专栏：面向大模型的网络技术

面向智算中心的新型以太网需求与关键技术

段晓东，李婕妤，程伟强，李晗，王瑞雪，王豪杰
（中国移动通信有限公司研究院，北京 100053）

摘　要：AI大模型正引领下一个十年的信息与通信技术（information and communications technology，ICT）产业发展热点。智算中心网络是支撑AI大模型分布式训练的通信底座，是决定AI集群效能的关键要素之一。AI大模型的数据量和参数量不断扩张，给智算中心网络带来了严峻的挑战，同时给关键网络技术进行代际性创新带来了机遇。在AI大模型训练和推理过程中，提供数据的高性能和高安全传输是AI业务对智算中心网络的两大核心需求。高效的负载均衡、拥塞控制技术和网络安全协议是其中的关键网络技术。为应对大规模AI业务带来的严峻挑战，提出全调度以太网（global scheduled Ethernet，GSE）作为对应的解决方案，并搭建真实的测试环境对GSE和RoCE（remote direct memory access over converged Ethernet）网络进行性能对比测试。测试结果证明，GSE相较RoCE网络显著改善了任务完成时间（job completion time，JCT）。

关键词：AI大模型分布式训练；全调度以太网；负载均衡；拥塞控制；网络安全协议
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2024171

Challenges and key technologies of new Ethernet for intelligent computing center

DUAN Xiaodong, LI Jieyu, CHENG Weiqiang, LI Han, WANG Ruixue, WANG Haojie
China Mobile Research Institute, Beijing 100053, China

Abstract: AI large model is leading the hot ICT(information and communications technology) industry in the next decade. Intelligent computing center network is a communication base to support the distributed training of AI large model, and it is one of the key factors to determine the efficiency of AI clusters. The data volume and the number of parameters of AI large model are expanding continuously, which brings the network of intelligent computing centers serious challenges, and also brings an opportunity for intergenerational innovation of key network technologies. In the process of AI large model training and inferencing, providing high performance and high security transmission of data are the two core requirements of AI business for intelligent computing network. Efficient load balancing, congestion control technologies and network security protocols are the key network technologies. To address the challenge brought by large-scale AI business, global scheduling ethernet (GSE) was proposed as a corresponding solution, and realistic test environment was built to compare the performance of GSE and RoCE. The test results show that GSE significantly improves JCT compared with RoCE network.

收稿日期：2024−04−01；修回日期：2024−06−13

Key words: large model AI distributed training, GSE, load balancing, congestion control, network security protocol

0　引言

近年来，ChatGPT、Sora 等生成式人工智能的迅猛发展获得广泛关注。从 Transformer 问世到 2023 年 ChatGPT 爆火，人们逐渐意识到随着模型参数规模的增加，模型的效果越来越好，且两者之间符合尺度定律。即当模型的参数规模超过数百亿后，AI 大模型的语言理解能力、逻辑推理能力以及问题分析能力迅速提升，这一巨大的潜力吸引业内领军企业竞相推出万亿、十万亿级参数量的大模型。为实现大规模训练集群高效的分布式计算，AI 大模型训练流程通常会包含数据并行、流水线并行及张量并行等多种并行计算模式，不同并行模式下均需要多个计算设备间进行集合通信操作，带来大量节点间通信损耗，单纯提升单个 GPU 的性能和 GPU 集群的节点规模无法直接获得集群算力的线性提升，节点间网络通信能力将成为制约 GPU 集群算力扩展的瓶颈。如何构建高吞吐、低时延的高性能智算中心网络，提升 GPU 有效计算时间占比（GPU 计算时间/整体训练时间），对于 AI 大模型分布式训练集群的效率提升至关重要。同时数据安全的重要性也随着 AI 大模型的发展日益凸显，在 AI 大模型的训练与推理过程中，大量数据资源通过网络传递，提供高效且安全的数据传输、防止敏感数据泄露是 AI 产业健康发展的基石。

智算中心网络当前主要有 IB（InfiniBand）和 RoCE（remote direct memory access over converged Ethernet）两种类型。其中，IB 是满足高性能交换设计的专用网络技术，方案成熟且性能优越，但产业封闭且成本高昂。RoCE 则采用标准以太网，产业开放，但传统以太网并非为高性能业务而设计，难以适应正在高速发展的 AI 大模型对智算中心网络的多方面需求。对以太网进行技术革新以满足 AI 业务的需求是智算网络领域的主流趋势，本文将重点从网络性能与网络安全两方面展开讨论 AI 业务对网络的需求，以及当前 RoCE 网络的现状与挑战，并给出全调度以太网（global scheduling Ethernet，GSE）对应的解决方案和 GSE 的初步测试结果。

1　智算中心新型以太网技术需求与挑战

1.1　AI 大模型分布式训练对智算中心网络的需求

（1）高性能

AI 大模型训练是一个具备同步性特征的周期迭代过程，基本过程如图 1 所示，GPU 间的通信占据整个分布式训练不可忽视的部分，随着分布式训练任务规模的增加，若智算中心网络不能提供足够的有效带宽，GPU 之间的通信时间可能成为瓶颈[1]，在某些情况下，GPU 在总训练时间中有高达 90% 的比例在等待网络完成数据的传输[2]，因此压缩大模型训练过程中的通信时间是提升 AI 集群计算效率的关键一环。同时，大模型训练存在同步性特征，即需多机多卡间完成该轮所有集合通信操作后才可进行训练的下一轮迭代，这就要求网络提供极低的长尾时延，避免出现木桶效应。另外，大模型训练流量还具备突发性的大带宽流特征，训练过程中通信五元组的总数量较少，但单流的带宽占用大，不管何种并行模式都对网络带宽提出 GB 级别的需求。AI 大模型流量总体呈现周期性突发、低熵、大带宽流等特点，要求网络提供高吞吐，且保障低长尾时延。

（2）高安全

随着 AI 大模型的发展，模型训练的环境逐渐由封闭性较好的专网数据中心环境演进到多租户的公有云网络，数据安全的重要性与挑战也日益

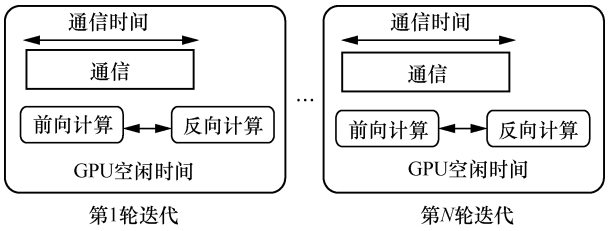

图1 AI大模型训练基本过程

凸显[3]。智算中心内的模型训练和推理过程涉及大量用户数据、训练后的模型参数也具备高价值，若数据安全保护不当，则存在参数及数据等敏感资产泄露的风险。通信链路、设备端口暴露，网络扩容升级、频繁运维、多租户等典型场景都需要加强安全防护。然而，保护数据安全的措施（如加/解密）会带来一定的性能损耗，AI场景又对网络性能有着极致的要求。如何在保证网络性能的前提下保障数据安全是AI业务对智算中心网络安全提出的核心诉求。

