专题：新型网络

基于通信云和承载网协同的算力网络编排技术

曹畅，张帅，刘莹，唐雄燕
（中国联合网络通信有限公司网络技术研究院，北京 100048）

摘　要：面向未来网络中计算与网络紧密结合、“算网一体”的技术发展趋势，提出了基于集中式和分布式两种控制方案的算力网络编排模型，并分别介绍了实现过程的关键技术。从方案与技术分析来看，基于电信运营商通信云和承载网协同的算力网络编排方案可以较好地适应未来移动边缘计算（MEC）站点成网后边边协同与云边协同的业务需求，增强了网络对业务的感知与调度能力，而集中式或分布式控制方案的具体选择与运营商通信云能力和承载网的演进阶段密切相关。

关键词：软件定义网络；网络功能虚拟化；算力网络；网络编排
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2020197

Convergence of telco cloud and bearer network based computing power network orchestration

CAO Chang, ZHANG Shuai, LIU Ying, TANG Xiongyan
Network Technology Research Institute of China United Network Communications Co., Ltd., Beijing 100048, China

Abstract: Facing the combination of computing and network connection in the future network, and the development trend of computing and network integration, a computing power network orchestration model based on centralized and distributed control schemes was proposed. From the perspective of solution analysis, the computing power network orchestration scheme based on the convergence of telco cloud and bearer network can meet the future business needs of mobile edge computing (MEC) site better. With scheduling capabilities coordinated for both computing and networking, the network’s perception of services can be optimized. The specific choice of centralized or distributed control scheme is closely related to the operator’s telco cloud capabilities and the evolution stage of the bearer network.

Key words: SDN, NFV, computing power network, network orchestration

1　引言

随着物联网、自动驾驶、智能工厂等新兴业务的逐步发展，其大带宽和低时延的特征需要数据进行就近处理和分析，驱动计算从云端下移到接近数据源的边缘，边缘计算、移动边缘计算应

收稿日期：2020−05−10；修回日期：2020−06−30
基金项目：国家重点研发计划基金资助项目（No.2019YFB1802800，No.2019YFB1802600）
Foundation Items: The National Key Research and Development Program of China (No.2019YFB1802800, No.2019YFB1802600)

运而生并且迅速发展，将在可见的未来逐步走向泛在计算。在边缘计算、乃至泛在计算的场景中，由于边缘或者终端侧单个站点的算力资源有限，往往难以单独完成一个复杂任务，需要多站点协作完成，这需要未来大量分布化的边缘计算节点、泛在计算节点连接成网并与云计算协调以实现全网算力的高效使用。

基于上述原因，近年业界提出了算力网络的概念与技术架构。算力网络是为应对算网融合发展趋势而提出的新型网络架构，基于无处不在的网络连接，将动态分布的计算与存储资源互联，通过网络、存储、算力等多维度资源的统一协同调度，使海量的应用能够按需、实时调用不同地方的计算资源，实现连接和算力在网络的全局优化，提供一致的用户体验。算力网络是动态、分布式的计算与网络深度融合的网络模型，通过提供连接与计算一体化服务，提供网络即服务（NaaS）功能，用户通过网络直接获取计算的结果。

算力网络编排管理方案通过网络、存储、算力等多维度资源的统一管理和协同调度，实现连接和算力在网络全局优化。用户通过算力网关（如边缘计算站点）接入网络，设备节点根据应用服务的需求，综合考虑实时的网络和计算资源状况，将不同的应用调度到合适的计算节点处理，保证业务体验。

