计算机研究与发展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(12): 2889−2894, 2025

观　点

领域定制方法促进网络与计算发展的思考

苏金树
（军事科学院　北京　100091）

Thinking on Networking and Computing Promoted by Domain-Specific Method

Su Jinshu
（*Academy of Military Science*, *Beijing* 100091）

Abstract　Intelligent computing characterized by LLMs (large language models) has become a focus point in the development of computer science. The exponential growth of parameters in LLMs poses significant challenges to both computing power and network infrastructure. In the era when Moore's Law is slowing down, domain-specific customization approaches have emerged. Based on author's over 30 years of experience in computer architecture research, this paper offers bold insights into the current development of computing architecture and presents preliminary practical results, aiming to spark further discussion and innovation.

Key words　domain-specific; computer architecture; network; computing node

摘　要　以大语言模型为特征的智能计算成为计算机学科的发展焦点. 大模型参数的指数级增长，给结点算力和互联网络同时带来极大挑战. 在摩尔定律放缓时代，领域定制方法脱颖而出. 本文根据30多年从事计算机系统结构研究的经验，对当前计算体系结构的发展，提出若干思考，报告初步实践结果，希望起到抛砖引玉的作用.

关键词　领域定制；体系结构；网络；计算结点

中图法分类号　TP393

DOI: 10.7544/issn1000-1239.202550787　　**CSTR:** 32373.14.issn1000-1239.202550787

随着人工智能技术的进步，智能计算得到强劲发展，并极大推动了以大模型为核心的人工智能技术的广泛应用. 在这次人工智能浪潮中，计算机体系结构的发展有哪些特点？网络技术有哪些特点？我们应该如何根据这些特点，更好地设计大规模智能计算系统？

1　Cluster和MPP成为大规模超算/智算主导结构，通用网络成为互连主流技术

首先，当前计算机系统发展面临的显著趋势是摩尔定律变缓、异构计算崛起和基于多核实现高性能. 图灵奖获得者Patterson在2018年获奖演讲时提出，在体系结构的历史调整与机遇中，未来10年将迎来体系结构的黄金时代[1]. 从实际发展来看，摩尔定律发展最顺利的阶段是从1985年到2005年，这20年的CPU处理器技术推进非常迅速，完全符合每18个月翻一番的规律. 随后主要采用多核技术、异构计算技术，单CPU的计算性能基本保持摩尔定律的发展速度，但是势头逐渐减弱.

计算机系统的发展呈现3个重要现象. 一是世界TOP500计算机系统的体系结构呈现高度的一致化，主要采用集群（Cluster）和大规模并行处理（massively parallel processing, MPP）结构，2000年前占

收稿日期：2025-10-31；**修回日期**：2025-11-10
基金项目：国家自然科学基金项目（62372462）
This work was supported by the National Natural Science Foundation of China (62372462).

主流的单指令多数据(single instruction multiple data, SIMD)、单处理器、向量计算机、SMP等逐步淡出;二是计算结点异构化,CPU+GPU几乎成为TOP500计算机系统的标配;三是通过与CPU、GPU主流厂商合作,系统制造商采用或者定制市场上性能最优的CPU、GPU,逐步放弃自主研制CPU的做法[2].

其次,大规模超算/智算系统的互连技术在向互联网技术靠拢.TOP500的互连从早期巨型机SMP,到专用互连或定制互连,再到大多数系统互连转向互联网技术,其中以太网技术占据半壁江山,高性能网络技术标准InfiniBand(IB)也成为高性能计算的主要互连技术标准[3-5].Mellanox和QLogic是2个主要的IB技术厂家.由于Intel在收购QLogic后放弃了IB的发展,IB技术成为Mellanox的独家产品.互联网技术在过去50年间,特别是近20多年来,呈现持续的"演化蝶变"与"筑基泛在"."演化蝶变"指的是从互联网延伸出物联网、移动互联网、工业互联网、数据中心网络、智能计算网络、智能体互联网等形态,核心技术内涵仍以互联网为基础."筑基泛在"是指互联网为云计算、大数据和人工智能的发展提供了推动和基础支撑作用,网络技术深入泛化到计算结点、存储结点、工业系统、航天器系统、大型装备等内部.过去多种多样的总线技术或者特殊专用互联技术,逐步转化为以太网为主的网络技术,使得互联网技术无处不在.

