算力网络架构与场景分析

黄光平 罗 鉴 周建锋
中兴通讯股份有限公司 南京 210012

摘 要 为了发挥分布式部署的算力资源池化优势，需要构建一张用于连接这些算力资源的网络，打破孤岛壁垒，为用户应用提供最优算力服务，并最优化全网算力资源利用率。算力网络架构设计必须在应用的算力需求满足和网络改造部署成本之间保持均衡。文章重点分析了集中式算力网络架构、分布式算力网络架构和混合式算力网络架构三种方案在收敛速度、全网资源视图、应用算力资源映射、设备和协议升级等方面的不同特征，并描述各自适用的应用场景。经分析，混合式架构结合了分布式与集中式架构的优势，在实际网络部署中有更大的参考意义。此外，在新型的网络路由架构(比如ICN)下，算力网络将获得更加优异的调度和转发性能。

关键词 算力网络；集中式架构；分布式架构；混合式架构

引言

算力网络，又名算网融合(Computing and Network Convergence)或算力感知网络(Computing Awareness Network)，是随着5G时代更多的算力资源下沉到边缘的行业发展趋势而提出的一种全新解决方案，旨在将终端、边缘、云数据中心的算力资源通过网络连接起来，以全网算力资源池的形态为更趋多样化的应用提供更加灵活优质的算力服务，同时通过算力网络打破分布式算力资源孤岛，提升全网算力资源利用率，避免传统模式下算力节点面对算力服务负荷增加只能进行资源扩容的被动和低效陷阱。

在算力网络体系架构下，应用的算力服务请求将不再局限于特定服务节点的能力，而是会依据应用的算力资源需求和算力服务质量，结合可用的网络路径，将应用请求或应用流路由至最优的算力节点进行处理。在全网算力资源和应用算力需求的双重约束下，算力网络可以依据应用的时延、安全等需求，对应用服务进行分类分解，并将不同类别的算力应用请求路由至不同的算力节点去处理，以提供精准的差异化算力服务，同时最优化应用算力服务质量和全网算力资源利用率。比如在车联网中，将车辆控制算力服务请求路由至最近的算力服务节点，以最大化满足时延需求，而可以将娱乐级算力服务请求路由至远端云数据中心处理。

如上所述，算力网络涉及依据节点算力资源状态和网络路径资源状态联合进行路由决策，但与传统路由转发处理不同的是，算力资源对应的最小处理颗粒度是某种基础或者叫做原子算力服务，而传统路由转发处理的颗粒度是报文或类型流，前者需要网络感知原子算力服务及其算力资源需求，这是传统网络架构体系中尚未涉及的，需要在算力网络架构中进行扩展。

本文将分别就三种算力网络架构方案进行利弊分析，并描述各自适应的典型应用场景。第1部分对算力网络的架构设计考量进行综合阐述和分析，第2部分对算力网络集中式架构进行具体分析，第3部分对算力网络分布式架构进行具体分析，第4部分对算力网络混合式架构进行具体分析，第5部分进行总结性综述。

1 算力网络架构综述

算力网络是对当前IP网络路由机制的一种扩展和延伸，其涉及的功能模块包括管理面、控制面和数据面(也称为转发面或用户面)，节点算力资源的收集和分发由控制面实现，算力原子服务的封装、解封装由数据面实现，算力服务性能监测、算力网络故障监测、算力网络参数配置由管理面实现。

1.1 算力网络控制面

算力节点，包括终端、边缘、云数据中心，其算力资源状态，比如CPU处理能力及占有率、队列状态、缓存状态、算力节点地址、算力节点原子算力实例等信息，需要通过控制面网络机制实现收集和分发，即在全网算力网络转发节点创建全局算力路由表，以备应用请求在全网算力资源池内进行最优路由调度。这种算力资源状态收集机制在具体的部署实践中，往往通过算力节点主动上报或注册的形式实现，算力网络再通过控制器或IGP&BGP路由协议进行分发。

