

非正交多址接入系统中基于受限马尔科夫决策过程的网络切片虚拟资源分配算法

唐 伦* 施颖洁 杨希希 陈前斌

(重庆邮电大学移动通信技术重点实验室 重庆 400065)

摘 要: 针对无线接入网络切片虚拟资源分配优化问题, 该文提出基于受限马尔可夫决策过程(CMDP)的网络切片自适应虚拟资源分配算法。首先, 该算法在非正交多址接入(NOMA)系统中以用户中断概率和切片队列积压为约束, 切片的总速率作为回报, 运用受限马尔可夫决策过程理论构建资源自适应问题的动态优化模型; 其次定义后决策状态, 规避最优值函数中的期望运算; 进一步地, 针对马尔科夫决策过程(MDP)的“维度灾难”问题, 基于近似动态规划理论, 定义关于分配行为的基函数, 替代决策后状态空间, 减少计算维度; 最后设计了一种自适应虚拟资源分配算法, 通过与外部环境的不断交互学习, 动态调整资源分配策略, 优化切片性能。仿真结果表明, 该算法可以较好地提高系统的性能, 满足切片的服务需求。

关键词: 5G网络切片; 资源分配; 受限马尔可夫决策过程; 非正交多址接入

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2018)12-2962-08

DOI: 10.11999/JEIT180131

Network Slice Virtual Resource Allocation Algorithm Based on Constrained Markov Decision Process in Non-orthogonal Multiple Access

TANG Lun SHI Yingjie YANY Xixi CHEN Qianbin

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

Abstract: An adaptive virtual resource allocation algorithm is proposed based on Constrained Markov Decision Process (CMDP) for wireless access network slice virtual resource allocation. First of all, this algorithm in the Non-Orthogonal Multiple Access (NOMA) system, uses the user outage probability and the slice queues as constraints, uses the total rate of slices as a reward to build a resource adaptive problem using the CMDP theory. Secondly, the post-decision state is defined to avoid the expectation operation in the optimal value function. Furthermore, aiming at the problem of “dimensionality disaster” of MDP, based on the approximate dynamic programming theory, a basis function for the assignment behavior is designed to replace the post-decision state space and to reduce the computational dimension. Finally, an adaptive virtual resource allocation algorithm is designed to optimize the slicing performance. The simulation results show that the algorithm can improve the performance of the system and meet the service requirements of slicing.

Key words: 5G virtual network slices; Resource allocation; Constrained Markov Decision Process (CMDP); Non-Orthogonal Multiple Access (NOMA)

1 引言

随着通信不断地发展, 通信业务不断的扩大, 网络切片技术应运而生^[1]。网络切片是一种相互独

立且具有灵活性的虚拟逻辑网络, 可以为不同应用场景下的用户需求提供有力保障。

传统基于切片的资源分配, 可以考虑频谱调度等方面为切片分配资源^[2]。文献^[3]在受到功率限制的特定区域内最大化虚拟无线网络的总吞吐量。然而, 在切片用户得到的虚拟资源没有考虑到资源的功率域。文献^[4]利用Q学习设计一个动态的资源分配优化方案, 提高系统性能。而此方案在面临大量级用户时, 可能会面临“维数灾难”问题。

收稿日期: 2018-01-30; 改回日期: 2018-08-16; 网络出版: 2018-08-23

*通信作者: 唐伦 tangl@cqupt.edu.cn

基金项目: 国家自然科学基金(61571073)

Foundation Item: The National Natural Science Foundation of China (61571073)

为了满足5G网络海量接入、超高带宽的要求^[5],同时充分利用频谱资源,基于NOMA(非正交多址接入)的资源分配也成为研究的热点。文献[6-9]在NOMA系统下,为用户制定了最优的功率分配方案,但是都没有考虑到用户的不同业务需求。

针对以上问题,本文提出了基于CMDP(受限马尔科夫决策过程)的网络切片自适应虚拟资源分配算法。本文的贡献主要包括:(1)在NOMA系统中以用户中断概率和切片队列积压为约束,切片的总速率作为回报,运用CMDP理论构建资源自适应问题的动态优化模型;(2)定义后决策状态,规避了最优值函数中的期望运算;针对MDP的“维度灾难”问题,基于近似动态规划理论,定义基函数,替代决策后状态空间,减少计算维度;(3)设计了一种自适应虚拟资源分配算法,通过与外部环境的不断交互学习,动态调整资源分配策略,在满足切片QoS需求的同时优化切片性能。

