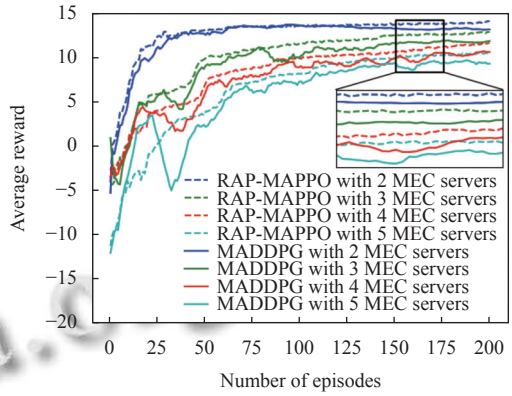

图 1 不同算法的收敛性曲线

5.2 收敛性

图 1 为所提出的 RAP-MAPPO 和 MADDPG 在不同 MEC 服务器数量下的收敛曲线. 随着训练次数的增加, MEC 服务器的平均奖励逐渐上升, 最终变为积极和稳定. 我们首先研究了 MEC 服务器的数量对收敛性的影响. 随着 MEC 服务器的增加, 这两种算法都需要更多的时间来收敛. 这是因为更多的服务器会导致更大的状态空间. 这两种算法需要对状态空间进行更多的探索, 才能获得可观的奖励. 此外, MEC 服务器的平均奖励随着 MEC 服务器的增加而降低. 这是因为更多的服务器会在竞争期间降低价格, 也就是说, 每个服务

器都希望出售其资源. 然后, 我们比较了两种算法在收敛性方面的性能. 我们很容易观察到, 与 MADDPG 相比, RAP-MAPPO 具有收敛速度更快、平均奖励速度更高的特点.

5.3 MEC 服务器和 MDs 在 Stackelberg 均衡下的收益

我们在两种场景下比较了这 4 种算法. 在第 1 种场景下, $B_{j,r}$从 5 到 40 不等, $Q_{i,r}$保持不变. 在第 2 种场景下, $Q_{i,r}$均匀分布在 [5, 40] 的范围内, $B_{j,r}$固定为 20. 为了说明预算约束的影响, 我们在实验过程中固定了 MD 的质量权重, 并将M, U分别设置为 4 和 8.

在第 1 种场景下, RAP-MAPPO 在没有对 MEC 服务器提供的资源质量的先验知识的情况下, 获得了接近最优的性能, 如图 2 所示. 当 MEC 服务器的资源不足时, MD 之间的激烈竞争导致了卖方市场. 同时, RAP-MAPPO 在社会福利方面优于 MADDPG 和随机算法 (见图 3). 这是因为随机定价不能充分利用 MEC 服务器的资源, 而 RAP-MAPPO 则鼓励 MEC 服务器动态调整其单价, 以达到提高资源效率的目的.

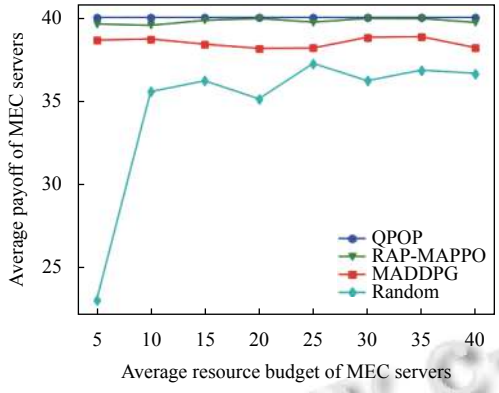

图 2 不同 MEC 资源数量下 MEC 服务器收益

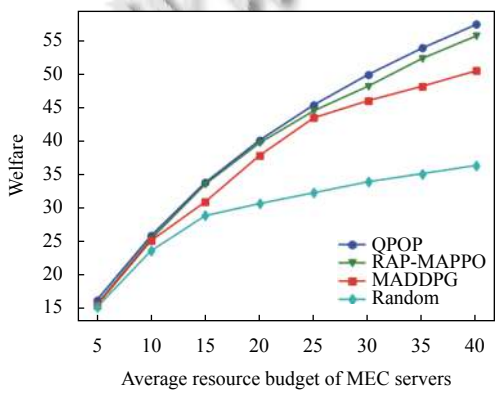

图 3 不同 MEC 资源数量下社会福利

在第 2 种场景下, 从图 4 和图 5 可以看出, 在大多数情况下, RAP-MAPPO 在 MD 收益和社会福利方面比 MADDPG 和随机算法获得了更好的性能. 随着 MD 的平均货币预算的增长, MEC 服务器更有可能提高价格, 以响应 AP 的资源购买策略. 因此, MD 需要降低他们的资源需求或支付更多的费用来鼓励 MEC 服务器销售更多的资源, 从而减少 MD 的回报, 即卖方市场.

