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

唐 伦 魏延南* 马润琳 贺小雨 陈前斌

(重庆邮电大学通信与信息工程学院 重庆 400065)

(重庆邮电大学移动通信技术重点实验室 重庆 400065)

摘 要：针对现有研究中缺乏云无线接入网络(C-RAN)场景下对网络切片高效的动态资源分配方案的问题，该文提出一种虚拟化C-RAN网络下的网络切片虚拟资源分配算法。首先基于受限马尔可夫决策过程(CMDP)理论建立了一个虚拟化C-RAN场景下的随机优化模型，该模型以最大化平均切片和速率为目标，同时受限于各切片平均时延约束以及网络平均回传链路带宽消耗约束。其次，为了克服CMDP优化问题中难以准确掌握系统状态转移概率的问题，引入决策后状态(PDS)的概念，将其作为一种“中间状态”描述系统在已知动态发生后，但在未知动态发生前所处的状态，其包含了所有与系统状态转移有关的已知信息。最后，提出一种基于在线学习的网络切片虚拟资源分配算法，其在每个离散的资源调度时隙内会根据当前系统状态为每个网络切片分配合适的资源块数量以及缓存资源。仿真结果表明，该算法能有效地满足各切片的服务质量(QoS)需求，降低网络回传链路带宽消耗的压力并同时提升系统吞吐量。

关键词：5G网络切片；云无线接入网络；资源分配；马尔可夫决策过程

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2019)07-1533-07

DOI: 10.11999/JEIT180771

Online Learning-based Virtual Resource Allocation for Network Slicing in Virtualized Cloud Radio Access Network

TANG Lun WEI Yannan MA Runlin HE Xiaoyu CHEN Qianbin

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

(*Key Laboratory of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: To solve the problem of lacking efficient and dynamic resource allocation schemes for 5G Network Slicing (NS) in Cloud Radio Access Network (C-RAN) scenario in the existing researches, a virtual resource allocation algorithm for NS in virtualized C-RAN is proposed. Firstly, a stochastic optimization model in virtualized C-RAN network is established based on the Constrained Markov Decision Process (CMDP) theory, which maximizes the average sum rates of all slices as its objective, and is subject to the average delay constraint for each slice as well as the average network backhaul link bandwidth consumption constraint in the meantime. Secondly, in order to overcome the issue of having difficulties in acquiring the accurate transition probabilities of the system states in the proposed CMDP optimization problem, the concept of Post-Decision State (PDS) as an “intermediate state” is introduced, which is used to describe the state of the system after the known dynamics, but before the unknown dynamics occur, and it incorporates all of the known information about the system state transition. Finally, an online learning based virtual resource allocation algorithm is presented for NS in virtualized C-RAN, where in each discrete resource scheduling slot, it will allocate appropriate Resource Blocks (RBs) and caching resource for each network slice according to the observed current system state. The simulation results reveal that the proposed algorithm can effectively satisfy the

收稿日期：2018-08-03；改回日期：2019-02-20；网络出版：2019-03-19

*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)

Quality of Service (QoS) demand of each individual network slice, reduce the pressure of backhaul link on bandwidth consumption and improve the system throughput.

Key words: 5G Network Slicing (NS); Cloud Radio Access Network (C-RAN); Resource allocation; Markov Decision Process (MDP)

1 引言

近年来，随着移动用户的数据需求激增，移动数据业务经历了大幅度增长。因此，移动运营商需要降低成本的可扩展解决方案，以满足未来5G网络在容量和时延等方面的性能指标。在现已提出的众多具有前景的技术和新型网络框架中，网络切片(Network Slicing, NS)和云无线接入网(Cloud Radio Access Network, C-RAN)获得了学者们的广泛关注和深入研究[1]。网络切片是指利用虚拟化技术将网络基础设施资源虚拟化为多个专用的虚拟网络，其实现了业务场景、网络功能和基础设施平台间的适配，可以更好地支持多样化的业务需求。C-RAN架构有助于在整个网络内交换业务和信道信息，其可在降低功耗的同时进一步提升网络的整体性能[2]。

已有大量工作深入研究了虚拟化技术和C-RAN架构。文献[3]提出了一种前传容量受限的C-RAN下的资源共享策略，其中一个网络运营商将无线电资源租借给多个服务提供商(Service Providers, SPs)并控制用户接入和关联。文献[4]将虚拟化技术与C-RAN相结合，以实现最大化系统吞吐量且最小化时延的目标。

