面向多算力中心协同的广域智算网络仿真架构设计

Wide-Area Intelligent Computing Network Simulation Architecture for Multi-Computing Centre Collaboration

边彦晖/BIAN Yanhui[1]，刘明远/LIU Mingyuan[2]，
虞红芳/YU Hongfang[1]
（1. 电子科技大学，中国 成都 611731；
2. 北京交通大学，中国 北京 100044）
(1. University of Electronic Science and Technology of China, Chengdu 611731, China;
2. Beijing Jiaotong University, Beijing 100044, China)

DOI：10.12142/ZTETJ.202502006
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.tn.20250427.1424.002.html
网络出版日期：2025-04-28
收稿日期：2025-03-22

摘要：针对智算仿真难以满足广域网时空动态性需求的情况，提出了一种面向多算力中心协同的广域智算网络仿真架构。该架构的主要创新点包括：基于属性图模型的拓扑抽象方法，实现异构算力间不规则连接建模和不稳定网络还原；基于流感知框架的广域通信模拟架构，提供高精度网络通信仿真；事件触发的多算力中心动态调度协议，通过逻辑时钟保障跨域操作因果一致性。本架构的提出弥补了广域多算力中心背景下仿真工具的缺失，为广域智算领域的相关研究人员提供高效、可靠的仿真支持。

关键词：多算力中心协同；广域环境；算网融合；仿真架构

Abstract: In response to the situation that intelligent computing simulation is difficult to meet the requirements of the spatio-temporal dynamics of the wide-area network, a wide-area intelligent computing network simulation architecture oriented to the collaboration of multiple computing power centre is proposed. The key innovations of this architecture comprise: 1) an attributed graph model-based topology abstraction method for modeling irregular connections among heterogeneous computing resources and restoring unstable networks; 2) a flow-aware framework-based wide-area communication simulation architecture enabling high-precision network communication emulation; 3) an event-triggered dynamic scheduling protocol for multi-computing centers that ensures cross-domain operation causal consistency through logical clocks. The proposal of this architecture makes up for the lack of simulation tools in the context of wide-area multiple computing power centers. It provides efficient and reliable simulation support for relevant researchers in the field of wide-area intelligent computing.

Keywords: multi-computing centre collaboration; wide-area environment; computing network convergence; simulation architecture

引用格式：边彦晖，刘明远，虞红芳 . 面向多算力中心协同的广域智算网络仿真架构设 [J]. 中兴通讯技术, 2025, 31(2): 39-46. DOI: 10.12142/ZTETJ.202502006
Citation: BIAN Y H, LIU M Y, YU H F. Wide-area intelligent computing network simulation architecture for multi-computing centre collaboration [J]. ZTE technology journal, 2025, 31(2): 39-46. DOI: 10.12142/ZTETJ.202502006

近年来，智能化浪潮席卷各行业，人工智能（AI）大模型训练需求呈爆发式增长。与传统的ResNet、视觉几何组（VGG）等模型相比，AI大模型的参数量实现指数级跃升，由此带来的算力需求也呈几何级数增长，给计算资源带来了前所未有的挑战[1]。在自然语言处理领域，算力需求的攀升尤为显著。以生成式预训练变换器（GPT）系列模型为例，其训练过程需调用约10 000张Nvidia A100计算卡，历经超过60 d方能完成训练。尽管DeepSeek通过创新系统架构优化策略，大幅降低训练所需算力，仅使用2 048张H100计算卡即可在54 d内完成训练任务，但这一量级的算力需求，对于当前大多数已建成的算力中心而言，依然难以满足。在此背景下，广域网络环境下的多算力中心协同训练技术凭借独特优势，成为业界焦点。该技术通过整合地理分布的分散算力资源，实现数据本地化处理，在保障训练效率的同时，有效解决隐私保护难题。其核心在于构建广域跨算力中心的动态资源池，突破单一算力节点的性能瓶颈；同

时，借助智能调度机制，对异构计算资源进行全局优化配置，显著提升跨地域、跨架构计算资源的使用效率，为AI大模型训练提供更高效、安全的解决方案。

然而，广域环境下的多算力中心协同训练面临时空动态性导致的双重挑战：1）空间维度：算力资源地理分布广泛，网络拓扑呈现不规则、异构化特征；2）时间维度：跨域链路状态时变显著，导致梯度同步等关键操作的不确定性激增。

上述时空动态性特征对训练仿真工具提出了双重要求：需要精准刻画地理分布带来的拓扑异构性，又需捕获网络状态的变化规律。然而，Astra-Sim[8-9]、SimAI[10]等主流单算力中心仿真工具在该复杂场景下存在显著缺陷：1）拓扑搭建局限性强：依赖规则拓扑的建模方法无法描述广域网络的不规则复合结构，导致跨域通信模拟失准；2）网络建模缺陷：现有方法将时延、带宽视为静态参数，未考虑丢包率与抖动的时空相关性；3）分布式时序混乱：基于全局时钟的假设与广域网络的异步特性相矛盾，造成分布式状态同步错位。

