超以太网技术的现状与展望

Status and Prospect of Ultra-Ethernet Technology

厉俊男/LI Junnan，李韬/LI Tao，杨惠/YANG Hui
（国防科技大学，中国 长沙 410073）
(National University of Defense Technology, Changsha 410073, China)

DOI：10.12142/ZTETJ.202406008
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20250108.1617.002.html
网络出版日期：2025-01-08
收稿日期：2024-10-17

摘要：随着数据中心、智算中心规模的急剧增长，传统以太网技术在通信带宽和延时等方面面临巨大挑战。深入分析传统以太网的优缺点，从物理层、链路层、传输层和软件层4个方面梳理了超以太网技术，并对其中的关键技术展开详细的介绍和研究。此外，还分析了超以太网相关技术在中国的发展现状。最后，探讨了超以太网技术发展面临的机遇与挑战。

关键词：超以太网；网络协议；人工智能

Abstract: With the rapid growth of data centers and intelligent computing centers, traditional ethernet protocols face enormous challenges in terms of communication bandwidth and latency. The advantages and disadvantages of traditional ethernet are analyzed. Then the key technologies of super Ethernet are introduced from four aspects: physical layer, link layer, transmission layer, and software layer. In addition, the development status of related technologies in China is also analyzed. Finally, the opportunities and challenges faced by the development of Ultra Ethernet technology are discussed.

Keywords: ultra ethernet; network protocols; artificial intelligence

引用格式：厉俊男, 李韬, 杨惠. 超以太网技术的现状与展望 [J]. 中兴通讯技术, 2024, 30(6): 48-53. DOI: 10.12142/ZTETJ.202406008
Citation：LI J N, LI T, YANG H. Status and prospect of ultra-ethernet technology [J]. ZTE technology journal, 2024, 30(6): 48-53. DOI: 10.12142/ZTETJ.202406008

随着人工智能（AI）技术的飞速发展，AI对算力的需求呈现指数级增长的态势，这促使围绕大规模分布式计算的基础设施建设迅猛发展。在庞大的分布式系统中，网络作为连接各个节点的“神经脉络”，不仅是数据流通的管道，更是基础设施互联的粘合剂。网络将算力资源、存储资源以及各类智能应用紧密地结合在一起。然而，面对AI技术日益增长的复杂性和高性能要求，现有的网络技术如以太网和IB（InfiniBand）网络，在各自领域内表现出色，但逐渐显露出局限性。

AI技术的发展迫切需要一种更加先进、高效、灵活且成本可控的网络解决方案。这种网络需要具备更大规模的扩展能力，以应对不断增长的算力需求；需要更高的带宽，以确保数据传输的畅通无阻；需要支持多路径传输，以提高网络的可靠性和容错性；需要实现对拥塞的快速反应和智能调度，以保证数据传输的实时性和稳定性；同时，还需要充分考虑单个数据流执行度的相互依赖关系，特别是尾延迟这一关键因素，以确保AI应用的整体性能和用户体验。

基于这样的背景，我们有必要重新审视和评估现有的网络技术，积极探索和研发能够满足智能计算需求的新型网络技术。这不仅是一次对网络技术的重大挑战，更是推动AI技术持续发展的关键所在。

1 超以太网技术发展背景

超以太网传输（UET）架构主要是从物理层、链路层、传输层与软件层4个方面来改进以太网技术，既兼容现有的以太网生态，又能提升以太网的交换转发性能，从而改进存储、管理、安全结构，提升遥测能力。超以太网传输技术由业界领军企业组成的非盈利性组织——超以太网联盟（UEC）提出，其目的是优化现有以太网技术，开发高性能全栈架构，满足当前人工智能（AI）对网络性能、灵活性和成本效益的严苛需求，推动相关技术的研发、标准制定及市场推广，以引领未来网络技术的发展方向。

基金项目：国家重点基础研究发展计划项目（2010CB328200、2010CB328201）；国家高技术研究发展计划项目（2006AA01Z257）；国家自然科学基金项目（60602058、60572120）；国家科技重大专项项目（2009ZX03003-002-02）

