

DOI: 10.7652/xjtuxb201804017

一种 5G 网络低时延资源调度算法

王琛^{1,2}, 汤红波^{1,2}, 游伟^{1,2}, 王晓雷^{1,2}, 袁泉^{1,2}

(1. 国家数字交换系统工程技术研究中心, 450002, 郑州; 2. 移动互联网安全技术国家工程实验室, 100876, 北京)

摘要: 针对现有资源调度算法难以满足 5G 低时延业务需求的现状,提出了一种基于联合遗传和禁忌搜索算法的资源调度(GATS)算法。首先利用整数线性规划建立了虚拟链路的动态带宽分配策略,然后在传统柔性车间调度模型的基础上引入了数据流量在虚拟链路中的传输时延,建立了相应的 5G 网络资源调度模型。为了求解这一调度模型,设计了采用联合遗传和禁忌搜索算法的启发式调度算法,该算法通过在遗传算法寻优过程中引入禁忌搜索,平衡全局搜索和局部搜索能力,有效解决了遗传算法早熟的问题,而且能够获得更好的调度方案。仿真实验表明,与 GA-BA 算法相比,GATS 算法将服务完成时间减少了 17%,不仅满足了 5G 低时延业务的需求,而且提高了用户体验和移动运营商的收益。

关键词: 5G 网络;网络功能虚拟化;调度;带宽分配;遗传算法;禁忌搜索

中图分类号: TN915.81 **文献标志码:** A **文章编号:** 0253-987X(2018)04-0117-08

A Resource Scheduling Algorithm with Low Latency for 5G Networks Based on Effective Hybrid Genetic Algorithm and Tabu Search

WANG Chen^{1,2}, TANG Hongbo^{1,2}, YOU Wei^{1,2}, WANG Xiaolei^{1,2}, YUAN Quan^{1,2}

(1. National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China;

2. National Engineering Laboratory for Mobile Network Security, Beijing 100876 China)

Abstract: A resource scheduling algorithm based on hybrid genetic algorithm and tabu search (named GATS) is proposed to solve the problem that the existing schedule methods are difficult to meet the requirement of the mobile communication with low latency. First, a dynamic bandwidth allocation policy of virtual links is established using an integer linear programming. Then, the transmission delay of data traffic in virtual links is introduced based on a traditional flexible job shop scheduling model, and the corresponding resource scheduling model for 5G is established. Owing to the complexity of the scheduling problem, the resource scheduling algorithm based on hybrid genetic algorithm and tabu search is developed for solving the problem efficiently. The algorithm introduces tabu search in optimization process of the genetic algorithm to balance capabilities of global and local searches, solves the problem of premature convergence of the genetic algorithm, and obtains better scheduling solutions. Simulation results show that the GATS algorithm outperforms the GA-BA algorithm in reducing the scheduling makespan by 17%, and caters to 5G service with stringent delay requirements, thereby increases users' experience and operators' revenues.

收稿日期: 2017-06-20。 作者简介: 王琛(1988—),男,硕士生;汤红波(通信作者),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(61521003);国家重点基础研究发展规划资助项目(2016YFB0801605)。

网络出版时间: 2018-02-26

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20180226.1945.004.html>

Keywords: 5G network; network function virtualization; scheduling; bandwidth allocation; genetic algorithm; tabu search

5G 移动通信网将是一个全方位服务、多技术融合的网络,通过信息通信技术的演进和创新,来满足未来用户多样化的业务需求^[1]。以软件定义网络 (software defined network, SDN) 和网络功能虚拟化 (network function virtualization, NFV) 为代表的网络新技术将成为 5G 网络的关键技术^[2]。NFV 技术使网络资源分配更加灵活,实现了异构的网络架构和服务的聚合^[3]。在传统的网络架构中,网络功能通过昂贵的硬件中间件来实现,当移动网络运营商部署新业务时,需要大量的资本支出和运营支出。而在 NFV 环境下,网络功能通过实例化在商用服务器上的虚拟机 (virtual machine, VM) 中的软件化的虚拟网络功能 (virtual network function, VNF) 来实现,网络服务则是由多个虚拟网络功能以一定的依赖关系构成服务功能链 (service function chaining, SFC) 来完成。移动网络运营商通过 NFV 技术能够有效地减少资本支出和运营支出,为用户提供先进和灵活的网络服务^[4],而 SDN 作为一种新型的网络架构,能够控制数据流通过相应的网络功能以执行相应的策略,实现灵活的网络管理^[5]。SDN 和 NFV 的结合为虚拟网络功能的协调控制和调度提供了有效的架构,促进了网络的创新^[6]。在网络虚拟化下,资源分配分为服务功能链的构建、服务功能链的映射和虚拟网络功能调度这 3 个阶段。虚拟网络功能调度作为 NFV 环境下资源分配的第 3 个阶段,主要解决在给定相应的计算资源约束和虚拟网络功能执行顺序的基础上,为多个服务功能链中的每个虚拟网络功能在虚拟机中选择相应的实例化时间,从而达到最小化完成所有网络服务总时间的目的^[7]。

