

能耗和时延感知的虚拟化云无线接入网络资源分配机制

王汝言 徐宁宁* 吴大鹏

(重庆邮电大学通信与信息工程学院 重庆 400065)
(重庆高校市级光通信与网络重点实验室 重庆 400065)

摘要: 针对现有虚拟化云无线接入网络(C-RAN)资源利用率低、能耗高、用户服务质量无法得到保证等问题, 该文提出一种能耗和时延感知的虚拟化资源分配机制。根据虚拟化C-RAN的网络特点及业务流量特征, 考虑资源约束和比例公平, 建立能耗和时延优化模型。进而, 利用启发式算法为不同类型虚拟C-RAN和用户虚拟基站分配资源, 完成资源的全局优化配置。仿真结果表明, 所提资源分配机制在提高网络资源利用率的同时, 不但使能耗节省了62.99%, 还使时延降低了32.32%。

关键词: 云无线接入网; 网络虚拟化; 时波分复用无源光网络; 资源分配

中图分类号: TP393.04

文献标识码: A

文章编号: 1009-5896(2019)01-0083-08

DOI: 10.11999/JEIT180063

Energy Consumption and Delay-aware Resource Allocation Mechanism for Virtualization Cloud Radio Access Network

WANG Ruyan XU Ningning WU Dapeng

(School of Telecommunication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

(Optical Communication and Network Key Laboratory of Chongqing, Chongqing 400065, China)

Abstract: To solve the problems of low resource utilization rate, high energy consumption and poor user service quality in the existing virtualized Cloud Radio Access Network (C-RAN), a virtual resource allocation mechanism based on energy consumption and delay is proposed. According to the network and traffic characteristics of the virtualized C-RAN, considering the resource constraints and proportional fairness, an energy consumption and delay optimization model is established. Furthermore, a heuristic algorithm is used to allocate resources for different types of virtual C-RAN and user virtual base stations to complete resource global optimization configuration. Simulation results show that the proposed resource allocation mechanism can effectively save energy by 62.99% and reduce the latency by 32.32% while improving the network resource utilization.

Key words: Cloud Radio Access Network (C-RAN); Network virtualization; Time and Wavelength Division Multiplexing-Passive Optical Network (TWDM-PON); Resource allocation

1 引言

随着用户对标清及高清视频、在线游戏、电话视频等应用的需求不断增加, 无线接入网的业务负载急剧上升, 从而引起能耗大幅度增长^[1,2]。为了

满足“绿色通信”的要求, 在第5代移动网络中引入云无线接入网(Cloud Radio Access Network, C-RAN)以解决能耗问题^[3]。C-RAN将传统基站配置分解成集中式基带处理单元(BaseBand Unit, BBU)池和分布式无线射频头端(Remote Radio Head, RRH)两个部分^[4], 并通过低成本、低时延、低能耗且具有丰富带宽的时波分复用无源光网络(Time and Wavelength Division Multiplexing Passive Optical Network, TWDM-PON)连接BBU池和RRH^[5]。此种方式不但可以降低系统升级和维护的资本和运营支出, 还可以提高频谱利用率。

然而, C-RAN仍然存在一些问题: 仅实现了

收稿日期: 2018-01-16; 改回日期: 2018-09-26; 网络出版: 2018-10-19

*通信作者: 徐宁宁 753113493@qq.com

基金项目: 国家自然科学基金(61371097, 61771082), 重庆市高校创新团队建设计划(CXTDX201601020)

Foundation Items: The National Natural Science Foundation of China (61371097, 61771082), Chongqing Funded Project of Chongqing University Innovation Team Construction (CXTDX201601020)

BBU资源的集中化处理, 对前传波长资源和无线资源无法实现共享; 各服务提供商(Service Provider, SP)独立运行C-RAN, 造成资源利用率不高。为了应对业务负载的频繁变化, 必须考虑可扩展且灵活的资源分配技术实现资源配置, 但是这种动态资源配置方式却难以部署在带有基础设施的网络上^[6]。此时, 网络虚拟化(Network Virtualization, NV)技术通过对计算资源、网络资源、频谱资源、基础设施等进行抽象, 实现不同SP之间的资源共享, 不但可以节省成本、提高资源利用率、降低能耗、易于资源部署, 还可以增强网络扩展性和灵活性^[7]。因此, 将虚拟化技术和C-RAN相联合的架构成为未来无线接入网发展的趋势^[8-12], 此种架构提出虚拟化基站(Virtualized Base Station, VBS)和虚拟化无源光网络(Virtualized Passive Optical Network, VPON)的概念, 并依照业务流量变化联合配置虚拟无线、光网络和BBU计算资源, 不仅提高了系统吞吐量和资源利用率, 而且降低了能耗。

但是, 随着虚拟化C-RAN网络规模增大, 资源的高度集中调度引起严重的延迟, 导致时延敏感业务的服务质量下降。因此, 如何合理地分配虚拟资源使时延和能耗最小, 成为国内外研究人员的关注重点。文献^[13]提出一个两阶段小区干扰感知调度算法最小化系统总体时延, 但没有考虑节省能耗。文献^[14]考虑不同的时延敏感服务, 提出一种基于排队理论的BBU选择新方法, 在资源成本约束下尽量减少BBU池的响应时间, 但没有考虑波长和无线资源分配。文献^[15]针对时延敏感服务的C-RAN下行链路传输, 利用随机梯度算法在前传消耗约束下分配混合协作多点传输的队列感知功率和速率, 实现协作增益与前传能耗之间的灵活折中。文献^[16,17]将网络虚拟化与C-RAN架构进行联合, 实现不同移动网络运营商之间的资源共享, 可以有效提高吞吐量和降低端到端时延, 然而, 所提方法并没有考虑BBU资源的共享调度。