综上所述，资源的有效编排管理是算力网络的关键技术问题。算力网络编排管理方案的实现有集中式和分布式两种。集中式编排方案是基于数据中心 SDN 集中调度的算力网络编排方案，即在云数据中心中，由多个分布式应用服务器节点构成集群，分担业务计算与存储请求，需要从多个实例中选择一个合适的实例提供服务，同时云数据中心向城域网扩展，与边缘云相连接，通过集中式的 SDN 控制器和 NFVO MANO 实现中心云及边缘云间的算力网络的统一管理和协同调度；分布式编排方案即基于电信运营商承载网分布式控制能力，结合承载网网元自身控制协议扩展，复用现有 IP 网络控制平面分布式协议的方式实现算力信息的分发与基于算力寻址的路由，同时综合考虑实时的网络和计算资源状况，将不同的应用调度到合适的计算节点处理，实现连接和算力在网络的全局优化。

2 通信云中 SDN 和 NFV 的协同方案

云化网络为算力网络提供敏捷和开放的虚拟管道，用于云和端的级互联。运营商通过通信云的建设，充分利用现有机房，构建 NFV 统一资源池，支撑网络云化演进。通信云云化网络总体架构如图 1 所示，沿用传统通信网络接入、城域、

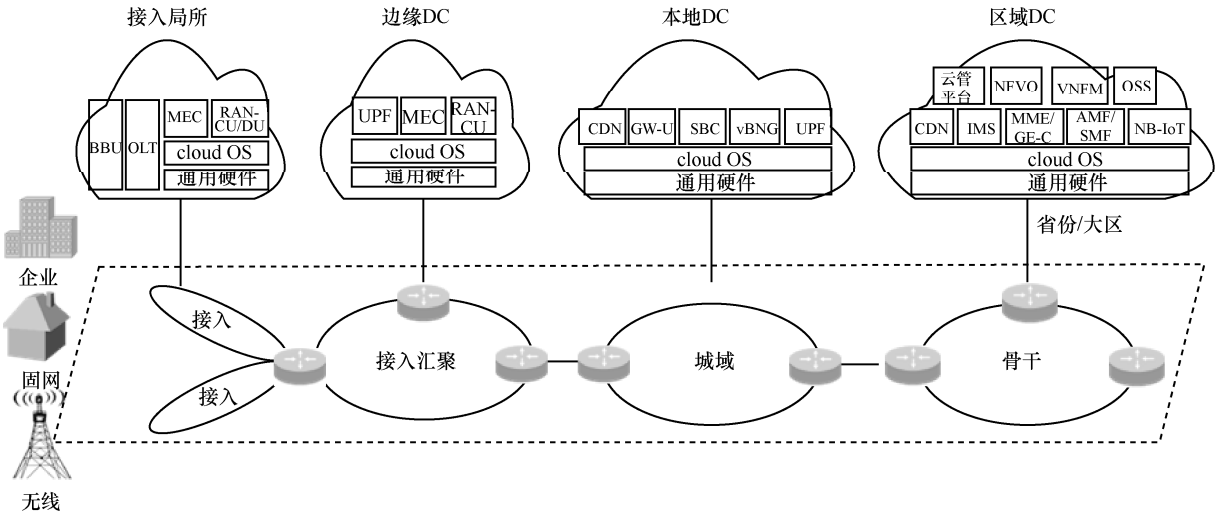


图 1 运营商通信云云化网络总体架构

骨干网络架构，与现有通信局所保持一定的对应和继承关系，在不同层级边缘、本地、区域进行分布式 DC 部署，实现面向宽带网/移动网/物联网等业务的统一接入、统一承载和统一服务。

通信云 DC 内按照叶脊（leaf-spine）架构组网，每部 leaf 节点（接入交换机）同所有 spine 节点（核心交换机）相连构建全连接拓扑，增强网络的可扩展性，也提高了通信的可靠性，提高了整个网络的吞吐量；扁平的网络结构保证了任意节点间较高的连接速率，同时对任意类型流量均拥有较低的时延。

通信云中采用 SDN 和 NFV 的融合部署网络架构，通过 NFVO 实现通信云 DC 内的资源管控，通信云的分层架构如图 2 所示，其中 VIM 是虚拟化基础设施管理系统，主要负责基础设施层硬件资源、虚拟化资源的管理，面向上层 VNFM 提供虚拟化资源池；PIM 对基础设施中的通用设备进行管理。

