专题：智算网络

面向异构算力互联的智算网络关键技术研究

苏昱臻[1]，王子潇[1]，钟驰量[2]，寇晓淮[1]，刘圆[1]，陈映[1]
（1. 中国电信股份有限公司研究院，北京 102209；
2. 网络与交换技术全国重点实验室（北京邮电大学），北京 100876）

摘　要：算力供给的代际异构性与供应链安全需求，促使异构算力成为AI基础设施的新趋势。然而，在异构混合训练场景中，基于融合以太网的RDMA版本2（RDMA over converged Ethernet version 2，RoCEv2）方案存在负载均衡与拥塞控制缺陷，在模型训练的并行通信中性能欠佳；而现有高性能同构智算网络方案因设备异构与集合通信库（collective communication library，CCL）闭源难以部署。为此，提出了面向异构算力场景的高性能智算网络解决方案——智能控制以太网（intelligent control Ethernet，ICE）。该方案基于RoCEv2协议体系，在避免对设备、CCL进行深度定制的前提下，将异构通信库信息采集、集中控制器与端侧自主控制相结合，实现全局最优路径规划及全局主动拥塞控制，显著提升异构并行通信性能。真实物理环境实验表明，ICE可提升集合通信性能最高达47%。ICE为异构智算网络建设提供了开创性、易部署的解决方案。

关键词：异构算力；智算网络；RoCEv2；通信调度；拥塞控制；负载均衡
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025183

Research on key technologies of intelligent computing network for heterogeneous computing power interconnection

SU Yuzhen[1], WANG Zixiao[1], ZHONG Chiliang[2], KOU Xiaohuai[1], LIU Yuan[1], CHEN Ying[1]
1. China Telecom Research Institute, Beijing 102209, China
2. State Key Laboratory of Networking and Switching Technology (BUPT), Beijing 100876, China

Abstract: The intergenerational heterogeneity of computing supply and the demand for supply chain security have made the driving forces behind heterogeneous computing becoming an emerging trend in AI infrastructure. However, in heterogeneous hybrid training scenarios, the RoCEv2 (RDMA over converged Ethernet version 2) solution suffered from deficiencies in load balancing and congestion control, resulting in suboptimal parallel communication performance during model training. Meanwhile, existing high-performance homogeneous intelligent computing network solutions were faced with deployment barriers due to the heterogeneity of devices and closed-source CCL (collective

收稿日期：2025−06−09；修回日期：2025−07−19
通信作者：王子潇，wangzx15@chinatelecom.cn
基金项目：中央企业算力网络创新联合体–新型智算高速互联技术研究任务项目（No.2024SWLHT-024）
Foundation Item: Innovation Consortium for Computing Power Network of Central State-owned Enterprises - Research Task on New High-speed Interconnection Technology for Intelligent Computing (No.2024SWLHT-024)

communication library). To address these challenges, ICE (intelligent control Ethernet), a high-performance intelligent computing network solution for heterogeneous computing scenarios, was proposed. Based on the RoCEv2 protocol framework, ICE was designed to avoid deep customization of devices and CCL. Through a combination of heterogeneous communication library information collection, a centralized controller, and autonomous control at the end side, global optimal path planning and global active congestion control were achieved, significantly enhancing heterogeneous parallel communication performance. Experiments conducted in real-world physical environments demonstrate that ICE improves performance by up to 47%. Thus, ICE presents as a pioneering and easily deployable solution for constructing heterogeneous intelligent computing networks.

Key words: heterogeneous computing power, intelligent computing network, RoCEv2, communication scheduling, congestion control, load balancing

0 引言

历经十余年演进，AI模型的参数量级经历了从深度学习（deep learning，DL）时代千万级别到大语言模型（large language model，LLM）万亿量级的跨越式发展[1]。这一演进直接导致算力需求模式发生根本性转变：传统模型的线性需求曲线已演变为指数级攀升态势，其增速远超摩尔定律驱动的单芯片算力提升幅度。当前，大模型预训练的单集群图形处理单元（graphics processing unit，GPU）部署规模已突破十万量级[2]。在此超大规模环境下，分布式混合并行策略已成为提升模型训练效率的核心技术方案[3]。

远程直接内存访问（remote direct memory access，RDMA）网络凭借其低延迟、高带宽以及零拷贝等显著优势脱颖而出，成为分布式大模型训练机间通信的首选技术。当前，RDMA技术方案的代表有IB[4]（InfiniBand）和RoCEv2[5]。IB协议因为需要专用硬件、部署成本高、技术复杂、封闭和扩展受限等局限，未能成为大规模商用的主流方案。RoCEv2协议可利用现有以太网设备，兼顾部署成本和网络兼容性，是目前主流的商用解决方案。

然而，RoCEv2在智算场景下面临流量特征适配性问题，具体表现为低熵流量的哈希极化效应与高速突发流量的拥塞控制失效。当前已经规模化建设的智算集群以同构集群为主，针对上述问题，业界提出的一系列智算网络解决方案有效弥补了RoCEv2性能的不足，且对模型训练效率有显著提升。

近年来，异构算力互联成为AI基础设施建设的新命题，其发展受到双重驱动影响：一方面，主流GPU厂商的产品更新周期已缩短至12～18个月[6]，促使算力供给呈现代际异构特征；另一方面，国际地缘政治格局演变[7]驱动国家战略层面提出自主可控智算供应链体系的建设要求，使得算力资源供应链呈现多元化。异构算力场景下，机间网络通信呈现出GPU设备和集合通信库（collective communication library，CCL）异构的特性，致使现有智算网络解决方案失效。

对此，本文在满足异构算力互联场景约束条件下，创新地提出了一种高效的异构算力互联智算网络方案——智能控制以太网（intelligent control Ethernet，ICE）。该方案依托现有RoCEv2协议体系，不需要进行设备定制化改造，且在非深度定制CCL的前提下，便可实现异构算力混合训练高性能网络传输。ICE通过集中控制器实现业务流量的全局最优路径规划和全局拥塞控制。经物理环境验证表明，ICE将集合通信的性能最高提升到了47%。

本文的主要贡献如下。

（1）系统归纳现有智算网络解决方案的技术

特征与异构算力互联对智算网络的新挑战，论述了现有同构智算网络解决方案无法直接实现异构算力的高效互联。

（2）提出异构智算网络解决方案 ICE。该方案基于 RoCEv2 协议构建，通过异构通信库信息采集、集中式智能选路和全局主动式拥塞控制架构提升传输效率，并保障异构设备兼容性。

（3）通过真实物理环境对 ICE 方案进行了实测验证。结果表明，相较 RoCEv2，ICE 可将同构 GPU 的集合通信性能提升 47%，异构 GPU 的集合通信性能提升 38%。

1　研究动机与现状

1.1　智算流量对 RoCEv2 网络性能的挑战

当前，大模型并行训练机间通信普遍使用 RoCEv2 网络。商用 RoCEv2 方案默认部署 DCQCN[8] 拥塞控制算法与等价多路径路由（equal-cost multi-path routing，ECMP）负载均衡机制，二者在具有高熵、持续性流量特征的通用计算场景中表现良好，但在智算场景的低熵、突发传输流量特征下存在性能瓶颈。