1.2 RoCE网络关键技术的现状与挑战

RoCEv2是IBTA于2014年发布的RoCE网络的第二代协议，是当前基于以太网的智算中心网络的事实标准。RoCEv2的传输层采用IB协议体系的传输层，网络层则用UDP/IP替换了第一代RoCE协议（RoCEv1）采用的IB网络层协议，以支持更大规模的3层互连能力，链路层和物理层使用以太网标准。RoCEv2的设计构建在传统以太网上，在面向AI业务规模扩展的趋势和对网络高性能、高安全的需求时，传统以太网与RoCE在性能、安全等方面已难以适应[4]。

（1）面向高吞吐、低时延的高性能需求，控制网络丢包的同时提升网络整体利用效率是需要解决的核心问题，负载均衡和拥塞控制是其中的关键技术，RoCEv2当前基于以太网和等价多路径路由（equal-cost multi-path routing，ECMP）的经典逐流负载均衡机制进行多路径负载分担，并通过以太网的优先级流控制（priority flow control，PFC）机制和基于显式拥塞通知（explicit congestion notification，ECN）算法的端到端数据中心量化拥塞通知（data center quantized congestion notification，DCQCN）机制来共同控制网络拥塞，第1.2.1节和第1.2.2节将分别面向负载均衡、拥塞控制两项关键技术来具体分析相关现状与挑战。

（2）面向数据安全的需求，如何保护数据传输过程的机密性、完整性是网络安全的主要问题，RoCE协议并未设计原生的端到端安全协议，缺乏端到端的数据传输保护，IPSec和MACSec等传统网络常用的安全协议开销较大，不适用智算场景，第1.2.3节将具体分析网络传输安全的现状与挑战。

1.2.1 负载均衡技术

当前数据中心的主流拓扑采用经典的Spine-Leaf胖树结构，在Leaf与Spine交换机之间提供多条等价的网络路径，负载均衡技术则是一类为提升网络整体利用率，基于一定策略与粒度将流量分散到多路径传输的技术。根据网络多路径转发的粒度，可将现有负载均衡技术分为基于流、基于子流（Flowlet）、基于报文3种类型。

（1）流级负载均衡

传统以太网主要采用流的粒度进行均衡，ECMP[5]是经典的等价多路径路由算法，以五元组为哈希因子，相同五元组的报文选择同一路径。但在低熵、大带宽流的AI大模型场景下，逐流多路径易造成哈希极化，容易出现某些“大流”集中选择同一路径，易造成少量路径流量超载，而其他路径出现空闲的状态，导致链路带宽利用率出现两极分化的局面。

（2）子流级负载均衡

Flowlet路由[6]是一种中等粒度的负载均衡方法，将一条流根据时间间隙切分为多个包含若干数据包的子流，当时间间隙超出设定阈值后，前后报文被划分为不同的子流，同一个子流内的报文沿同一条路径转发，具有一定的保序能力。通

算数据中心场景中的流量特征为大量基于事务应答模式的老鼠流混合少量大象流，较易找到流片段间隙，但智算中心主要由少量大象流组成，很难通过时间间隙来切分子流[7]。

（3）报文级负载均衡

报文级负载均衡也称数据包喷洒[8]，是理论上粒度最细的均衡方式，同一个流的不同数据包通过逐数据报文选择不同路径转发，将导致每条流的不同报文到达目的端时发生乱序。传统的远程直接数据存取（remote direct memory access，RDMA）协议采用Go-back-*N*机制实现丢包重传，依赖网络提供保序能力，报文乱序到达将导致吞吐显著下降。端侧即使采用选择性重传机制提供乱序接收能力，也需要消耗缓存资源实现乱序处理，当连接数超过数百，性能就明显下降[9]，难以支持更大规模的组网。

1.2.2　拥塞控制技术

区别于负载均衡技术旨在提升网络内部的链路利用率，消除负载不均导致的热点，拥塞控制技术则是控制进入网络的流量总量，防止过多的数据注入网络中造成拥塞，使设备缓存或链路容量不会过载。当前主流的拥塞管理方法从网络层次上划分，可分为二层的逐跳流控和端到端拥塞控制。RoCEv2结合以太网的链路级流控PFC机制和端到端DCQCN机制共同实现控制拥塞。

（1）链路级流控机制PFC

PFC技术是IEEE 802.1Qbb标准中定义的一种基于端口队列的链路级流控技术，通过在交换机上配置PFC水线，在本地入接口队列发生拥塞或者即将发生拥塞时，向上一级设备发送反压信号，实现对应端口队列停止发送流量，直至解除网络拥塞。但是由于PFC流控技术基于队列实现反压停流，无法精细化到流识别，容易产生队头阻塞[10]、拥塞传递[11]以及死锁[12]等问题，这些问题会对网络的可靠性和稳定性造成影响。业界进行了诸多解决PFC弊端的研究，但因交换机队列资源限制以及拥塞流识别精度受限等多方面原因，难以全面解决所有问题。当前主流的拥塞管理策略是尽量避免触发PFC机制，优先使用端到端拥塞控制机制对流量进行相对细粒度的控制。

（2）端到端拥塞控制技术

基于显式拥塞通知ECN的DCQCN[11]拥塞控制是RoCE网络经典的端到端拥塞控制机制，其基本技术理念是通过在设备上配置ECN门限值，在本设备出接口发生拥塞时，以一定概率启动ECN标记，最终由接收端根据ECN标志通知发送端进行降速。该技术方案具有较高公平性，但是需要沿路交换机支持ECN功能，且需要手动配置ECN参数，无法满足网络中流量模型复杂多变的场景，且控制环路较长，至少1个往返时延（round trip time，RTT）后才能调节发送端速率。同时，1 bit的ECN信号仅能定性地表示网络产生拥塞，无法定量地表示拥塞程度，端侧需要探测式调整发送速率，收敛速度慢，导致网络吞吐性能下降。

拥塞控制机制的关键要素之一在于端侧通过何种方式感知网络拥塞，除了ECN标记的方式，还可以将报文在网络的RTT作为拥塞信号，以RTT梯度调整传输速率，其典型技术方案是TIMELY技术[13]，该技术路线对网络设备无任何要求，但报文的往返时间易受队列的影响导致信息滞后而影响网络对拥塞的判断。基于带内遥测（inband network telemetry，INT）的拥塞控制是另一种实时网络预先感知的拥塞控制技术，通过直接获取网络信息，实时指导发送端调整速率，以HPCC技术[14]为典型代表，基于INT的拥塞控制技术能够实时感知网络状态并调整发送端速率，精确性较高，但需要沿路所有交换机支持INT技术，对网络设备要求较高。

1.2.3 网络传输加密技术

国际互联网工程任务组（IETF）为降低RDMA环境中的安全风险，提出在RDMA的网络层实现端到端的数据加密机制（RDMA-Sec）[15]。同时，业界已提出基于互联网安全协议（IPSec）的改进方案，以满足智算中心的安全需求[16]。此外，基于IEEE 802.1AE标准的MACSec（media access control security）可以为以太网设备之间提供数据链路层逐帧的安全加密通信，在园区办公场景得到较广泛应用[17]。然而RDMASec及MACSec应用于智算中心场景时仍存在如下问题。

