

面向移动边缘计算中多应用服务的虚拟机部署算法

李光辉 周辉* 胡世红

(江南大学人工智能与计算机学院 无锡 214122)

摘要: 移动边缘计算(MEC)通过在用户近端以虚拟机(VM)形式部署应用服务,能有效降低服务响应延迟并减少核心网络数据流量。然而,当前MEC中虚拟机部署的大多数研究尚未具体考虑用户对多种应用服务的需求。因此,该文针对MEC中多应用服务的虚拟机部署问题,提出两种启发式算法,即基于适应度的启发式部署算法(FHPA)和基于分治的启发式部署算法(DCBHPA),通过在边缘网络中配置支持多种应用服务的虚拟机来最大限度地减少网络中的数据流量。FHPA和DCBHPA分别基于边缘服务器的网络连接特征和用户对应用请求的差异性,定义了不同的适应度计算模型。在此基础上,通过子问题划分机制实现VM配置。仿真结果表明,相比于基准算法,所提算法能更好地控制系统数据流量,有效地提高边缘网络服务资源的利用率。

关键词: 移动边缘计算; 虚拟机部署; 多应用; 核心网流量

中图分类号: TN919.2; TP393

文献标识码: A

文章编号: 1009-5896(2022)07-2431-09

DOI: 10.11999/JEIT210415

Virtual Machine Placement Algorithm for Supporting Multiple Applications to Mobile Edge Computing

LI Guanghui ZHOU Hui HU Shihong

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

Abstract: In Mobile Edge Computing (MEC) environment, deploying application services in the form of Virtual Machines (VM) at the edge of the network can effectively reduce the service response delay and reduce the data traffic of core network. There have been many solutions to the problem of optimal allocation of edge network resources, but few studies consider the optimal deployment of VM that provide users with multiple application services to mobile edge networks. To this end, two heuristic algorithms are proposed, Fitness-based Heuristic Placement Algorithm (FHPA) and Divide-and-Conquer Based Heuristic Placement Algorithm (DCBHPA). By distributing VMs that support multiple application services to the MEC network, these two algorithms aim to minimize the data traffic in MEC architecture. Besides, FHPA and DCBHPA define respectively different fitness computing models based on the network connection characteristics of edge servers, as well as the differences in users' application requests. Thus, VM configuration can be realized through the sub-problem division mechanism. Compared with the baseline algorithms, the simulation results show that the proposed algorithms can better control the system data traffic and improve effectively the utility of edge network service resources.

Key words: Mobile Edge Computing (MEC); Virtual Machines (VM) placement; Multiple applications; Core network traffic

1 引言

近年来,5G技术和物联网技术快速发展,移动设备(如智能手机和可穿戴设备等)的数量急剧增

长。各种应用服务的爆炸性激增对计算能力和实时处理提出了严格的要求。受限于计算资源和能源供给^[1]的移动设备已经无法应对这一挑战。移动边缘计算^[2](Mobile Edge Computing, MEC)作为一种有潜力的新兴计算范式,可通过在用户附近部署计算和存储等资源来提高服务质量^[3]。AR和视频游戏等实时交互应用的增多导致移动边缘网络中的数据流量迅速增长,造成严重的网络资源消耗,使用户访问应用服务时延迟增加。因此,为了更好地控制

收稿日期: 2021-05-12; 改回日期: 2021-10-29; 网络出版: 2021-12-24

*通信作者: 周辉 6191910044@stu.jiangnan.edu.cn

基金项目: 国家自然科学基金(62072216)

Foundation Item: The National Natural Science Foundation of China (62072216)

网络时延并减少潜在的核心网堵塞, 必须减少边缘网络的数据流量。一个有效的解决方案是在边缘网络中部署支持多种应用服务的虚拟机(Virtual Machine, VM)为移动用户提供各种近端服务^[4], 依靠高带宽和脱机可用性来缓解核心网络的拥塞, 同时通过满足延迟和能耗等关键应用需求进一步提高计算服务质量。

VM是一个逻辑容器, 可使用数据运行软件实例。在MEC中使用VM技术, 能使MEC服务器成为特定应用的应用服务器, 即特定的应用请求只能由部署有相关应用VM的MEC服务器提供服务。目前, 已有许多关于MEC中VM管理问题的研究。由于用户的移动性, MEC系统需要在用户位置随时间变化时进行服务迁移, 文献^[5]提出了一种代理VM迁移方案, 采用基站与雾节点在本地提供计算资源的方式来为用户提供服务。为了减少服务迁移过程中出现的性能下降问题, 文献^[6]提出了一种自适应VM切换方法来减少用户移动时进行服务迁移需要传输的数据量。考虑到用户多种服务需求, 文献^[7]提出3种启发式算法寻找VRC最优部署位置来最小化平均请求响应时间。此外, 文献^[8]设计了一种VM容错部署机制来针对可能出现的边缘服务器故障问题。然而, 很少有研究考虑边缘网络中关于数据流量的绿色通信问题。

尽管现有工作对MEC中边缘网络的资源配置问题提出了很多解决方案, 但存在一个共同的局限性: 没有具体考虑用户对多种应用服务的需求。在边缘网络中提供多种应用服务将面临许多挑战, 如不同应用对传输带宽的需求差异会导致不同的平均链路时延, 所需服务资源的多样性会导致边缘服务器不同的工作负载等。文献^[9]采用云边资源协作的方式为用户提供服务, 提出一种云边协同模式下的任务分配机制将应用负载分配给合适的微云(Cloudlet)来最小化用户响应时延。文献^[10]通过图分割的方式将实际的物理网络抽象为一个完全图, 用服务器之间的通信时延表示边缘网络中的费用, 提出一种数据访问延迟最小化的虚拟机部署最优近似算法。然而他们没有考虑为用户提供多种不同特性的应用服务。文献^[11]设计了一种有效的任务调度和资源管理策略最小化任务完成时间, 文中提到可以支持多种应用类型, 但是没有明确指出支持的不同应用之间的区别。多数工作在部署边缘网络中的资源提供服务时, 没有详细考虑提供多种应用的差异性^[12], 而是集中于研究边缘服务器和移动用户之间的交互及协作^[13,14], 他们简单地假设所有移动用户的资源需求是统一的^[15], 而实际情况中, 必须考虑不同应

