

MEC 中基于改进遗传模拟退火算法的 虚拟网络功能部署策略

陈卓^{1,2}, 冯钢², 刘怡静², 周杨³

(1. 重庆理工大学计算机科学与工程学院, 重庆 200433; 2. 电子科技大学通信抗干扰技术国家级重点实验室, 四川 成都 710077;
3. 奥本大学计算机科学与软件工程学院, 奥本 36849)

摘要: 为了有效改善多集群共存的移动边缘网络中业务流端到端服务时延, 提出了一种基于改进遗传模拟退火算法的虚拟网络功能部署策略。通过开放 Jackson 排队网络对移动业务流的时延进行最优化建模, 在证明其 NP 性的基础上提出了将遗传算法与模拟退火算法相结合的求解策略, 该策略通过对服务节点的提前映射机制避免了可能带来的网络拥塞, 并通过个体的约束性判断和纠正遗传的方法避免了局部最优的出现。在不同的服务请求量、服务节点规模、集群数量及虚拟网络功能之间的逻辑连接关系等参数下的对比实验表明, 该策略能提供更低时延的端到端服务, 使时延敏感类移动业务获得更好体验。

关键词: 移动边缘计算; 虚拟网络功能; 服务时延; 遗传模拟退火算法

中图分类号: TP393.0

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020074

Virtual network function deployment strategy based on improved genetic simulated annealing algorithm in MEC

CHEN Zhuo^{1,2}, FENG Gang², LIU Yijing², ZHOU Yang³

1. College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 200433, China

2. National Key Laboratory of Science and Technology on Communications,
University of Electronic Science and Technology of China, Chengdu 710077, China

3. Department of Computer Science and Software Engineering, Auburn University, Auburn 36849, USA

Abstract: In order to effectively improve the end-to-end service delay of the flow in multi-clusters coexisting mobile edge computing (MEC) network, a virtual network function deployment strategy based on improved genetic simulated annealing algorithm was proposed. The delay of mobile service flow was mathematically modeled through the open Jackson queuing network. After proving the NP attribute of this problem, a solution combining genetic algorithm and simulated annealing algorithm was proposed. In this strategy, the advance mapping mechanism avoids the possibility of network congestion, and the occurrence of local optima was avoided through using the methods of individual judgment and corrective genetic. Extensive simulation was set up to evaluate the effectiveness of the proposed strategy under different parameter settings, such as different volume of requests, different scale of service nodes, different number of MEC clusters, and logical link relationships between virtual network functions. Results show that this strategy can provide lower end-to-end services delay and better service experience for latency-sensitive mobile application.

Key words: mobile edge computing, virtual network function, service delay, genetic simulated annealing algorithm

收稿日期: 2019-05-07; 修回日期: 2020-03-16

基金项目: 国家自然科学基金资助项目 (No.61471089, No.61401076); 重庆市技术创新与应用发展基金资助项目 (No.cstc2018jszx-cyztzx0088)

Foundation Items: The National Natural Science Foundation of China (No.61471089, No.61401076), The Technology Innovation and Application Development Project of Chongqing (No.cstc2018jszx-cyztzx0088)

1 引言

移动边缘计算 (MEC, mobile edge computing) 作为一种新的云服务模式, 将传统集中式部署和管理的云端资源分布式地部署至无线接入网 (RAN, radio access network), 使移动业务就近得到处理, 从而获得良好业务体验^[1-2]的同时降低了回程网络的网络负载^[3]。将网络功能虚拟化 (NFV, network function virtualization) 技术^[4]应用于 MEC, 运营商能将提供服务所需的 IT 资源以虚拟网络功能 (VNF, virtual network function) 的形式快速实例化, 这进一步提升了 MEC 的服务弹性。将服务节点成簇互联形成集群化的 MEC 网络^[5-7]能根据业务所需的网络功能类型及业务量的变化情况动态调整 VNF 的实例化规模, 使业务流尽可能地在 MEC 集群内完成端到端的服务从而达成就近高效服务的目标。但集群化的 MEC 的部署需要结合移动应用的请求位置分散且对资源的需求动态变化等特点, 同时还存在着单个 MEC 集群的 IT 资源受限、MEC 集群之间的网络资源受限及多个 VNF 需根据业务类型进行逻辑关联等诸多限制, 因此, 在集群化 MEC 网络中合理部署 VNF 和进行业务流传输路径的优选, 为移动业务提供最优的端到端服务时延颇具挑战。目前, 缺乏针对性的研究工作, 亟需深入探讨。

在集群化的 MEC 网络中, 本文以为时延敏感类移动业务提供低时延服务为目标, 基于开放 Jackson 排队网络建立了业务流的端到端时延的数学优化模型。通过将该优化问题归结为一个二维背包问题 (2KP, two-dimensional knapsack problem), 从而证明其 NP 性, 进一步提出了一种集群化部署的 MEC 网络中通过合理部署 VNF 及业务流路径选择的策略——iGSA (improved-genetic and simulated annealing), 该策略结合了遗传算法和模拟退火算法分别在全局解和局部解的搜索能力的优势, 通过对服务节点的提前映射机制避免了在节点部署时可能带来的 MEC 网络拥塞, 同时通过个体的约束性判断和纠正遗传的方法避免了局部最优的出现。在多个网络场景下的对比实验表明, iGSA 策略均能在单个 MEC 集群内或多个 MEC 集群之间, 通过优化改善 VNF 部署和业务流的路径选择提供更低的业务流端到端时延, 有效地改善了移动业务的体验。

2 相关研究工作

在集群化部署的 MEC 网络中提供低时延服务

所面临的挑战及具有智能特征的算法在大规模系统中快速求得优化解方面所具有的独特优势形成了本文研究方法的依据。与本文相关的工作可按照 MEC 网络的部署场景和 VNF 的部署方法进行分类讨论。

在 MEC 网络场景类似的工作介绍如下^[5-7]。文献[5-6]将 NFV 引入边缘节点并建立起虚拟化 MEC 网络架构, 通过构建服务功能链 (SFC, service function chain) 为移动边缘应用提供就近的 IT 弹性服务。文献[5]侧重于通过 SFC 实现缓存服务改善移动业务的体验, 而文献[6]则着重从主动故障恢复机制设计方面, 探讨提升虚拟化 MEC 网络的系统可靠性问题。文献[5-7]在集群化部署的 MEC 网络场景下, 力图通过计算最优的 MEC 集群数量来提高业务流的服务质量。但局限在独立的 MEC 集群中开展研究, 缺乏对更一般化的 MEC 部署场景中的相关问题进行探讨。另外, 和中心化的云服务相比较, MEC 最显著的优势是能够为用户就近提供低时延服务, 因此, 本文从 MEC 的主要功能特性出发, 研究并提出改善集群化部署的 MEC 网络中端到端服务时延的策略。针对 VNF 的优化部署方法, 相关工作参考文献[8-12]。其中文献[8-10]针对不同类型的业务请求, 分别从效用最大化、能耗最低及能够容纳的业务流最大化等角度进行了研究, 并提出了将预先定义了次序的多个 VNF 进行链接部署, 建立起 SFC 的策略。文献[11]通过优化多个微服务提供服务的时序, 以达到改善移动应用服务体验的目的。文献[12]则借助强化学习方法研究了 VNF 部署过程中虚拟节点到物理节点的优化映射和实例化的问题。上述研究工作^[8-12]的网络场景是 IT 资源(节点计算资源、节点存储资源和网络带宽)相对充裕的数据中心网络或移动核心网络, 其关注的重点集中在有效提升虚拟化的 IT 资源使用效率问题或提高整个系统服务能力的问题上, 所采用的方法通常是将多个物理指标转化为统一的系统开销和系统收益, 并建立优化模型加以分析。与之相比, 本文所研究的处于网络边缘的 MEC 节点 IT 资源相对稀缺, 网络系统在提供低时延端到端服务的同时, 合理分配 IT 资源显得尤其重要。而这涉及在 MEC 集群内和 MEC 集群之间的 VNF 优化部署及业务流虚拟路径的合理选择。

