

5G 边缘计算环境下资源效用与隐私保护 权衡的服务迁移方法

李邦源¹, 张春辉¹, 常 荣¹, 陈 君¹, 许小龙²

(1. 云南电网有限责任公司玉溪供电局, 云南 玉溪 653100)

(2. 南京信息工程大学计算机与软件学院, 江苏 南京 210044)

[摘要] 针对当前 5G 环境下边缘计算场景存在硬件资源浪费以及服务隐私泄露的问题, 本文设计了一种基于资源效用和隐私保护权衡的服务迁移方法 (A Service Offloading Method, SOM) 以实现边缘节点资源的高效利用与用户的隐私保护. 该方法首先对边缘节点在服务迁移场景中资源利用和隐私保护的需求进行了分析与建模, 并将其定义为一个多目标优化问题. 然后利用改进的强度帕累托进化算法 (Improving the Strength Pareto Evolutionary Algorithm, SPEA2) 生成问题的可行解, 采用简单加权法 (Simple Additive Weighting, SAW) 和多准则决策法 (Multi-Criteria Decision-Making Method, MCDM) 从多个可行解中挑选出全局最优解. 最后对提出的服务迁移方法 SOM 进行实验对比并证明其有效性.

[关键词] 5G, 服务迁移, 边缘计算, 负载均衡, 隐私保护

[中图分类号] G301 **[文献标志码]** A **[文章编号]** 1001-4616(2021)04-0102-09

A Service Offloading Method for Resource Utilization and Privacy Preservation Trade-offs in 5G-Enabled Edge Computing

Li Bangyuan¹, Zhang Chunhui¹, Chang Rong¹, Chen Jun¹, Xu Xiaolong²

(1. Yunnan Power Grid Corporation, Yuxi Power Supply Bureau, Yuxi 653100, China)

(2. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In view of the problems of hardware resource waste and service privacy leakage in the current 5G environment, a service offloading method (SOM) for resource utilization and privacy preservation trade-offs has been designed to achieve efficient utilization of edge nodes and protection of user privacy. Specifically, resource utilization and privacy protection requirements of edge nodes in service offloading scenario have been firstly analyzed and modeled, and then defined as a multi-objective optimization problem. Afterward, the feasible solutions of the problem are generated by utilizing the improving the strength pareto evolutionary algorithm (SPEA2). Furthermore, simple additive weighting (SAW) and multi-criteria decision-making method (MCDM) are deployed to select the global optimal solution from multiple feasible solutions. Finally, the experimental results show the effectiveness of the proposed SOM.

Key words: 5G, service offloading, edge computing, load balance, privacy preservation

近年来移动设备普及率不断提高, 用户对移动应用程序的需求量也不断增加, 这使得移动数据流量开始呈指数型增长^[1-2]. 根据思科网络视觉指数 (Cisco Network Vision Index) 报告, 预计截止 2020 年底数据流量需求的增速将达到 57%, 数据规模将是 2014 年的 10 倍. 而第五代无线通讯系统 (fifth-generation wireless systems, 5G) 的广泛应用可以在一定程度上缓解数据流量的负担. 5G 网络的架构包括分布式单元 (distributed unit, DU) 和中心单元 (central unit, CU), 但是这两个单元无法直接处理移动应用的计算任务^[3-4]. 边缘计算是一种将传统云计算下沉到网络边缘以提供高效服务的新型计算范式. 在边缘计算的辅助下, 5G 网络的处理能力得到了补充与增强^[5-6]. 考虑到 DU 的广泛使用性, 将 CU 增强为边缘节点 (Edge

收稿日期: 2020-06-29.

基金项目: 国家自然科学基金项目 (61702277).

通讯作者: 许小龙, 博士, 教授, 研究方向: 边缘计算, 服务计算. E-mail: xlxu@nuist.edu.cn

Node, EN) 更适合于协助 DU 处理数据^[7]. 其中, 时间敏感型的边缘服务可以迁移到 EN 上进行处理, 而非敏感型服务既可以按需迁移到 EN 处理, 也可以迁移到远程云数据中心执行^[8]. 在服务迁移中需要使用大量的计算资源, 而资源利用率和负载均衡是衡量服务迁移方案的重要指标. 其中, 资源利用率反映了边缘节点对计算资源的使用情况, 负载均衡则反映了网络总体的吞吐量、处理能力等^[9-10]. 在上述的处理方式下如何合理利用边缘节点的计算资源并提高资源利用率是本文的一个主要研究点.

