未来网络发展大会
FUTURE NETWORK
DEVELOPMENT CONFERENCE
NANJING · CHINA / 中国 · 南京

未来网络技术发展系列白皮书（2025）

中国移动云智算
新一代网络基础设施白皮书

第九届未来网络发展大会组委会
2025年8月

中国移动 China Mobile | 移动云

中国移动云智算新一代网络基础设施白皮书

引领云智算网络发展新范式

CHINA MOBILE NEXT-GENERATION NETWORK INFRASTRUCTURE FOR CLOUD AI WHITE PAPER

2025

编写说明

主要编写单位：

中国移动云能力中心

主要编写人员：

徐小虎、范文韬、姚军

目　录

第 1 章　AI 时代云计算发展新趋势

随着人工智能技术的迅猛发展，以大语言模型（LLM）为代表的 AI 应用场景不断拓展，对云计算基础设施提出了前所未有的性能与规模挑战。AI 不仅正在重塑云计算的使用方式，也正在推动公有云服务进入新一轮技术革新周期。

1.1　云计算市场趋势洞察

AI 云服务爆发增长

全球 AI 云市场呈现井喷式发展态势。据相关研究报告预测，至 2030 年，AI 云服务市场规模将突破 6476 亿美元，年复合增长率（CAGR）高达 39.7%。这主要受到两个因素驱动：1）大模型训练需求激增；2）AI 原生应用加速普及。然而，AI 基础设施的高性能网络、算力资源和技术体系仍存在巨大挑战，亟需云服务商加快布局和创新。

多云战略加速落地

企业在追求业务连续性、成本优化与数据安全的多重目标下，正全面拥抱多云部署策略。数据显示，已有 86%的企业采用多云架构，其中混合云仍是主流。如何实现多云环境下的网络互通与安全隔离，成为云网络架构面临的关键技术难题。

网络安全仍是上云首要关注点

在复杂多变的网络安全环境中，云上安全问题始终是企业上云的首要顾虑。Gartner 预测，全球云安全支出将从 2024 年的 115.12 亿美元增长至 2028 年的 217.73 亿美元，CAGR 达到 17.27%。多云架构带来的安全策略碎片化和合规复杂性，亟需新的网络安全解决方案进行系统性应对。

全球化云服务成为企业出海刚需

伴随中国企业加速“出海”，东南亚、中东等新兴市场的数字化转型迅猛推进，催生了跨区域云服务的强劲需求，跨国企业的云服务支出将不断提升。如何在保障数据主权和合规性的前提下，提供低延迟、可视化、广覆盖的云网服务，成为云基础设施全球化部署的核心挑战。特别的，云游戏作为低延迟、高并发的应用代表，其市场规模将于 2030 年突破 210.4 亿美元，年 CAGR 高达 44.3%。以中国游戏出海东南亚市场为例，网络延迟与覆盖能力成为制约用户体验关键。

1.2 云计算技术趋势洞察

AI 扩张定律持续生效

伴随大模型向多模态进化，其参数规模正以每年 10 倍的速度增长，已迈入十万亿级阶段。同时，万卡集群成为训练大模型的最低标配，十万卡级训练集群已成为主流趋势（如 xAI 基于 20 万卡集群训练 Grok3 模型）。这一趋势对智算网络和集群架构提出了超大规模、超高吞吐、超低延迟的极致要求。

多云部署技术日趋成熟

容器化技术（如 Docker）和编排调度系统（如 Kubernetes）已成为多云部署的基础设施标准。同时，Terraform 等基础设施即代码（IaC）工具的广泛应用，使得跨云资源管理实现了高度自动化与标准化。

AI 赋能网络安全成为新趋势

AGI 为网络犯罪分子提供了提升攻击复杂度的工具，ACL+安全组传统安全防护手段难以应付。Gartner 预测，到 2028 年，60%的零信任安全技术将集成 AI 功能（Predicts 2025: Scaling Zero-Trust Technology and Resilience），实现主动识

别威胁并实时响应，为云环境构建更加智能、精准的防护体系。

云网协同效应持续放大

随着企业上云与出海的不断推进，利用覆盖全球的高质量广域网提供企业虚拟广域网服务，相对 SDWAN1.0 更安全、更可靠。广域网流量工程实现网络资源的精细化运营，为客户（如云游戏客户）提供优质网络服务的同时，优化网络资源成本。

1.3 云计算产业竞争格局变化

AI 基础设施军备竞赛加速升级

全球头部云厂商纷纷加大对 AI 基础设施的研发和部署投入，以抢占智能时代的算力制高点。AWS 推出第二代自研 AI 芯片，并规划建设 40 万卡超大集群“Rainer”；GCP 发布第六代 TPU，服务于 10 万卡集群；阿里云则计划未来三年投资 3800 亿元用于云与 AI 基础设施建设，投资额超越过去十年总和。国内厂商亦积极推进自研智算网络方案，力求实现 AI 大模型训练所需的十万卡集群部署能力。

多云互联升级

AWS Cloud WAN：2022 年，升级 Transit GW 架构为 Cloud WAN，多云互联自动化和可视化能力大幅提升。

谷歌云 Cloud WAN：2025 年 4 月份，发布 Cloud WAN，为全球化客户提供便捷的虚拟广域网和多云互联方案。

云上安全能力持续强化

网络安全已成为云服务价值的重要组成部分。微软、谷歌等国际巨头持续通过高额收购扩大云安全版图：谷歌继斥资 54 亿美元并购 Mandiant 之后，2024 年拟以 320 亿美元收购 Wiz，成为其有史以来最大一笔收购案，同时也刷新全球网络安全领域的并购纪录。国内云厂商则通过“自研产品+第三方市场”双轮驱动，依托 MarketPlace 平台引入行业知名安全厂商，打造丰富灵活的云安全生态体系。

云网协同能力日益成为核心竞争力

差异化的网络服务能力正成为云服务商打造竞争优势的新焦点。Azure、GCP借助全球广域网与流量工程能力，为企业客户提供跨区域、高品质、低延迟的定制化服务。阿里云、腾讯云则基于广域流量调度系统，聚焦出海游戏等高价值客户，提供精细化、场景化的网络服务方案，提升客户体验与资源运营效率。

第 2 章　云智算发展方向与网络技术体系构建

2.1　云智算发展方向

在 AI 技术快速演进的时代背景下，云基础设施正从通用计算平台向以 AI 为中心的云智算形态加速转型。中国移动云智算顺应趋势，从战略层面聚焦“AI 优先、全球化、内生安全、差分服务”四大发展方向，全面推动云网架构升级和能力体系重塑，打造面向未来的智能基础设施底座。

2.1.1　AI 优先：打造算力驱动的核心能力体系

随着大模型与生成式 AI 快速演进，AI 原生需求正重构云网算基础设施体系。面向未来，AI 优先的发展路径将以算力为核心、以网络为底座，推动算力供给体系和网络架构深度融合，形成支持智能调度、高效服务和弹性编排的基础平台。

核心技术突破

面向 AI 流量高爆发、高带宽需求的演进趋势，技术体系需围绕网络、计算、存储等关键模块加快突破。在芯片层面，推进自研 AI 加速芯片、智能 DPU、RDMA

NIC 等核心部件优化升级；在设备层面，推动高通量交换机、低时延拓扑架构适配 AI 集群需求。通过端到端软硬协同，打通数据处理瓶颈，为 AI 训练与推理提供高性能承载平台。

标准牵引产业生态

为实现多厂商设备与系统的协同演进，有必要推动形成统一开放的 AI 基础设施标准体系。面向 AI 集群组网结构、通信协议、调度接口、性能指标等方向构建规范标准，提升产业间协同效率。通过标准牵引，联动芯片商、设备商、主机商、网络厂商、调度平台等上下游生态共同参与建设，推动形成开放、兼容、灵活的智算产业生态体系。

构建一流智算集群

面向未来 AI 大模型演进趋势，应规划超大规模智算集群的布局方向。智算资源配置将向“集中+分布”融合演进，集中承载模型训练、分布支撑任务推理。训练集群将具备百万卡级别规模、超高网络带宽域和灵活任务调度能力；推理资源池将按服务域动态部署，实现算力节点与用户流量的灵活适配。依托智算平台统一调度，推动算力高效供给和资源弹性使用。

2.1.2 全球化&多云互联：构建全域覆盖、多云互联的网络资源体系

在“出海战略”与“全球服务”的持续推进背景下，构建全球可用、体验一致、路径可控的云智算网络基础设施，正成为全球业务发展的关键支撑。云网一体架构需面向全球广域资源能力的融合发展，全面提升覆盖能力与业务支撑能力。

全球骨干网络一体化

推动境内骨干网与中移国际网络的架构融合是全球化能力建设的重要方向。通过统一的控制平面与调度策略，实现境内外路径互通、策略统一、服务一致。境外 POP 节点与境内数据中心之间应具备高速、安全、低时延的传输能力，为全球业务提供端到端的路径保障。同时，增强跨区域链路质量感知与带宽弹性调配能力，实现多业务并发承载下的稳定服务输出。

多云互联能力升级

面向全球范围内的多云环境，需持续强化跨云、混合云、异构云的网络互联能力。未来互联能力建设将从基础连通走向智能调度，支持 BGP 多路径互联、动态路径切换、QoS 策略传递等机制。基于云联网架构演进，构建支持多云互联、云边协同、可编排调度的广域连接体系，为全球范围的混合部署场景提供灵活的网络底座支撑。

2.1.3 内生安全：构筑多层次可编排的安全能力

随着数据要素价值持续提升与攻击手段愈发复杂，构建内生安全能力已成为云基础设施演进的关键方向。未来的安全体系将不再是被动防御的附加模块，而是与网络、计算、存储等能力深度融合的原生组成部分。基于“云网安一体、能力即服务”的理念，安全能力的体系化构建可从以下几个方向推进：

产品体系融合发展

将安全能力体系化、标准化、平台化是未来云安全服务演进的重要方向。一方面，可通过整合自研能力与第三方生态能力，形成覆盖 IaaS、PaaS、SaaS 等层级的多样化安全产品池；另一方面，应重点推动产品间策略统一、接口兼容与协同编排能力，提升整体安全产品生态的可插拔性与服务灵活度。

构建安全资源池

面向多租户、多业务场景的安全能力交付，需探索将各类安全功能资源池化、服务化的组织方式。通过统一抽象防护能力单元，建立可动态扩缩、策略隔离、安全隔区灵活组合的安全资源池，有助于提升资源利用率与响应效率。同时，结合自动化调度与平台化运维，可增强大规模弹性安全能力的服务支撑能力。

云网安一体化部署

安全能力与网络能力融合将是下一阶段能力演进重要路径。通过安全服务链等机制，在流量路径中按需加载安全能力，打破传统单点设备部署模式，推动安全能力随业务动态编排。未来应重点提升安全功能模块插入、并发承载、状态同步等能力，实现更高粒度、更强隔离、更可控的网络安全策略实施。

2.1.4 差分服务：提供可预期的优质服务体验

在多样化业务场景和细分行业快速发展的趋势下，传统统一化的网络能力难以同时满足对性能、成本、体验的多元化需求。面向未来网络服务体系，可从能力分级、调度智能化、策略可编排等方向开展差异化服务能力构建：

按需定制网络能力

网络能力应具备面向业务特征的定制交付能力，支持在带宽、时延、隔离性、SLA 保障级别等维度灵活组合。结合租户自服务能力和可视化配置平台，业务方可自主选择服务参数并完成在线配置。网络能力按需生成、定向激活，将成为通用云网络向行业专用网络演进的关键抓手。

网络资源精细调度

未来网络调度将从链路级别走向业务级别。需构建以用户等级、业务类型、实时负载等为输入的多因子调度引擎，实现链路资源按策略动态分配与路径选择。同时，应提升调度闭环能力，支持调度策略实时回调与效果反馈，在性能可控基础上提升调度效率和调度稳定性。

多层次服务能力体系

构建统一架构下的多等级服务供给模型，是实现差异化服务能力的核心路径。可在标准网络服务之上构建增强型、高保障型能力模块，提供带宽预留、低时延链路选择、优先调度等功能。面向行业用户，还可支持 SLA 协议签署、专属资源预置、定制化路径控制等能力，实现通用服务与特定场景的兼容覆盖。

2.2 云智算网络技术体系构建

为支撑云智算在 AI 原生、多云协同与全球部署等多场景下的持续演进，中国移动云智算构建了“物理网络+虚拟网络”双层协同的技术体系，全面满足高性能、高可靠、高灵活的云网融合需求。技术体系示意图如图 1 所示。

图 1　云智算网络技术体系示意图

2.2.1　物理网络：云智算高性能算网承载基础

物理网络作为云智算运行的底层支撑体系，承担着算力、存储、服务等核心资源的互联互通与调度保障职能。中国移动云智算聚焦“高吞吐、低延迟、高可靠、可扩展”的网络能力要求，从数据中心内部网络延伸至广域承载与边缘接入，形成了结构清晰、功能完备、性能领先的物理网络体系。

智算网络面向大模型训练与推理等高密度计算场景，围绕“超高吞吐、超低延迟、超高可靠性”三大特征进行优化，通过 Scale-Out 网络实现大规模 GPU 集群间的高速互联，通过 Scale-Up 网络支持高带宽域内部跨 GPU 的高性能通信，构建起满足 AI 原生需求的智算网络。

数据中心网络是算力基础设施的核心承载。数据中心网络面向通用计算资源池建设，强调可扩展性、高可用性、低延迟与低成本的综合均衡，采用 SHALL（Scalability、High Availability、Low latency、Low cost）设计理念，支撑大规模资源的灵活部署与高效调度。