IMT-2020 将增强移动宽带、海量机器通信和高可靠低时延通信作为 5G 的主要应用场景^[8]。为了使移动终端完美支持交互式游戏、3D 虚拟现实等在线交互应用,5G 提出了毫秒级时延和 10 Gbit/s 峰值速率的要求^[9]。因此,设计高效的虚拟网络功能调度方法是将 NFV 应用在 5G 当中的关键问题之一。目前仅有少量的文献研究了虚拟网络功能调度问题。文献[10]将虚拟网络功能调度问题定义为柔性车间调度问题 (flexible job shop problem, FJSP),然而未提出多项式时间内的解决方案,而且在其调度模型里仅考虑了虚拟网络功能的处理时

延。文献[11]提出了 NFV 环境下具有较低复杂度的多资源数据包调度算法,该算法考虑了数据包队列的特性,然而未考虑服务功能链的特性。文献[12]对服务功能链的端到端时延进行了分析,然而其假定时延具有固定的值,这在实际中并不适用,不能应用在动态的网络服务中。文献[13]在 FJSP 模型的基础上引入数据流量在虚拟链路上传输时延,建立了虚拟网络功能调度问题的混合整数线性规划模型,并设计了基于遗传算法的虚拟网络功能调度方法,但其提出的遗传算法容易早熟,陷入局部最优,无法获得更好的调度方案。虚拟化技术将物理服务器虚拟化为多个虚拟机节点,这些虚拟机节点之间由虚拟链路相互连接^[14],流入服务功能链中的数据流量在具有有限带宽的虚拟链路中传输时会带来不可忽视的传输时延,而完成一个服务功能链需要利用多个虚拟机,这些虚拟机可以放置在位于一个数据中心或多个数据中心的物理实体上,瓶颈传输链路带来的时延将会降低整个网络服务的用户体验,因此虚拟网络功能调度问题需要考虑高效的带宽资源分配策略。文献[15]阐明了因为缺少对虚拟机的带宽保证,传统的传输层协议是不公平的,其将带宽分配建模为协作博弈模型,设计了相应带宽分配策略。文献[16]针对现有的带宽分配方案不能有效应对高动态的流量变化状况,通过考虑大量短流和突发流的影响,在控制论框架下基于符号逻辑模型设计了一种分布式速率分配算法。

针对以上研究的不足,本文通过考虑虚拟网络功能实例化在虚拟机上的处理时延和数据流量在虚拟链路上传输时延,建立了相应的虚拟网络功能调度模型,并设计了动态链路带宽分配策略,提出了基于联合遗传和禁忌搜索 (GATS) 的虚拟网络功能调度方法。该算法在带宽分配上具有更高的细粒度,充分利用了遗传算法的全局搜索能力和禁忌搜索的局部搜索能力,有效的避免了早熟问题,具有更快的收敛速度,可以得到更优的调度方案。

1 网络模型与问题建模

1.1 网络模型

定义服务功能链集合 S , 对于每个服务功能链 $s_i (i \in [1, S])$, 具有一组数据流量需要以固定顺序通过的虚拟网络功能; $F_i (i \in [1, S])$ 为服务功能链

s_i 中虚拟网络功能的集合; f_{ij} 表示服务功能链 s_i 中的第 j 个虚拟网络功能; N 为虚拟网络中虚拟机的集合; L 为虚拟网络中虚拟链路的集合。假设每个虚拟机节点具有实例化不同虚拟网络功能的能力, 且在同一时间段内至多实例化一个虚拟网络功能, $M_{f_{ij}} (i \in [1, S], j \in [1, F_i])$ 为具有实例化虚拟网络功能 f_{ij} 能力的虚拟机节点集合, 即计算资源约束。本文考虑了 2 种类型的时延, 即服务功能链的数据流通过虚拟机之间的虚拟链路的传输时延和虚拟网络功能实例化在虚拟机中的处理时延。假定每个服务功能链从流入端到流出端具有固定的流量需求, 定义为 $D_i (i \in [1, S])$, 2 个虚拟机之间的虚拟链路的带宽为 $B_{k_1, k_2} (k_1, k_2 \in [1, N])$, 因此, 服务功能链 s_i 在虚拟机节点 k_1 和 k_2 之间虚拟链路的传输时延为 $D_i / B_{k_1, k_2}$ 。 $\tau_{k, f_{ij}} (k \in [1, N])$ 为虚拟网络功能 f_{ij} 实例化在虚拟机节点 k 上的处理时延。

1.2 问题建模

本文针对多个具有计算资源约束的服务功能链, 求解在虚拟网络上完成所有服务功能链请求的最短时间。定义 t_{ij} 表示虚拟网络功能 f_{ij} 开始处理到达数据流的时刻, $t'_{ij} (i \in [1, S], j \in [1, F_i - 1])$ 表示虚拟网络功能 f_{ij} 向流出链路开始转发数据流的时刻。定义二值变量 $x_{ijk}, x_{ijk} = 1$ 表示虚拟网络功能 f_{ij} 实例化在虚拟机节点 k 上; 反之, $x_{ijk} = 0$ 。通过考虑虚拟网络功能调度中的 2 种时延, 即虚拟网络功能实例化在虚拟机上的处理时延和虚拟网络功能之间的数据流在虚拟链路上的传输时延, 建立了相应的虚拟网络功能调度模型, 其优化目标表示为

$$\delta = \max_{i \in [1, S], j \in [1, F_i]_z} \left\{ t_{ij} + \sum_{k \in M_{f_{ij}}} x_{ijk} \tau_{k, f_{ij}} \right\} \quad (1)$$

