



计算机科学与探索

Journal of Frontiers of Computer Science and Technology

ISSN 1673-9418, CN 11-5602/TP

## 《计算机科学与探索》网络首发论文

题目: 基于 Stackelberg 博弈的边缘云资源定价机制研究  
作者: 刘荆欣, 王妍, 韩笑, 夏长清, 宋宝燕  
网络首发日期: 2020-10-23  
引用格式: 刘荆欣, 王妍, 韩笑, 夏长清, 宋宝燕. 基于 Stackelberg 博弈的边缘云资源定价机制研究. 计算机科学与探索.  
<https://kns.cnki.net/kcms/detail/11.5602.TP.20201023.0941.007.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于 Stackelberg 博弈的边缘云资源定价机制研究

刘荆欣<sup>1</sup>, 王 妍<sup>1,2+</sup>, 韩 笑<sup>1</sup>, 夏长清<sup>2,3,4</sup>, 宋宝燕<sup>1</sup>

1. 辽宁大学 信息学院, 沈阳 110036

2. 中国科学院沈阳自动化研究所 机器人学国家重点实验室, 沈阳 110016

3. 中国科学院沈阳自动化研究所 网络化控制系统重点实验室, 沈阳 110016

4. 中国科学院 机器人与智能制造创新研究院, 沈阳 110169

**摘 要:** 移动边缘计算 (mobile edge computing, MEC) 支持终端设备将任务或应用程序卸载到边缘云服务器处理, 边缘云服务器处理外来任务会消耗本地资源, 为激励边缘云提供资源服务, 构建向终端设备收费以奖励边缘云的资源定价机制尤为重要。现有的定价机制依赖中间商的静态定价, 费用高且终端任务处理不及时, 难以实现边缘云计算资源的有效利用。针对上述问题, 本文提出一种基于 Stackelberg 博弈的边缘云资源定价机制。首先, 针对资源定价时终端设备因资金不足而导致的本地任务搁置问题, 提出包含贷款和激励的辅助机制, 实现终端设备任务的及时处理; 其次, 提出影响资源定价的 4 种价格导向因素, 制定了一致性与弹性两种定价方案, 提高定价的准确性和效率, 并为动态定价做准备; 然后, 为了使终端设备与边缘云直接进行动态定价, 构建基于斯坦克伯格 (Stackelberg) 博弈的资源定价机制模型, 将资源需求与定价问题转化为边缘云收益最大与终端设备支付成本最小问题; 最后, 通过改进的强化学习 State Action Reward State Action (SARSA) 算法得到资源需求及定价的最优策略。实验表明, 本文提出的定价机制在边缘云收益最大化方面优于其他定价算法 12% 以上, 同时弹性定价方案下边缘云的收益优于一致性定价方案的 24%。

**关键词:** 移动边缘计算; 资源定价机制; Stackelberg 博弈; 强化学习

**文献标志码:** A      **中图分类号:** TP399

刘荆欣, 王妍, 韩笑, 等. 基于 Stackelberg 博弈的边缘云资源定价机制研究[J]. 计算机科学与探索

LIU J X, WANG Y, HAN X, et al. Research on Edge Cloud Resource Pricing Mechanism Based on Stackelberg Game[J]. Journal of Frontiers of Computer Science and Technology

\*The National Key Research and Development Program under Grant No. 2019YFB1406002 (国家重点研发计划); the National Natural Science Foundation of China under Grant No. 61903356 (国家自然科学基金); the State Key Laboratory of Robotics under Grant No. 2019-022 (机器人学国家重点实验室开放基金资助); the National Natural Science Foundation of Liaoning province under Grant No. 20180520029 (辽宁省自然科学基金计划重点项目); Economic and social development project of Liaoning Province: Research on Intelligent Manufacturing Transformation Scheme and Key Technologies of Liaoning Equipment Manufacturing Industry for Internet No. 2019lslktqn-023 (辽宁省经济社会发展课题“面向‘互联网+’的辽宁装备制造业智慧创造转型方案及关键技术研究”).

## Research on Edge Cloud Resource Pricing Mechanism Based on Stackelberg Game

LIU Jingxin<sup>1</sup>, WANG Yan<sup>1,2+</sup>, HAN Xiao<sup>1</sup>, XIA Changqing<sup>2,3,4</sup>, SONG Baoyan<sup>1</sup>

1. College of Information, Liaoning University, Shenyang 110036, China

2. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

3. Key Laboratory of Networked Control System, Shenyang Institute of Automation, Shenyang 110016, China

4. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China

**Abstract:** Mobile Edge Computing (MEC) supports terminal devices to offload tasks or applications to edge cloud server for processing. Edge cloud server will consume local resources when processing external tasks, so it is particularly important to build a resource pricing mechanism that charges terminal devices to reward edge cloud. The existing pricing mechanism relies on the static pricing of intermediaries, and the high cost and the late processing of terminal tasks make it difficult to realize the effective utilization of edge cloud computing resources. Aiming at the above problems, this paper proposes an edge cloud resource pricing mechanism based on Stackelberg game. Firstly, in view of the local task shelving problem of terminal devices due to insufficient funds during resource pricing, an auxiliary mechanism including loans and incentives is proposed to realize the timely processing of terminal devices tasks; Secondly, four price-oriented factors that affect resource pricing are proposed, and two pricing schemes, consistency and elasticity, are formulated to improve the accuracy and efficiency of pricing and prepare for dynamic pricing; Then, in order to make the dynamic pricing between terminal devices and edge cloud directly, a resource pricing mechanism model based on Stackelberg game is built, and the resource demand and pricing problem is transformed into the problem of maximum revenue of edge cloud and minimum payment cost of terminal devices; Finally, through improved reinforcement learning State Action Reward State Action (SARSA) algorithm to obtain the optimal strategy of resource demand and pricing. Experiments show that the pricing mechanism proposed in this paper is more than 12% better than other pricing algorithms in terms of edge cloud revenue maximization, and the edge cloud revenue under the elasticity pricing scheme is superior to 24% of consistency pricing.