IP 广域网络承担着数据中心之间和用户终端之间的跨域连接任务，体系分为 Internet 广域网与 DCI 广域网两个部分。前者服务于公网访问场景，提供低延迟、广覆盖、高可用的网络体验；后者面向数据中心互联场景，支持 SLA 感知的路径调度与带宽保障，实现跨地域、跨园区智能算力调度。两者协同构建了具备差分保障能力的广域承载体系，为多业务类型提供多级别的网络服务支持。

光传输广域网络为数据中心与边缘节点之间提供高速、稳定的物理链路。中国移动采用分布式集群互联与开放解耦架构，实现大容量、低时延的传输通道建设，全面支撑超大规模 AI 训练流量、数据同步与跨节点访问需求，夯实算力全球化布局的底层传输能力。

2.2.2 虚拟网络：云智算灵活调度与安全编排核心

基于统一的物理承载底座，虚拟网络作为云智算的服务交付层，承担着资源调度、流量管理与安全隔离的核心功能。中国移动云智算围绕“云内网络”与“云间网络”两大关键模块，构建了具备灵活编排、弹性调度与全球互通能力的虚拟网络体系，全面支撑多租户、多业务、多云环境下的高质量云网服务。

云内网络是支撑云智算计算与服务调度的关键基础。依托 SDN 架构实现控制与转发解耦，使网络具备集中管控、灵活编排与高可扩展性，满足多租户环境下的高性能通信与资源隔离需求。在此基础上，构建网络安全服务链能力，将防火墙、入侵检测、DDoS 防护等安全功能以服务链方式灵活插入业务路径，提升整体网络环境的安全性、可控性与弹性。

云间网络聚焦于实现不同区域、不同云环境之间的互联互通与统一调度。依托“云联网”平台，中国移动云智算构建了横跨全国乃至全球的多云互联能力，支持跨区域 VPC 打通、异构云资源融合、路径策略编排与 QoS 保障等能力。云间网络不仅提升了资源的使用效率，还为多云部署、混合云协同与全球业务出海提供了高品质、可预期的连接保障。

通过物理网络与虚拟网络的双层协同，中国移动云智算构建了具备“强承载、广覆盖、易编排、高可靠”能力的新型网络技术体系，为智算网络、多云协同、智能算力调度与全球业务部署提供了坚实的底座与灵活的网络服务能力。

第 3 章　云智算物理网络架构：智算网络

随着 AI 大模型参数规模突破十万亿级，训练数据集规模迈入数十万亿 Token，AI 训练集群的计算强度与通信复杂度呈指数级提升。智算网络作为云智算基础设施的核心组成部分，支撑超大规模 AI 训练任务，是保障大模型训练效率与稳定性的关键底座。智算网络主要由 Scale-Out 网络和 Scale-Up 网络两部分构成，分别服务于 GPU 服务器间以及单服务器或超节点内部 GPU 之间的高效互联通信。

3.1　Scale-Out 网络：构建超大规模训练集群

Scale-Out 网络主要用于实现 GPU 服务器或超节点之间的互联，是大规模集群数据并行、流水线并行等通信模型的基础支撑网络。

3.1.1　Scale-Out 网络需求

Scale-Out 智算网络作为 GPU 服务器之间通信的主干网络，需承担海量数据并行、流水线并行等任务中的高频参数同步，其性能直接决定集群整体训练效率和可扩展性。

当前 AI 训练集群对 Scale-Out 网络提出如下核心诉求：

超大规模

随着集群规模迈向十万卡级别，网络体系必须具备大规模横向扩展能力，

支持海量 GPU 节点的高效互联，同时确保系统在扩展过程中的可管理性与可靠性。在如此规模下，训练任务依赖于 GPU 之间的大量并行通信进行梯度同步与参数交换，集合通信成为高频操作。网络架构需要支持扁平化拓扑、低阻塞比的三层 CLOS 结构，确保在高并发通信场景中维持稳定的吞吐和较低的路径开销，支撑超大规模 AI 集群稳定高效运行。

超高可靠

AI 模型训练通常持续数小时至数天，训练中断将导致整个流程被迫停止，并需从上一次 checkpoint 进行断点恢复并重新训练当前轮次，不仅严重影响整体训练效率，还显著增加计算资源浪费与系统管理复杂度。因此，智算网络在架构设计中需高度关注连续性保障能力，通过路径冗余、设备高可用、故障快速切换等机制，提升系统在大规模长周期训练任务中的稳定性和容错能力。

当前主流的集合通信广泛采用 RDMA 技术（InfiniBand/RoCE），以实现高性能、高并发的并行通信。然而，RDMA 对网络丢包极其敏感，即便仅出现 1% 的丢包，也会导致通信吞吐性能下降一半以上。为保障大规模集群通信的可靠性，网络系统需具备完整的无损传输能力，构建端到端稳定可控的通信路径，降低训练中断风险，确保训练任务在复杂网络环境下仍能高效完成。

超高吞吐

AI 训练过程呈现“计算—通信”交替进行的模式，通信时长直接决定整体训练周期。提高网络系统的吞吐能力，有助于缩短通信阶段耗时，降低通信占比，从而释放 GPU 算力资源，提高集群整体运行效率。在训练规模不断扩大的背景下，网络高吞吐能力成为系统扩展性的核心基础。

AI 训练负载呈现流数少、流量大、并发度高的特征，若缺乏精细的调度机制，易导致链路资源利用不均，引发局部路径拥塞。为此，网络体系需支持精细化流量调度机制，结合高效负载均衡与拥塞控制算法，实现通信任务的合理分发和链路负载均衡，避免单路径瓶颈影响全局训练性能。

超低延迟

在大模型训练过程中，参数同步通常采用全节点参与的集合通信方式，训练任务必须等待所有 GPU 完成当前轮次的通信，才能进入下一轮计算。这种同步机制使得网络延迟成为决定训练效率的关键因素。一旦某些 GPU 因通信路径

延迟较高而拖慢整体同步进度，成为“木桶效应”中的短板，导致其他 GPU 处于等待状态，形成整体性能瓶颈。为有效应对“木桶效应”带来的性能损失，网络系统需从多个层面提升通信路径的时延控制能力。

3.1.2 技术路线对比：以太网 VS InfiniBand（IB）

当前，构建 AI 训练集群的网络互联方案主要面临两条技术路线选择：一是以 IB 为代表的高性能专用网络方案，二是基于开放以太网进行协议与架构升级的技术路线。两者在性能表现、产业生态、成本控制及可扩展性等方面各有特点，适用场景与发展路径也存在明显差异。

性能：IB 在 AI 训练场景中长期占据主流地位，具备低延迟、高带宽的通信优势，且其原生支持 RDMA（远程直接内存访问）机制，适用于集合通信密集的计算任务。但随着训练规模扩展至万卡甚至十万卡级，其网络调度灵活性和系统稳定性面临更高挑战。

产业生态：IB 技术相对封闭，其核心芯片与设备长期被国外厂商垄断，国内替代方案不足，存在一定的技术依赖风险。相比之下，以太网产业生态开放，拥有广泛的应用基础，涵盖数据中心、企业网络与互联网等多个领域，国内厂商具备较强的设计与制造能力，为系统建设与国产化提供更大自主空间。

成本与运维：IB 网络设备价格较高，网络建设与运维成本昂贵，尤其在大规模部署下，对网络管理经验要求较高。而以太网方案在设备采购、部署、故障排查等环节更为成熟，运维体系完善，具有明显的成本优势，适合在大规模训练场景中推广应用。

负载与流控机制：IB 支持自适应路由和基于信用的流控机制，能够动态应对链路拥塞，保障传输稳定性。而传统以太网则依赖静态的 ECMP（等价多路径）负载均衡和 PFC（优先级流控）机制，在面对 AI 集合通信这类流量大、并发高的场景时，容易出现拥塞传播、队头阻塞等问题。

下表总结了两种技术路线在主要维度下的差异：

表 1　IB 与以太网技术路线对比表

比较维度	IB 技术路线	以太网技术路线
产业生态	技术封闭，长期被海外厂商垄断，国产替代难度大	产业生态开放，国内具备完整链条，国产替代潜力强
成本结构	网络设备价格昂贵，运维成本高	建设与运维成本大幅低于 IB，经济性优越
负载均衡能力	支持自适应路由，链路动态调度能力强	传统 ECMP 为静态均衡，对集合通信流量支持不足
流控机制	基于信用的流控机制，拥塞控制效果佳	基于 PFC，易出现队头阻塞等问题，需协议优化
性能扩展性	性能强，但协议私有，扩展性与成本面临挑战	借助协议创新，可满足百万卡集群需求，具备长期可演进性

在 AI 大模型训练对网络规模、可靠性和成本提出更高要求的背景下，中国移动云智算选择以开放以太网为基础，通过自研协议和网络架构创新，构建具备强可扩展性与国产可控能力的智算网络技术体系，规避核心技术依赖风险，为支撑百万卡级别的训练集群奠定坚实基础。

3.1.3　Scale-Out 网络架构

面向 AI 大模型训练规模不断扩展的趋势，中国移动云智算围绕万卡至十万卡级 GPU 集群的互联需求，设计了基于开放以太网的新一代 Scale-Out 智算网络架构，如图 2 所示。该架构通过拓扑优化、设备升级、协议创新与冗余保障，全面支撑大规模训练任务对网络性能的极致要求。

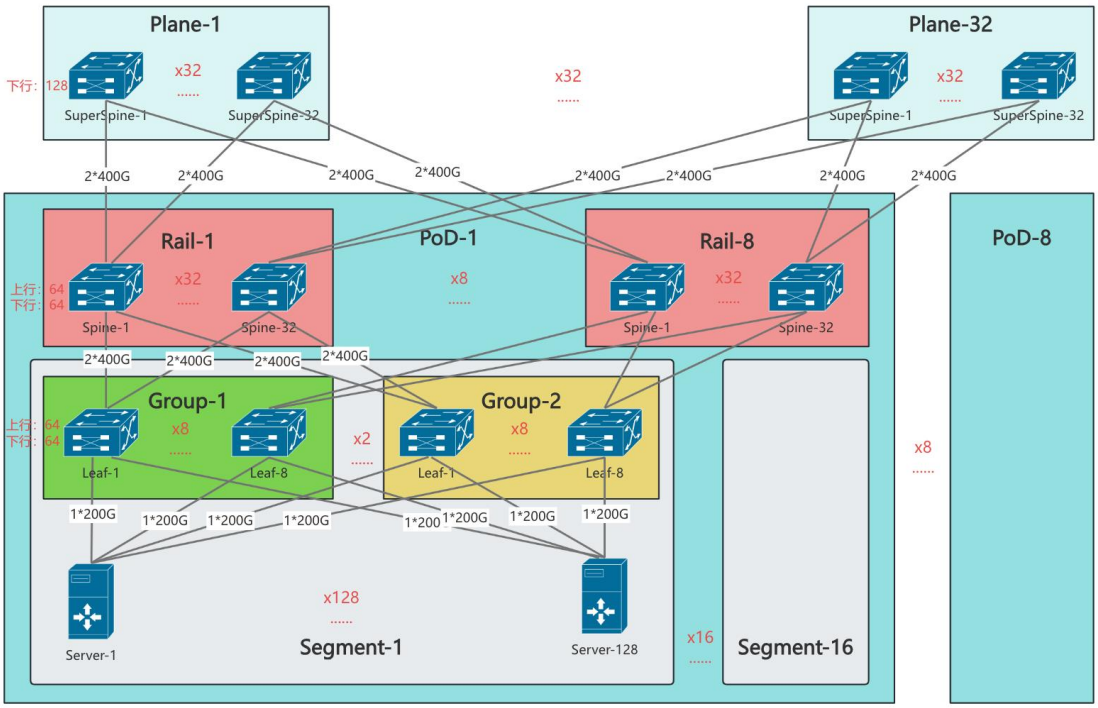

图 2　Scale-Out 网络架构示意图

拓扑结构设计：多层 CLOS 与多轨道多平面组网

在拓扑结构设计上，Scale-Out 网络采用了多层 CLOS 架构，并引入多轨道多平面组网策略。相比传统 CLOS 网络布局，多轨道设计通过连接不同服务器内的同号 GPU，有效降低了跨层数据转发，降低了整体网络延迟。多平面连接则通过增加单卡对外接口数量，使得集群规模倍增。本拓扑结构设计，单 PoD 内部可支持多达 6 万张 GPU 互联，显著优于当前主流集群架构，为未来百万卡级集群扩展预留了充分空间。

设备与链路配置：高带宽支撑

在设备配置方面，核心交换设备基于最新一代 51.2Tbps 单芯片交换机，具备高端口密度与超大交换容量，充分释放集群内部网络能力。每张 GPU 配置双 200G 端口，分别接入两个独立的网络平面，扩大集群规模的同时，通过物理隔离实现双平面互不干扰，确保大模型训练中梯度同步、参数传输等高频集合通信场景下的高效传输需求。

协议创新：FARE 协议提升负载均衡性能与带宽利用率

为进一步突破以太网在集合通信流量模型下存在的负载均衡瓶颈，中国移动云智算自主创新提出 FARE（Full Adaptive Routing Ethernet，全自适应路由以太

网）协议。FARE 协议针对 AI 训练中流数少、单流大、高并发的流量特征，支持多路径动态包喷洒（packet spraying）机制，能够根据链路实时拥塞状态灵活选择转发路径，极大提升了网络带宽利用率，有效缓解了传统以太网 ECMP 静态均衡导致的链路资源浪费问题。基于 FARE 协议，Scale-Out 网络实测带宽利用率达到 95%以上，显著优化了大规模训练集群的通信效率与资源利用水平。

高可靠性设计：多平面冗余容灾

为增强网络的整体可靠性与业务连续性，Scale-Out 网络通过多平面冗余机制提升系统韧性。GPU 服务器双网口分别接入两套独立交换平面，在任一链路、交换设备或平面发生故障时，另一平面能够无缝接管流量，确保训练任务不中断。同时，在跨设备连接中引入端口故障转移机制、链路状态实时探测与快速切换策略，提升训练网络的稳定性与可用性。