式(1)的约束条件为

$$\left. \begin{aligned} t_{ij} + \sum_{k \in M_{f_{ij}}} x_{ijk} \tau_{k, f_{ij}} &\leq t'_{ij} \\ i &\in [1, S]_z, j \in [1, F_i - 1]_z \end{aligned} \right\} \quad (2)$$

$$\left. \begin{aligned} t'_{ij-1} + \sum_{k_1 \in M_{f_{ij-1}}} \sum_{k_2 \in M_{f_{ij}}} x_{ij-1k_1} x_{ijk_2} \frac{D_i}{B_{k_1, k_2}} &\leq t_{ij} \\ i &\in [1, S]_z, j \in [2, F_i]_z \end{aligned} \right\} \quad (3)$$

$$\left. \begin{aligned} \sum_{k \in M_{f_{ij}}} x_{ijk} &= 1 \\ i &\in [1, S]_z, j \in [1, F_i]_z \end{aligned} \right\} \quad (4)$$

式中: δ 是所有服务功能链中最后一个虚拟网络功能的处理完成时间, 即多个网络服务的完成时间。式(1)表示虚拟网络功能调度的目标, 即最小化所有服务功能链请求的最大完成时间; 式(2)确保了当服务功能链中的数据流量在虚拟网络功能处理完成后, 相应的虚拟网络功能才能开始转发数据流; 式(3)确保了只有从上行虚拟机中接收到数据流后, 下行虚拟机才能开始处理相应的数据流; 式(4)确保了有且仅有一个虚拟机实例化虚拟网络功能 f_{ij} 。

1.3 带宽分配策略

合理的带宽分配策略能够有效的降低数据流在虚拟链路上的传输时延。定义 $C_{in}(k)/C_{out}(k)$ 为虚拟机节点 k 的流入/流出带宽。根据流经虚拟机的数据流量动态的分配带宽资源, 将带宽分配建模为整数线性规划模型, 其优化目标表示为

$$\begin{aligned} \min \eta \\ \eta = \sum_{i \in [1, S]_z} \sum_{j \in [1, F_i]_z} \frac{D_i}{\sum_{k_1 \in M_{f_{ij}}} \sum_{k_2 \in M_{f_{ij+1}}} x_{ijk_1} x_{ij+1k_2} B_{k_1, k_2}} \end{aligned} \quad (5)$$

式(5)的约束条件为

$$\sum_{k_2} B_{k_1, k_2} \leq C_{out}(k_1) \quad (6)$$

$$\sum_{k_1} B_{k_1, k_2} \leq C_{in}(k_2) \quad (7)$$

式中: η 表示虚拟链路总的传输时延。式(5)表示带宽动态分配的目标, 即最小化总传输时延; 式(6)确保从一个虚拟机流出的所有链路带宽之和不大干该虚拟机的流出带宽; 式(7)确保流入一个虚拟机的所有链路带宽之和不大干该虚拟机的流入带宽。

2 联合遗传和禁忌搜索算法(GATS)

根据上述虚拟网络功能调度模型, 虚拟网络功能调度问题可看做是柔性车间调度问题的扩展, 该问题是 NP 难问题, 引入动态的链路带宽分配后, 进一步增加了其复杂性, 因此在多项式时间内无法利用传统方法求得精确解^[7]。为了有效地解决虚拟网络功能调度问题, 本文设计了一种联合遗传算法 (genetic algorithm, GA) 和禁忌搜索 (tabu search, TS) 的启发式算法 (GATS), 通过利用遗传算法的全局搜索能力和禁忌搜索的局部搜索能力, 能够得到更优的调度方案。

2.1 染色体编码和解码

采用多层编码的方法对染色体进行编码, 每个染色体表示一种虚拟网络功能调度的可行解。染色

体的长度为 $2 \sum_{i \in [1, S]_z} F_i$, 对于任意染色体, 其基因被分为 2 部分, 第 1 部分表示虚拟网络功能的调度顺序, 第 2 部分表示实例化相应虚拟网络功能的虚拟机的序号。下式给出某个染色体编码示例

