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

未来网络白皮书系列

智算网络技术与产业白皮书

第八届未来网络发展大会组委会

2024 年 8 月

版权声明

主要编写单位：

紫金山实验室、北京邮电大学、华为技术有限公司、中兴通讯股份有限公司、中移（苏州）软件有限公司、中移（杭州）信息技术有限公司、天翼云科技有限公司、上海邮电设计咨询研究院有限公司、北京交通大学移动专用网络国家工程研究中心、浙江大华技术股份有限公司、科大讯飞股份有限公司、苏州盛科通信股份有限公司

主要编写人员（排名不分先后）：

黄韬、汪硕、高新平、肖玉明、徐鹃、李振红、时定兵、赵芷晴、杨彩云、韩红平、黄文浩、袁辉、胡秀丽、郑晓龙、徐峰、龚翔宇、吴涛、符哲蔚、陆振善、张佳玮、谷志群、李和松、段威、陆诗莹、贾玉、赵怡、成伟、王俊杰、罗远、刘静、马玉寅、彭天皓、吕宵双、杨志逵、刘耀华、史银妹、刘文斌、王国栋、周春旭、张涛

前　言

当前，以生成式人工智能为代表的通用人工智能技术在全球范围内引起了广泛关注，并以前所未有的速度、广度与深度催动经济和社会发展，掀起了新一轮科技革命与产业变革。在人工智能产业发展过程中，智算网络发挥了基础性支撑作用。业界基于高性能网络构建算力集群，从而突破单点算力的性能极限，实现智算中心内外的算力协同与数据交互，并进一步打破智算中心的烟囱式孤立局面，实现更大规模的算力互联，为 AI 技术发展与科技创新提供强有力的支撑。

智算算力互联的实现依赖于一个能够支持高性能计算任务的网络环境，这要求智算网络必须具备超大带宽、超低时延、零丢包和稳定可靠的数据传输能力，以确保数据传输的及时性、完整性与准确性，从而满足智算业务对算力资源的按需取用与高效利用需求，并支持面向未来多样化智算应用场景提供定制化的网络服务。

针对上述挑战，本白皮书首先系统性梳理了当前智算网络领域的政策背景、产业动态以及技术发展脉络，并深入探讨了未来智算产业对网络能力的核心诉求，分析了高性能智算环境在网络带宽、时延、抖动、丢包等方面存在的挑战，由此引发对智算集群内与集群间核心支撑技术的讨论，涉及新型网络架构、超宽可编程转发、负载均衡、光电融合组网与路由、广域 RDMA 等关键技术。随后，结合智算网络产业的典型案例，阐释了上述关键技术在智算基建建设中的应用。最后针对智算网络提出了技术与产业发展建议，旨在为行业从业者、

决策者及研究者提供一定参考，以推动智算网络技术的创新与应用。

目　　录

一、智算网络技术与产业发展概况

近年来，全球对智能算力的需求急剧增长，推动智算服务进入新一轮爆发期。据统计，2022 年全球智能算力规模已达 142 EFLOPS，并预计 2030 年将达到 16 ZFLOPS，年均增速超 80%，这种增速奠定了智能算力将成为全球算力规模增长主要驱动力的地位。在此背景下，本章将围绕智算政策态势、产业形式与技术趋势等方面展开深入分析。

（一）政策态势

随着全球科技革命与产业变革的加速，我国高度重视数字基础设施的建设，尤其在智能计算领域。国家通过《“十四五”国家信息化规划》明确了未来几年加强数字基础设施的基调，特别是智能算力基础设施的建设，将成为推动经济高质量发展的核心支撑。

（1）加强政策引导与支持

2017 年，国家工信部颁布了《促进新一代人工智能产业发展三年行动计划（2018-2020 年）》，明确指出要将人工智能与制造业深度融合，并推动智慧工厂的发展；同年，国务院发布了《新一代人工智能发展规划》，提出要构建以人工智能为主攻方向的创新机构，并逐步增加在该领域的投入；2021 年发布的《新型数据中心发展三年行动计划（2021-2023 年）》和《“十四五”数字经济发展规划》指出，要推动智能计算中心有序发展，打造智能算力、通用算法和开发平台一体化的新型智能基础设施，提供体系化的人工智能服务；2023 年

是 AI 大模型元年，该年两会报告中多次提及 ChatGPT 等大模型的人工智能词汇，并提出了关注数据安全与提升产业质量的核心建议和提案；2024 年，《政府工作报告》中首次提出开展“人工智能+”行动，标志着人工智能向大规模落地应用发展的态势。

（2）加快数字基础设施建设

在《“十四五”国家信息化规划》指导下，我国正在加快建设泛在智联的数字基础设施体系，包括部署高速可靠的 5G 网络与大规模卫星互联网，以及建立全国一体化大数据中心。上述措施已在多地实施，显著提升了区域间的数据处理能力与网络响应速度，为经济社会数字化发展提供了强有力的支撑。此外，还优化了全国互联网骨干直联点并加快了 IPv6 的规模部署，新建了国家级互联网交换中心提升网络效率与数据处理能力。通过发布系列政策，加强智算设施的建设与升级，支撑新感知和新算力设施的快速发展。

（3）强化规划与管理

国家发改委发布的通知中指出，必须制定跨地域、跨系统的数字基础设施建设规划，以确保东西部算力协同发展。通过优化资源配置与推动区域平衡发展，使国内多个地区实现了更为高效的数字基础设施管理。为加强统筹监测，引导东西部算力协同发展，构建全国一体化算力体系，政策已着力制定跨地域、跨系统的数字基础设施建设规划。通过加大对智算资源的规划投资，确保各地区、各行业的数字化转型需求得到有效满足。

（4）推动数字化产业升级

各地方政府正在积极抢占智算先机，推动产业的数字化升级。例如，北京正在建设亦庄等 E 级智能算力高地，并计划到 2027 年实现智算基础设施软硬件产品的全栈自主可控；上海在推进“算力浦江”智算行动实施方案，打造高质量智算发展格局；贵州通过与华为云、科大讯飞等企业合作，推动盘古、星火基础大模型在本省落地，并建立公共数据目录“一本账”，力争在数据训练与行业大模型培育方面取得领先优势。

（二）产业形势

我国正在积极推进智算网络标准化进程，以满足人工智能与高性能计算需求。国内智算产业链涵盖从核心技术研发、资源整合到广泛应用的全链条。各大云服务商和电信运营商正在加速构建 AI 大模型与智算平台，以提升业务流程的智能化水平和效率。

在国内标准化方面，中国通信标准化协会正在主导国内的智算网络标准化工作。当前阶段主要集中在互联互通与基础支撑方面，系统化地推动智算网络的总体技术要求、无损协议、广域网能力要求、存算一体、设备平台互联互通、安全等标准化研究进程。2023 年，中国联通、中国电信、信通院、紫金山实验室围绕下一代网络演进（NGNe, Next Generation Network Evolution）在 SG13 启动智算立项；在国际标准化方面，智算网络的标准化工作主要由 ITU 和 IETF 等国际组织推动。为满足人工智能和高性能计算（HPC, High Performance Computing）对智能算力需求的急速增长，2023 年 7 月，Linux 基金

会联合 AMD、Arista、博通、思科等公司共同成立了超以太网联盟（UEC , Ultra Ethernet Consortium），该联盟旨在通过改进以太网技术的物理层、链路层、传输层和软件层，提升其转发性能，同时兼容当前以太网生态。

此外，全国各地正在推进智算中心的建设。据统计，目前全国超过 30 个城市正在建设或提出建设智算中心，建设总数超 100 个，总投资规模超百亿元。这些项目的建设主体包括政府机构、三大电信运营商以及部分互联网企业。典型的智算中心包括中国电信京津冀大数据智能算力中心、阿里云张北超级智算中心、腾讯长三角（上海）人工智能先进计算中心、南京智能计算中心等，其中 12 个位于“东数西算”八大枢纽。2024 年，武昌智算中心、中国移动智算中心（青岛）、华南数谷智算中心、郑州人工智能计算中心、博大数据深圳前海智算中心等也已相继开工或投产使用。

图 1-1 我国智算中心及大模型分布

我国智算产业链已形成完整的上游核心技术研发、中游资源整合服务到下游广泛应用的链条：

（1）上游产业是智算的技术源头与核心支撑，包括芯片、软件及硬件供应商。目前，AI 芯片领域呈现多元竞争格局，GPU 和 FPGA 因其高技术壁垒，已形成稳固的寡头垄断市场；与同时，TPU、NPU 等 ASIC 芯片崭露头角，如华为昇腾 NPU 和阿里平头哥 NPU，凭借其在吞吐量、能效及算力等方面的突出表现，已在 AI 领域得到大量应用。

（2）中游产业是智算资源的整合者与服务提供者，主要由云服务商、电信运营商及第三方数据中心服务商组成。云商及科技公司利用技术积累，提供大模型及平台服务，一方面将部分传统数据中心改造为专为人工智能设计的智算中心，另一方面加速构建 AI 大模型。IDC 服务商则依托云网资源，深度参与智算建设。例如，中国电信已推出息壤智能计算平台，提供智算、超算、通算等多元化算力服务，为大模型训练、无人驾驶、生命科学等领域提供软硬件一体化解决方案，其 RDMA 吞吐能力高达 1.6Tb。

（3）下游产业涵盖互联网、交通、金融、工业等众多行业用户。通过引入智算技术，实现业务流程智能化、产品与服务创新以及决策支持优化，推动行业数字化转型与智能化升级。例如，百度文心大模型助力浦发银行与泰康保险在投资决策、理赔信息检索等业务中提升效率；华为盘古大模型为国家电网提供智能电力巡检解决方案；小鹏汽车在乌兰察布设立了自动驾驶智算中心“扶摇”，基于阿里飞天智

算平台，拥有 600 PFLOPS 算力，使自动驾驶核心模型训练速度提升近 170 倍。

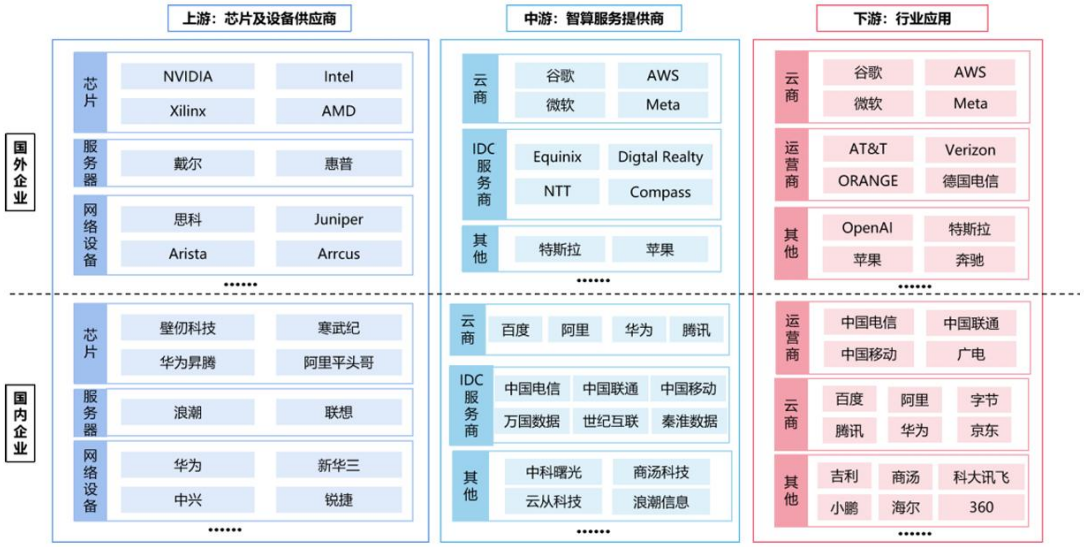

图 1-2 智算产业链

（三）技术趋势

（1）AI 模型参数规模将超百万亿，超长序列成为主流标配

从 2024 年 AI 行业的发展趋势来看，大模型 Scaling Law 依然保持旺盛生命力，万亿甚至百万亿参数规模的大模型成为必然趋势。以 OpenAI 为例，其下一代模型 GPT-5 的参数规模将达到 2 万亿以上，而更远期的 Q*系列模型将采用多模态自我演进训练机制，使模型训练不再局限于有限的人类数据，实现从数据驱动向算力驱动的转变。同时，超长序列也逐渐成为未来模型的主流标配，以 Sora 为例，视频生成场景需要使用长达百万长度的序列，例如 60 秒的视频需要 1M 的序列长度、10 分钟视频则需要 10M 序列长度，这标志着序列长度将成为衡量模型能力的重要指标。

模型规模与序列长度的急速递增推动了对算力的高需求，也激发了企业在智算基础设施领域的投入。例如，Tesla 计划在 2024 年投入 100 亿美元建设 AI 算力集群，而微软联合 OpenAI 启动的星际之门项目更计划投资 1000 亿美元，打造数百万 GPU 规模的算力集群。

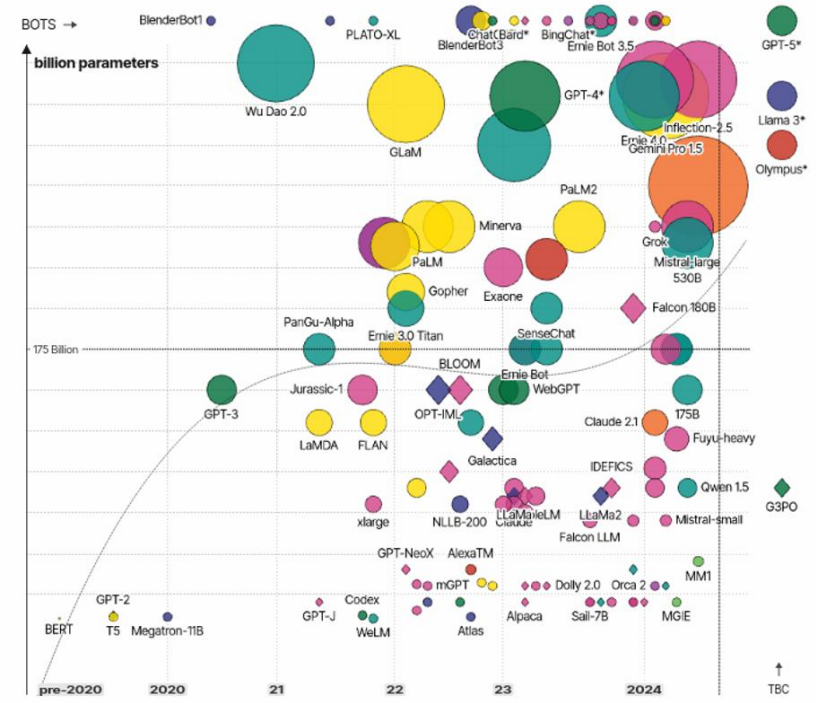

图 1-3 AI 大模型的发展趋势

（2）以太推动智算网络开放互联，百万卡集群成为共识

在产业界共同努力下，智算网络呈现出两方面的演进趋势：一是以太将成为智算网络开放互联的基础，二是百万卡集群规模成为行业共识。

行业正逐步认识到以太网在 AI 与 HPC 场景中的强劲生命力。更多的 GPU 厂商选择以太作为其算力芯片的 IO 接口形态，如 Intel Gaudi 及众多国产芯片商。同时，中国移动牵头的全调度以太网技术体系（GSE），以及由海内外头部厂商组成的超级以太网联盟，正在

突破基于以太网构建超大规模高性能 AI 集群的技术瓶颈。事实上，“以太网或 InfiniBand”这道命题已经有了答案，以太已成为构建超大规模开放互联网络的技术基石。

其次，百万卡集群将成为未来几年智算行业发展的重要方向。随着模型规模逐渐逼近甚至超过人脑水平，相应的 AI 集群规模也将从之前的千卡或万卡级别，迅速发展到十万卡甚至百万卡规模。2024 年是百万卡集群需求元年，并将在未来数年内持续推动智算网络技术的发展与创新。

（3）融合将成为智算网络演进的主路径

从宏观技术发展趋势来看，“融合”将成为智算网络演进的关键驱动。其中，总线与网络的融合、电互联和光互联的融合是两大重要技术趋势。

首先，传统的总线技术（如 PCIe、NVlink）和网络技术（如 Ethernet、Infiniband）之间的界限将变得更为模糊，总线网络化和网络总线化的趋势将同步进行。总线网络化指传统总线技术将借鉴网络大规模扩展的经验和技术特性，以实现更大规模的垂直扩展。而在网络总线化趋势下，传统网络技术也将借鉴总线技术低延迟和高可靠特性，以进一步提升互联性能并扩展新应用场景。因此，总线与网络技术的融合将成为未来数年内智算网络发展的主要趋势之一。

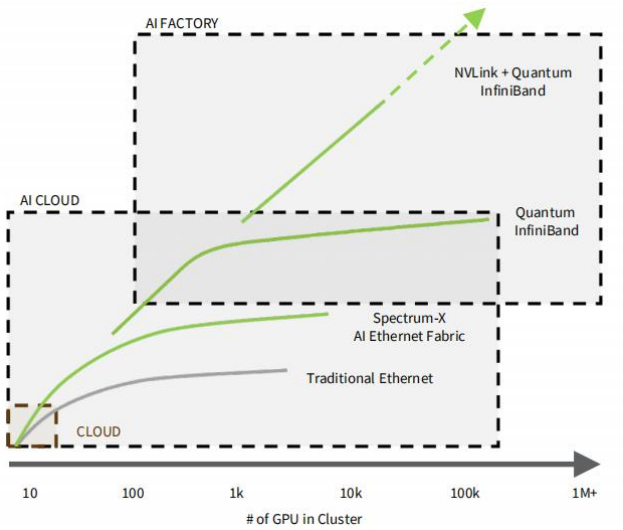

图 1-4 英伟达的总线和网络技术

其次，电互联和光互联的技术融合将推动智算网络在架构和成本方面的革新。若按当前算力芯片的发展速度来看，IO 密度与功耗将成为难以突破的瓶颈。对此，OCS 光交换机和基于硅光的 CPO/OIO 技术将在组网架构与单比特功耗等方面深刻影响未来数年智算网络的发展。可以预见，未来智算网络将紧密融合电互联与光互联技术，优化网络架构与功耗成本，以达到极致的应用效果。

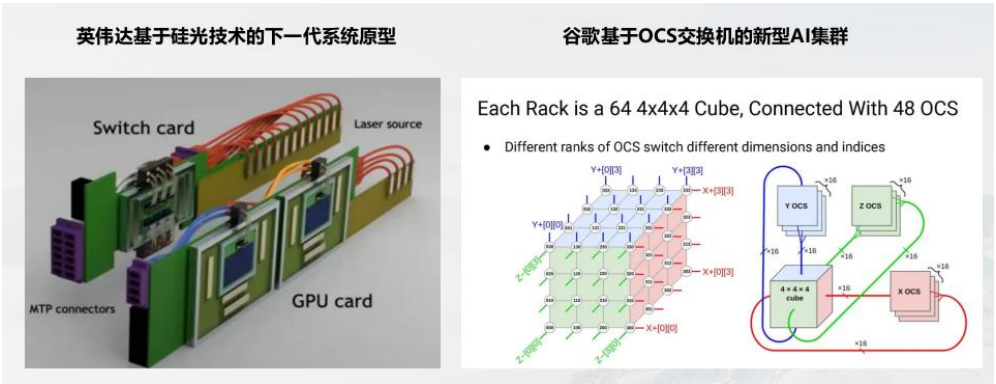

图 1-5 电互联技术和光互联技术的融合

（4）新型大容量网络芯片将成为智算网络发展的基石

随着智算业务对高速互联需求的持续攀升，新型大容量网络芯片正处于高速发展阶段，呈现如下趋势：

容量持续增长，单比特功耗不断降低。过去 10 年，以太网交换芯片的容量从百 G 迅速提升到 51.2T，增长近 100 倍，且单比特功耗下降 90%以上。在 AI 驱动下，未来网络芯片预计将迅速突破 100T 容量，单比特功耗将进一步降低。同时，400G/800G DPU 网卡需求也将迎来井喷。