综上所述, 先前文献主要通过对部分虚拟化C-RAN资源进行调度分配去降低时延或者能耗, 然而部分资源分配只考虑局部资源状况, 无法全局最小化时延和能耗。因此, 本文提出了一种能耗和时延感知的虚拟化C-RAN资源分配机制。根据虚拟化C-RAN的网络特点及其业务流量特征, 考虑资源约束和比例公平, 以最小化时延和能耗为目标函数建立多目标整数线性规划模型; 进而, 利用启发式算法基于负载为不同服务类型虚拟化C-RAN(Virtualization CRAN, V-CRAN)公平地分配资源, 然后针对用户的不同服务请求, 在相应V-

CRAN中基于能耗和时延最小的原则为之构建VBS, 从而完成C-RAN中的全局资源优化配置。

2 CRAN-NV系统架构

本文C-RAN与网络虚拟化(Cloud Radio Access Network and Network Virtualization, CRAN-NV)的联合架构使用基于TWDM-PON的C-RAN作为物理网络。RRH接收小区用户各类型服务的射频信号, 然后转发给与其直接相连的光网络单元(Optical Network Unit, ONU)。接着, ONU使用通用公共无线电接口将数字基带信号发送到光线路终端(Optical Line Terminal, OLT)处, 并由代表不同波长的线卡(Line Card, LC)服务。BBU池和OLT位于中心局(Central Office, CO), 在BBU池中的每个BBU与一个LC相关联。交换机位于BBU池之后, 且每个端口都连接到BBU, 可以将流量从一个BBU重定向到其他BBU。显然, 重定向的源和目的BBU之间的数据迁移会持续一段时间, 因此使用直通交换机快速转发数据包而不存储。

接着, 通过虚拟化技术将底层物理网络资源进行虚拟化。总集中控制器通过调度资源分配策略, 组成具有不同特点的V-CRAN。传统SP固定租用需求最大化时的资源并独自占用, 这使业务流量少的SP资源浪费, 业务流量过载的SP得不到满足, 造成资源利用率低下。因此, 本文考虑不同SP组成不同V-CRAN, 并依照负载情况公平地分配C-RAN资源。然后, 为了简化一级集中控制器的功能及降低处理时延, 采用分布式的方式为每个V-CRAN配置一个2级集中控制器。在V-CRAN中, BS被抽象为由虚拟网络资源构成的多个VBS。虚拟化网络资源包括BBU基带处理, 交换机端口, LC和RRH无线资源。其中, BBU基带处理被虚拟化为两种类型的功能实体, 即小区处理(Cell Processing, PHY-cell)和用户处理(User Processing, UP)^[18]。集中控制器为每个RRH建立一个PHY-cell, 并通过改变小区服务VPON的形式完成PHY-cell迁移。此外, 控制器还为具有PHY-cell的RRH所服务的每个用户建立UP, 从而提供基带处理。当小区用户UP需要改变时, 可以通过打开交换机端口去连通原UP所在BBU和新UP所在BBU的方式完成UP迁移。

由于TWDM-PON波长信道的高容量, 一个TWDM-PON支持多个VPON, 每个VPON都由一个连接BBU的LC和多个共享相同波长的ONU组成。接着, 通过配置虚拟化资源, 为每个用户的每一类服务构建一个VBS, 它由一个PHY-cell模块,

一个UP模块、一个通过VPON提供的虚拟前传链路以及小区端RRH和用户组成。其中，PHY-cell模块和UP模块不必在同一个BBU中。

3 能耗和时延感知的虚拟资源分配问题

3.1 能耗和时延模型

CRAN-NV架构的底层物理网络能耗主要来自基础冷却设施、CO网络设备和小区网络设备。基础冷却设施的能耗为 P_{cp} 。CO中的有源网络设备包括内部以太网交换机，BBU和LC。交换机的能量消耗 P_{sw} 分为基础运行能耗 P_b 和业务相关能耗 P_{pt} 两个部分，其中业务相关能耗随交换机激活的端口数量 n_p 变化，是端口能耗 P_p 和激活端口数目之积。对于BBU能耗，本文不考虑处理器动态电压和频率缩放，所以BBU仅在激活时消耗恒定的能量 P_B 。类似地，有源LC也消耗恒定的能量 P_L 。小区的活跃网络设备包括ONU和RRH，仅考虑激活时所消耗的能量 P_O 和 P_R 。因此，CRAN-NV架构的能耗为

$$P = P_{cp} + sw \cdot (P_b + P_p \cdot n_p) + (P_R + P_O) \cdot g + \sum_{s=1}^{\|S\|} \sum_{w=1}^{\|W_s\|} (P_L \cdot lc_w^s + P_B \cdot up_w^s) \quad (1)$$