2 网络切片调度模型

图1所示为系统场景图。在该系统中,有 L 个网

络切片,支持 L 种不同类型的业务,每个切片有 k_l 个用户。在 T 时间段内,资源调度器根据监测到的切片信息实时分配资源。

2.1 切片调度模型

假设,在 t 时刻的用户到达数为 $A(t)$, $A(t)$ 可以表示为

$$A(t) = \sum_{i=1}^l A_i(t) \quad (1)$$

其中, $A_i(t)$ 为 t 时刻切片 i 的到达用户数, i 表示不同的网络切片。

如图2所示,在 t 时刻,切片管理器根据机会调度策略^[10]将所有切片用户分配至各个不同的资源块。在NOMA系统中,采用功率复用技术为不同用户分配资源,为信道增益较高的用户分配较少的功率^[11]。将切片 l 的用户接入资源块 RB_n 的请求队列定义为 $Q_{l,n}^s(t) < \infty$ 。定义 $Q^s(t) = (Q_{l,n}^s(t))$ 为时隙 t 内资源块 RB_n 的切片 l 的用户队列长度。在 t 时刻的资源块上缓存切片 l 的队列积压可表示为

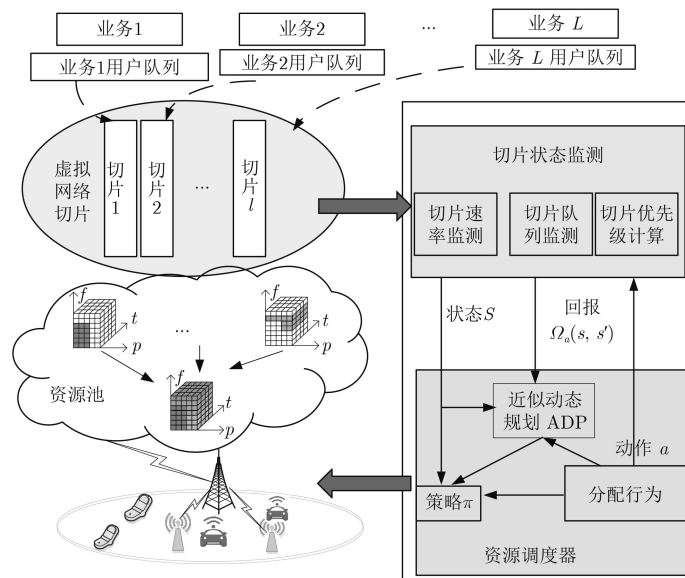


图1 系统场景图

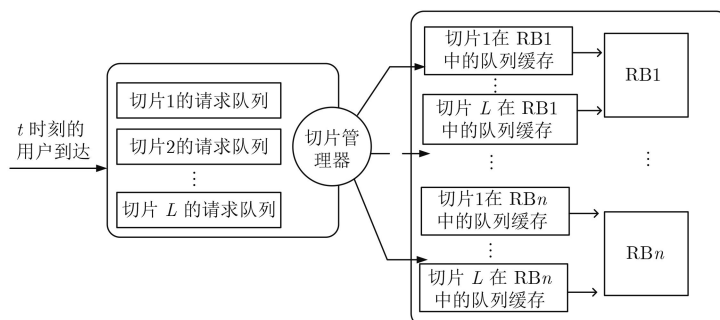


图2 切片调度模型

$$Q(t) = \sum_{l \in L} \sum_{n \in N} Q_{l,n}^S(t) \quad (2)$$

进一步定义平均队列积压 $\bar{Q}$ 为

$$\bar{Q} = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t \in T} E\{Q(t)\} \quad (3)$$

2.2 NOMA用户速率

该系统中的所有信道服从独立同分布的瑞利衰落信道，信道噪声为高斯白噪声。假定切片 l 接入一个资源块内同一个子载波 m 中的用户有 $k_{l,m}$ 个。设总发射功率为 P ，即 $P = \sum_{i=1}^{k_{l,m}} p_{i,m}$ ， $p_{i,m}$ 表示在子载波 m 上为用户 i 分配的功率。

在NOMA的下行链路中，在子载波 m 上 i 用户的接收信号为

$$y_{i,m} = h_{i,m} S_m + w_{i,m} \quad (4)$$

其中， $h_{i,m}$ 和 $w_{i,m}$ 分别表示在子载波 m 上基站到接收端的信道系数和噪声，噪声的均值为零，方差为 σ_m^2 。

根据香农公式，可以计算子载波 m 上的用户 i 的速率为 $r_{i,m}$ ：

$$r_{i,m} = B \log_2 \left(1 + \frac{p_{i,m} \Gamma_{i,m}}{k_{l,m} + \sum_{j=1}^{k_{l,m}} p_{j,m} \Gamma_{j,m}} \right) \quad (5)$$

其中， B 为子载波 m 的带宽， $\Gamma_{i,m}$ 为用户 i 的载波干扰噪声比(CINR)。

假定切片 l 接入的子载波集合为 M ，定义 R_l^M 为切片总速率：

$$R_l^M = \sum_{m=1}^M \sum_{k=1}^{k_{l,m}} r_{k,m} \quad (6)$$

NOMA系统的全局中断概率为

$$P_{\text{tot}}^m = 1 - \Pr \{ \forall r_{i,m} > R_l^{\min} \} \quad (7)$$

3 网络切片中资源分配CMDP建模

本文将无线网络切片中的虚拟资源分配问题描述为一个受限马尔科夫决策过程(CMDP)，其基本元素包括：系统状态、分配行为、状态转移关系和目标函数。

3.1 系统状态

本文在 t 时刻的状态由内部状态 $u(t)$ 和外部状态 $w(t)$ 组成。其中，内部状态 $u(t)$ 描述切片 l 在资源块 RB_n 上的子载波和功率分配情况。定义 $N_{ln}(t)$ 为资源块 RB_n 分配给切片 l 的子载波数， $P_{ln}^m(t)$ 为分配的子载波 m 的功率粒度，则系统的内部状态 $u(t)$ 可以

