一种面向服务的算网路由架构方案

An Architecture Solution of Service-Oriented Routing for Computing and Networking

黄光平/HUANG Guangping[1,2]，谭斌/TAN Bin[1,2]，吉晓威/JI Xiaowei[1,2]
（1. 中兴通讯股份有限公司，中国 深圳 518057；
2. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055）
(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518055, China)

DOI：10.12142/ZTETJ.202304008
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.tn.20230724.1536.006.html
网络出版日期：2023-07-25
收稿日期：2023-05-25

摘要：网络感知算力并据此执行算网融合路由，是算网融合在网络基础设施侧的一种重要技术方案。相对于传统基于主机地址的网络路由机制，算网路由最主要的增量是在网络侧引入分布式多算力实例的优选，因此位置和归属无关的服务标识将成为新的寻址和路由标的。阐述了一种端到端的算网路由解决方案及其对路由协议的影响，并基于典型场景和测试用例，分析了面向服务标识的算网路由架构方案在功能和性能维度的收益。

关键词：算网融合；算力路由；服务标识

Abstract: Routing based upon convergent computing and networking where the network is enabled to be aware of the computing, as far as the network infrastructure of convergence of computing and networking is concerned, is a cornerstone technique solution. Contrary to the conventional host address-based routing scheme, the above-mentioned routing brings a process of computing selection among multiple and distributed sites and nodes. Therefore, service identification which is both location and homing independent should be employed in terms of addressing and routing. The comprehensive solutions as well as their impacts on the existing routing protocols are discussed, and the gains in both function and performance of the new architecture solution are demonstrated with contexts of typical scenarios and testing analysis.

Keywords: convergence of computing and networking; routing of computing; service identification

引用格式：黄光平, 谭斌, 吉晓威. 一种面向服务的算网路由架构方案 [J]. 中兴通讯技术, 2023, 29(4): 38-42. DOI: 10.12142/ZTETJ.202304008
Citation: HUANG G P, TAN B, JI X W. An architecture solution of service-oriented routing for computing and networking [J]. ZTE technology journal, 2023, 29(4): 38-42. DOI:10.12142/ZTETJ.202304008

在网络路由和调度体系中引入算力是算网融合架构中的重要增量，网络因此扩展了对算力的感知能力。同时，路由和调度流程实现了算力和网络两个维度资源状态的融合考量，即行业通常所说的算力路由[1]。算力路由是算网融合的主要技术锚点[2-3]。算力路由从端到端协议和流程方面，打破了网络和算力这两个传统上相互隔离的技术和资源体系壁垒，实现了“网中有算，以网强算”。算力路由的本质是：面向同类算力服务的分布式等价多实例，基于算力和网络的资源状态以及业务需求，执行网络和算力联合优选寻址，即“一对多”的算网寻址路由。其中的“多”表示网络和算力均存在多路径、多实例的权衡优选。而面向用户的算力服务是位置无关的，甚至可能是归属无关的。用户对算力服务的请求仅表达意图，无须关心服务的提供方和部署位置。这是算力路由跟传统基于主机位置的IP路由最本质的区别，也是算力路由协议体系存在的主要变量之一[4]。

算力路由引入位置无关的服务标识，将其作为路由和寻址的全新对象，在使能全新的算力感知和路由功能的同时，也为现有网络路由寻址协议带来新的扩展需求和挑战。因此，新架构功能在引入的同时，需要保持与现网架构兼容。服务标识的引入在客观上打通了业务和网络之间的高效感知接口。网络通过服务标识可以精细化识别业务，并提供相应的细颗粒度网络连接服务。

面向服务的算网融合路由技术在典型的业务场景下，有独特的业务和资源应用价值。而当前的业务场景还存在一些亟待解决的问题，仍需要充分发挥面向服务的算网融合路由技术的优势。

1 算力路由在IP分组网络中面临的主要问题

算力路由是叠加在传统IP分组网络基础上的一种增强性路由。在主机IP地址路由的基础上，网络需要增强算力感知的能力，并在此基础上执行算网融合路由。这既包括对算力服务的路由寻址，也包括对算力服务节点的主机路由寻址。算力感知和算力路由引入了全新的算力因子，这给IP分组网络带来4个方面的问题。