此外, 在服务的迁移过程中存在隐私信息泄露问题. 当执行用户目标服务(包括远程视频、语音聊天、实时定位等)时, 服务所需的隐私信息将随着服务本身一起被迁移到边缘节点上. 如果这些隐私信息被不法分子窃取并破解, 这将严重损害用户的利益^[9-10]. 因此, 在 5G 边缘环境下执行服务迁移时防止服务数据发生隐私泄露是本文另一个主要研究点.

针对现存的移动应用服务迁移问题, 国内外学术界以及工业界都展开了深入研究. 在服务从移动设备迁移到 EN 或远程云端的过程中, 如何在保证网络高性能的同时保护隐私信息是一个亟待解决的重要问题. Qian 等^[11]设计出一种基于隐私感知服务的放置方案用以在服务放置时保护用户的隐私. Lu 等^[12]设计了一种雾计算下用于增强物联网的隐私保护服务聚合方案. Du 等^[13]使用差分隐私算法设计了一种机器学习策略. 该策略包含输出扰动算法和目标扰动算法, 可以有效保证隐私信息的安全性. Ding 等^[14]为文档分类制定了一套体系结构用来增强安全性. 该体系结构可以帮助处理和分析边缘计算中数据资源周围的原始数据流. 而针对隐私泄露问题, He 等^[15]提出了一种将隐私感知任务卸载算法与约束马尔可夫决策过程相结合的方法.

将边缘任务迁移到 5G 边缘节点以优化网络的整体性能同样有着广泛研究. Mao 等^[16]提出了一种带有能量采集装置的移动边缘计算系统, 并引入了一种高效的任务卸载策略. Chen 等在^[17]将计算卸载决策问题定义为一个计算卸载博弈问题以更好地优化服务迁移过程中的通讯延迟. Tao 等^[18]研究了如何在满足所有用户的服务体验的同时实现网络中心的节能以降低能耗. Tian 等^[19]提出了一种集成云、边缘云和物联网设备的新架构来解决网络的可扩展性问题. 为了解决密集交通环境下的任务计算和迁移问题, Zhang 等^[20]提出了一种基于边缘云的车辆计算卸载体系结构.

此外, 服务迁移的效率与自适应机制也是值得考虑的重要问题. Wang 等^[21]提出了一种改进边缘计算性能的创新框架, 以及一种优化的资源卸载方案, 该方案的目的是使接入节点的整体能量消耗最小化. Wang 等^[22]将服务迁移场景中的能耗最小化问题转化为凸问题, 并对非凸和非光滑的延迟最小化问题提出了单变量搜索局部最优算法, 以获得最优结果. Chen 等^[23]将服务迁移问题归纳为 NP-hard (non-deterministic polynomial, 非确定性多项式) 问题, 并设计了一个有效的解决任务分配和资源分配子问题的方案. Yu 等^[24-25]提出了一种基于缓存改进的加载模型, 该模型通过在资源共享场景中移动终端节点的联合加载策略来最小化总体实施延迟.

随着相关研究的不断推进, 服务迁移场景中存在的问题在一定程度上得到解决, 但是很少有工作考虑在进行服务迁移时, 边缘设备如何在存在隐私保护约束的条件下进行资源利用率与负载均衡方差指标的联合优化. 因此, 本文针对性地提出了一种边缘计算环境下资源效用与隐私保护权衡的服务迁移方法, 在隐私保护的约束下, 能够保证边缘设备的资源利用率和负载状态保持理想水平.

1 系统模型

1.1 资源模型

本文旨在对边缘用户的服务请求进行迁移, 并进行合理的分配以实现边缘节点整体的高效利用与隐私保护. 本文提出的服务迁移框架如图 1 所示, 主要由 3 部分组成, 底层是由产生计算服务的各种实际场景与分布在场景周围的分布式单元构成; 中间层是由边缘计算节点与接受服务传输的基站构成, 这些边缘节点只处理所属基站覆盖范围内的服务请求; 上层由远端云数据中心构成.

在迁移框架中, 分布节点周围存在大量不断产生服务请求的设备, 这些设备产生的请求任务首先由分布式单元接收. 再由分布式单元传输到最近的 EN. 然后, 会对是否在当前节点处理任务进行判断. 此判断由整体的资源利用率与负载均衡方差决定. 此外, 针对传输任务到远端 EN 的情况, 该框架引入了时间约束. 在特殊情况下, 任务还会由 EN 直接传输到远端的云数据中心进行处理以满足框架的整体要求.