（1）难以兼容全部存量设备。业界现有芯片硬化的RDMASec及MACSec方案，需要在端口物理层（PHY）芯片中进行比特流到包或帧的背靠背转换，将引入额外的实现复杂度与转换时延，也需要对设备硬件进行替换。

（2）引入封装开销，时延吞吐受限。现有安全机制均需逐帧独立加解密，需要逐帧插入标签，导致小包带宽利用率低，明显挤占业务带宽，影响AI业务算效。

（3）无法掩盖用户流量特征。现有安全机制均会暴露以太帧头部信息，无法完全掩盖报文长度、发包频率等流量特征，易被利用进行流量分析攻击[18]。同时无法保护PFC或pause帧等以太帧。

（4）管控机制复杂，难以部署。现有密钥管理机制安全复杂度高，需要消耗大量的CPU资源及网卡内存资源来维护节点间建立的安全会话，影响算效。

2 GSE架构与关键技术

针对RoCEv2网络在性能、安全两方面的问题，本节给出GSE中对应的解决方案。

（1）基于报文容器的负载均衡方案：首次同时考虑路由粒度和端侧性能开销问题，报文容器结合主动拥塞避免加全局调度技术，使智算网络在负载均衡做到接近最优的均匀分配，为降低业务长尾时延、提高有效带宽提供了保障。

（2）基于全局动态授权的主动拥塞避免方案：通过实现基于端到端的动态调度控制器和基于动态分配的流量控制方案，旨在基于网络实时变化，进行动态调整，为数据流目标设备端口动态分配资源，在有限的资源下，为更大规模的AI网络提供支持。

（3）物理层安全方案：通过在物理层构建数据加解密等安全传输能力，可以将加/解密时延降低至百纳秒级，且安全加密实例数低，不占用用户带宽，实现极低性能损伤的数据传输安全。

下文将先给出GSE的整体架构，并针对负载均衡、拥塞控制、数据传输安全3方面的关键创新技术展开介绍。

2.1 整体架构

GSE是具备无阻塞、高吞吐、低时延的新型以太网，服务于高性能计算，满足AI大模型部署及训推需求。GSE架构自上而下分为3层，分别为控制层、网络层和计算层，GSE网络架构如图2所示。

（1）控制层：集中式全调度操作系统（global scheduling operating system，GSOS）收集网络全局信息，实现基于全局信息编址、动态全局调度队列（dynamic global scheduling queue，DGSQ）统一授权、日常运维管理等功能。

（2）网络层：全调度网络处理节点（global scheduling processor，GSP）作为网络边缘处理节点，用于接入计算流量，流量上行时具备动态负载均衡能力，流量下行时具备流量排序能力，并对流量做全局调度。全调度交换网络（global scheduling fabric，GSF）作为网络核心交换节点，用于汇聚GSP节点，具备动态负载均衡能力以及反压信息发布能力。通过GSP和GSF的分工协作，构建具备全网流量有序调度、各链路间负载均衡、精细反压等技术融合的交换网络。

（3）计算层：计算层支持灵活的接入方案，图形处理器（graphics processing unit，GPU）服务器可以通过普通以太网卡接入新型智算中心网络，此时接入侧的网络交换设备承担GSP角色；也可以通过支持GSP功能的以太网卡接入新型智算中心网络，接入侧的网络交换设备承担GSF角色；对于GPU直出网口的情况，接入侧网络交换设备承担GSP角色。

与传统以太网基于流进行负载分担的机制不同，新型智算中心网络采用定长的报文容器（packet container，PKTC）进行报文转发及动态负载均衡，通过构建基于PKTC的DGSQ全调度机制、精细的反压机制和无感知自愈机制，实现微突发及故障场景下的精准控制，全面提升网络有效带宽和转发延迟稳定性。GSE网络端到端转发流程如图3所示。

（1）源端GSP设备从计算层收到报文后，通过转发表找到目的端GSP上的最终出口，并向最终出口申请DGSQ资源。

（2）源端GSP设备获得DGSQ授权后，将报文装入PKTC，并选择质量最优的链路将报文发送出去。归属于相同报文容器的所有报文，在网络中将被负载均衡到相同路径进行转发，以保证该PKTC内报文之间不乱序，降低出口GSP节点乱序重排的压力。

