



云边协同中的资源调度优化

王淑玲, 孙杰, 王鹏, 杨爱东

(亚信科技(中国)有限公司, 北京 100193)

摘要: 随着业务类型的丰富和多样化, 低时延、高带宽、数据私密性、高可靠性等成为业务普遍的要求。边缘计算、雾计算、分布式云、算力网络等方案相继被提出, 并在产学研各界引发了深度的研究和探索。针对“多级的算力分布以及算力的协同将是未来算力结构的主流”这一观点, 产业内外达成了共识, 算力管理、分配、调度等与资源优化相关的问题也成为当下的研究热点和重点攻关方向。为此, 面向未来的算力供给结构, 首先描述了学术界、产业界资源调度优化问题的最新进展, 总结了当前的主要方法论和工程实施架构; 然后, 针对两种典型的云边协同场景, 从场景拆分、调度目标、求解方案依次进行分析, 给出了适应场景特性的资源调度优化参考方案。

关键词: 云边协同; 边缘计算; 算网融合; 资源调度优化; 算网联合优化

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2023027

Resource scheduling optimization in cloud-edge collaboration

WANG Shuling, SUN Jie, WANG Peng, YANG Aidong

Asia Info Technologies (China) Co., Ltd., Beijing 100193, China

Abstract: With the enrichment and diversification of business types, low latency, high bandwidth, data privacy and high reliability have become common requirements. Edge computing, fog computing, distributed cloud, computing power network and other solutions have been proposed, and have triggered in-depth research and exploration in industry, academia and research. There is a consensus within and outside the industry on the view that “multi-level computing power distribution and collaboration of computing power will be the mainstream of computing power structure in the future”. The problems related to resource scheduling optimization, such as computing power management, allocation, scheduling, have also become the current research hotspot and key research direction. Therefore, for the future computing power supply structure, focuses on the latest progress of resource scheduling optimization in academia and industry, the current main methodology and engineering implementation architecture was summarized. And then, for the two typical cloud edge collaboration scenarios, the analysis was carried out from the perspective of scene splitting, scheduling objectives, and solutions in turn, and the resource scheduling optimization reference schemes that adapted to the characteristics of the scenarios were analyzed and discussed respectively.

Key words: cloud-edge collaboration, edge computing, computer and network convergence, resource scheduling optimization, computing and networking joint optimization



0 引言

当前,社会正加速进入以数字化、网络化、智能化为特征的信息社会。时低延、高带宽、数据私密性、高可靠性等新时代业务的普遍诉求,推动基础设施进入后云计算时代,边缘计算、雾计算(fog computing)、分布式云、算力网络等方案相继被提出,并在产、学、研各界引发了深度的研究和探索。

针对边缘计算的不同部署范式,多种概念被提出。多接入边缘计算(multi-access edge computing, MEC)^[1]由欧洲电信标准化协会(ETSI)率先提出,倡导在靠近业务的网络边缘侧部署计算处理能力,以满足未来业务对低时延、大带宽、高可靠的要求。雾计算是一种分布式的计算基础设施,通过将计算和存储设备配置在互联网边缘减少互联网数据传输量,从而降低时延、节省带宽及相关费用。文献[2]系统地阐述了雾计算的关键技术,列举了常见的雾计算应用。分布式云由ITU-SG13^[3]提出,强调将公有云服务分布到不同的物理位置,且服务的更新、管理仍由云服务商提供。算力网络是2019年由中国运营商在ITU-SG13^[4]上首次提出来的概念,其宗旨在于提供一体化的算力资源和网络资源的联合调度优化模型,以提升基础设施资源的使用效率。

类似的概念还有很多,尽管各自强调的特点、提出的背景都不尽相同,但是,针对“多级的算力分布以及算力的协同将是未来算力结构的主流”这一观点,产业内外达成了共识,算力管理、分配、调度等资源调度优化相关的问题也成为当下的研究热点和重点攻关方向。

