

doi:10.13756/j.gtxyj.2023.05.006

光通信系统与网络技术

黄旭,成梦虹,成芝言. 基于 SRv6 的算网基础管控体系研究[J]. 光通信研究,2023(5):38—44.

Huang X, Cheng M H, Cheng Z Y. Research on Basic Management and Control System of Computing Force Network based on SRv6[J]. Study on Optical Communications, 2023(5):38—44.

基于 SRv6 的算网基础管控体系研究

黄 旭,成梦虹,成芝言

(中国移动通信集团设计院有限公司,北京 100080)

摘要:随着算力网络的全新理念提出,算力和网络融合后如何管理、协同和应用部署成为了运营商当下亟待解决的问题。为了更好地解决网络、计算和存储等多维度资源的统一管理、感知调度和协同编排等问题,文章从算网管控方向展开研究,提出算力网络管控系统总体架构设计、核心能力研究以及底层网络技术演进等内容。以“网随算动、以网强算”为原则,借助管控系统能力形成算力网络“横向拉通、纵向调度”的协同模式,最终,有效支撑算力网络业务发展需求,助力国家“东数西算”工程发展。

关键词:算力网络;算网管控;架构设计;功能研究

中图分类号: TN915

文献标志码: A

文章编号: 1005-8788(2023)05-0038-07

Research on Basic Management and Control System of Computing Force Network based on SRv6

HUANG Xu, CHENG Meng-hong, CHENG Zhi-yan

(China Mobile Group Design Institute Co., Ltd., Beijing 100080)

Abstract: With the new concept of computing force network being put forward, how to manage, cooperate and deploy after the integration of computing power and network has become an urgent problem for telecom operators. To better solve the problems of unified management, perceptual dispatch, and collaborative arrangement of multi-dimensional resources such as network, computing, and storage, we study from the direction of computing force network management and control. Then we investigate the overall architecture design of computing force network management and control system, the core capabilities, and the evolution of underlying network technology. Based on the principle of "the network changes with the computing force, and the network strengthens the computing force", the collaborative mode of "horizontal connection and vertical dispatch" of the computing force network is formed with the help of the management and control system capabilities. Eventually, we believe it will effectively support the development demands of computing force network business and help the development of the national "Eastern Data and Western Computing" project.

Key words: computing force network; management and control; architecture design; functional research

0 引 言

算力网络是面向计算服务与网络承载融合的新架构、新协议和新技术,是响应国家数字经济发展、加速科技创新和落实“东数西算”工程部署的重要支撑。同时,也为运营商面向个人、企业和垂直行业等用户服务带来新机遇。随着第五代移动网络(5th Generation Mobile Network Technology, 5G)、物联网和人工智能等新技术和新场景的不断涌现,海量的数据存储、敏捷计算、高速传输和实时调度等需求也将对运营商网络及应用服务提出新的挑战。2021年中国移动算力网络白皮书指出,将以算为中心、网络为根基,实现“算力泛在、算网共生、智能编排、一

体服务”的目标,达成“网络无所不达、算力无所不在、智能无所不及”的愿景。

基于算力网络的发展目标及愿景,网络将作为用户需求与算力的桥梁,提供深度的网络覆盖与广泛的边缘汇聚能力,为用户提供大带宽、低时延的高可靠算网接入服务。同时,网络承载将融合算力资源,通过上层应用实现端到端算网业务跨域、跨专业和跨厂家的统一管控、协同编排、智能感知和按需调度等能力。

运营商作为国内最大的基础网络服务提供商,目前处于云网融合阶段,可提供一站式入云、云服务互联等业务,但还无法满足算力网络“云、边、端”资源协同发展、统一编排、感知调度和泛在应用等综合

收稿日期:2022-09-07; 修回日期:2022-09-21; 纸质出版日期:2023-10-10

作者简介:黄旭(1990—),男,辽宁凌源人。学士,主要研究方向为 IP 网络新技术研究。

通信作者:成梦虹,教授级高级工程师。E-mail: chengmenghong@cmdi.chinamobile.com

© Editorial Office of Study on Optical Communications. This is an open access article under the CC BY-NC-ND license.