现有的研究工作大多是分开讨论虚拟化技术与C-RAN架构，很少有工作将二者结合起来探讨，也没有考虑具有多样化性能需求的5G网络切片共存的情况。然而，网络切片与C-RAN结合具有明显的优势，一来无线资源可以实现跨小区的动态可扩展分配，二来可以较容易且灵活地重新部署云资源，以专注于高需求区域，从而提升网络覆盖范围并增强用户体验质量(Quality of Experience, QoE)。此外，当前针对网络切片的资源分配方案大多考虑的是频谱或功率资源，很少有文献考虑缓存资源与其他无线电资源的联合动态分配。针对5G网络切片的动态资源分配问题，本文联合考虑了频谱与缓存资源，并将虚拟化C-RAN网络下的资源分配看作一个无穷时间马尔可夫决策过程。本文的主要贡献包括：

(1)本文基于CMDP理论建立了一个虚拟化C-RAN场景下的随机优化模型，该模型以最大化平均切片和速率为目标，同时受限于各切片平均时延约束以及网络平均回传链路带宽消耗约束。

(2)针对CMDP优化问题中难以准确掌握系统状态转移概率的问题，本文引入决策后状态的概念，将其作为一种“中间状态”用于描述系统在已知动态发生后，但在未知动态发生前所处的状态，其包含了所有与系统状态转移有关的已知信息。

(3)本文提出了一种基于在线学习的网络切片虚拟资源分配算法。在该算法中，在每个离散的资源调度时隙内，算法会根据当前系统状态为每个网络切片分配合适的资源块数量以及缓存资源。

2 系统模型

2.1 系统场景

如图1所示，考虑一个虚拟化C-RAN网络的下行传输场景，假设原底层物理网络中的每个基站均带有一定的缓存空间，将所有缓存空间通过“云

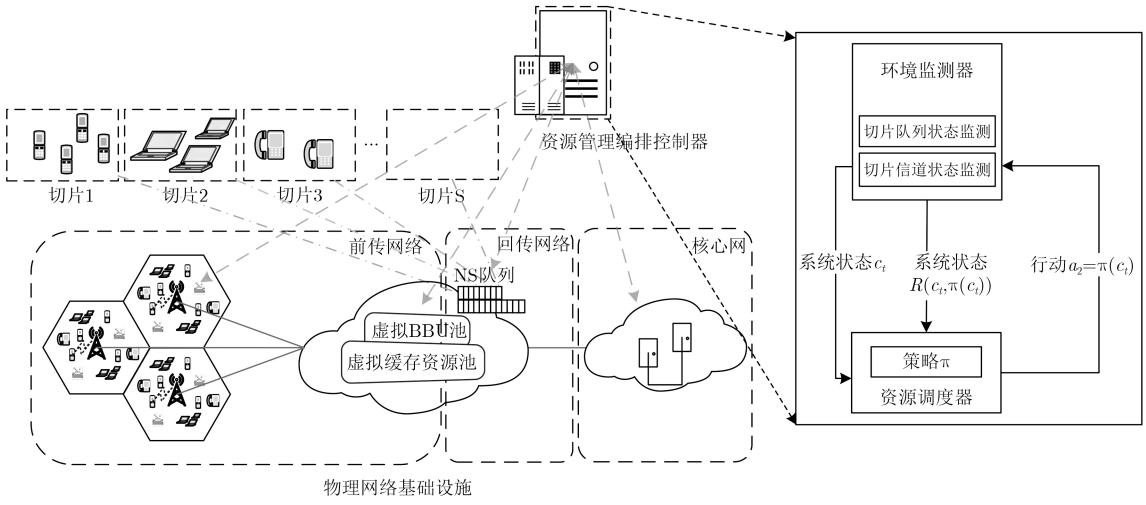

图 1 虚拟化C-RAN网络系统场景

化”操作集中形成虚拟缓存资源池(virtual cache resource pool)，并与虚拟基带单元(Base Band Unit, BBU)池部署在一起。整个网络共为K个用户提供S种不同的应用服务。令$\boldsymbol{K}=\{1,2,\cdots,K\}$，$\boldsymbol{S}=\{1,2,\cdots,S\}$，其中，$\boldsymbol{K}$和$\boldsymbol{S}$分别表示用户集合和业务集合。

网络中考虑一种时变随机信道模型，令$h_{ks}(t)$为用户k在时隙t请求业务s时的信道增益。$h_{ks}(t)\in\boldsymbol{H}=\{h_1,h_2,\cdots,h_H\}$ 且$\sum_{i=1}^{H}P(h_i)=1$，其中，$\boldsymbol{H}$为有限信道状态集合，$P(h_i)$表示信道状态为h_i的概率。假定用户请求各业务时的信道状态在每个时隙内保持不变，在时隙间随机变化，并且用户在不同时隙间请求业务时的信道状态是相互独立的。假设时隙t内用户间请求相同业务s的平均信道增益为$\bar{h}_s(t)$，对应的频谱效率为$\gamma_s(t)$。