1.2 算力网络数据面

如上所述，算力网络需要对应用算力服务请求(或算力原子服务)直接感知，并在数据面基于这种算力服务的算力需求信息，对应用流进行相应的路由决策，这就需要在数据面中封装算力服务及其算力需求信息，算力网络节点解析数据报文中对应的算力应用及其算力需求信息，并据此结合可用的网络路径进行应用请求流的路由。

1.3 算力网络管理面

1.1节所述算力节点的算力资源信息通过统一的算力标度，形成算力资源并配置相关的算力YANG模型，算力节点据此对算力网络转发和路由节点进行通告和配置。

此外，算力网络的性能和故障检测可在既有的网络OAM机制上扩展实现，并通过北向接口实现与算力网络管理面的通信和交互。

特别的，对算力节点算力资源信息收集和分发的不同机制，在收敛速度和实现复杂度上各有利弊，并对应不同的应用场景和需求。如通过NFVO或控制器集中收集和分发算力资源信息，具有全局资源视图和全局编排的优势，但是请求响应速度上需要做出牺牲。相反，通过分布式网络协议IGP(Internal Gateway Protocol)&BGP(Border Gateway Protocol)的方式洪泛算力资源信息，其收敛速度更快，尤其是在边缘网络中，更加有利于满足时延敏感业务的需求。

2 算力网络集中式架构分析

从算力网络架构和算力网络路由及传输的角度讲，算力资源的分布分成两种场景，或者说两种分布模式，一种是当前以及未来可预期的5G时期的终端、边缘计算、云数据中心三级算力分布模式，这也是当前算力网络架构要着重考虑的；另一种是算力资源也部署在除端、边、云之外的网络转发和路由节点，这是一种极端的理想模式，在这种模式下，算力资源将真正“无处不在”。由于网络转发和路由节点参与计算(及存储)，传统网络设备以及与此相关的通讯协议、流程都将发生根本性的演变，以适应网络从“传统转发+路由”到“转发+路由+计算”的模式转变，包括国际知名标准组织IRTF在内的众多研究机构都在对“Computing in Network (COIN)”[1]这一议题进行较有成效的研究。本文的架构分析仅限于端、边、云的三级算力分布模式。

集中式算力网络架构下，端、边、云的算力和网络资源及节点信息由集中式编排器(Network Function Virtualization，NFVO)或MEAO(Multi-access Edge Application Orchestrator)统一收集和分发，集中式编排器按照应用需求，结合全网算力和网络资源状态，编排最优的转发和路由路径，并下发至算力网络路由和转发节点，如图1所示。

算力资源分布部署于网络基础设施上，如边缘计算节点、云数据中心等，算力资源节点通过北向接口与集中编排器(或控制器)进行南北向垂直交互。应用从边

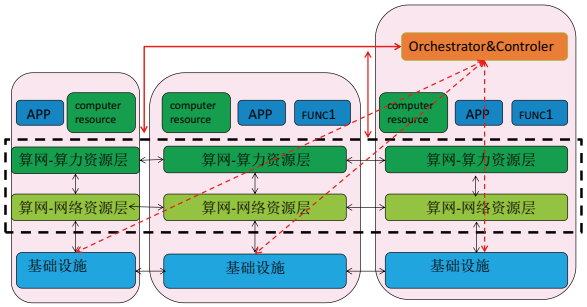

图1 算力网络集中式控制面架构示意图

缘节点接入，集中编排器(或控制器)需要在感知应用及其算力和网络需求的基础上，进行路由策略编排，对于少数典型应用，可通过集中编排器预编排预配置，并预下发至算力网络节点，入口节点将应用与预配置的路径做映射，进行相应的应用流量路由转发。对于非典型应用，算力网络入口节点(或算力网关)需要通过信令接口通告集中编排器，由后者进行相应的策略编排和下发。

2.1 全局资源视图

集中编排器(或控制器)与算力网络基础设施进行南北向垂直交互，与其他控制器进行东西向水平交互，加上层次化编排器的垂直交互，集中编排器对全网算力和网络资源状态拥有全局视图，在响应应用的算力和网络请求的时候，可以执行全网路由优化。