延迟优化与性能指标：极致低时延与高带宽利用

在延迟控制方面，通过多轨道组网缩短了跨服务器转发路径，结合 FARE 协议优化动态负载分配，Scale-Out 网络端到端通信延迟控制在 10 微秒以内。集合通信密集型的训练场景下，网络尾延迟得到显著压缩，有效避免了因单节点通信延迟导致的整体训练进度拖慢，进一步提升了 GPU 利用率和训练吞吐能力。

3.2 Scale-Up 网络：构建更大规模超高带宽域

Scale-Up 网络主要面向服务器内部或超节点内部 GPU 间的高速通信，是实现张量并行、MoE 专家并行、远端内存访问等 AI 模型通信需求的关键通道。

3.2.1 Scale-Up 网络需求

Scale-Up 网络主要面向超节点内部的高性能 GPU 互联，是满足 AI 模型张量并行、专家并行等深度融合计算需求的关键承载体系。随着模型规模的不断扩大，集群内部对带宽、延迟、语义兼容性及通信效率的要求持续提升。未来 Scale-Up 网络需面向以下关键技术指标进行持续演进与能力强化。

超高带宽

当前主流的 AI 大模型均采用 Transformer 架构，且逐步引入 MoE（Mixture of Experts）框架，以支撑万亿级参数规模。MoE 的引入虽然显著提升了参数稀疏性与模型效果，但也使得通信需求急剧增长，尤其在专家路由和反向梯度传播中产生大量 All-toAll 通信。随着参数规模突破单机承载能力，跨服务器、跨节点的专家并行需求已成为刚性诉求。为了保障 MoE 训练效率，Scale-Up 网络需构建支持 64 卡以上的高带宽域，实现数十至上千卡 GPU 的高速互联。

超低延迟

在 AI 模型训练过程中，为实现高效集合通信，GPU 间需进行频繁的数据交换与远端内存访问。为了满足跨 GPU 远程访问需求，网络系统需具备亚微秒级（如百纳秒）通信延迟控制能力。过高的转发延迟将直接影响到内存访问效率和计算通信重叠能力，进而导致 GPU 资源空闲与整体训练效率下降。

内存语义

Scale-Up 网络的性能优化不仅依赖带宽与延迟，还涉及语义支持。在 GPU 服务器内部，跨卡通信通常通过 Load/Store/Atomic 等原生内存语义访问操作实现直接交互，在性能和编程模型统一方面均具有显著优势。为延续此种原生内存语义通信，Scale-Up 网络需尽可能提供对内存语义的支持能力，尤其是在 RoCE 或新型以太传输机制下扩展语义能力接口，在保持应用生态无感知迁移、简化通信编程复杂度方面具备重要意义。

在网计算

随着集群规模扩大与通信流量增加，集合通信算子所带来的网络压力不断上升。部分 Scale-Out 智算网络系统，如 NVLink 与 IB 系统，已在交换设备中实现了在网计算能力，支持基于数据包的加法运算操作，在交换过程中完成部分集合通信逻辑。Scale-Up 网络域具有带宽高交换节点密集的特点，同样适合在网计算架构部署，以降低集合通信流量。

3.2.2 网络技术选型：以太网 VS PCIe

在 Scale-Up 通信中，传统方案通常依赖 PCIe 交换芯片作为节点内部互联手

段，但随着规模扩大，其局限性日益明显。以太网则在带宽能力、生态开放性和国产可控性方面展现出更大潜力。

具体来看，PCIe 在互联延迟方面具有天然优势，通常可控制在 100ns 以内，并且原生支持内存语义操作，如 Load/Store/Atom。但 PCIe 受限于带宽扩展速度，目前主流商用芯片（如 PCIe Gen5）每个 Lane 仅支持 32Gbps，总交换容量仅为 4.6Tbps，难以满足大规模超节点互联需求。同时，PCIe 交换芯片市场长期由少数海外厂商主导，国产替代难度较大，存在显著的技术风险。

相比之下，以太网已经实现了 224G SerDes 商用，下一代单芯片交换容量可达 102.4Tbps，远超 PCIe 体系。虽然以太网原生只支持消息语义（Message Semantics），不直接支持内存语义，但通过增加适配层，可以实现对 Load/Store 接口的兼容。此外，通过优化转发流程，先进以太网交换芯片可将延迟压缩至 300ns 以内，基本满足跨 GPU 高效通信需求。更重要的是，以太网产业生态开放，国内已具备从芯片到设备完整的产业链基础，支持长期可控发展。

综合比较如下表所示：

表 2　PCIe 与以太网技术路线对比表

比较维度	PCIe 技术路线	以太网技术路线
带宽能力	Gen5 速率 32Gbps/Lane，总容量约 4.6T	商用 224G SerDes，交换容量达 51.2/102.4Tbps
产业生态	技术封闭，主导厂商少，国产替代困难	开放标准，国内芯片与设备产业链成熟
内存语义支持	原生支持 Load/Store/Atom 操作	需通过适配层实现内存语义映射
延迟特性	极低，约 100ns 以内	以太网可优化至 300ns 以内
可扩展性与演进潜力	带宽增长缓慢，扩展受限	带宽提升迅速，支持大规模超节点布局

综上，基于开放以太网的技术路线，在大规模训练集群建设中更具发展潜力和系统弹性，成为 Scale-Up 智算网络的优先选择。

3.2.3 超节点硬件选型：开放架构 VS 封闭架构

在超节点系统设计方面，中国移动云智算围绕开放、模块化、灵活部署的理念，构建了面向未来 AI 大模型训练的硬件选型方案，示意图如图 3 所示。

计算节点采用轻量化定制的 8 卡 OAM 2.0 GPU 服务器，单机柜内部署 4 台服务器，合计 32 卡。交换节点选用标准以太网交换机，具备高密度高速接口，支持 AEC 铜缆连接计算节点、CPO 光模块连接交换节点之间高速互联。这样的设计既保证了高带宽互联需求，又大幅提升了硬件部署灵活性和系统扩展性。

在互联方式上，服务器与交换机之间采用标准 AEC 铜缆实现 L1 层高速互联，交换机之间则通过 CPO（Co-Packaged Optics）光纤互联构建 L2 扩展层。该模式充分利用了以太网的带宽扩展优势，降低了整体互联系统的延迟与功耗。

图 3　超节点硬件架构示意图

在散热方式上，超节点硬件系统根据机房环境条件灵活适配风冷或液冷散热方案，单机柜功耗控制在 40kW 至 60kW 区间，大幅降低对数据中心供电与制冷改造的要求，具备更好的适配性与部署灵活性。

在超节点硬件架构选择上，开放架构方案相比封闭式超节点具有明显优势：

封闭架构（如 NVIDIA NVL72 方案）采用高度集成设计，将计算单元与交换

模块封装于一体，虽然初期性能强大，但存在硬件绑定、扩展受限、运维复杂、功耗极高（120kW 以上）且必须液冷改造等问题，后期升级与维护成本巨大，缺乏长期演进能力。

开放架构（中国移动提出方案）则将计算与交换节点物理解耦，采用标准化组件与接口互联，不仅支持按需扩展，灵活部署，而且单柜功耗适中，无需强制液冷改造，具有更优的成本结构与国产替代潜力，能够更好适配未来 AI 基础设施发展的需求。

表 3　封闭架构与开放架构技术路线对比表

比较维度	封闭架构（如 NVIDIA NVL72）	开放架构（移动云方案）
系统集成	计算与交换节点高度集成	计算节点与交换节点物理分离，标准接口互联
扩展灵活性	扩展受限，需整机柜整体升级	支持按需横向扩展，灵活叠加资源
功耗控制	单柜功耗高达 120kW 以上，需大规模机房改造	单柜功耗控制在 40-60kW，适配常规 IDC
散热方式	必须采用液冷系统，成本高、维护复杂	支持风冷/液冷灵活切换，适配多种环境
厂商锁定风险	定制化严重，绑定单一供应链，升级受限	开放标准，避免锁定，支持国产自主可控

因此，中国移动云智算选择基于开放以太网+标准硬件组件的开放架构路线，既实现了超节点内部高效互联，又为后续系统扩展与演进奠定了坚实基础。

3.2.4　Scale-Up 网络架构

基于上述硬件与网络技术选型，Scale-Up 网络架构设计如图 4 所示。

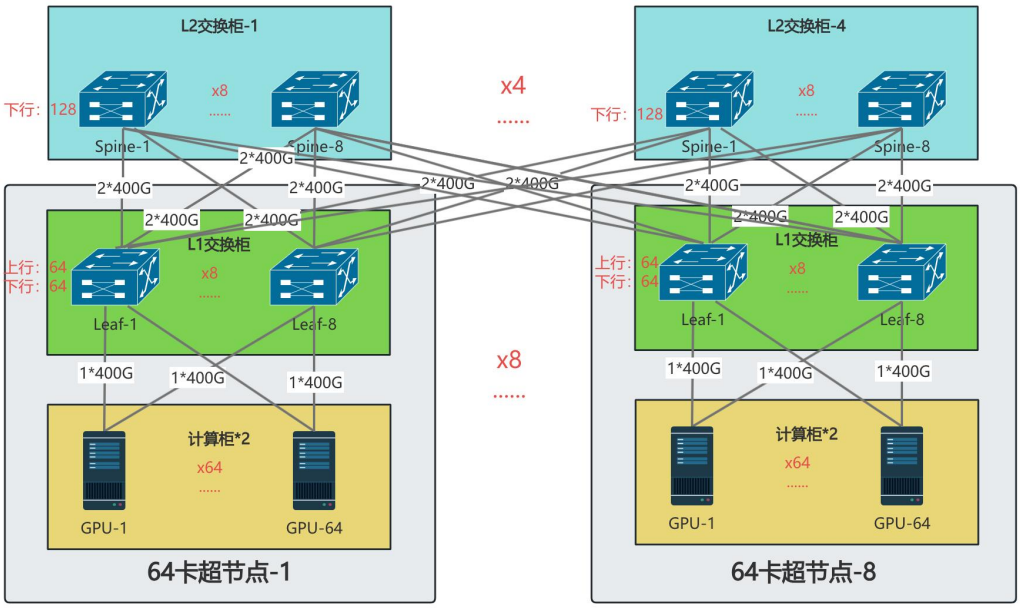

图 4　Scale-Up 网络架构示意图

局部互联（L1）：以 32 卡为基础单元，服务器通过高速铜缆连接至本地交换节点，构建高带宽通信域。

跨单元互联（L2）：多个基础单元通过 CPO 光纤高速互联，形成支持最多 1024 卡的大规模超节点集群。

通信协议支持：基于优化的 RoCE 协议实现远端内存访问，同时通过适配层支持内存语义访问，兼容 AI 大模型训练通信需求。

性能指标：跨 GPU 远端访问延迟控制在 300ns 以内，满足超大规模模型推理和训练中对高速同步的一致性要求。

通过开放解耦的 Scale-Up 网络架构，移动云开放超节点方案能够在满足高性能通信需求的同时，保持系统的开放性、灵活性与长期演进潜力，为 AI 智算时代的超大规模基础设施建设提供坚实支撑。

3.3 Scale-Out 与 Scale-Up 融合组网方案：支撑百万卡级 AI 集群的下一代智算网络

面对大模型训练中跨节点高带宽、高并发、低延迟通信的复合需求，急需构建一种融合两类网络优势、具备统一调度能力与弹性扩展能力的新型网络架构。为此，中国移动云智算提出基于开放以太架构的 Scale-Out 与 Scale-Up 融合组网方案。

3.3.1 Scale-Out 与 Scale-Up 融合组网需求

单一 Scale-Out 或 Scale-Up 网络体系在大规模训练任务中存在明显瓶颈：Scale-Out 适合大规模节点之间的数据并行任务，但在跨节点专家访问与远程读写方面延迟偏高；而 Scale-Up 擅长低延迟互联，但扩展能力受限，难以支撑百万卡规模部署。因此，需要兼顾超高带宽、极致低延迟与规模可拓展性的融合网络架构，打通 AI 集群内部与外部的通信瓶颈，全面释放 AI 大模型的训练潜力。

3.3.2 Scale-Out 与 Scale-Up 融合组网架构

融合组网方案将智算网络划分为超节点通信域、Segment 通信域与 Segment 互联域三个层级，实现覆盖三个维度的高性能网络架构。

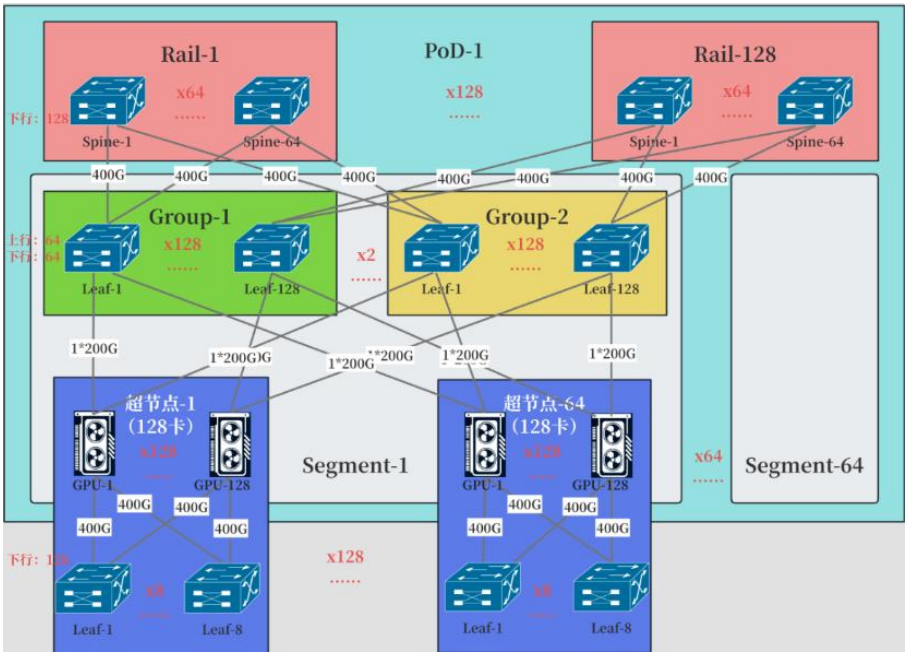

图 5　Scale-Out 与 Scale-Up 融合组网架构示意图

在超节点通信域，采用 128 卡液冷开放架构超节点作为基础计算单元，节点内通过优化 RoCE 协议配合内存语义支持，构建高性能高带宽通信域。高带宽铜缆与高容量以太交换芯片构成局部互联网络，通信延迟控制在 300 纳秒以内，满足亚微秒级别访问需求。