1.1 以太网的优势与面临的挑战

以太网自1973年诞生至今，获得了巨大成功：速率从最早的10 Mbit/s发展到如今的100 Gbit/s、200 Gbit/s甚至400 Gbit/s；广泛应用于各类AI训练的大型集群中。以太网/IP协议族具有众多优势：

1）具有极好的通信生态。以太网协议已经十分成熟，拥有广泛的应用市场。支持以太网协议的包括以太网交换机、网卡、线缆、收发器、光电转换等设备厂商，以及相应的以太网管理工具厂商。以太网使用标准的网络设备和标准化的通信协议，这使得部署和维护成本较低。

2）支持高带宽互连。以太网高达每秒数百吉比特的传输速率，可为数据中心提供计算资源之间的高速互联，也可为用户提供高速的网络资源访问能力，以满足现代网络用户对速度和效率的需求。

3）具备较高的可靠性。以太网通信是一种可靠的通信技术，采用错误检测和冗余机制，可以保证关键任务或者敏感数据传输的完整性和正确性。

4）易于管理。以太网不仅管理结构相对简单，同时具有丰富的网络管理工具和配置协议，能够有效简化网络管理员的配置和网络监管，提升管理效率。

5）配套使用的IP协议也非常成熟。IP网络支持大规模的路由寻址，能够支持机架级、园区级和数据中心级网络。

以太网的众多优势造就了其在AI计算领域的广泛应用。随着AI模型对算力需求的急剧增加，网络成为互联分布式计算资源的关键，并在AI大模型训练中变得越来越重要。

大型语言模型（LLM）如GPT-3、Chinchilla和PALM，推荐系统如深度学习推荐模型（DLRM）、深度和层次化集成网络模型（DHEN），都是在数千个图形处理器（GPU）的集群上进行训练的[1]。这些大型语言模型通常采用分布式训练方式，不同计算节点间存在频繁数据交互过程，即每启动新的一轮计算需要等待所有计算节点完成上一轮计算和数据交互。不同节点间数据交互过程最后一个消息到达的时间决定了下一轮计算阶段启动的时间。因此，尾部延迟通常是AI分布式计算系统性能的关键指标。

大型模型的参数数量持续增加，上下文窗口范围持续扩大。例如，2020年GPT-3拥有1 750亿参数[2]，而最近发布的GPT-4模型已有近一万亿参数[3]，DLRM更是拥有数万亿参数[1]，并仍会继续增长。这些规模愈发庞大的AI模型需要更大的集群以提供相应的训练算力，配套更高的通信带宽以实现数据交互。与此同时，网络时延也愈发重要，如网络的延时拥塞会造成集群中昂贵计算资源的闲置。

1.2 超以太网联盟

UEC由AMD、博通、思科、英特尔、Meta和微软等10家来自芯片、通信、互联网行业的领导者于2023年牵头成立，旨在完善以太网标准，以更好地满足人工智能、机器学习和高性能计算不断增长的需求[4]。

目前，UEC发展迅速，除了牵头的10家厂商外，已有超80家知名厂商加入该联盟，包括芯片设计、计算、通信、互联网等主流厂商，如IBM、Candence、Synopsys、瞻博网络、戴尔等。中国厂商也积极加入该联盟，如中兴通讯、华为、新华三、百度等。其中，阿里巴巴加入UEC技术委员会，与Meta、AMD、博通和微软等其他12名成员，一同推进以太网核心计算的研发工作和相关标准制定工作。

UEC成立之初划分了4个工作组，分别是物理层、链路层、传输层和软件层工作组。

1）物理层工作组。该工作组制定以太网物理层规范、电气和光信号特性规范，开发应用程序接口和定义相关数据结构，以提高物理层传输性能，降低传输延迟，改善以太网物理层配置管理。当前物理层工作组主要制定100 GbE和200 GbE速率端口物理层协议（PHY）规范。目前已经确定了100 GbE介质类型、PHY支持的速率和类型，200 GbE的规范还在制定中。