新增节点算力资源状态信息的上报和算力与网络联合路由策略的下发功能，仅涉及到北向接口特征参数的扩展，无需扩展协议。路由和路径编排策略由编排器执行并下发，与转发设备完全解耦。因此，从实际部署的角度讲，集中式算力网络控制面架构对当前在网网络设备和协议的影响最小，部署的代价最小，周期也最短。

2.2 应用及其算力和网路需求映射

在集中编排器(或控制器)的全局算力和网络视图下，应用对应的需求得以精准匹配和映射，并编排下发现网可执行的转发和路由策略。如上所述，非典型应用及其需求信息需要算力网络入口节点通过信令接口向集中编排器通告，这将不可避免地带来应用请求响应的延迟。因此，这种架构方案更加适合对时延没有严苛要求的算力应用。

2.3 层次化编排器解决方案

2.3.1 区县边缘和接入边缘

由于节点数目众多，传统集中编排器(或控制器)无法下沉到边缘，针对区县边缘和接入边缘，可通过轻量级虚拟化VIM(Virtual Infrastructure Manager)部署边缘编排器，进行算力和网络资源的收集和路由策略的下发。区县和接入边缘云数据中心通常以地市边缘云为中心进行二级管理。因此，边缘节点的算力和网络资源信息首先上报(或注册)到区县轻量级边缘云数据中心，集中编排器的全网路由策略也由该边缘云下发至边缘网络节点，避免边缘节点直接与大区核心云数据中心的集中编排器进行交互，对交互负荷进行了有效的层次化分担。对于可在本地执行路由决策的应用场景，则边缘云数据中心的虚拟化集中编排器就地执行路由策略编排和下发，客观上也提高了部分应用请求响应的效率和速度。

多接入边缘计算(Multi-access Edge Computing，MEC)第三方业务由部署于地市边缘云的多接入边缘应用编排器(Multi-access Edge Application Orchestrator，MEAO)进行编排，负责地市内第三方业务的发放、变更和撤销。

2.3.2 大区核心云数据中心

大区核心云数据中心的集中编排器可采用基于VIM的NFVO，NFVO仅与区县边缘和接入边缘云数据中心的轻量化集中编排器进行交互，即接收边缘云的算力和网络资源信息上报，进行全局算力和网络路由编排并下发。全网的CT(通信业务)业务统一由大区核心云数据中心的NFVO编排并下发。

算力网络集中式控制面架构如图2所示。这是基于运营商及第三方MEC现网拓扑的一个算力网络部署方案示意图，旨在阐明层次化编排器架构的运行模式，以及与现网的对应关系。可以明显看到，除了虚拟化部署的层次化NFVO、MEAO等之外，集中式算力网络架构对现网设备和协议的影响非常小。

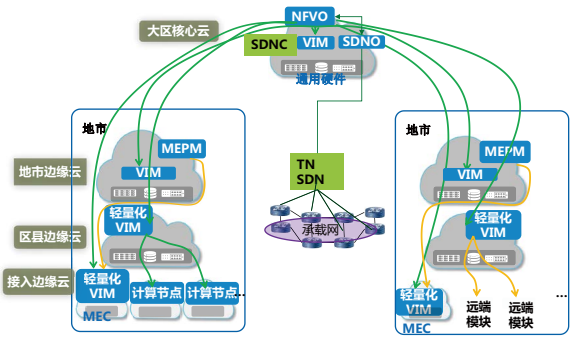

图2 现网层次化算力网络编排器架构示意图

3 算力网络分布式架构分析

算力资源节点就近向算力网络节点注册(含更新、删除)其算力资源状态信息，即算力网络转发和路由边缘节点对算力节点资源信息进行本地化管理，本地算力资源信息的全网通告则通过IGP&BGP实现。IGP&BGP对算力资源信息的洪泛通告，将在网络域内(IGP)或域间(BGP)构建全网算力资源状态数据库，以供算力网络转发和路由设备据此进行算力资源维度的路由和转发决策。