服务需求,尚不具备确定性网络管控能力。因此,算力网络通信技术演进、管控体系搭建等内容在现阶段具有较强的研究价值及重要意义。

1 网络技术演进

1.1 网络技术选型

传统网络多基于内部网关协议 (Interior Gateway Protocol, IGP) 或边界网关协议 (Border Gateway Protocol, BGP) 等动态路由完成业务转发,数据包的 control 与转发都要依赖底层设备能力。随着业务的多样化,需求的个性化,分段路由 (Segment Routing, SR) 被广泛应用。虽然该技术具备流量工程 (Traffic Engineering, TE) 能力,但基于标签分发协议 (Label Distribution Protocol, LDP) 完成转发,能力仍受限于 IGP 技术。因此,算力网络需引入新的路由技术,将算力能力与底层网络相结合,让网络架构、业务逻辑更简单,转发行为融入更多的人为意愿。

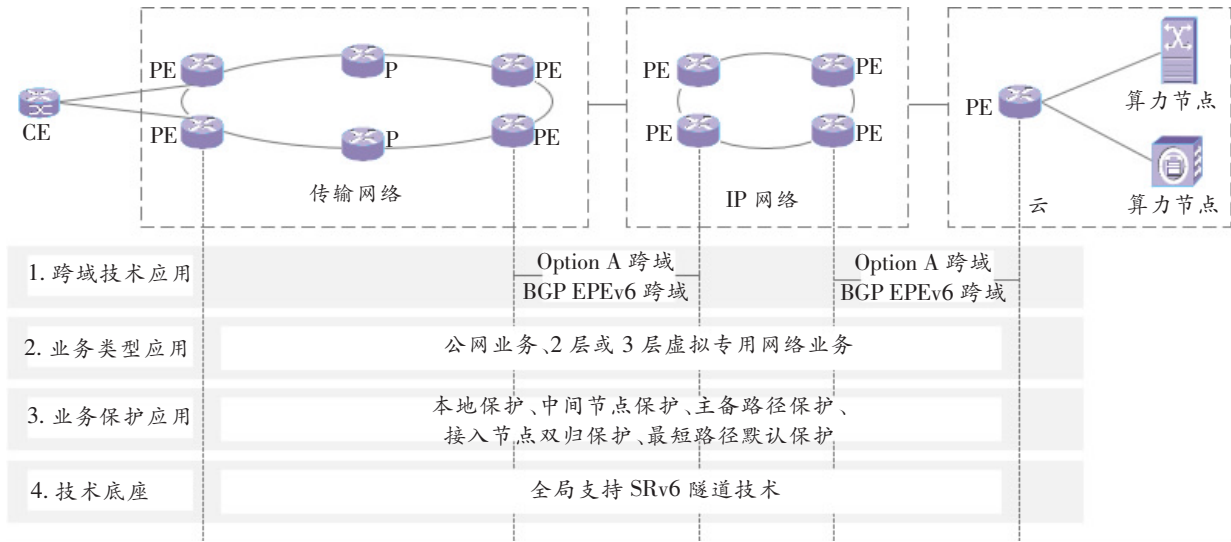
算力网络需打造以 SRv6 为技术底座的新型网络服务体系,解决传统网络中多层协议栈维护压力、跨域拉通复杂和端到端无法按需选路等难题。基于互联网协议 (Internet Protocol, IP) v6 的 SRv6 技术是以 IPv6 为转发平面的 SR 技术演进^[1],具备网络可编程、转发行为可编程和业务可编程能力。纯 IPv6 转发使网络更简单,中间节点支持 IPv6 技术

栈即可完成转发。同时,SRv6 的引入也解决了虚拟化网络环境中服务节点不支持传统多协议标签转换 (Multi-Protocol Label Switching, MPLS) 问题,使端到端互通性更好。

1.2 网络架构及技术应用

理想环境下,算力网络端到端节点均要具备 SRv6 能力,但运营商级别的基础网络环境较复杂,设备厂家和型号多且数量也大,全局部署难度大。因此,在算力网络规划和建设初期,网络架构及相关技术应用应至少具备两套设计方案。图 1 所示为全局 SRv6 技术应用,是算力网络演进的目标架构。

