光电技术及应用

DOI：10.16818/j.issn1001－5868.2024082702

基于深度强化学习的5G电力虚拟专用光网络切片时延优化算法研究

张　导[1]，秦　骁[1]，周之楠[1]*，邹喆旻[1]，钟淘淘[1]，唐　伦[2]

（1. 国网重庆市电力公司信息通信分公司，重庆 400014；2. 重庆邮电大学，重庆 400065）

摘　要： 针对5G电力虚拟专用光网络切片中业务端到端时延大及网元状态同步时延影响仿真策略准确性的问题，提出一种基于深度强化学习的5G电力虚拟专用光网络切片时延优化算法。首先，建立5G电力虚拟专用光网络切片系统模型，该模型包括一个用于网元状态更新的同步节点和其余业务节点，其中，同步节点通过专用特种光纤与软件定义网络控制器直接连接。在此基础上，在保证业务服务质量的同时，提出一个最小化包括业务时延和网元同步时延的优化问题。该优化问题的变量既有离散变量又有连续变量，因此，分别使用离散和连续的深度强化学习算法求解。仿真结果表明，所提出的算法能够有效降低电力虚拟专用光网络切片网络时延，在满足服务质量要求的同时能有效保证仿真策略的实时性。

关键词： 5G电力虚拟专用光网络；网络切片；时延优化；资源分配；深度强化学习

中图分类号： TN929.5　**文章编号：** 1001－5868(2024)06－0977－09

Research on Delay Optimization Algorithm in 5G Power Virtual Private Network Slicing Based on Deep Reinforcement Learning

ZHANG Dao[1], QIN Xiao[1], ZHOU Zhinan[1], ZOU Zhemin[1], ZHONG Taotao[1], TANG Lun[2]

(1. Information and Telecommunication Branch of State Grid Chongqing Electric Power Com. Ltd, Chongqing, 400014, CHN; 2. Chongqing University of Posts and Telecommunications, Chongqing 400065, CHN)

Abstract: Aiming at the issue of high end-to-end latency and the inaccurate of simulation strategy because of the large synchronization delay in 5G power virtual private optical network slicing, a latency optimization algorithm based on deep reinforcement learning is proposed. Firstly, a 5G power virtual private optical network slicing system model is established, which includes a synchronization node for network element updates and other service nodes, where the synchronization node is directly connected to the software defined network controller through a dedicated special fiber. Then, an optimization problem is proposed to minimize the total latency, including business and network element update latency. As this problem involves discrete and continuous variables, both discrete and continuous deep reinforcement learning algorithms were employed for solution. Simulation results show that the proposed algorithm can effectively reduce the latency of the power virtual private optical network slicing network, meet the requirements of service quality, and ensure the real-time performance of the simulation strategy.

Key words: 5G power virtual private optical network; network slicing; latency optimization; resource allocation; deep reinforcement learning

收稿日期：2024－08－27.

基金项目：国家电网有限公司科技项目(2023渝电科技6#).

* **通信作者：**周之楠

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，电力虚拟专用光网络通信技术逐渐显现其局限性，难以满足新兴业务愈发差异化的需求[1]。因此，亟需探索新的通信技术，以满足复杂多变、差异巨大的新型电网业务需求[2]。网络切片作为5G的一项关键技术，实现了一种按需组网的方式。该技术通过让运营商在统一的基础设施上分离出多个虚拟的端到端(End to End，E2E)网络，每个网络切片在无线接入网、承载网和核心网上进行逻辑隔离，以适配各种各样类型的应用[3]。光网络切片是在以光纤作为主要传输介质的光网络中进行切片。相较于传统切片，光网络切片能够实现更高速率和更远距离的传输。光切片由网络中部署在通用基础设施上的各种虚拟网络功能(VNF)动态组合而成。光网络切片将各通用基础设施抽象为节点，各节点的计算和存储资源总量各不相同。同时，部分节点之间通过光纤链路进行连接，而这些光纤链路的带宽存在一定差异[4]。网络业务的请求以服务功能链(Service Function Chaining，SFC)的形式到达网络切片，该SFC由通过光纤映射的虚拟链路按顺序连接的虚拟网络VNF组成。这些请求需要根据SFC的排列顺序，依次在相应节点的VNF上处理任务[5]。因此，网络切片首先按照SFC中VNF和虚拟链路顺序需求，将VNF部署在不同的节点上，同时选择虚拟链路实现信息传输，即VNF部署过程。由于VNF部署问题通常兼具NP难和非凸性，精确优化算法和启发式算法均无法求出其解析解。针对此类问题，一种可行解决方案是使用深度强化学习(Deep Reinforcement Learning，DRL)算法[6]。

目前，研究者对基于DRL的网络切片VNF的部署问题进行大量深入研究，且取得多项重要研究进展。文献[7]针对不同的网络拓扑中VNF部署策略的泛化能力，联合考虑资源利用问题和流量的服务质量(Quality of Service，QoS)保障，提出了结合DRL和图神经网络(Graph Neural Network，GNN)优势的VNF布局方案，命名为DeepOpt。该方案在不同的网络拓扑中均能表现出更好的泛化能力。文献[8]引入信息年龄(Age of Information，AoI)的概念，在考虑AoI的条件下研究物联网中VNF的部署问题，提出最小化目的节点的长期AoI均值的优化问题。针对该问题提出一种基于AoI感知的VNF部署算法，寻找最优VNF布局策略。文献[9]研究VNF的自主部署策略，提出了一种自主VNF部署的自适应增强学习VNF性能预测模块，命名为AREL3P。该模块利用端到端服务级别性能预测来部署VNF，并且基于适应预测的特定形式的强化学习，即Q学习，实现部署决策的可移植性，提高VNF部署效率。文献[10]研究考虑动态网络负载下的VNF部署问题，首先，将该问题表述为二进制整数线性规划模型，提出最小化包括VNF部署成本、VNF实例运行成本和拒绝SFC请求惩罚组成的总成本的优化问题。然后，提出了一种基于双深度Q网络的VNF布局算法，该算法包括离线训练和在线运行过程，更好地实现负载均衡。文献[11]研究联合VNF部署和路由问题，提出最小化服务时延和资源消耗成本的加权和的优化问题。之后，将该优化问题分解为部署子任务和路由子任务，每个子任务由多个并行的顺序决策过程组成。最后，设计一种多智能体深度强化学习(Multiagent Deep Reinforcement Learning，MADRL)部署和路由框架和相应算法，在此基础上提出基于参数迁移的模型再训练方法来处理变化的网络拓扑。