其中，式(1)右边第1项是基础冷却设施能耗；第2项是交换机能耗， $n_p = \sum_{s \in S} \sum_{w \in W_s} pt_w^s$ 是激活的交换机端口数目，sw是布尔变量，当交换机是激活状态时为1；第3项是小区网络设备能耗， g 是小区数目；第4项是激活LC和BBU的能耗， $\|S\|$ 是服务类型数目， $\|W_s\|$ 表示 s 类型服务中BBU的数目， lc_w^s 和 up_w^s 分别是两个布尔变量，表示服务类型 s 的第 w 个LC和BBU是否被激活，值得注意的是当BBU被激活时LC不一定被激活，而LC被激活时相连接的BBU一定被激活。

接着，在CRAN-NV架构中，用户请求的响应时延包括调度、处理和传输时延。本文将集中控制器中的总平均等待时延表示为调度时延。又因为本文考虑两级控制器进行调度，且用户资源请求在2级控制器得到分配，所以总控制器调度时延忽略不计，仅考虑2级控制器调度时延。依据排队理论，总调度时延为

$$T_d = \sum_{s \in S} \frac{\lambda_s}{\mu_s - \lambda_s} \quad (2)$$

其中， λ_s 表示 s 类型服务的业务到达率， μ_s 表示在 s 类型服务控制器中的平均业务处理率。

基带处理时延是所选BBU完成基带单元处理

所需的时间。本文将分为两种情况。当小区 i 中 s 类型服务请求 j 在主BBU w 进行服务，即 $z_{i,j,w}^s = 0$ 时，处理时延为服务请求带宽资源 $b_{i,j}^s$ 与处理BBU UP剩余资源 $RB_w^{i,j}$ 的比值；若小区 i 中 s 类型服务请求 j 在主BBU w 中发生UP迁移，即 $z_{i,j,w}^s = 1$ 时，处理时延为请求带宽与处理BBU UP剩余资源的比值加上交换机端口打开时延 T_{sw} 。因此，总基带处理时延如式(3)：

$$T_p = \sum_{s \in S} \sum_{w \in W_s} \sum_{i \in I} \sum_{j \in J_i} \left[(1 - z_{i,j,w}^s) \cdot u_{i,j,w}^s \cdot \frac{b_{i,j}^s}{RB_w^{i,j}} + z_{i,j,w}^s \cdot u_{i,j,w}^s \cdot \left(\frac{b_{i,j}^s}{RB_w^{i,j}} + T_{sw} \right) \right] \quad (3)$$

其中， $u_{i,j,w}^s$ 是一个布尔变量，当小区 i 中服务请求 j 的 s 服务是在BBU w 中进行UP基带处理时，为1。

最后，传输时延是在前传链路和无线链路上进行基带传输的时间。由于RRH位于小区站点，无线链路传输时延可以忽略不计。又因为采用低时延TWDM-PON作为前向传输，所以假设每MB传输时延 T_m 是一个常量。那么，总传输时延表示为

$$T_t = \sum_{s \in S} \sum_{i \in I} \sum_{j \in J_i} b_{i,j}^s \cdot T_m \quad (4)$$

因此，所有用户请求的总响应时延表示为

$$T = T_d + T_p + T_t \quad (5)$$

3.2 数学模型

3.2.1 问题描述

本文提出CRAN-NV架构，并通过集中控制器共享调度虚拟化C-RAN资源。由于集中式调度复杂度过高，根据服务类型将C-RAN划分为多个V-CRAN，且各V-CRAN之间相互独立。1级集中控制器基于负载流量对各V-CRAN公平地分配虚拟资源。接着，为了方便在V-CRAN中进一步完成独立资源配置，本文为其各配备一个2级集中控制器。2级集中控制器根据用户负载，对VBS的构建以及资源的合理分配等所有决策做出决定。在静态场景中，以能耗最少为原则为各小区分配VPON，接着根据总体激活BBU最少和时延限制对BBU UP的负载进行重新平衡。但是实际生活中业务流量随机变化，所以进一步考虑符合实际的复杂动态资源分配，而静态机制用于执行动态分配前业务资源的初始化。因此，下文结合能耗和时延两种因素，使用多目标整数线性规划模型来解决资源分配问题。

3.2.2 目标函数及约束条件

本文以最小化总时延和能耗为目标函数：

$$\min T \quad (6)$$

$$\min P \quad (7)$$

下面为虚拟资源分配时所考虑的约束条件:

$$\sum_{i \in I} x_{i,w}^s \cdot b_i^s \leq \rho_1 C_V, \forall w \in W_s, s \in S \quad (8)$$

$$\sum_{w \in W_s} x_{i,w}^s = 1, \forall i \in I, s \in S \quad (9)$$

$$\sum_{i \in I} \sum_{j \in J_i} u_{i,j,w}^s \cdot b_{i,j}^s \leq \rho_2^s C_U, \forall w \in W_s, s \in S \quad (10)$$

$$\sum_{w \in W_s} u_{i,j,w}^s = 1, \forall i \in I, j \in J_i, s \in S \quad (11)$$

$$\sum_{i \in I} \sum_{j \in J_i} k_{i,j}^s \cdot b_{i,j}^s \leq \rho_3 C_S, \forall s \in S \quad (12)$$