用对资源的需求差异, 这些将显著影响最优路由选择和工作负载分配。

因此, 本文以最小化移动边缘网络中云边数据流量和最小化服务配置总成本为目标, 构建了一种云边协同的计算架构, 在此架构基础上对支持多种应用服务的虚拟机部署问题进行研究, 在边缘服务器服务能力一定且提供多种应用服务前提下, 将问题形式化描述为资源约束下的最小化云边数据流量和部署总成本的优化问题。本文主要工作如下:

(1) 基于MEC架构, 本文首次研究了在提供多种应用服务的边缘网络中, 将每种应用的 k 个VM部署到 S 个边缘服务器上, 以最小化边缘网络中的平均数据流量(Minimizing the Average Data Traffic, MADT)为目标。

(2) 针对上述问题, 本文首先从网络拓扑结构特性出发, 计算每个边缘服务器的中心性特征, 根据定义的适应度模型, 提出了一种基于适应度的启发式部署算法(Fitness-based Heuristic Placement Algorithm, FHPA)。另外, 从用户对不同应用服务需求差异化考虑, 定义适应度模型, 通过子问题划分策略, 设计了一种基于分治的启发式部署算法(Divide-and-Conquer Based Heuristic Placement Algorithm, DCBHPA)。

(3) 仿真结果表明, 相比于基准算法和FHPA, DCBHPA能最大限度地减少数据流量, 有效地缓解远端云的流量压力, 减少潜在的核心网拥塞。同时, 对边缘网络中提供服务的配置总成本问题进行了研究。

2 系统模型

2.1 移动边缘计算架构

基于MEC的系统框架如图1所示, 该框架由移

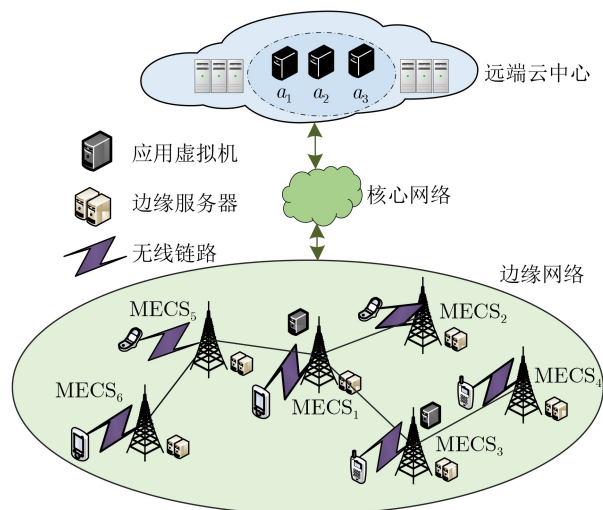


图1 提供多种应用服务的移动边缘计算架构

动设备、MEC服务器组成的边缘网络和远端云中心构成, 其中, 边缘网络采用的是MEC服务器与基站共同部署在无线接入网内的结构。各种资源密集型和时间敏感型的移动应用源服务器, 例如 a_1, a_2, a_3 , 运行在远端云计算中心, 同时在边缘端, 边缘服务器MECS₁和MECS₃分别部署有应用 a_1 的VM, 可充当应用 a_1 的应用服务器(Application Server, AS), 为移动用户提供特定的服务请求。

所有来自用户的请求首先由本地边缘服务器进行处理, 每个边缘服务器会根据当前边缘网络的资源利用情况为服务请求做出最优路由抉择, 确保其从边缘网络相应AS或远端云中心获取服务。例如, MECS₂服务范围内的移动用户对应用 a_1 的请求将被路由到部署有应用 a_1 相应VM的MECS₁, 由服务器MECS₁提供服务。同时对每个边缘服务器的服务容量进行约束, 当网络中AS的服务资源耗尽时, 如MECS₁和MECS₃全部过载, 新接收的应用 a_1 的请求由云中心提供服务。

2.2 网络模型

边缘网络拓扑结构表示为图 $G(S, L)$, 其中, 点集 $S = \{s_i, i = 1, 2, \dots, |S|\}$ 表示边缘服务器集合, 边集 $L = \{l_i, i = 1, 2, \dots, |L|\}$ 为边缘服务器之间的链路, 图 G 的直径为 E 。由不同边缘服务器覆盖的服务范围, 将边缘网络划分为不同区域, p_s 表示边缘服务器 $s \in S$ 服务域内的移动用户数, 则 $U = \{p_s, \forall s \in S\}$ 表示边缘网络中所有的用户。为了满足用户对不同应用服务的需求, 在边缘网络中部署支持多种应用服务的VM, 应用服务集由集合 $A = \{a_i, i = 1, 2, \dots, |A|\}$ 表示, 使用指示变量 I_s^a 表示MEC服务器 s 是否配置有应用 a 的VM, 若是, 则 $I_s^a = 1$, 反之, $I_s^a = 0$ 。

边缘网络提供的服务资源通常有限, MEC服务器 s 的处理容量定义为 c_s , 则整个边缘网络的服务能力表示为集合 $C = \{c_s, \forall s \in S\}$ 。对于不同的应用服务, 对请求应用 a 的服务资源需求量表示为 r_a , 并使用变量 B_a 表示访问应用 a 所要求的网络传输带宽, 同时, 将来自MEC服务器 v 服务域内对应用 a 的请求由MEC服务器 u 提供服务所占的请求比例由变量 $\eta_{v,u}^a$ 表示。