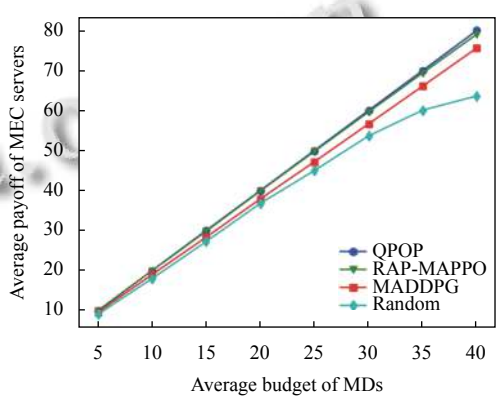

图 4 不同 MD 预算下 MEC 服务器收益

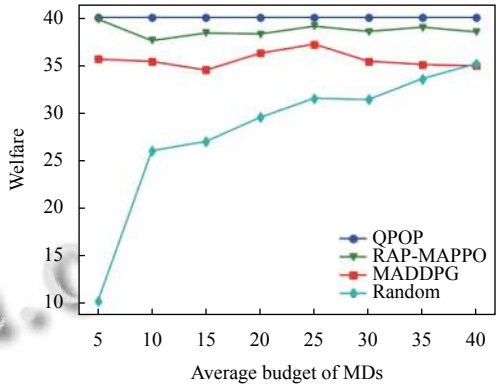

图 5 不同 MD 预算下社会福利

综上所述, RAP-MAPPO 在 MEC 服务器的收益和社会福利方面的表现优于 MADDPG 和随机算法. 它的性能与 QPOP 类似. QPOP 需要知道 MEC 服务器的质量权重信息和 MEC 服务器之间的无条件合作, 而我们的方法只是基于与环境相互作用的局部观察.

6 结论

本文研究了基于 Stackelberg 博弈的资源分配和定价问题, 其中 AP 和 MEC 服务器是领导者和追随者. 这个问题被分解为多个可以单独解决的单一资源类型的子问题. 我们采用 MAPPO 来解决这个问题. 对于任

意的 MEC 服务器, RAP-MAPPO 不需要知道其他 MEC 服务器所采取的操作, 这有助于减少信令开销. 此外, RAP-MAPPO 可以通过一系列的状态-行动-奖励观察来指导竞争智能体实现收益最大化. 仿真结果表明, 在 RAP-MAPPO 中, MD 和 MEC 服务器在满足前者严格的货币预算约束的同时, 学习了接近最优的回报. 此外, RAP-MAPPO 在收益和社会福利方面都优于 QPOP、随机和 MADDPG.

参考文献

1 Ding HC, Zhang C, Cai Y, *et al*. Smart cities on wheels: A newly emerging vehicular cognitive capability harvesting network for data transportation. IEEE Wireless Communications, 2018, 25(2): 160–169. [doi: 10.1109/MWC.2017.1700151]

2 Asir TRG, Manohar HL. Key challenges and success factors in IoT—A study on impact of data. Proceedings of 2018 International Conference on Computer, Communication, and Signal Processing. Chennai: IEEE, 2018. 1–5.

3 Shi WS, Cao J, Zhang Q, *et al*. Edge computing: Vision and challenges. IEEE Internet of Things Journal, 2016, 3(5): 637–646. [doi: 10.1109/JIOT.2016.2579198]

4 Yang B, Li ZY, Jiang SL, *et al*. Envy-free auction mechanism for VM pricing and allocation in clouds. Future Generation Computer Systems, 2018, 86: 680–693. [doi: 10.1016/j.future.2018.04.055]

5 Bonacquisto P, di Modica G, Petralia G, *et al*. A procurement auction market to trade residual cloud computing capacity. IEEE Transactions on Cloud Computing, 2015, 3(3): 345–357. [doi: 10.1109/TCC.2014.2369435]

6 Tanaka M, Murakami Y. Strategy-proof pricing for cloud service composition. IEEE Transactions on Cloud Computing, 2016, 4(3): 363–375. [doi: 10.1109/TCC.2014.2338310]

7 Yu C, Velu A, Vinitsky E, *et al*. The surprising effectiveness of PPO in cooperative, multi-agent games. arXiv: 2103.01955, 2021.

8 Dong SQ, Li HL, Qu YB, *et al*. Survey of research on computation unloading strategy in mobile edge computing. Computer Science, 2019, 46(11): 32–40.

