

面向5G电力虚拟专网的确定性自适应资源管理策略

夏炳森, 唐元春*, 陈霖锋, 游敏毅

(国网福建省电力有限公司经济技术研究院, 福州 350012)

摘要: 为保障电力用户性能稳定与服务质量(QoS), 文章提出一种面向5G电力虚拟专网的确定性自适应资源管理策略。首先, 针对电力业务经5G接入并依托光通信承载网络进行回传与汇聚的典型场景, 为保证电力服务资源的细粒度划分, 建立5G电力虚拟专网切片系统模型, 在大时间尺度上动态调整为业务保留的切片资源并建立资源约束, 在小时间尺度上采用信息年龄(AoI)来保证传输时延的确定性, 灵活配置功率及带宽。其次, 在保证业务QoS的前提下, 提出最大化长期系统效用的问题, 并将该问题拆分为双时间尺度子问题。最后, 为了求解资源管理最优策略, 在大时间尺度上, 采用双深度Q网络(DDQN)算法输出资源分配决策; 在小时间尺度上, 采用优先经验回放多智能体复合动作演员评论家(PER-MACA2C)算法动态调度资源。仿真结果表明, 所提算法能够有效降低时延, 并实现负载惩罚、成本的优化。该策略可为5G电力虚拟专网与光通信承载网络协同场景下的确定性资源管理提供参考。

关键词: 网络切片; 资源分配; 5G电力虚拟专网; 深度强化学习; 光通信承载网络

中图分类号: TN929.5 **文章编号:** 1001-5868(2026)04-0724-09

Deterministic Adaptive Resource Management Strategy for 5G Electricity Virtual Private Networks

XIA Bingsen, TANG Yuanchun, CHEN Linfeng, YOU Minyi

(Economic and Technological Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou 350012, CHN)

Abstract: In this study, a deterministic adaptive resource management strategy for 5G power private networks is developed to ensure stable performance and quality of service (QoS) for power users. First, for typical power-service scenarios where traffic is accessed through 5G and transported over optical communication bearing networks, to achieve fine-grained segmentation of electric service resources, a slicing system model for 5G power virtual private networks is developed. This model dynamically adjusts the slicing resources reserved for services and establishes resource constraints on a large time scale. Additionally, the age of information is adopted on a small time scale to ensure deterministic transmission delay and flexible configuration of power and bandwidth. Second, the problem of maximizing the long-term system utility while ensuring QoS is discussed herein. This problem is split into dual time-scale subproblems. Finally, to solve the optimal resource management strategy, the Double Deep Q Network algorithm is used to improve resource allocation decisions over large time scales, and the Preferred Empirical Replay Multi-Agents Composite Action Actor Critic algorithm is used to dynamically schedule resources over small time scales. The results show that the proposed algorithm can effectively reduce the delay and optimize load penalty and cost. The proposed

收稿日期: 2026-04-22.

基金项目: 国网福建经研院2024年基于多类业务的5G确定性网络仿真研究(52130N240008).

* 通信作者: 唐元春

strategy can provide a reference for deterministic resource management in coordinated scenarios of 5G power virtual private networks and optical communication bearing networks.

Key words: network slicing; resource allocation; 5G electricity virtual private networks; deep reinforcement learning; optical communication bearer network

0 引言

随着增强移动宽带、低时延高可靠通信和海量机器类通信等新业务的发展,传统电力虚拟专用光网络通信技术已难满足日益多样化的需求^[1],亟需探索新技术以应对复杂多变的电网业务。在此背景下,5G局域网(Local Area Network, LAN)逐渐成为助力垂直行业应用蓬勃发展的重要技术之一。5G LAN是利用5G技术将终端进行“分组”,组成一个局域网。其具有高可靠、低时延、抗干扰和高安全性的特点,是3GPP R16中最具前景的技术之一,能够满足企业园区网络通信的便捷管理、灵活互通和可靠通信的三类需求。此外,网络切片作为5G关键技术之一,通过在统一的基础设施上分离出多个虚拟的端到端网络实现按需组网,以适配各种类型的应用^[2]。网络切片能够为每种应用提供专用的网络资源,保障其服务质量(Quality of Service, QoS)和性能,为任务提供超低时延的通信服务,确保及时响应和操作。在电力业务通信场景中,5G虚拟专网主要解决终端灵活接入与无线侧隔离保障问题,而业务数据在完成无线接入后,通常仍需依托电力光纤通信专网、光传送网络或分组承载网络进行回传与汇聚,最终实现主站、边缘节点与现场终端之间的端到端互联。因此,面向新型电力系统的通信网络正呈现5G无线接入+光通信高可靠承载的光电融合发展趋势。

文献[3]提出一种节能的强化学习方法,为5G网络中的无线接入网切片分配无线资源和功率。为了实现细粒度频谱分配快速收敛的目的,文献[4]通过从Q值函数中获取优势项和状态值项,将离散归一化优势函数引入深度Q网络(Deep Q-Network, DQN),利用确定性策略梯度下降算法避免了不必要的计算。同样,文献[5]提出一种生成对抗网络驱动的双深度Q网络(Double Deep Q Network, DDQN),将不断变化的服务需求视为环境状态,将分配的资源视为行动。与文献[4]和[5]不同的是,文献[6]和[7]采用强化学习与深度学习相结合的复合学习架构。文献[6]的目标是

确保切片之间的性能隔离,而文献[7]的目标是最大化频谱效率和服务水平协议满意度。然而,这些工作仅根据实时的网络信息进行资源分配以保证QoS需求,而忽视了网络的动态性对服务性能的影响。