与已有工作相比较, 本文的创新性主要体现在以下 2 个方面。1) 在多个 MEC 集群共存的边缘网络场景下, 面向移动业务请求位置分散及对 IT 资

源需求动态变化的特点,通过排队网络模型对端到端的服务时延进行了形式化分析并建立最优化模型。2) 分析求证了 1) 中优化问题的 NP 性,并提出了一种易于部署的快速求解策略——iGSA。该策略通过将遗传算法和模拟退火算法合理的结合,在进行全局最优解搜索的同时有效地提高了求解效率并降低了运行时间,这对在大规模 MEC 网络中进行快速的 VNF 部署决策具有积极的借鉴意义。

3 模型化分析

3.1 网络场景描述及形式化定义

本文考虑集群化部署的 MEC 网络场景。如图 1 所示,MEC 集群化部署在移动通信网络边缘且和一个或多个 eNode B 连接^[6],MEC 集群通过 PDN-GW (packet data network gateway)和云化的 5G 移动核心网连接。一个 MEC 集群可以包括若干个虚拟化 MEC 节点。另外,MEC 集群之间通过网络连接,具备多个 MEC 集群之间协作的能力。与云化的数据中心网络或移动核心网中的 IT 资源可近乎认为无限不同的是,MEC 集群中节点数量和能提供的 IT 资源都是受限的。MEC 集群优先在本集群内完成对请求业务的服务,当资源无法满足时,则利用其他 MEC 集群的可用 IT 资源构建新的 VNF 完成服务。整个 MEC 集群的资源统计和分配由位于移

动核心网中的网络控制器实现^[13]。集群化的 MEC 部署方式,能够跨越多个 eNode B 和网络区域,为时延敏感类移动业务提供端到端的低时延服务,这对智能车/无人驾驶这类时延敏感类应用尤其重要。

将一个集群化部署的 MEC 网络定义为 $G = \{G_1, \dots, G_g\}$, 其中 g 为集群的数量, G_n 表示第 n 个 MEC 集群。定义一个无向图 $G_n = (V_n, E_n)$, 其中 V_n 和 E_n 分别为 MEC 集群 G_n 中的边缘节点和集群内网络链路。 (u, w) 表示 2 个边缘节点 u 和 w 之间的链路,此外, u 和 w 可以属于同一个或不同的集群 MEC。 $l_{u,w}$ 表示链路 (u, w) 的可用网络带宽资源,边缘节点 u 和 w 的距离表示为 $D_{u,w}$ 。 $n_v (v \in V)$ 表示用于构建 VNF 的通用服务节点(即虚拟机)数量。 M_n 表示在 MEC 集群 G_n 中由边缘节点经虚拟化后的通用服务节点集合,对于 MEC 集群 G_n 中某个通用服务节点 $m (m \in M_n)$, 当前可用计算资源表示为 W_m^n 。集群化部署的 MEC 网络为各类移动业务提供服务,以 H 表示 MEC 网络在时间 T 内收到 h 个移动服务请求, $H = \{d_1, d_2, \dots, d_h\}$ 。对于服务请求 $d_i (d_i \in H)$ 的入口节点和出口节点分别用 I_i 和 E_i 表示, I_i 到 E_i 的路径表示为 P_i , d_i 的数据率为 R_i 。根据不同移动应用业务的需求,多个 VNF 按照某种次序从逻辑上连接成一种串行结构或并行结构的 SFC^[13], 其中采用并行连接有助在 MEC 网络中提高时延敏感类

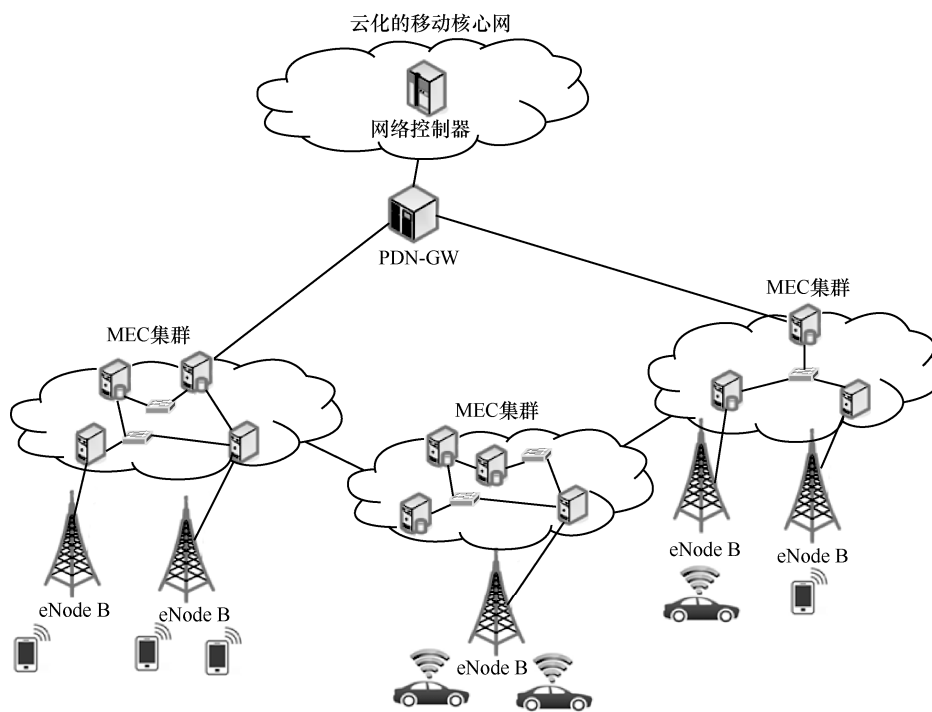


图 1 集群化部署的 MEC 网络框架

移动业务的服务效率。如图 2 所示, 为服务请求 d_i 提供服务的 SFC 表示为 $S_i = \{S_{i,1}, S_{i,2}, S_{i,3}, \dots, S_{i,K}\}$, 其长度表示为 $|S_i|$, 其中 $S_{i,j} (1 \leq j \leq K)$ 表示 SFC 上的某一个 VNF 或经并行连接后为 d_i 提供服务的多个 VNF 组成的集合, 例如图 2 中的 $S_2 = \{S_{2,1}, S_{2,2}, S_{2,3}, S_{2,4}\}$, $S_{2,1} = \{f_{21}^1\}$, $S_{2,3} = \{f_{23}^1, f_{23}^2\}$ 。假设在时间 T 内在 MEC 网络中能建立不同类型的 VNF, 用集合 F 表示。对于某一类型的 VNF $f (f \in F)$ 能最多被实例化建立 $|N_f|$ 个, 定义 f_k 为建立类型为 VNF $f (f \in F)$ 的第 k 个实例, 其计算资源占用量表示为 $\omega_{f_k}^i$, $\beta_{f_k, f_{k'}}^i$ 表示 f_k 和 $f_{k'}$ 之间的通信所占用网络带宽, $\forall f, f' \in F, 1 \leq k \leq |N_f|, 1 \leq k' \leq |N_{f'}|$ 。定义矩阵 B 为 G 中各条链路的带宽占用量。定义 $I(S_{i,j})$ 为 d_i 提供服务的 SFC 的路径 P_i 中对应的通用服务节点的索引。对于 d_i 的业务流按指定的顺序遍历多个 VNF, 即 $I(S_{i,j}) \leq I(S_{i,j'})$, $\forall S_{i,j}, S_{i,j'} \in S_i, j < j'$ 。