其中, W_s 是服务类型 s 的 VPON 波长集合, 也是 LC 和 BBU 的集合; S 是所有的服务类型集合; I, J_i 分别是小区集合、小区 i 的用户集合; C_V, C_U, C_S 分别表示 VPON 波长容量、UP 带宽容量、交换机最大负载容量; b_i^s 表示小区 i 中用户 j 的 s 类型服务请求带宽资源; b_i^s 是小区 i 中所有用户的 s 类型服务总请求带宽资源; ρ_1, ρ_3 分别是为了保证 VPON 和以太网交换机可靠性设置的容量约束参数; ρ_2^s 是为了保证 s 服务类型处理时延限制设置的容量约束参数; $x_{i,w}^s$ 是二进制变量, 若小区 i 的 s 类型服务在 VPON w 进行基带数据传输, 为 1; $u_{i,j,w}^s$ 也是二进制变量。若小区 i 属于 s 类型的服务请求 j 是在 BBU UP w 进行基带处理, 为 1; $k_{i,j}^s$ 表示若小区 i 属于 s 类型的服务请求 j 发生了 UP 迁移, 为 1。

约束式(8)–式(12)表示 VBS 构建时各类型基础设施资源的容量约束限制。约束式(8)保证了包含多个 PHY-cell 标识小区的 VPON 负载不大于 VPON 最大容量限制。约束式(9)表明一个小区的一种服务类型只能在一个 VPON 中进行基带数据传输。约束式(10)保证在 BBU 上的处理负载不大于 UP 容量限制。约束式(11)表明一个小区用户的一种服务类型请求只能在一个 BBU UP 中进行基带数据处理。约束式(12)表明交换机服务的所有 BBU UP 中迁移的总负载不大于交换机能承受的最大容量。

$$lc_w^s = \left\lceil \sum_{i \in I} x_{i,w}^s / \|I\| \right\rceil, \forall w \in W_s, s \in S \quad (13)$$

$$up_w^s = \left\lceil \sum_{i \in I} \sum_{j \in J_i} u_{i,j,w}^s / \left(\|I\| \cdot \sum_{i \in I} \|J_i\| \right) \right\rceil, \forall w \in W_s, s \in S \quad (14)$$

$$pt_w^s = \left\lceil \sum_{i \in I} \sum_{j \in J_i} z_{i,j,w}^s / \left(\|I\| \cdot \sum_{i \in I} \|J_i\| \right) \right\rceil, \forall w \in W_s, s \in S \quad (15)$$

$$sw = \left\lceil \frac{\sum_{s \in S} \sum_{i \in I} \sum_{j \in J_i} k_{i,j}^s}{\left(\|I\| \cdot \sum_{i \in I} \|J_i\| * \|S\| \right)} \right\rceil \quad (16)$$

其中, lc_w^s, up_w^s, pt_w^s 是二进制变量, 分别表示若当前 s 类服务的 LC w , BBU UP w , 交换机接口 w 是激活状态, 为 1; $z_{i,j,w}^s$ 表示若小区 i 属于 s 类型的服务请求 j 是在 BBU w 进行 UP 处理迁移, 为 1; sw 表示若当前以太网交换机是激活状态, 为 1; $\|\bullet\|$ 表示某集合元素的总数; $\lceil \bullet \rceil$ 表示向上取整。

约束式(13)–式(16)用于判断主要基础设施设备是否被激活。基础设施设备包括以太网交换机, BBU, LC, ONU 和 RRH。本文假设给定的 ONU 和 RRH 均处于激活状态, 所以只需对剩下的 3 种设备进行判断即可。约束式(13)表示只要有任意的小区在 VPON 上进行数据传输, 则 LC 始终保持激活状态。约束式(14)表示若任意小区的一个请求在 BBU UP 上进行基带数据处理, 则 BBU 保持激活状态。约束式(15)和式(16)表明无任何 UP 发生迁移时, 所有交换机端口处于休眠状态, 且以太网交换机也退出激活状态, 进入休眠。

$$z_{i,j,w}^s = |u_{i,j,w}^s - x_{i,w}^s|, \forall w \in W_s, s \in S \quad (17)$$

$$k_{i,j}^s = \left\lceil \sum_{w \in W_s} z_{i,j,w}^s / \|W_s\| \right\rceil, \forall i \in I, j \in J_i, s \in S \quad (18)$$

约束式(17)和式(18)用于描述 BBU UP 是否发生迁移, 增加额外时延。其中, $|\bullet|$ 表示取绝对值。约束式(17)表示当用户服务请求的主 BBU 与请求基带处理 UP 所在 BBU 不是同一个 BBU 时, 请求服务 UP 发生迁移。约束式(18)表示小区的请求服务 UP 发生迁移, 即不在主 BBU UP 进行基带处理。

$$\begin{aligned} & \sum_{w \in W_s} \left((1 - z_{i,j,w}^s) \cdot u_{i,j,w}^s \cdot \frac{b_{i,j}^s}{RB_w^s} \right. \\ & \quad \left. + z_{i,j,w}^s \cdot u_{i,j,w}^s \cdot \left(\frac{b_{i,j}^s}{RB_w^s} + T_{sw} \right) \right) \\ & \quad + \frac{b_{i,j}^s}{\mu_s - \lambda_s} + b_{i,j}^s \cdot T_m \leq T_{\max}^s, \\ & \quad \forall i \in I, j \in J_i, s \in S \end{aligned} \quad (19)$$