针对上述问题，为了还原真实广域环境下的分布式协同场景，本文提出了一种面向多算力中心协同的广域智算网络仿真架构。首先，该架构采用基于属性图模型的拓扑抽象方法与流感知框架的流量建模方法，实现对真实广域网络通信的高精度还原；其次，通过多中心间事件触发的动态调度协议，结合逻辑时钟实现跨域操作因果的一致性；最后，该架构凭借自动化模型转换框架与可视化交互界面等设计，兼具高易用性与扩展性，为用户提供端到端的便捷使用体验，为相关领域研究人员提供功能完备且操作简便的仿真工具。

1 面临的挑战

广域网络具有链路可保障带宽低、链路稳定性差的特点，这也是广域智算网络主要的瓶颈。与局域网相比，广域网因跨域、分布式等特点，其网络拓扑结构、性能参数等呈现高度复杂性。因此，现有针对局域网优化设计的智算仿真架构，难以在广域网环境下实现精准仿真效果，存在三大挑战：广域复杂拓扑与动态多维链路建模挑战、多中心控制面协同与同步挑战，以及系统易用性瓶颈挑战。如何应对这些挑战，以还原广域环境下的多算力中心的真实运行情况，是我们重点解决的问题。

1.1 广域复杂拓扑与动态多维链路建模挑战

作为智算仿真领域的开创之作，Astra-Sim将智算试验拆解为网络、计算和内存模块的设计思路，被SimAI[10]、vTrain[11]等后续工具广泛借鉴。

然而，Astra-Sim和SimAI在网络模块设计上存在显著不足，在广域场景下局限性凸显：

1）规则拓扑假设与广域复杂性存在矛盾。现有工具仅支持Ring、FullyConnected、Switch等规则拓扑的组合建模，其节点连接模式与真实广域网结构存在本质差异。这种结构性偏差在分布式训练中会导致梯度同步路径计算失准，进而造成端到端训练时间估算误差。

2）静态参数模型与广域网动态性存在矛盾。现有工具的网络通信模型采用公式（1）[9]所示的静态参数组合（LinkLatency为链路延迟，Hops为链路跳数，MessageSize为数据量，LinkBandwidth为链路带宽）。该模型未考虑时延抖动、丢包率等动态参数间的关联性，导致广域通信时间模拟存在误差。

$$\text{Time} = \left(\text{LinkLatency} \times \text{Hops}\right) + \frac{\text{MessageSize}}{\text{LinkBandwidth}} \quad (1)。$$

考虑到真实广域网的特点，必须从链路拓扑抽象、链路带宽、时延、时延抖动、流量分布等多个维度对广域通信链路进行还原。因此，设计智算网络仿真架构时面临的首要挑战在于：如何实现拓扑自由搭建，以还原广域环境下节点间不规则的逻辑连接关系；同时，如何从多个维度模拟网络链路，以最大程度逼近真实网络通信场景。

1.2 多中心控制面协同与同步挑战

在广域网环境下的多算力中心协同计算场景中，算力中心之间同时存在数据面和控制面的通信。数据面负责各算力中心间的计算结果同步，例如在“数据并行”模式下同步前向传播的损失值，以及在“流水线并行”模式下前向传播切片结果的传递等；控制面则承担计算、通信策略与时序同步任务，以实现高效协同计算。

然而，现有的面向算力中心网的智算试验工具在设计时并未充分考虑多算力中心架构的特殊性。其通信模型仅关注数据面行为，却忽略了通信算法切换、路由策略调整等控制面操作；此外，诸如Astra-Sim等模拟器均采用全局时钟假设，该假设在广域网的高延迟环境下存在时间同步错误。因此，如何在仿真过程中模拟广域网环境下多算力中心间的控制面，体现控制面的操作逻辑并提供相应的控制行为模拟能力，成为设计多算力中心协同的仿真架构面临的第二大挑战。

1.3 系统易用性瓶颈挑战

考虑到仿真系统的用户需求多元化，提升系统的易用性也是设计目标之一。当前智算仿真工具存在严重的端到端流

程割裂问题。以Astra-Sim为例，其输入流程高度依赖第三方工具链：用户需通过PyTorch Profile工具和Chakra工具[13]预生成执行轨迹文件（Trace模式），或借助Scale-Sim工具[14]重构模型代码以生成硬件参数配置文件（Txt模式）。这种碎片化的工作流程与智算范式对灵活性的需求严重脱节。

仿真系统的易用性瓶颈还显著体现在交互设计层面。用户不仅需跨多个工具完成预处理工作，还面临可视化支持不足的难题：现有仿真系统缺乏动态链路参数配置、时序依赖关系呈现等关键交互功能，导致用户在数据预处理流程上耗费大量时间资源。因此，设计多算力中心协同的仿真架构面临的第三大挑战在于：如何构建无需第三方依赖、支持动态模型适配且符合认知效率的端到端仿真框架。