面向 AI 场景优化将成为网络芯片发展的基本要求。过去两年中，实现面向 AI 场景优化已成为新一代网络芯片的重要特征，并将在未来五至十年内推动网络芯片的更新迭代。典型技术包括超低延迟、故障预测、智能流分析引擎、基于容器/包的负载均衡、在网计算等。

业务场景融合正成为新型网络芯片的发展方向。当前，总线网络化和网络总线化趋势明显，下一代芯片需要综合考虑总线和网络两个场景的需求，以实现架构统一和网络极简目标。

二、智算产业对于网络的核心要求

（一）网络带宽要求

网络带宽是 AI 大规模训练中的一个关键要素。以深度学习为例，训练复杂的神经网络模型可能需要处理 TB 级甚至 PB 级的数据。为保证训练效率，网络高带宽对存储设备、计算节点和内存之间的快速数据传输至关重要。尤其是在分布式训练场景下，多个计算节点之间需频繁交换大量中间结果与梯度信息，若带宽不足则将产生数据传输瓶颈，进而影响训练速度。目前，主流 AI 训练平台通常采用高带宽的网络连接，如 10Gbps、40Gbps 甚至更高的带宽，以满足大规模数据传输需求。随着 AI 模型复杂度和数据规模的持续增加，未来对网络带宽的需求将进一步提升。

网络带宽对于 AI 推理同样至关重要。推理过程通常需要快速响应用户请求，并在短时间内返回结果。例如，在自动驾驶、实时视频分析等场景中，系统需要在毫秒级的时间内处理和传输大量数据。即使采用边缘计算方式，仍需高带宽网络来实现用户与数据中心间的大体量数据传输，特别是在多边缘节点协同工作的情况下，该需求尤为突出。此外，随着多模态技术的快速发展，大量图像、声音、视频、传感等数据的传输，也对 AI 推理所需网络带宽提出了更高要求。

（二）网络时延要求

低时延是支撑 AI 大模型分布式训练的关键要素。分布式训练要求

在多个计算节点之间频繁交换数据，若网络时延过高，则将导致数据传输速度减慢，进而影响整体训练效率。特别是在同步训练模式下，所有计算节点必须等待最慢节点的数据传输完成，才能进行下一轮计算。因此，网络时延的增加将直接导致训练时间的延长。另外，随着AI 模型的复杂度与参数规模持续增加，其对低时延网络的需求将更为迫切。

低时延对于 AI 推理同样重要。推理过程通常要求快速响应用户请求，并在短时间内返回结果。例如，在智能客服系统中，需对用户提问实现秒级反馈；在自动驾驶系统中，车辆需要实时处理传感器数据，并做出及时的决策，任何延迟都可能导致严重后果。此外，随着 AI 技术发展及应用扩展，其对时延的需求将进一步严苛化，因此攻克低时延技术成为打造智算网络的一项核心挑战。

（三）网络抖动要求

通算与智算在流量特征方面存在显著区别。通算中心的特征是流数量多（通常超过 10W），但以小流为主，通信模式通常为点对点。相比之下，智算中心的特征为流数量少（通常低于 10K），但以周期性突发的大流为主，通常采用集合通信的模式，且流间存在同步效应。该同步效应对网络的抖动/长尾延迟极为敏感，可能引发大量 GPU 资源空转的问题。

然而，控制网络抖动相比平均时延更具挑战性。即使在无拥塞丢包的情况下，不合理的负载均衡与随机排队也可能引发抖动劣化，导

致应用性能下降。相关测试数据表明，在 AI 场景中，相比传统基于流的负载均衡技术，采用逐包负载均衡可显著减低时延抖动，同时可提升 40%的任务完成时间（JCT, Job Completion Time）增益。因此，有效控制时延抖动是 AI 高性能网络的重要需求，通过合理的技术手段可弥补当前网络抖动控制能力的不足。

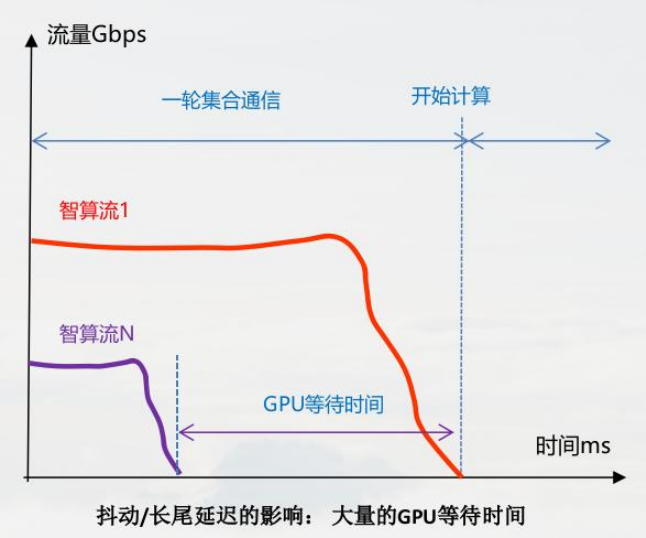

图 2-1 抖动带来的性能代价

（四）网络丢包要求

网络丢包在 AI 大规模训练中是一项极其重要的问题。分布式训练要在不同计算节点间频繁交换数据，若发生丢包则将导致数据传输失败，进而影响模型训练的准确性。尤其在同步训练模式下，任一节点的数据丢失都可能导致整个训练过程的中断，严重拖累训练进度。为减少丢包对训练的影响，需采用高可靠性的网络技术与协议，同时优化网络拓扑结构，增强网络设备稳定性，并在出现问题时快速进行路径切换，保证数据传输的连续性。

网络丢包同样会对 AI 推理性能产生影响，因为推理过程需要时

响应用户请求，若发生丢包则将导致数据传输失败，进而影响推理结果反馈的及时性。例如，在自动驾驶系统中，传感器数据的丢失可能导致系统无法正确识别路况，进而影响车辆安全行驶。

综上所述，智算网络的发展与应用亟需高带宽、低时延/抖动、轻丢包的网络支持，并通过不断创新与发展智算集群内与智算集群建的网络互联技术，为 AI 技术的研究与应用提供强有力的支撑。

为探索上述要求的解决方案，本白皮书将从智算集群内与智算集群外两个维度展开技术分析，详细阐述新型网络架构、无损网络、负载均衡等集群内关键技术，以及光电融合组网与路由、广域拥塞控制、广域 RDMA 等集群间关键技术。

三、智算集群内网络关键技术

（一）新型网络架构

在 AI 大模型训练场景中，GPU 数量与模型训练时长通常呈正比关系。多卡训练可极大缩短训练时间，尤其对于千亿级甚至万亿级参数规模的大语言模型，智算集群需支持万卡及以上的并行能力。智算集群内网络架构的优劣对 GPU 服务器内外的集合通信存在极大影响。因此，设计大规模、高可靠、低成本、易运维的优质网络架构，对于满足大模型训练的大算力、低时延和高吞吐需求具有重要意义。

（1）Clos 网络架构

胖树（Fat-Tree）Clos 无阻塞网络架构由于其高效的路由设计、良好的可扩展性及方便管理等优势，成为大模型训练常用网络架构。对于中小型规模的 GPU 集群网络，通常采用 Spine-Leaf 两层架构，如图 3-1 所示。对于较大规模的 GPU 集群则使用三层胖树（Core-Spine-Leaf）进行扩展组网，由于网络的层次增加，其转发跳数与时延也相应增加。

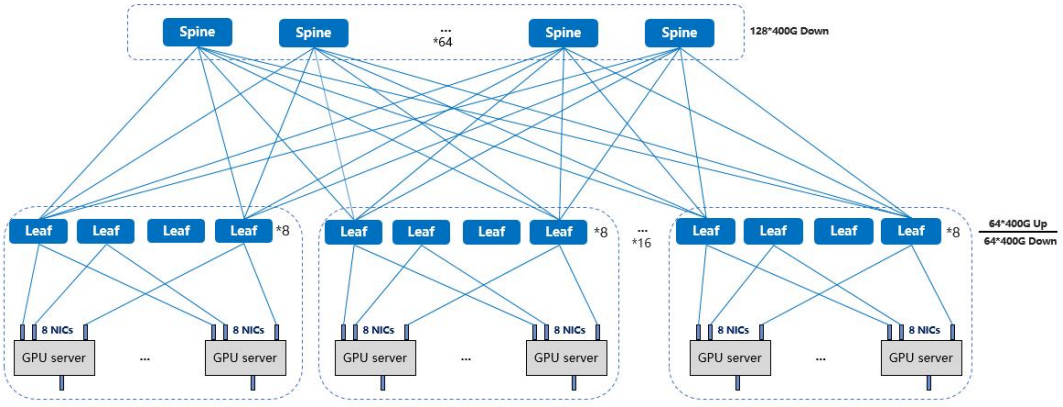

图 3-1 Spine-Leaf 两层 Fat-Tree 组网架构

GPU 服务器接入分为多轨和单轨两种方式。图 3-1 为多轨接入方式，其 GPU 服务器上的 8 张网卡依次接入 8 台 Leaf 交换机，该方式集群通信效率高，大部分流量经一级 Leaf 传输或者先走本地 GPU 服务器机内代理再经一级 Leaf 传输。图 3-2 为单轨接入方式，1 台 GPU 服务器上的网卡全部接入同一台 Leaf 交换机，该方式集群通信效率偏低，但在机房实施布线中有较大优势。此外，若 Leaf 交换机发生故障，多轨方式所影响的 GPU 服务器数量将多于单轨方式。

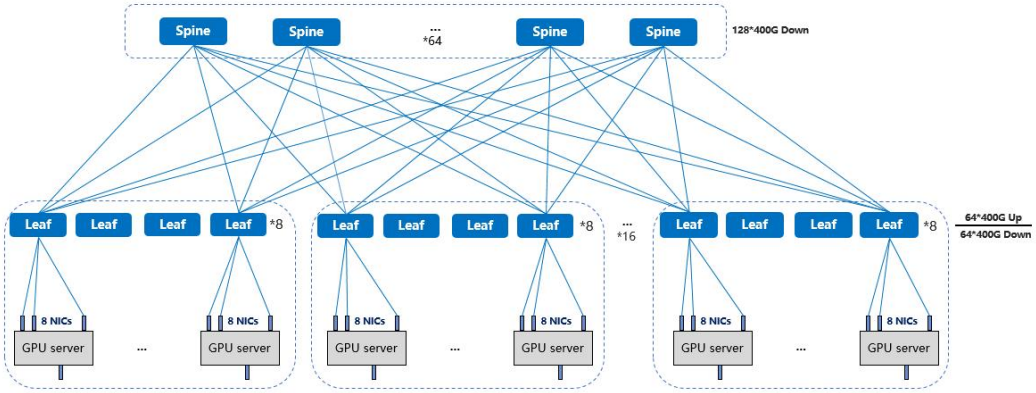

图 3-2 Spine-Leaf 两层 Fat-Tree 单轨组网架构

业内典型的大模型组网架构有腾讯星脉与阿里巴巴 HPN 网络。星脉网络采用无阻塞胖树（Fat-Tree）拓扑，分为 Cluster-Pod-Block 三级。如图 3-3 所示，以 128 端口 400G 交换机为例，其中 Block 为最小单元，各 Block 包含 1024 个 GPU，各 Pod 支持最大 64 个 Block，即 65536 个 GPU。多个 Pod 构成一个 Cluster 集群，支持 524288 个 GPU。

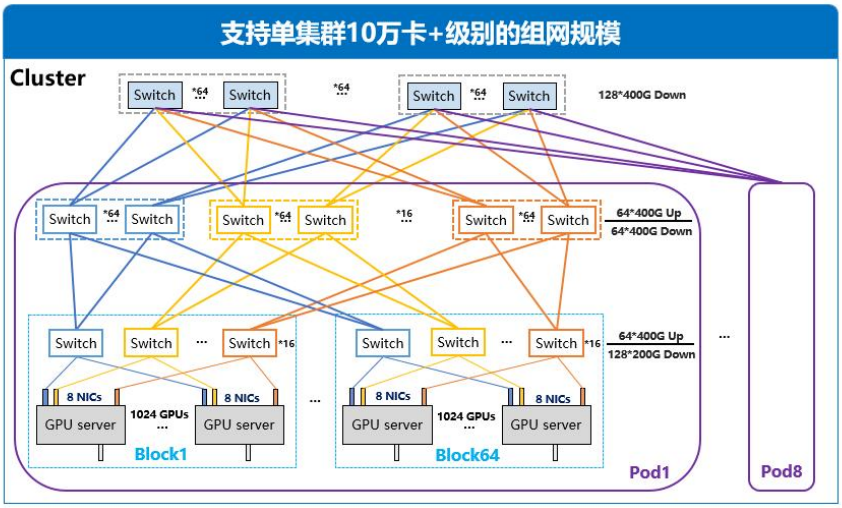

图 3-3 腾讯星脉 Cluster-Pod-Block 组网架构

阿里云大模型训练网络（HPN，High-Performance Networking）引入一种双平面两层架构，如图 3-4 所示。每台 GPU 服务器配置了 8 个 GPU，对应 8 个 NIC，各 NIC 提供 2×200Gbps 带宽，并上行连接到不同 Leaf 设备，形成双平面设计，从而避免单 Leaf 故障对训练任务的影响。若交换机为 128 端口，每台 GPU 服务器分别上行连至 16 台 Leaf，组成一个 Segment（包含 1024 个 GPU）。每台 Leaf 预留了额外 8×200G 端口接入 GPU 服务器，便于 GPU 服务器发生硬件等故障后可快速替换。Spine 层面连接多个 Segments 组成一个 Pod，每台 Leaf 上行有 60×400G 端口连接 Spine，因此一个 Pod 可容纳 15 个 Segments，即 15360 个 GPU。对于更大规模的训练任务，则会涉及到 Core 层面的连接进而组成算力规模更大的 GPU 集群。阿里根据其训练任务流量特性，选择 Spine-Core 之间采用 15:1 的收敛比设计，集群可支持 245760 个 GPU。

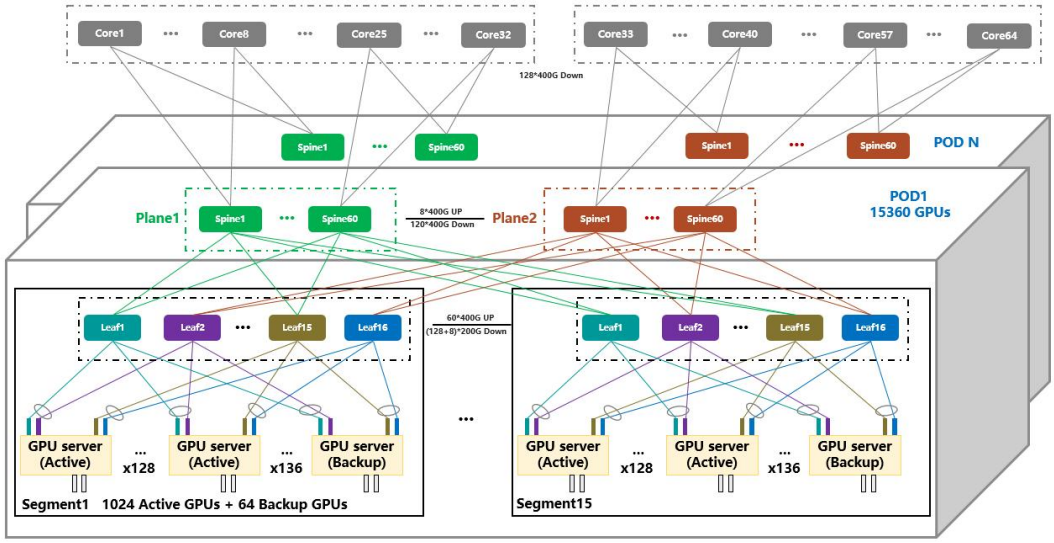

图 3-4 阿里巴巴 HPN7.0 组网架构

如图 3-5 所示，MIT 和 Meta 团队提出了 Rail-Only 大语言模型网络架构设计。相对于三层 Fat-Tree 组网，其剔除了 Spine 层交换机实现网络架构精简，仅使用一层 Rail 交换机用于高带宽域内 GPU 卡之间互联，其中每个高带宽域内 256 个 GPU 都通过 NVLink Switch 进行连接。Rail-Only 将网络转发路径跳数减小为 1，从而有效降低业务时延，并节约大量的网络设备建设成本。如图 3-5 所示，为实现 32768 个 GPU 的组网，三层 Fat-Tree 总共需要 1280 台交换机，即 512 台（64×8）Leaf 交换机、512 台（64×8）Spine 交换机和 256（64×4）台 Core 交换机，而 Rail-Only 组网仅需 256 台 Rail 交换机，可极大降低网络建设成本。

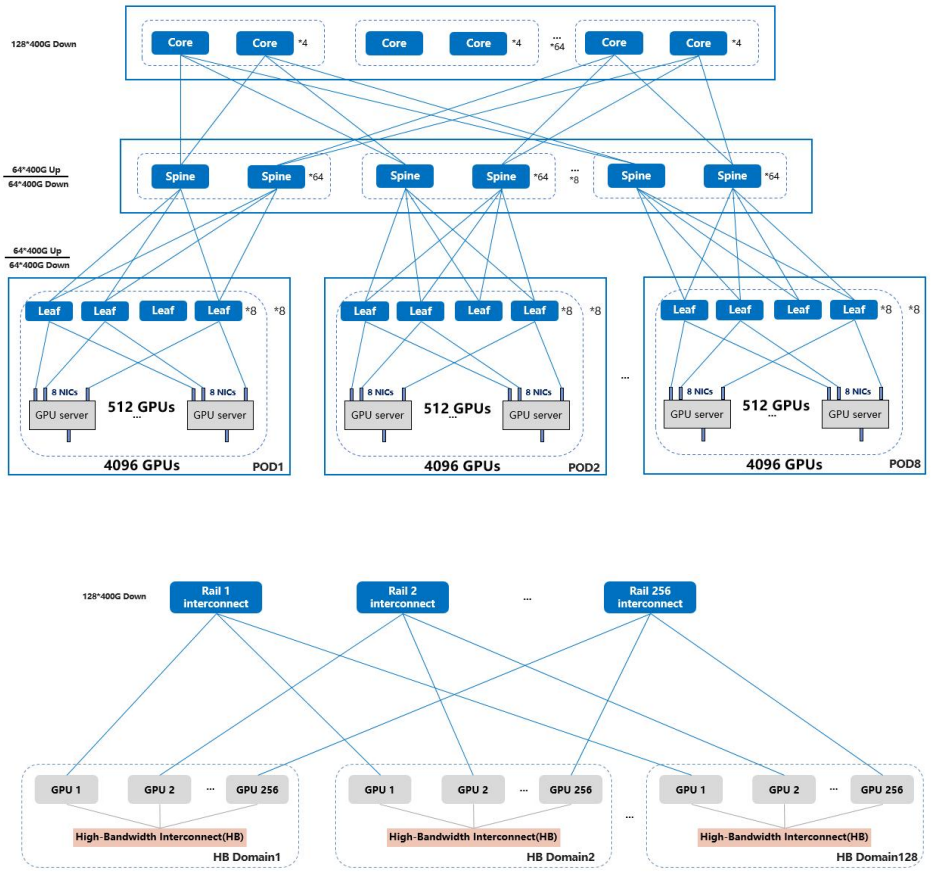

图 3-5 (a) 三层 Fat-Tree 与 (b) Rail-Only 组网架构对比(32768 GPUs)