传输、IP 以及云网络域内设备均支持 SRv6 技术。不同自治系统 (Autonomous System, AS)、业务域跨域互通时可应用背靠背 (Option A) 或基于 SRv6 的 BGP 出口对等工程 (BGP Egress Peer Engineering, BGP EPE) 等技术方案。通过应用发现 Option A 背靠背虚拟路由转发 (Virtual Routing Forwarding, VRF) 方案可以让网络配置更简单,且各专业运维人员只需维护域内的隧道标签栈,通过虚拟局域网 (Virtual Local Area Network, VLAN) 标签划分业务便可实现管理。但面向端到端业务全局拉通时,需应用 BGP EPEv6 跨域组网方案,此方案采用多业务域网络统一管控方式,可支撑用户端到端访问路径全局编排与下发,同时满足用户业务逐跳质量保障需求。



注:CE 为用户设备;PE 为网络边缘设备;P 为网络中间设备。

图 1 全局 SRv6 技术应用

Figure 1 The application of global SRv6 technology

算力网络用户接入应支持公网 IP 业务以及私网 IP 的虚拟专用网络 (Virtual Private Network, VPN) 业务,包括 L3 VPN、L2 VPN 和以太虚拟专

用网络 (Ethernet VPN, EVPN) 等接入方式,从而满足个人、企业、垂直行业等用户多方式入网,实现泛在接入目标。

业务保障需叠加节点故障保护、主备路径/隧道保护等技术,确保高质量算网业务稳定运行。同时,依托 SRv6 特性,加强业务保护环节设计。当 SRv6 TE 隧道失效时,设备将通过 SRv6 默认转发(Best Effort,BE)模式完成最终保护。从而提升算力网络底层通信架构可靠性,实现高价值用户业务多重保护能力。

在 SRv6 技术架构上应用 SRv6 策略(Policy), 通过 color 属性标记不同特征隧道, 实现高保障服务级别协议(Service Level Agreement, SLA)路径规划。例如, color A 表示时延属性、color B 表示丢包属性、color C 表示带宽属性等。

图 2 所示为边缘 SRv6 技术应用,适用于算力网络发展初期底层网络设计,当网络无法全局部署 SRv6 时,采用边缘节点 SRv6 技术应用的设计方案。在不同网络域、业务域的边缘节点应用 SRv6 技术,用于算网的用户接入及跨域协同。因此,承载的业务、跨域方案等应用与全局 SRv6 技术应用相同。不同之处在于,不支持 SRv6 技术的中间节点应用传统的 SR TE 或分段路由传输配置文件(Segment Routing Transport Profile, SR-TP)技术,可支持传统标签隧道建立与数据转发。最终,通过 SRv6 隧道与传统隧道标签栈粘合完成端到端路径全局编排。

以上两种组网方案可实现网络业务多接入、高可靠和全局路径可选可编排等能力。通过底层网络架构设计及相关技术应用,有效支撑算力网络建设初期及未来全局演进需求。算力网络底层架构发展

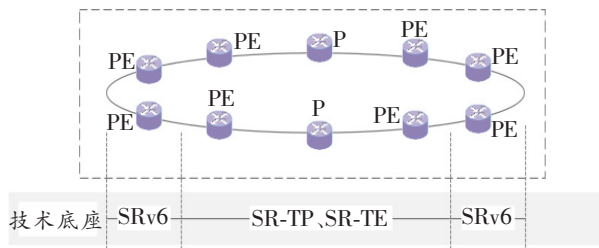


图 2 边缘 SRv6 技术应用

Figure 2 Application of edge SRv6 technology

目标是全节点 SRv6 部署,现阶段由于业务需求、网络升级规模和设备厂商能力等诸多问题,算网初期可应用边缘 SRv6 技术完成业务接入及多业务域协同等目标。

2 算网管控体系研究

管理、感知、编排和调用是算力网络面向用户业务应具备的核心能力^[2],如何管控、协同底层网络有效支撑上层应用,是运营商构建算力网络过程中要解决的问题。因此,现阶段对算网管控体系展开一定研究,包括算网管控系统架构、核心能力以及协同调用方式^[3]。

2.1 算网管控系统架构设计

运营商多以层级划分网络架构,包括接入层、汇聚层和核心层,同时,国家“东数西算”工程建设 8 大区为算力枢纽节点^[4]。因此,在算网架构设计上需考虑网络层级架构与区域空间结构融合,从而满足未来算网服务演进。图 3 所示为算网管控整体架构,采用分布式架构设计,具备一体化管控、分布式协同能力^[5]。