表示为

$$\begin{aligned} u(t) &= (u_{ln}(t))_{l \in L, n \in N} \\ &= (N_{ln}(t), P_{ln}^m(t))_{l \in L, n \in N, m \in M} \end{aligned} \quad (8)$$

外部状态 $w(t)$ 为更新的业务切片权值。每个切片的权值由切片业务需求和切片队列状态共同决定：

$$w(t) = (w_l(t))_{l \in L} \quad (9)$$

其中， $w_l(t)$ 为 t 时刻更新的业务切片 l 权值，具体地，利用式(10)计算切片在每个子载波的实时权重：

$$\omega_l(t) = -\frac{\lg \delta_l}{\tau_l} \times D_{\text{HOL},l} \times \frac{\bar{R}_l(t)}{R_l(t)} \quad (10)$$

其中， δ_l 是介于0到1之间的值， τ_l 表示切片所允许的最大时延门限， $D_{\text{HOL},l}$ 表示切片 l 所在队列队首的等待时延， $\bar{R}_l(t)$ 为切片 l 在当前时隙的理想瞬时速率， $R_l(t)$ 为切片 l 在当前时隙的速率。

3.2 分配行为

定义 A_s 为状态 s 下可采取的功率分配行为的集合：

$$A_s = \{\alpha_{ln}(t), \beta_{ln}(t)\} \quad (11)$$

其中， $\alpha_{ln}(t)$ 为 t 时刻在资源块 RB_n 内切片 l 的分配功率粒度调整行为； $\beta_{ln}(t)$ 为 t 时刻为切片 l 分配子载波的调整行为。

3.2.1 调整功率分配粒度 该行为通过改变功率分配的粒度，控制一个子载波内的用户数。 $P_{ln}^m(t)$ 为 t 时刻子载波 m 的功率分配粒度。假设资源块 n 内的子载波 m 服务用户数为 k ，则其总功率满足以下条件：

$$P_{\max} \geq \sum_{i=1}^k i(P_{ln}^m(t) \pm \alpha_{ln}(t)) \quad (12)$$

在每一时刻确定功率分配粒度后，可以得到该子载波最大服务用户数 $k_{\max}^m(t)$ ：

$$k_{\max}^m(t) = \left\lceil -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2P_{\max}}{P_{ln}^m(t) \pm \alpha_{ln}(t)}} \right\rceil \quad (13)$$

当前切片 l 的队列可以表示为

$$Q_{ln}^S(t) = Q_{ln}^S(t-1) - \sum_{m=1}^M k_{\max}^m(t) + A_l(t) \quad (14)$$

3.2.2 切片子载波分配 状态 s 下可采取的另一种功率分配行为是改变切片的子载波数。当切片 l 的队列积压超过最大值，且功率分配粒度已到达最小值，则考虑为该切片分配额外的子载波。本文定义 $\beta_{ln}(t) = \{-1, +1\}$ 。这里要说明，分配的子载波的功率粒度与原子载波的功率分配粒度一致。

3.3 状态转移方程

系统的状态转移方程可表示为

$$s(t+1) = s(s(t), a_s(t), w(t+1)) \quad (15)$$

系统内部状态 $u(t)$ 的转移关系如式(16)所示:

$$\begin{aligned} u_{ln}(t+1) &= (N_{ln}(t+1), P_{ln}^m(t+1))_{l \in L, n \in N, m \in n} \\ &= (\Psi^N(u_{ln}(t), \beta_{ln}(t)), \Psi^P(u_{ln}(t), \alpha_{ln}(t))), \\ &= \begin{cases} (N_{ln}(t), P_{ln}^m(t) \pm \alpha_{ln}(t)), \\ P_{\min} \leq \sum_{i=1}^k k(P_{ln}^m(t) \pm \alpha_{ln}(t)) \leq P_{\max} \\ (N_{ln}(t) + \beta_{ln}(t), P_{ln}^m(t) \pm \alpha_{ln}(t)), \\ P_{\min} > \sum_{i=1}^k k(P_{ln}^m(t) \pm \alpha_{ln}(t)) \text{ 或} \\ P_{\max} < \sum_{i=1}^k k(P_{ln}^m(t) \pm \alpha_{ln}(t)) \end{cases} \quad (16) \end{aligned}$$

其中, Ψ^N 和 Ψ^P 分别表示 $N_{ln}(t)$ 和 $P_{ln}^m(t)$ 的转移方程。系统的外部状态 $w(t)$ 取决于切片中用户到达过程以及资源分配后的切片状态, 与行为无关, 在系统不断和外部环境交互学习的过程中不断更新。