9 Liu MY, Liu Y. Price-based distributed offloading for mobile-edge computing with computation capacity constraints. IEEE Wireless Communications Letters, 2018, 7(3): 420–423. [doi: 10.1109/LWC.2017.2780128]

10 Li ZT, Kang JW, Yu R, *et al*. Consortium Blockchain for secure energy trading in industrial Internet of Things. IEEE Transactions on Industrial Informatics, 2018, 14(8): 3690–3700.

11 Zhang HL, Guo FX, Ji H, *et al*. Combinational auction-based service provider selection in mobile edge computing networks. IEEE Access, 2017, 5: 13455–13464. [doi: 10.1109/ACCESS.2017.2721957]

12 Dong LQ, Han SY, Gao Y, *et al*. A game-theoretical approach for resource allocation in mobile edge computing. Proceedings of the 2020 IEEE 20th International Conference on Communication Technology. Nanning: IEEE, 2020. 436–440.

13 Varian HR. Intermediate Microeconomics: A Modern Approach. 9th ed, New York: William Warder Norton & Company, 2014.

14 Lowe R, Wu Y, Tamar A, *et al*. Multi-agent actor-critic for mixed cooperative-competitive environments. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6382–6393.

15 Sunehag P, Lever G, Gruslys A, *et al*. Value-decomposition networks for cooperative multi-agent learning based on team reward. Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. Stockholm: ACM, 2018. 2085–2087.

16 Schulman J, Wolski F, Dhariwal P, *et al*. Proximal policy optimization algorithms. arXiv: 1707.06347, 2018.

附录 A. 定理 1 的证明

问题 1 对应的拉格朗日方程是:

$$L(x,\lambda,\mu)=-\sum_{j\in U}\log\left(\sum_{i\in M}\omega_{i,j,r}x_{i,j,r}\right)+\sum_{i\in M}\lambda_i\left(\sum_{j\in U}x_{i,j,r}-Q_{i,r}\right)+\sum_{j\in U}\mu_j\left(\sum_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\right) \quad (18)$$

KKT 条件如下:

$$\begin{cases}\dfrac{\partial L(x_{i,j,r},\lambda_i,\mu_j)}{\partial x_{i,j,r}}=0,\forall i\in M,\forall j\in U\\ \lambda_i\left(\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}\right)=0,\forall i\in M\\ \mu_j\left(\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\right)=0,\forall j\in U\\ x_{i,j,r}\geqslant 0,\forall i\in M,\forall j\in U\\ \lambda_i\geqslant 0,\forall i\in M\\ \mu_j\geqslant 0,\forall j\in U\end{cases} \quad (19)$$

消去λ_i, 可得:

$$\begin{cases}\left(\dfrac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\mu_j p_{i,r}\right)\left(\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}\right)=0,\forall i\in M\\ \mu_j\left(\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\right)=0,\forall j\in U\\ x_{i,j,r}\geqslant 0,\forall i\in M,\forall j\in U\\ \dfrac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\mu_j p_{i,r}\geqslant 0,\forall i\in M\\ \mu_j\geqslant 0,\forall j\in U\end{cases} \tag{20}$$

当$\dfrac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\mu_j p_{i,r}=0$时, 我们有:

$$\mu_j=\frac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}\geqslant 0 \tag{21}$$

因为$\mu_j(\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r})=0$, 所以可得:

$$B_{j,r}=\sum_{i\in M}p_{i,r}x_{i,j,r} \tag{22}$$

换句话说, 如果$\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}\neq 0$, 我们有$\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}=0$. 同理可得, 当$\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\neq 0$时$\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}=0$.

因此, 问题 1 可以被分解为如下两个子问题.

子问题 1:

$$\begin{cases}\max\limits_{x_r=\{x_{i,j,r}\}_{i\in M,j\in U}}\ \phi_r(x_r,p_r)\\ \text{s.t.}\ \sum\limits_{j\in U}x_{i,j,r}\leqslant Q_{i,r},\forall i\in M\\ \sum\limits_{i\in M}p_{i,r}x_{i,j,r}=B_{j,r},\forall j\in U\end{cases} \tag{23}$$

子问题 1 对应的拉格朗日形式的 KKT 条件如下:

$$\begin{cases}\dfrac{\partial L\left(x_{i,j,r},\lambda_i,\mu_j\right)}{\partial x_{i,j,r}}=0,\forall i\in M,\forall j\in U\\ \lambda_i\left(\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}\right)=0,\forall i\in M\\ x_{i,j,r}\geqslant 0,\forall i\in M,\forall j\in U\\ \lambda_i\geqslant 0,\forall i\in M\\ \sum\limits_{i\in M}p_{i,r}x_{i,j,r}=B_{j,r},\forall j\in U\end{cases} \tag{24}$$

