



专题：AI赋能通信网络

一种面向资源高效利用的无蜂窝 RAN 分层协同资源分配算法

王博业¹, 夏玮玮¹, 缪巍巍², 张明轩², 潘裕庆³, 燕锋¹, 沈连丰¹

(1. 东南大学移动通信国家重点实验室, 江苏 南京 210096;

2. 国网江苏省电力有限公司信息通信分公司, 江苏 南京 210024;

3. 国网苏州供电公司, 江苏 苏州 215031)

摘要: 为实现对6G无蜂窝无线电接入网 (radio access network, RAN) 切片资源的高效管理, 提出了一种基于强化学习分层协同优化的资源分配算法。该算法采用多时间尺度的分层协同架构, 在上层大时间尺度, 以各切片的资源匹配度作为优化反馈指标, 利用双深度Q网络算法动态调整切片的资源配置。在下层小时间尺度, 构建了基于近端策略优化的用户接入决策模型, 在满足用户服务质量的前提下, 通过用户协作簇选择和资源分配机制最小化切片资源消耗量。仿真结果表明, 所提算法通过面向用户的动态接入决策显著减少资源消耗量, 并通过切片资源的周期性重配置提升资源匹配度, 从而实现了系统资源的高效利用。

关键词: 无蜂窝; 网络切片; 深度强化学习; 资源分配; 多时间尺度

中图分类号: TN92

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025243

A layered collaborative resource allocation algorithm for cell-free RAN aimed at efficient resource utilization

WANG Boye¹, XIA Weiwei¹, MIAO Weiwei², ZHANG Mingxuan²,

PAN Yuqing³, YAN Feng¹, SHEN Lianfeng¹

1. National Mobile Communication Research Laboratory of Southeast University, Nanjing 210096, China

2. Information & Telecommunication Branch, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China

3. State Grid Suzhou Power Supply Company, Suzhou 215031, China

Abstract: To achieve efficient management of 6G cell-free radio access network (RAN) slice resources, a layered collaborative optimization resource allocation algorithm based on reinforcement learning (RL) was proposed. This algorithm adopted a multi-timescale hierarchical collaborative architecture. At the upper layer with a large timescale, the

收稿日期: 2025-08-05; 修回日期: 2025-10-21

通信作者: 夏玮玮, wxia@seu.edu.cn

基金项目: 国家电网总部科技项目“超五代无线接入理论研究”(No.5700-202499320A-1-3-ZB)

Foundation Item: Science and Technology Project of State Grid Corporation of China “Research on Beyond-5G Wireless Access Theory” (No.5700-202499320A-1-3-ZB)



resource matching degree of each slice was used as the optimization feedback metric, and the dual-depth Q-network (DDQN) algorithm was employed to dynamically adjust the resource configuration of slices. At the lower-level small time scale, a user access decision-making mechanism based on a proximity policy optimization (PPO) algorithm was established. Under the premise of meeting user quality of service (QoS) requirements, the algorithm minimized slice resource consumption through user collaboration cluster selection and resource allocation mechanisms. Simulation results demonstrate that the proposed algorithm significantly reduces resource consumption through user-centric dynamic access decisions and enhances resource matching through periodic reconfiguration of slice resources, thereby achieving efficient utilization of system resources.

Key words: cell-free, network slicing, deep reinforcement learning, resource allocation, multi-timescale

0 引言

6G 作为新一代移动通信系统，不仅具备超高速率、超高可靠性和超低时延等显著特征，还强调对人工智能、智能网络和物联网等多样化业务需求的灵活响应，是构建未来智能社会的核心技术之一^[1]。随着移动数据流量的爆发式增长，传统蜂窝架构在应对海量终端接入与动态业务负载方面存在边缘服务能力不足、跨小区干扰严重以及覆盖范围受限等问题。为此，无蜂窝网络架构通过大量部署分布式接入点（access point, AP），在同一时频资源中为大量用户提供服务^[2]，实现更广泛的系统覆盖^[3]和更高的网络吞吐量，是 6G 主要聚焦的技术方案^[4-5]。

网络切片技术通过将物理网络资源按需划分为多个逻辑切片，为不同类型的业务场景提供定制化的网络服务，从而实现资源的差异化配置与服务质量（quality of service, QoS）的保障^[6-7]，非常契合 6G 多类型业务对网络性能的差异化需求。无线电接入网（radio access network, RAN）切片作为网络切片的重要组成部分，其切片资源的动态调整需综合考虑信道状态的随机波动、无线资源的有限性以及网络架构的异构性，是当前 6G 无线通信领域的核心研究难点之一^[8-9]。然而，传统的启发式算法与组合优化分析等数学模型^[10-11]在应对这类非线性、非凸性、高维度的问题时，无法完整获取实时的环境信息，难以完成

实时且高效的资源优化，严重限制了系统 QoS 的进一步提升。鉴于此，深度强化学习（deep reinforcement learning, DRL）作为一种能够从环境中不断交互学习并动态调整策略的人工智能方法^[10,12]，正逐步成为解决 6G 无线资源分配问题的重要工具。

目前，已有不少基于 DRL 的方法用于解决 RAN 切片资源分配问题。为提升算法对动态业务的适应性，文献[13]引入双层深度 Q 网络（double deep Q-network, DDQN），构建了一种面向 5G 网络的业务需求感知资源分配算法。该算法能够根据切片的实时业务请求量动态调整其资源配置，突破了资源固定的限制。然而，该算法以用户请求变化为重配置触发条件，导致资源调整过于频繁，进而增加了计算开销。为缓解频繁重配置的问题，部分研究引入资源预留机制。文献[14]提出 5G RAN 切片之间共享资源的方法，从全局资源池中划分部分可调节资源，用于切片的资源调整，以应对终端用户设备服务需求的变化。文献[15]同样采用资源预留策略，根据切片的服务满意度和资源利用率反馈，利用 DQN 算法动态调整切片资源。然而，当场景中同时存在多个切片时，资源预留机制容易造成切片间的资源竞争，从而影响切片的性能隔离。

为保障用户的长期 QoS 需求并降低算法的计算开销，研究者提出了分层优化机制，即采用多时间尺度协同优化的方法对网络切片进行资源调

整。文献[16]以响应5G网络切片业务流变化为出发点,提出了基于一个深度学习(deep learning, DL)与DRL的协作框架。该框架将资源分配划分为大时间尺度和小时间尺度两阶段,其中,在大时间尺度上,根据长短期记忆(long short-term memory, LSTM)时序预测网络预测的资源需求量执行面向切片的资源配置;在小时间尺度上,采用DRL下的异步优势动作评价(asynchronous advantage actor-critic, A3C)执行面向用户的资源分配。然而,该方法对资源需求的预测精度依赖于历史样本的丰富性和正确性,难以在实时动态环境中保持预测的高精度。文献[17]同样采用多时间尺度协同的框架,以最大化网络长期效用和频谱利用率为目标,在大小时间尺度上分别采取DDQN和深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法执行资源的配置与分配,但所有切片资源的分配均由同一智能体决策,难以保障不同类型用户的QoS需求。文献[18]基于6G无蜂窝架构提出了分层优化的算法框架,采用多智能体深度确定性策略梯度(multi-agents deep deterministic policy gradient, MADDPG)接入以用户为中心的协作簇,有效降低了用户间干扰并提升了信道增益。文献[19]同样基于无蜂窝架构,在小时间尺度内设置共享智能体,为资源不足的网络切片执行资源调整,同时保障了用户在面临突发业务时的服务质量。上述方法主要以效用最大化为优化目标,虽然实现了切片资源的重配置,却容易出现频谱资源过度分配以满足QoS要求的问题,从而导致资源匹配度偏低,增加资源使用成本。在实际网络中,尤其是在频谱资源高度紧张的场景下,提升资源利用效率、避免资源浪费是确保系统可持续运行的关键。此外,现有算法未能充分考虑无蜂窝架构中用户接入协作簇的动态特性,缺乏协作接入与资源分配策略的联合优化机制,也未能系统性地应对增强移动宽带(enhanced mobile broadband,

eMBB)与超高可靠低时延通信(ultra-reliable and low latency communications, URLLC)等多业务切片共存场景中的资源冲突与分配挑战。

因此,本文提出了一种面向资源高效利用的无蜂窝RAN分层协同资源分配算法,具体贡献如下。

(1)以最大化无蜂窝RAN系统的资源匹配度为目标,在满足eMBB与URLLC两类业务传输速率与平均时延等QoS指标约束的前提下,构建分层协同的资源分配框架。

(2)在大时间尺度上,将面向切片的资源配置问题构建为以切片的资源匹配度作为优化反馈指标的DRL分配问题,利用DDQN算法动态调整切片的资源配置;在小时间尺度上,将面向用户的资源分配问题建模为满足QoS约束的最小化资源消耗的用户协作簇选择和资源分配问题,采用近端策略优化(proximal policy optimization, PPO)算法动态为用户接入最优协作簇。

