

H-CRAN网络下联合拥塞控制和资源分配的网络切片动态资源调度策略

唐伦^① 魏延南^{*①} 谭颀^① 唐睿^② 陈前斌^①

^①(重庆邮电大学通信与信息工程学院移动通信技术重点实验室 重庆 400065)

^②(重庆大学光电工程学院 重庆 400044)

摘要: 针对异构云无线接入网络(H-CRAN)网络下基于网络切片的在线无线资源动态优化问题, 该文通过综合考虑业务接入控制、拥塞控制、资源分配和复用, 建立一个以最大化网络平均和吞吐量为目标, 受限于基站(BS)发射功率、系统稳定性、不同切片的服务质量(QoS)需求和资源分配等约束的随机优化模型, 并进而提出了一种联合拥塞控制和资源分配的网络切片动态资源调度算法。该算法会在每个资源调度时隙内动态地为性能需求各异的网络切片中的用户分配资源。仿真结果表明, 该文算法能在满足各切片用户QoS需求和维持网络稳定的基础上, 提升网络整体吞吐量, 并且还可通过调整控制参量的取值实现时延和吞吐量间的动态平衡。

关键词: 异构云无线接入网络; 网络切片; 资源分配; 李雅普诺夫随机优化

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2020)05-1244-09

DOI: 10.11999/JEIT190439

Joint Congestion Control and Resource Allocation Dynamic Scheduling Strategy for Network Slices in Heterogeneous Cloud Radio Access Network

TANG Lun^① WEI Yanan^① TAN Qi^① TANG Rui^② CHEN Qianbin^①

^①(Key Laboratory of Mobile Communication Technology, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

^②(School of Measurement and Control Technology and Instruments, Chongqing University, Chongqing 400044, China)

Abstract: For online dynamic radio resources optimization for network slices in Heterogeneous Cloud Radio Access Network (H-CRAN), by comprehensively considering traffic admission control, congestion control, resource allocation and reuse, the problem is formulated as a stochastic optimization programming which maximizes network average total throughput subject to Base Station (BS) transmit power, system stability, Quality of Service (QoS) requirements of different slices and resource allocation constraints. Then, a joint congestion control and resource allocation dynamic scheduling algorithm is proposed which will dynamically allocate resources to users in network slices with distinct performance requirements within each resource scheduling time slot. The simulation results show that the proposed algorithm can improve the network overall throughput while satisfying the QoS requirement of each slice user and maintaining network stability. Besides, it could also flexibly strike a dynamic balance between delay and throughput by simply tuning an introduced control parameter.

Key words: Heterogeneous Cloud Radio Access Network (H-CRAN); Network Slice (NS); Resource allocation; Lyapunov stochastic optimization

1 引言

移动运营商期望通过下一代网络提供颠覆式的

速率、极低延迟和动态配置能力, 以满足日益增长的用户需求、更好地支持增强现实、物联网和智慧医疗等新兴应用案例, 更好地迎合跨移动、住宅和商业市场的全新服务趋势^[1]。众多科研学者和科技公司均提出了新颖的网络技术和架构, 其中, 网络切片技术和异构云无线接入网络(Heterogeneous Cloud Radio Access Network, H-CRAN)受到了广泛关注和深入研究。

基于虚拟化技术的网络切片是指在通用的底层物理网络构建逻辑上相对独立的虚拟网络, 每个虚

收稿日期: 2019-06-17; 改回日期: 2020-01-03; 网络出版: 2020-01-11

*通信作者: 魏延南 weiyannan_cqupt@163.com

基金项目: 国家自然科学基金(61571073), 重庆市教委科学技术研究项目(KJZD-M201800601)

Foundation Items: The National Natural Science Foundation of China (61571073), The Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M201800601)

拟网络称为一个网络切片。3GPP已经明确网络切片作为实现5G网络宏伟目标的关键技术之一^[2]，国际电信联盟ITU (International Telecommunication Union)也为5G定义了3大应用场景：移动增强宽带(enhanced Mobile BroadBand, eMBB)、大规模机器类通信(massive Machine-Type Communication, mMTC)和超可靠低时延通信(Ultra-Reliable and Low-Latency Communications, URLLC)^[3]。H-CRAN即在传统C-RAN网络中引入大功率节点(High Power Node, HPN)，并将控制平面从BBU池抽离到HPN中，实现了控制平面和数据平面的分离、避免了小区间的频繁切换和相应的控制信令开销并缓解了前向回传(fronthaul)链路容量压力^[4]。

已有许多研究工作深入探究了网络切片技术和H-CRAN架构，文献^[5]研究了下行非正交多址接入(Non-Orthogonal Multiple Access, NOMA)系统中高效率的动态资源优化方案，其联合子信道和功率分配，实现在维持网络稳定的条件下最大化系统能效。但是，文献^[5]没有考虑用户间存在各异的QoS需求，也没有探究NOMA技术和H-CRAN架构相结合所具有的潜在优势。文献^[6]研究了雾无线接入

网络中基于时延感知的网络切片无线资源优化问题，建立了一个多目标优化模型以实现在维持网络稳定和满足各切片用户QoS约束的条件下最大化网络平均加权效用。然而，只有当随机业务到达率在网络的容量域内时，文献^[6]中的系统才是稳定的，其没有考虑接入控制等控制策略以限制接入到网络中的流量，因而在随机业务到达突发过大时，网络仍有拥塞的可能。

为了解决上述问题，针对H-CRAN网络中基于网络切片的在线无线资源动态优化，本文通过联合考虑接入控制、拥塞控制、资源分配和复用以研究网络平均和吞吐量最大化问题，进而提出了一种联合拥塞控制和资源分配的网络切片动态资源调度算法。该算法能在满足各切片用户QoS需求和维持网络稳定的前提下，提升网络整体吞吐量，并且还可通过调整控制参数的取值实现时延和吞吐量间的动态平衡。

2 系统模型和问题描述

2.1 物理层模型

如图1所示，考虑一个H-CRAN网络的下行传输场景^[7-9]，令 m 表示高功率节点HPN， $\mathcal{K} = \{1, 2, \dots, K\}$ 表示网络中所有RRHs的集合。网络服务的所有用