**Key words:** mobile edge computing; resource pricing mechanism; Stackelberg game; reinforcement learning

## 1 引言

随着移动计算和物联网时代的发展,传统的云-端集中处理方式逐渐显现出诸多弊端。相比于云计算的不足,移动边缘计算(mobile edge computing, MEC)<sup>[1-4]</sup>解决了网络负载大、高时延、带宽不足等问题。MEC可以将计算资源丰富的边缘设备转化为边缘云,资源受限的终端设备可以通过无线网络将任务卸载到边缘云服务器<sup>[5,6]</sup>,实现云-边-端的分布式管理模式。

在 MEC 系统中,计算资源的管理对于提高系统资源利用率和优化系统资源效益起着关键作用<sup>[7]</sup>。边缘云服务器处理外来任务会消耗本地计算资源,因此,终端设备需要依据边缘云的资源定价机制支付一定的服务费用,旨在激励边缘云提供充足的计算资源。

现有定价机制如拍卖机制依赖于中间商的静态定价,不足之处一方面表现为拍卖双方需要向中间商支付额外的服务费用,总成本增加使得资源交

易双方无法实现最优效益；另一方面静态定价不能很好的满足终端设备资源需求的动态变化，终端设备对资源的总需求量减少，边缘云服务器难以实现本地资源的有效利用。为解决上述问题，本文构建基于 Stackelberg 博弈<sup>[8]</sup>的边缘云资源定价机制，在最大化资源供需双方效益的同时实现计算资源的有效利用。主要研究贡献如下：

(1) 针对资源定价时终端设备存在因资金不足而导致的本地任务搁置问题，本文提出包含贷款和激励的辅助机制，用以增加终端设备的参与积极性与资源需求量，实现终端设备任务的及时处理；

(2) 针对博弈时间过长与价格制定不准确带来的问题，提出影响资源定价的 4 种价格导向因素，根据价格导向因素制定了一致性与弹性两种定价方案，为后续的动态定价做准备；

(3) 针对传统静态定价策略成本高且无法满足终端设备的动态性需求问题，构建基于 Stackelberg 博弈的资源定价机制，为终端设备提供不同资源需求下的价格差异化服务。该模型将资源需求与定价问题转化为边缘云收益最大与终端设备支付成本最小问题；

(4) 移动边缘计算对资源定价的实时性要求高，本文设计出一种改进的强化学习 SARSA 算法，可以较快获得边缘云资源定价的最优解。实验表明，改进的 SARSA 算法的性能较优于其他算法。

本文后续章节内容为：第 2 节介绍相关工作，阐述了现有文献在资源分配与定价方面的研究进展以及待解决问题；第 3 节介绍系统模型，包括终端设备的资源需求辅助机制和边缘云的资源定价机制；第 4 节基于 Stackelberg 和强化学习算法得到最优资源需求与定价决策。第 5 节将本文所提定价方案与其他多个方案进行对比实验。第 6 节总结全文。

## 2 相关工作

近年来，很多学者对边缘系统效益的最优资源定价问题进行了相关研究，主流的研究分为定价导向因素与最优定价求解两个方向。

在资源定价导向因素的研究中，文献[9]提出一种云计算资源定价算法，该算法通过分析资源的历史利用率，不断迭代得到当前的资源价格。但是该定价算法仅考虑资源利用率的使用情况，未分析影响定价的其他导向因素。在文献[10]中提出基于价格的分布式算法。该文献仅以任务调度作为资源定价的研究点。文献[11]为云中心设定了静态的资源定价方案，虽然定价操作简单但不能满足终端设备的动态性需求，但是难以实现资源的有效利用。文献[10,11]都没有考虑用户需求与资源价格的实时关系，无法依据用户需求动态调整资源价格。

在资源定价机制中另一主要研究点为资源价格的最优求解方法。现有最优求解大多采用拍卖思想与博弈理论。在文献[12]中，作者通过拍卖算法实现系统效益和多维资源的联合优化，定价方式为系统性能的提升与单位效益的乘积。但该算法以每轮拍卖结果为优化目标，导致结果难以趋近全局最优且系统的执行成本高。在云计算或 MEC 的资源定价与分配问题中，文献[4,11,13]均采用 Stackelberg 博弈理论。在文献[4]中，作者将提供商和需求者的相互作用建模为 Stackelberg 游戏，并通过 Wolf-PHC 算法搜索最优定价策略。但由于 Wolf-PHC 为确定性策略，易受外部环境的影响，同时需要遍历整个 Q 表，导致时间复杂度高。在文献[11,13]中，均通过博弈理论的逆向归纳方法求得纳什均衡解。但逆向归纳方法是从博弈的最后阶段对每种可能路径进行比较，若出现不同路径却有相同利益时，该方法无法确定唯一的最优路径，适用性失效。

通过分析可以得出，目前的研究未深入分析影响资源定价的导向因素，而资源定价时定价导向因素的变动将直接影响定价准确性。在资源定价的最优求解方法上还存在求解结果难以趋近全局最优、资源利用率低、执行时间复杂度及费用高等问题亟待解决。

## 3 边缘云资源定价系统模型

针对资源定价场景中存在的终端设备任务处理不及时与边缘云资源定价不准确等问题，本文构



建了边缘云资源定价系统模型,旨在最大化边缘云效益的同时为终端设备提供计算资源服务。如图 1 所示,资源需求辅助机制包括贷款机制和激励机制。贷款机制可以促使终端设备尽可能多的使用边缘服务器并实现本地任务的实时处理。设计激励机制的目的方面在于减轻终端设备的支付成本,另一方面可以增加终端设备对计算资源的需求,进而提升系统的计算资源利用率。为了缩短边缘云资源定价时间并提高定价的准确性,设计了包含价格导向因素与定价方案的资源定价机制,实现边缘云的最优效益。本文将终端设备资源需求行为与边缘云资源定价行为设计成两阶段的 Stackelberg 边缘云博弈模型,该博弈模型摒弃了第三方的参与,打破了静态定价无法满足资源动态性需求的局面,实现资源供需双方直接交互。为满足移动边缘计算的实时性需求,本文设计出一种改进的强化学习 SARSA 算法,可以较快获得边缘云资源定价与终端设备资源需求的最优解。

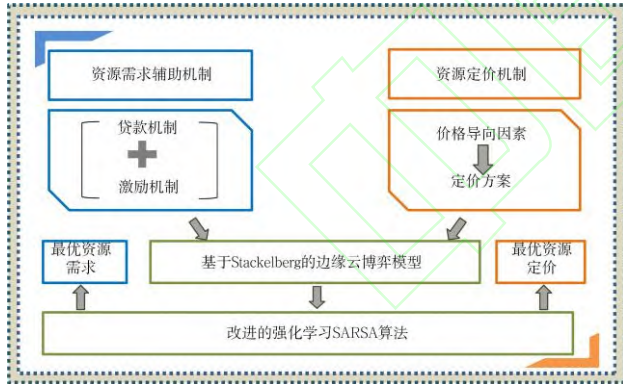


Fig.1 Edge cloud resource pricing system model.

