基于异步优势演员-评论家学习的服务功能链资源分配算法

唐 伦 贺小雨* 王 晓 谭 颀 胡彦娟 陈前斌

(重庆邮电大学通信与信息工程学院 重庆 400065)

(重庆邮电大学移动通信重点实验室 重庆 400065)

摘 要：考虑网络全局信息难以获悉的实际情况，针对接入网切片场景下用户终端(UE)的移动性和数据包到达的动态性导致的资源分配优化问题，该文提出了一种基于异步优势演员-评论家(A3C)学习的服务功能链(SFC)资源分配算法。首先，该算法建立基于区块链的资源管理机制，通过区块链技术实现可信地共享并更新网络全局信息，监督并记录SFC资源分配过程。然后，建立UE移动和数据包到达时变情况下的无线资源、计算资源和带宽资源联合分配的时延最小化模型，并进一步将其转化为马尔科夫决策过程(MDP)。最后，在所建立的MDP中采用A3C学习方法，实现资源分配策略的求解。仿真结果表明，该算法能够更加合理高效地利用资源，优化系统时延并保证UE需求。

关键词：网络切片；服务功能链资源分配；马尔科夫决策过程；异步优势演员-评论家学习；区块链

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2021)06-1733-09

DOI: 10.11999/JEIT200287

Resource allocation Algorithm of Service Function Chain Based on Asynchronous Advantage Actor-Critic Learning

TANG Lun HE Xiaoyu WANG Xiao TAN Qi HU Yanjuan CHEN Qianbin

(*School of Communication and Information Engineering*, *Chongqing University of Posts and Telecommunications*, *Chongqing* 400065, *China*)

(*Key Laboratory of Mobile Communication*, *Chongqing University of Posts and Telecommunications*, *Chongqing* 400065, *China*)

Abstract: Considering the fact that global network information is hard to obtain, and the slice resource allocation optimization problem caused by mobility of User Equipment (UE) and dynamics of packet arrival in the radio access network slice, a Service Function Chain(SFC)resource allocation algorithm based on Asynchronous Advantage Actor-Critic (A3C) learning is proposed. Firstly, a resource management mechanism based on blockchain technology is established, which can credibly share and update the global network information, also supervise and record SFC resource allocation process. Then, a delay minimization model based on joint allocation of radio resources, computing resources and bandwidth resources is built under the circumstance of UE moving and time-varying packet arrival, and further transformed into an Markov Decision Process(MDP) problem. At last, A3C learning method is adopted to obtain the resource allocation optimization strategy in this MDP. Simulation results show that the proposed algorithm could utilize resources more efficiently to optimize the system delay while guarantee the requirement of each UE.

Key words: Network slice; Service Function Chain(SFC) resource allocation; Markov Decision Process(MDP); Asynchronous Advantage Actor-Critic(A3C) learning; Blockchain

收稿日期：2020-04-21；改回日期：2020-09-28；网络出版：2020-09-30

*通信作者： 贺小雨 Hexy1995@163.com

基金项目：重庆市教委科学技术研究项目(KJZD-M20180601)，重庆市重大主题专项(cstc2019jscx-zdztzxX0006)

Foundation Items: The Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M20180601), The Major Theme Special Projects of Chongqing (cstc2019jscx-zdztzxX0006)

1 引言

网络切片指在一个完整通用的物理基础设施中，逻辑地分离网络功能和资源，以保证不同通信应用场景中的服务质量(Quality of Service, QoS)需求[1]。每个切片络包含有若干条相同服务类型的服务功能链(Service Function Chain, SFC)，每条SFC由若干有序虚拟网络功能(Virtual Network Function, VNF)组成。系统需要根据用户需求和相关约束，合理地将VNF放置在底层网络并为其分配CPU、内存、带宽等物理资源。在接入网中，用户终端(User Equipment, UE)的移动性使得接入网切片环境更具动态性和未知性[2,3]。

在接入网切片网络中，UE通过远端无线单元(Remote Radio Unit, RRU)将数据传输到对应的SFC进行处理，形成了特殊的UE-RRU-SFC 3层关联架构。因此，当UE移动时，首先会涉及到无线资源的重分配问题。文献[4,5]考虑了UE的移动性和时变的数据到达率，通过优化无线资源分配来降低时延。但实际上，由于形成的UE-RRU-SFC 3层关联架构，当UE从一个RRU覆盖范围移动至另一个RRU时，若当前RRU无法直接为其提供所需的SFC，则需要一条新路径将UE的数据从当前RRU传输到对应的SFC。在这一过程中，不仅需要进行无线资源分配的调整、物理链路带宽资源的重分配以及部分VNF迁移带来的物理设备上计算资源的重分配[6]。同时，这些资源的分配方案会对系统时延产生影响。因此，在资源有限的网络中，如何合理地设计SFC资源分配算法，从而提高资源利用率、降低系统时延是亟待解决的问题。

另一方面，由于UE的移动性和时变的数据到达率，需要适时地对资源进行调整。绝大多数文献中，在调整或分配资源前，实际上都是在已知网络资源状态、UE信息以及当前VNF放置和资源分配的前提下，而事实上，这些全局信息往往很难获得甚至无法获得。文献[7]提出了一种“共享账簿”的概念，用于记录和共享切片资源分配过程中所需的一些必要信息，且各个切片都具有修改和维护该账簿的权限。但是并未对该种“共享账簿机制”的实现展开讨论。文献[8]提出了一种基于区块链的云架构，实现网络各类资源信息在多台设备上的共享和分布式管理。由于区块链技术本身所具有的去中心化、集体维护性、自信任性、可验证性和可追溯性等特点，提升了资源管理的可靠性和可信度[9]。