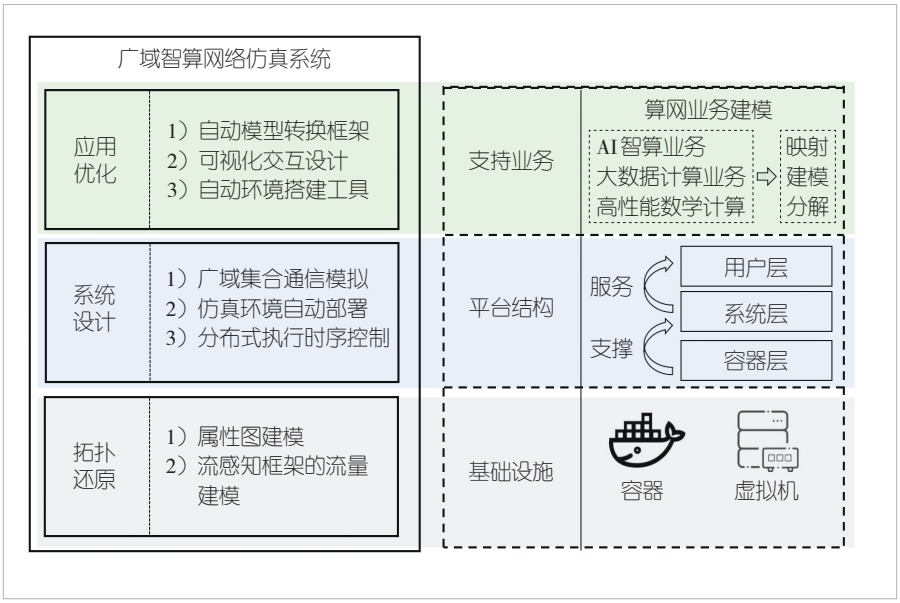

图1　广域智算网络仿真架构

2 方案设计

多算力中心协同的广域智算网络呈现网络链路不稳定、网络拓扑不规则、多算力中心调度复杂的特征。为了真实还原这些特点，本仿真架构采用自下而上的设计思路，按照“拓扑还原、系统设计、应用优化”3个层次展开架构设计，具体如图1所示。

首先，基于虚拟化技术开展图属性建模和流量建模，对广域网拓扑特性和流量特性进行高保真还原，赋予仿真系统符合广域网特性的网络通信模拟能力；其次，构建“容器层、系统层、用户层”的3层架构，其中“系统层”作为中间层承上启下，向上解析用户层输入范式，向下调用容器层服务，实现算力中心间通信调度、时序同步等多种控制面功能，为仿真架构提供完整的算力仿真和网络模拟能力；最后，为提升仿真系统易用性，开发自动化模型转换框架与仿真环境自动搭建工具，满足多元化用户需求。

2.1 广域网络时空动态性联合建模机制

鉴于广域网络的复杂性和不稳定性，仿真架构从拓扑搭建和流量模拟两个维度构建对广域网通信的高保真还原能力。首先，基于属性图建模技术，仿真架构实现对广域环境下不规则拓扑结构的还原；其次，通过对流感知框架下的流发生器进行结构性优化，使其能够从传输层协议、带宽、数据包特征等多个维度动态生成流量，既模拟广域网可用资源的时变特性，又还原智算业务中算力中心间集合通信流量的动态变化特征。该小节公式中所涉及的参数及其含义见表1。

2.1.1 属性图建模

传统仿真工具（如Astra-Sim）因割裂处理拓扑建模与链路参数，导致广域环境下的拓扑与链路关联失真。为此，本架构提出基于属性图模型的网络抽象方法：将每个算力节点建模为图的顶点，顶点属性包含节点向量 $\boldsymbol{H}=(M_{\text{mem}}, IP, N_{\text{num}})$，分别标识节点内存资源、IP地址、中央处理器（CPU）核数量；节点间链路建模为图边，边属性定义为参数张量 $\boldsymbol{L}_t=\left[B_t, \tau_t, \sigma_t, \lambda_t, D_t\right]$，分别标识链路带宽、链路时延、链路时延抖动率、链路丢包率、队列大小。结合团队自研平台Klonet[16]的自动编排技术，对抽象的属性图模型进行实体化重构，将属性图的顶点构建为容器，属性图的边

表1　属性图建模和流量建模相关的参数对照表

参数	含义	参数	含义
M_{mem}	内存资源	D_t	队列大小
IP	IP地址	$\mathcal{F}$	流量模型
N_{num}	CPU核数量	S	源地址
B_t	链路带宽	D	目的地址
τ_t	链路时延	P	传输协议
σ_t	链路时延抖动	$\boldsymbol{C}$	流量特征矩阵
λ_t	链路丢包率	$\mathcal{G}$	流依赖关系

CPU：中央处理器　IP：互联网协议

构建为TC-Qdisc控制的链路，其映射关系如图2所示。

通过属性图模型与流量控制（TC）规则的深度绑定，本架构实现了拓扑结构与链路参数的多维度解耦与协同控制。在拓扑构建阶段，用户可任意定义顶点连接关系，Klonet编排引擎将根据顶点属性选择合适的容器镜像，并基于边属性张量$\boldsymbol{L}_t$动态生成TC配置脚本。该方案保障了拓扑自由度与链路参数扩展的独立性：用户能够通过增删顶点或边灵活调整拓扑规模，彻底解决Astra-Sim等工具因硬编码拓扑模板导致的广域适应性缺陷；同时，每条链路凭借独立维护的五维参数$\boldsymbol{L}_t$进行模拟，确保仿真链路对广域真实链路的高保真还原。