但是,互联网在"演化蝶变"与"筑基泛在"过程中,内部技术体系面临如何发展的问题.2000年前后,互联网领域的权威专家、互联网之父,Clark曾提出,网络空间发展的关键问题是如何定义未来的互联网,呼吁建立新的互联网体系结构.从互联网技术发展的现实看,3个典型案例可作为我们思考互联网发展的参考:一是谷歌针对数据中心网络研发了软件定义网络(software defined network, SDN),掀起了前几年网络技术发展的热潮;二是美国国防部为满足军事需求,开展了AGN/MNP等项目研究,将成果落实到美军网络部署中;三是针对大规模AI计算,英伟达创新了NVLink技术,首次直接从计算部件引出网络的实现结构,而不再是通过PCI-e等外部总线,使得网络带宽超过总线带宽.

最后,在本轮以智能计算为核心的网络超级计算阶段中,发达国家的部分做法值得我们学习和借鉴.美国政府层面采取2个重要宏观举措:一是通过放松监管刺激技术创新,以维持全球领先地位;二是采取出口管制、"长臂管辖"等手段,拉大与其他国家的技术差距.美国企业层面则有2个微观举措值得关注:一是在技术上分工协作,企业定位清晰,深耕细分领域,例如英伟达、AMD、Intel等企业主要负责硬件芯片研发,OpenAI、谷歌、Meta等公司则主导AI模型与基础支撑,应用领域的分工则更为细致,很少有企业单独提供全栈解决方案;二是投资上,企业之间相互投资,做到有来有往,你中有我,我中有你,利益共享,风险共担,共同迎接AI时代的机遇和风险.OpenAI早期投资主要来自微软、马斯克等,在关键时期微软投入100亿,助推了OpenAI的发展.OpenAI的成果为微软找到了继PC操作系统、云计算后的又一新的"第二曲线",助推微软成为少数能够从PC互联网,一直到移动互联网、社交互联网、智能互联网时代,持续保持高速发展的明星企业.与微软同时期的不少企业现在则逊色很多,甚至已经不存在了.OpenAI又投资很多创新企业.现在英伟达、谷歌、亚马逊、Intel等企业之间相互投资,形成一个严密的生态共同体.最新数据表明由OpenAI、软银、甲骨文主导,聚合英伟达、微软、AMD等技术伙伴,及政府、国际资本等,筹集5 000亿美元,联合推进全球最大AI基础设施Stargate"星际之门"网络建设.

2 典型系统发展启示

2.1 智能计算结点的领域定制呈现演进式与革命式齐头并进格局

智能计算结点的发展可分为2类,第一类是演进式,第二类是革命式.前者以英伟达为代表.英伟达服务器GPU体系结构演化了10多代.若干重要里程碑时的做法值得我们关注,可以认为是智能计算结点体系结构发展的缩影.

1)持续耕耘图形处理的领域定制.英伟达从做图形处理开始,就创新了图形计算架构,开辟CUDA新赛道,与Intel一直保持兼容x86结构的做法大相径庭.一直到本轮AI热潮之前,英伟达虽然不能算是非常成功的公司,但在图形处理定制体系结构上已经默默耕耘了20多年.例如在2017年的第五代芯片Volta体系结构集成RT Core,负责处理光线追踪加速,是典型的"按需定制"技术方案.