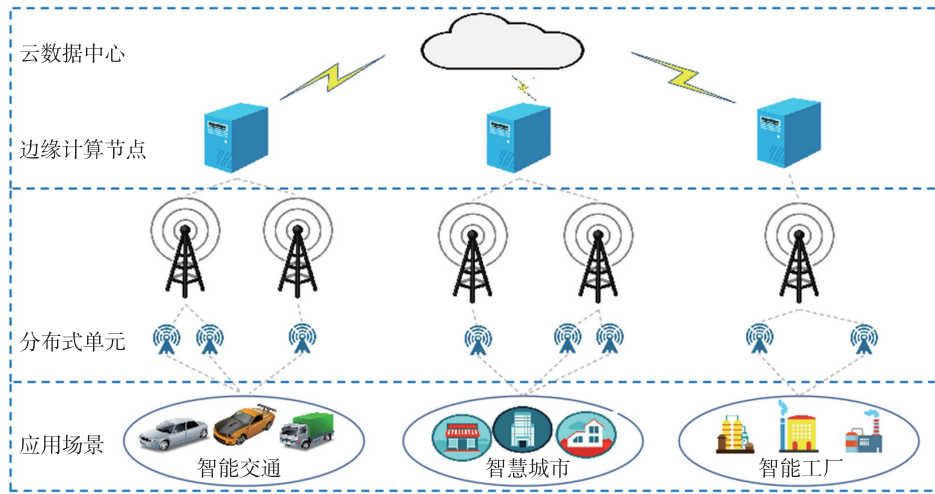


图 1 边缘计算环境下的服务迁移框架

Fig. 1 A service offloading framework in edge computing

在如图 1 所示的边缘计算任务迁移框架中,将 $AC = \{ac_1, ac_2, ac_3, \dots, ac_g\} (1 \leq g \leq G)$ 定义为 CU 集合,其中 G 代表 CU 的数量. CU 被改造为边缘节点以缓冲处理卸载任务所带来的负担. 得益于 EN 中使用的虚拟化技术,虚拟机(virtual machine, VM)作为计算单元被使用. 服务卸载场景中部署了大量的 DU 和 EN,它们的集合分别被定义为 $AD = \{ad_1, ad_2, ad_3, \dots, ad_f\} (1 \leq f \leq F)$ 和 $SE = \{se_1, se_2, se_3, \dots, se_g\}$. 移动设备的数量被预先设置为 E ,移动设备集合被定义为 $AM = \{am_1, am_2, am_3, \dots, am_e\} (1 \leq e \leq E)$. 相应地,被卸载的计算服务数量也被设置为 E ,该计算任务集合被定义为 $AP = \{ap_1, ap_2, ap_3, \dots, ap_e\}$.

1.2 平均资源利用率模型

在本模型中,资源利用率代表各个 EN 的使用程度. 而平均资源利用率则是重要的模型判断指标,其计算过程如下:

首先,使用变量 α_g 判断 se_g 是否已经被占用,其公式定义如下:

$$\alpha_g = \begin{cases} 1, & se_g \text{ 被占用,} \\ 0, & se_g \text{ 未被占用.} \end{cases} \quad (1)$$

然后,使用二进制变量 β_e^g 判断数据集 $ap_e (1 \leq e \leq E)$ 是否放置在 se_g 上,其公式定义如下:

$$\beta_e^g = \begin{cases} 1, & ap_e \text{ 被放置在 } se_g, \\ 0, & ap_e \text{ 未被放置在 } se_g. \end{cases} \quad (2)$$

本模型根据已使用节点中虚拟机的数量来计算 EN 的平均资源利用率. 已使用的边缘节点数量(记作 MS)的计算公式定义如下:

$$MS = \sum_{g=1}^G \beta_e^g. \quad (3)$$

进一步地,使用 UE_g 表示 VM 的资源利用率,其公式定义如下:

$$UE_g = \frac{1}{\chi_g} \cdot \sum_{e=1}^E \delta_e \alpha_e, \quad (4)$$

式中, χ_g 表示第 g 个 EN 中的 VM 数量, δ_e 表示数据集 ap_e 中所使用的 VM 数量.

最后,所有处于使用状态的 EN 的平均资源利用率被表示为 AVE,其公式定义如下:

$$AVE = \frac{1}{MS} \cdot \sum_{g=1}^G UE_g. \quad (5)$$

1.3 负载均衡模型

负载均衡方差是重要的评价标准,用于表示 EN 之间使用程度的差异,其计算需要基于资源利用率. se_g 的负载分布表示为 LB_g ,其公式定义如下:

$$LB_g = (AVE - UE_g)^2. \quad (6)$$

基于上述分析,所有已使用的 EN 的平均负载均衡方差,记作 AL ,其公式定义如下:

$$AL = \frac{1}{MS} \cdot \sum_{g=1}^G LB_g \beta_e^g. \quad (7)$$

1.4 隐私保护模型

隐私保护是所有网络中不可忽视的一点. 在本场景中,从各种移动设备卸载的边缘服务请求需要相应的数据集来处理. 然而,不同的数据集会存在不同程度的隐私冲突. 为了解决这一问题,这些存在隐私冲突的边缘服务需要不同的 EN 来执行处理.