图 1 边缘云资源定价系统模型

### 3.1 资源需求辅助机制

在资源定价场景中,当终端设备的剩余资金难以负担资源的支付费用时,本地任务将无法得到及时的处理,因此,提出包含贷款机制与激励机制的资源需求辅助机制,以缓解终端设备的任务执行压力,实现任务的及时处理。

#### 3.1.1 基于信用度的贷款机制

针对终端设备在资源定价场景下存在因资金不足导致任务被搁置问题,提出基于信用度的贷款

机制,以保证终端设备参与任务卸载的积极性。令  $a_n \in \{0,1\}$  表示终端设备的资金状况,  $a_n = 1$  表明资金充足,等待资源提供者定价;  $a_n = 0$  表明没有足够的资金购买边缘云资源,致使本地任务无法及时处理。为解决此问题,提出基于信用度的贷款机制,允许终端设备向临近的边缘云服务器发送贷款请求。

当边缘云服务器接收到贷款请求后,会立即对贷款者  $G_i$  的信用度进行多方位的审核,审核内容包括: 1) 贷款者需持有一些本地资源,可以服务外来小额任务; 2) 贷款者是否成功偿还系统规定的押金数额; 3) 根据终端设备以往贷款纪录计算的信用度为非负值。终端设备的信用度计算公式为:

$$cr(i)_n = cr(i)_{n-1} - \kappa \times num(i) \quad (1)$$

其中  $cr(i)_n$  为当前计算的信用度,设定  $cr(i)$  的初值  $cr(i)_0 = 10$ 。  $cr(i)_{n-1}$  为上轮信用度,  $num(i)$  为本轮贷款金额,  $\kappa$  为常量。

边缘云服务器会给审核通过的终端设备发放贷款金额  $num(i)$ , 并为其设定贷款利率  $\theta$  和未按时还款的惩罚因子  $\beta$ 。若贷款者在规定时间内偿还  $num(i)$  与贷款利息,则  $\varepsilon = 0$  且  $cr(i)$  值加 1。未按时还款时  $\varepsilon = 1$ , 信用度减 1, 并产生额外的惩罚金额。惩罚金额  $T(i)$  的计算公式为:

$$T(i) = \varepsilon \times num(i) \times v \quad (2)$$

$$v = \theta \left(1 + \frac{t_n}{t_i} \beta\right)$$

其中  $v$  代表逾期利率,  $t_i$  是规定的还款时间,  $t_n$  是逾期时间,  $\theta$  是贷款利率,  $\beta$  是未按时还款的惩罚因子。若  $G_i$  拒绝还款或未在系统可容忍的最大时间内还款,边缘云服务器将没收终端设备的押金,并向全网广播  $G_i$  的个人信息,所有设备都将拒绝为  $G_i$  提供资源服务。

#### 3.1.2 基于诱导因素的激励机制

终端设备进行任务卸载时,除了存在因资金不足导致任务无法执行问题外;还将面临由支付资源费用所引发的资金压力,而这会在一定程度上削弱

终端设备增加资源需求的动机。为此,提出基于诱导因素的激励机制,该激励机制鼓励终端设备向云中心共享数据以获得相应的奖励,旨在通过终端设备、边缘云服务器、云计算中心三者间的协同合作增加终端设备的总资源需求并提高资源利用率。首先,边缘云服务器作为任务处理商,会将任务处理结果相继反馈给终端设备及云数据中心。而后,云中心充当数据收集站,将给予作为原始数据提供者的终端设备相应的奖励。激励机制计算公式如下:

$$\begin{cases} v(G_i) = \frac{e_i}{\sum_{G_i \in G} e_j} \\ U_{mov} = R \times v(G_i) \end{cases} \quad (3)$$

其中  $R$  表示云中心设定的固定奖励,  $v(G_i)$  表示  $G_i$  的资源需求  $e_i$  相对于系统总资源需求的占比,  $U_{mov}$  表示终端设备基于激励机制所得奖励,其值与资源需求量  $e_i$  成正比。基于诱导因素的激励机制一方面可以减少终端设备的支付成本,增加终端设备的任务卸载积极性;另一方面,激励机制与终端设备的资源需求量相关,一定程度上促进终端设备增强资源需求的动机。

### 3.2 边缘云资源定价机制

在移动边缘计算网络中,边缘云服务器持有有限的计算资源,为激励其提供资源服务,需要为边缘云制定合理的资源价格。本文提出影响资源定价的4种导向因素,并根据导向因素为边缘云制定一致性与弹性两种定价方案,以提高定价的准确性并为后续的动态定价做准备。

#### 3.2.1 价格导向因素

价格导向因素的分析是实现边缘云资源准确定价的基石。同样,价格的准确性将进一步影响交易双方的利益,本研究提出如下4种价格导向因素。

(1) 资源总需求导向定价。边缘云服务器可根据本地资源剩余量与各终端设备资源总需求量的实时关系,动态设置资源价格。当边缘云资源剩余量较少且终端设备的总需求较大时,可适当提高资源价格以赚取更多奖励。而当资源剩余量多且

终端设备总需求较小时,可以适当降低资源价格刺激资源消费。总需求导向定价的价格调整因子  $\alpha_a$  如式4所示,其中  $Z$  为当前边缘云服务器的资源剩余量,  $E$  为各终端设备本轮的总资源需求。

$$\alpha_a = \begin{cases} -(\frac{Z-E}{Z} - 1), Z - E \geq \frac{Z}{2} \\ (\frac{Z-E}{Z} + 1), Z - E < \frac{Z}{2} \end{cases} \quad (4)$$

(2) 竞争环境导向定价。由于终端设备会优先选择可以提供相同服务且资源价格低廉的边缘云服务器,因此边缘云服务器必须考虑其他竞争者的定价策略。边缘云服务器本轮资源价格将以竞争者上轮的平均定价为依据,定义价格调整因子  $\alpha_b$  如式5所示,其中  $p$  代表服务器设定的初始资源价格,  $p_{ave r}$  代表上轮竞争者的平均资源定价。

$$\alpha_b = \frac{p}{p_{ave r}} \quad (5)$$

(3) 对象属性导向定价。边缘云服务器作为一种位置比较固定的基础设施,其身份信任度是可靠的。而终端设备因具有较强的移动性,其身份信任度很难确认。为此,本文将终端设备的行为可信度作为定价导向因素之一,边缘云通过对比各终端设备的当前可信度属性值,有选择性的为各终端设备设定资源价格,保证资源管理的安全性。定义价格调整因子  $\alpha_c$  如式6所示,其中  $m$  为常数,  $cr(i)$  表示终端设备的信用度。