3 问题描述

3.1 平均数据流量最小化问题

移动用户的服务请求在云边协同的MEC架构获取服务产生的数据流量包括: (1)由边缘网络中的AS为移动设备提供服务的边缘流量; (2)远端云提供服务产生的云边通信流量。将边缘网络中任意两台边缘服务器间的最短路径作为二者间的路由选择, 使用变量 $d_{v,u}$ 表示边缘服务器 v 到 u 的最短路径

跳数, 则MEC服务器 v 服务区域内的用户请求路由到MEC服务器 u 获取对应用 a 请求服务的数据流量定义为

$$D_{v,u}^a = B_a \cdot \eta_{v,u}^a \cdot R_v^a \cdot d_{v,u} \quad (1)$$

其中, $0 \leq \eta_{v,u}^a \leq 1, \forall v, u \in S, \forall a \in A$, 且 $R_v^a = p_v \cdot \delta_v^a \cdot \lambda_a^a$, 满足 $0 \leq \delta_v^a \leq 1, \forall v \in S, \forall a \in A$, R_v^a 表示MEC服务器 v 服务范围内发送对应用 a 的请求数, δ_v^a 为移动用户在服务器 v 范围内对 a 的请求速率, 并且用户对应用的请求平均生成速率服从泊松分布, 使用变量 λ_a 表示对应用 a 的平均请求生成率。

当边缘资源无法满足用户需求时, 即边缘网络中的AS过载, 超额的用户请求由云端提供服务。根据边缘网络中AS进行处理的服务请求数, 可计算出由远端云提供服务的请求占全部请求的比例, 即 $1 - \sum_{u \in S} \eta_{v,u}^a$ 。因此, 在边缘网络中部署提供多种应用服务的VM以最大限度地减少网络中的平均数据流量的目标优化函数为

$$\left. \begin{aligned} & \min \sum_{a \in A} \left(\sum_{v \in S} \left(\sum_{u \in S} D_{v,u}^a \right) \right. \\ & \quad \left. + B_a \cdot \sum_{v \in V} \left(\left(1 - \sum_{u \in S} \eta_{v,u}^a \right) \cdot R_v^a \cdot d \right) \right) \\ \text{s.t. C1: } & \sum_{v \in S} I_v^a = k, \forall a \in A \\ \text{C2: } & \sum_{a \in A} I_v^a \leq |A|, \forall v \in S \\ \text{C3: } & I_v^a = \{0, 1\}, \forall v \in S, \forall a \in A \\ \text{C4: } & \sum_{v \in S} \eta_{u,v}^a \leq 1, \forall u \in S, \forall a \in A \\ \text{C5: } & \eta_{u,v}^a - I_v^a \leq 0, \forall u, v \in S, \forall a \in A \\ \text{C6: } & \sum_{a \in A} R_v^a \cdot r_a \leq c_v, \forall v \in S, \forall c_v \in C \end{aligned} \right\} \quad (2)$$

其中, C1声明支持每种应用服务的VM数量为 k 。C2和C3确保在任一MEC服务器中至多部署同一应用的1个VM。C4确保MEC服务器 u 服务范围内用户对应用 a 的请求数不会超过该服务区内应用 a 的总请求量。C5将请求分配和VM部署进行连接, 规定只有当MEC服务器 v 部署有应用 a 的VM才能向移动用户提供该应用的服务。C6保证分配给MEC服务器 v 的所有应用总体请求的资源需求不会超出MEC服务器 v 的服务阈值。

3.2 应用服务配置总成本问题

配置总成本由VM部署成本和请求服务的数据流量成本构成。将相同应用相应VM间的状态同步和部署VM到边缘服务器的资源需求作为部署成本, 定义为所有应用种类数和支持每种应用VM数

的2次函数, 即 $\varepsilon \cdot k \cdot |A|$, 其中 $0 < \varepsilon < 1$ 。数据流量成本体现了边缘网络的拥堵情况, 可由所取得的边缘端流量 D_e 和云端流量 D_c 的线性函数表示, 分别为 $\sigma_1 \cdot D_e$ 和 $\sigma_2 \cdot D_c$, 其中 $\sigma_1, \sigma_2 > 0$ 。对部署成本和数据流量成本经过标准化函数 f, N 计算后, 定义整个系统的服务配置总成本为

$$\tau_k^A = f(N(\sigma_1 \cdot D_e) + N(\sigma_2 \cdot D_c) + N(\varepsilon \cdot k \cdot |A|)) \quad (3)$$

4 算法设计

4.1 基于适应度的启发式部署算法

为了解决MADT问题, 本文提出一种基于适应度的启发式部署算法。对于应用 a , FHPA寻找 k 个适应度最高的MEC服务器部署应用虚拟机, 其中每个MEC服务器的适应度评价由4种网络连接特征构成, 包括接近中心性(closeness centrality)、度中心性(degree centrality)、中介中心性(betweenness centrality)和特征向量中心性(eigenvector centrality), 再根据熵权法对每个连接特征的权重进行计算。下面首先介绍上述连接特征的概念。

