ZTE中兴

IP网络未来演进技术白皮书4.0
——服务感知网络（SAN）

中兴通讯股份有限公司

IP 网络未来演进技术白皮书 4.0 —— 服务感知网络(SAN)

版本	日期	作者	备注
V1.0	2021/05/22	ZTE	新建
V2.0	2022/08/28	ZTE	更新：提出开放服务互联网络解决方案和三大关键技术：服务感知网络（SAN）、增强确定性网络（EDN）、网络内生安全
V3.0	2023/08/22	ZTE	更新：提出增强确定性网络（EDN）架构及其关键技术
V4.0	2024/09/27	ZTE	更新：提出服务感知网络（SAN）架构及其关键技术

重要贡献单位：

中国信息通信研究院

中国电信研究院

中国联通研究院

北京交通大学

目录

图

1 前言

云计算、大数据、AI 训练等新型业务驱动下，算网融合成为新型数字基础设施的典型场景和核心需求。算网融合在资源协同调度效率、业务性能优化等方面的需求和收益，近年来已经成为行业初步共识。随着云原生、AI 训练等新型业务的飞速发展，对基础网络增量需求的内涵和外延，均出现了全新的变化。本文延续此前系列白皮书，聚焦算网融合场景的技术和架构演进视角，并就数据中心内及数据中心间东西向业务流量场景进行延伸覆盖，同时扩展了算网层次化性能检测、业网协同、独立语义服务标识等方向的架构和方案阐述。从服务感知网络整体架构视角来看，本文将继承基于服务标识索引的算力路由和精细化网络连接这两大基础功能。其中，服务标识的独立语义将会得到强化表述，即基于服务标识构建覆盖多场景的统一技术架构。服务标识是算力与网络、业务与网络、端侧与网络融合的关键功能单元和架构接口。

2 IP 网络未来演进趋势

从分组网络的业务承载类型来看，纯话音业务模式下，业务和网络一体化融合，网络内生支撑话音业务，业务得到确定性保障，业务就是网络，网络就是业务。互联网数据业务模式下，业务和网络在架构和协议层面分离，多种业务流量在承载网络之上混合共存，网络不再感知和识别业务，而是对所有业务流量进行基于统计复用的“尽力而为”转发，网络成为纯粹中立的承载基础设施，这种简明清晰的架构和部署模式，凭借其优异的可扩展性和成本优势，助推了互联网业务的爆发式增长。云计算的兴起和广泛部署，改变了业务数据的流量模型和算网交互模式，并对网络提出了延伸感知算力和业务的新型需求。网络、算力和业务从彼此独立割裂到适度融合以因应新型应用场景，遂成为新型数字基础设施模式下的演进趋势。

尽管过去 20 年以来，业界涌现出多种试图替代 IP 网络的技术架构（如 NDN/ICN 等），但由于牵涉海量的网络基础设施存量投资，颠覆性替代方案很难被广泛接受并部署。更重要的是，IP 网络架构并非封闭和一成不变，IPv6 提供了丰富的扩展机制，支持平滑演进的功能增强，SRv6、BIER（新型组播）等即为经典案例。同时，IETF 也已经正式启动 MPLS 2.0（又称 MNA）的标准推进工作，基于 MPLS 的扩展和增强也将成为现实可能。因此，基于轻量级数据面和智能控制面的功能增强和扩展，将是 IP 网络技术平滑演进的主流趋势。

本白皮书在《IP 网络未来演进技术白皮书 2.0》提出的开放服务互联网络及其关键技术的基础上，详细阐述服务感知网络（SAN）的场景需求、架构理念、关键技术、测试验证、技术及产业展望等内容。

3 SAN 关键场景及需求

3.1 算网融合

3.1.1 算网融合下的广域服务部署与协同场景

当前，构成互联网的两大基础设施——IT 基础设施和 CT 基础设施都在经历重大的融合演进变化，即“云化”和“网络化”。“云化”是指包括应用、网络和云设施等都在虚拟化，通过虚拟化实现应用、云和网共享基础资源池，如计算、存储、转发等，实现一体化编排和调度。而“网络化”可以理解为去中心化、分布式计算、边缘计算等。网络化主要是应对广泛部署的无线网络和 IoT 设施，原来基于集中式架构的云化模型无法满足泛在的应用需求和分布式的云资源部署，从而向由网络连接的分布式算力和存储等虚拟化云资源提供模式演进。

一方面，云原生作为一种面向云计算时代的软件架构和开发理念，旨在充分利用云服务的弹性、可伸缩性和高可用性，为应对云计算环境下的复杂性和需求提供全新的解决方案和开发范式，其自身具备丰富、鲜明和独特的特性，如容器化，微服务架构，自动化部署与调度，弹性资源伸缩，故障隔离及高可用性，服务网格等。

随着云原生应用开发和运维模式的成熟，未来上云的应用将可以分解为更细颗粒度的原子服务，通过调用分布式的原子服务，或者原子服务的组合，从而可以支持更复杂的应用场景，如元宇宙等。因此，以原子服务为颗粒度的服务和感知技术，将能够在精确服务能力和分布式资源调度方面满足未来云原生应用发展的需要。另一方面，边缘计算作为一种分布式计算范式，其核心思想是将计算、存储和应用程序的处理能力尽可能靠近数据源和最终用户，以降低延迟、提高响应速度和数据安全性。其特征与优势包括分布式部署，低时延，高资源利用率，带宽成本优势等。应用的原子化解构与基于广域的服务与算力部署、连接和协同，共同组成了新时代下的算网融合新场景。

3.1.1.1 广域跨边缘的服务连接场景

子场景 1：跨集群服务连接的边车与代理拦截

以 Istio 为例，其作为一种服务网格架构，使用 Envoy 代理服务网格中的所有服务协调入站和出站流量。Istio 使用 Envoy 的各种特性，如动态服务发现、负载均衡、TLS 终止、HTTP/2 和 gRPC 远程过程调用（gRPC）代理、熔断、健康检查、分阶段推出基于百分比的流量划分等。当存在流量流入或从应用和服务容器流出时，流量总是会被边车和代理拦截，进而由边车和代理根据控制面管理的路由和路由规则引导流量。

可以预料到，当部署在多个边缘云和集群中的服务需要连接和通信时，服务的拦截和引导将会变得更加复杂。数次的边车与代理拦截也将使得服务的连接能力下降。

子场景 2：异构集群的部署与广域协同

当处在不同连接域的资源和服务之间存在连接和协同的需要时，如图 1 所示，数个 K8S 集群中和另一个虚机集群中部署的服务存在连接需要，如何适应性地管理多种连接域的资源和服务实例成为了亟待解决的问题，其中包括但不限于：数个集群中的部分实例存在连接需要，但另一部分实例不被允许通信时，如何能够更高效地为每个集群配置规则；当处在 K8S 连接域的服务实例希望访问处在一个虚机集群中的实例中，如何统一服务语义。

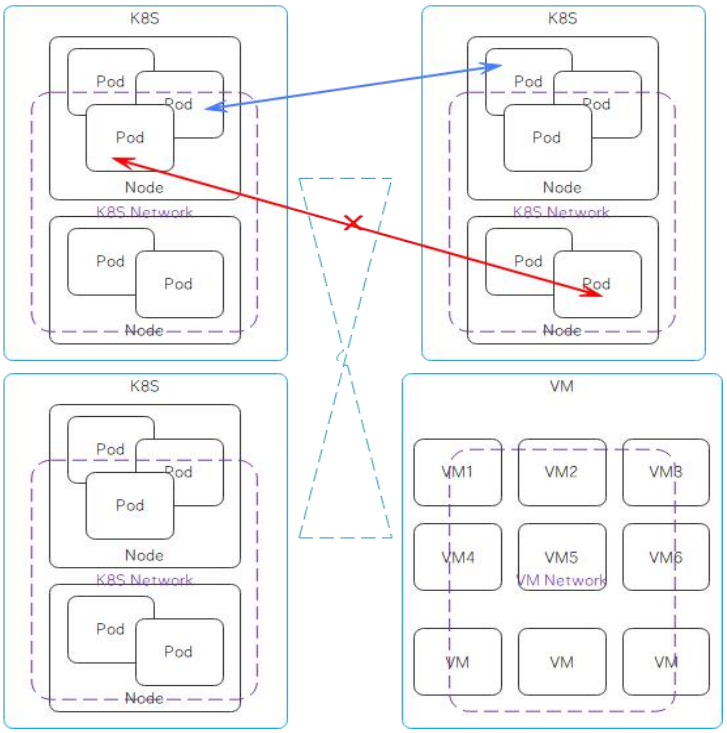

图 1 异构服务集群的互联

3.1.1.2 广域多边缘的服务调度场景

子场景 1：管理大量、动态的远端服务实例

在云原生的背景下，应用程序往往通过名称（例如 Service IP）对服务进行寻址，来避免直接通过 IP 地址对服务进行寻址的脆弱性。以集群 1 访问集群 2 中的服务为例，需要在集群 1 中配置集群 2 中的服务名以及其所对应的的 endpoints 中，如集群 2 的 Ingress Gateway 地址。即，该将集群 1 访问该服务名的流量路由到集群 2 的 Ingress Gateway。

而 Serverless 技术是一种云计算服务模型，其核心理念是开发者无需关心服务器的管理和维护，可以专注于编写和部署代码。其中，云服务提供商会根据负载自动扩展和收缩函数实例，以满足当前的服务请求量，而开发者则无需手动干预。当请求进入 FaaS 平台后，请求处理模块会在实例管理模块中查询是否有可用（空闲）状态的实例，如有空闲实例，则将对应的请求调度到该实例中，相应的加载和运行业务代码，并返回结果。当业务请求激增时，资源池中无可用实例，则去资源调度模块中申请扩容，资源调度模块相应创建新的实例加入资源池。

因此，在算网融合的背景下，网络连接的多边缘系统中可能存在着大量的、动态唤起或释放的远端服务实例。由于本集群访问远端集群实例的前提是远端服务实例在本端 API 中的注册和配置，而远端服务往往是大量的和动态的，这意味着本端集群需要动态地处理大量的远端服务实例在本端的配置，相应的，这会对本端集群的管理和运维造成极大的负担和困难。

子场景 2：算网一体调度

以 Istio 为例，Istio 中的流量操纵和路由的实践主要包括 API 的方式影响部署中的流量。如通过 DestinationRule 配置根据标签将单个服务拆分为子集。并为每个子集分别配置特性。常用的负载均衡方式包括：Round Robin（轮询）、Least Request（最小请求数）、Weighted Round Robin（加权轮询）等。

可以发现，服务路由的决策发生和执行在服务端侧，由集群的网络 API 配置指导执行，然而服务端侧的业务流量调度往往不具备网络侧传输能力的信息。如图 2，远端服务 Service B 部署在两个边缘，Service B 的多实例在本地集群的 ServiceEntry 中注册。Service B 的多实例之间在本地集群配置为 Round Robin 或 Weighted Round Robin 等可行的负载方法。然而，可以注意到，负载均衡方式的配置往往是静态的，当连接 Service B

的多实例的网络路径具备不同能力时，如图 2 中，两条网络传输路径的基本能力为（30ms，100M）和（60ms，400M），静态配置的负载均衡方式往往很难为用户提供极致的服务体验。

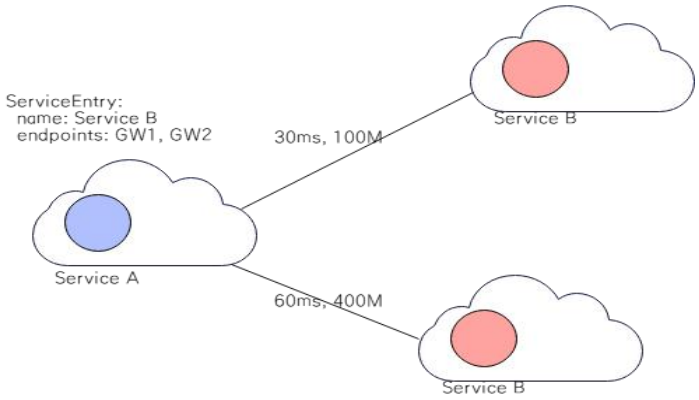

图 2 多边缘多实例的服务路由

子场景 3：多原子服务和功能的协同调度

在云原生应用的大背景下，应用程序往往不再以单体服务的形式存在，而是被分解为一系列云原生服务部署在 Kubernetes 集群中，服务之间存在调用链路、共同对外提供服务。如流量泳道技术应运而生，用于在灰度发布的场景下，对服务的整条请求链路进行环境隔离与流量控制。而事实上，流量泳道的创建依然由集群的各种网络 API 配置执行。配置于集群和服务端侧的服务路由标识了在一个服务名下的 endpoints 和在这些可选实例之间的调度策略和负载均衡方法。然而，在服务之间存在多跳的请求与调用链路的场景下，服务端侧的下一跳服务路由和该跳服务请求或调用在请求与调用链路中的相对位置和对应需求无法建立关联关系。

如图 3 多边缘之间部署了 Service A-D 的四种服务，它们之间存在着两种请求与调用链路，分别是 Service A-Service B-Service C 和 Service A-Service B-Service D。针对每种调用链路，其中的服务相互协同，共同提供对外服务，其中前者组成了一种端到端的

时延敏感型服务，而后者组成了一种对端到端传输带宽具有较高需求的服务。此时，对于配置在 Service A 所在集群的服务路由策略和规则，无法区分出这两种服务请求和调用链路，因为对于 Service A 的服务实例而言，这两种服务请求与调用链路所对应的对外服务具有着一个相同的服务的 endpoint。

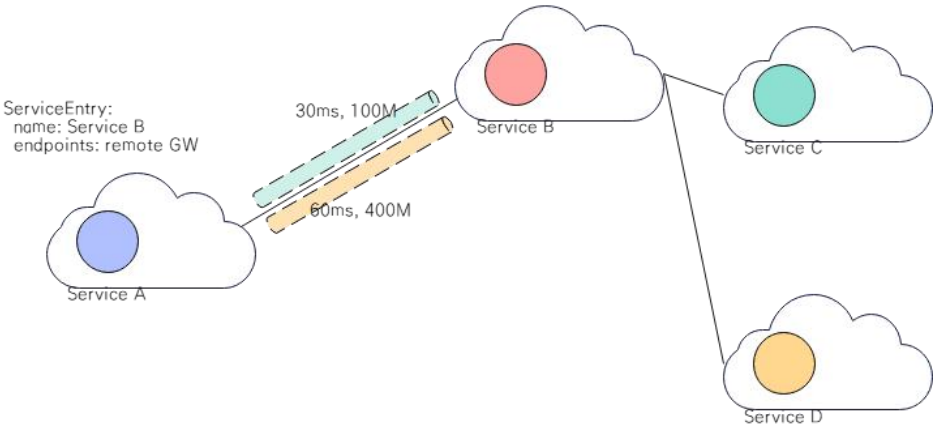

图 3 处在不同请求与调用链路的服务实例

3.1.1.3 广域多边缘的服务可观测场景

当应用和服务被分解为由多个分布式部署的服务实例协同和组合对外提供时，如何跨越多个服务端点和连接服务端点的传输网络提供端到端服务流量的可观测和保护特性成为了新的亟待解决的问题。

在网络中，已经具备了基本成熟的 OAM 技术实现传输网络和链路的检测。其中，如 STAMP（Simple Two-way Active Measurement Protocol，简单双向主动测量协议）可以通过配置实现 Session-Sender 和 Session-Reflector 之间 STAMP 会话的建立。STAMP 会话由（源端 IP、目的端 IP、源端 UDP 端口、目的端 UDP 端口）四元组组成一个 STAMP session。STAMP Session-Sender 向 STAMP Session-Reflector 周期性发送测量报文，STAMP Session-Reflector 针对每个接收到的测量报文回送响应测量报文，STAMP Session-Sender 根据接收到的响应测量报文计算出丢包率、时延和抖动等 SLA