然而，上述研究仅考虑业务的端到端时延，并没有考虑仿真过程所需节点和链路的信息是否与物理网络同步。由于5G电力虚拟专用光网络传输距离极长、传输速率极高的特点，若网络中部分节点和链路同步时延过大，仿真使用的节点和链路的数据未及时更新，会给切片策略准确性带来严重影响，无法满足业务QoS需求。

基于上述问题，首先，建立包含网元同步过程的电力虚拟专用光网络切片VNF部署系统模型，在此基础上将其表述为最小化业务同步时延和网元同步时延的优化问题。针对此问题，设计流程如下：选择一个节点作为同步节点，对网络中其余网元信息进行同步，并通过专用特种光纤链路直接连接位于网络边缘的软件定义网络(Software Defined Network，SDN)控制器；将其余节点设置为业务节点，部署业务所需VNF，实现业务SFC映射和资源分配；通过链路和部分节点连接，计算业务时延和同步时延。其次，由于优化变量兼具离散和连续变量，分别使用对应深度强化学习算法进行求解。

1　5G 电力虚拟专用光网络切片模型

1.1　网络场景

实验网络模型如图 1 所示，业务提供商向网络提供商租用物理资源，为电网用户提供不同的网络业务。网络拓扑建模为无向图 $G(N,L)$，其中 $N=\{n_1,n_2,\cdots,n_N\}$ 和 $L=\{l_1,l_2,\cdots,l_N\}$ 分别代表网络节点和链路的集合。节点 $n_i \in N$ 的资源为 $R_i=\{C_i,M_i\}$，其中 C_i 表示节点的计算资源总量，M_i 表示节点的存储资源总量。两物理节点 n_1 和 n_2 通过光纤连接，表示为物理链路 $l_{1,2} \in L$，长度为 $d_{1,2}$。链路资源和状态为 $R_l(t)=\{B_l(t),\eta_l(t),d_l\}$，其中 $B_l(t)$ 表示时隙 t 链路带宽资源，$\eta_l(t)$ 表示链路频谱效率，d_l 为链路长度。

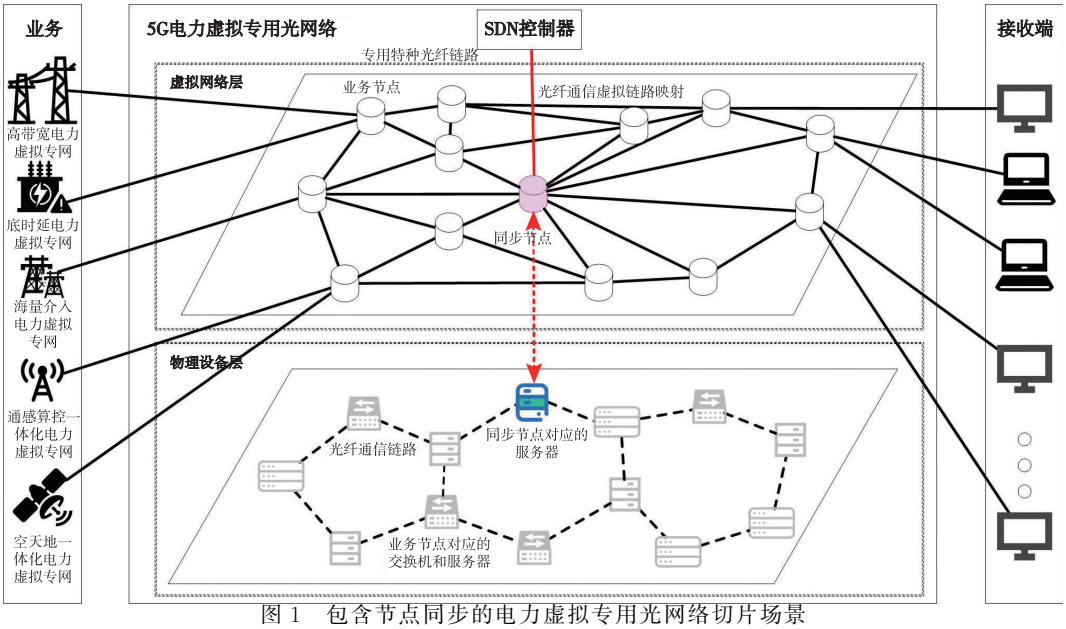


图 1　包含节点同步的电力虚拟专用光网络切片场景

假设时间系统是离散的，其最小间隔为 τ。为方便表述，使用时隙表述时间系统，时隙的集合表述为 $T=\{1,2,\cdots,t_T\}$。时隙 t 网络业务请求集为 $S(t)=\{s_1(t),s_2(t),\cdots,s_s(t)\}$，其中，业务 $s(t) \in S(t)$ 的资源需求以 SFC 的方式表示为集合 $s(t)=\{SFC_s(t),r_s(t),D_s,\tau_s^{\mathrm{m}}\}$，其中 $SFC_s(t)=\{f_{sf} \mid f=1,2,f_s\}$ 表示时隙 t 业务服务功能链，即业务所需功能集合，$r_s(t)$ 表示用户要求的业务最低传输速率，D_s 表示任务处理数据包大小，τ_s^{m} 表示用户允许的最大时延。