为了解决上述问题,本文提出一种面向5G电力虚拟专网的确定性自适应资源管理策略。首先,为了支持具有不同QoS需求的多样化电力服务并保证资源的细粒度划分,建立了5G电力虚拟专网切片系统模型。该模型在大时间尺度上动态调整为业务保留的切片资源,小时间尺度上采用信息年龄(Age of Information, AoI)保证传输时延的确定性,通过更加灵活的资源调度进一步为用户提供可靠通信。其次,在保证业务服务质量的前提下,提出最大化长期系统效用的问题。由于不同时间尺度之间存在耦合性,该问题被拆分为双时间尺度子问题。最后,为了解决同时间尺度的优化问题,提出一种分层深度强化学习(Deep Reinforcement Learning, DRL)框架。在大时间尺度上,采用DDQN算法来做出更好的资源分配决策;在小时间尺度上,采用优先经验回放多智能体复合动作演员评论家(Preferred Empirical Replay Multi-Agents Composite Action Actor Critic, PER-MACA2C)算法,用于解决包含离散动作和连续动作的问题,动态调整资源。

1 系统模型

1.1 电力虚拟专网-切片场景

本文考虑5G电力虚拟专网结合5G LAN技术下的智能资源管理。如图1所示,在物理网络中,由若干基站、切片、电力终端以及光通信承载网络组成电力虚拟专网-切片系统架构。联网电力终端可以接入覆盖范围内的基站来处理任务,从而减少相应的处理时间。终端业务在通过5G无线链路接入后,还需依托光纤通信链路完成回传、汇聚与上送,从而形成面向电力业务的端到端通信路径。假设基站集合表示为 $M=\{1, \dots, m, \dots, M\}$,其中 m 表示第 m 台基站。电力终端集合表示为 $I=$

$\{1, \dots, i, \dots, I\}$, 其中 i 表示第 i 个终端用户, 每个终端用户 $i \in I$ 都有一个基站提供服务, 并且每个基站 $m \in M$ 都有一组关联的用户终端, 用 I_m 表示。切片集合可以表示为 $N = \{1, \dots, n, \dots, N\}$, 每个切片都有一组关联的最终用户, 用 I_n 表示。

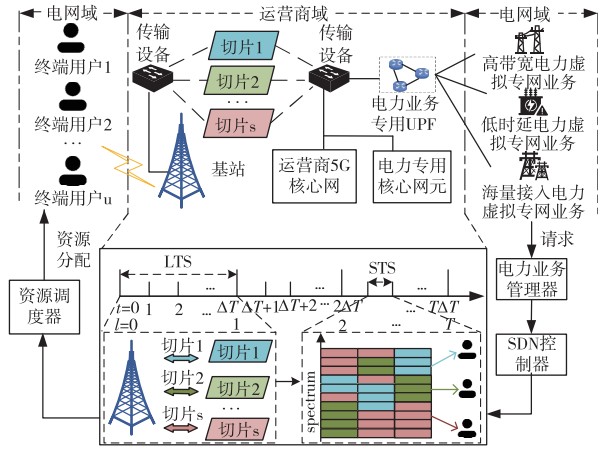


图1 电力虚拟专网-切片系统架构

因此, 切片资源管理不仅需要关注无线接入侧资源分配, 还需兼顾光承载侧资源保障, 以满足电力业务差异化的时延与可靠性需求。资源管理过程分为切片间与切片内两个级别。在切片间级, SDN可控制基站根据资源需求为切片分配带宽资源; 在切片内级, 由切片执行RB分配过程, 为每个最终用户分配必要的RB。

1.2 双时间尺度资源管理模型

为了适应网络动态拓扑变化需要进行细粒度资源调度, 以及避免频繁的切片重配置的成本过高及资源浪费问题, 在双时间尺度框架下进行资源管理。小时间尺度定义为STI, 用 $t \in \{1, 2, \dots, T\}$ 表示, 时隙 t 与下一时刻 $t+1$ 之间的固定持续时间为 τ ; 大时间尺度定义为LTI, 用 $l \in \{1, 2, \dots, L\}$ 表示, 持续时间为 ΔT 个STI。在每个LTI中, 系统重新确定用户与基站的关联策略, 以及预留给每个切片的带宽资源与计算资源。考虑到电力业务在完成5G无线接入后仍需通过光通信承载网络进行回传与汇聚, 大时间尺度上的切片资源预留不仅影响无线侧接入能力, 也影响业务在光承载侧的传输保障能力。因此, 本文所述切片资源可进一步理解为面向端到端业务传输的综合承载资源, 其预留结果为小时间尺度下的细粒度调度提供边界条件。在每个STI中, 为保证用户在每一时隙 t 内满足稳定性及动态性, 系统会根据终端用户状态进行切片内部的资源调度。

源调度。

在大时间尺度下, W_n^l 为在第 l 个LTI中系统为切片 n 预留的带宽资源, 得到预留给不同切片的带宽资源 W_n^l 后可以计算出切片 n 中RB数量, 即 $num_n^{\max} = W_n^l/B$ 。在小时间尺度下, 切片 n 中的物理资源块RB剩余情况表示为 R_n ; $A_n = \{A_{i,k}^n(t)\}_{i \in I_n, k \in K}$

表示RB的分配情况, 定义二进制变量 $A_{i,k}^n(t)$ 用于表示在时隙 t 切片 n 中第 $k \in K$ 个RB的分配情况, 若分配给第 i 台设备, 则 $A_{i,k}^n(t) = 1$, 否则 $A_{i,k}^n(t) = 0$ 。

假设终端用户 i 与基站 m 的通信包括将用户动态业务请求传送至基站的上行链路通信和将结果发送给用户的下行链路通信, 本文在通信模型中只考虑上行链路通信延迟。在时隙 t 切片 n 上的用户 i 与基站 m 可实现的SINR为:

$$SINR_i(t) = \frac{p_i(t) |\tilde{h}_{i,m}|^2 d_{i,m}^{-\alpha}(t)}{\tilde{I} + \sigma^2} \quad (1)$$