$$\alpha_c = m \cdot cr(i)^{-1}, cr(i) > 0 \quad (6)$$

(4) 对象需求导向定价。各边缘云服务器可根据终端设备资源需求量的不同,调整售卖给各终端设备的资源价格,旨在激励终端设备的增加资源需求量。当终端设备资源需求量增多导致资源竞争加剧时,可适当调高资源价格获取收益;对于资源需求量少的终端设备,可通过降低资源价格来提升系统的总资源需求。定义价格调整因子  $\alpha_d$  如式7所示,其中,  $e_i$  为终端设备的本轮资源需求,上轮资源需求为  $e_0$ ,  $p_0$  为上轮资源价格。

$$\alpha_d = \frac{(e_1 - e_0) / e_0}{p_0} \quad (7)$$

### 3.2.2 定价方案

合理的定价策略不仅能够激励供需双方进行有序的资源交易，而且将促进双方利益的优化和资源的有效利用。依据不同导向因素的价格调整因子为边缘云服务器制定了一致性与弹性两种定价方案。

一致性定价下，边缘云服务器对本地资源的需求者设定相同的资源价格  $p_j$ ，该定价方案对终端设备而言是公平的，没有价格差别，更容易实施。由于在单位时间内，每个边缘云的总需求导向定价和竞争环境导向定价的各价格调整因子  $\alpha_a$ 、 $\alpha_b$  都是固定的，因此可以为接入网络的各终端设备设定一致的资源价格。一致性定价方案  $p_j$  的计算如式 8 所示。

$$p_j = \alpha \times p, \alpha = \frac{\alpha_a + \alpha_b}{2} \quad (8)$$

弹性定价方案下，由于每个终端设备的属性值与资源需求量不同，因此，边缘云服务器可根据对象属性与对象资源需求导向定价的价格调整因子  $\alpha_c$ 、 $\alpha_d$  的差别性为终端设备制定不同的资源价格。弹性定价方案  $p_j$  的计算如式 9 所示。

$$p_j = \alpha \times p, \alpha = \frac{\alpha_c + \alpha_d}{2} \quad (9)$$

## 4 最优资源需求与定价决策

为获得终端设备资源需求及边缘云定价的最优策略，首先根据用户需求量与系统资源定价之间的实时关系，引入斯坦克伯格 (Stackelberg) 动态博弈模型，并分别为资源供需双方设置效用函数。其次，验证了非合作博弈中终端设备与边缘云服务器各纳什均衡点的存在性。最后，采用强化学习算法求解最优策略问题，旨在实现边缘云收益最大与终端设备支付成本最小。

### 4.1 基于 Stackelberg 的边缘云博弈模型

由于边缘云服务器的资源定价策略是一个动

态优化过程，现有的静态定价策略可能无法满足终端设备的动态性资源需求。因此，有必要设计一种合理的动态资源定价策略，以联合优化边缘云服务器的资源效益和终端设备的支付成本。本文构建了基于 Stackelberg 博弈的资源定价模型，以实现资源需求与资源价格的动态调整。

#### 4.1.1 Stackelberg 博弈制定



Fig.2 The Stackelberg game between resource supply and demand.

图 2 资源供需双方的 Stackelberg 博弈过程

终端设备与边缘云的博弈过程分为两个阶段，边缘云作为领导者，移动终端作为追随者。边缘云服务器集合记作  $Y = \{Y_1, Y_2, \dots, Y_j, \dots, Y_n\}$ ,  $j \in n$ ，接入边缘云系统的终端设备集合记作  $G = \{G_1, G_2, \dots, G_i, \dots, G_N\}$ ,  $i \in N, N > n$ 。如图 2 所示，在第 I 阶段，边缘云服务器  $Y_j$  首先根据价格导向因子的不同为每个终端设备制定初始的资源价格集  $p_j = \{p_{G_1}, p_{G_2}, \dots, p_{G_i}, \dots, p_{G_N}\}$ ，其中  $p_{G_i}$  表示边缘云服务器  $Y_j$  对终端设备  $G_i$  制定两种定价方案下的资源价格， $p_j \in \{\underline{p}, \bar{p}\}$ ， $\underline{p}$  是资源的成本价格， $\bar{p}$  是终端设备所能接受的最高价格。边缘云服务器将根据终端设备需求策略的变化，动态调节资源定价以最大化资源收益， $p_j^*$  表示  $Y_j$  的最优资源定价。定义  $\{U_{Y_j}\}_{j \in n}$  为边缘云服务器  $Y_j$  的效用函数，它主要由终端设备的资源支付费用和自身能耗损失成本  $c$  两部分构成。边缘云服务器的资源效用函数如式 10 所示：



$$U_{Y_j} = \begin{cases} \sum_{G_i \in G} p_j e_i - \sum_{G_i \in G} c e_i, a_n = 1 \\ \sum_{G_i \in G} [p_j(1 + \theta)e_i + \varepsilon p_j e_i(1 + \frac{t_n}{t_i} \beta) \theta] - \sum_{G_i \in G} c e_i, a_n = 0 \end{cases} \quad (10)$$

在 II 阶段, 终端设备根据领导者制定的资源价格调整资源需求量, 以最小化支付成本。终端设备的预估资源需求集合为  $e = \{e_1, e_2, \dots, e_i \dots e_N\}$ ,  $e_i \in [\underline{e}, \bar{e}]$ , 其中  $\underline{e}$  为终端设备任务卸载的最小资源需求,  $\bar{e}$  为服务器所能分配的最大资源量,  $e_i^*$  表示各终端设备的最优资源需求。终端设备支付成本的效用函数为  $\{U_{G_i}\}_{i \in N}$ , 效益函数由支付给边缘云服务器的资源费用和激励机制的奖励两部分组成。终端设备支付成本的效益函数如式 11 所示:

$$U_{G_i} = \begin{cases} p_j e_i - R \frac{e_i}{\sum_{G_j \in G} e_j}, a_n = 1 \\ p_j(1 + \theta)e_i + \varepsilon p_j e_i(1 + \frac{t_n}{t_i} \beta) \theta - R \frac{e_i}{\sum_{G_j \in G} e_j}, a_n = 0 \end{cases} \quad (11)$$

阶段 I、II 共同构成 Stackelberg 动态博弈, 博弈的目标旨在证明纳什均衡点  $(e^*, p^*)$  的存在性与唯一性。判定 Stackelberg 博弈成立的条件为:

$$U_{G_i}(e^*, p^*) \leq U_{G_i}(e, p^*), U_{Y_j}(e^*, p^*) \geq U_{Y_j}(e^*, p) \quad (12)$$