本文考虑一个离散时间排队系统，设每个时隙的长度为τ，每个用户在任一时隙内都能请求多种不同的业务，假设用户k在时隙t内业务s到达的数据包为$A_{ks}(t)$，其服从某种分布$E\{A_{ks}(t)\}=\lambda_{ks}$且在时隙间是独立同分布的。网络为每种业务构建一个网络切片及其相应的排队队列，切片与业务是一一对应的(使用同一索引)，令$Q_s(t)$表示切片s在时隙t开始时的队列长度，且有$Q_s(t)=\sum_{k\in\boldsymbol{K}}Q_{ks}(t)$，其中，$Q_{ks}(t)$为时隙$t$用户$k$的业务$s$的队列长度。$Q_s(t)$按式(1)动态更新

$$Q_s(t+1)=\max[Q_s(t)-D_s(t),0]+A_s(t) \tag{1}$$

其中，$A_s(t)=\sum_{k\in\boldsymbol{K}}A_{ks}(t)$为业务$s$在时隙$t$内到达的数据包数，$D_s(t)=\gamma_s(t)\cdot B\cdot X_s(t)\cdot\tau/L$表示时隙$t$从业务$s$的排队队列中离开的数据包数，$B$为单个RB(Resource Block)的带宽，$X_s(t)$为时隙t分配给切片s的RB个数，L为每个数据包的大小。进一步，令$\boldsymbol{Q}(t)=\{Q_1(t),Q_2(t),\cdots,Q_S(t)\}$表示系统在时隙$t$的全局队列状态信息(Queue State Information, QSI)，$\boldsymbol{H}(t)=\{\bar{h}_1(t),\bar{h}_2(t),\cdots,\bar{h}_S(t)\}$为时隙$t$的全局信道状态信息(Channel State Information, CSI)。

虚拟缓存资源池通过主动缓存“较优”的业务内容来减少网络回传链路的带宽消耗，并降低网络运营成本。另一方面，通过将基站边缘较小的缓存空间通过“池化”形成云端具有较大容量的缓存空间这一操作，可使得网络缓存更多的流行内容，从而可以被来自不同运营商的众多用户所共享，极大地提高了资源利用率，进一步降低了网络整体时延。为了便于理解和后续讨论，本文假设虚拟缓存资源池在每一时隙可完整缓存任一业务的全部内容，令$Z_s(t)\in\{0,1\}$表示时隙t的缓存策略，$Z_s(t)=1$意味着系统在时隙t内缓存业务s的全部内容，否则，$Z_s(t)=0$。进一步有

$$\sum_{s\in\boldsymbol{S}}Z_s(t)=1,\ \forall t \tag{2}$$

令$\boldsymbol{Z}(t)=\{Z_s(t),s\in\boldsymbol{S}\}$为时隙$t$内的缓存资源分配行为。类似地，令$\boldsymbol{X}(t)=\{X_s(t),s\in\boldsymbol{S}\}$为时隙$t$内的RB分配行为，其中，$X_s(t)$满足

$$X_s(t)\geq 0,\sum_{s\in\boldsymbol{S}}X_s(t)\leq N \tag{3}$$

其中，N为网络中的RB总数。

由上所述，时隙t内网络切片的和速率可以表示为

$$R(t)=\sum_{s\in\boldsymbol{S}}R_s(t)=\sum_{s\in\boldsymbol{S}}\gamma_s(t)\cdot B\cdot X_s(t) \tag{4}$$

其中，$R_s(t)$为切片s在时隙t的瞬时传输速率。进一步，本文假设回传链路的带宽消耗与前传链路瞬时传输速率相同，因此网络在时隙t内消耗的回传链路带宽可按式(5)计算

$$B(t)=R(t)-\sum_{s\in\boldsymbol{S}}Z_s(t)R_s(t) \tag{5}$$

2.2 问题描述

一个受限马尔可夫决策过程(Constrained Markov Decision Process, CMDP)问题可由1个4元组$<C,A,r_a(c'|c),R_a(c,c')>$描述，其中，$C$为状态空间，$A$为行动空间，$r_a(c'|c)=\Pr(c_{t+1}=c'|c_t=c,a_t=a)$表示系统在当前时隙$t$处于状态$c$下，执行动作$a$后，在下一时隙$t+1$转到状态$c'$的概率。$R_a(c,c')$表示系统在状态$c$下执行动作$a$并转移到状态$c'$时的即时成本/回报。