此外，随着未来网络规模的不断增大且部署更加灵活，传统方法难以解决高维度和高动态性的资源优化问题，因此智能资源管理成为当前研究的热点。文献[10]采用强化学习算法对SFC中VNF的调度问题进行研究，但由于该方案利用有限的离散值对连续动作进行量化，会破坏动作空间的完备性。文献[11]采用基于策略梯度的算法(Policy-Based Algorithm, PBA)对SFC部署问题进行研究，其能够在连续的动作空间中有效学习随机策略，并获得较好的收敛性，但易收敛到局部最优。文献[12]首次提出了演员-评论家(Actor-Critic, A-C)学习算法，它结合了策略方案和值函数方案，使得在连续随机策略方面有较好的优越性。然而，A-C学习只适用单智能体进行样本采集，可能导致得到的样本是高度相关的，从而随空间维度增加，算法将难以收敛。

针对接入网切片中SFC资源分配所存在的诸多问题，本文提出了一种基于异步优势演员-评论家学习(Asynchronous Advantage Actor-Critic, A3C)的SFC资源分配方案。主要贡献包括：

(1) 考虑SFC资源分配过程需获悉网络全局信息但难以获得的实际情况，包括UE位置信息、QoS需求、数据包到达信息，物理基础设施中的无线资源、计算资源、链路带宽资源信息以及目前VNF放置和资源分配情况信息等，提出一种基于区块链的资源管理机制。通过引入区块链技术，实现网络全局信息的“分布式账本式”存储和管理，并进行可信可靠的共享、同步及更新，完成SFC资源分配过程的监督和记录。

(2) 考虑接入网切片场景下形成的UE-RRU-SFC 3层关联架构，建立UE移动和数据包到达过程时变情况下的无线资源、计算资源和链路带宽资源的联合分配模型，以优化系统时延并满足UE的QoS需求。

(3) 将优化模型转化为马尔科夫决策过程(Markov Decision Process, MDP)进行求解。考虑到该MDP的状态和动作空间连续且维度较大，状态转移概率也未知，采用A3C方法实现SFC资源分配策略的求解。

2 系统架构

2.1 接入网切片场景的SFC资源分配框架图

如图1所示，基于5G C-RAN上行条件下，切片内的每个UE都拥有一条SFC进行数据传输。但是考虑到UE的移动性和数据包到达的时变性，需要考虑对SFC的资源分配进行适当地调整。UE的移动伴随着SFC中的VNFs迁移，因此需要重新为迁移的VNFs分配计算资源、链路带宽资源等，因此还会涉及到无线资源的调整。VNF迁移引起的网络资源重配置这一过程也会带来额外的时延。在图1所示的物理层中，基于区块链的分布式网络资源管理思想，各个UE,RRU以及物理设备之间会以