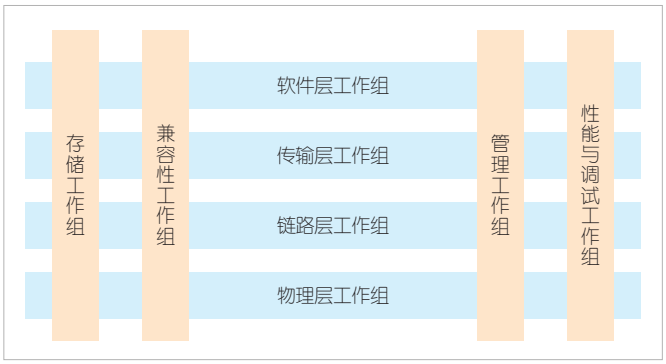
2）链路层工作组。该工作组主要研究链路层可靠性、多路径与报文喷洒策略、链路特性协商机制，以提升链路层传输的可靠性、传送效率和遥测能力。

3）传输层工作组。通过研究拥塞控制算法、安全策略、宽松的报文重排序机制，传输层工作组避免基于以太网的远程直接内存访问（RoCE）传输可能存在尾延时大的缺点，解决报文可靠传输、数据安全传送、应用程序扩展等难题。

4）软件层工作组。该工作组采用了兼容现有通信库的方法。现有通信库包括集合通信库（CCL）、信息传递接口（MPI）、共享内存（SHMEM）通信库等。它们使用libfabric作为数据平面框架的应用程序编程接口（API），这有助于上层应用的快速开发和部署。同时还定义了加速器和高速扩展接口（FEP）之间的交互方式，即各类加速器API。通过同一交换机、FEP以及聚合管理器（AM）的控制平面和数据平面接口定义，解决不同UEC供应商之间的互联互通、互操作问题。

近期，UEC又成立了存储、管理、兼容性与测试、性能与调试工作组，如图1所示，旨在完善超以太网系统应用、互操作等方面的能力。在此基础上，超以太网联盟已发布超以太网白皮书1.0版本[5]。

此外，UEC与其他开源联盟关系密切，如开放计算项目



▲图1 超以太网联盟工作组划分

联盟（OCP）、全球网络存储工业协会（SNIA）、OpenFabric联盟（OFA）、IEEE 802.3工作组等。

2 超以太网关键技术

为提升以太网传输带宽，降低尾延时，我们从物理层、链路层、传输层和软件层4个方面来研究各层协议的优化技术。图2展示了超以太网协议栈整体架构，物理层旨在提升以太网速率，支持100～200 Gbit/s；链路层则是优化传输可靠性和传输性能；传输层主要从拥塞控制、可靠性传输、数据安全、链路遥测4个方面优化设计；软件层/应用层通过拓展协议库为上层应用提供相应服务。

2.1 物理层技术

物理层除了上文提及的制定以太网物理层规范、电气和光信号特性规范，开发应用程序接口和定义相关数据结构外，还研究链路质量预测与评估的概念，并制定相关指标如误码率（UCR）、PHY平均错误时长（MTBPE）、平均误包被接受时长（MTTFPA），以更精确地预测和度量物理层链路质量。其中，误码率用于标识链路上数据报文发生错误的频率，PHY平均错误时长用于标识PHY接口的错误率，平均误包被接受时长用于标识误收错误报文的比例。