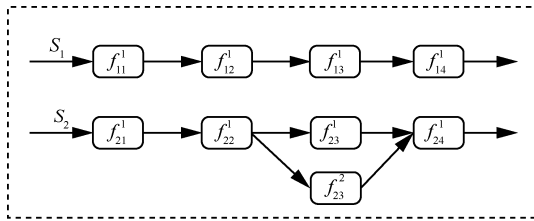


图 2 MEC 集群中串行或并行逻辑连接的多个 VNF

3.2 服务时延模型

移动业务通过集群化 MEC 网络实现端到端服务的过程中产生的时延包括业务流数据分组在 VNF 处等待处理的排队时延、业务流数据分组接受 VNF 的处理时延及业务流数据分组在 MEC 集群内和 MEC 集群之间传输产生的时延。特别说明的是, 对于 VNF 在通用服务节点之上的运行部署可能带来不同影响程度的处理时延, 在评估集群化部署的 MEC 网络中端到端服务时延时必须将其纳入进行考虑^[14-15]。

首先业务流从入口节点到出口节点需经过一条 SFC 中的多个 VNF 进行处理, 每个 VNF 在处理业务流时将产生处理时延。本文基于 M/M/1/c 类型的 Jackson 排队网络^[15]对业务流在 MEC 网络中的处理时延进行建模, 将一个 MEC 网络中的 VNF 视作服务节点, 假设业务数据分组到达该 VNF 的过程是一个泊松过程。对于 VNF f_k , 定义 $\lambda_{f,k}$ 和 $u_{f,k}$ 分别

表示其平均到达率和平均服务率, 且 $\rho_{f,k} = \frac{\lambda_{f,k}}{u_{f,k}}$ 。为保证 VNF 服务的系统稳定性, 有 $\rho_{f,k} < 1$ 。对于 MEC 网络中的某个 VNF, 输入流量来自入口节点直接导入, 也可能来自同集群或邻居集群的 VNF 流量输出, 因此, 定义 P_{jw} 表示业务流在 VNF $_j$ 完成处理并输出至 VNF $_w$ 的概率, 定义 λ_w^0 表示从 MEC 入口节点到 VNF $_w$ 的流量。VNF $_w$ 的输入流量速率可表示为

$$\lambda_w = \lambda_w^0 + \sum_{j=1}^{n_v} \lambda_j P_{jw}, \quad i=1,2,\dots,n_v \quad (1)$$

则在 VNF f_k 的缓存队列中的数据分组平均数可表示为

$$M_{f,k} = \frac{\rho}{1-\rho} = \frac{\lambda_{f,k}}{u_{f,k} - \lambda_{f,k}}, \quad \forall f \in F \quad (2)$$

VNF f_k 可运行部署在节点 w 之上, 则 f_k 的服务率导入至物理节点 w 可表示为 $u_{f_k} = R(w)$, 其中 $R(w)$ 由对业务流的传输能力决定, 更详细的计算方法可参考文献[15]。由 Little 定理^[16]可得到 VNF

$$f_k \text{ 的处理时延 } T_{f,k} = \frac{M_{f,k}}{\lambda_{f,k}}, \text{ 而 } \frac{M_{f,k}}{\lambda_{f,k}} = \frac{\lambda_{f,k}}{\sum_{i \in H} \lambda_i Z_{f,k}^i}, \text{ 则}$$

有

$$T_{f,k} = \frac{\lambda_w}{\sum_{i \in H} R_i Z_{f,k}^i \varepsilon_{f,k}^{n,w}} \quad (3)$$

其中, λ_i 表示第 i 条业务流的速率, $Z_{f,k}^i = 1$ 表示 d_i 需要 f_k 提供服务, $Z_{f,k}^i = 0$ 则表示 d_i 不需要 f_k 提供服务。则对于移动业务请求 d_i 的处理时延可表示为

$$t_1^i = \sum_{i \in H} \sum_{f \in F} \sum_{k=1}^{|N_f|} Z_{f,k}^i T_{f,k} \quad (4)$$

基于 M/M/1/c 排队网络进行模型化分析, 单个 VNF 的平均排队时延可表示为

$$t_{\text{queue}}^{f,k} = \sum_{j=1}^c \frac{j q_j}{\mu} \quad (5)$$

其中, q_j 表示当有 j 个用户到达系统时的平滑概率^[16], 则业务请求的排队时延可表示为

$$t_2^i = \sum_{i \in H} \sum_{f \in F} \sum_{k=1}^{|N_f|} Z_{f,k}^i t_{\text{queue}}^{f,k} \quad (6)$$

本文定义了一个由二元决策变量组成的矩阵 \mathbf{A} , 矩阵中的任意元素为 $\varepsilon_{f,k}^{n,u}$, 且有 $\varepsilon_{f,k}^{n,u} \in \{0,1\}$, $\{\varepsilon_{f,k}^{n,u} \mid f \in F, 0 < k \leq |N_f|\}$, $\varepsilon_{f,k}^{n,u} = 1$ 表示业务流 f_k 经由 MEC 集群 n 中的通用服务节点 u 进行处理, $\varepsilon_{f,k}^{n,u} = 0$ 表示业务流 f_k 没有流经 MEC 集群 n 中的通用服务节点 u . 另外, 定义二元决策变量 $\sigma_{i,f,k,f',k'}^{u,w}$, $\sigma_{i,f,k,f',k'}^{u,w} = 1$ 表示 $\beta_{f,k,f',k'}^i$ 映射到底层网络链路 $l_{u,w}$, $\sigma_{i,f,k,f',k'}^{u,w} = 0$ 表示 $\beta_{f,k,f',k'}^i$ 没有映射到底层网络链路 $l_{u,w}$.

业务请求的传播时延可表示为

$$t_3^i = \frac{\sum_{i \in H} \sum_{(u,w) \in E} D_{u,w} \sigma_{i,f,k,f',k'}^{u,w}}{p} \quad (7)$$

其中, p 表示信号在物理链路上的传输速率, $D_{u,w}$ 表示物理链路的长度, E 表示整个 MEC 集群内的网络链路组成的整体集合.