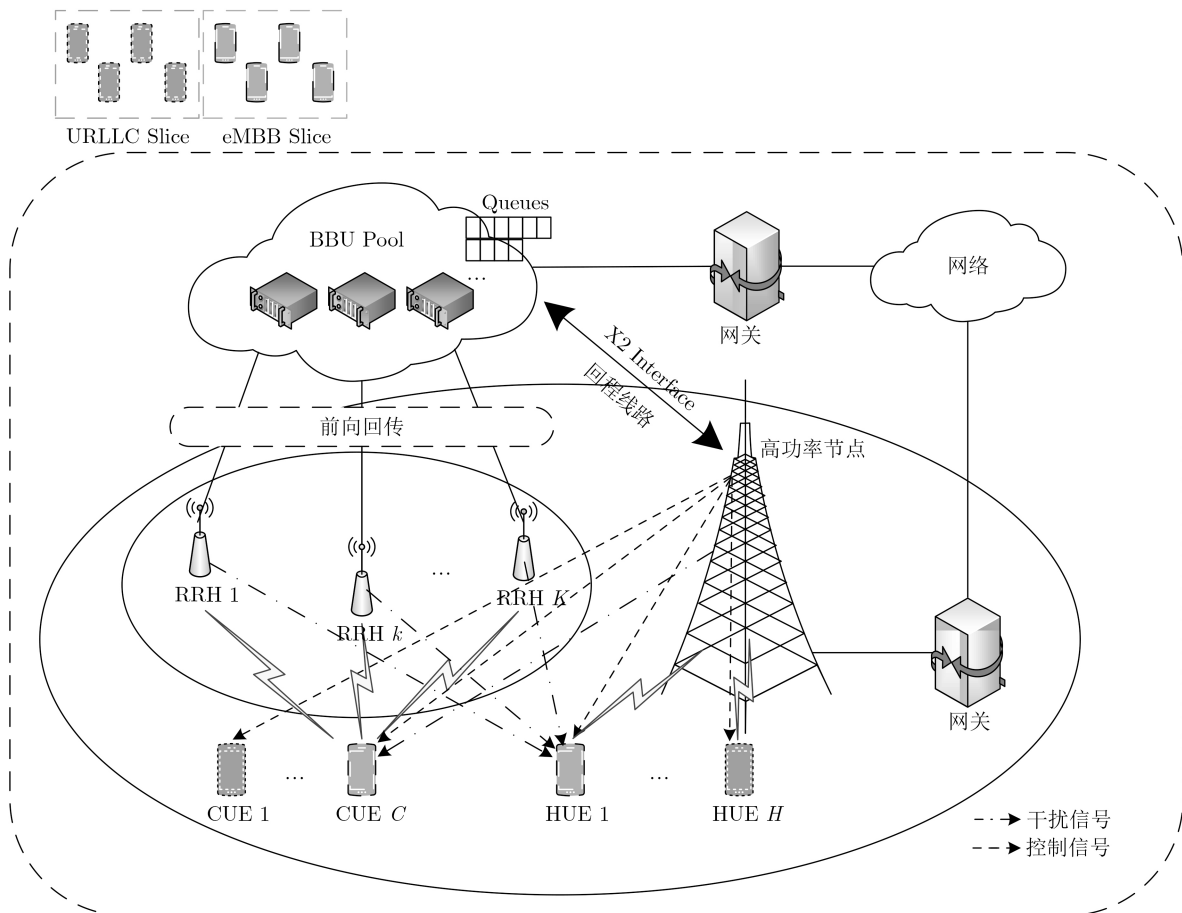


图 1 基于网络切片的H-CRAN下行传输场景

户(User Equipment, UE)用 $\mathcal{U} = \mathcal{C} \cup \mathcal{H}$ 表示, 其中 $\mathcal{C} = \{1, 2, \dots, C\}$ 和 $\mathcal{H} = \{1, 2, \dots, H\}$ 分别表示网络中RRHs和HPN的服务用户集合, RRHs服务的用户用CUE表示, HPN服务的用户用HUE表示。网络共有 S 个子载波, 令 $\mathcal{S} = \{1, 2, \dots, S\}$ 表示可用的子载波集合, 每个子载波 $s \in \mathcal{S}$ 的带宽为 W_0 , 且子载波在CUE与HUE间动态分配。本文考虑CUEs与HUEs间通过复用子载波来提高频谱资源利用率。具体地说, 令 $\alpha_s(t) = 1$ 表示时隙 t 子载波 s 在RRHs和HPN间复用, $\alpha_s(t) = 0$ 则表示时隙 t 子载波 s 至多只能分配给一个用户 $u \in \mathcal{U}$, 即正交资源分配。令 $\beta_{sc}(t)$ 和 $\beta_{sh}(t)$ 分别表示CUE用户和HUE用户的子载波分配指示变量, $\beta_{sc}(t) = 1$ 或 $\beta_{sh}(t) = 1$ 表示时隙 t 子载波 s 分配给CUE c 或HUE h , 反之则不分配。由于任一子载波在未被复用的情况下至多只能分配给一个CUE或HUE, 所以任一时隙 $\beta_{sc}(t)$ 和 $\beta_{sh}(t)$ 满足关系

$$C1: \alpha_s(t) \in \{0, 1\}, \forall s \in \mathcal{S} \quad (1)$$

$$C2: \beta_{sc}(t) \in \{0, 1\}, \sum_{c \in \mathcal{C}} \beta_{sc}(t) = \{0, 1\}, \forall s \in \mathcal{S} \quad (2)$$

$$C3: \beta_{sh}(t) \in \{0, 1\}, \sum_{h \in \mathcal{H}} \beta_{sh}(t) = \{0, 1\}, \forall s \in \mathcal{S} \quad (3)$$

$$C4: \sum_{c \in \mathcal{C}} \beta_{sc}(t) + \sum_{h \in \mathcal{H}} \beta_{sh}(t) = \begin{cases} 2, & \alpha_s(t) = 1 \\ \{0, 1\}, & \alpha_s(t) = 0 \end{cases} \quad (4)$$

网络中考虑eMBB和URLLC两种应用场景共存的情况, 并分别为每种业务构建一个网络切片^[10]。令eMBB切片中CUE与HUE用户集合分别表示为 \mathcal{C}_e 和 \mathcal{H}_e , URLLC切片中CUE与HUE用户集合分别为 \mathcal{C}_u 和 \mathcal{H}_u 。进一步, 令 $\mathcal{U} = \mathcal{U}_e \cup \mathcal{U}_u$, 其中, \mathcal{U}_e 和 \mathcal{U}_u 分别表示eMBB切片和URLLC切片的服务用户集合。

