doi:10.11959/j.issn.1000-436x.2017173

基于 Q-learning 算法的 vEPC 虚拟网络功能部署方法

袁泉[1,2]，汤红波[1,2]，黄开枝[1]，王晓雷[1,2]，赵宇[1,2]
（1. 国家数字交换系统工程技术研究中心，河南 郑州 450002；
2. 移动互联网安全技术国家工程实验室，北京 100876）

摘 要：针对虚拟化演进分组核心网(vEPC)环境下，现有虚拟网络功能(VNF)部署方法无法在优化时延的同时保证服务链部署收益的问题，提出一种改进的基于 Q-learning 算法的 vEPC 虚拟网络功能部署方法。在传统 0-1 规划模型的基础上，采用马尔可夫决策过程建立了 vEPC 服务链部署的空间—时间优化模型，并设计了改进的 Q-learning 算法求解。该方法同时考虑了空间维度下的 EPC 服务链虚拟映射和时间维度下的 VNF 生命周期管理，实现了 VNF 部署的收益—时延多目标优化。仿真结果表明，与其他 VNF 部署方法相比，该方法在降低网络时延的同时提高了 VNF 部署的收益和请求接受率。

关键词：5G；虚拟网络功能；服务功能链部署；Q-learning
中图分类号：TN915.81 **文献标识码**：A

Deployment method for vEPC virtualized network function via Q-learning

YUAN Quan[1,2], TANG Hong-bo[1,2], HUANG Kai-zhi[1], WANG Xiao-lei[1,2], ZHAO Yu[1,2]
(1. National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China;
2. National Engineering Laboratory for Mobile Network Security, Beijing 100876, China)

Abstract: In the context of vEPC, a method of virtualized network function (VNF) deployment via an improved Q-learning algorithm was proposed to solve the problem that the existing methods cannot achieve the optimization of time delay and revenue of VNF deployment simultaneously. To get the optimal deployment policy in both space dimension and time dimension, a Markov decision process model of vEPC service function chain deployment on the basis of the traditional 0-1 programming model was established and a solution with an improved Q-learning algorithm was proposed. The method had taken full consideration of both virtual network embedding in space dimension and orchestration of VNF life cycle in time dimension, and thus, the multi-objective optimization of revenue and delay could be attained. Simulation shows that the method can reduce network delay while increasing the revenue and the ratio of request acceptance compared with other deployment methods.

Key words: 5G, VNF, service function chain deployment, Q-learning

1 引言

移动互联网的蓬勃发展是 5G 移动通信的主要驱动力。移动互联网将成为未来各种新型业务的基础性平台，现在固定互联网将越来越多的业务通过无线方式提供给用户，云计算及后台服务的广泛应用将对 5G 移动通信系统提出更高的传输质量与系统容量要求。为了满足未来网络需求，驱动网络架构创新，演进分组核心网（EPC, evolved packet core）引入了以软件定义网络（SDN, software-defined

收稿日期：2017-04-14；**修回日期**：2017-07-15
通信作者：汤红波，b101180153@smail.nju.edu.cn
基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（No.2015AA01A706）；国家自然科学基金资助项目（No.61521003）；科技部支撑计划基金资助项目（No.2014BAH30B01）
Foundation Items: The National High Technology Research and Development Program of China (863 Program) (No.2015AA01A706), The National Natural Science Foundation of China (No.61521003), Ministry of Science and Technology Support Plan (No.2014BAH30B01)

network）和网络功能虚拟化（NFV, network function virtualization）为代表的虚拟化技术。虚拟化技术在时间和空间维度上为网络提供了集中管理和编排虚拟网络功能（VNF, virtualized network function）的能力，在此背景下，VNF 部署问题旨在通过灵活的部署策略提高网络基础设施的资源利用率和服务质量，使虚拟化网络平台能够满足新型业务的各项需求。

为了使移动终端完美支持交互式游戏、3D 虚拟现实等在线交互应用，5G 提出了毫秒级时延和 10 Gbit/s 峰值速率的需求。但是，目前 4G 长期演进（LTE, long term evolution）系统的端到端（E2E, end-to-end）时延只能达到 10~100 ms，其中，EPC 系统时延即网络侧时延超过了 10 ms。虚拟化技术为网络功能的编排提供了更加灵活的解决方案，本文旨在设计动态在线的部署方法，在保证网络业务流量的前提下，降低网络服务时延。文献[1]提出一种新型 EPC 网络架构，实现了数据转发和功能特征的分离，网络只负责数据转发，功能特征转移到云平台，可根据用户的需求自动化迁移或伸缩，在逻辑上实现了 VNF 的灵活部署。文献[2]提出应用 NFV 技术管理 5G 网络，在 NFV 网络架构的编排与管理模块（MANO, management and orchestration）中完成 VNF 部署策略的动态调整。文献[3]从空间维度出发提出了一种 vEPC 场景下的 VNF 部署模型，该模型将网络的功能分解为不同的 VNF 组件，并设计了基于决策树算法的动态调度策略实现组件间的组合映射。该方法细化了 VNF 映射的颗粒度，通过将不同 VNF 组合映射到相同的物理节点，优化服务链的总带宽开销，降低部署策略的传输时延，但是该方法没有考虑同一物理节点上不同 VNF 的调度顺序和生命周期的编排策略，仅能获得时间维度优化的次优解，且该方法无法兼顾映射算法的部署收益。文献[4]针对高动态网络及其服务环境的管理问题，从时间维度出发提出了一种 VNF 生命周期编排模型，通过监控系统资源状态，实现 VNF 生命周期的动态自适应管理。文献[5]提出了分离式部署（DD, decomposed deployment）和虚拟化式部署（VD, virtualized deployment）这 2 种 VNF 部署方式：DD 方法采用控制与转发分离的方式部署，避免了节点内部控制与承载之间的纵向信令传输，降低了处理时延；VD 方法将 VNF 的控制面和数据面全部虚拟化，减少了骨干网负载，降低了网络的传输时延。该模型通过选择不同的部署方式统筹优化网络的总体时延，但是该模型不能保证 EPC 服务链的部署收益。文献[6]提出一种网络虚拟化环境下的资源监控方法，通过选择监控代理优化状态信息上报所需的时延和带宽，将监控代理的部署转化为 0-1 规划问题并利用改进的量子遗传算法求解。该方法实现了时延—带宽开销多目标优化，但是该方法仅优化了 VNF 部署在空间维度上的资源分配，没有考虑 VNF 生命周期的编排。文献[7]针对 vEPC 网络不同的场景需求分别提出了基于 ε-Greedy 算法和动态规划算法的部署策略，先从空间维度确定 VNF 和物理节点之间的映射关系，再设计调度算法从时间维度优化各个节点上 VNF 的生命周期编排，以实现 VNF 部署的收益—时延多目标优化。其中，基于贪婪算法的部署策略，具有算法实现简单，计算资源需求小的优点，但是该方法仅能获得空间维度优化和时间维度优化的局部最优解，算法性能存在很大的提升空间。而基于动态规划的部署策略，计算复杂度较高，可以获得全局最优解，但是该方法没有考虑 2 个阶段优化方法中空间维度优化对时间维度优化的限制，降低了时延优化的求解性能。

