

ICS 33.040.40

CCS I631

T/NIDA

全球固定网络创新联盟

T/NIDA 010-2025

基于 AI 大模型的远程诊疗分布式训推广域 网技术要求

Technology Requirements of Distributed Training-Reasoning WAN for Telemedicine
based on AI Large-Scale Models

2025-10-26 发布

2026-1-1 施行

全球固定网络创新联盟（NIDA）发布

目 次

前 言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 面向远程诊疗分布式训推场景概述 2

5.1 专业联盟内医院通过医疗大模型进行辅助诊疗 2

5.2 医疗集团进行医疗模型训练开发与推理部署 2

5.3 社区医院/诊所租赁云上医疗模型进行辅助诊疗 2

6 面向远程诊疗分布式训推广域网络架构与部署要求 2

6.1 面向远程诊疗分布式训推广域网络架构 2

6.2 面向远程诊疗分布式训推广域网络部署要求 3

7 面向远程诊疗分布式训推广域关键技术能力要求 5

7.1 广域无损 5

7.2 高吞吐 6

7.3 安全 6

图 1 面向远程诊疗分布式训推广域网络架构2

图 2 专业联盟内医院通过医疗大模型进行辅助诊疗场景3

图 3 医疗集团进行医疗模型训练开发与推理场景（分院无IT机房）4

图 4 医疗集团进行医疗模型训练开发与推理场景（分院有IT机房）4

图 5 社区医院/诊所租赁云上医疗模型进行辅助诊疗场景 5

图 6 广域xSEC提供高性能自适应多点加密安全保证6

前 言

本文件按照 GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利权和著作权。本文件的发布机构不承担识别专利和著作权的责任。全球固定网络创新联盟不对标准涉及专利的真实性、有效性和范围持有任何立场；不涉足评估专利对标准的相关性或必要性；不参与解决有关标准中所涉及专利的使用许可纠纷等。

本文件由全球固定网络创新联盟技术委员会提出并归口。

本文件由全球固定网络创新联盟拥有版权，未经允许，严禁转载。

本文件起草单位：中南大学湘雅医院“移动医疗”教育部-中国移动联合实验室、中国联通、中国电信、中移（成都）信息通信科技有限公司、飞腾信息技术有限公司、北京协和医院、贵州省人民医院、西安交通大学第一附属医院、江苏省人民医院、兰州大学第一医院、山东省立医院、深圳市坪山区人民医院、南昌大学第一附属医院、华为技术有限公司

本文件主要起草人：黄伟红、庞冉、曹畅、朱永庆、胡泽华、种璟、范金鹏、朱雯、贺松、卫荣、王忠民、郗群、包国峰、王逸欣、曹磊、彭书萍

面向远程诊疗分布式训推广域网技术要求

1 范围

本文件规定了全球固定网络创新联盟中面向远程诊疗分布式训推广域网技术要求，包括面向远程诊疗分布式训推广域网场景、架构与部署要求、关键技术能力要求。

本文件适用于面向远程诊疗分布式训推广域网，主要应用于指导面向远程诊疗分布式训推广域网规划、设计和验收。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

3 术语和定义

3.1 面向智算业务的广域网 WAN for intelligent computing

广域网WAN (Wide Area Network) 对应的是跨域问题，智算广域具体到智算业务场景，是指跨智能计算中心互联和协同等场景。面向人工智能时代的新型广域网，智能IP广域网（AI WAN），与AI深度融合，由AI路由器、AI新网络、AI新大脑等构成，面向政企、行业、公众用户提供内生智能、多维感知、差异体验、安全可信、绿色低碳的网络服务，促进网络和业务融合向更智能、更高效、更安全的方向演进。

4 缩略语

下列缩略语适用于本文件。

AI: 人工智能 (Artificial Intelligence)

BGP: 边界网关协议 (Border Gateway Protocol)

CPE: 用户驻地设备 (Customer-Premises Equipment)

ECN: 显式拥塞通告 (Explicit Congestion Notification)

GRE: 通用路由封装 (Generic Routing Encapsulation)