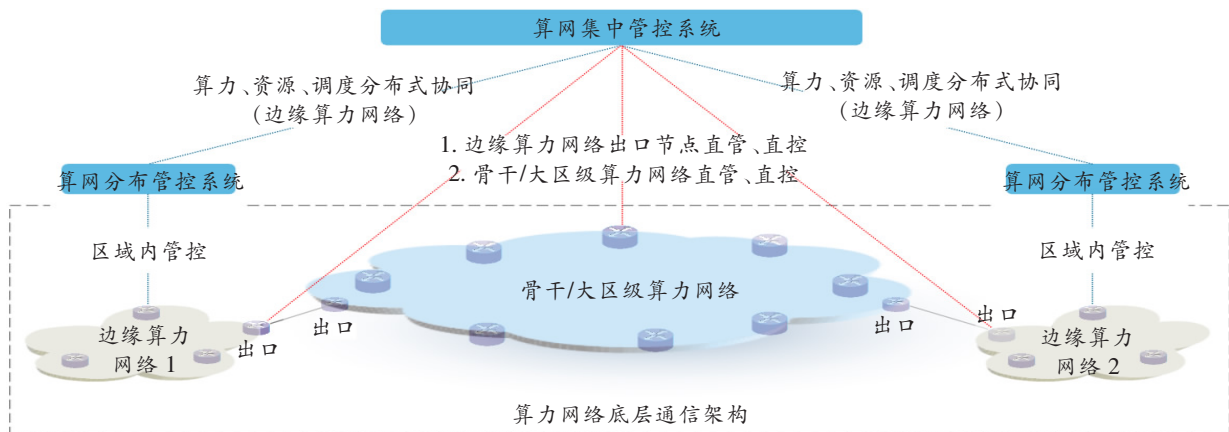


图 3 算网管控整体架构

Figure 3 Overall architecture of computing force network management and control

算网集中管控系统面向骨干/大区级算力网络全局设备、边缘算力网络出口设备,具备集中管理、算力全局感知和业务统一调度等能力;算网分布管

控系统面向边缘算力网络或接入层网络,具备设备、算力和业务等本地管控以及跨区域调度需求和算力资源上报等能力。管控系统采用集中与分布协同模

式,可实现算力网络泛在接入、端到端路径灵活编排和算力全局调优等目标。

图 4 所示为算网管控系统接口,基于软件定义网络(Software Defined Network, SDN)技术架构,实现底层通信网络控制面上移^[6-7],具备算力网络基础设施直采和直控能力。结合 SRv6 技术应用,使算力网络具备意图编程、智能管控等能力。北向采用表述性状态传递(Representational State Transfer, REST)接口对接外部系统实现能力共享及实时消息调用;南向接口采用多协议对接,其中,网络配置协议(Network Configuration Protocol, NETCONF)完成设备管控,其余协议主要实现底层算网资源采集与系统间协同调用。

