光电技术及应用

DOI：10.16818/j.issn1001－5868.2023061303

5G电力虚拟专网和数字孪生相结合的系统架构和构建

吴维农[1*]，莫　婷[1]，赵妍妍[1]，谢秋洪[1]，付泉泳[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 通信与信息工程学院，重庆 400065）

摘　要： 设计了一种5G电力虚拟专网和数字孪生相结合的系统架构，并考虑了终端数字孪生时同步的构建方法，提出了一种分层多智能强化学习算法。该算法决定分配给电力用户终端的资源和数字孪生体的放置，上层通过深度Q网络实现切片通信资源的分配，下层通过多智能体深度强化学习实现电力用户终端的数字孪生体放置。实验结果表明，所提出的分层多智能体算法在电力用户终端数字孪生体的同步强度上能获得较好的系统收益。

关键词： 5G电力虚拟专网；数字孪生；深度强化学习

中图分类号： TP3；TM73　**文章编号：** 1001－5868(2023)05－0803－08

Research on the System Architecture and Construction Method of Combining 5G Power Virtual Private Network and Digital Twin

WU Weinong[1], MO Ting[1], ZHAO Yanyan[1], XIE Qiuhong[1], FU Quanyong[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: A system architecture that combined 5G power virtual private network and digital twins was designed in this paper, and the construction method of synchronization during the digital twins of terminals was considered. Then, a hierarchical multi-agent reinforcement learning algorithm was proposed to determine the allocation of resources to power user terminals and the placement of digital twins. The upper layer achieved the allocation of sliced communication resources through deep Q network, and the lower level achieved the placement of digital twins in power user terminals through multi-agent deep reinforcement learning. The experimental results show that the proposed hierarchical multi-agent algorithm can achieve better system benefits when the synchronization strength of the digital twin of the power user terminal is reached.

Key words: 5G power virtual private network; digital twin; deep reinforcement learning

0　引言

5G电力虚拟专网是指在5G网络中，通过软件定义网络(Software Defined Network，SDN)技术将物理设施划分为多个逻辑隔离的软切片，此外还可通过物理隔离形成硬切片，进而在无线接入网、传输网和核心网虚拟出面向电力行业的专用网络[1]。

5G电力虚拟专网包含两种类型虚拟专网：局域虚拟专网和广域虚拟专网[2]。其中局域虚拟专网主要面向封闭场景的局域任务，通过5G电力虚拟专网和基站、移动边缘计算实现电力业务的需求。广域虚拟专网面向开放场景的任务，通常通过电力切

收稿日期：2023－06－13.

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：**吴维农　**E-mail：**wuweinong2023@163.com

片专网实现端到端的差异化服务。然而5G电力虚拟专网的构建前提是能实时获取网络中网元和电力业务终端(以下简称终端)运行状况,数字孪生技术近年来在工业界得到成功应用[3],数字孪生电网是基于数字孪生基础设施的电网数字化转型方法,通过在数字空间建设电网、环境、人员和业务四要素,实现数据驱动的全局、全生命周期的数字孪生体,从而达到不断改善的应用目的[4]。尽管数字孪生技术已在电网中得到研究,但目前并没有将5G电力虚拟专网与数字孪生相结合的架构。通过在数字空间建设5G电力虚拟专网和终端的实时镜像,可实现数据驱动的全局、全生命周期的数字孪生体,从而实现实时的5G电力虚拟专网管理[5]。

另一方面,对于源侧的5G电力虚拟专网,由于无线信道的动态变化性、无线接入网拥塞情况以及终端数字孪生体的放置位置,终端的数字孪生体同步会受到影响,进而可能导致构建的5G电力虚拟专网资源管理出现偏差。因此终端的数字孪生体构建需保证其要求的同步强度。目前对数字孪生体的同步研究文献较少,文献[6]提出了一种物联网终端的数字孪生体放置和迁移方案,旨在最小化物联网终端数字孪生体的同步时间,并采用深度确定性策略梯度算法和迁移学习实现数字孪生体的数字孪生体放置和迁移,然而在实际应用中,物联网终端在上传数字孪生体更新数据时并不能保证有足够的通信资源。作者在文献[7]中提出了一种数字孪生体同步方案,然而这种方案是在元宇宙的基础上实现的,即作者的目标是最大限度减少真实世界和分布式元宇宙之间的同步时间,同时满足数字孪生体的同步强度要求,然后作者提出了一种基于最优传输理论的方案解决元宇宙的区域划分和计算资源的分配问题。为终端构建数字孪生体时,考虑到需实时捕获其运行状态,因此其数字孪生体通常放置在边缘服务器上[8],以此减少数字孪生体同步时延。

综上所述,为解决源侧的5G电力虚拟专网资源管理可能出现的偏差,本文首先就5G电力虚拟专网和数字孪生相结合的系统架构进行研究,构建了数字孪生使能的5G电力虚拟专网管理架构,并为了保证5G电力虚拟专网能及时获取网络中实体的实时信息,为终端构建数字孪生体并保证其数字孪生体同步时延,提出一种基于深度Q网络(Deep Q Network,DQN)的分层多智能体算法来求解。