3.1 原子算力服务分解

分布式算力网络控制面架构下，应用(或服务)与网络和算力资源的匹配需要路由设备完成并在算力网络入口转发，当前及可预见的未来一段时间内，转发和路由设备的全局资源规划和应用智能匹配资源的能力是受限的，即转发设备只能按照既定的规则在可以穷举的数据、参数、流程范围内进行资源的调度，而无法对没有边界的开放参数类型进行处理。算力网络体系下的算力应用和服务就属于这种没有特定范式特定规则的业务类型，直接将其在转发路由层进行资源匹配，代价高昂，甚至在多数情况下不可行。因此，将用户的无边界算力应用分解为网络层设备可识别可处理且可穷举的基础应用(又名原子应用)，比如矩阵运算、傅里叶变换、编解码算法等，是算力网络分布式架构下需要重点解决的架构性问题。在算力网络的用户网络接口(UNI)处部署算力应用分解网关或者模块，将用户应用分解成原子算力服务，并给出各原子服务之间的执行时序，算力网络入口节点据此对原子服务进行算力资源映射，结合网络资源为用户应用流进行路径和路由编排并转发。

原子算力服务的算力资源需求可以通过少数参数进行量化确定，因此算力网络入口节点仅需在此量化算力需求的基础上，进行全网算力资源调配和路径编排，比如将应用流量路由至当前最能满足该原子算力服务算力需求的最近的算力节点。这是一种应用算力需求与算力资源的直接匹配和路由机制，算力路由设备需要维护较大的节点算力资源信息库，并需要承担原子服务算力需求计算与算力路由匹配的任务，负荷较重。另一种更加灵活高效的机制，是各算力节点为原子算力服务匹配对应的算力资源，并以原子算力服务实例的形式注册和洪泛，在这种机制下，算力网络入口节点只需要执行原子算力服务和原子算力服务实例之间的简单映射和路由策略转发[2]。

在已知的网络转发面机制中，SFC、SRv6、ICN等均可以通过一定的扩展，较好地支持上述算力原子服务转发。当然，这些转发面机制下，算力原子服务实例、算力节点资源状态信息等的上报、注册和路由策略下发，既可以通过分布式机制实现，也可以通过集中式和混合式(第4章将详述)实现。

应用请求分解为原子算力服务，并由网络路由设备进行算力资源映射和路由策略编排，即在应用和网络层之间增加一个新的适配层模块，该模块执行对应用请求的解构，这不仅需要该模块具有智能识别应用请求算力维度的子功能解构的能力，还需要兼顾适配现网算力微服务库的部署情况及其颗粒度，后者决定分解原子算力服务的深度。无论是软件还是硬件模块，同时支持上述两种能力，挑战性无疑是十分巨大的。

3.2 算力资源状态同步

在实际网络部署中，算力资源节点向最近的算力网络节点注册(含更新、删除)其算力资源信息，包括节点算力资源状态(CPU占用率、队列和缓存资源等)、节点

标识、算力原子服务实例标识等。算力网络转发节点通过分布式路由协议IGP&BGP将本节点管理的节点算力资源状态信息洪泛通告至邻居节点，以使算力网络转发和路由设备创建全网算力资源状态信息数据库。

算力节点承担的算力服务的多样性，决定了其算力资源状态的高度动态性，各种算力服务的生命周期长短不一，短则毫秒级，长则分钟甚至小时级，因此很难对算力资源状态同步的触发点进行严格意义上的量化，高频率的更新和同步，虽然能带来较高精度较细颗粒度的算力状态信息同步，但也给算力网络转发和路由设备带来不菲的同步负荷，依据具体的应用场景，按照网络运营和运行的经验数据，确定算力资源状态同步的周期值，是比较可行的方案。

算力节点资源分布式同步的网络拓扑示意图如图3所示。从图3可以看到，支持同样算力原子服务的实例或节点可能分布在多个算力网络转发节点管理域内，转发和路由设备根据网络和算力资源状态进行联合路由决策，选择最优的算力节点进行应用流转发。