L3EVPN: 三层以太网虚拟专用网络 (Layer3 Ethernet Virtual Private Network)

IP: 互联网协议 (Internet Protocol)

IT: 信息技术 (Information Technology)

PFC: 基于优先级的流量控制 (Priority-based Flow Control)

RDMA: 远程直接存储器访问 (Remote Direct Memory Access)

xSEC: 扩展安全 (Enhanced Security)

SRv6: 基于IPv6数据面的段路由 (Segment Routing over IPv6 Data Plane)

WAN: 广域网 (Wide Area Network)

5 面向远程诊疗分布式训推场景概述

5.1 专业联盟内医院通过医疗大模型进行辅助诊疗

专业联盟内医院通过医疗大模型进行辅助诊疗，医院训练大模型使用的样本不能出医院，联盟内其他医院使用推理服务时数据不出医院，训推/推理使用的算力可灵活扩缩。

5.2 医疗集团进行医疗模型训练开发与推理部署

医疗集团进行医疗模型训练开发与推理部署，分院无IT机房，训练样本及推理数据安全回传总院统一处理。分院样本数据/推理请求可通过安全通道传输给总院，总院训练医疗大模型使用的样本以及推理数据不能出集团，训推/推理使用的算力可灵活扩缩。

医疗集团进行医疗模型训练开发与推理部署，分院有IT机房，训练样本安全回传总院，推理数据不出分院园区。分院样本数据可通过安全通道传输给总院，总院训练大模型使用的样本以及推理数据不能出集团，分院的推理数据不出园区，训推/推理使用的算力可灵活扩缩。

5.3 社区医院/诊所租赁云上医疗模型进行辅助诊疗

社区医院/诊所租赁云上医疗模型进行辅助诊疗，社区医院/诊所按需部署算力，社区医院/诊所内诊疗数据不出医院/诊所。

6 面向远程诊疗分布式训推广域网络架构与部署要求

6.1 面向远程诊疗分布式训推广域网络架构

面向远程诊疗分布式训推广域网络架构，如图 1 所示，对模型进行模型切分，在边侧部署部分切分后的模型，云端智算中心部署部分切分后的模型，模型的输入输出层部署在医院园区。

通过边云协同训推，Prompt 在本地输入，Token 在本地生成，保证数据安全不出域；中间传输隐变量数据，经模型变化的高维向量，保证传输数据可用不可见；在云端部署模型中间层，无法还原完整模型，保证云端模型安全。模型参数规模越大、输入/输出序列越长、部署在客户侧的层数越多，数据安全的防御能力越强。