1 系统模型

1.1 数字孪生使能的5G电力虚拟专网管理框架

如图1所示,数字孪生使能的5G电力虚拟专网管理框架主要包含3部分:物理网络、5G电力虚拟专网和数字孪生体,其中数字孪生体又包含物理网络孪生、5G电力虚拟专网孪生和孪生体管理。物理网络孪生通过采集网元、终端等实体信息,进而构建物理网络的数字模型,而5G电力虚拟专网孪生在构建的时候需要融合当前物理网络孪生和已有虚拟专网,因为物理网络孪生的变化会导致源侧虚拟专网资源变化。而5G电力虚拟专网孪生模型的变化可引起5G电力虚拟专网的重配置,然后重配置操作下发到5G电力虚拟专网。

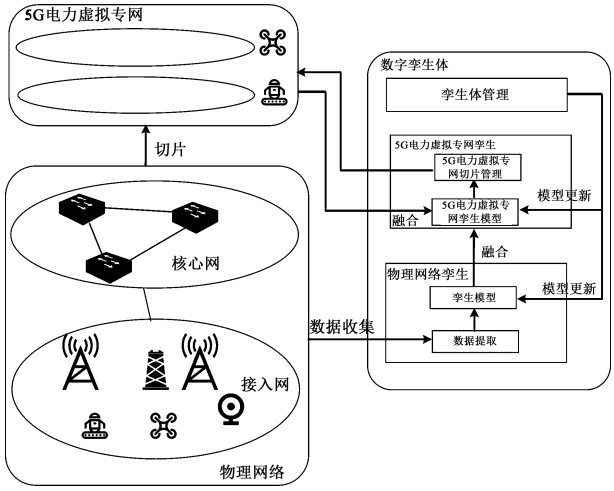


图1 数字孪生使能的5G电力虚拟专网管理框架

孪生体管理用于更新物理网络孪生模型和5G电力虚拟专网模型。

由上述分析可知,物理网络孪生是5G源侧电力虚拟专网构建的重要前提,因此本文基于数字孪生使能的5G电力虚拟专网管理框架,并重点聚焦在物理网络中无线接入网侧的终端数字孪生模型的构建,以保证在终端数字孪生体同步强度的前提下,最大化系统收益。

1.2 同步时延模型

如图1所示,数字孪生使能的5G电力虚拟专网管理框架中的物理网络包含基站、多个终端和边缘服务器等物理实体。

令无线接入网侧需构建数字孪生体的终端 $u \in Y=\{1,2,\cdots,U\}$,边缘服务器 $e \in E=\{1,2,\cdots,G\}$,其中,Y,E 分别表示终端和边缘服务器的有限集合。

为保证图 1 中终端在边缘服务器中的数字孪生体实更新时有充足的通信资源，我们为终端定制切片，令切片集合为 $N=\{1,2,\cdots,N\}$，$n\in N$ 表示第 n 个切片。令二进制指示变量 $\beta_{u,n}=1$ 表示终端 u 通过切片 n 获得通信资源。因此，切片 n 中的终端到基站可获得的平均上行传输速率可表示为：

$$r_n=x_nB\log\left(1+\frac{p_ug}{\sigma^2}\right) \tag{1}$$

式中，x_n 表示切片控制器分配给切片 n 的通信资源比例，B 为基站预留给数字孪生体数据上传的最大带宽，p_u 表示终端的发射功率，g 为终端到基站的信道增益。

对于同类型的终端，其数字孪生体更新的包大小相似，故假设接入切片 n 内的终端 u 更新的数据包大小为 l_n 且服从到达率为 $\lambda_{u,n}$ 的泊松过程，则切片 n 中需要更新的数据包总到达率为：$\lambda_n=\sum_{u=1}^{U}\beta_{u,n}\lambda_{u,n}$，将切片 n 中数据包建模为一个 $M/M/1$ 队列[9]，则根据利特尔定理可知切片 n 中终端上传数字孪生体更新数据的平均时延为：

$$t_{\text{com}}^{u}=\frac{1}{r_n/l_n-\lambda_n} \tag{2}$$

终端将更新的数字孪生体数据传输到基站后，基站需将数据传输到数字孪生体所在的边缘服务器进行模型更新。数字孪生体所在的边缘服务器到基站的距离以及边缘服务器提供给终端数字孪生体更新的计算资源影响数字孪生体更新数据的时间。因此将终端的数字孪生体放置在合适的边缘服务器尤为重要。

令 $y_{e,u,n}=1$ 表示接入切片 n 的终端 u 的数字孪生体被放置在边缘服务器 e 上。则基站将终端 u 的更新数据传输到数字孪生体所在边缘服务器上的时延为：

$$t_{\text{tr}}^{u}=\sum_{n=1}^{N}\sum_{c=1}^{E}\chi\beta_{u,n}y_{e,u,n}l_n\lambda_n\cdot D(p_e,p_b) \tag{3}$$

式中，χ 为单位距离传输单位比特数据所需的单位时延，$D(p_e,p_b)$是边缘服务器 e 到基站的距离。

基站将终端数字孪生体更新数据传送到放置数字孪生体的边缘服务器 e 后，边缘服务器中的数字孪生体需要运行计算和分析接收到的数据，并更新数字孪生模型。则数字孪生体更新时计算数据大小为 $l_n\lambda_n$ 比特的时延为：

$$t_{\text{cal}}^{u}=\sum_{e=1}^{E}\sum_{n=1}^{N}\frac{\beta_{u,n}y_{e,u,n}l_n\lambda_nO}{f_{e,u,n}} \tag{4}$$