在集群化部署的 MEC 收到多个业务请求的情况下, 对于存在各种资源限制的集群化部署 MEC 网络中, 本文通过 VNF 的部署和业务流路径的选择优化业务流端到端服务时延. 该最优化模型可表示为

$$\min \sum_{f,k} t_j^i \quad (8)$$

$$\text{s.t.} \quad \sum_{i \in H} \sum_{f \in F} \sum_{f' \in F} \sum_{k=1}^{|N_f|} \sum_{k'=1}^{|N_{f'}|} \beta_{f,k,f',k'}^i \sigma_{i,f,k,f',k'}^{u,w} \leq l_{u,w}, \forall (u,w) \in E \quad (9)$$

$$\sum_{i \in H} \sum_{f \in F} \sum_{f' \in F} \sum_{k=1}^{|N_f|} \sum_{k'=1}^{|N_{f'}|} R_i \sigma_{i,f,k,f',k'}^{u,w} \leq l_{u,w}, \forall u \in E_n, w \in E_m, m \neq n \quad (10)$$

$$\sum_{i \in H} \sum_{f \in F} \sum_{k=1}^{|N_f|} \omega_{f,k}^i \varepsilon_{f,k}^{n,u} \leq W_u^n, \forall u \in V, 1 \leq n \leq g \quad (11)$$

$$\sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{p,u} - \sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{u,w} = 0 \quad (12)$$

$\forall i \in H, \forall p, w \notin \{I_i, E_i\}, \forall u, w \in V$

$$\sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{u,E_i} - \sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{E_i,u} = 1 \quad (13)$$

$$\forall i \in H, u \neq E_i$$

$$\sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{I_i,u} - \sum_{f \in F} \sum_{k=1}^{|N_f|} \sum_{f' \in F} \sum_{k'=1}^{|N_{f'}|} \sigma_{i,f,k,f',k'}^{u,I_i} = 1 \quad (14)$$

$\forall i \in H, u \neq I_i$

$$\varepsilon_{f_k}^{n,u} \varepsilon_{f_{k'}}^{n',w} = \sigma_{i,f_k,f_{k'}}^{u,w}, 1 \leq n, n' \leq g, u, w \in V \quad (15)$$

$$\sum_{j=1}^3 t_j^i \leq t^i, \forall 1 \leq i \leq h \quad (16)$$

$$\sigma_{i,f_k,f_{k'}}^{u,w} \in \{0,1\}, \forall i \in H, \forall f_k, f_{k'} \in F, \forall (u,w) \in E \quad (17)$$

$$\varepsilon_{f_k}^{n,u} \in \{0,1\}, 1 \leq n \leq g, \forall u \in V, \forall f_k \in F \quad (18)$$

该优化模型中, 式(9)表示同一个 MEC 集群中的 2 个通用服务节点间的链路带宽资源限制. 式(10)表示 MEC 集群之间的链路带宽资源限制. 式(11)表示 MEC 集群 n 中的节点 u 上当前已部署的一个或多个 VNF 占用的计算资源不能超过其计算资源总量. 式(12)表示 MEC 网络中除入口节点和出口节点之外所有活动节点需要满足流量守恒. 式(13)和式(14)表示服务请求仅从一个入口节点进入 MEC 网络且仅从一个出口节点离开集群化部署的 MEC 网络. 式(15)要求一条 SFC 业务流需经过其预定义的所有 VNF. 式(16)表示业务请求的时延约束. 式(17)和式(18)分别表示需要具备可行的链路和节点映射.

3.3 NP 性讨论

在运筹学领域, 多维背包问题 (MKP, multi-dimensional knapsack problem) 是一个经典的优化问题, 其求解目标是在满足各项资源约束的前提下, 从候选对象集中找出可以使目标价值达到最大 (或最小) 的对象子集. 该问题已被证明是一种 NP-Hard^[17] 问题.

在 3.2 节建立的最优化模型中, 集群化的 MEC 网络同时为多条业务流提供端到端服务, 并为业务流服务的多个 VNF 需求同时占用计算处理资源和转发业务流的链路带宽资源. 由于 VNF 部署于 MEC 网络中的通用服务节点之上, 因此这 2 种资源的占用分别不能超过通用服务节点的可用计算资源和通用服务节点之间的物理链路带宽 (包括 MEC 集群内和 MEC 集群间的链路带宽). 假设 MEC 集群中能为第 i 条业务流提供服务的通用服务节点为 n 个, $r_{1,j}$ 表示为业务流提供服务需占用第 j 个通用服务节点的计算资源, $r_{2,m}$ 和 $r_{3,m}$ 分别表示为业务流提供服务第 j 个通用服务节点需占用的 MEC 集群内和 MEC 集群间的带宽资源. 定义 x_j 表示通用服务节点 x_j 是否被选中用于部署为业务流提供服务的 VNF. 由于业务流的端到端服务需经过多个部署了 VNF 的通用服务节点, 因此定义 p_j 表示当业务流通

过通用服务节点 x_j 产生的时延, 则式(8)所描述的最优化问题可简化为 $\omega = \min \sum_{j=1}^n p_j x_j$, 满足约束条件 I:

$\sum_{j=1}^n r_{m,j} x_j \leq b_m (m=1,2,3)$ 和约束条件 II: $x_j \in \{0,1\} (j=1,2,\dots,n)$, $x_j=0$ 表示通用服务节点被选中, $x_j=1$ 表示通用服务节点未被选中。若进一步将服务业务流对于 MEC 集群内和 MEC 集群间的链路带宽资源占用视作一类资源, 则式(8)简化后的最优化问题是一个 2KP。根据 MKP 问题的 NP 性, 因此本文所描述的在集群化部署的 MEC 网络中的业务流端到端时延最小化问题也是一个 NP-Hard 问题。

4 基于遗传模拟退火算法的部署策略

第 3 节所定义的问题难以在多项式时间复杂度内找到全局最优解。而当问题的规模较大时, 若采用贪心法或者枚举法等精确算法的运行时间代价较高, 很难在集群化 MEC 网络中进行实际部署。因此需要采用相应的近似算法对其进行求解。而以模拟退火算法^[18]和遗传算法^[19]为代表的智能算法近年在多个领域得到了广泛有效的应用。其中, 模拟退火算法模拟固体物质退火过程的热平衡问题与随机搜索寻优问题的相似性来达到寻找全局最优或近似全局最优的目的。若在模拟退火算法的运行过程中融入遗传算法, 称为遗传模拟退火算法^[20-21]。

本文所研究的问题从本质上是根据业务流类型和性能约束条件, 按照某种预定义顺序将多个 VNF 映射到多个 MEC 通用服务节点之上, 实现时延最优的离散优化问题。在求解该优化问题时, 需要权衡好求解方法的执行效率和求解质量。本文所提的 iGSA 策略做了 2 个规定: 1) 依据业务请求到达 MEC 网络的时间先后顺序进行分析, 如果多个业务请求同时到达, 则针对这些请求所形成的业务流的总时延为优化目标; 2) 当一个 MEC 群集中已实例化的 VNF 资源不足以满足服务需求时, 则采用在当前 MEC 集群中新开启通用服务节点并实例化 VNF 或将业务流引导至相邻 MEC 集群。iGSA 策略设计时在 2 个方面做了性能改进, 一方面在 K 最短路径的选择和业务流的引导之前进行在通用服务节点的映射 (即虚拟机的映射), 以避免网络拥塞; 另一方面对不符合约束的个体加以纠正, 然后将纠正的个体放入下一代个体中, 以避免陷入局

部最优。iGSA 策略的主要步骤包括编码、选择复制、交叉、变异和可行性检测。

4.1 编码

设第 k 代种群中的个体数目为 N , 这里每个个体即为 $n_v \times m_f$ 染色体矩阵, m_f 表示需要使用多少个 VNF, n_v 表示当前已经开启的通用服务节点数量。矩阵 $\mathbf{Q}_k = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k\}$ 表示个体, 第 k 代种群的第 r 个个体表示为

$$\mathbf{A}_k^r = \begin{bmatrix} a_{11}^r & a_{12}^r & \cdots & a_{1m_f}^r \\ a_{21}^r & a_{22}^r & \cdots & a_{2m_f}^r \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_v,1}^r & a_{n_v,2}^r & \cdots & a_{n_v,m_f}^r \end{bmatrix}$$