2.2 链路层技术

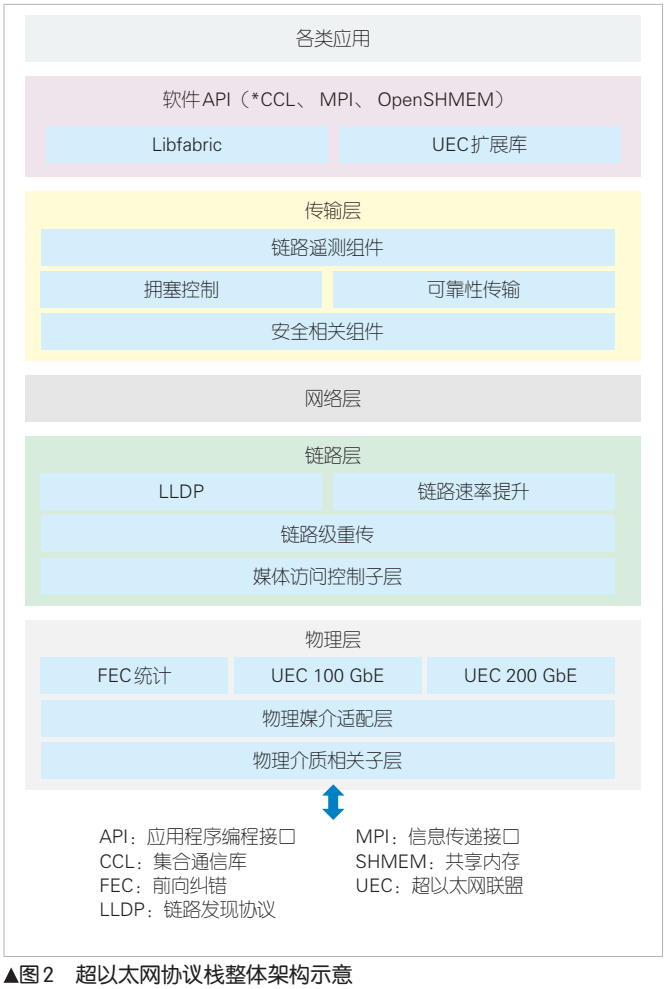
链路层工作组从可靠性、报文传输效率以及多路径与报文喷洒3个方面来提升链路层传输可靠性和传输性能。

1）链路可靠性保证机制。该机制通过在链路层的逻辑链路管理（LLC）和MAC管理之间设计和插入新的子层——链接级别重试（LLR），以构建链路层端到端错包重传。

2）报文传输效率提升策略。该策略针对智能计算大量消息有效载荷在16字节的特点，以及传统以太网短报文有效载荷比率过小的问题，采用以太网报文头压缩策略，增加帧的传送效率。为兼容现有以太网协议，报文头中设计了压缩标识信息，用于区分压缩报文和非压缩报文，从而允许两类报文可在网络中共存，而不影响原有的功能。

3）多路径与报文喷洒。传统的以太网网络基于生成树协议，确保从A到B的单一路径，以避免网络中的环路。随后出现了多路径技术，例如等价多路径（ECMP）[6]，网络尝试利用尽可能多的链路来连接通信对象。ECMP通常使用“流哈希”，它将不同五元组的流量映射到不同路径上，也可以将不同五元组的流量映射到同一条路径上。然而，这种方法可能将高吞吐量流量限制在一条路径上，当过多的流量映射到单一网络路径时，网络性能会下降，因此需要对负载均衡进行精细管理以获得最佳性能。

超以太网的基本思路是将单条流的不同报文同时分散到所有可以通往目的地的路径上，这种技术被称为“报文喷洒”，可以更加均衡地利用所有的网络路径。这种更灵活的多路径策略会引入报文频繁乱序的问题。如果仍然采用严格的报文排序要求，则会阻止乱序报文直接从网络传输到应用程序缓冲区，最终限制传输效率。



▲图2 超以太网协议栈整体架构示意

在AI工作负载中，大量GPU或者加速器之间数据需要频繁交互。这其实是一种“集合”通信，包含All-Reduce和All-to-All两种模式，其中All-Reduce通过单节点上获取所有节点信息，并执行Reduce操作；All-to-All作为全交换操作，通过All-to-All通信，可以让每个节点都获取其他节点的信息。考虑到AI应用程序只关心给定消息的最后部分何时到达目的地，集合通信快速完成的关键是节点间快速完成批量传输。针对上述两种交互方式，采用报文喷洒可有效降低数据交互的尾延时。