式中，O 表示单位比特的计算密度，$f_{e,u,n}$ 为边缘服务器 e 分配给接入切片 n 的终端 u 的计算资源。因此，终端 u 的数字孪生体在更新数据时的总同步延迟为：

$$t_{\text{tot}}^{u}=t_{\text{com}}^{u}+t_{\text{tr}}^{u}+t_{\text{cal}}^{u} \tag{5}$$

1.3　成本和收益模型

尽管数字孪生体能实时反映电网中各个网元、终端的运行状态，然而创建、更新各网元、终端的数字孪生体需要消耗大量的通信、计算等资源。因此在保证数字孪生体同步强度的同时应充分考虑资源的消耗问题。

令 $\kappa_{\text{com}},\kappa_{\text{cal}}$ 分别为通信资源和计算资源的单位成本，则数字孪生体同步更新所消耗的成本如下：

$$c_{\text{res}}=\kappa_{\text{com}}\sum_{n=1}^{N}x_nB+\kappa_{\text{cal}}\sum_{u=1}^{U}\sum_{e=1}^{E}\sum_{n=1}^{N}\beta_{u,n}y_{e,u,n}f_{e,u,n} \tag{6}$$

本文假设系统收益与用户终端获得的同步时延有关，且系统收益计算如下：

$$r_{\text{rev}}=\kappa_{\text{in}}\sum_{u=1}^{U}(\gamma_u-t_{\text{tot}}^{u}) \tag{7}$$

式中，κ_{in} 表示单位收益，单位收益应是一个相对较大的值，以鼓励较低的同步时延。

2　问题构建

在数字孪生使能的 5G 电力虚拟专网管理框架下，保证终端数字孪生体同步强度时权衡资源的消耗和同步时延，即以最小的资源成本获得最小的同步时延，构建如下目标函数：

$$P:\max_{\boldsymbol{x},\boldsymbol{y}}(r_{\text{rev}}-c_{\text{res}}) \tag{8}$$

$$\text{s.t}\quad r_n/l_n\geqslant\lambda,\forall n\in N \tag{8a}$$

$$\sum_{n=1}^{N}\sum_{u=1}^{U}\beta_{u,n}y_{e,u,n}f_{e,u,n}\leqslant F_e,\forall e\in E \tag{8b}$$

$$t_{\text{tot}}^{u}\leqslant\gamma_u,\forall u\in Y \tag{8c}$$

$$x_n\in(0,1],\forall n\in N \tag{8d}$$

$$\sum_{n=1}^{N}x_n\leqslant 1 \tag{8e}$$

$$y_{e,u,n}\in\{0,1\},\forall e\in E,u\in Y,n\in N \tag{8f}$$

$$\sum_{e=1}^{E}y_{e,u,n}=1,\forall u\in Y,n\in N \tag{8g}$$

其中问题 P 中变量 $\boldsymbol{x}=\{x_n\}_{n\in N}$ 且对于接入切片 $n\in N$ 中用户，放置矩阵 $\boldsymbol{y}=\{y_{e,u,n}\}_{e\in E,u\in Y}$。约束(8a)表明对于任意切片 n 要保证 $M/M/1$ 队列稳定性需满足的条件，(8b)表明放置在边缘服务器 e 上的数字孪生体计算容量不能超过其预留给数字孪生

体的计算容量;(8c)表示终端 u 的总同步时延不能超过其同步强度要求 γ_u;约束(8d)和(8e)分别表示切片资源分配变量取值范围及各切片的切片比例和不能超过 1;(8f)和(8g)分别表明数字孪生体放置只能取值 0 或 1 且数字孪生体只能放置在一个边缘服务器上。

问题 P 是一个混合 0-1 整数规划问题,通常采用穷举法或启发式算法求解,然而随着终端数量和边缘服务器数量的增加,穷举法中需列举的组合数呈指数增长,而启发式算法通常只能获得近似解且解与所提出的启发思想有关。深度强化学习将深度学习和强化学习相结合,不断以试错的方式与环境进行交互,通过最大化累积奖赏的方式来获得最优策略[10]。因此本文同样采用深度强化学习求解问题 P,尽管现有的集中式深度强化学习算法能直接训练出问题 P 的近似解,但在执行集中式强化学习模型的前提是集中控制器能实时获得所有环境状态,然而这在实际场景中,不仅增加了信令的开销,也面临状态传输时的隐私泄露问题,因此本文提出一种分层的多智能体算法求解问题 P。

3 问题求解

对于问题(8),当终端的数字孪生体放置策略确定后,用户终端的数字孪生体同步时延仅取决于基站切片控制器为各个切片 n 分配的通信资源,因此问题(8)解耦为切片资源分配子问题:

$$P_1:\max_{\boldsymbol{x}}(r_{\mathrm{rev}}-c_{\mathrm{res}}) \tag{9}$$
$$\text{s.t (8a),(8c),(8d),(8e)}$$