目前，多数 VNF 部署方法仅从单一维度优化部署问题，而在现有的“先空间、后时间”2 个阶段策略中，固定的空间维度部署策略限制了时间维度下时延优化的可行解空间，导致时延优化问题无法获得全局最优解。为了满足 5G 移动通信在网络侧的高流量和低时延需求，本文提出了一种改进的基于 Q-learning 算法的 VNF 自适应部署方法，该方法在一阶段内完成时间维度和空间维度部署，解决 VNF 部署收益—时延多目标优化问题。该方法首先在空间维度上采用 0-1 规划模型描述 VNF 服务链部署的时延—收益多目标优化问题，然后进一步结合马尔可夫决策过程建立 VNF 部署的时间—空间优化模型，实现了 VNF 生命周期的动态编排。最后，设计改进的 Q-learning 算法求解该模型下的多目标优化问题，并引入基于 BP 神经网络的行为值函数近似方法解决了大规模网络中 Q 矩阵的“维度灾”问题。

2　网络模型

2.1　EPC 网络虚拟化模型

EPC 网络服务的实现需要一组有序的网元功能，该有序的功能集合被称为服务功能链[8]（简称“服务链”）。vEPC 网络服务功能链如图 1 所示，

其中，RAN 表示无线接入节点，S-GW 表示服务网关，P-GW 表示分组数据网关，MME 表示移动管理实体，HSS 表示归属用户服务器。各个网元功能按照业务流程规定的执行顺序组成 vEPC 服务功能链向用户提供服务。SDN&NFV 的应用使 EPC 网络具备了集中控制和自适应编排网络功能的能力。如图 2 所示，SDN 实现了控制与转发分离，将网络实体中异构的网元抽取分离，在控制层集中部署同质网元。NFV 将网元功能与专属硬件平台解耦，使服务提供商能够在通用分布式云平台上动态实例化 VNF 功能链向租户提供服务[9]。

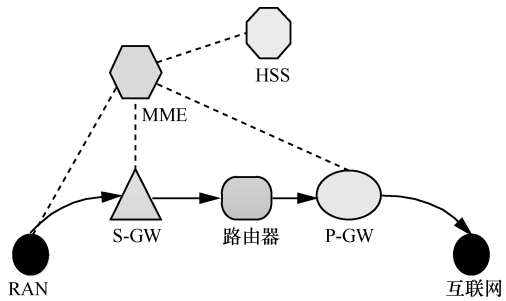


图 1　4G 网络业务流模型

2.2　VNF 部署的空间维度优化模型

VNF 服务链功能部署模型如图 3 所示，物理网络由通用的硬件设备平台组成，表示为一个赋权无向图 G_S，虚拟网络请求是由一组功能组件形成的 VNF 服务链，表示为一个赋权有向图 G_V。虚拟化资源管理和编排模块根据服务链的请求信息和底层资源的状态信息，完成映射 $f:G_V \to G_S$ [10]。VNF 部署的空间维度优化采用的一般性方法为：根据指定的目标函数设计相应的虚拟映射算法，在空间维度上找到 $f:G_V \to G_S$ 映射中，VNF 和虚拟链路所应部署的最优位置。

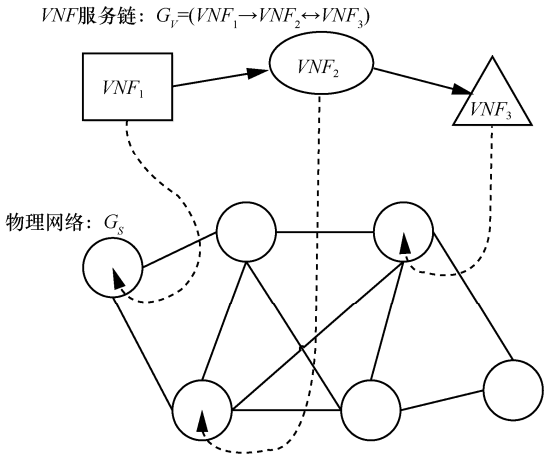


图 3　VNF 服务链功能部署模型

底层网络。图 4 是典型的分布式云平台拓扑结构，表示为一个由物理节点和节点间链路组成的赋权无向图 $G^S=(N^S,E^S)$，其中，N^S 表示物理节点集合，E^S 表示底层链路集合。N^S 中包含服