式中, $p_i(t)$ 为用户 i 的发射功率, $h_{m,i}(t)$ 为用户 i 的信道功率增益, $d_{i,m}^{-\alpha}(t)$ 表示路径损耗, α 是路径损耗的指数, $\tilde{I}$ 和 σ^2 分别表示干扰噪声和加性高斯白噪声功率; $d_{i,m}(t)$ 表示终端用户 i 和基站 m 之间的距离, 可以表示为: $d_{i,m}(t) = \sqrt{(x_u - x_m)^2 + (y_u - y_m)^2}$, 其中 (x_m, y_m) 表示基站 m 的坐标位置, (x_i, y_i) 表示用户 i 的坐标位置。

假设一个资源块上用户设备可以实现的传输速率为 r_B , 可以表示为:

$$r_B(t) = B \log_2 [1 + SINR_i(t)] \quad (2)$$

式中, B 表示一个RB的带宽。

因此用户终端 $i \in I_n$ 的无线通信速率可表示为:

$$r_i^n(t) = \sum_{k=1}^K A_{i,k}^n(t) r_B(t) \quad (3)$$

接入到切片 n 中第 i 个用户的传输速率要求表示为 $R_n^{\min}$, 则第 i 个用户所需要的RB数为 $num_{i,n}^{\min} = \lceil R_n^{\min}/r_B(t) \rceil$ 。为了满足所需RB数为整数, 定义 $\lceil x \rceil$ 表示为向上取整函数, 即若 x 为整数, 则 $\lceil x \rceil$ 为自己 x 本身, 若 $\lceil x \rceil$ 为小数, 则为 $\lceil x \rceil + 1$ 。

假设发送至基站的由应用程序和数据组成的数据包大小为 $D_i(t)$, 基站接收 $D_i(t)$ 后返回反馈信号ACK, 用户设备接收到反馈信号后进行下一次数据包的传输。若接收到数据包出现误码则会丢弃该数据包, 同时反馈给终端重新发送出错的数数据包。

数据传输过程中的差错可由 SINR 来衡量,根据文献[8],如果 SINR 小于阈值 Γ^{th} ,则采用二进制变量 $\gamma_i(t)=1$ 表示,即在时隙 t 用户设备 i 发送的信息能经受住干扰,数据能够成功在基站侧解码;否则 $\gamma_i(t)=0$,由式(4)给出,即:

$$\gamma_i(t)=\begin{cases} 1, \text{SINR}_i(t) \geq \Gamma^{\text{th}} \\ 0, \text{SINR}_i(t) < \Gamma^{\text{th}} \end{cases} \quad (4)$$

此外,传输速率直接影响数据包能否成功到达基站。速率过低可能导致数据无法在规定时间内传输完整,造成丢包或分包,从而影响传输成功率。采用二进制变量 $o_i(t)=1$ 表示传感数据在单位时隙内成功发送,否则 $o_i(t)=0$,由式(5)给出,即:

$$o_i(t)=\begin{cases} 1, r_i^n(t)\tau \geq D_i(t) \\ 0, r_i^n(t)\tau < D_i(t) \end{cases} \quad (5)$$

假设虚拟电力专网中用户之间的任务是独立的^[9],用户 i 在时隙 t 的任务到达率遵循泊松过程,表示为 $\lambda_i(t)$ 。用户具有不同的任务到达率,则总体到达率可以表示为 $\lambda(t)=\sum_{i \in I} \lambda_i(t)$ 。为了评估平均传输延迟,将用户设备传输信息到基站的过程建模为一个 M/M/1 队列^[10]。由于无线信道的随机性,假设平均服务率服从指数分布^[11],平均服务率可表示为: $\mu(t)=E[r_i^n(t)/D_i(t)]$

根据 M/M/1 排队理论^[12],可以定义平均传输时延为:

$$T^{\text{up}}(t)=\frac{1}{\mu(t)-\lambda(t)} \quad (6)$$

为了保证传输队列的稳定性,必须满足以下条件:

$$E[r_i^n(t)/D_i(t)] - \sum_{i \in I} \lambda_i(t) \geq 0 \quad (7)$$

AoI 是关于信息时效性的度量指标,被定义为目标节点最新接收到的数据包自其生成以来所经过的时间^[13-14],用 $\Delta_i(t)$ 衡量用户设备 i 产生的数据包 AoI。当用户 i 向基站传输数据包时,时隙 t 用户 AoI 值表示为:

$$\Delta_i(t)=\begin{cases} T^{\text{up}}(t), \gamma_i(t) \cdot o_i(t)=1 \\ \Delta_i(t-1)+1, \gamma_i(t) \cdot o_i(t) \neq 1 \end{cases} \quad (8)$$

定义用户设备 i 在 DT 同步时的平均 AoI 为:

$$\bar{\Delta}_i(t)=\frac{1}{\tau_0} \sum_{\tau=t-\tau_0+1}^t \Delta_i(\tau)。假设 \xi_n^{\text{max}} 为 AoI 约束,则所有用户的 AoI 应当满足约束 \Delta_{i_n}(t) \leq \xi_n^{\text{max}}, i \in I_n。$$

2 问题构造

假设系统关于时延的收益与传输 AoI 有关,收益与系统所需和获得的 AoI 之差成正比关系,随着 AoI 差的增加,收益也相应增加,即:

$$RE_{\text{age}}(t)=\chi_{\text{in}} \max \left\{ \xi^{\text{max}} - [\Delta_i(t)]_{i \in I}, 0 \right\} \quad (9)$$

式中, χ_{in} 为系统关于时延的单位收入。同理,系统关于传输差错收益与用户数据包传输过程有关,即:

$$RE_{\text{tr}}(t)=\eta_{\text{in}} \max \left\{ [\text{SINR}_i(t)]_{i \in I} - \Gamma^{\text{th}}, 0 \right\} + \epsilon_{\text{in}} \max \left\{ [r_i(t)\tau - \Delta D_i^{\text{sensor}}]_{i \in I}, 0 \right\} \quad (10)$$