另一方面,当基站的切片资源确定后,用户终端的数字孪生体同步时延仅取决于数字孪生体的放置策略,因此问题(8)解耦为数字孪生体放置子问题:

$$P_2:\max_{\boldsymbol{y}}(r_{\mathrm{rev}}-c_{\mathrm{res}}) \tag{10}$$
$$\text{s.t (8b),(8c),(8f),(8g)}$$

3.1 MDP 构建

为了求解问题 P_1 和 P_2,需分别将两个问题转化为马尔科夫决策过程(Markov Decision Processes,MDP)。MDP 由五元组$\langle \Sigma, A, \boldsymbol{P}, R, \gamma\rangle$组成,其中 Σ 表示有限的状态集合,A 表示有限的动作集合,$\boldsymbol{P}$ 为状态转移矩阵,R 为奖励函数,γ 表示折扣因子[11]。

1)对于问题 P_1,基站的切片控制器作为智能体观测当前的环境状态,记智能体在时隙 t 观测到的状态如下:

$$s_t=(B(t),\{p_u(t)\}_{u\in Y},g(t),\{\beta_{u,n}(t)\}_{n\in N,u\in Y})\in\Sigma \tag{11}$$

其中,$B(t)$表示基站的切片控制器在 t 时刻观测到的可用带宽资源,$\{P_u(t)\}_{u\in Y}$ 表示在时刻 t 切片控制器观测到的 U 个终端的发射功率,$g(t)$表示切片控制器在 t 时刻观测到的信道增益,$\{\beta_{u,n}(t)\}_{u\in Y,n\in N}$ 表示切片控制器在时刻 t 观测到的 U 个用户接入基站的情况。

切片控制器根据 t 时刻观测到的状态 s_t 采取如下动作:

$$a_t=\pi(t) \tag{12}$$

其中,$\pi(t)\in A=\{1,2,3,4,5,6,7,8,9,10\}$,1~10 分别表示 10 种不同比例的通信资源分配方案。

在采取资源分配动作后,切片控制器获得立即奖励 r_t,r_t 的设计可使智能体不断改进策略实现收益最大化,因此将问题 P_1 的目标函数转化为切片控制器的奖励:

$$r_t=\kappa_{\mathrm{in}}\sum_{u=1}^{U}(\gamma_u-t_{\mathrm{tot}}^u(t))-\kappa_{\mathrm{com}}\sum_{n=1}^{N}x_n(t)B(t) \tag{13}$$

当违背 P_1 中约束时,r_t 取一个极大负值,同时该值还受终端数字孪生体放置策略的影响,若终端的放置策略导致负奖励,该值也为极大负值。

2)对于问题 P_2,终端的数字孪生体放置策略受其他终端放置策略的影响,而对每一个终端,其仅能观测到部分环境状态,记各个终端 u 在时隙 t 观测到的部分状态如下:

$$o_t^u=(\beta_{u,n}(t),t_{\mathrm{tr}}^u(t),G)\in O^u \tag{14}$$

其中,$\beta_{u,n}(t)$表示在时刻 t 终端 u 观测到的接入基站的情况,$t_{\mathrm{tr}}^u(t)$为切片控制器进行通信资源分配后带来的数据传输时延,该值由切片控制器广播给其服务的终端。E 表示时刻 t 终端用户所观测到的可放置数字孪生体的边缘服务器总数量。

各终端 u 根据时刻 t 观测到的状态 o_t^u 采取数字孪生体放置动作:

$$a_t^u=(y_{e,u,n}(t))\in A^u \tag{15}$$

其中,动作集合 $A^u=\{1,2,\cdots,G\}$。

终端 u 在时刻 t 观测到的状态 o_t^u 并采取动作 a_t^u 后获得如下立即奖励:

$$r_t^u=\kappa_{\mathrm{in}}(\gamma_u-t_{\mathrm{tot}}^u)-\kappa_{\mathrm{cal}}\sum_{e=1}^{E}\sum_{n=1}^{N}x_{u,n}y_{e,u,n}f_{e,u,n} \tag{16}$$

与基站切片控制代理一样,当终端采取的动作违背

P_2 中的约束时，r_t'' 取一个极大负值。

3.2 基于 DQN 的分层多智能体算法

深度 Q 网络(Deep Q Network，DQN)结合 Q 学习和深度学习[12]，Q 学习是一种基于价值的算法，即通过判断每一步动作的价值来进行下一步的动作。动作值函数有两个输入：状态和动作，它返回在某一状态下执行某一动作的未来奖励的期望，即：

$$Q_\pi(s_t,a_t)=E(r_{t+1}+\gamma r_{t+2}+\gamma^2 r_{t+3}+\cdots+|\ s_t,a_t) \tag{17}$$

其中，r_t 表示时刻 t 在状态 s_t 时执行动作 a_t 所能获得的立即奖励。在 Q 学习中需要遍历所有状态、动作组合情况，当状态空间、动作空间巨大时，Q 值表的收敛计算量随空间的增加呈指数增加，因此 DQN 引入了深度学习，利用 Q 值神经网络去逼近 Q 值函数，即：

$$Q(s,a;w)\approx Q^*(s,a) \tag{18}$$

此外，DQN 采用了两个关键技术：(1)回放缓存。在每个时间步中，DQN 将智能体获得的经验(s_t，a_t，r_t，s_{t+1})存入回放缓存中，然后从该缓存中采样小批量样本用于 Q 学习的更新，通过回放缓存可以重用每个时间步的经验来学习 Q 函数，提高数据使用效率，降低样本的相关性，此外经验回放能平滑学习减少参数的震荡或发散。(2)目标网络。目标网络每 C 步复制主 Q 网络的参数，并使用旧参数生成 Q 学习目标，使目标值的产生不受最新参数的影响，从而避免 Q 网络产生的训练目标值过估计[13]。