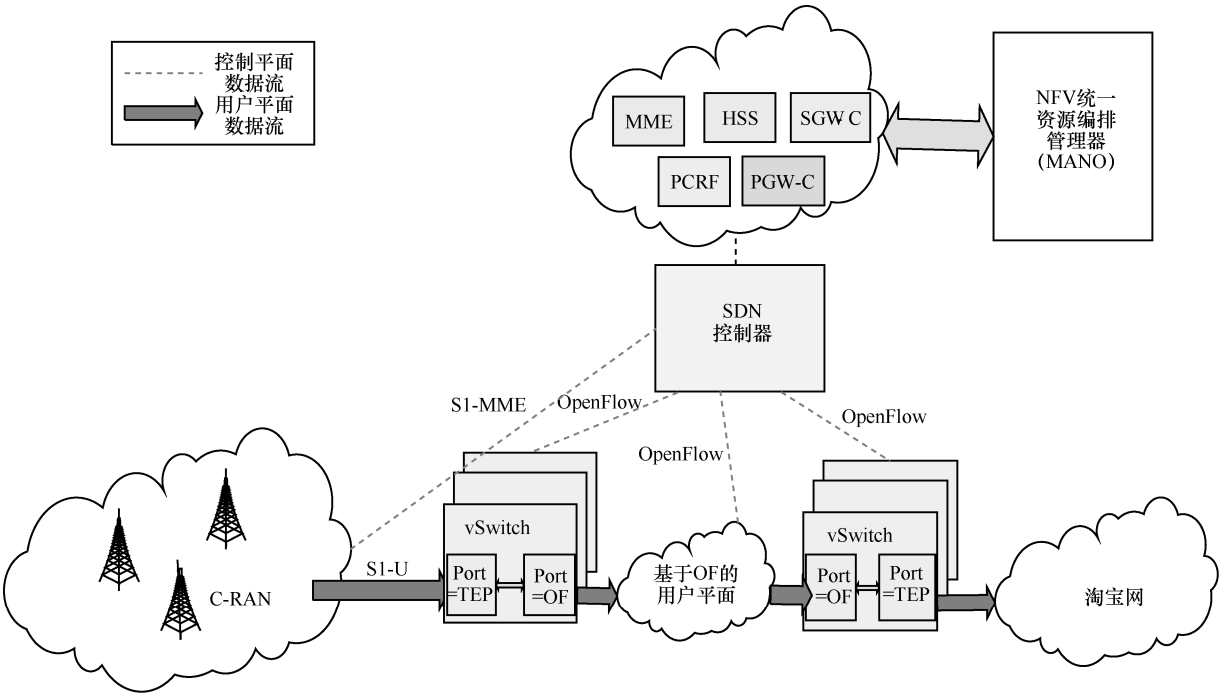


图 2　EPC 网络虚拟化模型

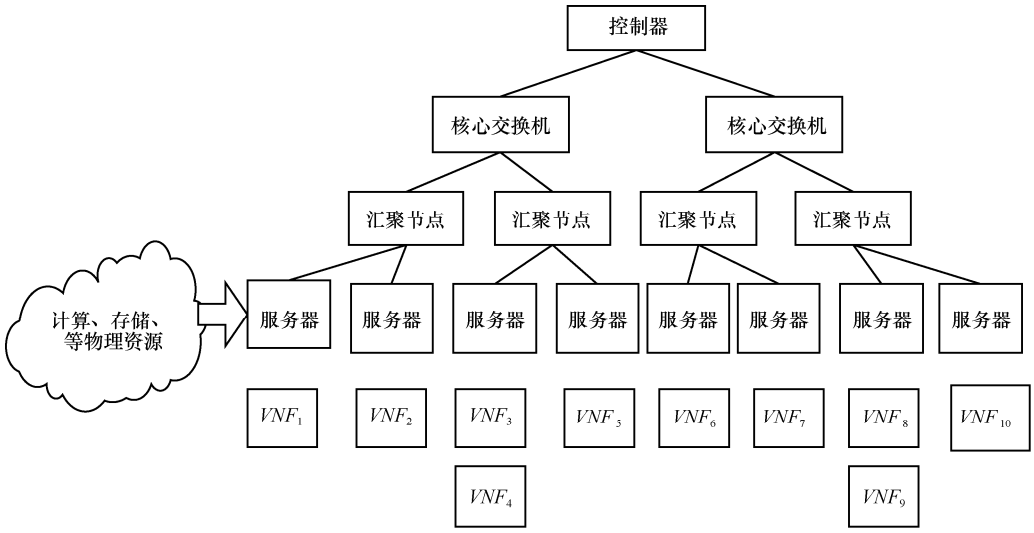


图 4　分布式云平台拓扑结构

务节点（服务器）集合和转发节点（核心交换机和汇聚节点）集合。服务节点用于部署 vMME、vHSS、S/PGW-C 等网元控制平面 VNF，转发节点用于部署 S/PGW-U 等网元转发平面 VNF。n 表示底层网络中物理节点的总数量，k 表示底层网络中提供的物理资源类型数量（如计算、存储、带宽等）。

定义矩阵 $\boldsymbol{C}_{n\times k}$ 为底层网络资源容量矩阵，表示所有物理节点上的各类型资源容量，其中，元素 $C_{i,j}$ 表示节点 i 上第 j 类资源的资源容量。定义矩阵 $\boldsymbol{B}_{n\times n}$ 为物理节点的邻接矩阵，表示物理节点之间的邻接关系，其中，元素 $B_{i,j}$ 表示物理节点 i 和 j 之间通信回路的链路带宽。

EPC 服务链。采用赋权有向图 $G^V=(N^V,E^V)$ 表示服务链中全部 VNF 节点及其关系的逻辑视图，N^V 表示 VNF 的逻辑节点集合，E^V 表示逻辑链路集合。租户请求的 EPC 服务链由一组有序的 VNF 组成，m 表示请求的最大 VNF 数量，l 表示服务链请求的 VNF 类型总数，用数字 $1,2,\cdots,l$ 分别表示 $\text{vMME},\text{vHSS},\cdots,\text{vSGW}$ 及各功能组件等不同类型的 VNF。

定义行向量 $\boldsymbol{L}_{1\times m}$ 为服务链组成向量，表示租户请求服务链的 VNF 组成，其中，元素 $L(i)\in\{1,2\cdots l\},i=1,2\cdots,m$，表示服务链上第 i 个 VNF 的类型。定义矩阵 $\boldsymbol{R}_{m\times k}$ 为 VNF 服务链请求的资源矩阵，其中，矩阵的行表示部署该 VNF 所需的资源向量，列表示服务链请求的 VNF 序列，元素 $R_{i,j}$ 表示服务链上第 i 个 VNF 所请求的第 j 类资源的数量。

定义矩阵 $\boldsymbol{V}_{m\times m}$ 为 VNF 邻接矩阵，其中，元素 $V_{i,j}$ 表示服务链中 VNF_i 和 VNF_j 之间的虚拟链路带宽。

服务链部署。定义二值化矩阵 $\boldsymbol{X}_{m\times n}$ 为映射关系矩阵，表示 VNF 与服务节点之间的映射关系，矩阵中的元素 $X_{i,j}\in\{0,1\}$，矩阵的行表示服务链中的 VNF，列表示底层网络中的服务节点，$X_{i,j}=1$ 表示在服务节点 j 上部署了 i 类型的 VNF，且每一个 VNF 能且仅能部署在一个物理节点上。定义向量 $\boldsymbol{E}_{1\times m}$ 为部署收益向量，表示服务链中所有 VNF 的部署收益，其中，元素 $E(i)$ 表示服务链上的第 i 个 VNF 的部署收益。VNF 部署收益采用该 VNF 占用的所有类型资源的加权和表示，$E(i)=\sum_{j}^{k}\lambda_j R_{i,j}$，为了消除不同类型资源之间由量纲引起的数量级差异，统一衡量各类型资源的部署收益，引入了一组归一化的资源权重系数 $\lambda_1,\cdots,\lambda_k$，$\sum_{i=1}^{k}\lambda_i=1$ 表示各类型资源在部署收益中所占的权重。定义向量 $\boldsymbol{W}_{1\times l}^{p}$ 为 VNF 的处理时延开销向量，表示服务链中各个类型 VNF 的处理时延。文献[6]实验得出，基于 SDN 的网络架构将 VNF 的控制平面和数据平面抽取分离，在控制器北向应用层中单独实例化控制平面功能，VNF 的处理时延将不受数据平面流量的影响，仅与 VNF 的类型有关，同一类型的 VNF 其处理时延近似为一个常数，因此采用元素 $W^p(L(i)),i\in\{1,\cdots,m\}$ 表示服务链 L 上第 i 个 VNF（类型为 $L(i)$）的处理时延开销。定义矩阵 $\boldsymbol{W}_{n\times n}$ 为物理节点之间通信的传输时延开销矩阵，其中，元

素 $W_{i,j}$ 表示物理节点 i 和节点 j 之间的传输时延，如表 1 所示。

表 1 主要参数符号定义

参数	定义
m	租户请求的最大 VNF 数量
n	底层网络中物理节点的总数量
k	表示底层网络中提供的物理资源类型
l	服务链中请求的 VNF 种类
$\boldsymbol{C}_{n\times k}$	底层网络资源容量矩阵
$\boldsymbol{R}_{m\times k}$	VNF 服务链请求的资源矩阵
$\boldsymbol{B}_{n\times n}$	物理节点的邻接矩阵
$\boldsymbol{V}_{m\times m}$	VNF 邻接矩阵
$\boldsymbol{W}_{n\times n}$	传输时延开销矩阵
$\boldsymbol{W}_{1\times l}^{p}$	处理时延开销向量
$\boldsymbol{L}_{1\times m}$	服务链的 VNF 组成向量
$\boldsymbol{E}_{1\times m}$	服务链的部署收益向量
$\boldsymbol{X}_{m\times n}$	服务链的部署关系矩阵

