面向边缘-云协同的联邦学习安全架构研究

摘要：随着人工智能、物联网与云边端协同计算的快速发展，面向多机构、多终端的数据协同建模正在成为信息安全技术中的重要前沿方向。联邦学习通过“数据不出域、模型跨域协同”的方式缓解了数据孤岛问题，但在边缘-云协同环境下仍面临身份伪造、梯度泄露、模型投毒、中间人攻击、半可信聚合器越权分析等安全挑战。本报告围绕信息安全技术课程“新技术研究”的要求，从系统工程视角分析边缘计算、联邦学习、零信任、差分隐私、安全多方计算、同态加密、密文检索和可信执行环境等关键技术的协同关系，并归纳出“边缘层本地训练、云协调层安全聚合、安全控制层持续验证”的融合安全框架。随后结合医疗、车联网与工业互联网场景，讨论工程落地中的性能开销、治理边界、法规约束与社会可持续发展影响。综合现有研究可见，面向边缘-云协同的联邦学习安全体系不应再依赖单一的边界防护，而应构建覆盖身份、数据、模型、系统和合规全流程的内生安全机制，在保障数据价值释放的同时实现可信、可控与可审计的智能协作。

关键词：联邦学习；边缘计算；零信任；隐私计算；密文检索；可信执行环境

1 引言

人工智能模型的能力越来越依赖海量、多源、持续更新的数据，但现实中的数据往往分散在医院、工厂、车联网节点、移动终端以及各类行业信息系统中。传统集中式建模需要将数据汇聚到统一平台，这一方式虽然便于管理，却会显著增加跨域传输、集中存储和统一处理带来的泄露风险，也容易与个人信息保护、行业监管和业务竞争边界产生冲突。在数据要素化与智能化应用同步推进的背景下，如何在“不汇集原始数据”的前提下完成可信协同计算，已经成为信息安全研究与工程实践共同面对的核心问题[2,3]。

联邦学习为这一矛盾提供了重要的技术路径。其基本思想是由各参与节点在本地完成模型训练，仅上传参数更新、梯度或中间统计量，由聚合方整合为全局模型。这种“数据不动模型动”的机制缓解了数据出域风险，使跨机构协同建模具备现实可行性[1,2,3]。与此同时，边缘计算的发展使大量数据在接近数据源的位置被处理，云平台则承担统一协调、弹性算力和全局优化职能，二者结合形成了具有低时延、高扩展性和多域协同特征的边缘-云计算体系[9]。将联邦学习部署在边缘-云协同架构中，是当前智慧医疗、智能交通、工业互联网和城市治理等场景的重要发展方向。

然而，联邦学习并不天然安全，边缘-云协同也并不天然可信。边缘节点种类繁杂、能力异构、运行环境开放，攻击者既可能潜伏于终端和边缘节点，也可

能借助网络链路与聚合器位置发起攻击。模型更新中蕴含的数据特征可能被重构，恶意节点可以通过投毒和后门操纵全局模型，聚合器还可能利用其视角推断参与方行为。传统“内网可信、边界防护”的安全模式在这种分布式、动态、多主体场景下已明显失效。因此，本报告聚焦“面向边缘-云协同的联邦学习安全架构”这一主题，重点讨论零信任与隐私计算融合的设计思路，分析其在安全防护、工程应用、标准法规和可持续发展层面的综合价值。

2 相关技术基础与发展现状

2.1 边缘计算与云计算协同架构

边缘计算强调将计算、存储和安全能力下沉到靠近数据产生源的位置，以减少广域传输时延、降低核心网络压力，并提升本地自治处理能力。与之相对，云计算在集中算力、全局视野、弹性伸缩和统一编排方面具有优势。单纯依赖云端，难以满足车联网控制、工业生产调度、医疗设备实时分析等毫秒级或近实时场景；单纯依赖边缘，则又会受限于资源碎片化、管理复杂度高和全局模型难以收敛等问题。边缘-云协同的价值，正在于通过合理的任务切分，将“低时延处理”和“高强度聚合优化”统一在一个分层体系中[9]。

从信息安全视角看，边缘-云协同并非只是资源调度模型，更是多信任域并存的安全环境。边缘节点位于用户侧或行业现场，易受物理接触、恶意软件和配置漂移影响；云端具备强管理能力，却也因汇聚了训练任务、模型版本和审计数据而成为高价值攻击目标。因此，安全研究不能只关注传输加密是否开启，还必须关注节点身份可信、运行环境可信、数据使用边界可信和聚合过程可信等更深层问题。

2.2 联邦学习基本原理与发展

联邦学习通常由初始化、本地训练、参数上传、安全聚合和全局更新五个步骤构成。根据数据分布形式不同，可分为横向联邦学习、纵向联邦学习和联邦迁移学习三类。横向联邦学习适用于特征空间相似但样本主体不同的场景，如多家医院共享相同类型病历字段但患者不同；纵向联邦学习适用于样本主体重叠但特征不同的场景，如银行与电商联合开展信用评估；联邦迁移学习则用于样本和特征都重叠较少的复杂跨域任务[1,2,3]。