3.4 回报函数

对于策略 $\pi \in \Pi$, 由值函数 $V^\pi(s)$ 进行衡量:

$$R_l(t) = \begin{cases} \sum_{n=1}^N \sum_{m=1}^{N_{ln}(t)} \sum_{k=1}^{k_{n,m}} B_m \log_2 \left(1 + \frac{k(P_{ln}^m(t) \pm \alpha_{ln}(t)) \Gamma_{i,m}}{1 + \sum_{j=1}^{k_{l,m}} j(P_{ln}^m(t) \pm \alpha_{ln}(t)) \Gamma_{j,m}} \right), \\ P_{\min} \leq \sum_{i=1}^k i(P_{ln}^m(t) \pm \alpha_{ln}(t)) \leq P_{\max} \\ \sum_{n=1}^N \sum_{m=1}^{N_{ln}(t) \pm 1} \sum_{k=1}^{k_{n,m}} B_m \log_2 \left(1 + \frac{k(P_{ln}^m(t) \pm \alpha_{ln}(t)) \Gamma_{i,m}}{1 + \sum_{j=1}^{k_{l,m}} j(P_{ln}^m(t) \pm \alpha_{ln}(t)) \Gamma_{j,m}} \right), \\ P_{\min} > \sum_{i=1}^k i(P_{ln}^m(t) \pm \alpha_{ln}(t)) \text{ 或 } P_{\max} < \sum_{i=1}^k i(P_{ln}^m(t) \pm \alpha_{ln}(t)) \end{cases} \quad (21)$$

资源管理器的最终目标是在平均队列和中断概率的约束下, 最大化回报函数:

$$\begin{aligned} &\max_{\alpha, \beta} E^\pi \left\{ \sum_{t=0}^{\infty} \Omega_a(s, s') \right\} \\ &\text{s.t. } \begin{cases} \bar{Q}(t) < Q_{\max} \\ \frac{1}{L} \sum_{l \in L} \bar{P}_l^{\text{out}}(t) < P_{\max}^{\text{out}} \end{cases} \end{aligned} \quad (22)$$

在式(22)的约束条件中, $\bar{Q}(t)$ 为 t 时刻系统的平

$$V^\pi(s) = E^\pi \left\{ \sum_{t=0}^{\infty} \gamma^t \Omega_a(s, s') \mid s(t) = s, a_s(t) = a \right\} \quad (17)$$

其中, γ 为折扣因子, $\Omega_a(s, s')$ 为与切片速率和权值有关的回报函数。

最优值函数如式(18)所示:

$$V^*(s) = \max_{a_s(t) \in A_s} E^\pi \{ \Omega_a(s, s') + \gamma V^*(s') \mid s(t) = s, a_s(t) = a \} \quad (18)$$

对应地, 状态 s 下的最优策略为

$$\pi^*(s) = \arg \max V^*(s) \quad (19)$$

定义回报函数 $\Omega_a(s, s')$ 为

$$\Omega_a(s, s') = \sum_{l=1}^L \omega_l(t) R_l(t) \quad (20)$$

其中, $\omega_l(t)$ 为系统的切片外部状态 $w(t)$, $R_l(t)$ 为资源分配完成后的切片速率。

资源分配完成后的切片速率, 即内部状态转换后的切片速率, 由式(16)可得, $R_l(t)$ 如式(21)所示。

均队列, $Q_{\max}$ 为系统允许的最大队列。 $\frac{1}{L} \sum_{l \in L} \bar{P}_l^{\text{out}}(t)$

为系统的平均中断概率, $P_{\max}^{\text{out}}$ 为系统可允许的最大中断概率。

4 资源自适应算法设计

本文的CMDP模型中无法得到系统转移概率, 当规模较大时容易陷入“维数灾难”问题^[12]。针对以上问题, 本文采用基于ADP(近似动态规划)算法, 避免由转移概率而产生的期望运算, 提高算法的效率。

4.1 后决策状态

定义后决策状态 $S^a(t)$, 将系统的状态转移方程等价的拆分为以下两个步骤^[13]:

$$\begin{aligned} S^a(t) &= u_{ln}^a(t) = (N_{ln}^a(t), P_{ln,m}^a(t))_{l \in L, n \in N, m \in n} \\ &= u_{ln}(t+1) \\ &= (N_{ln}(t+1), P_{ln}^m(t+1))_{l \in L, n \in N, m \in n} \\ &= S^a(S(t), a_s(t)) \end{aligned} \quad (23)$$

$$\begin{aligned} S(t+1) &= (u(t+1), w(t+1)) \\ &= S^w(S^a(t), w(t+1)) \end{aligned} \quad (24)$$

式(23)、式(24)中状态 $S_{ln}^a(t)$ 和 $S(t+1)$ 分别为 t 时刻的后决策状态和 $t+1$ 时刻的决策前状态。系统状态的转移过程如图3所示。

根据后决策状态的定义, 可以将最优值函数式(18)改写为式(25)