（2）Dragonfly 网络架构

传统 Clos 树形架构作为主流的智算网络架构，重点突出其普适性，但在时延与建设成本方面并非最优。在高性能计算网络中，Dragonfly 网络因其较小的网络直径与较低的部署成本被大量使用。如图 3-6 所示，Dragonfly 网络分为 Switch 层、Group 层和 System 层。其中，Switch 层包含一台交换机，并通过终端链路接入 p 个计算节点；每个 Group 层包含 a 个 Switch，各 switch 通过 a-1 条本地链路与其它 a 个设备节点进行全连接；每个 System 层包含 g 个 Group，各 Group 通过 h 条全局链路与其它 Group 内设备节点进行全连接。为实现负载均衡，通常取值：a=2p=2h。在组网性能方面，以 64 端口交换机为例，

Dragonfly 可支持超过 27 万个 GPU 卡，相当于三层 Fat-Tree 架构所容纳 GPU 数量的 4 倍以上，而交换机数量及传输跳数可降低 20%。尽管 Dragonfly 网络可提供较高的性价比与更低的传输时延，但 GPU 集群每次扩展都需重新部署链路，因此其可维护性相对较差。

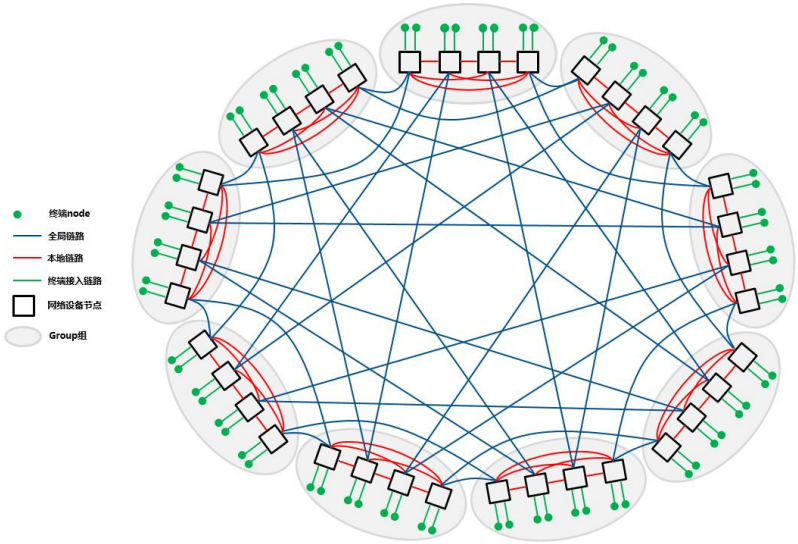

图 3-6 Dragonfly 组网架构

（3）Group-wise Dragonfly+网络架构

当规模需求超过十万卡时，最直接的组网方式是引入无收敛三层 Fat-Tree 架构。 以单端口为 400G 的 51.2T 盒式交换机为例，三层盒盒盒组网，最大支撑 50 万+节点组网。然而此架构存在以下两个主要问题：1、系统复杂度，三层组网的负载均衡、拥塞控制等网路技术的难度和复杂度将大幅提升； 2、成本和功耗，对比二层 Fat-Tree 组网网络成本和功耗开销提高。为了应对以上两个挑战，在此场景下可以有两种架构选择：

架构一为第二层带收敛的三层 Fat-Tree 架构，即下图中 L2 层交换机的下行带宽：上行带宽为 N：M（N>M）。在同等规模下此架构可降低 L3 层的设备数量，节省成本和功耗。

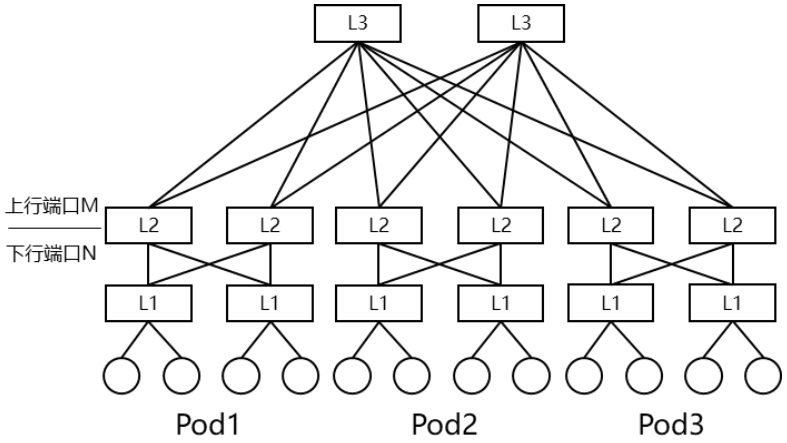

图 3-7　L2 带收敛的三层 Fat-Tree 架构示意图

架构二为 Group-wise Dragonfly+（GW-DF+）直连架构。如下图所示，每个 Pod 内设备通过二层 Fat-Tree 架构互联。Pod 间，同位置或同号的 L2 设备两两直连。以单端口为 400G 的 51.2T 盒式交换机为例，此架构最大可支持 20 万+节点规模。如果 L2 替换为框式交换机，规模可超 100 万。

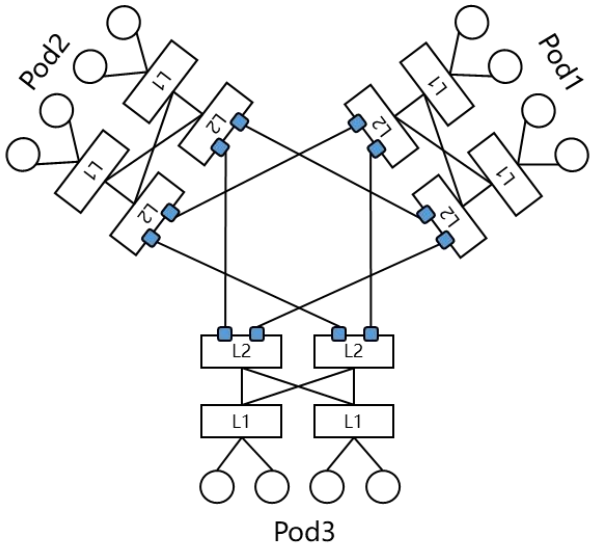

图 3-8 Group-wise Dragonfly+（GW-DF+）直连架构示意图

对比二层 Fat-Tree 架构，此架构可大幅提高组网规模；对比三层 Fat-Tree 架构，此架构可节省一层交换机带来的成本和功耗开销；对比传统 DF+架构，此架构可避免上下设备绕行，简化路由复杂度和提

升系统效率

（4）Torus 网络架构

Torus 网络架构是一种完全对称的拓扑结构，具备低时延、低网络直径等特性，适合集合通信使用，可显著降低建设成本。图 3-7 呈现了一维边长为 3 的 Torus 及二维边长为 3 的 Torus 网络。Torus 网络环面拓扑特性可使得其在邻居节点之间拥有最优通信性能。然而，Torus 网络扩展可能涉及拓扑重新调整，且维护复杂度较高。

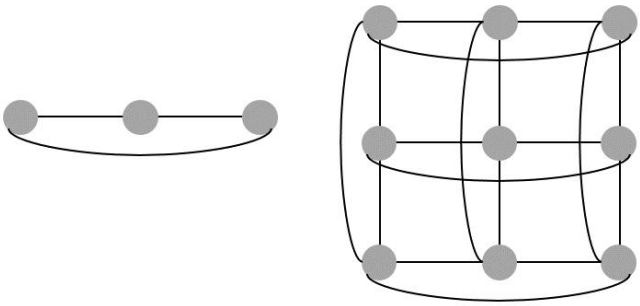

图 3-9 一维与二维 Torus 网络

（二）超宽可编程转发技术

超宽可编程转发技术是一种高度灵活、可定制的网络转发技术，可在不同层级上进行数据处理与转发，包括物理层、数据链路层、网络层和应用层等。这种灵活性使得它在智算网络中具备广泛的应用前景，可适应不同应用场景需求，并应对未来网络及应用需求的发展变化。

超宽可编程转发主要采用 RoCE 网络方案，当前以博通 Tomahawk 5 芯片的白盒交换机为主流，转发容量可达 51.2Tb/s，支持 64×800G/128×400G/256×200G 端口接入以及多种流量调度策略（如

逐包负载分担、全局负载分担等）。当前国内 OTT 厂商（阿里、腾讯等）、运营商（中国移动等）以及紫金山实验室，都在积极自研基于 Tomahawk 5 芯片的白盒交换机，构建布局智算中心超宽无损网络解决方案。英伟达同时提出了基于 RoCE 与 IB 的产品方案，其 InfiniBand NDR Quantum-2 交换机当前只支持 64×400G，相对于业界最新 RoCE 交换机，其性能上存在一定劣势。而 Spectrum-X 以太网交换机转发容量已达 51.2Tb/s，支持 64×800G/128×400G 端口接入，与博通同处业界领先水平；国内厂商以华为为例，自研芯片盒式交换机转发容量达到 12.8Tbps，可支持 32×400G 端口接入，并提出全局负载均衡（NSLB，Network Scale Load Balance）调度方案，以实现智算中心网络超宽无损承载。

可编程转发的实现主要涉及控制面、转发面的可编程操作：

（1）控制面可编程：实现集中化的流量调度

控制面主要对底层网络交换设备进行集中管理，包括状态监测、转发决策以及调度数据平面流量。控制面可编程技术的发展将引入如下优势：*i*）白盒交换机采用类似服务器的网络操作系统，可利用现有的服务器管理工具实现网络自动化，支持对开源服务器软件包的便捷访问，实现在交换机上使用与服务器完全相同的配置管理接口，从而加快技术创新；*ii*）将传统交换机的专有网络环境转变为更通用的环境，有助于高效地拓展与管理网络服务，提升白盒交换机的可编程性和网络可见性；*iii*）通过 API 和控制器，在交换机的网络操作系统中实现网络功能的按需编写（如网络分流器），从而减少每个交换机

上的硬件部署，实现对网络的集中管理和监控。

在智算网络场景中，低熵大流量可能导致负载不均衡问题，进而影响大模型的训练效率。对此，业界提出了控制面与 AI 平台联动的解决方案，实现对网络的整体感知与高效承载。例如，紫金山实验室提出了集中化流量调度方案，构建网络控制器与 AI 平台的协同任务调度机制，设计任务流量模型，制定策略路由并下发至白盒设备，通过集中化调度显著提升整网链路的带宽利用率，实现高达 97%以上的有效吞吐。

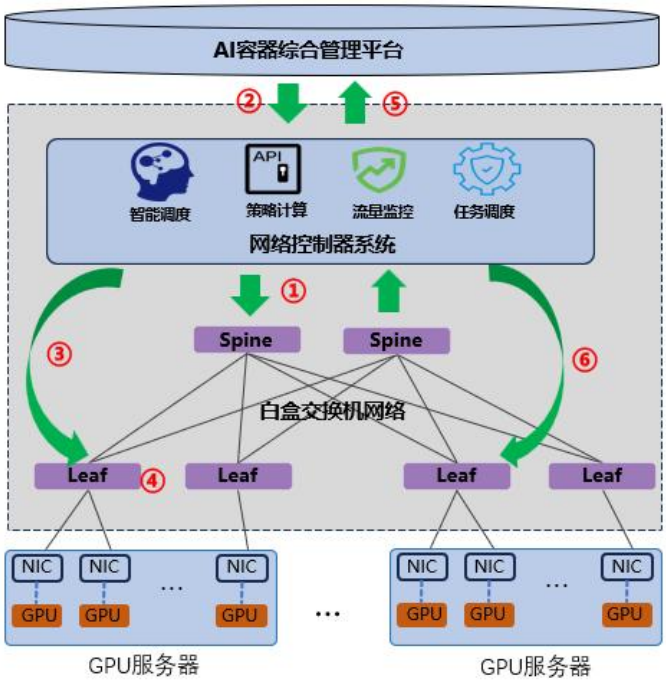

图 3-10 集中流量调度方案架构图

集中化流量调度方案操作步骤如下：

① 白盒部署上线，开启 ZTP 下载版本及基础配置，实现 Underlay 配置自动化部署；

② AI 平台创建任务，并将任务模型下发至网络控制器；

③ 控制器解析任务模型，规划流量路径，并通过策略路由下发

至白盒设备；

④ Leaf 策略路由生效，指导业务流量均匀转发；

⑤ AI 平台删除任务，并将任务删除事件下发至网络控制器；

⑥ 控制器下发删除任务至白盒设备，删除任务对应的策略路由。

（2）转发面可编程：实现自定义的节点转发逻辑

除了控制面可编程，数据面的可编程同样重要。然而，传统的网络芯片采用固定的流水线方式，包括固化的报文头解析、转发逻辑以及报文封装，导致整个处理过程无法更改。因此，许多新协议或新特性难以通过灵活编程得以快速实现。

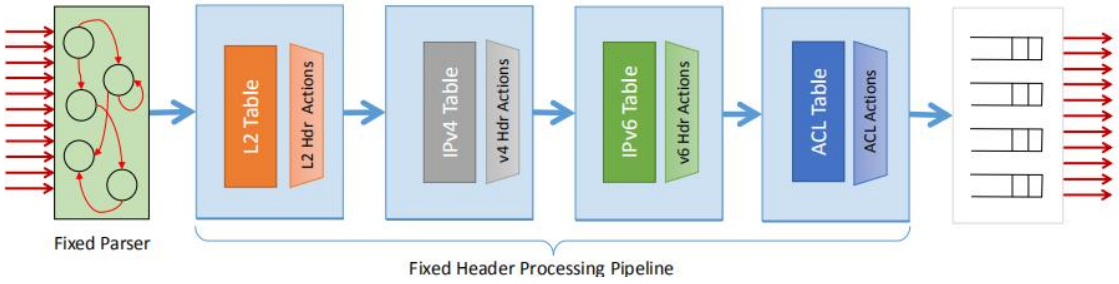

图 3-11 传统芯片的固定流水线

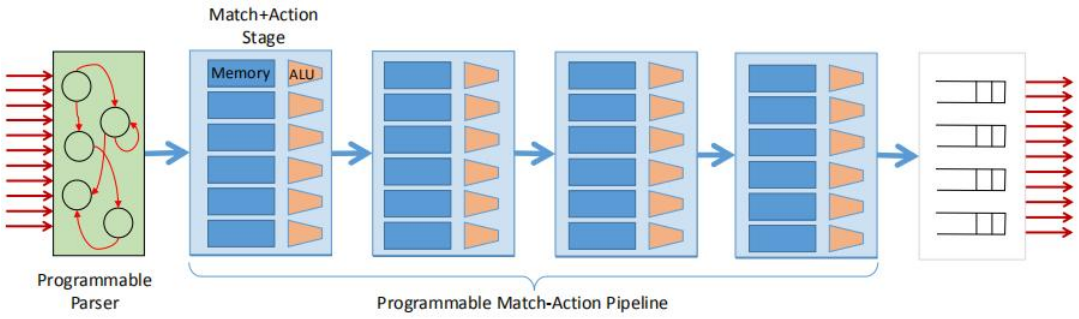

图 3-12 基于 PISA 架构的可编程流水线

PISA（Protocol Independent Switch Architecture）架构的引入使得硬件芯片具备了可编程能力。PISA 架构涵盖解析器、逆解析器、匹配和动作表、元数据总线等组件。解析器将报文转化为元数据，逆解析器将元数据转化为报文发送，匹配和动作表用于操作元数据并实现

所需的转发逻辑。数据面程序使用高级语言 P4 编写，经由 P4 语言编译器进行编译后在 PISA 设备上运行。

数据面可编程已在网络各个领域得到了广泛应用。基于 PISA 架构的“带内网络遥测”可实时收集数据面的转发路径、时延、抖动、拥塞等信息，为网络可视化和智能运维提供数据基础；基于 PISA 架构的“传输层负载均衡”实现了四层负载均衡器的功能，解决了软件负载均衡器所面临的带宽性能瓶颈，以及硬件负载均衡器面临的高成本等问题；基于 PISA 架构的“应用加速卸载”针对网络应用性能问题，采用协议无关的 P4 语言及底层可编程硬件，实现网络应用性能优化与关键功能卸载，例如 VNF 卸载等。

（三）无损网络技术

RDMA 技术相比传统网络具有显著优势，其实现了内核旁路机制，允许应用程序直接与网卡进行数据读写，无需操作系统和 TCP/IP 协议栈的介入，可将数据传输时延降低至 1μs。此外，RDMA 的内存零拷贝机制允许接收端直接从发送端的内存读取数据，大幅减少了 CPU 负担，提高了 CPU 效率。

虽然，RDMA 技术显著降低了服务器侧处理时延，提升了计算和存储效率，但同时会加剧网络拥塞，从而引发如下问题：增加网络处理时延以及导致业务丢包。业务丢包引发的重传将进一步增加时延，严重影响计算和存储效率。因此，需要构建无损网络技术体系，为 RDMA 提供低时延、零丢包与高吞吐的网络承载环境。

无损网络技术旨在确保网络不丢包的情况下实现高吞吐转发，其中包括流量控制、拥塞控制及负载均衡等。由于原生 IB RDMA 技术依赖于专用且昂贵的网络设备，加之从网络设备和线路的利旧角度考虑，基于以太网的 RoCEv2 技术将拥有广阔的应用前景。在 RoCEv2 网络中，业界通常采用 PFC（Priority-based Flow Control）技术来处理拥塞场景下的丢包和重传时延问题，提高计算和存储效率。然而，过多的 PFC Pause 可能降低吞吐量，甚至引发 PFC 死锁。为在低时延、无丢包网络中提高吞吐量，业界进一步引入 ECN（Explicit Congestion Notification）和 DCQCN（Data Center Quantized Congestion Notification）技术。ECN 用于感知设备内部的队列拥塞情况，并配合 DCQCN 调整发送端速率。