优化目标。为了满足新型业务高流量和低时延的需求，本文提出了时延—收益多目标优化，同时将部署收益和网络总体时延开销作为服务链 VNF 部署的优化目标，在保证资源约束的前提下，求解部署关系矩阵 $\boldsymbol{X}_{m\times n}$。最优的部署策略应满足：1) 最小化全网的服务时延开销；2) 最大化网络部署收益；3) 满足虚拟网络映射的资源约束条件。综上，可以将服务链 VNF 部署转化为 0-1 规划问题。

优化目标为

$$\max(\gamma_b benefit - \gamma_d delay_{\text{overall}}) \tag{1}$$

约束条件为

$$\boldsymbol{X}_{m\times n}\boldsymbol{I}_{n\times 1} = \boldsymbol{I}_{m\times 1} \tag{2}$$

$$(\boldsymbol{X}_{m\times n})^{\mathrm{T}}\boldsymbol{R}_{m\times k} \leqslant \boldsymbol{C}_{n\times k} \tag{3}$$

$$\sum_{i\in N^V}\sum_{j\in N^V} X_{i,s_1}X_{\mathrm{j},s_2}V_{i,j} \leqslant B_{s_1,s_2}, \forall 1\leqslant s_1,s_2\leqslant n, s_1\neq s_2 \tag{4}$$

$$X_{i,s_1}\in\{0,1\}, \forall 1\leqslant s_1\leqslant n, \forall 1\leqslant i\leqslant m \tag{5}$$

式(1)为目标函数，$benefit$ 表示服务链的长期平均部署收益，$benefit = \dfrac{\sum_T\sum_{i=1}^{m}E(i)}{T}$，用于定量描述单位时间内网络因承载业务而消耗的资源数量；$delay_{\text{overall}}$ 表示网络总时延开销，由处理时延开销和传输时延开销 2 个部分组成，$delay_{\text{overall}} = delay_{\text{p}} + delay_{\text{t}}$，其中，网络处理时延开销 $delay_{\text{p}} = \sum_{i=1}^{m} W^{\text{p}}(L(i))$，网络传输时延开销 $delay_{\text{t}} = \sum_{s_1,s_2\in N^S}\sum_{i,j\in N^V} X_{i,s_1}X_{j,s_2}W_{s_1,s_2}, s_1\neq s_2$；引入权重因子 γ_b 和 γ_d 以调节不同场景下 $benefit$ 和 $delay_{\text{overall}}$ 对部署策略的影响，例如，当网络对时延要求较高时，可适当增加 γ_d 的权重。式(2)表示每一个 VNF 能且仅能部署在一个底层服务节点上，其中，$\boldsymbol{I}_{n\times 1}$ 和 $\boldsymbol{I}_{m\times 1}$ 为归一化单位列向量。式(3)表示服务链中 VNF 请求的节点资源不超过底层网络各个服务节点的资源容量。式(4)表示 VNF 之间的虚拟链路带宽不超过底层网络中相应的服务节点间最小链路带宽，其中，X_{i,s_1} 表示第 i 种类型的 VNF 与物理节点 s_1 之间的部署关系，$V_{i,j}$ 表示服务链中第 i 种类型和第 j 种类型的 VNF 之间虚拟链路请求的带宽资源数量。式(5)表示部署关系矩阵元素二值化，当 $X_{i,s_1}=1$ 时，VNF_i 部署在服务节点 s_1 上，反之，则表示 VNF_i 部署没有在服务节点 s_1 上。

2.3 VNF 部署的空间—时间维度优化模型

VNF 部署的时间维度优化是指网络管理编排器通过动态调度服务链中各个 VNF 的生命周期，提高已部署 VNF 在单位时间内的资源利用率。考虑到 EPC 业务流程中网元调度的序贯性，本节在空间维度优化的基础上，采用离散时间马尔可夫决策过程对 VNF 部署的空间和时间维度优化问题建模。传统部署方法认为服务链中的每个 VNF 都具有与服务请求相同的生命周期，但在真实网络环境中，VNF 请求的到达和离开符合泊松过程，泊松过程满足以下 2 个条件：1) 不同 VNF 请求的到达或离开是互相独立的事件，即不同 VNF 的生命周期相互独立；2) 在单位时间内，有且仅有不超过一个 VNF 请求到达或离开，其数学描述如式(6)所示，其中，$R(t)$ 表示 $[0,t]$ 时间内 VNF 请求的数量，h 表示时间维度上的一个无穷小量，$\lambda(t)$ 为 t 时刻泊松分布的强度。式(6)表明服从泊松分布的 VNF 请求在单位时间内最多仅可能有一个请求到达[11]。因此，一个 VNF 服务链部署请求可以被拆分成有序的多个 VNF 部署请求，编排器在单位时间内只选择单一 VNF 节点的部署位置，再由多个 VNF 有序连接形成 VNF 服务链。

$$\begin{cases} P\{R(t+h)-R(t)=1\} = \lambda(t)h + o(h) \\ P\{R(t+h)-R(t)\geqslant 2\} = o(h) \end{cases} \tag{6}$$

综上，VNF 服务链节点部署模型可采用离散时

间平稳马尔可夫决策过程描述，表示为五元组 $\{S,A,r,P,J\}$ [12]，其中，S 定义为有限的 VNF 部署状态空间，空间中的基本事件如式(7)所示，$\boldsymbol{X}(t)$ 定义为 t 时刻空间维度优化模型中的部署状态矩阵，其中，i 表示 VNF，j 表示物理节点，$X_{i,j}$ 为矩阵 $\boldsymbol{X}(t)$ 中的元素，表示 VNF 与服务节点之间的部署关系；A 表示有限的行为空间，其基本事件为服务节点上任意 VNF 的实例化或移除；r 表示收益函数，如果时刻 t 的状态—行为对 (s_t,a_t) 满足 0-1 规划的约束条件，则该部署收益由空间维度优化的目标函数(式(1))表示，$r(s_t,a_t)=\gamma_b benefit-\gamma_d delay_{\text{overall}}$，否则部署收益由惩罚因子表示，$r=-\dfrac{1}{\varepsilon}$，$\varepsilon$ 为一个足够小的正实数；P 为马尔可夫决策过程的状态转移概率，满足式(8)所示的马尔可夫性和齐次性；J 表示折扣总收益 $J=\sum_{t=0}^{T}r(t)$。

$$\boldsymbol{X}(t)\in S,\boldsymbol{X}(t)=\begin{pmatrix} X_{11} & \cdots & X_{1n} \\ & \ddots & \\ \vdots & X_{i,j} & \vdots \\ & & \ddots \\ X_{m1} & \cdots & X_{mn} \end{pmatrix},$$
$$X_{i,j},i\in\{1,\cdots,m\},j\in\{1,\cdots,n\} \tag{7}$$

$$\begin{cases} P\{X_{i,j}(t+1)=\alpha\mid X_{i,j}(t)=\beta, \\ A_n=a,X_{n-1},A_{n-1},\cdots,X_0,A_0\} \\ =P\{X_{i,j}(t+1)=\alpha\mid X_{i,j}(t)=\beta, \\ A_n=a\}=P(i,j,a,\alpha,\beta) \\ \alpha,\beta\in\{0,1\},X_{i,j}\in S,a\in A,\forall t\geqslant 0 \end{cases} \tag{8}$$