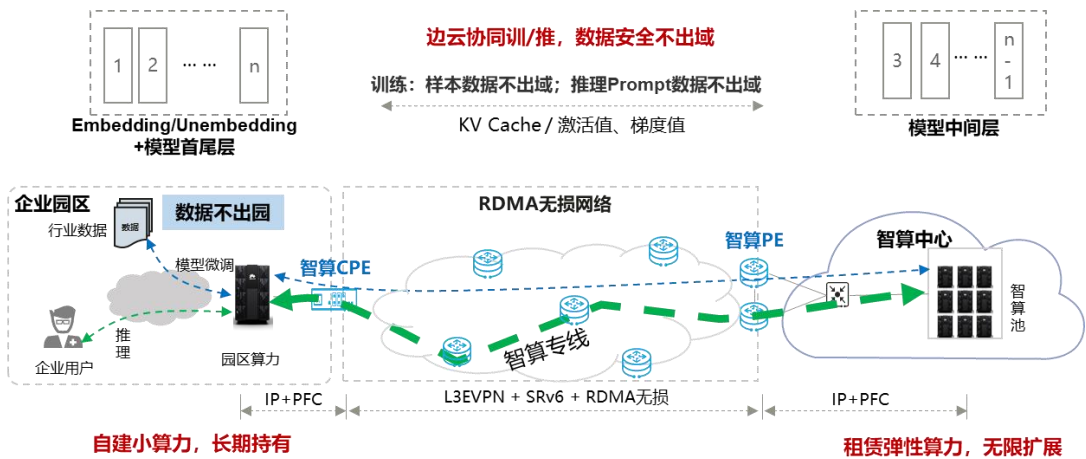


图 1 面向远程诊疗分布式训推广域网络架构

- 模型分层部署，保障数据不出域：模型部分部署在企业侧，部分部署在云端，中间交互梯度值、激活值等，保障样本/Prompt 数据不出域。

- RDMA 无损网络，保障业务 SLA：智算广域网通过 L3EVPN + SRv6 + RDMA 无损技术对训推参数面 RDMA 业务流高性能传输，训练算效高，推理时延、吞吐有保障。

本文中面向智算业务的广域网可以是智能IP广域网（AI WAN），即面向人工智能时代的新型广域网，与AI深度融合，由AI路由器、AI新网络、AI新大脑等构成，面向政企、行业、公众用户提供内生智能、多维感知、差异体验、安全可信、绿色低碳的网络服务。

对于运营商综合承载广域网，如果支持基于AI大模型的远程诊疗分布式训推，可以通过FlexE、切片等技术与 SRv6 可编程路径能力相结合，面向远程医疗客户提供网络切片服务，为业务提供端到端差异化保障，保障确定性的低时延、低抖动和高可靠数据传输服务。

6.2 面向远程诊疗分布式训推广域网部署要求

6.2.1 专业联盟内医院通过医疗大模型进行辅助诊疗场景

专业联盟内医院通过医疗大模型进行辅助诊疗，如图 2 所示，医院A基于自身行业数据训练医疗行业大模型，模型训练完成后部署推理服务给专业联盟内其他医院使用。

专业联盟使用行业医疗模型进行安全推理，如图2所示，联盟内其他医院使用医院A发布的医疗行业模型进行实时推理，要求各自医院内的医疗数据不出园区，不共享。

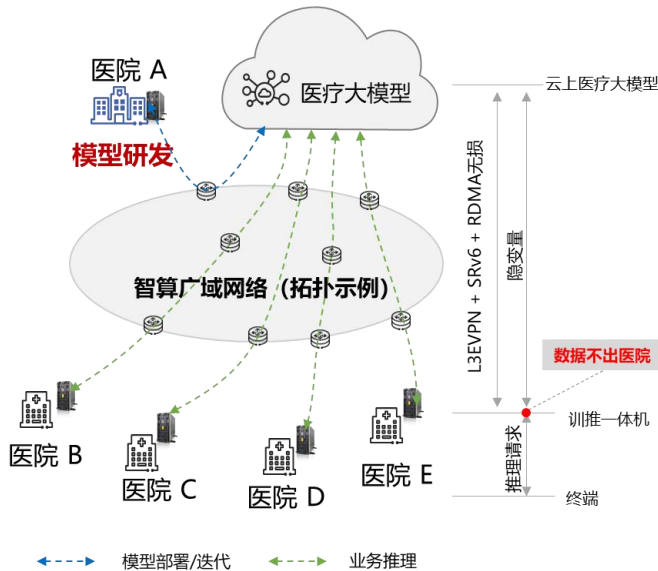


图 2 专业联盟内医院通过医疗大模型进行辅助诊疗场景

医院A园区部署训推一体机，通过智算广域网调度算力中心算力，例如，通过模型切割将模型首尾层部署在医院园区，中间层部署在智算中心，基于医院行业数据进行模型训练，训练完成后部署推理服务。

联盟内其他医院园区部署训推一体机，医院A部署医疗行业模型推理服务，云上为各医院分别部署推理实例，进行实时推理，医疗数据不出园区。

在联盟内医院及智算中心间建设智算广域网，通过L3EVPN + SRv6 + RDMA无损技术为联盟内医院提供无损高吞吐网络连接。

6.2.2 医疗集团进行医疗模型训练开发与推理场景

医疗集团进行医疗模型训练开发与推理场景，在分院无IT机房时，如图 3 所示，各分院将医疗数据汇集到总院，需网络提供大带宽、高吞吐、安全的数据通道。总院基于汇集行业数据，租赁/购买算力进行行业模型训练/微调。总院/分院基于训练的医疗行业模型进行实时推理，医疗数据不出集团。