(1)度中心性。度中心性的基本思想是网络中重要节点为与其他节点的连接边数较多的节点。相应地, MEC服务器 s 的度中心性为与他直接相连的MEC服务器个数, 归一化后服务器 s 的度中心性计算表达式为

$$\alpha_s^d = \frac{\deg(s)}{|S| - 1} \quad (4)$$

其中, $\deg(s)$ 表示 s 通过1跳连接到其他MEC服务器的跳数。

(2)接近中心性。接近中心性是基于网络最短路径的一种衡量节点重要程度的中心性方法。一个节点的接近中心性值是该节点到网络其余节点经过边的最小数目。相应地, MEC服务器 s 的接近中心性为

$$\alpha_s^c = \frac{1}{\sum_{j \in S} d(s, j)} \quad (5)$$

α_s^c 值越大, 代表在边缘网络中服务器 s 与其他MEC服务器间的最短路径越短。由式(1)可知, 部署应用虚拟机在接近中心性值较大的MEC服务器上, 为用户提供服务产生的数据流量会更小。

(3)中介中心性。中介中心性是基于平均最短路径的常用中心性测试指标, 指的是在边缘网络中两个MEC服务器间经过 s 的最短路径所占二者间最短路径总数之比, 其计算表达式为

$$\alpha_s^b = \sum_{i \neq s \neq j \in S} \frac{\sigma_{ij}(s)}{\sigma_{ij}} \quad (6)$$

其中, σ_{ij} 表示MEC服务器 i 到 j 的最短路径数, $\sigma_{ij}(s)$ 表示 i 到 j 经过 s 的最短路径数。

(4)特征向量中心性。特征向量中心性从节点的邻接节点的重要性来衡量其在网络中的影响力^[6], 即一个节点的特征向量中心性值由其邻接节点的值决定。MEC服务器 s 的特征向量中心性为

$$\alpha_s^e = \sum_{j=1}^{|S|} m_{js} \alpha_j^e \quad (7)$$

令 M 为邻接矩阵, 其元素 m_{js} 表示MEC服务器 j 对服务器 s 中心度的贡献量。

使用 F 来表示所有连接特征集, $F = \{\alpha^c, \alpha^d, \alpha^b, \alpha^e\}$, 则MEC服务器 s 的连接特征指标值分别由 $\alpha_s^c, \alpha_s^d, \alpha_s^b, \alpha_s^e, \forall s \in S$ 表示。基于每个MEC服务器的连接特征, 使用信息熵来确定连接特征中各指标值的离散化程度, 离散化度越高, 表示其熵值越低, 意味着该连接特征对MEC服务器的适应度计算影响更大。经过每个MEC服务器连接特征取值的归一化过程, 可求得每个连接特征的熵值, 其计算公式为

$$H_\alpha = \frac{-\sum_{s \in S} \alpha_s \ln \alpha_s}{\ln |S|}, \forall \alpha \in F \quad (8)$$

此处规定, 若 $\alpha_s = 0$, 则 $\ln \alpha_s = 0$ 。基于所求熵值, 连接特征的熵权为

$$\omega_\alpha = \frac{1 - H_\alpha}{|F| - \sum_{\alpha \in F} H_\alpha}, \forall \alpha \in F \quad (9)$$

熵权 ω_α 用来从整体上度量每个连接属性对每个MEC服务器部署VM的适应度。基于每个MEC服务器的连接特征和相应熵权, MEC服务器 s 的适应度定义为

$$W_s = \sum_{\alpha \in F} \omega_\alpha \alpha_s, \forall s \in S \quad (10)$$

对于应用 $a \in A$, 基于每个MEC服务器的适应度降序排序结果, 选择前 k 个服务器为应用 a 的初始AS, 分别分配到集合 $X_i, i = 1, 2, \dots, k$, 结合文献[17], 设计一个聚类机制如表1, 轮询地将全部候选MEC服务器划分到与其最近的初始AS所在集合 X_i 。对集合 X_i 中的每一个MEC服务器, 将其作为该集合中应用 a 的AS为集合中其他边缘服务器服务范围内的应用请求提供服务, 并计算所产生的数据流量, 最后选择能最小化该集合中平均数据流量的MEC

服务器, 部署应用VM。FHPA通过将MADT问题划分为 k 个子问题, 在每个子问题中寻找应用服务 a 的一个VM最优部署位置, 然后通过整合子问题的最优解求解原问题, FHPA详情如表2所示。

4.2 基于分治的启发式部署算法

由于应用请求在本地MEC服务器进行处理不会注入流量到边缘网络, 因此, 选择覆盖范围内应用请求数较大的MEC服务器作为AS, 可以减少产生到边缘网络的流量。然而, 由于AS容量受限, 这样会导致该AS附近的边缘服务器的应用请求发送到云端的比例将上升, 进而增加系统流量成本。因此, 本节进一步提出一种基于分治的启发式部署算法, 与文献[18]只考虑用户请求密度不同, DCBHPA将当前已配置应用 a 相应VM的边缘服务器对网络中其他将部署同应用服务VM的候选MEC服务器的影响进行了考虑, 从而避免应用 a 的

AS集中在网络某一局部地区, 提高服务资源的利用率。

变量 Q 表示已部署应用 a 相应VM的MEC服务器集, 与当前已部署的AS越远, 则候选MEC服务器所受影响越小, 设当前已部署的AS j 对候选MEC服务器 $i \in S$ 的影响因子为 $\theta \cdot t(i, j)$, 其中 $0 < \theta < 1$, $t(i, j)$ 为 i 和 j 归一化后的距离。每个候选MEC服务器将受到当前所有已部署AS的影响, 影响因子为

$$\beta_i = \begin{cases} 1 / \sum_{j \in Q} \theta \cdot t(i, j), & Q \neq \phi \\ 1, & Q = \phi \end{cases} \quad (11)$$