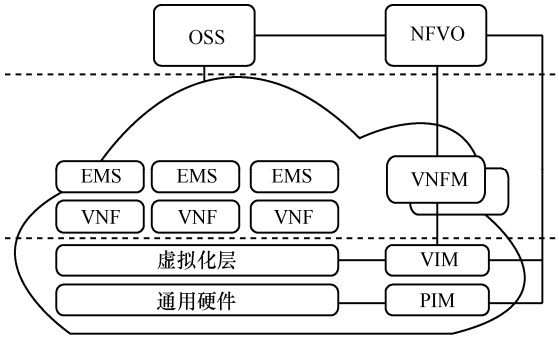


图 2　通信云的分层架构

通信云 SDN 方案有两种，分别是软件 overlay SDN 方案和硬件 SDN 方案。软件 overlay 方案中 SDN 控制器作为集中式控制面，还需纳管 vSwitch，控制粒度更细，同时 SDN 控制器可靠性直接影响网络可靠性，且 vSwitch 复杂性提升，存在转发性能风险。硬件 SDN 方案中 SDN 控制器只管理硬件网络设备，完成配置下发，控制粒度较粗，但网络采用传统成熟硬件设备，可靠性高，vSwitch 做 VLAN 转发，性能高，服务器资源耗费少，由硬件交换机做 VxLAN 转发，可以做到线速转发。

NFVO 和 SDN 控制器协同可实现整网 ICT 资源（物理资源、虚拟资源、VNF）的统一控制和逻辑抽象，并根据业务需要灵活地对 ICT 资源进行细粒度、自定义的编排，使得业务流量按照编排的顺序经过抽象网络功能节点，为云用户提供可靠、可调、差异化的网络服务。

3　基于数据中心 SDN 集中调度的算力网络编排方案

云网融合已经成为 ICT 发展的趋势，伴随着云计算应用的不断落地，云和网络正在打破彼此的界限，相互融合，云在网内、网在云中、网随云动。

在通信云的整体架构中，边缘 DC 可以看作一个运行在电信网络边缘的、运行特定任务的云服务器，它将对应的网络功能部署在最靠近用户的边缘位置，流量在本地被卸载，节省了边缘节点到核心网和 Internet 的传输资源。同时这种边缘计算节点可以灵活部署在不同的网络位置，以适应对时延、带宽有不同需求的计算业务。云数据中心的网络通过 VxLAN+EVPN 技术向城域的扩展，边缘云作为公有云的延伸，将云的部分服务或者能力扩展到边缘节点上，构建用户接入云（中心或边缘）一跳直达的扁平化网络，通过减少跳数，实现网络拥塞点的减少。中心云和云边缘相互配合，实现全网资源共享、全网统一管控等能力。

如同网络资源的集中管理方案，同样需要一个算力网络编排管理平台将算力节点与网络统一纳管，实现算力与网络的调度、匹配，每个边缘节点、每个云池，甚至是每个端设备，都是其管理的一个算力节点。算力网络编排管理平台是通过集中式的控制单元来统一收集全网的算力资源、网络资源以及其他资源信息，用户将业务需求发送给集中控制单元，由该单元利用全局视角进行最优化的资源选择与分配。由于算力网络编

排管理平台不但要收集各类资源信息，同时还要进行相应的抽象与计算，最后需要将算力分配策略发送给用户和算力资源池，并调度网络建立相应的传送通道。因此算力网络编排管理平台需要集成原有网络的 SDN 控制功能、NFV 编排功能等，实现控制与数据平面的分离。这种集中式的控制方案，不改变当前底层网络架构与 IP 协议实现，通过算力网络编排管理平台实现资源的协同管理，调度参数通过松耦合的调度逻辑实现。一方面可以兼容当前单体业务、应用的架构设计资源调度，同时也能支撑各类轻量化微服务架构的细粒度调度。