性能参数值。

在应用侧，如 eBPF（Extended Berkeley Packet Filter）技术可以在 Linux 内核中动态注入代码，从而实现高效的系统监测和调试。它可以使用钩子和事件触发器，如通过在关键位置（例如系统调用、网络数据包处理、内核事件等）安装钩子（hooks），以及定义事件触发器（event triggers），来捕获特定的系统活动。如可以在进程创建、调度、终止等关键事件时设置钩子，从而记录或者分析这些事件的发生情况。eBPF 可以分析进程或线程的运行时行为，例如测量系统调用的响应时间、资源使用情况等，以帮助识别性能瓶颈或优化潜力。其他的，如 APM 技术（Application Performance Management，应用性能管理），提供了监控与数据采集、性能分析与诊断、实时监控与报警和可视化与报告特性与能力。APM 通常提供直观的仪表板和报告，用于展示应用程序的性能指标和趋势，有助于运维人员和开发团队快速了解应用程序的健康状况和潜在问题。

网络侧的 OAM 技术提供了对传输网络的监测和可观测手段，但是缺乏网络基础设施关联的服务与应用的信息；eBPF 通过在内核中动态注入代码，利用系统调用和内核事件的钩子机制，实现了对 Linux 系统中线程和进程活动的细粒度监测，但是这些线程与进程活动如何与服务逻辑关联仍然是缺失的一环；APM 提供了应用的仪表与大盘，却没有手段采集和获知跨广域网络的连接情况与能力。

3.1.2 算网融合下广域服务协同的需求与挑战

算网融合下的广域服务部署域协同场景下的需求与问题，对新时代下的算网基础设施与系统提出了新的需求和挑战：

- 一致和高效的服务连接能力：实现同或异服务提供和运营商的跨集群、单或多网络、可能不同运行态和连接域场景下的服务实例间高性能和一致性连接。

- 有效和智能的服务调度能力：实现有效和高效的广域多边缘、多集群的分布式服务实例的系统性管理和运营；实现多样化管理和调度策略下，结合算网资源能力并满足服务 SLA 需求服务流量调度和负载均衡；实现跨越多边缘、多服务端点的全程服务的差异化服务提供与一致性服务能力保障。
- 端到端的全程服务可观测能力：实现跨越多边缘、多服务端点的全程服务的可观测能力，提供包括服务粒度的多层次可观测与劣化、失能定位与维护手段，进而实现端到端全程服务的调优和失能保护。

3.2 算网一体调度和运维

3.2.1 OTT 和智能边缘服务调度主要问题和痛点

在算网一体服务模式下，确保端到端协同是维持高质量与稳定性的核心。然而，在传统 OTT 业务中，基于 DNS+GSLB 的传统算力服务模式与网络状态分离，难以满足全面的服务需求。随着 AI 大模型的发展，受限于单芯片算力提升速度（摩尔定律），未来算力供给将更多依赖于集群计算。入算推理结合边缘计算与 AI 技术以提升用户体验、优化资源分配并增强数据安全性。在这种背景下，算网分离的调度模式将面临以下关键问题和痛点：

- 性能短板：算力服务与网络状态不匹配时，可能导致数据传输延迟或失败，影响系统性能和效率。例如，在算力资源充足的情况下，网络拥堵仍会导致数据处理速度变慢。
- 资源浪费：算网分离状态下资源分配不合理，可能导致在网络负载较低时未能充分利用算力资源，或在网络拥堵时过度分配算力资源。
- 服务质量下降：高延迟或数据包丢失等现象可能影响用户体验，如视频流不流畅、在线游戏体验差或文件传输失败。

- 成本增加：资源使用效率低下可能导致成本上升，例如过度使用云服务资源或产生不必要的数据传输费用。
- 安全风险：算网状态分离可能导致安全策略执行不一致，从而增加数据泄露和其他安全威胁的风险。
- 管理复杂性增加：需要分别监控和优化算力服务及网络状态，增加了管理复杂性和工作量。

3.2.2 算网一体调度和运维场景和需求

针对 OTT 和智能边缘服务调度的问题，算力网络提出了一体化调度策略，以网为中心感知网络和算力状态，拉通算网资源运营和调度体系。这一策略旨在获取面向服务访问的最优解，并重点满足以下关键场景需求：

- 一致最优体验：端到端算网调度过程中，全面考虑网络延迟、服务器响应时间等关键因素。高延迟会使得页面加载迟缓、视频播放不流畅；低速网络则显著降低数据传输效率。因此，在算网联合决策时，必须重视这些因素以确保用户获得流畅、高效的使用体验。
- 算网负载均衡：作为复杂的系统工程，实现计算和网络资源的有效利用需要综合考量资源状态（计算节点和网络链路的负载、可用性和健康状态）、任务特性、负载均衡策略与算法、动态调整机制、故障恢复能力以及预测优化技术，以提高系统整体性能和响应速度，增强系统稳定性和可靠性。
- 一体化运维：打破算和网信息孤岛的关键，在于构建统一的监控平台以实时监测计算、存储与网络资源状态；利用自动化工具和流程提高运维效率；建立快速响应机制确保故障迅速定位与服务恢复；通过动态资源调整和优化调度策略避免资源浪费并提升利用率。

3.2.3 算网一体调度和运维主要挑战

传统网络的 Traffic Engineering (TE)、运维(Operations)与 Operations, Administration, and Maintenance (OAM) 三者在现代网络体系中扮演着至关重要的角色。TE 致力于优化网络资源分配和提升传输效率，运维负责日常操作与管理，确保服务连续性，而 OAM 则专注于网络监控、故障检测与性能管理。这三个方面相互依存，协同作业：TE 为运维和 OAM 提供基础网络优化策略；运维通过执行 TE 策略并利用 OAM 工具监控网络状态；OAM 反馈的实时信息支持 TE 和运维做出进一步优化决策。这种协同模式保障了网络资源的有效利用和持续的高服务质量。

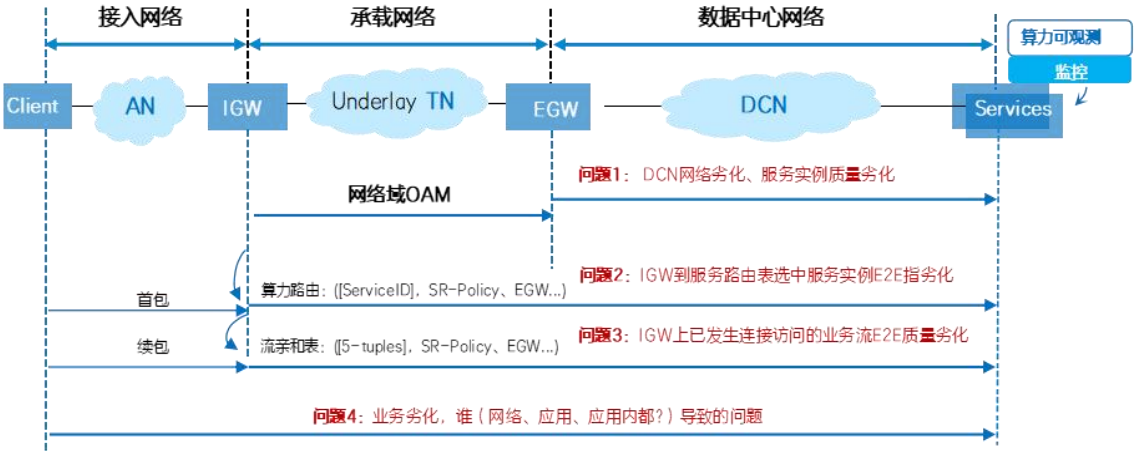

图 4 算力网络 TE 和运维示意图

如图 4 所示，在算力网络环境下，网络设备需感知算力资源状态，并结合网络资源状态进行智能决策，选择最佳计算实例和路径。为了提供更优质的服务，一方面要基于网络和算力资源实时状态进行联合计算；另一方面需快速检测算网服务质量是否达标，以触发业务动态调整。此外，故障根因分析对于提升运维效率至关重要。在分布式场景中，IGW（Internet Gateway）需综合考虑网络与算力状态，精准计算最优路径及资源分配。传统 TE 技术通过链路 OAM 收集信息，并经 IGP-TE 扩展至 IGW 以满足网络 SLA 需求。然而，在算网融合场景下，TE 机制必须处理网络与算力数据，在周期性调度、逐流分配及负载均衡等场景

下优化资源分配，但目前的 TE 技术尚不具备算力感知能力。

此外，实现基础算力网络功能尚面临四大挑战：

- DCN 网络与计算实例性能波动：DCN 网络劣化或中断以及计算实例负载上升均会影响连接质量。虽然算力上报机制可以监控实例状态，但慢周期上报机制在 DCN 网络故障或性能下降时会导致算力重路由延迟，形成路由黑洞。
- 算网 SLA 一致性验证缺失：IGW 基于服务标识的算力路由表难以确保算网联合服务达到 SLA 要求。为解决此问题，需要实时 OAM 检测以动态调整路由策略，但现有电信级 OAM 技术无法直接支持算力路由级别的监测。
- 用户体验保障难题：算力服务访问依赖流亲和表转发至计算实例。流量增加会使实例处理延迟加剧，影响用户体验。现有电信级 OAM 技术缺乏针对流级别的精准监测与动态路径优化能力，难以保证终端用户的一致体验。
- 故障排查困难：当业务端到端业务会话质量下降时，需要从运维层面判断问题源自网络还是计算资源。如果问题是计算资源导致的，需进一步定位具体环节。

尽管当前网络侧 OAM 技术已相对成熟，但尚未延伸至计算实例节点监控，加之算力领域缺乏统一的 OAM 标准，在构建以网为中心的算力网络时，确保算网服务质量和实现高效运维成为两大关键挑战。为应对这些挑战，需要将网络域 OAM 的专业知识扩展至计算资源状态分析，同时优化资源分配与应用部署，并从用户角度出发确保服务健康及性能。通过 AI-OPS 和大数据分析等先进技术的融合，可以提升 IT 服务的可靠性、性能及用户体验，实现基础架构健康监控与问题的快速定位。

3.3 网络能力和业务需求协同

3.3.1 未来 IP 网络中业务和网络协同

第一阶段——无 QoS 保障的业网分离工作模式：

在过去 10 多年中，互联网业务的快速发展主要归功于“以业务为中心”的网络设计理念，业务开发和设计不用过多考虑底层承载网络的约束条件。因而互联网业务发展的第一阶段，以 IP 为中心的网络架构便设计成业务层逻辑和网络层传输是解耦的，因此各种业务数据流得以无差别的在一个标准的底层网络通道中进行传输，如图 5 所示：

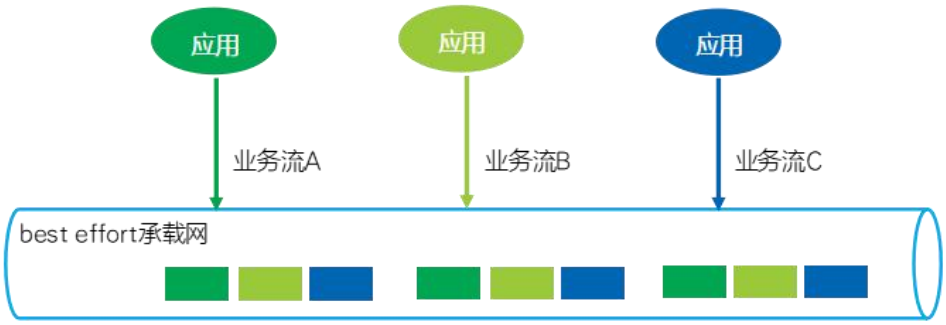

图 5 传统互联网基于“尽力而为”的无差别传输服务

大多数业务都是以 OTT（over-the-top）形式运行，OTT+互联网的组合推动了互联网电子业务的初期的蓬勃发展。

第二阶段 --- Overlay 网络支持基于内容的差异化传输工作模式：

在互联网业务发展的第一阶段中，IP 承载网虽然支持多样化的业务数据传输，但是无法提供针对特定 QoS 业务和用户分类的数据传输保障。因此，在互联网业务发展的第二阶段，对于传统业务 QoS/QoE 的保障，一般使用应用层优化冗余机制，如传输协议的优化，或者叠加应用层网络（overlay），以及专线等方式来进行业务支撑（CDN, RTN, QUIC 是 overlay 解决方案的经典代表），如图 6 所示。

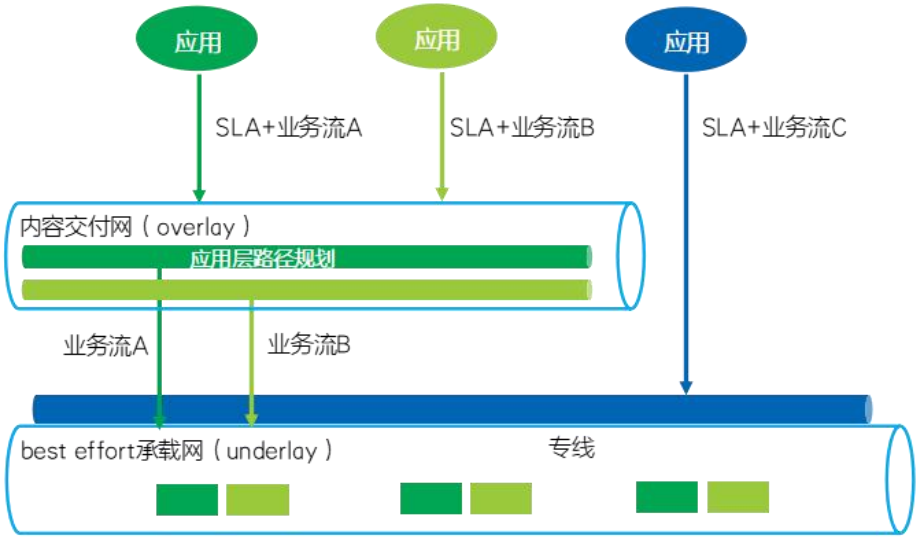

图 6 overlay 应用层和专线提供差异化内容交付服务

但从 underlay 承载网络方面来看，当前流量轻载的现状和专线业务，均意味着存量资源和服务成本还存在较大的挖掘空间。

第三阶段 --- 基于业务感知的业网协同差异化网络传输工作模式：

随着未来更多类型业务的发展，一些应用服务都出现了基于业务特性的数据精细化传输需求。然而在目前的互联网尽力而为的模式下，仅仅依靠应用层的各种补偿方案，没有相关的网络资源的配合，是无法彻底的解决差异化服务问题的。因此在互联网业务第三阶段的发展方向上，这类有别于传统意义上的非 OTT 业务则是对于业务和网络的协同工作提出了需求。