（3）当报文到达目的端GSP设备后，先进行

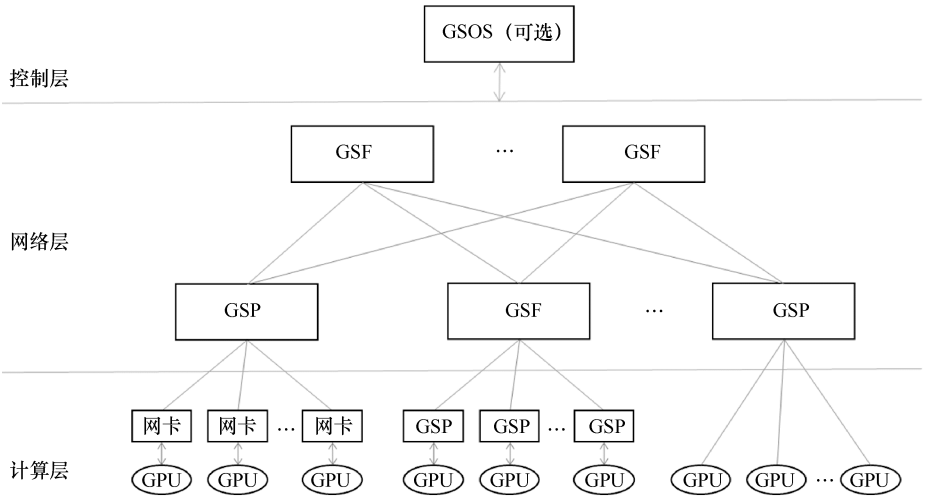

图2　GSE网络架构

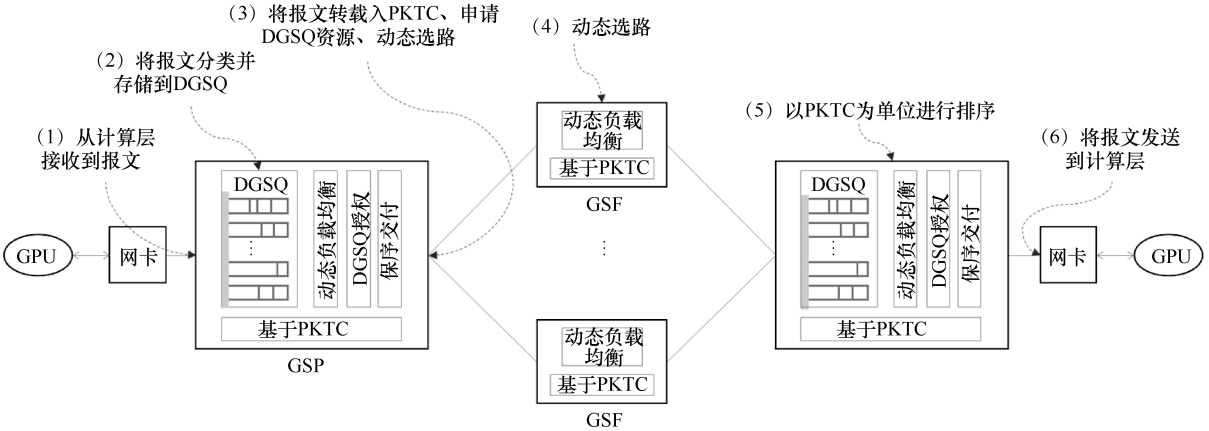

图3　GSE网络端到端转发流程

PKTC级别的排序，再通过转发表将报文存储到物理网卡的队列，最终通过端口调度将报文发送到计算层。

2.2 关键技术

2.2.1 基于PKTC的负载均衡

与传统以太网基于流进行负载分担的机制不同，GSE交换网络采用定长的PKTC进行报文转发及动态负载均衡。PKTC转发是区别于CELL转发的一种核心转发机制，该机制下以太网报文根据最终设备或者设备出端口被逻辑分配并组装成等长的虚拟PKTC，并以该容器为最小单元在交换网络中传输。

PKTC转发示意图如图4所示，PKTC的实现是逻辑虚拟的，GSE网络各节点均直接转发报文，无须缓存报文和构建实际容器。当一个报文进入源GSP节点时，GSP节点将记录其归属的报文容器编号、在该容器中占用的字节数等信息，当报文字节数超过虚拟报文容器设定长度时，将该报文调度并记录到下一个报文容器中，直接为报文封装GSE Header信息，基于GSE Header信息，将各个报文容器负载分担到各个上行链路，对于归属于相同报文容器内的所有报文均调度到同一条物理链路，使其沿着相同路径进行转发，以保证同属于一个报文容器的数据包保序传输。而不同报文容器经过负载分担后，形成多路径传输，由于不同路径的传输时延存在一定差异，当不同路径的流量到达目的GSP时存在报文容器之间的乱序，目的GSP设备收到报文之后，需要进行基于报文容器粒度的重排序处理。

2.2.2 基于DGSQ的拥塞控制

分布式高性能应用存在大量多对一通信（Incast）流量。如果这个现象是短暂的，在出口处可以通过一定的Buffer进行吸收。如果时间持续过长且多个入口的流量相加远大于出口的线速带宽，出口设备需启用反压机制避免丢包，而反压一旦出现，网络的转发性能就会大幅度下降，从而损害分布式应用的性能。GSE提出基于DGSQ的全局调度技术来解决上述的Incast问题，基于DGSQ调度流程如图5所示，在GSP上建立网络中所有设备出口的虚拟队列，用以实现本GSP节点到对应所有出端口的流量调度。本GSP节点的DGSQ调度带宽依赖授权请求和响应机制，由最终的设备出口、途经的设备统一进行全网端到端授权。由于中间节点的流量压力差异，GSP去往最终目的端口不再通过ECMP路径授权权重选择路径，而是需要基于授予的权重在不同的路径上

报文容器编号	报文1长度	报文2长度	…	报文*N*长度
1	2 KB	4 KB		1 KB
2	4 KB	4 KB		1 KB
4	3 KB	3 KB		3 KB

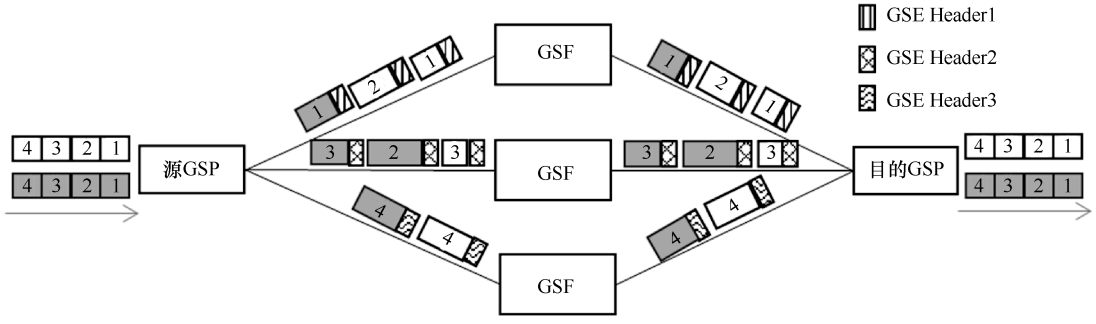

图4 PKTC转发示意图

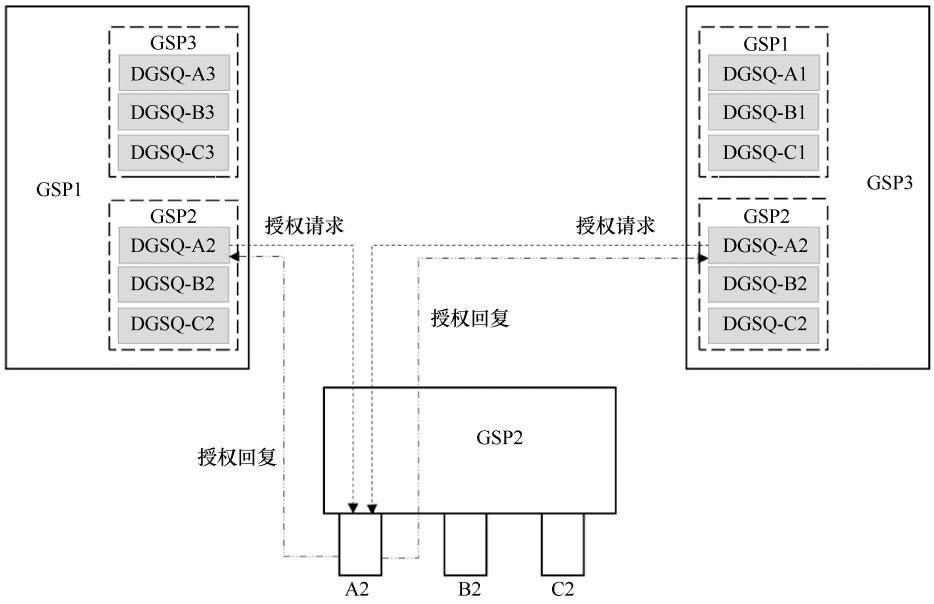

图5　基于DGSQ调度流程

进行流量调度。通过这种方式，可保证全网去任何一个端口的流量不但不会超过该端口的负载能力，也不会超出中间任一网络节点的转发能力，可降低网络中Incast流量产生的概率，减少全网内部反压机制触发。