2.2 算网管控系统能力研究

算力网络的意图是将网络资源与计算资源、存储资源融合,实现全局资源协同、路径编排、调度统一和质量感知等核心能力^[8]。图 5 所示为算网管控

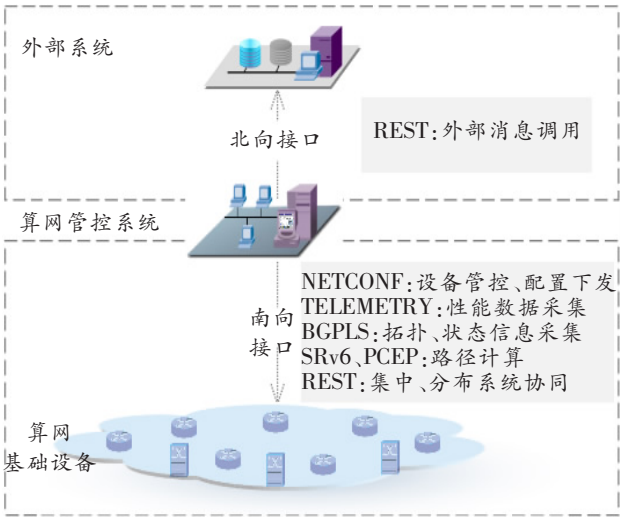
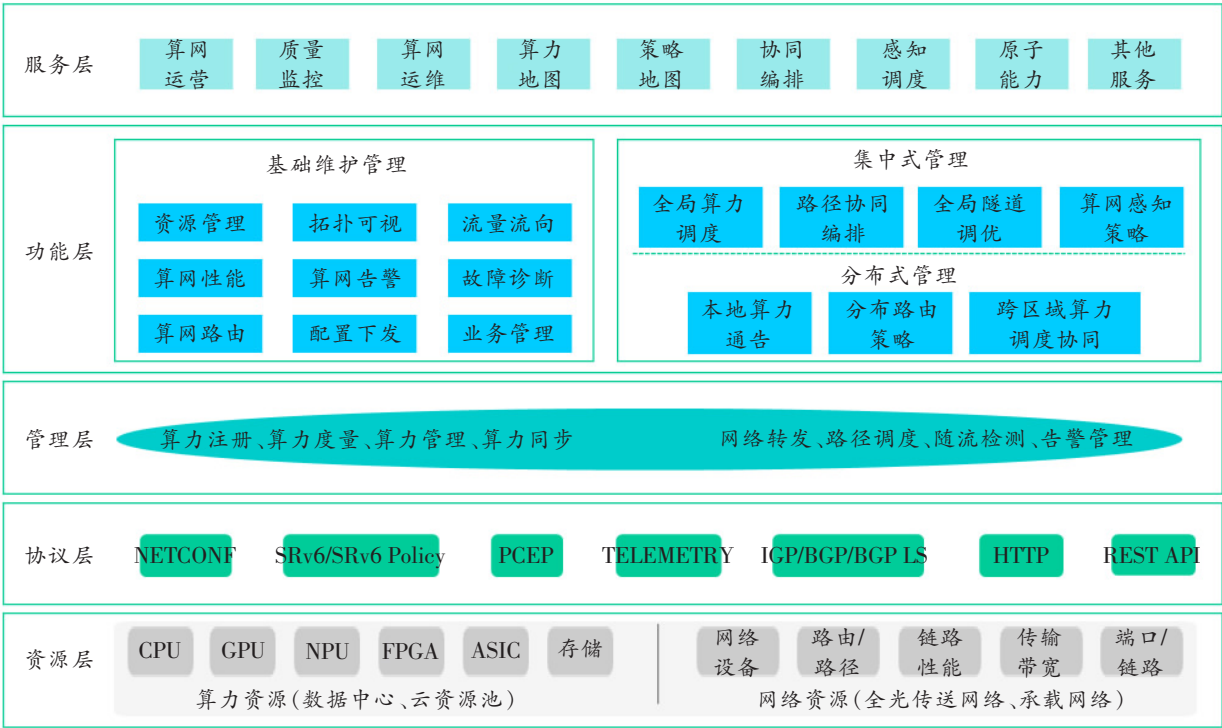


图 4 算网管控系统接口

Figure 4 Interface of computing force network management control system

系统功能架构,通过资源、协议、管理、功能和服务逐层应用,完成管控系统整体功能设计。



注:CPU 为中央处理器;GPU 为图形处理器;NPU 为网络处理器;FPGA 为可编程逻辑门阵列;
ASIC 为专用集成电路;PCEP 为路径计算单元通信协议;TELEMTRY 为远程高速采集协议;
HTTP 为超文本传输协议;REST API 为状态传输接口协议。

图 5 算网管控系统功能架构

Figure 5 Functional architecture of computing force network management and control system

2.2.1 资源层

资源层是算力网络基础设施面向上层管控系统需上报以及被采集的算力网络基础资源,也是算力网络最底层、最小粒度的原始数据,主要由算力基础

设施的算力资源和网络基础设施的网络资源组成。其中,算力资源包括软/硬件基础设施、通用与高性能计算、存储容量与性能、算力安全与分析等数据;网络资源包括网络设备、物理及逻辑端口、链路、路

由转发及业务隧道路径、最小集链路性能及带宽等数据。通过资源数据采集,构筑算力侧与网络侧基础信息,支撑算力网络管控系统上层应用。

2.2.2 协议层

协议层主要服务图 4 算网管控系统中南向接口,可实现算力网络南向采集与管控功能,诸多的协议应用也有效支撑了南向增、删、改、查等行为的灵活性与时序编排能力。同时,新技术的应用使南向交互的时效性更高。