在 Segment 通信域，以 128 个超节点构建一个 Segment 单元，支持最多 16,384 张 GPU 的高效互联。内部采用多轨双平面 CLOS 结构，提升集合通信的并发处理能力；通过冗余路径设计和多平面隔离机制，实现训练任务的通信稳定性与高可用性。

在 Segment 互联域连接所有 Segment，全局网络部署 FARE 协议，实现端到端的动态路径选择与全局负载均衡。网络支持包级粒度的动态调度与路径喷洒机制，能够根据链路负载与拓扑变化智能调整数据传输路径，确保百万卡规模集群通信稳定高效。

3.3.3　技术挑战性、创新性和先进性

技术挑战性

Scale-out 网络支持百万卡集群：当前业界最大规模为十万卡集群，网络规

模扩张十倍将面临稳定性挑战增加、网络转发延迟增大、网络吞吐性能变差等严峻挑战。

Scale-up 网络支持千卡超节点：对标英伟达私有 NVLINK 协议，通过以太网技术优化，实现超高带宽（相对 scale-out 高一个数量级）、超低延迟（相对 scale-out 降低一个数量级）的高带宽域，存在巨大挑战。

技术创新性

网络架构创新：以超节点作为基本建设单元，采用双层多轨道、单层多平面的 CLOS 网络架构，可基于两层网络构建百万卡集群。

网络协议创新：主导 IETF 个人草案-FARE（draft-xu-idr-fare，draft-xu-lsr-fare，draft-xu-rtgwg-fare-in-sun)，确保智算网络高吞吐、低延迟。主导 IETF 国际标准（RFC9793），助力高效 MoE 通信。

硬件工程创新：业界首创开放解构超节点架构，遵循 OCP 倡导的开放解构理念。

技术先进性

百万卡集群规模：两层网络支持百万卡集群规模，单 PoD 可以容纳更大集群（收敛比 15:1，6 万卡，是阿里 HPN7.0 的 4 倍）。

业界最佳网络性能：采用 FARE（全自适应路由以太网）协议，支持多路径包喷洒，带宽利用率可达 95%以上，与业界最佳水平即英伟达以太网方案看齐。

开放解构系统架构：消除厂商锁定风险，AI infra 朝着更加开放方向发展。

第 4 章　云智算物理网络架构：数据中心网络

随着云计算服务从资源即服务向能力即服务加速演进，数据中心网络作为支撑通用计算、存储与平台服务的基础连接架构，正面临前所未有的扩展需求与性能挑战。一方面，超大规模数据中心不断涌现，百万级服务器与多可用区异构资源的统一调度成为常态；另一方面，云原生、微服务等新型应用架构带来网络流量模型深刻变化，对网络的可扩展性、可靠性、低时延与低成本提出更高要求。

为应对这一趋势，中国移动面向云智算新型基础设施，系统性提出了数据中心网络的 SHALL 架构设计理念，即可扩展（Scalability）、高可靠（High Availability）、低延迟（Low latency）与低成本（Low cost），构建具备未来导向的数据中心网络演进目标体系。该架构不仅回应了云智算时代通用算力承载的核心诉求，也为大规模异构计算资源的灵活调度、高效接入与敏捷部署提供坚实网络支撑。

4.1　数据中心网络需求

在云智算持续演进和超大规模云数据中心快速建设的背景下，数据中心网络作为承载海量通用算力的底层连接架构，面临着从容量、可靠性到性能与成本的多重挑战。中国移动聚焦“SHALL”四大核心特性，明确提出数据中心网络的新一代需求体系。

S：可扩展（Scalability）

数据中心网络需具备超大规模水平扩展能力，满足未来百万核级别通用算力节点的统一接入与管理。面对资源池化趋势，数据中心网络应支持多可用区、分区部署，实现大二层或多租户间灵活互联，构建具备超强弹性与横向拓展能力的基础网络架构。多平面可扩展架构成为新型数据中心的重要支撑。

HA：高可靠（High Availability）

传统二层网络难以应对大规模环境下的收敛与稳定性要求，数据中心网络需采用“全三层组网+集中控制”模式，消除二层广播域带来的瓶颈风险，提升网络的稳定性与自愈能力。同时，需具备跨节点、跨区域的多活冗余机制，提升业务承载连续性与服务可用性，满足云上多租户高可用场景下的连接保障需求。

L：低延迟（Low latency）

随着微服务架构和多容器部署的普及，服务间频繁调用带来了更敏感的网络响应要求。数据中心网络需具备端到端低时延能力，确保关键业务流程在 10 微秒量级内完成通信转发，避免延迟瓶颈拖累业务处理效率。同时，网络需具备智能拥塞感知与缓解能力，保障延迟稳定性，降低长尾延迟对任务完成时间的影响。

L：低成本（Low cost）

为应对算力快速增长带来的成本压力，数据中心网络架构需基于白盒交换设备与分布式控制平台构建，支持开放协议与自动化运维，降低设备采购与运维成本。在满足高性能与高可靠基础上，通过扁平化组网结构与资源调度优化实现 TCO 压降，为通用计算场景提供更加经济高效的网络支撑能力。

4.2 数据中心网络架构

面向云智算多样化通算场景，中国移动提出构建具备可扩展（Scalable）、高可靠（High Availability）、低延迟（Low Latency）、低成本（Low Cost） 的 SHALL 架构型数据中心网络，满足大规模算力资源池的承载需求，并具备面向未来演进的灵活性与可持续性，架构示意图如图 6 所示。

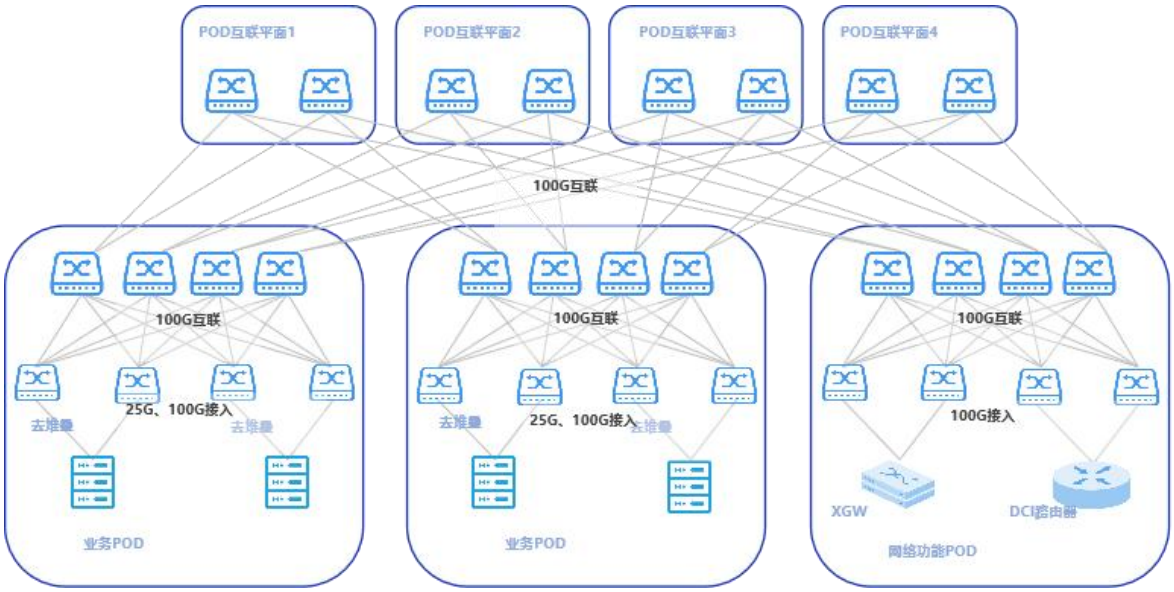

图 6　数据中心网络架构示意图

可扩展（Scalable）：支撑十万级服务器的横向扩展能力

为满足超大规模算力集群的接入与东西向高密通信需求，数据中心网络采用五级 CLOS 架构，具备高度对称性与横向扩展能力。单集群网络规模可支持 10 万台服务器并发连接，通过集群间横向拼接实现跨区域、跨资源池的算力统一承载能力。同时，每个可用区（AZ）支持部署多个独立集群，形成多集群联动、弹性扩容的网络体系，满足未来多区域算力调度需求。

高可靠（High Availability）：全三层网络保障业务连续性

架构设计全面摒弃传统堆叠式与大二层组网方式，全面采用全三层网络架构，实现控制面与转发面的解耦，显著提升系统稳定性与故障隔离能力。通过多平面冗余、多路径保护、业务无损倒换等机制，实现分钟级的网络故障自愈与业务快速恢复，构建具备电信级可用性的云底座网络。

低延迟（Low latency）：确保关键业务的数据转发效率

数据中心内部网络采用单芯片全盒式设备构建，替代传统多芯片框式设备，消除内部多级背板转发延迟。配合全对等多轨网络拓扑设计，显著减少网络层次与中转路径，使得端到端通信时延降低百倍以上，为高性能计算、实时推理、数据库分布式处理等低时延业务提供强有力支撑。

低成本（Low cost）：白盒硬件与集中控制驱动资源效率最大化

从设备选型到系统建设，数据中心网络坚持“简洁高效”的设计理念，采用单芯片全盒式白盒交换机，搭配第三方光模块与通用化布线体系，有效降低初始设备投资（CapEx）。在运维层面，结合统一数据中心网络控制器，实现跨厂商设备的集中控制与自动化运维，提升管理效率，降低运营成本（OpEx）。此外，网络架构实现计算、存储、管理三网合一，简化整体部署方案，进一步优化资源使用效率。

4.3 技术挑战性、创新性和先进性

在中国移动提出的 SHALL 架构体系指导下，数据中心网络不仅面向十万级服务器的超大规模集群场景，同时需应对运维自动化、设备异构、协议标准化等方面的系统性挑战，并在架构与协议层面持续推动创新，构建面向未来的数据中心网络竞争优势。

技术挑战性

超大规模网络自动化难题：超大规模网络环境下，人工运维成本高，难度大，需要引入自动化手段实现网络的自治。

跨厂商设备的自动化管理：由于不同厂商采用各自商业 NOS，导致不同厂商的网络设备需要配置独立的网管系统，在多厂商的网络环境下，网络运维工作量大。

技术创新性

网络架构创新：采用 O/U 全解藕、全盒式设备、全三层组网架构，实现可扩展、高可靠、低延迟和低成本目标（SHALL）。

技术生态创新：采用灰盒+白盒技术路线，实现多厂商设备统一管理，建立良好产业生态。

网络协议创新：主导 IETF 个人草案- 面向 OCS 的数据中心网络自动化（draft-xu-idr-neighbor-autodiscovery，draft-acee-idr-lldp-peer-discovery），助力网络自动化。

技术先进性

集群规模：支持十万服务器集群规模，与业界最佳水平看齐，为百万卡 GPU 集群建设坚定基础。

网络自动化：采用移动云自主创新的 BGP 邻居自动发现机制，实现 BGP 配置的自动化，并降低交换机之间的 BGP 会话数量，极大提升网络收敛性能。采用统一 NOS 适配不同厂商的硬件设备，实现统一 DCN 控制器跨不同厂商设备的自动化管控。

第 5 章　云智算物理网络架构：广域网络

随着企业出海步伐加快和 AI 原生场景加速落地，云智算网络的边界正由数据中心内部不断向全球延展。跨区域、跨境的业务需求对网络的性能确定性、全球可达性和资源调度能力提出了全新挑战。广域网络作为云智算向全球化发展的关键承载底座，正面临从传统“尽力而为”模式向“可预期、可编排、可保障”新型架构转型的迫切需求。

中国移动云智算基于广域 IP 可预期网络与广域光网络，构建全球一体化的可预期广域网络服务。在 IP 网络层面，通过 DCI-TE、EPE-TE 和 SGA 等关键技术，面向出海企业和 AI 服务场景，提供 SLA 驱动的路径调度、边界路径智能选控以及跨境接入加速能力；在光网络层面，依托 800G/1.6T 等高阶传输技术和 OpenConfig 解耦架构，打造开放、高弹性、易演进的光承载平台，满足大规模智算跨 DC 部署和全球业务高效互联需求。

本章将围绕上述两大核心能力展开，重点介绍广域网络架构的需求分析、架构构成、能力优势与创新亮点，全面展现云智算广域网络基础设施架构。

5.1　广域 IP 网络

可预期网络是面向云智算多场景互联接入的网络架构，旨在解决当前网络在多业务、多租户环境下面临的可预期性不足、服务保障能力有限的问题。该架构通过引入可预测、可度量、可保障的技术体系，为多类型业务提供差异化、确定性和端到端的网络服务能力。