\mathbf{A}_k^r 的每个元素为一个二元值, 其中 $a_{ij}^r=1$ 表示第 j 个 VNF 被部署在通用服务节点 i 之上, 而通用服务节点是在第 k 代种群中的第 r 个个体; $a_{ij}^r=0$ 表示第 j 个 VNF 没有被部署在通用服务器节点 i 上。 \mathbf{A}_k^r 的限制条件为 $\sum_{i=1}^{n_v} a_{ij}^r=1$ 和 $a_{ij}^r \in \{0,1\}, \forall i \in \{1,2,\dots,n_v\}, j \in \{1,2,\dots,m_f\}$ 。

4.2 选择复制

当以概率值 $P(\mathbf{A}_k^r)$ 对个体进行选择复制时, 本文采用 K 最短路算法。根据此个体 \mathbf{A}_k^r , 检测当前系统是否存在满足所有业务流时延要求的路径, 如果不存在, 则选择一个满足时延和链路带宽限制的个体 \mathbf{A}_k^r 并纠正当前个体 \mathbf{A}_k^r , 将纠正后的 \mathbf{A}_k^r 放入种群的下一代种群中; 如果存在, 直接复制该个体 \mathbf{A}_k^r 到下一个代种群中。个体 \mathbf{A}_k^r 被选中的概率 $P(\mathbf{A}_k^r)$ 和适应度成正比, 可表示为

$$P(\mathbf{A}_k^r) = \frac{\text{fit}(\mathbf{A}_k^r)}{\sum_{i=1}^N \text{fit}(\mathbf{A}_k^i)} \quad (19)$$

其中, $\text{fit}(\mathbf{A}_k^r)$ 表示个体 \mathbf{A}_k^r 的适应度函数。本文采用的遗传模拟退火算法在适应度函数定义时引入了温度参数。相比之下, 单纯的遗传算法不涉及温度参数变化, 而遗传模拟退火则通过引入温度参数, 使随着温度降低种群个体跳出局部最优平均次数增加。采用的适应度函数为

$$\text{fit}(\mathbf{A}_k^r) = \exp \left\{ -\frac{\text{fun}(\mathbf{A}_k^r) - \text{fun}(\mathbf{A}_k^{\min})}{t_k} \right\} \quad (20)$$

其中, t_k 表示状态 k 下的温度, 而初始的温度值定义为 $t_0 = Z\pi_0$, Z 为一个常量, $\pi_0 = \text{fun}(A_k^r)_{\max} - \text{fun}(A_0^r)_{\min}$, 其中 $\text{fun}(A_k^r) = \sum_{i=1}^h \sum_{j=1}^3 t_j^i$ 。

4.3 交叉

本文使用多行矩阵杂交, 例如使 Q_k 中的 A_k^i 和 A_k^j 进行配对, 并且将 A_k^i 和 A_k^j 相对应的行以概率 $P_c = 0.6$ 进行互换。种群中 2 个染色体之间的交叉过程如下

$$A_k^i = \begin{bmatrix} a_{11}^i & a_{12}^i & \cdots & a_{1m_f}^i \\ a_{21}^i & a_{22}^i & \cdots & a_{2m_f}^i \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_v,1}^i & a_{n_v,2}^i & \cdots & a_{n_v,m_f}^i \end{bmatrix}$$

$$A_k^j = \begin{bmatrix} a_{11}^j & a_{12}^j & \cdots & a_{1m_f}^j \\ a_{21}^j & a_{22}^j & \cdots & a_{2m_f}^j \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_v,1}^j & a_{n_v,2}^j & \cdots & a_{n_v,m_f}^j \end{bmatrix}$$

在 A_k^i 和 A_k^j 的第二行中, 进行交叉操作后的染色体矩阵分别为

$$A_k^i = \begin{bmatrix} a_{11}^i & a_{12}^i & \cdots & a_{1m}^i \\ a_{21}^j & a_{22}^j & \cdots & a_{2m}^j \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_v,1}^i & a_{n_v,2}^i & \cdots & a_{n_v,m}^i \end{bmatrix}$$

$$A_k^j = \begin{bmatrix} a_{11}^j & a_{12}^j & \cdots & a_{1m}^j \\ a_{21}^i & a_{22}^i & \cdots & a_{2m}^i \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_v,1}^j & a_{n_v,2}^j & \cdots & a_{n_v,m}^j \end{bmatrix}$$

4.4 变异和可行性检测

本文假设突变的概率 $P_m=0.01$ 。对于任何个体, 需要判断是否发生了突变。如果发生了突变, 就需要对该个体进行可行性检测。

在交叉和变异操作之后, 需要判断这些新个体是否满足资源限制条件。对于一个给定的染色体, 可以得到 VNF 映射策略, 从中可进一步获得决策变量 $\varepsilon_{f,k}^{n,v}$ 的值, 从而判断是否满足通用服务节点映射的约束。如果不满足, 则节点不满足计算资源限制, 并且将对应于这些节点的行元素从 1 随机校正为 0, 直到满足节点计算约束。通过 VNF 的映射可

以获得相应的最优网络链路映射, 然后为每条业务流在 K 条最短路径中选择合适的路径匹配, 在满足链路和时延约束的同时最小化适应度函数。以这种方式, 在第一阶段中将纠正的个体和链接映射策略重复至下一代。在集群化部署的 MEC 网络中, iGSA 策略的时间复杂度为 $O(n^2)$, 其中 n 表示通用服务节点的数量。对于通用服务节点规模较小的 MEC 网络而言, 该算法的计算复杂度在可接受的程度内。iGSA 策略的伪码如算法 1 所示。

算法 1 基于遗传模拟退火的启发式策略——iGSA

输入 集群化部署的 MEC 网络通用服务节点和链路信息, 各条链路的带宽占用量矩阵 B , 某个 MEC 集群 G_n 中某个通用服务节点 m ($m \in M_n$) 当前可用计算资源 W_m^n , VNF 的类型集合 F , SFC 集合 (包括 SFC 长度、SFC 类型和 R_i)。

输出 SFC 的部署策略 M_{best} 和 B_{best} , 以及对于在 $[t, t + T]$ 时间间隔内到达的业务请求的端到端时延预估值 t_{avg}

步骤 1 记录 f 、 $\omega_{f,k}^i$ 、 β_{f_k, f_k}^i 、 $Z_{f_k}^i$, 记录类型为 f 的 VNF 能够部署的数量 $|N_f|$

步骤 2 初始化 $\text{fit}_{\min} \leftarrow 0$, $m \leftarrow 0$, $M_{\text{best}} \leftarrow \emptyset$, t_k 和循环次数 K

步骤 3 以种群大小 N 、交叉概率 P_c 、变异概率 P_m 、温度范围 Temp_{\min} 和温度变化系数 ξ 为初始值初始化种群 Q_0

步骤 4 计算 $P(A_k^r)$, 在 Q_0 中找到最大适应度 fit 值, 并记录对应的 M_{best} 和 B_{best}

步骤 5 以概率 $P(A_k^r)$ 在 Q_m 中选择复制个体, 重复 N 次得到 Q'_m

步骤 6 while($m' < K$ && $t_k \geq \text{Temp}_{\min}$)

```
{
for  $A_m^i \in Q'_m$  do
    得到  $\varepsilon_{f,k}^{n,v}$ ;
```

得到满足链路带宽限制和时延限制的 K 最短路径, 优先在同一个 MEC 集群中选择计算适应度 $\text{fit}(A_m^i)$ 并排序, 得到 fit_{\min} ;

得到 VNF 的部署策略 $M_{\text{best}} \leftarrow A_m^k$;

得到链路映射策略 $B_{\text{best}} \leftarrow M_{\text{best}}$;

```
end
```

```
for  $i, i' \in Q'_m$  do
```