式中, η_{in} 和 ϵ_{in} 分别为系统的传输过程单位收入。因此系统总收益为:

$$RE_{\text{to}}(t)=\beta_1 RE_{\text{age}}(t)+\beta_2 RE_{\text{tr}}(t) \quad (11)$$

式中, β_1 和 β_2 为权重系数,满足 $\beta_1 + \beta_2 = 1$ 。

切片为用户预留通信资源,资源成本是管理中的关键指标。对于电力业务而言,该类通信资源不仅包括无线接入侧带宽资源,也包含支撑业务回传与汇聚的光通信承载资源,因此资源成本可理解为面向端到端切片保障的综合开销。尽管目标是尽可能低延迟,但前提是充足的通信资源,而这些资源是有代价的,资源分配成本计算式为:

$$Co_{\text{re}}^l = c_{\text{sl}} \sum_{n \in N} W_n^l + l_{\text{sl}} \sum_{n \in N} |W_n^l - W_n^{l-1}| \quad (12)$$

式中, Co_{re}^l 为切片在第 l 个 LTI 需要支付资源的成本函数, c_{sl} 为单位带宽的使用成本, l_{sl} 为切片缩放的单位成本。

为减轻负载、提升均衡性并便于调度,需惩罚资源分配不当。考虑到切片预留资源不同,应基于剩余资源与实际调度资源的相对比例进行分析。假设 η_n^l 为在第 l 个 LTI 时切片 n 的带宽资源预留率,其计算公式为:

$$\eta_n^l = \frac{W_n^l - B \cdot \sum_{t=l\Delta T+1}^{(l+1)\Delta T} \sum_{k=1}^K \sum_{i \in I_n} A_{i,k}^n(t)}{W_n^l} \quad (13)$$

假设 α_{B} 表示切片资源预留警戒值,对资源预留率不足该门限的部分给予惩罚,与警戒值相差越多则应受惩罚越多,则切片 n 在第 l 个 LTI 时的超载惩罚 E_n^l 可表示为:

$$E_n^l = \begin{cases} \epsilon_{\text{B}} [\eta_n^l - \alpha_{\text{B}}]^2, \eta_n^l < \alpha_{\text{B}} \\ 0, \eta_n^l \geq \alpha_{\text{B}} \end{cases} \quad (14)$$

式中, ϵ_B 表示资源预留率不足部分所受到的单位惩罚。设置切片资源预留率最小阈值 ϵ_B , 各切片的 η_n^l 不得小于 α_B 。当 α_B 为 0 时表示不考虑未来情况而允许对切片内的资源进行不保留分配。

综上, 在第 l 个 LTI 的系统效用可以表示为总收益与总成本的差值, 即:

$$U^l = \omega_1 \sum_{t=l\Delta T+1}^{(l+1)\Delta T} RE_{\text{total}}(t) - \omega_2 Co_{\text{re}}^l - \omega_3 \sum_{n \in N} E_n^l \quad (15)$$

实验目标是在最大化系统收益的同时控制系统成本, 寻求最优双时间尺度切片资源管理最优策略。优化问题表述为:

$$\begin{aligned} \text{P1: } \max_{A, P, W} \sum_{l=0}^T U^l / T \\ \text{s.t. C1: } A_{i,k}^n(t) \in \{0, 1\}, \forall i \in I_n, k \in K \\ \text{C2: } E[r_i(t) / \Delta D_i^{\text{sensor}}(t)] - \sum_{i \in I_n} \lambda_i(t) \geq 0 \\ \text{C3: } \sum_{n \in N} \sum_{i \in I} A_{i,k}^n(t) \leq 1, k \in K \\ \text{C4: } \Delta_{i,n}(t) \leq \xi_n^{\max}, i \in I_n, n \in N \\ \text{C5: } \sum_{n \in N} W_n^l \leq W_{\text{re}}^l, l = 1, 2, \dots \\ \text{C6: } \sum_{i \in I_n} \sum_{k=0}^{\text{num}_k^n} A_{i,k}^n(t) \leq \text{num}_n^{\max} \\ \text{C7: } \sum_{k=0}^{\text{num}_k^n} A_{i,k}^n(t) \geq \text{num}_{i,n}^{\min}, \forall i \in I_n, \forall n \in N \\ \text{C8: } R_n^{\min} \leq r_i(t), \forall i \in I_n, \forall n \in N \\ \text{C9: } 0 \leq p_i \leq P_{\max}, \forall i \in I_n, \forall n \in N \end{aligned} \quad (16)$$

在上述约束条件中, C1 表明资源块分配变量是一个二进制变量; C2 保证传输队列的稳定性; C3 保证每个资源块在单个时隙内只分配给唯一用户; C4 保证数据包传输过程满足 AoI 需求; C5 保证切片占用频带资源总和不超过总频带资源; C6 表示用户带宽之和不能高于切片的带宽资源; C7 保证所有设备满足 QoS 要求; C8 表示用户传输速率约束; C9 表示用户发射功率约束。集合 A, P, W 分别表示大时间尺度下带宽预留, 小时间尺度下资源块分配和用户功率的分配。

根据双时间尺度特性, 可直接将 P1 拆分为不同尺度的子问题。考虑到在大时间尺度上关注的是总体资源分配策略, 无需进行细粒度的考虑和分配, 可以接受连续变量转换为离散变量产生的微小损失。因此, 将连续变量 W_n^l 转为离散变量 $w_{D,n}^l$, 其集合为 W^D 。离散变量定义为使用资源占总资源的比例, 即: $w_{D,n}^l = w\%$, $w = 1, 2, \dots, 100$ 。