其中, β_i 值越高, 表明在受其他已部署AS影响情况下, 候选MEC服务器 i 更适合部署应用VM。MEC服务器 i 覆盖范围内对应用 $a \in A$ 的请求数 R_i^a 归一化后用变量 γ_i^a 表示, 类似FHPA, 则MEC服务器 i 的适应度为

$$P_i = \beta_i \cdot \gamma_i, \forall i \in S \quad (12)$$

在每次迭代中, DCBHPA选择 P 值更高的MEC服务器作为初始AS, k 次迭代更新后, 选择 k 个初始AS加入集合 $X_i, i = 1, 2, \dots, k$ 进行初始化, 之后的求解MADT过程同FHPA, DCBHPA详情见表3。

表 1 聚类过程

初始化图 G 的直径 E , 根据适应度定义集合 X_i , 令 Y 为空集
(1) for $j \leftarrow 1$ to E do
(2) for $i \leftarrow 1$ to k do
(3) 对于距离集合 X_i 中初始AS跳数为 j 的元素 y
(4) if $y \notin Y$ then
(5) 更新集合 $Y \leftarrow Y \cup \{y\}$
(6) 更新集合 $X_i \leftarrow X_i \cup \{y\}$
(7) end if
(8) end for
(9) end for

表 2 基于适应度的启发式部署算法

输入: G, k, A, U, δ, E, X
输出: 最优VM分配策略
(1) 计算和极值正规化 $\alpha_s, \forall s \in S, \forall \alpha \in F$
(2) 根据式(9)计算熵权 $\omega_\alpha, \alpha \in F$
(3) 根据式(10)计算 $W_s, \forall s \in S$
(4) 对 W 进行降序排序, 选择前 k 个MEC服务器分别分配到集合 $X_i, i = 1, 2, \dots, k$
(5) 执行表1中的聚类过程将所有MEC服务器划分到 $X_i, i = 1, 2, \dots, k$
(6) for $a \in A$ do
(7) for $i \leftarrow 1$ to k do
(8) for $s \in X_i$ do
(9) 将应用 a 的一个VM配置到 s 上, 通过式(2)计算集合 X_i 内的数据流量
(10) end for
(11) 将应用服务 a 的一个VM部署在最小化 X_i 内数据流量的边缘服务器上
(12) end for
(13) end for

表 3 基于分治的启发式部署算法

输入: G, k, A, U, δ, E, X
输出: 最优VM分配策略
(1) 计算并归一化 $R_s^a, \forall s \in S, \forall a \in A$
(2) for $a \in A$ do
(3) $Y = \emptyset$
(4) for $i \leftarrow 1$ to k do
(5) for $s \in S$ do
(6) 通过式(11)计算影响因子 β_s
(7) end for
(8) $P = \{P_s = 0, \forall s \in S\}$
(9) for $s \in S$ do
(10) 根据式(12)计算 P_s
(11) end for
(12) 对 P 降序排序, 选择第1个 P_s , 将满足 $s \notin Y$ 的 s 加入集合 X_i
(13) 更新集合 $Y = Y \cup \{s\}$
(14) end for
(15) 执行表1中的聚类过程将所有MEC服务器划分到 $X_i, i = 1, 2, \dots, k$
(16) end for
(17) 根据表2的步骤(7)—步骤(12)找到最优VM部署位置
(18) end for

4.3 算法分析

FHPA是一个简单有效的算法。它首先计算每个MEC服务器的连接特征。其中,每个MEC服务器的接近中心性表明它与边缘网络中其余MEC服务器的平均距离。度中心性越高,则该MEC服务器能在相同的距离内与更多的MEC服务器相连,距离越短则服务请求产生的数据流量越小。因此具有更高度中心性的MEC服务器更适合部署应用虚拟机。当网络中所有移动设备的应用服务请求路由到相应的应用服务器时,通过每个MEC服务器的路径个数能显示该MEC服务器上需传输数据的到达速率,边缘网络中路由路径数越多经过该MEC服务器,表明该MEC服务器节点拥堵程度越高,则请求数据在该服务器节点的等待时间将很长,这一特性通过中介中心性表现出来。为了避免在网络中影响程度较低的MEC服务器成为应用服务器,将每个MEC服务器的特征向量中心性考虑进来。因此,基于归一化的连接特征值,计算每个MEC服务器上的所有连接特征的权重之和,得出每个MEC服务器部署应用虚拟机的适应度。最后FHPA选择适应度最高的 k 个MEC服务器部署应用虚拟机。

然而, FHPA也存在两个缺点。第1个缺点是选择虚拟机部署的MEC服务器时,没有考虑用户的请求数。来自用户的服务请求由本地边缘服务器处理将不会注入流量到边缘网络,从而选择覆盖范围内有大量应用请求的MEC服务器部署应用虚拟机,并将其作为AS可以减少网络的流量负载。因此,边缘服务器范围内移动用户的服务请求数也是影响其部署应用虚拟机成为AS的重要因素。然而, FHPA从网络拓扑结构的连接特征出发来确定应用虚拟机部署位置,没有利用移动用户请求数来提升部署效益。第2个缺点是当大量用户集中在边缘网络同一局部地区时, FHPA容易造成部分AS过载。因此,本文从用户请求角度出发进一步提出DCBHPA算法求解MADT问题。

DCBHPA考虑了当前已部署应用虚拟机的边缘服务器对边缘网络中其余候选MEC服务器的影响因素。第1个是每个MEC服务器与当前已部署虚拟机的MEC服务器间的平均距离,为了避免应用服务器集中部署在边缘网络的某一局部地区,候选MEC服务器与当前已部署应用虚拟机的MEC服务器距离越远越合适。第2个是在当前已部署的应用服务器提供服务的情况下,每个边缘服务器接收的应用请求数量。为了减少边缘网络中的平均数据流量,接收应用请求数量更多的MEC服务器更适合部署应用虚拟机作为下一个AS。DCBHPA在每次