2.1.2 流感知框架的流量建模

为实现对广域通信的真实模拟，除了设计更加符合广域网络真实情况的链路外，还需生成与之匹配的流量模式。本仿真架构针对广域网背景设计了流感知框架，并在其背景下将广域网流量抽象为五元组，即：$\mathcal{F}=\langle S,D,P,\boldsymbol{C},\mathcal{G}\rangle$。其中，$S/D$为源/目的节点的网络属性，从顶点向量$\boldsymbol{H}$中获取；$P$为传输协议且具备拓展性，支持自定义传输协议，即$P\in${TCP, UDP,Custom}；$\boldsymbol{C}$为流量特征矩阵，包含包大小、包间隔、流量分布等流量特征描述，$\mathcal{G}$为流间依赖关系图，用于描述集合通信操作。基于该五元组描述，本仿真架构设计了适配广域网场景的专用流发生器。该流发生器采用多线程分离与实时监控技术，提高系统的并发能力，并为用户提供细粒度的流量实时监控；同时，支持包大小、包间隔和流量分布等多维空间设计，实现对广域流量的真实还原。

通过属性图建模与流感知框架的深度协同，仿真架构实现了广域网络时空动态性的高保真还原：属性图建模以顶点-边结构解耦拓扑与链路参数，灵活构建不规则拓扑并精准调控带宽、时延、抖动等五维链路参数；流感知框架则基于五元组$\langle S,D,P,\boldsymbol{C},\mathcal{G}\rangle$流量模型，还原广域网流量时变性。二者通过拓扑-流量联合仿真机制，共同为仿真过程提供了对广域网背景下的通信流量的高保真还原。

2.2 事件驱动的多算力中心系统层调度

为了应对广域网环境下算力中心间数据同步、时序同步等多种挑战，本仿真架构自下而上设计了“容器层的系统层”“系统层的系统层”“用户层的系统层”3层结构。其中，顶层用户层接收并解析来自用户输入，将其转化为系统层规定的输入范式；底层容器层构建仿真过程的底层实体，接收并执行来自系统层的指令；而系统层作为两者之间的过渡层，在不同的任务中承担多个角色。

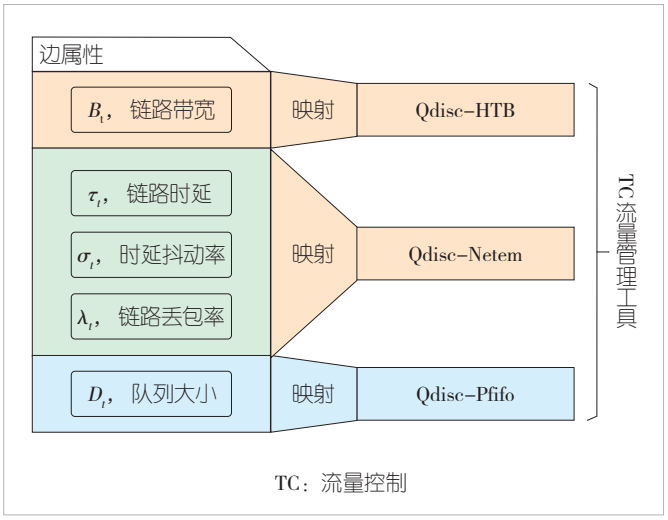

图2　边属性与流量控制工具的映射关系

1）在集合通信模拟任务中，系统层将来自用户层的集合通信算法映射为点对点通信，调用容器层的通信接口进行模拟。

2）针对仿真环境自动化部署任务，系统层接收来自用户层的仿真环境部署任务，将其重构为容器层搭建拓扑所需要的范式，调用容器层的拓扑接口进行模拟。

3）面对分布式执行时序控制任务，系统层采用逻辑时钟对分布式任务执行进行时序控制，确保各节点执行时序的正确性。

2.2.1 广域集合通信操作模拟机制

系统层基于容器层构建的物理拓扑感知能力，通过动态调用容器层的接口，实现跨算力中心集合通信行为的模拟。

1）集合通信分解映射：采用拓扑感知的通信图分解算法，将集合通信操作转化为有向无环图。图中节点表征通信任务，边表示时序依赖关系，以此确保通信仿真与真实场景的时序一致性。

2）点对点通信建模：基于五元组流量特征，系统层调用容器层流量发生器动态生成数据流，实时采集吞吐量、时延、丢包率等指标，用于后续仿真分析。

3）事件驱动的控制面模拟：针对集合通信操作中的控制面行为，如集合通信算法切换、路由策略切换等，系统层借助预设事件触发器，动态调整通信模式与资源分配策略，实现数据面与控制面的联合仿真。