1）IP主机地址路由体系下的算力服务路由寻址问题

基于IP分组网络的算力路由，本质上是面向服务的分布式多算力实例寻址路由，即基于算力资源状态和网络资源状态，在多实例多路径中根据服务等级协议（SLA）需求进行算力节点优选或引流。这种面向服务的分布式路由机制，跟面向IP主机地址的路由机制完全不同。后者指向全局唯一主机，且基于前缀的寻址机制是基于物理上的子网部署模式；而算力路由从语义上并不指向特定算力服务主机，而是指向特定算力服务，并且同一类算力服务可能部署在不同的物理子网内。因此，基于IP前缀的子网模式并不适用于算力服务的部署模式。同一类算力服务与多服务实例及多实例主机地址关联，算力服务仅仅充当一种抽象类型索引。网络需要在这个服务索引与它对应的算力、网络资源、服务实例主机地址之间构建动态的映射关系。

2）算力感知对IP路由协议造成的震荡问题和表项膨胀问题

网络对算力资源状态的感知，需要针对相应的接口和协议进行扩展，并且在网络路由和转发节点引入新的算力路由表项。然而，算力资源类型及其状态变更频率都非常多样化，全颗粒度算力资源状态向网络暴露，将不可避免地导致现有网络协议（如边界网关协议）收敛震荡，对现网运行造成破坏性冲击。除此之外，海量的算力资源状态必将导致网络路由和转发节点对应数量级的算力路由表项，对节点性能造成严重影响。

3）算力对网络暴露的参数类型及颗粒度问题

对IP分组网络控制面而言，算力参数可以分为算力原始状态数据和网络链路维度的算力度量折算值（即网络路由域的Metric）。

a）算力原始状态数据。算力系统通过预定接口向网络管控系统直接通告算力原始运行状态数据，如服务实例会话负荷、CPU/GPU占用率、内存占用比等。网络管控系统会对这些原始数据按照特定规则或算法进行处理，并生成对应的算网路由策略，指导网络路由和转发节点进行流量引流和路由。这种模式将显著增加网络管控系统的处理复杂度和运行负荷。

b）网络链路维度的算力Metric。算力系统将自身运行的动态数据折算成网络链路维度的Metric并向网络管控系统通告，后者据此执行传统IP路由。但是这种模式势必引入巨量的头结点路由开销，比如需要维护每实例、每出口节点、每链路的路由条目[5-6]。

4）算力与网络融合路由带来的多因子多策略问题

基于分组网络系统执行算力路由时，网络和算力融合路由将带来多因子联合优化的策略问题。算网双维度因子的全面融合将导致路由协议体系及其算路流程复杂度翻番，并严重冲击当前既有的路由和转发性能，无法实现与现网的平滑兼容。

2 基于服务标识的算力路由技术

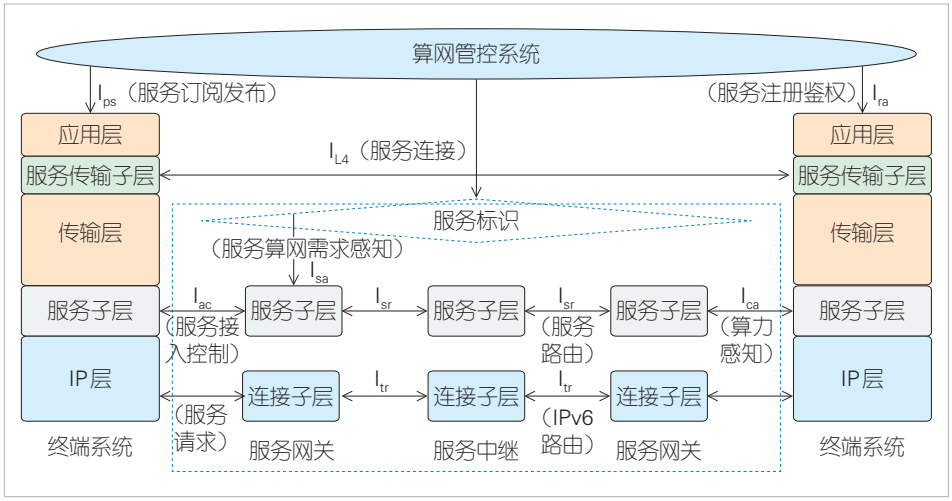
在IP路由协议体系中引入一个拥有独立语义的服务标识，将从根本上解决前文所述算力路由在IP网络中面临的主要问题，并提供统一的端到端架构解决方案。当然，服务标识也给IP网络带来一些新问题。这是在进行架构设计尤其是服务标识设计与界定的过程中需要特别考量的。