2)2016年首次引入深度学习引擎.英伟达在图形处理单元GPU的第四代芯片PASCAL首次引入深度学习引擎,将图形处理与深度学习融合.虽然英伟达未公开具体决策细节,但从时间线可以看出,其敏锐捕捉到了全球人工智能技术发展趋势,开始推进

深度学习与图形处理的融合. 根据黄仁勋在 2025 年 11 月在获得伊丽莎白女王工程奖圆桌论坛上的叙述，在 2010 年前后发现多伦多、斯坦福、纽约等大学的实验室研究人员都在用深度学习开放软件. 因为 2012 年辛顿提出深度学习网络 AlexNet，在李飞飞团队的 ImageNet 数据集上的图像识别竞赛中取得显著突破. AlphaGo 的发展引发了全世界对人工智能的关注. 上述 2 个方面凸显了人工智能发展的热潮序曲. 2017 年，在第五代芯片 Volta 中，在智能计算方面集成了 Tensor Core 执行张量计算，支持 FP32、INT32 运算，根据实际需求进一步优化计算部件.

3）2020 年融合智能计算与科学计算. 2018 年英伟达在第六代芯片 Turing 体系结构中，移除了 64 位浮点计算功能，转而增加 8 位、4 位整数计算支持；第七代芯片 Ampere（安培）中升级了计算核心，新增了 TF32, BF16 两种数据格式，同时重新支持 64 位浮点计算. 个人推测，这种看似矛盾的调整，实则根据一线实践的反馈，为了谋划智能计算与超级计算的融合，让芯片同时满足科学计算、智能计算需求.

4）持续优化智能计算部件. 2024 年 3 月，第九代体系结构 Blackwell 引入了 FP4、FP6 等专门针对智能计算的新结构. 为适应大模型巨量参数的需求，网络方面也进行了升级，具体将在网络部分展开阐述. 从整体发展来看，无论是结点的计算能力还是网络互联能力，每一代芯片都在持续提升，这种逐步迭代的模式我们将之归类为“演进式”发展.

第二类体系结构是与之相对应的“革命式”发展，典型代表是谷歌的张量处理单元（Tensor Processing Unit, TPU）. Patterson 认为采用领域定制体系结构的核心原因有 4 点：一是针对特定领域可大幅提升并行计算效率；二是能更高效地利用存储带宽；三是可消除不必要的计算精度冗余；四是通过定制化领域语言，降低软件开发难度[1]. 目前已发展到第六代 TPU 的“Ironwood”，单个 POP（这里的 POP 概念来自电信领域，类似后面的结点簇）可支持超过 9 000 个计算核心，浮点计算能力达 4 614 TFLOPS, 相比早期产品提升了 3 600 倍，带宽也增至 1.2 TBps. 谷歌认为，其 Iron Wood TPU 与英伟达 Blackwell GPU 的性能不相上下，但由于 TPU 采用定制化架构，生态建设上与英伟达尚有不少距离.

从实际应用案例来看，无论是“演进式”，还是“革命式”，领域定制体系结构具备显著优势. 谷歌在训练 AlphaGo 时，仅用 48 个 TPU 就达到了 1 200 个 GPU+176 个 CPU 的计算效果；2023 年 3 月英伟达在 GTC 大会上也提到，台积电公司将 4 万台服务器的计算任务，替换为 500 个 DGX H100 设备完成，使得设备数量大大减少，功耗也从 3 500 万瓦降至 500 万瓦，只有原来的 1/7，既减少了设备数量，也节省了大量电费.