节点 $n_i \in N$ 具有两种属性：其一是业务节点，用于部署用户所需 VNF，完成 SFC 映射和资源分配。对于此类节点，假设各节点支持业务 VNF 类型和数量不同，并且可以部署多个业务 VNF，但是假设一个时隙只能完成一种业务的任务处理，即一个时隙只能使用一个 VNF；其二是同步节点，用于实现网络切片的同步。对于此类节点，不部署业务 VNF，不处理业务任务和流量，而是收集网络其余节点和链路的信息，并通过专用特种光纤直接和 SDN 通信，同步网络中各节点和链路的信息。

1.2　同步节点选择模型

为保证网络切片策略的时效性，防止因网络节点链路信息不同步导致的切片策略滞后或错误，首先需要选取网络中的一个节点作为网络同步节点。在其余网络切片研究中，SDN 控制器的部署位置通常被忽略，并且研究中 SDN 控制器掌握所有节点和链路的实时变化信息。然而，在现实网络切片中，SDN 控制器部署于网络边缘，通过光纤与网络中少量节点和链路连接，因此，SDN 收集到的节点和链路信息不是实时的，即距离 SDN 连接节点越近的网络节点信息同步时延越低，SDN 获得节点数据越准确，而部分距离 SDN 连接节点较远的节点信息同步时间越长，信息同步不及时。当距离 SDN 连接节点较远的节点执行任务计算时，节点信息同步不及时容易导致切片资源分配超过节点剩余资源，甚至节点出现故障时仍为其分配计算任务，严重影响计算质量。为解决此问题，首先对同步节点进行选择，并使其余节点和链路向其发送同步信息。通过选取适当位置的同步节点降低网络最大节点同步时延，保证策略的有效性。

假设实验中 SDN 只能选择一个节点通过虚拟光纤链路直接连接，在此基础上同步节点选择变量为 $\boldsymbol{\Theta}(t)=\{\theta_1(t),\theta_2(t),\cdots,\theta_N(t)\}$，其中二进制变量 $\theta_n(t)$表示节点 $n_i \in N$ 的类型，当 $\theta_n(t)=0$ 时表示节点 $n_i \in N$ 是业务节点，执行业务 VNF 部署和资源分配；当 $\theta_n(t)=1$ 时表示节点 $n_i \in N$ 是同步节点，同步网络中剩余节点的信息。在此基础上，网络同步节点的部署需要满足以下约束：

$$\sum_{n_i \in N}\theta_n(t)=1 \tag{1}$$

该约束具备两层含义，其一为网络中必须有同步节点，其二是网络中同步节点的数量不能超过一个。在此基础上，所有节点和链路通过最短路径到达同步节点，并传输同步信息，将上述过程称为关联过程，具体建模如下。

假设各节点和链路向同步节点发送数据包大小分别为 d_n 和 d_l，数据包内容分别为节点部署的 VNF 数量、类型、剩余计算资源、存储资源以及链路剩余带宽资源，同步数据包大小远小于业务传输数据包。在此基础上，通过查找路由表，节点和链路将同步数据包按照最短路径依次传输至同步节点处，假设节点 $n_i \in N$ 和链路 $l_1 \in L$ 传输路径分别为 $G_i^N=\{l_j,j=1,2,\cdots,J,\forall l_j \in L\}$ 和 $G_1^L=\{l_k,k=1,2,\cdots,K,\forall l_k \in L\}$，由于数据包很小，且传输过程中不需要进行处理，不会造成链路堵塞。因此，影响同步的主要因素是链路的传播时延，假设同步时延不超过 τ_d^{m} 时，则认为策略实时性得到保障。在此基础上，假设业务资源分配和网络同步资源分配过程不相互影响。

1.3 业务 VNF 部署和资源分配模型

1.3.1 业务 VNF 部署模型

VNF 部署模型是 SFC 映射的重要阶段，时隙 t 的各节点 VNF 的部署情况使用二进制变量向量 $\boldsymbol{P}(t)$表示，其中变量 $p_{s,f}^i(t)=1$ 表示业务 s 的 SFC 中第 f 个 VNFf_{sf} 映射到节点 n_i。向量 $\boldsymbol{P}(t)$如式(2)所示：

$$\begin{aligned}\boldsymbol{P}(t)=[&p_{1,1}^1(t),p_{1,2}^1(t),\cdots,p_{1,f_1}^1(t),p_{1,1}^{n_N}(t),\\&\cdots,p_{1,f_1}^{n_N}(t),\cdots,p_{s_S,f_{s_S}}^{n_N}(t)]\end{aligned} \tag{2}$$

在此基础上，为了保证业务的服务质量以及网络正常运行，约束式(3)需要得到保证：

$$\sum_{n_i \in N}\sum_{f \in SFC_s}p_{s,f}^i(t)=1 \tag{3}$$

约束式(3)表示业务 s 的 SFC 中任意 VNFf_{sf} 至多部署在一个节点上，该约束可以防止 SFC 所需节点重复部署，避免造成资源浪费。

约束式(4)表示节点部署的 VNF 应该为其支持的 VNF 种类。

$$\sum_{s \in S}\sum_{f_{sf} \notin VNF_i}p_{s,f}^i(t)=0 \tag{4}$$

约束式(5)节点 n_i 部署的 VNF 不应超过节点 VNF 部署数量上限 F_i。

$$\sum_{s \in S}\sum_{f \in SFC_s}p_{s,f}^i(t) \leqslant F_i \tag{5}$$

1.3.2 资源分配模型

VNF 部署完成后，节点和链路应根据业务需求分配节点计算资源、存储资源以及链路的带宽资源。对于节点 n_i，时隙 t 为业务 s 分配的资源使用集合 $r_i(t)=\{c_{if}^s(t),m_{if}^s(t)\}$ 表示，其中 $c_{if}^s(t)$表示节点分配的计算资源，$m_{if}^s(t)$表示节点分配的存储资源。链路 l_1 分配的资源使用集合 $r_l(t)=\{b_l(t)\}$ 表示，其中，$b_l(t)$表示时隙 t 链路分配的带宽资源，各资源集合分别为 $C(t),M(t),B(t)$。资源分配过程应该满足以下约束条件，以保证业务服务质量和网络资源正常运行。