2.1 基于服务标识的算力路由架构

在算网融合调度和路由系统中引入服务标识，为IP分组网络提供了一个面向业务和算力系统的新型接口，使网络得以提供面向服务标识的路由和寻址功能。如图1所示，基于IP分组网协议的服务标识在数据面扩展定义和封装，并在控制面经由服务标识，打通算力系统动态资源和业务系统精细化SLA需求的感知接口，从逻辑上构成一个在IP分组网上的OverLay服务子层。传统分组数据网作为连接子层，为服务子层提供连接支撑能力。服务子层与连接子层之间以控制面服务标识为索引进行交互[7]。

如前文所述，服务标识在语义上与主机位置无关，因此传输层有可能通过服务标识保持业务连接，从而解决传统传输层终端或服务迁移连续性的问题，即主机地址变更导致的链路迁移仅在L3层执行，而L4层面向终端和用户的业务链接因为服务标识的位置无关属性得以维持不变，从而保障用户在这类场景下的业务体验。

数据面的服务标识是面向用户的一种轻量级算网服务能力集合表征。服务标识关联的算网质量和能力在特定算网运



▲图1 基于服务标识的算网路由架构

营管理域内可管可控，比如拥有端到端20 Mbit/s保障带宽的某种视频业务、10 ms端到端时延保障的渲染业务等。因此，服务标识内生支持精细化算网SLA需求的表征和接口。控制面基于这种服务标识的算网SLA Profile以及算网资源状态，生成以服务标识为索引的路由和转发策略，并将其下发到服务网关，指导业务流量转发。以入口服务网关对业务流量的转发和路由流程为例，用户侧报文通过服务标识表达对算力系统中特定算力服务的访问意图，以及这种服务在算网系统中的SLA需求。这里的服务标识并不指向特定的主机，而是由服务网关根据控制面的算网策略表选择特定的服务主机和网络链路，从而同时实现多服务实例间的算力优选和多网络路径中的路径优选，为对应的业务提供精细化的算网策略编排。由此可见，服务标识在东西向充当网络和算力系统之间的资源感知接口，在南北向充当网络和业务之间定制化业务SLA需求的高效感知接口。

需要指出的是，服务标识本身并不需要包含业务SLA需求的信息和参数，它仅需要在数据面和控制面之间充当映射接口即可。业务SLA需求语义由控制面来维护和表征。服务标识的可管可控、轻量级设计在解决安全性问题的同时，不会给服务网关硬件处理性能带来额外负担[8]。

1）服务标识的治理

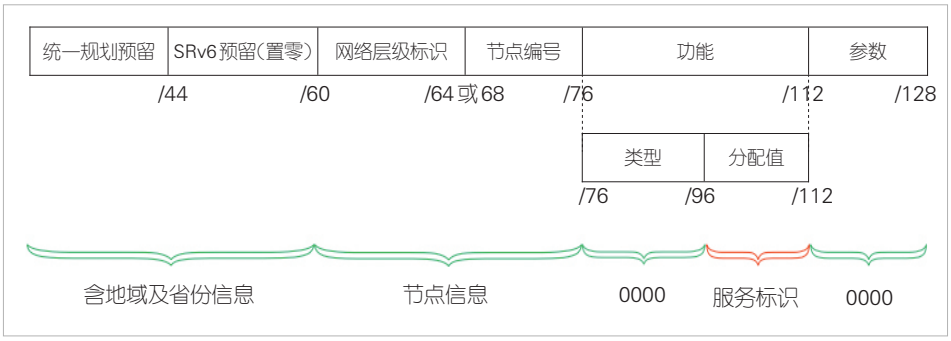
如前文所述，服务标识对用户、业务、算力和网络系统而言是一种接口。对于算网基础设施资源和业务运营方而言，服务标识是一种服务能力承诺。算网运营方应该对服务标识的全生命周期可管可控，即服务标识的注册、发布、订阅、更新和中止均应在算网运营系统的闭环治理范围内。在不同的算网运营管理域之间，服务标识的互通需要经过协商、映射甚至标准化，而这取决于特定算力服务的部署和运营模式。除了部分获得行业高度共识的基础服务涉及全网互通标准化之外，大部分服务标识的治理在单运营管理域内完成，无须标准化。