约束式(19)用于描述时延限制, 即小区中服务请求的总时延均不大于标准最大时延 T_{\max}^s 。其中, 式(19)左边第 1 项表示基带处理时延, 第 2 项和第 3 项分别表示集中控制器调度时延和传输时延。

$$\begin{aligned} & \|W_1\| : \|W_2\| : \dots : \|W_{\|S\|}\| \\ & = (1 \pm \alpha) (b_1 : b_2 : \dots : b_{\|S\|}) \end{aligned} \quad (20)$$

$$\sum_{s \in S} \|W_s\| = \|W\| \quad (21)$$

$$b_s = \sum_{i \in I} \sum_{j \in J_i} b_{i,j}^s, \forall s \in S \quad (22)$$

约束式(20)–式(22)用于描述比例公平限制。约束式(20)，式(21)表示不同V-CRAN分配的虚拟资源与其负责的服务类型负载成比例，且不大于C-RAN虚拟总资源。其中， α 是一个整数，可以使严格比例约束得到松弛。约束式(22)表示某类型服务的总负载。

对待具有多目标的线性规划问题，通常采用传统优化、进化和粒子群算法等方法在非劣解中寻找最优解，但这些算法均具有较高的复杂性，一旦网络规模稍大，将造成严重的时延损耗甚至计算不可行。因此，本文采用简单的启发式算法将资源分配问题分为几个小问题解决，从而减少算法复杂性。

4 资源分配算法

4.1 静态资源分配机制

本节提出能耗和时延感知的静态VBS构建机制(Energy consumption and Delay-Aware Static VBS Formation Mechanism, EDA-SVFM)。在获知小区业务信息后，1级集中控制器依照业务类型将其划分为S类，且每一类业务由一个V-CRAN和2级控制器负责。为了保证比例公平，本文基于各服务负载为V-CRAN分配虚拟资源。接着，2级控制器完成V-CRAN中VBS的构建，对各小区首先按照能耗最少的原则分配PHY-cell；然后依据处理时延约束和激活BBU最少的原则对过载BBU UP进行迁移、交换、激活新BBU服务，从而完成UP重新平衡操作。接着，将已经完成PHY-cell分配和UP重新平衡的BBU按照用户UP负载请求降序排序，而BBU依照这个顺序对用户请求进行基带数据处理。最后，统计活动BBU、端口和LC的数目，以及各小区用户服务请求的处理时延、调度时延和传输时延，利用式(1)和式(5)分别计算能耗和时延。值得注意的是，加入参数 ρ 来保证BBU UP处理的可靠性，且不同类型的服务具有不同的 ρ 。

4.2 动态资源分配机制

在实际生活中，小区业务流量不是固定不变的，所以考虑与静态情况不同的动态资源分配，即本节提出能耗和时延感知的动态VBS构建机制(Energy consumption and Delay-Aware Dynamic VBS Formation Mechanism, EDA-DVFM)。当一个请求到达时，1级控制器先判断是否对V-CRAN重新配置虚拟资源，然后根据服务类型将请求分配

给相应2级控制器。2级控制器接收到请求信息后，先判断其所在小区的PHY-cell或者新建一个PHY-cell，然后基于能耗和BBU处理时延最少选择合适的UP进行基带处理。

4.2.1 比例公平资源配置机制

为了降低复杂度，将虚拟C-RAN划分为S个V-CRAN，进行区分业务服务。然后，为了适应业务流量的动态变化，按照各类型业务的负载公平地为V-CRAN分配虚拟资源。当一个用户请求到达后，1级集中控制器先判断连接和服务类型，并记录上一时刻各类型服务的负载信息和V-CRAN资源分配信息。如果请求属于中断请求类型， $m=1$ ，则将减去请求负载大小的当前时刻各类型服务负载比例与上一请求到达后的资源分配比例作比较，若大致相等，则资源配置保持不变，且采用动态VBS构建机制构建用户服务VBS；若不等，则根据当前负载比例和先前配置信息重新进行V-CRAN资源分配。如果请求属于连接请求类型， $m=-1$ ，则当前时刻服务负载为加上请求负载后的结果，其具体操作与中断请求类型类似，不再加以赘述。

4.2.2 PHY-cell选择机制

PHY-cell选择机制主要是为未配置PHY-cell的小区选择适当的主BBU和VPON。具体算法为：首先对属于j类服务的VPON按照剩余带宽资源由大到小排列。然后，将最大剩余带宽资源与请求带宽进行对比，若小于，则激活新的BBU和与之相连的VPON进行服务，并将小区PHY-cell写入到BBU中；否则将与具有最大剩余带宽资源的VPON相连的BBU作为主服务BBU。

4.2.3 UP选择机制

UP选择机制主要是为请求选择在BBU中的基带处理单元。本文结合能耗和时延两种因素选取合适的UP，从而实现数据的高效传输，同时降低运营商的成本。在算法中，首先对属于j类服务的激活BBU按照UP剩余带宽资源由大到小排列。然后，计算请求数据在活动BBU中的处理时间。考虑将请求数据大小与排序后的BBU UP剩余带宽资源进行对比，若出现第n个BBU剩余资源小于请求数据最小带宽的情况，则将n至W_J_ACTIVE的BBU处理时延均设置为无穷大，这样可以有效降低算法复杂度。接着，如果存在处理时延小于无穷大的BBU，将请求的主BBU的处理时延与最小的其他BBU处理时延加上交换机端口时延作对比，从而选择合适的BBU进行处理；否则，当所有的激活BBU处理时延均是无穷大时，激活一个新BBU专用于UP服务，其中，这个BBU没有PHY-cell模块，且与之相连的PON模块不被激活。