将 A_m^i 和 A_m^j 以概率 P_c 交叉处理得到 $A_{m_new}^i$ 和 $A_{m_new}^j$, 并用 A_{new} 表示交叉后得到的新个体;

```

对  $A_{new}$  进行检测;
    if 可行测试失败
        纠正个体;
    end
    计算适应度  $fit(A_{new})$ 
end
for  $A_m^i \in Q_m$  do
    if (突变产生)
        检测和纠正个体;
    end
end
 $t_k \leftarrow \xi t_k$ ;
 $Q_{m+1} \leftarrow Q_m$ ;
 $m' \leftarrow m + 1$ ;
}

```

步骤7 输出集群化部署的 MEC 网络中 VNF 的部署 M_{best} 和路径选择 B_{best} , 以及平均时延预估值 t_{avg} 。

5 实验与性能评估

5.1 实验方法及仿真参数设置

本文通过 Matlab 建立数值仿真环境评估 iGSA 策略的性能。实验基于 Congent^[22]生成 MEC 集群网络的拓扑。实验生成了 2 种类型的网络(网络 I 和网络 II)以评估算法在不同 MEC 集群网络规模下的表现, 其中网络 I 包括 3 个 MEC 集群, 部署 30 个通用服务节点、55 条集群内和集群间的链路, 集群内的带宽资源参数在 10~200 Mbit/s 内随机选取, 集群间的带宽资源参数在 10~100 Mbit/s 内随机选取。网络 II 的 MEC 集群数量为 10 个, 部署 200 个通用服务节点、355 条集群内和集群间的网络链接, 集群内的带宽资源参数在 1~10 Gbit/s 内随机选取, 集群间的带宽资源参数在 200 Mbit/s~1 Gbit/s 内随机选取。2 种网络中的服务功能链 SFC 的平均长度为 4。SFC 的类型、MEC 网络中通用服务节点的计算资源、不同类型 VNF 所需计算资源和 VNF 之间的关联关系等参数从各自的区间内随机选择。与文献[23]类似, 集群化部署的 MEC 网络节点之间的传播时延与其链路

距离成比例, 通过乘以在区间[0.8,1.5]中取得的随机数引入适当的随机性。本文参考文献[24], 服务请求所需的流量服从幂率分布, 设置 $\alpha = 2.1$ 产生了 $x_{min}=10$ Mbit/s 的服务请求。随机地在 MEC 集群中选择一个服务请求的入口网元和出口网元以模拟不同类型的移动业务。另外, 突变概率为 0.01, MEC 群集之间和内部的交叉概率分别设置为 $P_c=0.6$ 和 $P_c=0.8$, 温度变化系数 $\xi=0.45$ 。

实验中使用以下几种典型策略进行对比以客观评估 iGSA 策略的性能。1) AH 策略 (AH strategy)^[25], 该算法在一个 MEC 网络中优先选择具有最多剩余资源的可用节点部署 VNF 和建立 SFC。和 iGSA 策略相比较, AH 策略减少了 VNF 的处理时延, 但没有考虑处理时延和传播时延的整体优化, 特别没有考虑多个 MEC 集群共存的一般化场景。2) 贪心策略 (greedy strategy): 集群化部署的 MEC 网络逐个处理时间 T 内的服务请求, 并依次最小化每个服务请求的端到端时延。3) 随机策略 (random strategy) 在时间 T 内收到 i 个服务请求, 该策略在满足计算资源限制、链路带宽资源限制和服务请求的时延限制的通用服务节点用于部署 VNF 和建立 SFC。

5.2 实验结果与分析

本文首先探讨在同一时间段内, 到达集群化部署 MEC 网络中的服务请求数与平均端到端服务时延之间的关系。图 3 显示了在网络 I 场景下, 获得服务的平均时延与服务请求数量之间的关系。从图 3 可以看到, 服务的平均时延随着服务请求的数量增加而增加, 这主要是因为随着业务请求的增加, 当一个 MEC 集群中的计算和网络资源不足以开启部署新的 VNF 时, 业务流将被引导至相邻 MEC 集群, 而跨 MEC 的带宽资源受限从而导致服务的整体时延增加。和典型的 AH 策略、贪心策略和随机策略相比, 基于遗传模拟退火算法的 iGSA 策略在降低服务时延方面有更好的表现。同时由于贪心策略考虑了多个 MEC 集群共存的情况, 优先在同一个 MEC 集群中选择通用服务节点以降低服务时延, 因此其表现优于 AH 策略。在网络 I 场景中, 相较于其他策略, iGSA 策略都具有稳定的性能表现, 分别比贪心策略、AH 策略及随机策略有平均 10.62%、23.94%和 71.36%的端到端服务时延优势。

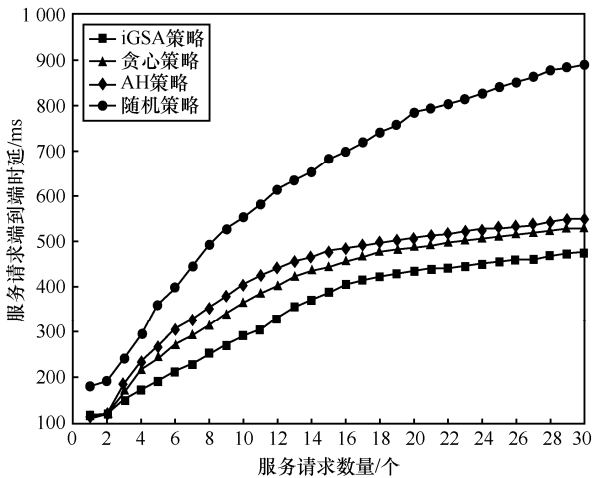


图 3 不同服务请求量下的服务时延对比(网络 I)

为业务请求提供持续高质量的服务,多个不同类型的 VNF 从逻辑上链接成 SFC 或服务功能图 (SFG, service function graph), 本文还深入探讨了 SFC 的长度与服务请求的平均时延之间的关系。从图 4 中可以观察到, 服务请求的平均时延随着 SFC 长度 (即一条 SFC 链上的 VNF 的个数) 的增加而同步增加, 其主要原因是 SFC 越长, 表示业务流经的 VNF 实例也越多, 则在网络资源固定的情况下需要更多的计算、网络带宽及存储资源, 从而导致可用的节点和链路变得更稀缺。本文在 3.1 节图 2 中所描述的并行逻辑连接的多个 VNF 构成了 SFG, 该逻辑结构是一种特殊的 SFC。和 SFC 的主要区别在于, SFG 通过分析业务流经过的多个 VNF 之间的依赖关系, 引入 VNF 的并行化执行思想实现对业务流的更高效服务, 基于 SFG 模式的业务流服务, 能显著降低业务的端到端时延^[26]。实验场景中设置了部分服务请求不能以并行化只能以串行化 SFC 的方式提供服务, SFC 和 SFG 的总数量为 100。在一个 SFC 或一个 SFG 中, 业务流需经过 8 个 VNF 的处理才能完成端到端的服务。该实验通过增加 SFG 的数量来评估对于服务请求的端到端平均时延的影响。从图 5 中可以发现, 在 SFG 的数量从 10 增到 100 的过程中, 即 SFG 的占比增加, 而 SFC 的占比下降, 服务请求的平均时延减少。其主要原因是当 SFG 占比增加时, 越来越多的 VNF 都在并行处理业务流数据, 完成端到端的服务时延减少。该实验结论表明, 本文所提的 iGSA 策略能很好地适应采用不同的 VNF 逻辑链接关系建立的 MEC 应用。