(3)仿真实验评估本文所提算法的性能,并与已有算法进行比较。本文所提算法在用户接入策略优化和切片资源重配置的协同作用下,能够有效降低资源消耗并提升系统资源匹配度,实现资源的高效利用。

1 系统模型

本文考虑的6G无蜂窝RAN架构如图1所示。从上往下依次为中央处理单元(central processing unit, CPU)、边缘分布式单元(edge distributed unit, EDU)、以用户为中心的分布式单元(user-centric distributed unit, UCDCU)以及AP。在该架构中,CPU作为控制核心,在负责设备管理的同时执行物理资源块(physical resource block, PRB)分配的决策任务。相较于传统蜂窝架构,每个AP均配备多根天线,并以均匀分布的方式部署在系统覆盖区域,从而缩短用户与AP之间的平均距离。各AP通过回程链路与CPU



连接：CPU 向各 AP 传输用户数据和功率控制系数，AP 则将上行接收到的用户数据与信道状态信息反馈给 CPU。假设 RAN 切片中存在 N 个用户、 J 个 AP、 I 个切片，分别用 $\mathcal{N} = \{1, 2, \dots, n, \dots, N\}$ 、 $\mathcal{J} = \{1, 2, \dots, j, \dots, J\}$ 和 $\mathcal{I} = \{1, 2, \dots, i, \dots, I\}$ 表示用户集合、AP 集合和切片集合，并用 U_i 表示切片 i 中的所有用户集合。在无蜂窝网络的实际部署中，若所有 AP 同时为所有用户提供服务，虽然能够提升系统的覆盖范围与增加接收信号的强度^[4,20]，但同时也引发了资源浪费、用户间干扰增强等问题，还会引发导频污染。为缓解这一问题，Emil 等^[21]提出了协作簇的概念，即将所有 AP 划分为多个协作簇，使其在不同时刻为不同用户群体提供服务，从而在保证用户接收到足够信号强度的同时，显著降低冗余干扰与能源消耗，进一步提升系统的整体性能与能效。已有研究使用群智能优化算法、图论等方法^[20,22]探讨协作簇的划分，由于本文重点关注无蜂窝 RAN 切片内用户接入协作簇，因此假设切片中的协作簇已划分，用集合 $M_i = \{1, 2, \dots, m_i, \dots, M_i\}$ 表示切片 i 中的协作簇，用 J_{m_i} 表明协作簇 m_i 中的 AP 集合。同一协作簇内的用户将共享相同的时频资源。本

文假设由正交频分复用技术将总频谱资源划分为 F 个带宽相同的 PRB，其中每个 PRB 的带宽为 $B = B^{\max}/F$ 。

无蜂窝架构的上下行链路通常采用时分双工 (time-division duplex, TDD) 模式进行通信，该模式下将相干间隔 τ_c 分为 3 个阶段：上行导频训练阶段、上行链路数据传输阶段和下行链路数据传输阶段。考虑到上下行传输的目标与需求通常存在差异^[23]，且当前研究主要聚焦于 AP 向用户终端提供服务的过程，因此本文在不失一般性的前提下，仅关注下行链路数据传输阶段的建模与优化^[18]。在上行导频训练阶段，用户会向所有 AP 发送预设的导频序列，AP 接收后可基于导频信号估计所有用户的信道状态信息 (channel state information, CSI)。本文用 $g_{i,m,n,t}^j$ 表示协作簇 m_i 中用户 n 与 AP j 之间在传输时间间隔 (transmission time interval, TTI) t 的信道增益，其数学式可表示为：

$$g_{i,m,n,t}^j = \sqrt{\beta_{i,m,n,t}^j} h_{i,m,n,t}^j \quad (1)$$

其中， $\beta_{i,m,n,t}^j$ 表示 AP j 与用户 n 之间的大尺度衰落，涵盖路径损耗与阴影效应的影响； $h_{i,m,n,t}^j$ 则表示小尺度衰落特性。

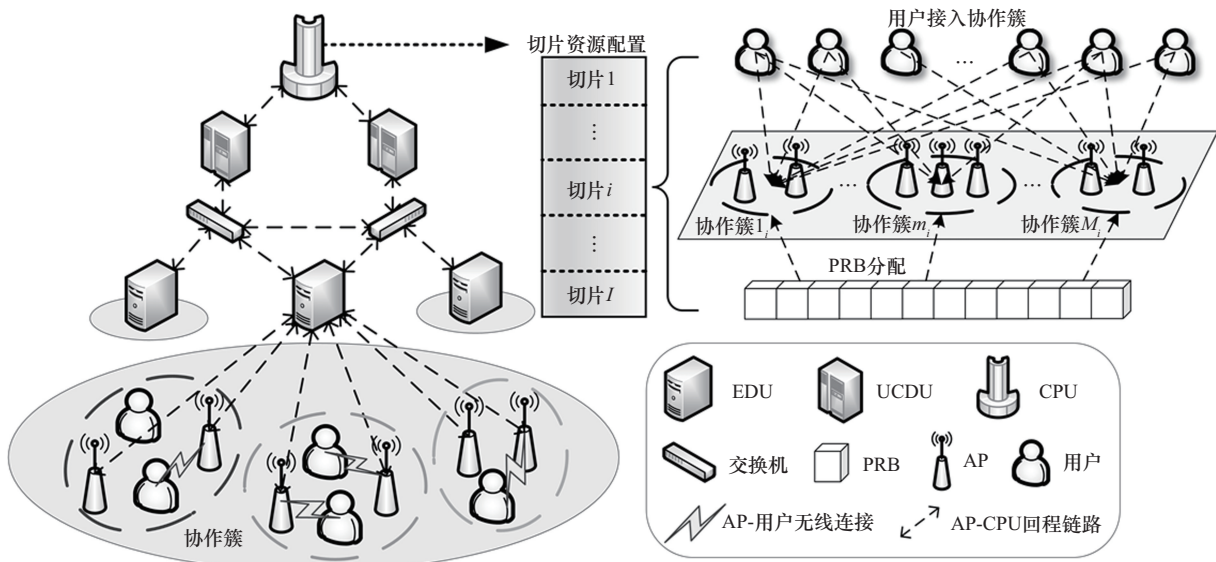


图1 本文考虑的6G无蜂窝RAN架构

假设正交导频序列长度为 τ ，定义切片 i 的协作簇 m_i 中用户 n 向所有 AP 发送的导频序列为 $(\tau)^{1/2}\zeta_{i,m,n} \in \mathbb{C}^{\tau \times 1}$ ，其中 $\zeta_{i,m,n}$ 是导频序列向量，满足 $\|\zeta_{i,m,n}\|^2=1$ 。假设同一协作簇内不同用户之间的导频序列具有正交性，即 $\zeta_{i,m,n}^H \zeta_{i,m,n'}=0, \forall n \neq n'$ ，可得到 t 时刻 AP j 在估计用户 n 信道时接收到的有效信号^[24]，由式（2）表示。

$$y_{i,m,n,t}^j = \sqrt{\tau p^p} g_{i,m,n,t}^j + \mathbf{d}^j \zeta_{i,m,n} \quad (2)$$

其中， p^p 是导频序列归一化后的发射功率， $\mathbf{d}^j \in \mathbb{C}^{J_i \times \tau}$ 是服从复高斯分布的加性噪声。

采用最小均方误差方法对式（2）进行信道估计，可估计出用户 n 与 AP j 之间的信道增益为：

$$\hat{g}_{i,m,n,t}^j = \frac{E[g_{i,m,n,t}^j y_{i,m,n,t}^{j*}]}{E[|y_{i,m,n,t}^j|^2]} y_{i,m,n,t}^j = \frac{\sqrt{\tau p^p} \beta_{i,m,n,t}^j}{1 + \tau p^p \beta_{i,m,n,t}^j} y_{i,m,n,t}^j \quad (3)$$

由于 TDD 模式下信道具有上下行互易性，因此上行信道估计得到的 CSI 可用于下行链路中的预编码^[25]设计，得到预编码矢量 $\mathbf{v}_{i,m,n,t}^j$ 。同时，根据 CPU 设定的功率系数 $p_{i,m,n,t}^j$ 和用户 n 在 TTI t 的传输数据 $\zeta_{i,m,n,t}$ ，得到 AP j 向用户 n 的发送信号 $x_{i,m,n,t}^j$ ，如式（4）所示。

$$x_{i,m,n,t}^j = \sqrt{p_{i,m,n,t}^j} \mathbf{v}_{i,m,n,t}^j \zeta_{i,m,n,t} \quad (4)$$

$$\mathbf{v}_{i,m,n,t}^j = \frac{\hat{\mathbf{g}}_{i,m,n,t}^j}{\sqrt{E[\|\hat{\mathbf{g}}_{i,m,n,t}^j\|_2^2]}} \quad (5)$$