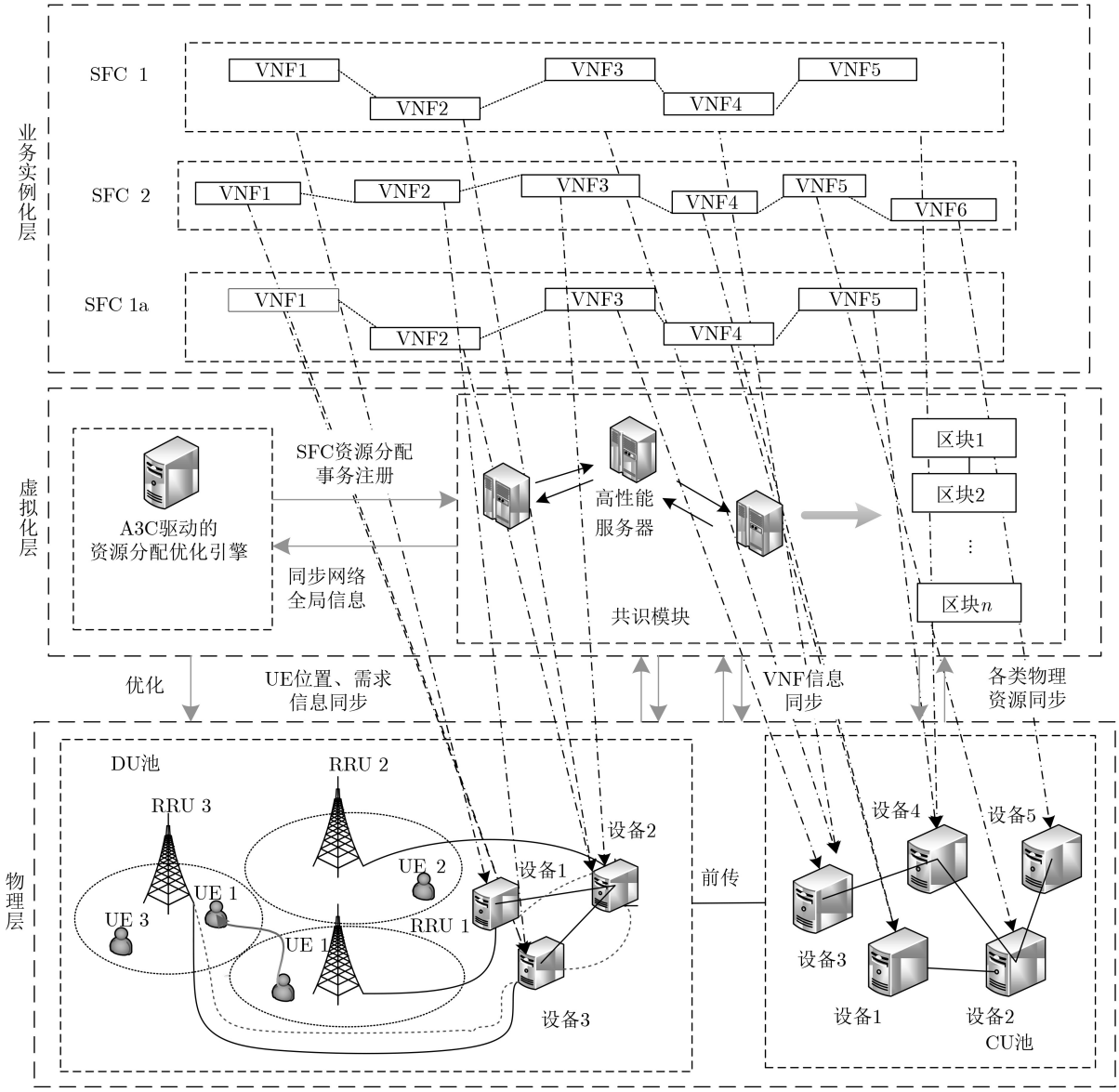

图 1 接入网切片SFC资源分配框架

点对点(Peer to Peer, P2P)网络进行信息泛洪，并通过共识过程保证各个物理设备上的信息同步且一致，实现网络全局信息可信可靠的分布式存储记录。本文以联盟区块链的形式构建分布式账本，相比于公有区块链更加高效[13]。物理节点分为联盟成员和轻节点两类，目前存在的共识算法包括工作证明，股权证明，拜占庭容错(Practical Byzantine Fault Tolerance, PBFT)等等[14–16]。对于不需要货币体系的联盟链而言，常采用PBFT算法。进一步，为了减少区块链网络压力和时间开销，可省去传统PBFT算法的确认阶段[17]，因此本文采用此种优化的PBFT算法完成共识。

2.2 基于区块链的资源管理机制

SFC资源分配过程主要分为全局信息同步和资源分配两个模块，分为3个步骤：

(1) 全局信息同步。不同UE会由自身私钥对最新的位置信息、QoS信息以及数据包到达信息等进行签名，不同的物理设备也会由其自身私钥对最新的各类物理资源容量信息以及VNF放置和资源分配情况信息进行签名。而后，这些信息经过P2P网络进行泛洪，联盟成员基于优化的PBFT算法执行共识过程，生成一个新的区块并将包含该事务的新区块加入到区块链中。

(2) 基于A3C的SFC资源分配。基于A3C的网络优化引擎通过同步区块链数据查询到最新的全局信息，而后通过A3C算法实现SFC资源分配策略的求解。

(3) 服务供应。根据策略优化结果，完成物理设备、链路上的VNF放置和各类资源分配，为UE提供服务。

3 问题建立

3.1 物理网络模型

用带权无向图$G=\{N,L\}$表示基础设施网络，其中，$N=N_{\mathrm{D}}\cup N_{\mathrm{C}}$为物理设备节点集，由DU池节点集$N_{\mathrm{D}}$和CU池节点集$N_{\mathrm{C}}$组成。$L=L_{\mathrm{D}}\cup L_{\mathrm{C}}\cup L_{\mathrm{FH}}$代表物理链路集，由DU池链路集$L_{\mathrm{D}}$、CU池链路集$L_{\mathrm{C}}$和前传网络$L_{\mathrm{FH}}$构成。物理设备节点$n$的计算资源容量为$C_n$，链路$l$的带宽资源容量为$B_l$，l.head和l.tail代表连接$l$的两个相邻物理节点。此外，系统中RRU集合为$K$, RRU$k$的物理资源块(Physical Resource Block, PRB)容量表示为RB_k, SFC集合为M, UE集合为U。UEu请求服务的QoS需求通过两个参数来描述：最小可接受传输速率$r^u_{\min}$和最大可容忍时延Dl_u。

3.2 资源分配模型

当检测到系统内有UE移动时，首先需要对无线资源进行调整。令2元变量$x^u_k(t)$=1代表UEu与RRUk相连，否则$x^u_k(t)$= 0。$n^u_k(t)$代表UEu在RRUk上占用的PRB个数。因此，RRUk的无线资源分配策略表示为$a^r_k(t)=\{x^u_k(t)\cdot n^u_k(t)|u\in U\}$。在时隙$t$系统内所有RRU的无线资源分配方式为$a_r(t)=\{a^r_k(t)|k\in K\}$。

在资源分配方案调整过程中，会涉及到部分VNF的迁移，因此需要调整物理设备上的计算资源分配方式。设UEu对应的SFC表示为m_u，SFCm_u的VNF集合表示为F_{m_u}。令2元变量$y^n_{m_u,j}(t)=1$代表SFCm_u的VNFj放置在物理设备n上，否则$y^n_{m_u,j}(t)=0$。$c^n_{m_u,j}(t)$则为VNFj在物理设备n上所占用的计算资源。假定每台设备上包含有多个CPU，单个CPU的计算能力为C_{cpu}(CPU cycles/s)。从而，SFCm_u的计算资源分配策略表示为$a^c_{m_u}(t)=\{y^n_{m_u,j}(t)\cdot c^n_{m_u,j}(t)|j\in F_{m_u},n\in N\}$。因此，系统内所有SFC在时隙$t$的计算资源分配方式表示为$a_c(t)=\{a^c_{m_u}(t)|u\in U\}$。

除了上述无线资源和计算资源的分配，还需要进行链路资源的分配，包括物理设备之间的链路带宽分配、前传网络带宽分配以及UE移动后可能产生的新路径$p_u(t)$上的带宽资源分配。令2元变量$z^l_{m_u,j}(t)=1$代表时隙t中SFCm_u的VNFj选择链路l向下一个VNF传输数据，否则$z^l_{m_u,j}(t)=0$。$b^l_{m_u,j}(t)$则代表VNFj在链路l上占用的带宽资源。因此，令$a^b_{m_u}(t)=\{z^l_{m_u,j}(t)\cdot b^l_{m_u,j}(t)|j\in F'_{m_u}\}$表示SFC$m_u$在物理链路上的带宽资源分配策略。其中，$F'_{m_u}$代表不包括DU池和CU池末端VNF的集合。此外，为了表示统一起见，令SFCm_u在前传链路上的带宽分配方式表示为$a^{\mathrm{fh}}_{m_u}(t)=\{b^{\mathrm{fh}}_{m_u}(t)\}$。