$$\sum_{s \in S}\sum_{f_{if} \in VNF_i}c_{if}^s(t) \leqslant C_i \tag{6}$$

$$\sum_{s \in S}\sum_{f_{if} \in VNF_i}m_{if}^s(t) \leqslant M_i \tag{7}$$

$$\sum_{s \in S}\sum_{f_{if} \in VNF_i}\sum_{f_{i'f'} \in VNF_{i'}}b_{if,i'f'}^s(t)+\sum_{s \in S}\sum_{f_{if} \in VNF_i}\cdot\sum_{f_{i'f'} \in VNF_{i'}}b_{i'f',if}^s(t) \leqslant B_{ii'} \tag{8}$$

约束式(6)～(8)分别表示节点分配的计算、存储资源以及链路的带宽资源不能超过该节点和链路的总资源。同时，为了保证业务的最低速率需求，约束式(9)也需要满足。

$$\sum_{f_{if} \in VNF_i}\sum_{f_{i'f'} \in VNF_i'}b_{if,i'f'}^s(t)\cdot\eta_{ii'}(t) \geqslant r_s(t) \tag{9}$$

1.4 时延建模和优化问题表述

满足上述约束条件，可以保证业务正常传输和网络正常运行。在此基础上，本节描述各业务和网络中同步的性能指标。首先，建立业务时延模型评价业务 VNF 部署和资源分配方案的有效性；其次，建立同步时延模型对同步性能进行量化；最后，将上述模型表述为满足约束服务质量的同时最小化切片业务时延和同步时延的优化问题。

1.4.1 业务时延模型

业务时延是业务服务质量的重要指标之一，因

此保证业务时延不大于业务最大时延是保证业务服务质量的关键之一。由于业务以 SFC 的形式映射在网路中，业务时延为经过所有 VNF 和链路时延的总和。首先，假设当前 VNF 需要进行任务处理，定义业务 s 从 VNFf_{sf} 到 VNFf_{sf+1} 的时延为：

$$\tau_s^{f,f+1}(t)=\tau_s^p(t)+\tau_s^c(t) \tag{10}$$

式中，$\tau_s^p(t)$和 $\tau_s^c(t)$分别表示业务的任务处理时延和传输传播时延，计算方式为：

$$\tau_s^p(t)=p_{s,f}^i(t)\cdot\left[\frac{D_s(t)\cdot c_i}{\sum_{f\in VNF_i}c_{if}^s(t)\cdot r_s(t)}+\frac{D_s(t)\cdot m_i}{\sum_{f\in VNF_i}m_{if}^s(t)\cdot r_s(t)}\right] \tag{11}$$

$$\tau_s^c(t)=p_{s,f}^i(t)\cdot p_{s,f+1}^{i'}(t)\cdot\left[\frac{r_s(t)}{\eta_{ii'}(t)\cdot\sum_{f_{if}\in VNF_i}\sum_{f_{i'f'}\in VNF_{i'}}b_{if,i'f'}^s(t)}+\frac{d_{ii'}}{c}\right] \tag{12}$$

其中，任务处理时延为节点执行计算任务和存储任务的时延之和；传输传播时延为业务流量经过链路的时延总和。则业务 s 的总时延计算式为：

$$\tau_s(t)=\sum_{f_{sf}\in SFC_s}\tau_s^{f,f+1}(t) \tag{13}$$

为了保证业务的服务质量，必须保证业务时延不超过业务最大时延限制，即：

$$\tau_s(t)\leqslant\tau_s^{\mathrm{m}} \tag{14}$$

1.4.2　网络同步时延模型

满足业务服务质量的同时，实验目标的一部分为最小化同步时延。同步时延如式(15)所示。

$$\tau_{\mathrm{D}}(t)=\max_{n_i\in N}\sum_{l_{ii'}\in G_1^L}\sum_{n_i\in G_i^N}\left(\frac{d_{\mathrm{n}}+d_{\mathrm{l}}}{B_{ii'}(t)\cdot\eta_{ii'}(t)}+\frac{d_{ii'}}{c}\right)+\frac{d_{\mathrm{SDN}}}{c}+\sum_{n_i\in N}\theta_n(t)\cdot(1-\theta_{n-1}(t-1))\cdot\frac{\tau_{\mathrm{D}}(t)\cdot(t-1)}{c} \tag{15}$$