2.3 传输层技术

传输层工作组从可靠性传输、数据安全、拥塞控制、链路遥测4个方面展开研究。四者的关系如图3所示，拥塞控制与链路遥测密切配合，通过感知链路状态以准确、及时反馈拥塞状态；可靠传输与拥塞控制构成报文传输模块；安全模块负责数据的加解密以及密钥的分发管理。

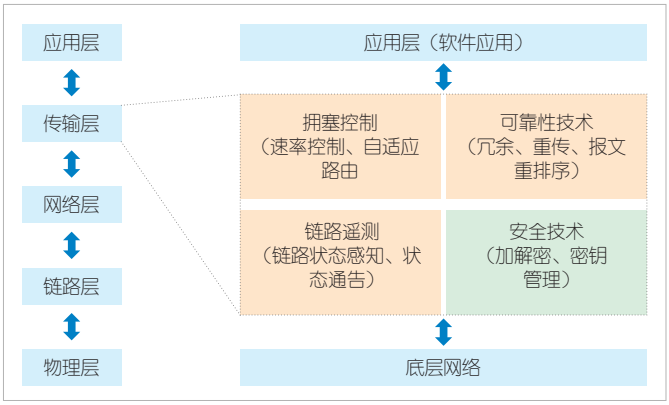
1）支持报文乱序的可靠传输技术。全链路的报文喷洒必然会引入更多的报文乱序，为此针对不同应用需求可设置3类不同的报文传输模式：

（1）可靠，有序传输（ROD）。该模式按照顺序传输报文，用于需要消息有序传输的应用。

（2）可靠，无序传输（RUD）。该模式只能向语义层传输一次报文，但可以忍受网络中的乱序传输。可靠性传输层需检测重复报文，以确保每个报文只能向语义层传送一次。

（3）不可靠，无序传输（UUD）。不可靠报文可以承载许多UET的新语义，用户无须可靠传输，通过其他方式就可保障可靠性。

2）安全传输。安全传输机制作为超以太网传输技术的重要研究内容，可针对业务的需求，以及任务对延时、吞吐率的要求，选择报文头、部分报文或者全部报文数据负荷加密和认证。



▲图3 传输层关键设计

3）拥塞控制技术。网络拥塞可能发生在3个地方：发送方到第一跳交换机的出站链路、第一跳交换机和最后一跳交换机之间的链路、最后一跳交换机到接收方的最后链路。对于AI来说，发送方出站链路上的拥塞主要可以通过发送主机上的调度算法进行控制，因为主机可以看到所有出站流量。上文中提到的多路径报文喷洒通过均匀分配所有路径上的负载，最小化了第一跳和最后一跳交换机之间的热点和拥塞。拥塞的另一种形式——“Incast”，发生在多个发送者同时向同一目标发送流量时[7]，即最后一条到接收方的链路上。Incast既可能发生在“All-Reduce”过程，也可能发生在“All-to-All”过程。

近年来，学术界和工业界对拥塞控制展开了广泛研究，提出了许多优秀的拥塞控制算法，如数据中心量化拥塞通知（DCQCN）[8]、数据中心TCP拥塞控制协议（DCTCP）[7]、简单有效的拥塞控制（SWIFT）[9]、Timely[10]等。但是上述拥塞控制算法无法同时满足为AI优化的传输协议的所有需求，这些需求包括：

（1）在高带宽、低往返时间（RTT）网络中，当链路无拥塞时，整个网络可快速达到线速，而不对路径已存在的流量造成影响，即降低已有流量的吞吐率。

（2）感知整个网络的拥塞程度，并充分利用多路径最大限度提升传输效率。

（3）公平共享最后一跳链路来避免incast、报文丢失、重传或尾部延迟。

（4）与流量特点、硬件架构解耦，无须随着流量组合的变化、计算节点的发展、链路速度的提高和网络硬件的发展而进行调优和配置。

为此，UEC考虑在链路层采用端到端基于信用的流控机制（CBFC）来管理链路间帧的无损传输。CBFC机制用来替换基于优先级流量控制（PFC）流控。接收者周期性发送缓存空间给对端，发送者基于报文优先级和缓存大小发送报文。缓存空间也可以用于自适应路由选路，同时配合链路遥测技术，准确感知整个网络的不同链路空闲和拥塞程度，及时调度流量，快速响应链路拥塞。