1 资源调度优化问题研究进展

资源调度优化问题是学术界和产业界的经典问题,在多种系统及场景下普遍存在。

针对云一边算力协同场景,学术界的工作聚

焦于:在算力和网络的联合资源约束、经济性和绿色节能等效能要求下,用数学工具对资源调度优化问题进行建模并求解。文献[5]和文献[6]重点阐述了近3年内该领域的工作。文献[5]针对“云一边一端”的计算系统,全面阐述了单一任务的卸载或分配决策、多任务分配及调度、边缘计算环境中任务优先级约束等几种场景下的问题求解方法。文献[6]将资源调度优化问题进一步细分为资源映射、任务调度、任务流调度3个子问题。结合子问题的数学模型类型,系统地总结了问题求解的方法和适用场景,包括:(1)经典的先来先服务(first come and first service, FCFS)、基于优先级等的调度方法^[7-8];(2)基于模糊理论的调度方法^[9],重点解决边端的资源不确定性和动态性问题;(3)启发式的调度方法^[10-11],以有限代价求解NP难问题;(4)强化学习算法^[12],应对环境的不确定性,提升系统的自优化、自学习能力。

在学术界,对于资源调度和优化问题,学者们提供了非常多的参考模型,但是,这些模型在走向工业生产时遇到了一些问题。这些问题主要体现在以下两个方面:第一,参考模型基于的前提和假设与实际场景存在的差异;第二,模型以及求解过程中涉及的数据,在实际的系统中较难收集。因此,需要将实际工程的场景与学术参考模型的前提进行映射,进而选择适宜的方案。

在产业界,云一边协同的概念和实现方案尚未形成一个共识。首先,在概念上,国内和国际已经分化。国际上,以Gartner为首,将云一边协同中的算力资源调度和网络资源调度分别映射至分布式云(distributed cloud)^[13]和云网络(cloud networking)^[14]两个分支,推荐的供应商也分别属于运营商和云商两个阵营。在国外,算力和网络分属于两个不同的产业角色,要将其联合调度难度较大,因此,从全局资源统筹的角度上来看,两类资源是分开调度的。在国内,算力基础设施

的供应一直存在着较为激烈的争夺。运营商借助有网络、有算力的优势,在2018年就已经提出云网融合^[15]的概念,以撼动云厂商算力运营服务的头部地位。在2019年,中国运营商开始在ITU提出算力网络^[4]的概念,将云网资源统一协调和调度的想法又一次抛了出来,这一概念最初仅是云资源、网络资源的联合编排以及运营层的融合,但是随着算力网络热潮的迅速到来、“东数西算”工程的快速推进,产业界已经着手考虑云、网资源联合调度的技术攻关。在商业实现上,Kubernetes是当前云边资源管理的主流框架。这一系统中的资源调度,重点还是考虑云资源的特性以及云内的算网协同,在其调度算法中更多考虑云资源的状态等信息。针对这一问题,产业界针对边缘算力的需求和特性做了一定的改进,如Kuberedge、Superedge等,但是算网资源效能、跨管理域中的资源协调等方面考虑得较少。

算力和网络的底层基础设施资源联合调度是一个技术门槛较高的领域。在过去的十多年里,基础资源的调度和优化一直以云商主导的IT路线为主。因此,整体以云和分布式云这套体系展开工作,云边协同中的数据传输、镜像传输等网络需求是以一种较为粗放的方式满足的。尽管做了一些工作,但是没有太流行起来,对于成熟商用的系统,常常也都讳莫如深,并不过多暴露细节。为此,本文将面向未来的算力供给结构,结合学术界和产业界的研究进展,给工程实施提供一些框架性的思考。

2 云边协同中的资源调度优化问题定义

尽管雾计算、MEC、分布式云、算力网络等概念存在管理结构、算力形态等方面的差异,但是多级的、协同一体化的算力结构是一致的。

过去,在多级算力协同的场景下,针对算力如何组织、管理和调度,是通过集中式还是分布自治的方式,如何能够达到资源效率的最优等关

键问题,产业界内外存在不同的观点和解决方案。文献[18]从量化的角度证实,在多权属的算力资源池中进行资源调度时,资源池的整体使用效率会得到提升。因此,面向未来的数字化经济,需要打破集中式云算力的限制,利用网络连接多级、多权属算力,获取强大的算力基础设施资源。