由于无蜂窝架构下用户可由多 AP 协同提供

$$R_{i,n,t} = \begin{cases} \frac{\tau_c - \tau}{\tau_c} B_{i,n,t} \text{lb}(1 + \Gamma_{i,m,n,t}), & \text{eMBB切片} \\ \frac{\tau_c - \tau}{\tau_c} B_{i,n,t} \left[\text{lb}(1 + \Gamma_{i,m,n,t}) - \sqrt{\frac{V_{i,m,n,t}}{\rho_i}} \frac{Q^{-1}(\varpi)}{\ln 2} \right], & \text{URLLC切片} \end{cases} \quad (11)$$

其中， $B_{i,n,t}$ 表示用户带宽， $Q^{-1}(\cdot)$ 是逆高斯 Q 函数， ϖ 为有效解码错误概率， ρ_i 是 URLLC 切片传

服务的，因此用户 n 与协作簇 m_i 中所有相连的 AP 接收到的信号可表示为：

$$y_{i,m,n,t} = \sum_{m_i \in \mathbf{M}_i} \sum_{j \in \mathbf{J}_{m_i}} w_{m_i,n,t} g_{i,m,n,t}^{j,H} x_{i,m,n,t}^j + o_{i,m,n,t} = \text{DS}_{i,m,n,t} \zeta_{i,m,n,t} + \text{NI}_{i,m,n,t} \zeta_{i,m,n,t} + \sum_{n' \in m_i, n' \neq n} \text{UI}_{i,m,n',t} \zeta_{i,m,n',t} + o_{i,m,n,t} \quad (6)$$

其中， $w_{m_i,n,t}$ 是二进制变量， $w_{m_i,n,t}=1$ 表明切片 i 内的用户 n 在 t 时刻接入协作簇 m_i ， $w_{m_i,n,t}=0$ 则说明不接入协作簇 m_i ； $\text{DS}_{i,m,n,t}$ 表示期望信号； $\text{NI}_{i,m,n,t}$ 表示预编码增益的不确定性； $\text{UI}_{i,m,n',t}$ 表示同一协作簇中其他用户的干扰，具体表示为：

$$\text{DS}_{i,m,n,t} = E \left[\sum_{m_i \in \mathbf{M}_i} \sum_{j \in \mathbf{J}_{m_i}} w_{m_i,n,t} \sqrt{p_{i,m,n,t}^j} g_{i,m,n,t}^{j,H} \mathbf{v}_{i,m,n,t}^j \right] \quad (7)$$

$$\text{NI}_{i,m,n,t} = \sum_{m_i \in \mathbf{M}_i} \sum_{j \in \mathbf{J}_{m_i}} w_{m_i,n,t} \sqrt{p_{i,m,n,t}^j} g_{i,m,n,t}^{j,H} \mathbf{v}_{i,m,n,t}^j - \text{DS}_{i,m,n,t} \quad (8)$$

$$\text{UI}_{i,m,n',t} = \sum_{m_i \in \mathbf{M}_i} w_{m_i,n',t} \sqrt{p_{i,m,n',t}^j} g_{i,m,n',t}^{j,H} \mathbf{v}_{i,m,n',t}^j \quad (9)$$

由此，用户的信干噪比（signal to interference plus noise ratio, SINR）可表示为式（10），并根据香农定律和有限长块码理论^[26]可以分别推导出 eMBB 切片与 URLLC 切片中用户的下行传输速率，如式（11）所示。

$$\Gamma_{i,m,n,t} = \frac{|\text{DS}_{i,m,n,t}|^2}{E[|\text{NI}_{i,m,n,t}|^2] + \sum_{n' \in m_i, n' \neq n} E[|\text{UI}_{i,m,n',t}|^2] + 1} \quad (10)$$

输的平均数据包长度， $V_{i,m,n,t}$ 表示信道弥散，具体表示为：



$$V_{i,m,n,t} = 1 - (1 + \Gamma_{i,m,n,t})^{-2} \quad (12)$$

在数据传输过程，假设每个用户都有一个虚拟数据队列用于缓冲传入的数据包，并遵循先到先服务的策略传输。定义变量 $A_{i,n,t}$ 表示切片 i 中用户 n 的数据包在 TTI t 的到达过程，假设该到达过程在每个 TTI 独立同分布，服从泊松分布，到达率满足 $\lambda_{i,n,t} = E[A_{i,n,t}]$ 。同时，用变量 Ω_i 表示切片 i 中用户的平均包长， Δt 表示每个 TTI 的时长， $Q_{i,n,t}$ 表示切片 i 中用户 n 在 TTI t 的队列长度，队列更新过程可描述为^[27-28]：

$$Q_{i,n,t+1} = \max \{Q_{i,n,t} - R_{i,n,t} \Delta t / \Omega_i, 0\} + A_{i,n,t} \quad (13)$$

本文考虑的数据包时延由传输时延 $\bar{D}_{i,n,k}^{\text{trans}}$ 和排队时延 $\bar{D}_{i,n,k}^q$ 两部分组成，其中传输时延取决于式 (11) 中的用户传输速率；排队时延取决于数据包的到达率和传输速率。根据文献[28]中的速率稳定定理，当所有离散队列 $Q_{i,n,t}$ 满足均值速率稳定性时，即 $\lim_{t \rightarrow \infty} \{ |Q_{i,n,t}| \} / t = 0$ ，网络处于稳定状态。基于此，根据 Little 定理^[29]推导出切片 i 中的用户 n 在 PW 时间周期 k 的平均排队时延，表示为：

$$\bar{D}_{i,n,k}^q = \frac{\bar{Q}_{i,n,k}}{\bar{\lambda}_{i,n,k}} \quad (14)$$

其中， $\bar{Q}_{i,n,k} = \frac{1}{\Delta T} \sum_{t=(k-1)\Delta T+1}^{k\Delta T} E[Q_{i,n,t}]$ 是切片 i 中用户 n 在时间周期 k 内的平均队列长度， $\bar{\lambda}_{i,n,k}$ 是用户 n 在时间周期 k 的数据包平均到达率。

因此，可以推导平均时延的表达式为：

$$\bar{D}_{i,n,k} = \bar{D}_{i,n,k}^{\text{trans}} + \bar{D}_{i,n,k}^q = \frac{1}{\Delta T} \sum_{t=(k-1)\Delta T+1}^{k\Delta T} \left(\frac{\Omega_i}{R_{i,n,t}} + \frac{E[Q_{i,n,t}]}{\lambda_{i,n,t}} \right) \quad (15)$$

2 问题定义与建模

为实现切片资源长期效率与用户短期性能的协同保障，本文设计了一种多时间尺度的分层资源优化框架。该框架在时间域上引入明确的划分标准，以支撑面向切片与面向用户两层资源管理

机制的有序协同。具体而言，将大时间尺度的基本时间单位定义为决策时间间隔 (decision time interval, DTI)，表示切片资源周期性重配置的周期；将小时间尺度的基本时间单位定义为 TTI，假设每个 DTI 包含多个连续的 TTI，即 $\text{DTI} = \Delta T (\text{TTI})$ 。在每个 DTI 的起始阶段，系统基于上一个周期内各切片的资源使用与需求反馈，执行一次面向切片的资源重配置决策，确定后续 TTI 内各切片的 PRB 配置数。在整个 DTI 周期内，该配置数将保持不变，以保障资源分配的稳定性与切片间的隔离性。在每一个 TTI 内，切片根据当前用户 CSI 与可用 PRB 数目，执行面向用户的协作簇选择和资源分配策略。此过程中，系统为用户动态选择最优协作簇，并以最优协作簇下的 SINR 计算满足 QoS 要求所需的最小 PRB 数量，实现小时间尺度内的资源高效利用。

2.1 大时间尺度的优化问题

在大时间尺度，系统周期性地评估各网络切片的资源使用情况，以资源匹配度 η_i 作为优化反馈指标，对各切片的 PRB 资源进行动态重配置，以确保长期资源的分配更符合业务负载的变化，其数学式表示如下：

$$\eta_i(k) = \frac{\bar{C}_i(k)}{A_i(k)} \quad (16)$$

其中， $\bar{C}_i(k)$ 是切片 i 在第 k 个 DTI 用于满足其内部所有用户的传输速率和时延指标的平均资源需求值，由小时间尺度反馈得到； $A_i(k)$ 则表示切片 i 在第 k 个 DTI 的资源配置值。

