

基于需求预测的云无线接入网计算资源分配策略研究

王志朋^{1,2}, 曹斌^{1,2}, 张钦宇^{1,2}

(1. 哈尔滨工业大学(深圳)电子与信息工程学院, 广东 深圳 518055;

2. 鹏城实验室网络通信研究中心, 广东 深圳 518055)

摘要: 云无线接入网利用网络功能虚拟化和软件定义网络技术以支持端到端的网络切片, 使得接入网可共享无线、终端和网络等资源, 已成为 5G 网络中优先采用的网络架构。针对云无线接入网中端到端的网络切片场景, 通过控制平面建立的数据驱动运维框架来收集网络信息并进行数据处理。预测未来一段时间内计算业务量的需求, 设计了一种基于虚拟化网络功能(VNF)的计算资源分配方案, 提出了基于降序最佳适应(BFD)的离散粒子群算法。仿真结果表明, 所提出的策略和算法可实现云无线接入网计算资源的动态灵活分配, 并能有效降低 VNF 迁移能耗和迁移次数。

关键词: 云无线接入网; 计算资源分配; 离散粒子群算法; 网络切片

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2019.00126

Research on computing resource allocation strategy for cloud radio access network based on demand forecasting

WANG Zhipeng^{1,2}, CAO Bin^{1,2}, ZHANG Qinyu^{1,2}

1. School of Electronic and Information Engineering, Harbin Institute of Technology(Shenzhen), Shenzhen 518055, China

2. Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen 518055, China

Abstract: Cloud radio access network(CRAN) is a new network architecture commonly used in 5G network. It uses network function virtualization and software defined network technology to support end-to-end network slicing so that access network can share the same infrastructure. A computing resource allocation scheme based on virtualized network function(VNF) computing resource demand forecasting was proposed for cloud wireless network supporting end-to-end network slicing. Data-driven operation and maintenance framework established by the control plane collects network information and performs data processing to predict the demand for computing traffic in a period of time, which used the discrete particle swarm optimization based on best-fit-decreasing to dynamically and dynamically allocate CRAN computing resources, reducing the migration energy consumption and migration times of VNF.

Key words: cloud radio access network, computing resource allocation, discrete particle swarm optimization, network slicing

1 引言

近年来, 便携式移动终端设备数量及产生的数据流量呈急剧增加的态势, 为了满足终端和数据量

的快速增长需求, 如采用原有的接入网架构进行密集化建设并开放更高频谱, 需要更换新的通信设备, 从而增加了资本支出和运营成本, 并且会带来更严重的信道干扰。为了支撑自动驾驶、智慧城市^[1]等

收稿日期: 2019-06-17; 修回日期: 2019-08-26

基金项目: 广东省自然科学基金资助项目(No.2017A030313372); 深圳市科技计划项目(No.JCYJ20170307151148585); 海洋立体通信网络示范验证平台项目(No.PCL2018KP002)

Foundation Items: NSF Guangdong(No.2017A030313372), The Basic Research Project of Shenzhen(No.JCYJ20170307151148585), Stereoscopic Coverage Communication Network Verification Platform for China Sea(No.PCL2018KP002)

新型产业的落地，基于网络功能虚拟化（NFV，network functions virtualization）和软件定义网络（SDN，software defined network）的思想，中国移动在 2010 年首次提出了云无线接入网（CRAN，cloud radio access network）的概念和架构^[2]，旨在支撑移动互联网业务，提供可扩展性，降低运营商的资本支出及运营成本。

根据 3GPP 标准^[3]，采用 NFV 的接入网架构主要包含两类网络功能组件：虚拟化网络功能（VNF，virtualized network function）和物理网络功能（PNF，physical network function）。其中，VNF 是部署在网络功能虚拟化基础设施（NFVI，network function virtualization infrastructure）中的网络功能函数，VNF 可采用虚拟机（VM，virtual machine）、容器等方式实现，而 NFVI 通常采用通用硬件，造价较低。PNF 利用软、硬件的紧密耦合以实现相应的网络功能。为了支持端到端通信，NFV 通常由一组规定顺序的网络功能来支撑，网络功能可以是 VNF 或 PNF，亦可由 VNF 和 PNF 结合形成。VNF 和 PNF 的联合部署可有效保证通信服务质量，并支持网络的可扩展性。

CRAN 采用集中式方式，并引入 NFV 和 SDN 等关键技术，可兼容多种空口标准并支持基础设施的共享。多切片 CRAN 架构如图 1 所示，CRAN 主要由 3 部分组成：分布式的无线远端射频单元（RRH）、高性能的基带处理（BBU）池以及连接 BBU 和 RRH 的光传输网络。具体而言，CRAN 将基站的处理和控制单元 BBU 汇聚到 BBU 池中，将功能简单的 RRH 部署在边缘网络区域，而 BBU 和 RRH 之间采用光纤和公共射频接口协议进行通信。可以看出，BBU 的集中管控及虚拟化技术的引入，有效地支持了基带通信和计算资源的共享，因此，可方便利用协作多点（CoMP）技术抑制小区间的干扰，并减少机房的数目。此外，RRH 可利用基站休眠等技术以降低网络的部署代价和运维能耗，有效提高系统容量和频谱效率。