大语言模型训练中普遍采用 3D/4D 并行策略加速模型训练[3]，4D 并行示意图如图 1 所示。其中，数据并行和专家并行这两类典型策略的通信特征是带来挑战的关键因素。

（1）数据并行（data parallelism，DP）将完整的模型复制到多个计算设备并行执行计算，以提升训练效率，其采用周期性的跨机全规约（AllReduce）通信同步梯度。

（2）专家并行（expert parallelism，EP）将多个专家网络分布在不同的计算设备上，以提高计算效率和缓解内存压力，其通过跨机全对全（All-to-all）通信实现样本−专家匹配。

具体来说，DP 中的 AllReduce 操作表现出周期性并发、节点队列对（queue pair，QP）数量少、跨脊（Spine）流量等低熵特征。此类特性使得叶（Leaf）交换机在使用 ECMP 进行上行转发端口选择时，极易引发哈希冲突问题，导致流量无法在各上行链路间均匀分布。在无收敛组网环境下，这将导致部分链路空闲而部分链路拥塞，有效吞吐降至线速的 66%，严重影响网络传输效率。再者，EP 中的 All-to-all 流量呈现出并发的多对多模式，对拥塞控制的性能提出了挑战。DCQCN 采用被动启发式的控速算法，在面对 400 Gbit/s 链路场景下 All-to-all 流量时，存在收敛速度慢、控速效果差等问题[9]。更甚者，跨 Leaf

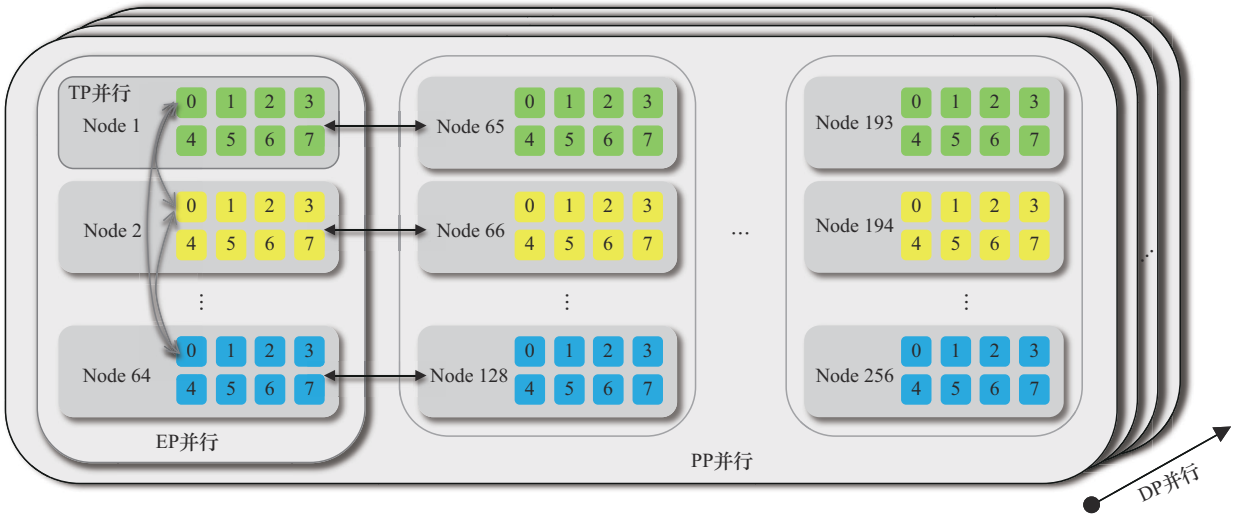


图 1　4D 并行示意图

交换机的All-to-all流量还会遭遇ECMP哈希冲突问题，有效吞吐降至线速的70%。现有研究表明，使用RoCEv2网络，部分大模型训练中通信占比最大可达50%，导致GPU资源利用率和模型训练效率显著受限[10]。

1.2 业界智算网络解决方案

业界现有智算网络解决方案主要针对大规模同构GPU算力，用于解决RoCEv2在并行训练中的性能不足问题，且在实际应用中实现了显著的性能收益[10]。基于技术演进路径差异，现有解决方案可以划分为两大技术体系：一是与自家硬件能力强绑定的设备厂商方案；二是以设备定制化为显著特征的互联网企业方案。本节着重分析两类方案的技术细节及局限性。

1.2.1 设备厂商解决方案

针对大模型训练的智算网络需求，设备厂商提出了差异化方案，其代表如下。

华为推出的星河AI解决方案[11]，采用集中式控制器动态修改路由表，实现流量的逐跳路径规划。该方案基于RoCEv2协议，旨在实现“零冲突、零拥塞”目标。然而，该方案存在双重局限：（1）完全禁用拥塞控制机制，在应对混合专家[12]（mixture of expert，MoE）模型All-to-all流量引发的多对一（Incast）问题时存在局限性；（2）其集中式路由控制算法与动态路由表修改功能深度集成于华为训练解决方案，缺乏跨平台通用部署能力。

英伟达的Spectrum-X方案[13]沿用RoCEv2协议，融合逐包自适应路由（adaptive routing，AR）技术解决ECMP哈希冲突问题，并保留DCQCN算法实现拥塞控制。然而，该方案存在3点不足：（1）逐包AR机制引入额外的分组头开销及网卡侧乱序重排性能损耗，在All-to-all拥塞场景下，端侧需要并行执行多QP乱序重排，性能表现反而劣于RoCEv2方案；（2）AR功能强制依赖英伟达自有网卡与交换机的协商机制，存在厂商锁定问题；（3）DCQCN固有的被动响应特性在极端拥塞场景下仍存在收敛迟滞问题。

博通的分布式解耦机框[14]（distributed and disaggregated chassis，DDC）同样基于RoCEv2协议，采用等长信元粒度喷洒式AR与DCQCN拥塞控制，理论上可实现趋近最优的负载均衡并消除无收敛网络中的链路拥塞。DDC将数据报文拆分与重组下沉至网络侧，不仅实现了异构网卡兼容，还彻底消除了对网卡乱序重排能力的依赖。然而，该方案存在以下局限性：（1）信元级负载均衡需设备维护有状态表项，受限于硬件表项容量，实际可扩展集群规模存在上限；（2）DDC信元喷洒功能依赖报文切片传输机制，存在潜在数据安全风险；（3）DDC尚未实现大规模商业部署，其真实性能与稳定性还需进一步验证。

1.2.2 互联网企业解决方案

国内外领先的互联网企业纷纷布局大模型训练网络优化，通过定制化的集合通信库和网络协议，实现网络流量路径规划和高效拥塞控制，以弥补RoCEv2的缺陷。

阿里云自主研发的ACCL[15]（Alibaba Collective Communication Library）集合通信库通过路径探测机制构建流量路径映射，进而实现集合通信流量的亲和性调度与动态路径优化。其创新的solar-RDMA协议[16]集成自研HPCC[17]（high precision congestion control）拥塞控制算法，基于细粒度网络内遥测[18]（in-network telemetry，INT）技术获取实时链路状态，并结合窗口控速机制，相较于传统DCQCN性能更优。