本节建立了 vEPC 服务链部署的时间—空间维度优化模型。在空间维度采用 0-1 规划模型描述服务链在某一时刻的部署状态；利用马尔可夫决策过程将时间维度上 VNF 生命周期优化描述为不同时刻部署状态变化的序贯优化问题。与传统服务链部署采用的隐马尔可夫模型相比，该模型将 VNF 生命周期管理的粒度从“服务链级”降低到“网元级”，在时间维度上降低了空闲 VNF 的资源占用率，提高了单位时间内的基础设施收益。

3　算法描述

3.1　基于 Q-learning 算法的虚拟网络功能部署

对于参数已知的马尔可夫决策过程可以利用基于动态规划的值迭代或策略迭代算法求解，但是在现实网络环境中，考虑到时延参数的动态随机性和状态—行为空间的规模，传统的动态规划方法无法在短时间内对模型准确求解，此时增强学习方法成为一种有效的求解手段。增强学习将动态规划的有关理论与学习心理学的机制相互结合，以求解具有延迟收益的序贯优化决策问题为目标，如图 5 所示，Q-learning 算法模型是一个自适应闭环控制系统，在 t 时刻，Agent 对 VNF 部署时—空优化模型的当前状态 s_t 执行行为 a_t 后到达状态 s_{t+1}，同时反馈回路会向 Agent 上报 t 时刻的收益函数 $r(s_t,a_t)$ 并据此更新行为值函数 $Q(s_t,a_t)$，Agent 将根据 $Q(s_t,a_t)$ 更新行为值函数估计表 $Q(s,a)$，再对状态 s_{t+1} 重复上述操作，如此循环直到 $Q(s,a)$ 到达最优行为值 $Q^*(s_t,a_t)$，Agent 根据 $Q^*(s_t,a_t)$ 中各个策略的折扣收益和 J 选择最优策略 $\pi_Q{}^*$。

$$Q(s_t,a_t)=Q(s_t,a_t)+\alpha_t[r(s_t,a_t)+\max Q(s_{t+1},a_{t+1})-Q(s_t,a_t)] \tag{9}$$

$$Q^*(s_t,a_t)=E[r(s_t,a_t)+\max Q^*(s_{t+1},a_{t+1})] \tag{10}$$

Q-learning 算法的收敛性已在文献[13]中证明，行为值函数更新规则如式(9)所示。其中，(s_t,a_t) 为马尔可夫决策过程在时刻 t 的状态—行为对，s_{t+1} 为时刻 $t+1$ 的状态，$r(s_t,a_t)$ 表示 t 时刻 Agent 收到的部署收益，$0<\alpha_t<1$ 为学习因子。最优行为值函数满足 Bellman 最优值方程如式(10)所示，其中，$Q^*(s_t,a_t)$ 表示 t 时刻的最优行为值。

在迭代过程中，假设 t 时刻的部分网络视图如图 6 所示，其中，节点内部数字表示底层节点编号，$d_{ij}(t)$ 表示 t 时刻链路 (i,j)，$i,j\in[1,n]$，节点下方数字表示 VNF 部署到该节点后的业务处理时延，带箭头实线表示当前策略 π_Q。此时，基于 Q-learning 算法的服务链部署方法在 t 时刻对部署状态 s_t 执行动作 a_t，根据部署收益和网络总时延开销计算 Q 学习收益函数瞬时值 $r(s_t,a_t)$，并据此更新 $Q(s_t,a_t)$ 的估计值。在 $t+1$ 时刻更新网络视图，重复上述过程，直到 $Q^*(s_t,a_t)$ 满足式(10)所示的 Bellman 方程，输出最优策略。该方法在时间维度上利用动态规划的思想求解网络时延，采用单位时间内时延的采样值近似瞬时时延，提高了时延优化的精确度。

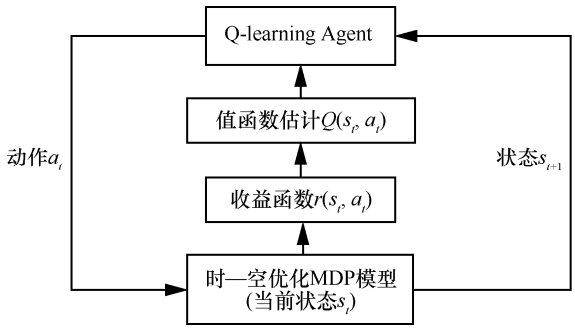


图 5 增强学习模型

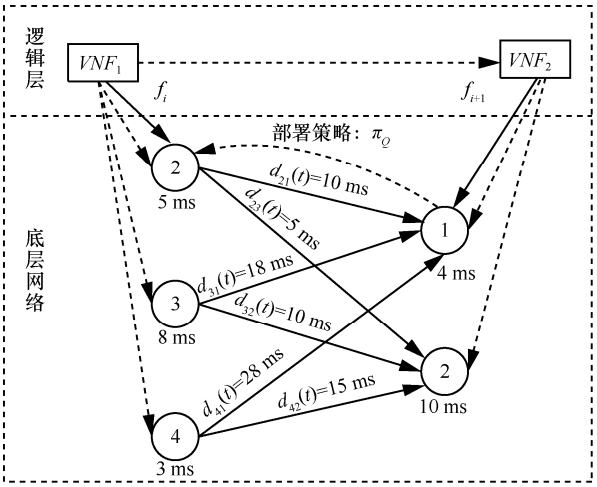


图 6 t 时刻网络视图

3.2 基于 BP 神经网络的行为值函数逼近

在现实环境中，由于网络拓扑复杂，Q-learning 算法通常会面临“维数灾”问题。“维度灾”是指当马尔可夫决策过程拥有庞大的状态空间 S 和行为空间 A 时，由于 Q-learning 算法单一学习周期的值函数估计表 $Q(s,a)$ 规模为 $|S|\times|A|$，随着学习周期的增加，$Q(s,a)$ 将持续占用大量的存储资源，导致学习过程无法完成。为了解决“维数灾”问题，本节引入 BP 神经网络实现对行为值函数 $Q(s,a)$ 的逼近。

BP 神经网络的基本结构如图 7 所示，网络由输入层、隐层和输出层组成，每一层包含多个神经元，前层和后层通过权值连接。BP 网络的学习过程由正向传播和反向传播 2 个部分组成，在正向传播过程中，每一层神经元的状态只影响下一层神经元结构，如果输出层的实际输出与期望输出之间存在误差，网络转向反向输出过程，通过梯度下降法逐层调节权值，逼近输出误差的极小值。在学习过程中，2 个传播过程反复作用，直到误差达到设定范围后，输出函数估计值。