定义 $\varphi=(S,C)$ 用以描述边缘服务之间的隐私冲突,其中 S 表示为边缘服务集, C 表示为冲突关系的集合. 为了保证边缘服务中隐私信息的安全,定义冲突关系 $(b_i, b_{i'})$ ($b_i, b_{i'} \in S$) 表示两个有冲突的任务不能放在同一个节点上执行. b_g 的冲突边缘服务公式定义如下:

$$cb_i = \{b_{i'} | (b_i, b_{i'}) \in C, i' \in \{1, 2, 3, \dots, I\}\}. \quad (8)$$

边缘服务的计算卸载策略集被表示为 $OS = \{os_1, os_2, os_g, \dots, os_G\}$ ($os_g \in G$), 其中 os_g 表示要处理的 b_g 的目标 EN. 根据获取的一组需要处理的冲突服务, b_g 存在冲突的 EN 集合,其公式定义如下:

$$cd_i = \{cs_j | cs_j \in cb_i, j \in \{1, 2, 3, \dots, |cb_i|\}\}. \quad (9)$$

1.5 时间消耗模型

如果最先卸载的 EN 中存在服务作业,则需要将边缘服务卸载空闲节点进行处理. 这一过程将带来额外时间消耗,需要将其考虑到整体的时间消耗模型中.

首先,使用 VF_e 判断 ap_e 是否被卸载到 se_g 上或者云上,其公式定义如下:

$$VF_e = \begin{cases} 1, & ap_e \text{ 被迁移,} \\ 0, & ap_e \text{ 未被迁移.} \end{cases} \quad (10)$$

当边缘服务 ap_e 被卸载到空闲节点上,其虚拟机与接入点之间的时间消耗,记作 TT ,其公式定义如下:

$$TT = \sum_{e=1}^E \sum_{g=1}^G \beta_e^g VF_e \frac{MS_e}{\mu}, \quad (11)$$

式中, MS_e 代表第 e 个服务 se_g 的任务大小, μ 表示 VM 与 AP 之间的传输效率.

其次,使用 BE 表示服务卸载中的 EN 数量. EN_s 和 EN_F 分别表示迁移过程中最初到达的 EN 与最后处理服务的 EN,它们之间的迁移时间被表示为 PT ,其公式定义如下:

$$PT = \sum_{e=1}^E \sum_{g=1}^G \beta_e^g VF_e \frac{MS_e}{v} (BE-1), \quad (12)$$

式中, v 代表 EN 之间的传输效率.

如果边缘服务从 EN 迁移到远程云,则从 EN 传递到云上所需的时间消耗被表示为 CT ,其公式定义如下:

$$CT = \sum_{e=1}^E \sum_{g=1}^G \beta_e^g VF_e \frac{MS_e}{\sigma}, \quad (13)$$

式中, σ 表示 EN 和远程云之间的传输效率.

定义二进制变量 λ 判断 ap_e 被迁移到哪个目的地,其公式定义如下:

$$\lambda = \begin{cases} 1, & ap_e \text{ 迁移到 } EN_F \text{ 中,} \\ 0, & ap_e \text{ 迁移到云中.} \end{cases} \quad (14)$$

基于上述时间消耗的计算,最后使用 OT 表示总体时间消耗,其公式定义如下:

$$OT = \begin{cases} \sum_{g=1}^G 2TT + PT, & \lambda = 1, \\ \sum_{g=1}^G 2TT + CT, & \lambda = 0. \end{cases} \quad (15)$$

1.6 目标函数

在上述分析的基础上,本节对各边缘节点的资源利用率和负载均衡方差进行了建模和量化.如式(5)和(7)所示,资源利用率和负载均衡方差是本文两个主要优化目标.标准的多目标优化模型表示为:

$$\max AVE, \quad (16)$$

$$\min AL. \quad (17)$$

$$\text{s.t. } cb_i \notin cd_i, \quad (18)$$

$$OT \leq LT. \quad (19)$$

式(18)表明需要考虑隐私约束以确保边缘服务中的隐私信息.式(19)表明整个时间消耗根据传输效率、不同 EN 之间的距离和边缘服务的大小计算,将其限制为 LT .