借助 NFV 和 SDN 技术，网络切片能够定制不同性能要求的网络，并满足多样化的应用需求，有望进一步降低运营商的运维成本。与 CRAN 的不同之处在于，网络切片侧重于业务之间资源的隔离机制，以降低甚至消除不同业务间的影响为目标。此外，通过开放接口，网络切片能够向第三方租户提供网络资源，以满足更加智能、多样化的业务需求。

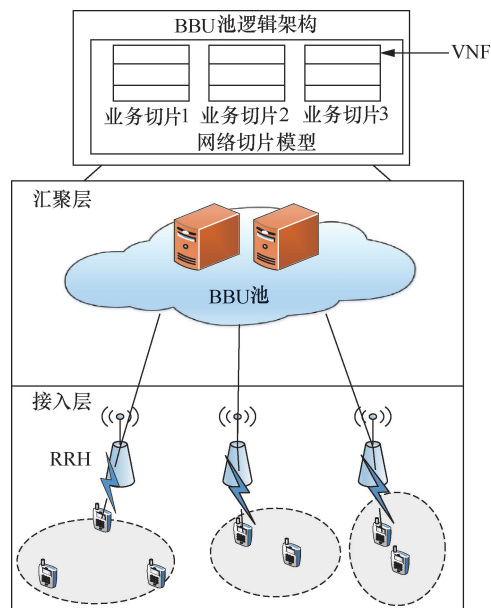


图 1 多切片 CRAN 架构

5G 需要支持连续广域覆盖、热点大容量、低功耗大连接以及低时延、高可靠的应用场景，这要求网络架构必须满足差异化的业务需求，而基于网络切片和 CRAN 的 5G 网络架构可有效支持这种差异化的应用需求。在按需定制的准则下，以计算资源的分配为例，因为 VNF 需要的计算资源随时间动态变化，为了满足不同时刻的需求并降低系统资源池的能耗，需要对计算资源进行按时、按需分配。Zhang 等^[4]采用首次适应算法和遗传算法进行 CRAN 中 BBU 池的计算资源分配。Lyazidi 等^[5]将 BBU 池资源分配建模为背包问题，并采用 CPLEX 进行求解。Wang 等^[6]采用基于降序最佳适应 (BFD, best fit decreasing) 分配计算资源，以最小化 BBU 池能耗。Aqeeli 等^[7]将计算资源分配问题建模为 0-1 整数规划问题，并采用启发式算法进行求解。

上述工作主要根据 RRH 的实时业务量来实现 BBU 池中的计算资源分配，并未考虑 NFV 和 SDN 的特性。值得注意的是，SDN 和 NFV 的引入虽然降低了运营和资本支出，但是也带来了其他问题。Yousaf 等^[8]指出，当业务的需求发生变化时，为了满足服务的 KPI，需要进行 VNF 的 Scale-down、Scale-out、迁移或更新等操作。这表明，当 RRH 的通信业务量发生变化时，会造成 BBU 池中每个 VNF 所需计算资源发生变化，需要重新分配资源。即当 VNF 进行 Scale-down 或 Scale-out 操作导致 BBU 单元的利用率低或资源不足时，会发生 VNF 的合并和负载的均衡，即 VNF 迁移。但性能影响最小的

在线实时迁移（热迁移）也会导致 VNF 性能急剧下降。

为了满足动态需求，本文提出了一种基于流量预测的 CRAN 计算资源分配策略，根据预测信息，进行合理地 VNF 放置，以减少不必要的 VNF 迁移。具体而言，建立了基于数据驱动的 CRAN 资源分配框架，采用支持向量回归算法进行业务预测，给出基于 BFD 的离散粒子群分配策略和算法，并进行仿真验证和分析。

2 数据驱动的 CRAN 资源分配框架

在 CRAN 的控制平面中引入数据驱动，数据驱动的 CRAN 资源分配框架如图 2 所示，主要包含数据收集层、数据处理层、网络优化层和控制层。

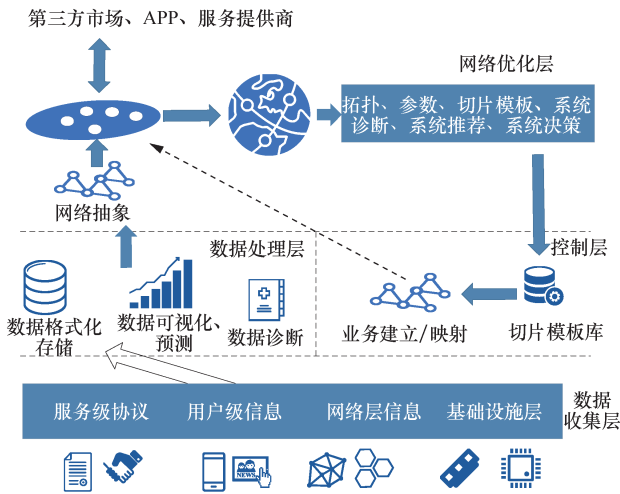


图 2 数据驱动的 CRAN 资源分配框架

2.1 数据收集层

数据收集层主要依靠网络监控功能收集 CRAN 的各种参数信息，主要分为 4 类：1) 服务级协议，是运营商与客户间的服务协议，包括业务速率、覆盖范围、失败重传时间间隔、时延等；2) 用户级信息，包括用户接收机的类型、电池限制、用户位置等；3) 网络层信息，包括频谱、用户个数、QoS、网络功能的 KPI、调度周期等；4) 基础设施层信息，包括 CPU、内存、存储、路由等。