腾讯云依托星脉2.0网络架构，构建高性能智算网络方案，网络通信效率提升60%[10]。在流量调度方面，其TCCL[19]（Tencent Collective Communication Library）基于交换机哈希模拟器实现流量路径感知，进而实现流量全局负载均衡；拥塞控制方面，其自研TiTa2.0协议采用主动式拥塞控制算法，有效提升All-to-all传输能力。

Meta[9]侧重软件层网络传输优化创新，通过流量工程和增强型 ECMP（enhanced ECMP，E-ECMP）分别优化小规模与大规模训练任务的负载均衡性能，并在集合通信库层实现了接收端驱动的拥塞控制机制，以代替传统 DCQCN 拥塞机制。

需要注意的是，上述方案存在以下特性。（1）基础设施深度绑定。上述方案深度耦合企业私有技术栈，关键能力（如哈希模拟器、INT）依赖定制化交换机实现，自研智算网络协议需要配合专用定制网卡方能启用完整功能。（2）CCL 深度二次开发。上述方案要求对集合通信库进行深度开发修改以适配特定网络功能。

1.3　异构算力互联对智算网络解决方案的新挑战

1.3.1　异构算力混合训练方案

异构算力场景下，不同 GPU 的算力差距较大，各厂商 GPU 性能对比见表 1。关键性能指标方面，英伟达（NVIDIA）H100 的 32 位浮点数（float point 32，FP32）算力最高可达国产 GPU 的 2.5 倍；16 位浮点数（float point 16，FP16）算力最高可达到国产 GPU 的 19.8 倍。在此条件下，采用传统均质拆分策略存在算力资源利用率低、异构集合通信效率差等问题[20]。

表 1　各厂商 GPU 性能对比

GPU	FP32（归一化）	FP16（归一化）
NV A100	1.0	32.0
NV H100	3.1	102.6
国产 A	1.5	12.3
国产 B	1.2	5.2
国产 C	2.5	9.8
国产 D	1.6	26.3
国产 E	3.3	13.1

目前，业界采用非均质拆分的策略来实现异构算力混合训练。异构流水线并行（pipeline parallelism，PP）非均质拆分训练方案架构如图 2 所

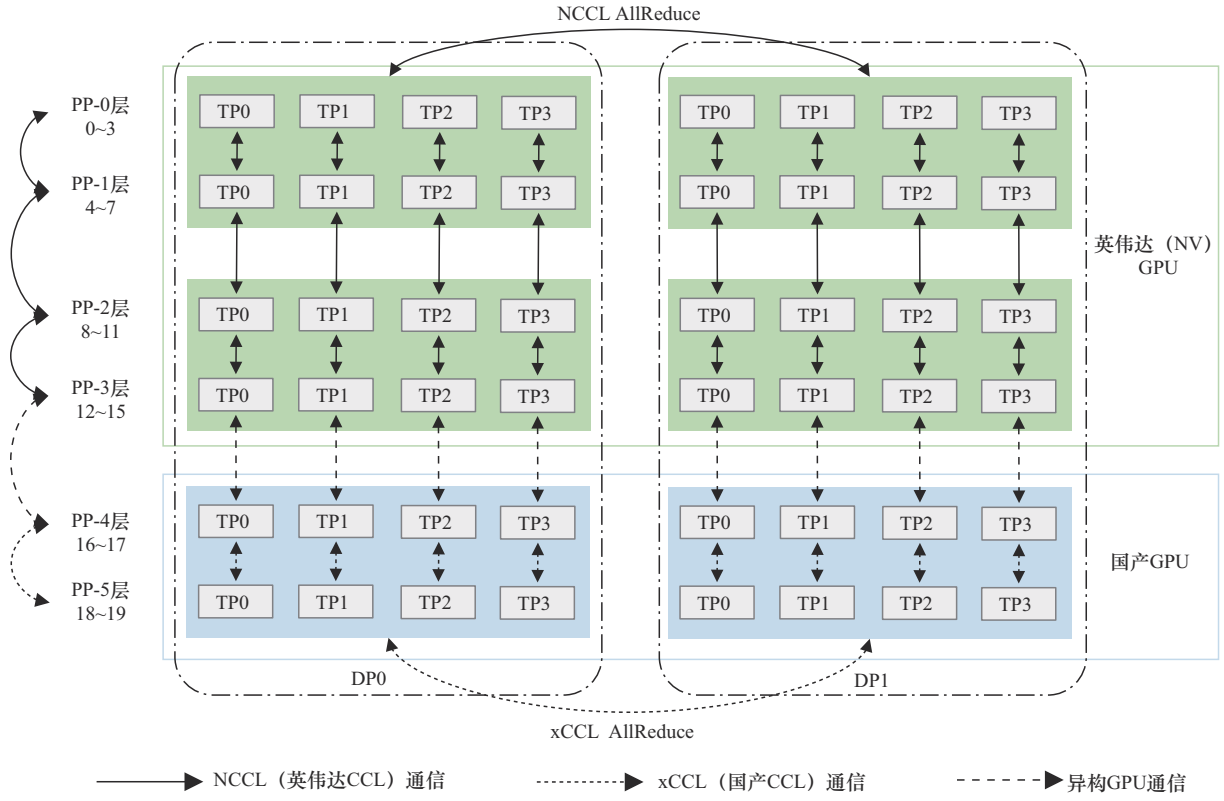


图 2　异构 PP 非均质拆分训练方案架构

示，该架构根据异构GPU的显存和算力灵活分配模型层数，实现最佳的资源利用率。此外，PP非均质拆分只涉及PP并行异构GPU通信，可最小化异构GPU通信复杂度，提升传输效率。相较于同构均值拆分策略，异构PP非均质拆分策略可达到其性能的97%[20]。

1.3.2 现有方案的异构兼容性问题

异构算力混合训练方案中，涉及GPU异构、集合通信库异构，这给智算网络解决方案带来多维技术挑战。

（1）异构GPU对定制化网络设备的兼容性难以保证。当前国产异构算力主要针对英伟达标准商用网卡进行适配，若使用定制化网卡，一些会对性能带来重要影响的功能，如GPU直连RDMA[21]（GPU direct RDMA，GDR）可能无法直接使用，需要GPU与网卡厂商进行联合适配。因此，当前异构智算集群建设仍以标准商用网卡为主，依赖于定制网络设备的现有智算网络解决方案无法使用。

（2）异构集合通信库生态与厂商深度锁定。绝大多数国产GPU的CCL闭源，缺乏底层接口可编程性，导致自研RDMA协议无法穿透硬件抽象层，依赖深度修改CCL代码的流量调度方案（如阿里云、腾讯星脉等）也无法使用。