在第 l 个 LTI 时刻切片 n 带宽资源分配用离散

变量表示为 $w_{D,n}^l$, 连续变量表示为 W_n^l , 系统剩余带宽资源表示为 W_{re}^l , 忽略离散变量和连续变量转换的精度差, 两者关系表示为: $W_n^l = w_{D,n}^l \cdot W_{\text{re}}^l$ 。因此 P2 表达式为:

$$\begin{aligned} \text{P2: } \max_{W^D} \sum_{l=0}^T \left(-\omega_2 Co_{\text{re}}^l - \omega_3 \sum_{n \in N} E_n^l \right) / T \\ \text{s.t. C1: } \sum_{n \in N} w_{D,n}^l \leq 1, l = 1, 2, \dots \\ \text{C2: } w_{D,n}^l \in [0, 1], n \in N, l = 1, 2, \dots \end{aligned} \quad (17)$$

在小时间尺度上, 主要考虑细粒度切片的资源调度及用户发射功率。目标子问题 P3 表达为:

$$\begin{aligned} \text{P3: } \max_{A, P} \left[\sum_{z=0}^T \sum_{t=z\Delta T+1}^{(z+1)\Delta T} \omega_1 RE_{\text{total}}(t) \right] / T \\ \text{s.t. (16a)-(16d), (16f)-(16i)} \end{aligned} \quad (18)$$

3 双时间尺度下智能资源管理算法

3.1 双时间尺度马尔科夫决策过程

在问题 P2 中, 状态集合由通信资源剩余情况、资源块分配状态及切片集合构成, 其中通信资源剩余情况可综合表征无线接入侧与光通信承载侧的可用资源状态, 可以表示为 $\mathcal{S}_L = \{\mathcal{W}(z-1), \mathcal{N}, W_{\text{re}}^l\}$; 动作集合为在第 l 个 LTI 时刻完成切片预留通信资源划分, 可以表示为 $\mathcal{A}_L = \{\{\mathcal{W}(l)\}_{l \in T}\}$; 奖励函数为目标函数 $U_L = -\omega_2 Co_{\text{re}}^l - \omega_3 \sum_{n \in N} E_n^l$ 。

在问题 P3 中, 状态集合由切片预留资源情况、用户数量、用户到达率构成, 可以表示为 $\mathcal{S}_s = \{\mathcal{W}(z), \mathcal{I}, (\lambda)_{i \in \mathcal{I}}\}$; 动作集合由 t 时隙切片频带资源块调度和用户设备发射功率分配构成, 可以表示为 $\mathcal{A}_s = \{\mathcal{A}(t), \mathcal{P}(t)\}$; 奖励函数为目标函数 $U_s = \omega_1 RE_{\text{total}}(t)$ 。

3.2 大时间尺度下基于 DDQN 的切片间资源分配算法

基于 DDQN 的切片间资源分配算法如表 1 所示。其通过引入目标网络与训练网络的双网结构, 缓解 DQN 奖励高估问题, 进而提升了训练稳定性。其损失函数表示为:

$$\begin{aligned} L(\mu) = E \left\{ \left\{ Q(s, a, \mu) - \left\{ r + \right. \right. \right. \\ \left. \left. \left. \gamma_s Q \left[s', \arg \max_a Q(s', a, \mu), \mu' \right] \right]^2 \right\} \right\} \end{aligned} \quad (19)$$

式中, μ 表示主 Q 网络参数, μ' 代表目标网络的参

数。采用adam优化算法来最小化损失函数梯度并完成Q网络参数更新,损失函数的梯度计算公式为:

$$\nabla_{\mu} L(\mu) = E \left\{ r + \gamma_s Q \left[s', \arg \max_a Q(s', a, \mu), \mu' \right] - Q(s, a, \mu) \nabla_{\mu} Q(s, a, \mu) \right\} \quad (20)$$

表1 基于DDQN的切片间资源分配算法

输入: 学习率 λ_Q , 软更新参数 τ , Q网络参数 μ , 目标网络参数 μ' , 探索概率 ϵ , 折扣因子 γ_s
输出: 资源分配策略
(1) 初始化: 回放缓存池 R_1 , 目标网络参数 $\mu' \leftarrow \mu$
(2) for $episode = 1 \sim episode_max$ do
(3) 智能体观察环境并获得初始值 s
(4) for $step = 1 \sim step_max$ do
(5) 选择 ϵ -贪婪概率选择动作, 即 $a = \begin{cases} \arg \max_a Q(s, a, \mu), 1 - \epsilon \\ random, \epsilon \end{cases}$
(6) 智能体获得与环境交互的惩罚与奖励, 过渡到下一状态 s'
(7) 将经验 (s, a, r, s') 存储到回放池 R_r , 并从经验池 R_r 中抽取样本sample
(8) 根据公式(19)和(20)计算损失函数梯度, 完成网络参数 μ 的更新
(9) 每隔 K 步更新目标网络参数 $\mu' \leftarrow \tau\mu + (1 - \tau)\mu'$
(10) end for
(11) end for

因此,大时间尺度下的资源分配决策不仅决定切片无线带宽预留水平,也为业务在光通信承载网络中的稳定传输提供资源保障基础。

3.3 小时间尺度下基于PER-MACA2C的切片内资源调度算法

基于PER-MACA2C的切片内资源调度算法子算法如表2所示。小时间尺度上的资源划分更细致灵活,不能接受连续变量离散化带来的损失。为解决同时涉及离散与连续动作的问题,基于文献[15],本文提出PER-MACA2C算法,用于解决既包含离散动作又包含连续动作的问题。算法框架如图2所示。

把每个基站视为一个智能体,则智能体 m 在当前状态 $s(t)$ 下使Q函数值最大的离散动作作为:

$$a_m(t) = \arg \max_{a_m(t)} Q_m^* \left(s(t), \{a_{i,m}(t), \pi_m^*[s(t), a_{s,m}(t); \mu_m]\}; \vartheta_m \right) \quad (21)$$