对于每个CUE用户 $c \in \mathcal{C}$, 为了提高其传输速率, 本文假设所有RRHs协作地传输信号并在接收端采用最大比合并技术。因此, 对于任意CUE $c \in \mathcal{C}$, 当子载波 s 在RRHs与HPN间复用时, 时隙 t 内其在子载波 s 上的信干噪比(Signal to Interference plus Noise Ratio, SINR)可表示为

$$\gamma_{cs}^{\text{on}}(t) = \frac{\left(\sum_{k \in \mathcal{K}} P_{ksc}(t) |g_{ksc}(t)|^2 \right)}{\left(P_{ms}(t) |g_{msc}(t)|^2 + \sigma^2 \right)} \quad (5)$$

其中, $P_{ksc}(t)$ 表示时隙 t RRH k 对CUE c 在子载波 s 上的发送功率, $P_{ms}(t)$ 表示时隙 t HPN m 在子载波 s 上的发送功率, $g_{ksc}(t)$ 和 $g_{msc}(t)$ 分别表示时隙 t RRH k , HPN m 对CUE c 在子载波 s 上的信道增益, σ^2 为加性高斯白噪声的功率。

类似地, 对任意 $c \in \mathcal{C}$, 当子载波 s 不在RRHs与HPN间复用时, 即正交分配时, 时隙 t 内其在子载波 s 上的信干噪比可表示为

$$\gamma_{cs}^{\text{off}}(t) = \frac{\left(\sum_{k \in \mathcal{K}} P_{ksc}(t) |g_{ksc}(t)|^2 \right)}{\sigma^2} \quad (6)$$

同理, 对于任意HUE $h \in \mathcal{H}$, 当子载波 s 在RRHs与HPN间复用/不复用时, 其在时隙 t 内的信干噪比可分别表示为

$$\gamma_{hs}^{\text{on}}(t) = \frac{\left(P_{msh}(t) |g_{msh}(t)|^2 \right)}{\left(\sum_{k \in \mathcal{K}} P_{ks}(t) |g_{ksh}(t)|^2 + \sigma^2 \right)} \quad (7)$$

$$\gamma_{hs}^{\text{off}}(t) = \frac{\left(P_{msh}(t) |g_{msh}(t)|^2 \right)}{\sigma^2} \quad (8)$$

其中, $P_{msh}(t)$ 为时隙 t HPN m 在子载波 s 上对HUE h 的发送功率, $g_{msh}(t)$ 和 $g_{ksh}(t)$ 分别表示时隙 t HPN m 和RRH k 在子载波 s 上与HUE h 间的信道增益, $P_{ks}(t)$ 为时隙 t RRH k 在子载波 s 上的发送功率。

由式(5)–式(8)可得, CUE与HUE在时隙 t 的瞬时传输速率可作计算为

$$R_u(t) = \sum_{s \in \mathcal{S}} R_{us}(t) = \sum_{s \in \mathcal{S}} \left\{ [\alpha_s(t) W_0 \log_2(1 + \gamma_{us}^{\text{on}}(t)) + (1 - \alpha_s(t)) W_0 \log_2(1 + \gamma_{us}^{\text{off}}(t))] \beta_{su}(t) \right\}, \quad \forall u \in \mathcal{U} \quad (9)$$

因此, 网络在时隙 t 内的切片用户和速率为

$$R(t) = \sum_{c \in \mathcal{C}} R_c(t) + \sum_{h \in \mathcal{H}} R_h(t) \quad (10)$$

进一步, RRH k 和HPN m 在时隙 t 内消耗的功率可分别表示为

$$P_k(t) = \sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}} \beta_{sc}(t) P_{ksc}(t), \forall k \in \mathcal{K} \quad (11)$$

$$P_m(t) = \sum_{h \in \mathcal{H}} \sum_{s \in \mathcal{S}} \beta_{sh}(t) P_{msh}(t) \quad (12)$$

RRHs与HPN的发射功率是有限的, 因此有

$$C5: 0 \leq P_{ksc}(t) \leq P_k^{\text{max}}, P_k(t) \leq P_k^{\text{max}}, \forall k \in \mathcal{K}, t \quad (13)$$

$$C6: 0 \leq P_{msh}(t) \leq P_m^{\text{max}}, P_m(t) \leq P_m^{\text{max}}, \forall t \quad (14)$$

其中, P_k^{max} , $\forall k \in \mathcal{K}$ 和 P_m^{max} 分别为RRH k 与HPN m 的最大发射功率。

对于eMBB切片用户来说, 网络需要满足各用户关于瞬时传输速率的QoS需求, 即有

$$C7: R_u(t) \geq R_u^{\text{thr}}, \forall u \in \mathcal{U}_e, t \quad (15)$$

其中, R_u^{thr} 为eMBB切片用户 $u \in \mathcal{U}_e$ 的时隙最低速率需求, 也即瞬时速率下界。

对于URLLC切片用户, 本文考虑各用户关于可靠性的QoS需求, 即对于任一URLLC切片用户 $u \in \mathcal{U}_u$, 其在任何子载波上传输的信干噪比不能低于某一阈值^[1], 具体可描述为

$$C8: \beta_{su}(t)\gamma_{us}^\theta(t) \geq \beta_{su}(t)\gamma_u^{\text{thr}}, \theta = \{\text{on}, \text{off}\}, \forall u \in \mathcal{U}_u, t \quad (16)$$

其中, γ_u^{thr} 为URLLC切片用户 $u \in \mathcal{U}_u$ 的最低SINR需求, θ 用以表示子载波 s 当前是否被复用。

2.2 队列模型和队列稳定性

令 $Q_c(t), c \in \mathcal{C}$ 和 $Q_h(t), h \in \mathcal{H}$ 分别表示CUE与HUE用户在时隙 t 的队列长度。令 $A_c(t)$ 和 $A_h(t)$ 分别表示时隙 t 内CUE c 和HUE h 的随机业务到达量, 其分别在每一时隙内存在峰值业务到达量, 即满足 $A_c(t) \leq A_c^{\max}, A_h(t) \leq A_h^{\max}$ 。进一步, 令 $r_c(t)$ 和 $r_h(t)$ 分别表示时隙 t 内CUE c 和HUE h 允许接入的数据量, 因而有 $0 \leq r_c(t) \leq A_c(t), 0 \leq r_h(t) \leq A_h(t)$ 。业务队列 $Q_u(t), u \in \mathcal{U}$ 按式(17)动态更新

$$Q_u(t+1) = \max\{Q_u(t) - R_u(t)\tau, 0\} + r_u(t), \forall u \in \mathcal{U} \quad (17)$$