资源匹配度通常存在 3 种情况：(1) $\eta_i(k) > 1$ ，即切片平均资源需求值大于实际配置值，说明该周期切片资源配置不足，难以保障用户的 QoS 指标；(2) $\eta_i(k) = 1$ ，表明切片的实际配置值恰好满足切片资源需求，是最理想情况；(3) $\eta_i(k) < 1$ ，切片配置的资源足够支撑用户需求，但为实现资源高效利用的目标，需要重配置切片资源，使得 η_i 越接近 1 越好。因此，大时间尺度的优化

问题可表示如下：

$$\arg \min_{\eta_i(k)} \sum_{k=0}^{\infty} \sum_{i \in I} |1 - \eta_i(k)| \quad (17)$$

$$\text{s.t. } R_{i,n,t} \geq R_i^{\min}, \forall i \in I, n \in U_i, t \in [(k-1)\Delta T + 1, k\Delta T], k=1, 2, \dots \quad (18)$$

$$\bar{D}_{i,n,k} \leq D_i^{\max}, \forall i \in I, n \in U_i, k=1, 2, \dots \quad (19)$$

$$\sum_{i \in I} A_i(k) \leq F, k=1, 2, \dots \quad (20)$$

其中，式 (18) 表明需要保障用户的传输速率不可低于用于满足该切片业务需求的最小速率 R_i^{\min} ，式 (19) 表明需要保障用户在任意 DTI 的平均时延 $\bar{D}_{i,n,k}$ 不可超过阈值 D_i^{\max} ，式 (20) 表

明任意 DTI 的切片配置资源总和不可超过总资源量 F 。

2.2 小时间尺度的优化问题

在小时间尺度下，为满足多业务切片在传输速率和平均时延方面的 QoS 要求，以最小化实际消耗的 PRB 数量为优化目标，系统根据当前用户的 CSI 动态为其选择最优协作簇，使用户的 SINR 最大，从而减少 PRB 的消耗数量。同时，根据香农容量公式和切片的最小速率 R_i^{\min} 与时延阈值 D_i^{\max} 要求反推各用户的最小 PRB 需求量 $h_{m_i,n}(t)$ ，其数学表达式如下：

$$h_{m_i,n}(t) = \begin{cases} \frac{\tau_c}{\tau_c - \tau} \frac{R_{i,t}^{\text{new}}}{B \text{lb}(1 + \Gamma_{i,m,n,t})}, & \text{eMBB切片} \\ \frac{\tau_c}{\tau_c - \tau} \frac{R_{i,t}^{\text{new}}}{B \left[\text{lb}(1 + \Gamma_{i,m,n,t}) - \sqrt{\frac{V_{i,m,n,t}}{\rho_i} \frac{Q^{-1}(\varpi)}{\ln 2}} \right]}, & \text{URLLC切片} \end{cases} \quad (21)$$

$$R_{i,t}^{\text{new}} = \max \left(R_i^{\min}, \frac{\Omega_i}{D_i^{\max} - \bar{D}_{i,n,t-1}^q} \right) \quad (22)$$

结合式 (15)，式 (22) 的 $R_{i,t}^{\text{new}}$ 表示在 TTI t 时刻，为满足业务速率需求，同时避免时延超过上限的所需 PRB 最小值。因此，小时间尺度的优化问题可表示如下：

$$\min \sum_{m_i \in M_i} r_{m_i}(t) \quad (23)$$

$$\text{s.t. } r_{m_i,n}(t) \geq \lceil h_{m_i,n}(t) \rceil, \forall i \in I, n \in U_i \quad (24)$$

$$\sum_{m_i \in M_i} r_{m_i}(t) \leq A_i(k), \forall i \in I, t \in [(k-1)\Delta T + 1, k\Delta T], k=1, 2, \dots \quad (25)$$

$$\sum_{m_i \in M_i} w_{m_i,n}(t) = 1, \forall i \in I, n \in U_i \quad (26)$$

其中，式 (24) 表明为保障用户的 QoS，用户的分配值 $r_{m_i,n}(t)$ 应大于或等于最小需求值 $\lceil h_{m_i,n}(t) \rceil$ ，由于同一协作簇内的用户将共享时频资源，因此切片 i 中协作簇 m_i 在时刻 t 的 PRB 分配值 $r_{m_i}(t)$ 需要满足 $r_{m_i}(t) = \max(r_{m_i,1}(t), r_{m_i,2}(t), \dots, r_{m_i,n}(t), \dots, r_{m_i,|U_i|}(t))$ ，

$n \in U_i$ ；式 (25) 表明为各协作簇分配的 PRB 总和不可超过当前 DTI 切片配置的 PRB 数 $A_i(k)$ ；式 (26) 表明各用户在同一时刻只能接入一个协作簇。

3 无蜂窝 RAN 分层协同资源分配算法

由第 2.1 节和第 2.2 节的优化问题可知，6G 无蜂窝 RAN 场景下的切片资源分配属于典型的 NP 难问题，且该场景的无线信道状态和业务负载呈现时变特性，进一步增加了资源调整的挑战，传统静态优化方法在应对此类动态环境时表现出适应性不足。为此，引入 DRL 作为求解工具，以构建具备感知与决策能力的智能资源分配机制，将 RAN 切片资源分配问题建模成马尔可夫决策过程 (Markov decision process, MDP)，通过智能体与环境之间的交互获取对象状态，进行动作选择，并通过累积奖励函数值更新策略，从而实现长期性能最优。面向资源高效利用的无蜂窝 RAN 分层协同资源分配算法流程如图 2 所

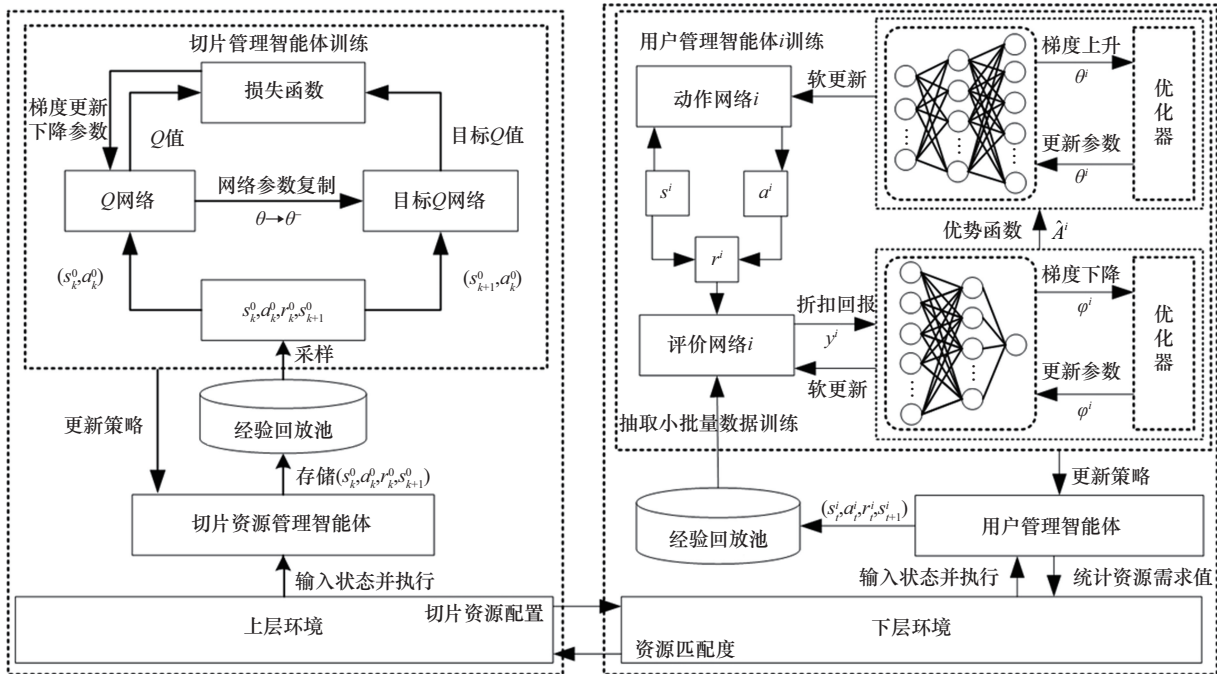


图2 面向资源高效利用的无蜂窝RAN分层协同资源分配算法流程

示，其中，上层大时间尺度通过切片资源管理智能体执行DDQN算法，以优化切片的资源分配；下层小时间尺度通过用户管理智能体执行PPO算法，实现用户接入策略的优化。

3.1 大时间尺度算法设计

3.1.1 MDP建模

大时间尺度的切片资源配置任务由部署在CPU中的切片资源管理智能体0负责，其状态空间 \mathcal{S} 、动作空间 \mathcal{A} 和奖励函数 \mathcal{R} 的定义如下。