2.2 数据处理层

数据处理层主要包含两个任务：将收集的信息进行预处理和数据分析。数据预处理对原始数据进行整合、清洗、过滤等操作；数据分析可采用 3GPP 的网络数据分析^[9]和欧洲电信标准化协会（ETSI, European Telecommunications Standards Institute）的

实验性网络智能模型。网络数据分析主要满足 3 种需求：数据的可视化、预测、诊断。统计学习、强化学习以及深度学习等是支撑该模块的关键技术。

2.3 网络优化层

根据数据处理层的输出数据，优化器主要实现网络资源分配的策略功能，因此，优化器需要具备有效的预测模型、合理的数据以及和控制器模块交互的接口，从而形成具有反馈作用的体系。优化器主要分为 3 部分功能：1) 垂直方和第三方业务的网络参数接口；2) 业务抽象，主要为网络性能抽象、QoS 以及网络约束；3) 可采用智能计算、启发式算法、优化算法等设计切片模板。模板包括 VNF 种类、网络功能的拓扑和顺序、可靠性、安全性等要素。同时，网络优化层使用 AI 技术实现网络的诊断、推荐和决策。

2.4 控制层

控制层的核心部件控制器主要根据优化器设计的切片模板使用 SDN 控制器和 MANO (management and network orchestration) 进行网络定制化，并实现切片的实例化和性能维护^[10]。其中，根据切片模板可确定资源映射、网络拓扑和实例化切片。切片的性能维护依靠切片的监控、分析和重配置。通过切片的监控将切片实例进行抽象化，传递给优化器进行分析，由优化器决定是否重配置或结束。

3 基于流量预测的 CRAN 计算资源分配

3.1 系统模型和假设

将图 1 作为一个典型的 CRAN 架构，包含 I 个 RRH 和 Q 种业务。在 BBU 池中采用网络切片隔离，根据网络中 Q 种业务将其划分为 Q 个计算切片。为了保证一定的切片隔离度，每个 BBU 仅服务于一个切片。在每个切片内，一个 VNF 仅服务于单个 RRH 中的一个业务。 t 时刻时，第 i 个 RRH 的第 q 个业务所需要的计算资源为 $C_{i,q}^t$ ， $C_{\max,q}$ 表示单个 BBU 可为业务 q 提供的计算资源上限。为了减少迁移次数，并尽可能地减少 BBU 的开启个数，对每种业务切片建立装箱优化问题 (bin packing problem)。

$$\min_{x_i^t} \sum_{t=1}^{T-1} \sum_{i \in I} C_{i,q}^t d(x_{i,q}^t - x_{i,q}^{t+1}) + \sum_{t=1}^T \sum_{k \in K} C_{\max,q} z_k^t \quad (1)$$

s.t.

$$\sum_{i \in I'_{k,q}} C'_{i,q} \leq C_{\max,q} z'_k \quad \forall t \in T, \forall k \in K_q, I'_{k,q} = \{i | x'_{i,q} = k\} \quad (2)$$

$$x'_{i,q} \in \{1, 2, \dots, K_q\} \quad \forall t \in T, \forall i \in A \quad (3)$$

$$d(z) = \begin{cases} 0, & z = 0 \\ 1, & \text{其他} \end{cases} \quad (4)$$

其中, K_q 表示业务 q 可用 BBU 的集合, z'_k 表示 t 时刻第 k 个 BBU 开启, $I'_{k,q}$ 表示在 t 时刻第 k 个 BBU 服务的业务类型为 q 的 VNF 的集合。式(2)表示每个 BBU 所服务的 VNF 计算需求总和不能超过其所能提供的计算能力上限, 式(3)表示 t 时刻 RRH i 中业务 q 所选择的 BBU 集合。

3.2 流量预测

通过数据收集层对每个 RRH 的 VNF 所需计算量进行格式化记录, 传递给数据处理层, 以估计未来 T 时刻 VNF 所需计算量。

支持向量回归 (SVR, support vector regression) 能够较好地解决小样本、非线性问题, 具有良好的泛化性能和较低的时间复杂度。通过 ε -不敏感函数, 可将回归问题和支撑向量问题联系起来。为了解决非线性回归问题, SVR 应使用适当的核进行表示, 如多项式或高斯核, 将非线性样本映射到高维特征空间中, 使样本线性可分。因此, SVR 模型的参数可通过求解具有线性等式和不等式约束的二次规划问题而获得。