在本文中，定义系统在时隙t的状态为$c_t=(\boldsymbol{Q}(t),\boldsymbol{H}(t))\in C$，定义系统在时隙$t$的行动为$a_t=(\boldsymbol{X}(t),\boldsymbol{Z}(t))\in A$。令$\pi:C\to A$代表一个稳定的确定性策略，其将状态空间映射到行动空间上，即$a=\pi(c)$。令$\boldsymbol{\Phi}$表示所有可能的策略集合，给定初始状态c，以及策略$\pi\in\boldsymbol{\Phi}$，则期望累积折扣回报(切片和速率)、期望累积折扣切片时延以及期望累积折扣回传链路带宽消耗可以分别表示为

$$\begin{aligned}\bar{R}^{\pi}(c)&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t R(c_t,\pi(c_t))|c_0=c\right\}\\&=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t\sum_{s\in\boldsymbol{S}}\gamma_s(t)\cdot B\cdot X_s(t)|c_0=c\right\}\end{aligned} \tag{6}$$

$$\bar{D}_s^{\pi}(c)=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t d_s(t)|c_0=c\right\}=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t\frac{Q_s(t)}{\lambda_s}|c_0=c\right\},\forall s\in\boldsymbol{S}\tag{7}$$

$$\bar{B}^{\pi}(c)=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t B(t)|c_0=c\right\}=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t[R(t)-\sum_{s\in\boldsymbol{S}}Z_s(t)R_s(t)]|c_0=c\right\}\tag{8}$$

其中，$\lambda_s=\mathrm{E}[A_s(t)]=\sum_{k\in\boldsymbol{K}}\lambda_{ks}$为切片$s$的数据包到达过程的均值，$Q_s(t)/\lambda_s$根据文献[5]可以理解为切片$s$的平均时延，$\gamma\in(0,1]$为折扣因子，其指示了未来的回报对当前行为选择的影响程度。本文的目的是通过合理动态的分配频谱资源与缓存资源，在满足各网络切片平均时延约束以及网络平均回传链路带宽消耗约束的前提下，最大化平均网络切片和速率，因而建立式(9)所示的随机优化模型

$$\begin{aligned}&\min\ -E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t\sum_{s\in\boldsymbol{S}}\gamma_s(t)\cdot B\cdot X_s(t)|c_0=c\right\}\\&\text{s.t.C1 }E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t\frac{Q_s(t)}{\lambda_s}|c_0=c\right\}\le\delta_s,\forall s\in\boldsymbol{S}\\&\quad\text{C2 }E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t[R(t)-\sum_{s\in\boldsymbol{S}}Z_s(t)R_s(t)]|c_0=c\right\}\le\delta_0\\&\quad\text{C3 }X_s(t)\ge0,\sum_{s\in\boldsymbol{S}}X_s(t)\le N,\forall t\\&\quad\text{C4 }Z_s(t)\in\{0,1\},\sum_{s\in\boldsymbol{S}}Z_s(t)=1,\forall t\end{aligned}\tag{9}$$

其中，$\delta_s,s\in\boldsymbol{S}$为各切片的时延约束，$\delta_0$为网络回传链路带宽消耗约束。

2.3 问题转换

CMDP问题式(9)可以通过拉格朗日理论转化为不受限的MDP问题，定义问题式(9)对应的拉格朗日函数为

$$L(\boldsymbol{\beta},c,\pi)=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t g^{\beta}(c_t,\pi(c_t))|c_0=c\right\}-\sum_{s\in\boldsymbol{S}}\beta_s\delta_s-\beta_0\delta_0\tag{10}$$

其中，$g^{\beta}(c_t,\pi(c_t))=-\sum_{s\in\boldsymbol{S}}\gamma_s(t)BX_s(t)+\sum_{s\in\boldsymbol{S}}\beta_s\frac{Q_s(t)}{\lambda_s}+\beta_0\left[R(t)-\sum_{s\in\boldsymbol{S}}Z_s(t)R_s(t)\right]$为时隙$t$的拉格朗日回报，$\beta_i\ge0\ (i=0,1,\cdots,S)$为拉格朗日乘子，令$\boldsymbol{\beta}=\{\beta_i,i=0,1,\cdots,S\}$。进一步，定义状态值函数为

$$V^{\pi,\beta}(c)=E^{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t g^{\beta}(c_t,\pi(c_t))|c_0=c\right\}\tag{11}$$

因此，优化问题式(9)可转化为式(12)的无约束MDP问题

$$\min_{\pi\in\Phi}\max_{\boldsymbol{\beta}:\beta_i\ge0}\left(V^{\pi,\beta}(c)-\sum_{s\in\boldsymbol{S}}\beta_s\delta_s-\beta_0\delta_0\right)\tag{12}$$