2.2 结点互连向互联网技术体系靠拢，互连结构带宽按需定制

在最近 10 年的 TOP500 超算榜单中，以太网（Ethernet）与 IB（InfiniBand）逐步成为 TOP500 的主流互连①技术. 以太网最初是通用的网络技术，由于速率低，很少作为超级计算的主要连接技术. 但自 10 Gbps 的以太网问世后，局势很快出现变化，10 Gbps/100 Gbps/400 Gbps 以太网逐步成为超级计算的主要连接手段，尤其是 MPP 和集群计算主导 TOP500 体系结构后[3-5]. IB 早期是由 Mellanox 公司与 QLogic 公司共同主导的网络新技术，QLogic 公司被 Intel 收购，推进两年后，Intel 公司放弃了 IB 技术. IB 从此成为 Mellanox 独家主导的技术，经常被误认为是 Mellanox 的私有网络技术. 2020 年前后，英伟达以略高于微软和 Intel 的价格，成功收购 Mellanox. 英伟达高管曾公开表示，收购 Mellanox 是其在智能计算网络技术领域保持领先的关键原因，也是当前发展占优的重要因素. 除了以太网和 IB，Intel 的 Omni-Path 目前在 TOP500 中占据一定市场份额. TOP500 榜单的前 10 台超级计算机几乎都采用定制化或专用互联方案.

与超算相对应的是，大规模智能计算系统的网络技术，目前主要分为结点簇（有时也被称为超结点）内部互联（被称为 Scale-Up）和结点簇之间的网络互联（被称为 Scale-Out）. 结点簇在物理上可能是单机柜，也可能是多机柜. 例如英伟达的 NVL72、AMD 的 MI400 Helios 是单机柜的，英伟达 DGX-200 系统、华为 CloudMatrix384 系统都是多机柜的.

我们通过智能计算系统的 3 个实际案例，可以管窥智能计算系统网络互联的做法.

1）英伟达簇内簇间的“极高带宽”. 英伟达 DGX-

① 早期超级计算结点间的连接主要是链路层的技术，一般写为互连. 计算机网络，特别是互联网技术中，结点之间都是计算机系统的互联，一般写为“互联”. 随着超算、智算采用协议与互联网技术融合，两者的区分不明显. 本文中，除超级计算沿用习惯写为“互连”，其他一般采用“互联”.

200 系统包含 256 张 GPU 卡, 支持一级、二级互联的 NVLink 4 互联, 单 GPU 间互联带宽达 900 GBps, 网络性能极强; 2025 年推出的 MVL72 主要提升了集成密度, 单机柜包含 72 个 GPU 卡. 按计划 2025 年底将推出 NVLink 5, 2026 年将推出 NVLink 6. 英伟达打破了传统计算机体系结构的设计, 包括超算结点常规设计, 从 GPU 计算部件直接对外提供 NVLink 连接, 其带宽超越了 PC 总线带宽. 2014 年第三代芯片 Maxwell 第一次推出 NVLink 1 时, GPU 间的带宽达到 160 GBps, 约为当时最高总线 PCIe3x16 的 5 倍. 英伟达在簇内采用 NVLink 技术, 簇间采用 IB 技术, 一个实际的英伟达智能计算系统一般包含 4~5 个网络, 除结点簇内部 NVLink、结点簇间 IB 外, 还有控制管理的千兆或万兆以太网、计算存储之间的 200/400 G 以太网技术等. 为了实现最佳性能价格比, 英伟达会根据具体应用场景, 配置 400 Gbps 或 200 Gbps 的 IB 网络.

2)华为 CloudMatrix384 的“以量取胜”. 受工艺水平的限制, 国内单芯片能力普遍低于英伟达, 华为 910B/C 的单芯片性能与英伟达 Blackwell200 存在较大差距. 为此, 必须采用更多芯片, 构成一个结点簇, 华为公司称之为超结点. 在构建 CloudMatrix384 时, 华为采用与传统网络不同的设计思路, 不追求单机柜大容量, 也不刻意追求减少端口浪费. 早期互联网设备追求“尽量提高单交换机端口密度, 以减少端口浪费”. 由于计算结点的单接口带宽已达当前网络技术的上限, 交换机无法提供更高带宽的上联接口, 上联速率与下联速率处于持平状态. 因此华为采用“标准化、模块化”设计, 这也是当前智能计算互联的趋势.