无损网络技术旨在确保网络不丢包的情况下实现高吞吐转发，其中包括流量控制、拥塞控制及负载均衡等。由于原生 IB RDMA 技术依赖于专用且昂贵的网络设备，加之从网络设备和线路的利旧角度考虑，基于以太网的 RoCEv2 技术将拥有广阔的应用前景。在 RoCEv2 网络中，业界通常采用 PFC（Priority-based Flow Control）技术来处理拥塞场景下的丢包和重传时延问题，提高计算和存储效率。然而，过多的 PFC Pause 可能降低吞吐量，甚至引发 PFC 死锁。为在低时延、无丢包网络中提高吞吐量，业界进一步引入 ECN（Explicit Congestion Notification）和 DCQCN（Data Center Quantized Congestion Notification）技术。ECN 用于感知设备内部的队列拥塞情况，并配合 DCQCN 调整发送端速率。

下文将围绕无损网络中的流控、拥塞控制等技术展开详细阐述：

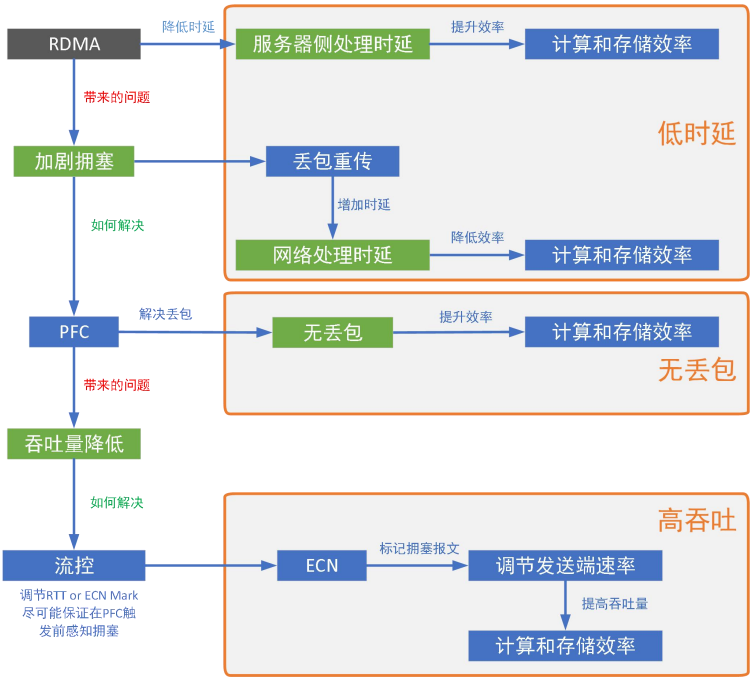

图 3-13 无损网络关键技术

（1）流控技术

流量控制是保障网络零丢包的基础技术，其提供了一种作用于接收方的机制，由接收方来控制数据传输速率，以防止高速的发送方压倒慢速的接收方。本节主要介绍流控相关技术，以及如何解决 PFC 死锁问题。

1）PFC 技术

PFC 是对 Pause 机制的一种增强，其原理如下：当下游设备发现其接收能力低于上游设备发送能力时，会向上游设备发送 Pause 帧至以暂停流量发送，并在等待一定时间后恢复发送。传统以太 Pause 机制是针对整条链路的流量暂停，而 PFC 支持在一条链路上创建 8 个虚拟通道，各虚拟通道对应一个优先级，从而支持单独暂停或重启任意虚拟通道，同时允许其它虚拟通道流量的无中断传输。

2）PFC 死锁

PFC 死锁是指当多个交换机之间因为环路等原因同时出现拥塞，各自端口缓存消耗超过阈值，而又相互等待对方释放资源，从而导致所有交换机上的数据流都永久阻塞的一种网络状态。

正常情况下，PFC 中流量暂停只针对某一个或几个优先级队列，不针对整个接口进行中断，每个队列都能单独进行暂停或重启，而不影响其他队列上的流量。然而，在某些特殊情况下，如链路故障或设备故障时，在路由重新收敛期间可能会出现短暂环路，导致出现一个环形依赖缓存区。如下图所示，当 4 台交换机都达到 PFC 门限，则将同时向对端发送 PFC 反压帧，此时拓扑中所有交换机都处于停流状态，整网吞吐量将变为零。虽然经过修复可使短暂环路很快消失，但其造成的死锁不是暂时的，即便重启服务器中断流量，死锁也无法自动恢复。

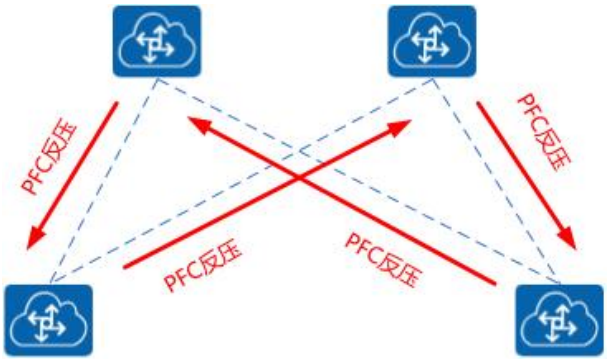

图 3-14 循环缓冲区依赖形成 PFC 死锁

3）PFC 死锁监测

服务器网卡故障可能引发其不断发送 PFC 反压帧，网络内 PFC 反压帧进一步扩散，导致出现 PFC 死锁，最终导致整网受 PFC 控制的业务瘫痪。

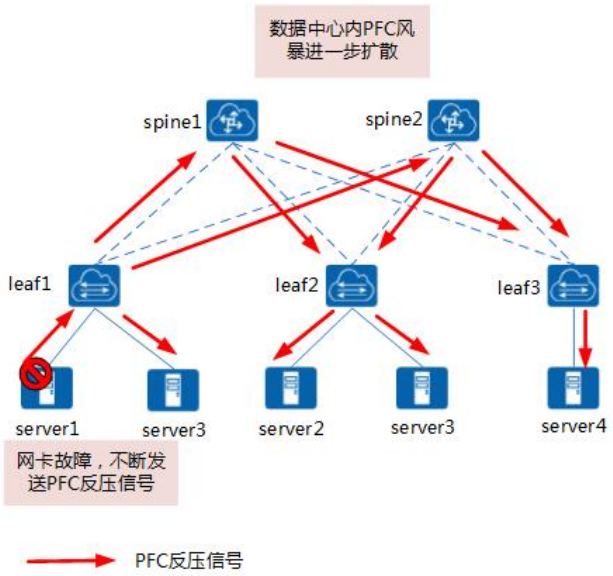

图 3-15 服务器网卡故障引起 PFC 风暴并形成 PFC 死锁

一旦出现 PFC 死锁，若不及时解除，将威胁整网的无损业务。无损以太网为每个设备提供了 PFC 死锁检测功能，通过如下过程对 PFC 死锁进行全程监控，当设备在死锁检测周期内持续收到 PFC 反压帧时，将不予响应。

- 死锁检测

如图 3-14 所示，Device2 的端口收到 Device1 发送的 PFC 反压帧后，内部调度器将停止发送对应优先级的队列流量，并开启定时器，根据设定的死锁检测和精度开始检测队列收到的 PFC 反压帧。

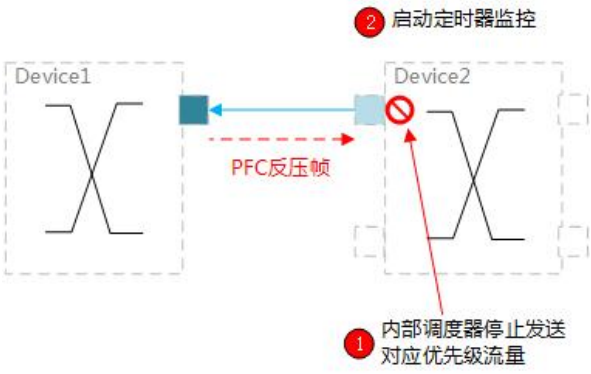

图 3-16 开启死锁检测

- 死锁判定

如图 3-15 所示，若在设定的 PFC 死锁检测时间内该队列一直处于 PFC-XOFF（即被流控）状态，则认为出现了 PFC 死锁，需要进行 PFC 死锁恢复处理流程。

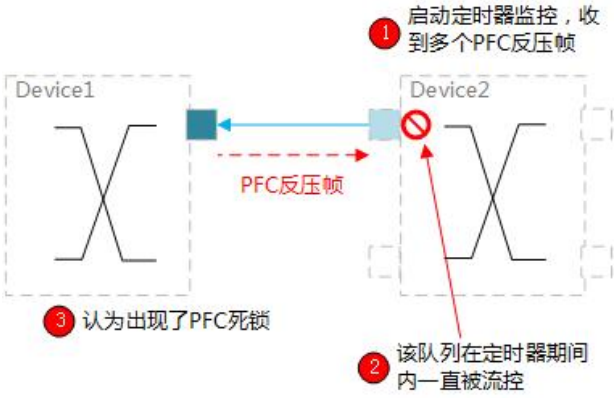

图 3-17 死锁判定

- 死锁恢复

在 PFC 死锁恢复过程中，会忽略端口接收到的 PFC 反压帧，内部调度器会恢复发送对应优先级的队列流量，也可选择丢弃对应优先级的队列流量，在恢复周期后恢复 PFC 的正常流控机制。若下一次死锁检测周期内仍然判断出现了死锁，那么将进行新一轮周期的死锁恢复流程。

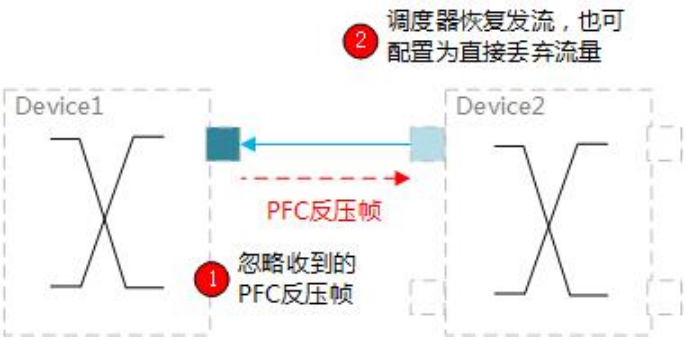

图 3-18 死锁恢复流程

- 死锁控制

若上述死锁恢复流程未起作用，仍然不断出现 PFC 死锁现象，

则可强制进入死锁控制流程。比如，在设定的时间段内，PFC 死锁触发次数达到某一阈值，则认为网络中频繁出现死锁现象，存在极大风险，此时进入死锁控制流程，设备将自动关闭 PFC 功能，需用户手动恢复。

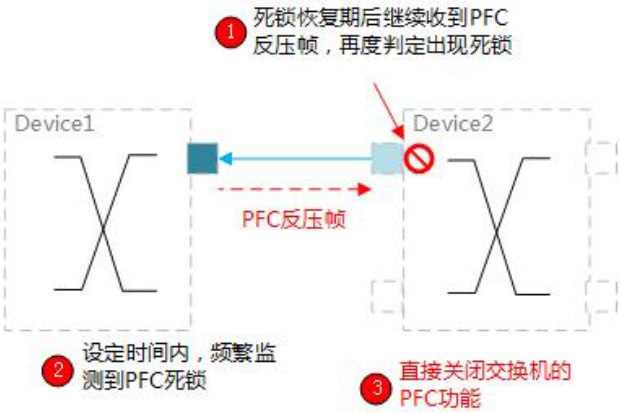

图 3-19 频繁出现死锁可关闭 PFC 功能

4）PFC 死锁预防

PFC 死锁预防是针对 Clos 组网的一种解决方案，通过识别易造成 PFC 死锁的业务流，并修改队列优先级，从而改变 PFC 反压路径，避免 PFC 反压帧形成环路。

如下图所示，某业务流沿 Server 1 → Leaf 1 → Spine 1 → Leaf 2 → Server 4 路径转发数据，正常转发过程不会引起 PFC 死锁。然而，若 Leaf 2 与 Server 4 间出现链路故障，或 Leaf 2 因某些原因未学习到 Server 4 的地址，均将导致流量不从 Leaf 2 下游端口转发，而从 Leaf 2 的上游端口转发。此时，Leaf 2 → Spine 2 → Leaf 1 → Spine 1 就形成了一个循环依赖缓冲区，当 4 台交换机的缓存占用都达到 PFC 反压帧触发门限时，都会同时向对端发送 PFC 反压帧停止发送某个优先级的流量，将形成 PFC 死锁状态，最终导致该优先级的流量在网

络中被停止转发。

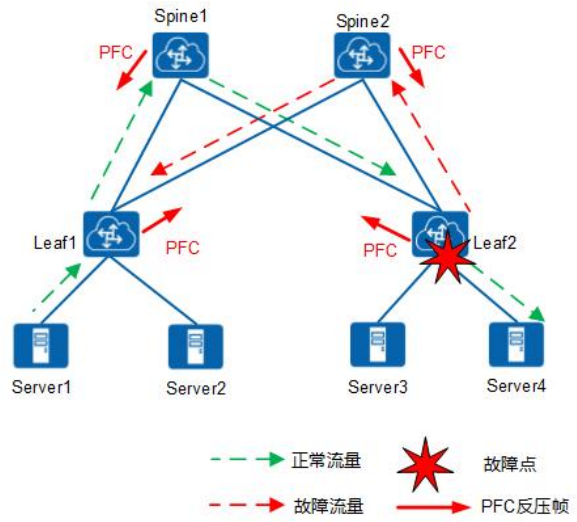

图 3-20　Clos 架构下的 PFC 死锁

PFC 死锁预防功能中定义了 PFC 上联端口组，用户可将一个 Leaf 设备与 Spine 相连的接口（如图 3-21 中 Interface 1 与 Interface 2）都加入 PFC 上联端口组。一旦 Leaf 2 设备检测到同一条业务流从属于该端口组的接口内进出，即说明该业务流是一条高风险的钩子流，易引起 PFC 死锁现象。

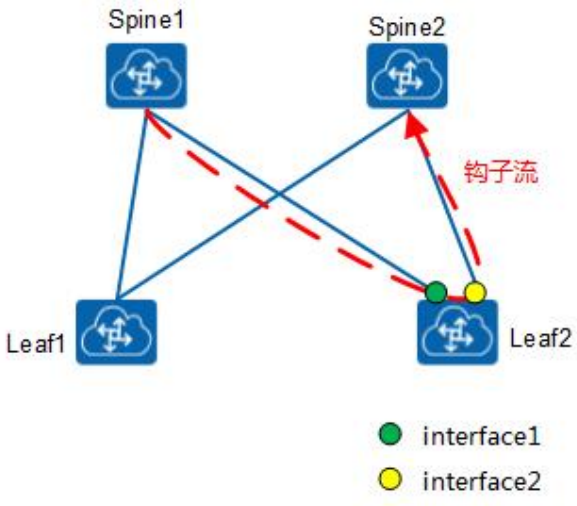

图 3-21 PFC 钩子流

为解决该问题，如下图所示，当 Device 2 识别出从 Device 1 发来的一条钩子流后，Device 2 会修改该流的优先级及其 DSCP 值，使其

从其它队列转发（即从队列 *a* 转移至队列 *b*），若该流在下游设备 Device 3 处发生拥塞并触发 PFC 门限，则将对 Device 2 的队列 *b* 进行反压使其停止发送流量，而不会影响到队列 *a*，进而避免形成循环依赖缓冲区的可能，防止 PFC 死锁的发生。

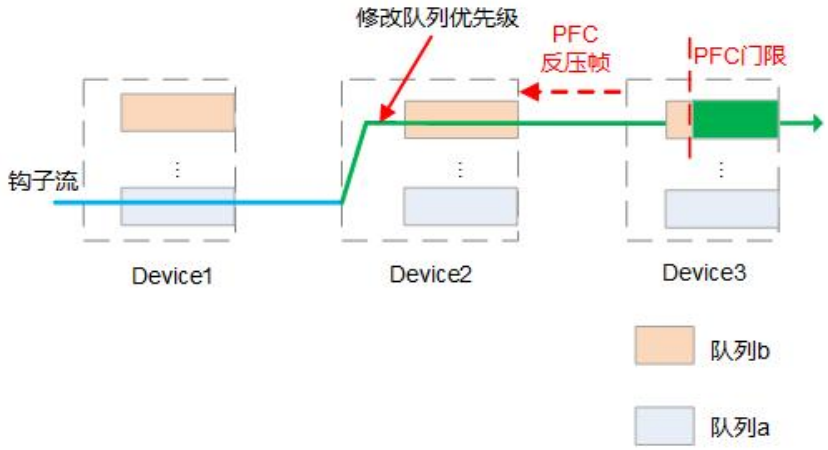

图 3-22 PFC 死锁预防原理

5）点刹式 PFC

传统 PFC 需要较大的缓存来保证不丢包和不欠吞吐。在长距数据中心互联场景中，两个数据中心的入口设备间距离很远，若一端设备出现缓存拥塞，从该设备发送 PFC 反压帧给对端设备到停止接收对端设备发来流量的时间差内，设备需要有足够的缓存空间吸收这段时间内对端设备发来的流量，以保证长距无损。在缓存空间大小和带宽一定的情况下，点刹式流控依靠短周期、高频率、持续少量调节流量发送与暂停的机制，能够比传统 PFC 支持更长距离的长距无损场景。

点刹式流控通过周期性扫描无损队列的缓存占用情况，计算一个周期 *t* 内需要上游设备停止发送流量的时长 t_{stop}。若 t_{stop}>0，则通过向

上游设备发送带反压定时器的 PFC 反压帧，控制上游设备在对应周期内停流时长 t_{stop} 后再发送流量。（一个周期时长为 t，t 远小于两设备间的单向转发时延，即 $0 \leqslant t_{stop} \leqslant t$）

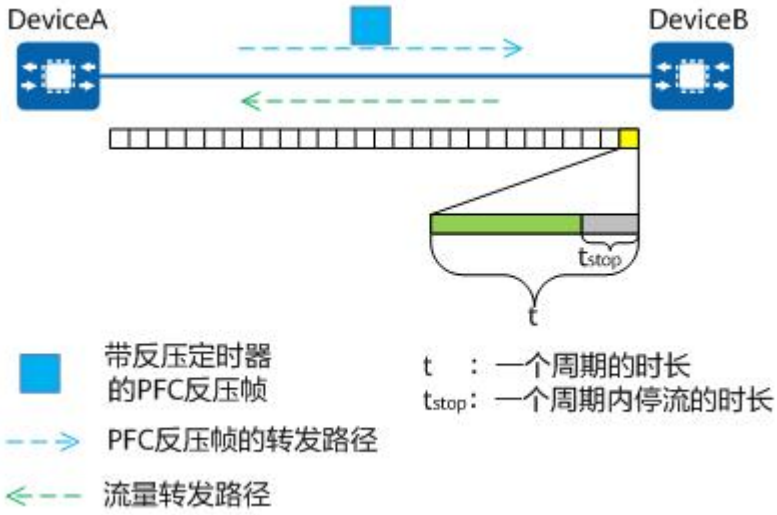

图 3-23 点刹式流控工作原理图

无损队列的 Headroom 缓存空间用于存储本队列发送 PFC 反压帧之后到停止接收上游设备发送的报文这段时间内收到的报文，以防这段时间内的报文被丢弃。根据上文对传统 PFC 流控机制的分析，传统 PFC 的反压帧触发门限值大小至少为 BPFC×2TPFC，Headroom 缓存空间大小至少为 BPFC×2TPFC，因此缓存空间占用至少需要 2×（BPFC×2TPFC）。（BPFC 表示接口带宽，TPFC 表示两端设备之间的单向转发时延）