图 6 所示为网络设备配置交互策略图,采用 NETCONF 协议定义南向接口,完成与网络设备的管控交互。该策略需具备多厂家、多类型和多版本的设备对接能力,通过统一建模完成多类 Yang 策略匹配及解析,屏蔽私有性和单一性问题。

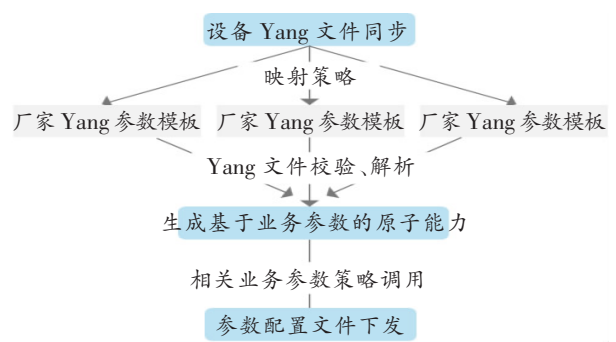


图 6 网络设备配置交互策略

Figure 6 Interactive policy of network device configuration

2.2.3 管理层

主要负责算网资源的全局管理,是算力资源与网络资源的第一层融合。例如,某区域算力设施的计算、存储和分析资源在管理层完成统一注册与度量^[9],并关联对应网络实体的端口、链路、带宽和性能等参数,通过内部接口最终实现底层算网资源统一建模并发布。该过程通过意图驱动或系统内策略编排的形式支撑上层功能应用,包括业务开通、路径编排和按需调度等场景。

2.2.4 功能层

算网管控系统的核心能力主要通过功能层呈现,由于系统整体架构采用分布式设计,因此,功能层应用需要分层级规划、部署。

基础维护管理是集中或分布管控系统功能层的通用部分,具备算网资源、性能、告警、业务和配置等功能管理,多面向算网管理、控制、分析和可视化等基础能力。集中式管理系统在云、边、端、骨干/大区等多网络环境中应用,因此功能层增加全局算力调度、路径协同编排、全局路径/隧道调优以及全局

算网感知策略功能,具备多网络、多区域的算网全局协同编排、统一调度能力;分布管控系统应用在单网络、单区域环境中,核心能力主要体现在跨区域算力调度协同,以消息形式上报跨区域调度请求至集中管控系统。其余功能只需具备本地管控、调度和算力实时通告功能即可,属于轻量化部署、灵活分布,主要实现管控能力下沉。

(1) 集中式核心功能

全局算力调度:是满足算网业务接入与调度的核心功能,主要体现多网络结构中算力资源全局管理、实时匹配,底层网络跨域互通、灵活调度。同时,实现了算力与网络的协同一体化能力。

路径协同编排:基于 SRv6 技术底座,按需完成业务路径全局编排,包括时延、带宽、丢包和链路利用率等条件下多个域内、域间的分段路由头部(Segment Routing Header,SRH)全局编排与统一封装。

算网感知策略:包括网络状态感知与算力状态感知,当业务关联的算网状态、指标发生变化时,通过策略驱动的方式触发感知预警与感知调度。

全局隧道调优:按不同业务需求建立高质量 SLA 隧道,基于感知策略完成业务隧道自动调优,实现业务端到端转发不受故障影响,是算网可靠性的保障。

(2) 分布式核心功能

本地算力通告:本地算力资源、算力状态实时上报,支撑集中管控系统实现全局算力资源管理。

分布路由策略:是云、边、端等本地接入层的路由发布与调度功能,完成本地算力网络感知调度或路径调优。

跨区域算力调度协同:当本地算网不具备某业务接入或调度能力时,对上层管控系统发布算力协同调度消息,从而完成本地算力业务跨区域调度。

2.2.5 服务层

结合多层次、多维度算网功能,为算网管理人员或外部系统提供诸多运维及开放性原子能力。支持面向个人、企业和垂直行业等用户的门户系统对接,通过接口直接调用的方式提供北向运营服务,结合时序编排能力,实现算力泛在、一体服务的目标。

2.3 算网管控系统协同

算网管控体系架构的核心是面向全局算网集中管控、本地算网分布协同,图 7 所示为算网管控系统协同架构,目的在于多网络、多区域算网集中管控不冗余,保证在全局算力需求实时感知的同时,依靠算网资源完成全局路径编排,支撑分布管控系统实现