目前 ETSI 已经定义了 NFV MANO（网络功能虚拟化管理和编排）和 MEC 的架构。NFV MANO 是用于管理和协调虚拟化网络功能（VNF）和其他软件组件的架构框架，MEC 定义为具备无线网络能力开放和运营能力开放的平台，通过 MEAO 可以实现 MEC 资源的管理与调度。将算力网络编排管理平台添加到 ETSI NFV 和 MEC 框架上，网络能力向算力业务开放，依据业务需求，实现算力服务和网络服务相结合的多层异构分布通信网络服务管理和调度，算力网络集中式方案参考架构如图 3 所示。算力网络编排管理平台功能模块包含算力网络编排（computing power network orchestrator，CPNO）模块、算力网络编排管理（computing power network manager，CPNM）模块、算力资源管理（computing power infrastructure manager，CPIM）模块以及算力网络算法管理（computing power network algorithm manager，CPNAM）模块，各模块间相互协作，完成算力资源灵活调度，提高计算资源利用率。

算力网络管理编排平台的主要模块功能如下。

（1）算力网络编排模块：与 NFVO（网络功能虚拟化编排）模块对接，根据业务需求，结合算力资源、网络资源以及其他资源信息，运行优化算法，进行弹性调度，使得资源全局最优。

（2）算力网络管理模块：与 VNFM（虚拟化网络功能管理）对接，管理业务部署节点，包含算力资源基础设施和 NFV 基础设施。

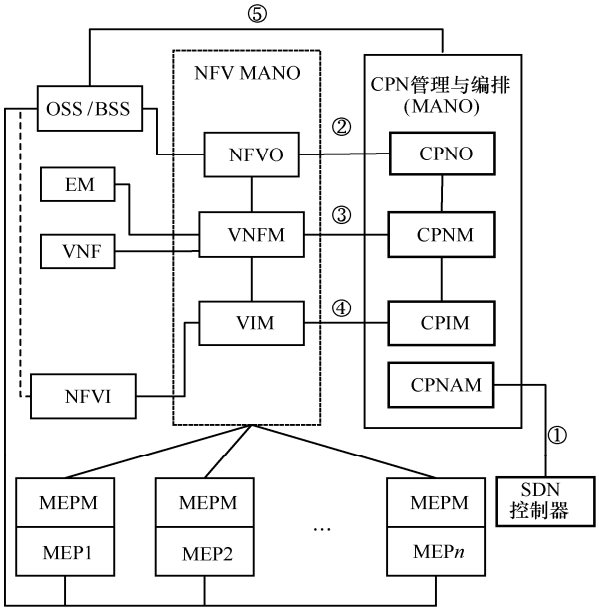


图 3　算力网络集中式方案参考架构

（3）算力资源管理模块：管理通信云（中心与边缘）节点的算力资源，与 VIM（虚拟设施管理）模块对接，根据业务需求为用户分配相应的计算、存储、网络资源，并根据策略对业务部署位置、业务算力进行弹性调整。

（4）算力网络算法管理模块：维护和管理不同的网络节点联合优化算法，根据用户的业务类型和网络实际状况的解析结果，为用户选择合适的部署算法。与网络控制面（SDN 控制器或网管）连接，集中管理用户、边缘云、核心云的网络基础设施，当用户业务部署或调整之后，配置用户到业务处理节点之间的网络，将用户流量路由到处理节点。

此方案基于数据中心 SDN 扩展实现，在云网协同之上进行扩展，广域网能力向算力业务开放，集成算力信息与算力策略，提高算力利用率。

4　基于承载网分布式控制的算力网络编排方案

除了上文介绍的通过扩展数据中心 SDN 控制

器能力，实现核心 DC 与边缘 DC 的统一调度和控制外，也可以采用承载网网元自身控制协议扩展，复用现有 IP 网络控制平面分布式协议的方式实现算力信息的分发与基于算力寻址的路由，具体实现方式如下。