2）服务标识的封装

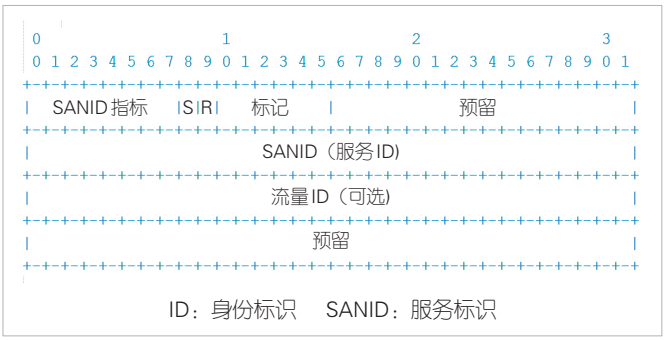
服务标识的表征对象是分布式多云部署的基础通用算力服务，因此，标识的对象空间有限。通常而言，16 ~ 32 bit足够覆盖既存的、可预见的服务类型。具体到IPv6报文头接口，服务标识的封装分为重用IPv6固定字段和扩展报文头定义封装两大类。

a）重用IPv6固定字段。源、目的地址以及流标签均可重用部分或者全部字段空间表征服务标识语义。如图2所示，基于SRv6地址结构的服务标识封装可重用功能中的低位16 bit。这种封装模式充分保留了SRv6地址原有的语义和功能。在重用IPv6固定字段的模式下，终端接口、业务请求流程以及协议栈均保持不变。此时方案落地环境兼容性较好。

b）扩展报文头定义封装。在IPv6标准扩展报文头目的选项头（DOH）、逐跳（HBH）、路由扩展头（SRH）中单独定义和封装服务标识头结构，如图3所示。服务标识头结构在服务标识之外封装了其他可选字段，用于特定场景。这种封装模式的优势是独立封装，不受服务网关节点本地处理机制的影响。服务标识可以直通算力服务系统，为算力系统网络提供增值功能，如可视化操作维护管理（OAM）、基于服



▲图2 基于SRv6地址结构的服务标识封装示例



▲图3 服务标识在IPv6扩展报文头中的独立封装示例

务标识的云内均衡和引流等。

2.2 层次化算力路由机制

将算力系统的全颗粒度算力资源状态信息通告同步到网络管控系统，将会导致现有IP分组网络协议收敛震荡和表项膨胀。为保持算力路由与现网路由协议体系的平滑兼容，需要对算力资源状态进行分类和聚合，在不同的网络节点维护不同类型的算力资源类别以及对应的算力路由表项，从而确保算力路由通告与现有IP路由之间的平滑兼容。全局算力路由表项条目数量仅与网络边缘节点有关，与云侧算力服务实例无关。这将压缩远端网络节点维护的算力路由表项空间，减轻节点的查表和处理负荷。高频变化的算力服务实例资源状态仅维护在本地网络边缘节点。这种层次化算力路由的机制，将控制面的端到端算网路由决策分层分布在网络远端和本地边缘节点，在转发流程上涉及两段路由转发：从网络远端到本地边缘节点、从本地边缘节点到算力服务实例。当然，这种层次化表项维护机制，将可能导致网络头结点算力资源信息的部分失真，可以满足绝大部分算力业务路由场景需求，但在极端异常场景下，仍需要引入丢弃或保护策略机制。