边缘算力业务跨区域调度。单网络、单区域的分布管控则满足算力泛在接入理念,以管控下沉的方式完成边缘侧、端侧算网本地管控、本地调度。同时,分布系统与集中系统实时交互,具备算力资源、算力

状态通告,跨区域算力调度请求等能力。最终,通过统一与协同的双重服务模式,驱动算力网络实现全局编排、全局调度等目标。

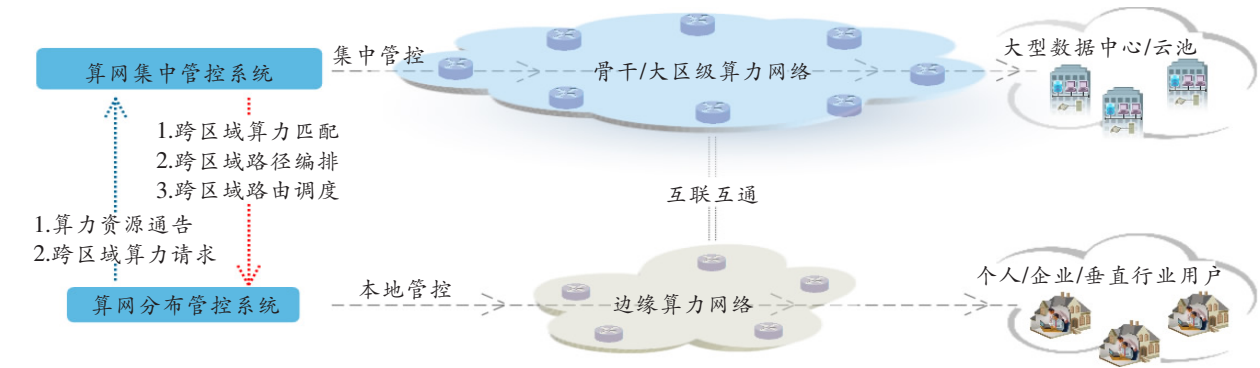


图7 算网管控系统协同架构

Figure 7 Coordination architecture of computing force network management and control system

2.3.1 分布管控系统协同策略

- (1) 资源同步:分布管控系统完成本区域边缘算力网络资源管理,实时上报算力资源和算力状态等消息至集中管控系统。
- (2) 调度请求:当边缘算网因故障和算力资源不足等原因导致本地算网业务无法正常开通、合理调度时,需上报跨区域算力请求消息至集中管控系统。

2.3.2 集中管控系统协同策略

- (1) 跨区域算力匹配:集中管控系统动态管理全局算力资源,根据分布管控系统上报的算力请求

完成算力调度目标选择。

- (2) 跨区域路径编排:以请求调度消息中的业务需求驱动路径编排策略,执行全局路径编排。编排策略应包括就近区域、路径时延最优、路径丢包最少、跨节点最少调度策略以及组合策略等。

- (3) 跨区域路由调度:调度前进行业务仿真,验证路径连通性、性能满足情况等,依靠图4算网管控系统中南向接口以及算网底层SRv6技术特性,完成源/目的节点相关配置下发。图8所示为跨区域算力调度策略示意图。