3.3 算力网络入口节点源路由编排

从3.2可知，算力网络入口节点通过分布式算力资源状态信息的同步创建了全网算力资源信息数据库，在接收到来自用户的算力原子服务算力需求后，为其进行全网最优算力资源映射，并据此转发应用流量。因此，在分布式算力网络架构下，入口节点执行路径和路由规划，即源端路由规划，并完成数据面的路由封装，后续转发和路由节点据此进行流量转发。因此，SR(Segment Routing，分段路由)技术可以支撑该场景下的算力网络数据面的封装和转发机制，基于MPLS和IPv6网络的算力网络可以分别构建在SR-MPLS和SRv6的数据面上。

3.4 应用场景简析

由于全网算力资源状态信息数据库在分布式协议机制下实现了准实时的状态同步和更新，无需与集中编排器或控制器等中央节点交互，收敛速度更快，对应用的算力服务响应速度也更快。

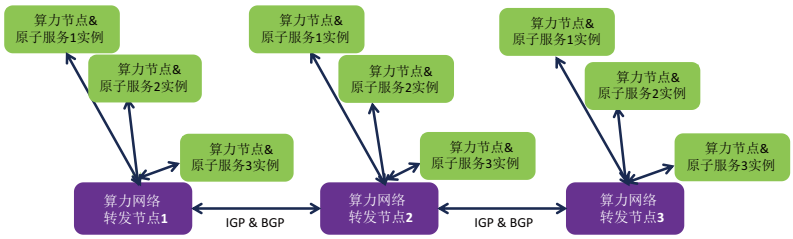

图3 算力资源状态分布式同步网络拓扑示意图

4 算力网络混合式架构分析

集中式算力网络架构拥有全局资源视图优势，并且对设备和现行协议的影响较小，但是由于大量计算节点和网络节点需要频繁与控制器或编排器进行交互，收敛速度慢，效率较低，无法适应时延敏感的算力应用。相反分布式算力网络架构能够较好地解决集中式架构的弊端，但是它涉及到现网设备和现行协议的大幅度调整，代价高昂，落地周期更长。

因此，一种既有集中式交互机制又有分布式交互机制的混合式架构，在很多应用场景能比较好地平衡部署代价、收敛速度等方面的需求。

4.1 局部及本地算力资源状态分布式架构方案

边缘算力节点就近注册(含更新、删除)其算力信息到对应的算力网络转发和路由节点，由后者创建本地算力资源信息数据库，即把算力资源节点和算力网络节点从逻辑上分开，前者的资源信息由本地算力网络节点代理管理和维护，从而避免大量边缘节点或MEC与集中编排器或控制器交互，将算力资源状态的同步本地化，提升收敛速度，尤其是对于只需要本地化算力资源调度的应用和业务，响应速度更快。

对于中小规模的边缘算力网

络，如网络节点在20个左右，算力网络转发和路由节点之间的算力资源通告和同步通过分布式协议IGP&BGP实现，从而发挥小规模网络分布式同步快速收敛的优势。

4.2 关键算力网络节点集中式架构方案

对于更大范围的全局算力网络资源同步，则由局部网络中的代理节点集中向集中编排器或控制器交互，发挥集中式架构大规模全网资源视图的优势。一些重要的算力节点比如运营商边缘云数据中心、第三方云数据中心等，则可以选择两种资源同步方式共存，已适应更加灵活多样的应用算力需求，它们既可以通过局部分布式协议进行资源同步，也可以通过与集中编排器或控制器交互，实现资源同步和策略下发，如图4所示。