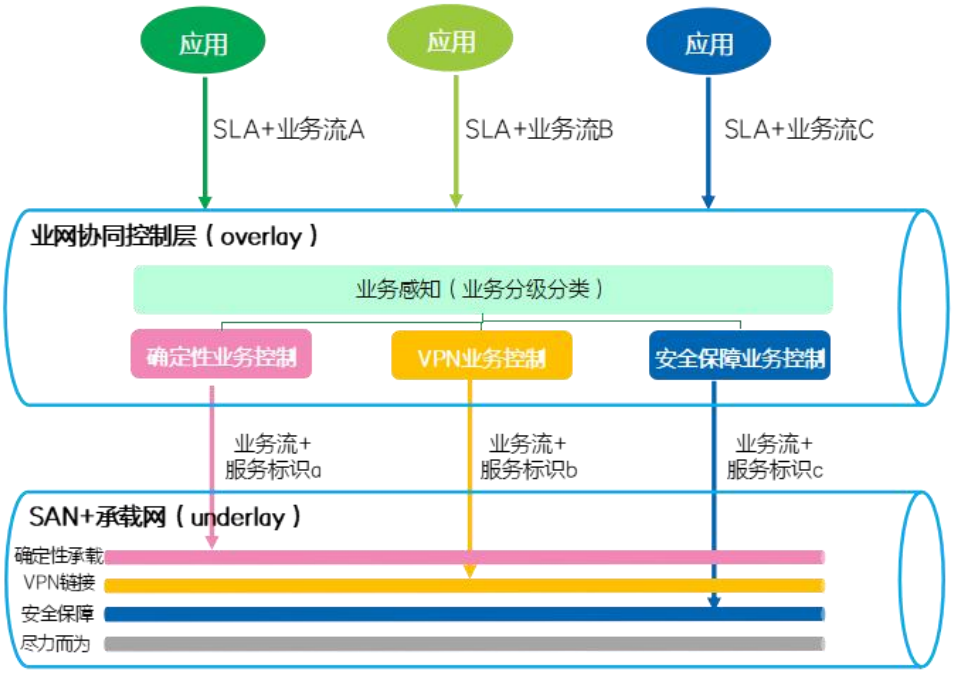

图 7 业网协同功能协同模式为应用层提供匹配业务特征和传输需求的差异化网络服务

如图 7 所示，支持非 OTT 业务的业网协同控制层架构在核心设计上包含了一个能够感知用户业务需求，同时能感知网络资源状态的可运营的中间层设计。此中间层可以为上层的应用层提供可运营的网络层的业务，同时也能为网络层传递提供符合用户业务需求特征的业务控制指令，以便网络层能够理解并转化为控制网络资源和传输策略的传输控制指令，实现对业务的精细化控制，从而将传统的“尽力而为”的管道型互联网转变为“智能化”的业务感知型传输网络。

因此，具有业网协同功能的增强性业务控制层可以更好的服务于多种对于端到端的网络连接和内容传输具有定制化要求的应用场景，例如要求极低延时的工业控制服务场景，要求带宽保障的全息通信服务场景，以及要求具有即时性，安全性和零丢包的数据快递服务场景等等。同时，业网协同在应用于上述场景也催生出了应用和业务需求感知，具有增值服务的网络连接控制，业网一体化的安全保障等等新的网络能力演进需求。

3.3.2 业网协同应用场景

3.3.2.1 支持确定性业务的业网协同应用场景

确定性的业务由于自身业务的特性需求对于数据传输的 QoS/QoE 都有一些定制化的 SLA 等级保障。例如，在面向全景高清音视频媒体流传输的场景下对于带宽保障的需求，对于一些实时双向互动要求比较高的业务场景，例如云游戏，则是对时延有限制需求。而在实时的工业控制上，则对于时延和丢包都有着极度严格的上下边界阈值要求。因此，确定性的业务的核心诉求是需要网络在传输此类业务流数据时严格遵守业务提出的定制化约束条件。

业网协同功能在这一场景中需要实现的能力至少应包括业务的感知能力，其中业务感知包括了能够感知业务流的类型，具体的 SLA 约束参数以及现有网络能够提供服务的能力、资源的匹配等。

3.3.2.2 支持增值型连接业务的业网应用场景

除了应用于数据传输上的确定性业务场景之外，为了保障业务的质量，业务运营方对于业务的连接管控也存在一些差异化的需求。例如一些对于特定业务的访问需要提供业务隔离的 VPN 隧道服务，对业务覆盖区域不同的场景可以提供具有地域广度的单播/组播切换服务，一些特殊应用可以提供固移融合的双通道接入服务，对于服务连续性要求比较高的场景可以提供具有故障秒级切换的自愈网络保障服务，以及能够针对有计算能力要求较高的服务场景提供计算服务的算力网络资源使用服务。

因此，业网协同功能在这一场景中需要能够为不同应用和用户的业务连接提供增值性的连接服务，即同一对节点间不仅仅为宽带用户提供基础的“尽力而为”网络连接能力，还可提供可运营的低延时、大带宽、确定性、安全加密、VPN 连通等不同特性的增值性业务连接和管控能力，从而支撑可运营的多种连接业务模式，并能够对应用层业务进行能力开放和

使用。

业网协同增强层的设计在这一场景中至少需要实现对于网络资源的按需编排以及对于定制化业务连接需求在网络层面上的调度和控制能力，以便将业务数据按需引导入不同的转发通道中。

3.3.2.3 网络支持业务安全的业网协同应用场景

在传统的互联网服务形态上，对于业务的安全保障，由于存在业务和网络解耦的现状，因此对于业务的安全保障和网络的安全保障也是相对解耦的状态。业务层的安全保障机制无法确保在数据传输过程中得到相应的保障，也无法直接控制网络层的安全策略。

业网协同在这一场景中需要能够将业务层的安全认证机制和安全需求结合网络层的安全保障机制进行协同，可以将网络安全服务能力和业务安全服务能力结合后为上层应用提供一个 E2E 的网络连接安全服务，在诸如智能电网，智能远程医疗的应用场景下，构筑业务和网络的协同的安全链接，利用网络的内生安全机制，形成单点登录，零信任安全架构体系，数据加密等综合安全保护方案等。

业网协同功能这一场景中至少需要实现对于业务层安全需求的感知和网络层安全机制的匹配，并能够将安全保障策略在业务层和网络层实现一体化的配置，减少冗余安全处理流程。

3.3.3 业网协同功能的设计需求

上一章节描述了应用业网协同增强能力层的三个应用场景，根据对于应用场景中的需求总结，可以将具体的需求细化如下：

1、网络感知业务需求

业务感知是业务和网络协同工作的关键能力之一，它包含了业务类型感知和业务需求感

知两个方面。在业务类型感知方面，业网协同功能需要能够按业务应用分类分级的要求，精确识别不同应用产生的不同类别特征的流量数据，并能够对于不同的流量进行标识化的管理。

在业务需求感知方面，业网协同功能需要能够深入理解不同业务对网络承载的具体需求，如确定性传输需求、低时延传输需求和大带宽传输需求等。这些需求可以通过时延、带宽、抖动、丢包、乱序等指标的上下界来量化描述。业网协同功能应该能够根据这些需求指标，动态、智能地适配业务需求和网络资源，通过全程全网统一的标识体系和规划，使用统一的精细化标记来携带应用信息、租户信息和业务意图信息等。通过贯穿终端、网络、应用的算力资源和存储资源以及各类操作系统的端到端资源配合，实现算网业务端、网、算、存的一体化调度和确定性和差异化保障。

2、网络资源匹配业务需求

业网协同功能需要能够通过与 Underlay 承载网络之间的交互可以获得资源感知或预占能力。通过获取承载网的性能和资源情况从而提供业务需求和网络资源的精确匹配，并且根据用户的业务状态和当前网络状态按需实时为需要通讯的终端建立连接会话，生成网络能力组合和业务能力组合，对满足 SLA 要求所需要的网络资源量进行系统性的编排。

3、业务协同的运营调度和管理需求

业网协同功能需要提供标准业务承载模型，将具有相同需求属性的业务统一纳入管理，统一调配，并将业务流引导入相应的承载平面，确保业务的快速部署和高效运行。同时，并将业务流引导可以进行多维度的精细化计费，并需要在连接会话终止后，实时回收所分配的 Underlay 资源。

此外，业网协同功能可以实例化成一种具有网络和业务协同工作能力的业务层，实现对网络业务的开通、管控和精细调度能力，能够提供对网络业务的生命周期管理，对传统网络

连接的管理进行自动化、智能化的升级和改造。

4、业务层与传输层可信安全协同需求

业网协同功能需把业务应用的认证、信任关系传递到网络传输层，实现业务与网络安全互信。需提供灵活的用户和业务策略管理功能，根据强大的准入机制，通过身份验证、许可证管理和访问控制等机制，确保只有经许可的业务和用户能够访问到网络。

需要注意的是，不同的应用场景可能实现一个或者多个功能需求的组合。

3.4 智算网络端网协同

3.4.1 网络多路径控制问题及需求

随着分布式存储、高性能计算以及 AI 分布式机器学习场景的兴起，这些新型业务场景对网络时延、抖动以及吞吐性能极其敏感。为了尽量避免拥塞并优化流量转发性能，由发送端主导进行网络路径控制为业界的主流技术路线之一。网络多路径控制要求端侧为业务流量选定满足业务要求的路径，并且在业务流量质量劣化时，需要及时对路径进行切换。

现有现有方案存在如下问题：

- 由端侧探测路径，进行路径的精确发现和信息维护的方案中，通常由发送端主动发送探测报文，通过改变报文 TTL 值以及流量特征值（如源端口号等）实现网络路径的发现，形成发送端的路径数据库。但由于哈希冲突的存在，该种方式有可能导致对网络路径探测不全，此外当出现网络链路变更时，无法及时获取变更后的路径信息。
- 基于流量特征值和网络侧流量负载分担算法，由端侧对网络上各设备选路进行模拟计算，从而得出相应流量的网络路径信息。该方案效率更高，也能够对网络事件更快速的做出反应。但需要端侧提前预知网络设备转发逻辑，并预置算法，端网耦合较紧密，且增加了端侧计算资源的开销和实现的复杂性。

发挥端侧和网络侧原生优势，简化路径控制方案，是实现高效、低成本的网络多路径控制方案的可能路线之一，主要需求如下：

- 明确端和网的能力边界，利用网络侧较强的路径探测和控制能力，满足端侧业务路径需求，降低端侧实现复杂度；
- 提供端网交互标准接口，实现端网解耦，端网相对独立，提升方案兼容性。

3.4.2 拥塞控制问题及需求

在 AI 网络中，需要把网络拥塞控制在一个极低的水平，以最大程度降低端到端的时延和抖动，同时避免由于拥塞引发的丢包，以满足极致性能和超高稳定的需求。拥塞控制技术由拥塞控制算法和拥塞信令机制两部分组成。

当前拥塞控制机制的问题在于：

- 不同的拥塞控制算法一般需要不同类型的网络拥塞指标信息，通常要求相应的拥塞信令机制进行匹配；
- 不同算法对于网络侧需要反馈的拥塞指标信息的需求不一，也需要网络支持不同检测参数的组合。

为了尽量避免定制化的解决方案，一方面部分的拥塞控制算法在设计之初就考虑对于支持多种拥塞信令，另一方面，网络侧的改进需求如下：

- 拥塞信令支持的参数覆盖主流拥塞控制算法的需求；
- 网络修改范围尽可能小，避免整网的改动或升级。

4 SAN 整体架构及设计理念

4.1 算网一体流量工程

传统网络流量工程（Traffic Engineering, TE）旨在通过智能路由、负载均衡、延迟控制、拥塞预防、冗余设计、成本控制、安全增强及动态适应等策略，实现流量和路径调优。其核心目标是优化网络资源利用率与服务质量，以缓堵保畅，满足不断增长的网络性能需求。

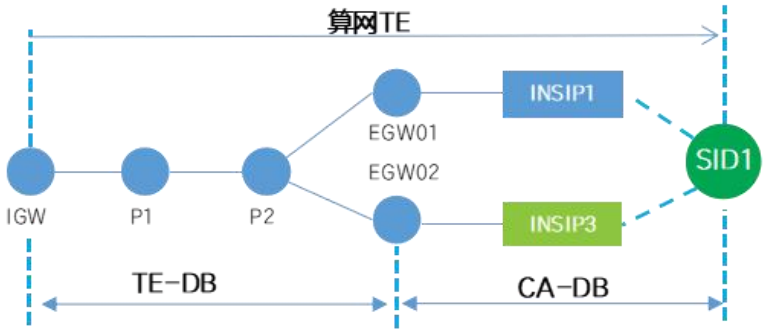

图 8 算网 TE 计算模型

算网一体调度是将用户业务请求匹配到合适的网络路径和计算实例，确保用户体验的同时，提升算网资源利用率，算网一体调度并提供服务的过程即为算网流量工程， 如图 8 所示，计算 IGW 到服务标识 SID1 的最佳路径和计算实例并执行业务请求引流即为算网 TE 过程，算网一体 TE（Integrated Computing and Networking Traffic Engineering）进一步提升了网络 TE 的能力，实现 TE 能力升维（图 9）。它通过感知网络和计算资源的状态，实现计算和网络资源以及业务请求的一体化控制与编排。这种集成化的方法能够实现算力实例的优选，以及流量和路径的精细调优，从而构建出高效、灵活且可扩展的算力网络环境。

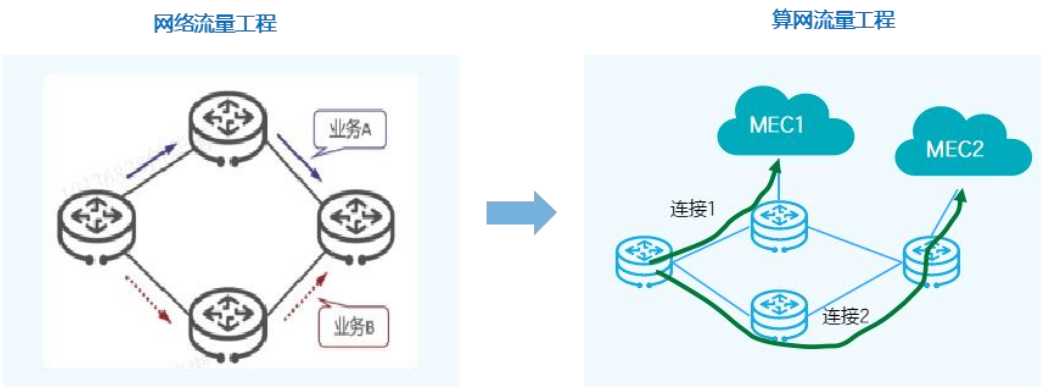

图 9 网络流量工程和算网流量工程

4.1.1 算网一体 TE 三要素

算网一体 TE 通过扩展网络 TE 的三要素，实现算网统一调度：

- 资源管理：网络 TE 的核心在于确保网络符合业务需求的同时，实现网络资源的有效利用。其中，关键在于资源策略需适应网络的动态变化，如流量波动和资源可用性。算网 TE 通过感知计算资源状态，与网络 TE 管理的资源相结合，综合考量网络带宽、缓冲区和队列状态、丢包率和延迟等网络参数，同时考虑处理时延和资源占用等计算度量，形成 TE-DB（Traffic Engineering Database）和 CA-DB（Compute Awareness Database）两大数据库。这些数据库为算网 TE 决策提供关键感知数据集合和决策依据。
- 路径和实例计算：基于 TE-DB 和 CA-DB 的数据，算网 TE 能够根据算网 SLA 要求，计算出当前网络和计算资源状态下满足需求的网络路径和实例。这一过程充分考虑了计算和网络资源的状态信息，有效避免了算力网络一体化服务中的性能反转问题，确保满足特定的服务访问网络质量（QoS）需求和带宽要求。值得注意的是，所选路径和实例不一定是网络中的最短路径。为了简化算网一体计算的复杂度，计算度量信息可被折算成网络同量纲，但这可能会带来一定程度的信息失真，在某些特定场景下，这种模型

的选择是必要的。

- 路径和实例引流：当前，以 SR-Policy 路径为基础，叠加切片（带宽划分）、确定性资源预留（确保抖动上界），并提供多种引流方式，是网络基于 SR 体系提供服务（含算网）的主要方式。在传统的 BSID、Color、DSCP 三种基本引流方式基础上，算网一体 TE 还扩展了服务标识引流机制。利用服务标识，可以将业务精准引入对应的 SR-Policy 和计算实例，实现算网需求的精准匹配。

总结而言，服务连接路径和实例的计算和生成仅与感知到的网络和计算状态以及 SLA 目标相关。通过引流机制，则解决了“谁”（即特定的目的 IP、路由前缀、包含 Service-ID 的 ANYCAST IP、五元组流或聚合流如 Service-ID）将使用这些路径和实例的问题，实现路径和实例生成与使用对象的完全解耦。