5.1.1 可预期网络需求

随着全球化、AI 化和多云架构的普及，云智算场景对网络提出了更高要求。传统的“尽力而为”网络已无法满足多样化、复杂化的业务需求，必须引入可预期网络架构，为不同场景提供性能优先、成本优先的差异化服务。主要需求包括以下四个方面。

性能优先

在企业出海、全球扩张背景下，低延迟、高可靠的网络连接成为核心需求。出海企业的延迟敏感型业务，如跨境视频会议、云游戏、实时交互直播等，需要通过优化的跨境网络服务提供低延迟、低抖动、低丢包的连接体验。例如，视频会议要求毫秒级延迟、稳定带宽和快速恢复能力，而出海游戏则需基于覆盖广泛的边缘节点，保障玩家的流畅体验。这类场景的网络需求突出强调性能优先，需要具备高质量、可预测的传输能力。

成本优先

对于延迟不敏感的业务，如海量数据备份、系统日志归档、非实时文件同步等，网络成本是首要考虑因素。这类业务对带宽消耗大，但对时延和抖动要求相对宽松。可预期网络需提供低成本链路、空闲链路绕行等能力，将高质量出口资源保留给高价值、时延敏感业务，从而实现“好钢用在刀刃上”的资源优化策略。此外，互联网带宽资源的优化调度也是降本增效的重要组成部分。

全局负载均衡

随着全球范围的计算资源分布愈加广泛，企业亟需实现跨地域的全局负载均衡调度。通过就近接入云服务商的全球广域网，结合跨域流量工程能力，企业可以动态感知不同区域的计算和网络资源状态，智能分配请求到最优的节点或数据中心。这不仅能提高服务响应速度，还能在突发流量场景下有效分摊负载，保障用户体验的稳定性和一致性。

全球覆盖

在企业全球化运营中，跨境、跨区域网络连接的广覆盖能力不可或缺。可预期网络需为企业出海、全球拓展提供全球一体化的网络接入与传输服务，确保业务在全球范围内的顺畅运行。尤其是对于出海游戏、跨境直播等对体验要

求极高的业务场景，必须通过部署覆盖广泛的边缘节点，结合智能路由和加速策略，显著优化跨境访问质量和用户体验。

5.1.2 DCI-TE：跨境数据中心互联场景下的可预期网络能力

在企业出海、数据全球部署与跨国业务协同日益深化的背景下，跨境数据中心互联（DCI）成为云智算网络架构中的核心能力之一。尤其在游戏出海、跨境电商、视频会议、跨境直播等场景中，用户对跨境数据传输的时延、稳定性和可预测性提出了远高于传统网络的要求。然而，受制于多运营商接入、不统一的传输协议、复杂的互联链路结构等问题，传统国际传输网络普遍存在路径不稳定、链路易拥塞、跨域路由不可控等问题。为破解上述难题，中国移动依托算网一体化技术体系，构建了面向出海业务的跨境 DCI-TE 技术能力，实现端到端、面向业务意图的高性能路径调度与智能流量工程。

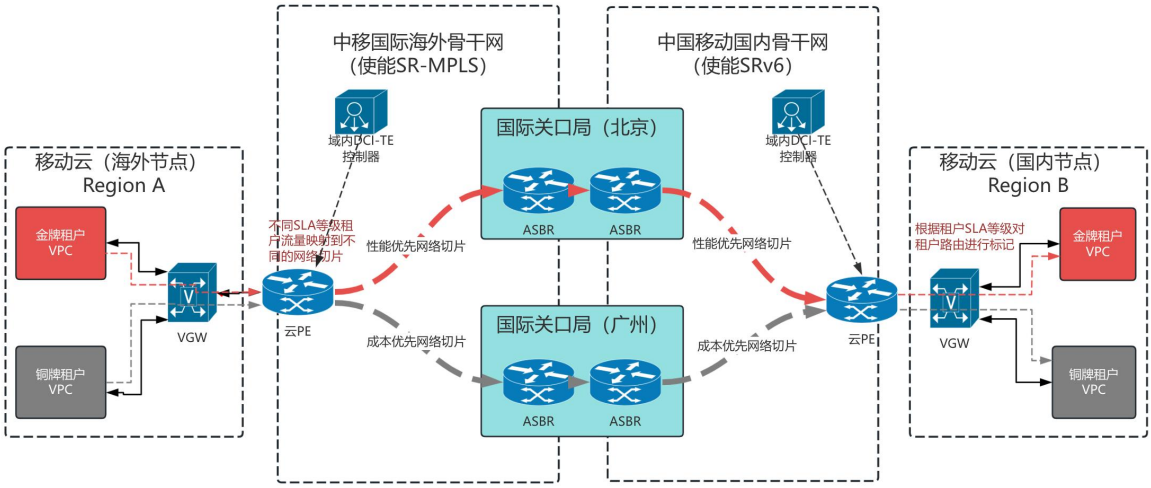

图 7　跨境 DCI-TE 架构示意图

跨境 DCI-TE 主要服务于企业级出海客户，尤其是对传输性能敏感的业务，如实时音视频、云游戏、AI 推理分发等。此类业务往往具有以下特征：流量波动大、交互频次高、容错空间小。传统公网或 VPN 方式在应对突发网络劣化时往往难以保证业务连续性。因此，DCI-TE 设计的目标，是通过算网协同、路径调度与切片编排，实现“业务驱动的网络能力保障”，在广域骨干网络中构建具备高性能、高可用、可预期特性的跨境承载路径。DCI-TE 架构图如图 7 所示，包含以下关键技术：

独立域内 SR-TE 架构：简化部署，独立演进

为适应跨境网络环境下多运营商、多技术体系并存的现实条件，跨境 DCI-TE 架构采用“自治域内独立演进”的设计理念，在国内与境外分别构建相对独立、自治可控的 SR-TE 能力体系，形成“双段骨干、分层调度”的跨境互联技术结构。

如图 7 所示，在骨干网构成方面，国内部分由中国移动境内骨干网络承载，负责从用户侧数据中心或云 VPC 出发，经省际核心网至国际出入口节点的高速承载任务。该网络采用 SRv6 协议体系，具备原生 IPv6 编址能力、路径指令可编排能力与服务链扩展能力，为路径调度和服务质量保障提供灵活基础。境外部分则由中移国际统一承载，涵盖从境外 PoP 节点至海外公有云、边缘节点、本地运营商等目标区域的网络段。该网络根据不同地区实际部署情况，采用 SR-MPLS、传统 MPLS 或混合技术协议，重点保障跨境业务的广域到达率、节点可控性和服务连续性。

在控制架构方面，两个网络自治域分别部署域内 SR 控制器，独立完成拓扑发现、路径计算、QoS 策略应用与服务链配置。控制面与转发面解耦设计提升了架构灵活性，既支持 SR 域内路径的可视、可调、可编排，又能适配各自的运维体系与演进节奏，降低跨域部署耦合度。

这一架构具备高度的工程可落地性与阶段性演进能力，既满足现阶段中国移动“境内-境外”分段部署策略，也为未来统一调度、全球扩展提供平滑升级路径。通过独立域内 SR-TE 架构，跨境 DCI-TE 能够在不打通全局控制平面的前提下，实现各域内的稳定可控演进和业务调度能力保障。

性能感知 BGP 路由(PAR)：实现跨域低延迟选路网络增值服务能力

在跨境网络环境中，路径状态受物理距离、运营商接入差异、国际互联链路质量等因素影响，业务性能易受波动干扰，传统路由机制难以提供稳定、可预期的连接保障。为解决这一问题，DCI-TE 引入了具备性能感知能力的 BGP 路由机制，在国内与境外两个自治域之间实现可调度、可控制的路径拼接与流量引导。

系统可对自治域之间的多条可用路径进行持续状态监测，包括链路可用带宽、时延、丢包率、抖动变化等指标，并以此为基础构建跨域路径性能评估体

系。与传统静态选路机制不同，DCI-TE 将业务侧 SLA 约束作为路径调度的主导依据，明确业务对网络的性能诉求，建立起“业务需求—路径能力”之间的动态映射关系。

在路径拼接方面，系统根据目标 SLA 选取满足要求的多个路径段进行组合，构建端到端的高性能连接通道。例如，延迟敏感型业务可优先拼接低延迟路径段，确保整体传输满足毫秒级时延目标。路径拼接过程完全由控制器驱动完成，具备高自动化、无人工干预的部署特性。

在流量引导方面，DCI-TE 结合业务入网时携带的服务等级信息，自动识别其 SLA 目标，动态匹配对应路径并实施引流策略。该机制支持分级引流控制：对于关键业务可引导至高保障路径，并启用路径容灾保护机制；对于可容忍业务则可引流至成本较优但性能适配的路径，实现整体资源效率最大化。

这一机制不依赖集中式 TE 控制器，而是基于 BGP 协议扩展的分布式选路能力，有效提升了系统部署灵活性和系统的健壮性。通过基于 SLA 驱动的路径拼接和引流策略，DCI-TE 实现了从“可达性导向”向“性能保障导向”的路径调度演进，是出海业务获得稳定、高质量连接体验的关键保障能力。

5.1.3 EPE-TE：BGP 出口智能选路能力

在企业出海部署过程中，出口路由策略直接影响用户访问的连通性与体验质量。尤其在游戏、音视频等实时性业务场景中，公网跨境路径存在显著性能差异，稳定性难以保障。中国移动面向出海企业构建的 EPE-TE（Egress Peer Engineering – Traffic Engineering）能力体系，基于既有的多线 BGP 资源优势，结合标准化架构与智能策略调度机制，为企业提供差异化、可编排的 BGP 出口控制能力，满足高性能业务的出境保障与全局资源的效率调度需求。EPE-TE 架构示意图如图 8 所示。

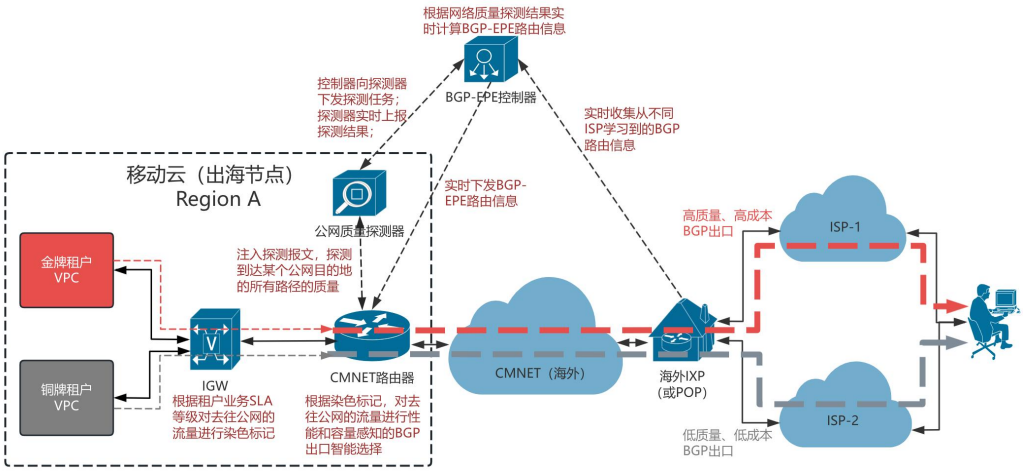

图 8　EPE-TE 架构示意图

多线 BGP 组网：出口调度能力的基础前提

在海外区域，特别是东南亚、中东、拉美等出海重点区域，当地 ISP 众多，网络质量参差不齐，不同链路的稳定性、时延、带宽能力差异显著。为适应这种复杂环境，中国移动已在多个海外节点建设了具备多线 BGP 能力的接入体系，接入多个主流本地运营商线路，形成丰富的跨境链路资源池。这一多线 BGP 组网能力，为出海企业提供了跨运营商、多路径、多质量等级的可选出口，是构建智能调度能力的前提条件。

多线 BGP 能力本身并不直接构成调度机制，但为 EPE-TE 的差异化出口策略提供了必要的基础资源保障。在此基础上，才能实现基于业务属性的路径匹配与资源优化配置。

BGP-EPE 架构：标准化、可扩展的流量工程方案

EPE-TE 基于 BGP-EPE 架构，构建了相对 PBR（策略路由）更为稳定、简洁和可控的 BGP 出口流量工程能力。该架构利用边界路由器对等会话的独立标识能力，在控制平面实现对每条出境路径的可视化和策略控制，从而实现业务流量的灵活调度。

与 PBR 方式相比，BGP-EPE 避免了复杂 ACL 规则和手动策略维护，提升了系统的稳定性与运维效率。同时，EPE-TE 具备良好的可扩展性，可覆盖多租户、跨区域、不同业务类型的应用需求，适配企业出海在不同阶段对网络策略控制的差异化诉求。

性能与容量感知的智能选路：实现服务保障与资源优化协同

在调度策略层面，EPE-TE 引入性能感知与容量感知并行驱动的智能 BGP 出口选路机制。系统通过实时监测各 BGP 出口链路时延、丢包率、可用带宽等核心指标，结合各类业务的 SLA 需求，为不同租户、不同业务动态匹配最优路径。

例如，对于时延敏感、体验要求高的游戏、直播类业务，系统可优先引导至性能优的 BGP 出口；而对于成本敏感、带宽占用大但对实时性要求较低的业务，如数据归档、内容同步等，则可匹配至价格更优、容量富余的链路。这种策略使网络服务能够在保障体验的同时，实现链路资源的动态平衡与 BGP 带宽的精细化运营，有效提升网络总体利用效率，降低出口带宽使用成本。