新路径$p_u(t)$的起点是UEu所连接的RRUk_u，终点则是SFCm_u的第1个VNF所在的物理设备。$p_u(t)$可通过链路状态路由算法得到，则UEu将沿着所得路径$p_u(t)$将数据从其连接的RRUk_u传输到对应的SFCm_u，因此，还需要为这些新路径分配带宽资源。令2元变量$z^l_{p_u}(t)=1$代表时隙t物理链路l在路径$p_u(t)$上，否则$z^l_{p_u}(t)=0$。$b^l_{p_u}(t)$则代表路径$p_u(t)$在链路l上占用的带宽资源。因此，路径$p_u(t)$的带宽资源分配策略表示为$a^b_{p_u}(t)=\{z^l_{p_u}(t)\cdot b^l_{p_u}(t)|l\in L\}$，进而$a_b(t)=\{(a^b_{m_u}(t),a^{\mathrm{fh}}_{m_u}(t),a^b_{p_u}(t))|u\in U\}$表示系统中所有SFCs在时隙$t$的链路带宽资源分配方式。

3.3 优化模型

将系统的时间维度分为若干个时隙，用$\Gamma=\{1,2,\cdots,t,\cdots,T\}$表示时隙集合，$T_s$为每个时隙$t$的持续时间。UE$u$的数据包到达过程为服从时变参数为$\lambda_u(t)$的泊松分布。首先，对于无线传输时延，由香农公式可得与RRUk关联的UEu的可达无线传输数据率为

$$r^u_k(t)=\left(\sum_{k\in K}x^u_k(t)\cdot n^u_k(t)\right)\cdot B\cdot\log_2\left(1+\mathrm{SNR}^u_k(t)\right) \tag{1}$$

其中，B为单个PRB的带宽，$\mathrm{SNR}^u_k(t)$为信噪比。

因此，UEu在无线信道上的传输时延$d^u_r(t)$可表示为

$$d^u_r(t)=\lambda_u(t)/r^u_k(t) \tag{2}$$

其次，对于物理设备处理时延，不同UE的SFC请求处理单位比特数据所需的CPU cycles有所差异，设为x_{m_u}，则UEu在物理设备上的处理时延$d^u_c(t)$表示为

$$d^u_c(t)=\sum_{j\in F_{m_u}}\frac{\lambda_u(t)\cdot x_{m_u}}{\sum\limits_{n\in N}y^n_{m_u,j}(t)\cdot c^n_{m_u,j}(t)} \tag{3}$$

再者，对于链路传输时延，包括物理链路、前传链路和新路径$p_u(t)$3部分传输时延，则UEu在物理链路上的传输时延$d^u_{\mathrm{pl}}(t)$表示为

$$d^u_{\mathrm{pl}}(t)=\sum_{j\in F_{m_u}}\frac{\lambda_u(t)}{\sum\limits_{l\in L}z^l_{m_u,j}(t)\cdot b^l_{m_u,j}(t)} \tag{4}$$

同理，UEu在前传链路上的传输时延$d^u_{\mathrm{fh}}(t)$表示为

$$d^u_{\mathrm{fh}}(t)=\lambda_u(t)/b^{\mathrm{fh}}_{k_u}(t) \tag{5}$$

进一步，数据从连接的RRU传输到对应的SFC需要一条新路径$p_u(t)$，UEu在路径$p_u(t)$上的传输时延$d^u_p(t)$表示为

$$d^u_{p_u}(t)=\sum_{l\in L}z^l_{p_u}(t)\cdot\frac{\lambda_u(t)}{b^l_{p_u}(t)+\Delta} \tag{6}$$

其中，Δ是一个极小的常数，以避免分母为0的情况。

综上，UEu请求数据的链路传输时延$d_b^u(t)$表示为

$$d_b^u(t)=d_{\mathrm{pl}}^u(t)+d_{\mathrm{fh}}^u(t)+d_{p_u}^u(t) \tag{7}$$

最后，不容忽略的是VNF迁移将引起额外的网络重配置时延[18]，因此需要控制网络中的VNF迁移动作。随着各个设备上VNF的迁移，各条物理链路的状态也将发生变化，若在时隙$t+1$，SFCm_u的VNFj迁移到了其他物理设备，则$z_{m_u,j}^l(t)=0$。链路状态的切换需要在链路两端的物理设备上重新配置服务和路由，设一台设备上的重配置时延为q_{re}，则UEu所对应SFCm_u由于VNF迁移引起的网络重配置时延$d_{\mathrm{re}}^u(t)$表示为

$$d_{\mathrm{re}}^u(t)=2\cdot\sum_{j\in F'_{m_u}}\sum_{l\in L}q_{\mathrm{re}}\cdot\max\{z_{m_u,j}^l(t+1)-z_{m_u,j}^l(t),0\} \tag{8}$$

从而，在时隙t内UEu传输数据的时延$d_u(t)$表示为

$$d_u(t)=d_r^u(t)+d_c^u(t)+d_b^u(t)+d_{\mathrm{re}}^u(t) \tag{9}$$

则系统中所有UE传输数据包的接入网切片总时延$d(t)$表示为

$$d(t)=\sum_{u\in U}d_u(t) \tag{10}$$

因此，系统中所有UE传输数据包的总平均接入网切片时延d表示为

$$d=\lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left\{\sum_{t=0}^{T}d(t)\right\} \tag{11}$$