其中, τ 为每个时隙的持续时间。

由文献[12]可知, 一个排队队列强稳定的条件是它有一个有界的时间平均队列积压。进一步, 当一个网络是强稳定时, 平均可达吞吐量可以由时间平均的接入业务量计算, 因此, H-CRAN网络中各用户的实际平均吞吐量可表示为

$$\bar{r}_u = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} r_u(t), \forall u \in \mathcal{U} \quad (18)$$

2.3 问题描述

本文研究通信需求不同的应用场景(eMBB, URLLC)共存的H-CRAN网络下的网络切片动态资源调度, 首先将所考虑的问题建模为一个随机优化模型, 具体如式(19)所示

$$\begin{aligned} & \max_{\alpha, \beta, P, r} \sum_{u \in \mathcal{U}} \bar{r}_u \\ & \text{s.t. C1 - C8} \\ & \text{C9: } 0 \leq r_c(t) \leq A_c(t), 0 \leq r_h(t) \leq A_h(t), \\ & \quad \forall c \in \mathcal{C}, h \in \mathcal{H}, t \\ & \text{C10: } \bar{Q}_u < \infty, \forall u \in \mathcal{U} \end{aligned} \quad (19)$$

随机优化问题式(19)中, 约束条件C4说明了资源复用和资源分配间的关系, 当子载波 s 在RRHs与HPN间复用时, 即 $\alpha_s(t) = 1$, 则RRHs与HPN服务的用户中各有一个用户分配到子载波 s , 当子载波 s 正交分配时, 即 $\alpha_s(t) = 0$, 则RRHs与HPN服务的所有用户中至多只有一个用户得到子载波 s , 否则, 该子载波当前处于空闲态; C10保证了系统的稳定性。

3 基于Lyapunov优化的动态网络控制与资源管理

接下来, 本文首先利用Lyapunov随机优化理论对问题式(19)进行等效转化, 将其转换成单时隙内的优化问题, 并进而设计一种H-CRAN网络下联合拥塞控制与资源分配的网络切片动态资源调度策略。

3.1 问题转换

首先, 令 $r(t) = \sum_{u \in \mathcal{U}} r_u(t)$ 表示时隙 t 内网络和吞吐量, $\mathbf{Q}(t) = \{Q_u(t), \forall u \in \mathcal{U}\}$ 为时隙 t 系统全局队列状态向量。为了刻画队列拥塞程度的标量度量,

定义二次Lyapunov函数 $L(\mathbf{Q}(t)) = \frac{1}{2} \left(\sum_{u \in \mathcal{U}} Q_u^2(t) \right)$,

其中, 较小的 $L(\mathbf{Q}(t))$ 意味着当前用户排队队列中积压的数据包相对较少。进一步, 定义Lyapunov偏移为 $\Delta(\mathbf{Q}(t)) = E[L(\mathbf{Q}(t+1)) - L(\mathbf{Q}(t)) | \mathbf{Q}(t)]$, 因此, 偏移与加权代价函数可表示为

$$\begin{aligned} & \Delta(\mathbf{Q}(t)) - VE[r(t) | \mathbf{Q}(t)] \\ & = \Delta(\mathbf{Q}(t)) - VE \left[\sum_{u \in \mathcal{U}} r_u(t) | \mathbf{Q}(t) \right] \end{aligned} \quad (20)$$

其中, $V \geq 0$ 是引入的非负控制参量, 其表示网络平均和吞吐量最大化相比于队列稳定性的重要程度。

引理1 对于任意流量接入控制、子载波分配和复用以及功率分配策略, 对于任意的控制变量 $V \geq 0$ 和所有可能的全局队列状态向量 $\mathbf{Q}(t)$, 式(20)中偏移与加权代价函数具有上界^[13]为

$$\begin{aligned} & \Delta(\mathbf{Q}(t)) - VE[r(t) | \mathbf{Q}(t)] \leq B \\ & + E \left[\sum_{u \in \mathcal{U}} Q_u(t) (r_u(t) - R_u(t)\tau) - Vr_u(t) | \mathbf{Q}(t) \right] \end{aligned} \quad (21)$$

其中, B 是一个有限正数。

根据随机优化理论^[12], 求解一个随机优化问题, 可以通过最小化其偏移与加权代价函数的上界, 同时受限于除去稳定性约束以外的其他约束条件。在本文中, 为了求解随机优化问题式(19), 需最小化式(21)不等号右边的项, 同时受限于约束C1-C9。因此, 优化问题式(19)可进一步转化为单时隙优化问题

$$\begin{aligned} & \min_{\alpha(t), \beta(t), P(t), r(t)} \sum_{u \in \mathcal{U}} (Q_u(t) - V) r_u(t) \\ & - \sum_{u \in \mathcal{U}} Q_u(t) R_u(t)\tau, \quad \text{s.t. C1 - C9} \end{aligned} \quad (22)$$

3.2 问题分解

优化问题式(22)可进一步分解为接入控制 $\mathbf{r}(t)$ 和无线资源分配 $\alpha(t), \beta(t), P(t)$ 两个独立的子问题。

(1) 流量接入控制: 最优流量接入控制可以通过最小化问题式(22)的目标函数的第1项获得, 由于用户间的最优数据接入量是相互独立的, 因此, 对于任意 $u \in \mathcal{U}$, 其在每个时隙内的最优业务接入量可按式(23)计算

$$\min_{r_u(t)} (Q_u(t) - V) r_u(t), \quad \text{s.t. } 0 \leq r_u(t) \leq A_u(t) \quad (23)$$

这是一个简单的线性问题, 其最优解为

$$r_u(t) = \begin{cases} A_u(t), & Q_u(t) - V < 0 \\ 0, & \text{其他} \end{cases}, \quad \forall u \in \mathcal{U} \quad (24)$$