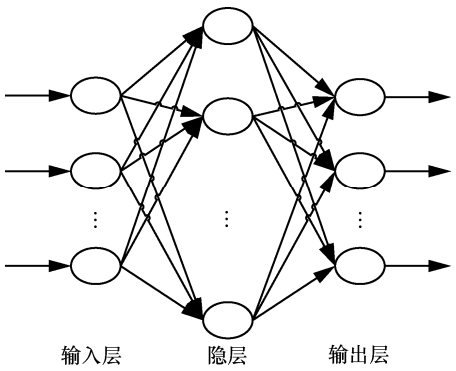


图 7 BP 神经网络结构

目前，BP 神经网络作为一种有监督的函数估计方法已经被证明可以以任意精度逼近值为实数的目标函数。根据文献[14]中提出的一种基于神经网络的强化学习算法以及文献[15]提出的一种基于线性值函数逼近的 Q-learning 算法，本节改进了图 5 中的 Q 学习系统，设计了如图 8 所示的 BP-Q 学习系统。在 t 时刻，Agent 对处于状态 s_t 的 VNF 部署时—空优化模型执行动作 a_t，获得即时收益 $r(s_t,a_t)$ 后，系统向神经网络模块输入 MDP 当前的状态—动作对 (s_t,a_t) 和即时收益 $r(s_t,a_t)$，神经网络模块根据输入的 (s_t,a_t) 和即时收益 $r(s_t,a_t)$ 对行为值函数进行逼近，输出行为值函数的估计值 $Q_{NN}(s_t,a_t)$ 至 Agent，Agent 使用估计值执行行为值函数迭代，并将学习结果 $Q(s_t,a_t)$ 反向传输回神经网络模块实现权值向量调节。

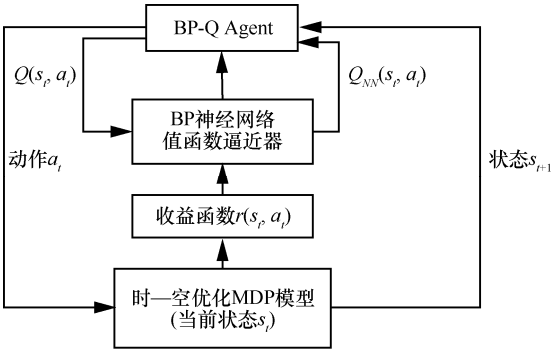


图 8 BP-Q 学习系统

根据上述流程，式(9)中行为值函数更新规则可修改为如式(11)所示的形式，改进后的算法称为 BP-Q 学习算法，改进后的系统不再存储并更新规模为 $|S|\times|A|$ 的行为值函数估计表 $Q(s,a)$，而仅仅需要存储神经网络中的各个神经元的权值，其存储规模仅与所设计神经网络的拓扑结构有关。基于改进的 BP-Q 学习系统，相应的 VNF 自动化部署算法如算法 1 所示。

$$Q(s_t,a_t)=Q_{NN}(s_t,a_t)+\alpha_t[r(s_t,a_t)+\max Q_{NN}(s_{t+1},a_{t+1})-Q_{NN}(s_t,a_t)] \quad (11)$$

算法 1　基于 BP-Q 学习的 VNF 自动化部署算法

输入　服务链 l 的请求信息；底层网络图 G_S；VNF 的参数

输出　服务链 l 的时延—收益优化部署方案 π_Q^*

1) 根据输入初始化马尔可夫决策过程状态空间 S、行为空间 A 和转移概率矩阵 P；

2) 根据服务链 l 请求信息，组合 VNF 得到功能链 l 的顺序约束 φ_l；

3) 根据顺序约束 φ_l 初始化选择策略 π_Q；

4) 初始化 Q-learning 学习因子 α_0，误差允许阈值 ε，最大学习周期 Max_t；

5) 初始化神经网络权重向量 W 和样本集合 $dataset$，最大学习步长 Max_step；

6) 建立 0-1 规划模型；

7) 根据 0-1 规划初始化即时收益函数 $r(s_t,a_t)$；

8) 设置 $t=0$，随机初始化起始状态 s_0；

9) while($|Q_{NN}(s_t,a_t)-Q^*(s_t,a_t)|\geqslant\varepsilon$ && $t\leqslant Max\_t$);

10) for s_t 中所有的 a_t；

11) $step = 0$；

12) 计算 $Q_{NN}(s_t,a_t)$；

13) while ($step$!= Max_step // s_t is not terminal);

14) 将样本 $\{s_t,a_t,Q_{NN}(s_t,a_t)\}$ 加入样本集 dataset

15) 执行动作 a_t 获得即时收益 $r(s_t,a_t)$，进入状态 s_{t+1}；

16) $step=step+1$；

17) if $rand<1-\varepsilon$；

18) $a^*=\arg\max_a Q_{NN}(s_{t+1},a_{t+1})$；

19) $\delta=r(s_t,a_t)+Q_{NN}(s_{t+1},a_{t+1})-Q_{NN}(s_t,a)$；

20) else

21) 根据 π_Q 随机选取当前动作 a_t；

22) $Q(s_t,a_t)=Q_{NN}(s_t,a_t)+\alpha\delta$，更新 $dataset$；

23) end if

24) if $dataset$ 存储空间已满

25) BP 神经网络开始训练，更新权值向量 W；

26) end if

27) end while

28) $t=t+1$,更新学习因子 α_t；

29) end for

30) end while

31) 根据 $Q^*(s,a)$ 计算服务链部署方案 P 的折扣收益和 J；

4　实验结果及性能分析

为了全面评估模型可行性及算法有效性，本文采用服务请求的处理时间、接受率，底层网络的收益和网络总时延作为性能评价指标进行仿真，并与基于量子遗传算法（AQGA）[5]、贪婪（ε-Greedy）算法和动态规划（DP）算法[7]的部署方法进行对比。

4.1　实验环境和参数设置

本实验在 Intel Core i7-4790、3.60 GHz CPU、8 GB 内存的 Linux 系统 PC 机上运行，底层网络拓扑由 GT-ITM 工具生成[16]，仿真程序在 Matlab 环境下实现并运行。本实验所用拓扑结构取自 SNDlib 库中的 Polska12 测试例子[17]，如图 9 所示。图 9 中节点位置部署了运营商的云数据中心，假设所有节点都可以承载除了 PGW 之外的任意核心网功能，PGW 作为移动核心网和互联网之间的锚点，位置通常是固定的。在 Polska12 节点测试例子中，设定只有 Rzeszow 和 Bydgoszcz 节点可以承载 PGW。在底层网络中，节点网络资源的数量和链路带宽由 SNDlib 给出，底层节点的其他类型资源根据节点网络资源按比例随机生成。考虑到物理链路时延的随机性，选择期望（单位：ms）取值范围为 (0,1] 的平稳随机过程表示时延。每个服务请求由一条 VNF 服务链表示，服务链中 VNF 的类型和数量满足 $2\leqslant l\leqslant m\leqslant 10$。

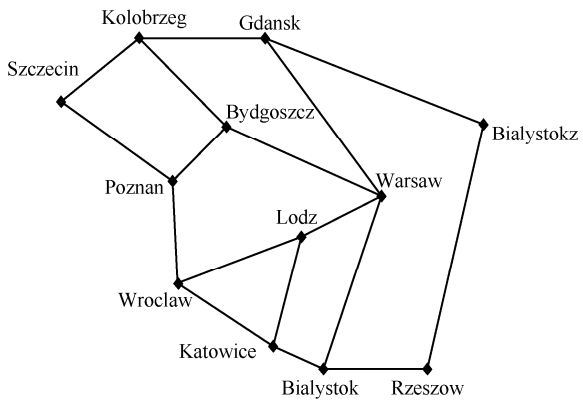


图 9　Polska12 节点拓扑

优化目标中权重因子 γ_b 和 γ_d 都设为 1。实验采

用 3 层 BP 神经网络逼近 Q 函数，拓扑结构为 $1 \leftrightarrow 2n \leftrightarrow 1$，其中，$n$ 表示底层节点的总数量。在改进的 Q-learning 算法中，惩罚因子 $\varepsilon = 0.01$，最大学习周期 $Max\_t = 1\,500$，最大学习步长 $Max\_step = 1\,400$，初始状态下学习因子 $\alpha_0 = 0.8$，在迭代过程中，学习因子采用动态更新策略提高 agent 在学习过程中的泛化能力。