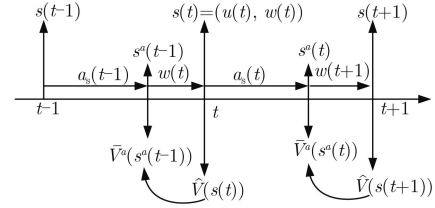


图3 状态转移图

$$V(s) = \max_{a_s(t) \in A_s} \left\{ \Omega^a(S^a(t)) + \gamma \bar{V}^a(S_{ln}^a(t)) \right\} \quad (25)$$

其中, $\bar{V}^a(S^a(t))$ 为值函数的估计值。 $\Omega^a(S^a(t))$ 为效用函数关于 $S^a(t)$ 重新定义的函数, 如(26)所示。

$$\begin{aligned} \Omega^a(S^a(t)) &= E\{\Omega_a(s, s')\} \\ &= \sum_{l=1}^L \omega_l(t+1) R_l^a(S^a(t)) \end{aligned} \quad (26)$$

其中, $R_l^a(S^a(t))$ 如式(27)所示。

$$R_l^a(S^a(t)) = \sum_{n=1}^N \sum_{m=1}^{N_{ln}(S^a(t))} \sum_{k=1}^{k_{n,m}} B_m \log_2 \left(1 + \frac{k [p_{ln}^m(S^a(t)) \pm \alpha_{ln}(S^a(t))] \Gamma_{i,m}}{1 + \sum_{j=k+1}^{k_{l,m}} j [p_{ln}^m(S^a(t)) \pm \alpha_{ln}(S^a(t))] \Gamma_{j,m}} \right) \quad (27)$$

将后决策状态下的队列信息以及中断概率定义为式(28)和式(29):

$$Q^a(S^a(t)) = \bar{Q}(t) = \lim_{T \rightarrow \infty} \frac{1}{T} Q(t) \quad (28)$$

$$\begin{aligned} P^a(S^a(t)) &= \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{N} \sum_{i \in A} P_i^{\text{out}}(t) \\ &= \frac{1}{N} \sum_{i \in A} \bar{P}_i^{\text{out}}(S^a(t)) \end{aligned} \quad (29)$$

4.2 资源自适应算法设计

本节基于后决策状态设计资源自适应算法。对于优化模型式(22)中, 本文利用拉格朗日算法将约束条件代入目标函数:

$$\begin{aligned} L(\pi, \lambda_1, \lambda_2) &= \sum_{t=0}^{\infty} \Omega_a(s, s') - \lambda_1 (\bar{Q}(t) - Q_{\min}) \\ &\quad - \lambda_2 \left(\frac{1}{N} \sum_{i \in A} \bar{P}_i^{\text{out}}(t) - P_{\max}^{\text{out}} \right) \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \left[\sum_{t=0}^{\infty} \gamma^t \Omega_a(s, s') - \lambda_1 (Q(t) - T Q_{\min}) \right] \\ &\quad - \lim_{T \rightarrow \infty} \frac{1}{T} \left[\lambda_2 \left(\frac{1}{N} \sum_{i \in A} P_i^{\text{out}}(t) - T P_{\max}^{\text{out}} \right) \right] \\ &= \Omega^a(S^a(t)) - \lambda_1 (Q^a(S_{ln}^a(t)) - Q_{\min}) \\ &\quad - \lambda_2 (P^a(S_{ln}^a(t)) - P_{\max}^{\text{out}}) \end{aligned} \quad (30)$$

引入拉格朗日因子 λ_1, λ_2 , 值函数模型式(25)可以转化为式(31):

$$(\pi^*, \lambda_1^*, \lambda_2^*) = \arg \max_{\lambda_1 \geq 0, \lambda_2 \geq 0, \pi} L(\pi, \lambda_1, \lambda_2) \quad (31)$$

利用式(32)和式(33)更新拉格朗日因子:

$$\lambda_1^{t+1} = \lambda_1^t + \varsigma_1 (\bar{Q}^{\pi_{\lambda_1, \lambda_2}}(t) - Q_{\min}) \quad (32)$$

$$\lambda_2^{t+1} = \lambda_2^t + \varsigma_2 \left(\frac{1}{N} \sum_{i \in A} \bar{P}_i^{\pi_{\lambda_1, \lambda_2}}(t) - P_{\max}^{\text{out}} \right) \quad (33)$$

将拉格朗日函数代入最优值函数中, 最优值函数如式(34)所示。

$$V(s) = \max_{a_s(t) \in A_s} \left\{ L^a(S^a(t)) + \gamma \bar{V}^a(S^a(t)) \right\} \quad (34)$$

定义基函数如表1所示。

根据表1利用行动表达的基函数, 引入参数向量 $\eta = \{\eta_1, \eta_2, \dots, \eta_h\}^T$, $h \in \mathcal{h}$ 表示资源分配行为的