#### 4.1.2 Stackelberg 博弈分析

为了验证 Stackelberg 博弈纳什均衡点的存在性, 本文分析了边缘云服务器效用函数  $U_{Y_j}$  和终端设备的效用函数  $U_{G_i}$  的一阶、二阶导数。

$$\frac{\partial v(G_i)}{\partial e_i} = \frac{\sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^2} \quad \frac{\partial^2 v(G_i)}{\partial e_i^2} = \frac{-2 \sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^3} < 0 \quad (13)$$

$$\frac{\partial U_{G_i}}{\partial e_i} = p - R \cdot \frac{\sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^2} \quad (14)$$

$$\frac{\partial^2 U_{G_i}}{\partial e_i^2} = -R \cdot \frac{\sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^3} = \frac{2R \sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^3} > 0$$

$$\frac{\partial U_{G_i}}{\partial e_i} = 0, e_i^* = \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p}} - \sum_{G_j \in G} e_j \quad (15)$$

$$U_{Y_j} = \sum_{G_i \in G} (p_j - c) \left( \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p_j}} - \sum_{G_i \in G} e_j \right)$$

$$\frac{\partial U_{Y_j}}{\partial p_j} = \sum_{G_i \in G} \left( \frac{1}{2} \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p_j}} - \sum_{G_i \in G} e_j + \frac{c}{2} \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p_j^3}} \right) \quad (16)$$

$$\frac{\partial^2 U_{Y_j}}{\partial p_j^2} = \sum_{G_i \in G} \left( -\frac{1}{4} \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p_j^3}} - \frac{3c}{4} \sqrt{\frac{R \cdot \sum_{G_j \in G} e_j}{p_j^5}} \right) < 0$$

综上可知, 当  $a_n = 1$  时,  $U_{G_i}$  相对于  $e_i$  是严格凹,  $U_{Y_j}$  相对于  $p_j$  是严格凸的, 证明了博弈双方纳什均衡解的存在性。同理可证  $a_n = 0$  时 Stackelberg 博弈纳什均衡点的存在性。

$$\frac{\partial^2 U_{G_i}}{\partial e_i^2} = 2R \frac{\sum_{G_j \in G} e_j}{(\sum_{G_j \in G} e_j)^3} > 0$$

令  $l = 1 + \theta + a$  (17)

$$\frac{\partial^2 U_{Y_j}}{\partial p_j^2} = \sum_{G_i \in G} \left( -\frac{1}{4} \sqrt{\frac{lR \sum_{G_j \in G} e_j}{p_j^3}} - \frac{3c}{4} \sqrt{\frac{R \sum_{G_j \in G} e_j}{p_j^5 l}} \right) < 0$$

#### 4.2 基于改进的 SARSA 算法纳什均衡求解

通过凹凸函数证明了终端设备与边缘云纳什均衡解的存在性, 本节提出改进的强化学习算法得到博弈双方最优需求与定价决策。强化学习中很多场景涉及多智能体<sup>[14,15]</sup>系统的交互, 比如多玩家的博弈游戏。本节将终端设备与边缘云服务器的行为建模为多智能体交互, 旨在实现所有设备的效益最大化<sup>[16]</sup>。

##### 4.2.1 多边缘设备智能体设计

多边缘智能体设备进行强化训练时, 每个智能体的状态和行为易受其他设备影响, 因此需要将多个设备同步强化学习。强化学习算法的基本要素为状态  $s$ , 动作  $a$ , 策略  $\pi(s, a)$ , 即时奖励



$r(s^t, a^t, s^{t+1})$ 。定义终端设备的动作空间为  $A_G$ ，资源需求  $e \in A_G$ 。边缘云服务器的动作空间为  $A_Y$ ，资源定价  $p \in A_Y$ 。边缘云服务器和终端设备的状态空间分别为  $S_Y$ 、 $S_G$ ，策略函数分别为  $\pi(s_Y, a_Y)$ 、 $\pi(s_G, a_G)$ ，即时奖励分别为  $U_{Y_j}$ 、 $U_{G_i}$ 。首先，边缘云服务器基于终端设备的需求状态  $s_Y^t$  做出本轮定价动作。同样，终端设备基于状态  $s_G^t = [p^t]$  响应其资源需求行动  $e$ ， $p^t$  代表本轮资源定价。

#### 4.2.2 基于改进的 SARSA 算法策略

针对现有最优求解方案存在适用性低、实现全局最优难度大及时间复杂度高的问题，提出改进的 State Action Reward State Action (SARSA) 算法解决上述问题，算法思想如表 1 所示。由于 SARSA 算法的采样和优化策略都是  $\pi$ ，不需要对所有策略列举并计算，时间复杂度低，但这也导致了学习效果可能不理想。为此，本文引入平均状态动作函数  $\bar{Q}$ ，通过监测  $\bar{Q}$  值不断调整 SARSA 算法的学习效率  $\lambda$ ，以尽可能快的学习到最优解。SARSA 算法作为一种确定性策略，无法很好地应对多变环境的挑战，因此，设计了状态转移概率  $P_a$ 、动作转移概率  $P_b$ ，将确定性策略转化为随机性策略，旨在更好地适应复杂多变的环境。

强化学习的目标是通过训练得到最优策略  $\pi^*(s, a)$ ，并通过状态值  $V^\pi(s)$ 、状态动作值  $Q^\pi(s, a)$  评估期望回报。若满足  $Q^\pi(s, a) > V^\pi(s)$ ，则可以调整  $p_a$ 、 $p_b$  增加策略  $\pi(s, a)$  的概率。终端设备及边缘云服务器  $Q$  值的更新策略分别为：

$$\begin{aligned} Q_G^\pi(s_G^t, e^t) &\leftarrow Q_G^\pi(s_G^t, e^t) + \alpha(r + \gamma Q_G^\pi(s_G^{t+1}, e^{t+1}) - Q_G^\pi(s_G^t, e^t)) \\ Q_Y^\pi(s_Y^t, p^t) &\leftarrow Q_Y^\pi(s_Y^t, p^t) + \alpha(r + \gamma Q_Y^\pi(s_Y^{t+1}, p^{t+1}) - Q_Y^\pi(s_Y^t, p^t)) \end{aligned} \quad (18)$$

对于策略  $\pi(s^t, a^t)$ ，根据状态转移概率、动作转移概率产生新的随机性策略  $\pi(s^{t+1}, a^{t+1}) = p(s^t, a^t)$ 。若新的随机策略满足