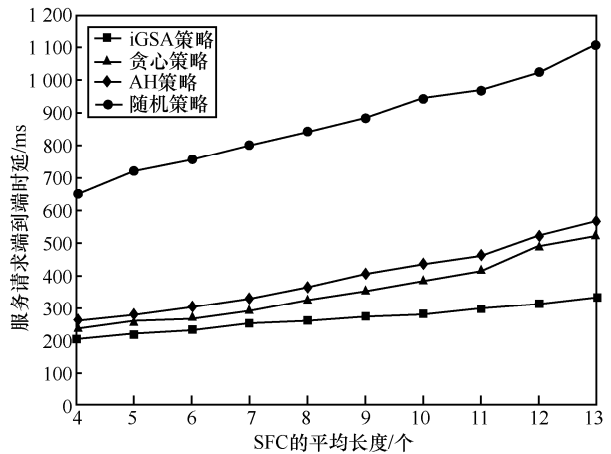


图 4 不同 SFC 长度下的服务时延对比(网络 II)

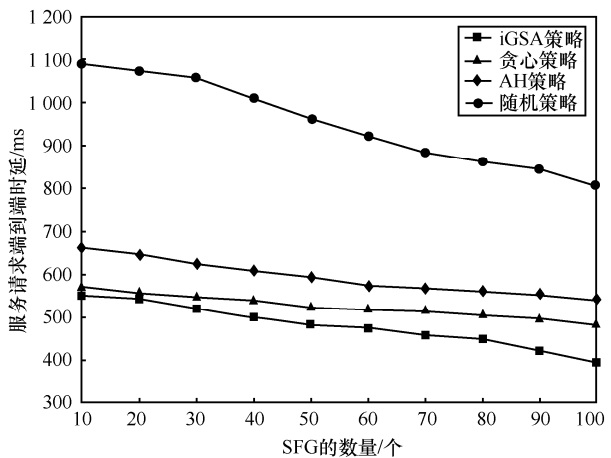


图 5 不同 SFG 下的服务时延对比(网络 II)

将 NFV 技术引入 MEC 网络后, MEC 能够根据服务的请求量动态调整通用服务节点的数量以实现在服务质量和资源开销之间的平衡。本文评估了当 MEC 集群中根据业务请求开启不同规模的通用服务节点时, 不同算法的端到端服务时延对比, 并比较了在网络 II 的 10 个 MEC 群集中从开启 80 个通用服务节点到 260 个通用服务节点时端到端的平均服务时延, 从图 6 可以观察到, 平均时延随着 MEC 集群中开启的通用服务节点数量的增加而下降, 其原因主要是随着每个 MEC 集群开启的通用服务节点数量的增加, 实例化的 VNF 数量也同步增加, 服务请求更多地可以在入口网元所在的 MEC 集群中得到处理, 而不需要导入至其他相邻的 MEC 集群。进一步地可以看到, 在一个 MEC 集群中开启相同的通用服务节点的情况下, 得益于 iGSA 策略能够在多个 MEC 集群中进行跨区域的路径选择, 该策略总能提供更低的端到端服务时延。相对于集中式部署的云

数据中心或云化的移动核心网, MEC 的计算及网络等 IT 资源都相对受限,因此本文继续对比了 IT 资源受限情况下不同的策略所提供的端到端服务时延,包括通用服务节点的计算资源以及 MEC 集群之间的平均链路带宽受限。

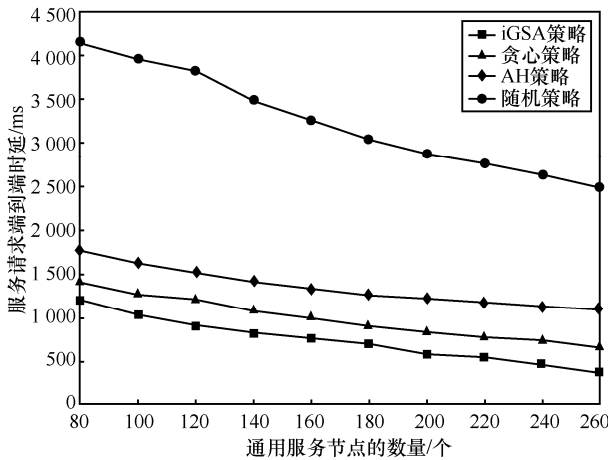


图6 MEC 集群中不同的通用服务节点规模下的服务时延对比(网络 II)

图7和图8分别对比了 MEC 集群之间的带宽资源和 MEC 集群计算资源处于不同稀缺情况下的业务流端到端时延。MEC 集群之间的带宽资源从 200 Mbit/s 增加到 1 Gbit/s, 服务请求数为 100 个。当 IT 资源增加时, 服务请求的平均服务时延逐渐降低。这是因为计算资源增加后更多的 VNF 可以在本地 MEC 集群中的通用服务节点生成。同时, 由于 MEC 集群之间的链路带宽增加, 可以将业务流所需的更多的计算和网络资源引导至相邻 MEC 集群。可以看到, 随着 MEC 集群的计算资源的逐渐增加, 更多的服务请求将在本地集群中得到处理, 服务的端到端路径将有效缩短。在图7和图8中, 当 MEC 的计算资源处于不同稀缺程度时, 相对于其他策略, iGSA 均能取得更好的性能。在不同的带宽资源稀缺程度下, iGSA 策略分别比贪心策略和 AH 策略平均有 13.32% 和 48.76% 的性能优势, 在不同计算资源稀缺程度下, iGSA 策略则分别比贪心策略和 AH 策略平均有 22.72% 和 35.29% 的性能优势。

上述实验对比中, 通过设置不同的参数, 包括服务请求数量、MEC 网络中服务节点规模、MEC 集群数量及 VNF 之间的逻辑连接关系等, 详细对比了 iGSA 策略和几个相关策略性能。实验结果表明, iGSA 策略通过将模拟退火和遗传算法相结合,

在保证算法效率的同时获得了更优的求解方案, 使时延敏感类移动业务获得更好体验。实验结论有力地支撑了本文通过改进遗传模拟退火算法解决 VNF 在 MEC 网络中的优化部署问题上所做的创新。

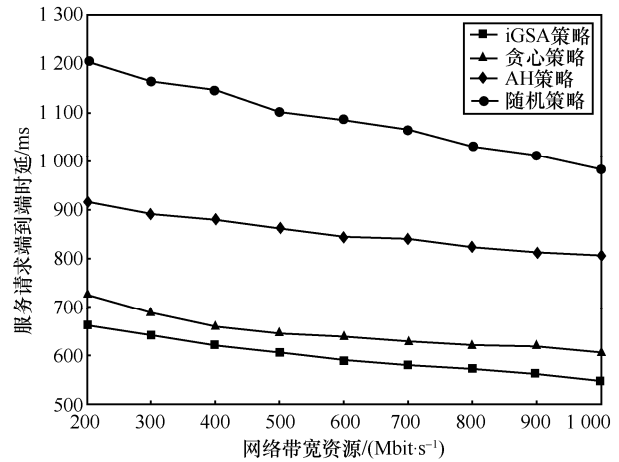


图7 不同 MEC 集群间的网络带宽资源下服务时延对比(网络 II)