消去λ_i, 可得:

$$\begin{cases}\left(\dfrac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\mu_j p_{i,r}\right)\left(\sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}\right)=0,\forall i\in M\\ \dfrac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\mu_j p_{i,r}\geqslant 0,\forall i\in M\\ x_{i,j,r}\geqslant 0,\forall i\in M,\forall j\in U\\ \sum\limits_{i\in M}p_{i,r}x_{i,j,r}=B_{j,r},\forall j\in U\end{cases} \tag{25}$$

由式 (25) 可知:

$$\frac{\omega_{i,j,r}}{\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}\geqslant\mu_j p_{i,r} \tag{26}$$

$$x^*_{i,j,r}=\begin{cases}\dfrac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}\mu^*_j},\ \mu^*_j\geqslant 0\\ 0,\ \text{otherwise}\end{cases} \tag{27}$$

另一方面, 式 (23) 的可行解可推出:

$$\sum_{i\in M}p_{i,r}\frac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}\mu^*_j}=B_{j,r} \tag{28}$$

因此可知:

$$x^*_{i,j,r}=\frac{B_{j,r}\omega_{i,j,r}}{p_{i,r}\sum_{i\in M}\omega_{i,j,r}}$$

子问题 2:

$$\begin{cases}\max\limits_{x_r=\{x_{i,j,r}\}_{i\in M,j\in U}}\ \phi_r(x_r,p_r)\\ \text{s.t.}\ \sum\limits_{j\in U}x_{i,j,r}=Q_{i,r},\forall i\in M\\ \sum\limits_{i\in M}p_{i,r}x_{i,j,r}\leqslant B_{j,r},\forall j\in U\end{cases} \tag{29}$$

子问题 2 对应的拉格朗日形式的 KKT 条件如下:

$$\begin{cases}\dfrac{\partial L\left(x_{i,j,r},\lambda_i,\mu_j\right)}{\partial x_{i,j,r}}=0,\ \forall i\in M,\forall j\in U\\ \mu_j\left(\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\right)=0,\ \forall j\in U\\ x_{i,j,r}\geqslant 0,\ \forall i\in M,\forall j\in U\\ \sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}=0,\ \forall i\in M\\ \mu_j\geqslant 0,\ \forall j\in U\end{cases} \tag{30}$$

消去μ_j, 可得:

$$\begin{cases} \dfrac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\dfrac{\lambda_i}{p_{i,r}}\geqslant 0,\ \forall j\in U \\ \left(\dfrac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}-\dfrac{\lambda_i}{p_{i,r}}\right)\left(\sum\limits_{i\in M}p_{i,r}x_{i,j,r}-B_{j,r}\right)=0 \\ x_{i,j,r}\geqslant 0,\ \forall i\in M,\forall j\in U \\ \sum\limits_{j\in U}x_{i,j,r}-Q_{i,r}=0,\ \forall i\in M \end{cases} \tag{31}$$

由式 (30) 可知:

$$\frac{\omega_{i,j,r}}{p_{i,r}\sum\limits_{i\in M}\omega_{i,j,r}x_{i,j,r}}\geqslant\frac{\lambda_i}{p_{i,r}} \tag{32}$$

$$x^*_{i,j,r}=\begin{cases}\dfrac{\omega_{i,j,r}}{\lambda^*_i\sum\limits_{i\in M}\omega_{i,j,r}},\lambda^*_i\geqslant 0\\ 0,\quad \text{otherwise}\end{cases} \tag{33}$$

另一方面, 式 (29) 的可行解可推出:

$$\sum_{j\in U}x_{i,j,r}=Q_{i,r} \tag{34}$$

因此可得:

$$x^*_{i,j,r}=\frac{Q_{i,r}\omega_{i,j,r}}{\sum_{j\in U}\omega_{i,j,r}} \tag{35}$$

综上所述, 问题 1 的解如下:

$$x^*_{i,j,r}=\begin{cases}\dfrac{B_{j,r}\omega_{i,j,r}}{p_{i,r}\sum_{i\in \mathrm{M}}\omega_{i,j,r}},\ \text{if } I\geqslant 0\\ \dfrac{Q_{i,r}\omega_{i,j,r}}{\sum_{j\in \mathrm{U}}\omega_{i,j,r}},\quad \text{otherwise}\end{cases}$$

其中, $I=\sum\limits_{j\in \mathrm{U}}\log\left(\sum\limits_{i\in \mathrm{M}}\dfrac{B_{j,r}\omega^2_{i,j,r}}{p_{i,r}\sum_{i\in \mathrm{M}}\omega_{i,j,r}}\right)-\sum\limits_{j\in \mathrm{U}}\dfrac{Q_{i,r}\omega^2_{i,j,r}}{\sum_{j\in \mathrm{U}}\omega_{i,j,r}}$.

(校对责编: 孙君艳)