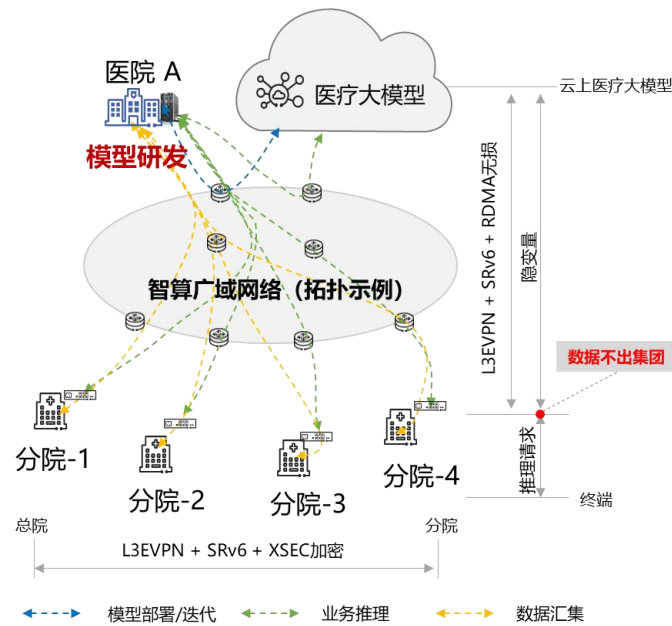


图 3 医疗集团进行医疗模型训练开发与推理场景（分院无IT机房）

总院园区部署训推一体机，通过智算广域网络调度算力中心算力，例如，通过模型切割将模型首尾层部署在医院园区，中间层部署在智算中心，基于医院行业数据进行模型训练，训练完成后部署推理服务。

分院园区部署智算CPE，通过智算广域网络与总院互联，传输医疗数据及实时推理请求。

在总院、分院及智算中心间建设智算广域网络，通过L3EVPN + SRv6 + RDMA无损技术为总院与智算中心间提供无损高吞吐网络连接，通过L3EVPN + SRv6 + XSEC加密实现总院/分院间数据安全传输。

医疗集团进行医疗模型训练开发与推理场景，在分院有IT机房时，如图 4 所示，各分院将医疗数据汇集到总院，需网络提供大带宽、高吞吐、安全的数据通道。总院基于汇集行业数据，租赁/购买算力进行行业模型训练/微调。总院/分院基于训练的医疗行业模型进行实时推理，医疗数据不出集团。

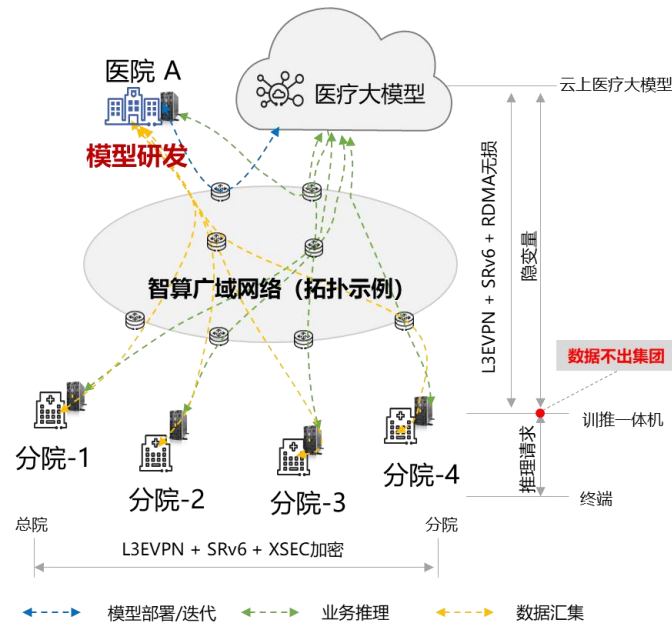


图 4 医疗集团进行医疗模型训练开发与推理场景（分院有IT机房）

总院园区部署训推一体机，通过智算广域网络调度算力中心算力，例如，通过模型切割将模型首尾层部署在医院园区，中间层部署在智算中心，基于医院行业数据进行模型训练，训练完成后部署推理服务。