其对偶问题为

$$\max_{\boldsymbol{\beta}:\beta_i\ge0}\min_{\pi\in\Phi}\left(V^{\pi,\beta}(c)-\sum_{s\in\boldsymbol{S}}\beta_s\delta_s-\beta_0\delta_0\right)\tag{13}$$

对于一个给定的$\boldsymbol{\beta}:\beta_i\ge0$，无约束优化问题式(13)对应的最优策略$\pi^{*,\beta}$满足式(14)的贝尔曼最优性方程

$$V^{*,\beta}(c)=\min_{a\in A}\left\{g^{\beta}(c,a)+\gamma\sum_{c'\in C}r_a(c'|c)\,V^{*,\beta}(c')\right\}\tag{14}$$

其中，$V^{*,\beta}:C\to R$称为最优状态值函数。类似地，定义$Q^{*,\beta}:C\times A\to R$为最优行动值函数，其满足

$$Q^{*,\beta}(c,a)=g^{\beta}(c,a)+\gamma\sum_{c'\in C}r_a(c'|c)\,V^{*,\beta}(c')\tag{15}$$

由式(14)和式(15)可得

$$V^{*,\beta}(c)=\min_{a\in A}Q^{*,\beta}(c,a)\tag{16}$$

因此，最优策略$\pi^{*,\beta}$可得

$$\pi^{*,\beta}=\arg\min_{a\in A}Q^{*,\beta}(c,a),\forall c\in C\tag{17}$$

为了叙述方便，接下来将符号中的$\boldsymbol{\beta}$省略，拉格朗日乘子$\boldsymbol{\beta}:\beta_i\ge0$的问题将在下一节讨论。

3 一种虚拟化C-RAN网络下基于在线学习的网络切片虚拟资源分配算法

本文接下来引入决策后状态(Post-Decision State, PDS)的概念并进而提出一种基于在线学习的网络切片虚拟资源分配算法，该算法能够很好地利用系统已知动态信息，提高学习算法的性能。

PDS描述了系统在已知动态发生后，但在未知动态发生前所处的状态，令$\tilde{c}$表示PDS且所有的PDSs均包含于状态空间C中。当时隙t在状态c_t下采取行动a_t，那么时隙t的PDS可表示为

$$\tilde{c}_t=(\tilde{\boldsymbol{Q}}_t,\tilde{\boldsymbol{H}}_t)=(\tilde{\boldsymbol{Q}}_t,\boldsymbol{H}_t)\tag{18}$$

其中，$\tilde{\boldsymbol{Q}}_t=\left\{\tilde{Q}_s(t),s\in\boldsymbol{S}\right\}$，$\tilde{Q}_s(t)=\max[Q_s(t)-D_s(t),0]=\max[Q_s(t)-\gamma_s(t)\cdot B\cdot X_s(t)\cdot\tau/L,0]$。

时隙$t+1$的系统状态为

$$c_{t+1}=(\boldsymbol{Q}_{t+1},\boldsymbol{H}_{t+1}) \tag{19}$$

其中，$\boldsymbol{Q}_{t+1}=\{Q_s(t+1),s\in S\}$，$Q_s(t+1)=\tilde{Q}_s(t)+A_s(t)$。

PDS $\tilde{c}_t$包含了所有的与从状态c_t执行动作a_t再转到c_{t+1}有关的已知信息，下一时隙状态c_{t+1}则包含了所有未知动态，即切片数据包到达过程$A_s(t)$和信道状态$\boldsymbol{H}_{t+1}$。

根据PDS的定义，一般地，系统状态转移概率$r_a(c'|c)$可以被分解为已知和未知两部分，已知部分给出从状态c到PDS $\tilde{c}$的转移概率，未知部分给出从PDS $\tilde{c}$到下一状态c'的转移概率。令$r_a^k(\tilde{c}|c)$和$r^u(c'|\tilde{c})$分别表示已知转移概率和未知转移概率，则有

$$r_a(c'|c)=\sum_{\tilde{c}} r_a^k(\tilde{c}|c)r^u(c'|\tilde{c}) \tag{20}$$

类似地，即时回报函数也可分解为已知和未知两部分，如式(21)所示

$$g(c,a)=g^k(c,a)+\sum_{\tilde{c}} r_a^k(\tilde{c}|c)g^u(\tilde{c}) \tag{21}$$

值得注意的是，本文中不存在未知成本，即$g^u(\tilde{c})=0$，$g^k(c,a)=g(c,a)$。为了便于叙述，定义系统状态与PDS的状态转移方程分别为

$$\tilde{c}_t=S^{M,a}(c_t,a_t) \tag{22}$$

$$c_{t+1}=S^{M,W}(\tilde{c}_t,\boldsymbol{A}_t,\boldsymbol{H}_{t+1}) \tag{23}$$

其中，$\boldsymbol{A}_t=\{A_s(t),s\in\boldsymbol{S}\}$，式(22)与当前采取的行动有关，式(23)与外部随机事件有关，包括业务数据包到达过程与信道状态变化等。