4.2.4 动态VBS构建机制

VBS构建机制是2级集中控制器的主控算法,完成虚拟化基站的总体构建。其具体如下:首先利用前一节中的静态VBS构建算法对某一时间的业务进行初始化。然后,当一个用户请求到达时,判断请求的服务和连接类型,并执行V-CRAN比例公平资源分配算法为其V-CRAN重新配置资源。如果请求属于中断请求类型,那么首先撤销请求在BBU的UP和VPON中所占的资源,然后将撤销请求后计算的理论BBU激活数目与撤销请求前激活的BBU数目相比较,如果小于,且当请求所在的BBU在撤销请求资源后的剩余资源与UP容量相等且理论数目与撤销前数目相差1时,就关闭请求所在的BBU;如果大于或者等于,就保留原来的配置不变,等待下一请求的到达。如果请求属于连接请求类型,先判断请求所在小区是否存在主BBU,即PHY-cell是否存在,若存在,观察请求所在的VPON在加上请求资源后的剩余资源是否大于VPON约束容量,如果大于,则调用PHY-cell选择算法为请求小区重新选择主服务BBU,否则调用UP选择算法对请求UP进行设置;若不存在,则调用PHY-cell选择算法为请求所在小区选择合适的主BBU。

5 性能测试与分析

5.1 仿真参数设置

本文采用MATLAB R2016a仿真平台对所提出的资源分配机制进行性能评估。考虑4种服务,即VoIP业务、IPTV业务、网页搜索、视频会议。由C-RAN白皮书可知,网页和视频业务分别占总流量的38%和45%。为了简便起见,假定3种视频业务分别各占15%。考虑两种网络场景,在小(或大)规模网络中,设置11(或100)个BBU和小区。每个波长容量为10Gbps, BBU UP容量为5Gbps,每个小区中的用户流量服从[100Mbps, 1Gbps]上的均匀分布。业务请求到达时间服从泊松分布,到达时间间隔独立且服从负指数分布。通过文献[9]和文献[14]可得, P_b, P_p, P_{cp} 分别为50W, 15W及500W; P_R, P_O 分别为20W及7.7W; P_L, P_B 分别为5W和100W; $b^{Vo}, b^{IP}, b^w, b^{vi}$ (MB) 分别为3, 5, 2及7; $T_m^{Vo}, T_m^{IP}, T_m^w, T_m^{vi}$ (ms) 分别为150, 200, 1000及150; T_m 为250 μ s。为了有效衡量所提资源分配机制的性能, 本文将其与其他3种机制进行对比, 即不具备VBS构建的C-RAN(C-RAN without VBS Formation, CRAN-NVF)^[9], D-RAN和能耗感知的VBS构建机制(Energy consumption-Aware VBS Formation Mechanism, EA-VFM)^[10]。下面就从不

同机制随时间的能耗、时延以及时延和能耗随容量限制的变化等3个方面进行仿真分析。

5.2 静态资源分配的能耗和时延性能

5种算法在不同业务流量下的能耗比较如图1所示。显然, 5种算法的能耗均随流量的增加而增加。另一方面, D-RAN由于在小区端单独设置BBU冷却设施, 导致其能耗是最大的。CRAN-NVF由于BBU池的集中, 所以只有一个冷却设施, 其能耗比D-RAN少; 但是, CRAN-NVF没有合理的资源分配策略, 导致所有的BBU在整个过程中始终处于激活状态, 造成多余的能量消耗。显然, EA-VFM随各时段业务负载的变化进行资源配置, 有效地降低了能耗。而本文的EDA-SVFM和EA-VFM具有相似的能耗性能, 即与CRAN-NVF相比, 能耗最多降低62.99%; 与D-RAN相比, 能耗平均降低75.52%。此外, EDA-SVFM机制通过启发式算法所得的次优解能耗与线性规划算法所得的最优解能耗相比, 最多相差8.37%。

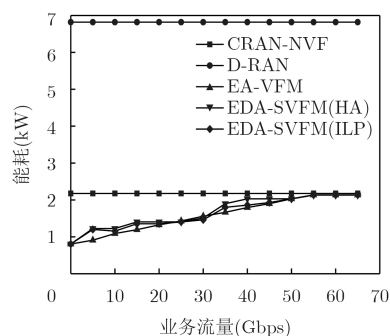


图1 小规模网络中不同业务流量下的能耗比较

为了清晰地观察5种算法的时延变化趋势, 将时延性能进行归一化处理, 即均一化平均时延等于实际平均时延/最大平均时延。如图2所示, 5种算法的时延均随流量的增加而增加。另一方面, D-RAN由于BBU和RRH均在小区端, 没有传输时延和集中控制器调度时延, 所以时延是最小的。CRAN-NVF将所有BBU集中在中心局的BBU池, 通过光纤传输基带信号, 所以具有传输时延和调度时延; 但是BBU池的BBU没有共享资源, 每个小区仍需要一个专用的BBU和LC为之服务, 导致所有的BBU均处于激活状态, 所以CRAN-NVF时延是次小的。接着, EA-VFM通过关闭一些不必要的BBU去节省能耗, 但同时也使相同BBU处理更多的业务请求, 导致时延增加。而本文EDA-SVFM在考虑能耗的同时, 又考虑时延的影响, 采用分布式机制合理分配资源, 从而有效降低时延。与EA-VFM相比, 时延平均降低了32.32%。此外, EDA-