分院园区部署训推一体机，部署医疗行业模型推理服务，云上为各分院分别部署推理实例，进行实时推理，医疗数据不出园区。

在总院、分院及智算中心间建设智算广域网络，通过L3EVPN + SRv6 + RDMA无损技术为总院与智算中心间提供无损高吞吐网络连接，通过L3EVPN + SRv6 + XSEC加密实现总院/分院间数据安全传输。

6.2.3 社区医院/诊所租赁云上医疗模型进行辅助诊疗场景

社区医院/诊所使用行业医疗模型安全推理，如图 5 所示，社区医院/诊所使用云上医疗大模型进行辅助诊断，社区医院/诊所内的医疗数据不出园区。

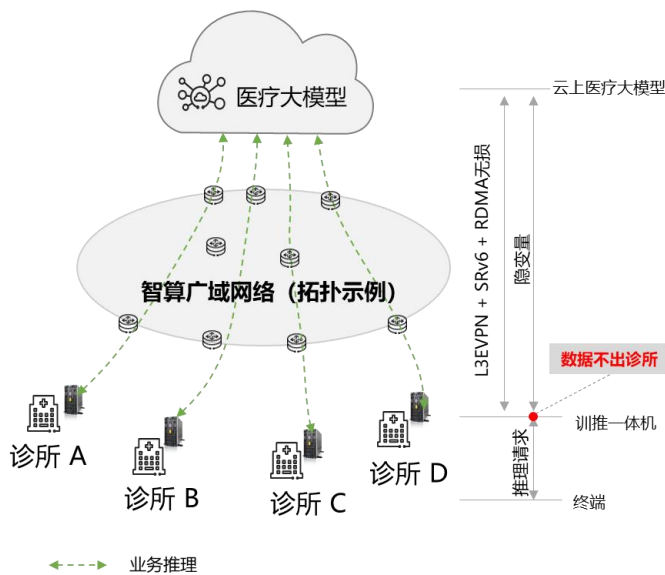


图 5 社区医院/诊所租赁云上医疗模型进行辅助诊疗场景

社区医院/诊所内部署训推一体机，通过租赁云上医疗大模型服务，进行安全推理，例如，将进行模型切割的医疗大模型首尾层部署在社区医院/诊所的一体机内，模型中间层部署在云上智算中心，进行云边协同的安全推理。

在社区医院/诊所及云上智算中心间建设智算广域网络，通过L3EVPN + SRv6 + RDMA无损技术为社区医院/诊所与智算中心间提供无损高吞吐网络连接。

7 面向远程诊疗分布式训推广域网关键技术能力要求

7.1 广域无损

人工智能大规模模型分布式训推场景对广域网络提出了“零丢包”的广域无损挑战。现有的基于优先级的流量控制（PFC）和显式拥塞通知（ECN）等技术难以应对人工智能业务的突发流量，并可能造成头阻、死锁、风暴等问题。通过流级精准流控，避免在模型训练的过程中发生丢包，保障边侧算力与云侧算力协同过程中算效不下降。

智算广域网络需要具有广域无损的能力：

- 用户优先级反压流控能力：基于广域流队列构筑微秒级拥塞感知和反压流控，通过为每租户或每业务创建队列，并感知队列拥塞情况，当拥塞时向上游设备精准反压，实现用户队列的零丢包弹性传输能力；
- 网络路径缓存能力：在不影响其他用户队列流量的情况下，利用网络路径上的缓存能力，突破单一网元的缓存限制，有效应对网络突发等不确定因素，解决传统流控技术面临的挑战，保障广域拥塞不扩散，实现广域长距零丢包无损传输；
- 边云协同能力：避免在模型训练的过程中发生丢包，保障边侧算力与云侧算力协同过程中算效不下降。