DQN 通过一个主神经网络来生成状态-动作值函数 $Q(s,a;w)$，其中 w 为主神经网络需训练的参数，然后智能体选择最大 Q 值对应的动作执行。$Q(s,a;w)$神经网络的训练是一个最优化问题，目标是最小化如下损失函数：

$$L(w)=E[(y_j-Q(s_j,a_j;w))^2] \tag{19}$$

其中，$y_j=r_j+\gamma\max\limits_a Q(s_{j+1},a_j;w^-)$ 为时间差分(Temporal Difference，TD)目标，$Q(s_{j+1},a_{j+1};w^-)$为神经网络参数取 w^- 的目标网络生成的 Q 值。

为求解问题 P，设计了一种基于 DQN 的分层多智能体算法，该算法整体框架如图 2 所示。该框架分为两部分，其中上部分为切片控制器代理与环境交互的过程，下部分为多个终端与环境交互的过程。需要注意的是切片控制器在执行动作后将其切片情况广播给接入它的所有终端，如图 2 中短虚线所示，另一方面，多个终端组成的多智能体环境在执行数字孪生体放置动作后需要将其获得的奖励反馈给切片控制器，如图中长虚线所示，以便切片控制器知晓其切片动作是否违背约束。

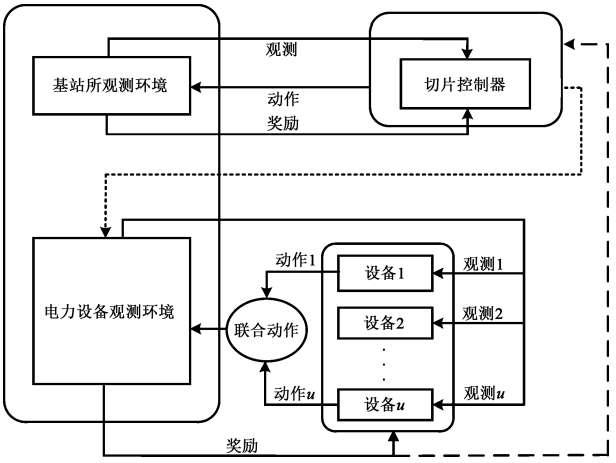


图 2　基于 DQN 的多智能体算法框架

基站的切片控制器作为单智能体根据 3.1 节中 MDP 要素进行切片资源分配，其训练算法如算法 1 所示。

算法 1：　切片控制器资源分配训练算法

输入：环境

输出：训练好的切片控制器 DQN 网络

1　初始化：经验回放池 D，切片控制器 DQN 主网络和目标网络参数

2　for 回合=1，2，…，do

3　初始化环境并获得观测状态 s_0；

4　for 时间步=1，2，…，do

5　切片控制器通过 ε-贪心策略选择动作 a_t；

6　切片控制器执行切片动作 a_t，获得奖励 r_t 并进入下一个状态 s_{t+1}，同时切片控制器将该分配动作结果下发给用户终端所处的多智能体环境；

7　将经验(s_t，a_t，r_t，s_{t+1})存入经验回放池 D；

8　从经验回放池随机采样小批量经验样本；

9　计算时间差分目标：

$$y_j=r_j+\gamma\max_a Q(s_{j+1},a;w^-);$$

10　以梯度下降法最小化损失函数(19)，进而更新 DQN 主网络参数 w；

11　每时间 C 步将 DQN 主网络参数 w 复制给目标网络参数 w^-。

12　end for

13　end for

在算法 1 中，首先初始化切片控制器的 DQN 中的主网络参数 w 和目标网络参数 w^- 以及经验回放池 D(行 1)，每个回合开始初始化训练环境，切片控制器获得初始观测状态 s_0(行 3)，在每个时间步，切片控制器根据 ε-贪心策略选择切片动作，ε-贪心策略是指切片控制器以概率 ε 选择一个随机动作 a_u^t，否则选择动作 $a_t=\text{argmax}_{a\in A}Q(s_t,a;w)$[14](行 5)，切片控制器执行切片动作 a_u^t 后获得立即奖励 r_t，并进入下一状态 s_{t+1}。与传统的 DQN 算法不同，本文提出的改进的多智能体算法中切片控制器资源分配算法和终端的数字孪生体放置算法是关联

的，因为切片控制器的资源分配会影响数字孪生体的放置，因此在切片控制器对通信资源分配后，还需要将分配结果即 $t_{\mathrm{tr}}^{u}(t)$ 广播给接入其的终端(行 6)。然后切片控制器智能体将经验存入经验回放池，当存储到一定量时，切片控制器从经验回放池中随机采样批量经验样本进行训练(行 7-8)，在计算 TD-目标后，通过式(19)计算切片控制器智能体 DQN 网络的损失函数，再通过梯度下降法最小化损失函数 $L(w)$ 并更新 DQN 网络参数 w(行 9-10)。由于切片控制器智能体的目标网络与主网络结构一样，因此需要每隔 C 时间步将主网络参数 w 赋给目标网络 w^{-}。