目前具有代表性的智算网络方案对比见表2。

2 异构智算网络解决方案

2.1 ICE总体架构

针对前文提到的异构算力互联对智算网络解决方案的挑战，提出基于算网联合优化的异构智算网络解决方案ICE。具体而言，ICE具备以下关键能力。

（1）异构兼容流量调度能力：基于RoCEv2网络功能，杜绝依赖定制化网络功能和深度修改闭源CCL代码，实现高效统一的异构集合通信流量路径规划，系统性解决哈希冲突问题。

（2）异构兼容拥塞控制能力：基于RoCEv2协议与标准商用网卡的通用可编程框架，构建高效主动拥塞控制。

ICE整体架构如图3所示，主要包括作为决策中心的ICE集中控制引擎，以及部署在各算力节点上的异构集合通信库TeleCCL和ICE单元。

- 路径建模单元。集成在ICE集中控制引擎内，具备参数面网络拓扑探测与生成功能，接收各算力节点上报的网络路径和通信域信息，为网络控制单元提供实时网络与通信域状态信息，支撑其决策过程。
- 网络控制单元。集成在ICE集中控制引擎内，该单元基于全局网络拓扑、网络路径和通信域信息，通过融合动态负载均衡算法和全局拥塞控制机制，构建面向服务质

表2 智算网络解决方案对比

方案	核心思路	负载均衡	拥塞控制	异构设备兼容
华为星河AI	集中式路径规划	NSLB集中算路控制器分配路径（自家设备锁定）	禁用拥塞控制	×
英伟达Spectrum-X	自适应路由优化	逐包AR（自家设备锁定）	DCQCN	×
博通DDC	交换芯片级流量调度	交换机等长信元自适应路由（自家设备锁定）	DCQCN	×
阿里云	全栈自研软硬协同	ACCL层路径规划（定制网卡、交换机实现）	HPCC（定制网卡实现）	×
腾讯星脉网络	全栈自研软硬协同	TCCL层路径规划（定制网卡、交换机实现）	TiTa2.0（定制网卡实现）	×
Meta	着重软件层优化	增强ECMP+小规模集中算路（深度定制CCL实现）	接收端驱动流量控制（深度定制CCL实现）	×
异构智算网络方案诉求	异构设备兼容、高性能	基于标准网络功能和非深度定制CCL	可基于标准网卡实现	√

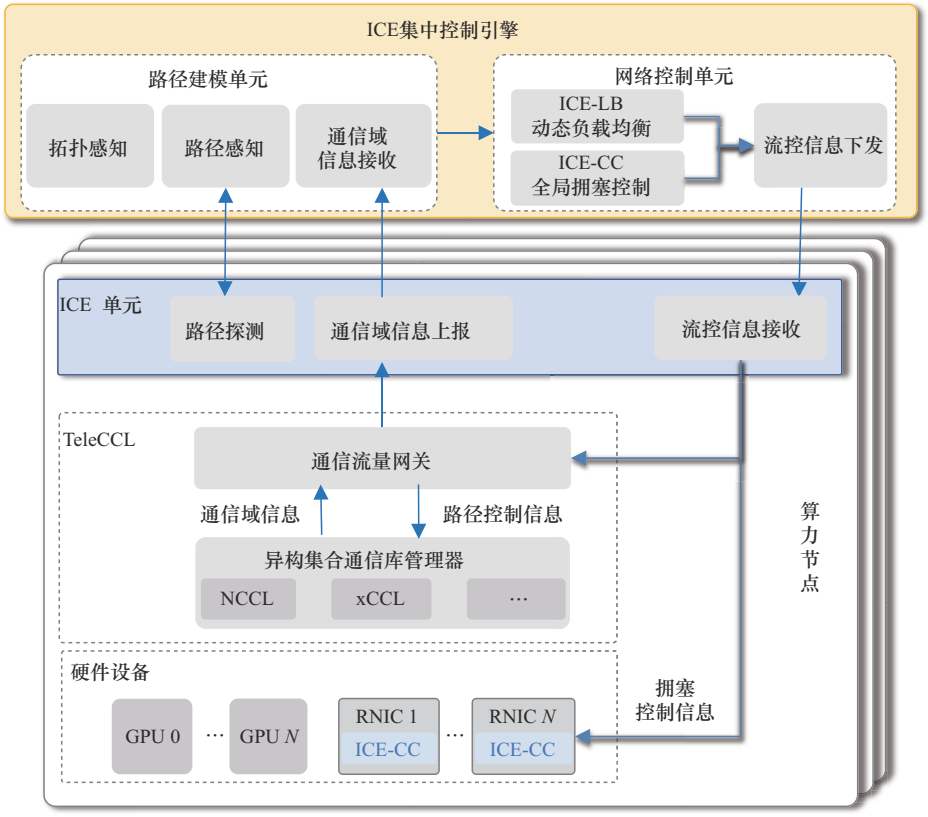


图3　ICE整体架构

量保障的流量调度模型，确保业务无冲突转发与高吞吐传输。

- TeleCCL。是自研异构集合通信库，它通过集成多厂商CCL，实现异构算力混合训练场景下的统一通信调度。基于非侵入式设计理念，TeleCCL联合主流GPU厂商完成CCL接口适配，在不修改各厂商CCL核心代码的前提下，实现了通信域信息实时上报与流量传输路径动态配置功能，有效保障了异构集群间的高效协同通信。
- ICE单元。部署在各算力节点上，负责端侧网络路径探测和通信域信息上报，同时接收来自ICE引擎的流控信息。对于流控信息，该单元将动态负载均衡信息下发至TeleCCL，实现流量转发零冲突；将拥塞控制信息通过主机驱动层标准接口下发到各RDMA网卡（RDMA NIC，RNIC）生效，构建全局主动拥塞控制体系，有效保障异构混合训练场景下的网络传输效率与稳定性。

因不涉及设备的定制和改造，ICE方案可以直接在异构算力集群中进行增量部署。

2.2　ICE关键技术

2.2.1　网络拓扑感知

ICE动态负载均衡算法依赖全局网络拓扑为流量规划路径，因此ICE引擎需实时感知集群全局网络拓扑。ICE实现了融合链路层发现协议（link layer discovery protocol，LLDP）和简单网络管理协议（simple network management protocol，SNMP）的拓扑发现方案。该方案利用网络设备原生支持的标准化协议实现拓扑信息的分布式采集与聚合，解决异构算力环境下基础设施的兼容性问题，最终形成支持流量智能调度的全局拓扑视图。

LLDP基于数据链路层报文交互自动发现邻居节点，SNMP通过标准化方式获取邻居节点网络设备详细信息。如算法1所示，ICE通过广度优先遍历算法将LLDP发现的邻接关系与SNMP采集的设备属性进行图结构融合，最终构建出包含物理连接特征的集群全局拓扑图谱。

算法1 ICE拓扑感知算法

输出 集群全局网络拓扑

```
searched_nodes=[]
wait_search_nodes=[]
local_node=snmp_get_node_info(local)
wait_search_nodes.append(local_node)
while wait_search_nodes:
    tmp_node=wait_search_nodes.pop()
    rem_servers = lldp_get_rem_servers(tmp_ node)
    for rem_server in rem_servers:
        tmp_rem_node = semp_get_node_info(rem_server)
        struct_link(tmp_node, tmp_rem_node)
        wait_search_nodes.append(tmp_rem_node)
    searched_nodes.append(tmp_node)
```