5.1.4 SGA：云网一体的跨境超级全球加速

为提升出海用户访问境内算力资源的体验质量，中国移动依托中移国际在海外广泛部署的骨干网络，有效屏蔽了境外 ISP 路径不稳定所带来的影响。在此基础上，构建了超级全球加速（SGA, Super Global Acceleration）技术体系，旨在实现全球范围内端到端的跨境算力接入优化。SGA 通过路径选择与源站调度的协同机制，全面提升跨境算力访问的服务质量，助力用户获得更低延迟、更高可靠性的一体化全球接入能力。

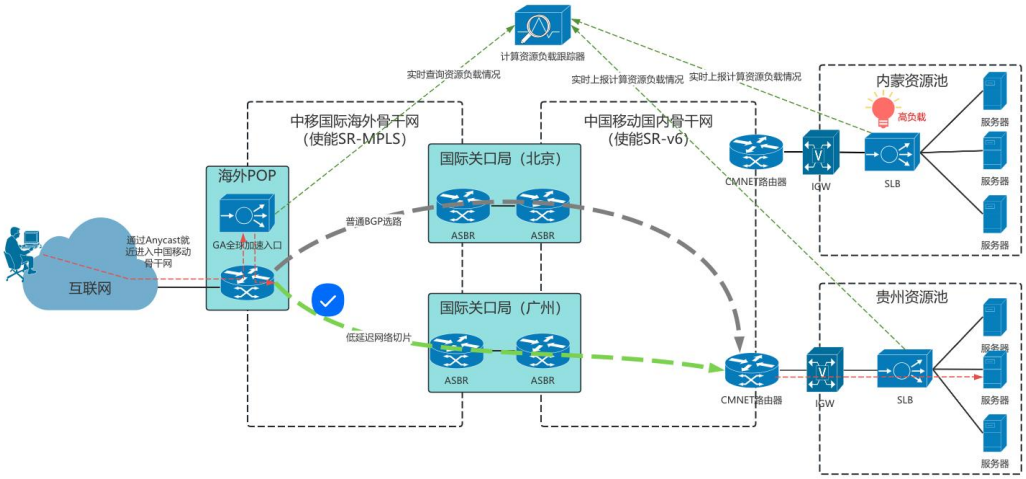

图 9　SGA 架构示意图

超级全球加速（SGA）的核心理念是，借助中移国际在海外部署的骨干网络，实现与境内算力资源的高效互联，并通过路径优化与源站选择的协同调度，为

跨境算力访问提供端到端的低时延、高可靠性保障，SGA 架构示意图如图 9 所示。SGA 创新性地将路径优化与源站调度相结合，通过联合感知网络性能与算力资源状态，实现面向业务目标的源站+路径协同决策。相较于传统加速方案路径与算力分离决策的模式，SGA 避免了优质源站搭配低性能路径的情况，构建具备端到端可预期能力的全球加速服务网络。SGA 技术体系的构建依赖于多项核心能力的协同配合：

全球化任播服务

SGA 在全球范围内部署了 Anycast 接入节点，用户流量可快速接入距离最近的 GA 加速节点，从而提升初始接入效率。在此基础上，GA 节点通过与中移国际及中国移动国内广域网的骨干网络协同，构建了覆盖全球的 POP 节点体系。

跨境低延迟选路能力

基于 BGP 性能感知路由（PAR）能力，为全球加速流量提供跨境的低延迟网络连接能力，进一步实现跨境骨干网范围内的网络加速。

源站负载实时跟踪与全局负载均衡

SGA 在 GA 节点部署了集中式计算资源负载跟踪器，可对接入的全部源站进行实时监测。系统能够获取各源站的 CPU、GPU 利用率、内存占用及任务响应能力等运行指标，并与算力资源池深度整合，构建全局可视的资源状态图谱。当系统检测到某一源站出现负载过高或资源瓶颈时，将触发动态负载均衡机制，结合用户任务的紧急程度与切换代价，引导部分请求切换至其他低负载源站，从而规避热点资源瓶颈、提升整体算力利用效率和用户访问体验。

协同源站选择与路径规划

SGA 通过协同优化路径选择与源站调度，实现跨境接入服务质量的整体提升。系统在 GA 节点侧同时感知各条跨境路径的状态信息和源站资源的负载情况，在此基础上对“路径+源站”的组合进行综合评估与最优决策。GA 节点对所有候选路径和源站组合，基于路径时延+源站负载综合权重排序，最终选择权重值最大的组合进行调度。这一协同机制兼顾了网络质量与算力负载两方面的约束，确保每一次接入都基于全局最优视角作出判断，提升系统整体效率的同时增强了鲁棒性与动态适应能力。

5.1.5 技术挑战性、创新性和先进性

技术挑战性

广域网流量工程：通常要求不同域采用相同的 TE 隧道技术（比如 MPLS-SR 或 SRv6），且不同域的 TE 控制器需要协同并实现跨域隧道路径的集中计算和隧道转发层面的拼接，技术方案复杂，可扩展性差。

BGP 出口流量工程：不仅需要考虑不同租户流量的需求，同时需要考虑多 BGP 出口的带宽容量、成本和性能因素，还要考虑出口故障的快速检测和回退机制，技术架构相当复杂。

技术创新性

广域网流量工程：通过 BGP 性能路由实现多段域内独立的 TE 隧道的自动化拼接，不同域内 TE 隧道技术方案独立演进，极大降低跨域 TE 方案部署的技术门槛以及后续运营的复杂性。

标准协议创新：主导的 IETF 工作组草案-BGP 性能感知路由(draft-ietf-idr-performance-routing)、服务功能自动发现（draft-xu-dnssd-sf-discovery）以及参与的 IETF 国际标准-BGP-EPE 网络故障快速检测机制(RFC9703)。

技术先进性

广域网流量工程：相对谷歌的 B4（基于 Openflow 下发 PBR+GRE 隧道拼接），DCI-TE 采用无状态的 SR，技术方案的可扩展性和稳定性更好，相对谷歌的 B2 方案，引入基于标准的 BGP-EPE 技术方案，技术方案极简开放，系统稳定性和可运维性极大提升。

超级全球加速：相对 AWS GA 方案，除了实现跨不同地域的资源池的计算资源的全局负载均衡，同时结合跨域低延迟选路能力，为跨境的 GA 提供云网一体感知的超级全球加速体验。

5.2 广域光网络

随着人工智能训练集群从单数据中心向多数据中心演进，跨区域分布式训练成为新常态，广域光网络作为智算互联的核心承载平台，亟需实现大带宽、低时延、开放解耦和成本优化等关键能力。中国移动聚焦云智算广域网络能力建设，提出基于开放架构的下一代光传输网络体系，依托单波 800G/1.6T 传输技术、光电解耦能力与 OpenConfig 控制接口，打造具备超高容量、可编程、智能化的广域光承载网络。

5.2.1 光网络发展趋势

支撑跨区域分布式训练集群互联

随着 AI 大模型训练从单一数据中心向多数据中心分布式架构迁移，集群间大容量数据同步对光网络提出更高要求。例如谷歌已基于跨 DC 架构完成 Gemini Ultra 大模型训练，产业界对大容量 WDM（波分复用）技术的需求迅速提升。

超高速 WDM 加速商用部署

800G 及以上传输技术正在逐步完成从试验向商用的过渡。中国移动等运营商已在长距离传输与 AI 智算场景中开展 80×800G WDM 网络验证，有望在未来形成普遍部署能力，支撑区域间高密互联。

开放控制标准日益成熟

以 OpenConfig 为代表的开放标准接口从传统 IP 网络逐步渗透至光网络领域，实现控制面和管理面的统一编排。国际云服务商（如谷歌、阿里、腾讯）已在大规模网络中完成部署，运营商也逐步试点推进。

光电解耦与自研光模块加速落地

光层与电层设备之间的解耦正在成为主流趋势，厂商锁定风险显著降低。互联网企业推动自研光模块能力发展成熟，光模块成本占比逐步下降。

5.2.2 广域光网络技术架构

广域光网络核心技术路线为“高带宽、低成本、全开放”，构建以 800G/1.6T 为基础的开放光传输架构，提升网络弹性与长期演进能力，技术架构下图所示。

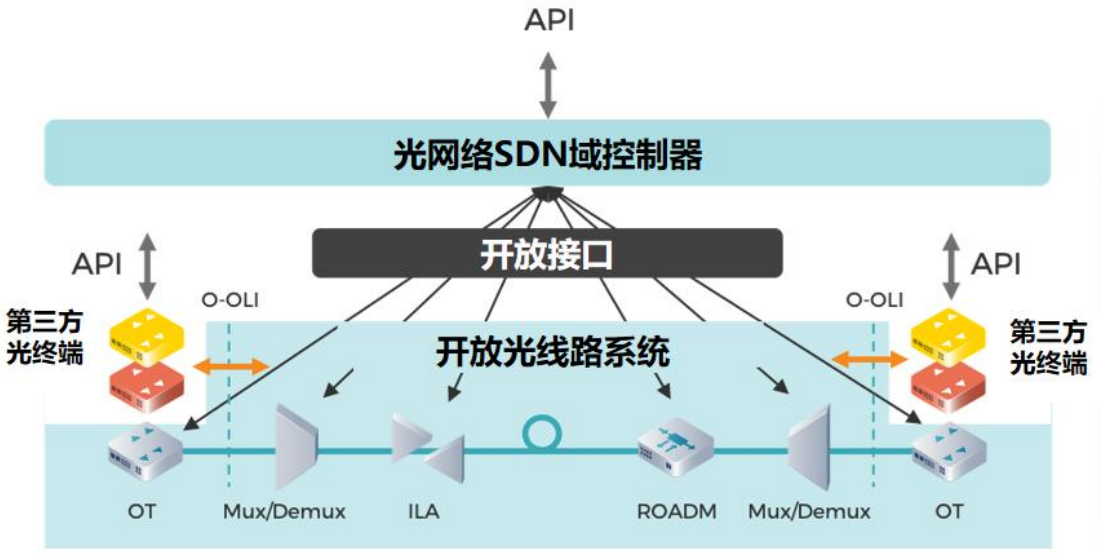

图 10　广域光网络架构示意图

（*参考 TIP/OOPT 工作组 ”MUST Optical SDN Controller NBI Technical Requirements Document”*）

高带宽传输能力

网络引入单波 800G 乃至 1.6T 光传输能力，实现“80×800G”长距离稳定传输，满足跨区域训练集群大规模数据交互对带宽容量的极致需求。

开放解耦体系结构

基于 OpenConfig 控制器实现光网络软硬件解耦，进一步实现光层与电层解耦。多厂商系统可统一管理，有效解决专有协议、私有配置带来的运维难题，全面提升网络可控性与演进灵活性。

成本优化设计

通过引入自研或定制化光模块，显著降低单位比特传输成本。以标准化光接口为基础，结合产业链成熟的器件供应体系，构建低成本、高兼容性的网络系统，提升光网络大规模部署的经济性。

5.2.3 技术挑战性、创新性与先进性

技术挑战性

多厂商兼容挑战：多厂商设备存在大量私有 YANG 模型，对集中控制器管纳多厂商设备存在一定挑战，此外，光层电层跨厂商的互通性验证也需要进过广泛验证。

800G 长距传输：中短距 800G 已经基本成熟，长距超 800G 的技术仍处在前瞻研究阶段，在噪声高容忍调制技术、低损耗光纤选择、频谱波段扩展等方面仍有技术挑战。

技术创新性

Ai+光网络：AI+赋能光网络，实现智能化运维、光功率进行动态调优。

光电联动：基于 OXC 的光电联动新型全光网。光电联动重在光层和电层业务之间的互通，利用电交叉矩阵完成小颗粒业务汇聚和调度，利用光交叉完成波长级业务调度，拉通光电 OAM 机制，实现光电组网联动以支持大规模组网和灵活调度。

技术先进性

超大容量：中国移动基于空芯光纤首次完成了 160 波 ×800G 传输系统技术试验，单芯光纤实现 128Tb/s 超大容量传输。

开放解构：基于控制器与设备解耦，光电设备解耦，实现完全开放的光网络。极大降低光网络的建设和运维成本。

第 6 章　云智算虚拟网络架构

云智算网络的演进不仅需在物理层面构建高性能的数据中心与广域承载网络，更需在虚拟层面提供灵活、可编排的网络能力。虚拟网络作为连接租户资源、承载多云互联、支撑安全防护的关键基础，是云智算网络架构不可或缺的组成部分。

中国移动云围绕云内网络（SDN）、云间网络（云联网）与安全服务链三大方向系统构建和持续优化虚拟网络产品能力体系，全面提升资源池内部的网络效率、云间连通能力与云上业务的安全保障水平。通过自研 SDN 控制器与智能化调度平台，支持万级服务器规模集群的高性能网络管理；基于 NaaS 架构和标准化协议，构建跨 Region、跨云的多租户互联能力；引入安全服务链机制，突破传统网络防护在可扩展性和可编程性方面的限制。

本章将分别从 SDN 网络、云联网和安全服务链三个维度展开，介绍云智算虚拟网络架构中的核心技术与创新突破，助力云智算打造更加敏捷、安全、可持续演进的虚拟网络底座。

6.1　云内网络：SDN

云内网络作为承载虚拟计算、存储和多租户服务的关键基础，正面临规模高速增长、资源调度复杂、服务敏捷性不足等新挑战。传统网络架构已难以满足超大规模、多租户、高弹性云服务的需求，亟需构建具备集中控制、精细编排、智能感知与灵活服务能力的全新网络体系。中国移动提出 SDN 架构，致力

于打造统一、开放、智能的云内网络基础设施。

6.1.1 SDN 网络需求

随着云服务规模不断扩大、租户网络复杂度不断提升，云内网络正面临集群资源规模剧增、转发性能瓶颈、成本控制压力以及网络可用性保障等多方面挑战。传统网络方案难以支撑高弹性、强隔离、敏捷部署方面的关键需求，迫切需要引入新一代 SDN 架构，以实现网络资源的集中控制、自动编排与智能调度。云智算 SDN 网络的演进需求主要体现在以下四个方面：