式中,策略 $\pi_m^*(s(t), a_{s,m}(t); \mu_m)$ 提供给定状态 $s(t)$ 和离散动作资源块分配下的最佳功率。在Actor-Critic体系结构下来估计 π_m^* 和 Q_m^* ,其中 μ_m 和 ϑ_m 分别表示Actor和Critic网络参数。

表2 基于PER-MACA2C的切片内资源调度算法子算法

输入: 学习率 α_E, β_E , 探索概率 ϵ , 软更新参数 τ , 策略网络参数 $(\mu_m)_{m \in M}$, Q网络参数 $(\vartheta_m)_{m \in M}$, 目标网络参数 $(\hat{\mu}_m, \hat{\vartheta}_m)_{m \in M}$
输出: 资源块调度及发射功率调度策略
(1) 初始化: 回放缓存池 R_S , 目标网络参数
(2) for $episode = 1 \sim episode_max$ do
(3) 所有智能体观察初始状态
(4) for $step = 1 \sim step_max$ do
(5) 所有智能体选择资源块调度动作 $\left\{ a_s^m(t) = \begin{cases} \text{公式(21)}, 1 - \epsilon \\ random, \epsilon \end{cases} \right\}_{m \in M}$
(6) 确定观测频率及发射功率分配动作 $a_{s,m}(t) = \pi_m(s(t), a_{i,m}(t); \mu_m)$
(7) 智能体获得与环境交互的惩罚与奖励,过渡到下一状态 s_m'
(8) 将经验存储到回放池 R_S
(9) for 智能体 $m = 1, 2, \dots, M$ do
(10) 从经验池 R_S 抽取样本
(11) 根据公式(24)计算损失函数和梯度
(12) 更新Q网络参数 $\vartheta_m = \vartheta_m - \beta_E \nabla_{\vartheta_m} L_m(\vartheta_m)$
(13) 更新策略网络参数 $\mu_m = \mu_m - \alpha_E \nabla_{\mu_m} J_m(\mu_m)$
(14) 更新目标网络参数
(15) end for

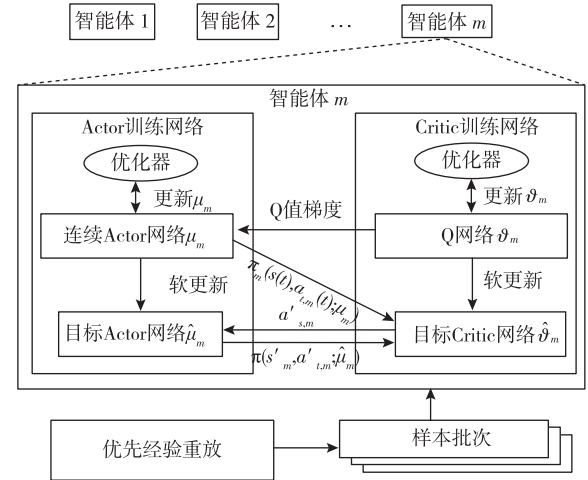


图2 PER-MACA2C 框架图

在小时间尺度上,切片内资源调度主要面向无线接入侧进行细粒度优化,但其调度结果受到大时间尺度端到端资源预留状态约束,从而保证无线接入与光承载过程之间的协同一致性。

对于行动者网络训练,智能体 m 首先结合状态 $s(t)$ 和资源块调度决策,将其作为Actor网络 $\pi_m(s(t), a_m(t); \mu_m)$ 的输入,Actor网络根据这些输入计算出功率和观测频率的选择。然后智能体 m 将状态、选择的连续动作和做出的资源块调度决策结合起来,作为评价者网络的输入。

智能体 m 朝着最大化期望奖励采取对应动作,根据策略梯度,Actor网络进行更新,梯度为:

$$\nabla_{\mu_m} J_m \approx \frac{1}{|r|} \sum_r \nabla_{\mu_m} \pi_m(s_m^r) \times \nabla_{a_m} Q\left(s_m^r, \left\{a_{t,m}, \pi_m[s_m^r, a_{s,m}(t); \mu_m]; \vartheta_m\right\}\right) \quad (22)$$

式中, $|r|$ 为最小批数据的数量, r 为经验的指标。

对于 Critic 网络, 其参数通过梯度下降的方式更新, 以最小化评论家网络的目标 Q 值和实际获得 Q 值之间的误差。Critic 网络的训练目标是逼近真实的动作值函数, 以提供准确的动作评估。目标 Critic 网络输出为:

$$y_m = r_m + \epsilon Q\left(s_m', \left[a_{t,m}', \pi(s_m', a_{s,m}'; \hat{\mu}_m)\right]; \hat{\vartheta}_m\right) \quad (23)$$

式中, $\hat{\mu}_m$ 和 $\hat{\vartheta}_m$ 分别为行动者和批评家的目标网络参数, 以减少 Q 值估计过高的倾向。

Q 网络的均方误差通常被采用来衡量评论家网络的差值, 即:

$$L_m(\vartheta_m) = \frac{1}{|r|} \left[y_m^r - Q(s_m^r, a_m^r | \vartheta_m) \right]^2 \quad (24)$$

同时, 采用优先经验回放 (Preferred Empirical Replay, PER) 能够增加高 TD-error 值的体验的重播概率, 从而导致高样本效率和更好的策略。在 PER 中, 第 r 次经验的重要性通过 TD-error 的绝对值来衡量, 即: $\delta_r = |y_m^r - Q(s_m^r, a_m^r | \vartheta_m)|$ 。

TD-error 值越大, 经验重放的可能性就越大, 采样概率定义为: $P(r) = p_r^\alpha / \sum_r p_r^\alpha$, 其中 α 决定执行多少优先级。采用比例优先级, 其中 $p_r = |\delta_r| + \zeta$, ζ 是一个很小的正值, 以防止样本 r 本抽取到的概率为 0。