迭代中选择适应度最高的MEC服务器来部署应用虚拟机,对于应用 $a \in A$,在 k 次迭代后,获得 k 个最适合部署关于应用 a 相应虚拟机的边缘服务器。因此, DCBHPA在共经过 $k \cdot |A|$ 次迭代后,完成所有应用的虚拟机部署。

5 仿真分析

本文利用Python工具完成了仿真实验。不失一般性,从Topology Zoo获得真实的在线网络拓扑用于VM部署实验^[19],网络拓扑图结构如表4所示。不同应用服务的网络带宽需求设置为[800 kB, 1600 kB],并假定分布在整个边缘网络的移动用户数在500~3000。默认设置3种应用以及每种应用部署4个VM和noel通信网络进行实验。同时,选择文献^[20]中的RPA算法和文献^[21]中的DBC算法进行性能比较。

5.1 网络数据流量

如图2所示,各算法产生的平均数据流量随着支持每种应用的VM数增加而单调减少。这是因为随着边缘网络中支持同一应用服务的VM数增多,即该应用的AS数量增加,可以在本地进行更多的请求处理,从而减少注入到边缘网络中的数据流量。同时,配置更多的VM可以减少边缘网络中移动用户与请求的目的AS之间的平均距离,导致边缘网络更小的数据流量产生。从图2可以发现,DCBHPA在最小化平均数据流量方面相比其他算法可以获得最优性能, FHPA性能表现要优于DBC,且二者都优于RPA。

表4 网络拓扑表

拓扑图	节点数	链路数
spiralight	15	16
noel	19	25
agis	25	30
shentel	28	35
evolink	37	45

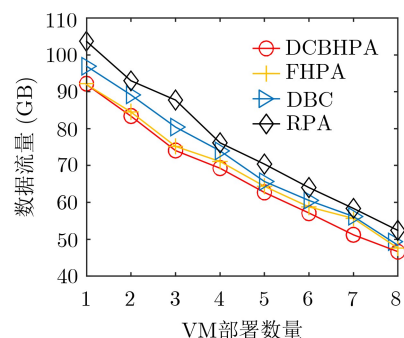


图2 不同VM数对数据流量的影响

从图3可以清楚地发现，在相同数量的移动用户情况下，与FHPA, DBC和RPA相比，DCBHPA总是可以获得最小的平均数据流量。此外，在移动用户数较小的情况下，各算法性能差异较小，随着用户数不断增加，所有4种算法提供服务产生的平均数据流量都会增加，其中RPA数据流量增加最为明显。当移动用户数为3000时，DCBHPA相比RPA可以减少25.8%的流量生成，反映了DCBHPA在大规模移动边缘网络中为移动用户提供多种应用服务进行VM配置的优势。

图4为不同网络拓扑下FHPA, DCBHPA, DBC和RPA在边缘网络中最小化请求数据流量方面的性能比较结果，可以看到4种算法提出的服务配置方案产生的数据流量会随着网络规模的扩大而上升。在同一网络下，DCBHPA在减小平均数据流量方面能获得最佳性能表现。同时，图5进一步比较了FHPA, DCBHPA, DBC和RPA在应用服务类型不断增多情况下的性能。实验结果表明，DCBHPA要优于其他3种算法。与此同时，随着不同应用种类的增加，各算法进行VM配置提供服务所产生的数据流量都会增长，且DCBHPA与FHPA, DBC和RPA的性能差异也更加明显。此外，FHPA和DBC的性能表现基本相近，且二者都要优于RPA。

接着，在支持每种应用服务的VM数量不同且应用服务种数不断增加的前提下，进一步分析了

DCBHPA的性能表现。仿真结果如图6所示，当应用服务数量增加时，对于支持各应用服务的每组VM设置，DCBHPA获得的平均数据流量呈指数增加。这是因为更多的应用服务种类意味着边缘网络中将有更多的应用请求需要提供服务，而这将导致AS可用服务资源的短缺，进而增加系统的数据流量产生。同时，更多的服务请求将发送到云端获得服务从而增加流量成本。从图6可以看出，当应用服务种类一定时，随着支持每种应用服务的VM增加，系统的平均数据流量将越少，这与图2的实验结果相符。

5.2 服务配置总成本

一个好的部署方案不仅要保证较低的平均数据流量，而且要考虑边缘网络中提供服务配置的成本因素，因此，我们进一步分析了DCBHPA提供服务时的总成本。如图7所示，同一应用提供更多的VM部署到边缘网络将减少流量成本，同时增加部署成本。通过将归一化后的流量成本和部署成本相加，可以得到服务提供总成本。从图中可以发现，当VM部署数量不超过3时，总成本随着VM配置数量的增加而下降，这是因为当VM的数量很小时，主导服务配置总成本的是数据流量成本，如图2所述，因此总成本将随着支持每种应用服务的VM数量的增加而减少。当部署VM的数量继续增加时，VM的配置成本就会逐渐占据服务配置总成本的主导地位，导致当支持每种应用服务的VM数量超过