2.2.2 声明式仿真环境自动部署框架

系统层提出基于拓扑语义的声明式部署框架，其核心为两个抽象阶段和一个执行阶段：

1）拓扑抽象阶段：将用户输入的顶点–边结构映射为属性图 $\mathcal{G}=(V,E,\varphi_v,\varphi_e)$，其中顶点属性 φ_v 对编码算力资源（CPU/GPU 配额）进行编码，边属性 φ_e 定义多维度链路参数，涵盖带宽、时延抖动、丢包率等。

2）资源抽象阶段：构建多维度资源联合抽象模型，通过归一化处理，将异构硬件参数映射为统一虚拟资源单元，以支持跨架构资源混合部署场景，如张量处理单元（TPU）和 GPU 集群的联合仿真。

3）搭建执行阶段：系统层根据阶段 1 和阶段 2 的抽象结果，借助容器层提供的仿真环境部署框架，完成部署搭建。

2.2.3 分布式执行时序控制模型

为保障多算力中心操作的时序正确性，系统层引入分布式执行时序控制模型，其具体功能如下：

1）基于逻辑时钟的因果排序：鉴于广域网络环境下的高延迟与强不稳定性，采用全局时钟进行多算力中心控制面调度易产生较大误差。本仿真架构基于 Lamport 逻辑时钟，构建因果排序机制，以逻辑计数代替物理时间，为系统内所有事件分配具备因果确定性的时间戳，有效解决分布式环境中因缺乏全局时钟引发的时间顺序混乱问题。

2）乱序消息重排序处理：针对网络抖动导致的消息乱序问题，通过引入暂态因果缓冲机制，按逻辑时间戳对迟到消息重新调度，确保处理顺序严格遵循事件因果关系。

3）关键操作的原子性保障：对于梯度同步、参数更新等关键操作，系统层通过逻辑时钟确保其原子性。通过逻辑时钟的有序递增和消息传递中的时钟更新策略，避免关键操作不受其他操作干扰，从而保证训练结果的正确性和一致性。

如图 3 所示，优化后的系统层架构通过三层协同机制，实现以下核心功能：

1）实现广域网集合通信操作中数据面与控制面的联合仿真，模拟真实多中心调控机制。

2）具备拓扑无关的自动化部署能力，支持分钟级千节点试验环境构建。

3）确保多算力中心操作的强时序一致性。

2.3 仿真系统的高易用性设计方案

为满足仿真系统的用户的多样化需求，本系统通过自动化模型转换框架、可视化图形界面和自动试验环境搭建工具等多个设计提升仿真系统的易用性。具体而言，自动化模型转换框架实现用户输入模型向算力中心模拟所需输入格式的自动转化，为用户提供端到端的试验能力；可视化交互设计

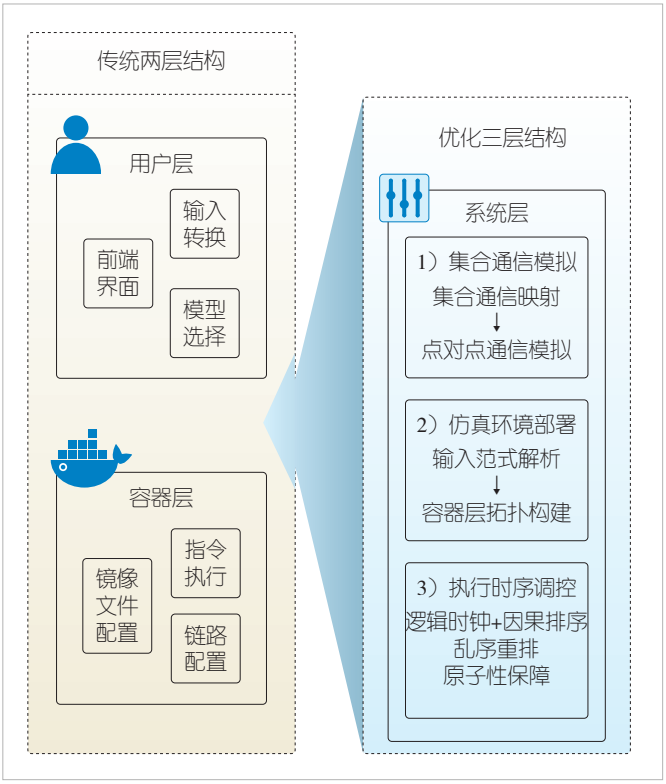

图 3　优化后的三层架构

助力研究人员直观、高效地完成试验环境搭建；自动试验环境搭建工具则显著降低超大规模试验场景的搭建难度，大幅减少用户工作量。

1）自动化模型转换框架

为消除用户对第三方预处理工具的依赖，我们提出了提出分层模型转换架构。该架构支持直接解析 PyTorch、TensorFlow 等框架的模型文件，并通过内部集成的层尺寸转换器、Scale–Sim 转换器等多层转化模块，将模型文件直接输出为算力中心内部模拟训练需要的输入格式，从而为用户提供从模型文件到模拟结果的端到端全周期仿真过程。该自动化模型转换框架如图 4 所示。