可以看出, 用户的最优流量接入量是一个简单的基于阈值的接入控制策略。

(2) 子载波分配和复用、功率分配: 每一时隙内的子载波分配和复用以及功率分配策略可以通过最小化问题式(22)的目标函数的剩余项得到, 具体表示为

$$\max_{\alpha(t), \beta(t), P(t)} \sum_{u \in \mathcal{U}} Q_u(t) R_u(t) \tau, \quad \text{s.t. C1 - C8} \quad (25)$$

为了表述方便, 本文在接下来的阐述中省略时隙索引 t 。下面, 利用拉格朗日对偶分解法对优化问题式(25)进行转换, 不考虑约束C1—C4, 问题式(25)的部分拉格朗日函数可表示为^[14,15]

$$\begin{aligned} \mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu) = & - \sum_{u \in \mathcal{U}} Q_u R_u \tau \\ & + \sum_{k \in \mathcal{K}} \lambda_k (P_k - P_k^{\max}) \\ & + \lambda_m (P_m - P_m^{\max}) \\ & + \sum_{u \in \mathcal{U}_e} \eta_u (R_u^{\text{thr}} - R_u) \\ & + \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{U}_u} \mu_{su} \beta_{su} (\gamma_u^{\text{thr}} - \gamma_{us}^\theta) \end{aligned} \quad (26)$$

其中, λ, η 和 μ 均为拉格朗日乘子向量。接着, 拉格朗日函数 $\mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu)$ 的对偶函数为

$$\mathcal{J}(\lambda, \eta, \mu) = \min_{\alpha, \beta, P} \mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu), \quad \text{s.t. C1 - C4} \quad (27)$$

进一步, 与优化问题式(25)相对应的对偶问题为

$$\max_{\lambda, \eta, \mu} \mathcal{J}(\lambda, \eta, \mu), \quad \text{s.t. } \lambda \geq 0, \eta \geq 0, \mu \geq 0 \quad (28)$$

对偶函数 $\mathcal{J}(\lambda, \eta, \mu)$ 是一个凸函数, 可以通过经典的次梯度法求得优化问题式(28)的最优解。对于问题式(27), 可以利用拉格朗日对偶分解法得到最优的子载波分配和复用以及功率分配策略。具体地, 首先, 问题式(27)可被分解为 S 个相互独立的子问题, 拉格朗日函数 $\mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu)$ 也可改写为

$$\begin{aligned} \mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu) = & \sum_{s \in \mathcal{S}} \left\{ - \sum_{u \in \mathcal{U}} Q_u R_{us} \tau \right. \\ & + \sum_{k \in \mathcal{K}} \sum_{c \in \mathcal{C}} \lambda_k \beta_{sc} P_{ksc} \\ & + \sum_{h \in \mathcal{H}} \lambda_m \beta_{sh} P_{msh} \\ & - \sum_{u \in \mathcal{U}_e} \eta_u R_{us} \\ & + \sum_{u \in \mathcal{U}_u} \mu_{su} \beta_{su} (\gamma_u^{\text{thr}} - \gamma_{us}^\theta) \left. \right\} \\ & - \sum_{k \in \mathcal{K}} \lambda_k P_k^{\max} - \lambda_m P_m^{\max} \\ & + \sum_{u \in \mathcal{U}_e} \eta_u R_u^{\text{thr}} \end{aligned} \quad (29)$$

因此, 对于任意子载波 $s \in \mathcal{S}$, 与其相应的子问题可表示为

$$\begin{aligned} \min_{\alpha_s, \beta_s, P_s} & - \sum_{u \in \mathcal{U}} Q_u R_{us} \tau + \sum_{k \in \mathcal{K}} \sum_{c \in \mathcal{C}} \lambda_k \beta_{sc} P_{ksc} \\ & + \sum_{h \in \mathcal{H}} \lambda_m \beta_{sh} P_{msh} - \sum_{u \in \mathcal{U}_e} \eta_u R_{us} \\ & + \sum_{u \in \mathcal{U}_u} \mu_{su} \beta_{su} (\gamma_u^{\text{thr}} - \gamma_{us}^\theta), \quad \text{s.t. C1 - C4} \end{aligned} \quad (30)$$

其中, β_s 和 P_s 分别表示与子载波 s 相关的子载波分配和功率分配向量。进一步, 由约束条件C1—C4可知, 对于任意子载波 $s \in \mathcal{S}$, 一共存在以下几种情况:

- (a) $\alpha_s = 0$: (i) $\beta_{sc} = 1, \forall c \in \mathcal{C}_e$; (ii) $\beta_{sc} = 1, \forall c \in \mathcal{C}_u$; (iii) $\beta_{sh} = 1, \forall h \in \mathcal{H}_e$; (iv) $\beta_{sh} = 1, \forall h \in \mathcal{H}_u$
- (b) $\alpha_s = 1$: (i) $\beta_{sc} = 1, \beta_{sh} = 1, \forall c \in \mathcal{C}_e, h \in \mathcal{H}_e$; (ii) $\beta_{sc} = 1, \beta_{sh} = 1, \forall c \in \mathcal{C}_e, h \in \mathcal{H}_u$; (iii) $\beta_{sc} = 1, \beta_{sh} = 1, \forall c \in \mathcal{C}_u, h \in \mathcal{H}_e$; (iv) $\beta_{sc} = 1, \beta_{sh} = 1, \forall c \in \mathcal{C}_u, h \in \mathcal{H}_u$ 。

接下来, 本文分别讨论上述每种情况, 求解出各情况下最优的子载波分配和功率分配策略, 然后将各情况所对应的优化问题式(30)的最优目标函数值和子载波处于空闲状态下的目标函数值相比较, 从而确定当前时隙最终的子载波复用和对应的子载波分配与功率分配方案。

(a) $\alpha_s = 0$: 当子载波 s 正交分配给 eMBB 切片中的 CUE c 时, 问题式(30)的目标函数可以简写为

$$\varphi_{\text{off},sc}^{\mathcal{C}_e}(P_{\text{off},s}^{\mathcal{C}_e}) = -Q_c R_{cs} \tau + \sum_{k \in \mathcal{K}} \lambda_k P_{\text{off},ksc}^{\mathcal{C}_e} - \eta_c R_{cs} \quad \forall c \in \mathcal{C}_e \quad (31)$$