点刹式流控没有反压帧触发门限，当缓存占用超过阈值 threshold 时，点刹式流控开始生效。从一个周期结束时发送反压帧到该反压帧控制的流量发送到本设备需要经过的时长为 2TABS，为保证无损队列无丢包且不欠吞吐，Headroom 缓存空间大小至少需要 BABS×2TABS，因此缓存空间占用至少为 BABS×2TABS+threshold。

为了达到最小缓存支持最大长距的效果，建议阈值 threshold 设置为 0，此时点刹式流控的缓存空间占用约为 BABS×2TABS。（BABS 表示接口带宽，TABS 表示两端设备之间的单向转发时延），相对于传统 PFC，TABS 可以支持更大的传输时延，即支持更长的距离。

（2）拥塞控制技术

拥塞控制是指对进入网络的数据总量进行控制，使网络流量维持在可接受水平的一种控制方法。其与流量控制的区别在于，流量控制作用于接收者，而拥塞控制作用于网络，典型技术如 ECN。

1）ECN 技术

ECN 在接收端感知到网络中发生拥塞后，通过协议报文通知发送端降低发送速率，从而在早期避免拥塞而产生丢包。为使接收端能够感知网络拥塞，IP 报文中定义了 ECN 字段，并由中间交换机修改 ECN 字段以实现对接收端的拥塞通知。根据 RFC 791 定义，IP 报文头 ToS（Type of Service）域由 8 比特组成，其中比特 0～5 为 IP 报文的 DSCP（Differentiated Services Code Point），比特 6～7 为 ECN 字段，如图 3-21 所示。协议对 ECN 字段进行了如下规定：

- ECN 字段为 00，表示该报文不支持 ECN。
- ECN 字段为 01 或者 10，表示该报文支持 ECN。
- ECN 字段为 11，表示该报文的转发路径上发生了拥塞。

因此，中间交换机通过将 ECN 字段置为 11，以通知接收端本交换机处发生了拥塞。

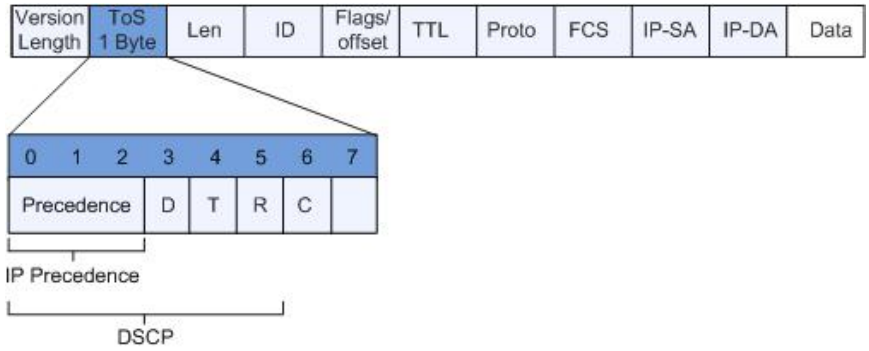

图 3-24 IP Precedence/DSCP 字段

当接收端收到 ECN 字段为 11 的报文时，表明网络中发生了拥塞，则向发送端发送协议通告报文，发送端在收到通告报文后，则降低报文发送速率，避免网络中拥塞加剧。当网络中拥塞解除后，接收端不会再收到 ECN 字段为 11 的报文，则停止向发送端发送协议通告报文，因此发送端收不到协议通告报文，则认为网络中无拥塞，并会恢复报文的发送速率。

2）AI ECN

无损队列的动态 ECN 门限可根据网络流量 N 对 1 的 Incast 值、大小流占比来动态调整，从而在避免触发 PFC 流控的同时，尽可能兼顾时延敏感小流和吞吐敏感大流。然而，现网中的流量场景复杂多变，传统动态 ECN 门限功能难以覆盖所有场景。对此，AI ECN 功能被提出，其则可根据现网流量模型进行 AI 训练，从而对网络流量的变化进行预测，并根据队列长度等流量特征调整 ECN 门限，进行队列的精确调度，保障整网的最优性能。

如下图所示，设备会对流量特征进行采集并上送至 AI 业务组件，AI 业务组件将根据预加载的流量模型为无损队列设置最佳 ECN 门限，保障无损队列的低时延和高吞吐，让不同流量场景下的无损业务性能

都能达到最佳。具体操作如下：

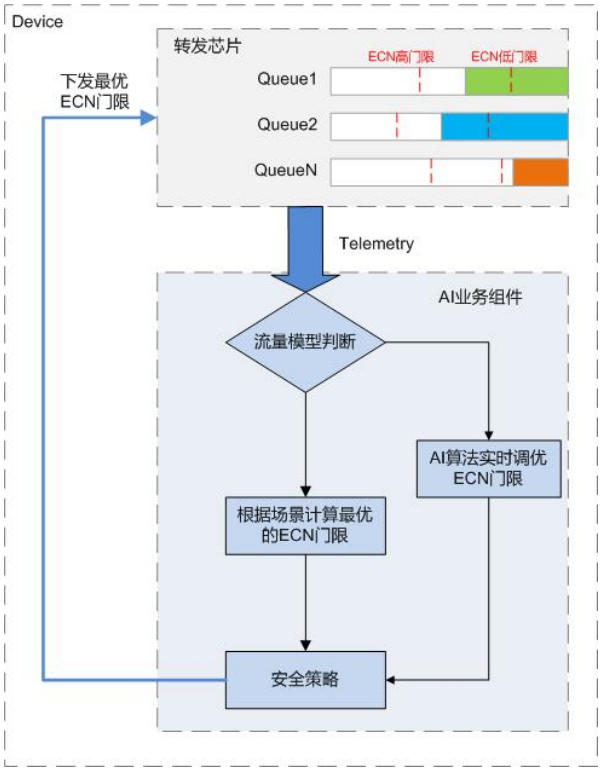

图 3-25 无损队列的 AI ECN 功能实现

- Device 设备内的转发芯片对当前流量特征进行采集，如队列缓存占用率、带宽吞吐、当前 ECN 门限配置等，然后通过 Telemetry 技术将流量实时状态信息推送给 AI 业务组件。
- AI 业务组件收到流量状态信息后，根据预加载的流量模型文件对当前流量场景进行识别。若为已知流量场景，AI 业务组件将基于大量的 ECN 门限配置记忆样本，推理出与当前网络状态匹配的 ECN 门限配置。
- 若为未知流量场景，AI 业务组件将结合 AI 算法，在保证高带宽、低时延的前提下，对当前的 ECN 门限不断进行修正，最终计算出最优的 ECN 门限配置。
- AI 业务组件将符合安全策略的最优 ECN 门限下发至设备，

设备完成无损队列的 ECN 门限调整。

- 对于获得的新流量状态，设备将重复上述操作以保障无损业务的最佳性能。

同时，与拥塞管理技术（队列调度技术）配合使用时，无损队列的 AI ECN 门限功能可实现网络中 TCP 流量与 RoCEv2 流量的混合调度，在保障 RoCEv2 流量无损传输的同时实现低时延和高吞吐。

3）端网协同拥塞控制

高性能计算、AI 模型训练、以及数据中心网络，均要求网络传输排队时延和吞吐能获得进一步提升。为达到极低传输时延，应尽可能降低网络设备内的排队时延，同时维持接近瓶颈的链路满吞吐。对此，端网协同拥塞控制技术采用“端（智能网卡）网（交换机）”配合的方式实现交换机内“近零排队”时延，进而使得端到端传输时延接近静态时延。

早期的端到端拥塞控制方案都属于被动控制，即源端在拥塞发生前不断尝试提高发送速率，但这可能导致网络排队进而形成拥塞。在拥塞情况反馈到源端后，源端才被动降低发送速率。由于反馈存在一定的延迟，因此被动拥塞控制往往导致网络在拥塞和非拥塞状态之间震荡。相比之下，端网协同根据网络可用带宽来调整端侧发送速率，这种源端与交换机之间的密切配合使得网络中的队列近乎为空，并且能够保持接近 100%的带宽利用率。

根据实验室测试：采用典型拥塞场景哑铃状拓扑，瓶颈链路带宽为 100Gbps，瓶颈链路上存在 200 条长流，评估不同算法在瓶颈链路

上的排队时延（μs）。如下表可见，与业界主流拥塞控制算法 HPCC、DCQCN 相比，端网协同算法 C-AQM 能够显著降低排队时延，同时使瓶颈链路达到接近 100%利用率。

表 3-1 不同拥塞控制算法性能对比对比

流数 200	C-AQM	HPCC	DCQCN
50%-ile	0.155	3.023	116.612
90%-ile	0.238	6.662	121.82
99%-ile	0.321	8.204	125.48
99.9%-ile	0.401	9.094	127.131

（四）网络负载均衡技术

在 AI 大模型场景中，业务流量呈现出大象流、低熵、同步效应等特征，并进一步导致传统 ECMP 基于流的五元组哈希机制失效，其产生的哈希冲突将引发链路拥塞直至丢包，进而劣化集群的整体性能。针对该问题，业界提出了许多新型负载均衡技术，按照技术原理大致可分为两类：基于网络状态感知的负载均衡技术，以及基于流切分技术的负载均衡技术，下文将对这两类技术进行阐述：

（1）基于网络状态感知的负载均衡技术

针对传统 ECMP 机制的不足，一种解决思路是将“网络状态”作为负载均衡选路决策的依据之一。这种“网络状态”可以是设备本地的负载状态，也可以是“网络全局”的负载状态。该负载均衡技术的特点是保证同一流走相同路径的基本特征，不会造成数据包的乱序

问题，其技术原理如下：

- 基于本地负载状态的均衡技术：区别于基于五元组 HASH 结果选路的传统 ECMP 机制，基于本地负载状态的均衡技术会读取本地出接口的队列、发包统计等信息作为报文转发的依据，通过感知拥塞状态的方式保持网络流量分布的均衡性。典型技术是动态负载均衡（DLB，Dynamic Load Balancing），其基本原理是交换机在进行 ECMP 选路时，不再随机或者轮询，而是通过出接口负载辅助选路，如选择综合负载最小的链路，也可通过队列深度、接口带宽利用率等作为拥塞程度的量化依据。这类负载均衡技术带来的性能收益是当网络存在等价多路径时，使不同路径上的负载更为均衡。

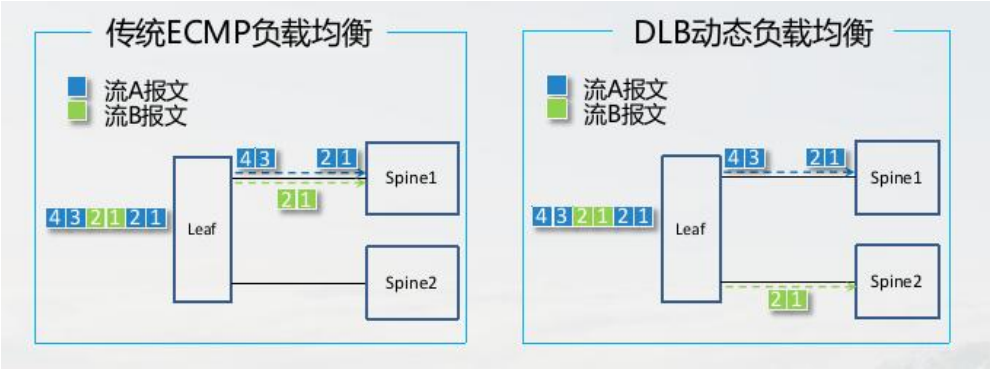

图 3-26 DLB 技术的整体效果

- 基于全局负载状态的均衡技术：在相对复杂的多级网络中，流量发送端通常无法感知网络的全局状态信息，这种状态包括下游的拥塞状态、网络整体的流量分布及带宽利用率等。基于全局负载状态的均衡技术的基本原理是先收集网络全局状态（状态收集可通过集中式的网络控制器或分布式的数据面协议实现），然后再基于全局状态信息进行流量规划，避免

局部拥塞的同时最大化整网的吞吐性能。基于全局的负载均衡涉及到一些私有协议和算法的制定，目前尚缺乏统一的标准，大多为厂商私有化实现，典型技术方案如中兴的智能全局负载均衡技术，其利用 AI 训练可预测、周期性迭代的特点，控制器通过 API 接口被动接收 AI 调度平台的流信息（如五元组、通信数据量等），通过集中 TE 算法将活跃的数据流均衡预规划到 Fabric 网络中，再下发配置到网络设备以实现流量工程。

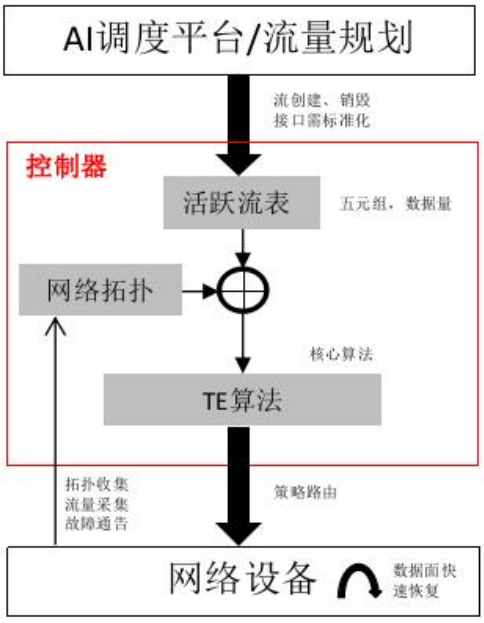

图 3-27 智能全局负载均衡基本原理

此外，基于负载状态的均衡技术通常和 Flowlet 负载均衡技术组合使用，例如网络识别大象流，对大象流进行 Flowlet 切分后依据负载状态进行均衡转发。

（2）基于流切分技术的负载均衡技术

另一种有效的负载均衡优化思路是将数据流切分为更小的单元，

不同单元可走不同的网络转发路径，从而达到网络负载均衡的目的。其优势在于可实现更佳的负载均衡效果，但会引发数据包的乱序问题，因此需要额外引入乱序处理功能。按照切分的粒度不同，当前常用实现方式包括 Flowlet、容器、包、以及信元等。

- 基于 Flowlet 的负载均衡

Flowlet 负载均衡用于解决数据中心网络内流量不平衡问题。它通过将流量切分成更小的单位（即 Flowlet），然后在多条路径上进行均衡分发。

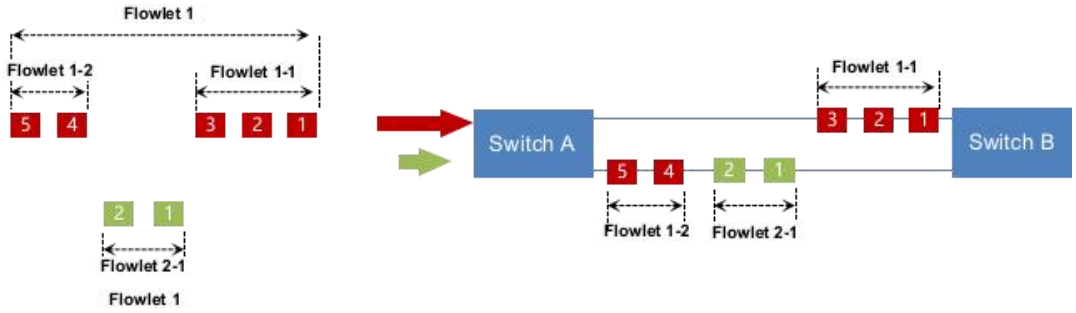

图 3-28 Flowlet 的基本原理

Flowlet 负载均衡的核心思想是利用流的规模与持续时间来分割流量。当流量到达网络设备时，设备会将其分割成多个 Flowlet，每个 Flowlet 包含一部分数据包。然后，设备会使用负载均衡算法将这些 Flowlet 分配到不同的路径上进行传输。

Flowlet 技术提出较早，是一种被普遍支持的负载均衡技术，但也存在其局限性。首先，为避免引入报文乱序，Flowlet 切分的时间间隔与流量模型紧密相关。但在实际应用中很难避免完全乱序，且 Flowlet 自身并不提供乱序处理能力。其次，这种基于时间间隔的子流切分方式在特殊的流量场景中可能失效，如 AI 模型训练场景等。

- 基于容器的负载均衡

基于容器的负载均衡是中国移动提出的全调度以太网技术的核心思想，其引入了一个虚拟的报文容器概念。报文容器是一个逻辑上的定长单元，其容纳的报文数量可依据业务报文长度的分布情况进行调整，要求至少能够容纳 1 个最长的业务报文，且总长度在芯片转发能力和乱序能力允许的情况下尽可能短，以达到精细切分数据流、充分提高瞬间负载均衡度的目的。报文在转发过程中仅依据 GSE 头部的控制信息实现快速转发，确保同一个容器内不同报文走相同的网络路径，不同容器的报文走不同的网络路径，在网络出口再根据控制信息实现最小代价的乱序重排。基于容器的负载均衡技术希望在网络的均衡度和报文乱序度之间寻找最佳的平衡。

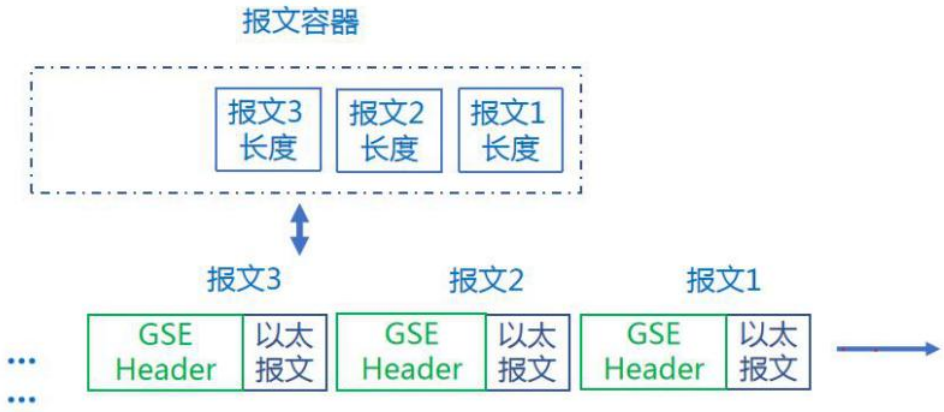

图 3-29 GSE 容器调度的基本原理

- 逐包负载均衡技术

在网络流量转发过程中，包通常是最小的转发单元，因此基于逐包的负载均衡技术理论上可达到最优的均衡效果，但同样会面临报文乱序问题。在典型的 AI 训练场景，报文乱序处理通常由支持 RDMA