令$\tilde{V}^*$表示最优的PDS状态值函数

$$\tilde{V}^*(\tilde{c})=E\{V^*(c')|\tilde{c}\}=\sum_{c'\in C} r^u(c'|\tilde{c})V^*(c') \tag{24}$$

则最优状态值函数V^*可改写为

$$V^*(c)=\min_{a\in A}\left\{g(c,a)+\gamma\tilde{V}^*(\tilde{c})\right\} \tag{25}$$

因此，最优策略π^*_{PDS}可按式(26)选择

$$\pi^*_{\mathrm{PDS}}=\arg\min_{a\in A}\left\{g(c,a)+\gamma\tilde{V}^*(\tilde{c})\right\} \tag{26}$$

式(25)是式(14)等效的改写形式，所以π^*_{PDS}与π^*也同样等价。因此，PDS状态值函数适用于学习最优策略。从式(25)可以看出，基于PDS的学习算法可通过学习系统未知动态来获得最优值函数V^*和最优策略π^*，可利用迭代的方式逐渐逼近最优的PDS状态值函数$\tilde{V}^*$。本文提出的基于在线学习的网络切片虚拟资源分配算法具体如表1所示。

表 1　虚拟化C-RAN网络下基于在线学习的网络切片虚拟资源分配算法

输入　系统状态空间C，动作空间A，拉格朗日回报函数$g(c_t,\pi(c_t))$，有限信道状态集合$\boldsymbol{H}$。

初始化：初始化决策后状态的状态值函数$\tilde{V}_0(\tilde{c})\in R,\forall\tilde{c}\in C$，令$t\leftarrow 0$，$c_t\leftarrow c\in C$。

学习阶段：

(1) 求解

$$a_t=\arg\min_{a\in A}\left\{g(c_t,a)+\gamma\tilde{V}_t(S^{M,a}(c_t,a))\right\}; \tag{27}$$

(2) 观察PDS状态$\tilde{c}_t$和下一时隙状态c_{t+1}：$\tilde{c}_t=S^{M,a}(c_t,a_t)$，$c_{t+1}=S^{M,W}(\tilde{c}_t,\boldsymbol{A}_t,\boldsymbol{H}_{t+1})$；

(3) 计算c_{t+1}的状态值函数：

$$V_t(c_{t+1})=\min_{a\in A}\left\{g(c_{t+1},a)+\gamma\tilde{V}_t(S^{M,a}(c_{t+1},a))\right\}; \tag{28}$$

(4) 更新$\tilde{V}_{t+1}(\tilde{c}_t)$：

$$\tilde{V}_{t+1}(\tilde{c}_t)=(1-\alpha_t)\tilde{V}_t(\tilde{c}_t)+\alpha_t V_t(c_{t+1}); \tag{29}$$

(5) 利用随机次梯度法更新拉格朗日乘子$\boldsymbol{\beta}:\beta_i\geq 0$。

输出　最优策略π^*_{PDS}。

表1中，α_t是第t次迭代时的学习速率，遵循随机近似条件[6]，其应满足$0<\alpha_t<1$，$\sum_t\alpha_t=\infty$和$\sum_t\alpha_t^2<\infty$。可以证明，当$t\to\infty$时序列$\tilde{V}_t(\tilde{c})$会以概率1收敛到最优PDS状态值函数$\tilde{V}^*(\tilde{c})$。

4　仿真结果与分析

为了评估本文所提出的基于在线学习的网络切片虚拟资源分配算法的有效性，本节将其与文献[4]提出的启发式(heuristic)算法、文献[6]中的Q学习(Q-Learning)以及文献[7]中的比例公平静态共享算法(Static Sharing with PF)作比较，根据仿真结果进行详尽地分析。启发式算法中，在每个离散的资源调度时隙上，对于任一RB，首先计算各网络切片的当前权重，其中权重与当前时隙各切片的最低资源需求、信道状态以及队列状态有关。接着，算法将RB分配给对应权重最大的网络切片。比例公平静态共享算法中，在每个离散的资源调度时隙上，首先根据当前时隙的系统状态计算各网络切片的最低资源需求。接着，在分配给各网络切片最低需求的资源块数后，算法将依据各切片队列长度比例公平的原则分配剩余的资源块。