时间序列是可观测变量 y 在相同时间间隔内的历史测量序列。研究时间序列的目的包括: 基于过去的知识对未来的预测, 对时间序列进行简明描述或分析其背后的现象。单步和多步时间预测原理如图 3 所示, 图 3(a)为单步时间预测。对于已记录的时间序列, 若将该序列 $t-N$ 时刻到 $t-1$ 时刻的数据作为模型的输入, 将 t 时刻作为模型的输出, 则单步时间序列预测可转化为监督学习回归问题。而多步时间预测可采用迭代的单步时间预测来实现, 如图 3(b)所示, 将上次所预测的数据 \hat{y}_t 作为模型的输入以预测数据 \hat{y}_{t+1} 。

训练集 $D = \{(\mathbf{x}_i, y_i) \in \mathbb{R}^N \times \mathbb{R}, i = 1, \dots, l\}$, 其中, 输入变量 \mathbf{x}_i 为 N 维向量, 表示过去 N 个时刻的值; 输出 y_i 为连续变量, 表示下一时刻的值。核函数将非线性的数据映射到合适的维度, 使数据线性可分。根据训练集 D , 引入核函数的线性回归模型可表示为

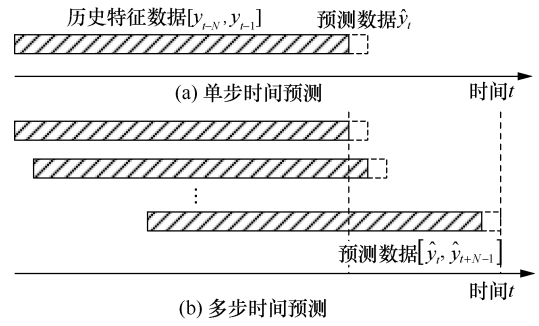


图3 单步和多步时间预测原理

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^l w_i \times \phi(\mathbf{x}) + b \quad (5)$$

其中, $\mathbf{w} = [w_1, w_2, \dots, w_l]$ 代表模型权重系数, b 为模型偏差, $\phi(\mathbf{x})$ 为待映射函数, 一般将 N 维空间的点映射到更高维的空间。

核方法和支撑向量机是两个独立的概念, 可通过引入 ε -不敏感函数 $g_\varepsilon(\bullet)$ 将回归模型的求解转化为借助支撑向量机的方法求解。因此, 模型的求解可以采用 SVR 模型^[11]

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l g_\varepsilon(y_i - f(\mathbf{x}_i, \mathbf{w})) \quad (6)$$

常用的 ε -不敏感函数表示为

$$g_\varepsilon(z) = \begin{cases} 0, & |z| \leq \varepsilon \\ |z| - \varepsilon, & \text{其他} \end{cases} \quad (7)$$

一般 SVR 模型采用软间隔, 即允许样本存在一定的误差, 因此, 引入松弛变量 ξ_i 和 $\hat{\xi}_i$, 由此可得

$$\min_{\mathbf{w}, b, \xi_i, \hat{\xi}_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \hat{\xi}_i) \quad (8)$$

s. t.

$$f(\mathbf{x}_i, \mathbf{w}) - y_i \leq \varepsilon + \hat{\xi}_i \quad (9)$$

$$y_i - f(\mathbf{x}_i, \mathbf{w}) \leq \varepsilon + \xi_i \quad (10)$$

$$\hat{\xi}_i \geq 0, \xi_i \geq 0, i = 1, 2, \dots, l \quad (11)$$

通过采用拉格朗日乘子和 KKT 条件可得 SVR 的解为

$$f(\mathbf{x}) = \sum_{i=1}^l (\hat{\alpha}_i - \alpha_i) \phi(\mathbf{x}_i) \phi(\mathbf{x}) + b \quad (12)$$

其中, α_i 和 $\hat{\alpha}_i$ 为式(8)中第一个和第二个约束条件的拉格朗日乘子。

3.3 基于 BFD 离散粒子群的资源分配

本节介绍一种结合流量预测和 BFD 的离散粒子群算法。网络优化层接收由数据处理层预测的 T 时刻计算需求量，并结合 BFD 和离散粒子群算法进行资源分配。

由于不同的业务需求存在较大的不平衡性，如蜂窝网络中的语音通话、视频通话、网页浏览等业务的需求中，仅有少数业务具有较大的需求，而大多数业务的需求量较少。因此，需要根据业务需求日常流量进行资源分配。对于需求波动较小的业务，可采用 BFD 算法进行资源分配，从而降低整体资源分配的复杂度。 T 时刻的 CRAN 计算资源分配流程示意图如图 4 所示。

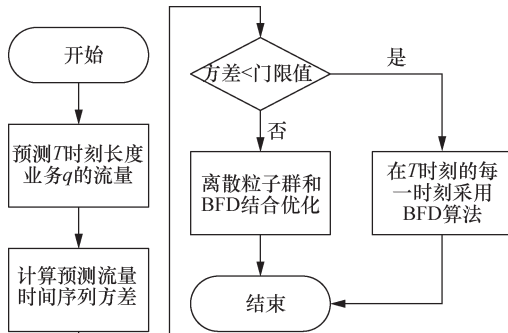


图 4 T 时刻的 CRAN 计算资源分配流程示意图

CRAN 中的计算资源分配包含以下步骤。

首先对业务波动进行分类。计算 CRAN 中以某种业务所需计算资源总和和时间序列的方差，设置判决门限。当计算资源需求波动较大时，即超过该门限，则执行基于 BFD 离散粒子群算法分配；当计算资源需求波动较小时，即低于该门限，则执行 BFD 分配。