通过算法 1，切片控制器一旦观察到某一状态立即执行切片资源分配。

在每个时间步基站切片控制器分配切片资源给终端后，多个终端根据部分观测环境自主决定采取的动作，其训练算法由算法 2 给出。不同于算法 1 中 Q 值的计算仅取决于自身的动作，算法 2 中的终端 u 的 Q 值还取决于其他智能体的联合动作，

记除终端 u 以外的联合动作为 a^{-u}，因此终端 u 的时间差分目标和损失函数分别为：

$$y_j^u = r_j^u + \gamma \max_{a^u} Q(o_{j+1}^u, a^u, a^{-u}; w^{u-}) \quad (20)$$

$$L(w^u) = E[(y_j^u - Q(o_j^u, a_j^u, a_j^{-u}; w^u))^2] \quad (21)$$

算法 2： 多终端数字孪生体放置训练算法

输入：环境
输出：训练好的各终端 DQN 网络
1 初始化：每个终端 u 的经验回放池 D_u、DQN 主网络和目标网络参数
2 for 回合=1,2,…,do
3 初始化环境，各终端获得观测状态 o_0^u；
4 for 时间步=1,2,…,do
5 for 每个终端 u do
6 终端通过 ε-贪心策略选择动作 a_t^u；
7 终端智能体执行动作 a_t^u，获得奖励 r_t^u，并进入其能观测到的下一个状态 o_{t+1}^u，同时终端将其所获奖励反馈至切片控制器，并将选择的动作告知其他终端；
8 将经验 $(o_t^u, a_t^u, a_t^{-u}, r_t^u, o_{t+1}^u)$ 存入经验回放池 D_u；
9 从经验回放池采样小批量经验样本；
10 根据式(20)计算时间差分目标；
11 以梯度下降法最小化损失函数(21)，并更新主网络参数 w^u；
12 每时间 C 步将终端 DQN 主网络参数 w^u 复制到目标网络。
13 end for
14 end for
15 end for

算法 2 中每个终端 u 根据自身观测到的环境部分状态决定自己的数字孪生体放置。算法首先初始化每个终端 u 的经验回放池以及其 DQN 主网络和目标网络参数(行 1)。在每个训练的回合开始，初始化训练环境，每个终端获得其观测状态 o_0^u(行 3)，在每个时间步，对每个终端 u 执行相同的操作：首先终端通过 ε-贪心策略选择动作 a_t^u 并执行，然后获得立即奖励 r_t^u，并进入下一个观测状态 o_{t+1}^u，同时终端还需将获得的奖励反馈给切片控制器，以便切片控制器调整其资源分配方案，并将执行的动作告知其他终端，让其他终端知道联合的放置动作带来的价值(行 6-7)。终端 u 将经验 $(o_t^u, a_t^u, a_t^{-u}, r_t^u, o_{t+1}^u)$ 存入经验回放池 D_u。当经验回放池中的样本量存到指定的最大值时终端开始从经验回放池中随机采样批量经验样本，并计算 TD-目标，并以梯度下降法最小化损失函数进而更新主网络参数 w^u(行 8-11)。最后每隔时间 C 步将终端 u 的主网络参数 w^u 赋值给目标网络参数 w^{u-}。

在算法 2 中，终端需要将自己选择的动作告知其他终端，尽管在训练过程中终端需要获得其他终端的放置动作，然而训练完成后的实施阶段，终端只需要根据自己观测的动作执行放置策略即可。因此算法 3 给出了终端 u 的放置实施算法。

算法 3： 基于 DQN 的终端放置实施算法

输入：训练好的各终端 DQN 网络
输出：终端的放置策略
1 加载每个终端的 DQN 网络
2 for 回合=1,2,…,do
3 初始化环境；
4 for 时间步=1,2,…,do
5 for 每个终端 u do
6 观测当前状态 o_t^u；
7 根据 DQN 网络选择动作 a_t^u；
8 end for
9 end for
10 end for

从算法 3 中可以看到在终端的放置实施阶段，终端不需要知道其他终端的放置动作，因此所提方案可实现终端数字孪生体放置的分布式执行。

4 实验验证

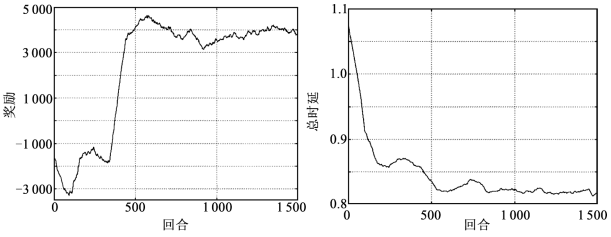
在实验验证阶段，我们借助 Python 3.7 和 Pytorch 1.4.0 版本进行仿真验证，并构建了一个仿真环境，表 1 给出了仿真环境中的主要参数。

为验证所提出的基于 DQN 的分层多智能体算法的性能，本节将以下两种方案作为对比方案：1)DQN+就近放置[2]，该方案利用 DQN 算法分配切片资源，终端总是选择最近的边缘服务器放置其数字孪生体；2)DQN+最大算力放置，该方案同样利用 DQN 算法分配切片资源，但终端总是选择算力最大的边缘服务器放置数字孪生体。