可扩展性

集群规模扩展能力：为支撑超大规模云智算资源池，网络需具备单 Region 万级服务器集群接入能力。网络架构需支持水平扩展与多可用区集群纳管，满足未来异构算力资源池统一接入的需求。

租户规模扩展能力：网络需支持百万级虚拟私有云（VPC）实例，满足大型政企客户在一云多租、一租多 VPC、多 VPC 灵活互通等多样化部署场景中的资源隔离与弹性管理需求。

高性能

软硬一体能力融合：传统 NFV 方案依赖 x86 通用服务器运行虚拟网络功能，具备灵活性，但转发性能受限。面对大带宽、高并发业务场景，现有纯软件转发面临瓶颈，需引入软硬一体技术，实现高性能虚拟网络。

低成本

虚拟网络资源利用效率：在传统架构下，NFV 方案需要消耗大量 CPU 资源用于实现基础网络功能，带来较高运营成本。为降低资源损耗与能耗水平，需构建具备精细调度能力、动态按需供给机制的网络基础架构，实现网络能力与计算能力的解耦和优化分配，提升整体 TCO（总拥有成本）效率。

高可靠

网络状态可视化能力：在多租户并发运行场景下，需实现租户虚拟网络资源的实时可视与运行状态可观测，便于网络异常预警、故障快速定位与网络运

维决策支持。

网络异常自动化闭环处理：网络需具备对租户故障的快速感知、精确定位、自动隔离与动态恢复能力，实现“分钟级”或“秒级”故障处理，确保关键业务链路不中断、服务连续性可保障。

6.1.2 SDN 技术架构

面向超大规模云服务集群的资源管理与网络调度需求，中国移动构建了具备“开放解耦、软硬一体”特征的新一代 SDN 技术架构，架构如图 11 所示。

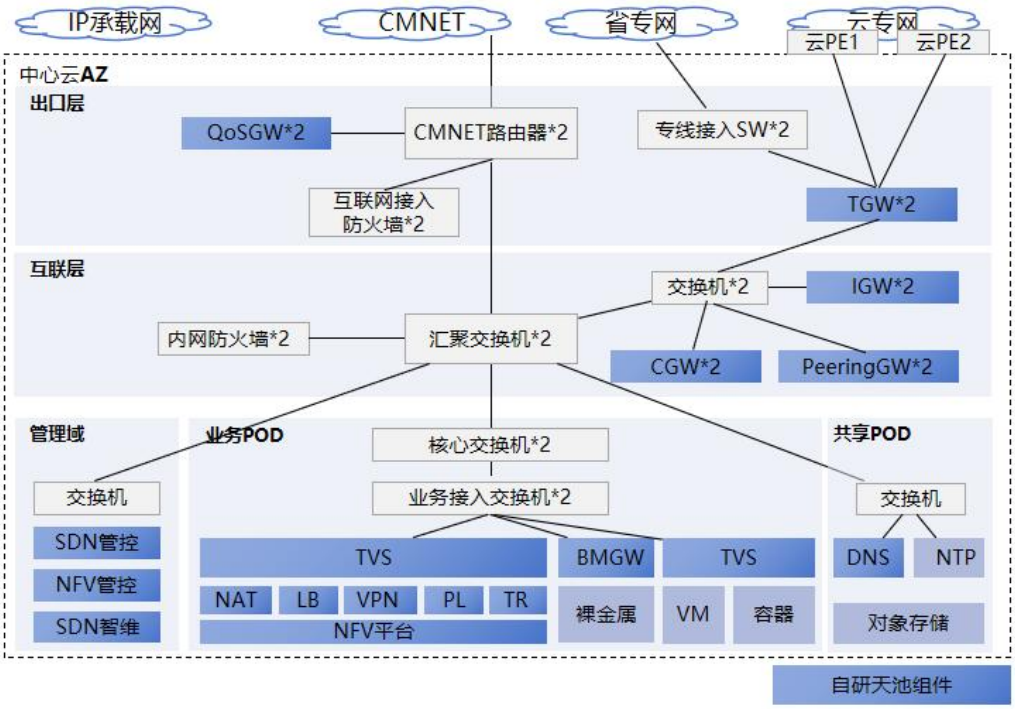

图 11　SDN 技术架构示意图

超大规模

SDN 网络采用“管控分离、分域控制”的架构设计，控制平面支持基于租户粒度、VPC 粒度、计算节点粒度灵活划分，实现资源视图分级解耦与独立调度。系统可支持单资源池纳管超 2 万节点，满足百万级 VPC 租户资源部署与弹性管理需求，具备面向超大规模算力集群部署能力。

超高性能

通过引入 SONiC 与 P4 为代表的开放可编程网络体系，实现关键节点交换功能向下卸载。支持亚微秒级转发时延、十亿级 PPS 包转发能力，可满足运营商级公网、专线、云互联等高密度、高速接入场景下的服务需求。同时，支持精细化 QoS 控制与差异化服务编排，保障多租户高性能业务运行。

极致弹性

基于统一资源建模体系与中国移动自研云控制平台，构建标准化、弹性扩展的网络服务能力。NFV 平台支持 NAT、负载均衡、VPN 等网络功能的秒级开通与五分钟内部署，具备“灵活调度、极简交付、统一接入”的编排能力，为公有云、私有云及三方云资源融合提供统一支撑。

稳定可靠

SDN 系统具备统一视图观测与多层网络闭环保护能力。通过与底层网络及 Overlay 虚拟网络双向联动，结合海量网络数据采集与告警策略，实现端到端链路状态监控、Overlay 隧道路径探测与租户级异常预警。系统支持秒级感知、分钟级联动、租户级服务保障，构建面向多租户的高可用网络体系。

6.1.3 技术挑战性、创新性与先进性

技术挑战性

高性能 SDN 网关技术复杂：当前主流高性能 SDN 网关多基于可编程芯片（如 Tofino）或 FPGA 构建，具备高转发能力和灵活功能编排能力，但产品实现门槛高、定制化程度强，存在一定的技术壁垒。

自动化运维难度大：在超大规模租户网络部署环境中，面向租户的虚拟网络状态动态变化频繁，链路状态与健康度可感知能力薄弱，传统依赖人工运维方式难以实现快速故障定位和高效恢复。网络异常排障效率低，租户体验受损，成为影响云服务稳定性的重要因素。

技术创新性

开放解耦高性能网关架构：通过构建开放解耦的高性能网关架构，分别采用不同技术路线满足灵活性与性能需求：有状态网关引入 DPU 进行数据平面加

速，适用于需深度包处理的 NAT、负载均衡等应用场景；无状态网关基于可编程网络芯片，实现转发逻辑灵活编排与性能极致优化。

网络可视化与智能化运维能力：结合虚拟化网络与底层物理网络的联动信息，构建覆盖租户网络、网络功能实例、Overlay 隧道、物理链路等多维的可视化监控体系。通过引入智能诊断引擎，实现故障原因自动识别、定位与隔离，提升网络稳定性与租户保障能力，为网络“自感知、自决策、自修复”演进提供基础支撑。

6.2 云间网络：云联网

在云智算和算网一体发展的新阶段，企业客户对于多云、混合云环境的网络互联需求愈发迫切。传统网络方案存在配置复杂、扩展困难、缺乏灵活性等问题，无法满足智能时代对敏捷、可靠、弹性的云网一体化需求。中国移动云智算依托自研的新一代云联网架构，以“一点接入、全域可达”为核心目标，面向多云、混合云场景提供高可用、高扩展、低复杂度的云网互联服务，为企业数字化升级和智算需求提供坚实的网络底座。

6.2.1 云联网需求

混合云互联

针对企业本地数据中心（IDC）、总部与分支机构与云端 VPC 之间的互联需求，云联网能够打通云上云下的网络链路，实现本地与云端业务、数据的统一调度、弹性扩展与协同计算，构建灵活可扩展的混合云架构。

跨云互联

企业越来越多采用多云策略以规避单一云厂商风险，云联网提供移动云与其他云之间的高速、稳定、安全的互通。

安全隔离与共享

针对大型企业多部门、多业务线并行运作的复杂需求，云联网支持单租户下多 VPC 组网，实现 VPC 间的安全隔离。与此同时，通过共享 VPC 配置，可实

现部分公共资源（如网关、安全组）的跨 VPC 复用，在保证隔离的同时提升资源使用效率。

易用性与服务化

传统云联网架构基于 Full-mesh 组网，需手动搭建隧道、配置路由，运维复杂。NaaS（Network as a Service）服务化模式，用户仅需选择互联区域和 VPC 实例，即可自动完成连接，降低网络使用门槛，减少部署和运维成本。

可视化与自动化运维

随着组网规模扩大，运维复杂度随之提升。云联网提供端到端的网络拓扑可视化、流量监控、链路健康检测、告警与自动切换等能力，确保网络稳定性，降低故障响应时间。

6.2.2 云联网架构

通过集中管控与分布式调度结合的架构，云联网为企业提供“一点接入、全域可达”的高性能网络连接能力，简化网络配置，提升跨 VPC、跨区域组网灵活性，实现从网络基础设施到服务层的整体升级。云联网架构如图 12 所示。

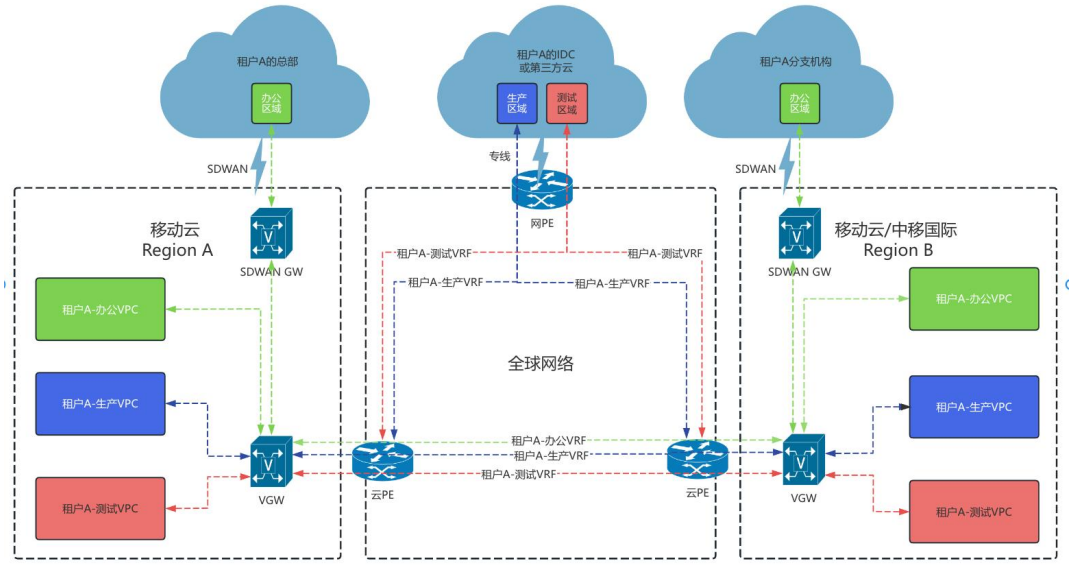

图 12　云联网架构示意图

单租户多 VPC 组网：满足隔离与扩展需求

支持单租户下划分多个 VPC 实例，适用于多部门、多业务线的企业架构。各 VPC 既可实现相互隔离，也可通过灵活配置实现按需互通，便于不同业务单元实现独立部署或资源协作。通过这一能力，企业能够更好地实现精细化资源管理、灵活的业务隔离策略，并减少整体运维复杂度。

NaaS 服务化模式：降低运维门槛、提升配置效率

采用 NaaS（Network as a Service）模式，将传统网络组网、隧道搭建、路由配置等工作全面云化、自动化。用户只需通过控制台或 API 选择互联区域与 VPC，系统即可自动完成隧道部署与路由优化，无需关注底层网络细节。这一模式大幅降低网络运维门槛，尤其适合快速扩张、敏捷部署的中大型企业。

标准协议支撑：实现开放互联与兼容扩展

基于 BGP 等标准协议，实现 Underlay 与 Overlay 协同，有效支持有效支持多云及混合云互联。通过标准协议设计，不仅降低了多云部署复杂性，还增强了网络系统的扩展性、开放性与异构兼容性。此外，业界主导的 IPsec 隧道负载均衡、BGP 多实例等创新方案，为云联网大规模应用提供了强有力的技术支撑。

6.2.3 架构对比：云联网架构 VS TR 架构

在多云、混合云场景日益普及的背景下，不同架构设计的优劣差距逐渐拉大。传统 TR（Transit Router）方案因依赖 Full-mesh 手动隧道搭建与静态路由配置，在面对节点数量增长时，复杂度、成本和风险迅速放大，已无法满足当前多租户、大规模、多区域的企业需求。

相比之下，NaaS（Network as a Service）架构通过自动路由、动态调度、集中控制的设计，显著降低了运维门槛，具备出色的扩展性、灵活性与智能化能力，成为当前及未来云网互联发展的主流趋势。

表 4　TR 架构与云联网架构对比表

对比维度	TR 架构	云联网架构
组网复杂度	高，需要手动配置隧道与路由	低，自动完成拓扑和路由配置

扩展能力	节点数增加导致隧道数量指数增长	节点扩展线性增长，系统自动完成路由更新
运维成本	高，需专人维护、排查	低，提供自动化运维工具、告警与自愈能力
兼容性	异构环境支持弱，对接复杂	开放架构，兼容自研与第三方 NFV、PE 设备
安全防护	基础安全，防护粒度粗	动态防护、细粒度微隔离、零信任机制