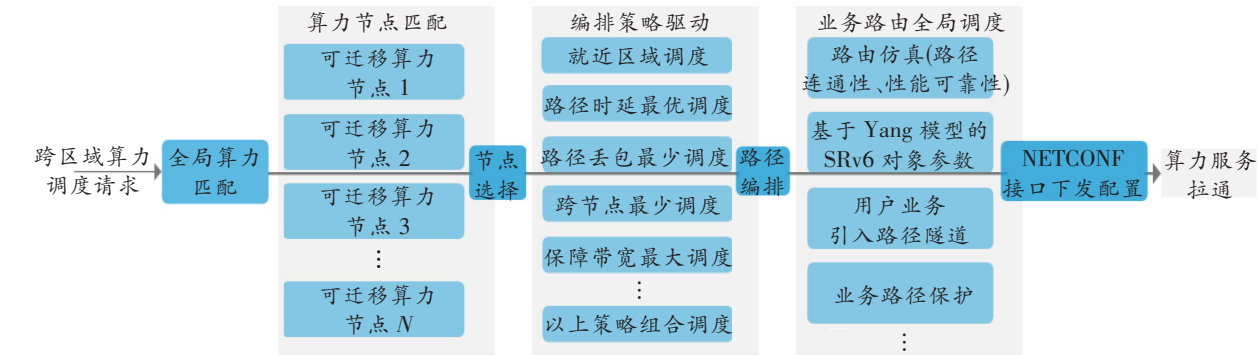


图8 跨区域算力调度策略

Figure 8 Cross regional computing power scheduling policy

基于分布与集中管控系统协同能力以及图8跨区域算力调度策略,完成不同业务域间算力调度需求。通过算力调度请求触发全局算力匹配、算力节点选择、按需路径编排以及业务隧道配置下发等相关策略,最终实现算力服务全局协同能力。

3 结束语

算力网络不仅是算力与网络的技术融合,更是面向个人、垂直行业,乃至社会的一种无需关心底层网络结构、路径性能、接入方式的算力“随用即取”的服务能力。而算网管控体系的搭建将构造算力网络

“管控大脑”，有效支撑算网应用及服务。因此，在算网建设初期，本文从底层网络演进和上层系统功能等内容展开研究，分析了算网应具备的能力及如何更好地对算网进行管控、协同。随着一系列新技术的融合与不断创新，算力网络架构、能力以及管控体系将更加完善，从而加快国家实现“东数西算”、“数字强国”等目标。

参考文献：

[1] IPv6+技术创新工作组. SRv6 技术与产业白皮书[R]. 北京:推进 IPv6 规模部署专家委员会,2019.
IPv6+ Technology Innovation Working Group. SRv6 Technology and Industry White Paper[R]. Beijing: Expert Committee on Promoting IPv6 Scale Deployment, 2019.

[2] 中国移动通信集团有限公司. 算力网络白皮书[R/OL]. (2021-11-01)[2022-08-10]. <http://www.econsortium.org/Uploads/file/20211108/163652251472904.pdf>. China Mobile Communications Group Co., Ltd. Computational Power Network White Paper [R/OL]. (2021-11-01) [2022-08-10]. <http://www.econsortium.org/Uploads/file/20211108/163652251472904.pdf>.

[3] 曹畅. 算力网络——云网融合 2.0 时代的网络架构与关键技术[M]. 北京:电子工业出版社,2021.
Cao C. Network Architecture and Key Technologies in the Era of Computing Force Network Cloud Network Integration 2.0 [M]. Beijing: Electronic Industry Press, 2021.

[4] 王晓云,段晓东,张昊. 算力时代[M]. 北京:中信出版社,2022.
Wang X Y, Duan X D, Zhang H. Computing Age

[M]. Beijing: CITIC Press, 2022.

[5] 王学光,韩晓非,杨明福. 协同分布式应用系统的架构研究[J]. 计算机应用与软件,2004,21(5):74—76.
Wang X G, Han X F, Yang M F. The Research of Framework for Collaborative Distributed Application System [J]. Computer Applications and Software, 2004, 21 (5):74—76.

[6] GB/T 40696-2021,信息技术 系统间远程通信和信息交换 基于 SDN 的网络联合调度[S].
GB/T 40696-2021, Information Technology. Telecommunications and Information Exchange between Systems. Network joint Scheduling based on SDN[S].

[7] 张凤军,罗广军,邱帆,等. 基于 SDN 的分组传送网架构设计及实现[J]. 中国电子科学研究院学报,2020,15(7):665—671.
Zhang F J, Luo G J, Qiu F, et al. Design and Implementation of Packet Transport Network Architecture based on SDN[J]. Journal of China Academy of Electronics and Information Technology, 2020, 15 (7): 665—671.

[8] 付月霞,姚惠娟,陆璐,等. 面向全网算力的算力感知网络关键技术研究[R]. 北京:中国通信标准化协会,2022.
Fu YX, Yao H J, Lu L, et al. Research on Key Technologies of Computing Power Aware Network for the Whole Network[R]. Beijing:China National Communication Standardization Association. 2022.

[9] 韦传讲,庄毅. 面向移动云计算的自适应虚拟机调度算法[J]. 小型微型计算机系统,2021,42(1):96—104.
Wei C J, Zhuang Y. Adaptive Virtual Machine Scheduling Algorithm for Mobile Cloud Computing [J]. Small Microcomputer System, 2021, 42 (1):96—104.