表1 基函数定义

基函数	描述
$P_{ln}^m(t) + \alpha_{ln}(t)$	切片 l 功率分配粒度
$N_{ln}(t) + \beta_{ln}(t)$	切片 l 的子载波数
$(P_{ln}^m(t) + \alpha_{ln}(t))^2$	切片 l 功率分配粒度平方
$(N_{ln}(t) + \beta_{ln}(t))^2$	切片 l 的子载波数平方
$(N_{ln}(t) + \beta_{ln}(t))(P_{ln}^m(t) + \alpha_{ln}(t))$	切片 l 中功率分配粒度与子载波数的乘积

关键特征。根据基函数向量 $\chi = \{\chi_1(S^a), \chi_2(S^a), \dots, \chi_h(S^a)\}$ ，可以利用线性函数式(35)得到近似值函数。

$$\bar{V}^a(S^a(t)) = \eta^T \chi = \sum_{h \in h} \eta_h \chi_h(S^a(t)) \quad (35)$$

基于式(35)的值函数近似，可以将优化问题式(34)改为

$$\bar{V}^*(s) = \max_{a_s(t) \in A_s} \left\{ L^a(S^a(t)) + \gamma \sum_{h \in h} \eta_h \chi_h(S^a(t)) \right\} \quad (36)$$

在式(36)中，只需更新参数向量 η ，即可得到估计值。利用梯度算法更新参数向量 η ，如式(37)所示。

$$\eta^* = \arg \min E \left\{ \frac{1}{2} \left[\bar{V}^a(\eta) - \hat{V} \right]^2 \right\} \quad (37)$$

其中， $\bar{V}^a(\eta)$ 为估计值， $\hat{V}$ 为样本值。近似函数值 $\bar{V}^a(\eta)$ 关于参数向量 η 的随机梯度如式(38)所示：

$$\begin{aligned} \nabla_{\eta} \bar{V}^a(S^a(t) | \eta) &= \left(\frac{\partial \bar{V}^a(S^a(t) | \eta)}{\partial \eta_1}, \dots, \frac{\partial \bar{V}^a(S^a(t) | \eta)}{\partial \eta_{|h|}} \right)^T \\ &= (\chi_1(S^a(t)), \dots, \chi_{|h|}(S^a(t)))^T \\ &= \chi(S^a(t)) \end{aligned} \quad (38)$$

参数向量 η 沿梯度方向更新，更新式如式(39)：

$$\begin{aligned} \eta &\leftarrow \eta - \mu_{t-1} (\bar{V}^a(S^a(t-1) | \eta) \\ &\quad - \hat{V}^a(S(t))) \chi(S^a(t-1)) \end{aligned} \quad (39)$$

式(37)中的样本函数 $\hat{V}^a(S(t))$ 表达式如式(40)，该样本值根据决策前状态 $S(t+1)$ 的值函数所得：

$$\hat{V}^a(S(t)) = \max_{a_s(t) \in A_s} \left\{ L^a(S^a(t)) + \gamma \bar{V}^a(S_{ln}^a(t)) \right\} \quad (40)$$

则当前最优策略行为可以表示为

$$\pi^*(s) = \arg \max \left\{ L^a(S^a(t)) + \gamma \sum_{h \in h} \eta_h \chi_h(S^a(t)) \right\} \quad (41)$$

算法的具体步骤如表2所示。

5 性能仿真与结果分析

本小结对本文提出虚拟资源自适应分配算法参照文献[13,14]中NOMA系统的仿真数值，利用MATLAB工具进行数值仿真。为了验证该算法的性能，仿真实验分为两个阶段：学习阶段和测试阶段。仿真时间一共持续600周期，其中学习阶段占400周期，测试阶段200周期。具体系统仿真参数如表3所示。

为了更好体现本文所提基于NOMA的网络切

表 2 基于近似动态规划的资源自适应算法

输入：	$\chi_h(S^a(t))$ ：基函数； γ ：折扣因子；
输出：	η ：参数向量； λ_1, λ_2 ：拉格朗日因子；
(1)	while a new time period starts do
(2)	$t \leftarrow 0$; $\eta \leftarrow 0$; $\lambda_1, \lambda_2 \leftarrow 0$; //初始化
(3)	for ($t = 1$; $t \leq T$; $t++$)
(4)	while
(5)	while
(6)	根据式(40)更新样本函数值
(7)	if $t > 0$ then
(8)	根据式(39)更新参数向量 η
(9)	End if
(10)	采样外部随机变量 $w(t+1)$ 的样本值
(11)	代入更新参数向量 η ，根据式(35)更新决策后状态的近似函数值
(12)	end while
(13)	根据式(34)代入最优策略行为计算目标函数
(14)	根据式(32)和式(33)更新 λ_1, λ_2
(15)	end while
(16)	end for
(17)	end while

表 3 系统仿真参数

仿真参数	仿真值
子载波数	64
基站发射功率	33 dBm
路径损耗	$133.6 + 35 \lg(d)$
传输天线数	1
接收天线数	1
基站服务范围	500 m
单个子载波叠加用户数	1~4 (个)
分配行为：调整功率粒度	$\alpha = \{0.25, 0.50, 1.00\}$
分配行为：调整子载波数	$\beta = 1$
切片1需求	(5 ms, 200 kbit/s)
切片2需求	(10 ms, 500 kbit/s)
切片3需求	(50 ms, 1 Mbit/s)