$\pi^*(s) = \arg \min_a Q^\pi(s, e)$ ，则  $\pi^*(s)$  为最优需求策略。若满足  $\pi^*(s) = \arg \max_a Q^\pi(s, p)$ ，则  $\pi^*(s)$  为最优定价策略。

引入状态动作均值  $\bar{Q} = (s^t, a)$ ，作为评判学习效果依据，将  $Q$  值与  $\bar{Q}$  值进行对比，学习率  $\lambda \in (0, 1]$ 。对于边缘云服务器而言，若  $Q > \bar{Q}$ ，表明学习效果良好，可以适当降低  $\lambda$  使其他设备适应当前策略，并增加  $p_a$ 、 $p_b$ ，使得策略  $\pi(s, a)$  的概率增加。若  $Q < \bar{Q}$ ，适当增加  $\lambda$  值以提升学习效果。终端设备  $\alpha$  的调整策略与边缘云服务器恰好相反。 $\bar{Q}$  随时间动态调整，引入  $Num$  来记录并更新状态的改变次数。

$$\bar{Q}(s^t, a^t) \leftarrow \bar{Q}(s^t, a^t) + \Delta, a^t \in A \quad (19)$$

$$\Delta = \frac{\bar{Q}(s^t, a^t) - Q(s^t, a^t)}{Num} \quad (20)$$

Table 1 The improved SARSA algorithm

表 1 改进的 SARSA 算法

Algorithm 1 The improved SARSA algorithm for edge cloud server	
<b>Input:</b> state space $S$ , action space $A$ , discount rate $\gamma$ , learning rate $\lambda$ , instant reward $r$	
1. Randomly initialize $V(s) = 0, Q(s, a)$ ;	
2. $\forall s, \forall a, \pi(s, a) = \frac{1}{ A }$ ;	
3. <b>Repeat</b>	
4. Initialize the random $s$ ;	
5. Select action $p$ with the probability policy $\pi(s, p)$ ;	
6. <b>Repeat</b>	
7. Execute action $p$ , get instant reward $r$ and new state $s^t$ according to $P_a$ ;	
8. In state $s^t$ , select action $p^t$ with the probability policy $\pi(s^t, p^t)$ ;	
9. According to $\pi(s^t, p^t)$ , compare and update $Q, \bar{Q}$ :	
$Q^\pi(s, p) \leftarrow Q^\pi(s, p) + \alpha(r + \gamma Q^\pi(s^t, p^t) - Q^\pi(s, p))$	
10. $\bar{Q}(s, p) \leftarrow \bar{Q}(s, p) + \frac{Q(s, p) - \bar{Q}(s, p)}{Num}, p \in A_Y$ ;	
11. <b>If</b> $Q < \bar{Q}$	
12. Increase $\lambda$ ;	
13. <b>Else</b>	
14. Decrease $\lambda$ ;	

---

```

15.   If  $Q^\pi(s, p) > V^\pi(s)$ 
16.     Increase the probability of  $\pi(s, p)$ ;
17.   Else
18.     Update  $\pi$ ;
19.    $\pi(s, p) = \arg \max Q(s, p)$ ;
20.    $s \leftarrow s', p \leftarrow p'$ ;
21.   Until  $s$  is the termination state;
22. Until  $\forall s, p, Q(s, p)$  convergence;

```

---

Output: policy  $\pi^*$

---

## 5 仿真实验

在本节中，首先评估改进的多智能体强化学习算法的性能。然后根据实验结果证明，所提定价机制在最大化资源供需双方利益方面优于现有的其他算法。最后，通过广泛的数值模拟两种定价方案对最优资源需求及定价的影响，旨在得到最优定价策略。

### 5.1 实验对比

仿真实验基于 Python 语言环境进行，在仿真模拟中，通过强化学习算法不断训练得到资源供需双方的最优策略。简单的将终端设备的资源需求动作空间定义为  $A_G = \{20, \dots, 90\}$ ，边缘云服务器的定价动作空间定义为  $A_Y = \{50, \dots, 90\}$ 。一些默认参数值设置如下：设定  $\kappa = 0.0005$ ， $m = 2$ ，边缘云的资源成本  $c = 2$ ，贷款利率  $\theta = 2\%$ ，惩罚因子  $\beta = 5\%$ ，折扣率  $\gamma = 0.6$ ，学习率  $\lambda \in (0, 1]$ 。为保证多 agent 能够收敛到最优解，将最大集数设定为 6000。首先通过仿真实验将现有最优策略的求解方法与所提强化学习方法对比，仿真实图像如下：

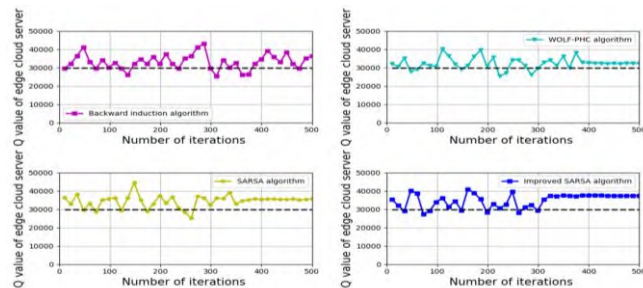


Fig.3 Comparison of four algorithms for solving optimal strategies

图 3 四种求解最优策略的算法比较

图 3 展示了 4 种算法的收敛性能与环境适应性。由图可知逆向回归算法未在有限迭代次数内收敛，其余三种算法均收敛于一个稳定值，且本文所提算法的迭代次数最少。这是因为逆向回归算法无法根据环境的改变及时调整路径，需进行全局搜索，导致收敛性能差。而改进的 SARSA 算法通过状态、动作转移概率，增加可能取得最大 Q 值的策略概率。在环境适应性方面，WOLF-PHC 与改进的 SARSA 算法优于另外两种算法，这是由于两种算法为尽快的训练出最优策略，引入学习因子  $\lambda$ ，并依据 Q 值得变化实时调整  $\lambda$ 。基于 SARSA 与改进的 SARSA 学习算法收敛性能好，但由于 SARSA 算法为确定性策略，易受环境影响，学习效果欠佳。通过与各种算法对比可知，改进的 SARSA 算法的性能优于其他算法。

图 4 是将本研究所提的 2 种定价方案与文献 [9]、文献[11]中的定价方案进行比较，旨在探究各定价方案对边缘云效益的影响。可以看出，随着迭代过程的推进，效用值曲线整体呈现上升规律。当迭代的次数达到 21 次，除了动态拍卖算法外，其余 3 种定价下服务器的效用值不再增加，收敛于最大效益值。这是因为动态拍卖算法将每轮的拍卖结果作为优化目标，导致结果难以趋近全局最优，收敛性能差。除此之外，弹性定价方案下边缘云效益值高于其他 3 种定价方案，效用值优于一致性方案 24%。这是因为本文通过价格导向因子制定资源价格，价格制定的准确性对博弈过程中边缘云实现最大化自身效益至关重要。