报文乱序重排的智能网卡实现，如 NVIDIA BlueField 网卡等。

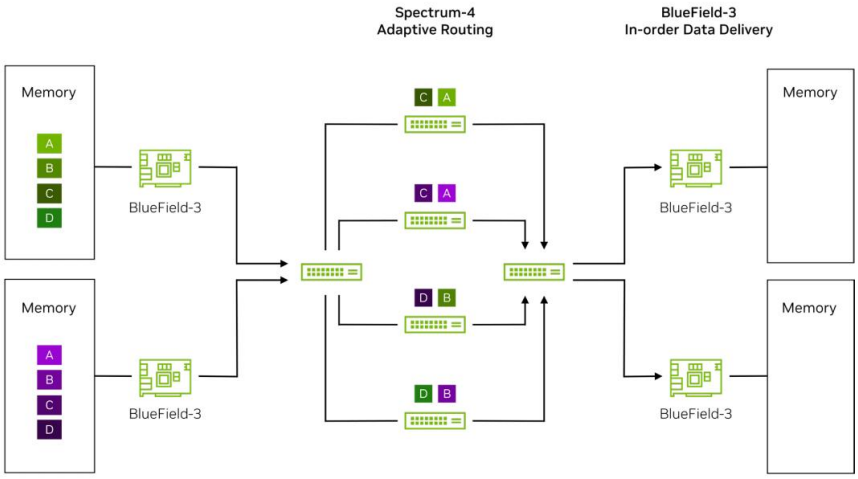

图 3-30 NVIDIA 逐包负载均衡的基本原理

- 基于信元交换的负载均衡技术

另一种代表性的切分方式是将报文切分成更小粒度的信元，然后以信元为单位进行负载均衡。这种负载均衡技术相对特殊，其打破了以太网转发的基本逻辑，本质上是一种全定制化的网络方案，依赖特定的芯片硬件实现，典型的代表是分布式解耦机框（DDC，Disaggregated Distributed Chassis）方案。DDC 通过网络硬件将数据包切分成理论上等长的小信元，并在不同的上行链路进行负载均衡转发，再由出口节点依据信元的控制信息进行报文重组和乱序处理。

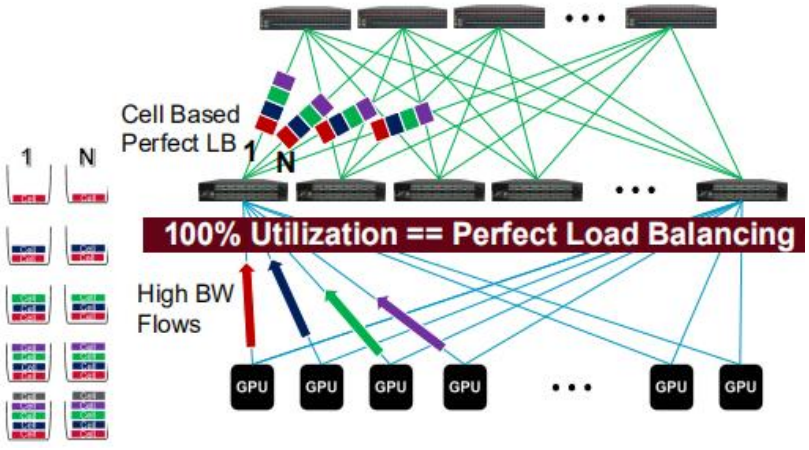

图 3-31 信元交换的基本原理

该技术优势是可实现近乎完美的负载均衡效果，技术成熟且有完善的国产芯片支持，如中兴通讯适用于 DDC 的自研芯片可提供高达 14.4T 的转发能力；缺点是技术的落地依赖特定厂商的芯片组，整体方案相对较封闭。2022 年，由紫金山实验室、中国电信研究院和北京邮电大学联合发布的《分布式解耦机柜技术白皮书》，也研发了 DDC 方案，并进行了部署实践。

- 先进哈希算法

静态 ECMP 由于只使用五元组进行哈希，由于哈希因子少导致熵值低，因此可以纳入更多的数据包字段作为哈希因子，例如 RDMA 头部中的 QP 对（Queue Pair）信息，甚至是用户自定义的字段，以增加哈希熵值，从而实现更加均衡的负载分配。

（五）端网协同的 NetMind 跨层通信架构

（1）NetMind 概述

智算场景中网络流量呈现出周期性、长流等特征，其体量随着大模型规模扩大而快速增长，因此需要多种并行策略的组合使用，同时尽可能降低通信开销。然而，传统的并行与优化策略只针对单一主体，如网络或 AI 模型，而未来智算网络中应将两者结合，实现网算协同的调度优化。对此，面向 AI 大模型集群的 MetMind 通信架构被提出，其通过网算协同实现系统整体性能的提升，主要涉及两个方向的协同：

1）水平方向：

- 从网到算：网络向 NetMind 提供动态拓扑（除静态网络拓扑

外，还包括光模块链路中断、网络资源占用情况等），再由NetMind 将全局拓扑提供给计算组件，进行拓扑亲和计算。

- 从算到网：由于 AI 大模型的流量规律特征明显，因此可根据该规律提前部署网络路由及 QoS 策略。具体而言，基于计算侧通信需求，经由 NetMind 通告网络进行主动均衡、拓扑调整及 QoS 资源分配等。从而确保计算侧的调度、并行策略、通信算法执行时，AI 模型通信流量与网络拓扑、当前链路状态相匹配。

2）垂直方向：

根据 NetMind 提供的集合通信算法的适用区间和通信效率进行作业调度，并基于 NetMind 提供的建模求解优化策略提升自动模型切分的速度与效果。

（2）NetMind 通信架构

NetMind 旨在为 AI 大模型系统中的不同用户提供统一框架以提升性能。如下图所示，整个系统分为 NetMind Server 和 NetMind Client 两个核心组件。其中，NetMind Client 部署于每台计算服务器的主机侧，包含一个部署在计算节点 CPU 上的 Agent 进程。Agent 从 AI 作业进程调用的 NetMind Runtime 获取到相关信息（如通信域、通信算子、通信算法等），并统一上报至 NetMind Server。NetMind Server 作为一个进程部署在一个独立的服务器内，负责聚合全局信息、与网络侧交互、以及提供各业务辅助决策模块。

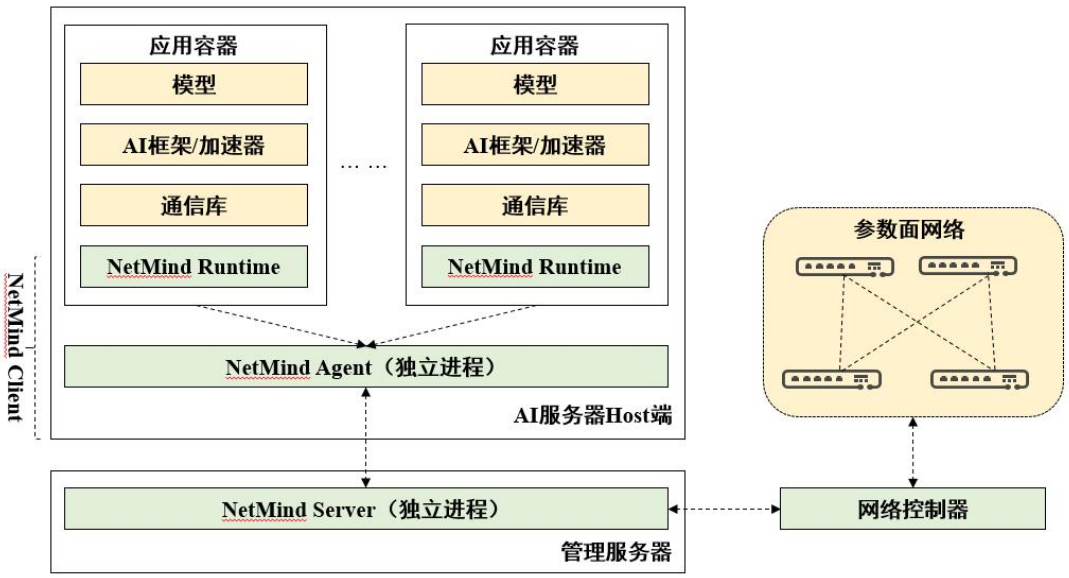

图 3-32 NetMind 通信架构

（3）业务辅助决策模块

1）拓扑感知模块

首先，NetMind 分别从网络侧和计算侧获取到参数面物理组网拓扑、链路资源占用率和模型切分信息，在拓扑感知模块中完成物理拓扑到模型通信域逻辑拓扑的映射。其次，基于该逻辑拓扑的分层关系，通过逻辑同号卡拆分解决同层 group 间不对称、层内训练卡数量非 2 的幂次方等造成的低效率通信问题，并通过时分复用原理解决分层间带宽收敛的问题。

2）NSLB 网络均衡模块

NetMind 从计算侧聚合出以训练任务为粒度的通信矩阵，再根据通信矩阵中每轮迭代的每个阶段内各训练卡间的通信关系，对网络资源进行逐流的全局算路规划，从而避免传统 HASH 负载不均导致的链路利用率低以及训练性能下降等问题。

3）AI QoS 模块

NetMind 分别从网络侧和计算侧获取网络状态信息和模型训练的并行策略、通信量等信息，并在 NetMind Server 应用感知 QoS 模块中计算不同并行策略间流量的最优 QoS 调度方式，然后结合业务逻辑下发至网络侧，实现对 QoS 调度的动态控制，同时减少模型训练过程中不同并行策略的流量对网络资源的竞争，以提升业务性能。

4）OXC 拓扑调整模块

针对智算场景中光电混合组网方案，先由 NetMind 感知计算侧的模型切分策略、通信域信息、通信算子、通信算法等信息，随后由光交叉连接（OXC，Optical Cross Connect）模块基于灵活无损的拓扑工程技术，快速计算出符合当前业务场景的最优拓扑，并指导光路进行无损重构。同时，基于网络级负载均衡（NSLB，Network Scale Load Balance）模块对流量进行全局路径规划，实现任务性能最优的流量路径编排。

四、智算集群间网络关键技术

（一）光电融合组网与路由技术

当前广域网在实现智算集群互联时面临设备成本与功耗高昂、容量受限、质量难承诺等挑战，究其原因在于光传送与数通领域长期独立发展，未能形成有效合力。对此，广域网技术体系应进一步面向光电融合演进升级，从传统带宽驱动的通道式网络向业务驱动的光电融合定制化网络演进。

光电融合承载已成为智算网络发展的必然趋势：一方面，随着相干光模块的持续小型化与市场化发展，使得路由器可直接通过可插拔彩光模块与光分叉复用器（ROADM，Reconfigurable Optical Add/Drop Multiplexers）相连，无需配置大量波长转换板卡，从而显著节约部署空间、降低设备成本与功耗，如图 4-1 所示。此外，通过引入先进的高阶调制、纠错算法以及光放技术，使得单波容量与无电中继传输距离得以提升；另一方面，业界在持续推动 ROADM 设备的软硬件解耦与接口开放，并支持设备功能的软件驱动，使得软件控制的 IP+光融合网络成为可能，从而极大提升系统灵活性与承载质量。因此，当前网络形态正经历从传统 IP+WDM 的光电复合式组网向软件定义 IPoWDM 的光电一体化组网转变。

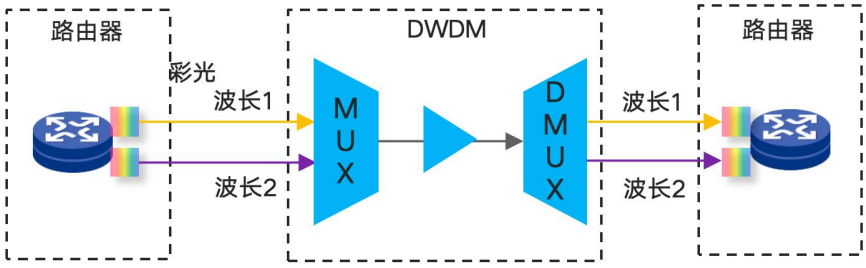

图 4-1 光电融合路由设备形态

如图 4-2 所示，融合网络虽从设备形态上趋于一体化，但从逻辑角度仍分为电层与光层，其中电层对应 IP 路由器，负责数据包的汇聚转发以及电层时隙（如 FlexE）的映射与交叉；光层对应光纤、光模块、ROADM、放大器等器件与设备，负责提供大容量、低时延传输通道与多波带光交叉能力。在设备上线后，控制面会将其所有端口能力纳入所在节点的资源池，并屏蔽底层技术细节，以一虚多或多虚一的方式为业务提供适配的资源粒度。在路由方面，融合网络将形成如 4-2 左图所示的光电一体化网络拓扑视图，其中 IP 层与光层节点间、光层节点间为物理链路（表征客观存在的设备连接与光纤连接），IP 层节点间为虚拟链路（表征 IP 层连接，可动态建立或释放），并针对不同网络状态与业务场景提供多元算路约束与拓扑权重（如时延优先、功耗优先等），支持光层直通与光电混合转发等多种自适应传输模式。

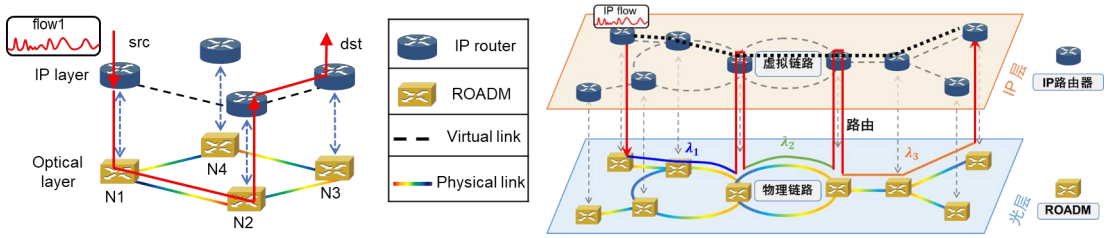

图 4-2 光电融合组网与路由技术

光电融合网络采用集中化控制方式，实现对全局光电资源的统一

编排与控制，提升网络节点部署敏捷性与运维简便性。控制平面开放网络可编程接口，其中南向接口控制转发面骨干网络设备，北向接口提供智算应用定制网络能力的入口，从而实现智算应用与网络的无缝集成。针对不同智算业务承载需求，光电融合控制平面提供多种网络调度与规划能力，支持快速业务部署、联合故障规避、动态光电调度等功能，具体如下：

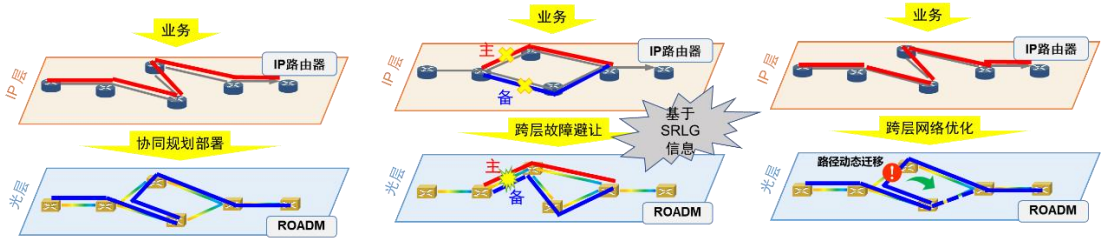

图 4-3（a）快速业务部署；（b）联合故障规避；（c）动态光电调度

（1）快速业务部署：传统网络中 IP 层与光层的调度规划相对独立，存在业务部署时耗长、资源利用低效、拓扑构建静态化、流量分配不均衡等问题，从而引入额外的建设投资与运维开销。如图 3-37（a）所示，通过采用光电融合路由技术，可实现高效的双层协同规划与快速业务部署。规划方可将当前网络配置数据导入编排系统，并根据智算业务需求选择所需的算路因子，进行双层一体算路。编排系统还可迅速响应业务需求变化，并快速调整端到端路由及光电资源配置策略，例如快速无损带宽调整等。

（2）联合故障规避：传统网络中 IP 层与光层保护相互独立，存在保护资源冗余的问题，进而增加网络运营成本。此外，传统网络还面临在同一个共享风险链路组（SRLG，Shared Risk Link Group）中部署主备路径的风险，导致主备路径同时遭受故障冲击而造成业务中

断。对此，光电融合路由技术基于一体化拓扑视图，一方面从全局视角统筹解决了保护资源冗余的问题，另一方面从根本上避免了主备路径的相交问题。此外，光电融合网络还基于数字孪生技术在虚拟环境中投射物理网络拓扑及连接，并通过不断模拟各类故障情况构建知识库，训练故障定位的 AI 模型，从而在发生故障时快速定位故障源，同时也可用于分析网络的潜在薄弱点。

（3）动态光电调度：传统网络更倾向于离线规划与静态调度，然而随着业务场景多元化以及新旧业务请求更替频繁化，使得网络环境趋于高动态性，因此引入面向光电域的动态联合规划尤为重要。如图 4-3（c）所示，一方面，通过推拉结合的方式监控网络流量变化，并根据预设门限自动进行带宽扩缩容或流量迁移。另一方面，对于新生业务请求，支持在已有路径中映射所需资源，或者快速触发网络新建光电路径。

（二）广域拥塞控制技术

随着国家“东数西算”战略的推进，东部数据可通过广域网传输至西部智算中心进行处理，实现各地区间算力资源的优化配置。然而，广域长距传输可能会对业务性能产生一定影响，这就要求引入拥塞控制技术以实现高效的数据搬移。下文将聚焦智算网络场景，对传统广域拥塞控制技术面临的新挑战进行分析：

- 高带宽利用率：在长距传输中，带宽利用率将直接影响数据传输效率和成本，提高利用率即可在单位时间内传输更多数

据，从而延缓扩容需求并降低成本；

- 低丢包率：丢包会导致数据重传，从而占用额外带宽资源并降低传输效率，该问题在超长距传输环境中尤为突出；
- 传输延迟及网络状态反馈滞后：数千公里的传输距离引入不可忽视的传输时延，这使得网络状态反馈存在一定滞后性。传统基于丢包的拥塞控制算法（如 Cubic 算法）在长距离传输中表现不佳，其带宽利用率偏低且丢包率较高；
- 光纤传输的错包问题：超长距光纤传输中，错包问题难以完全避免，这使得数据传输的完整性和可靠性面临挑战。

为解决上述问题，本节将从传输层与链路层角度对广域拥塞控制技术进行阐述。

（1）广域 TCP 拥塞控制

目前主流跨域传输算法均基于 TCP 协议实现，根据具体应用场景可分为跨智算中心传输和跨广域网传输。现有的一些拥塞控制算法有 Cubic、BBR 、PCC Vivace、Copa 等。Cubic 是当前 Linux 中默认的 TCP 拥塞控制算法，其将标准线性窗口增长函数修改为三次函数，以提高 TCP 在远距网络上的可扩展性。为应对长距传输面临的高带宽时延积问题，Cubic 将往返时延（RTT，Round-trip Time）大小与窗口增长解耦，实现公平带宽分配和稳定广域传输；BBR（Bottleneck Bandwidth and Round-trip Time）算法基于 RTT 和 ACK 进行传输控制，通过带宽估计提升带宽利用率并降低传输时延，其在 Google B4 骨干网络中已部署应用；PCC Vivace 通过结合 PCC 基本框架与机器学习

中的在线凸优化理论，提出了一种基于学习的网络拥塞控制算法，通过调整发送端速率的调整方向、调整步长与调整阈值，解决拥塞控制问题；Copa 提出了三种具体的控制方式，能够根据目标速率调整当前发送速率，并迅速将流量收敛至合适的公平速率。Copa 可有效应对长距传输中的交换机缓存限制，同时不影响存在交叉路径的其它流量传输。

