

确定性网络技术与算力网络 的应用思考

汇报人：张桂玉

日期：2024年7月10日





01

算力网络的发展情况

1.1 “东数西算” 战略促进算力网络成为业界发展热点

政策驱动：明确顶层建设目标

《算力基础设施高质量发展行动计划》

- 一、完善算力综合供给体系
- 二、提升算力高效运载能力
- 三、强化存力高效灵活保障
- 四、深化算力赋能行业应用
- 五、促进绿色低碳算力发展
- 六、加强安全保障能力建设

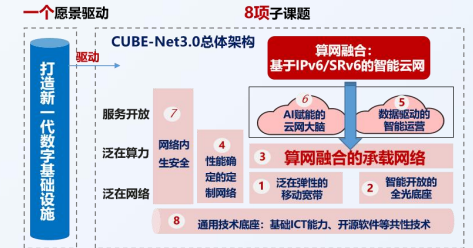
《深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》

- 一、统筹通用算力、智能算力、超级算力的一体化布局
- 二、统筹东中西部算力的一体化协同
- 三、统筹算力与数据、算法的一体化应用
- 四、统筹推动算力与绿色电力的一体化融合
- 五、统筹算力发展与安全保障的一体化推进
- 六、加强统筹、政策、技术等保障措施

运营商战略：算力网络建设是重要战略发展方向



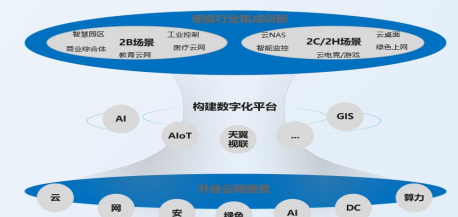
- 推进**架构先进、安全可靠、服务卓越的算力网络新布局**，为数字经济打造“第一算力引擎”



- 算网“**三个突破/三个转变/三个融合**”，包括算网基础设施构建、业务融合创新、创新技术引领



- 升级**云网融合3.0战略**，加速智算领域布局，加强算网相关平台研发



1.2 智算快速发展，对算力网络提出新的确定性要求

AI算力时代发展加速



OpenAI发布ChatGPT，5天用户量突破100万，2023年2月达到月活1亿

智能算力规模飞速增长，预计2024-2027年中国智能算力规模年复合增长率可达到33.9%，远超同期通用算力规模增长率



小模型

专用领域，感知理解

少量数据
单数据中心训练



大模型

通用领域，生成创造

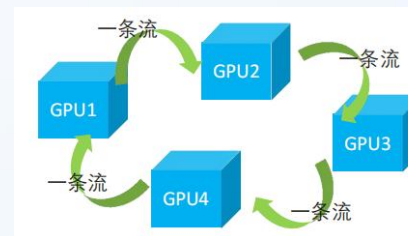
海量数据，高效协同
需考虑万卡以上规模

智算业务形态变化亟需网络演进升级，提供确定性保障能力

模型	训练数据	模型大小
盘古	40TB	500GB
GPT3	45TB	700GB
GPT4	2000TB	7200GB

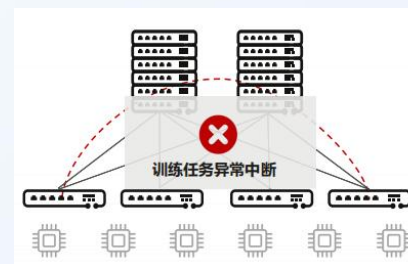
确定性带宽

模型规模增长催生海量训练数据上传，需要网络具备确定性带宽保障能力




确定性质量

多算力资源协同训练，为保证训练效率，需要网络具备确定性时延与丢包率保障能力



确定性可靠

AI训练的大规模组网协同，需要网络具备确定性的健壮性与可靠性保障能力



02

确定性网络技术的发展情况

2.1 确定性网络概念及技术特征

定义：确定性网络是指利用网络资源提供具有带宽确定性、时延确定性、抖动确定性和高可靠的业务体验的 IP 网络。

—2023年CCSA最新标准

发展：不同时期发展关注点也在不断变化，IEEE与IETF在2015年相继定义了局域“时间敏感网络”与广域“确定性网络”，之后**超低时延无损传输**也成为确定性网络的一个重要内容。目前关注的是**确定性和无损**两类。