2.2.2 ICE-LB动态负载均衡

大模型训练过程中，各节点QP数量少且为长连接，兼容异构算力的RoCEv2网络设备通常使用ECMP机制进行流量转发，该机制基于报文五元组（源IP地址、目的IP地址、源端口、目的端口、协议类型）的哈希运算确定转发路径。通常情况下，用户不会对五元组中的字段进行自定义，导致哈希运算的结果不可控，在智算流量低熵特性下极易造成哈希极化问题，使经Leaf交换机转发的多条流量抢占同一上行链路，或Spine交换机转发的多条流量抢占同一下行链路，进而损害整个集群的通信性能。

DP AllReduce哈希冲突如图4所示。在4节点8 GPU组成的DP流量交互模式中，每个节点的同号GPU构成一个DP组，组内的4卡会在每个批次的数据训练结束后通过AllReduce操作将各自的梯度进行汇总更新。此时，GPU 2到GPU 4、GPU 3到GPU 5的流量会通过Spine交换机，可能在Leaf交换机（交换机0）的上行链路发生哈希冲突。同理，GPU 7到GPU 1、GPU 6到GPU 0的流量可能在交换机1的上行链路发生哈希冲突。

为解决哈希极化问题，对控制ECMP哈希结果

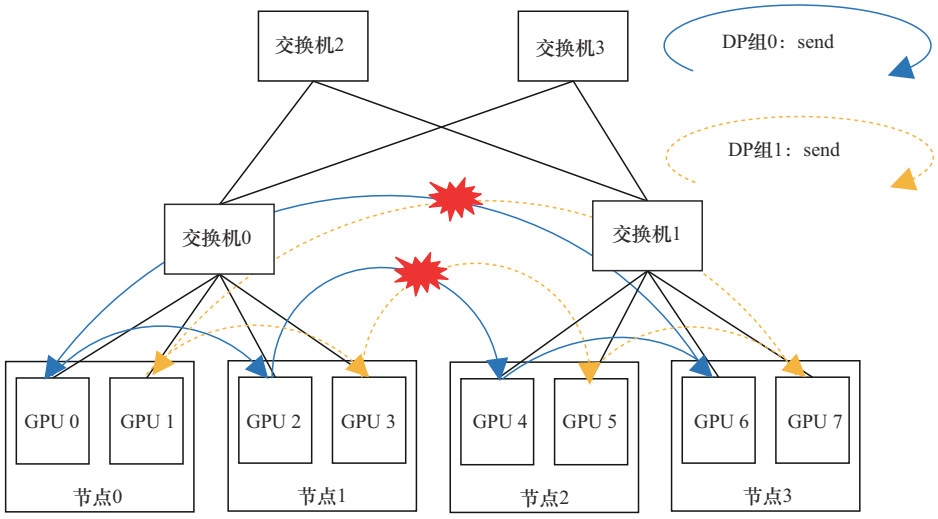


图4 DP AllReduce哈希冲突

的五元组进行了分析：源IP/目的IP地址在传输会话建立后即固化不可变更，目的端口固定为4791，协议类型固定为用户数据报协议（user datagram protocol，UDP），以此构成了RoCEv2协议的封装规范。值得关注的是，RoCEv2报文中的源端口字段与协议的标准封装无关，因此，利用该字段实现ICE动态负载均衡（ICE-LB）方案。具体而言，该方案通过预探测建立源端口与ECMP路径的映射关系，并依托全局算路构建基于源端口动态调整的路径规划机制，突破传统ECMP在动态负载均衡方面的局限性。ICE-LB的具体流程如下。

（1）可用路径探测。受相关研究工作[22]的启发，提出基于Traceroute的ICE非侵入式路径发现机制：针对特定的通信元组<源IP地址, 源端口, 目的IP地址, 目的端口>，通过递增式存活时间（time to live，TTL）探测包捕获各跳网络设备IP地址，并基于拓扑感知状态信息，构建源端口（sport）与物理路径的映射字典，其核心映射逻辑可形式化为$f(\text{sport})\rightarrow\{\text{dev}_1, \text{dev}_2, \cdots, \text{dev}_n\}$，最终实现ECMP多路径的端到端拓扑感知，得到ICE_Path。

（2）全局主动算路。训练过程中，各算力节点内ICE单元将上报集合通信的通信域信息。各节点周期性上报的通信域信息涵盖各QP中的源/目的GPU标识符，构成通信矩阵。ICE构建通信矩阵示例见表3。ICE路径建模单元通过聚合全局通信域信息，结合集群网络拓扑感知结果ICE_Topo、路径感知结果ICE_Path，将其整合至网络控制单元执行主动算路。

ICE-LB全局算路算法见算法2。ICE-LB基于贪婪算法为每个QP计算不产生哈希冲突的UDP源端口，并下发到各算力节点。得益于模型训练过程的时序规律性，在无网络设备故障期间，节点间通信关系可维持稳定状态，实现流量网络侧零冲突转发。

表3　ICE构建通信矩阵示例

连接ID	源GPU	目的GPU	端口	流经交换机ID
0	0	2	50 003	0
1	1	3	50 006	0
2	2	4	50 001	0，2，1
3	3	5	50 005	0，3，1
4	4	6	50 004	1
5	5	7	50 002	1
6	6	0	50 008	1，2，0
7	7	1	50 007	1，3，0

算法2　ICE-LB全局算路算法

输入　通信矩阵***M***，ICE引擎拓扑感知结果ICE_Topo、路径感知结果ICE_Path

输出　修改各连接端口后的通信矩阵***M***

（1） **for** QP **in** ***M***:

（2）　SW=GetLeafSwitch(ICE_Topo, QP.src)

（3）　cur_link=GetMinUpLink(SW)

（4）　sport=GetSport(ICE_Path, QP, cur_link)

（5）　QP.sport=sport

（6）　path=GetQPPath(ICE_Path, QP, sport)

（7）　**for** link **in** path:

（8）　　add QP into link.QPs

上述算法中，步骤2通过解析QP源地址与交换矩阵的对应关系，确定QP源所属的Leaf交换机节点SW，完成网络拓扑层级的精确定位；步骤3基于贪婪算法策略，选取当前QP占用率最低的上行链路作为最优传输路径；步骤4从ICE路径感知结果获取路径的ECMP映射源端口——sport；步骤5~8将得到的sport与对应的QP链路进行绑定，并对该QP的链路占用状态进行更新。

基于上述机制，ICE可以为各QP连接分配转发零冲突的UDP源端口（见表3的第4列）。随后，ICE引擎将源端口配置下发至对应算力节点内的TeleCCL，完成路径规划结果使能。

2.2.3　ICE-CC主被动拥塞控制

传统拥塞控制方案基于显式拥塞通告[23]