2 本文方法

本文设计了一种基于负载均衡和隐私保护联合优化的服务迁移方法(SOM),首先基于改进的强度帕累托进化算法(SPEA2),该方法在处理多目标问题时具有并行处理、全局最优和鲁棒良好等优点,且能精准地对边缘节点的资源使用率和负载均衡进行优化并得到多组 Pareto 均衡解;然后利用简单加权法(simple additive weighting,SAW)和多准则决策法(multi-criteria decision-making method,MCDM),从 SPEA2 计算所得的均衡解集合中获取到最优任务迁移策略.

2.1 基于 SPEA2 的服务迁移方法

为了使服务迁移策略在考虑隐私保护与执行时间约束的目标下实现资源利用率与负载均衡的联合优化,SOM 在求解服务迁移策略时利用 SPEA2 算法来解决该多目标优化问题. SPEA2 在解决多目标问题时具有更加高效的全局搜索能力,且具有更低的时间和空间复杂度. SPEA2 迭代求解服务迁移策略的步骤如下所示:

步骤 1:初始化(Initialization):在遗传算法中,需要设置种群大小 Z 、交叉概率 CP 、变异概率 MP 以及迭代次数 MM ,不适合的值会对最后结果的优越程度产生影响.

步骤 2:编码(Encoding):遗传算法的首要问题是将迁移策略编码到由基因组成的染色体中.染色体中的每个基因代表了每个边缘服务的迁移策略.基于对服务迁移的研究与分析,本文将染色体以整数数组的形式进行编码.经过算法迭代得到的计算任务迁移策略被定义为 $FS = \{FS_1, FS_2, \dots, FS_Z\}$,其中 Z 是设置的种群大小,即策略的个数.

步骤 3:选择(Selection):选择操作是一个比较的过程.从目前的策略中选出两个进行锦标赛选择,选择其中更优越的策略放入交配池中,而交配池中的较为优秀的个体将继续进行后续操作,从而加速迁移策略的生成.

步骤 4:交叉和变异(Crossover and Mutation):在交叉过程中采用单点交叉方法,以概率 CP 随机选择交叉点,围绕这个交叉点,两个父染色体进行交换,从而产生两个新的子染色体.当子染色体的表现不再优于父染色体但依然没有达到全局最优解时,就会发生过早收敛.此时就会以 MP 的概率进行突变操作,从而产生种群个体即新的任务迁移策略,其目的是加速策略生成同时保持策略的多样性.

步骤 5:适应性函数和约束条件(Adaptive Functions and Constraints):遗传算法选择基于适应度函数评价每个个体的优劣程度.在本优化问题中,两个对象组成了适应度函数:式(5)的资源利用率和式(7)的负载均衡方差.如何在最大化资源利用率的同时最小化负载均衡方差是本文的目标问题.此外,由于在边缘服务传输过程中,隐私保护和传输时间的约束也需要满足,因此定义式(18)和(19)实现隐私保护和时间约束.

经过执行此方法,一个服务迁移任务可能多产生多个可行的迁移策略.在本文中,上述迁移方法将产生 Z 个策略,这些策略都是可以满足约束(18)和(19)的可行解,因此,需要从这些可行解中选取一个最优解来充当最终执行策略.本文拟采用 SAW 和 MCDM 方法来获取该解.

2.2 基于 SAW 和 MCDM 的最优策略选择

在本文中,资源利用率为正准则,负载均衡为负准则.根据模型部分计算出资源利用率和负载均衡方差.正准则资源利用率的有效性随着其值的增大而提高,负准则负载均衡的有效性随着其值的减少而降

低. 假定最终经过迭代后得到的 Z 个迁移的策略集合为 FS , 每一个迁移策略所对应的 $UT_z (1 \leq z \leq Z)$ 代表第 z 个策略的资源使用率, BA_z 代表第 z 个策略的负载均衡. 正准则资源利用率的公式定义如下:

$$VL_{z,UT} = \begin{cases} \frac{UT_z - UT_{\min}}{UT_{\max} - UT_{\min}}, & UT_{\max} - UT_{\min} \neq 0, \\ 1, & UT_{\max} - UT_{\min} = 0, \end{cases} \quad (20)$$

式中, UT_{\max} 代表资源利用率的最大值, UT_{\min} 代表最小值.