1) 若需求波动较小，采用 BFD 分配

BFD 是基于贪心策略的优化算法，并非考虑全局的最优策略，而是保证每次迭代选择最优的策略。BFD 算法如图 5 所示，在一次 BBU 和 VNF 配对中，分多次迭代，每次迭代选择 VNF 最大计算需求量，与 BBU 按照剩余资源升序排列的顺序进行匹配，选择首个满足选中 VNF 的 BBU。如果不存在已开启的 BBU 满足 VNF，则开启一个未开启的 BBU。

2) 若需求波动较大，采用 BFD 离散粒子群算法进行分配

首先，由 BFD 分配算法确定 BBU 在一段时间

内的开启个数，并将该个数 $|\kappa_{\text{BFD}}|$ 与前一时刻分配方案开启的个数 $|\kappa_t|$ 进行比较，以确定离散粒子群 BBU 输入集合 κ_{in} 。如果 $|\kappa_{\text{BFD}}| < |\kappa_t|$ ，则将前一时刻分配方案中的 BBU 集合作为输入集合。否则添加 $|\kappa_{\text{BFD}}| - |\kappa_t| + 1$ 个 BBU 到前一时刻分配方案的 BBU 集合中作为输入集合。最后，使用离散粒子群进行求解，当满足终止条件时结束。结合 BFD 的离散粒子群算法如图 6 所示。

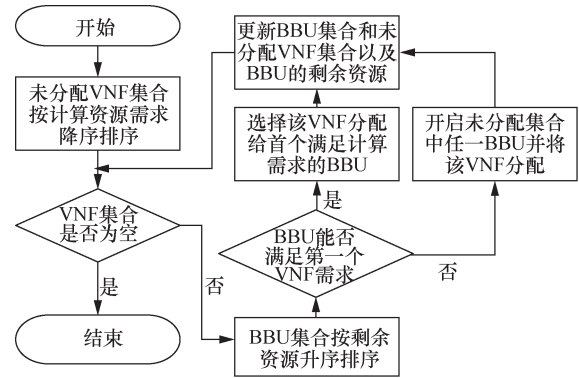


图 5 BFD 算法

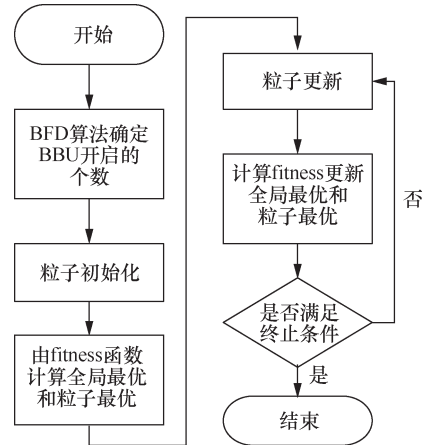


图 6 结合 BFD 的离散粒子群算法

本文所使用的离散粒子群算法是基于 Pan 等^[12]所提的基于遗传算法的粒子群算法。首先进行粒子的编码，通过对 κ_{in} 集合中的 BBU 进行从 1 到 $|\kappa_{\text{in}}|$ 的编号，利用 VNF 的个数和时间进行编码，编码长度为 $T \times I$ ，每位上的值为 1 到 $|\kappa_{\text{in}}|$ 的整数。

离散粒子群算法的更新计算式为

$$X'_i = c_2 \otimes F_3(c_1 \otimes F_2(w \otimes F_1(X_i^{t-1}), P_i^{t-1}), G^{t-1}) \quad (13)$$

其中， X'_i 和 X_i^{t-1} 是第 t 轮和第 $t-1$ 轮的解， P_i^{t-1} 为第 $t-1$ 轮时粒子 i 的最优解， G^{t-1} 为全部粒子历史中的最优解。 $F_1(\bullet)$ 操作代表变异，即 X_i^{t-1} 编码中任

一元素变异。 $F_2(\bullet)$ 代表杂交, 即和该粒子的历史最优解替换编码片段。 $F_3(\bullet)$ 表示和全局的历史最优解进行杂交, c_1 、 c_2 为区间 $[0,1]$ 的随机数。通过式(13)更新下一轮的粒子。其中, 变异操作作为粒子编码序列的某一元素变为 1 到 $|\kappa_m|$ 的任一值, 杂交是两个粒子的编码序列在某一相同位置处被切断, 前、后两串分别交叉合成两个粒子编码序列。

离散粒子群的 fitness 函数根据所提优化问题进行构造。由于所提优化问题包含限定条件, 可通过构造惩罚函数将原问题变为无约束问题, 即由式(7)和式(8)可得

$$F = \sum_{t=1}^{T-1} \sum_{i \in I} C_{i,q}^t d(x_i^t - x_i^{t+1}) + \sum_{t=1}^T \sum_{k \in K} C_{\max,q} z_k^t + \mu \sum_{t=1}^T \sum_{k=1}^K \max(0, (\sum_{i \in I_k} C_{i,q}^t - C_{\max,q} z_k^t)^2) \quad (14)$$