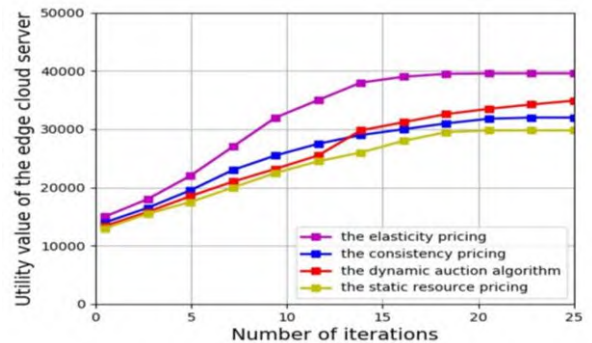


Fig.4 Comparison of convergence of edge cloud utility functions.

图 4 边缘云效用函数的收敛情况比较

## 5.2 实验分析

本节通过实验数据比较一致性定价与弹性定价对最优资源需求与定价策略的影响，旨在为终端设备及边缘云选择最大化自身利益的定价方案。

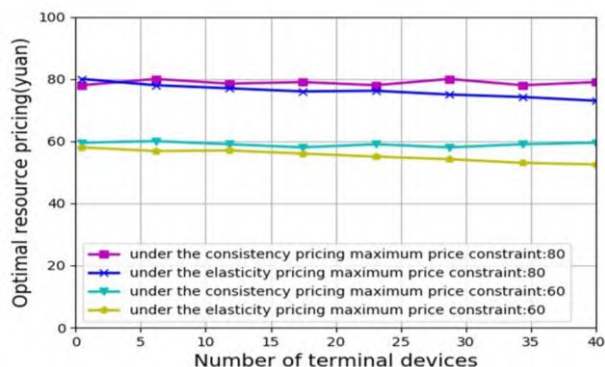


Fig.5 The impact of the number of terminal device on the optimal resource pricing.

图 5 终端设备数量对最优资源定价的影响

首先评估两种定价方案下终端设备数量对最优资源定价的影响，如图 5 所示。在最高资源定价分别为 80 和 60 的情况下，一致性定价的最优价格均近似于最大价格，这是因为边缘云服务器的效益随价格的增加而增加，并有意将一致性定价下的最优价格设置为临近最大价格。而弹性定价下，最优价格略低于一致性定价，且最优价格随终端设备数目的增加而缓慢下降。这是由于弹性定价下，边缘云服务器可以有选择的为终端设备设定不同的资源价格，旨在鼓励终端设备购买更多的资源。

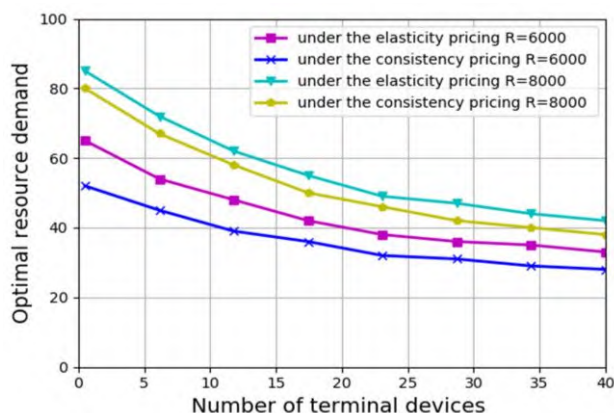


Fig.6 The impact of the number of terminal device on the optimal resource pricing.

图 6 终端设备数目对最优资源需求的影响

图 6 评估了终端设备数量对最优资源需求的影响。可以看出，随着终端设备的增加，单个设备的最优资源需求下降。这是因为终端设备增多会加剧设备间的资源竞争，导致自身获得资源量减少。但弹性定价下的最优资源需求高于一致性定价，这是因为弹性定价可以基于定价导向因素为终端设备设置不同的资源价格，使得终端设备有动力申请更多的资源服务。且  $R = 6000$  时的资源需求低于  $R = 8000$ ，这是由于奖励增加可以降低终端设备的支付成本，从而激励终端设备增加资源需求量。

图 7 展示了两种定价方案下终端设备效用函数的收敛情况，即支付成本的收敛情况。可以看出，当迭代次数为 18 时，2 种方案下终端设备的效用值达到最小且不再变化，并且弹性定价下的支付成本略低于一致性定价 7%。这是因为弹性定价会促使终端设备增加资源需求量，获得更多的云中心奖励，并且弹性定价下最优资源定价低于一致性定价。因此，终端设备在弹性定价方案下实现最小化支付成本，获得资源需求的纳什均衡解。

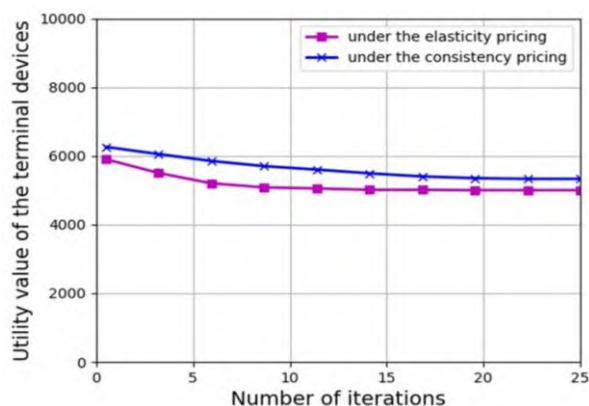


Fig.7 Convergence of utility functions of terminal devices under two pricing schemes

图 7 两种定价方案下终端设备效用函数的收敛情况

## 6 总结全文

针对现有定价机制在定价导向因素与最优定价求解方面存在价格制定不准确、最优解难以趋近全局最优、依赖中间商的静态定价费用高、资源利用率低、执行时间复杂度高等上述问题，本文构建了一种基于 Stackelberg 博弈的边缘云资源定价机制