(1) \mathcal{S} : 大时间尺度的 \mathcal{S} 定义为上一个DTI各切片的资源匹配度 $\eta_i(k)$ 和切片的平均资源需求值 $\bar{C}_i(k)$ ，表示为 $s_k^0 = \{\eta_k, \bar{C}_k\}$ 。其中， $\eta_k = \{\eta_i(k) | \forall i \in \mathcal{I}\}$, $\bar{C}_k = \{\bar{C}_i(k) | \forall i \in \mathcal{I}\}$ 。

(2) \mathcal{A} : 大时间尺度完成对各切片的PRB配置，故 \mathcal{A} 由离散的PRB分配值构成，表示为 $a_k^0 = \{A_1(k), \dots, A_i(k), \dots, A_I(k)\}$ 。该动作空间需要满足式(20)，即满足分配给切片的PRB总数不可超过总资源数 F 。

(3) \mathcal{R} : 考虑大时间尺度的目标是切片配置

量匹配实际需求，且切片配置策略的更新由奖励函数直接驱动，因此大时间尺度的 \mathcal{R} 将围绕资源匹配度 $\eta_i(k)$ 设计，可表示如下：

$$r_k^0 = \sum_{i \in \mathcal{I}} 1 - (1 - \eta_i(k))^2 \quad (27)$$

3.1.2 基于DDQN的切片资源配置

为实现大时间尺度下的切片资源自适应配置，引入DDQN算法，并结合经验回放机制提升策略学习的稳定性和泛化能力，其具体执行步骤如下。

步骤1~步骤3 初始化系统参数并设定训练回合。

步骤4~步骤8 为数据收集阶段。智能体执行动作后，获得交互数据，并存放至经验回放池 \mathcal{D} 中。

步骤9~步骤10 为参数更新阶段。智能体随机采样小批量样本，主网络根据式(28)选择动作，目标网络根据式(29)评估 Q 值，最后通过式(30)计算损失函数，通过梯度下降更新网络参数 θ 。

$$a^* = \arg \max_a Q(s_{k+1}^0, a; \theta) \quad (28)$$

$$y_k^0 = r_k^0 + \gamma Q(s_{k+1}^0, a^*; \theta^-) \quad (29)$$

$$L(\theta) = E_{(s_k^0, a_k^0, r_k^0, s_{k+1}^0) \in \mathcal{D}} \left[\left(y_k^0 - Q(s, a; \theta) \right)^2 \right] \quad (30)$$

算法 1 基于 DDQN 的切片资源配置

输入 各切片平均资源需求值 $\bar{C}_i(k)$

输出 各切片资源配置值 $A_i(k)$

- (1) 初始化主网络 $Q(s, a; \theta)$ 、目标网络 $Q'(s, a; \theta^-)$ 、经验回放池 \mathcal{D} ;
- (2) **for** episode= 1 to T_{do}
- (3) 初始化各切片资源配置值 $A_i(0)$, 设定 PRB 总资源上限 F ;
- (4) **for** step $k= 1$ to T_U **do**
- (5) 智能体根据 ϵ -贪心策略选择动作;
- (6) 小时间尺度算法模拟运行, 统计出各切片平均资源需求值 $\bar{C}_i(k)$;
- (7) 根据式 (14) 计算资源匹配度 $\eta_i(k)$, 并根据奖励函数计算奖励 r_k^0 , 环境状态由 $s_k^0 \rightarrow s_{k+1}^0$;
- (8) 存储四元组 $(s_k^0, a_k^0, r_k^0, s_{k+1}^0)$ 到经验回放池 \mathcal{D} ;
- (9) 从 \mathcal{D} 中采样小批量样本, 根据式 (28) ~ (30) 执行梯度下降策略, 更新主网络参数 θ ;
- (10) 每隔一定步数同步更新目标网络参数 $\theta^- \leftarrow \theta$;
- (11) **end for**
- (12) **end for**

3.2 小时间尺度算法设计

3.2.1 MDP 建模

小时间尺度的资源分配由同样部署于 CPU 的用户管理智能体 i 负责, 其状态空间 S 、动作空间 A 和奖励函数 R 的定义如下。

(1) S : 小时间尺度的状态包含各用户的大尺度衰落系数 $\beta_{i,m,n,t}^j$ 与上一时刻的 SINR $\Gamma_{i,m,n,t}$, 表示为 $s_t^i =$

$\{\beta_{i,m,n,t}^j, \Gamma_{i,m,n,t}\}$ 。其中, $\beta_{i,m,n,t}^j = \{\beta_{i,m,n}^j(t) | \forall j \in \mathbf{J}\}$ 。

(2) A : 小时间尺度的动作是为各用户接入协作簇, $a_t^i = \{w_{m_i,n,t} | w_{m_i,n,t} \in \mathbf{W}^i\}$, \mathbf{W}^i 是切片 i 内用户接入协作簇的所有动作组合。

(3) R : 小时间尺度的优化目标是在满足用户 QoS 的前提下, 最小化切片的资源消耗, 且需要满足约束式 (24), 因此奖励函数围绕用户的最小 PRB 需求量 $h_{m_i,n}(t)$ 进行设计。由于同一协作簇内用户共享 PRB, 因此协作簇需求值满足 $h_{m_i}(t) = \max(h_{m_i,1}(t), h_{m_i,2}(t), \dots, h_{m_i,n}(t), \dots, h_{m_i,|U_i|}(t))$, $n \in U_i$ 。

$$r_t^i = - \sum_{m_i \in M_i} h_{m_i}(t) \quad (31)$$

3.2.2 基于 PPO 的用户协作簇接入

本文引入 PPO 算法实现用户协作簇的动态接入策略优化。在每个 TTI 内, 用户管理智能体 i 根据 CSI 决定所有用户接入的协作簇选择, 从而影响 SINR 的大小与所需的最小资源消耗量。PPO 算法采用了策略-评价结构, 通过构建策略网络执行用户接入协作簇的策略 $\pi_{\theta}^i(a_t^i | s_t^i)$ 和评价网络估计当前状态的价值函数 $V(s_t^i)$, 其具体执行过程如下。

步骤 1~步骤 3 初始化系统参数并设定训练回合。

步骤 4~步骤 7 为数据收集阶段。智能体采样策略网络输出的动作空间概率分布, 获取动作 a_t^i , 并计算奖励 r_t^i , 再将四元组 $(s_t^i, a_t^i, r_t^i, s_{t+1}^i)$ 存储至经验回放池 \mathcal{D}^i 。

步骤 8~步骤 13 为参数更新阶段。策略网络通过式 (32) 更新策略网络参数 θ^i , 具体表达式如下:

$$\max_{\theta^i} E_t \left[\min(\xi_t^i(\theta^i) \hat{A}_t^i, \text{clip}(\xi_t^i(\theta^i), 1-\mu, 1+\mu) \hat{A}_t^i) \right] \quad (32)$$

其中, $\xi_t^i(\theta^i) = \pi_{\theta}^i(a_t^i | s_t^i) / \pi_{\theta_{\text{old}}}^i(a_t^i | s_t^i)$ 是当前策略 $\pi_{\theta}^i(a_t^i | s_t^i)$ 与旧策略 $\pi_{\theta_{\text{old}}}^i(a_t^i | s_t^i)$ 之间的比值; \hat{A}_t^i 是广



义优势估计的联合优势函数， $\hat{A}_t^i = \sum_{l=0}^{T-t} (\gamma\lambda)^l \delta_{t+l}^i$ ，而 $\delta_t^i = r_t^i + \gamma V(s_{t+1}^i) - V(s_t^i)$ 是时间差分误差， γ 是折扣因子， λ 是平滑系数； $\text{clip}()$ 是裁剪函数，用于限制更新幅度。 $\text{clip}()$ 的具体表达式如下：

$$\text{clip}(\xi_t^i(\theta^i), 1-\mu, 1+\mu) = \begin{cases} 1+\mu, \xi_t^i(\theta^i) \geq 1+\mu \\ 1-\mu, \xi_t^i(\theta^i) < 1-\mu \end{cases} \quad (33)$$

评价网络则通过最小化损失函数来更新评价网络参数 φ^i ，其计算式如下：

$$\text{loss}_t(\varphi^i) = \mathbb{E} \left[\left(V_\varphi(s_t^i) - y_t^i \right)^2 \right] \quad (34)$$

其中， y_t^i 是累计折扣回报值。

算法2 基于PPO的用户协作簇接入算法

输入 各切片资源配置值 $A_i(k)$

输出 用户接入协作簇策略 π_θ^i 与各切片平均资源需求值 $\bar{C}_i(k)$

(1) 初始化策略网络 π_θ^i 、策略网络参数 θ_{old}^i 、评价网络 V_φ^i 、评价网络参数 φ_{old}^i 、经验回放池 \mathcal{D}^i ；