4）链路遥测技术。获得理想的拥塞往往需要及时感知网络链路状态和拥塞程度，因此我们需要研究端到端遥测技术。使用该技术可以准确获得网络的拥塞情况，及时将链路拥塞信号反馈回发送端，从而实现更快的拥塞控制。无论是发送方还是接收方安排传输，现代交换机都可以通过快速传递准确的拥塞信息给调度器，促进响应式的拥塞控制，提高拥塞控制算法的响应速度和准确性。链路遥测技术减少了拥塞，降低了丢包率，缩短了队列长度，降低了尾部延迟。

此外，网络系统可通过扩展LLDP协议，方便网络设备之间协商各自支持的链路层功能。这些功能包括超以太网技术中提及的新链路层功能，如LLR、CBFC、PFC等。

2.4 软件层技术

软件层工作组除了利用现有通信库设计开发各类应用通信接口和数据结构外，还研究在网计算相关工作，包括但不限于：1）基于C语言定义在网计算（INC）所使用的软硬件交互API接口；2）描述和定义硬件在网计算能力以及软硬件关于卸载能力的协商机制；3）设计和定义相关库函数、API接口实现主机与网络节点的数据交互，调用网络节点计算资源；4）OpenConfig扩展，用于配置网络设备的前端处理器（FEP）进行集合通信卸载，并对性能和错误进行监控；5）INC在网络设备上的适配，根据INC功能特性设计配置文件，并引导UEC传输协议的开发，以便INC技术可以轻松地应用到硬件实现中。

超以太网在链路层、传输层涉及的关键技术已在业界有了相关的研究，例如链路层的报文喷洒[11]、传输层的拥塞控制[8–10]，以及软件层涉及的在网计算[12]。超以太网技术与现有以太网技术不同的是，其主要面向AI计算中分布式资源的高效数据交互，即高带宽、低延时（低平均延时、低尾延时）传输需求。为此，超以太网技术可以借鉴现有的网络协议和相关技术研究，包括但不限于可编程数据平面、可编程网络、网络虚拟化、智能拥塞控制、网络链路遥测等，并在此基础上针对应用场景的特点，设计更加高效的传输协议和技术。

3 超以太网技术在中国的相关研究

针对高性能智算需求，中国的相关企业、高校、研究机构也积极布局下一代以太网技术，成立高通量以太网联盟、人工智能算力网络推进联盟等。

3.1 高通量以太网联盟

为应对AI数据中心网络面临的挑战，阿里云与中科院计算所联合成立高通量以太网联盟，旨在利用现有的以太网生态，优化传统以太网技术，研究和定义新型以太网协议和规范，设计新型智算网络，满足AI数据中心网络对高性能和低传输延时的需求。截至目前，高通量以太网联盟已经集结了大量中国学术界知名大学、产业界各类厂商和机构，打通理论研究、试验验证、产品部署全链条。

高通量以太网联盟在2024年计算机学会高性能计算学术年会上，对外发布了高通量以太网（ETH+）协议规范（1.0版本）、基于ETH+协议的相关开源网卡等硬件和系统。高通量以太网ETH+协议通过优化以太网帧格式，有效提升以太网帧的有效载荷比（74%），大幅提高AI数据中心大量短数据报文的传输效率。此外，ETH+以太网在链路层、物理层配套设计报文重传机制，有效提升数据传输的可靠性。与此同时，ETH+还可以支持在网计算功能，将原先在单节点上实现的部分计算卸载到网络节点中实现，可有效提升集合通信性30%以上的性能，从而解决传统网络单节点计算所存在的通信、计算瓶颈问题。