综上，本文接入网切片场景的SFC资源分配问题可建立为基于无线资源、计算资源和链路带宽资源联合分配的时延最小化数学模型

$$\begin{aligned}
&\min_{a_r(t),a_c(t),a_b(t)}\{d\}\\
&\text{s.t.}\ \ \mathrm{C1}:\sum_{u\in U}x_k^u(t)\cdot n_k^u(t)\le \mathrm{PRB}_k,\forall k\in K\\
&\quad\mathrm{C2}:\sum_{u\in U}\sum_{j\in F_{m_u}}y_{m_u,j}^n(t)\cdot c_{m_u,j}^n(t)\le C_n,\forall n\in N\\
&\quad\mathrm{C3}:\sum_{u\in U}\left(\sum_{j\in F'_{m_u}}z_{m_u,j}^l(t)\cdot b_{m_u,j}^l(t)\right)+z_{p_u}^l(t)\cdot b_{p_u}^l(t)\le B_l,\forall l\in L\\
&\quad\mathrm{C4}:\sum_{u\in U}b_{m_u}^{\mathrm{fh}}(t)\le B_{\mathrm{fh}}\\
&\quad\mathrm{C5}:\sum_{k\in K}x_k^u(t)=1,\forall u\in U\\
&\quad\mathrm{C6}:\sum_{n\in N}y_{m_u,j}^n(t)=1,\forall m_u\in M,\forall j\in F_{m_u}\\
&\quad\mathrm{C7}:\sum_{l\in L}z_{m_u,j}^l(t)\le 1,\forall m_u\in M,\forall j\in F'_{m_u}\\
&\quad\mathrm{C8}:y_{m_u,j}^n(t)=\sum_{n=\mathrm{l.head}}z_{m_u,j}^l(t),\forall m_u\in M,\forall j\in F_{m_u}\\
&\quad\mathrm{C9}:y_{m_u,j}^n(t)=\sum_{n=\mathrm{l.head}}z_{m_u,j-1}^l(t),\forall m_u\in M,\forall j-1\in F_{m_u}\\
&\quad\mathrm{C10}:x_k^u(t)\cdot n_k^u(t)=n_k^u(t),\forall u\in U\\
&\quad\mathrm{C11}:y_{m_u,j}^n(t)\cdot c_{m_u,j}^n(t)=c_{m_u,j}^n(t),\forall m_u\in M,\forall j\in F_{m_u}\\
&\quad\mathrm{C12}:z_{m_u,j}^l(t)\cdot b_{m_u,j}^l(t)=b_{m_u,j}^l(t),\forall m_u\in M,\forall j\in F'_{m_u}\\
&\quad\mathrm{C13}:z_{p_u}^l(t)\cdot b_{p_u}^l(t)=b_{p_u}^l(t),\forall u\in U\\
&\quad\mathrm{C14}:r_k^u(t)\ge r_{\min}^u,\forall u\in U,\\
&\quad\mathrm{C15}:d_u(t)\le Dl_u,\forall u\in U
\end{aligned} \tag{12}$$

在上述约束条件中，C1～C4分别代表无线资源、计算资源和带宽资源分配约束；C5限制任意UE只能连接到1个RRU；C6限制任意VNF只能实例化在1台物理设备上；C7限制任意VNF至多只能选择1条物理链路传输数据；C8和C9确保任意SFC上相邻的两个VNF若是部署在不同的物理设备上，则这两台设备必须相邻；C10表示任意UE只有连接到RRU才分配无线资源；C11和C12分别表示任意SFC只有当其虚拟节点即VNF映射到物理节点、虚拟链路映射到物理链路时，才分配计算资源

和带宽资源；C13表示任意UE产生的新路径只有映射到了实际物理链路上才分配带宽资源；C14和C15确保任意UE的QoS得到满足，即无线传输速率高于最小可接受传输速率，数据传输时延低于最大可容忍时延。

4 基于A3C学习的SFC资源分配算法

4.1 MDP

前述SFC资源分配过程可以建模为离散时间的MDP。MDP定义为一个多元组$M=<\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R}>$，其中$\mathcal{S}$是状态空间，$\mathcal{A}$是动作空间，$\mathcal{P}$是转移概率，$\mathcal{R}$是奖励函数。

令$s^{(t)}$代表时隙t的系统状态，由基于UE请求QoS满意度的全体SFCs状态决定，表示为

$$s^{(t)}=\{s_1(t),s_2(t),\cdots,s_U(t)\} \tag{13}$$

其中$s_u(t)=\{0,1\}$。

根据网络当前的资源信息，智能体采取动作即各类资源的分配方式。则在时隙t采取的动作$a^{(t)}$表示为

$$a^{(t)}=\{a_r(t),a_c(t),a_b(t)\} \tag{14}$$

其中，$a_r(t)$代表所有UE在时隙t的无线资源分配方式，$a_c(t)$和$a_b(t)$分别代表所有SFC的计算资源和带宽资源分配方式。在状态$s^{(t)}$采取动作$a^{(t)}$后，系统中某些UE的QoS满意度可能会发生变化，即转移到下一状态$s^{(t+1)}$，此时系统会得到一个立即回报R_t，表示为

$$R_t=-d(t) \tag{15}$$

4.2 A3C学习过程

由于UE的移动性和数据包到达的动态性，系统需要支持需求驱动和自动调整的服务供应，同时考虑到动作空间的连续性，本文引入了A3C学习来优化SFC的资源分配策略。该强化学习算法能并行地在环境中执行多个智能体的概念，不需要经验池也能很好地进行更新[19]。

本文将每条SFC都视为一个智能体，则智能体集合为M。采用参数向量$\boldsymbol{\theta}_a=(\theta_a^1,\theta_a^2,\cdots,\theta_a^n)^{\mathrm{T}}$来构建策略$\pi(a|s)$，采用参数向量$\boldsymbol{\theta}_c=(\theta_c^1,\theta_c^2,\cdots,\theta_c^n)^{\mathrm{T}}$来构建状态值函数$V(s^{(t)})$。因此，对于当前网络状态$s^{(t)}$，各个智能体维护一个参数化随机动作策略$\pi(a^{(t)}|s^{(t)};\boldsymbol{\theta}_a)$，以及一个参数化状态值函数$V(s^{(t)};\boldsymbol{\theta}_c)$。