（explicit congestion notification，ECN）、往返时延（round-trip time，RTT）或INT等信号构建交换机队列状态感知体系，通过发送端速率调节实现拥塞缓解。该机制适用于对带宽与时延非敏感的传统通信场景，但在具有突发流量特征且尾时延决定系统性能的智算网络中，若等待队列堆积后启动调速机制，易因响应延迟触发优先级流量控制[24]（priority flow control，PFC）；而依据队列波动采取保守调速策略则可能导致过度降速，进而损害训练整体吞吐效率。

传统拥塞控制机制中，各发送节点因全局信息感知能力受限，无法实时获取其他节点的发送时序与流量规模等关键参数。这种分布式决策模式迫使各节点依赖启发式算法进行局部最优解搜索，并通过多轮次速率调整逐步逼近理论最优传输速率。究其本质，传统业务流量的突发性与非周期性特征导致全局控制器难以建立有效的通信行为预测模型，进而无法实现基于先验知识的主动式资源调度优化。

ICE-CC算法分为启发式拥塞控制、全局主动拥塞控制两种模式，如算法3所示，算法中涉及的参数符号定义见表4。在训练首轮迭代期间，RNIC执行启发式拥塞控制模式。已有研究表明，基于RTT的启发式调速可以达到比DCQCN更优的效果[25]。为此，ICE-CC构建目标往返时延阈值（T_{target}）作为链路队列深度与吞吐优化的约束条件，并创新地引入基于当前RTT观测值及RTT梯度（T_{grad}）特征的双参数预测机制，通过动态预估下一周期RTT提升网络状态感知精度。需说明的是，算法中设置当前RTT探测一定在上一次RTT探测结束后开始，因此T_{pred}不会为0。自次轮迭代起，ICE引擎开始向各算力节点中的ICE单元下发流控信息，进入全局主动拥塞控制模式。

算法3 ICE-CC主被动拥塞控制算法

输入 T_{base}、T、R_T，超参数R_{AI}、γ、w、b

输出 当前QP的发送速率R_c

（1） **if** R_T **is** within its validity period **then**

表4 ICE-CC算法中涉及的符号定义

算法符号	含义
T_{base}	链路空载往返时延
T	本次探测得到的往返时延
T_{prev}	上次探测得到的往返时延
t	本次探测开始的时刻
t_{prev}	上次探测开始的时刻
w	权重因子
b	偏置因子
R_{AI}	加性增速因子
γ	乘性减速因子
R_T	ICE引擎下发的目标速率
μ	平滑因子

（2） $R_c=\mu R_T+(1-\mu)R_c$

（3） **else if** NACK received **then**

（4） $R_c=R_c/2$

（5） **else**

（6） $T_{\text{target}}=T_{\text{base}}+\dfrac{w}{\sqrt{R_c}}+b$

（7） $T_{\text{grad}}=\dfrac{T-\mathrm{T}_{\text{prev}}}{t-t_{\text{prev}}}$

（8） $T_{\text{pred}}=T(T_{\text{grad}}+1)$

（9） **if** $T_{\text{pred}}<T_{\text{target}}$ **then**

（10） $R_c=\min\left(R_c+R_{\text{AI}},\dfrac{R_c}{1-\gamma}\right)$

（11） **else**

（12） $R_c=R_c\left(1-\gamma\left(\dfrac{T_{\text{pred}}-T_{\text{target}}}{T_{\text{pred}}}\right)\right)$

算法3中步骤1~2是全局主动拥塞控制模式，当收到来自ICE引擎下发的目标速率，并且处于有效时间内时，算法通过指数滑动平均实现发送速率的渐进式收敛。步骤6~12是启发式拥塞控制模式，当一段时间未收到目标速率后，算法会自动进入该模式，实现端侧自主调速。其中，步骤6中使用对数据包在交换机排队引入的延迟进行估计；步骤7~8使用RTT梯度预测机制，提升算法调速的准确性；步骤10和步骤12执行加性增和乘性减策略的双向调速策略。

3　实验与评估

为验证ICE方案对大模型训练网络通信的提升效果，本文进行了物理真机实验。物理实验分别在2种环境中开展。（1）基于英伟达环境进行ICE方案性能验证。该选择源于以下考量：英伟达GPU硬件架构及其配套的NCCL当前处于行业领先地位，可有效规避异构硬件差异及第三方软件缺陷对性能评估的干扰，从而客观呈现ICE的基准网络性能。（2）将实验扩展至英伟达-天数异构环境，重点验证ICE在异构算力的兼容性及其异构智算网络中的实际性能。

3.1　英伟达环境实验

英伟达环境集群拓扑如图5所示，拓扑为二层Leaf-Spine架构，无收敛组网，交换设备均为英伟达Spectrum-4 SN5600交换机。每个Leaf交换机下连2台GPU服务器，GPU型号为英伟达H800；网卡为英伟达BlueField-3（BF3） DPU，链路速率为400 Gbit/s。

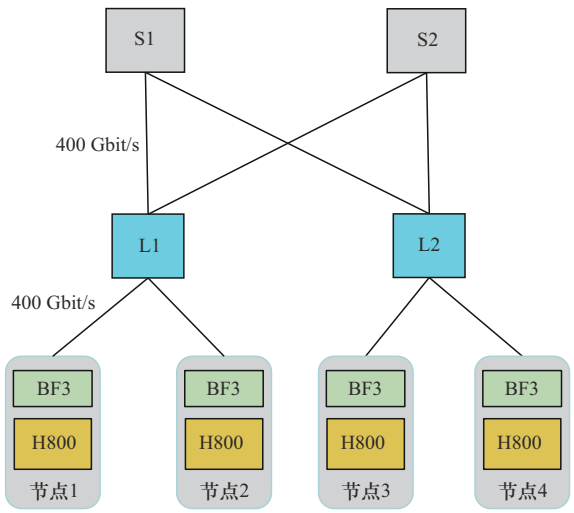


图5　英伟达环境集群拓扑

AllReduce和All-to-all分别对应DP、EP并行数据同步的过程，是导致网络传输瓶颈的关键集合通信操作，因此，本节重点测试导致RoCEv2方案性能不足的上述2种集合通信模式。使用NCCL-Tests进行测试，AllReduce流量从1 GB按2的幂次序列递增，直至16 GB；All-to-all流量从4 MB按2的幂次序列递增，直至512 MB。对比方案为RoCEv2方案（ECMP+DCQCN）和逐包自适应路由的NV Spectrum-X方案（NV-AR+DCQCN）。由于华为星河AI仅支持特定设备、博通DDC尚未大规模商用、互联网方案（如阿里云、腾讯星脉）技术闭源和租用其服务器无法实现自定义拓扑，故未纳入对比。

NCCL-Tests的AllReduce测试结果如图6所示。ICE方案平均吞吐量优于RoCEv2和NV Spectrum-X方案。相较于RoCEv2，ICE方案减少了跨交换机通信次数、彻底避免了哈希冲突，并且充分利用链路带宽，使平均吞吐量最高提升47%。尽管NV Spectrum-X基于逐包负载均衡解决了哈希冲突问题，但其引入了包粒度的分组头开销和乱序重排性能损耗，造成带宽浪费。相较之下，ICE通过软件层实现零冲突路径规划，仅需流粒度UDP源端口配置（开销可忽略），且无传输时额外损耗，故平均吞吐量更优。因此，相较于NV Spectrum-X，ICE仍可以取得10 Gbit/s（约2%）的平均吞吐量提升。