4.1.2 算网一体 TE 转控分离技术架构

算网一体架构主要新增了算网计算组件(C-PS)、服务度量代理组件(C-SMA)、算力感知数据库(CA-DB)，实现 IP 网络从一维升级到二维的算力路由，并利用 SDN 转控分离和可编程优势，通过扩展 BGP 协议，融入算力信息，并支持算网计算组、服务度量代理组件、算力感知数据库灵活部署，可以灵活适应感知方式和计算方式多种组合，实现集中式、分布式、混合式算力路由，根据业务和部署特点灵活选择。

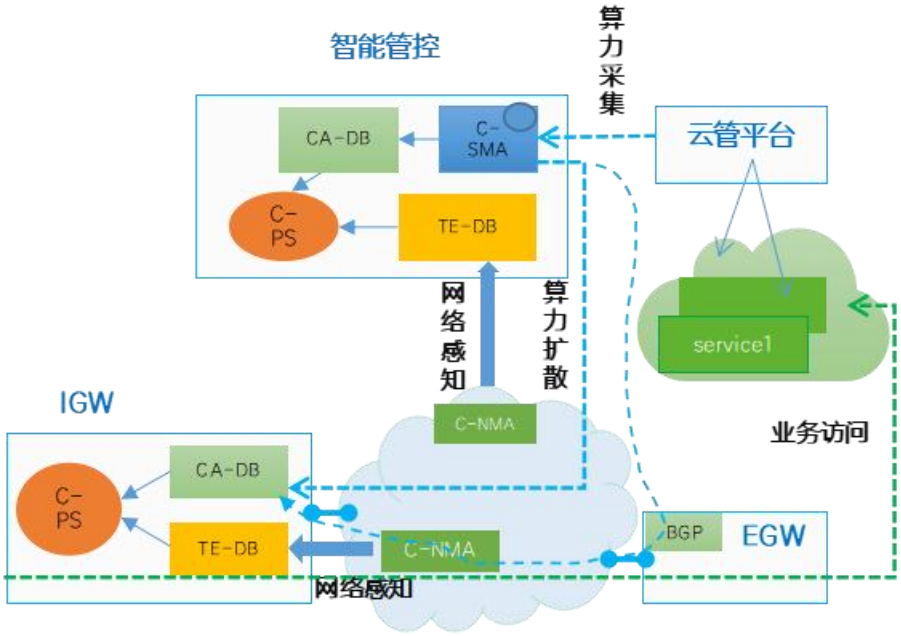

图 10 集中式算力采集+分布式/集中式计算架构

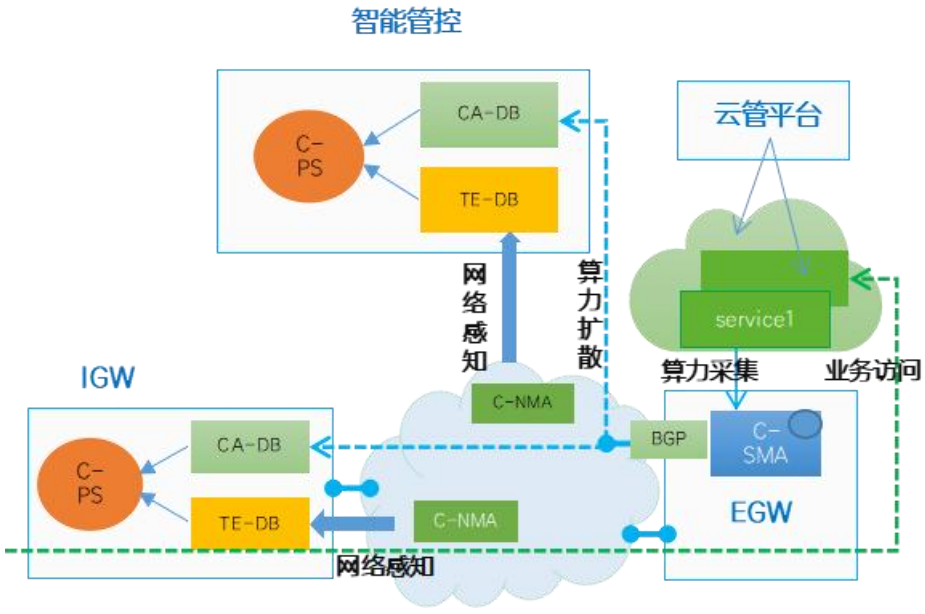

图 11 分布式算力采集+分布式/集中式计算架构

4.1.3 算网一体 TE 部署和应用演进

算网一体 TE 部署和应用遵循循序渐进的原则，以降低对现有网络的影响和侵入性。总体而言，可分为两个演进阶段：

阶段一：

面向尚未全面部署 SRv6（Segment Routing over IPv6）的网络环境。在这一阶段，通过在网络的两端增加算力路由网关，可以在不大幅改变现有网络架构的前提下，预先配置满足特定需求的 SR-Policy。基于这一策略，结合算力实例的状态信息，进行联合调度与计算，以确定符合要求的网络路径与计算资源。这种部署方式能够初步构建算力网络的原型，虽然在支持网络路径的粒度和差异化服务方面存在一定的局限性，但仍能满足基础的算力网络需求。

阶段二：

面向未来全面部署 SRv6 的网络基础设施。随着网络能力的显著增强，特别是可编程性的大幅提升，算网 TE 将能够基于逐跳（per-hop）和链路级别的信息，结合实时的算力实例状态，进行更为精细的联合计算。这一阶段的目标是实现一次性计算出最优的计算实例，并创建相应的 Policy 以满足特定的服务质量要求。这不仅能够提供高度灵活的差异化网络路径选择，还能够实现算力与网络资源的深度融合与开放共享，为未来 6G 时代算网一体化的愿景奠定坚实的基础。

需要说明的是，无论是在阶段一还是阶段二中，变化仅体现在 TE（流量工程）计算模型上。服务标识作为引流的锚点，在转发层面依旧沿用 Policy 的转发模型进行引流与转发操作。总之，在算网一体 TE 的演进过程中，从初步适应现有网络环境到全面拥抱未来网络技术的发展趋势，每个阶段都旨在逐步提升网络的服务质量和资源利用效率，以满足不断变化的业务需求。

4.2 SAN 核心设计理念

基于算网一体流量工程的新型算网协同场景，服务感知网络（service awareness network，SAN）采用基于服务标识的算网一体架构（见图 12），将服务标识引入算网融

合与路由系统，构建了面向业务和算力的 IP 网络新接口。该接口通过服务标识驱动，数据层面细化寻址与流量管理，控制层面关联动态分配的算力资源与业务 SLA，形成了基于 IP 分组网络的服务感知子层（Overlay）。

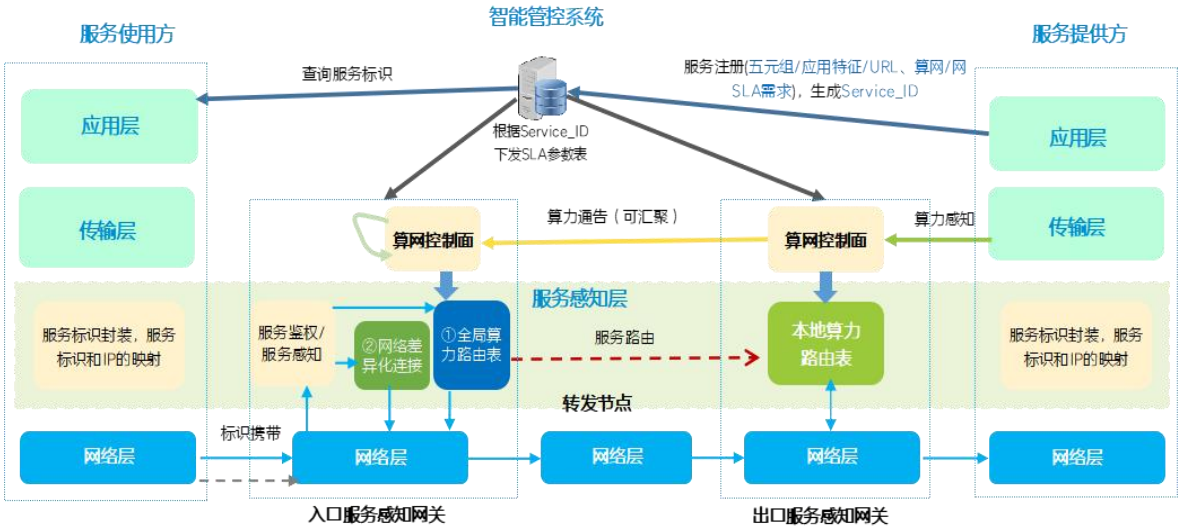

图 12 服务感知网络关键设计理念

传统分组数据网作为底层连接，为服务子层提供连接保障。服务子层与基础架构的交互依赖于服务标识的控制面索引，以实现终端用户与服务提供商的端到端高效交互。为了兼容现有设备并处理数据传输，SAN 在网络边界引入服务感知网关。

4.2.1 独立语义服务标识

SAN 引入独立语义服务标识， 其在用户、业务、算力和网络系统中扮演接口角色，对资源提供者和算网服务运营方而言，它是算网一体能力的承诺接口。服务标识由智能管控系统统一纳管（含注册/鉴权/校验/发布/订阅/策略配置），其生命周期涉及注册、发布、策略下发、订阅、服务请求（封装）、服务路由、服务分发、服务更新、服务撤销，这些流程都在运营系统的闭环治理流程中执行。服务标识拥有终端、网络、云算全局语义，所有接口均以服务标识为中心索引。在不同算网管理区域间，服务标识的互操作可经过协商、映射，

或标准化，取决于具体的部署和运营模式。基础服务通常会寻求跨域的标准化，而大部分服务标识治理则局限于单一管理域，则可以遵守服务标识企业标准。

独立语义服务标识有助于实现服务的发现、组合与重用，是构建可扩展、可互操作微服务架构的重要组成部分。

(1) 服务标识在端、网、云的语义

- 在终端，服务标识是用户对服务的请求接口。对用户而言，无需关心服务提供方及其位置，甚至无需关心服务参数，订阅服务并获取服务标识，向 SAN 发起业务请求。
- 在 SAN 网络域，服务标识是算网 SLA 策略的唯一索引，也是网络 UnderLay 服务策略的映射标识。
- 在云端，服务标识是云侧服务调度和路由的对象。

(2) 服务标识的设计原则：

- 服务标识在终端、网络、云端具有全局语义，以服务标识为中心无缝拉通业务、网络和云算系统，实现算网深度融合；
- 服务标识仅适用于基础通用的服务类型。服务标识的语义空间不是全覆盖，仅涵盖复用率较高的共性服务类型；
- 服务标识主要适用于对算网资源有高于“尽力而为”和“一般性计算处理”的服务类型，即 SAN 在 L3 层提供一体化算网统一服务；
- 服务标识支持类型聚合。聚合式标识结构有利于索引和查表效率提升；
- 服务标识可选支持同一应用下多种子业务的关联和同步，以及同一子业务会话的上下行数据流；
- 服务标识可选支持业务功能链的语义设计；

- 服务标识可选支持用户语义设计，以实现具体业务流量的精细化识别。

(3) 服务标识的封装：

- 主机侧方案下的服务标识封装。服务标识在 IPv6 扩展头或 SAN 专用转发头中封装，为兼容主流硬件的转发架构，推荐标识长度为 32，64，128 比特；
- 网络侧方案下的服务标识封装。服务标识重用目的地址字段，即使用 IPv6 地址结构标识服务语义，通过特定前缀唯一表征服务标识并区别于普通 IPv6 地址。

4.2.2 位置和归属无关

通过归属无关协议接口和统一服务命名体系实现泛在服务感知 IP 网络的演进。这一设计理念的核心在于对于网和业务，解耦应用与服务的物理位置绑定，确保用户能够随时随地请求并获取所需服务，服务标识语义本身无需关注服务提供者或其地理位置，后者将以服务标识关联属性和状态的模式在网络系统中运转和维护。

SAN 提出的泛在服务感知 IP 演进路径包括以下关键要素：

- 统一服务命名体系：引入逻辑标识符来表示服务，而非依赖于特定的 IP 地址或地理位置。这一命名体系允许智能控制面根据逻辑标识符自动解析服务的实际位置，并将请求导向正确的服务节点。
- 归属无关协议：负责建立服务连接，并管理和维护基于服务的连接状态。它提供面向服务的拥塞控制、移动性管理、保序、多路径/多归属以及内生安全等增强传输功能，使应用能够直接调用协议接口请求服务。

这一设计理念的核心优势在于它提升了服务的可访问性和移动性，优化了网络资源的利用，并简化了服务管理和运维，实现了无论服务提供者位置如何变化，用户都能无缝访问所需服务，资源分配根据实时需求动态调整，且通过逻辑标识符管理服务降低了管理复杂度。

4.2.3 端到端服务

基于服务标识的端到端设计理念通过解耦服务逻辑与实现细节，利用唯一标识符实现从用户到服务源头的无缝连接，极大提升了网络架构的灵活性、可管理和可扩展性，使用户无需关注服务位置或网络拓扑即可享受高效通信。

服务标识在端到端服务和调度上呈现如下增强功能：

- 打破网与云界限：通过服务标识作为统一索引，无缝连接网络域与计算域，使网络感知与调度超越传统维度，不仅优化南北向用户上云体验，实现东西向 AI 训练负载均衡调优，还能直连云内各子服务，强化 ServiceMesh 互联下的服务通信安全与控制力，此举简化服务治理与运维，提升微服务架构的灵活性和可扩展性，提供全面可观测性及追踪能力，加速 DevOps 与 CI/CD 流程优化，显著提高分布式系统的管理与运维效率。
- 实现端与网络协同：服务标识赋能网络感知端侧应用需求，确保精准匹配；同时，通过标识扩展网络节点信息（如出口队列深度、端口状态），反馈网络拥塞情况至端侧，实现动态调速与精准拥塞控制。

4.2.4 数据面轻量化

网络提供面向服务标识的路由和寻址功能。在数据面上，服务使用方携带服务标识发起服务连接请求，服务提供方则基于服务标识进行服务的请求侦听。网络边缘节点根据服务标识选择最优服务目的节点，以及对网络资源的编排，并执行对应的 SLA 策略。

服务感知网关通过简化的标识在报文中的传递，感知应用需求并为业务应用提供差异化的集成服务，减资源消耗，减轻数据处理压力，节省网络资源。

SAN 提供两种封装服务标识的方法：一是无主机变更的 Anycast IP，支持算网一体化，这种方法降低了对现有 IP 网络和主机协议栈侵入性；二是 IP 扩展头，支持算网一体化或差异化连接服务，该方法扩展性较好，可进一步支持端网协同能力。整体来说，SAN 通过智能化管理和灵活的标识策略，优化了服务交付和资源利用。

服务标识在数据面的主要功能：

- 应用及服务接入：携带服务标识的业务流量从服务终端传输到网络中，完成应用及服务的接入的端网控制。
- 服务层转发路由： 服务标识为索引进行标识路由调度，将业务流量调度到云算资源池或其他标识路由网关，以实现服务的转发和交付。
- 连接层 Underlay 路由：服务标识算力路由与传统 underlay 路由解耦，是服务层与连接层交互的接口。基于连接子层的路由功能，可在算力路由中实现各种增强型内生路由功能。

4.2.5 控制面极简化

基于 IP 网络的算力路由本质上是从一维升级到二维路由，理论上将导致乘数效应。SAN 采用 SDN 的控制分离和可编程优势，通过扩展 BGP 协议，在控制面融入算力信息，并扩展传统 TE（traffic engineering）到算网 TE 以支持多种灵活的调度策略（算力优先、资源优先、体验优先），支持分布式、集中式、混合式等算力路由方案，实现灵活的算力路由方案。