3)优化通信算法“按需定制”互联带宽. 马斯克的 xAI 构建了第一个十万卡的智能计算系统, 其网络系统是根据具体场景与大模型算法进行设计和配置. xAI 的单结点服务器配置了 9 个 400 G 网络接口, 总带宽达到 3.6 TBps, 但仍远低于英伟达单结点 900 GBps 的水平. DeepSeek 的大模型训练系统也是由于采用优化算法, 减少了网络通信开销. 如果采用类似 DGX-100 结构, 万卡系统需要 1 320 台交换机, 而 DeepSeek 减少到 120 台[6]. 目前尚无公开文献阐述 DeepSeek 网络带宽与训练速度的关系.

2.3 国内结点芯片现状

目前国内在算力、存储带宽、互联网络等方面与国际领先水平仍有差距, 短期内主要依靠“集群化”(通过增加结点数量)来弥补单芯片性能不足. 从技术底层来看, 受限于芯片制造工艺, 国内与美国的技术差距较为明显. 尤其是在浮点 64 位计算能力上, 国内多数 AI 芯片对 64 位计算的支持不足, 市场上尚未出现能同时覆盖浮点 64 位到浮点 4 位计算的芯片. 目前没有企业公开解释不支持 64 位计算的原因. 作者猜测为了尽量提高 32 位智算指标, 节省芯片面积, 只能牺牲 64 位计算能力. 长此以往, 会导致芯片、支撑库、应用软件等可能会出现偏离智能超算一体化的目标.

国内企业在积极追赶国际先进指标, 不同芯片各有优势. 例如华为的 910B/C 智算性能已经具有相当竞争力, 寒武纪、昆仑芯、摩尔线程等也都在积极追赶. 国防科技大学研发的“迈创”芯片除了智算常规 FP32/INT32 计算精度外, 已具备 64 位浮点计算支持能力. 天数智芯推出“智铠”、天垓架构, 希望通过通用 GPU 架构支持多类 AI 模型, 同时兼容现有 CUDA 生态的技术路线取得发展.

3 智能领域定制网络 iDSN 的探索

如前所述, 领域定制已经成为事实上的计算机系统结构发展道路. 国内外有一些研究人员也从不同维度提出针对特定应用场景, 开展定制网络技术研究工作, 如针对云网融合、多模态硬件支撑、超级计算内部网络等[7-9]. 我们希望提出一个通用的智能领域定制网络, 目的是利用大模型等人工智能技术, 为网络的领域定制探索一条道路. 为此, 我们提出了“智能领域定制网络 iDSN”的总体概念, 核心目标是做好坚实的网络算法库, 实时感知网络态势, 根据领域场景需求, 高效敏捷生成场景需要的网络协议. iDSN 的架构如图 1 所示.

针对智能时代“定制计算”与“定制网络”的需求,

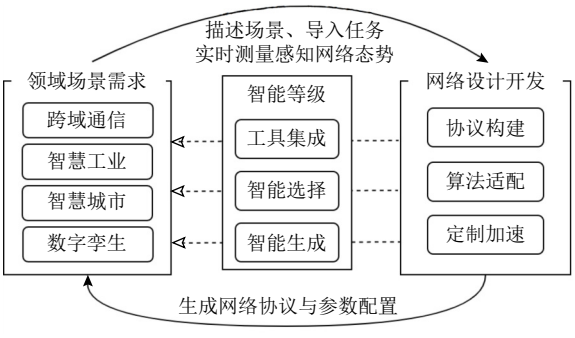


Fig. 1 Framework of intelligent domain-specific networking iDSN

图 1 智能领域定制网络 iDSN 总体框架

我们在设计 iDSN 时，提出 2 个基本原则.

原则一是“敏捷开放”的技术理念，解决传统网络设备“交互周期长、迭代速度慢”的问题. 在传统模式下，用户或需求方将需求提交给设备厂商，厂商再交由软件团队开发，软件开发至少需要 1 个季度，甚至 1 年，设备迭代周期则需数年时间，难以满足智能时代的发展之需.