一般地,资源调度优化问题定义如下。

定义1: 映射 $f: A \rightarrow I$, 使得 $O(f, A, \mathfrak{R}, I)$ 最优。

其中, $A = \{a_1, a_2, \dots, a_m\}$ 为用户提交的资源请求集合; $\mathfrak{R} = \{R_A, R_I\}$, $R_A = \{R_{a_1}, R_{a_2}, \dots, R_{a_m}\}$ 为用户提交的资源需求, $R_I = \{R_{I_1}, R_{I_2}, \dots, R_{I_n}\}$ 为资源节点 I 的资源约束; $O(f, A, \mathfrak{R}, I)$ 表示在给定 A 、 $\mathfrak{R}$ 和 I 的情况下,映射 f 带来的目标效能。

过去讨论资源调度优化时,需求方的需求 R_A 是直接描述资源需求的,如200 MHz的上行带宽或者2vCPU、64 GB内存、120 TB存储等可直接量化的资源集合。因此,资源的所有方在自有的资源池中,在满足各种约束条件的情况下,找到映射 f , 满足用户需求的同时,优化 O 。这种调度模式属于资源分配型调度,本质上仍是“规划—建设”的粗放式资源使用方式。随着云原生技术、容器技术等的发展,函数计算服务在云平台率先实现,使得根据业务的运行时态动态调整资源的调度和分配成为可能,实现了真正意义上的按需服务,属于服务型调度。但是,网络的调度和优化仍然属于资源分配型调度,动态适配业务的需求挑战较大。算力和网络不同维度、粒度的调度方式如何能够协作起来,以满足业务侧的弹性、动态、灵活等特性要求?这仅仅是云边协同场景中的调度面临的挑战之一,其另一挑战在于权属结构的多样性下的算力管理。过去无论是在网络侧还是在算力侧,厂商对于可调度的资源都有绝对的权属和控制权,拥有集中式的手段管理资源的状态、使用等信息。在云边协同场景下,算力的泛在特性也带来了管理的泛在性,如何构建一个分布式的管理结构也是一大课题。后文将试图



从两个典型的云边协同的场景中,找到一些参考。

3 云边协同场景中资源调度优化问题的实施参考

3.1 工程实施方案三步骤

在求解定义 1 描述的问题时,算力的权属、算力的结构、调度问题的决策主体等直接影响调度优化的维度,因此,在实际场景中,为了找到满足资源约束、目标约束的 f ,本文认为需要依次明确 3 个方面:场景、目标及约束、建模及求解。

(1) 场景

在云边多级算力协同的场景下,何方拥有算力的绝对管控权直接关系到问题的求解思路,具体地,会影响定义 1 中的目标函数 O 、资源约束 R 和映射 f 的设计等。

在集团企业自建的云边协同架构下,云边算力为企业私有,边缘算力调度目标 O 是在有限的资源内完成尽可能多的任务。集中式的调度方法所需的绝对控制权、数据使用权在此种场景下都能够得到充分的满足。在分布式云的架构下,尽管云商有算力的绝对控制和感知能力,但是调度目标受到了资源需求方的服务等级协定(service level agreement, SLA)、经济性等的约束,因此,调度问题则会转化为云商和需求方之间的博弈问题,从而目标函数 O 则转换为满足需求方 SLA 约束下最大化云商利润率。在运营商主导的算力网络架构体系下,目标函数 O 又会包含尽可能多地调度到指定的算力节点或者均衡化算力节点的资源利用率等。同样的情况,也会出现在 R 和 f 的设计上。

因此,在考虑问题的解决方案时,首先需要明确对算力结构、算力的管控权属等问题进行明确。

(2) 目标及约束

在支撑数字化业务的过程中,调度算法会根据参与主体对调度结果的评估维度,选择合适的资源。但在实际的场景中,不同参与方的评估维

度存在较大差异,甚至相悖,如在一个开放的算力供需市场下,算力的需求方期望的是以最低的算力成本得到最好的服务,而算力的供给方则是满足业务 SLA 的同时收益最高。

因此,第 2 个需要明确的是:满足何种约束条件,如何评估调度方案的优劣。

(3) 建模及求解

云边协同中的资源调度优化是一个比较复杂的问题。在算力节点规模较小时,简单模型即能在可接受的时间内得到可接受的解,但是在诸如算力网络这样的场景下,节点规模大、节点类型多、资源约束维度高,优化方案的求解相对不容易,已有多项学术研究^[16-20]证明了这是一个 NP 难问题。因此,需要对问题进行抽象和建模,并借助数学手段快速找到解。目前,针对这一问题,学术界的研究非常多,给产业界提供了很好的参考。文献[5-6]基于雾计算的场景,全面地综述和分析了当前的模型和求解方法,并对求解方法进行了分类。