该时延的物理含义由网络中各节点和链路将信息发送至同步节点所需时延的最大值(第一项)、同步节点传输数据至 SDN 控制器所需时延(第二项)以及迁移时延(第三项)之和决定。迁移时延是同步节点发生迁移时所产生的时延，近似表示为前$(t-1)$个时隙同步时延之和。近似时延的触发条件为时隙$(t-1)$和时隙 t 同步节点分别部署于网络不同位置，即同步节点发生了迁移。具体而言，如果两个时隙同步节点都位于节点 n_i，则对于该节点有 $\theta_n(t)\cdot(1-\theta_{n-1}(t-1))=1\times0=0$，对于其余节点有 $\theta_n(t)\cdot(1-\theta_{n-1}(t-1))=0\times1=0$，不存在迁移时延；如果两个时隙同步节点由 n_i 迁移至 n_j，对于节点 $\theta_n(t)\cdot(1-\theta_{n-1}(t-1))=1\times1=1$，对于节点 n_i，$\theta_n(t)\cdot(1-\theta_{n-1}(t-1))=0\times0=0$，对于其他节点，$\theta_n(t)\cdot(1-\theta_{n-1}(t-1))=0\times1=0$。因此，存在迁移成本，且迁移成本远大于部署成本。为了保证网络中同步时延的最小化，有必要尽量避免同步节点的迁移。在此基础上，假设同步时延只有满足不超过最大时延 τ_d^m 的约束，才能保证策略的实时性。

$$\tau_{\mathrm{D}}(t)\leqslant\tau_d^{\mathrm{m}} \tag{16}$$

1.4.3　优化问题表述

实验优化目标为保证用户服务质量的同时最小化业务时延和同步时延，因此，优化目标为：

$$\max-\tau(t)=-\left(\alpha\cdot\sum_{s\in S}\tau_s(t)+\beta\cdot\tau_{\mathrm{D}}(t)\right)$$
$$\text{s.t.}\quad (1),(3)\text{-}(9),(14) \tag{17}$$

其中，优化变量包括 $\boldsymbol{\Theta}(t)$，$\boldsymbol{P}(t)$，$C(t)$，$M(t)$，$B(t)$。

2　基于深度强化学习的电力虚拟专用光网络切片同步算法

2.1　状态、动作和奖励设计

由于优化问题(17)是一个 NP 难和非凸问题，启发式算法不能求解。因此，考虑使用深度强化学习算法对上述优化问题进行近似和迭代求解。首先，将上述问题建模为马尔可夫决策过程模型[12]，使用状态、动作和奖励三元组表示：

状态空间设计为：

$$S(t)=\{\boldsymbol{\Theta}(t-1),\boldsymbol{P}(t-1),C_i(t),M_i(t),B_{\mathrm{l}}(t),s(t)\} \tag{18}$$

式中，$\boldsymbol{\Theta}(t-1)$，$\boldsymbol{P}(t-1)$表示上一时隙网络的基本情况，包括同步节点部署情况和业务 VNF 部署情况；$C_i(t)$，$M_i(t)$和 $B_{\mathrm{l}}(t)$是时隙 t 节点剩余计算资源、存储资源和链路剩余带宽资源。如果时隙 t 计算任务、存储任务和链路传输能够完成，时隙 t 资源数量等于节点和链路最大资源数量，否则为总资源数量减去时隙 t 未完成任务消耗的资源数量；业务请求集 $s(t)=\{SFC_s(t),r_s(t),D_s,\tau_s^{\mathrm{m}}\}$ 决定到达业务的种类、最低速率、任务处理数据包大小和最大时延要求。

动作设计为：

$$A(t)=\{\boldsymbol{\Theta}(t),\boldsymbol{P}(t),C(t),M(t),B(t)\} \tag{19}$$

$A(t)$代表网络中同步节点和业务 VNF 的部署情况，以及各节点和链路为各业务分配的计算资源、存储资源和带宽资源。

奖励设计为：

$$R(t)=-\tau(t)=-\left(\alpha\cdot\sum_{s\in S}\tau_s(t)+\beta\cdot\tau_{\mathrm{D}}(t)\right) \tag{20}$$

当且仅当所有约束条件得到满足，求解方案有效，否则需要重新求解优化问题获取求解方案。

2.2 基于优势演员-评论家的同步节点选择和业务 VNF 部署算法

由于同步节点选择和业务 VNF 部署相关问题具有离散变量的优化变量，因此，动作空间和状态空间也是离散的。基于上述思路，提出了基于优势演员-评论家（Advantage Actor-Critic，A2C）[13] 的同步节点选择和业务 VNF 部署算法来求解上述优化问题。

首先，考虑到同步节点选择和业务 VNF 部署过程具有关联性，算法首先需要求解同步节点，确定其位置后才能在剩余节点中部署业务 VNF，即变量 $\boldsymbol{\Theta}(t)$ 和 $\boldsymbol{P}(t)$ 的求解具有先后性，因此，提出具有双重 A2C 网络结构的同步节点选择和业务 VNF 部署算法。

如图 2 所示，环境首先将状态发送给第 1 个 A2C 网络，该网络将状态输入策略网络中获取网络的同步策略 $\boldsymbol{\Theta}(t)$，将状态 $s(t)$ 和同步策略 $\boldsymbol{\Theta}(t)$ 发送给第二个 A2C 网络。网络首先根据同步策略 $\boldsymbol{\Theta}(t)$ 识别同步节点和业务节点，并通过剩余节点确定 VNF 部署策略 $\boldsymbol{P}(t)$，然后整合所有状态和动作，存入回放缓存池中。当算法进入更新阶段时，从回放缓存中抽样 minibatch 数量的样本，分别发送给两个 A2C 网络，两个网络分别通过状态和相关动作计算 TD 误差和损失函数，并更新自身策略网络和价值网络。