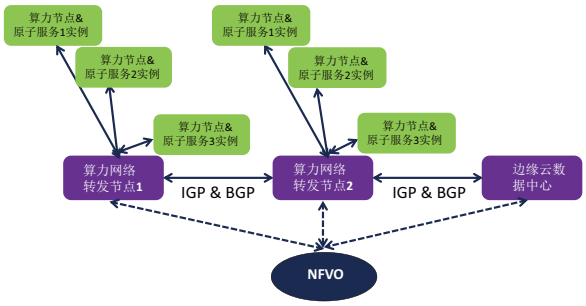

图4 算力网络混合式架构示意图

在混合式架构下，可以通过分布式本地化和集中式全局资源视图的融合优势，较好地实现时延敏感算力业务和全网资源优化业务的需求平衡。

4.3 算力应用资源映射方案

混合式算力网络架构下，应用流与算力及网络资源的映射将发生在两个层面。对于只需本地(含局部小规模算力网络)算力资源调度的应用，由网络入口节点进行本地化资源映射和路由策略编排并转发，机制类似分布式架构的算力原子服务流的资源映射和路由策略。对于需要进行全局资源调度的应用，则由集中编排器或控制器根据全网资源视图进行资源映射，并编排和下发路由转发策略到网络节点。

5 结束语

跟传统网络存储转发机制处理的颗粒度为报文或流(本质仍然是一种报文聚合)不同的是，算力网络中算力资源处理的颗粒度不再是报文，而是应用或服务，这就必然导致其资源状态的同步和映射机制跟传统网络大为不同，架构方案的设计和考虑将直接影响到不同的算力应用场景的需求满足情况。

算力网络架构设计过程中，在首先考量满足不同算力应用需求的同时，更要慎重考量对现网设备和协议调整和扩展的复杂度，以及由此而来的网络升级成本和收益平衡。如本文第2部分所述，集中式算力网络架构由于主要涉及北向接口的扩展和集中编排器或控制器(NFV)的功能升级，代价最小，且拥有全局资源视图，但是收敛速度慢。第3部分所述的分布式架构则拥有收敛速度快的优势，但设备和协议升级代价较高。综合两种架构的优劣势，第4部分阐述了混合式架构的本地化分布式架构和关键节点集中式架构的融合优势，是实际网络部署中将会重点考虑的架构方案。

算力网络的核心是为应用在全网进行最优资源匹配并路由，因此随着基础网络路由机制的丰富和发展创新，其架构也将随之优化和更新，比如ICN等全新网络路由机制下，算力网络的资源寻址机制也将随之发生根本性的变化。

参考文献

[1] I.Kunze,K.Wehrle.Transport Protocol issues of In-Network Computing system[EB/OL].(2020-03-09)[2020-07-21].https://datatracker.ietf.org/doc/draft-kunze-coinrg-transport-issues/?include_text=1

[2] S.Gu,G.Zhuang,H.Yao, et al.A Report on Compute First Networking (CFN) Field Trial [EB/OL].(2019-12-02)[2020-07-21].https://tools.ietf.org/html/draft-gu-rtgwg-cfn-field-trial-01

作者简介

黄光平

硕士研究生，主要研究方向为下一代分组路由架构和协议方案，发表论文5篇，获得国际国内通信类技术发明专利10余项。

罗　鉴

中兴通讯股份有限公司有线研究院总工程师、高级工程师，CCSA TC3副主席、CCF互联网专委会常务委员，主要研究方向为分组数据网络、网络架构演进等。

周建锋

本科，主要研究方向为核心网、NFV。

Analysis of Computation Network Architecture and According Scenarios

Huang Guangping
Luo Jian　　ZTE Corporation, Nanjing 210012, China
Zhou Jianfeng

Abstract To take advantage of pooling benefits of the distributed computing resources in the network, it's necessary to specifically design a network to connect these resources with the isolated resource barriers broken and service quality optimized upon the basis of the whole computation network. It's extremely important to strike a good balance between satisfying service computation requirements and cost of upgrading the devices and protocols. This paper focuses on the analysis upon three architecture solutions in terms of convergence speed, whole picture of network computation resource, mapping between service and computation resources and upgrading of devices and protocols, as well as the scenarios best suitable under each solution. Distributed architecture boasts of faster convergence speed while comes with higher cost for device and protocol upgrading, centralized architecture holds the benefit of whole view picture of the network computation resource and limited impacts upon the devices and protocols. The hybrid architecture combines the advantages of the two solutions and tries to minimize the disadvantages. In addition, under the new network routing architecture, the computational power network would obtain more excellent scheduling and forwarding performance.

Keywords Computation Network; Centralized Architecture; Distributed Architecture; Hybrid Architecture