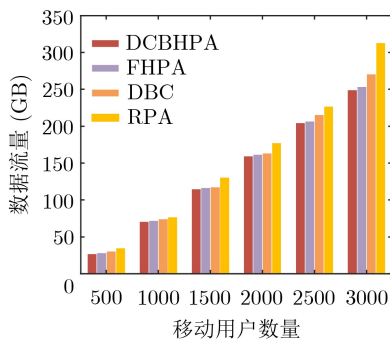


图3 移动用户数对数据流量的影响

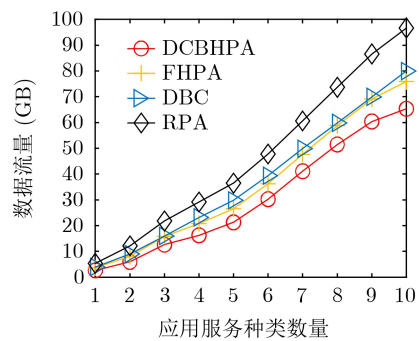


图5 应用服务种类数对数据流量的影响

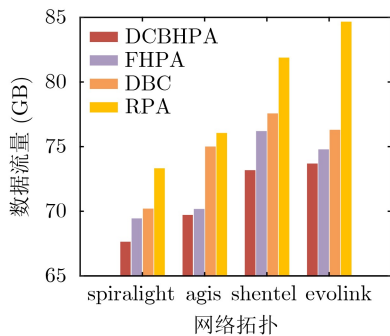


图4 网络规模对数据流量的影响

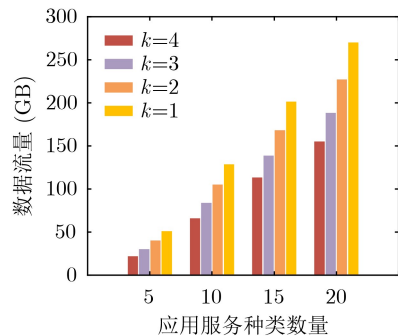


图6 不同应用服务下VM部署取得的数据流量

3时, DCBHPA提供服务配置的总成本会随着VM数的增加而增长。

如图8所示, 在应用服务种类不断增加且支持每种应用服务的VM数为4的情况下, 对4种算法在应用服务部署成本方面的性能进行了比较。图8表明, 在最小化部署总成本方面, DCBHPA可以获得最佳性能表现, 且FHPA要优于DBC, 同时二者都要强于RPA。根据定义, VM部署成本为支持每种应用的VM数和不同应用种类数的2次函数, 因此部署成本将会随着应用数的增加而上升。图5表明所有算法产生的数据流量会随着应用种类数的扩展而显著增加, 因此数据流量成本将随着算法产生的数据流量增加而线性增长, 所以服务配置总成本将随着应用种类数的增加而上升, 这与图8的结果一致。

6 结束语

本文研究了移动边缘网络中支持多种应用服务的VM部署问题, 在提供多种应用服务请求和MEC服务器处理能力受限约束下, 为了最小化移动边缘网络中的平均数据流量这一目标, 提出了两种启发式的部署算法, 并利用广泛的仿真实验来评估所提算法的性能。实验结果表明, DCBHPA算法性能在最大限度地减少平均数据流量方面, 要优于FHPA算法和基准算法, 能有效减少云边通信流量, 缓解潜在的网络堵塞问题。未来工作中, 将进

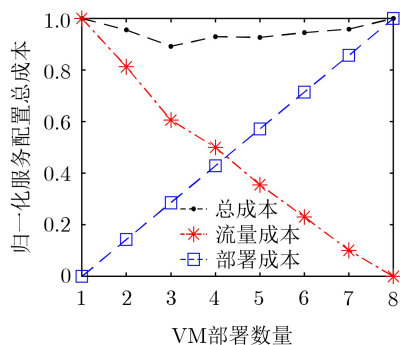


图7 不同VM数对归一化服务配置总成本的影响

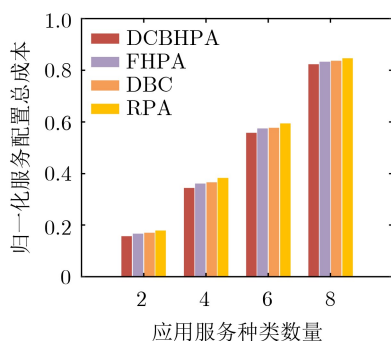


图8 应用服务种类数对归一化服务配置总成本的影响

一步深入研究多应用服务下不同VM部署方案对用户所需服务的平均请求响应时间的影响规律。

参考文献

- [1] JIANG Congfeng, FAN Tiantian, GAO Honghao, *et al.* Energy aware edge computing: A survey[J]. *Computer Communications*, 2020, 151: 556–580. doi: [10.1016/J.COMCOM.2020.01.004](https://doi.org/10.1016/J.COMCOM.2020.01.004).
- [2] KHAN W Z, AHMED E, HAKAK S, *et al.* Edge computing: A survey[J]. *Future Generation Computer Systems*, 2019, 97: 219–235. doi: [10.1016/J.FUTURE.2019.02.050](https://doi.org/10.1016/J.FUTURE.2019.02.050).
- [3] PAN Jianli and MCELHANNON J. Future edge cloud and edge computing for internet of things applications[J]. *IEEE Internet of Things Journal*, 2018, 5(1): 439–449. doi: [10.1109/JIOT.2017.2767608](https://doi.org/10.1109/JIOT.2017.2767608).
- [4] ZHOU Yuchen, YU F R, CHEN Jian, *et al.* Resource allocation for information-centric virtualized heterogeneous networks with in-network caching and mobile edge computing[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(12): 11339–11351. doi: [10.1109/TVT.2017.2737028](https://doi.org/10.1109/TVT.2017.2737028).
- [5] SUN Xiang and ANSARI N. EdgeIoT: Mobile edge computing for the internet of things[J]. *IEEE Communications Magazine*, 2016, 54(12): 22–29. doi: [10.1109/MCOM.2016.1600492CM](https://doi.org/10.1109/MCOM.2016.1600492CM).
- [6] HA K, ABE Y, EISZLER T, *et al.* You can teach elephants to dance: Agile VM handoff for edge computing[C]. The 2nd ACM/IEEE Symposium on Edge Computing, San Jose, USA, 2017: 12. doi: [10.1145/3132211.3134453](https://doi.org/10.1145/3132211.3134453).
- [7] ZHAO Lei and LIU Jiajia. Optimal placement of virtual machines for supporting multiple applications in mobile edge networks[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(7): 6533–6545. doi: [10.1109/TVT.2018.2808171](https://doi.org/10.1109/TVT.2018.2808171).
- [8] GAO Siyi, ZHOU Ao, CHEN Xican, *et al.* Redundant virtual machine placement in mobile edge computing[C]. The 1st International Conference on Blockchain and Trustworthy Systems, Guangzhou, China, 2019: 371–384. doi: [10.1007/978-981-15-2777-7_30](https://doi.org/10.1007/978-981-15-2777-7_30).
- [9] SUN Xiang and ANSARI N. Latency aware workload offloading in the cloudlet network[J]. *IEEE Communications Letters*, 2017, 21(7): 1481–1484. doi: [10.1109/LCOMM.2017.2690678](https://doi.org/10.1109/LCOMM.2017.2690678).
- [10] KUO J J, YANG H H, and TSAI M J. Optimal approximation algorithm of virtual machine placement for data latency minimization in cloud systems[C]. The IEEE Conference on Computer Communications, Toronto, Canada, 2014: 1303–1311. doi: [10.1109/INFOCOM.2014.6848063](https://doi.org/10.1109/INFOCOM.2014.6848063).
- [11] ZENG Deze, GU Lin, GUO Song, *et al.* Joint optimization