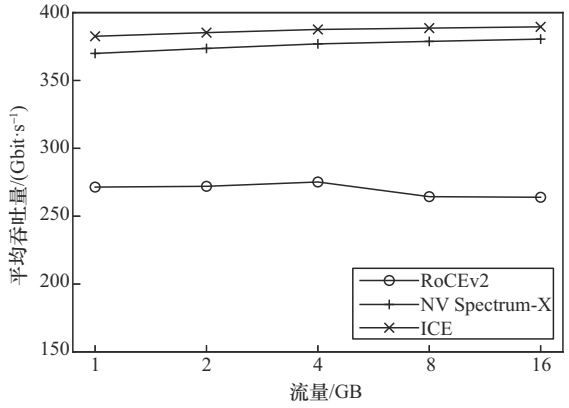


图6　NCCL-Tests的AllReduce测试结果

NCCL-Tests的All-to-all测试结果如图7所示。相较于RoCEv2方案，ICE在平均吞吐量上具有显著优势，最高提升达22%。ICE的性能

提升主要源于两方面：其一，全局算路机制有效解决了哈希冲突问题；其二，针对All-to-all通信中因流量启动异步暴露的“慢节点”问题，传统启发式拥塞控制需经多轮速率调整才能实现公平性，此过程会显著损害整体带宽。而ICE通过主动拥塞控制使“快节点”预先限速，确保所有节点加入时即达到公平发送速率，实现最后一跳零拥塞，从而有效提升吞吐量。

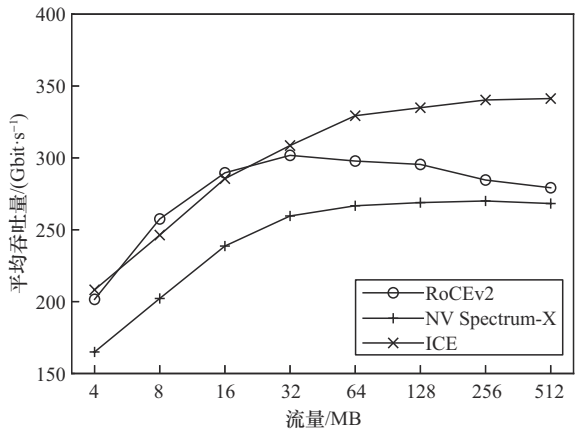


图7　NCCL-Tests的All-to-all测试结果

此外，NCCL-Tests的All-to-all测试结果揭示了Spectrum-X在All-to-all通信中存在显著性能损失。其可能的原因在于：（1）All-to-all流量模式导致逐包打散引发大量接收端乱序报文，乱序重排降低处理效率；（2）逐包负载均衡致使拥塞反馈信息无法准确反映瓶颈链路状态，引发调速决策偏差。这表明，相较现有商用方案，ICE采用的逐流控制机制具有更高简洁性与效能。

3.2　英伟达–天数异构环境实验

英伟达–天数异构环境集群拓扑如图8所示，其在英伟达环境的基础上进行扩展，每个Leaf交换机下部署天数BI-V150 GPU服务器，构建异构PP非均质拆分训练环境；增设独立Spine交换机构建无阻塞网络。该拓扑有效模拟了异构算力训练场景。选用天数BI-V150 GPU服务器进行实验，因其厂商是完成ICE方案适配的代表，且该型号为最新商用产品。

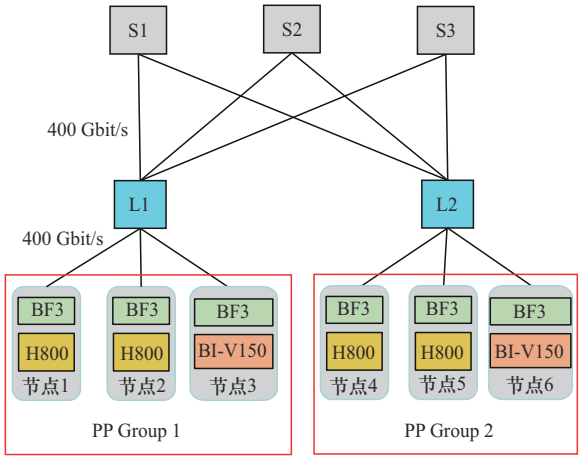


图8　英伟达–天数异构环境集群拓扑

使用TeleCCL-Tests基准测试工具对英伟达–天数异构算力集群进行性能评估。在异构PP非均质拆分训练场景中，涉及的异构跨机通信主要为DP并行梯度更新的AllReduce通信。因此，重点评估该场景下的梯度更新AllReduce通信性能。实验设计中，基于H800与BI-V150 4:1的算力比例关系，设定其通信流量配比为4:1。测试参数配置方面，H800节点的通信流规模按2的幂次序列从1 GB扩展至16 GB。对比方案为RoCEv2方案（ECMP+DCQCN）和逐包自适应路由的NV Spectrum-X方案（NV-AR+DCQCN）。

TeleCCL-Tests异构AllReduce测试结果如图9所示。相较于RoCEv2方案，ICE将平均吞吐量最高提升了38%，主要得益于路径规划机制有效规避了ECMP哈希冲突引发的负载不均问题，从而提升了链路利用率。相比NV Spectrum-X方案，ICE通过降低分组头开销与乱序重排损耗，实现平均吞吐量提升了8%，这与NCCL-Tests测试结果趋势一致。实验表明，ICE可为异构算力混合训练场景提供高性能智算网络传输能力。

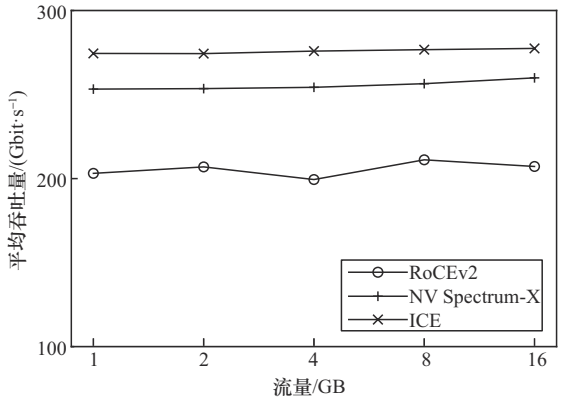


图 9　TeleCCL-Tests 异构 AllReduce 测试结果

4　结束语

本文对当前智算网络的研究进展、存在的挑战进行了分析，针对智算网络在异构算力互联场景下的技术挑战，创新性地提出了智能控制以太网解决方案——ICE。ICE 让网络具备感知智算业务的能力，实现从“尽力而为”到“确定性服务”、从“被动传输”到“主动协同”的范式转换。此外，本文通过物理环境实验验证了 ICE 方案的有效性。