2）可视化交互设计

基于团队自研的 Klonet 网络仿真系统[16]，我们设计了多算力中心智算仿真系统的可视化界面，具体如图 5 所示。用户借助拖拽、点击等交互操作，即可直观完成试验环境搭建，该方式在小规模试验场景中展现出良好适用性。

3）自动环境搭建工具

针对图形化界面难以胜任的大规模复杂试验场景搭建难题，本仿真架构配备自动搭建工具。该工具以脚本形式运行，可依据用户需求自动化构建大规模场景，有效规避了单一可视化图形界面在处理复杂大规模场景时操作繁琐的弊

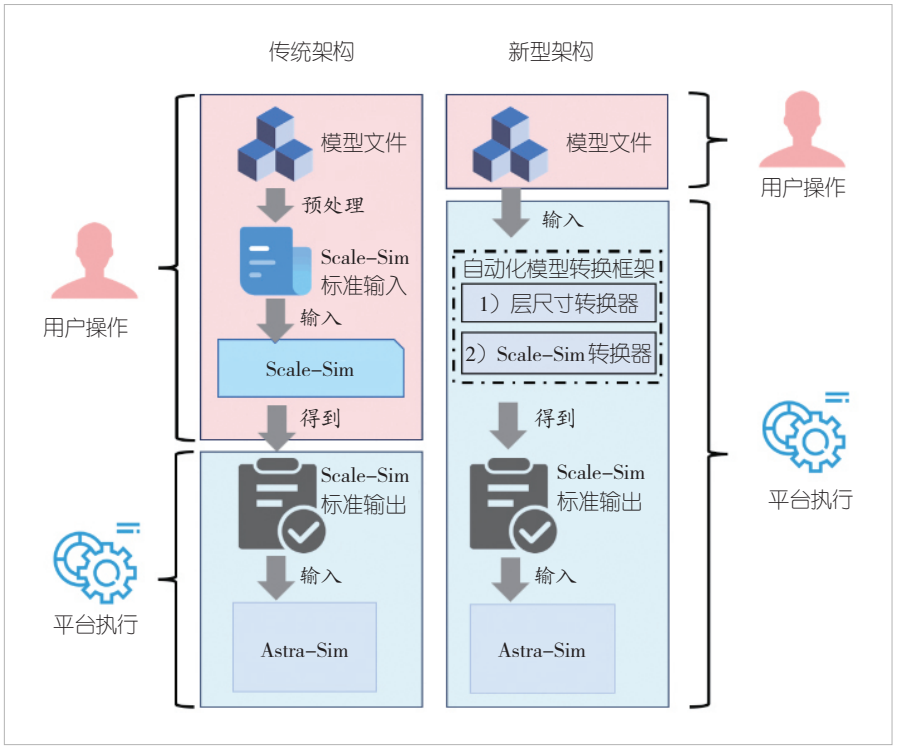

图4　自动化模型转换架构

图5　可视化图形界面

端，为用户提供了多样化的场景搭建解决方案。

3 仿真实验

本文中，我们以广域环境下大模型多算力中心分布式训练为研究对象，开展自由拓扑搭建、试验环境配置、算力中心协同调度及广域环境多算力中心通信模拟等试验。上述试验有效验证了仿真架构在不规则拓扑构建、系统层协同调度以及广域网络高保真模拟等方面的性能。

在智算试验环节，用户可从并行方式、网络拓扑、硬件配置、通信算法等维度进行参数设置，进而获取不同配置下的模拟训练结果。本文以 Llama-3-8B[17]大语言模型的数据并行训练为例，系统演示仿真架构的完整工作流程。

3.1 输入配置

在准备阶段，用户需要准备好本次试验的配置信息，具体包括：模型文件（Llama-3-8B）、数据集（WuDao[18]）、并行方式（数据并行）、广域网络拓扑信息（此次示例中以星型结构为例）、广域集合通信算法选择（Ring-AllReduce）、算力中心内部配置信息。