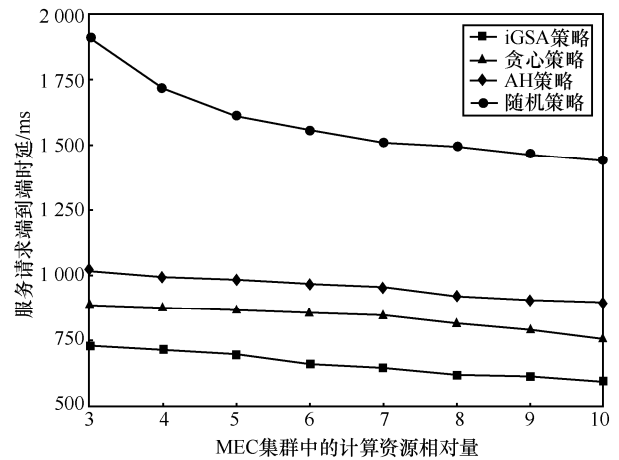


图8 不同 MEC 集群计算资源下的服务时延对比(网络 II)

6 结束语

本文研究了多集群 MEC 网络中的 VNF 的优化部署策略, 首先将业务流经过部署在 MEC 通用服务节点的多个 VNF 的过程形式化为一个开放 Jackson 排队网络, 并进一步得到业务流的服务时延最优化模型。在证明了该优化问题是一个 NP-Hard 问题的基础上, 通过将遗传算法和模拟退火算法结合, 提出一种 MEC 集群网络中基于遗传模拟退火算法的 VNF 部署及路径优选策略。通过在不同参数条件下, 将所提策略与类似策略进行了详细对比, 结果表明所提出策略能为移动业务提供更低的端到端服务时延, 有效改善业务的体验。本文所提

出的算法及结论能为优化 MEC 的资源部署提供有借鉴意义的参考。在将来的工作中, 将把本文的场景继续扩展到融合网络切片的 MEC 网络, 研究 MEC 节点和网络切片中的虚拟资源部署和联合调度。

参考文献:

- [1] LIU H, ELDARRAT F, ALQAHTANI H, et al. Mobile edge cloud system: architectures, challenges, and approaches[J]. IEEE Systems Journal, 2018, 12(3): 2495-2508.
- [2] ESSWIE A A, PEDERSEN K I. Opportunistic spatial preemptive scheduling for URLLC and eMBB coexistence in multi-user 5G networks[J]. IEEE Access, 2018(6):38451-38463.
- [3] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3):1628-1656.
- [4] CHATRAS B, OZOG F F. Network functions virtualization: the portability challenge[J]. IEEE Network, 2016, 30(4):4-8.
- [5] LEI L, XIONG X, HOU L, et al. Collaborative edge caching through service function chaining: architecture and challenges[J]. IEEE Wireless Communications, 2018, 25(3):94-102.
- [6] HUANG H W, GUO S. Proactive failure recovery for NFV in distributed edge computing[J]. IEEE Communications Magazine, 2019, 57(5): 131-137.
- [7] NAM Y, SONG S, CHUNG J M. Clustered NFV service chaining optimization in mobile edge clouds[J]. IEEE Communications Letters, 2017, 21(2): 350-353.
- [8] GU S J, LI Z P, WU C, et al. An efficient auction mechanism for service chains in the NFV market[C]// IEEE International Conference on Computer Communications. Piscataway: IEEE Press, 2016:1-9.
- [9] XU Z C, ZHANG X N, YU S, et al. Energy-efficient virtual network function placement in telecom networks[C]//IEEE International Conference on Communications. Piscataway: IEEE Press, 2018:1-7.
- [10] MA W, SANDOVAL O, BELTRAN J, et al. Traffic aware placement of interdependent NFV middle-boxes[C]//IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017:1-9.
- [11] BHAMARE D, SAMAKA M, ERBAD A, et al. Multi-objective scheduling of micro-services for optimal service function chains[C]//IEEE International Conference on Communications. Piscataway: IEEE Press, 2017:1-6.
- [12] 魏亮, 黄韬, 张娇, 等. 基于强化学习的服务链映射算法[J]. 通信学报, 2018, 39(1): 90-100.
WEI L, HUANG T, ZHANG J, et al. Service chain mapping algorithm based on reinforcement learning[J]. Journal on Communications, 2018, 39(1): 90-100.
- [13] YOUSAF F Z, BREDEL M, SCHALLER S, et al. NFV and SDN- key technology enablers for 5G networks[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(11):2468-2478.
- [14] ZENG C, LIU F, CHEN S, et al. Demystifying the performance interference of co-located virtual network functions[C]// IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 765-773.
- [15] LIU Z, YANG M, DAI J. Performance improvement based on path delay analysis in WiMax mesh networks[C]//Proceedings of the Second International Conference on Communications and Networking in China. Piscataway: IEEE Press, 2007: 958-962.
- [16] KHINCHIN A Y, ANDREWS D M, QUENOUILLE M H. Mathematical methods in the theory of queuing[M]. North Chelmsford: Courier Corporation, 2013.
- [17] CHU P C, BEASLEY J E. A genetic algorithm for the multidimensional knapsack problem[J]. Journal of Heuristics, 1998, 4(1): 63-86.
- [18] LIN C C, SHU L, DENG D J. Router node placement with service priority in wireless mesh networks using simulated annealing with momentum terms[J]. IEEE Systems Journal, 2016, 10(4):1402-1411.
- [19] HOU N, HE F Z, ZHOU Y, et al. A parallel genetic algorithm with dispersion correction for HW/SW partitioning on multi-core CPU and many-core GPU[J]. IEEE Access, 2017(6): 883-898.
- [20] YANG L Y, DENG Y H, YANG L T, et al. Reducing the cooling power of data centers by intelligently assigning tasks[J]. IEEE Transactions on Automation Science and Engineering, 2018,5(3):1667-1678.
- [21] 孟凡超, 初佃辉, 李克秋, 等. 基于混合遗传模拟退火算法的 SaaS 构件优化放置[J]. 软件学报, 2016, 27(4):916-932.
MENG F C, CHU D H, LI K Q, et al. Solving SaaS components optimization placement problem with hybrid genetic and simulated annealing algorithm[J]. Journal of Software, 2016, 27(4): 916-932.
- [22] JIA Y, WU C, LI Z, et al. Online scaling of NFV service chains across geo-distributed datacenters[J]. IEEE/ACM Transactions on Networking (TON), 2018,26(2):699-710.
- [23] JIA Y, WU C, LI Z, et al. Online scaling of NFV service chains across geo-distributed datacenters[J]. IEEE/ACM Transactions on Networking, 2018, 26(2): 699-710.
- [24] LI X, QIAN C. Low-complexity multi-resource packet scheduling for network function virtualization[C]// IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2015: 1400-1408.
- [25] XIA M, SHIRAZIPOUR M, ZHANG Y, et al. Network function placement for NFV chaining in packet/optical datacenters[J]. Journal of Lightwave Technology, 2015, 33(8): 1565-1570.
- [26] SUN C, BI J, ZHENG Z L, et al. NFP: enabling network function parallelism in NFV[C]// The Conference of the ACM Special Interest Group on Data Communication. New York : ACM Press, 2017:43-56.

[作者简介]



陈卓 (1980-), 男, 重庆人, 博士, 重庆理工大学副教授、硕士生导师, 主要研究方向为网络虚拟化、物联网应用及网络性能分析和评估。



冯钢 (1964-), 男, 重庆人, 博士, 电子科技大学教授、博士生导师, 主要研究方向为无线通信网络、网络虚拟化及网络资源分配。

刘怡静 (1994-), 女, 四川成都人, 电子科技大学硕士生, 主要研究方向为无线电通信网络、网络虚拟化。

周杨 (1975-), 男, 重庆人, 博士, 奥本大学助理教授、博士生导师, 主要研究方向为最优算法、大数据技术。