片自适应虚拟资源分配算法(Power Granularity Uncertain - Approximate Dynamic Programming, PGU-ADP)的性能，将本文算法与文献[4]中的PGU-QL算法和文献[15]中的PGC-F算法进行比较。

图4为在整个仿真过程中近似值和样本值的对比。由仿真图可以看出，在前400周期即学习阶段，近似值与样本值最初差距较大，通过不断学习，近似值与样本值的差距渐渐减小，在400周期附近接近重合。在后200周期即测试阶段，参数向量仍然根据当前状态不断更新，但波动较小，基本

趋于稳定,且近似值与样本值基本重合。

图5为当功率分配粒度分别为0.25, 0.50, 1.00且最大中断概率 $P_{\max}$ 分别为0.3和0.6时,中断概率的变化。在 $P_{\max}=0.3$ 条件下,可以看出,随着分配粒度 α 的增加,中断概率越来越大,但是,当用户数大于170左右时, $\alpha=0.25$ 的中断概率大于其它所有的。在 $P_{\max}=0.6$ 条件下,中断概率增加较快, α 分别为0.25, 0.50, 1.00的上升趋势差距不大,但当 $\alpha=0.25$ 时,中断概率变化波动幅度较大,直至服务用户数稳定,但该情况下平均中断概率达到0.5970,大于 $\alpha=1.00$ 和 $\alpha=0.50$ 的情况。

图6为当切片中用户数增加,不同方案下总速率的比较。PGC-F方案中总速率稳步上升,当切片中用户数到100左右由于服务用户数的限制,增速变缓。PGU-QL方案在用户数较少时无法完成收敛,速率较低,但当有一定用户时,增速较快,与PGU-ADP方案持平。本文PGU-ADP方案,收敛速度快,速率一直处于上升状态,当用户数到达180时增速变缓。

图7为当用户数增加,不同方案下平均队列积压的比较。明显可以看出,PGC-F方案的队列积压最大,PGU-QL方案随着用户的增加,增速变

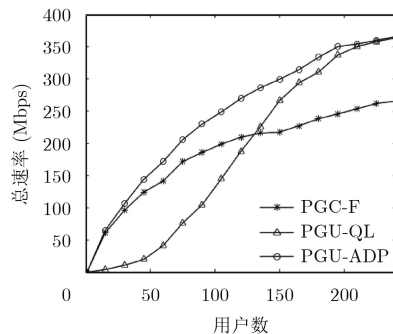


图6 不同方案下,总速率的比较

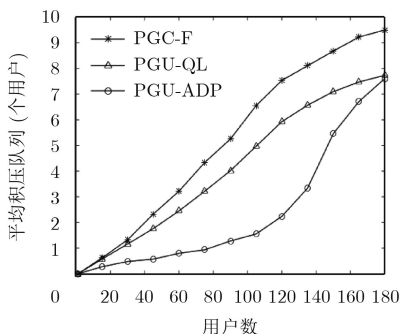


图7 不同方案下,平均队列积压的比较

缓。本文PGU-ADP方案,当用户数大于140时,队列积压由于中断概率的增加会极速增加,但在用户数小于140时有明显优势。

6 结束语

本文针对无线接入网络切片虚拟资源分配优化问题,提出了基于CMDP的网络切片自适应虚拟资源分配算法PGU-ADP。该算法首先针对切片业务需求的多样性,分析切片状态(切片速率和切片队列长度);然后,运用CMDP理论构建资源自适应分配问题的动态优化模型,通过改变功率粒度大小或改变带宽,在满足功率约束的条件下,满足切片的速率与时延需求;最后利用近似动态理论提出一种资源自适应分配算法,并仿真验证其性能。仿真结果表明,该算法可以较好地提高系统的性能,满足用户的服务需求。

参考文献

- [1] 唐伦, 张亚, 梁荣, 等. 基于网络切片的网络效用最大化虚拟资源分配算法[J]. 电子与信息学报, 2017, 39(8): 1812-1818. doi: 10.11999/JEIT161322.
- TANG Lun, ZHANG Ya, LIANG Rong, et al. Virtual resource allocation algorithm for network utility maximization based on network slicing[J]. *Journal of Electronics & Information Technology*, 2017, 39(8): 1812-1818. doi: 10.11999/JEIT161322.
- [2] SALLENT O, PEREZ-ROMERO J, FERRUS R, et al. On