2.2.3　物理层安全协议PHYSec

以太网物理层处于网络协议栈的更低层次，将安全加密与以太网物理层特性相融合来构建全新的以太网安全机制，可实现低时延、低开销、高吞吐、高安全的数据加密，满足智算中心场景对安全技术的要求。PHYSec技术体系架构主要包括3个层次：认证通道层、密钥管理层和数据加/解密层。PHYSec技术体系架构如图6所示。

（1）认证通道层：负责设备及光模块的身份认证与身份管理，确保相互通信的两端是合法的以太网设备。认证通过后，需要对认证通道进行保活。认证通道层的功能主要由平台业务软件实现。

（2）密钥管理层：负责运行过程中密钥的派生与管理、密钥定期更新分发以及密钥超期等异常状态处理。密钥分发完成后，还需要对使用该密钥的加密链路进行保活。密钥管理层的功能主要由平台业务软件实现。

（3）数据加/解密层：分为链路级加/解密与通道级加/解密。基于系统下发的密钥，分别通过加密引擎和解密引擎对信号进行加密和解密操作。数据加/解密层可以在光模块或PHY芯片实现。

PHYSec链路级加/解密方案和通道级加/解密方案在以太网物理层的不同层次实现，以200 Gbit/s/400 Gbit/s/800 Gbit/s为例，PHYSec的部署层级架构如图7所示。链路级加解密技术将多个对齐标识（alignment marker，AM）数据段复合形成复帧比特流，作为PHYSec的最小加/解密单元，优先在模块内部署，PHYSec链路级加/解密架构如图8所示。数据帧从发送端MAC层经过协调子层（reconciliation sublayer，RS）进入发送端PHY芯片后变成数据比特流，经过编码、扰码、AM插入、前向纠错（forward error correction，FEC）等物理层处理流程后进入光模块。发送端光模块的光数字信号处理器（optical digital signal processor，oDSP）先对收到的比特流进

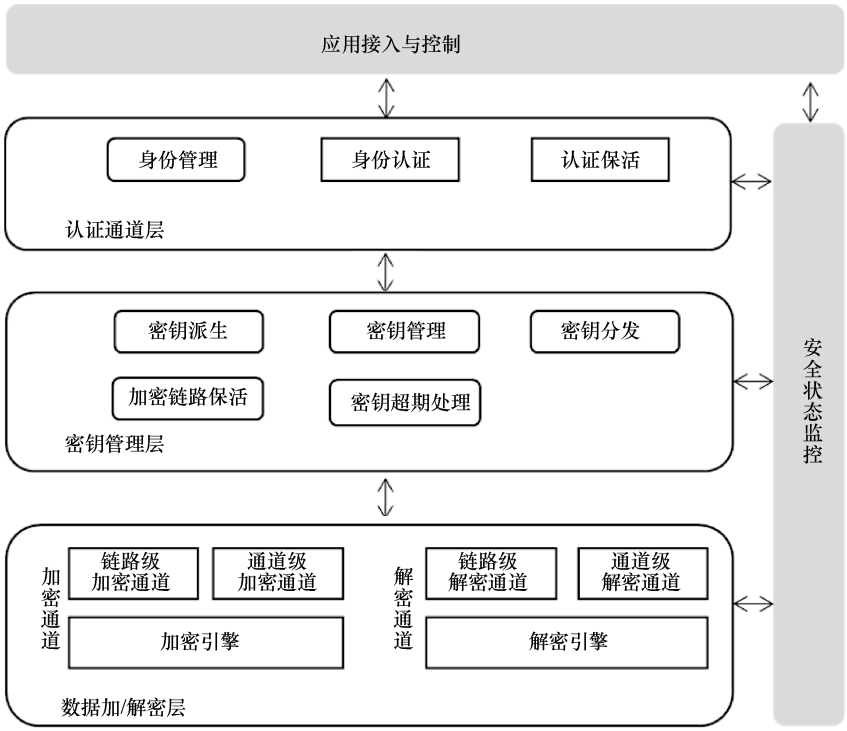

图6　PHYSec技术体系架构

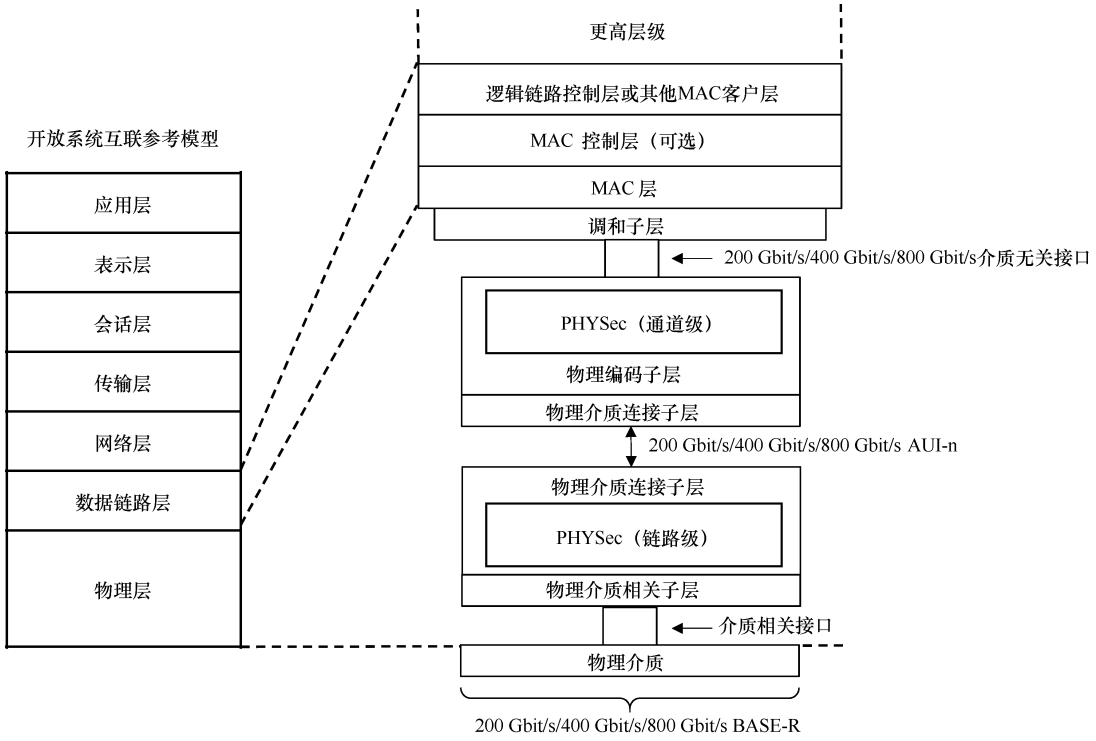

图7　PHYSec部署层级架构

行AM锁定，然后将AM锁定后的比特流组合为复帧进入oDSP内的加密模块进行加密变成密文比特流后发送到对端。接收端光模块收到密文后先进行AM锁定恢复复帧比特流，然后oDSP内