3.2 典型场景的方案建议

基于前文的三步骤法,本节选择 AR/VR 业务以及企业云边协同数据治理业务作为云边协同的典型场景,讨论资源调度优化的实现架构和方案。

(1) AR/VR 业务

AR/VR 业务本质上是一种交互式的在线视频流,AR/VR 业务交互模式如图 1 所示,对边、云侧有较强的算力和低时延的网络需求。在公共的边/云侧算力完成渲染后,通过高速、高可靠的网络传送给用户,如手机、个人计算机、平板计算机、机顶盒等终端设备;用户通过输入设备(虚拟键盘、手柄等)对业务进行实时的操作。另外,AR/VR 业务在高铁、地铁等高速移动的场景下,用户侧终端设备将会在多个基站甚至多个地域进行网络切换,这样与初始连接的公共边、云侧节点网络时延增加,因而需要在多个公共的算力节点进行切换,根据用户的实际情况进行统一的调

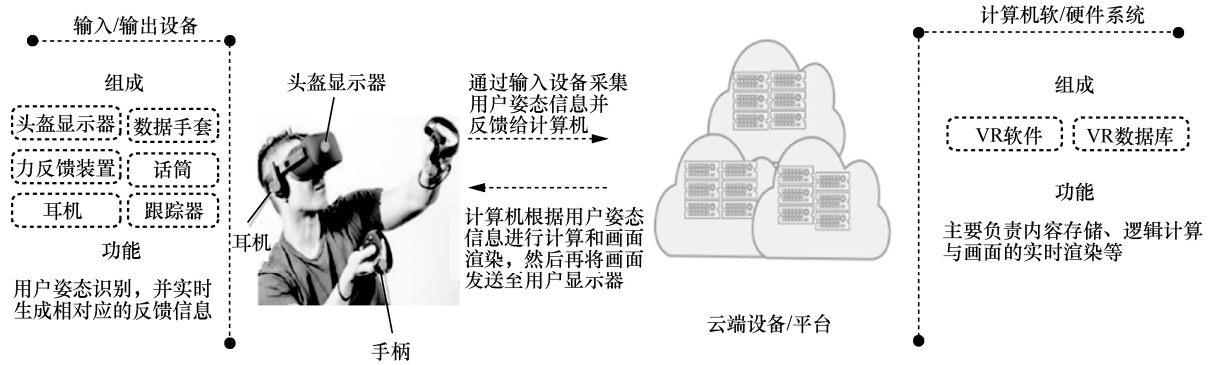


图1 AR/VR 业务交互模式

度和管理，将计算能力在多个节点之间无缝迁移，且保障用户无感的服务切换。

将 AR/VR 业务与上述的模型进行映射。首先，在场景方面，AR/VR 业务所需的算力是典型的多方供给模式。端侧的算力属于终端用户，由业务方进行调度和控制；考虑业务接入的随机性和高速需求，边侧的算力和网络资源大概率由运营商所控制；云侧的算力属于云商。因此，在 AR/VR 场景中的资源调度优化需要将服务商和资源方的权责进行解耦。在基础设施资源层，由运营商依照基础设施的布局，为业务选择合适的算力资源和网络资源的组合；在基础设施服务层，云商在自身的资源约束下，依据调度优化目标的指引，为业务服务选择合适的资源；在业务服务层，业务服务方在应用层定义子业务和资源的映射。

目标函数 O 的设定也是多维度的。在基础设施资源层，调度的目标是优化资源的使用效率，包括资源的电量消耗、资源负载的均衡、资源供应方的偏好、业务的 SLA 等。在基础设施服务层，服务商的考虑包括资源类型、服务优先级、集群负载等。在业务服务层，考虑的是业务的体验质量（quality of experience, QoE）。文献[19]是在 Kubernetes 调度机制上的改进，在云算力的基础上增加了边缘算力相关的时延、数据传输等指标，以适应云边协同中存在的镜像同步、数据同步等场景。