(2) **for** episode= 1 to T **do**

(3) 智能体观察初始环境状态 s^i ；

(4) **for** step $t = 1$ to T_L **do**

(5) 智能体根据策略 π_θ^i 为各用户接入协作簇，根据式(10)和式(16)计算各用户的SINR和所需PRB数目，并根据奖励函数计算奖励 r_t^i ，环境状态由 $s_t^i \rightarrow s_{t+1}^i$ ；

(6) **end for**

(7) 计算优势函数 \hat{A}_t^i 并将该回合的五元组 $\left\{ \left[s_t^i, a_t^i, r_t^i, s_{t+1}^i, \hat{A}_t^i \right] \right\}_{t=1}^{T_L}$ 存储至经验回放池 \mathcal{D}^i 中；

(8) **for** epoch $z = 1$ to Z **do**

(9) 从经验回放池 \mathcal{D}^i 中抽取小批量数据；

(10) 根据式(32)计算损失函数，对策略网络参数 θ^i 进行梯度上升更新；

(11) 根据式(34)计算损失函数，对评价网络参数 φ^i 进行梯度下降更新；

(12) **end for**

(13) 智能体更新旧策略网络参数 $\theta_{\text{old}}^i \leftarrow \theta^i$ ，评价网络参数 $\varphi_{\text{old}}^i \leftarrow \varphi^i$ ；

(14) **end for**

3.3 时间复杂度

本文采用离线训练与在线推理相结合的部署方式：模型在GPU环境下利用交互数据完成策略训练，待策略收敛后，将参数下发至系统CPU执行在线推理，并同步收集交互数据，以供GPU端更新策略。由于本研究在RAN切片的上下两层资源分配中均引入了DRL，因此需要分别对两层算法的时间复杂度进行分析。下层资源分配中所采用的PPO算法，由部署在CPU中的用户管理智能体执行，其时间复杂度^[30]可表示为：

$$O \left(T_L |I| \left(\sum_{l=0}^{L_{\text{actor}}} n_l^{\text{actor}} n_{l+1}^{\text{actor}} + \sum_{l=0}^{L_{\text{critic}}} n_l^{\text{critic}} n_{l+1}^{\text{critic}} \right) \right) \quad (35)$$

其中， T_L 是下层策略训练步数， $|I|$ 是用户管理智能体的数量， n_l^{actor} 是策略网络在第 l 层的神经元数量， n_l^{critic} 是评价网络在第 l 层的神经元数量， L_{actor} 和 L_{critic} 是策略和评价网络的隐藏层数量。

上层资源分配中所采用的DDQN算法，由部署在CPU中的切片资源管理智能体执行，其时间复杂度^[18]如表达式(36)所示。

$$O \left(T_U \left(\sum_{l=0}^{L_{\text{DDQN}}} n_l^{\text{DDQN}} n_{l+1}^{\text{DDQN}} \right) \right) \quad (36)$$

其中， T_U 是上层策略训练步数， n_l^{DDQN} 是DDQN在第 l 层的神经元数量， L_{DDQN} 是DDQN的隐藏层数量。

4 仿真分析

为验证本文所提分层资源分配算法在无蜂窝RAN架构下的性能，基于Python3.9的仿真环境，模拟6G典型下行通信场景，开展仿真实验。仿真场景为500 m×500 m的正方形区域，含eMBB

和URLLC两类业务切片。每类业务切片数量为3，每个切片内用户数为6，协作簇数量为3，且各切片间独立并行，无交互。大时间尺度中，DDQN的神经网络包含2层全连接层，神经元个数均设置为128，每个训练回合步长为 $T_U=10$ ，其学习率为 5×10^{-3} ，折扣因子为0.98， ε 的衰减率为0.99；小时间尺度中，PPO的策略网络和评价网络均具有2层全连接层，神经元个数为128，每个训练回合步长为 $T_L=50$ ，采用Adam优化器，其策略网络学习率为 3×10^{-4} ，评价网络学习率为 10^{-3} ，折扣因子为0.99，clip系数为0.2。相关仿真参数的设置参考了文献[17-19]，具体见表1。

表1 相关仿真参数设置

参数	值
传输时间间隔 (TTI) /ms	1
决策时间间隔 (DTI) /ms	50
AP数量(J)/个	6
每个AP的天线数量/根	32
PRB数量/块	20
每块PRB的带宽/kHz	180
噪声功率谱密度/(dBm·Hz ⁻¹)	-174
eMBB、URLLC平均数据包长(Q)/B	800, 200
eMBB、URLLC数据包到达率 (packet·ms ⁻¹)	0.1, 0.02
eMBB、URLLC平均时延上限(D_i^{\max})/ms	4, 2
eMBB、URLLC传输速率阈值(R_i^{\min})/(Mbit·s ⁻¹)	1, 0.2

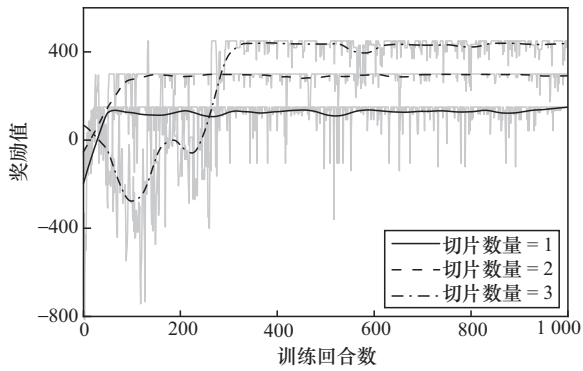
多时间尺度下智能体训练奖励变化如图3所示。其中，图3(a)展示了在大时间尺度下进行切片资源自适应配置的训练过程。在训练初期，由于DDQN算法处于策略探索阶段，系统奖励波动较大，策略尚未稳定。随着训练推进，智能体逐步学习环境特征，策略趋于收敛，奖励曲线也逐渐平稳，表明DDQN算法具备良好的收敛性。随着切片数量增加，动作空间由单一维度扩展为多维组合，策略学习所需的训练轮数相应增多，收敛速度呈下降趋势，但最终均能达到收敛状态。同时，由

于多切片最优配置带来的累积奖励提升，系统奖励值也随切片数量增多而上升。图3(b)展示了小时间尺度下用户协作簇动态接入的训练过程。随着用户数量增加，PPO算法训练回合数从800增至1500，收敛速度明显放缓。其原因主要在于：一方面，用户数上升使动作空间维度增大，有效动作占比下降，探索效率降低；另一方面，维度提升引起梯度方差增大，并受剪切目标的限制，单位迭代的有效更新幅度减小，因此需要更多训练回合迭代才能达到相同性能水平。为此，采用策略热启动方法，将已收敛的PPO策略作为初始化参数，从而显著缩短大规模场景下的前期探索时间。与此同时，系统奖励随用户增加呈下降的趋势。由于奖励函数以所需PRB数的负值构建，用户数量增多将导致同一协作簇内的干扰加剧，整体SINR下降，进而使得每位用户所需的PRB数量增加，系统总体资源消耗上升，最终表现为奖励值的下降。

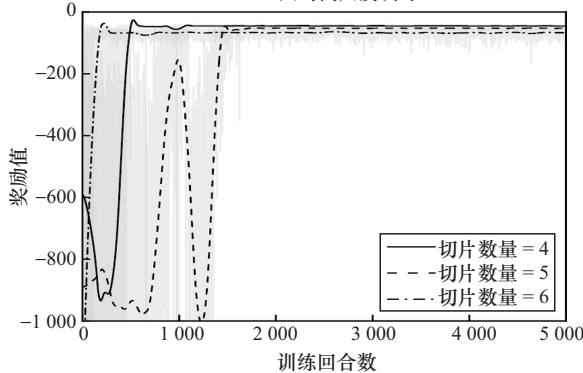
最小速率需求与PRB总消耗量变化如图4所示。该图展示了本文所提算法、MADDPG、固定簇及不分簇算法^[23]在eMBB切片中不同最小速率需求下的PRB消耗对比。不分簇算法不涉及协作簇的概念，所有用户连接至切片内全部AP。这尽管提高了连接数密度，但也加剧了用户间的干扰，且所分配的PRB必须满足信道增益最差用户的需求，容易导致系统资源消耗激增。固定簇算法通过预设接入簇在一定程度上降低了部分干扰，但缺乏对信道动态变化的适应能力，资源利用效率仍有限。相比之下，本文所提算法根据实时信道与业务需求动态调整用户接入簇，实现干扰抑制与资源匹配的联合优化。MADDPG算法同样采用动态接入协作簇架构，但其优化目标为切片效用最大化，因此分配给用户的资源量通常高于实际需求。本文在后续对比中提到的智能资源规划策略(intelligent resource scheduling strategy, iRSS)^[16]本质上属于不分簇的架构，因此不再单独展开讨论。设定eMBB切片在本DTI周期内配置的PRB