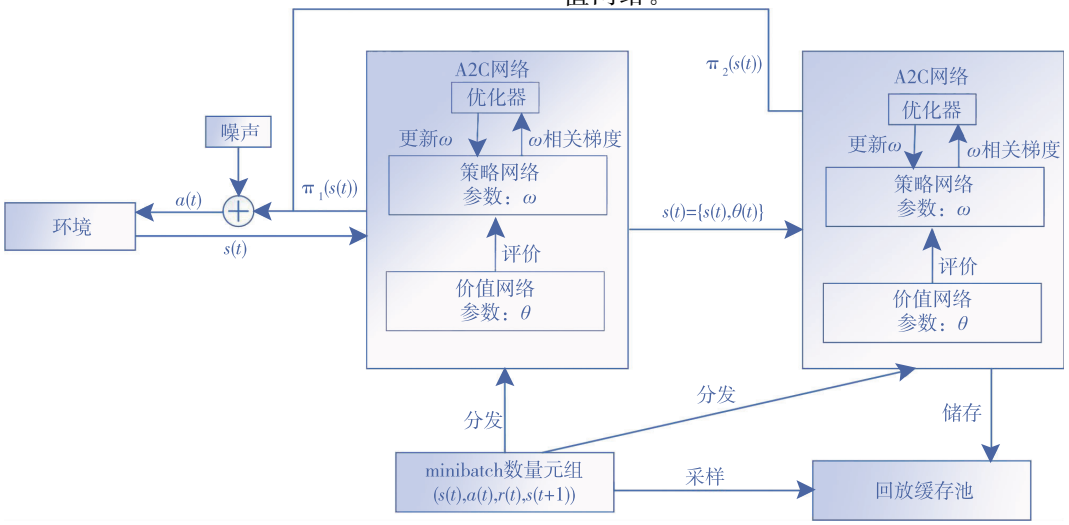


图 2 基于 A2C 的同步节点选择和业务 VNF 部署算法

2.3 基于深度确定性策略梯度的资源分配算法

通过基于 A2C 的同步节点选择和业务 VNF 部署算法获得节点同步策略和业务 VNF 部署策略后，需要根据业务需求完成资源分配。由于资源分配变量都是连续的变量，无法和上述变量联合求解。因此，本节提出了基于深度确定性策略梯度（Deep Deterministic Policy Gradient，DDPG）[14] 的资源分配算法，如图 3 所示。

如图 3 所示，DDPG 算法由两个独立的神经网络组成，即策略网络和评论家网络，网络参数分别为 ω 和 θ。策略网络通过确定性策略 $\pi(a(t)\mid s(t);\omega(t))$ 输出连续动作，评论家网络通过评估策略网络输出动作。由于很难获得 $\pi(a(t)\mid s(t);\omega(t))$ 和 $Q(s(t),a(t);\theta(t))$ 的显性表达，策略网络输出近似为确定性策略 $\pi(a(t)\mid s(t);\omega(t))$，评论家网络输出近似为值函数 $Q(s(t),a(t);\theta(t))$。对于当前状态 $s(t)$，策略网络基于确定性策略 $\pi(a(t)\mid s(t);\omega(t))$ 生成动作 $a(t)$。评论家基于估计的 Q 值 $Q(s(t),a(t);\theta(t))$ 对损失函数进行计算，表述式为：

$$\Delta\theta=\alpha_{\mathrm{c}}\left[R(t)+\gamma\max_{a(t+1)}Q(s(t+1),a(t+1);\tilde{\theta})-Q(s(t),a(t);\theta)\right]\cdot\nabla_{\theta}Q(s(t),a(t);\theta) \tag{21}$$

式中，$Q(s(t),a(t);\theta(t))=E_{a(t)\sim\pi(a(t)\mid s(t);\omega(t))}[R$

$(t)|s(t),a(t)]$，并基于确定性策略 $\pi(a(t)|s(t);\omega(t))$，对动作 $a(t)$ 进行预期。策略网络的损失函数定义为 $-J(\omega(t))=-E[Q(s(t),a(t);\theta(t))]$。评论家网络使用策略梯度方法，即计算梯度 $\nabla_{\omega}J(\omega(t))=E[\nabla_{\omega}\log\pi(a(t)|s(t);\omega(t))Q(s(t),a(t);\theta(t))]$，并使用梯度下降更新策略网络，即 $\omega(t):\omega(t)\leftarrow\omega(t)-\alpha_p(-\nabla_{\omega}J(\omega(t)))$，$\alpha_p$ 表示策略网络的学习率。由于 θ 的损失函数连续可微，损失函数的梯度可用于更新评论家网络的参数 θ，如式(22)所示。其中，α_c 表示评论家学习率。在网络更新中，使用 minibatch 数量的样本更新参数 θ，如式(22)所示：

$$\Delta\theta=\frac{\alpha_c}{minibatch}\sum_{i=1}^{minibatch}\left[R^i(t)+\gamma\max_{a^i(i+1)}Q\left(s^i(t+1),a^i(t+1);\bar{\theta}\right)-Q(s^i(t),a^i(t);\theta)\right]\cdot\nabla_{\theta}Q(s^i(t),a^i(t);\theta)\cdot \tag{22}$$

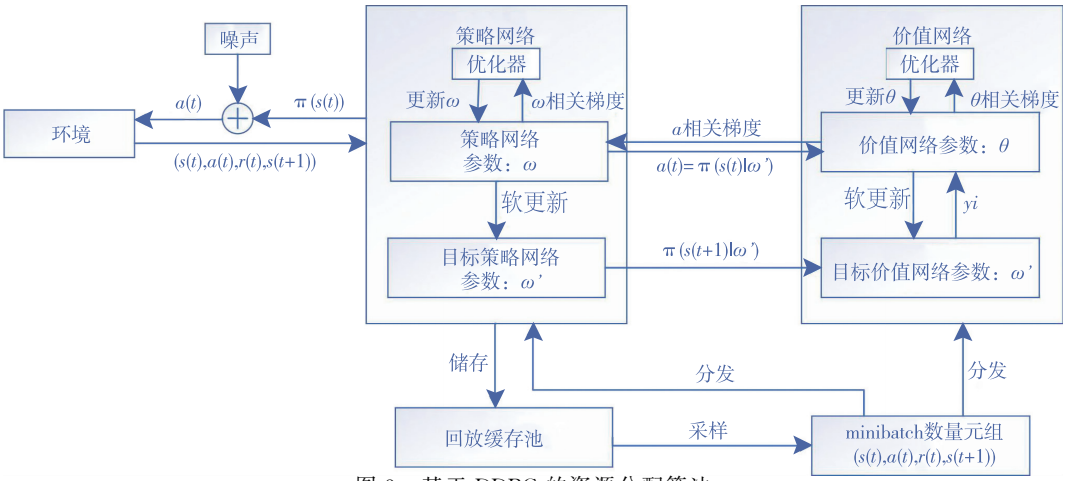


图 3　基于 DDPG 的资源分配算法

2.4　基于深度强化学习的电力虚拟专用光网络切片同步算法

综上考虑，将 2.2 节所提的基于 A2C 的同步节点选择、业务 VNF 部署算法和基于 DDPG 的资源分配算法整合，可得基于深度强化学习的电力虚拟专用光网络切片同步算法。具体流程如表 1 所示。