$$c_i = [322131231231421324] \quad (8)$$

c_i 染色体中包含 3 条服务功能链, 每条服务功能链由 3 个虚拟网络功能构成, 共有 4 个虚拟机实例化所有虚拟网络功能。前 9 位基因表示虚拟网络功能的调度顺序, 可被解码为 $f_{31} \rightarrow f_{21} \rightarrow f_{22} \rightarrow f_{11} \rightarrow f_{32} \rightarrow f_{12} \rightarrow f_{23} \rightarrow f_{33} \rightarrow f_{13}$, 后 9 位基因表示实例化前 9 位虚拟网络功能的虚拟机序号, 可解码为 $M_{f_{31}}(2) \rightarrow M_{f_{21}}(3) \rightarrow M_{f_{22}}(1) \rightarrow M_{f_{11}}(4) \rightarrow M_{f_{32}}(2) \rightarrow M_{f_{12}}(1) \rightarrow M_{f_{23}}(3) \rightarrow M_{f_{33}}(2) \rightarrow M_{f_{13}}(4)$, 其中 $M_{f_{31}}(2)$ 为可实例化虚拟网络功能 f_{31} 的虚拟机节点集合 $M_{f_{31}}$ 中的第 2 个元素。当完成染色体编码后, 需要将其解码, 以生成可行的调度方案。下列伪代码表示相应的解码算法:

```

1  输入: 染色体  $c$ 
2  输出: 可行的调度方案  $\{t_{ij}, t'_{ij}, x_{ijk}\}$ 
3  定义:  $T = \{t_{ij}\}, T' = \{t'_{ij}\}, Z = \{x_{ijk}\}$ 
4  根据  $c$  计算每个虚拟机的路由
5  根据动态链路带宽分配模型, 执行整数规划算法
   进行链路带宽分配
6  for 从前往后选择  $c$  中的基因  $o$  do
7       $i = \text{get\_service}(c(o))$ 
8       $j = \text{get\_function}(c(o))$ 
9       $k = \text{get\_VM\_node}(c(o))$ 
10      $x_{ijk} = 1$ 
11     if  $j = 1$  then
12          $t_{ij} = \text{starting\_time\_VM}(T, T', Z, k)$ 
13     else
14          $t'_{ij-1} = \text{starting\_time\_link}(T, T', Z, k)$ 
15          $t_{ij} = \text{starting\_time\_VM}(T, T', Z, k)$ 
16     end if
17 end for
```

算法的第 1、2 行表示, 由染色体得到虚拟机之间的链路分配, 根据链路带宽分配模型, 执行整数线性规划算法, 求得虚拟网络中各个虚拟机之间的链路带宽。第 3~7 行表示, 得到基因 o 的服务功能链序号、虚拟网络功能序号和分配的虚拟机序号。第 8~12 行表示, 求解得到相应虚拟网络功能 f_{ij} 在虚拟机中开始处理达到数据流的时间 t_{ij} 和在虚拟链路中开始转发数据流的时间 t'_{ij} , 最终得到可行的调

度方案。

2.2 适应度计算和选择

染色体的适应度值为所有服务功能链请求完成时间 δ 的倒数, 如式(9)所示。

$$f_{\text{fit}}(c_i) = \frac{1}{\delta} \quad (9)$$

对于任意染色体, 计算其适应度值, 然后采用轮盘赌法进行排序, 个体的选择概率如式(10)所示, 其中 N_{NIND} 表示种群个数。

$$p(c_i) = \frac{f_{\text{fit}}(c_i)}{\sum_i f_{\text{fit}}(c_i)} \quad (10)$$

2.3 交叉

种群通过交叉获得新的个体, 从而推动整个种群向前进化, 交叉操作采用整数交叉法。首先, 从种群中随机选择两个染色体, 并取出每个染色体的前 $\sum_{i \in [1, S]_z} F_i$ 位基因, 然后随机选择交叉位置进行交叉。交叉后生成新的染色体中, 一些服务功能链中会增加虚拟网络功能, 有些服务功能链会缺少虚拟网络功能, 因此, 把增加虚拟网络功能的服务功能链序号调整为缺失虚拟网络功能的服务功能链序号, 并按照交叉前相应虚拟网络功能的虚拟机分配来调整个体 ($\sum_{i \in [1, S]_z} F_i + 1$) 到 $2 \sum_{i \in [1, S]_z} F_i$ 位基因, 染色体交叉过程如图 1 所示。例如, 交叉位置为 5, 交叉 2 个染色体 c_1 和 c_2 的前 4 位基因获得 c_3 和 c_4 。交叉后染色体 c_3 上出现了 2 个问题: 一是服务功能链 s_1 缺失一个虚拟网络功能而服务功能链 s_2 上增加了一个虚拟网络功能; 二是虚拟机分配与虚拟网络功能的调度顺序不再匹配。根据以上 2 个问题, 对 c_3 做出调整, 首先将 5 号位基因, 即增加了虚拟网络功能的服务功能链 s_2 序号, 将其由 2 改为 1。然后, 和 c_2 交叉第 10 到第 13 位基因, 以还原相应虚拟网络功能所分配的虚拟机。

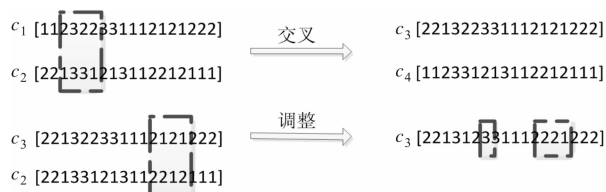


图1 染色体交叉过程

2.4 变异

首先从种群中随机选择变异个体, 然后在个体中随机选择 2 个虚拟网络功能, 最后交换这 2 个虚拟

网络功能及实例化虚拟网络功能的虚拟机基因位置。染色体变异过程如图 2 所示,首先,随机选择第 2 和第 3 位基因,然后交换该位置基因和相应虚拟机分配基因(第 11 和第 12 位基因)的顺序,完成变异。