原则二是“持续升级”. 传统网络面临“可扩展性不足”的问题，SDN 的引入解决了部分问题. 以企业级负载均衡（SLB）为例，早期需要在前端服务器处理请求并分发，通过 SDN 技术，可在网络设备中直接集成负载均衡功能，但是协议升级、核心功能动态更新等问题依然没有解决.

智能领域定制网络 iDSN 以网络领域知识库、算法库为支撑，根据应用领域的场景需求描述，实时测量感知网络态势，自动选择有关算法或者参数，最终实现算法参数的实时智能生成. iDSN 计划通过 3 个阶段，逐步达到设计目标. 第 1 阶段是工具与算法集成：集成网络领域的各类核心算法与技术，形成标准化工具库、知识库，为 iDSN 建立技术基础；第 2 阶段是参数与算法智能选择：根据具体场景，自动选择适配的参数或算法；第 3 阶段是协议智能生成：未来可根据场景需求，自动生成所需的协议与功能模块.

iDSN 的实现结构如图 2 所示. 最底层是平台支撑层，包括 CPU/GPU 处理器、操作系统、网络模拟环境、AI 大模型等网络系统的支撑要素；第 2 层是硬件加速层，按照通用架构构建的解决方案会面临速度瓶颈，加速层主要解决面向领域的性能问题；第 3 层是领域定制网络的核心基础层，目标是建立一个完善的网络算法库；第 4 层是智能控制层，为了适应具体应用场景，能够自动选择算法、协议，直至自动生成协议的控制要素. 最高层是应用场景层，随着应用不断丰富发展，推动其他层次技术的发展和丰富.

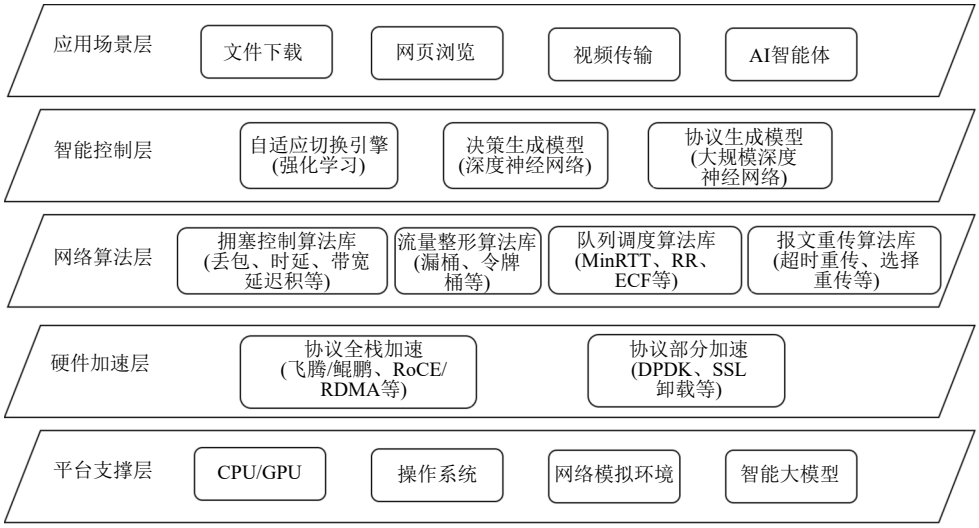


Fig. 2 Implementation structure of intelligent domain-specific networking iDSN

图 2 智能领域定制 iDSN 网络的实现结构

我们已开展了初步探索. 以多路径调度问题为出发点，研究开发了“多路径报文调度库 FlexPath”[10]；围绕智能传输实现了“网络协议智能配置”，可以动态选择，例如采用 iDSN 理念，构建多路径 QUIC，在部分场景替代 TCP、QUIC 等协议，并且在实际场景中应用部署，可根据场景需求自动选择算法或配置参数[11-14]. 上述算法已在实际场景得到初步部署[13]. 与国防科技大学、东北大学、中国科学院计算技术研究所、桂林电子科技大学等科研院校和有关头部企业开展合作，共同推进 iDSN 的发展.