上文针对端侧的广域拥塞控制机制进行了论述，下面将从算网协同角度出发分析多种新型拥塞控制机制，包括 Annulus、Gemini、GTCP、IDCC 等。。为解决广域网中长往返时延和智算中心交换机缓存有限的问题，Annulus 使用双控制回路进行拥塞控制。一条控制回路处理广域网或智算中心内的流量瓶颈，另一条控制回路管理近源端广域网和智算中心流量的共存问题，实现跨智算中心的拥塞反馈。在发送端，选择双控制循环中较低的发送速率作为实际发送速率，并通过第二条循环迅速感知智算中心内部的拥塞情况以做出快速反应；由于广域网与智算中心内的拥塞信号存在差异性，因而单独使用任一种信号都无法实现拥塞响应，对此 Gemini 整合 ECN 和延迟信号进行跨智算中心拥塞控制，实现了高吞吐的跨域传输；为解决智算中心内网和智算中心间网络的异构问题，混合拥塞控制（GTCP，General Transmission Control Protocol）将反应式和主动式两种拥塞控制进行结合。前者由发端驱动，通过不断探测网络可用带宽，并在收到拥塞信号（丢包、时延增大、冗余确认报文等）后对发送速率进行调整。后者由收端驱动，通过令牌方式进行网络带宽“预留”，从而尽量避免拥塞发生。

GTCP 对智算中心内流量使用主动式拥塞控制，而对智算中心间流量先使用反应式机制探测网络带宽，当发端收到网络 ECN 信号后切换为主动式拥塞控制，从而避免交换机上积累过多数据包；IDCC（INT and Delay based Congestion Control）是一种基于时延的跨域拥塞控制机制，通过带内遥测与 RTT 分别测量广域网与智算中心内的排队时延，并通过比例积分微分（PID，Proportional Integral Derivative）调控排队深度，最终提升算力中心间吞吐量。

（2）广域链路流量控制

为实现精细化调控，广域流量控制技术正向细粒度化方向发展。现有的一些流量控制机制有 BFC、Floodgate、CaPFC、P-PFC、GFC 等。BFC（Backpressure Flow Control）提出了在交换机上维护各流状态以实现逐跳的流量控制，交换机根据流标识的哈希值将数据包放入不同队列，在哈希冲突较少时，各流可占据独立队列，从而减少队头阻塞带来的影响；Floodgate 与 BFC 采用相似思想，但 Floodgate 是以目的 IP 地址为对象进行队列隔离，来对入播流量进行快速监测和控制。Floodgate 采用信令机制，下游交换机定期向上游发送累计信令值来通告队列长度，以此控制上游交换机的发送和暂停；CaPFC（Congestion aware Priority Flow Control）是基于 PFC 的改进型流控机制，同时监测入队列与出队列的长度。当出队列长度超过阈值后，对各端口进入交换机的数据包数量进行统计，从中选出值最大的入队列来发送暂停帧。同时使用出队列与入队列长度进行流控，可快速地在交换机内部传递拥塞信息，以提升流控对拥塞的敏感程度；P-PFC

（Predictive PFC）指出瞬时的队列长度难以完全反应当前拥塞情况，对此提出使用队列变化速率（即队列长度随时间变化函数的导数）与队列长度相结合来进行流控，因为通过速率变化可以更快地发现拥塞的发生和解除，从而实现更高效的流量控制；GFC（Gentle Flow Control）是针对智算中心网络死锁问题的流控机制。区别于 PFC 完全暂停流发送的方式，GFC 基于预设函数从下游入队列长度来推导上游发送速度并通知上游，因此尽可能避免了流被完全暂停，打破了死锁发生条件，提高了无损网络的健壮性。

（三）广域 RDMA 技术

RDMA 是从 DMA 技术演进而来，最早由 IBTA（InfiniBand Trade Association）组织基于 InfiniBand（IB）架构而提出。其允许计算机系统直接访问远程计算机内存而无需 CPU 参与，从而显著减少通信延迟和 CPU 开销，是一种高性能网络传输技术。但由于原生 RDMA 技术需要专用且昂贵的网络设备才能运转，这极大限值了其普及率，因此基于以太网的 RoCE、RoCEv2 和 iWARP 协议相继出现，并在局域网中得到了广泛应用。随着国家推行东数西算、海量科学计算等应用场景，智算中心间的互传数据呈爆发式增长，使得传统网络传输技术劣势越发凸显，因此 RDMA 从局域网迈向广域网已成为一个重要趋势。

（1）广域 RDMA 技术架构

广域 RDMA 技术为最大限度地利旧现有网络设备与线路，因此

在 IEEE802.3 基础上使用 IP 协议进行传输，采用 RoCEv2 或 iWARP 协议封装 RDMA 数据载荷，具体应用架构如图 4-4 所示。

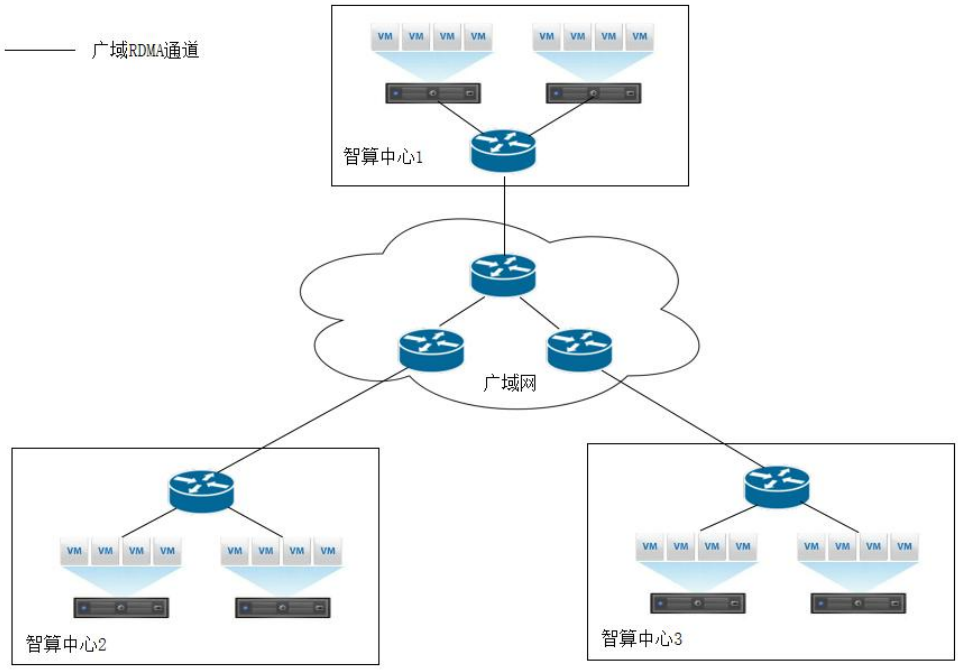

图 4-4 广域 RDMA 应用场景

智算中心间基于现有广域网进行互联，采用广域 RDMA 技术承载海量数据。RDMA 网卡（含 DPU）直接部署在物理机中，通过虚拟化等技术供虚拟机使用，以减少操作系统协议栈、Hypervisor 层的性能开销。出于对广域网复杂多变的传输环境及“尽力而为”的承载方式考虑，在广域网上进行长距离 RDMA 传输容许轻微有损，以保证带宽和时延的高要求。

（2）广域 RDMA 技术

原生 RDMA 是基于 InfiniBand 架构的网络传输技术，不兼容现有主流的以太网网络传输协议栈，因此在广域 RDMA 传输上优先使用 RoCEv2 或 iWARP 进行传输。

- **RoCEv2**

RoCEv2 是基于 UDP 协议的 RDMA 传输技术，是由 IBTA 组织主导的一套 RDMA 传输协议，具体协议栈如图 4-5 所示。RoCEv2 在广域网上进行传输遵循 IBTA 组织定义的协议栈格式，可概括为 IB Over UDP，该结构决定了其可基于以太网协议进行传输。

图 4-5 RoCEv2 协议栈

由于 UDP 的不可靠性，在传输过程中会面临丢包问题，而且实验显示，进入广域网后，RTT 随着传输距离变长而增加，丢包对 RoCEv2 的吞吐的影响变得更大。因此 RoCEv2 需要承载于丢包较少的广域网中，配合 Go-Back-N 或 SACK 机制来检测丢包和重传，并使用大缓存的网络设备，以减少丢包概率。同时，为充分发挥 RDMA 的高吞吐量传输优势，还需及时感知网络负载、拥塞等情况，及时调整网络负载和流量控制等策略。

此外，在广域多路径传输情况下，一方面需引入包括最大带宽、最小时延、链路丢包率等在内的多因子算路机制来选择最优路径；另一方面需选择合适的负载均衡技术，由于广域网不同路径间距离差别较大，会导致数据包传输时延差距过大，而基于逐包的负载分担会给端侧设备引入较大的乱序处理压力，因此多路径负载分担技术推荐基

于逐流机制。同时基于对大象流的识别，合理调整多路径上的流量分布，并配合 PFC、ECN 及更精细的流级别拥塞控制算法等流控技术，可有效避免拥塞，提升传输效率。

- **iWARP**

iWARP 是由 IETF 组织发布的一套 RDMA 传输协议，区别于 IBTA 组织，其 iWARP 协议和 InfiniBand 无法兼容，具体协议如图 4-6 所示。

图 4-6 iWARP 协议栈

iWARP 是基于以太网与 TCP/IP 协议的 RDMA 技术，由于 iWARP 并未指定物理层信息，因此能够运行在任何使用 TCP/IP 协议的网络上。由于 iWARP 基于 TCP 进行传输，对于质量较差的广域网可优先选用此协议，通过 TCP 面向连接的可靠传输机制，减少广域网丢包，并配合 DCTCP 等拥塞控制算法，保证传输的高效性。

（四）新型低损光纤技术

为满足不断升级的算力业务需求，业界正在持续探索 200G/400G、

800G 及以上的高速光传输方案。同时，为满足高速光传输下的超长距离、低损耗、低延时需求，部署新型超低损光纤也成为必然趋势。

（1）G.654.E 低损光纤

G.654.E 光纤兼具超低损耗和大有效面积特性，相较于常规的 G.652 光纤，可显著提高 100G、200G、400G 及未来更高速率网络长距离传输性能。因此，G.654.E 光纤被公认为是下一代超高速长距离光传输性能优化的有效解决方案。

G.654.E 光纤属于新型截止波长位移单模光纤，符合 G.654.E 标准。该标准由国际电信联盟电信标准化部（ITU-T）于 2016 年 11 月发布，是 ITU-T G.654《截止波长位移单模光纤光缆的特性》的最新版。该标准自 1988 年发布以来，历经多次修订，其中包括 G.654.A、G.654.B、G.654.C、G.654.D，主要应用于海缆通信系统。同 G.654.A、G.654.B、G.654.C、G.654.D 光纤一样，G.654.E 光纤具备超低损耗、大有效面积的特点，但其独特优势在于工作温度、宏弯损耗等方面。前面四类光纤主要应用于温度恒定在-1℃~2℃之间的海洋环境，而 G.654.E 光纤适用于陆上网络，环境温度可从-65℃变化至 85℃。此外，G.654.E 光纤可抵抗各类应力，具备极佳的抗弯性能，以应对陆地复杂环境中的环境压力、弯曲应力、机械冲击等。根据上述特点，G.654.E 光纤尤其适用于陆上长距离高速传输网络：

1）提高光信噪比值：光信噪比是影响光传输质量的重要参数之一。由于 G.654.E 光纤具有极小的宏弯衰减和大有效面积，能够有效保持纤芯中的光功率，并使其更为分散地传播，从而一定程度缓解光

信噪比随传输距离降低的问题。

2）延长无电中继传输距离：G.654.E 光纤增加了纤芯尺寸，由此实现了大有效面积，使得光纤可传输更高的光功率。因此，与常规 G.652 光纤相比，该光纤可将光传输距离延长 70%-100%。

3）降低网络部署成本：单从光纤本身而言，G.654.E 比 G.652 造价更高，但网络部署并非只有光纤，因此整体成本反而会降低。这是因为 G.652 光纤的无电中继传输距离较短，网络中需部署更多的光中继站，而 G.654.E 光纤则可有效减少中继站数量与成本。

（2）空芯光纤

传统实芯光纤发展成熟、应用广泛，但因基质材料的本征限制，如材料的吸收、色散、非线性等属性，逐渐难以适应相关行业的发展。为突破传统光纤的局限，业界提出了纤芯为空气的空芯光纤（Hollow Core Fiber，HCF）。空芯光纤的结构相对于传统光纤较为特殊，其通过特定的包层结构，可将光限制在空气纤芯中进行传输，从根本上避免了由于在材料本征限制而引发的一系列问题。

从组成结构而言，空芯光纤与传统实芯光纤一样，由纤芯、包层与涂覆层组成，区别在于纤芯与包层。传统光纤是基于光的全反射原理实现光在玻芯中传播，而空芯光纤的芯为空气，由于空气折射率低于包层介质折射率，因此不满足全反射条件，需采用特殊设计的包层结构。空芯光纤包层是基于微结构的设计，通常由一系列微小的空气孔构成，这些空气孔沿光纤的长度方向排列，具有精确设定的孔径大小、孔间距与周期。当光入射至纤芯与包层界面时，会受到包层中周

期排列的空气孔的强烈散射，这种多重散射产生相干，使得满足特定波长与入射角的光波能够回到芯层中继续传播。因而，通过该种结构，可在纤芯层的折射率低于包层的情况下，实现对光的引导与传播。

空芯光纤对于网络传输可提供如下优势：

1）低延时特性：根据光在折射率为 n 的介质中的传输速度公式 $v=c/n$ 可知，当介质折射率越大时，光传输速度越小。由于空气折射率为 1，因此光在空芯光线中以光速传输，远超在玻璃介质中的速度，从而可极大降低链路传播时延（从 5μs/km 下降至 3.46μs/km）。

2）超低损耗特性：目前空芯光纤可实现 0.174dB/km 损耗性能，与现有最新一代实芯光纤性能持平。但空芯光纤在通信窗口理论最小极限可低至 0.1dB/km 以下，低于普通实芯光纤的理论极限 0.14dB/km，从而支持更长距离的传输。

3）低色散特性：空芯光纤的传输介质是空气，极大降低了材料色散带来的传输损耗。一般而言，空芯光纤的材料色散要低于实芯光纤三个数量级。

4）超低非线性特性：空气芯中光与介质的相互作用减弱，从而减少了非线性效应的产生，可比常规玻芯光纤低 3~4 个数量级，使得入纤光功率可大幅提高。业界设备厂家已基于这一特性展开相关光系统研究，如高阶调制及高功率放大器技术等，预期至少可提升系统容量及传输距离 2 倍以上。

五、智算网络产业典型案例

（一）天翼云昇腾智算项目

1．项目背景

随着生成式人工智能大模型的驱动与发展，教育、电商、游戏、影视、医疗、汽车等行业领域均产生了相关应用，如 ChatGPT、自动驾驶、智慧问诊等。为了迎接 AI 时代，众多云厂商除了提供通算算力服务外，正纷纷入局开拓人工智能市场，建设大规模 GPU 智算计算资源池。

2．网络方案

本项目分为多个不同业务平面的物理组网，包括虚拟私有云（VPC，Virtual Private Cloud）网络、参数面网络、服务器 BMC 网络、交换机管理网络等，整体 AI 智算网络方案主要针对参数面网络进行设计。

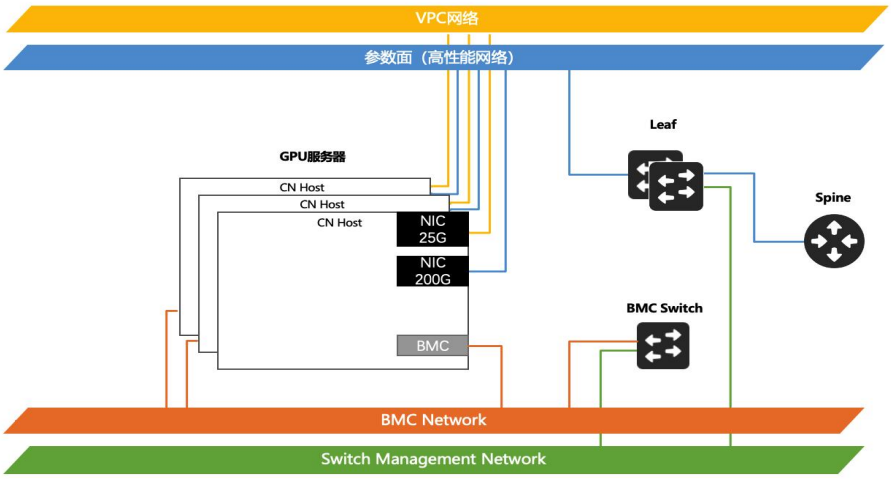

图 5-1 天翼云智算网络方案

GPU 服务器通过不同的网络接口连接至不同的网络平面中，其中参数面网络用于智算集群分布式训练的参数同步，实现将多台训练服务器互联并组建一张高带宽、低延迟、无丢包的高性能网络，满足 GPU 跨机互访要求，进而形成 AI 训练集群。

（1）底层协议栈：支持 RDMA 应用的底层协议通常有 IB 和 RoCEv2。本项目选用 RoCEv2，实现 RDMA 在以太网网络中的传输，仅使用 IB 的“轻量级”传输层，从而降低设备成本和对 IB 网络环境的需求。

（2）无损以太网络：RoCEv2 使用 UDP 头部来封装 RDMA 相关协议栈内容，网络上不仅可通过二层的 PFC pause 帧，还可通过三层头部字段中的 ECN 标记位，两者结合保证流量在传统以太网内的低时延、无损转发。

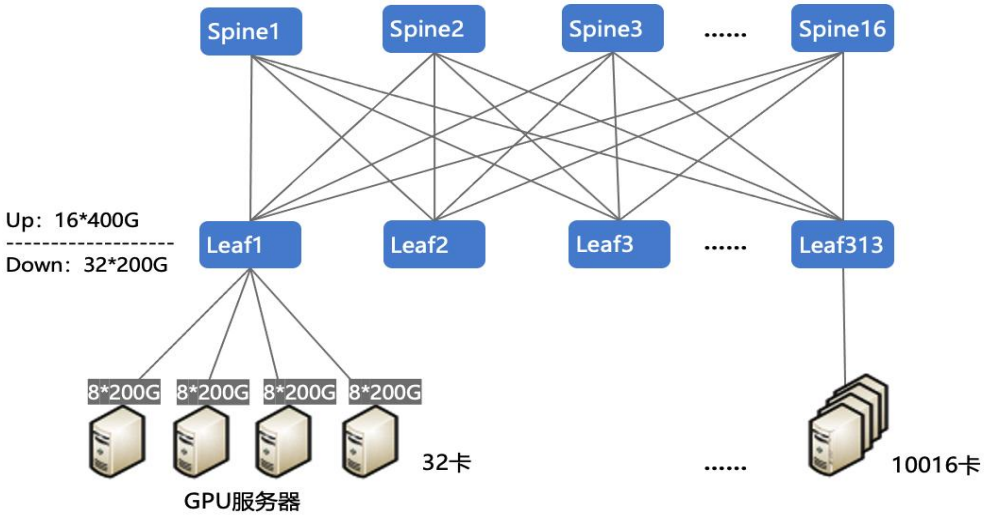

图 5-2 天翼云智算集群内组网结构

（3）硬件选型：

- Leaf：华为 4 槽 CE9860 盒式交换机，搭配 8×400GE

QSFP-DD 接口子卡；

- Spine：华为 8 槽 CE16808\16 槽 CE16816 框式交换机，搭配 36×400G QSFP-DD 接口业务板卡；
- GPU 服务器：芯片选用昇腾 910B。