将 $\varphi_{\text{off},sc}^{\mathcal{C}_e}(P_{\text{off},s}^{\mathcal{C}_e})$ 对 $P_{\text{off},ksc}^{\mathcal{C}_e}$ 求导并令其结果为0可得

$$P_{\text{off},ksc}^{C_c} = \left[\begin{aligned} & (Q_c \tau + \eta_c) W_0 / (\lambda_k \ln 2) \\ & - \left(\sigma^2 + \sum_{k' \in \mathcal{K}, k' \neq k} P_{\text{off},k'sc}^{C_c} |g_{k'sc}|^2 \right) / |g_{ksc}|^2 \end{aligned} \right]^+ \quad (32)$$

该功率分配具有多级注水的形式^[4], 其中, 注水水平由业务队列状态决定。将式(32)代入式(31)可求得相应的目标函数值, 且此情况下最优的子载波分配和功率分配可按式(33)–式(35)确定

$$c_{\text{off},sc}^{C_c*} = \arg \min_{c \in C_e} \left\{ \varphi_{\text{off},sc}^{C_c}(P_{\text{off},s}^{C_c}) \right\} \quad (33)$$

$$\beta_{\text{off},sc}^{C_c*} = \begin{cases} 1, & c = c_{\text{off},sc}^{C_c*} \\ 0, & \text{其他} \end{cases} \quad (34)$$

$$P_{\text{off},sc}^{C_c*} = \left(P_{\text{off},s}^{C_c} \right)_{\beta_{\text{off},sc}^{C_c*}} \quad (35)$$

与上述情况的分析过程类似, 当子载波 s 正交分配时, 其他3种情况下最优子载波分配和功率分配的计算步骤与式(33)–式(35)有相似的形式, 本文不再赘述, 并分别用 $(\beta_{\text{off},sc}^{C_u*}, P_{\text{off},s}^{C_u*}), (\beta_{\text{off},sh}^{H_c*}, P_{\text{off},msh}^{H_c*})$ 和 $(\beta_{\text{off},sh}^{H_u*}, P_{\text{off},msh}^{H_u*})$ 表示。

(b) $\alpha_s = 1$: 当子载波 s 以复用的方式分别分配给eMBB切片中的CUE c 和HUE h 时, 问题式(30)的目标函数可以简化为

$$\varphi_{\text{on},sch}^{ee}(P_{\text{on},s}^{ee}) = -Q_c R_{cs} \tau - Q_h R_{hs} \tau + \sum_{k \in \mathcal{K}} \lambda_k P_{\text{on},ksc}^{ee} + \lambda_m P_{\text{on},msh}^{ee} - \eta_c R_{cs} - \eta_h R_{hs} \quad (36)$$

由于函数 $\varphi_{\text{on},sch}^{ee}(P_{\text{on},s}^{ee})$ 关于功率分配变量 $P_{\text{on},ksc}^{ee}$ 和 $P_{\text{on},msh}^{ee}$ 的凹凸性难以判定, 且其分别对 $P_{\text{on},ksc}^{ee}$ 和 $P_{\text{on},msh}^{ee}$ 求导并令其结果为0后的式子是一个复杂的隐函数。因此, 本文参考文献^[16], 采用基于相似度变异的粒子群算法获得近似最优的功率分配策略, 进而此情况下近似最优的子载波分配和功率分配可按式(37)–式(39)确定

$$(c_{\text{on},sch}^{ee*}, h_{\text{on},sch}^{ee*}) = \arg \min_{c \in C_e, h \in \mathcal{H}_e} \left\{ \varphi_{\text{on},sch}^{ee}(P_{\text{on},s}^{ee}) \right\} \quad (37)$$

$$\beta_{\text{on},sc}^{ee*} = \begin{cases} 1, & c = c_{\text{on},sch}^{ee*} \\ 0, & \text{其他} \end{cases}, \beta_{\text{on},sh}^{ee*} = \begin{cases} 1, & h = h_{\text{on},sch}^{ee*} \\ 0, & \text{其他} \end{cases} \quad (38)$$

$$P_{\text{on},s}^{ee*} = \left(P_{\text{on},s}^{ee} \right)_{\beta_{\text{on},sc}^{ee*}, \beta_{\text{on},sh}^{ee*}} \quad (39)$$

与上述情况的分析过程类似, 当子载波 s 在RRHs与HPN间复用时, 其他3种情况下近似最优的子载波分配和功率分配的计算步骤与式(37)–

式(39)有相似的形式, 本文不再赘述, 并分别用 $(\beta_{\text{on},sc}^{eu*}, \beta_{\text{on},sh}^{eu*}, P_{\text{on},s}^{eu*}), (\beta_{\text{on},sc}^{ue*}, \beta_{\text{on},sh}^{ue*}, P_{\text{on},s}^{ue*})$ 和 $(\beta_{\text{on},sc}^{uu*}, \beta_{\text{on},sh}^{uu*}, P_{\text{on},s}^{uu*})$ 表示。

最后, 当子载波 s 处于空闲态时, 问题式(30)的优化目标为0。

如前所述, 对任一子载波 s , 通过比较上述各情况下和子载波处于空闲状态时问题式(30)的(近似)最优目标函数值, 可以确定其在当前时隙最终的复用、分配和其上相应的功率分配策略。特别地, 在逐一分配 S 个子载波的过程中, 应在每分配一个子载波后, 判断得到该子载波的用户是否已经获得了足够的资源, 若是, 则将其从接下来的子载波分配进程中剔除, 避免浪费资源。综上, 针对原始优化问题式(19), 本文提出的H-CRAN网络下联合拥塞控制和资源分配的网络切片动态资源调度算法具体如表1所示。

4 仿真结果与讨论

本节中, 将表1的算法与文献^[5]中NOMA系统下高效动态资源优化算法和文献^[6]中基于时延感知的网络切片无线资源分配算法相比较(分别用本文方法, NOMA-based和Fog-based表示)以评估其性能, 并对仿真结果进行详尽地讨论分析。

4.1 参数设置

仿真中所考虑的H-CRAN网络包含1个HPN、3个RRHs、10个CUEs和10HUEs。网络中共有100个可用子载波, 每个子载波的带宽为15 kHz。时隙长度为10 ms。RRHs和HPN的路径损耗模型分别采用 $31.5 + 40 \lg(d)$ 和 $31.5 + 35 \lg(d)$ ^[4], 其中, d 表示接入节点和发送节点间的距离, 以m为单位。HPN和RRHs的最大发射功率分别为46 dBm和30 dBm, 噪声功率为-102 dBm。CUE和HUE的业务到达服从泊松分布, 且平均业务到达率满足 $\lambda_c = \lambda$ 和 $\lambda_h = 0.5\lambda$, 其中, λ 的取值根据具体仿真场景而取不同的值。