对于每一个智能体m，首先，定义状态值函数$V(s^{(t)})$，实际上就是折扣奖励期望值，表示为

$$V(s^{(t)})=\mathrm{E}_\pi[R_t+\beta V(s^{(t+1)})] \tag{16}$$

其中$\beta\in(0,1)$是衡量当前和未来决策的折扣因子，$\mathrm{E}\{\cdot\}$表示期望。式(16)意味着当前状态$s^{(t)}$的回报值为立即回报R_t与转移到下一状态$s^{(t+1)}$的折扣值函数之和。

此外，与状态值函数$V(s^{(t)})$类似，还需要定义状态-动作值函数$Q(s^{(t)},a^{(t)})$，表示为

$$Q(s^{(t)},a^{(t)})=\mathrm{E}_\pi[R_t+\beta Q(s^{(t+1)},a^{(t+1)})] \tag{17}$$

A-C学习过程中，采用策略梯度法对参数进行更新，为了有效降低梯度计算的方差，并进一步提高评论家部分函数近似的精度，引入了优势函数$A(s^{(t)},a^{(t)})$，表示为

$$A(s^{(t)},a^{(t)})=Q(s^{(t)},a^{(t)})-V(s^{(t)}) \tag{18}$$

为了对策略π进行更新，定义目标函数$J(\pi)$表示为

$$J(\pi)=\mathrm{E}_\pi(V(s^{(0)}))=\int_{\mathcal{S}}d^\pi(s)\int_{\mathcal{A}}\pi(a|s)V(s)\mathrm{d}a\mathrm{d}s \tag{19}$$

其中，$d^\pi(s)=\lim\limits_{t\to\infty}\Pr(s^{(t)}=s|s^{(0)},\pi)$是在策略$\pi$下的稳定状态分布。

演员部分负责更新策略参数向量$\boldsymbol{\theta}_a$，其策略梯度公式表示为

$$\nabla_{\boldsymbol{\theta}_a}J(\pi)=\mathrm{E}_{s\sim d^\pi(s),a\sim\pi(s)}\left[A(s,a)\cdot\nabla_{\boldsymbol{\theta}_a}\lg\pi(a|s)\right] \tag{20}$$

在A3C学习中，采用N步采样来加速收敛，状态值函数$V_m(s^{(t)})$满足如式(21)贝尔曼公式

$$V(s^{(0)})\to R_0+\beta R_1+\ldots+\beta^nV(s^{(N)}) \tag{21}$$

令$\boldsymbol{\theta}_a$和$\boldsymbol{\theta}_c$表示全局A-C网络中的参数向量，对应地，$\boldsymbol{\theta}_a^{'}$和$\boldsymbol{\theta}_c^{'}$表示本地A-C网络中的参数向量。将优势函数$A(s^{(t)},a^{(t)})$展开，则在演员部分更新策略参数向量$\boldsymbol{\theta}_a$的梯度表示为

$$\begin{aligned}\mathrm{d}\boldsymbol{\theta}_a\leftarrow{}&\mathrm{d}\boldsymbol{\theta}_a+\nabla_{\boldsymbol{\theta}_a^{'}}\lg\pi(a^{(t)}|s^{(t)};\boldsymbol{\theta}_a^{'})\\&\cdot\left[\sum_{i=1}^{T_{\mathrm{loc}}-1}\beta^iR_{t+i}+\beta^{T_{\mathrm{loc}}}V(s^{(t+T_{\mathrm{loc}})},\boldsymbol{\theta}_c^{'})\right.\\&\left.-V(s^{(t)},\boldsymbol{\theta}_c^{'})\right]+\delta\nabla_{\boldsymbol{\theta}_a^{'}}H[\pi(s^{(t)};\boldsymbol{\theta}_a^{'})]\end{aligned} \tag{22}$$

其中，T_{loc}为本地网络单次迭代时间序列最大长度，H代表策略的熵，δ为熵超参数。

评论家部分状态值函数中的参数向量$\boldsymbol{\theta}_c$通过TD模式的梯度下降进行更新，为

$$\begin{aligned}\mathrm{d}\boldsymbol{\theta}_c\leftarrow{}&\mathrm{d}\boldsymbol{\theta}_c+\partial\left[\sum_{i=0}^{T_{\mathrm{loc}}-1}\beta^iR_{t+i}+\beta^{T_{\mathrm{loc}}}V(s^{t+T_{\mathrm{loc}}};\nabla_c^{'})\right.\\&\left.-V(s^{(t)};\nabla_c^{'})\right]^2\Big/\partial\boldsymbol{\theta}_c^{'}\end{aligned} \tag{23}$$

5　仿真与性能分析

A3C模块基于Python3.6平台和Tensorflow工具实现。区块链系统在Docker 18.06环境下搭载Hyperledger Fabric 1.4版本实现[20]，并使用Caliper区块链性能测试框架进行测试。仿真场景中，UE数据包到达率(包/s)范围为$\lambda_u(t) \in [40, 80]$。数据包大小均值设置为$\bar{p} = 500$ kbit。处理单位比特数据所需CPU cycles取值为$\{5900, 6400, 7000\}$。SFC总数量取值范围为$[10, 50]$条，VNF序列长度取值范围为$[9, 11]$。仿真中利用函数随机生成DU池和CU池的基础设施网络，其物理设备数分别为12和18个。任意设备上的计算资源和任意链路上的带宽资源均随机取值，其中计算资源由CPU个数和单个CPU计算能力定义，取值范围分别为$[4, 10]$个，$[10^{11}, 10^{12}]$CPU cycles，链路带宽取值范围$[100, 200]$ Mbps，前传网络带宽为1000 Mbps。此外，DU池中RRU的数目为4个，各个RRU提供的PRB数量为250个，单个PRB的带宽为360 kHz。A3C学习过程设置折扣因子β=0.9，熵超参数δ=0.01。全局A-C网络最大步数$T_{g_\max}$=800，本地A-C网络最大步数$T_{l_\max}$=100，全局A-C网络更新频率$T_{\text{up_iter}}$=10。演员学习率ε_a=0.0001，评论家学习率ε_c=0.01。