图 2 染色体变异过程

2.5 禁忌搜索

禁忌搜索的思想是标记已搜索的局部最优解的一些对象,并在进一步的迭代搜索中尽量避开这些对象,从而保证对不同的有效搜索途径的探索。禁忌搜索涉及到邻域、禁忌表、禁忌长度、候选解、藐视准则等概念。禁忌搜索是解决调度问题的一种有效的局部搜索策略,本文采用了文献[17]提出的邻域结构,提出的禁忌搜索算法的流程如图 3 所示。

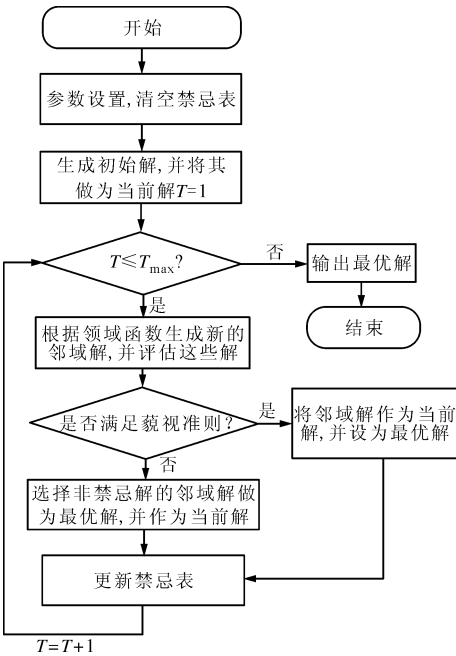


图 3 禁忌搜索算法流程图

本文提出的 GATS 算法充分利用了遗传算法的全局搜索能力和禁忌搜索的局部搜索能力,其算法流程如图 4 所示。

在提出的 GATS 算法中,选择个体执行禁忌搜索时,首先需要将其解码为一个可行的调度解,然后将这个解作为禁忌搜索的初始解,算法执行完后,将输出的最优解编码为染色体,继续执行遗传操作,当达到最大进化代数 G_{max} 时,终止 GATS 算法,当达到最大迭代次数 T_{max} 时,终止禁忌搜索算法。设置 $T_{max} = 80 \times (G/G_{max})$ 。在 GATS 算法的早期阶段,因为遗传算法不能为禁忌搜索提供较好的个体,禁

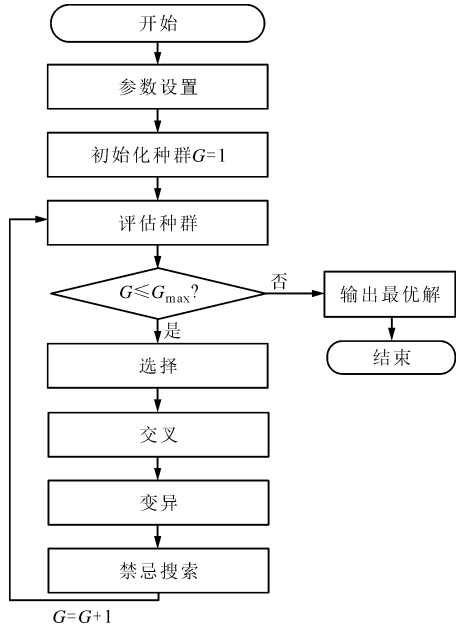


图 4 GATS 算法流程图

忌搜索算法很难找到更优的解,所以禁忌搜索算法的最大迭代次数此时较小,可以节省 GATS 算法的计算时间。在 GATS 算法的后期阶段,遗传算法可以为禁忌搜索算法提供较好的个体,因此增大禁忌搜索算法的最大迭代次数可以帮助其找到更优的解。因此,将禁忌搜索算法的最大迭代次数随着 GATS 算法的进化过程动态调节,可以有效的降低算法时间复杂度。

3 实验验证与结果分析

未来 5G 系统将采用基于 SDN/NFV 的网络架构,可以实时的掌握全局网络视图,得到算法所需的输入数据。为了评估模型的可行性及算法有效性,本文将提出的联合算法(GATS)与文献[13]中均匀链路带宽分配的遗传算法(GA-NBA)和动态带宽分配的遗传算法(GA-BA)在统一的实验平台上进行了比较。

3.1 实验设置

5G 系统将采用云数据中心架构,具有强大的并行处理能力。本文为了方便算法比较,使用了配置为 Intel Core i7-4790 3.60 GHz CPU、8 GB 内存的个人电脑,并利用 CPLEX 12.4 工具求解动态带宽分配整数规划模型,编写了算法的 Matlab 仿真程序。网络参数方面,设置最大服务功能链数量 $S = 100$,最大虚拟机数量 $N = 20$ 。虚拟网络功能在虚拟机中的处理时延在 $[2, 10]$ ms 内随机生成。虚拟网络为全连通网络,虚拟机之间的链路带宽初始化

为 2 Mb/s,虚拟机的流入/流出带宽可以由此得出。考虑移动核心网数据和控制平面服务功能链的数据流量需求,设置流经控制面服务功能链的数据流量在(0,2]kb 之间随机取值,而针对数据平面服务功能链请求,流经数据平面服务功能链的数据流量则在[10,20]kb 之间随机取值。可实例化虚拟网络功能的虚拟机节点集合随机生成,其中每个虚拟机节点实例化虚拟网络功能的数量不大于 3,任意虚拟网络功能至少可由 1 个虚拟机节点实例化。3 种算法参数统一设置为:种群数量 $N_{\text{NIND}}=40$,遗传代数 $\max G_{\text{en}}=50$,交叉概率为 80%,变异概率为 60%。