的解密模块将密文比特流解密恢复成明文比特流后送入接收端PHY芯片内进行后续处理。

通道级加/解密技术针对64B/66B码块流进行全部加密，增加D码块承载解密所需的参数，需要在PHY芯片内部署，PHYSec通道级加/解密架构如图9所示。数据帧从发送端MAC层经过RS进入发送端PHY芯片后被物理编码子层（physical coding sublayer，PCS）编码为64B/66B码块流，针对此码块流进行65B压缩，然后全部加密，密文封装到64B/66B码块流中类似为数据码块的64 bit区域，增加1个D码块承载解密所需的参数，最后首位添加S、T码块，构造一个完整的加密段，再实施PHY层其他处理。接收端收到加密段，从第一个D码块提取解密所需的参数，

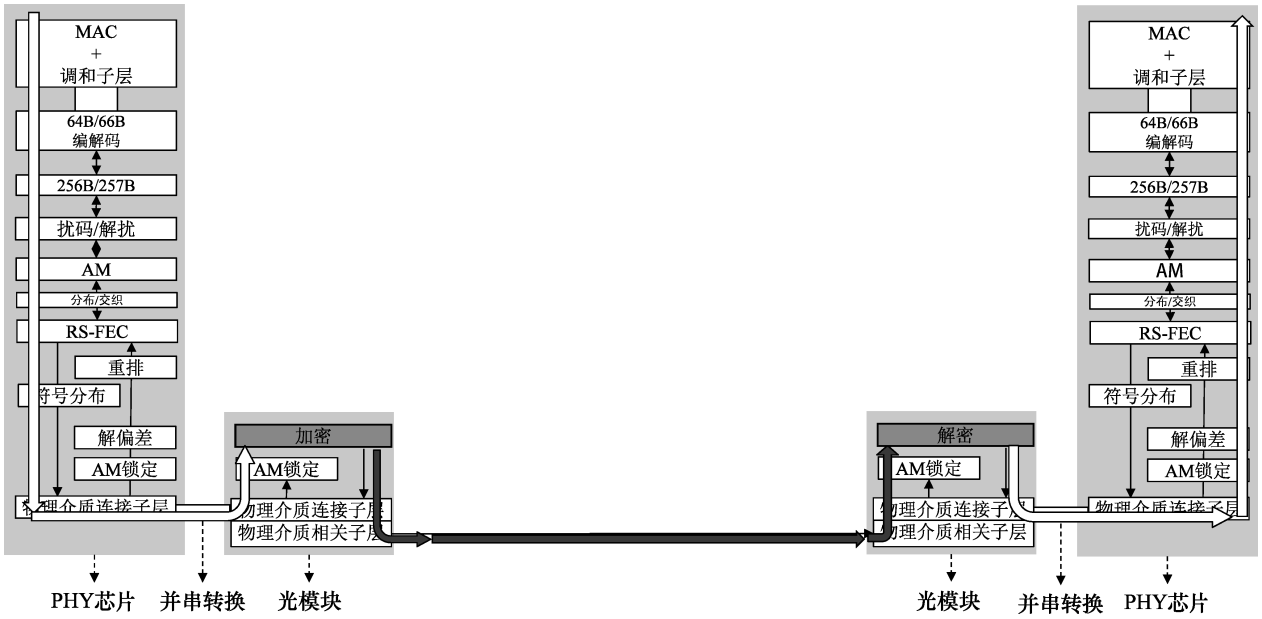

图8　PHYSec链路级加解密架构

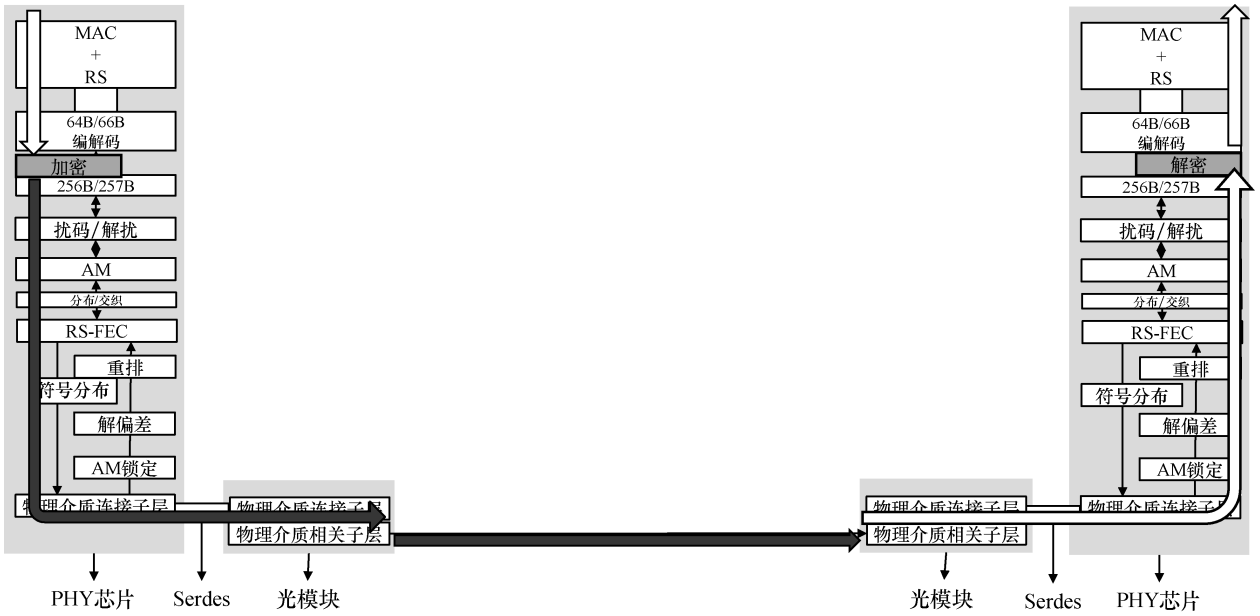

图9　PHYSec通道级加解密架构

对其余D码块内的用户信息实施解密，然后将其还原为65B码块流，再解压缩为64B/66B码块流。整个加密和解密流程都以IEEE 802.3标准规范的64B/66B码块流实施加密、解密，前向兼容IEEE 802.3 MAC/PHY标准功能。

3 GSE测试结果

3.1 网络拓扑

为对比AI大模型训练场景下GSE方案与传统RoCE方案下的网络通信性能，本文分别搭建GSE组网和传统RoCE组网两套环境进行实验室验证。GSE方案与传统RoCE方案网络拓扑如图10所示。图10中，GSE环境和传统RoCE环境的逻辑组网相同，但Leaf和Spine的设备类型不同。在GSE方案中，Leaf和Spine为GSE交换机；在传统RoCE方案中，Leaf和Spine则为传统RoCE交换机。