其中, μ 为惩罚因子。

为了跳出局部最优, 当全局最优解连续 S 次(总迭代次数的 5% 左右)没有改进时, 将 30% 的粒子进行重新初始化。算法的终止条件是全局最优解连续出现总迭代次数的 50% 或达到最大迭代次数时结束。

4 仿真结果与分析

4.1 评价指标

假设 BBU 池采用共享存储的虚拟机热迁移技术, 前期阶段 VNF 在源 BBU 上运行, 预复制阶段进行内存的预迁移。预复制进行一段时间后, 目标 BBU 具备运行 VNF 的必要资源, 此时预复制停止, 进入 VNF 停机复制阶段, 经停机复制后唤醒已迁移到目的 BBU 上的 VNF。

文献[13]对虚拟机热迁移的迁移能耗进行建模, 在仿真中, 假设 SDN 虚拟机分配的带宽 R 按照 VNF 的内存丢失率 D 的大小进行分配

$$\lambda = \frac{D}{R} \quad (15)$$

则整个迁移所传输的内存总和为

$$V_{\text{mig}} = V_m \frac{1 - \lambda^{n+1}}{1 - \lambda} = V_m \rho \quad (16)$$

其中, V_m 为 VNF 的内存大小。实验表明, 迁移的能耗与 VNF 的内存呈近似线性关系, 即

$$E_{\text{mig}} = \alpha V_{\text{mig}} + \beta \quad (17)$$

其中, $\alpha = 0.512$, $\beta = 20.165$ 。

4.2 仿真数据集的生成及计算资源定义

根据 2014 年意大利电信推出的数据集, 生成本文仿真的数据集^[14]。该数据集为基站 CDR 在时间和空间业务量的聚集, 其中, 在时间上以 10 min 为间隔, 在空间上是 $235 \times 235 \text{ m}^2$ 的 CDR 聚集。根据该数据集, 得到每个基站在每 10 min 内业务的平均到达速率。

为了表征计算资源需求, 假设数据业务的数据分组大小为 D_d , 数据分组的到达速率服从参数为 λ_d 的泊松分布, 将数据业务建模为 M/D/1 排队系统。语音业务的请求服从参数为 λ_c 的泊松分布, 通话时长服从参数为 μ_c 的指数分布, 消息业务的请求服从参数为 λ_s 的泊松分布, 消息长度服从参数为 μ_s 的指数分布, 将通话业务和消息业务建模为 M/M/1 排队系统。将 VNF 处理业务的速度作为 VNF 计算资源的需求数据。

一般而言, VNF 的内存大小由操作系统的底层算法进行分配, 本文假设 VNF 的内存分配按照队列模型长度的大小比例进行配置, 即对于同类型的 VNF, 其内存大小为

$$V_{\text{mig}}^{q,i} = \theta_q l_i + \delta_q \quad (18)$$

其中, θ_q 、 δ_q 为业务 q 的系数, l_i 为队列长度大小。假定数据切片的 VNF 中, $\theta_d = 1$ 、 $\delta_d = 200$; 语音切片采用 VoLTE 的 VNF 中, $\theta_c = 0.5$ 、 $\delta_c = 250$; 消息业务的 VNF 中, $\theta_s = 2$ 、 $\delta_s = 150$ 。

在实际的 CRAN 中, 通过虚拟化基础设施管理器 (VIM) 监控 VNF 的计算资源使用情况, 并统计资源需求量和 VNF 迁移的能耗。

4.3 仿真参数

不同业务的 QoS 指标不同, 根据文献[15]中的 QoS 指标, 仿真中数据业务的时延为 5 ms, 语音业务的时延为 50 ms, 消息业务的时延为 150 ms。假定语音业务和消息业务采用 VoLTE, 即语音数据和消息均基于 LTE 技术, 单个 BBU 的计算上限为 50 Mbit/s。单个用户语音业务速率为 50 kbit/s, 消息业务数据分组的平均长度为 0.56 kB, 数据业务的数据分组大小为 1 MB。

某一基站 3 种业务的一天仿真数据集如图 7 所示, 可知每种业务需求的计算资源随时间变化, 趋势大致与人们的日常作息时间一致。语音通话的计算资源需求的波动区间长度小于 0.025 Mbit/s, 数据

业务的计算资源需求的波动区间长度小于 6 Mbit/s, 消息业务的计算资源需求的波动区间长度小于 2×10^{-4} Mbit/s, 3 种业务差异明显。语音业务和消息业务的计算需求波动趋势相较于数据业务过于平稳, 原因在于只有提供一定量的计算资源, 语音业务和消息业务才能满足一定的时延, 而单个语音业务和消息业务的传输数据量少, 现实生活中的语音业务和消息业务所占比重远小于数据业务, 因此, 波动较低。