服务标识在控制面的主要功能：

- 服务注册：应用和服务通过此接口向网络或云计算资源池注册，传递服务属性，资源池据此生成基于服务标识的注册信息。

- 服务订阅/发布：连接控制管理单元与用户终端，终端通过此接口发送服务订阅请求。管理单元据此分配服务标识并回传给终端，终端将其嵌入业务流量报文中。
- 服务感知：1)控制管理单元与算网管理面间传递服务 SLA 需求，将需求与资源状态映射至服务标识;2)算网管理面反馈服务属性给控制管理单元，生成调度策略并与服务标识关联;3)算网管理面与控制面（或网关）间传递匹配 SLA 需求的调度策略;4)在集中式架构中直接从云资源池获取算力信息；分布式架构下，则通过出口网关获取算力信息。

4.3 SAN 参考架构

服务感知网络架构的核心要素是引入了网络层独立语义服务标识，并以此为接口感知算力和业务，即网络通过服务标识感知对应业务的算力状态，同时通过服务标识感知业务类型及其算网需求，基于面向服务标识的多维感知，由指定网络节点（如网络入口节点和出口节点）执行相应的路由和引流。服务感知网络仅需在特定网络节点维护服务标识转发表项，其他路径节点可按普通流量转发，无需识别服务标识，也无需维护状态。因此，服务感知网络是一种 OverLay 架构方案。

相对此前系列白皮书，本文在经典南北向业务流量场景的基础上，延伸覆盖了数据中心内和数据中心间的东西向业务流量场景。相应的，服务感知网络架构做了部分扩展和增强。如图 13 所示，服务标识依然是参考架构的核心功能要件，由于服务标识语义独立于 IP 网络既有的标识，如 IP 地址，端口号等，由扩展的服务标识转发面和控制面构成了逻辑上的服务标识子层。南北向业务流量场景下，业务节点分别为客户端设备和服务端设备，东西向业务流量场景下，业务节点延伸包含数据中心内的接入网关和代理网关。特别的，不同网络管理域之间，可以通过统一的服务标识进行一致性互通，也可以是独立标识体系，在域间执

行无损语义映射。

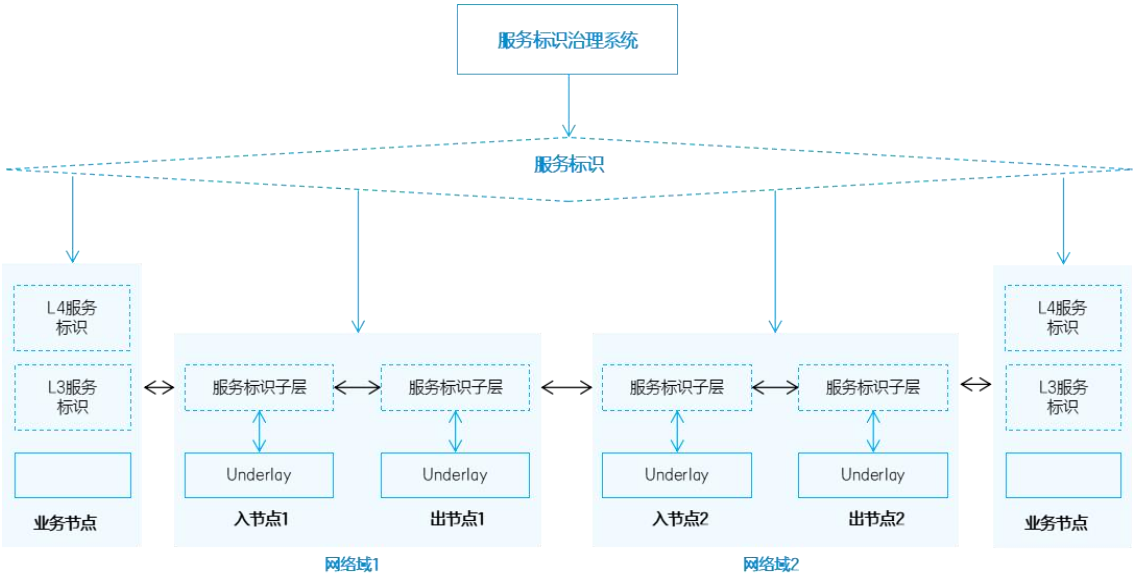

图 13 服务感知网络参考架构

4.4 SAN 与现有技术的 GAP 分析

4.4.1 基于 DNS 和 GSLB（全局负载均衡）服务调度模式

当前云服务产商对于算力资源的调度，大都采用 overlay 方案，如图 14 所示，通过服务域名(DNS)服务器解析为算力资源的 VIP 地址，同时采用统计学的方法对不同的 VIP 进行网络质量评估，选择出一个满足用户需求的 IP 地址反馈给用户，来实现算力服务资源池的选择，这种方式存在如下几方面的不足：

- DNS 地址在本地有缓存，当对应算力资源池或者网络存在故障时，本地缓存来不及及时更新，导致流量丢失。
- 服务首包通过 HTTP 连接到 DNS 服务器，当返回真实的算力资源 VIP 后，HTTP 重定向到真实的 IP 地址，存在一定的首包时延，对于时延敏感性的业务，满足不了客户的要求。

- 当前基于 DNS+GSLB 方式的调度，没有充分考虑网络和算力实时质量，当网络质量或者算力质量出现劣化时，整体业务体验出现明显下滑，但是服务的资源池未能及时切换。

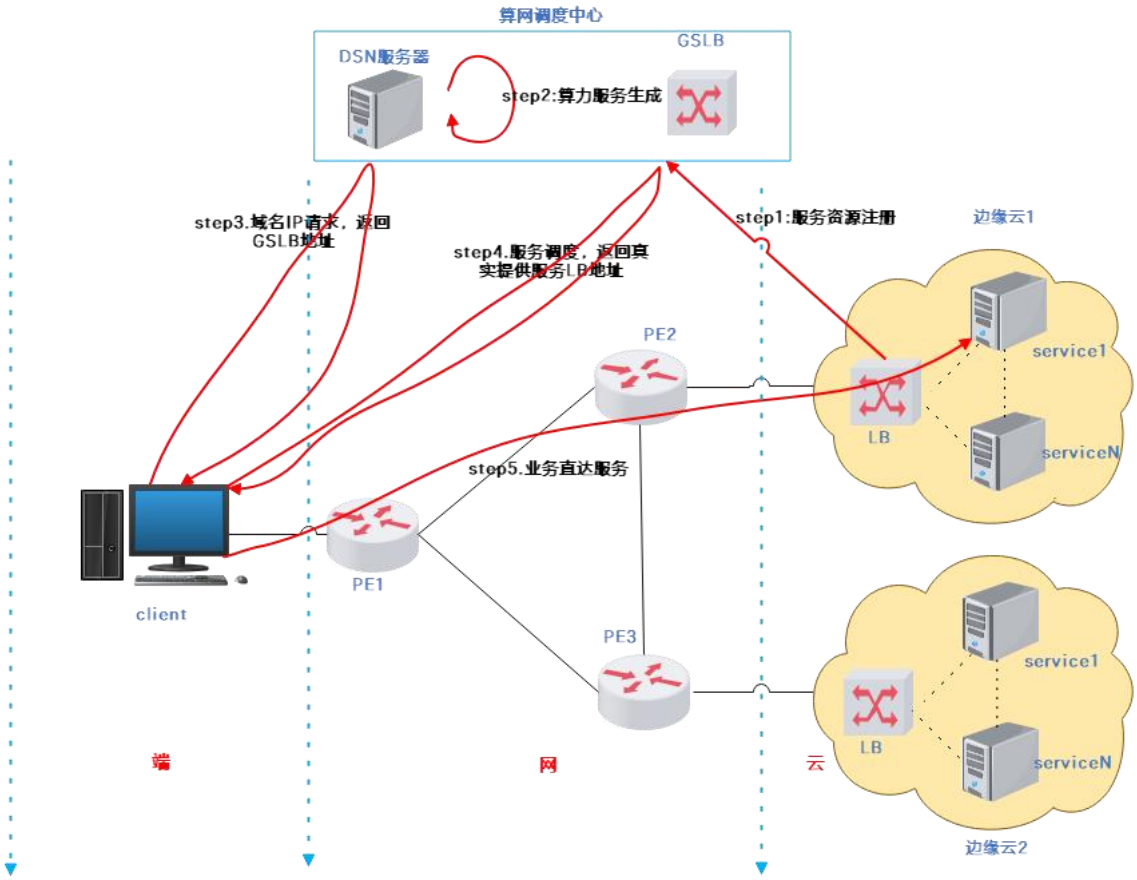

图 14 DNS+GSLB 的算力服务调度

4.4.2 基于 ICN 的服务调度

信息中心网络（Information-Centric Networking, ICN）是一种创新的网络架构，它将网络的焦点从传统的主机中心（Host-Centric）转移到了信息中心（Information-Centric）。这种转变意味着网络通信不再依赖于数据源或目的地的位置，而是依赖于数据内容本身。ICN 的设计理念主要是为了解决当前 TCP/IP 网络面临的一些挑战，如路由可扩展性、数据动态性、信息安全可控性等。ICN 的关键特点包括内容与位置分离，发布/订阅范式，内容命名与路由，内置缓存等。ICN 从架构上解决了移动性，服务调

度的失效性等问题，但是与 IP 网络存在诸多不兼容。

5 SAN 关键技术

基于 SAN 设计理念和参考架构，SAN 扩展增强网络技术，研究了 HFC 混合功能链、FSP 灵活调度策略、FS-OAM 全栈算网 OAM、端到端网络业务协同及基于统一标识的智能端网协同等关键技术，构建了 SAN 技术体系并研制了 SAN 样机。

5.1 HFC（Hybrid Function Chain）混合功能链

5.1.1 HFC 技术特征与内涵

伴随着应用和服务愈来愈复杂和多样化，一种对外服务中往往包含多个子服务协同及链式调度关系，而服务间互联意味着业务流量将经由网络通过多个服务端点。旨在为跨越多个服务端点和相应连接网络的对外服务提供全程与端到端的一致性服务需求保障能力，本文定义了一种混合功能链（HFC，Hybrid Function Chain）技术架构。

混合功能链（HFC，Hybrid Function Chain）的特性主要包含：

- 混合的服务功能类型与部署方式：基于部署态，服务与应用功能可以通过容器的形式部署在一个或多个集群，也可以基于虚机部署；服务与功能实例可以多实例部署，也可以基于 Serverless 架构动态申请与释放。基于运行态，微服务与单功能组成了多样化的对外服务，相应的，微服务与单功能对资源侧和网络侧也往往不同的需求。不同于传统网络领域服务功能链（SFC，Service Function Chain）中的服务都是是网络服务功能，HFC 中面向的服务功能也包括应用服务。
- 混合的服务功能间关系：服务与应用功能之间连接、请求、交互与协同方法往往呈现出多种形态，上下游服务与应用功能或微服务之间的协同与连接方式是多样化的，如单向

Notification 模式、双向 Request-Response 模式，并且单个上游服务可能请求单个下游服务，也可能同时需要调用多个下游服务。

- 混合的算网技术的应用：跨越多段连接网络，经历多个应用服务端点的 HFC 需要资源侧和网络侧高度协同，技术栈跨越应用和网络层。

5.1.2 HFC 系统架构

如图 15 所示，HFC 架构层次包括管理面、控制面和数据面。在以 Istio 系统层次和当前网络业务组件为例的当前云网控制面基础上，HFC 系统架构设计相应的补充与增量特性。在管理面中新增了服务分析和运营、服务和资源建模和服务编排和调度策略管理功能；在控制面中新增了服务注册和注入、服务发现和发布、服务路由生成和服务基础互联功能；在数据面中新增了服务标识管理、服务感知的转发和服务观测和保障功能。

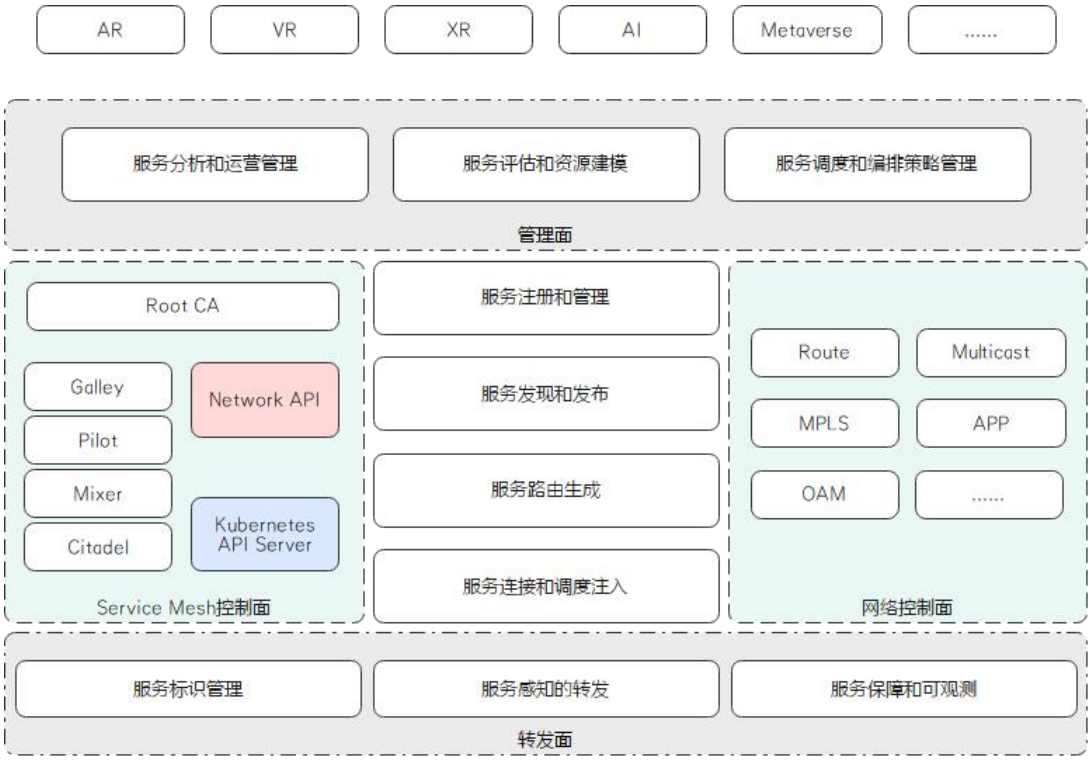

图 15 混合功能链（HFC，Hybrid Function Chain）架构

5.1.3 基于 SRv6 的 HFC 技术

SRv6 技术作为新一代 IP 承载协议，基于现有的 IPv6 转发能力，通过灵活的 IPv6 扩展头，实现网络可编程。而基于 SRv6 的能力扩展，可以将 HFC 的端到端服务路径信息、各服务感知网关 SID 信息或连接上下游服务的转发路径信息编码在报文中，从而实现跨越多个服务端点和相应连接网络的对外服务提供全程与端到端的一致性需求保障，并提供对应的算网流量工程与智能调度能力。

根据在始发服务感知网关处直接封装端到端服务路径或仅封装下一跳服务转发路径，可以将基于 SRv6 的 HFC 技术更具体地以紧耦合和松耦合的技术特征进行分类。

5.2 FSP(Flexible Scheduling Policy)灵活调度策略

5.2.1 支持多种约束条件的 TE 计算

基于 SAN-TE 统一算网调度架构，允许实现调度策略灵活选择和部署，兼顾集中全局优化（跨越多个区域、连结算力与网络）与快速收敛的分布式计算。SAN-TE 计算约束条件分为体验优先（如延迟、抖动和丢包）、成本优先（资源消耗和能耗）和效率优先（如资源利用率、均衡性）三类，各自在系统设计和运营中发挥着关键作用：

- 体验优先：这类约束确保服务质量（QoS），重点关注用户体验的关键指标，如延迟、抖动和丢包率。在算网一体化环境中，延迟是指数据从源到目的地所需的时间；抖动关注延迟变化的稳定性；丢包则反映网络传输的可靠性。通过优化这些指标，算网 SAN-TE 致力于提供流畅、无中断的服务体验。
- 成本优先：成本优先约束关注资源消耗和能耗的最小化。在资源有限的情况下，高效利用计算和网络资源对降低运营成本至关重要。此外，随着绿色计算和可持续发展目标的