与超以太网联盟组织架构不同，高通量以太网联盟成立了协议标准和产业项目两个工作组。其中，协议组设计的高通量以太网协议和规范设计能够兼容现有以太网协议，并解决传统以太网协议可扩展性不足、负载不均、性能欠佳等问题。产业项目工作组负责针对差异化应用场景，将高通量以太网协议、规范应用部署其中，并负责项目实施落地工作。同时，联盟特设产业咨询会，负责跟进产业需求、拉动产业资源；设置执行小组制定技术路线图，协同推进各小组工作，从而促进中国各个芯片公司之间的合作与交流，推动技术创新和成果转化。

3.2 人工智能算力网络推进联盟

随着人工智能技术的迅猛发展，人工智能模型规模愈发庞大，原先一些小模型逐渐消失，取而代之的是“大模型+大数据+大算力”的紧密配合模式。从2018年的GPT到现在的GPT–4，大模型对算力的需求呈现指数增长态势，传统实验室、小型数据中心提供的算力已无法满足需求。

人工智能算力中心作为智能时代的新型公共基础设施，是人工智能产业发展的基础资源保障。为发挥其公共基础设施作用，必须要构建能够支撑人工智能产业持续发展的新型管理运营机制。为了促进中国战略性新兴产业的迅速发展和繁荣壮大，发挥各行业各地方在推进人工智能技术和产业发展的积极性，在鹏城实验室的倡议和推动下，“人工智能算力网络推进联盟”（简称“智算网络联盟”[13]）成立。

智算网络联盟目前已经有鹏城实验室、华为、百度、讯飞、燧原、天数智芯、北京智源研究院、武汉智算中心、珠海横琴智算中心等近20家单位参与。智算网络联盟将会在“平等自愿、优势互补、资源共享、合作共赢”的基础上，诚挚邀请致力于推动中国人工智能算力中心发展的企事业单位、科研院所、投资机构等加入。联盟成立后将重点在“智算中心及智算网络标准的研究及标准化”“推进成立人工智能算力网络管理中心”“组织开发并建设算力网络管理信息系统”“打造品牌活动，拓展影响”4个方面开展工作，致力于构建具有中国特色的新一代信息基础设施。

中国高通量以太网联盟、人工智能算力中心等联盟的成立，与超以太网联盟、开放计算项目联盟有着相似的目标，即针对AI计算对算力需求指数增长趋势，通过优化现有以太网技术、数据中心计算架构来实现通信能力和算力提升。

4 结束语

以太网凭借其高传输带宽、低成本、随即接入能力和成熟的协议生态，已然成为数据中心和高性能计算中心内部互连互通的关键技术。然而，传统以太网的优化技术主要考虑传输带宽的提升，在传输延时方面仍存在缺陷。超以太网联盟则针对现有以太网技术存在的缺陷展开研究，对其进行优化而非彻底颠覆。这种方式使得超以太网技术更容易被现有数据中心、智算中心接受和使用。

考虑到超以太网技术目前仍在发展初期，还未形成统一的协议规范，加上以太网生态的复杂性，因此对协议标准、技术、应用进行升级是一个巨大工程。超以太网技术离真正落地部署还有较长的距离。我们认为，超以太网技术未来要取得成功不仅需要依靠技术革新，还需要构建开源开放的生态，正如博通副总裁VEKAGA所说“不会有一家公司提供所有GPU，也不会有一家公司提供所有互连解决方案”。因此，超以太网快速发展的重要途径应该是建立一个生态系统，由多个供应商提供加速器。这个生态系统的生存依赖于构建一个开放的、基于标准的、高性能的和具有成本效益的互连架构。我们可以借鉴RISC-V开源指令集的思路，制定超以太网或者高通量以太网技术中基础且必须支持的协议规范，允许各大数据中心、厂商根据差异化的应用场景自定义扩展自己的协议规范，以吸引更多厂商和机构加入其中，从而进一步推进以太网技术的落地部署。此外，超以太网联盟还必须重视商业应用所注重的低成本、低复杂度、互连互通，才有可能使得超以太网技术进一步延伸至AI计算甚至高性能计算领域。