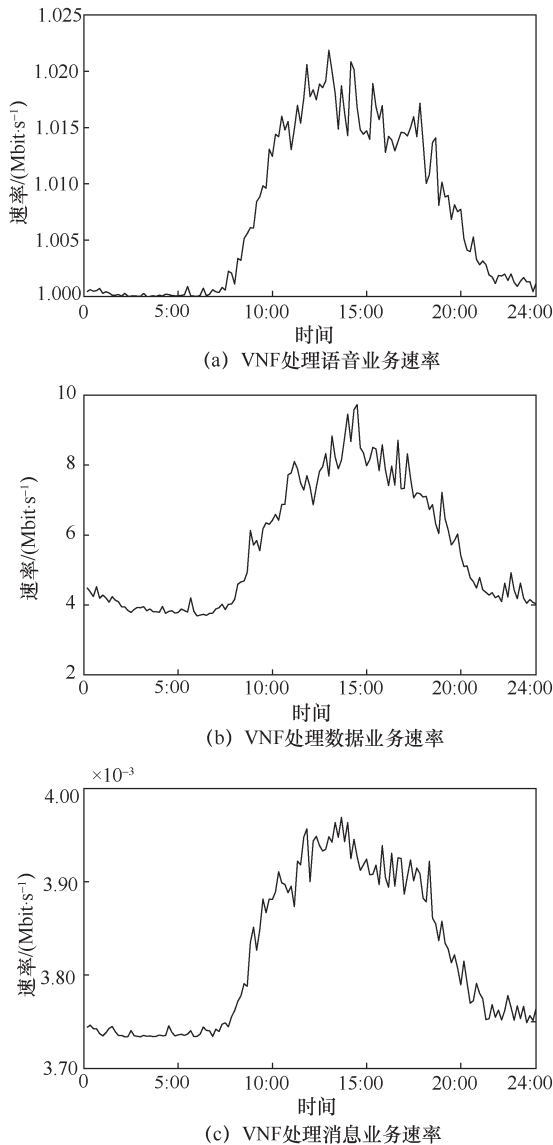


图 7 某一基站 3 种业务的一天仿真数据集

假设 9 个 RRH 均匀分布, 每个 RRH 有 3 种业务, 每个 VNF 的计算资源需求量由仿真数据集的计算需求来代替。VNF 迁移次数如图 8 所示, 采用基于 BFD 的离散粒子群算法和 BFD 算法进行比较,

图 8 中每一点为 1 h 迁移次数的总和。在凌晨 1 点到早上 8 点的业务量少、波动小, 分配的 BBU 能够服务所有 RRH, VNF 未发生迁移; 早上 8 点到晚上 24 点, 由于活动较多, 业务量波动大, 采用 BFD 算法, 仅考虑每次分配最优, 未考虑未来一段时间的业务量, 会造成不必要的迁移过多。迁移能耗如图 9 所示, 由于不同 RRH 在业务量上具有一定的相似性, 按照所提内存的分配方式, RRH 和其邻近 RRH 的 VNF 的内存大小相近, 因此, 其迁移的能耗和迁移次数走势相同。

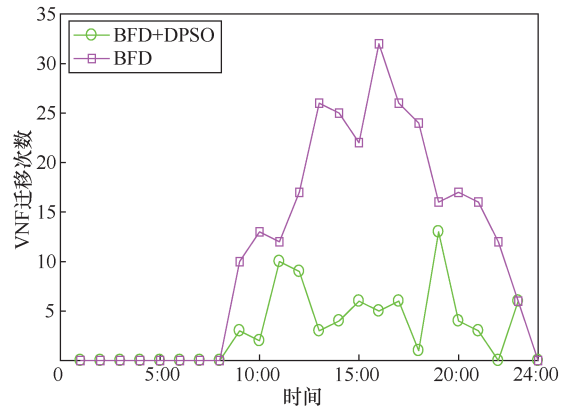


图 8 VNF 迁移次数

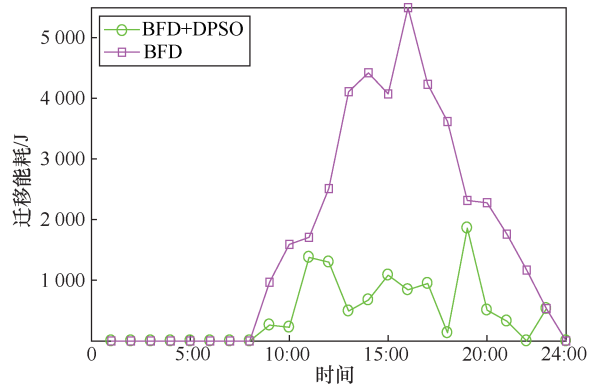


图 9 迁移能耗

单位 BBU 使用个数如图 10 所示, 由于所提算法是基于 BFD 算法的离散粒子群算法, 将单个 BBU 开启 10 min 定义为单位 BBU, 可看出所使用的单位 BBU 的个数与 BFD 接近。BBU 的固定能耗在 BBU 池的总能耗中所占比重较大, 采用 BFD 和离散粒子群结合的方式兼顾了 CRAN 的能耗。