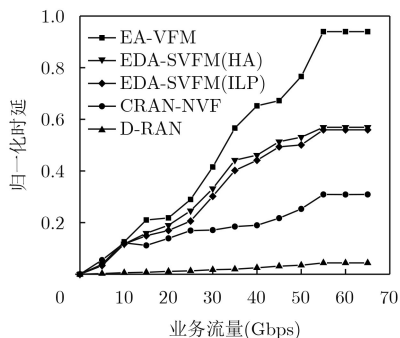


图2 小规模网络中不同业务流量下的总业务时延比较

SVFM通过启发式算法所得的次优解时延与线性规划算法所得的最优解时延相比，最多相差8.05%。

5.3 不同权重系数下的能耗和时延权衡

本节给出EDA-SVFM算法于12点时刻在小规模网络中时延和能耗的权衡。从图3可以看出，随着权重系数的增加，能耗呈现下降趋势，最后趋于稳定，这是因为权重系数越大，每个BBU UP的可用容量越大，从而可以激活较少的BBU去承载处理一定的业务，随之将会降低能耗。而另一方面，随着权重系数的增加，时延越来越大。因为BBU UP可用容量的增大，使更多的用户在同一BBU UP进行业务处理，导致处理时延增加。特殊地，当权重系数为0.5时，逐渐下降的能耗和逐渐上升的时延进行交汇，产生平衡点。

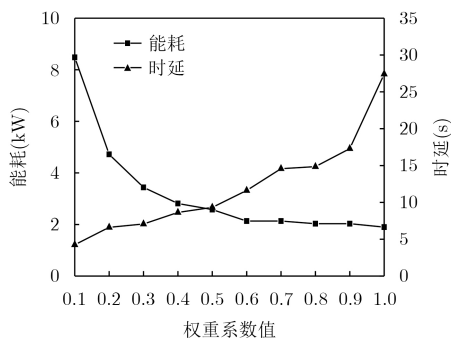


图3 12点时刻时延和能耗的权衡

5.4 动态资源分配的能耗和时延性能

本节给出EDA-DVFM和EDA-SVFM在大规模网络场景下一天期间的能耗和时延比较。从图4可以看出，特殊地，在一些业务流量低和空闲的时间段，静态算法和动态算法能耗完全相同。而在一些流量高的时间段，动态算法比静态算法更优，因为在同一时间段中，静态算法按照区间最大的业务流量固定分配资源，而动态算法根据业务请求的到达和服务时间进行动态资源分配，从而动态算法可以进一步节省能耗。

从图5可以看出，动态算法时延总体小于静态

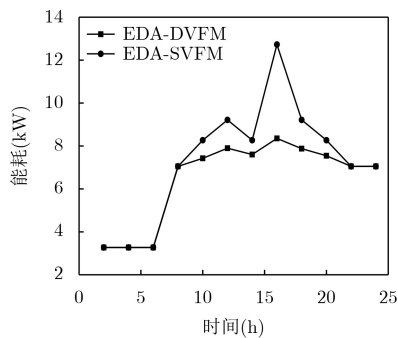


图4 动态和静态算法的能耗比较

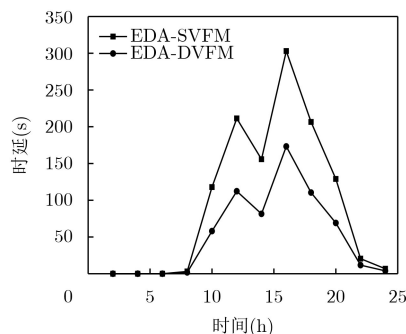


图5 动态和静态算法的总业务时延比较

算法时延，它主要有两方面原因导致这样的结果。首先，传输时延和调度时延随流量增加而增大，静态算法在每个时间段选取最高的流量固定分配资源，而动态算法依据动态变化的到达业务请求流量进行资源分配，所以静态算法产生更多时延；其次，静态算法是对某一时间段中的所有用户请求进行UP资源分配，而动态算法对每一个到达的用户请求根据当前资源状况选取最优的UP进行分配，所以，静态算法的基带处理时延比动态算法更大。因此，静态总时延大于动态总时延。

6 结束语

为了解决现有虚拟化C-RAN资源利用率低、能耗高、用户服务质量无法得到保证等问题，本文提出一种能耗和时延感知的虚拟化C-RAN资源分配机制。首先，给出基于TWDM-PON的C-RAN和NV相联合的新框架，并对VBS和VPON进行定义说明；然后考虑资源约束和比例公平，以最小化能效和时延为目标函数，构建多目标线性规划模型；最后提出简单的启发式算法解决多目标不能得出最优解的问题。结果表明，所提机制与D-RAN, CRAN-NVF, EA-VFM相比，有效降低了能耗和时延。

参考文献

[1] CISCO. Global mobile data traffic forecast update, 2016–2021 White Paper[R]. San Jose: CISCO/14652 72001663118, 2017.