关键指标：

确定性

■ 时延确定性

- 保证确定性的时延（时延的上界）和确定性的时延抖动（时延的上界和下界）。

■ 带宽确定性

- 轻载和拥塞场景下均能提供承诺的带宽，包括承诺保证带宽(CIR)和允许突发带宽（PIR）。

无损

■ 高可用性：

- 具备大于 99.999%及以上的高可用性，具备丢包确定性，低丢包率或长期无丢包特性（无损）。

■ 低时延：

- 降低消息传输中的处理延迟，增强数据传输的时效性，保证业务的低时延特性。

2.2 确定性网络技术发展情况

局域——标准成熟，园区/数据中心成熟商用

- **应用场景：**园区工业PLC、车载网络、智能变电站、移动前传等局域时延敏感场景；数据中心海量数据搬运及快速同步；
- **关键技术：**TSN（时延敏感网络），保证二层网络数据包到达时间、顺序的可控、稳定；局域无损通过IB、RoCE和RoCEv2协议提升系统吞吐，降低通信时延。
- **业界情况：**
 - 国内外市场提供TSN软硬件产品及服务、IB/RoCE交换机相对成熟。
- **运营商情况：**政企关注并拓展园区/智算中心市场。
- **标准组织：**2000-2010年IBTA先后标准化IB、RoCE协议；2015年IEEE定义TSN，2018-2022年国际国内标准先后确定，**标准成熟。**

广域——技术和标准尚不成熟

- **应用场景：**能源、医疗、工业等多行业均在探索确定性应用，如电力差动保护、远程手术等；AI大模型对智算广域海量数据**无损传输**提出要求。
- **关键技术：**IETF DetNet、CCSA DIP/EDN，其他：FlexE、SRv6，以实现**确定性保障效果为主**；广域RDMA通过网络测量、流量和拥塞控制、路径调优及确定转发等多种技术综合，保证**网络无性能损失**。
- **业界情况：**
 - 国内厂商有DetNet/DIP产品，但DIP暂未开放；广域RDMA暂无相关产品。
- **运营商情况：**
 - 运营商：以探索为主，认为技术值得研究和探讨。
- **标准组织：**IETF、CCSA并行立项DetNet/DIP标准，技术实现机制存在差异；国内发布部分广域RDMA标准，IETF暂无体系标准。

2.3 确定性网络技术也是业界关注的热点

紫金山实验室、三大运营商、厂商等均在积极推进确定性网络发展和探索相关应用

紫金山实验室

- 积极推动确定性网络技术的发展，并构建确定性广域网CENI；
- 21年进行云化PLC验证确定性技术；3年时间完成了一系列行业创新应用验证。



运营商

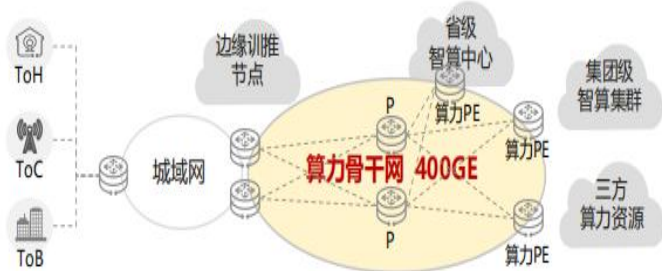
- **联通**：携手合作伙伴成立“中国联通5G-TSN联合攻关实验室”，发布轻量化5G-TSN 端到端解决方案及《中国联通5G-TSN技术发展白皮书》；结合多行业应用制定端到端确定性网络解决方案
- **移动**：率先发布“确定性网络服务1.0”，为业务提供在时延、带宽、路径、质量方面的确定性网络体验，有力支撑数据快递等新型业务快速发展。
- **电信**：联合宝钢集团打造“5G确定性网络在智能制造中的应用”项目荣获2024年亚洲5G行业创新奖；

厂商

- 主流厂商积极跟进研发定制化DIP/DetNet设备，推进相关技术实验及试点验证。

2.4 各种技术结合保障端到端的网络确定性

确定性带宽保障



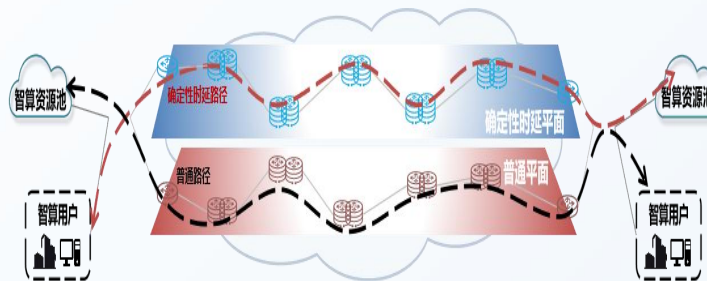
超大带宽网络

400G超宽骨干网，提供超宽调度网络管道

承诺带宽保证

通过FlexE和小颗粒切片等技术保障确定性带宽，实现无论轻载还是拥塞场景下广域连接承诺保证带宽（CIR）和容许保证带宽(PIR)