4.1　参数设置

本文考虑的仿真场景中有两个时延需求不同的网络切片，即$S=2$，将各切片的队列长度离散化为有限个等间距的区间，每个区间表示一种队列状态，因而系统状态空间C为有限状态集合。仿真中主要参数设置如表2所示。

4.2　仿真结果分析

图2和图3分别比较了不同平均时延约束下，4种不同算法在平均切片和速率以及平均切片总时

表 2 仿真参数

仿真参数	值
远端射频头(RRH)最大发射功率	20 dBm
各切片最大队列长度$Q_{s,\max}$	20 packets
噪声功率谱密度	–174 dBm/Hz
数据包大小L	4 kbit/packet
路径损耗模型	104.5+20lg(d) (d[km])
时隙长度τ	1 ms

延方面的性能。从图2和图3中可以看出，平均时延约束越大，不同算法下的平均切片和速率和平均切片总时延都将增大。这是因为平均时延约束越大，算法就会在满足平均时延约束条件下优先考虑平均切片和速率最大化的问题，会出现将信道条件较差的网络切片新到达的业务数据包积压进排队队列，从而能分配更多的资源给信道条件较好的切片以提高整体平均切片和速率的情况。此外，如图2和图3所示，本文提出的基于在线学习的网络切片虚拟资源分配算法具有最高的平均切片和速率，Q学习算法实现了与其相近的平均切片和速率和平均切片总时延性能，启发式算法的平均切片总时延最大，比例公平静态共享算法的平均切片和速率最小，但其平均切片总时延要优于其它3种算法。这是因为在比例公平静态共享算法中，在每个离散的资源调度时隙内，在分配给各网络切片最低需求的资源块数后，算法将按照各切片队列长度比例公平的原则分配剩余的资源块。因此，信道条件较差的网络切片也能分配到相对更多的资源，从而会降低切片平均总时延，但相应地也会影响平均切片和速率。

图4和图5分别比较了不同数据包到达率λ_1下，4种不同算法在平均切片和速率以及平均切片总时延方面的性能。从图中可以看出，随着切片1的业务数据包到达率λ_1增大，不同算法下的平均切片和速率将减小，平均切片总时延将增大。这是因为数据包到达率λ_1越大，每一时隙内切片1的排队队列中堆积的待传数据包也就越多，因而平均切片总时延也会相应增大。此外，为了满足切片1的平均时延约束，算法会在每个离散的资源调度时隙内为其分配更多的资源块，尤其是当信道条件较差的时候，相应地，切片2就会得到相对较少的资源块数