图 10 中的未更新学习因子曲线显示了当在改进的 Q-learning 算法中采用固定的 α 更新策略时，学习周期 t 和各个周期内学习步数 $step$ 之间的关系。在学习初期，曲线振荡的幅度和频率均较大，提供给神经网络的样本组误差较大，导致网络的权值偏差较大。为了提高算法效率，提高较优样本的学习经验，本节提出了基于步长的学习因子更新策略，令 $\alpha_t = \alpha_0 \dfrac{step}{Max\_step}$。图 10 中的更新学习因子曲线表示采用更新策略后，学习周期 t 和各个周期内学习步数 $step$ 之间的关系，实验证明，该策略能够有效提高网络的收敛速度。

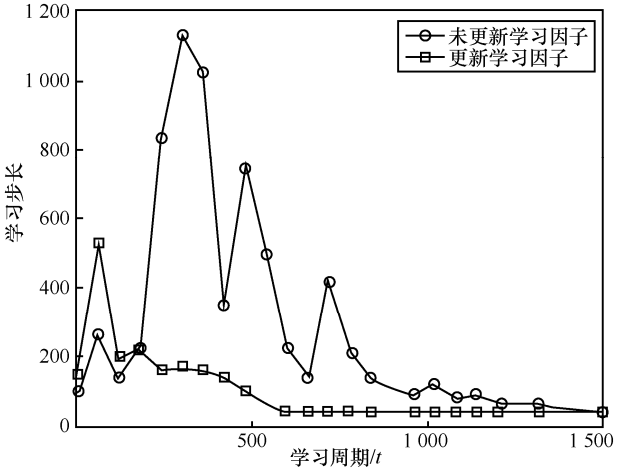


图 10 学习因子更新前后对比

4.2 性能分析

首先对各个部署方法的服务请求处理时间进行分析，如图 11 所示，Greedy 方法处理时间最长，因为只有该算法是 2 个阶段映射算法，第 1 阶段遍历地选择满足节点约束的 VNF 节点部署方案，第 2 阶段通过贪婪算法选择节点之间的时延最小链路，优先完成部署的 VNF 需要在缓存队列中等待服务链中的下一个 VNF 完成部署后才能与其建立链路。由于第 1 阶段将遍历到大量不符合约束条件的无效解，导致 VNF 在队列中等待的时间较长，降低了算法效率。基于 DP 的方法采用递归的方式无方向地搜索时延最短路径，利用动态规划的思想先将服务链部署问题分割成多条子链的部署问题，再按顺序组合各条子链，获取最优部署方案，算法复杂度为 $O(n^2m)$。基于改进的 Q-learning 算法的方法通过更新行为值 $Q(s,a)$ 不断地向折扣收益和最大的方向搜索部署策略，避免了盲目遍历全部状态空间，但是在学习初期，$Q(s,a)$ 包含的知识较少，神经网络训练效率较低，影响了的 $Q(s,a)$ 收敛速度。基于量子遗传算法的部署方法以式(1)作为适应度函数，不断地向适应度更高的方向进化，在相同的请求强度下，该方法的服务请求处理时间最短。

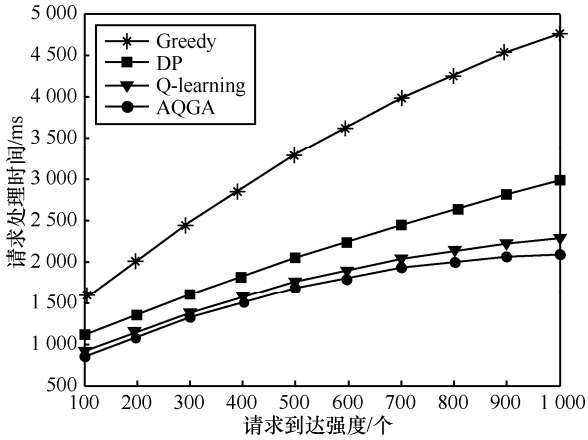


图 11 不同请求强度下的服务请求处理时间

图 12 表示不同请求强度下各方法的请求接受率性能。服务请求处理时间直接决定相同请求的资源占用周期，影响底层网络资源的利用率。Greedy 方法的请求接受率最低，且随着请求强度的增加呈明显下降趋势。这是因为该方法未能综合考虑底层网络节点和链路资源状态对部署策略的影响，往往赋予低时延链路较高的部署优先级，导致部分链路频繁复用，造成网络局部过载，降低请求接受率。DP 方法与 AQGA 方法在空间维度上实现了节点和链路资源的统筹分配，但在时间维度上，该方法把整个服务请求到达至离开的时间段作为所有 VNF 的生命周期，没有考虑 VNF 生命周期在时间维度上的顺序约束，导致大量空闲 VNF 占用底层资源，降低了请求接受率。Q-learning 方法实现了 VNF 生命周期的细粒度管理，在服务请求时间内将各个 VNF 的创建和移除看作独立的随机过程，提高了物理资源在时间维度上的利用率，因此 Q-learning 方法具有最高的请求接受率。

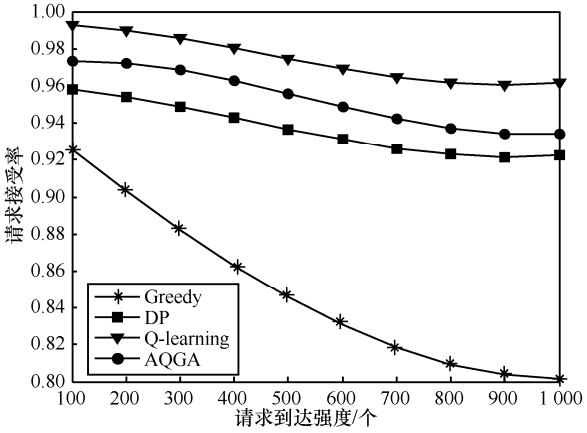


图 12　不同请求强度下的服务请求接受率

图 13 表示不同请求强度下各算法的长期平均收益，收益表达式由式(1)中的长期平均收益 *benefit* 定义。Greedy 方法和 DP 方法不支持多目标优化，在优化时延的基础上仅能获得部署收益的次优解。AQGA 由于进化方向性强，梯度下降迅速在寻优过程中容易陷入局部最优解。此外，Greedy 方法、DP 方法和 AQGA 都没有考虑时延的随机特性，时延优化精确度较低。此时，Q-learning 算法表现出在多目标优化问题中的良好性能，与 Greedy 方法、DP 方法和 AQGA 相比，Q-learning 算法实现了 VNF 生命周期的细粒度管理，在相同的请求强度下，基于 Q-learning 算法的 VNF 部署方法能够避免空闲 VNF 占用物理资源，降低单位时间内的通信开销，有效提高底层网络的基础设施收益。