3.2 实验结果分析

首先,设置网络为全连通的具有 5 个虚拟机的底层网络和 2 条控制面服务功能链(VNF4→VNF2→VNF1→VNF3→VNF5)和 (VNF3→VNF2→VNF4→VNF5→VNF1),对 3 种算法进行了仿真实验。图 6 和图 7 分别给出了在控制面流量下 GATS 算法(本例中 GA-BA 的仿真结果与其相同)和 GA-NBA 算法的仿真结果。图中,Y 轴表示不同的虚拟机,X 轴表示时间。不同颜色表示不同的服务功能链,色块中上方的数字表示虚拟网络功能所在的服务功能链的序号和在服务功能链中的位置,下方的文字表示虚拟网络功能的类型。可以看出,2 种具有带宽分配策略的方法获得了相同的仿真结果,这

是因为控制面流量较小,传输时延对调度结果的影响较小。表 1 给出了 3 种算法在算法时间复杂度和服务完成时间上的对比。GA-NBA 算法的服务完成时间最长为 17 s。引入带宽分配策略后其他 2 种算法将服务完成时间减少了 7%。对比 GA-BA 算法和 GATS 算法的时间复杂度,由于 GATS 算法在遗传算法中加入了禁忌搜索算法,增加了算法的时间复杂度。

表 1 控制平面流量结果比较

算法	服务完成 时间/ms	算法时间 复杂度/s	优化比例/%
GA-NBA	17.0	0.43	0
GA-BA	15.8	0.84	7
GATS	15.8	1.12	7

使用相同的虚拟机网络拓扑,针对 2 条数据平面服务功能链请求运行 3 种算法,运行结果如图 8~图 10 所示。从图 8~图 10 可见,由于数据流量较大,传输时延对服务完成时间的影响更加显著。表 2 给出了 3 种算法在算法时间复杂度和服务完成时间上的对比。实验结果表明,引入链路带宽分配后,服务完成时间明显减少,GA-BA 算法和 GATS 算法分别在带宽平均分配的 GA-NBA 方法的基础