图2描述了部署4, 6, 8个联盟成员对区块链共识时延的影响。一方面，随着系统中SFCs数量的增加，区块链共识时延随之升高，这是由每条SFC对应的UE信息、以及各自的VNF放置信息、资源分配信息等都属于网络全局信息需要进行同步所导致。另一方面，联盟成员数量的增加也会导致共识时延的升高。在优化的PBFT算法中，虽然部署更多的联盟成员可以提高安全性和容错性，但同时也会增大PBFT各个阶段的信息广播、交互过程的时间开销，从而导致共识时延升高。

图3描述了不同事务请求发送速率下，共识节点的CPU使用率。其中，在Caliper区块链性能测试框架中，共识请求发送速率单位为每秒传输的事务个数。随着事务请求发送速率的不断增加，由于需要进行更多的共识过程，因此CPU使用率逐渐升高，同时平均请求成功接受率下降。所部署的联盟成员的数量越多，意味着将进行更为复杂的共识过程，安全性和容错性也得到提升，因此会占用更多的CPU资源。

设置系统中SFC条数为50。取值δ=0.01, δ=0.05以及δ=0.001时的算法收敛过程如图4所示。在800个学习回合中，不同δ取值的最终收敛系统时延值较为接近，但是当δ=0.01，曲线的波动或突变程度较小，且收敛速度更快。因此，在后续仿真过程中采用熵超参数δ=0.01。

图5所示为不同算法在不同SFC数量下的节点计算资源。方差越小说明VNF的放置和互连以及多条SFC之间的资源分配更加合理。基于A3C学习

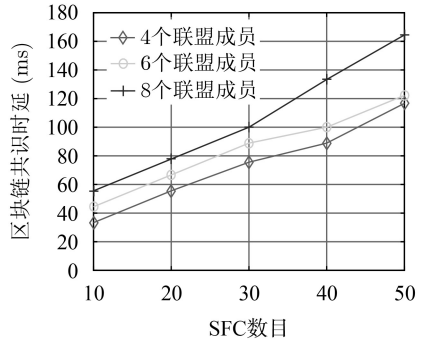

图 2 SFC数目与区块链共识时延关系图

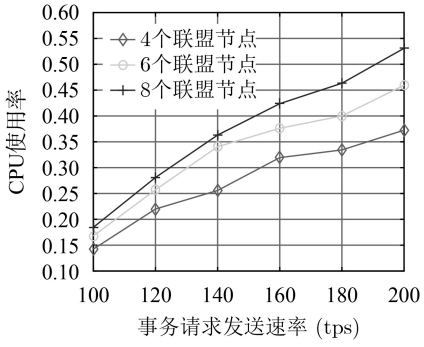

图 3 区块链节点CPU使用率

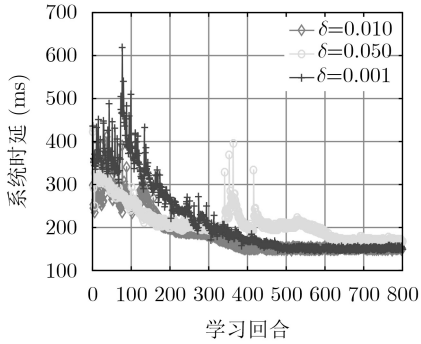

图 4 不同熵超参数δ的A3C算法收敛性

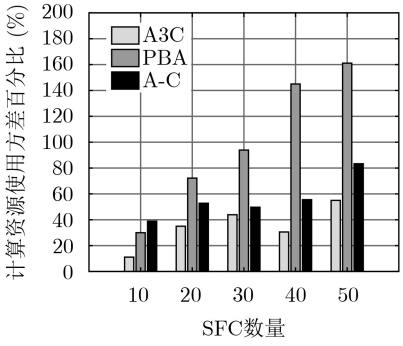

图 5 不同学习算法的资源使用方差百分比

的SFC资源分配算法的结果都明显低于基于A-C学习和PBA的算法，这是因为A3C采用多智能体并行学习，能够更好地与环境进行交互，制定出更为合理的资源分配策略，而更加均匀地资源分配也是系统时延性能更为优越的直接原因。

6 结束语

本文考虑网络全局信息难以获悉的实际情况，针对接入网侧UE的移动性以及业务到达的随机性和动态性引起的系统时延问题，提出了一种基于A3C学习的SFC资源分配算法。本算法通过引入区块链技术实现全局信息的“分布式账本式”存储和管理。考虑到UE的移动性，建立以最小化时延为目标的SFC多维资源联合优化模型，并采用A3C学习算法进行资源分配策略求解。仿真结果表明，本算法能够更加合理高效地利用资源，优化系统时延并保证UE需求。

参 考 文 献

[1] OTOKURA M, LEIBNITZ K, KOIZUMI Y, *et al*. Evolvable virtual network function placement method: Mechanism and performance evaluation[J]. *IEEE Transactions on Network and Service Management*, 2019, 16(1): 27–40. doi: 10.1109/TNSM.2018.2890273.

[2] CABALLERO P, BANCHS A, DE VECIANA G, *et al*. Network slicing games: Enabling customization in multi-tenant mobile networks[J]. *IEEE/ACM Transactions on Networking*, 2019, 27(2): 662–675. doi: 10.1109/TNET.2019.2895378.

[3] ALQERM I and SHIHADA B. Sophisticated online learning scheme for green resource allocation in 5G heterogeneous cloud radio access networks[J]. *IEEE Transactions on Mobile Computing*, 2018, 17(10): 2423–2437. doi: 10.1109/TMC.2018.2797166.