2.3 基于算力感知的算网路由解决方案

算力资源状态如何约束和影响网络边缘节点对算力和网络的选择，是算力路由的关键，也是基于IP路由的主要增量。因此，算力资源状态在网络控制面的呈现形态，是决定选择哪种端到端算力路由解决方案的关键因素。如前文第1节所述，算力参数主要有原始算力参数和网络维度算力Metric两种主要的呈现形态，与之对应的是两种不同的算力路由方案。

1） 基于算力映射的算力路由方案

算力系统向网络管控系统通告算力服务关联的原始算力状态数据。该原始算力状态数据与网络控制面路由决策系统之间的索引接口即为服务标识。网络控制面基于此类原始算力状态数据，结合网络资源状态、业务SLA需求，生成算网路由策略，完成原始算力状态数据到主机地址的映射。这个方案的优势是网络节点无须维护额外的算力路由表项。当然，在分布式路由协议方案下，算力原始状态数据的通告同样需要层次化状态维护机制，以平滑兼容现网路由协议。

2） 基于算力Metric的算力路由方案

算力系统通过一定的度量和折算机制，将算力服务关联的原始算力状态数据转换为网络维度的度量Metric，并通过特定协议接口向网络管控系统通告。具体来讲，这里的Metric可以是网络维度既有的Metric类型（如时延、带宽、等），也可以是新增的算力Metric类型。前者可以沿用既有的路由算法完成端到端算网路由编排，后者则需要扩展基于算力Metric的路由算法完成端到端路由编排。分布式路由协议方案下的层次化路由机制引入与上文所述类似，这里不再赘述。

2.4 基于算力与网络解耦的多因子多策略路由机制

在IP分组网络基础上执行算力路由，本质上是将传统IP网络的网络单维路由算法升级为算网二维路由算法。算力和网络两个维度的约束变量理论上是乘数关系，但在实际部署中，这种算网全维乘数算法将大幅增加路由算法的复杂度，甚至破坏现有IP路由协议机制的稳定性。远端网络边缘节点将“选算”和“选网”分离处理，使两类路由先决策再进行线性叠加，形成近似的算网融合优化路由策略。因此，算力和网络路由解耦，将算网二维乘数算法简化为一维线性叠加算法，并在算网融合的基础上，简化路由协议流程。需要说明的是，这种解耦机制不影响现有IP路由协议。

算网解耦以及算力、网络、业务SLA多种路由因子的引入，也为算网系统提供了多元调度机制，使能灵活的算网业务和资源运营模式。算网调度因子可以分为如下3类：

1）体验类：服务质量和体验相关的SLA指标，如时延、抖动、丢包等；

2）代价类：服务关联的算网资源成本、能耗等；

3）资源类：服务关联的算网资源的使用效率，如算网均衡度、算网利用率等。

相关算网调度策略有4种：

1）体验优先：体验类指标最优调度；

2）代价优先：体验类指标满足设定门限指标，代价类指标最优调度；

3）资源优先：体验类指标和代价类指标均设定门限指标，资源类指标最优调度；

4）资源均衡：体验类指标和代价类指标均满足设定门限指标，资源类指标均衡调度（资源使用率的方差最小）。

3 基于服务标识的算网路由评价体系及测试分析

相对于传统IP路由，算力路由带来了多方面的增量功能。这里我们从4个维度给出算力路由价值评价体系，并对方案的部分测试数据进行简要分析。

1）增强服务会话响应时延性能。算力路由通过数据面带内服务发现替代传统DNS带外服务发现。这里的服务响应时延是指：客户端首包发出到获得服务的时间间隔。DNS服务发现机制下的服务响应时延在100 ms ~ 1 s之间。本文以传输控制协议（TCP）3次握手为服务会话建立基准，使端到端响应时延低至2.76 ms，大大提高了服务会话的响应时延性能。