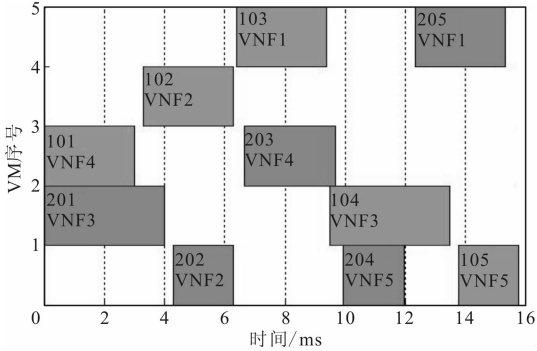


图 6 控制面流量 GATS 算法(GA-BA 算法)调度结果

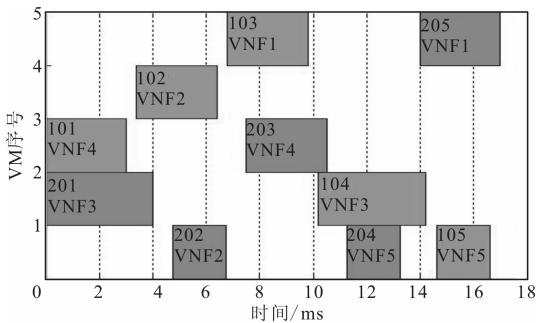


图 7 控制面流量 GA-NBA 算法调度结果

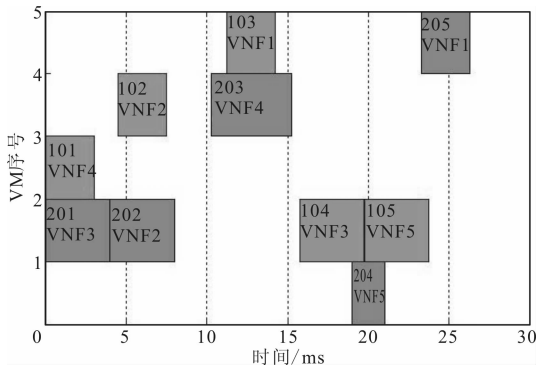


图 8 数据面流量 GA-NBA 算法调度结果

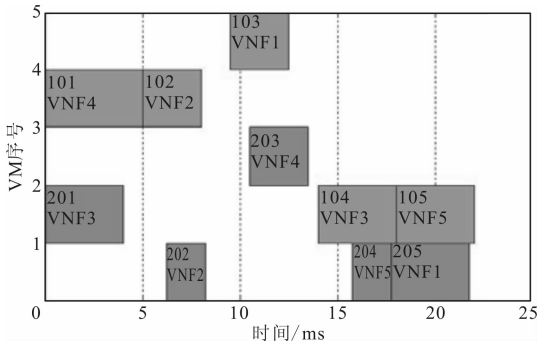


图 9 数据面流量 GABA 算法调度结果

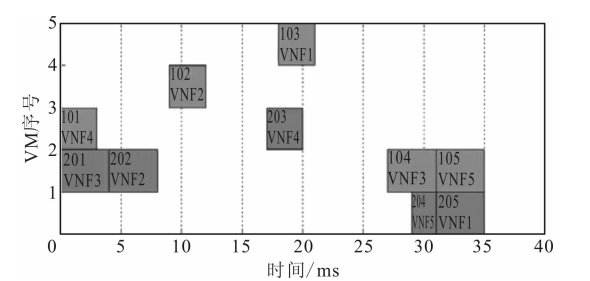


图 10 数据面流量 GATS 算法调度结果

上将服务完成时间降低了 24.3% 和 37.1%。因为本文 GATS 算法对带宽分配的细粒度更小,能够更好地带宽分配策略,因此,GATS 算法比 GA-BA 算法获得了更小的服务完成时间。通过比较表 1 和表 2 可以看出,算法的时间复杂度仅与虚拟网络的拓扑和服务功能链的数量有关。

表 2 数据平面流量结果比较

算法	服务完成 时间/ms	算法时间 复杂度/s	优化比例/%
GA-NBA	35.0	0.43	0.0
GA-BA	26.5	0.84	24.3
GATS	22.0	1.12	37.1

图 11 表示了在数据平面流量下,不同进化代数下服务完成时间的收敛过程。由图 11 可以看出,GATS 算法具有更优的收敛效率,而 GA-BA 算法则经常陷入局部最优解,这是因为,GATS 算法采用了禁忌搜索对染色体的解进行了进一步优化,具有较好的局部搜索能力,表现出较快的收敛速度,避免了早熟问题,同时采用了细粒度更高的带宽分配策略能够获得更短的服务完成时间。

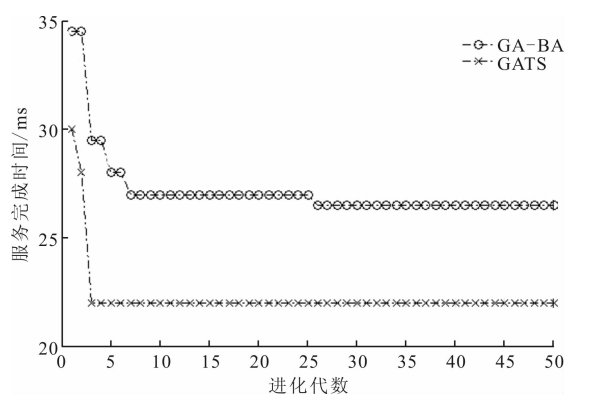


图 11 2 种算法的收敛过程

图 12 表示不同的服务功能链请求强度下的 2 种算法的服务完成时间,其中,网络参数设置为具有 20 节点的全联通网络,设置种群数量 N_{NIND} 设置为

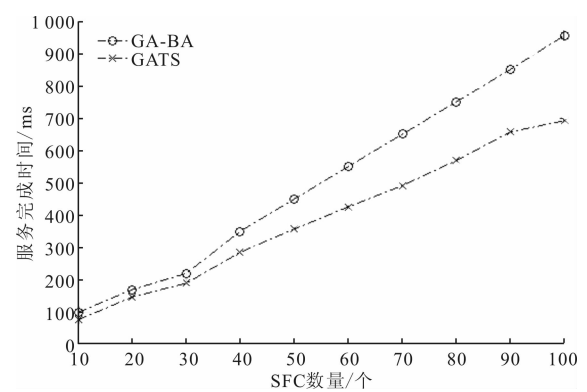


图 12 不同服务功能链请求强度下的服务完成时间

40,最大遗传代数 $\max G_{\text{en}}$ 为 50,服务功能链数据流量在 0~20 kb 均匀生成。可以看出,GATS 算法获得的服务完成时间更少,而且随着流量的上升,本文方法的优势更加明显,这是因为由于流量的增加,传输时延将成为影响服务完成时间的关键因素,GATS 算法在链路带宽分配上具有更高的细粒度,而且 GATS 算法将遗传算法和禁忌搜索算法有效的结合,能够搜索到更优的调度方案,在更短的时间内完成网络服务。

4 结 论

本文基于柔性车间调度问题提出了结合虚拟链路动态带宽分配的虚拟网络功能调度模型,并设计了联合遗传和禁忌搜索的启发式算法(GATS)进行求解。GATS 算法通过混合整数规划的方法准确求出链路带宽分配的闭式最优解,提高了带宽分配的细粒度,同时利用禁忌搜索算法的局部搜索能力,有效避免了遗传算法的早熟问题。实验结果表明,GATS 算法在收敛速度上具有较高优势,而且获得了更好的调度方案,能够满足 5G 低时延场景的需求,增加了用户的服务体验和运营商的收益。但是,本文提出的算法由于融合了 2 种算法,增加了算法的时间复杂度。下一步,将在保证求得更优调度方案的前提下,研究算法时间复杂度更低的算法。

参考文献:

[1] 张平,陶运铮,张治. 5G 若干关键技术评述 [J]. 通信学报,2016,37(7): 15-29.
ZHANG Ping, TAO Yunzheng, ZHANG Zhi. Survey of several key technologies for 5G [J]. Journal on Communications, 2016, 37(7): 15-29.
[2] MARTINI B, PAGANELLI F, CAPPANERA P, et al. Latency-aware composition of virtual functions in 5G [C] // IEEE Conference on Network Softwariza-