负准则负载均衡的公式定义如下:

$$VL_{z,BA} = \begin{cases} \frac{BA_{\max} - BA_z}{BA_{\max} - BA_{\min}}, & BA_{\max} - BA_{\min} \neq 0, \\ 1, & BA_{\max} - BA_{\min} = 0, \end{cases} \quad (21)$$

式中, BA_{\max} 代表负载均衡方差的最大值, BA_{\min} 代表最小值.

基于式(20)和(21), 效用值公式定义如下:

$$EY_z = VL_{z,UT} \cdot \tau_1 + VL_{z,BA} \cdot \tau_2, \quad (22)$$

式中, τ_1 代表资源使用率的权重, τ_2 代表负载均衡方差的权重.

最后, MCDM 在所有迁移策略中选择具有最大效用值的策略作为最优策略.

2.3 算法概况

SOM 方法旨在满足隐私保护和时间约束的限制下, 优化资源利用率以及降低负载均衡方差. 而在多目标优化问题中, SPEA2 算法有着良好性能, 常用来解决此类问题. SOM 的流程如下: 首先, 所有被卸载边缘服务的节点都需要被编码. 然后, 基于选择、交叉和变异操作, 得到相应迁移策略. 进一步地, 给出该方法的适应度函数和约束条件. 最后, 使用 SAW 和 MCDM 方法获得迁移最优策略. 在算法 1 中, 伪代码描述了本文所提出的 SOM 方法. 其中, MM 代表最大迭代次数, ST 代表最优迁移策略.

算法 1 SOM 求解过程

```

输入  MM
输出  ST
Begin
初始化种群
mm = 1
for mm < MM do
    执行交叉和变异操作
    for 种群中的每一个染色体 do
        根据式(5)计算资源使用率
        根据式(7)计算负载均衡
    end for
    执行选择操作
    根据式(20)、(21)和(22)评估效用值
    mm = mm + 1
end for
Return 最优服务迁移策略 ST
End

```

3 实验结果与讨论

3.1 实验环境

本实验的计算服务器是联想 Y7000, 基本配置包括英特尔 i5-8300H 2.3GHz、8GB 内存和 512GB SSD. 表 1 列出了本实验所使用的 6 个基本参数. 为了确保分析准确与合理, 本实验设置了 5 种不同规模的边缘服务的数量来生成 5 种不同规模的数据集. 边缘服务的数量分别设置为 50、100、150、200 和 250, 这些规模被统一定义为 N .

表 1 参数设置

Table 1 Parameter settings

参数	取值	参数	取值
每个节点上运行 VM 的数量	[1, 6]	边缘服务的大小	[0.3, 0.6]
AP 和 ENs 之间传播速率	1 600 Mb/s	VM 中运行单元数量	7
APs 之间传输速率	800 Mb/s	边缘服务数量	{50, 100, 150, 200, 250}

3.2 对比方法

为了充分展示 SOM 的性能, 在本文实验中选择以下 3 种方法作为对比方法^[21], 具体描述如下:

Benchmark: 边缘服务被卸载到最近的边缘节点进行处理. 如果边缘服务需要更多的虚拟机才能执行, 则这些服务将被迁移到当前服务附近的边缘节点. 此过程重复执行, 直到迁移完所有边缘服务.

FFD: 根据任务所需的虚拟机数量, 按降序排列边缘服务. 然后, 将第一个边缘服务迁移到可以提供足

够多虚拟机的第一个边缘节点. 这个过程会一直执行下去,直到所有边缘服务都迁移完.

BFD:边缘服务是根据所需的虚拟机数量按降序排序. 然后,将第一个边缘服务迁移到拥有最少虚拟机数的边缘节点中,但仍然足够执行此边缘服务的节点上. 这个过程重复执行,直到所有的边缘服务都迁移完.

3.3 实验结果与分析

3.3.1 平均资源利用率对比

平均资源利用率是评价整个系统综合利用程度与算法实际表现的重要指标之一. 在本实验中,资源利用率是指各个边缘计算节点中硬件的利用程度,越低的资源使用率意味着越多的硬件处于闲置状态,将极大程度上地造成资源的浪费、影响服务的执行,本文的目标是尽可能地提高各个节点的资源利用率. 而在实际实验中,不可避免的情况是不同机器的资源利用率处于不同的水平,因此本实验采用的评价指标是平均资源利用率,保证整个系统所有资源的整体利用率保持在较高水准.

如图 2 所示,在 5 种边缘服务规模的对比中,可以直观地看出 SOM 在平均资源利用率指标上的表现均优于 Benchmark、FFD 和 BFD. 而表 2 展示了上述 3 种基准方法的对比,即 SOM 在资源利用率指标上性能的提高程度,可以看出,在任务较少时,SOM 方法在提高资源利用率性能优越程度上表现较好. 随着任务数量的增多,SOM 的性能提高程度降低,这也符合现实依据,在硬件资源一定的条件下,资源利用率逐渐趋近阈值,提高程度逐渐降低.