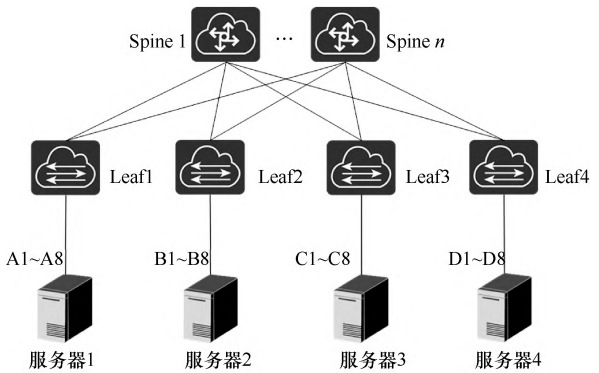

图10 GSE方案与传统RoCE方案网络拓扑

GSE组网与传统RoCE组网环境均采用二层Fat-Tree（胖树）组网，使用4台华为昇腾GPU服务器，每台服务器有8张GPU卡，每张网卡带宽均为100 GE。Leaf上行出400 GE口接入Spine设备，Leaf下行出100 GE口接入服务器网卡，同服务器的网卡接入同Leaf。

3.2 测试方案

本实验通过测试GSE方案和RoCE方案，对比在基于逐Container和基于逐流下的网络均衡能力。本实验的测试方案主要涉及加速比设计、通信流量设计、Container设计3部分内容。

（1）加速比设计：考虑当前被测的GSE设备尚未支持PFC功能，为了尽量避免网络拥塞和丢包问题，本实验分别在Leaf加速比为1.5:1和加速比为2:1的环境下进行测试，即Leaf节点通过n个100 Gbit/s×4的网口分别与n个Spine节点连接（n为加速比）。

（2）通信流量设计：由于机内通信流量会占用带宽，影响网络性能数据的结果分析，本实验GSE测试和传统RoCE测试采用不同服务器的GPU间通信，以保证机内GPU间不通信，机间GPU跨网络通信。本实验设定8个Job，即Job1为A1~D1为一组GPU通信，Job2为A2~D2为一组GPU通信，以此类推，Job8为A8~D8为一组GPU通信。

（3）Container设计：在GSE组网方案测试中，GSE Container的长度设置为16 KB，网卡侧每个数据包的大小为4 KB左右，即每一个Container可以容纳4个GSE数据包。

本实验采用AlltoAll和AllReduce两种典型通信模型，比较GSE和RoCE组网在不同通信模型下，随着报文长度（data size）的增加，两种组网的模型通信性能差异。

3.3 测试结果

本实验在AlltoAll和AllReduce两种通信模型下GSE和RoCE性能对比的测试结果如下。

3.3.1 AlltoAll通信模型下GSE和RoCE性能

本实验在AlltoAll通信模型下进行，在加速比分别为1.5和2.0时，分别统计报文大小为1～512 MB下的任务完成时间（job completion time，JCT）。AlltoAll模型下的JCT柱状图如图11所示。

通过测试结果发现，在AlltoAll通信模型下，随着报文长度的提升，GSE与传统RoCE组网的JCT都随着报文大小的增长呈指数级增长的趋

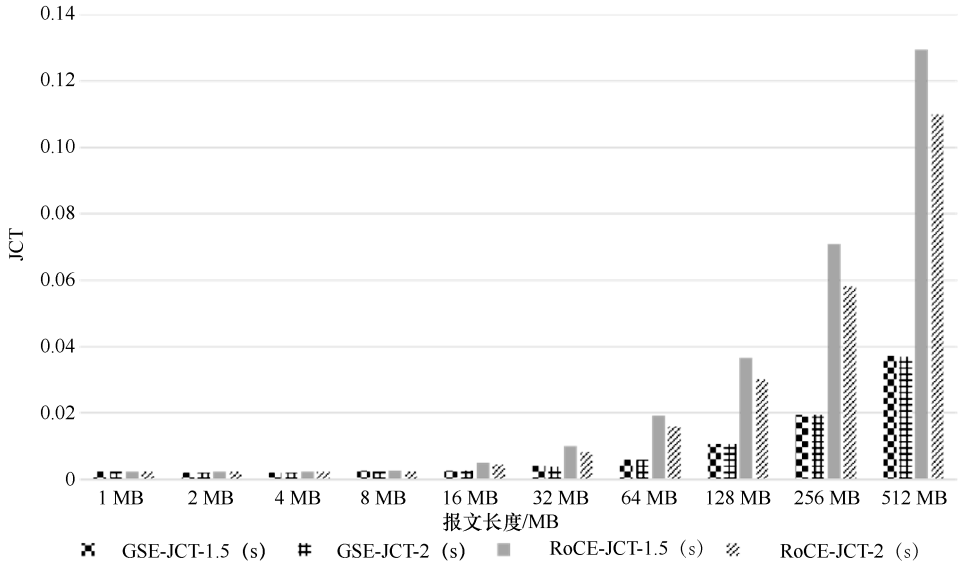

图 11　AlltoAll 模型下的 JCT 柱状图

势，但传统 RoCE 模式的 JCT 增长速度明显大于 GSE 模式的 JCT 增长速度。其中，在报文大小为 512 MB 时，GSE 模式的 JCT 约为传统 RoCE 模式的 1/3，JCT 提升 3.5 倍左右。

3.3.2　AllReduce 通信模型下 GSE 和 RoCE 性能

本实验在 AllReduce 通信模型下进行，在加速比分别为 1.5 和 2.0 时，分别统计报文长度为 1~512 MB 下的 JCT。AllReduce 模型下的 JCT 柱状图如图 12 所示。

通过测试结果发现，在 AllReduce 通信模型下，随着报文长度的提升，GSE 与传统 RoCE 组网的 JCT 都随着报文长度的增长呈指数级增长的趋势，但传统 RoCE 模式的 JCT 增长速度明显大于 GSE 模式的 JCT 增长速度，AllReduce 模型下的实验现象与 AlltoAll 模型的实验现象基本相同。其中，在报文长度为 512 MB 时，GSE 模式的 JCT 约为传统 RoCE 模式的 1/3，JCT 提升 2.6 倍左右。

结合两种通信模型下的实验结果，可知随着数据包长的增加，基于包均衡的负载均衡方案的

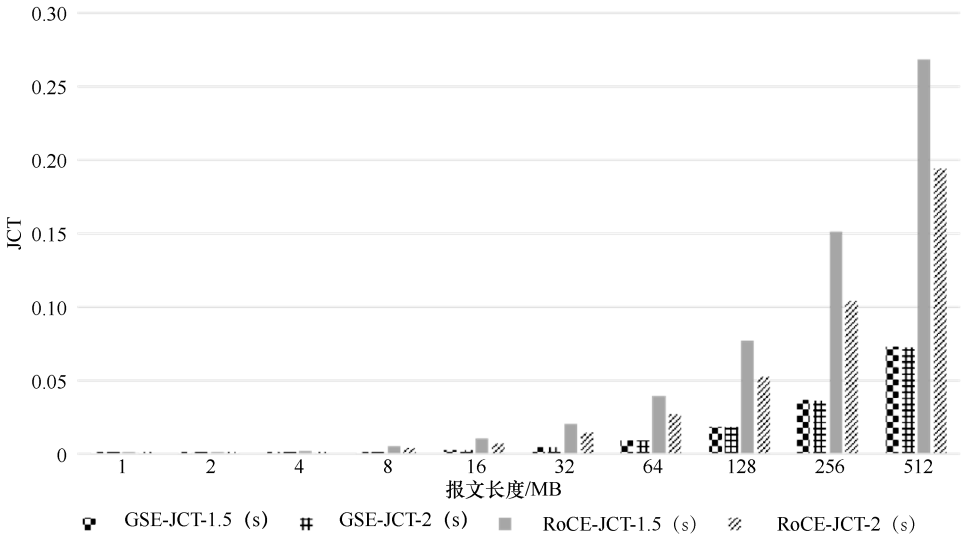

图 12　AllReduce 模型下的 JCT 柱状图

传输时延逐渐优于基于ECMP的负载均衡方案，且随着包长增加带来的网络负载加剧，ECMP的JCT的涨幅明显大于GSE。因此，进一步验证GSE在AI流量模式下的性能表现更优。