确定性时延/抖动保障



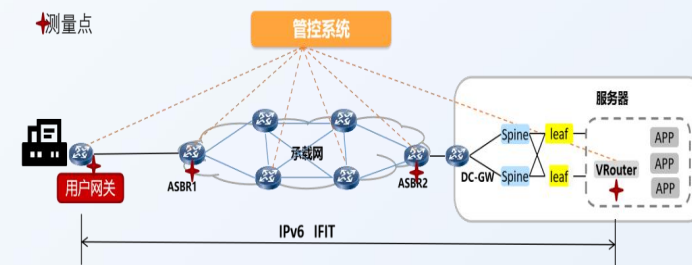
SRv6路径随选

通过SRv6满足算力之间差异化路径需求（最小时延等）

确定性技术确保时延稳定

通过DetNet/DIP的周期时隙映射保障确定性时延

确定性可靠性保障



多发选收保障高可用

在入口设备上将报文复制多份，经多条路径转发，确保至少有一份报文准确到达对端，确保网络高可用性

随流感知业务质量

通过端到端随流检测检测业务级性能指标，如丢包率、抖动等，逐跳随流检测定位故障

2.5 各种技术结合保障端到端的网络确定性

- 通过实时测量网络与业务指标，多维度感知网络与业务状况，进行性能指标分析并关联触发告警，基于告警对网络进行优化调整，保障业务的确定性服务质量，实现确定性网络**全周期闭环智能管控**。
- 实现从传统网络“**尽力而为**” → 确定性网络“**按时**” → “**实时**” “**确保所需**”的突破。



2.6 确定性网络技术可在多行业场景应用

- 确定性网络技术可在智慧交通、远控医疗、会议电视等典型场景提供**差异化、确定性的带宽、时延、抖动服务**，并可实现与非确定性业务的综合承载。

智慧交通

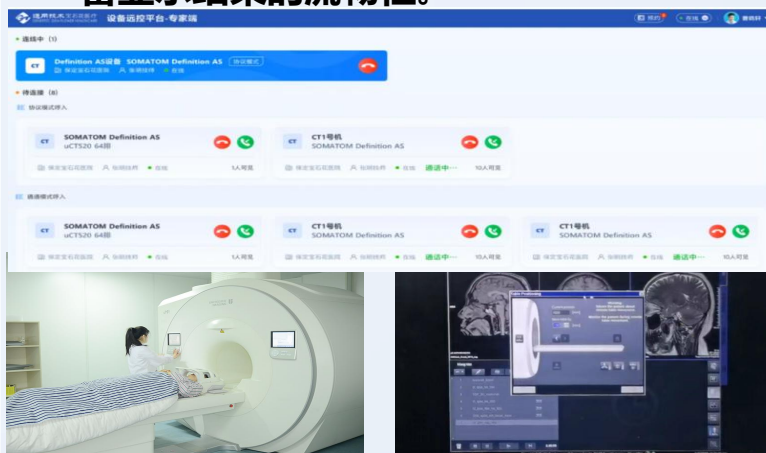
- 利用视频摄像机、毫米波/激光雷达、高精度地图等技术，形成**准确、全面的车辆、行人、非机动车时空轨迹数据**。
- 交警总队指挥中心实时监控**交通流量、运行态势和重点道路视频等**。



- **大带宽，中等时延：带宽100M，时延 $\leq 30\text{ms}$ ，抖动 $< 15\text{ms}$ 。**

远控医疗

- 面向基层医疗影像能力不足场景，基层技师申请远程专家协助，专家通过语音、视频、远程操控等方式指导操作相关仪器。
- 保证专家**远程操控的实时性和查看医疗设备显示结果的流畅性**。



- **中等带宽，低抖动：单终端20M，时延 $\leq 10\text{ms}$ ，抖动 $< 5\text{ms}$ 。**

会议电视

- 会议电视终端采集音视频信息，转化成数据报文传输给异地会议电视终端。
- 视频具备较高清晰度，保证与会者**实时、顺畅互动**，音视频同步传输，无不同步、画面卡顿等问题。