4 仿真实验

基于 Python 和 Pytorch 进行模拟来验证所提出算法的性能, 仿真参数见表 3。仿真场景中默认电力业务经 5G 虚拟专网接入后由光通信承载网络完成回传与汇聚, 因此本文对切片资源管理效果的评估可视为面向光无线协同承载场景下的性能验证。在 $[0, 10] \text{ km} \times [0, 10] \text{ km}$ 的工业物联网拓扑下, 基站数量为 5, 用户设备数量为 20。对于 DDQN 算法, 网络设置两个全连接层, 每层含有 64 个神经元, 激活函数为 Relu。对于 MACA2C 算法, 演员和评论家神经网络是三层全连接网络, 每层有 512 个神经元, 并使用 Relu 作为激活函数。

表 3 仿真参数设置

参数	值	参数	值
基站数量	5	AoI 约束	70 ms
用户数量	20	用户数据包	400 kB
带宽	50 MHz	经验池容量及最小批	512/32
平均任务到达率	1	子算法 1、2 折扣因子	0.95/0.9
切片传输速率约束	$1.6 \times 10^4 \text{ bps}$	软更新参数	0.01
最大用户发射功率	700 mW	子算法 1 学习率	0.001
LTS/STS 长度	5 s/500 ms	子算法 2 Actor/Critic 学习率	0.01/0.001

不同算法的总效用如图 3 所示, 对比算法为 DDQN-MACA2C 算法、DQN-MADDPG 算法及 SSP^[16] 算法。其中, SSP 资源根据需求按比例分配给切片, 并且随时间保持不变。用户资源通过切片提供, 资源保证最少。由图可知, 所提出的算法总效用最高, SSP 算法效用最低。该算法采用 PER, 收敛速度更快且采用多智能体之间进行合作机制, 因此学习效率更高。此外, 该算法在小时间尺度中联合考虑了离散与连续动作, 相比其他算法具有更高的效用。由于 SSP 策略资源分配相对固定, 不会频繁变化, 不适合动态变化网络。

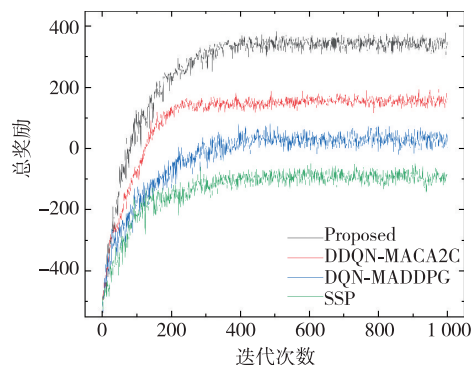


图 3 不同算法的总效用对比

不同算法资源成本如图 4 所示。SSP 将资源固定分配给各个切片, 无法根据实际需求和网络状态进行动态调整, 易造成资源浪费, 增加额外的成本, 因此资源成本最高。所提算法在大时间尺度通过 DDQN 双网络结构能更好训练策略, 解决了 DQN 估计奖励高于实际奖励的问题, 因此成本最低。

图 5 和图 6 分别为各算法在不同带宽资源和功率阈值下的平均 AoI。无论采取何种策略, 随着带宽资源及功率阈值的增加, 平均 AoI 都随之减小, 下降的幅度趋于平稳。所提算法的 AoI 最小, 在大时间尺度下通过 DDQN 算法合理分配不同切片资源后, 小时间尺度下通过多智能体协作进一步进行资源的细粒度划分。同时, 考虑了连续动作与离散动

作,对有限的资源能够进行合理的分配,证明所提算法能够对资源进行细粒度分配,并适应连续与离散动作混合的环境。

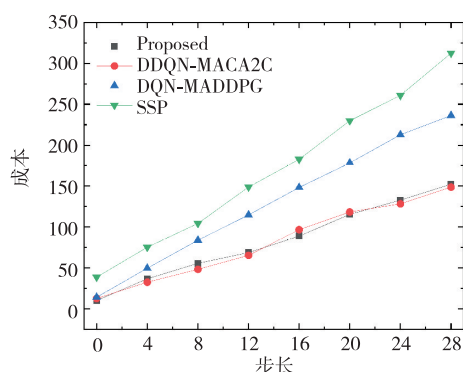


图4 不同算法资源成本对比

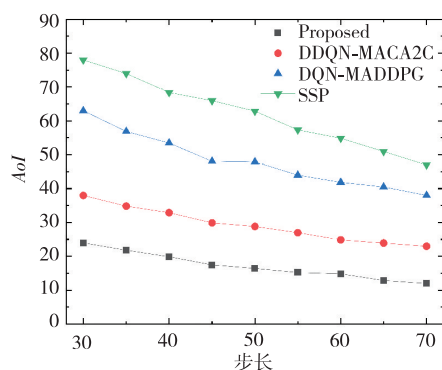


图5 不同带宽资源下的平均AoI对比

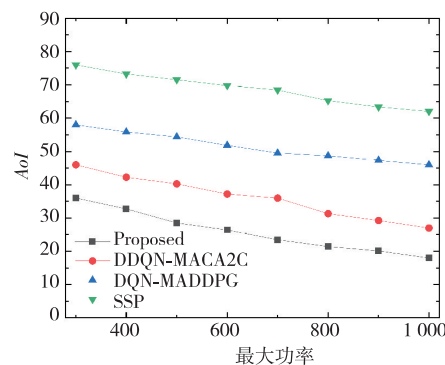


图6 不同功率阈值下的平均AoI对比

不同算法的网络超载惩罚如图7所示。网络超载惩罚随着用户数量的增加而增加。终端用户数量的增加会导致对网络资源需求的增加,切片能够预留的资源减少,进而容易造成网络超载。由于SSP为静态资源管理,在动态变化网络下资源超载惩罚最高,且MADDPG忽略了离散动作,相比之下所提算法的惩罚最小。