用户需按照规定的范式将上述信息填入配置文件，具体如图 6 所示。完成配置文件后，用户可在前端界面上传该文件。

3.2 环境配置

仿真架构的系统层在接收到用户层的配置信息后，对配置文件进行解析，并依次执行以下操作：

1）确定并行模式为数据并行。系统层根据用户选择的并行模式（数据并行），调用相应的处理逻辑。例如，完成模型文件的完整分发和数据集的切分操作。

2）搭建试验环境。系统层根据用户需求调用容器层接口，在容器层部署试验环境。具体操作包括：

（1）搭建不规则算力中心拓扑；

（2）选择算力中心所需的 Ubuntu 20.04 版本镜像和 Astra-Sim 预设镜像；

（3）根据配置文件中的信息配置对应的链路网络信息（如带宽、延迟、丢包率等），具体如图 7 所示。

3）配置文件分发。系统层将各个算力中心的配置文件分发到对应节点，并根据用户在各个算力中心的配置生成对应的执行指令。

4）确定执行时序。系统层确定试验的整体执行时序，具体包括步骤：模型文件与数据集文件分发、单算力中心进行训练模拟、算力中心间进行梯度汇总、各算力中心进行参

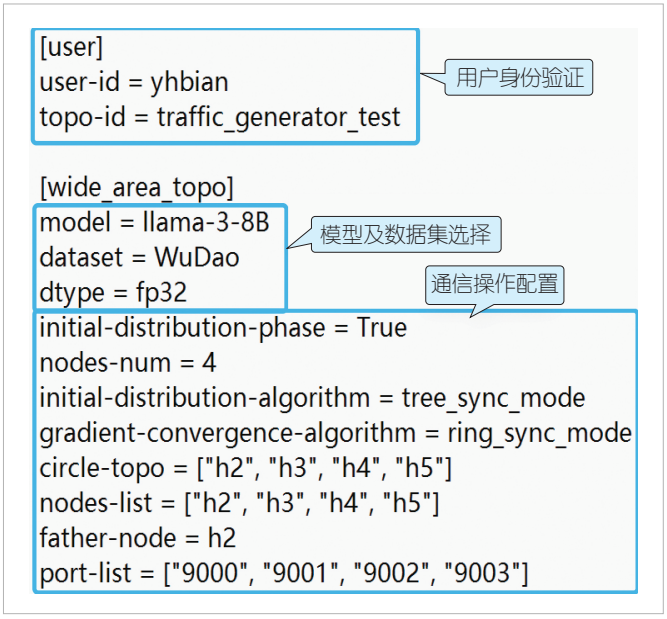

图6 用户配置文件预览

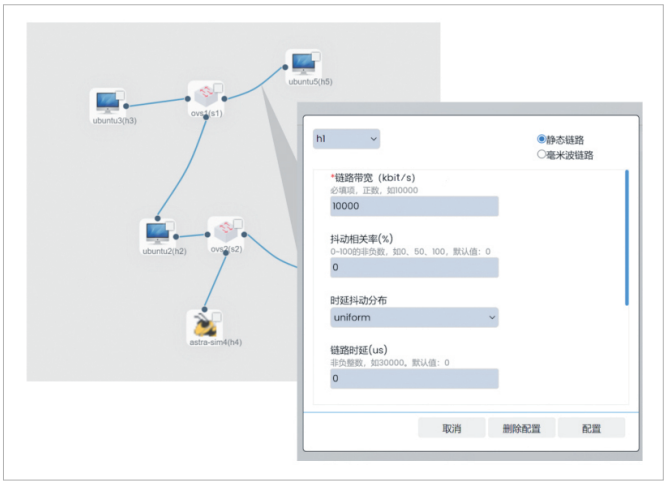

图7 搭建试验拓扑实例

数更新、进入下一个Epoch。系统层根据该执行时序依次执行对应的指令和操作。

3.3 仿真结果

仿真架构将模拟训练过程的结果保存至用户指定的路径，供用户后续分析和优化使用[19]。试验结果包括通信开销、计算开销、训练时间等关键指标，具体如图8所示。

1）广播阶段的试验结果呈现了数据并行模式下，系统在初始模型分发和训练集分发过程中的通信细节。用户可通过调整网络拓扑、变更广播通信算法等方式，实现对广播阶段性能的优化。

2）单个算力中心内部的模拟结果反映了各个算力中心内部的独立仿真情况。用户可以通过调整算力中心内部的算力配置、调整内部并行计算模式等方法，对算力中心的训练过程进行优化。

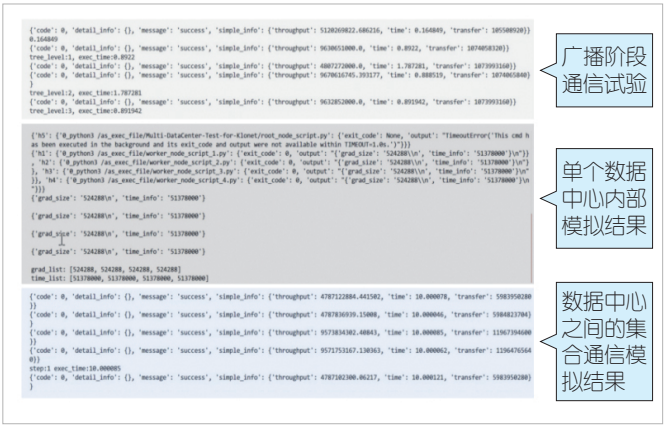

图8 仿真架构输出结果

3）算力中心之间的集合通信模拟结果，展示了在梯度汇聚与参数分发阶段广域环境下的集合通信操作特性。用户可以通过调整集合通信算法、优化网络配置等方法，对这一部分的通信时间进行优化，以提升整体训练效率。

4 结束语

本文中我们提出了一种面向多算力中心协同的广域智算网络仿真架构。该架构充分借鉴现有主流智算仿真架构的设计经验，在拓扑搭建、系统架构及易用性设计等关键层面进行创新性优化。通过构建高自由度拓扑搭建机制，设计高还原度广域网络模拟系统，并融入高易用性系统设计，为广域智算领域研究人员提供了功能完备、操作便捷的仿真平台。后续研究将持续提升架构的仿真精度与运行性能，探索新型网络技术与调度算法，并针对多样化应用场景进行适应性优化，助力推动中国AI产业发展，丰富全球技术生态。