模型。首先,提出了包含贷款和激励的辅助机制以促使终端设备任务的及时处理;其次,提出了影响资源定价的4种价格导向因素,并据此制定一致性与弹性两种定价方案,提高了定价的准确性和效率;然后,构建了基于Stackelberg博弈的资源定价机制模型,实现了资源需求与资源价格的动态调整,联合优化了边缘云服务器的资源效益与终端设备的支付成本;最后,通过改进的强化学习SARSA算法得到了资源需求及定价的最优策略。实验表明,改进的SARSA算法在收敛性能、学习效果、边缘云效益等方面优于其他最优求解算法,同时终端设备和边缘云服务器的效用函数均在弹性定价方案下实现最优。本文所提定价方法在多终端设备同时与同一边缘云服务器博弈的情况下,博弈进程的时间可能略有增加,后续工作将针对现有方案存在的不足之处加以完善。

## 参考文献:

- [1] Shi Weisong, Cao jie, Zhang Quan, et al. Edge computing: Vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646
- [2] 谢人超, 廉晓飞, 贾庆民, 等. 移动边缘计算卸载技术综述[J]. 通信学报, 2018, 039(011): 138-155.  
XIE RC, LIAN XF, JIA QM, et al. Survey on computation offloading in mobile edge computing[J]. Journal on Communications, 2018, 039(011): 138-155.
- [3] 张开元, 桂小林, 任德旺, 等. 移动边缘网络中计算迁移与内容缓存研究综述[J]. 软件学报, 2019, Vol.30 Issue(8): 2491-2516.  
Zhang KY, Gui XL, Ren DW, et al. Survey on computation offloading and content caching in mobile edge networks[J]. Journal of Software, 2019, 30(8): 2491-2516
- [4] 邓晓衡, 关培源, 万志文, 等. 基于综合信任的边缘计算资源协同研究[J]. 计算机研究与发展, 2018, 55(3): 5-33.  
Deng XH, Guan PY, Wan ZW, et al. Integrated Trust Based Resource Cooperation in Edge Computing[J]. Journal of Computer Research and Development, 2018, 55(3): 449-477.
- [5] PAN Jian li and MCELHANNON J. Future edge cloud an edge computing for internet of things applications[J]. IEEE Internet of Things Journal, 2018, 5(1): 439-449.
- [6] Yao H, Mai T, Wang J, et al. Resource Trading in Blockchain-based Industrial Internet of Things[J]. IEEE Transactions on Industrial Informatics, 2019, 2(3): 1-1.
- [7] Zhang, Yongmin, Lan, Xiaolong, Lan, Xiaolong, et al. Efficient Computation Resource Management in Mobile Edge-Cloud Computing[J]. IEEE Internet of Things Journal, 2018, 5: 1-1.
- [8] 张伊璇, 何泾沙, 赵斌, 等. 一个基于博弈理论的隐私保护模型[J]. 计算机学报, 2016, v.39; No.399 (3): 177-189.  
ZHANG YX, HE JS, ZHAO B, et al. A privacy protection model based on game theory[J]. Chinese Journal of Computers, 2016, v.39; No.399 (3): 177-189.
- [9] 董思岐, 李海龙, 屈毓铨, 等. 移动边缘计算中的计算卸载策略研究综述[J]. 计算机科学, 2019(11): 1-12.  
DONG SQ, LI HL, QU YB, et al. A review of computational unloading strategies in moving Edge computing[J]. Computer Science, 2019(11): 1-12.
- [10] LIU M, LIU Y. Price-based distributed offloading for mobile-edge computing with computation capacity constraints[J]. IEEE Wireless Communications Letters, 2018, 7(3): 420-423.
- [11] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang. Consortium blockchain for secure energy trading in industrial internet of things[J]. IEEE Transactions on Industrial Informatics, 2018, 14(8): 3690-3700.
- [12] Zhang H, Guo F, Ji H. et al. Combinational auction-based service provider selection in mobile edge computing networks [J]. IEEE Access. 2017, 5: 13455-13464.
- [13] Xiong Z, Feng S, Wang W, et al. Cloud/fog computing resource management and pricing for blockchain networks[J]. IEEE Internet of Things Journal, 2019, 6(3): 4585-4600.
- [14] 杜威, 丁世飞. 多智能体强化学习综述[J]. 计算机科学 2019, 46(8): 1-8.  
DU W and DING SF. Overview on Multi-agent Reinforcement Learning[J]. Computer Science, 2019, 46(8): 1-8.
- [15] 张文旭, 马磊, 贺荟霖, 等. 强化学习的地-空异构多智能体协作覆盖研究[J]. 智能系统学报, 2018, 13(02): 202-207.  
ZHANG Wenxu, MA Lei, HE Huilin, et al. Air-ground heterogeneous coordination for multi-agent coverage based on reinforced learning[J]. CAAI Transactions on Intelligent Systems, 2018, 13(02): 202-207.
- [16] L. Busoniu, R. Babuska, and B. De Schutter. A comprehensive survey of multiagent reinforcement learning[J]. IEEE Transactions on Systems Man and Cybernetics, 2008, 38(2): 156-172.





LIU Jingxin was born in 1996. She is an M.S. candidate at Liaoning University, and the student member of CCF. Her research interests include edge computing resource management and task scheduling, etc.

刘荆欣(1996-), 女, 山东招远人, 辽宁大学硕士研究生, CCF 学生会员, 主要研究领域为边缘计算资源管理与任务调度等。



WANG Yan was born in 1978, Ph.D. degree. She is a professor and M.S. supervisor at Liaoning university, and the Senior Member of CCF. Her research interests include IIoT data processing, task scheduling and big data technology, etc.

王妍(1978-), 女, 辽宁抚顺人, 博士学位, 现为辽宁大学信息学院教授、研究生导师, CCF 普通会员, 主要研究领域为工业物联网数据处理, 任务调度, 大数据技术等。



HAN Xiao was born in 1994. He is an M.S. candidate at Liaoning university. His research interests include task scheduling under edge computing, etc.

韩笑(1994-), 男, 辽宁锦州人, 辽宁大学研究生, 主要研究方向为边缘计算下的任务调度等。



XIA Changqing was born in 1985, Ph.D. degree. He is an assistant researcher of Shenyang Institute of Automation. His research interests include industry network and task scheduling in Edge Computing, etc.

夏长清(1985-), 男, 山东威海人, 博士学位, 沈阳自动化研究所助理研究员, CCF 普通会员, 主要研究方向为工业物联网, 边缘计算下的任务调度等。



SONG Baoyan was born in 1965, Ph.D. degree. She is a professor and M.S. supervisor at Liaoning university, and the Senior Member of CCF. Her research interests include database theory and techniques and big data management, etc.

宋宝燕(1965-), 女, 辽宁开原人, 博士学位, 现为辽宁大学信息学院教授、研究生导师, CCF 高级会员, 主要研究方向为数据库理论和技术, 大数据管理等。