总数为50，仿真结果表明，随着最小速率需求的增加，本文所提算法的PRB消耗始终低于MADDPG、固定簇和不分簇算法，分别平均减少23.4%、86.1%和87%。当速率需求超过1.2 Mbit/s时，固定簇和不分簇算法已触及资源上限，而本文所提算法的PRB消耗仅缓慢增长，最大增幅不超28.1%，仍能持续满足服务要求，充分体现了其在资源效率和QoS保障方面的优势。



(a) 大时间尺度训练



(b) 小时间尺度训练

图3 多时间尺度下智能体训练奖励变化

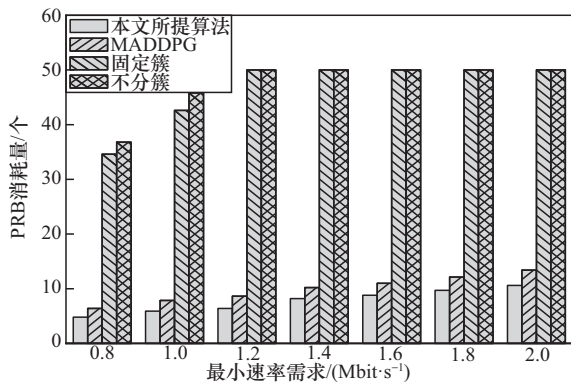


图4 最小速率需求与PRB总消耗量变化

不同算法下URLLC切片的数据包到达率与用户平均时延的关系如图5所示。随着数据包到达率的提高，各算法的用户平均时延均呈增长趋势，这主要源于负载增加导致的队列积压。本文所提算法引入了强时延约束机制，能够在TTI内实时监测平均排队时延与传输时延，并基于表达式(22)动态调整最小速率需求。当负载超出当前切片承载能力时，系统通过大时间尺度下的DDQN算法启动切片资源重配置，从而持续保持资源匹配度与时延稳定性。相比之下，MADDPG和iRSS算法未考虑时延反馈，其资源分配策略难以应对排队增长，因此时延上升快速。固定簇与不分簇算法在PRB资源划分与用户接入上存在固有局限，表现出更高的平均时延。仿真结果显示，在高负载场景下，本文所提算法的时延性能较MADDPG、iRSS、固定簇和不分簇算法分别提升了11.89%、26.49%、36.07%和33.52%，验证了其在动态业务需求变化下保障QoS方面的优越性。

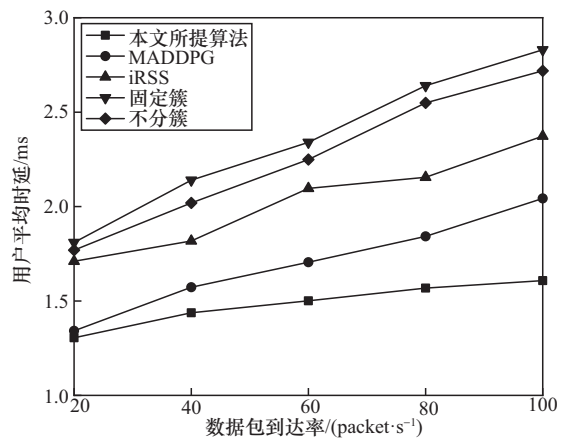


图5 数据包到达率与用户平均时延的关系

在增加eMBB切片内用户数量的情形下，各算法的切片和速率变化趋势如图6所示。为确保实验的公平性，所有对比算法在切片配置的PRB数量上均以本文所提算法配置的资源数量为基准。上文提到，不分簇算法将用户与所有AP的全连接，削弱了SINR，导致切片和速率增长受

限。固定簇算法虽通过静态划分降低了干扰，但未根据用户的时变信道状态动态接入协作簇，SINR、切片和速率的提升也受限制。相比之下，本文所提出的动态接入协作簇算法能够实时接入最优协作簇，有效提升 SINR。因此，切片和速率的增长幅度明显优于其他算法，分别较固定簇和不分簇算法提升了 9.75% 与 28.08%。iRSS 算法的切片和速率增长幅度虽高于固定簇和不分簇算法，但这主要得益于其具备切片资源重配置和共享机制，通过额外消耗更多 PRB 以满足最小速率需求。然而，该方法未能在 SINR 层面实现最优，整体资源利用效率相对较低，而本文所提算法以更少的 PRB 消耗实现更高的速率性能，在切片和速率上较 iRSS 算法提升了 3.04%，而 PRB 用量则减少了 17.3%，展示出更高效的资源优化能力。尽管 MADDPG 同样具备动态接入簇的能力，整体性能接近本文所提算法，但由于策略学习的收敛与鲁棒性限制，仍出现部分协作簇选择与资源分配的次优情形，在相同资源基准下，文本所提算法的切片和速率仍高出了 2.32%。

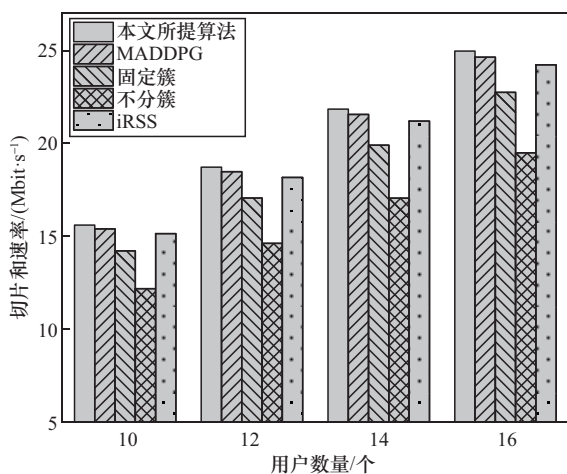


图6 切片和速率的变化趋势

本文所提算法与 MADDPG、iRSS 算法的切片资源匹配度随着周期重配置的变化如图 7 所示。对比算法均采用大、小时间尺度协同优化架构，iRSS 在大时间尺度使用 LSTM 预测切片资源需

求，在小时间尺度通过 A3C 进行用户资源分配。然而，iRSS 未考虑时延约束，资源配置仅基于基础速率需求，导致用户在接入相同协作簇结构下，其计算的资源需求通常偏低，进而引发资源配置不足，存在匹配度 > 1 的问题。此外，iRSS 依赖 LSTM 进行序列建模，对历史数据敏感，一旦样本偏差积累，预测误差将难以修正，进而影响长期匹配效果。MADDPG 算法侧重效用最大化，在大时间尺度上配置的切片资源值通常高于实际需求值，因此其资源匹配度通常 < 1。相比之下，本文所提算法融合时延约束与 DDQN 重配置机制，能够基于每个 DTI 的资源使用情况实现切片资源的自适应调整。尽管初期因策略探索出现了资源匹配度较低的情形，但随着训练推进，资源匹配度逐步逼近 1，显著提升了资源利用效率。为验证鲁棒性，实验逐步增加用户数量，以模拟负载波动。结果显示，本文所提算法仅在负载初增时略有下降，但始终维持在 93% 以上，并能快速恢复至近 100%；而 iRSS 因缺乏时延反馈，资源配置滞后，训练后期阶段的匹配度长期维持在 110%~120%；MADDPG 资源数量存在冗余，使得资源匹配度稳定在 90% 左右。综上所述，本文所提算法在资源匹配度方面展现出更强的精度与稳定性，能够在不同业务负载条件下始终维持较高的资源利用效率，验证了其在 6G 无蜂窝 RAN 场景中的高鲁棒性。