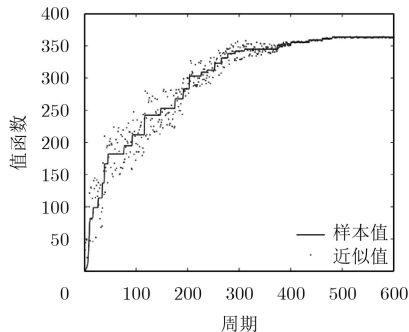


图4 连续600个周期内值函数近似值与样本值比较

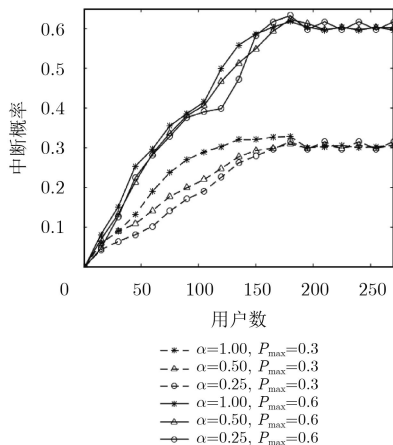


图5 不同分配行为及约束条件下,中断概率比较

- radio access network slicing from a radio resource management perspective[J]. *IEEE Wireless Communications*, 2017, 24(5): 166–174. doi: [10.1109/MWC.2017.1600220WC](#).
- [3] PARSAEEFARD S, DAWADI R, DERAKHSHANI M, *et al.* Joint user-association and resource-allocation in virtualized wireless networks[J]. *IEEE Access*, 2016, 4: 2738–2750. doi: [10.1109/ACCESS.2016.2560218](#).
- [4] BEGA D, GRAMAGLIA M, BANCHS A, *et al.* Optimising 5G infrastructure markets: The business of network slicing[C]. IEEE Conference on Computer Communications, Atlanta, USA, 2017: 1–9. doi: [10.1109/INFOCOM.2017.8057045](#).
- [5] AKPAKWU G A, SILVA B J, HANCKE G P, *et al.* A survey on 5G networks for the internet of things: Communication technologies and challenges[J]. *IEEE Access*, 2018, 6: 3619–3647. doi: [10.1109/ACCESS.2017.2779844](#).
- [6] ZHANG Zihan, XIA Qinghong, YU Guanding, *et al.* Power control, user scheduling and resource allocation for downlink NOMA systems with imperfect channel state information[C]. 2017 9th International Conference on Wireless Communications and Signal Processing, Nanjing, China, 2017: 1–6. doi: [10.1109/WCSP.2017.8171063](#).
- [7] ZHU Jianyue, WANG Jiaheng, HUANG Yongming, *et al.* On optimal power allocation for downlink non-orthogonal multiple access systems[J]. *IEEE Journal on Selected Areas in Communications*, 2017, 35(12): 2744–2757. doi: [10.1109/JSAC.2017.2725618](#).
- [8] FANG Fang, ZHANG Haijun, CHENG Julian, *et al.* Joint user scheduling and power allocation optimization for energy efficient NOMA systems with imperfect CSI[J]. *IEEE Journal on Selected Areas in Communications*, 2017, 35(12): 2874–2885. doi: [10.1109/JSAC.2017.2777672](#).
- [9] DAWADI R, PARSAEEFARD S, DERAKHSHANI M, *et al.* Power-efficient resource allocation in NOMA virtualized wireless networks[C]. IEEE Global Communications Conference, Washington D.C., USA, 2016: 1–6. doi: [10.1109/GLOCOM.2016.7842162](#).
- [10] ZHANG Qi, ZHU Quanyan, ZHANI M F, *et al.* Dynamic service placement in geographically distributed clouds[C]. 2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, 2012: 526–535. doi: [10.1109/ICDCS.2012.74](#).
- [11] ISLAM S M R, AVAZOV N, DOBRE O A, *et al.* Power-domain Non-Orthogonal Multiple Access (NOMA) in 5G systems: potentials and challenges[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(2): 721–742. doi: [10.1109/COMST.2016.2621116](#).
- [12] CHEN Tianyi, MOKHTARI A, WANG Xin, *et al.* Stochastic averaging for constrained optimization with application to online resource allocation[J]. *IEEE Transactions on Signal Processing*, 2017, 65(12): 3078–3093. doi: [10.1109/TSP.2017.2679690](#).
- [13] POWELL W B. Approximate Dynamic Programming: Solving the Curses of Dimensionality[M]. NJ, Princeton: Wiley, 2007: 129–144.
- [14] FANG Fang, ZHANG Haijun, CHENG Julian, *et al.* Energy-efficient resource scheduling for NOMA systems with imperfect channel state information[C]. IEEE International Conference on Communications, Paris, France, 2017: 1–5. doi: [10.1109/ICC.2017.7996360](#).
- [15] RAI R, ZHU H, and WANG Jiangzhou. Resource scheduling in Non-Orthogonal Multiple Access (NOMA) based cloud-RAN systems[C]. 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York City, USA, 2017: 418–422. doi: [10.1109/UEMCON.2017.8249102](#).

唐 伦: 男, 1973年生, 教授, 主要研究方向为新一代无线通信网络、异构蜂窝网络、软件定义无线网络等。

施颖洁: 女, 1993年生, 硕士生, 研究方向为网络虚拟资源分配。

杨希希: 女, 1992年生, 硕士生, 研究方向为网络虚拟化。

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、下一代移动通信网络、异构蜂窝网络等。