- **大带宽，低时延：带宽100M，时延 $\leq 15\text{ms}$ ，抖动 $< 10\text{ms}$ 。**



03

智算场景下确定性网络技术的发展与应用

确定性大带宽与流量吞吐的跨域数据快递服务，提供训练数据的确定性时效传输

传统数据传输采用线下硬盘快递+人工拷贝方式

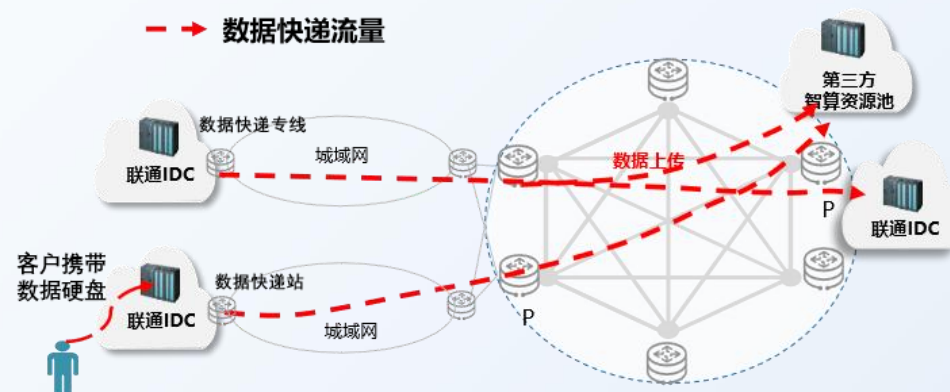
痛点1：成本高、损耗大

- **硬盘投资**：5K*28块硬盘*20台车 = 280万，每年损耗1/4
- **数据传输**：快递4万/月*12+专人24万/年+硬盘损耗280万*25% = 142万/年

痛点2：效率低、风险高

- 硬盘从寄出、上传到寄回，耗时两周
- 硬盘丢失/损坏+数据来不及上传（网联法规要求存储不超过7天），丢失率20%

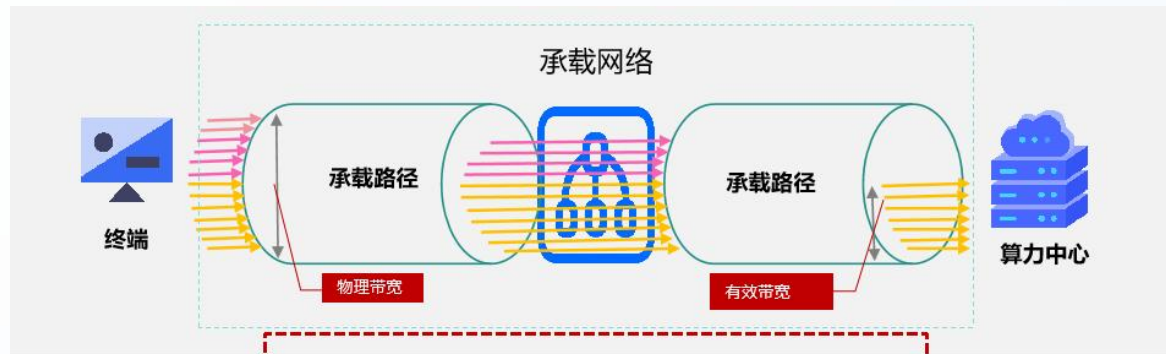
高通量数据传递服务：基于“数据量+完成时间”提供相对确定的数据传输能力保障，完成时效相对确定的数据传输



■ 确定性技术应用需求：

- **确定性带宽保障**：提供不同阶梯的高带宽保障弹性专线，实现大数据小时级传输

通过增大有效带宽，提高单位带宽下的数据传输体量，实现智算场景下的高通量数据传递



- 网络优化关键技术：**
- ① SRv6承载选路
 - ② 无损传输
 - ③ 负载均衡
 - ④ 带宽检测协议

有效带宽最大化

- 区分时效性与实时性业务；
- 在满足时效性要求的前提下，通过充分利用网络带宽潮汐特征以及网络多路径提升网络总吞吐，实现有效带宽最大化。

传输效率最大化

- 传输效率越高，单位数据量完成传输的时间越短，为用户提供的传输服务质量越高；
- 通过最小化待传输任务体量，以及传输层协议优化，提供传输效率。

网络丢包最小化

- 通过承载网络侧传输路径调优、负载均衡等技术实现最小化网络丢包；
- 在网络建设成本一定的前提下，最大化提高业务承载质量。

现网影响最小化

- 高通量数据网承载的业务除了超算、智算、东数西算业务外，还有运营商传统业务；
- 尽量减少对现有网络的改造。

中国联通2024年4月完成上海-宁夏广域传输测试验证，首次实现了基于IP承载网络的3000公里海量数据任务式长距传输，相同带宽下超大文件传输效率比现有能力提升4倍