2）提升服务算网质量。网络感知与计算、网络质量是SLA双维度保障。在服务体验方面，网络可能会出现丢包、卡顿等现象。业务有效通量为用户的实际业务量。

3）实现资源利用率均衡。这包括网络资源利用率均衡、算力资源利用率均衡（池间），涉及网络负载偏离度和算力负载偏离度。其中，网络负载偏离度是指：调度过程中同一时刻不同网络路径的资源利用率的最大差值，算力负载偏离度是指：调度过程中同一时刻不同算力池的资源利用率最大差值。本文中，我们测试了两种机制下的资源利用率均衡度。在非均衡调度条件下，4个用户的流量均为20 Mbit/s，由资源池A提供服务，资源池B空载，此时负载偏离度比较高（>40%）；在均衡调度条件下，4个用户的流量均为20 Mbit/s，由资源池A和资源池B提供均衡服务，此时负载偏离度较低（<11%）；

4）提高资源利用效率。资源总量相同，网络可以承载更多的用户会话。

4 总结

在传统IP路由的基础上扩展算力路由功能，是实现算网融合的关键技术要素。算力路由与IP主机路由之间在机理和目标方面存在较大的差距，从而给方案部署带来诸多挑战。本文聚焦4类算力路由带来的协议和调度策略问题，并以平滑兼容现网协议和架构为目标，针对性地提出基于服务标识的算网路由架构方案。该方案的核心是引入独立于IP主机地址的服务标识，并构建用户与算网系统之间、网络与业务之间、网络与算力系统之间的简明高效互通接口。在此基础上，本文创造性地提出层次化算力路由、算力映射与算力Metric路由机制、基于算网解耦路由的多因子多策略算法等解决方案，为IP分组网络提供兼容性较好的端到端算力路由方案；同时，基于4个维度的算力路由价值评价体系，对部分典型场景测试数据进行分析。

参考文献

[1] 陈晓, 黄光平. 微服务架构下的算力路由技术 [J]. 中兴通讯技术, 2022, 27(1): 70-74. DOI:10.12142/ZTETJ.202201014
[2] 唐雄燕, 张帅, 曹畅. 夯实云网融合, 迈向算网一体 [J]. 中兴通讯技术, 2021, 27(3): 42-46. DOI:10.12142/ZTETJ.202103009
[3] 周吉喆, 杨思远, 王志勤. 面向业务感知的算网融合关键技术研究 [J]. 中兴通讯技术, 2022, 27(5): 2-6. DOI:10.12142/ZTETJ.202205002
[4] 朱海东. 云网一体使能网络即服务 [J]. 中兴通讯技术, 2019, 25(2): 9-14. DOI: 10.12142/ZTETJ.201902002
[5] 刘铎, 杨涓, 谭玉娟. 边缘存储的发展现状与挑战 [J]. 中兴通讯技术, 2019, 25(3): 15-22. DOI: 10.12142/ZTETJ.201903003
[6] 雷波, 宋军, 曹畅. 边缘计算2.0: 网络架构与技术体系 [M]. 北京: 电子工业出版社, 2021
[7] 陈晓, 郭勇, 谭斌, 等. 面向算网一体的开放服务互联架构 [J]. 信息通信技术, 2022, 16(2): 53-59
[8] HUANG D, TAN B, YANG D. Service aware network framework [EB/OL]. (2021-06-015) [2023-03-06]. https://datatracker. ietf. org/doc/html/draft-huang-service-aware-network-framework-01

作 者 简 介

黄光平，中兴通讯股份有限公司资深架构师；主要研究方向为下一代IP网络架构及关键技术，先后从事增值业务消息系统设计和开发、确定性网络以及远程宽带接入网关全球标准工作，近年聚焦算力网络架构、路由协议、算力标识等技术研究；发表论文8篇，申请专利30余项。

谭斌，中兴通讯股份有限公司未来网络技术研究项目经理；主要研究方向为IP网络、SDN系统架构与技术，先后从事有线路由器、接入产品开发、产品规划和市场等工作；申请专利2项。

吉晓威，中兴通讯股份有限公司IP产品规划工程师；长期从事IP网络、SDN/NFV等产品的规划和系统设计，目前研究方向为云网融合、算力网络；获中国算力大会创新先锋奖、SDN/NFV/网络AI最佳实践案例奖等奖项。