参考文献

[1] JIANG Z, LIN H B, ZHONG Y M, et al. MegaScale: scaling large language model training to more than 10, 000 GPUs [EB/OL]. [2025-03-18]. https://arxiv.org/abs/2402.15627
[2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2025-03-19]. https://dl. acm. org/doi/10.5555/3295222.3295349
[3] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [EB/OL]. [2025-03-19]. https://cdn. openai. com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
[4] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners [EB/OL]. (2020-07-22) [2025-03-15]. https://arxiv.org/abs/2005.14165
[5] DeepSeek-AI, LIU A X, FENG B, et al. DeepSeek-V3 technical report [EB/OL]. (2025-02-18) [2025-03-18]. https://arxiv.org/abs/

2412.19437
[6] DeepSeek-AI, LIU A X, FENG B, et al. DeepSeek-V2: a strong, economical, and efficient mxture-of-experts language Model [EB/OL]. (2024-06-19) [2025-03-18]. https://arxiv. org/abs/2405.04434
[7] DeepSeek-AI, GUO D, YANG D, et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning [EB/OL]. (2025-01-22) [2025-03-16]. https://arxiv. org/abs/2501.12948
[8] RASHIDI S, SRIDHARAN S, SRINIVASAN S, et al. ASTRA-SIM: enabling SW/HW co-design exploration for distributed DL training platforms [C]//Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). IEEE, 2020: 81-92. DOI: 10.1109/ispass48437.2020.00018
[9] WON W, HEO T, RASHIDI S, et al. ASTRA-sim2.0: modeling hierarchical networks and disaggregated systems for large-model training at scale [C]//Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). IEEE, 2023: 283-294. DOI: 10.1109/ISPASS57527.2023.00035
[10] WANG X Z, LI Q X, XU Y C, et al. SimAI: unifying architecture design and performance tunning for large-scale large language model training with scalability and precision [EB/OL]. [2025-03-18]. https://ennanzhai.github.io/pub/nsdi25spring-simai.pdf
[11] BANG J, CHOI Y, KIM M, et al. vTrain: a simulation framework for evaluating cost-effective and compute-optimal large language model training [C]//Proceedings of 57th IEEE/ACM International Symposium on Microarchitecture (MICRO). IEEE, 2024: 153-167. DOI: 10.1109/MICRO61859.2024.00021
[12] RILEY G F, HENDERSON T R. The ns-3 network simulator [M]// Modeling and Tools for Network Simulation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 15-34. DOI: 10.1007/978-3-642-12331-3_2
[13] SRIDHARAN S, HEO T, FENG L, et al. Chakra: advancing performance benchmarking and co-design using standardized execution traces [EB/OL]. (2023-05-26) [2025-03-20]. https://arxiv.org/abs/2305.14516
[14] SAMAJDAR A, ZHU Y, WHATMOUGH P N, et al. SCALE-sim: systolic CNN accelerator [EB/OL]. [2025-03-20]. http://arxiv.org/abs/1811.02883
[15] HUBERT B. Linux advanced routing & traffic control [EB/OL]. [2025-03-20]. https://www. kernel. org/doc/ols/2002/ols2002-pages-213-222.pdf
[16] MA T, LUO L, YU H F, et al. Klonet: an easy-to-use and scalable platform for computer networks education [EB/OL]. [2025-03-20]. https://www. usenix. org/conference/nsdi24/presentation/ma
[17] DUBEY A, JAUHRI A, PANDEV A, et al. The llama 3 herd of models [EB/OL]. [2025-03-20]. https://arxiv.org/abs/2407.21783
[18] YUAN A, ZHAO H, DU Z, et al. WuDaoCorpora: a super large-scale Chinese corpora for pre-training language models [EB/OL]. [2025-03-20]. https://www. sciencedirect. com/science/article/pii/S2666651021000152
[19] 冯文佼, 李宗航, 虞红芳. 低资源集群中的大语言模型分布式推理技术 [J]. 中兴通讯技术, 2024, 30(2): 43-49. DOI: 10.12142/ZTETJ.202402007

作 者 简 介

边彦晖，电子科技大学在读本科生；研究方向为算网融合、分布式大模型训练模拟技术。

刘明远，北京交通大学电子信息工程学院讲师、信息与通信工程学院在读博士；研究方向为网络算网融合、低空网络、多路径传输调度等。

虞红芳，电子科技大学教授、博士生导师，信息与通信工程学院副院长；长期致力于智慧网络及应用研究；主持多个国家自然基金重大项目、国家重点研发计划、军科委等国家级项目；曾获教育部自然科学奖二等奖、电子学会科技进步二等奖、四川省教学成果二等奖等奖项，主持研发的“跨数据中心高性能分布式机器学习系统GeoMX”和“基于轻量级虚拟化的大规模网络创新平台Klonet”分别获中国通信学会2021年未来网络领先创新科技成果奖、2021年网络5.0创新科技成果奖；发表论文100余篇，授权中国发明专利30余项、美国发明专利2项，出版学术专著5本。