4.2 仿真结果分析

图2和图3分别描述了不同URLLC用户QoS需求下网络平均和吞吐量和平均队列时延随控制参量 V 的变化情况, 其中, 平均业务到达率 $\lambda = 4$ kbit/slot。可以看出, 随着 V 的取值增大, 平均吞吐量以 $O(1/V)$ 的速度逼近理论最优值 U^* , 平均队列时延则以 $O(V)$ 的速度呈线性增长。这是因为较大的 V 意味着系统更侧重于最大化网络平均和吞吐量, 会在维持网络稳定的基础上以牺牲用户平均队列时延为代价, 优先考虑网络效用最大化的问题。此外, 从图2和图3还可以看出, 控制参量 V 提供了一种灵活

表 1 H-CRAN网络下联合拥塞控制和资源分配的网络切片动态资源调度算法

- (1) 初始化控制参量 $V > 0$ 、各用户的初始队列长度 $Q_u(0), \forall u \in \mathcal{U}$ 和最大时隙数 T^{\max} 初始化最大迭代次数 T_0^{\max} 和允许误差 δ
- (2) for $t = 0, 1, \dots, T^{\max} - 1$
- (3) 根据式(24)分别计算各用户当前时隙最优的流量接入控制策略
- (4) Repeat:
- (5) 令迭代索引 $n = 1$, 初始化拉格朗日乘子 λ, η 和 μ
- (6) for $s \in \mathcal{S}$
- (7) 计算子载波 s 当前时隙(近似)最优的子载波复用、分配和功率分配策略 α_s^*, β_s^* 和 P_s^* , 进而更新各用户剩余的排队队列长度
- (8) 若某用户 $u \in \mathcal{U}$ 已经获得了足够的子载波(即其队列长度为0), 则将其从接下来的子载波分配过程中排除。
- (9) 若所有用户均分配到足够的子载波, 则break
- (10) end for
- (11) 根据得到的(近似)最优子载波复用、分配和功率分配策略 α^*, β^* 和 P^* 计算拉格朗日函数 $\mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu)^{(n)}$
- (12) Until $|\mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu)^{(n)} - \mathcal{L}(\alpha, \beta, P, \lambda, \eta, \mu)^{(n-1)}| \leq \delta$ or $n > T_0^{\max}$, then stop Otherwise, 利用次梯度法更新拉格朗日乘子 λ, η 和 μ , 令 $n = n + 1$ 并返回第6步
- (13) 根据式(17)更新各用户在下一时隙的业务队列长度
- (14) end for
- (15) 输出: (近似)最优流量接入控制、子载波复用和分配以及功率分配策略 r, α, β 和 $P, Q_u(t), \forall u \in \mathcal{U}, t$.

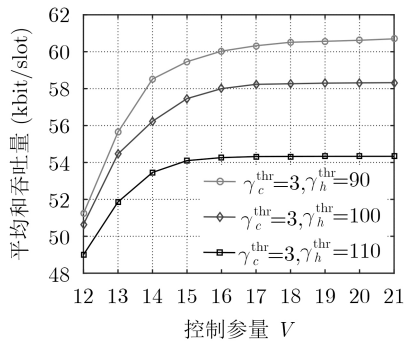


图 2 平均吞吐量与控制参量V

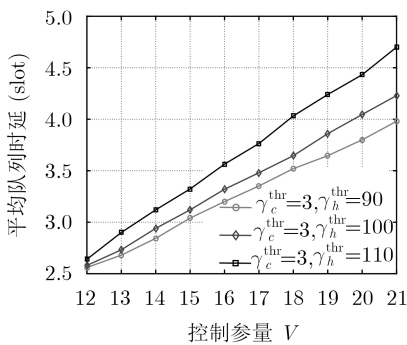


图 3 平均队列时延与控制参量V

实现时延和吞吐量间动态平衡的方法, 因此可以根据需要选择合适的V值以使系统工作在期望的性能状态下。

图4和图5分别比较了不同算法下平均速率和平均队列时延随平均业务到达率λ的变化情况, 其中, 控制参量 $V = 16$ 。从图4可以看出, 在λ较小时, 3种算法的平均速率几乎相同且不小于平均

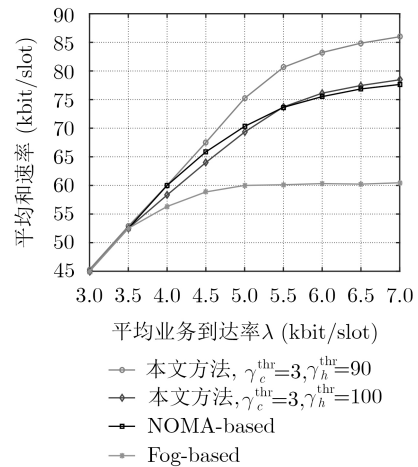


图 4 平均速率与平均业务到达率λ

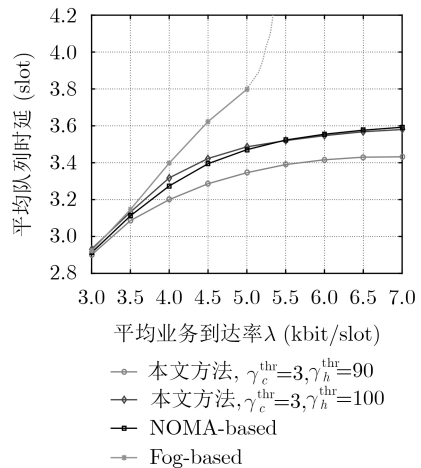


图 5 平均队列时延与平均业务到达率λ

总业务到达率,并随着 λ 的增加逐渐增大,最终趋于稳定。这是因为在 λ 较小时,3种算法均能及时地将用户业务队列中的数据包包发送出去,在维持网络稳定的条件下尽可能地接收更多的数据量。但是,当 λ 增大到一定程度时,本文所提算法和文献[5] NOMA-based算法中的流量接入控制机制开始发挥作用,通过拒绝部分业务数据的接入以保障系统的稳定性,此时平均和速率会小于平均总业务到达率。从图5可以看出,3种算法的平均队列时延均随着平均业务到达率 λ 的增大而增大,且文献[6]中 Fog-based算法的平均队列时延会在 λ 高于特定值之后开始剧烈增加并趋于无穷。这是因为当 λ 较小时,随着 λ 的增加,用户业务队列中积压的数据包也相对变多,从而平均队列时延也相应增加。而当 λ 持续增大到超过某个特定值之后,本文所提算法和文献[5]NOMA-based算法中的流量接入控制机制开始发挥作用,限制部分业务数据的接入以维持网络的稳定性。但是,对于Fog-based算法,由于文献[6]没有考虑接入控制等控制策略限制接入到网络中的流量,因而当 λ 增大到超出网络容量域时,网络会发生严重拥塞,平均队列时延急剧增大。