云联网聚焦国内多云、混合云场景，充分结合算网一体化需求，通过标准创新、架构优化和国产化能力建设，打造了具备国际竞争力的自主可控云网一体化解决方案。

6.2.4 技术挑战性、创新性与先进性

技术挑战性

多云互联的挑战：大多数云商缺少全球覆盖的网络，导致多云互联需要云商、全球化网络运营商多方对接，网络故障定位定界难度大。此外基于 TR 的 DIY 模式多云互联方案，使用门槛较高。

Overlay 与 Underlay 协同难题：现有 Overlay 控制器大多采用集中式 SDN 方案，Underlay 也多采用完全集中式流量工程控制器，二者之间缺乏标准化协同机制，网络故障定位与流量调度存在瓶颈，阻碍差异化网络服务的按需构建与故障联动响应。

技术创新性

网络服务化模式（NaaS）：基于 NaaS（Network as a Service）模式的云联网产品，避免租户的 TR 间 Full-mesh 的复杂隧道配置与路由维护工作负担。

标准协议接口：基于标准 BGP 协议的扩展能力，构建 Overlay 与 Underlay 自动协同控制通道，实现端到端路径资源的动态调度、跨云服务编排与多维可

视，支持更丰富的服务编排接口与跨平台服务一致性保障，提升网络智能调度能力与对多云环境的适配性。

技术先进性

NaaS 架构：中国移动构建的云联网 NaaS 方案在易用性与横向扩展能力上超越传统 TR 架构，具备统一纳管、跨云直连与分区隔离等能力，可对标 AWS Cloud WAN 等全球化连接方案，满足多云、跨区域高弹性资源调度的需求。

差异化网络服务：在 AWS 的 Cloud WAN 的基础之上，进一步实现 Overlay 和 Underlay 的智能选路能力的协同，为云互联的用户提供差异化的网络连接服务，比如低延迟广域网服务或低成本广域网服务。

6.3 内生安全：网络安全服务链

随着多云、混合云和算网一体化架构的发展，企业和机构的业务系统面临前所未有的安全挑战，包括跨租户攻击、DDoS 攻击、数据泄漏和勒索软件扩散等。这些复杂的威胁不仅威胁单一系统，还可能在网络中横向扩散，影响整个平台的稳定性和信任度。

在此背景下，安全服务链成为云联网架构重要组成部分。通过模块化、分布式、灵活编排的安全能力，安全服务链为多租户、多场景、多层次的网络环境提供端到端、动态化的纵深安全防护，提升云网一体化环境安全水平与韧性。

6.3.1 网络安全服务链需求

在多云、混合云和算网一体化架构日益普及的背景下，企业和机构的业务系统正面临前所未有的安全挑战。传统以边界为中心的安全防护体系，难以适应如今业务高度分布化、动态化的环境，这催生了对安全服务链的迫切需求。具体需求与挑战包括以下几个方面：

多租户安全隔离

随着企业 IT 架构的演进，云环境中往往承载多个业务部门、子公司或项目组的服务，这些业务单元之间在资源使用和数据流转上需要严格隔离。安全服

务链必须具备基于租户的隔离能力，不仅要从网络层实现租户间的隔离，还要在应用层和数据层确保无越权访问。同时，需要灵活支持企业根据组织架构调整业务的合并、拆分、重组时的隔离策略，确保安全和灵活性兼顾。

动态防护与弹性扩展

云计算环境的动态特性使得业务流量具有极强的波动性，例如电商促销、热点事件、线上教育等场景都可能出现流量激增。传统静态部署的安全设备往往无法应对流量骤增，容易成为瓶颈。安全服务链需要具备按需启用和扩展的能力，实现安全资源与业务流量的实时匹配，确保在业务高峰期也能提供持续稳定的防护能力。

多样化场景适配

现代业务场景下，企业不仅需要应对 DDoS 攻击、SQL 注入、跨站脚本攻击（XSS）、恶意代码传播等通用型威胁，还要针对金融、电信、能源等行业的特定威胁提供精准防护。安全服务链需要具备丰富的模块能力，能够支持 Web 安全、应用安全、网络安全、内容安全、主机安全等多样化场景，并根据业务需求灵活组合，实现多场景下的安全全覆盖。

可视化、可观测与可追溯性

在复杂云网环境下，安全运维面临海量数据、快速变化和多元威胁的挑战。单一的日志和告警无法满足运维需求，需要实现从链路、流量、威胁到用户行为的全链路可视化。安全服务链不仅要提供流量监控、告警通知、攻击识别等基本功能，还要支持攻击路径回溯、威胁溯源、日志审计等深度分析能力，帮助运维团队及时定位问题、精准处置、优化防护策略。

低延迟与高可靠性保障

在安全防护与业务体验之间，如何取得平衡是重要挑战。企业希望安全防护的介入对业务延迟最小、对带宽影响可控，同时在系统故障、链路异常时，安全服务链能够实现快速切换和高可用保障，避免防护系统本身成为业务稳定性的薄弱环节。

6.3.2 网络安全服务链架构

基于云联网架构，安全服务链通过多项核心能力实现对多租户、多场景、多业务环境的全面防护，安全服务链架构示意图如图 13 所示。这一架构具备灵活扩展、精细控制和高可靠性，以下从四个核心维度对安全服务链的技术架构与能力体系进行详细分析。

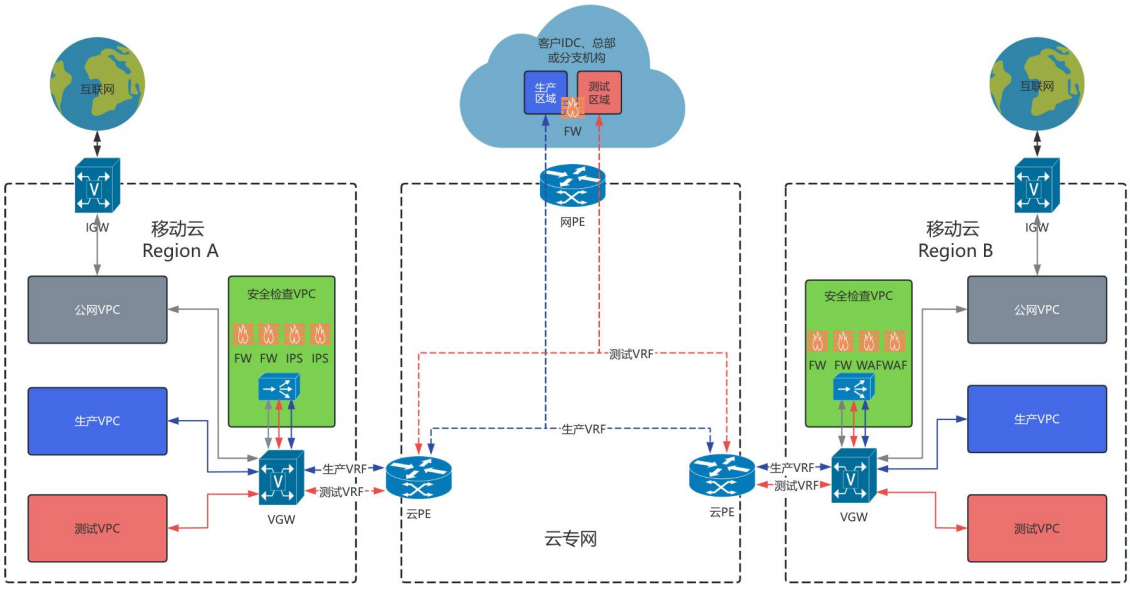

图 13　网络安全服务链架构示意图

开放架构：基于开放标准协议无缝接入

安全服务链采用开放架构设计，基于 BGP、VXLAN 等开放标准协议，能够无缝对接自有云网络安全产品，并与第三方安全产品协同工作。这一开放性架构提高了系统的兼容性和互操作性，使企业能够在多云、混合云环境中灵活引入防火墙、入侵防御、DDoS 清洗、威胁情报等能力模块，显著降低集成成本，为构建开放、协同、可持续演进的安全生态奠定了坚实基础。

水平扩展：基于 GWLB 与 SR-SFC 的弹性能力

安全服务链通过引入网关负载均衡（GWLB，Gateway Load Balancer）技术，打通各类安全网元的水平扩展能力。借助 GWLB，系统能够针对业务流量和安全需求实现按需扩缩容，有效应对高并发业务场景或突发攻击流量，提高整体系统的弹性和资源利用效率。进一步地，安全服务链结合 SR-SFC（Segment Routing – Service Function Chaining，无状态服务链）技术，通过在网络报文中携带服务

路径信息，实现网络与安全服务链的无状态化调度。SR-SFC 突破了传统服务链对硬件状态维护的依赖，使多安全网元间的编排更灵活、链路更高效，并降低了扩展复杂度。结合 GWLB 与 SR-SFC，安全服务链具备出色的水平扩展能力，能够满足大规模、多租户、多场景的动态防护需求。

最小权限访问控制：实现跨 VPC 安全隔离与受控访问

基于云联网的多 VPC 安全隔离机制，安全服务链引入跨 VPC 的受控访问控制能力，实现了面向租户、应用的最小权限访问控制。系统可灵活配置访问策略，确保不同 VPC、不同业务单元之间仅开放必要的最小权限访问，有效降低潜在攻击面。这一机制对提升租户隔离强度、强化安全边界、落地零信任理念起到了关键作用，为多租户环境提供了更高水平的安全保障。

高可靠性：动态路由驱动的安全资源池容灾切换

安全服务链采用动态路由机制构建安全资源池，当检测到节点故障或链路异常时，系统可实现快速流量切换，完成安全服务的容灾恢复与业务连续性保障。相比静态绑定方式，动态路由方案具备更强的自愈能力和更高的系统可靠性，能够显著降低因节点故障、攻击冲击带来的业务中断风险，满足金融、政务、电商等对业务可用性要求极高的行业需求。

6.3.3 技术挑战性、创新性与先进性

技术挑战性

安全服务链对租户 VPC 侵入性：当前主流安全服务链方案多依赖 PBR 策略路由或基于 NAT 的转发机制，需在用户 VPC 路径上强制引入安全检测路径，导致业务链路变更、路由复杂、运维成本高，且难以实现无感知接入，削弱云原生架构对弹性与自治的支持。

安全网元状态同步限制扩展性：当前 vFW、IPS 等安全服务大多为有状态网元，通常需要进行双机同步状态，导致系统复杂度高、部署效率低、资源利用率受限。在大规模租户并发接入和横向扩展场景下，状态一致性维护成为制约服务链弹性扩展能力的重要瓶颈。

技术创新性

云原生网络安全能力构建：依托云联网底座，中国移动实现东西向、南北向流量的无侵入式引流，安全检测路径不依赖租户 VPC 原生配置，实现安全能力对租户的“透明插入”和弹性挂载，全面支撑云原生环境下的网络安全能力弹性接入。

标准协议创新：主导 IETF 工作组草案-基于 SR 的无状态服务链(draft-ietf-spring-sr-service-programming)以及个人草案-面向 SDWAN 的高效 IPsec 隧道封装（draft-xu-ipsecme-esp-in-udp-lb）。

技术先进性

云原生安全架构对接简洁高效：相较于 AWS 的 Security VPC 方案，无侵入式云原生安全架构具备更高的接入简洁性与多租户兼容性，更易对接第三方网络安全网元，实现云端安全能力灵活引入与统一管理。

无状态安全服务链：采用集中式 GWLB 路径编排方案的传统 Hub-Spoke 服务链形态存在系统耦合度高、调度灵活性差的问题。无状态安全服务链具备良好的横向扩展性与链路动态重构能力，安全网络的增删调整更加灵活。

第 7 章　结语

面向 AI 时代的发展需求，中国移动针对智算与云计算深度融合趋势，系统性构建了云智算新型网络基础设施体系，持续推进关键技术创新。在智算网络侧，依托开放以太架构实现 Scale-Out 与 Scale-Up 网络高性能互联与架构统一融合，打造超低时延、超大带宽、超高可靠的智算网络；在物理广域网络侧，基于可预期 IP 网络以及开放光传输网络底座，构建全球一体化的可预期广域网络服务；在虚拟网络侧，强化云内网络编排与多云互联能力，构建灵活可编排、按需可调度的网络服务体系，并融合安全服务链防护机制，保障网络与业务安全。

面向未来，中国移动将通过持续的技术创新与规模化实践，不断迭代云智算网络基础设施的能力与架构，为 AI 模型演进、数字经济发展和全球业务拓展提供坚实的网络底座。

附录：术语与缩略语

中文名称	英文缩写	英文全拼
大语言模型	LLM	Large Language Model
远程直接内存访问	RDMA	Remote Direct Memory Access
图形处理器	GPU	Graphics Processing Unit
可扩展性、高可用性、低时延、低成本	SHALL	Scalability, High Availability, Low latency, Low cost
数据中心互联	DCI	Data Center Interconnect
软件定义网络	SDN	Software Defined Networking
网络即服务	NaaS	Network as a Service
网关负载均衡	GWLB	Gateway Load Balancer
段路由服务链	SR-SFC	Segment Routing - Service Function Chaining
虚拟私有云	VPC	Virtual Private Cloud
网络功能虚拟化	NFV	Network Functions Virtualization

虚拟防火墙	vFW	Virtual Firewall
超级全球加速	SGA	Super Global Acceleration
边界路由出口工程	EPE	Egress Peer Engineering
流量工程	TE	Traffic Engineering
基于优先级的流控	PFC	Priority Flow Control
等价多路径	ECMP	Equal-Cost Multi-Path
全自适应路由以太网	FARE	Fully Adaptive Routing Ethernet
人工智能	AI	Artificial Intelligence
容器编排系统	K8s	Kubernetes
基础设施即代码	IaC	Infrastructure as Code
软件定义广域网	SDWAN	Software Defined Wide Area Network
应用编程接口	API	Application Programming Interface


中国移动
China Mobile