7.2 高吞吐

通过智算广域网络传输海量数据样本到智算中心来进行计算，对网络吞吐率提出了更高的要求。训练数据和样本传输产生的流量呈现大象流特征，与普通流量相比，大象流具有流数少、单流带宽大、持续时间长的特点，单流带宽可达到10Gbps。大象流的出现会导致网络负载不均衡，流量拥塞，网络吞吐率大幅下降。

智算广域网络需要具有高吞吐的能力：

- 大象流智能识别：网络设备精准识别大象流，实时采集流信息并上报给网络控制器；
- 基于流的自适应调优：网络控制器实时监测网络负载，根据各个路径上的负载情况动态调整业务流的转发路径，实现网络全局负载均衡，保障网络高吞吐。
- 当网络中出现大象流导致某条路径上部分链路拥塞时：网络控制器需要重新规划流量转发路径，将部分流量调整至其他路径，保证各路径上负载均衡，网络维持高吞吐。
- 在多用户并发训练任务时：需要保证训练数据和样本的上传效率。

7.3 安全

面向远程诊疗分布式训推，模型、样本、训练数据安全尤其重要，一旦产生数据安全问题，医院将产生巨大的损失。医院内病例、影像等数据敏感，跨广域网络要求加密传输。基于IPsec加密机配置加密隧道，需要M*N方式进行P2P逐点配置。新业务百倍增长，基于ACL方式静态方式规划加密业务，复杂易出错。广域xSEC提供高性能自适应多点加密安全保证。

智算广域网络针对安全定义如下关键能力要求：

- 极简部署：基于BGP协议扩展自适应加密，效率提升30%；
- 精准加密：VPN业务加密，实现VPN业务级保护。
- 支持租户隔离：智算网络是多用户共享的网络，租户间需要隔离，避免数据泄露。

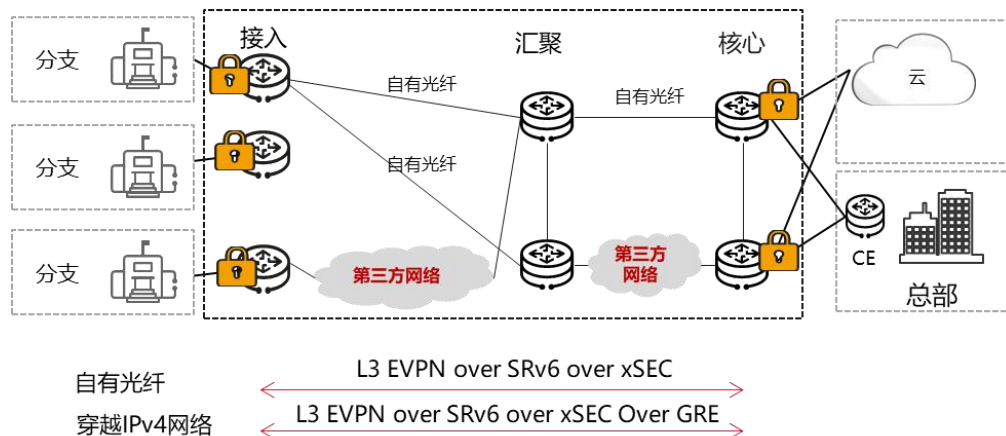


图 6 广域xSEC提供高性能自适应多点加密安全保证

7.4 高可靠

面向远程诊疗分布式训推广域网应防止因为链路拥塞、光模块故障、光纤质量下降等故障导致训练的中断，保证长期稳定。

面向远程诊疗分布式训推广域网需具备租户级故障隔离能力，实现租户级精准反压，避免故障扩散影响多租户算效。

面向远程诊疗分布式训推广域网需具备随流检测、高精仿真、网络自愈等智能运维能力，通过业务流级可视、秒级流量趋势展示、秒级故障感知和逐包故障定界等技术，实现业务流故障的快速定位与恢复，保障算力的高效能利用。