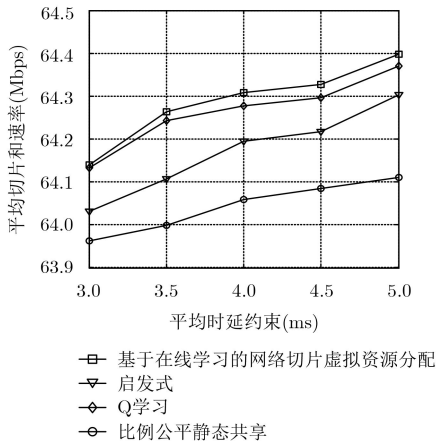

图 2 不同平均时延约束下的平均切片和速率

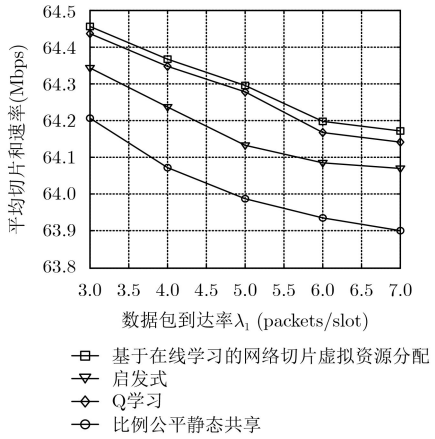

图 4 不同数据包到达率λ_1下的平均切片和速率

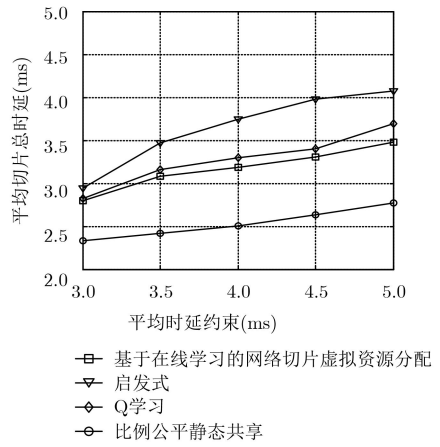

图 3 不同平均时延约束下的平均切片总时延

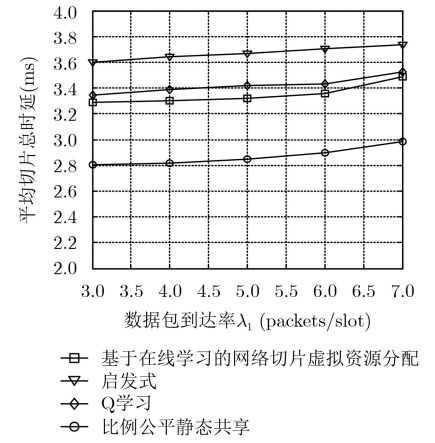

图 5 不同数据包到达率λ_1下的平均切片总时延

量，即使其当前的信道质量较好，即可以理解为以牺牲部分切片和速率为代价来满足切片时延约束，因此平均切片和速率会随着数据包到达率λ_1的增大而减小。

5　结束语

针对现有研究中缺乏C-RAN场景下网络切片的动态资源分配方案的问题，本文将虚拟化C-RAN网络下的资源分配看作一个无穷时间马尔可夫决策过程，建立了一个以最大化平均网络切片和速率为目标，同时受限于各切片平均时延约束以及网络平均回传链路带宽消耗约束的随机优化模型，进而设计了一种基于在线学习的网络切片虚拟资源分配算法。仿真结果显示，所提出的算法能够有效地满足各切片的服务质量需求，降低回传链路带宽消耗的压力并同时提升系统吞吐量。

传统的基于MDP或Q-Learning的强化学习算法是基于查找表来选择每一时刻的最优行动的，然而当状态空间较大时，会引起所谓的“维度灾难(Curses of Dimensionality)”问题，导致算法的可扩展性较差，且对于访问次数较少的状态来说，其值函数的收敛速度较慢。为了解决上述问题，许多新兴的技术受到了学者们的广泛关注和深入研究，如值函数近似策略[8,9]、基于Actor-Critic (AC)的强化学习方法[10]、深度强化学习(Deep Reinforcement Learning, DRL)[11]等，这也是进一步研究工作的重点。

参 考 文 献

[1] 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.

[2] CHECKO A, CHRISTIANSEN H L, YAN Ying, *et al*. Cloud RAN for mobile networks-A technology overview[J]. *IEEE Communications Surveys & Tutorials*, 2015, 17(1): 405–426. doi: 10.1109/COMST.2014.2355255.

[3] NIU Binglai, ZHOU Yong, SHAH-MANSOURI H, *et al*. A dynamic resource sharing mechanism for cloud radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(12): 8325–8338. doi: 10.1109/TWC.2016.2613896.

[4] KALIL M, Al-DWEIK A, SHARKH M F A, *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.

[5] BERTSEKAS D and GALLAGER R. Data Networks[M]. Englewood Cliffs: Prentice-Hall, 1991, 152–162.

[6] YANG Jian, ZHANG Shuben, WU Xiaomin, *et al*. Online learning-based server provisioning for electricity cost reduction in data center[J]. *IEEE Transactions on Control Systems Technology*, 2017, 25(3): 1044–1051. doi: 10.1109/TCST.2016.2575801.

[7] KALIL M, SHAMI A, and YE Yinghua. Wireless resources virtualization in LTE systems[C]. Proceedings of 2014 IEEE Conference on Computer Communications Workshops, Toronto, Canada, 2014: 363–368. doi: 10.1109/INFCOMW.2014.6849259.

[8] POWELL W B. Approximate Dynamic Programming: Solving the Curses of Dimensionality[M]. Hoboken, USA: Wiley, 2011, 289–388.

[9] LAKSHMINARAYANAN C and BHATNAGAR S. Approximate dynamic programming with (min, +) linear function approximation for Markov decision processes[J]. arXiv preprint arXiv: 1403.4179, 2014.

[10] LI Rongpeng, ZHAO Zhifeng, CHEN Xianfu, *et al*. TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(4): 2000–2011. doi: 10.1109/TWC.2014.022014.130840.

[11] HE Xiaoming, WANG Kun, HUANG Huawei, *et al*. Green resource allocation based on deep reinforcement learning in content-centric IoT[J]. *IEEE Transactions on Emerging Topics in Computing*, 2019. doi: 10.1109/TETC.2018.2805718.

唐　伦：男，1973年生，教授，主要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.

马润琳：女，1993年生，硕士生，研究方向为5G网络切片、网络功能虚拟化、无线资源分配.

贺小雨：女，1995年生，硕士生，研究方向为5G网络切片、无线网络虚拟化、智能优化理论.

陈前斌：男，1967年生，教授，博士生导师，主要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.