4 总 结

智能计算的崛起与摩尔定律的放缓给计算机体系结构发展带来挑战，领域定制是应对挑战的方法之一. 本文试图分析当前智能计算发展的特点，从主流芯片和系统互联两个维度深入剖析具体系统的构成、发展历程，来探究发展背后的规律. 根据发展趋势，提出智能领域定制网络 iDSN 的总体发展思路、实现框架，并进行了初步的实践探索.

致谢:感谢致力推动 iDSN 发展,并共同做出贡献的国内外学术界同仁.

参 考 文 献

[1] Hennessy John L, Patterson David A. A new golden age for computer architecture[J]. Communications of the ACM, 2019, 62(2): 48−60

[2] Clark D D, Wroclawski J, Sollins K R, et al. Tussle in cyberspace: Defining tomorrow's Internet[J]. IEEE/ACM Transactions on Networking, 2005, 13(3): 462−475

[3] Su Jinshu, Zhao Baokang, Yi Dai, et al. Technology trends in large-scale high-efficiency network computing[J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(12): 1733−1746

[4] Top500.org. Top 500 Vendors list[EB/OL]. (2025-11-03)[2025-11-03]. https://top500.org/statistics/list/

[5] NVIDIA. World leader in artificial intelligence computing[EB/OL]. (2025-11-03)[2025-11-03]. https://www.nvidia.com/en-us/

[6] An Wei, Bi Xiao, Chen Guanting, et. al. Fire-Flyer AI-HPC: A cost-effective software-hardware co-design for deep learning[J]. ArXiv: 2408.14158v2

[7] Yan Fei. Research on cloud-network synergy oriented cloud network customization scheme[J], Designing Technology of Posts and Telecommunications, 2024 (10): 60−65 (in Chinese)
(闫飞. 面向云网融合的云网定制网方案研究[J]. 邮电设计技术, 2024(10): 60−65)

[8] Hu Yuxiang, Cui Zixi, Li Ziyong, et al. Construction technologies of polymorphic network environment based on codesign of domain-specific software/hardware[J]. Journal on Communications, 2022, 43(4): 3−13 (in Chinese)
(胡宇翔, 崔子熙, 李子勇, 等. 基于领域专用软硬件协同的多模态网络环境构造技术[J]. 通信学报, 2022, 43(4): 3−13)

[9] Raghunath MT, Ranade A. Customizing interconnection networks to suit packaging hierarchies[OL].[2025-11-03].https://www2.eecs.berkeley.edu/Pubs/TechRpts/1991/6036.html

[10] Chen Yitang, Han Biao, Li Sudan, et al. FlexPath: A customizable and QoS-aware multipath transmission protocol for heterogeneous networks[C]//Proc of 2025 IEEE/CIC Int Conf on Communications. Piscataway, NJ: IEEE, 2025: 1−6

[11] Liu Jiahao, Wang Yi, Chen Rongmao, et al. Universally composable subversion-resilient authenticated key exchange[C]//Proc of ASIACRYPT 2025(LNCS15543). Singapore: Springer, 2025. DOI: 10.1007/978-981-96-4731-6_11

[12] Liu Jiahao, Wang Yi, Tang Xincheng, et al. srTLS: Secure TLS Handshake on Corrupted Machines[J]. IEEE Transactions on Dependable & Secure Computing, 2025, 22(3): 3137−3154

[13] Ji Xiaolan, Han Biao, Xu Cao, et al. Adaptive QoS-aware multipath congestion control for live streaming[J]. Computer Networks, 2023, 220: 109470

[14] Song Congxi, Han Biao, Ji Xiaolan, et al. AI-driven multipath transmission: Empowering UAV-based live streaming[J]. IEEE Network, 2023, 38(2): 202−210



Su Jinshu, born in 1962. Professor and PhD supervisor. Fellow of CCF. His main research interests include computer network and cyberspace security.

苏金树, 1962 年生. 研究员, 博士生导师. CCF 会士.主要研究方向为计算机网络、网络空间安全等.