日益重要，能耗成为 SAN-TE 设计中的另一个重要考量。通过智能化调度和优化策略，SAN-TE 致力于在满足服务需求的同时减少能源消耗和环境影响。

- 效率优先：这一约束集中在提高资源利用率和系统均衡性上。资源利用率高意味着系统能够更充分地利用现有资源，避免浪费；而均衡性则确保了系统负载分布均匀，防止局部过载导致的性能瓶颈。通过动态调整资源分配策略和优化网络流量管理，SAN-TE 力求实现整体系统效率的最大化。

综上所述，体验优先、成本优先和效率优先三类约束在 SAN-TE 中扮演着不可或缺的角色。它们不仅指导着系统的设计和优化方向，还直接关系到最终用户的服务质量和运营商的经济效益。在实际部署中，平衡这些约束之间的关系是一项挑战，需要综合考量业务需求、技术可行性和经济成本等因素。

5.2.2 业务调度和负载均衡技术

从业务视角出发，根据调度触发机制的不同，SAN-TE（SAN Traffic Engineering）的应用模型存在以下三种，其特点如下：

- 周期调度：该调度方式并不依赖于实时的业务流请求来触发。相反，它遵循预设的时间周期或特定事件（例如算力状态超出预设阈值）来自动执行。在这一机制下，算网 TE 会定期或在事件触发时计算面向特定服务标识的最优网络路径与计算实例，并将计算结果直接下发至转发平面。当业务请求首次抵达入口算力网关时，系统会根据服务标识进行匹配。一旦匹配成功，将进一步触发生成流亲和表，确保后续所有相关报文能够准确无误地被路由至同一计算实例，从而实现高效的数据处理与资源利用。

- 逐流调度：在逐流调度模式中，控制平面并不直接向转发平面推送详细的算网调度结果。相反，它仅下发服务标识表以识别特定的算力请求报文。当带有服务标识的业务请求进入算力网关时，报文会被上送至控制平面。随后，控制平面会基于当前的算力状态与网络状况，动态计算出最合适的路由路径或匹配已有的、符合需求的算力资源分配方案。计算完成后，相应的流亲和表将被生成并下发至转发平面，确保后续同一业务流的所有报文都能被导向至同一计算实例进行处理，以实现精准的业务流亲和与资源优化分配。
- 混合调度：在同一个支持多种服务的算网调度系统中，根据业务特点不同，基于服务标识定制调度方式，根据调度方式配置来灵活选择周期调度或逐流调度。通过智能化的路径计算与资源分配策略，提升算力网络的服务质量和效率。

基于以上三种调度策略，SAN 汲取了多种模式的优势，还创新性地引入了一种全新的负载均衡机制。这一机制旨在缓解 TE 计算带来的局限性。其主要特点包括：

- 控制面 TE 计算：通过一次 TE 计算，系统能够自入口网关至所有满足算网 SLA（服务等级协议）要求的服务实例及其可达路径进行编排，并计算出相应的分担比例。这些信息被整合进算力路由表中，形成了一系列可行的下一跳选项。
- 转发面 TE 转发：利用首包报文中的五元组或三元组 HASH 值确定多下下一跳路由表中唯一下一跳并引流和转发。这一过程同时确保了同一会话流中所有后续报文的连续性和一致性，实现了算力网络服务请求在会话级别上的精准负载均衡并大大降低控制面负载。

通过结合这一负载均衡机制与多种调度模式灵活组合，SAN 不仅保障了用户会话质量达到算网 SLA 的要求，还实现了算网资源的细粒度均衡利用。具体而言，在连续两次更新算力路由表的时间周期内，它有效避免了单一计算或网络资源因过度占用而导致的“极化”

现象。同时，该机制显著减少了因算力实例状态频繁变动对控制面造成的额外压力，从而确保了系统的稳定性和效率。

总之，在提升用户体验的同时优化资源分配是 SAN 的核心目标之一。通过智能调度策略和负载均衡技术的应用，SAN 有效地避免了资源过度集中或分散所带来的问题，在算网环境中展现出强大的适应性和高效性。

5.3 FS-OAM(Full-Stack OAM) 全栈算网 OAM

5.3.1 网络和计算可观测现状

可观测性是指对监控对象的状态、性能指标和事件的监控、测量、追踪和分析的能力。随着云计算、大数据以及 AI 技术和应用的迅猛发展，数据中心作为支撑各类关键业务运行的基础设施，其规模的快速发展带来了网络规模的显著扩大，网络架构日趋复杂。在这种背景下，可观测性变得尤为重要，它不仅关系到故障的快速发现、定位、和修复，也关系到性能优化，更直接影响到业务的连续性和服务质量。

5.3.2 FS-OAM 关键目标和技术

传统的网络侧与计算侧可观测性技术长期以来独立发展，互不兼容，缺乏必要的互操作性。面对日益显著的算力与网络融合趋势，构建一个全面覆盖、端到端的监测体系成为迫切需求。为此，整合网络域和计算域的运维管理（OAM）能力显得至关重要。

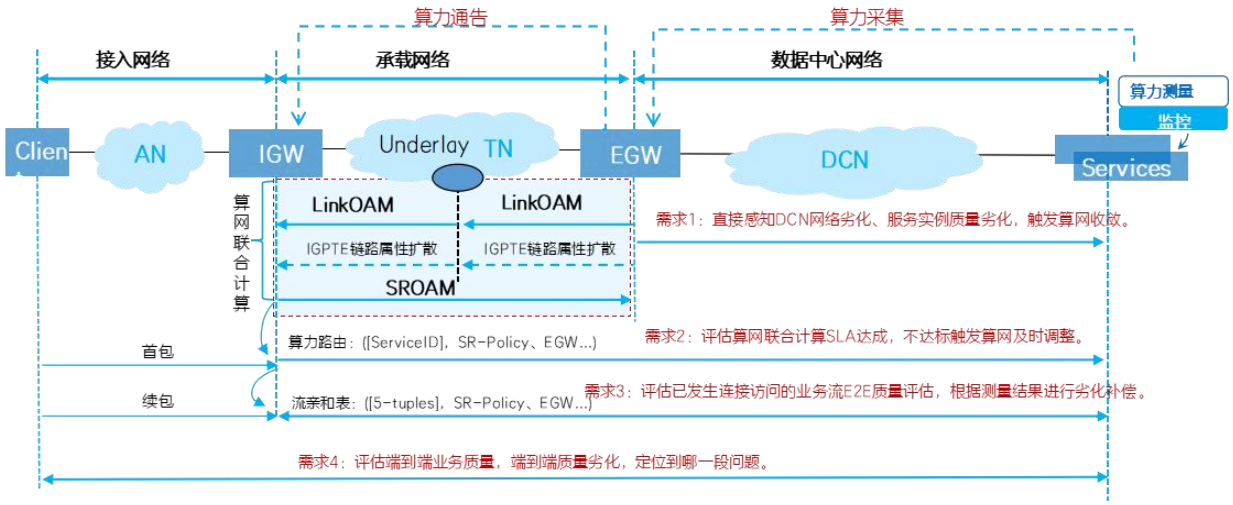

图 16 算网 OAM 需求和目标

SAN 提出 FS-OAM(Full-Stack OAM) 全栈算网 OAM，主要包括检测连通性故障、丢包、时延等指标，还包括新定义的算侧指标（连接数、资源占用等），并可以逐段、逐层覆盖， 特别关注从算力网关到计算实例的性能监测与优化，支持快速的路由收敛机制和业务质量保障。同时，它还具备强大的故障定位能力，以及高度的实用性和可扩展性。如图 16 所示，当前算网一体存在四类关键需求：

- 需求 1--加速算网控制面收敛：为快速响应计算实例状态的变化，当前控制面感知机制在处理速度上存在局限，难以跟上算力实例状态的迅速变动。因此，亟需在数据平面实现快速监测功能，对出口算力网关到算力实例端到端状态进行实时监控。一旦检测到故障或性能劣化，将立即触发算网流量工程（TE）的计算收敛机制，采取行动避免服务黑洞现象，确保资源的有效利用与快速恢复。
- 需求 2--算网 SLA 评估闭环保障：为了确保服务等级协议（SLA）得到持续满足，有必要从算力路由表中提取与网络到计算实例路径相关的测量数据。通过 TE 计算得出的算力路由信息需经过验证，确认其是否符合既定的算网 SLA 标准。

- 需求 3--业务流 SLA 闭环保障：服务访问激活后，后续数据包通过流亲和表转发至同一计算实例，期间计算与网络性能可能波动。为确保稳定高质量的用户体验，需持续监控从网关到实例的业务流性能。依据算网 SLA 要求，在网络层适时进行修复或调整，以维持服务稳定可靠。
- 需求 4--业务故障定界和排障：当用户体验下降时，迅速准确定位问题是关键。这需要快速识别从用户终端到计算实例全路径上的故障点。传统 OAM 技术仅限于传输层监控，应扩展至应用层以实现全面监控，并据此进行端到端的故障定位与排除。增强的 OAM 机制能更精准地识别故障源头，缩短恢复时间并提升服务质量。

FS-OAM 面向算网一体四大类需求提出综合解决方案，主要包括以下三个技术层面：

1. 针对需求 1 和 2，FS-OAM 分别提出算网实例 OAM 和算力路由级 OAM，相应的技术考虑如下：

   - **检测协议的选择与应用**：可选用 PING、TWAMP、OWAMP、STAMP 等协议或其组合进行网络到计算节点健康状况检测，建议优先采用反射型策略以提高检测效率和准确性。
   - **检测协议部署**：除了支持网络侧部署，在算侧支持上述检测协议的部署与运行，确保能够实时监测并及时响应各种变化和需求。
   - **检测新能力拓展**：为了全面覆盖算力网络的独特指标，应积极探索并引入新型检测协议，确保能够准确评估计算资源的状态和性能。

2. 针对需求 3，FS-OAM 提出算力会话级 OAM，相应的技术考虑如下：

- **检测协议的选择与应用**：为确保全面覆盖真实的业务流并获取精确的数据分析，我们应充分考虑采用随流检测技术，如 IOAM、IFA 等。
- **检测协议部署**：在算力网络中，从网关入口至计算实例的整个传输路径上，推荐网络节点和计算实例部署。
- **检测新能力拓展**：扩展覆盖算力网络的独特指标，比如计算时延、负载等增量信息。

3. 针对需求 4，FS-OAM 提出应用级 OAM，相应的技术考虑如下：

- **检测协议的选择与应用**：存在两种路线（1）为了实现高效且低开销的端到端业务性能监测，可基于创新的轻量级端侧协议栈利用探针机制，确保对业务流程进行全面而精准的追踪与分析；（2）在跨端、网、云的环境中，采用随流检测技术延伸到算侧，实现对传输业务的更加精细化监控定界。
- **检测新能力拓展**：将云内可观测性与网络随流检测紧密结合，具体措施是在随流检测头部携带 TraceID 和 SPAN ID,结合 eBPF 将助力我们实现从用户终端到云端应用层的全程性能评估与故障定位，有效提升运维效率及用户体验。

总之，FS-OAM 延伸到算侧并定义的四个层次（实例级、表项级、业务流级、应用级）和关键技术，特别说明的是，应用级 OAM 不仅打通了网络和算侧，还实现网络 OAM 和 eBPF、分布式追踪技术结合，实现端到端业务故障快速定界，实现算力网络“悬丝诊脉”，解决未来算网一体化运维的关键痛点。

5.4 端到端网络业务协同服务系统

5.4.1 基于 SAN 的增强型业务控制层

具有业网协同能力的业务控制层是一种具能感知应用需求并针对现有网络资源进行资源编排的增强性业务控制中间层系统，基于此层能实现多种网络服务能力对上层应用的开放。在本文中，基于业网协同工作域中最核心的业务感知能力，本章节通过综合上述业网协同功能层的核心能力与上下游系统的关系，展示了一种融合了 SAN 功能组件的增强型业务控制功能系统框架方案，如图 17 所示：

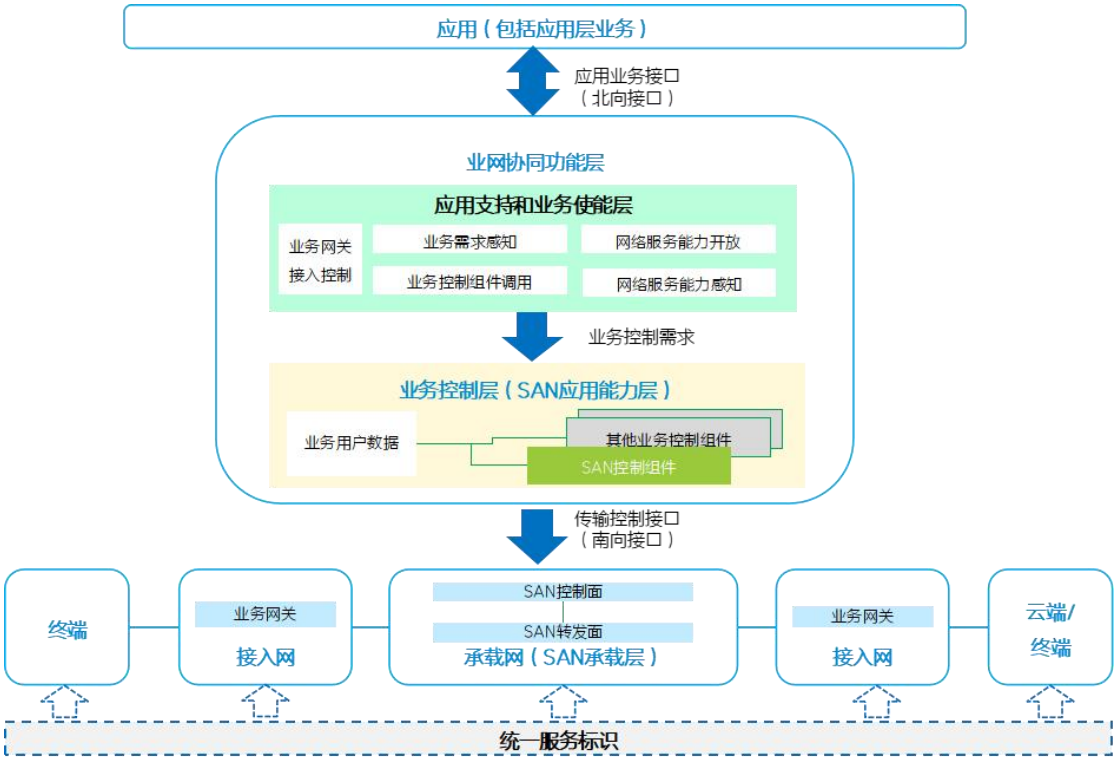

图 17 融合 SAN 的业网协同层的功能框架

如图所示，整个业网协同功能层的逻辑功能架构需要有两个子功能单元合作完成，包括了应用支持和业务使能功能（服务使能层），以及业务控制功能（业务控制层），并配合底层 SAN 网络通过统一的服务标识完成端到端的差异化数据转发保障服务。如图 18 所示：