按照第 2 节对于资源问题的定义，可将 AR/VR 业务建模如下。

$SP=\{SP_1, SP_2, SP_3\}$ 为服务提供商，分别代表终端应用开发商、边缘算力及服务提供商、云端算力及服务提供商。

$I=\{I_1, I_2, \dots, I_n\}$ 为所有的算力节点集合，其中，网络抽象为一类特殊的节点。

因此，基于节点和服务供应商的从属关系可定义为：

$$z_i^s = \begin{cases} 0, & I_i \notin SP_s \\ 1, & I_i \in SP_s \end{cases} \quad (1)$$

在 t 时刻，针对用户 j 发出的任务 k ，其资源调度结果为：

$$f=\{a_{1j}^k(t), a_{2j}^k(t), \dots, a_{nj}^k(t)\} \quad (2)$$

其中， $a_{ij}^k(t)$ 为节点 i 分配给用户 j 用来处理任务 k 的资源集合，并且满足：

$$\begin{cases} \sum_i a_{ij}^k(t) = R_{a_j}^k \\ \forall i, \text{满足} \sum_k \sum_j a_{ij}^k(t) \leq R_{I_i} \end{cases} \quad (3)$$

如此，用户 j 针对任务 k 付出的代价定义为：

$$\text{cost}_j^k(t) = \sum_i v_{ij}^k \pi_i(t) \cdot a_{ij}^k(t) \quad (4)$$

其中， $v_{ij}^k = \begin{cases} 0, & \text{用户 } j \text{ 的第 } k \text{ 个任务未分配至节点 } i \text{ 上} \\ 1, & \text{用户 } j \text{ 的第 } k \text{ 个任务分配至节点 } i \text{ 上} \end{cases}$ ，



$\pi_i(t)$ 表示 t 时刻节点 i 上的资源/服务报价节点。

服务提供商 s 在一个计费周期 T 内得到的效用为:

$$\text{Utility}_s = \sum_{t \in T} \sum_i \sum_k \sum_j z_i^s v_{ij}^k [\pi_i(t) * a_{ij}^k(t) - c_i(t, a_{ij}^k(t))] \quad (5)$$

其中, $c_i(t, a_{ij}^k(t))$ 为节点提供资源所需要付出的成本。

目标函数可定义为:

$$\begin{aligned} \min \text{cost}_j^k(t) &= \sum_i v_{ij}^k \pi_i(t) * a_{ij}^k(t) \\ \max \text{Utility}_s \end{aligned} \quad (6)$$

上述模型从经济效能的角度,给出了调度方案的评估模型示例。当然也可根据实际系统的设计重新设计。文献[16]提出的 Zenith 就是另一项值得参考的研究工作。Zenith 在服务管理、边缘资源管理解耦的前提条件下,提出了一种基于市场动态定价的机制的边缘算力调度和优化方法提升边缘数据中心 (data center, DC) 使用效率的同时提升了资源服务提供方的利润率。基本过程是: ①资源需求方 (App 层) 根据自身业务特点,规划好对于边缘 DC 的需求; ②根据地理分布特点,引入维诺图,并基于加权维诺图的理论为资源需求方选择位置临近的边缘 DC; ③边缘 DC 算力的多个算力服务提供方依据自身的成本和利润,向资源需求方报价; 需求方依据该 DC 的历史服务质量等级给出报价,并选择合适的算力供应方; ④双方就报价形成共识,则需求方的任务调度到边缘 DC 上。

支持 AR/VR 业务时,该方案值得参考的包括: 服务管理和边缘资源管理解耦的前提与实际情况较为符合; 在需求方和供给方之间,以市场化的手段,基于算力的历史服务水平指导供需双方合理定义量化的算力价格,对算力的供给、定价模式有较强的参考价值; Zenith 通过市场机制,桥接了资源型调度和服务型调度的差异。同时,尚有几点需要进一步细化: ①选择边缘 DC。文中

未细化描述边缘 DC 与业务接入点的映射。在业务实施中,可考虑接入点位置、边缘 DC 的网络接入带宽等因素; 亚信科技提出了算力网络中的一种资源调度方法,为云边协同业务选择合适的 DC; 文献[17]针对主播直播场景提出了一种边缘接入节点的选择方案; ②边缘算力提供报价时,如何评估资源自身的成本。在文献的基础上,可以考虑增加数据传输代价、镜像传输代价。