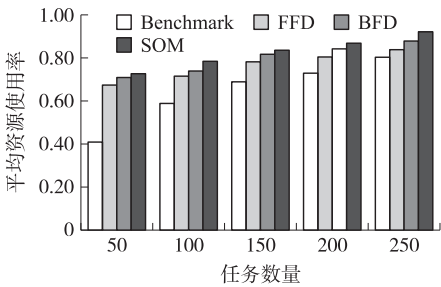


图 2 资源使用率对比图
Fig. 2 The chart of resource utilization comparison

表 2 资源利用率性能提高程度对比
Table 2 Comparison of resource utilization

任务数量	50	100	150	200	250
Benchmark	38.0%	33.9%	21.7%	19.2%	15.0%
FFD	7.3%	9.7%	7.7%	7.4%	9.5%
BFD	2.8%	6.8%	2.4%	3.6%	1.8%

3.3.2 负载均衡方差对比

负载均衡方差是作为评价整个系统灵活性和计算能力的重要指标. 传统的负载均衡机制按照指定的负载均衡算法分配任务到指定的边缘计算节点,当某些节点因不可抗力因素导致服务不可达时,该机制可以保证任务被合理分配,进而保证系统的健壮性. 当边缘节点之间的负载均衡得到保证时,就可以更大幅度地发挥出边缘节点的性能,并减少节点的整体能耗.

如图 3 所示,在 5 种边缘服务规模的对比中,可以直观地看出 SOM 在负载均衡方差指标上的表现均优于 Benchmark、FFD 和 BFD. 而表 3 展示了上述 3 种基准方法对比,即 SOM 在负载均衡方差指标上性能的提高程度. 可以看出在该指标上,SOM 方法的提高程度处于较高的水平,这表明 SOM 方法比 Benchmark、FD 和 BFD 更适用于对灵活系统有较高需求的场景,而服务迁移对灵活性正有着较高的需求,因此 SOM 的有效性得到了证明.

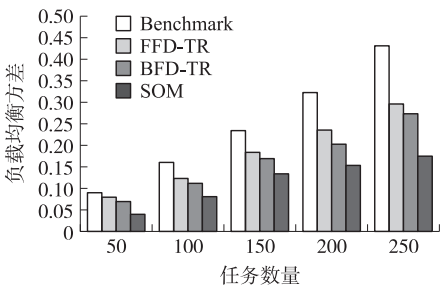


图 3 负载均衡方差对比图
Fig. 3 The chart of load balance variance comparison

表 3 负载均衡方差提高程度对比
Table 3 Comparison of load balancing

任务数量	50	100	150	200	250
Benchmark	55.6%	50.0%	43.4%	53.1%	60.5%
FFD	50.3%	33.3%	27.8%	34.8%	43.3%
BFD	42.8%	27.2%	23.5%	25.0%	37.0%

4 结论

在服务迁移的场景下,为了在满足隐私保护和时间限制约束的同时优化资源利用率和保持边缘节点的负载均衡,本文提出了一种基于资源效用和隐私保护的服务迁移方法. 首先分析边缘节点的资源利用和隐私保护现状,然后对问题进行建模,将其规约为约束条件下的双目标问题. 然后提出相应的服务迁移方法,利用 SPEA2、SAW 和 MCDM 等技术手段对问题进行求解,使得边缘用户的计算任务可以合理分配,并均衡部署在多个边缘节点中进行计算,实现边缘节点资源的高效利用. 最后,进行实验评估,验证了该方法的有效性.

将任务用户的不规则移动性考虑到服务迁移中并且扩展到现实世界是笔者未来的工作方向之一;用户具有复杂移动性的特征符合现实情况,该研究将具有更高的研究价值与更广的现实意义. 此外,鉴于当今深度学习技术不断推进与对人工智能技术需求不断提高的现实依据,将与深度学习有关的复杂异构任务考虑结合到该边缘计算的服务迁移场景下是笔者未来的另一研究方向;不断探究边缘智能场景中存在的问题与挑战,最终实现边缘智能的实际应用与落地实现.