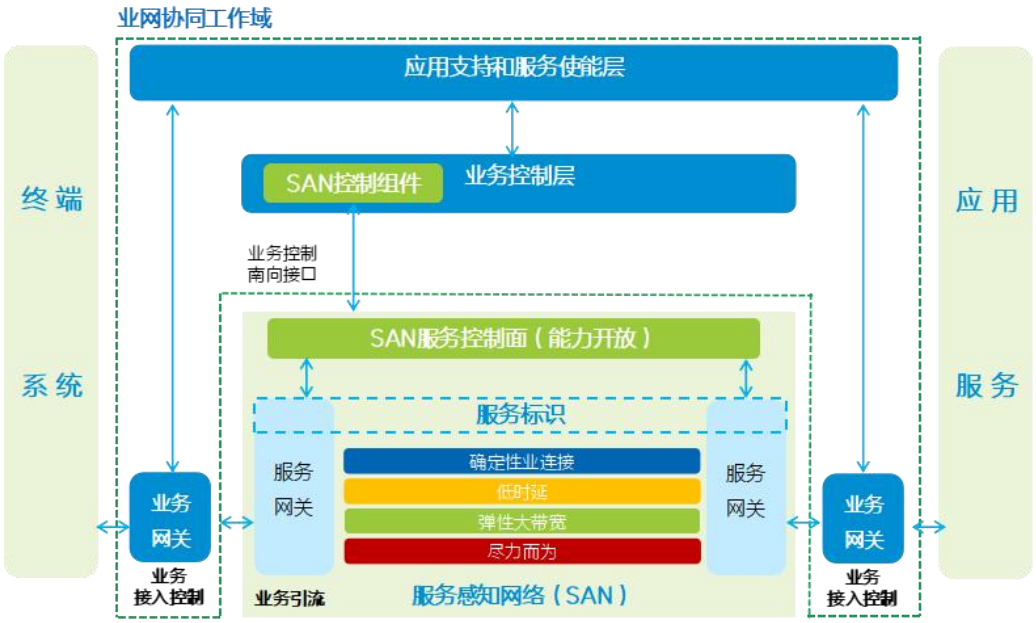

图 18 业网协同功能层和 SAN 网络的系统工作模式

业网协同功能层和 SAN 网络的协同过程中，SAN 网络系统需要将其服务能力通过业务控制组件的方式注册到业网协同功能层的业务控制单元中，并通过业网协同功能层的南向接口提供其可以调用的 SAN 服务标识（统一服务标识）。对于具有不同传输需求的业务流分类分级，通过对于需求的感知来确定需要提供 SAN 服务的业务流，并调用 SAN 服务组件来提供传输的保障服务。业网协同功能根据用户的签约信息和 SLA 需求来为用户匹配合适的接入节点，在建立业务控制会话的过程中维护用户信息，并提供增强性的传输协议，利用 SAN 网络的服务标识来确保数据传输过程中对应业务流的数据包的转发可以按照既定的转发策略来进行转发控制，为用户提供差异化的数据传输服务。

5.4.2 业网协同功能层的接口需求

业网协同功能层需要实现两类外部接口：

- 与业务应用的接口-北向接口

北向接口是网络提供商提供给上层业务应用的一种网络服务能力感知和调用的网络业务 API 接口。通过该接口，上层应用可以发现并选择订购某一类网络服务，并且可以通过

该接口将业务所需的网络连接和数据传输需求通知到业网协同功能，进而可以对可用的网络资源和能力进行调用或者预留。

业网协同功能需用能够提供一些属性来描述业务，定义业务的具体性能和参数。这些参数主要包括诸如业务类型，业务状态，业务问题等等可以对业务进行分类分级的因素。

该接口需要针对业网协同增强控制层设计新的应用接口 API，并需要提供通用化的业务描述信息格式。

- 与传输层的接口-南向接口

南向接口是用于业网协同功能层将业务需求对应的控制信息，通过相应的业务控制组件调用并下发给下层的承载网络，包括网络资源调用策略，业务流传输策略，安全管控策略等，完成对于业务逻辑控制和传输资源匹配调用的功能。

该接口可以通过扩展现有的业务控制功能和网络传输控制功能之间的接口实现。

5.5 基于统一标识的智算端网协同

5.5.1 基于统一标识的智算端网协同架构

基于统一标识的智算端网协同总体架构如图 19，其中：

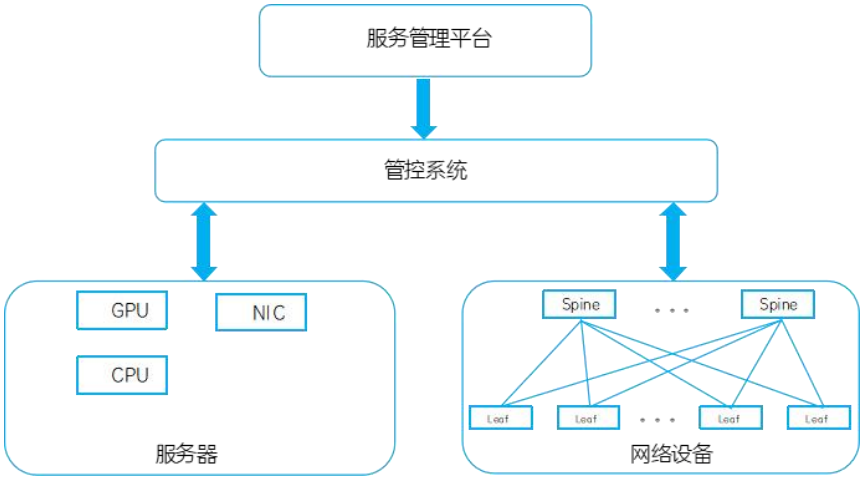

图 19 基于统一标识的智算端网协同总体架构

服务管理平台：维护业务流量特征和服务需求信息，在引入新的业务流量或服务需求后，需要向服务管理平台进行注册。

管控系统：纳管服务器和网络设备，感知网络侧服务提供能力。通过北向接口从服务管理平台获取业务流量特征和相应服务需求，并为服务需求分配服务标识。通过南向接口基于服务标识进行分别在端网进行规划配置，以满足相应的服务需求。

网络设备：网络接入设备接收管控系统下发的基于服务标识的报文处理策略，识别服务标识对于业务报文进行处理。

服务器：服务器网卡根据管控系统配置，在发出业务报文时，携带相应服务标识。

5.5.2 基于统一标识的多路径控制

管控系统根据不同的路径质量需求，在网络侧基于不同服务标识进行具体的路径规划和控制策略下发。网卡根据业务对于质量保障需求，在资源池中选择相应服务标识进行封装。网络接入侧设备识别接收报文中的服务标识，将业务流量引入到相应路径。网络侧包括负载均衡方式，精确转发路径等详细内部信息，均对端侧屏蔽。在网络拓扑、链路、配置等发生改变后，管控系统基于最新的网络状态，对各服务标识对应的路径进行重新调整和规划，在端侧不感知的前提下，维持网络对外的服务保障能力。

通过统一标识传递端侧路径控制需求，满足了端网解耦需求，由网络侧基于需求标识进行路径控制，发挥网络侧原生优势，也降低了端侧的复杂度和方案部署成本。

5.5.3 基于统一标识的拥塞控制参数适配

面对端侧不同的拥塞控制算法的需要网络反馈的拥塞指标参数需求，管控系统基于端侧需求和网络侧的对于各拥塞指标的检测能力进行提前匹配，为不同的参数组合需求分配服务

标识。网卡根据自身对于拥塞指标参数的需求，选择服务标识，封装在用于拥塞检测的报文中。网络接入侧设备负责识别标识，并根据管控系统预先下发的服务标识与网络参数需求的映射关系，将网络实际需反馈的相应字段写入拥塞检测信令中传递，以完成拥塞指标的收集。

通过统一标识在网络接入侧设备完成参数转换的方式，提升了网络侧对不同拥塞控制算法的兼容性，也满足了控制网侧关联设备范围的需求。

6 SAN 样机及测试总结

6.1 SAN 算力路由样机试点

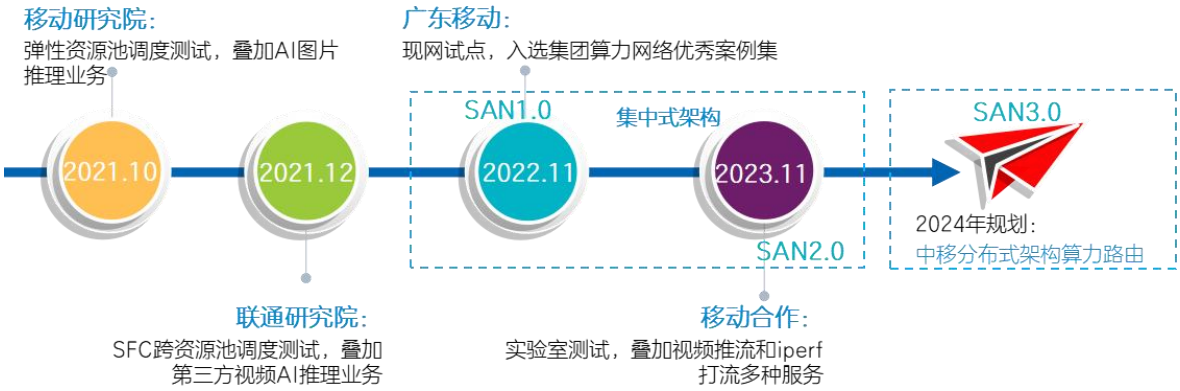

图 20 SAN 样机实践历程

围绕 SAN 的关键理念和技术，中兴通讯展开了一系列样机开发和试点测试项目。如图 20 所示，宏观来看，分为三个阶段：

第一阶段（2021 年）：这是 SAN 技术旅程的起点，我们主要专注于算力路由领域的关键技术原始积累，成功地实现了算力路由行业的早期原型开发与验证工作。这一阶段标志着 SAN 技术的萌芽。

第二阶段（2022~2023 年）：基于 SAN 的核心原则——独立语义服务标识，我们深入探索并构建了集中式算力路由样机。在这一阶段的技术演进中，“算网联合感知”、“算网一体资源效率提升”成为 SAN 1.0 的关键标签；而 SAN 2.0 则进一步拓展至“算网端到

端 SLA 保障”、“算网资源均衡”、“层次化路由”与“多种调度策略”（涵盖体验类与资源类调度）。我们的努力获得了业界的高度认可，在 2022 年 8 月 1 日的首届算力大会上荣获“创新先锋”奖，并在 2023 年 5 月的云网智联大会上摘得“最佳实践案例”奖。此外，在 2024 年 2 月 26 日的世界移动通信大会（巴塞罗那展）上，中国移动与中兴通讯联合发布了全球首台算力路由器。

第三阶段（自 2024 年起）：遵循 SAN 的全量设计原则，我们正致力于研发集成了集中式、分布式与混合式架构的算力路由样机。一方面，我们积极参与由中国移动牵头制定的互通标准，旨在促进不同网络之间的协作与融合；另一方面，我们也在积极拓展业务场景，探索算力路由技术在更广泛领域的应用潜力，以期最大化其商业价值。未来的发展令人期待，我们相信这一领域的创新将不断推动行业向前发展。

中兴通讯与中国移动携手，在算力网络领域开展深度合作。通过创新的算力路由机制和随流调度算法，实现了服务标识驱动的资源优化配置，显著提高了业务响应速度与转发效率。引入资源动态感知机制与多策略协同框架，有效提升了算网资源利用率和业务承载能力。借助 SAN1.0 与 SAN2.0 样机，双方成功将现网设备升级为算力路由网关。特别是在 2024 年 8 月江苏的 CDN+算力路由试点项目中，基于独立现网 CDN 资源的 SAN2.0 验证了集中式算力路由方案的关键技术和预期效果。

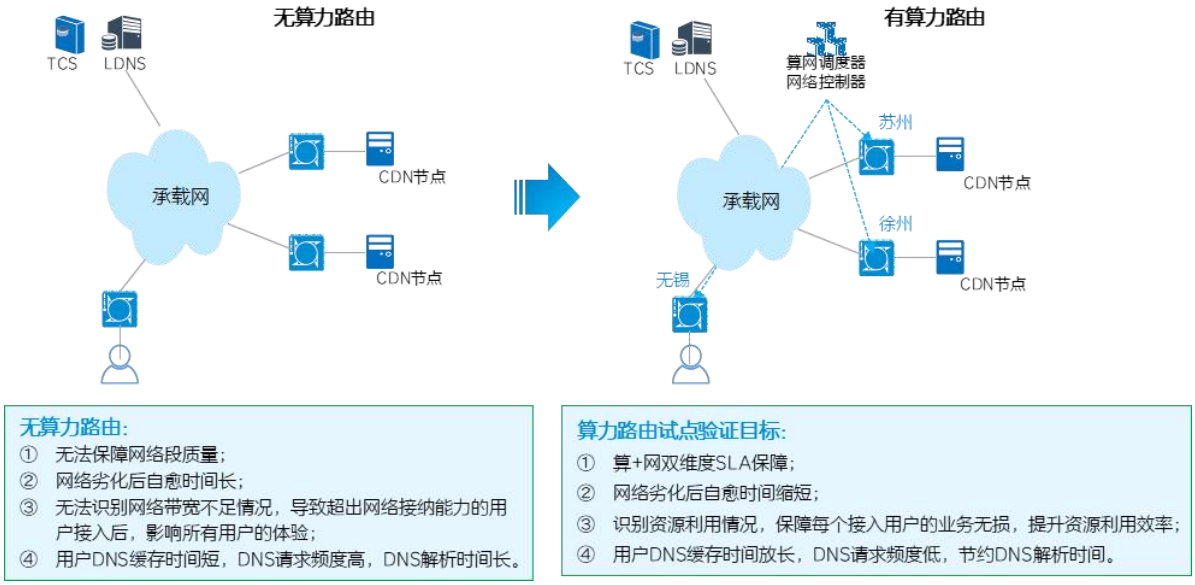

图 21 SAN2.0 样机试点组网和目标

测试环境介绍：本次测试方案及环境部署如图 21 所示， 在无锡、苏州、徐州三地协同，其中无锡到苏州和徐州两个方向，分别部署 2 条 10GE 传输专线，苏州和无锡分别对接独立的 CDN 资源服务器，用户从无锡发起 CDN 业务访问。

测试结论：

- 业务自愈能力：在网络质量下降的情况下，系统自动触发自愈机制。具体表现为：当检测到网络劣化时，用户的操作将重新发起链路建立请求。此时，新会话会被智能调度至其他可用的 CDN 节点，确保视频播放过程不受影响，维持流畅体验。
- 资源利用效率显著提升：通过优化调度策略和资源分配算法，实现了对更多用户访问需求的有效保障，尤其是在保证用户体验的前提下。直接促进了 CDN 资源利用率的跃升，相较于无算力路由提升了 35%以上。
- DNS 解析效率优化：针对 DNS 解析这一关键环节进行了深度优化，成功将平均解析时间缩短约 30 毫秒，大幅减少了用户等待时间。

综上所述，本次试点项目的结果完全符合预期目标：不仅显著降低了广域网服务响应时延，极大改善了终端用户的实际体验；同时，在算网感知技术的有力支撑下，实现了算网资

源利用率、利用率均衡度以及业务承载量的显著提升。这些成果充分验证了方案的有效性和先进性。

6.2 SAN 服务治理测试验证

如果说在以 Kubernetes 为基础构建起的云原生世界里，哪种设计模式最为经典，Sidecar 模式无疑是其中最有力的竞争者。当需要为应用提供与自身逻辑无关的辅助功能时，在应用 Pod 内注入对应功能的 Sidecar 显然是最 Kubernetes Native 的方式，主要实现手段则是通过在应用旁部署一个 Proxy，在 Kubernetes 场景下则为应用 Pod 注入 Sidecar，拦截应用流量至 Sidecar。Sidecar 根据获取的用户配置对应用流量进行处理，以一种对应用代码几乎无侵入的方式实现了服务治理。

而随着 Service Mesh 的落地规模不断扩大，传统 Sidecar 模式在云原生环境中的诸多挑战，如侵入性、资源利用率低及生命周期绑定等问题。针对以上缺陷，Sidecar-less 成为业界关注的焦点，而在其中 Ambient Mesh 已成为一个重要的方向，Ambient Mesh 实现了对应用的零侵入和独立演进，同时优化了资源占用，提高了整体性能。随着 Istio 社区对 Ambient Mesh 的持续投入和实验特性的逐步稳定，这一模式有望在未来成为云原生服务治理的重要选择，推动大规模生产环境的网格技术落地。