4 后续研究计划

以太网除了前文所论述的性能与安全两方面的问题与挑战，在可靠性方面也存在亟待解决的问题。AI业务是典型的带宽密集型负载，对应网络的接口速率将很快从200 Gbit/s/400 Gbit/s过渡到800 Gbit/s/1.6 Tbit/s的时代，与之对应的编码技术将从不归零码（NRZ）演进到4级脉冲幅度调制（PAM4）。然而，受香农定律约束，PAM4的误码率相比NRZ而言将高出几个数量级，可能达到惊人的1×10^{-4}量级[4]，如何解决接口速率提升引入的误码问题是GSE后续研究的重点问题之一。

5 结束语

AI大模型业务的迅猛发展给智算中心网络提出了新的性能需求和挑战，也带来了技术革新的机遇。本文重点面向AI业务在高性能和高安全两方面的需求，分析RoCEv2和传统以太网的技术现状和挑战。RoCEv2的负载均衡技术采用以太网传统的逐流负载均衡，难以适应AI的流量特征，包喷洒是负载均衡技术面向AI业务的未来趋势，但仍面临着报文乱序带来的严峻挑战，针对该现状，GSE提出基于报文容器的负载均衡技术，兼顾细粒度负载均衡和一定保序能力。然后针对RoCEv2基于PFC和DCQCN的拥塞控制技术所存在的不足， GSE提出基于DGSQ的全局调度技术，以更好地解决网络边缘的Incast拥塞问题。针对RoCEv2缺乏原生传输安全设计以及IPSec、MACSec等现有安全协议开销较大的问题，GSE提出新型的以太网物理层安全协议PhySec，兼顾高安全与高性能。最后，本文给出了GSE初步的测试结果以及后续的改进计划，初步证明了GSE相较RoCE网络显著降低了JCT。

参考文献：

[1] ZHANG Z, CHANG C K, LIN H B, et al. "Is network the bottleneck of distributed training?"[EB]. arXiv preprint arXiv, 2020: 2006.10103.

[2] LUO L, WEST P, KRISHNAMURTHY A, CEZE L, et al. PLink: discovering and exploiting datacenter network locality for efficient cloud-based distributed training[C]//Proceedings of 2020 MLSys. 2020.

[3] ROTHENBERGER B, TARANOV K, PERRIG A, et al. {ReDMArk}: Bypassing {RDMA} security mechanisms[C]//30th USENIX Security Symposium (USENIX Security 21). 2021: 4277-4292.

[4] HOEFLER T, ROWETH D, UNDERWOOD K, et al. Datacenter Ethernet and RDMA: issues at hyperscale[EB]. arXiv preprint arXiv, 2023: 2302.03337.

[5] HOPPS C. Analysis of an equal-cost multi-path algorithm[R]. Technical Report, 2000.

[6] QURESHI M A, CHENG Y C, YIN Q W,et al. PLB: congestion signals are simple and effective for network load balancing[C]//Proceedings of the ACM SIGCOMM 2022 Conference. New York: ACM Press, 2022: 207-218.

[7] SONG C H, KHOOI X Z, JOSHI R, et al. Network load balancing with in-network reordering support for RDMA[C]//Proceedings of the ACM SIGCOMM 2023 Conference. 2023: 816-831.

[8] DIXIT A, PRAKASH A, HU Y C, et al. On the impact of packet spraying in data center networks[C]//Proceedings of IEEE INFOCOM 2013. Piscataway: IEEE Press, 2013: 2130-2138.

[9] SCHARF M, KIESEL S. NXG03-5: Head-of-line blocking in TCP and SCTP: analysis and measurements[C]//Proceedings of the 49th IEEE Global Telecommunications Conference (GLOBECOM 2006). Piscataway: IEEE Press, 2006: 1-5.

[10] XUE J C, CHAUDHRY M U, VAMANAN B, et al. Dart: divide and specialize for fast response to congestion in RDMA-based datacenter networks[J]. IEEE/ACM Transactions on Networking, 2020, 28(1): 322-335.

[11] 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..

[12] HU S H, ZHU Y B, CHENG P, et al. Deadlocks in datacenter networks: why do they form, and how to avoid them[C]//Proceedings of the 15th ACM Workshop on Hot Topics in Networks. New York: ACM Press, 2016: 92-98.

[13] MITTAL R, THE L V, DUKKIPATI N, et al. TIMELY: RTT-based congestion control for the datacenter[C]//Proceedings of

SIGCOMM 2015. 2015: 537-550.
[14] LI Y, MIAO R, LIU H H, et al. HPCC: high precision congestion control[C]//Proceedings of the ACM Special Interest Group on Data Communication. NewYork: ACM Press, 2019: 44-58.
[15] PINKERTON J, DELEGANES E. RFC 5042:direct data placement protocol (DDP)/remote direct memory access protocol (RDMAP) security[R]. IEFT, 2007.
[16] Google. Google white paper: PSP architecture specification[R]. 2022.
[17] IEEE 802.1AE-2018: media access control (MAC) security[S]. 2018.
[18] HOPPS C. RFC 9347:aggregation and fragmentation mode for encapsulating security payload (ESP) and its use for IP traffic flow security (IP-TFS)[R]. IEFT, 2023.

[作者简介]

段晓东（1977– ），男，中国移动通信有限公司研究院副院长、“新世纪百千万人才工程”国家级人选、正高级工程师，主要研究方向为下一代互联网、算力网络、5G网络架构、6G网络架构、SDN/NFV等。

李婕妤（1994– ），女，博士，现就职于中国移动通信有限公司研究院，主要研究方向为数据中心网络。

程伟强（1980– ），男，中国移动通信有限公司研究院基础网络技术研究所副所长、正高级工程师，主要研究方向为下一代互联网、数据中心网络、传输网。

李晗（1975– ），男，博士，中国移动通信有限公司研究院基础网络技术研究所所长、正高级工程师，主要研究方向为光通信和承载。

王瑞雪（1990– ），女，现就职于中国移动通信有限公司研究院，主要研究方向为数据中心网络、SDN/NFV、算力网络等。

王豪杰（1992– ），男，博士，现就职于中国移动通信有限公司研究院，主要研究方向为高速以太网。