[参考文献]

- [1] 赵梓铭,刘芳,蔡志平,等. 边缘计算:平台、应用与挑战[J]. 计算机研究与发展,2018,55(2):327-337.
- [2] IBRAHIM H,IBRA Y,BADRUL A,et al. The rise of “big data” on cloud computing:review and open research issues[J]. Information systems,2015,47:98-115.
- [3] MARIA P R,MISCHA D,ALFREDO G,et al. Internet of things in the 5g era:enablers,architecture,and business models[J]. IEEE journal on selected areas in communications,2016,34(3):510-527.
- [4] Akhil G,Rakesh J K. A survey of 5G network;architecture and emerging technologies[J]. IEEE Access,2015,3:1206-1232.
- [5] HE Q,CUI G M,ZHANG X Y,et al. A game-theoretical approach for user allocation in edge computing environment[J]. IEEE transactions on parallel and distributed systems(TPDS),2019,31(3):515-529.
- [6] GAO H H,MIAO H K,LIU L,et al. Automated quantitative verification for service-based system design:a visualization transform tool perspective[J]. International journal of software engineering and knowledge engineering,2018,28(10):1369-1397.
- [7] WANG Y C,HE Q,YE D Y,et al. Formulating criticality-based cost-effective fault tolerance strategies for multi-tenant service-based systems[J]. IEEE transactions on software engineering,2017,44(3):291-307.
- [8] 林俊宇,王慧强,马春光,等. 一种基于 DAG 动态重构的认知网络服务迁移方法[J]. 软件学报,2014,25(10):2373-2384.
- [9] 彭长根,丁红发,朱义杰,等. 隐私保护的信息熵模型及其度量方法[J]. 软件学报,2016,27(8):1891-1903.
- [10] 朱永红,丁恩杰. 负载均衡的异构 WMSN 节点布局方法[J]. 通信学报,2015,36(10):157-164.
- [11] QIAN Y F,HU L,CHEN J,et al. Privacy-aware service placement for mobile edge computing via federated learning[J]. Information sciences,2019,505:562-570.
- [12] LU R X,KEVIN H,ARASH L H,et al. A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot[J]. IEEE access,2017,5:3302-3312.
- [13] DU M,WANG K,XIA Z Q,et al. Differential privacy preserving of training model in wireless big data with edge computing[J]. IEEE transactions on big data,2018,6(2):283-295.
- [14] DING L,MALEK S B. A novel architecture for automatic document classification for effective security in edge computing environments[C]//2018 IEEE/ACM Symposium on Edge Computing(SEC),Singapore;IEEE,2018:416-420. doi:10.1109/SEC.2018.00056.
- [15] HE X F,LIU J,JIN R C,et al. Privacy-aware offloading in mobile-edge computing[C]//GLOBECOM 2017-2017 IEEE Global Communications Conference,Singapore;IEEE,2017:1-6.
- [16] MAO Y Y,ZHANG J,KHALED L B. Dynamic computation offloading for mobile-edge computing with energy harvesting devices[J]. IEEE journal on selected areas in communications,2016,34(12):3590-3605.
- [17] CHEN X,JIAO L,LI W Z,et al. Efficient multi-user computation offloading for mobile-edge cloud computing[J]. IEEE/ACM transactions on networking,2015,24(5):2795-2808.
- [18] TAO X Y,KAORU O,DONG M X,et al. Performance guaranteed computation offloading for mobile-edge cloud computing[J].

- IEEE wireless communications letters,2017,6(6):774-777.
- [19] TIAN H,JIANG L,ALEXEY V,et al. Selective offloading in mobile edge computing for the green internet of things[J]. IEEE network,2018,32(1):54-60.
- [20] ZHANG K,MAO Y M,LENG S P,et al. Optimal delay constrained offloading for vehicular edge computing networks[C]// 2017 IEEE International Conference on Communications(ICC),Paris:IEEE,2017:1-6.
- [21] WANG F,XU J,WANG X,et al. Joint offloading and computing optimization in wireless powered mobile-edge computing systems[J]. IEEE transactions on wireless communications,2017,17(3):1784-1797.
- [22] WANG Y T,SHENG M,WANG X J,et al. Mobile-edge computing:partial computation offloading using dynamic voltage scaling[J]. IEEE transactions on communications,2016,64(10):4268-4282.
- [23] CHEN M,HAO Y X. Task offloading for mobile edge computing in software defined ultra-dense network[J]. IEEE journal on selected areas in communications,2018,36(3):587-597.
- [24] YU S,LANGAR R,FU X M,et al. Computation offloading with data caching enhancement for mobile edge computing[J]. IEEE transactions on vehicular technology,2018,67(11):11098-11112.
- [25] YUAN D,YANG R,LIU R,et al. A data placement strategy in scientific cloud workflows[J]. Future generation computer systems,2010,26(8):1200-1214.

[责任编辑:顾晓天]