为进一步考察不同算法下的资源超载影响,对比了不同算法在不同警戒值门限下的网络超载惩

罚。如图8所示,随着警戒值门限的设定值增大,已超载的部分资源预留率与警戒值差距越来越大,而未超载的部分资源余留率小于警戒值的可能性也增大。所提算法在不同阈值下的惩罚最小,证明其能够保证用户QoS的情况下合理分配资源,有效应对突发情况。

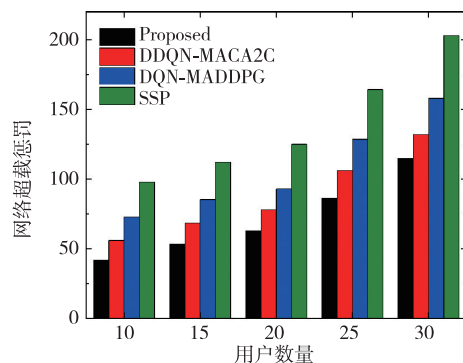


图7 不同算法的网络惩罚对比

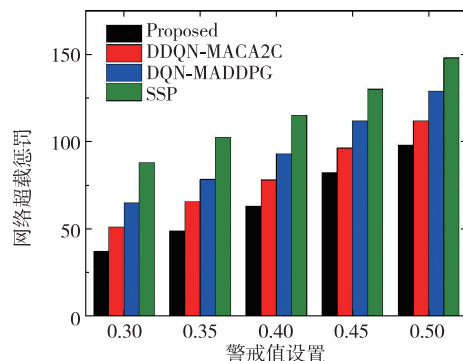


图8 不同警戒值门限下的网络超载惩罚对比

5 结论

本文提出一种面向5G电力虚拟专网的确定性自适应资源管理策略,以确保电力用户的性能稳定并满足QoS。首先,基于5G电力虚拟专网与切片结合的系统框架,建立5G电力虚拟专网切片系统模型,并将该模型分为大小时间尺度。在大时间尺度下动态调整为业务保留的切片资源,在大时间尺度的约束下,小时间尺度采用AoI保证传输时延的确定性,通过更加灵活的资源调度进一步为用户提供可靠通信。其次,提出最大化长期系统效用的问题,为解决双时间尺度的耦合性,将该问题拆分为两个时间尺度的子问题。最后,在大小时间尺度上分别采用DDQN算法和PER-MACA2C算法动态分配调度资源。结果表明,所提算法能够有效优化时延和成本,实现负载均衡。同时,本文结合电力业务5G无线接入+光通信承载的典型传输场景,

对双时间尺度下的切片资源管理进行了研究,可为5G电力虚拟专网与光通信网络融合建设中的确定性资源配置提供参考。

参考文献:

- [1] You X, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts[J]. Science China Information Sciences, 2020, 64(1): 110301.
- [2] Linguaglossa L, Lange S, Pontarelli S, et al. Survey of performance acceleration techniques for network function virtualization[J]. Proceedings of the IEEE, 2019, 107(4): 746-764.
- [3] Azimi Y, Yousefi S, Kalbkhani H, et al. Energy-efficient deep reinforcement learning assisted resource allocation for 5G-RAN slicing [J]. IEEE Transactions on Vehicular Technology, 2022, 71(1): 856-871.
- [4] Qi C, Hua Y, Li R, et al. Deep reinforcement learning with discrete normalized advantage functions for resource management in network slicing [J]. IEEE Communications Letters, 2019, 23(8): 1337-1341.
- [5] Hua Y, Li R, Zhao Z, 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.
- [6] Yan M, Feng G, Zhou J, et al. Intelligent resource scheduling for 5G radio access network slicing[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 7691-7703.
- [7] Li R, Wang C, Zhao Z, et al. The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility[J]. IEEE Communications Letters, 2020, 24(9): 2005-2009.
- [8] Mai T, Yao H, Zhang N, et al. Transfer reinforcement learning aided distributed network slicing optimization in industrial IoT [J]. IEEE Transactions on Industrial Informatics, 2022, 18(6): 4308-4316.
- [9] Wu W, Chen N, Zhou C, et al. Dynamic RAN slicing for service-oriented vehicular networks *via* constrained learning [J]. IEEE Journal on Selected Areas in Communications, 2021, 39(7): 2076-2089.
- [10] Xiang B, Elias J, Martignon F, et al. Joint network slicing and mobile edge computing in 5G networks[C]//ICC 2019 - 2019 IEEE International Conference on Communications (ICC), 2019: 1-7.
- [11] Zhang S, Luo H, Li J, et al. Hierarchical soft slicing to meet multi-dimensional QoS demand in cache-enabled vehicular networks[J]. IEEE Transactions on Wireless Communications, 2020, 19(3): 2150-2162.
- [12] Huang S, Sun Y, Feng D, et al. A resource allocation framework for network slicing with multi-service coexistence[C]//ICC 2021- IEEE International Conference on Communications, 2021: 1-6.
- [13] Yu B, Cai Y, Wu D, et al. Average age of information in short packet based machine type communication [J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 10306-10319.
- [14] Kadota I, Modiano E. Minimizing the age of information in wireless networks with stochastic arrivals [J]. IEEE Transactions on Mobile Computing, 2021, 20(3): 1173-1185.
- [15] Hu J, Zhang H, Song L, et al. Cooperative Internet of UAVs: distributed trajectory design by multi-agent deep reinforcement learning [J]. IEEE Transactions on Communications, 2020, 68(11): 6807-6821.
- [16] Nguyen D T, Le L B, Bhargava V K. A market-based framework for multi-resource allocation in fog computing[J]. IEEE/ACM Transactions on Networking, 2019, 27(3): 1151-1164.

作者简介:

夏炳森(1981—),男,工程师,主要从事电力通信规划设计及电力物联网建设研究;

唐元春(1973—),男,正高级工程师,主要从事电力通信规划设计及电力物联网建设研究。