4.1　面向异构算力资源载体的算力状态感知与传递

由于网络中存在着大量的边缘算力节点，对于不同边缘计算站点计算类型和计算能力均可能有差异，也就要求采用不同的部署方式，在网络中发布服务状态前，屏蔽其算力的异构属性。因此，对于各网元存储的算力服务状态的定义需要统一，对服务状态的感知方法需要针对不同部署方式进行设计。首先需要定义算力服务的模板，模板中算力服务的特征可以分为两类：一类是静态特征，一般内容在定义时就已经固化，主要包括服务 ID、服务 locator（IP 端口号）、节点算力总容量，包含计算节点类型、CPU/GPU 性能、存储容量、网络接口带宽等；另一类是动态特征，主要包括当前在线的服务实例数量、CPU/GPU/内存使用率以及当前连接数等。

在算力状态模板构建完成之后，就需要对算力的服务状态进行扩散与同步，这里定义了一种通过 BGP update 消息通告算力信息的方式，通过新定义的计算路经属性（computing path attribute）封装不同应用的算力信息，并且通过承载网网元之间互通信令消息将其传递出去，如图 4 所示，具体步骤如下。

步骤 1　不同应用可使用不同的 status options（sub-TLV 格式）的组合来发布自己的算力信息。算力信息可以有两种形式，其中原始算力指标可具有多个维度，是矢量形式；而融合算力指标是由原始算力指标信息计算得出的标量。

步骤 2　一个 BGP update 消息发布一个应用的算力信息，将不同应用的算力信息放在不同的 BGP update 状态中发送。

步骤 3　将服务状态路由属性封装在不同的 BGP update 报文中，以支持不同场景下的算力通告，包含普通 BGP 以及 BGP VPN、EVPN、labeled BGP 等。

基于 BGP 的算力消息发布，又可以分为两个步骤，分别是 MEC 站点到承载网网元的消息发布以及承载网网元之间的消息发布。前者需要 MEC 站点上的 BGP speaker 与邻近的承载网网元建立邻居关系，BGP speaker 将应用的算力信息封装在 BGP update 报文中，发送给该网元（如图 4 中的 IP 承载网节点），然后 BGP speaker 周期性地发送 update 报文。而后者需要承载网网元之间建立 iBGP 邻居，使用 MP-BGP 传递不同应用的算力信息，通过该过程，承载网路由器可以具有所有应用在所有 MEC 站点上的算力信息。