(2) 企业数据治理业务

当数据定义为企业的数字化资产后,通过数据治理逐步实现数据价值成为企业转型的关键环节。在面对企业数据治理业务的“海量数据”“即时交互”“稳定安全”的要求时,传统集中模式的云计算已不再是最优方案,云边端协同成为产业界公认的替代方案。在工业生产中,端侧部署现场的数采能力对接生产现场的各种异构数据源; 边缘侧以“服务器+虚拟化”的方式部署的服务包括业务数据处理能力、数据的建模和可视化能力、数据开放的三大能力以及本地应用的管理; 云侧采用“1+N”的集团——工厂的集约式管理,多层级联,一次定义多点分发的模式,提升企业管理效率和标准化程度。

企业自建云或者企业自建边缘云+公有云是支撑数据治理业务的典型算力结构。此方式下的场景建模可认为是 AR/VR 业务的简化。首先是服务提供商方面的简化,服务提供商收缩为企业、网络供应商和公有云服务商; 其次,是各类业务参与者的效用函数设计。在企业自建云的场景下,可设计为:

$$\begin{aligned} \max \sum_j \sum_k u_j^k, \\ u_j^k = \begin{cases} 0, & \text{用户}j\text{的第}k\text{个任务未能成功调度} \\ 1, & \text{用户}j\text{的第}k\text{个任务被成功调度} \end{cases} \end{aligned} \quad (7)$$

该类问题的求解还可参考文献[18]的相关工作。文献[18]基于有限的资源池,聚焦考虑数据密集型业务场景,提出了一种边缘服务部署以及服务

请求调度的方案。该方案值得参考的包括：①方案基于的前提和假设与企业自建云情形较为吻合，包括资源有限、调度目标设计、数据密集型、调度方案动态调整等；②方案采用启发式的方法求解问题，以有限的代价得到可接受的方案，便于工程实施。尚需进一步探讨的是该方案在当前的主流开发框架中的软件实现。目前，Kubernetes是云计算的主流工程框架，并且在随着云边协同的需求迭代和演进，出现了KuberEdge、SuperEdge等云边协同的开源框架，得到了产业界的关注。但是，主要集中在云边服务的管理和协同，对于云边协同的资源调度优化较少涉及，相关的工作可以参考文献[19]。

在企业自建边缘云+公有云的算力结构下，资源调度问题退化为结合网络代价选择云端算力，此种情况可参考文献[16]。在该工作中引入云代理的角色处理云端算力的选择问题，对云管平台相关厂商有较大的参考价值。基于不同地域、不同云服务商的算力价格不同的现实情况，该文献以算力需求方的目标为导向，综合考虑虚拟机镜像启动时延和迁移代价，构建资源调度优化模型，并采用启发式的方法求解问题。

4 结束语

云边协同的计算模式由原来的“中心-用户”的单边模式转变为“用户-中心”“用户-用户”的多边模式。伴随这种转变，在新的商业模式、技术架构调整的引领下，云边系统中资源调度优化问题需要产业界、学术界共同攻克。在中国，算力网络的概念引发了算力资源和网络资源联合调度的研究热潮。但是，纵观国内外，国际产业界鲜有相关的工程实施参考，学术界的研究成果对于工程实施的指导意义又有一定局限性。因此，本文结合产业界的典型场景，试图筛选出对工程实施具有较大参考价值的学术成果，以期为未来的云边协同架构的推进尽一份力量。

从上述云边协同的典型场景分析中可知，为了应对算力权属的多样性，资源调度和优化需要划分层次。在涉及多方权属问题时，采用资源分配型调度方式，以经济性换取可用性；在独立权属的资源池内，可采用服务型调度方式，满足用户对于灵活、动态、弹性等高级特性的需求。因此，面向未来的云网协同场景，资源调度优化针对调度的维度会划分成不同的阶段，并且这种多阶段的调度方式将会持续较长的一段时间。