- of task scheduling and image placement in fog computing supported software-defined embedded system[J]. *IEEE Transactions on Computers*, 2016, 65(12): 3702–3712. doi: [10.1109/TC.2016.2536019](https://doi.org/10.1109/TC.2016.2536019).
- [12] GU Lin, ZENG Deze, GUO Song, *et al.* Cost efficient resource management in fog computing supported medical cyber-physical system[J]. *IEEE Transactions on Emerging Topics in Computing*, 2017, 5(1): 108–119. doi: [10.1109/TETC.2015.2508382](https://doi.org/10.1109/TETC.2015.2508382).
- [13] 张海波, 刘香渝, 荆昆仑, 等. 车联网中基于NOMA-MEC的卸载策略研究[J]. *电子与信息学报*, 2021, 43(4): 1072–1079. doi: [10.11999/JEIT200017](https://doi.org/10.11999/JEIT200017).
ZHANG Haibo, LIU Xiangyu, JING Kunlun, *et al.* Research on NOMA-MEC-based offloading strategy in internet of vehicles[J]. *Journal of Electronics & Information Technology*, 2021, 43(4): 1072–1079. doi: [10.11999/JEIT200017](https://doi.org/10.11999/JEIT200017).
- [14] 张海波, 李虎, 陈善学, 等. 超密集网络中基于移动边缘计算的任务卸载和资源优化[J]. *电子与信息学报*, 2019, 41(5): 1194–1201. doi: [10.11999/JEIT180592](https://doi.org/10.11999/JEIT180592).
ZHANG Haibo, LI Hu, CHEN Shanxue, *et al.* Computing offloading and resource optimization in ultra-dense networks with mobile edge computation[J]. *Journal of Electronics & Information Technology*, 2019, 41(5): 1194–1201. doi: [10.11999/JEIT180592](https://doi.org/10.11999/JEIT180592).
- [15] YANG Song, LI Fan, SHEN Meng, *et al.* Cloudlet placement and task allocation in mobile edge computing[J]. *IEEE Internet of Things Journal*, 2019, 6(3): 5853–5863. doi: [10.1109/JIOT.2019.2907605](https://doi.org/10.1109/JIOT.2019.2907605).
- [16] AGRYZKOV T, TORTOSA L, VICENT J F, *et al.* A centrality measure for urban networks based on the eigenvector centrality concept[J]. *Environment and Planning B: Urban Analytics and City Science*, 2019, 46(4): 668–689. doi: [10.1177/2399808317724444](https://doi.org/10.1177/2399808317724444).
- [17] WANG Yamin, CAO Zhiying, ZHANG Xiuguo, *et al.* Clustering-based algorithm for services deployment in mobile edge computing environment[C]. The 2019 IEEE 25th International Conference on Parallel and Distributed Systems, Tianjin, China, 2019: 963–966. doi: [10.1109/ICPADS47876.2019.00145](https://doi.org/10.1109/ICPADS47876.2019.00145).
- [18] CHIN T L, CHEN Y S, and LYU K Y. Queuing model based edge placement for work offloading in mobile cloud networks[J]. *IEEE Access*, 2020, 8: 47295–47303. doi: [10.1109/ACCESS.2020.2979479](https://doi.org/10.1109/ACCESS.2020.2979479).
- [19] KNIGHT S, NGUYEN H X, FALKNER N, *et al.* The internet topology zoo[J]. *IEEE Journal on Selected Areas in Communications*, 2011, 29(9): 1765–1775. doi: [10.1109/JSAC.2011.111002](https://doi.org/10.1109/JSAC.2011.111002).
- [20] ZHAO Lei, LIU Jiajia, SHI Yongpeng, *et al.* Optimal placement of virtual machines in mobile edge computing[C]. 2017 IEEE Global Communications Conference, Singapore, 2017: 1–6. doi: [10.1109/GLOCOM.2017.8254084](https://doi.org/10.1109/GLOCOM.2017.8254084).
- [21] JIA Mike, CAO Jiannong, and LIANG Weifa. Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks[J]. *IEEE Transactions on Cloud Computing*, 2017, 5(4): 725–737. doi: [10.1109/TCC.2015.2449834](https://doi.org/10.1109/TCC.2015.2449834).
- 李光辉: 男, 1970年生, 教授, 博士生导师, 研究方向为边缘计算、物联网。
周 辉: 男, 1996年生, 硕士生, 研究方向为移动边缘计算。
胡世红: 女, 1993年生, 博士生, 研究方向为边缘计算。

责任编辑: 马秀强