这里通过与算力路由密切相关的云游戏/cloud VR 场景来说明其所必须的算力指标。云平台需要为每个用户分配独立的云渲染节点，保证

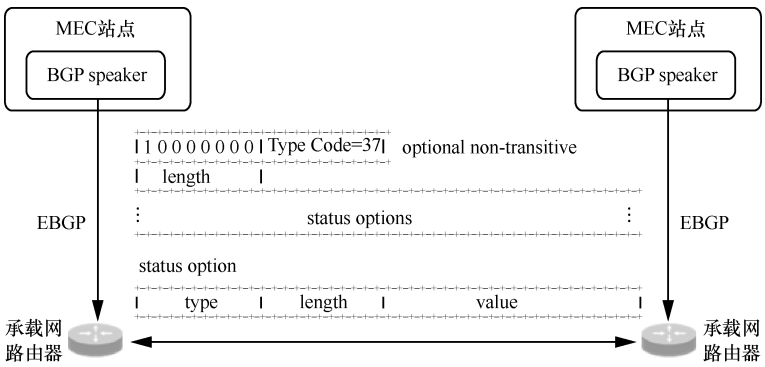


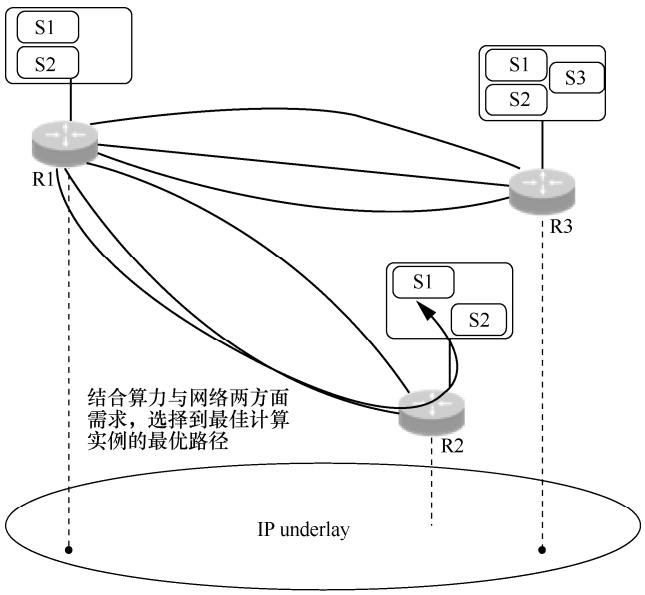
图 4　携带业务算力信息的 BGP 信令过程

每个用户可以独立地进行游戏，互不影响，这就需要借助硬件资源（CPU/GPU/内存）的虚拟化实现。其中，CPU 负责逻辑计算，GPU 负责渲染（细节勾勒），同时需要将硬件资源进行切片分配虚拟机/容器使用，实现多个虚拟机/容器共享一套硬件资源，切片包括时间片段的划分和GPU资源（显存）的划分。MEC 站点在抓取原始游戏画面后，由于数据量巨大，不宜直接传输，需要经过编码压缩，又分为软编码和硬编码两种方式。软编码使用 CPU 进行编码，实现直接、简单，同码率下质量较高，但由于对 CPU 的负载较重，性能相对较低。硬编码主要使用 GPU 进行编码，性能高，速度快，但同码率下质量不如软编码。单个 MEC 站点的算力上限，决定了其可运行的游戏实例数量。因此，在云游戏/云 VR 中，其主要的算力指标建议包含：CPU、GPU 和内存的剩余资源数决定是否能承载新的 VM/容器实例，CPU、GPU 和内存的使用率决定渲染和转码的处理时延，以及到 MEC 站点的网络质量决定音视频的呈现质量和指令的响应质量。

4.2 面向网络算力负载均衡的业务最优路径选择

首先，网络设备作为负载均衡器，目的是将计算任务调度到合适的边缘计算节点，保证业务体验，这就需要网络做到能够基于全局算力状态信息，选择最佳计算实例，特别是在首个数据分组到达检测、服务质量劣化的情况下，这种需求就会更加迫切。其次，负载均衡调度算法需要兼顾性能和稳定性，这就需要考虑计算、网络等多个因素。例如，网络成环，同一设备下挂多个站点（服务器集群），数据流会话保持（session affinity），支持不同协议层级，如 L3/L4（TCP/UDP）、L7（HTTP）等。

在承载网中的节点感知计算实例状态的基础上，可继续叠加应用需求感知（App-aware）和网络状态感知（path-aware），为转发提供多维输入。其中，应用需求感知主要包括基于 App-ID、anycast 地址等识别当前应用以及时延上限、带宽需求、分组丢失率上限等应用的需求匹配。而网络状态感知主要包括：到其他计算节点的 E2E 网络状态以及到其他计算节点的路径信息，基于网络测量确定每条路径性能以及时延、可用带宽、分组丢失率等。算力需求与应用需求、路径状态的对应关系如图 5 所示，是一个在承载网中根据算力信息选择业务最优路径