未来的工作将主要聚焦在泛在算力的纳管、多级算力结构下的调度协同、调度机制在工程项目中的软件实现。

参考文献：

- [1] ETSI. Multi-access edge computing (MEC)[EB]. 2022.
- [2] HU P F, DHELM S, NING H S, et al. Survey on fog computing: architecture, key technologies, applications and open issues[J]. Journal of Network and Computer Applications, 2017(98): 27-42.
- [3] MARTIN A. Distributed computing: utilities, grids & clouds ITU-T technology watch report 2009[R]. 2009.
- [4] ITU-T. Y.2501: computing power network-framework and architecture[S]. 2019.
- [5] 王凌, 吴楚格, 范文慧. 边缘计算资源分配与任务调度优化综述[J]. 系统仿真学报, 2021, 33(3): 509-520.
WANG L, WU C G, FAN W H. A survey of edge computing resource allocation and task scheduling optimization[J]. Journal of System Simulation, 2021, 33(3): 509-520.
- [6] JAMIL B, IJAZ H, SHOJAFAR M, et al. Resource allocation and task scheduling in fog computing and Internet of everything environments: a taxonomy, review, and future directions[J]. ACM Computing Surveys, 2022, 54(11s): 1-38.
- [7] IBRAHIM E, EL-BAHNASAWY N A, OMARA F A. Task scheduling algorithm in cloud computing environment based on cloud pricing models[C]//Proceedings of 2016 World Symposium on Computer Applications & Research (WSCAR). Piscataway: IEEE Press, 2016: 65-71.
- [8] ABDULLAHI C, GOUR K, JOARDER K. The co-evolution of cloud and IoT applications: recent and future trends[R]. 2019.
- [9] BENBLIDIA M A, BRIK B, MERGHEM-BOULAHIA L, et al. Ranking fog nodes for tasks scheduling in fog-cloud environments: a fuzzy logic approach[C]//Proceedings of 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). Piscataway: IEEE Press, 2019:



- 1451-1457.
- [10] ABDELMONEEM R M, BENSLIMANE A, SHAABAN E. Mobility-aware task scheduling in cloud-Fog IoT-based healthcare architectures[J]. Computer Networks, 2020(179): 107348.
- [11] NI L N, ZHANG J Q, JIANG C J, et al. Resource allocation strategy in fog computing based on priced timed petri nets[J]. IEEE Internet of Things Journal, 2017, 4(5): 1216-1228.
- [12] ZHAO X Y, ZONG Q, TIAN B L, et al. Fast task allocation for heterogeneous unmanned aerial vehicles through reinforcement learning[J]. Aerospace Science and Technology, 2019(92): 588-594.
- [13] Gartner. Gartner trends 2021: what they mean for retailers[R]. 2020.
- [14] Cloud Networking. The 2020 gartner magic quadrant for data center and cloud networking[R]. 2019.
- [15] 中国电信. 云网融合 2030 技术白皮书[R]. 2020. China Telecom. Computing and network convergence technical white paper[R]. 2020
- [16] RAUSCH T, RASHED A, DUSTDAR S. Optimized container scheduling for data-intensive server less edge computing[J]. Future Generation Computer Systems, 2021 (114): 259-271.
- [17] XU J L, PALANISAMY B, LUDWIG H, et al. Zenith: utility-aware resource allocation for edge computing[C]//Proceedings of 2017 IEEE International Conference on Edge Computing (EDGE). Piscataway: IEEE Press, 2017: 47-54.
- [18] CHEN J S, BALASUBRAMANIAN B, HUANG Z. Liv(e)-ing on the edge: user-uploaded live streams driven by “first-Mile” edge decisions[C]//Proceedings of 2019 IEEE International Conference on Edge Computing (EDGE). Piscataway: IEEE Press, 2019: 41-50.
- [19] FARHADI V, MEHMETI F, HE T, et al. Service placement and request scheduling for data-intensive applications in edge clouds[J]. IEEE/ACM Transactions on Networking, 2021, 29(2): 779-792.
- [20] ADDYA S K, SATPATHY A, GHOSH B C, et al. CoMCLoud: virtual machine coalition for multi-tier applications over

multi-cloud environments[J]. IEEE Transactions on Cloud Computing, 2021(99): 1.

[作者简介]



王淑玲(1988-), 女, 博士, 亚信科技(中国)有限公司研发中心规划部规划总监, 主要研究方向为网络通信、云网融合。



孙杰(1983-), 男, 亚信科技(中国)有限公司研发中心云网规划部经理, 主要研究方向为通信与 5G 网络智能化。



王鹏(1976-), 男, 亚信科技(中国)有限公司研发中心高级总监, 主要研究方向为通信业务支撑、大数据和人工智能。



杨爱东(1984-), 男, 博士, 亚信科技(中国)有限公司通信人工智能实验室首席数据科学家, 主要研究方向为 5G 无线通信、大数据挖掘、机器学习及其应用。