表 1　主要仿真参数

参数	值
基站预留带宽/MHz	1.8
终端数量	10
服务器数量	4
切片数量	2
基站至服务器距离/km	[5,20]
服务器给每个数字孪生体提供的计算资源/(cycle/s)	$[1,5]\times10^4$
数字孪生体同步强度/ms	{6,8}
学习率	0.001
贪心概率 ε	0.1
批量样本数	64
经验回放池大小	100 000

图 3 给出了在学习率为 0.001 时，本文提出的基于 DQN 的分层多智能体算法在 1 500 个训练回合的收敛性能图。其中图 3(a)给出了 1 500 个回合的平均奖励收敛性能图，在 1 000 回合后所提算法能获得 4 000 左右的平均奖励，图 3(b)给出了 1 500 个回合内的平均时延，在 1 000 回合后所提算法能获得约 0.6 s 左右的平均时延。通过图 3 可知我们所提基于 DQN 的分层多智能体算法尽管采用了分层机制，切片控制器资源分配层级的奖励受数字孪生体放置层奖励的影响，但所提算法仍能实现收敛。



(a) 平均奖励　　(b) 平均时延

图 3　基于 DQN 的分层多智能体算法的平均奖励和时延

图 4 给出了不同方案下 10 个时间步内的累积计算资源成本，由图 4 可见，所提的算法的计算资源成本高于 DQN＋最近放置的方案，略低于 DQN＋最大算力的放置方案，这是由于 DQN＋就近放置的方案总是选择距离终端接入基站最近的边缘服务器放置其数字孪生体，而近端的边缘服务器所能提供的计算资源有限，因此其累积计算资源成本低于所提的基于 DQN 的分层多智能体算法，另一方面，DQN＋最大算力的放置方案总是选择计算资源最大的边缘服务器放置数字孪生体，因此其累积计算资源成本大于本文所提算法。

图 5 给出了不同方案下 10 个时间步内的累积时延，所提算法的累积时延均低于其他两种方案，这是由于所提的算法通过神经网络训练能为各个终端找到最佳的放置策略，而 DQN＋最近放置不考虑边缘服务器能提供的计算资源，因此导致数字孪生体模型更新时耗费较长的计算时间，而 DQN＋最大算力不考虑边缘服务器与终端接入基站的距离，尽管边缘服务器能提供较大的计算能力给数字孪生体模型更新从而减少计算时间，但是由于提供最大算力的边缘服务器到终端接入基站的距离较远，导致了较大的数据传输时间，因此两种对比方案的累积时延均高于所提的基于 DQN 的分层多智能体算法。

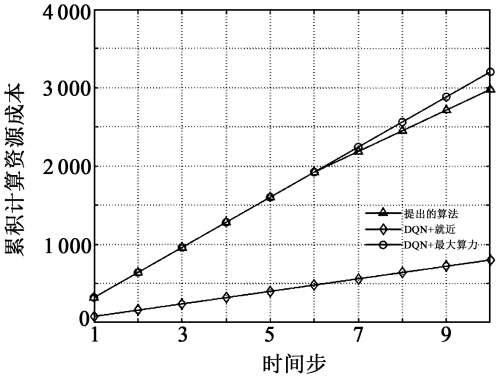


图 4　不同方案下的累积计算资源成本

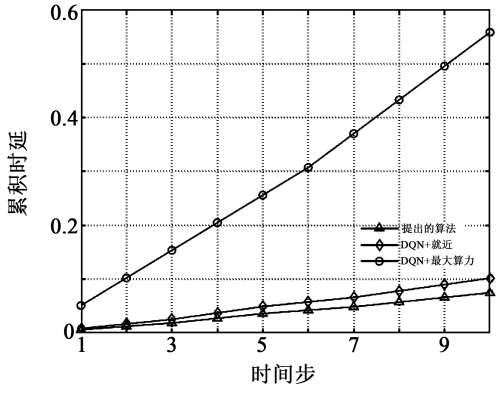


图 5　不同方案下的累积时延

图 6 给出了不同方案下 10 个时间步内的累积收益，所提算法的累积收益高于其他两种对比方案，虽然所提算法的计算资源成本高于 DQN＋最近放置的方案，但所提出的算法在每个时间步总能获得

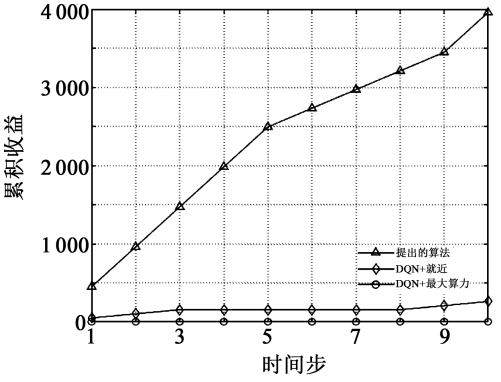


图 6　不同方案下的累积收益

最大的收益，这是因为所提算法通过训练能找到距离和计算资源适合的边缘服务器放置数字孪生体，然而其他两种方案只能得到次优方案，特别是在图6中，DQN＋最大算力的方案所获得的时延总不能满足终端数字孪生体同步要求，因此其总收益为0。