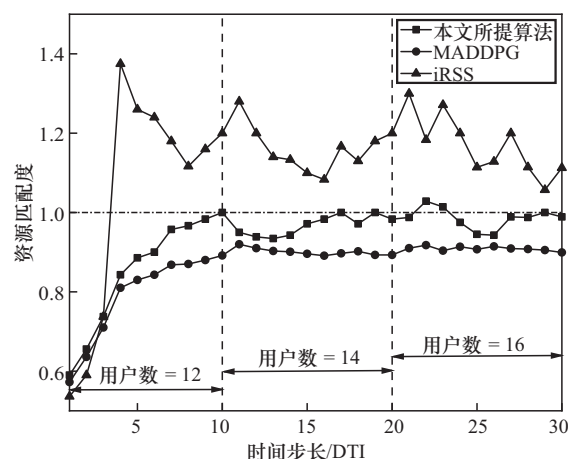


图7 切片资源匹配度随着周期重配置的变化



5 结束语

针对 6G 无蜂窝 RAN 系统中的资源高效利用问题, 本文综合考虑 eMBB 与 URLLC 两类业务需求, 提出了一种分层的资源分配框架。在大时间尺度上, 将切片的资源自适应配置建模为 MDP 过程, 并引入 DDQN 算法动态调整切片的 PRB 分配策略, 优化资源匹配度。在小时间尺度上, 构建以最小化资源消耗为目标的用户协作簇接入问题, 基于 PPO 算法实现用户与协作簇之间的动态关联。仿真结果表明, 本文所提算法能够显著降低资源消耗、提升资源利用率, 且在保障 QoS 约束条件下, 表现出良好的鲁棒性和适应性, 为面向未来 6G 网络的智能资源调整提供了可行的技术路径。此外, 在大规模机器型通信 (massive machine type communications, mMTC) 场景下, 该算法可以采用动态簇接入活跃设备进行分组和资源分配, 从而在高连接密度下保持资源匹配度。需要指出的是, 在 mMTC 场景下, 由设备数量密集、协作簇动态接入带来的动作空间膨胀, 可能导致决策优化中需要更多训练时间并影响强化学习算法的收敛性能。未来的研究可进一步探索动作空间压缩策略, 以增强本文所提框架在大规模多业务环境中的实用性与可扩展性。

参考文献:

- [1] ALSABAH M, NASER M A, MAHMMOD B M, et al. 6G wireless communications networks: a comprehensive survey[J]. IEEE Access, 2021, 9: 148191-148243.
- [2] NGO H Q, ASHIKHMIN A, YANG H, et al. Cell-free massive MIMO versus small cells[J]. IEEE Transactions on Wireless Communications, 2017, 16(3): 1834-1850.
- [3] BUZZI S, D'ANDREA C, ZAPPONE A, et al. User-centric 5G cellular networks: resource allocation and comparison with the cell-free massive MIMO approach[J]. IEEE Transactions on Wireless Communications, 2020, 19(2): 1250-1264.
- [4] NGO H Q, TRAN L N, DUONG T Q, et al. On the total energy efficiency of cell-free massive MIMO[J]. IEEE Transactions on Green Communications and Networking, 2018, 2(1): 25-39.
- [5] AKYILDIZ I F, KAK A, NIE S. 6G and beyond: the future of wireless communications systems[J]. IEEE Access, 2020, 8: 133995-134030.
- [6] WANG Z Y, WEI Y F, YU F R, et al. Utility optimization for resource allocation in edge network slicing using DRL[C]//Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2021: 1-6.
- [7] 孙君, 霁振宇. 基于差异性隔离和复用的网络切片无线资源分配方案[J]. 通信学报, 2025, 46(3): 109-121.
SUN J, AI Z Y. Wireless resource allocation scheme for network slicing based on differentiated isolation and multiplexing[J]. Journal on Communications, 2025, 46(3): 109-121.
- [8] TANG J H, SHIM B, QUEK T Q S. Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(4): 881-895.
- [9] ALWARAFY A, ABDALLAH M, ÇIFTLER B S, et al. The frontiers of deep reinforcement learning for resource management in future wireless HetNets: techniques, challenges, and research directions[J]. IEEE Open Journal of the Communications Society, 2022, 3: 322-365.
- [10] ZAPPONE A, DI RENZO M, DEBBAH M. Wireless networks design in the era of deep learning: model-based, AI-based, or both?[J]. IEEE Transactions on Communications, 2019, 67(10): 7331-7376.
- [11] 元伟敬, 宋清洋, 郭磊. 面向软件定义多模态车联网的双时间尺度 RAN 切片资源分配[J]. 通信学报, 2022, 43(4): 60-70.
QI W J, SONG Q Y, GUO L. Dual time scale resource allocation for RAN slicing in software-defined oriented polymorphic IoV[J]. Journal on Communications, 2022, 43(4): 60-70.
- [12] CHEN M Z, CHALLITA U, SAAD W, et al. Artificial neural networks-based machine learning for wireless networks: a tutorial[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3039-3071.
- [13] HUA Y X, LI R P, ZHAO Z F, et al. GAN-powered deep distributional reinforcement learning for resource management in network slicing[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(2): 334-349.
- [14] ZHANG H, PAN G J, XU S G, et al. A hard and soft hybrid

- slicing framework for service level agreement guarantee via deep reinforcement learning[C]//Proceedings of the 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring). Piscataway: IEEE Press, 2022: 1-5.
- [15] SUN G L, GEBREKIDAN Z T, BOATENG G O, et al. Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks[J]. IEEE Access, 2019, 7: 45758-45772.
- [16] YAN M, FENG G, ZHOU J H, et al. Intelligent resource scheduling for 5G radio access network slicing[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 7691-7703.
- [17] MEI J, WANG X B, ZHENG K, et al. Intelligent radio access network slicing for service provisioning in 6G: a hierarchical deep reinforcement learning approach[J]. IEEE Transactions on Communications, 2021, 69(9): 6063-6078.
- [18] YE F, WANG J, LI J M, et al. Intelligent hierarchical network slicing based on dynamic multi-connectivity in cell-free distributed massive MIMO systems[J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11855-11870.
- [19] 夏玮玮, 王博业, 夏雅星, 等. 基于多时间尺度协同的无蜂窝 RAN 切片资源分配算法[J]. 通信学报, 2025, 46(7): 60-77.
- XIA W W, WANG B Y, XIA Y X, et al. Cellular-free RAN slicing resource allocation algorithm based on multi-timescale collaboration[J]. Journal on Communications, 2025, 46(7): 60-77.
- [20] WANG J Y, DAI L, YANG L, et al. Clustered cell-free networking: a graph partitioning approach[J]. IEEE Transactions on Wireless Communications, 2023, 22(8): 5349-5364.
- [21] BJORNSON E, JALDEN N, BENGTTSSON M, et al. Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission[J]. IEEE Transactions on Signal Processing, 2011, 59(12): 6086-6101.
- [22] JIANG J, WANG J C, CHU H Y, et al. Whale swarm reinforcement learning based dynamic cooperation clustering method for cell-free massive MIMO systems[J]. IEEE Transactions on Vehicular Technology, 2023, 72(3): 4114-4118.
- [23] HOSSAIN A R, ANSARI N. Priority-based downlink wireless resource provisioning for radio access network slicing[J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 9273-9281.
- [24] CAO J, WANG D M, LI J M, et al. Uplink spectral efficiency analysis of multi-cell multi-user massive MIMO over correlated Ricean channel[J]. Science China Information Sciences, 2018, 61(8): 082305.
- [25] BJÖRNSON E, SANGUINETTI L. Scalable cell-free massive MIMO systems[J]. IEEE Transactions on Communications, 2020, 68(7): 4247-4261.
- [26] YANG H J, ZHENG K, ZHANG K, et al. Ultra-reliable and low-latency communications for connected vehicles: challenges and solutions[J]. IEEE Network, 2020, 34(3): 92-100.
- [27] SONG F, LI J, MA C, et al. Dynamic virtual resource allocation for 5G and beyond network slicing[J]. IEEE Open Journal of Vehicular Technology, 2020, 1: 215-226.
- [28] NEELY M J. Stochastic network optimization with application to communication and queueing systems[M]. Cham: Springer International Publishing, 2010.
- [29] BERTSEKAS D P, GALLAGER R G. Data networks: 2nd edition[M]. Englewood Cliffs, New Jersey: Prentice-Hall, 1992.
- [30] ZENG W B, HE Y G, LI B, et al. Pilot assignment for cell free massive MIMO systems using a weighted graphic framework[J]. IEEE Transactions on Vehicular Technology, 2021, 70(6): 6190-6194.

[作者简介]



王博业 (2001-), 男, 东南大学移动通信国家重点实验室硕士生, 主要研究方向为无线网络资源管理、6G通信等。



夏玮玮 (1975-), 女, 博士, 东南大学移动通信国家重点实验室副研究员, 主要研究方向为无线网络资源管理、边缘计算、泛在网络与短距离无线通信等。



缪巍巍 (1968-), 男, 国网江苏省电力有限公司信息通信分公司正高级工程师, 主要从事电力信息通信网络、电力物联网、电力无线专网工作。



张明轩（1985-），男，博士，国网江苏省电力有限公司信息通信分公司高级工程师，主要从事电力通信工作。



燕锋（1983-），男，博士，东南大学移动通信国家重点实验室副研究员，主要研究方向为无线传感器网络、异构网络、无人机网络、自组织网络等。



潘裕庆（1977-），男，国网苏州供电公司高级工程师，主要从事电力通信与信息化工作。



沈连丰（1952-），男，东南大学移动通信国家重点实验室教授、博士生导师，主要研究方向为宽带移动通信、短距离无线通信和泛在网络等。