参考文献

[1] ZHAO W X, ZHOU K, LI J Y, et al. A survey of large language models [EB/OL]. [2024-10-10]. https://arxiv. org/abs/2303.18223 v15

[2] FLORIDI L, CHIRIATTI M. GPT-3: its nature, scope, limits, and consequences [J]. Minds and machines, 2020, 30(4): 681-694. DOI: 10.1007/s11023-020-09548-1

[3] ACHIAM J, ADLER S, AGARWAL S, et al. GPT-4 technical report [EB/OL]. [2024-10-04]. http://splab.sdu.edu.cn/GPT4.pdf

[4] Ultra Ethernet Consortium. The new era needs a new network [EB/OL]. [2024-10-01]. https://ultraethernet.org/

[5] Ultra Ethernet Consortium. Overview of and motivation for the forthcoming ultra ethernet consortium specification [EB/OL]. [2024-10-01]. https://ultraethernet. org/wp-content/uploads/sites/20/2023/10/23.07.12-UEC-1.0-Overview-FINAL-WITH-LOGO.pdf

[6] ZHANG H L, GUO X, YAN J Y, et al. SDN-based ECMP algorithm for data center networks [C]//Proceedings of IEEE Computers, Communications and IT Applications Conference. IEEE, 2014: 13-18. DOI: 10.1109/comcomap.2014.7017162

[7] ZHU Y B, ERAN H, FIRESTONE D, et al. Congestion control for large-scale RDMA deployments [C]//Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. ACM, 2015: 523-536. DOI: 10.1145/2785956.2787484

[8] ALIZADEH M, GREENBERG A, MALTZ D A, et al. Data center TCP (DCTCP) [C]//Proceedings of the ACM SIGCOMM 2010 conference. ACM, 2010: 63-74. DOI: 10.1145/1851182.1851192

[9] KUMAR G, DUKKIPATI N, JANG K, et al. Swift [C]//Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication. ACM, 2020: 514-528. DOI: 10.1145/3387514.3406591

[10] MITTAL R, LAM V T, DUKKIPATI N, et al. TIMELY: RTT-based congestion control for the datacenter [J]. ACM SIGCOMM computer communication review, 2015, 45(4): 537-550. DOI: 10.1145/2829988.2787510

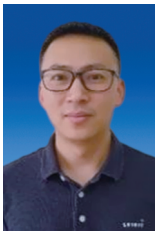
[11] ADDANKI V, GOYAL P, MARINOS I. Challenging the need for packet spraying in large-scale distributed training [EB/OL]. [2024-10-05]. https://arxiv.org/abs/2407.00550v1

[12] TOKUSASHI Y, DANG H T, PEDONE F, et al. The case for in-network computing on demand [C]//Proceedings of the Fourteenth EuroSys Conference 2019. ACM, 2019: 1-16. DOI: 10.1145/3302424.3303979

[13] 人工智能算力网络推进联盟 [EB/OL]. [2024-10-01]. https://c2net.openi.org.cn/

作 者 简 介

厉俊男，国防科技大学第六十三研究所助理研究员；研究方向为可编程网络处理器、低功耗嵌入式处理器；参与“863”计划、国家自然科学基金、武器装备预先研究等多项项目；发表论文10余篇，出版专著1部。



李韬，国防科技大学计算机学院网络空间安全系副研究员；研究方向为高性能网络芯片及系统；主持和参与“863”、重点研发、自然科学基金、武器装备预研等项目10余项，主持研制5款专用网络芯片；研究成果获4项科研成果奖；发表论文40余篇，出版专著2部，获授权专利20余项。

杨惠，国防科技大学计算机学院网络空间安全系副研究员；研究方向为高性能网络体系结构、网络处理器芯片；主持和承担芯片型谱、武器装备预先研究、重点研发、自然科学基金等国家及军队级项目10余项；发表论文30余篇，出版专著1部，获授权专利30余项。