- [2] AROUK O, KSENTINI A, and TALEB T. Group paging-based energy saving for massive MTC accesses in LTE and beyond networks[J]. *IEEE Journal on Selected Areas in Communications*, 2016, 34(5): 1086–1102. doi: [10.1109/JSAC.2016.2520222](https://doi.org/10.1109/JSAC.2016.2520222).
- [3] HOSSAIN E and HASAN M. 5G cellular: Key enabling technologies and research challenges[J]. *IEEE Instrumentation & Measurement Magazine*, 2015, 18(3): 11–21. doi: [10.1109/MIM.2015.7108393](https://doi.org/10.1109/MIM.2015.7108393).
- [4] China Mobile Research Institute. C-RAN-The road towards green RAN[R]. Beijing: CMRI/Version 3.0, 2013.
- [5] THOMAS P. Next generation mobile fronthaul architectures[C]. 2015 Optical Fiber Communications Conference and Exhibition, Los Angeles, USA, 2015: 1–3.
- [6] MOHAMAD K, ARAFAT A D, MOHAMED S, *et al.* A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks[J]. *IEEE Access*, 2017, 5: 20814–20827. doi: [10.1109/ACCESS.2017.2746666](https://doi.org/10.1109/ACCESS.2017.2746666).
- [7] LIANG Chengchao and YU F R. Wireless network virtualization: A survey, some research issues and challenges[J]. *IEEE Communications Surveys & Tutorials*, 2015, 17(1): 358–380. doi: [10.1109/COMST.2014.2352118](https://doi.org/10.1109/COMST.2014.2352118).
- [8] WANG Xinbo, CAVDAR C, WANG Lin, *et al.* Joint allocation of radio and optical resources in virtualized cloud RAN with CoMP[C]. 2016 IEEE Global Communications Conference (GLOBECOM), Washington, USA, 2016: 1–6.
- [9] WANG Xinbo, THOTA S, TORNATORE M, *et al.* Green virtual base station in optical-access-enabled cloud-RAN[C]. 2015 IEEE International Conference on Communications (ICC), London, UK, 2015: 5002–5006.
- [10] WANG Xinbo, THOTA S, TORNATORE M, *et al.* Energy-efficient virtual base station formation in optical-access-enabled cloud-RAN[J]. *IEEE Journal on Selected Areas in Communications*, 2016, 34(5): 1130–1139. doi: [10.1109/JSAC.2016.2520247](https://doi.org/10.1109/JSAC.2016.2520247).
- [11] WANG Xinbo, CAVDAR C, WANG Lin, *et al.* Virtualized cloud radio access network for 5G transport[J]. *IEEE Communications Magazine*, 2017, 55(9): 202–209. doi: [10.1109/MCOM.2017.1600866](https://doi.org/10.1109/MCOM.2017.1600866).
- [12] TAN Zhongwei, YANG Chuanchuan, and WANG Ziyu. Energy consume analysis for ring-topology TWDM-PON fronthaul enabled cloud RAN[J]. *IEEE Journal of Lightwave Technology*, 2017, 35(20): 4526–4534. doi: [10.1109/JLT.2017.2745998](https://doi.org/10.1109/JLT.2017.2745998).
- [13] LI Yi, GURSOY M C, and VELIPASALAR S. Intercell interference-aware scheduling for delay sensitive applications in C-RAN[OL]. <http://arxiv.org/abs/1708.00852v1>, 2017.
- [14] AMEL A, SOUMAYA H, LOUTFI N, *et al.* Minimization of delays in multi-service cloud-RAN BBU pools[C]. 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 2017: 1846–1850.
- [15] LI Jian, PENG Mugen, CHEN Aolin, *et al.* Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks[J]. *IEEE Systems Journal*, 2017, 11(4): 2267–2278. doi: [10.1109/JSYST.2014.2364252](https://doi.org/10.1109/JSYST.2014.2364252).
- [16] NIU Binglai, ZHOU Yong, HAMED S M, *et al.* A dynamic resource sharing mechanism for cloud radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2017, 15(12): 8325–8338. doi: [10.1109/TWC.2016.2613896](https://doi.org/10.1109/TWC.2016.2613896).
- [17] IMAD A S, MATTEO A, HENRIK C, *et al.* Envisioning spectrum management in virtualised C-RAN[C]. 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, USA, 2017: 1–6.
- [18] BERND H, FARIBORZ D, HEIDRUN G L, *et al.* Radio base stations in the cloud[J]. *Bell Labs Technical Journal*, 2013, 18(1): 129–152. doi: [10.1002/bltj.21596](https://doi.org/10.1002/bltj.21596).
- 王汝言: 男, 1969年生, 博士, 教授, 研究方向为泛在网络、全光网络理论与技术、多媒体信息处理等。
- 徐宁宁: 女, 1993年生, 硕士生, 研究方向为光无线融合网络。
- 吴大鹏: 男, 1979年生, 博士, 教授, 研究方向为泛在无线网络、社会计算、互联网服务质量控制等。