算力需求

service ID	site	load
S1	R1	90%
S2	R2	30%
S3	R3	20%

网络需求

service ID	requirement
S1	最大时延5 ms
S2	保障带宽大于300 Mbit/s

路径状态

site	path	performance
R1	path 1	latency=1 ms bandwidth=100 Mbit/s
R2	path 2	latency=3 ms bandwidth=500 Mbit/s
R3	path 3	latency=7 ms bandwidth=200 Mbit/s

图 5　算力信息与应用需求、路径状态的对应关系

的具体方式。

在图 5 中，在不同的 MEC 节点存在着 S1、S2 两种服务，均对算力和网络有一定的要求，如图 5 中的算力需求和网络需求所示。结合这两方面要素，图 5 中 3 个路由节点之间通过互通 BGP 消息的方式，业务得出了在 R1、R2 和 R3 3 个不同的节点进行服务的路径信息和相应的性能信息，经过综合分析，网络选择 R2 节点进行服务，业务进入网络后到 R2 的最优路由为 path 2，这样完成了一个根据算力需求和网络需求，通过分布式控制来确定业务路由的过程。

5　结束语

算力网络作为一种融合计算和网络的新型解决方案，在 5G 时代能够有效解决边缘节点之间算力协同，网络适配业务，以提升用户综合业务体验，是未来计算能力下沉后边缘节点提供规模化服务的重要能力抓手。本文介绍了算力网络中的算力路由这一关键技术，给出了 overlay 和 underlay 两种实现方案。其中 overlay 方案基于数据中心 SDN 扩展实现，该方案在数据中心云网协同之上进行扩展，目前已有一定的实现基础，广域网能力向算力业务开放，能做到算力节点的路由可达，配置通过集中式的 SDN 控制器可快速实现。但该方案的问题是计算节点无法与网络属性联动，不能感知途径的承载网节点的实际状态，因此可以作为算力网络初期的实现方案，或者更加适用于不具备基础网络资源但是有平台运营能力的云服务商的技术选择。

另一种 underlay 方式的基于承载网分布式路由的方案，目前主要通过复用网络控制平面，通过 BGP 路由扩展来实现。该方式需要根据具体的业务需求选择 BGP 扩展的种类和形式，实现比较复杂，目前也尚未标准化。但是该方案充分调动了承载网中 IP 路由器节点的控制能力，一方面减轻了核心网负担，另一方面应用可以感知路径中沿途的所有节点的质量，是真正意义上的计算需求向网络开放，实现了计算与网络协同、均衡的提供服务，因此较适用于远期，是一种更加灵活和高效的解决方案，同时也是适合具有基础网络资源的电信运营商采用的技术方案。

综上所述，针对 5G 边缘计算网络承载的具体需求场景，引入算力网络的编排技术是应对网络算力供给不均匀、流量负载不平衡的重要手段。电信运营商应该结合自身网络能力和边缘业务的实际需求，对于算力网络编排技术的具体实现方式进行评估，在不同阶段引入适合自身的技术方案并进行充分的验证，助力业务发展。

参考文献：

[1] 唐雄燕, 曹畅. 中国联通网络重构与新技术应用实践[J]. 中兴通讯技术, 2017(2): 22-23.
TANG X Y, CAO C. Network reconstruction and new technology applications of China Unicom[J]. ZTE Technology Journal, 2017(2): 22-23.

[2] 蒋林涛. 从云网融合到 ICT 基础设施[J]. 信息通信技术, 2019(2): 4-6.
JIANG L T. From cloud networks to ICT infrastructure[J]. Information and Communications Technologies, 2019(2): 4-6.

[3] 李彤, 马季春. 云化背景下运营商数据网演进思路探讨[J]. 邮电设计技术, 2017(10): 1-4.
LI T, MA J C. Discussion on the evolution of carrier's data networks affected by cloud computing[J]. Designing Techniques of Posts and Telecommunications, 2017(10): 1-4.

[4] 雷波, 刘增义, 王旭亮, 等. 基于云、网、边融合的边缘计算新方案：算力网络[J]. 电信科学, 2019, 35(9): 44-51.
LEI B, LIU Z Y, WANG X L, et al. Computing network: a new multi-access edge computing[J]. Telecommunications Science, 2019, 35(9): 44-51.