- tion: Software-Defined Infrastructures for Networks, Clouds, IoT and Services, NETSOFT 2015. Piscataway, NJ, USA: IEEE, 2015: 1-6.
- [3] ABDELWAHAB S, HAMDAOUI B, GUIZANI M, et al. Network function virtualization in 5G [J]. IEEE Communications Magazine, 2016, 54(4): 84-91.
- [4] 刘彩霞, 卢干强, 汤红波, 等. 一种基于 Viterbi 算法的虚拟网络功能自适应部署方法 [J]. 电子与信息学报, 2016, 38(11): 2922-2930.
- LIU Caixia, LU Ganqiang, TANG Hongbo, et al. Adaptive deployment method for virtualized network function based on Viterbi algorithm [J]. Journal of Electronics & Information Technology, 2016, 38(11): 2922-2930.
- [5] KREUTZ D, RAMOS F M V, ESTEVES P, et al. Software-defined networking: A comprehensive survey [J]. Proceedings of the IEEE, 2014, 103(1): 10-13.
- [6] CHENG G, CHEN H, HU H, et al. Enabling network function combination via service chain instantiation [J]. Computer Networks, 2015, 92: 396-407.
- [7] HERRERA J G, BOTERO J F. Resource allocation in NFV: A comprehensive survey [J]. IEEE Transactions on Network and Service Management, 2016, 13(3): 518-532.
- [8] OSSEIRAN A, BOCCARDI F, BRAUN V, et al. Scenarios for 5G mobile and wireless communications: the vision of the METIS project [J]. IEEE Communications Magazine, 2014, 52(5): 26-35.
- [9] 袁泉, 汤红波, 黄开枝, 等. 基于 Q-learning 算法的 vEPC 虚拟网络功能部署方法 [J]. 通信学报, 2017, 38(8): 172-182.
- YUAN Quan, TANG Hongbo, HUANG Kaizhi, et al. Deployment method for vEPC virtualized network function via Q-learning [J]. Journal on Communications, 2017, 38(8): 172-182.
- [10] RIERA J F, ESCALONA E, BATALLE J, et al. Virtual network function scheduling: Concept and challenges [C] // 2014 International Conference on Smart Communications in Network Technologies. Piscataway, NJ, USA: IEEE, 2014: 6867768.
- [11] LI X, QIAN C. Low-complexity multi-resource packet scheduling for network function virtualization [C] // Proceedings of 2015 IEEE Conference on Computer Communications. Piscataway, NJ, USA: IEEE, 2015: 1400-1408.
- [12] LUIZELLI M C, BAYS L R, BURIOL L S, et al. Piecing together the NFV provisioning puzzle: Efficient placement and chaining of virtual network functions [C] // Proceedings of 2015 IFIP/IEEE International Symposium on Integrated Network Management. Piscataway, NJ, USA: IEEE, 2015: 98-106.
- [13] QU L, ASSI C, SHABAN K. Delay-aware scheduling and resource optimization with network function virtualization [J]. IEEE Transactions on Communications, 2016, 64(9): 3746-3758.
- [14] MIJUMBI R, SERRAT J, GORRICHIO J L, et al. Network function virtualization: State-of-the-art and research challenges [J]. IEEE Communications on Surveys & Tutorials, 2016, 18(1): 236-262.
- [15] GUO J, LIU F, LUI J, et al. Fair network bandwidth allocation in IaaS datacenters via a cooperative game approach [J]. IEEE/ACM Transactions on Networking, 2016, 24(2): 873-886.
- [16] GUO J, LIU F, HUANG X, et al. On efficient bandwidth allocation for traffic variability in datacenters [C] // Proceedings of IEEE INFOCOM. Piscataway, NJ, USA: IEEE, 2014: 1572-1580.
- [17] GAMBARDELLA L M, MASTROLILLI M. Effective neighborhood functions for the flexible job shop problem [J]. Journal of Scheduling, 1996, 3(3): 3-20.

[本刊相关文献链接]

- 陈俊杰, 周晖, 王伟. 一种低成本的云资源提供算法. 2017, 51(10): 135-141. [doi: 10.7652/xjtuxb201710022]
- 王琳, 钱德沛, 王锐, 等. 高效支持负载聚合的程序资源敏感度获取及分析方法. 2017, 51(4): 79-84. [doi: 10.7652/xjtuxb201704012]
- 朱正仓, 赵季红, 唐睿, 等. 移动中继协助下终端直通中的模式选择和资源分配方案. 2016, 50(10): 111-117. [doi: 10.7652/xjtuxb201610017]
- 赵辉, 郑庆华, 张未展, 等. 多版本视频点播流媒体服务器集群资源分配方法. 2016, 50(6): 30-35. [doi: 10.7652/xjtuxb201606005]
- 周墨颂, 朱正东, 董小社, 等. 采用资源划分的云环境下 Hadoop 资源许可调度方法. 2015, 49(8): 69-74. [doi: 10.7652/xjtuxb201508012]

(编辑 刘杨)