[4] DEMIR M S, SAIT S M, and UYSAL M. Unified resource allocation and mobility management technique using particle swarm optimization for VLC networks[J]. *IEEE Photonics Journal*, 2018, 10(6): 7908809. doi: 10.1109/JPHOT.2018.2864139.

[5] DASTGHEIB M A, BEYRANVAND H, SALEHI J A, *et al*. Mobility-aware resource allocation in VLC networks using T-step look-ahead policy[J]. *Journal of Lightwave Technology*, 2018, 36(23): 5358–5370. doi: 10.1109/JLT.2018.2872869.

[6] 唐伦, 周钰, 谭颀, 等. 基于强化学习的5G网络切片虚拟网络功能迁移算法[J]. 电子与信息学报, 2020, 42(3): 669–677. doi: 10.11999/JEIT190290.

TANG Lun, ZHOU Yu, TAN Qi, *et al*. Virtual network function migration algorithm based on reinforcement learning for 5G network slicing[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 669–677. doi: 10.11999/JEIT190290.

[7] SHARMA P K, CHEN M Y, and PARK J H. A software defined fog node based distributed blockchain cloud architecture for IoT[J]. *IEEE Access*, 2017, 6: 115–124. doi: 10.1109/ACCESS.2017.2757955.

[8] XIE Lixia, DING Ying, YANG Hongyu, *et al*. Blockchain-based secure and trustworthy Internet of Things in SDN-enabled 5G-VANETs[J]. *IEEE Access*, 2019, 7: 56656–56666. doi: 10.1109/ACCESS.2019.2913682.

[9] SUN Yao, FENG Gang, QIN Shuang, *et al*. The SMART handoff policy for millimeter wave heterogeneous cellular networks[J]. *IEEE Transactions on Mobile Computing*, 2018, 17(6): 1456–1468. doi: 10.1109/TMC.2017.2762668.

[10] LI Junling, SHI Weisen, ZHANG Ning, *et al*. Reinforcement learning based VNF scheduling with end-to-end delay guarantee[C]. 2019 IEEE/CIC International Conference on Communications in China (ICCC), Changchun, China, 2019: 572–577. doi: 10.1109/ICCChina.2019.8855889.

[11] LI Guanglei, ZHOU Huachun, FENG Bohao, *et al*. Efficient provision of service function chains in overlay networks using reinforcement learning[J]. *IEEE Transactions on Cloud Computing*, To be pulished. doi: 10.1109/TCC.2019.2961093.

[12] GRONDMAN I, BUSONIU L, LOPES G A D, *et al*. A survey of actor-critic reinforcement learning: Standard and natural policy gradients[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, 42(6): 1291–1307. doi: 10.1109/TSMCC.2012.2218595.

[13] 朱立, 俞欢, 詹士潇, 等. 高性能联盟区块链技术研究[J]. 软件学报, 2019, 30(6): 1577–1593. doi: 10.13328/j.cnki.jos.005737.

ZHU Li, YU Huan, ZHAN Shixiao, *et al*. Research on high-performance consortium blockchain technology[J]. *Journal of Software*, 2019, 30(6): 1577–1593. doi: 10.13328/j.cnki.jos.005737.

[14] KIAYIAS A, RUSSELL A, DAVID B, *et al*. Ouroboros: A provably secure proof-of-stake blockchain protocol[C]. The 37th Annual International Cryptology Conference, Santa Barbara, USA, 2017: 357–388. doi: 10.1007/978-3-319-63688-7_12.

[15] YAO Yingying, CHANG Xiaolin, MIŠIĆ J, *et al*. BLA: Blockchain-assisted lightweight anonymous authentication for distributed vehicular fog services[J]. *IEEE Internet of Things Journal*, 2019, 6(2): 3775–3784. doi: 10.1109/JIOT.2019.2892009.

[16] CHEN Zhonglin, CHEN Shanzhi, XU Hui, *et al*. A security

authentication scheme of 5G ultra-dense network based on block chain[J]. *IEEE Access*, 2018, 6: 55372–55379. doi: 10.1109/ACCESS.2018.2871642.

[17] HE Li and HOU Zhixin. An improvement of consensus fault tolerant algorithm applied to alliance chain[C]. The IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2019: 1–4. doi: 10.1109/ICEIEC.2019.8784495.

[18] GUO Shaoyong, DAI Yao, XU Siya, *et al*. Trusted cloud-edge network resource management: DRL-driven service function chain orchestration for IoT[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 6010–6022. doi: 10.1109/JIOT.2019.2951593.

[19] WEI Qinglai, WANG Lingxiao, LIU Yu, *et al*. Optimal elevator group control via deep asynchronous actor-critic learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(12): 5245–5256. doi: 10.1109/TNNLS.2020.2965208.

[20] 戴鹏. 基于实用拜占庭共识算法(PBFT)的区块链模型的评估与改进[D]. [硕士论文], 北京邮电大学, 2019.

DAI Peng. Evalution and research of blockchain model based on practical byzantine consensus algorithm (PBFT)[D]. [Master dissertation], Beijing University of Posts and Telecommunications, 2019.

唐　伦：男，1973年生，教授，博士生导师，主要研究方向为新一代无线通信网络、异构蜂窝网络等.

贺小雨：女，1995年生，硕士生，研究方向为网络切片资源分配和强化学习.

王　晓：男，1995年生，硕士生，研究方向为网络切片资源优化和机器学习.

谭　颀：女，1995年生，硕士生，研究方向为5G网络切片、资源分配、随机优化理论.

胡彦娟：女，1992年生，硕士生，研究方向为移动边缘计算中的资源分配和任务卸载.

陈前斌：男，1967年生，教授，博士生导师，主要研究方向为个人通信、多媒体信息处理与传输、下一代移动通信网络、异构蜂窝网络等.

责任编辑：余　蓉