5 结束语

本文提出了一种 CRAN 端到端网络切片场景下基于需求预测的 BFD 离散粒子群算法。该场景

下所提数据驱动框架收集系统的各类信息,将 VNF 计算业务需求信息作为数据集,采用 SVR 模型训练,通过递归的多步时间预测未来一段时间内的 VNF 计算业务需求量。借助预测的业务需求信息,建立背包问题,使用基于 BFD 的离散粒子群算法分配计算资源。所提算法相较于单一的 BFD 算法,大幅度降低了迁移的次数,并且在 BBU 能耗方面,和单一的 BFD 算法性能相同。

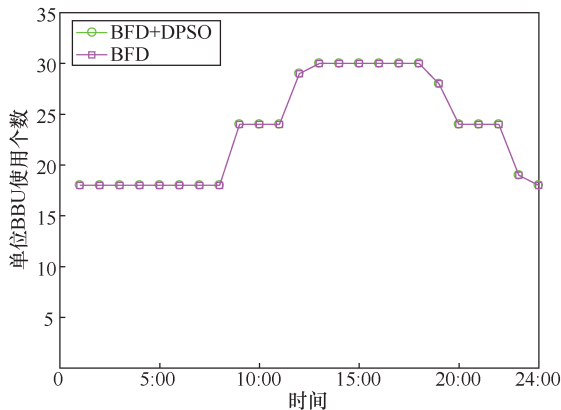


图 10 单位 BBU 使用个数

在未来的工作中,可以将数据业务进一步细化分配,并且考虑虚拟化对于不同数据业务类型的影响,对网络的资源进行更细致的分配,制定更加贴合实际的资源分配方案。

参考文献:

- [1] 杨靖,张祖伟,姚道远,等. 新型智慧城市全面感知体系[J]. 物联网学报, 2018, 2(3): 95-101.
YANG J, ZHANG Z W, YAO D Y, et al. Comprehensive sensing system of new smart city[J]. Chinese Journal on Internet of Things, 2018, 2(3): 95-101.
- [2] 中国移动通信有限公司研究院. 迈向 5G C-RAN: 需求、架构与挑战白皮书[R]. 2016.
Research Institute of China Mobile Communication Co., Ltd. Towards 5G C-RAN: demand, architecture and challenge white paper[R]. 2016.
- [3] 3GPP. Study on management aspects of virtualized network functions that are part of the New Radio (NR) [S]. 2017.
- [4] ZHANG F, ZHENG J, ZHANG Y, et al. An efficient and balanced BBU computing resource allocation algorithm for cloud radio access networks[C]//2017 IEEE 85th Vehicular Technology Conference (VTC Spring). IEEE, 2017: 1-5.
- [5] LYAZIDI M Y, AITSAADI N, LANGAR R. Dynamic resource allocation for Cloud-RAN in LTE with real-time BBU/RRH assignment[C]//2016 IEEE International Conference on Communications (ICC). IEEE, 2016: 1-6.
- [6] WANG K, ZHOU W, MAO S. On joint BBU/RRH resource allocation in heterogeneous cloud-RANs[J]. IEEE Internet of Things Journal, 2017, 4(3): 749-759.

- [7] AQEELI E, MOUBAYED A, SHAMI A. Power-aware optimized RRH to BBU allocation in C-RAN[J]. IEEE Transactions on Wireless Communications, 2017, (99): 1.
- [8] YOUSAF F Z, GRAMAGLIA M, FRIDERIKOS V, et al. Network slicing with flexible mobility and QoS/QoE support for 5G Networks[C]//IEEE International Conference on Communications Workshops(ICC Workshops). IEEE, 2017: 1195-1201.
- [9] Study on access traffic steering, switching and splitting support in the 5G system architecture[S]. 2017.
- [10] AFOLABI I, TALEB T, SAMDANIS K, et al. Network slicing & softwarization: a survey on principles, enabling technologies & solutions[J]. IEEE Communications Surveys & Tutorials, 2018, (99): 1.
- [11] SMOLA A J, SCHOLKOPF B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199.
- [12] PAN Q K, TASGETIREN M F, LIANG Y C. A discrete particle swarm optimization algorithm for the no-wait flowshop scheduling problem[J]. Computers and Operations Research, 2008, 35(9): 2807-2839.
- [13] LIU H, XU C Z, JIN H, et al. Performance and energy modeling for live migration of virtual machines[C]//ACM 20th International Symposium on High Performance Distributed Computing (HPDC'11). ACM, 2011: 171-182.
- [14] BOGOMOLOV A, LEPRI B, LARCHER R, et al. Energy consumption prediction using people dynamics derived from cellular network data[J]. EPJ Data Science, 2016, 5(1): 13.
- [15] JIN Y Q, XU X D, WANG Y T, et al. Multi-QoS mobile services guaranteed resource allocation with effective capacity[C]//2017 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2017: 1-6.

[作者简介]



王志朋(1994-),男,河南周口人,哈尔滨工业大学(深圳)电子与信息工程学院硕士生,主要研究方向为网络切片、云无线接入网和资源分配。



曹斌(1985-),男,江西景德镇人,博士,哈尔滨工业大学(深圳)电子与信息工程学院党委副书记、副教授,主要研究方向为无线通信和网络、通信信号处理。



张钦宇(1972-),男,江苏扬州人,博士,哈尔滨工业大学(深圳)电子与信息工程学院院长、教授,主要研究方向为空间通信和无线通信等。