表 1　基于深度强化学习的电力虚拟专用光网络切片同步算法

01：初始化基于 A2C 的同步节点选择和业务 VNF 部署算法网络参数 $\theta^{Q_p^{11}}$，$\theta^{Q_p^{12}}$，动作网络 $\theta^{\mu_p^{11}}$，$\theta^{\mu_p^{12}}$

02：初始化基于 DDPG 的资源分配算法网络参数 $\theta^{Q_p^{21}}$，$\theta^{Q_p^{22}}$，动作网络 $\theta^{\mu_p^{21}}$，$\theta^{\mu_p^{22}}$

03：for 迭代次数 $k=1,2,\cdots,k_{max}$ do

04：获取初始化状态 $s(1)$

05：for 时隙状态 $t=1,2,\cdots,T$ do

06：随机获取概率 p_1，p_2，根据概率 p_1 确定基于 A2C 的同步节点选择和业务 VNF 部署算法第一个 A2C 网络根据状态 $s(t)$ 获取动作 $a(t)$ 或随机选取行动

07：整合第一个 A2C 网络的节点同步策略和状态，根据概率 p_2 确定第二个 A2C 网络根据整合状态 $s(t)$ 获取动作 $a(t)$ 或随机选取行动

08：整合基于 A2C 的同步节点选择和业务 VNF 部署算法状态和动作，输入 DDPG 网络，策略网络根据整合状态 $s(t)$ 获取动作并加入噪声，获得奖励，将三元组存入回放缓存池中，进入下一状态

09：end for

10：从回放缓存池中抽取 minibatch 数量的样本，分发给基于 A2C 的同步节点选择和业务 VNF 部署算法和基于 DDPG 的资源分配算法

11：各算法根据样本数据更新自身网络参数

3 仿真结果与分析

在 Tensorflow 平台上对所提出的算法进行仿真验证，评估该算法的性能。将所提出的节点同步算法与随机节点同步算法（随机选取网络中的一个节点用于同步）进行对比，评估所提出的同步方案对切片策略和实时性的影响。各业务的具体要求和网络仿真参数如表 2 所示[15]。

表 2 仿真参数设置

仿真参数	仿真值
网络节点、链路和 VNF 数量	10，40，36
节点最大计算资源/GHz	2～8
节点最大存储资源/TB	2～8
链路最大带宽资源/Mbps	1 000～10 000
链路长度/km	50～250
业务数量	3
业务最大时延限制/ms	100
网络同步最大时延/ms	3
总时隙 T	16
策略网络和评论家网络学习率	0.003，0.003
折扣系数 γ	0.98
样本数量 minibatch	16
最大迭代次数	16 000
噪声类型	OU 噪声
激活函数	Softmax，Relu 和 tanh
单层网络神经元数量	256

图 4 对比了所提出的节点同步算法与随机选择同步节点的网络切片算法的奖励。从图中可以看出，所提出的节点同步算法拥有更高的奖励，即更低的业务时延和同步时延。因此，所提出的节点同步算法拥有更好的综合性能，能够获得更低的总时延，即能够更有效地保障业务的时延和业务的实时性，防止网络状态同步不及时造成业务策略滞后。

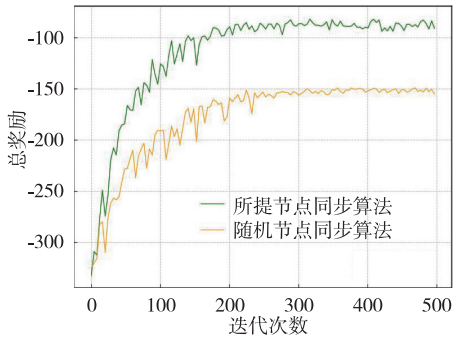


图 4 不同同步方案的总奖励对比

图 5 对比了所提出的节点同步算法和随机节点同步算法的业务时延。所提出的算法和随机节点同步算法业务时延相差不大，都能满足业务最大时延需求的同时获得较好的收敛效果，因此，所提出的算法增加的节点同步过程并不会对业务造成较大的影响，是一种保证业务服务质量同时提高业务实时性的有效方案。

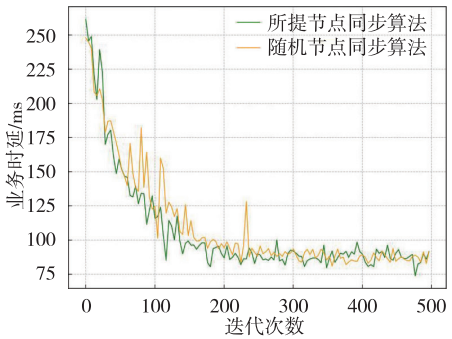


图 5 不同同步方案的业务时延

图 6 对比了所提出的节点同步算法和随机节点同步算法的同步时延。所提出的算法通过网络同步过程，选取较合理的节点作为同步节点，明显降低了网络中各节点和各链路的最大同步时延，保障了切片策略的实时性。另一方面，随机节点同步策略不满足公式(16)的网络同步最大时延，因此，策略实时性不能保证，极易导致切片策略滞后和错误；相反，通过选取适当节点进行同步，能够将同步时延降低至网络同步最大时延以内，保证切片策略的实时性。