[5] 付乔, 段晓东, 王路, 等. 边缘电信云架构与关键技术[J]. 电信科学, 2018, 34(7): 8-14.
FU Q, DUAN X D, WANG L, et al. Architecture and key technology of telco edge cloud[J]. Telecommunications Science, 2018, 34(7): 8-14.

[6] 唐雄燕, 周光涛, 赫罡, 等. 新一代网络体系架构 CUBE-Net2.0 研究[J]. 邮电设计技术, 2016(11): 11-12.
TANG X Y, ZHOU G T, HE G, et al. Research on next genera-

tion network architecture CUBE-Net2.0[J]. Architecture and Network Evolution, 2016(11): 11-12.

[7] Cisco Data Center. Spine-and-leaf architecture: design overview white paper[R]. 2019.

[8] 姚慧娟, 耿亮. 面向计算网络融合的下一代网络架构[J]. 电信科学, 2019, 35(9): 38-43.
YAO H J, GENG L. Trend of next generation network architecture: computing and networking convergence evolution[J]. Telecommunications Science, 2019, 35(9): 38-43.

[9] 王海军, 王光全, 郑波, 等. 5G 网络架构及其对承载网的影响[J]. 移动通信, 2018,42(1): 33-38.
WANG H J, WANG G Q, ZHENG B, et al. Architecture of 5G network and its influences on carrying networks[J]. Mobile Communications, 2018, 42(1): 33-38.

[10] 孙嘉琪, 李玉娟, 杨广铭, 等. 5G 承载网演进方案探讨[J]. 移动通信, 2018, 42(1): 1-6.
SUN J Q, LI Y J, YANG G M, et al. Discussion on evolution schemes of 5G carrying network[J]. Mobile Communications, 2018, 42(1): 1-6.

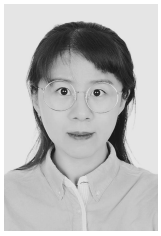
[11] 马季春, 孟丽珠. 面向云网协同的新型城域网[J]. 中兴通讯技术, 2019(4): 37-40.
MA J C, MENG L Z. New metropolitan area network for cloud network synergy[J]. ZTE Technology Journal, 2019(4): 37-40.

[12] 陈运清, 雷波, 谢云鹏. 面向云网一体的新型城域网演进探讨[J]. 中兴通讯技术, 2019(4): 2-8.
CHEN Y Q, LEI B, XIE Y P. Evolution of new metropolitan area network for cloud network convergence[J]. ZTE Technology Journal, 2019(4): 2-8.

[13] 曹畅, 张帅, 唐雄燕. 下一代智能融合城域网方案[J]. 电信科学, 2019, 35(9): 51-59.
CAO C, ZHANG S, TANG X Y. Next generation intelligent converged metro network[J]. Telecommunications Science, 2019, 35(9): 51-59.

[作者简介]

曹畅（1984– ），男，博士，中国联合网络通信有限公司网络技术研究院未来网络研究部高级专家、智能云网技术研究室主任，主要研究方向为 IP 网宽带通信、SDN/NFV、新一代网络编排技术等。



张帅（1990– ），女，中国联合网络通信有限公司网络技术研究院未来网络研究部工程师，主要研究方向为 IP 传输网、数据中心网络、SDN 技术等。

刘莹（1989– ），女，现就职于中国联合网络通信有限公司网络技术研究院未来网络研究部，主要从事 SDN/NFV 应用、网络白盒设备及开放网络融合设备方面的研究工作。

唐雄燕（1967– ），男，中国联合网络通信有限公司网络技术研究院首席科学家，北京邮电大学兼职教授、博士生导师，主要研究方向为宽带通信、光纤传输、互联网/物联网、SDN/NFV 与新一代网络等。