实验测试结果表明，在 RoCEv2 网络部署环境下，ICE 使 AllReduce 集合通信吞吐量较传统方案最高可提升 47%，All-to-all 场景中的吞吐量最高提升 22%。在真实异构算力环境中，ICE 实现异构集合通信性能最高提升 38%。实验结果验证了 ICE 在高性能、高兼容性及智能运维方面的技术可行性，为国产智算网络建设提供了可推广的工程范式。

在接下来的工作中，ICE 方案在技术演进与生态构建层面仍具有广阔发展空间。首先，在协议扩展性方面，当前方案通过动态修改 UDP 源端口实现 ECMP 选路控制，虽能兼容标准交换机，但缺乏逐跳路径规划的精细度。下一步拟联合设备厂商推进 RoCE 协议定制化开发，进一步提升网络资源利用率。

此外，加州大学伯克利分校近期基于 NCCL 开源了 UCCL 方案[26]，该方案将负载均衡、拥塞控制直接转移到主机侧，绕过网卡硬件能力的限制，为智算网络优化提供了另一种思路。该方案目前仅面向同构英伟达与 AMD GPU 集群，依赖深度修改 NCCL 源码，需厂商开放硬件接口（如 GPUDirect），尚不能与异构 GPU 算力适配。后续工作中，笔者也会探索将 ICE 与 UCCL 联合优化的可行性。

此外，针对当前需求愈发明确的高性能推理场景，需进一步研究集中路径规划面对多用户并发推理请求时的优化方案。笔者将联合科研机构、设备厂商与互联网企业，建立开放兼容的智算异构算力技术联盟，推动 ICE 核心模块的标准化进程。

参考文献：

[1] HOWARTH J. Number of parameters in GPT-4[EB]. 2025.

[2] PATEL D, NISHBALL D，ONTIVEROS J E. Multi-datacenter training: OpenAI's ambitious plan to beat Google’s infrastructure[EB]. 2024.

[3] NARAYANAN D, SHOEYBI M, CASPER J, et al. Efficient large-scale language model training on GPU clusters using megatron-LM[C]//Proceedings of the SC21: International Conference for High Performance Computing, Networking, Storage and Analysis. Piscataway: IEEE Press, 2021: 1-14.

[4] SHANLEY T. InfiniBand network architecture[M]. Boston: Addison Wesley Developer's Press, 2003.

[5] NVIDIA. RDMA over converged Ethernet (RoCE) [EB]. 2025.

[6] 王欣. 英伟达加快 AI 芯片路线图：黄仁勋透露 GPU 将一年一更[EB]. 2025.
WANG X. NVIDIA accelerates AI chip roadmap: Huang Renxun reveals GPU will be updated annually[EB]. 2025.

[7] The White House. Fact sheet: ensuring U. S. security and economic strength in the age of artificial intelligence[EB]. 2025.

[8] ZHU Y B, ERAN H, FIRESTONE D, et al. Congestion control for large-scale RDMA deployments[J]. ACM SIGCOMM Computer Communication Review, 2015, 45(4): 523-536.

[9] GANGIDI A, MIAO R, ZHENG S B, et al. RDMA over Ethernet for distributed training at meta scale[C]//Proceedings of the ACM SIGCOMM 2024 Conference. New York: ACM, 2024: 57-70.

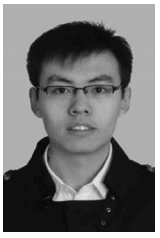
[10] 宋婧. 腾讯发布星脉网络 2.0, AI 大模型训练效率提升 20%[EB]. 2024.
SONG J. Tencent releases Star Pulse Network 2.0, increasing AI model training efficiency by 20%[EB]. 2024.

[11] 华为, 中国信通院. 星河 AI 网络白皮书[R]. 2025.

HUAWEI, CAICT. Xinghe network white paper[R]. 2025.
[12] JACOBS R A, JORDAN M I, NOWLAN S J, et al. Adaptive mixtures of local experts[J]. Neural Computation, 1991, 3(1): 79-87.
[13] NVIDIA. NVIDIA Spectrum-X networking platform[EB]. 2025.
[14] WU X G. Reducing job completion time in AI/ML clusters[EB]. 2025.
[15] DONG J B, WANG S C, FENG F, et al. ACCL: architecting highly scalable distributed training systems with highly efficient collective communication library[J]. IEEE Micro, 2021, 41(5): 85-92.
[16] MIAO R, ZHU L J, MA S, et al. From luna to solar: the evolutions of the compute-to-storage networks in Alibaba cloud[C]// Proceedings of the ACM SIGCOMM 2022 Conference. New York: ACM Press, 2022: 753-766.
[17] LI Y L, MIAO R, LIU H H, et al. HPCC: high precision congestion control[C]//Proceedings of the ACM Special Interest Group on Data Communication. New York: ACM Press, 2019: 44-58.
[18] JUNIPER NETWORK. Paragon insights data ingest guide[EB]. 2022.
[19] LI B J, WANG X L, WANG J Z, et al. TCCL: co-optimizing collective communication and traffic routing for GPU-centric clusters[C]//Proceedings of the 2024 SIGCOMM Workshop on Networks for AI Computing. New York: ACM Press, 2024: 48-53.
[20] 杨维. 多芯异构混池训练技术及实践[EB]. 2025.
YANG W. Multi core heterogeneous mixed pool training technology and practice[EB]. 2025.
[21] NVIDIA. Developing a Linux Kernel Module using GPUDirect RDMA[EB]. 2025.
[22] LIU K F, JIANG Z, ZHANG J, et al. R-pingmesh: a service-aware RoCE network monitoring and diagnostic system[C]// Proceedings of the ACM SIGCOMM 2024 Conference. New York: ACM Press, 2024: 554-567.
[23] RAMAKRISHNAN K, FLOYD S, BLACK D. The addition of explicit congestion notification (ECN) to IP: RFC3168-2011[S]. 2011.
[24] IEEE. Qbb－priority-based flow control：802.1-2011[S]. 2011.
[25] KUMAR G, DUKKIPATI N, JANG K, et al. Swift: delay is simple and effective for congestion control in the datacenter[C]// Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication. New York: ACM Press, 2020: 514-528.
[26] ZHOU Y, CHEN Z J, MAO Z M, et al. An extensible software transport layer for GPU networking[J]. arXiv preprint，2025, arXiv:2504.17307.

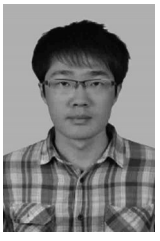
[作者简介]

苏昱臻（1998－），男，中国电信股份有限公司研究院工程师，主要研究方向为数据中心RDMA网络与异构计算。



王子潇（1995－），男，中国电信股份有限公司研究院工程师，主要研究方向为高性能RDMA网络与异构计算。

钟驰量（2001－），男，北京邮电大学硕士生，主要研究方向为可编程网络与RDMA。



寇晓淮（1989－），男，中国电信股份有限公司研究院工程师，主要研究方向为数据中心网络。

刘圆（1979－），男，博士，中国电信股份有限公司研究院工程师，主要研究方向为智能计算基础设施和大语言模型。

陈映（1993－），男，中国电信股份有限公司研究院工程师，主要研究方向为智算基础设施。