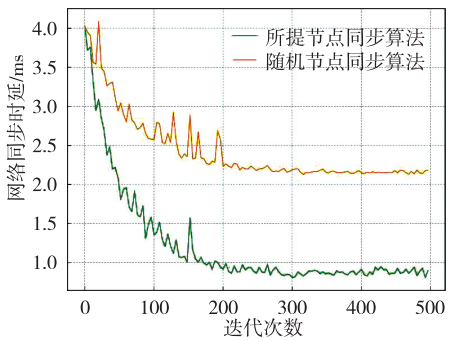


图 6 不同同步方案的网络节点同步奖励对比

综上所述，所提出的基于深度强化学习的电力虚拟专用光网络切片同步算法能够通过合理地同步节点选择降低网络各节点和链路的最大同步时延，保证策略的实时性和有效性，同时不会对业务 VNF 部署和资源分配造成较大影响，能够满足业务服务需求。

4 结束语

针对电力虚拟专用光网络切片中策略仿真时延

较大，导致仿真使用网络节点和链路状态更新不及时，严重影响仿真策略的准确性，无法满足业务 QoS 的问题，本文提出一种基于深度强化学习的电力虚拟专用光网络切片同步算法。首先，针对电网场景进行描述，对业务 VNF 部署和节点同步模型进行构建，在此基础上构建 VNF 部署模型和资源分配模型，描述节点同步和业务 VNF 部署过程。基于上述模型，本文构建资源分配模型，建立业务时延和同步时延模型，将其表述为保证用户服务质量的同时最小化业务时延和同步时延的优化问题。由于问题具有 NP 难和非凸性，采用深度强化学习算法进行求解，分别设计基于优势演员-评论家的同步节点选择和业务 VNF 部署算法和基于深度强化学习的电力虚拟专用光网络切片同步算法求解上述优化问题。仿真结果表明，所提出的节点同步算法能够有效改进电力虚拟专用光网络切片网络中各节点和链路的最大同步时延，同时节点同步不会对业务服务质量产生较大影响，因此，所提方案能够保证切片策略的实时性和有效性。

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, 2021, 64: 110301.

[2] 刘　珊，黄　蓉，王友祥. 全球 6G 研究发展综述[J]. 邮电设计技术，2021(3): 16-20.
Liu Shan, Huang Rong, Wang Youxiang. Overview of global 6G research and development[J]. Designing Techniques of Posts and Telecommunications, 2021(3): 16-20.

[3] 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.

[4] 孔凌辉，饶哲恒，徐彦彦，等. 基于深度强化学习的无线网络智能路由算法[J]. 计算机工程，2023，49(9): 199-207，216.
Kong Linghui, Rao Zheheng, Xu Yanyan, et al. Intelligent routing algorithm for wireless networks based on deep reinforcement learning[J]. Computer Engineering, 2023, 49(9): 199-207, 216.

[5] Chen H, Wang S, Li G, et al. Distributed orchestration of service function chains for edge intelligence in the industrial internet of things[J]. IEEE Trans. on Industrial Informatics, 2021, 18(9): 6244-6254.

[6] Ding L, Lin Z, Shi X, et al. Target-value-competition-based multi-agent deep reinforcement learning algorithm for distributed nonconvex economic dispatch[J]. IEEE Trans. on Power Systems, 2022, 38(1): 204-217.

[7] Sun P, Lan J, Li J, et al. Combining deep reinforcement learning with graph neural networks for optimal VNF placement[J]. IEEE Communications Letters, 2020, 25(1): 176-180.

[8] Chen Z, Li H, Ota K, et al. Deep reinforcement learning for AoI aware VNF placement in multiple source systems[C]// 2022 IEEE Global Communications Conference, 2022: 2873-2878.

[9] Bunyakitanon M, Vasilakos X, Nejabati R, et al. End-to-end performance-based autonomous VNF placement with adopted reinforcement learning[J]. IEEE Trans. on Cognitive Communications and Networking, 2020, 6(2): 534-547.

[10] Pei J, Hong P, Pan M, et al. Optimal VNF placement via deep reinforcement learning in SDN/NFV-enabled networks[J]. IEEE J. on Sel. Areas in Communications, 2019, 38(2): 263-278.

[11] Wang S, Yuen C, Ni W, et al. Multiagent deep reinforcement learning for cost-and-delay-sensitive virtual network function placement and routing[J]. IEEE Trans. on Communications, 2022, 70(8): 5208-5224.

[12] Tan X, Zhou L, Wang H, et al. Cooperative multi-agent reinforcement-learning-based distributed dynamic spectrum access in cognitive radio networks[J]. IEEE Internet of Things J., 2022, 9(19): 19477-19488.

[13] Chen Z, Hu J, Min G, et al. Adaptive and efficient resource allocation in cloud datacenters using actor-critic deep reinforcement learning[J]. IEEE Trans. on Parallel and Distributed Systems, 2021, 33(8): 1911-1923.

[14] Cui Y, Huang X, He P, et al. QoS guaranteed network slicing orchestration for internet of vehicles[J]. IEEE Internet of Things J., 2022, 9(16): 15215-15227.

[15] Gu L, Zeng D, Li W, et al. Intelligent VNF orchestration and flow scheduling via model-assisted deep reinforcement learning[J]. IEEE J. on Sel. Areas in Communications, 2019, 38(2): 279-291.

作者简介：

张　导(1983—)，男，重庆市人，高级工程师，研究方向为电力 OTN/SDH 传输、程控/软交换等电力通信系统；

周之楠(1990—)，女，重庆市人，研究方向为视频会议通信系统。