然而，值得注意的是，尽管 Ambient Mesh 在性能上有所突破，但它目前仍然依赖于基于 L4/L7 的代理机制来实现多级服务之间的互联。在跨域多级服务互联的场景中，这种机制可能会暴露出性能上的局限性。具体来说，随着服务层级的增加和跨域通信的复杂化，基于 L4/L7 的代理机制在处理多级服务互联时，每一级的代理都会引入额外的开销，这些开销会累积并呈现倍数增长的关系。这不仅会增加网络延迟，还可能对整体服务的性能和稳定性产生不利影响。

SAN 不仅能实时根据算力（服务负荷、算力资源利用等）和网络（网络拓扑、链路状态、带宽使用情况等）关键信息进行综合调度，实现高效、稳定地连接不同资源池的云网络。同时与 Ambient Mesh 模式相比，SAN 在跨资源池服务互联时，减少了两次 Socket 连接和两次 L7 代理处理，直接提升了服务请求的响应速度，使网络运行更加高效。这一改进不仅增强了云网络服务的灵活性，还提高了系统的稳定性，为用户提供了更加可靠的服务体验。

基于以上思路，对集成 SAN 的云网络跨资源池服务互联方案进行了评估，并与 Istio 的 Ambient 模式和 Sidecar 模式进行比较，具体说明如下：

对于集成 SAN 的云网络跨资源池服务互联方案测试，测试环境包含三个云集群，其中每个云集群由一个主节点和三个工作节点组成，每个节点都运行着 Ubuntu 20.04 系统，并配置了 8 核 CPU、16GB 内存以及 30GB 硬盘存储资源。

对于 Istio 的 Ambient 模式的测试，测试环境同样包含三个云集群，其中每个云集群由一个主节点和三个工作节点组成，每个节点都运行着 Ubuntu 20.04 系统，并配置了 8 核 CPU、16GB 内存以及 30GB 硬盘存储资源。同时，云集群还配置了 Istio 的 Ambient 模式。Isito 的 Sidecar 模式测试环境与 Ambient 模式测试环境类似，区别是云集群配置的是 Istio 的 Sidecar 模式。

在 SAN 集成方案测试过程中，我们在 SAN 方案测试每个集群的其中一个节点中部署了具有明确顺序依赖关系的子服务，每个子服务在任务完成后会将结果返回至子服务请求方，服务时间不计入统计。在 Cloud A 中，我们还部署了入口网关用于传递用户请求，其中入口网关与子服务部署在不同节点上。测试的请求传递与结果返回过程如图 22 所示。测试节点（User）首先借助 Cloud A 中 Node A 的 Ztunnel 代理向 Cloud A 的入口网关发起请求。随后，请求根据路由规则按照既定的顺序经过 SAN 网关和其他 Ztunnel 代理依次传递给 Cloud A 的 Service1、Cloud B 的 Service2 和 Cloud C 的 Service3。每个服务在接

收到请求后，都会完成其分配的任务，并将请求结果按原路径返回。

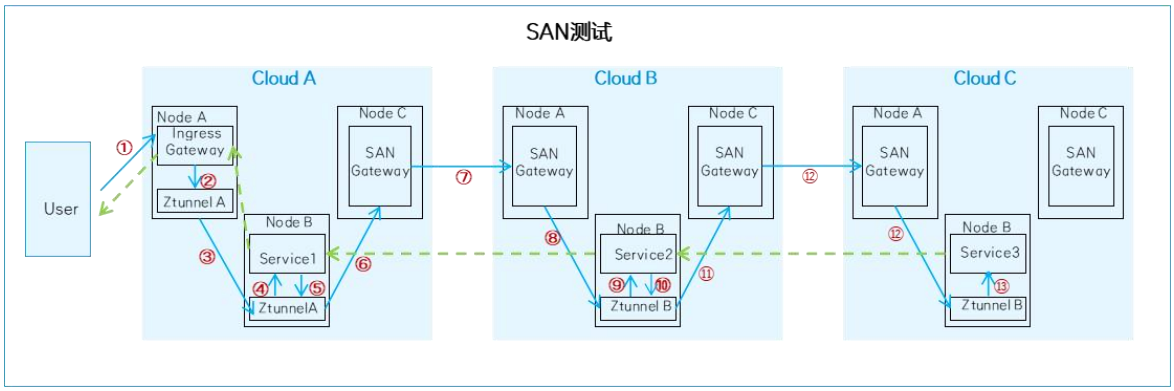

图 22 集成 SAN 的服务互联测试方案

在 Istio 的 Ambient 模式测试过程中，我们在三个云集群中分别部署了入口网关、出口网关与 SAN 测试过程中相同的子服务，同样对服务时间不计入统计，其中子服务被部署到与入口网关和出口网关不同的节点上。测试的请求传递与结果返回过程如图 23 所示，用户测试节点（User）通过 Cloud A 中 Node A 的 Ztunnel 代理向该云集群的入口网关发出请求。随后，根据预先配置好的路由规则依次经过其他节点的服务和代理与各个云集群的入口网关和出口网关最终完成用户的服务请求，并将请求结果按原路径返回。Istio 的 Sidecar 模式测试如图 24 所示，其过程与 Ambient 模式的测试相似，这里不再赘述。

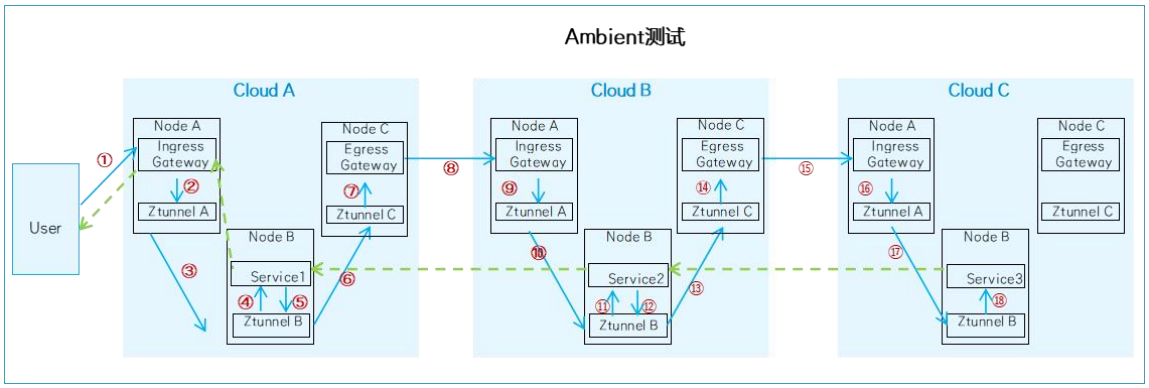

图 23 Istio 的 Ambient 模式服务互联测试方案

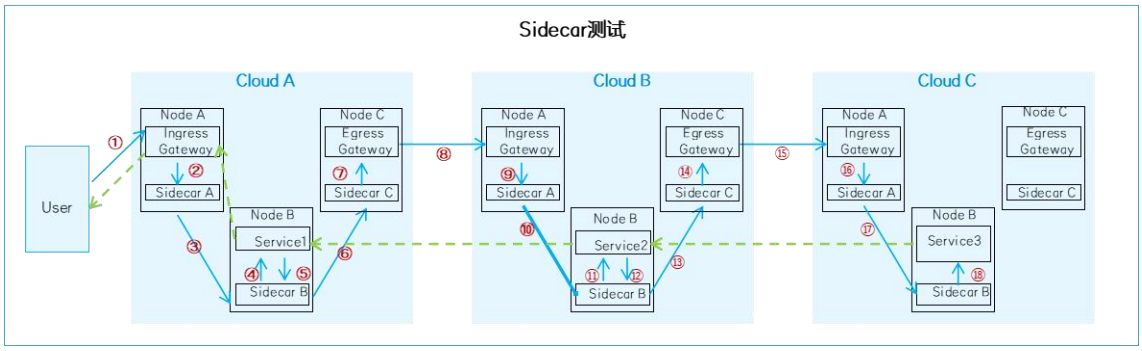

图 24 Istio 的 Sidecar 模式服务互联测试方案

我们在每次测试过程中通过用户节点顺序发送了 5000 次的服务请求。图 25 为各方案中测试过程服务完成时延分布图。可以看出，集成 SAN 方案的整体性能要明显高于 Istio 的 Ambient 模式和 Sidecar 模式。如图 26 所示的平均服务完成时延图，集成 SAN 方案的平均服务完成时延较 Istio 的 Ambient 模式和 Sidecar 模式分别提高了 29.5%和 37.1%。

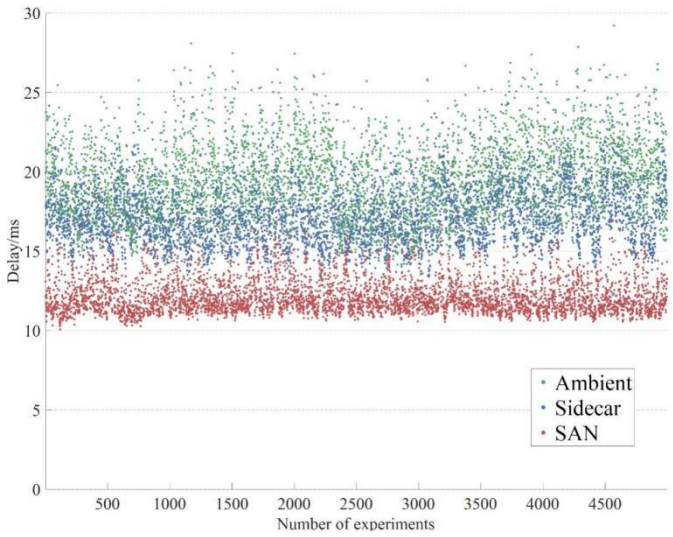

图 25 不同方案服务完成时延分布

图 26 不同方案平均服务完成时延

7 技术及产业展望

随着国内外算力路由标准的陆续落地，以及国内运营商的积极试点探索，算力路由已经进入产业规模落地的前夜。算网资源利用率在算力路由方案智能调度的支撑下得以大幅提升，已经得到初步试点验证。对业务而言，算力路由的内生功能优势在于服务会话发现的数量级时延改进，以及网络层业务的无缝热迁移，这都是现有网络架构完全无力支持的。随着高效交互型业务（如云游戏）及无状态云原生业务的逐步部署，传统应用层的部分功能下沉至网络由算力路由执行硬件级转发，势必将释放出巨大的功能和性能红利，从而使能更多新型业务。

网络对精选业务定向感知，并据此提供精细化网络连接服务，已成为行业初步共识。由于涉及应用、业务和网络的复杂生态，行业中涌现出多样化技术解决方案，其共性特征和需求是在网络中引入业务标识，使能网络对业务的精细化感知和连接服务。从互联互通的视角看，进一步凝聚需求和方案共识，推动行业统一标准方案的收敛，符合产业的共同利益，势必成为行业共同努力和推进的方向。

HFC 有望使能云原生规模化部署和全新业务体验和业务模式。HFC 技术为跨越多个服务端点和相应连接网络的对外服务提供了微服务解构背景下的全程与端到端的一致性需求

保障，并提供对应的算网流量工程与智能调度能力，有望使能云原生规模化部署和全新业务体验和业务模式。云上部署应用将可以分解为更细颗粒度的原子服务，原子服务与功能可以进一步地以按需的方式就近在边缘精细化和智能化部署。通过动态唤起和调用就近的边缘服务，进而利用 HFC 的穿越算网基础设施的服务保障与调度能力，更加有可能为用户提供有保障的极致服务体验；另一方面，动态的实例管理可以更好地提供算网一体基础设施的资源利用率。HFC 的算网一体可视也将为应用与服务提供更加全面和客观的管理视角。

基于端网一体的服务标识在支持未来高性能智算 DC 网络的基础服务能力–AI 大模型的训练和推理部署的不同过程中都能够发挥重要作用。大型语言模型(LLM) 如 ChatGPT、SORA 等需要很多的 GPU 卡(万卡或 10 万卡级别)，建设和运营成本很高，通常由大型数据中心进行训练任务。而且很多小型模型，如科学领域和专业领域，由于数据的规模、类型以及模型中高质量数据的版权和隐私性等，通常不需要很多的 GPU(如百卡级别)，可以被部署到更靠近用户的边缘侧。大语言模型能够返回更多通用类型的答案，而小语言模型通常返回特定领域内的答案。因此，在满足复杂 AI 训练和推理任务的场景下，需要采用分布式模型并行加速训练、推理互联模式。分布式大模型的核心本质在于通过开放泛在的算力资源，分担巨大的算力成本和功耗的开销。而基于端网一体的服务标识互联网络则能够提供这样灵活的分布式模型部署，通过网络互联完成更大模型的训练。不同的模型服务实例根据需求可以部署在不同的云服务节点，并且能够保证模型实例间的算网一体资源保证，保证应用的可靠性和高性能。

8 参考文献

[1] 中兴通讯股份有限公司. IP 网络未来演进技术白皮书. 2021.06

[2] 中兴通讯股份有限公司. IP 网络未来演进技术白皮书 2.0——开放服务互联网络. 2022.09

[3] 谭斌，黄兵，黄光平. 面向算网一体的服务感知网络. 中兴通讯技术(简讯), 2022,09

[4] 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)

[5] 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, 28(1)

[6] 吉晓威. 边缘算力网络架构及实践.中兴通讯技术(简讯), 2022, 07

[7] 段晓东，程伟强，王瑞雪，王雯萱. 面向新型智算中心的全调度以太网技术[J]. 中兴通讯技术, 2023, 29(4)

[8] 付华楷，黄光平. 服务感知网络技术和演进探讨.中兴通讯技术(简讯), 2024, 06

[9] 郭胜楠，雅承，庞冉等. IP 承载网络技术演进方向研究[J].邮电设计技术, 2024(4)

[10] 雷波，赵倩颖，凌泽军. 算力网络实现一体化服务的探索与实践[J]. 中兴通讯技术, 2021, 27(3)

[11] N Katta，A Ghag. Clove: Congestion-Aware Load Balancing at the Virtual Edge. ACM, 2017

[12] 雷波，陈运清等. 边缘计算与算力网络——5G+AI 时代的新型算力平台与网络连接. 电子工业出版社, 2020.11

[13] 曹畅，唐雄燕. 算力网络——云网融合 2.0 时代的网络架构与关键技术. 电子工业出版社, 2021.09

[14] 罗峰，张东飞，高智芳. 算力网络详解卷 3 算网大数据. 清华大学出版社, 2023.01

[15] 中国电信研究院. 云网一体信息基础设施——云网融合下的算力网络技术与实践白皮书. 2023.08

[16] 中国电信研究院. 云网一体信息基础设施——IP 网络 3.0 技术白皮书. 2023.08

[17] 中国移动研究院. 面向 AI 大模型的智算中心网络演进白皮书. 2023.04

[18] 中国移动.“九州”算力互联网（MATRIXES）目标架构白皮书. 2024.04

[19] 中国联通研究院. 2024 算力网络智能运营白皮书. 2024.08

[20] 中国电信研究院. 分布式智算中心无损网络技术白皮书. 2024.08

[21] 安捷诺. 2024 面向未来的算力网络连接-中国算力网络市场发展白皮书. 2024.06

[22] IETF. Computing-Aware Traffic Steering (cats). https://datatracker.ietf.org/doc/charter-ietf-cats/, 2023.03

[23] IETF. L3 Standalone Service ID Framework. https://datatracker.ietf.org/doc/draft-huang-rtgwg-standalone-sid-framework/, 2024.07

[24] IETF. HPCC++: Enhanced High Precision Congestion Control. https://datatracker.ietf.org/doc/html/draft-miao-ccwg-hpcc-02/, 2024.02

[25] 中国通信标准化协会 CCSA.算力网络 总体技术要求[S]. 2023.07

[26] 中国通信标准化协会 CCSA.算力网络 算力路由协议技术要求[S].2024.05

[27] 黄光平，史伟强，谭斌. 基于 SRv6 的算力网络资源和服务编排调度[J]. 中兴通讯技术, 2021, 27(3)