（4）组网设计：

- 采用二层 Clos 架构，Spine 和 Leaf 之间采用 Full-Mesh 全互联，运行 eBGP 协议组网；
- Leaf 交换机通过 32×200G 端口下行连接服务器，采用 Y 型一分二线缆与服务器 200G 接口对接，共接入 4 台服务器。其中，单台服务器通过 8 个 200G 网口连接至一台 Leaf 交换机，8 个网口分别配置独立的 IP 地址。
- Leaf 交换机通过 16×400G 端口上行连接至 Spine 交换机，Spine 交换机端口扇出决定了 AI 集群规模。例如万卡集群需要至少 313 台 Leaf 接入，则选用 16 台 16 槽的框式交换机，且单台 Spine 设备的 400G 端口数大于 313。

（5）负载均衡与拥塞控制：

- NSLB：AI 训练场景存在大量跨 Leaf 流量，依靠传统 ECMP 无收敛组网已无法规避流量冲突，本项目引入多路径规划解决全局负载均衡问题，对于突发极端拥塞场景，网络侧使能 PFC 兜底。
- HCCL（Huawei Collective Communication Library）：基于昇腾芯片的高性能集合通信库，提供单机多卡、多机多卡

集合通信原语，在 PCIe、HCCS 和 RoCE 高速链路实现集合通信功能，实现分布式训练。

当前，由于美国制裁禁令限制 NVIDIA 芯片对华销售，云公司通过综合评估国内芯片能力，以支持国产化芯片发展为目标，选择打造基于华为昇腾算力的智算资源池，并采用上述网络方案进行了万卡集群建设，在长三角、京津冀规划落地。

3．效益分析

传统通算算力需求已无法满足市场客户发展需求，智算资源池成为 AIGC 重要的算力基础设施。对此，云公司紧跟客户需求，建设高性能、高可靠的智算算力基础设施，并提供租赁服务降低客户对智算算力的使用门槛与成本投入，从而推动 AI 市场蓬勃发展，提升核心产品竞争力。

（二）紫金山新型无损数据中心项目

1．项目背景

人工智能技术被誉为“数字经济发动机”，全球各国纷纷围绕相关技术与产业展开积极布局。2023 年 2 月 24 日，国家科技部官员陈家昌发表讲话，提到科技部已将人工智能视为我国的战略性新兴技术，国家各部门后续将在政策和资金上给予人工智能发展更多支持。2023 年两会报告中，ChatGPT 等前沿科技词汇频繁出现，与会代表和委员们针对 AI 领域提出了多项发展建议和提案。其中，新型数据中心作

为支撑人工智能发展的核心基建，是实现制造强国、网络强国的重要动能。

在此背景下，紫金山实验室基于自研自主可控可编程承载设备和网络控制器，结合 RDMA、智能网卡、PFC/ECN 等技术，建设了集“存、算、训、服、推”为一体的新型无损数据中心，支撑实验室在未来网络、工业互联网、车联网等领域的技术创新。

2．网络方案

如图 5-3 所示，新型无损数据中心采用 Fat-Tree 的组网架构，根据服务场景不同，分为人工智能高性能专区、工业互联网专区和国产化专区。

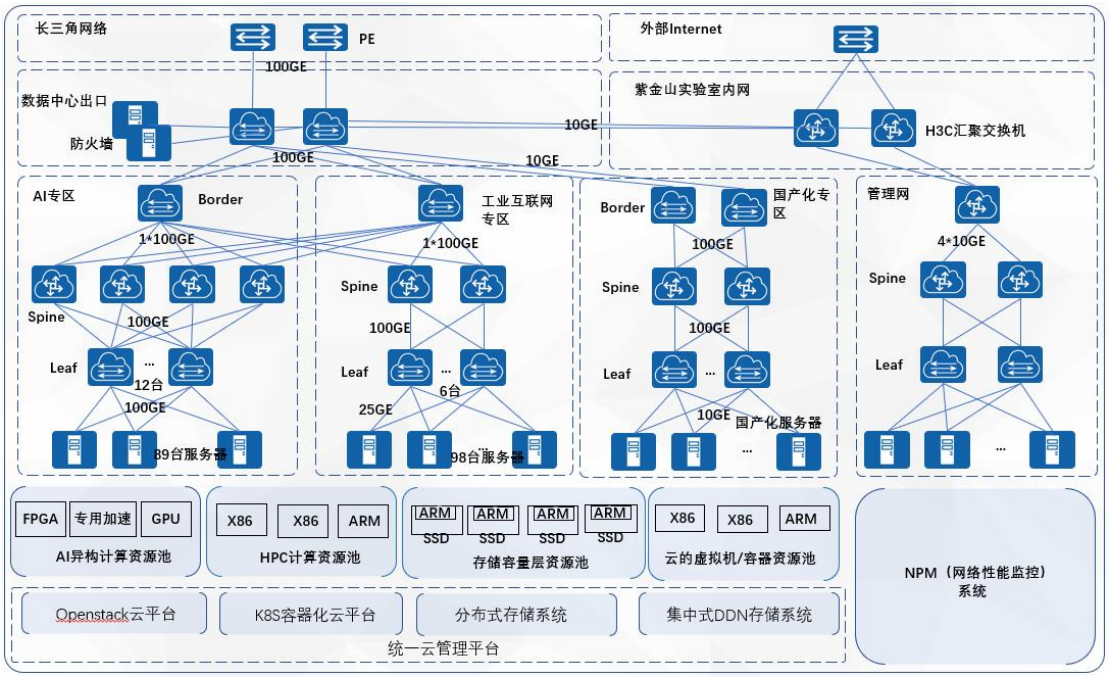

图 5-3 新型无损数据中心组网架构

本项目有如下特点：

（1）基于自研可编程交换机构筑多场景数据中心网络

基于自研的六款白盒化硬件设备，覆盖工业互联网专区可编程数据中心网络、AI 专区无损数据中心网络、国产化专区数据中心网络，为打造全场景端到端数据中心网络奠定基础，其优势如下：

- 建立开放标准的数据平面模型。使用开放标准的 OpenConfig YANG 模型定义网络设备数据模型，实现对网络资源的统一调度和感知；
- 支持大网级操作系统 CNOS 的集中控制。通过 RestFul、NetConf、P4 Runtime 等接口标准，动态部署和增删 UniNOS 中的网络业务功能；
- 异构厂商芯片支持机制。统一转发平台和标准 SAI 接口，屏蔽底层硬件差异，增强数据平面可扩展性，支持多种可编程芯片。

（2）基于自研 INT 技术构筑网络可视化平台

网络测量是网络管控的基础手段和数据来源。按照测量方式的不同，传统意义上的网络测量可分为主动测量、被动测量和混合测量。带内测量是近几年兴起的一种混合测量方法，通过路径中间交换节点对数据包依次插入元数据的方式完成网络状态采集。相较于传统网络测量方案，带内测量能够对网络拓扑、网络性能和网络流量实现更细粒度的测量。

基于工业互联网专区可编程数据中心网络进行实施验证，突破高精度网络测量核心机制，基于带内遥测技术研发可视化平台，实现数据面状态信息的实时采集、分析和展示；支持网络异常（拥塞、时延

过大）、丢包、事件（流产生和终止、时延变化）、队列最大深度、平均深度、设备最大时延、平均时延、端到端时延等可视化功能。

（3）基于自研 OVS 实现智能网卡硬件卸载提升性能

在服务器端，虚拟交换机（vSwitch）在处理网络流量时，会消耗大量的宿主机计算资源。为保证网络数据的转发性能，vSwitch 通常需要绑定多个 CPU 核来处理网络流量，但这样会消耗更多的 CPU 资源，导致服务器的运行成本和能耗增加。为解决该问题，可将某些任务卸载到网卡上处理，从而释放大量 CPU 资源，减少服务器的运行成本和能耗，同时保证网络流量的高效处理。具体而言，是指将原本在内核网络协议栈中进行的 IP 分片、TCP 分段、重组、checksum 校验等操作，转移到网卡硬件中处理，使得 CPU 的发包路径更短、消耗更低，提高处理性能。此外，智能网卡实现的网络加速有多种，除基本网络功能外，还包括 RoCEv2、VXLAN、OVSct 功能、TF-vRouter 虚拟路由、kTLS/IPSec 加速等技术。

3．效益分析

新型无损数据中心已稳定运营四年，支撑紫金山实验室科研任务的同时，积极向社会科研创新需求开放。目前已成功服务中国信通院、国家天文台/紫金山天文台/上海天文台、中国电信炫彩公司、南京大学、北京邮电大学、南京邮电大学、国家（深圳前海）新型互联网交换中心、中科院南京信息高铁、华为、浪潮等行业龙头，并联合华为、信通院斩获 CCF HPC CHINA2022 超融合无损以太技术创新奖。

六、智算网络技术与产业发展建议

智算网络作为新一代信息技术的核心支撑，正在加速对各行各业的数字化变革进程。为充分发挥其在经济和社会发展中的重要作用，技术与产业的协调发展至关重要。本章节将从多个维度探讨智算网络的技术进步与产业推进路径，旨在构建一个高效、智能、开放的智算网络生态系统，助力实现未来社会的全面数字化转型。

（1）深化硬件创新与技术优化

硬件创新与技术优化是智算网络深入发展的基础，应进一步：*i*）优化异构计算架构，聚焦于 CPU、GPU 和专用加速器（如 TPU、FPGA）的协同工作，提升整体计算效率，并优化硬件接口和数据传输路径，提高算力资源利用率；*ii*）研发新一代高速互连技术，提升数据中心内外的通信效率，采用先进的可编程交换机与路由器技术，支持更高的数据传输速率和更低的网络延迟；*iii*）发展超低损耗光纤技术，推广使用 G.654.E 等超低损耗光纤，减少中继站数量，降低整体部署成本，并延长无中继传输距离；*iv*）探索高效存储解决方案，提高存储访问速度和数据处理能力，优化存储层次结构，减少数据在不同存储介质之间的传输延迟；*v*）推进硬件加速器的集成与优化，如 AI 推理加速器、加密加速器等，提升特定任务的处理性能，同时优化加速器的驱动和软件栈，确保其高效运行；*vi*）注重能效优化，采用更高效的电源管理技术和冷却方案，降低数据中心的能耗，提高硬件设备的长期运行稳定性。

（2）推进软件与算法的智能化集成应用

软件和算法的智能化是提升智算网络性能的核心，应进一步：*i*）开发与推广智能编译器技术，实现代码的自动优化，使其更好地适应不同硬件平台。智能编译器能够根据特定的硬件架构调整代码，提高执行效率与性能；*ii*）深入研究并行计算算法，挖掘多核处理和异构计算资源的潜力，优化并行计算算法，实现计算任务的高效并行处理；*iii*）推进 AI 算法在资源分配和任务调度中的应用，以进一步提升系统性能，并支持适应不同的负载与任务需求，提升系统的整体效率；*iv*）研究先进的数据管理技术，包括数据去重、压缩、加密等，以提升存储效率和保障数据安全，同时减少存储空间的占用、加快数据访问速度。

（3）催化标准化与开放性的行业实践

标准化和开放性是智算网络技术发展的重要推动力，应进一步：*i*）推动行业标准的制定与推广，鼓励相关企业、科研机构和行业组织共同参与制定智算网络的技术标准和规范，包括计算架构、通信协议、接口标准等，以确保不同设备和系统之间的兼容性和互操作性；*ii*）倡导开放硬件与软件平台，促进建立开源社区，鼓励企业和开发者共享资源，加速技术创新，降低开发成本；*iii*）推动互操作性测试与认证机制的建立，确保不同厂商的设备与系统能够无缝协同工作；*iv*）呼吁政府和监管机构制定智算网络的相关政策与法规，通过政策引导和激励措施，推动行业朝着标准化和开放性的方向发展。

（4）加大政府资金支持，促进产业生态合作

政府资金的投入对于推动智算网络产业发展至关重要。通过增加财政预算、设立专项基金以及实施税收优惠政策等方式，降低企业在智算网络投资和运营的成本，并激励更多企业与研究机构投身于智算网络的技术创新与应用开发。同时，通过搭建跨行业合作平台，促进不同领域企业间的资源共享和协同创新，推动产学研用深度融合，加强智算网络技术的研发和应用示范，以形成强大的产业生态，共同推动智算网络产业的繁荣发展。

七、总结与展望

中国智算网络产业正站在新的历史起点上，面临着前所未有的发展机遇。在国家政策的引领与推动下，中国人工智能核心产业规模持续扩大，人工智能向大规模落地应用发展，业务发展呈现多样化趋势，智能算力需求急剧增加。从宏观技术发展的趋势来看，新型大容量网络芯片将成为智算网络发展的基石。同时，新型计算架构、高速传输技术、高效存储也将发挥重要作用。在核心技术层面上，大模型的崛起正引领人工智能产业的发展趋势，而以太网将成为构建超大规模智算集群的技术基础，本文着重对集群内与集群间的关键技术进行了详细讨论，这些技术对提升算力利用效率及优化数据传输具有重要意义。

未来，智算网络技术将继续沿着高性能、高效率和智能化的方向演进，新型网络架构、量子通信、边缘计算等前沿技术有望进一步提升智算网络的综合性能。此外，随着与人工智能、大数据、物联网等产业领域的紧密结合，将形成更加丰富的智算服务生态系统，这种跨

界融合将促进不同技术领域之间的协作与创新，不断扩展智算应用场景，持续为各行业提供更高效、智能化的解决方案，为中国数字经济发展和智能化社会建设注入新动力。

附录 A：术语与缩略语

中文名称	英文缩写	英文全拼
基于瓶颈带宽和往返时间的拥塞控制	BBR	Bottleneck Bandwidth and Round-trip Time
背压流量控制	BFC	Backpressure Flow Control
拥塞感知的流量控制	CaPFC	Congestion aware Priority Flow Control
光电合封装	CPO	Co-packaged optics
数据中心量化拥塞通知	DCQCN	Data Center Quantized Congestion Notification
分布式解耦机框	DDC	Disaggregated Distributed Chassis
动态负载均衡	DLB	Dynamic Load Balancing
差分服务码点	DSCP	Differentiated Services Code Point
等价多路径	ECMP	Equal-Cost Multiple-Path
显式拥塞通知	ECN	Explicit Congestion Notification
通用传输控制协议	GTCP	General Transmission Control Protocol
全调度以太网技术体系	GSE	Global Scheduling Ethenet
华为集合通信库	HCCL	Huawei Collective Communication Library
空芯光纤	HCF	Hollow Core Fiber
高性能计算	HPC	High Performance Computing
高精度拥塞控制	HPCC	High Precision Congestion Control
高性能网络	HPN	High-Performance Networking
基于延迟的拥塞控制	IDCC	INT and Delay based Congestion Control
任务完成时间	JCT	Job Completion Time
下一代网络演进	NGNe	Next Generation Network Evolution
网卡	NIC	Network Interface Card
网络级负载均衡	NSLB	Network Scale Load Balance
光电路交换	OCS	Optical Circuit Switching
封装内光学 I/O	OIO	In-Package Optical I/O

光交叉连接	OXC	Optical Cross Connect
基于优先级的流量控制	PFC	Priority-based Flow Control
比例积分微分	PID	Proportional Integral Derivative
协议无关交换机架构	PISA	Protocol Independent Switch Architecture
预测型 PFC 流控	P-PFC	Predictive PFC
服务质量	QoS	Quality of Service
队列偶	QP	Queue Pair
远程直接内存访问	RDMA	Remote Direct Memory Access
可重构光分叉复用器	ROADM	Reconfigurable Optical Add/Drop Multiplexer
往返时延	RTT	Round-trip Time
共享风险链路组	SRLG	Shared Risk Link Group
服务类型	ToS	Type of Service
超以太网联盟	UEC	Ultra Ethernet Consortium
虚拟网络功能	VNF	Virtual Network Function

参考文献

[1] “十四五”国家信息化规划. 2021-12. URL: https://www.cac.gov.cn/2021-12/27/c_1642205314518676.htm.

[2] 新华三，中国信通院. 2023 智算算力发展白皮书. 2023-08.

[3] 中国移动通信研究院. 新一代智算中心网络白皮书. 2022.

[4] Gavin. What is RDMA?RoCE vs. InfiniBand vs. iWARP Difference. 2023-12. URL: https://www.naddod.com/blog/what-is-rdma-roce-vs-infiniband-vs-iwar-difference.

[5] IEEE 802.1Q. Data Center Bridging WG [Online]. URL: https://www.ieee802.org/1/pages/dcbridges.html.

[6] Zhang Z, Zhang J, Ma H, et al., "ADMIRE+: curiosity-exploration-driven reinforcement learning with dynamic graph attention networks for IP/optical cross-layer routing," 49th European Conference on Optical Communications (ECOC), 2023.

[7] S. Ha, I. Rhee, and L. Xu, “Cubic: a new TCP-friendly high-speed TCP variant,” in Proc. ACM, 2008.

[8] N. Cardwell, Y. Cheng, C. S. Gunn, et al., “BBR: Congestion-based congestion control,” ACM Queue, vol. 14, no.5, pp. 50-83, 2016.

[9] M. Dong, T. Meng, D. Zarchy, et al., “PCC Vivace: Online-learning congestion control,” in Proc.USENIX NSDI, 2018.

[10] V. Arun, and H. Balakrishnan, “Copa: Practical delay-based congestion control for the Internet,” in Proc. USENIX NSDI, 2018.

[11] Saeed, Ahmed, Varun Gupta, Prateesh Goyal, et al., “Annulus: A Dual Congestion Control Loop for Datacenter and WAN Traffic Aggregates,” in SIGCOMM, 2020.

[12] Zeng, Gaoxiong, Wei Bai, Ge Chen, et al., “Congestion Control for Cross-Datacenter Networks,” IEEE/ACM Transactions on Networking, 2022.

[13] Zou, Shaojun, Jiawei Huang, Jingling Liu, et al., “GTCP: Hybrid Congestion Control for Cross-Datacenter Networks,” in ICDCS, 2021.

[14] Goyal, P., Shah, P., Sharma, N.K., et al., “Backpressure Flow Control,” in NSDI, 2022.

[15] Liu, K., Tian, C., Wang, Q., et al., “December. Floodgate: Taming incast in datacenter

networks,” in CoNEXT, 2021.

[16] Avci S N, Li Z, Liu F, “Congestion aware priority fow control in data center networks,” in IFIP Networking Workshops, 2016.

[17] Tian C, Li B, Qin L, et al. “P-PFC: Reducing tail latency with predictive PFC in lossless data center networks,” IEEE Transactions on Parallel and Distributed Systems, 2020.

[18] Qian K, Cheng W, Zhang T, et al, “Gentle Flow Control: Avoiding Deadlock in Lossless Networks,” in SIGCOMM, 2019.

[19] 中兴通讯. 未来通信的光速之路. 2024-03.

[20] Estrella. G.654.E 新 型 光 纤 . 2020-09. URL: https://community.fs.com/cn/article/g654e-fiber-for-400g-and-beyond.html.