5 结论

针对H-CRAN网络中基于网络切片的在线无线资源优化问题,本文通过联合考虑接入控制、拥塞控制、资源分配和复用,提出了一种联合拥塞控制和资源分配的网络切片动态资源调度算法,以实现在满足各切片(eMBB, URLLC)用户QoS需求和维持网络稳定的前提下,最大化网络平均和吞吐量。此外,该算法还可通过调整控制参量 V 的取值灵活地实现时延和吞吐量间的动态平衡。

参考文献

- [1] Cisco System. Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022 White Paper[R/OL]. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html>, 2019.
- [2] LI Xin, SAMAKA M, CHAN H A, *et al*. Network slicing for 5G: challenges and opportunities[J]. *IEEE Internet Computing*, 2017, 21(5): 20–27. doi: 10.1109/MIC.2017.3481355.
- [3] ITU-R. IMT vision-framework and overall objectives of the future development of IMT for 2020 and beyond[EB/OL]. <http://www.itu.int/pub/R-REC/en>. 2020.
- [4] LI Jian, PENG Mugen, YU Yuling, *et al*. Energy-efficient joint congestion control and resource optimization in heterogeneous cloud radio access networks[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(12): 9873–9887. doi: 10.1109/TVT.2016.2531184.
- [5] ZHANG Haijun, WANG Baobao, JIANG Chunxiao, *et al*. Energy efficient dynamic resource optimization in NOMA system[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(9): 5671–5683. doi: 10.1109/TWC.2018.2844359.
- [6] DANG Tian and PENG Mugen. Delay-aware radio resource allocation optimization for network slicing in fog radio access networks[C]. 2018 IEEE International Conference on Communications Workshops, Kansas City, USA, 2018: 1–6. doi: 10.1109/ICCW.2018.8403717.
- [7] AMANI N, PEDRAM H, TAHERI H, *et al*. Energy-efficient resource allocation in heterogeneous cloud radio access networks via BBU offloading[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(2): 1365–1377. doi: 10.1109/TVT.2018.2882466.
- [8] KIM T and CHANG J M. Profitable and energy-efficient resource optimization for heterogeneous cloud-based radio access networks[J]. *IEEE Access*, 2019, 7: 34719–34737. doi: 10.1109/ACCESS.2019.2904766.
- [9] ZHANG Biling, MAO Xingwang, YU J L, *et al*. Resource allocation for 5G heterogeneous cloud radio access networks with D2D communication: a matching and coalition approach[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(7): 5883–5894. doi: 10.1109/TVT.2018.2802900.
- [10] 唐伦, 魏延南, 马润琳, 等. 虚拟化云无线接入网络下基于在线学习的网络切片虚拟资源分配算法[J]. 电子与信息学报, 2019, 41(7): 1533–1539. doi: 10.11999/JEIT180771.
- TANG Lun, WEI Yannan, MA Runlin, *et al*. Online learning-based virtual resource allocation for network slicing in virtualized cloud radio access network[J]. *Journal of Electronics & Information Technology*, 2019, 41(7): 1533–1539. doi: 10.11999/JEIT180771.
- [11] MEI Jie, ZHENG Kan, ZHAO Long, *et al*. A latency and reliability guaranteed resource allocation scheme for LTE V2V communication systems[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(6): 3850–3860. doi: 10.1109/TWC.2018.2816942.
- [12] NEELY M J. Stochastic network optimization with application to communication and queueing systems[J]. *Synthesis Lectures on Communication Networks*, 2010, 3(1): 15–62. doi: 10.2200/S00271ED1V01Y201006CNT007.
- [13] MOKDAD A, AZMI P, MOKARI N, *et al*. Cross-layer energy efficient resource allocation in PD-NOMA based H-CRANs: implementation via GPU[J]. *IEEE Transactions on Mobile Computing*, 2019, 18(6): 1246–1259. doi: 10.1109/TMC.2018.2860985.

- [14] TANG Liya, ZHANG Xian, XIANG Hongyu, *et al.* Joint resource allocation and caching placement for network slicing in fog radio access networks[C]. The 18th International Workshop on Signal Processing Advances in Wireless Communications, Sapporo, Japan, 2017: 1-6. doi: [10.1109/SPAWC.2017.8227791](https://doi.org/10.1109/SPAWC.2017.8227791).
- [15] LEE Y L, LOO J, CHUAH T C, *et al.* Dynamic network slicing for multitenant heterogeneous cloud radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(4): 2146-2161. doi: [10.1109/TWC.2017.2789294](https://doi.org/10.1109/TWC.2017.2789294).
- [16] TANG Lun, YANG Xixi, WU Xiaolin, *et al.* Queue stability-based virtual resource allocation for virtualized wireless networks with self-backhauls[J]. *IEEE Access*, 2018, 6: 13604-13616. doi: [10.1109/ACCESS.2018.2797088](https://doi.org/10.1109/ACCESS.2018.2797088).
- 唐 伦: 男, 1973年生, 教授, 研究方向为下一代无线通信网络、异构蜂窝网络、软件定义无线网络等.
- 魏延南: 男, 1995年生, 硕士生, 研究方向为5G网络切片、虚拟资源分配、随机优化理论.
- 谭 颀: 女, 1995年生, 硕士生, 研究方向为5G网络切片、资源分配、随机优化理论.
- 陈前斌: 男, 1967年生, 教授, 博士生导师, 研究方向为个人通信、多媒体信息处理与传输、异构蜂窝网络等.