5 结语

为解决源侧的5G电力虚拟专网资源管理可能出现的偏差，本文首先构建了数字孪生使能的5G电力虚拟专网管理框架，同时为保证数字孪生使能的5G电力虚拟专网管理框架下5G电力虚拟专网能获取实时信息，对终端构建满足同步要求的数字孪生体，然后构建了一种结合无线接入网网络切片技术的数字孪生体放置模型，其中无线接入网网络切片为终端的数据同步保证通信资源，终端自主选择其数字孪生体放置位置以保证其同步强度。提出了一种分层多智能体算法选择通信资源分配和数字孪生体放置策略。最后，与其他放置方案对比，验证了所提出的分层多智能体算法在保证终端数字孪生体的同步强度时能获得最大的系统收益。

参考文献：

[1] 王甜甜，杨鸿珍，毛秀伟，等. 基于5G电力虚拟专网的电力业务典型应用[J]. 电力信息与通信技术，2021，19(9)：38-44.
Wang Tiantian，Yang Hongzhen，Mao Xiuwei，et al. The typical application scenarios based on 5G power virtual private network[J]. Electric Power Information and Communication Technology，2021，19(9)：38-44.

[2] 王　荣，张剑寅，宋　月，等. 5G电力虚拟专网架构及关键技术研究[J]. 电信科学，2022，38(S1)：240-249.
Wang Rong，Zhang Jianyin，Song Yue，et al. Architecture and key technologies of 5G power private network [J]. Telecommunications Science，2022，38(S1)：240-249.

[3] 孙　滔，周　铖，段晓东，等. 数字孪生网络(DTN)：概念、架构及关键技术[J]. 自动化学报，2021，47(3)：569-582.
Sun Tao，Zhou Cheng，Duan Xiaodong，et al. Digital twin network (DTN)：Concepts，architecture and key technologies [J]. Acta Automatica Sinica，2021，47(3)：569-582.

[4] Filali A，Mlika Z，Cherkaoui S，et al. Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services[J]. IEEE Trans. on Network Science and Engineering，2022，9(4)：2174-2187.

[5] Shen X，Gao J，Wu W，et al. Holistic network virtualization and pervasive network intelligence for 6G [J]. IEEE Communications Surveys & Tutorials，2021，24(1)：1-30.

[6] Lu Y，Maharjan S，Zhang Y. Adaptive edge association for wireless digital twin networks in 6G[J]. IEEE Internet of Things J.，2021，8(22)：16219-16230.

[7] Hashash O，Chaccour C，Saad W，et al. Towards a decentralized metaverse：Synchronized orchestration of digital twins and sub-metaverses[J/OL]. arXiv preprint arXiv：2211.14686，2022[2023-6-13]. http://arXiv.org/pdf/2211.14686.pdf.

[8] 刘　虹. 基于5G专网的智能电网业务实现方案[J]. 电信工程技术与标准化，2023，36(2)：20-27.
Liu Hong. Smart grid service implementation scheme based on 5G private network[J]. Telecom Engineering Technics and Standardization，2023，36(2)：20-27.

[9] Filali A，Mlika Z，Cherkaoui S，et al. Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services[J]. IEEE Trans. on Network Science and Engineering，2022，9(4)：2174-2187.

[10] 廖晓闽，严少虎，石　嘉，等. 基于深度强化学习的蜂窝网资源分配算法[J]. 通信学报，2019，40(2)：11-18.
Liao Xiaomin，Yan Shaohu，Shi Jia，et al. Deep reinforcement learning based resource allocation algorithm in cellular networks [J]. Journal on Communications，2019，40(2)：11-18.

[11] 董　豪，丁子涵，仉尚航，等. 深度强化学习：基础、研究与应用[M]. 北京：电子工业出版社，2021.
Dong Hao，Ding Zihan，Zhang Shanghang，et al. Deep Reinforcement Learning：Fundamentals，Research and Applications[M]. Beijing：Publishing House of Electronics Industry，2021.

[12] 黄岩松，姚锡凡，景　轩，等. 基于DQN的多起点多终点AGV路径规划[J/OL]. 计算机集成制造系统. https://kns.cnki.net/kcms/detail//11.5946.tp.20230104.0959.002.html.
Huang Yansong，Yao Xifan，Jin Xuan，et al. DQN-based AGV path planning for situations with multi-starts and multi-targets [J/OL]. Computer Integrated Manufacturing System. https://kns.cnki.net/kcms/detail//11.5946.tp.20230104.0959.002.html.

[13] 刘建伟，高　峰，罗雄麟. 基于值函数和策略梯度的深度强化学习综述[J]. 计算机学报，2019，42(6)：1406-1438.
Liu Jianwei，Gao Feng，Luo Xionglin. Survey of deep reinforcement learning based on value function and policy gradient[J]. Chinese Journal of Computers，2019，42(6)：1406-1438.

[14] 工业4.0研究院. 数字孪生电网白皮书——电力企业数字化转型之道[R]. 2021.

作者简介：

吴维农(1965—)，男，四川省人，正高级工程师，主要研究方向包括OPGW特种光缆、无线工业控制、北斗高精度监测等电力通信系统；

唐　伦(1973—)，男，重庆市人，博士，教授，研究方向为新一代无线网络基础理论和组网技术。