大模型远端训练与跨域协同训练，需要确定性的网络质量保障

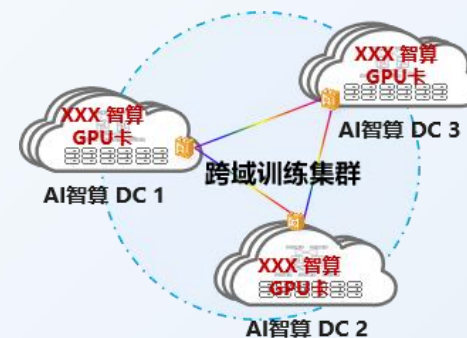
基于广域无损的大模型远端训练服务

远端数据安全协同训练服务：对于特定客户的敏感数据，满足数据不出域的管理规定，保证数据不在算力侧落地存储



基于广域无损的智算中心集群训练

跨域数据中心协同训练：受机房环境、用电等限制，单智算中心规模无法满足未来AI训练的十数万卡需求，需考虑不同地域的集群算力聚合为更大的算力资源池，进行协同训练



■ 确定性技术应用需求：

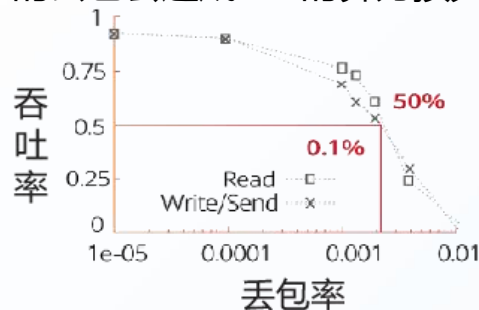
- **确定性带宽保障：**提供广域超大带宽能力
- **网络无损控制：**广域RDMA传输保证数据不丢包不重传

3.2 智算场景下确定性网络技术应用--广域无损 (2/2)

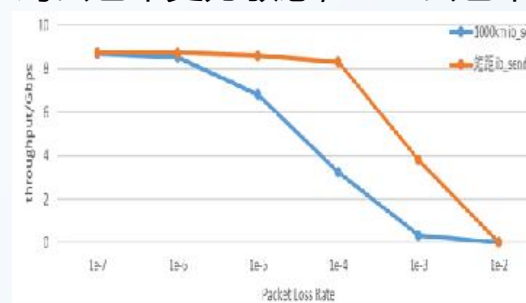
广域RDMA是指突破RDMA长距传输瓶颈，实现广域高性能互联无丢包、低延时的无损传输

长距RDMA丢包更敏感，距离与丢包影响算力效率

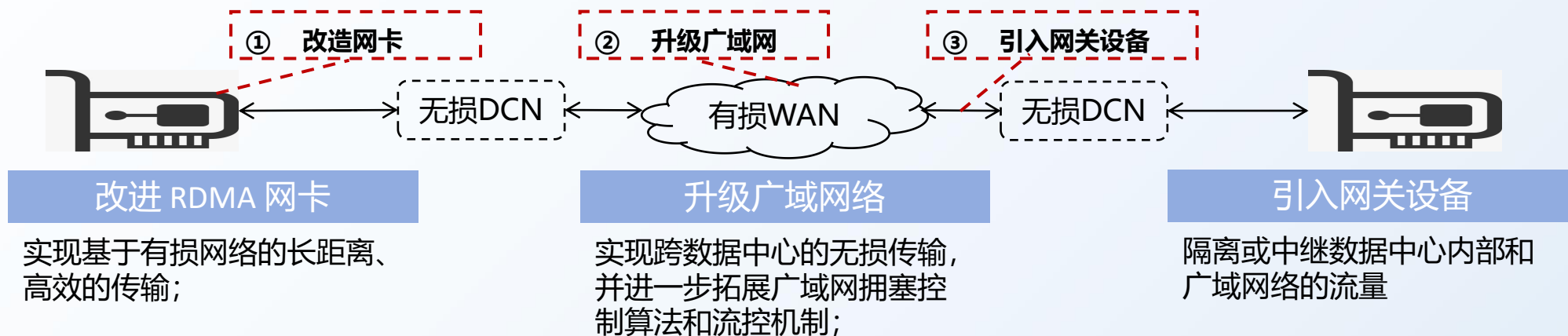
➤ 0.1%的丢包会造成50%的算力损失



➤ 长距RDMA (1000KM) 对丢包率更为敏感，1e-4丢包率吞吐下降60%以上



基于不同的场景与需求，广域 RDMA 的技术路线目前主要包括 3 类



RDMA作为广域无损的核心技术成为共识，但仍存在多项挑战需要技术攻关突破

开放·创新·融合·共赢

共同把握算网融合发展机遇