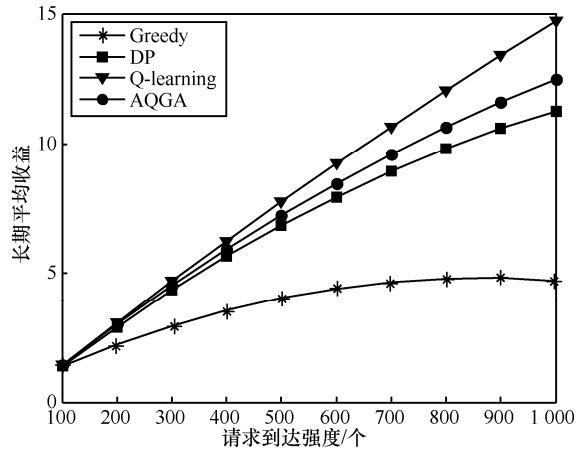


图 13　不同请求强度下的收益均值

图 14 表示不同请求强度下的网络总时延开销。Greedy 方法、DP 方法和 AQGA 方法未考虑时延参数的随机性，在一个迭代周期内采用固定参数描述链路时延，无法求得服务链时延的实时最优解。本次实验采用分布函数单调递减的平稳随机过程描述链路时延，在迭代过程中，基于 Q-learning 的算法采用单位步长时间内时延的采样值近似瞬时时延计算收益函数并据此更新当前部署策略下的行为值函数估计的延迟收益，提高了时延优化的精确度，结果表明 Q-learning 算法能够有效降低 vEPC 传输时延，提高服务质量。

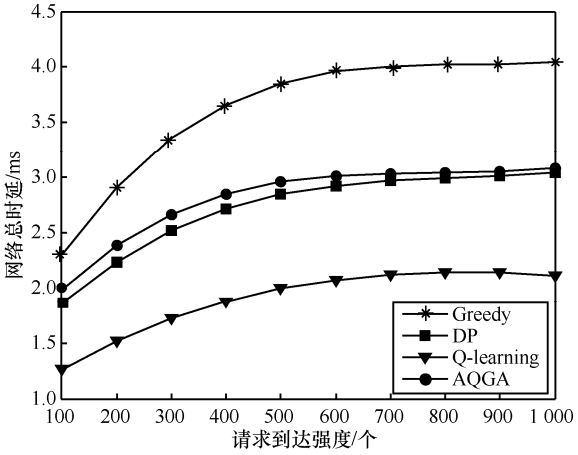


图 14　不同请求强度下的时延均值

5　结束语

本文主要研究了虚拟化环境下服务链 VNF 的部署问题，针对传统虚拟网络映射算法的不足和 5G 业务更高的时延要求，提出基于 Q-learning 算法的 VNF 部署方法，并验证了方法的有效性。为了进一步提高虚拟化背景下移动业务的服务质量，后续将针对可靠性条件下 VNF 部署问题进行研究，以满足移动业务的电信级可靠性要求。

参考文献：

[1] SAMA M R, CONTRERAS L M, KAIPPALLIMALIL J, et al. Software-defined control of the virtualized mobile packet core[J]. IEEE Communications Magazine, 2015, 53(2): 107-115.

[2] TALEB T, CORICI M, PARADA C, et al. EASE: EPC as a service to ease mobile core network deployment over cloud[J]. IEEE Network, 2015, 29(2): 78-88.

[3] MOENS H, DE F. VNF-P: a model for efficient placement of virtualized network functions[C]//IEEE International Conference on Network and Service Management. 2014: 418-423.

[4] CLAYMAN S, MAINI E, GALIS A, et al. The dynamic placement of virtual network functions[C]//IEEE International Conference on Network Operations and Management Symposium. 2014: 1-9.

[5] BASTA A, KELLERER W, HOFFMANN M, et al. Applying NFV and

SDN to LTE mobile core gateways, the functions placement problem[C]//ACM Workshop on All Things Cellular: Operations, Applications, & Challenges, 2014: 33-38.

[6] 江逸茗, 兰巨龙, 周惠琴. 网络虚拟化环境下的资源监控策略[J]. 电子与信息学报, 2014, 36(3): 708-714.
JIANG Y M, LAN J L, ZHOU H Q. Resource monitoring policy for network virtualization environment[J]. Journal of Electronics & Information Technology, 2014, 36(3): 708-714.

[7] MIJUMBI R, SERRAT J, GORRICHO J L, et al. Design and evaluation of algorithms for mapping and scheduling of virtual network functions[C]//IEEE Conference on Network Softwarization. 2015:1-9.

[8] HAN B, GOPALAKRISHNAN V, JI L S, et al. Network function virtualization: challenges and opportunities for innovations[J]. IEEE Communications Magazine, 2015, 53(2): 90-97.

[9] BASTA A, KELLERER W, HOFFMANN M, et al. A virtual SDN-enabled LTE EPC architecture: a case study for S-/P-Gateways functions[C]//IEEE International Conference on SDN for Future Networks and Services. 2013:1-7.

[10] FISCHER A, BOTERO J F, TILL BECK M, et al. Virtual network embedding: a survey[J]. IEEE Communications Surveys & Tutorials, 2013, 15(4): 1888-1906.

[11] ROSS S M. Stochastic processes[J]. John Wiley & Sons Inc New York, 1996, 48(1):528-529.

[12] SONG H, LIU C, LAWARRÉE J, et al. Optimal electricity supply bidding by Markov decision process[J]. IEEE Transactions on Power Systems, 2000, 15(2): 618-624.

[13] WATKINS C, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3-4): 279-292.

[14] BUSONIU L, BABUSKA R. Reinforcement learning and dynamic programming using function approximators[M]. Florida: CRC Press, 2010.

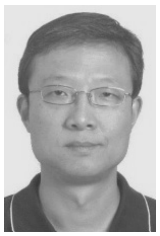
[15] DUONG T, CHU Y, NGUYEN T, et al. Virtual machine placement via Q-learning with function approximation[C]//2015 IEEE Global Communications Conference. 2015: 1-6.

[16] ZEGURA E W, CALVERT K L, ACHARJEE S B. How to model an internetwork[C]//IEEE Conference of Computer Societies, Networking the Next Generation. 1996: 594-602.

[17] ORLOWSKI S, WESSÄLY R, PIÓRO M, et al. SNDlib 1.0-survivable network design library[J]. Networks, 2010, 55(3): 276-286.

作者简介：

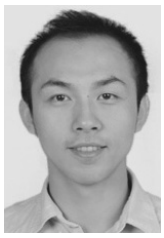
袁泉（1991-），男，山东青岛人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为移动通信网络、网络功能虚拟化。



汤红波（1968-），男，湖北孝感人，国家数字交换系统工程技术研究中心博士生导师，主要研究方向为移动通信网络、新型网络体系结构。

黄开枝（1973-），女，安徽滁州人，博士，国家数字交换系统工程技术研究中心博士生导师，主要研究方向为无线移动通信、无线物理层安全。

王晓雷（1982-），田，山东淄博人，国家数字交换系统工程技术研究中心讲师，主要研究方向为移动通信网络、新型网络体系结构等。



赵宇（1984-），男，吉林辽源人，国家数字交换系统工程技术研究中心讲师，主要研究方向为移动通信网络、新型网络体系结构等。