从协作机理看，横向联邦学习更强调同构字段条件下的大规模并行参与，纵向联邦学习更强调跨机构主体对齐和联合特征计算，迁移联邦学习则关注异构域之间的知识桥接与表示迁移。虽然三类模式在组织方式上存在差异，但其安全问题都围绕“初始化、本地训练、参数上传、聚合更新”这一基础闭环展开，这也是后续安全架构设计的共同落脚点。

近年来，联邦学习研究已从算法层面逐步扩展到系统、平台和治理层面。一

方面，研究者关注非独立同分布数据、通信瓶颈、客户端掉线和异构设备导致的训练不稳定问题；另一方面，也越来越重视模型更新的安全性、参与方身份真实性和聚合器可验证性[2,3]。随着大模型与端侧智能的发展，联邦学习正在从“单一算法框架”演变为“分布式可信智能基础设施”的组成部分，其边界不再局限于机器学习，而是与密码学、操作系统安全、可信硬件、访问控制、数据治理密切耦合。

2.3 零信任安全模型

零信任的核心并不是“不允许任何访问”，而是“不预设任何主体天然可信”。在边缘-云协同联邦学习中，用户、设备、应用、训练任务、模型版本乃至一次参数上传请求，都应被视为需要持续验证的对象。零信任安全模型通常强调统一身份、最小权限、动态授权、微隔离、持续评估和全链路审计[10]。这些原则与联邦学习的分布式特征高度契合，因为联邦学习的参与方常常是临时接入、跨组织协作、状态持续变化的节点集合。

对于联邦学习而言，零信任的价值主要体现在三个层面。第一，它能够解决“节点是否可以参与训练”的接入问题，通过设备证明、证书体系和上下文感知鉴别请求来源。第二，它能够解决“节点能做什么”的授权问题，根据任务敏感度、历史行为、位置环境和风险评分动态调整权限。第三，它能够解决“出现异常后如何快速收敛风险”的控制问题，例如隔离异常节点、限制其仅参与推理而不参与聚合，或要求额外的验证与人工复核。这种按会话、按任务、按风险持续收缩暴露面的思路，是当前分布式智能系统安全的重要发展方向。

2.4 隐私计算关键技术

隐私计算是实现“数据可用不可见”的一组关键技术集合，其中最常与联邦学习结合的包括差分隐私、安全多方计算、同态加密和可信执行环境。差分隐私通过向梯度、统计结果或输出模型引入可控噪声，降低攻击者从输出推断个体记录的能力，其优势在于机制清晰、易于与现有训练流程结合，但噪声注入会直接影响模型精度[5]。安全多方计算通过协议设计保证各方在不暴露原始输入的情况下协同完成目标函数计算，适合高敏感、多机构协作场景，但交互复杂、通信代价较高；在联邦学习实践中，安全聚合是其中最具代表性的工程化方案之一[4]。同态加密允许对密文直接计算，理论上能够让聚合方在看不到明文参数的情况下完成运算，但现实中对算力和延迟要求较高[6]。可信执行环境则通过硬件隔离和远程证明，为关键聚合或本地训练过程提供受保护的运行空间，兼具较高效率与较强工程实用性[8]。

在实际系统中，上述技术往往不是替代关系，而是组合关系。例如，在资源受限的边缘节点上采用轻量级差分隐私和安全传输，在区域边缘或云侧聚合阶段利用安全多方计算或可信执行环境进一步提升保护强度，在检索与审计环节

配合密文检索和加密日志机制，能够形成分层、按需、可扩展的隐私保护体系。当前前沿趋势不再追求“单一技术解决所有问题”，而是强调隐私计算与零信任、容器隔离、策略编排协同工作。

2.5 密文检索与可信执行基础

联邦学习系统不仅存在训练任务，还存在模型查询、版本回溯、策略检索和审计核验等数据访问需求。如果这些查询完全以明文方式进行，攻击者可能绕过训练保护链路，通过索引、标签和检索日志反推出敏感信息。密文检索技术可以让用户在不暴露查询关键词和明文内容的情况下完成受控检索，对于跨域模型仓库、训练样本元数据管理和安全审计具有实际意义[7]。尤其是在需要按疾病类型、设备类型、场景标签查找模型片段或策略模板时，密文检索能够减少“查询过程泄密”这一常被忽略的风险点。

与此同时，可信执行环境、容器隔离、安全启动和远程证明等底层安全机制，为联邦学习的“可信执行”提供了系统支撑。若底层运行环境已经被篡改，即使上传过程使用了加密协议，仍然可能在本地训练前后阶段发生数据窃取、模型替换或参数伪造。因此，联邦学习安全研究必须将操作系统和硬件根信任纳入整体架构，而不能只停留在算法与协议层面。

综合来看，边缘-云联邦学习的安全基础可以概括为一个自上而下逐层耦合的技术体系：上层由业务场景决定数据边界和协作目标，中层由联邦训练机制承接协同闭环，下层则由隐私计算、零信任、可信执行与合规治理共同提供支撑。课程研究若只讨论其中某一项技术，往往难以解释系统为何在真实环境中仍会失效，因此必须把这些能力放在同一架构视角下理解。

表 2-1 联邦学习常用隐私保护技术比较

技术	主要保护对象	典型优势	主要局限
差分隐私	梯度、统计结果、发布模型	机制成熟，易于融入现有训练流程，适合大规模节点	噪声影响精度，隐私预算分配复杂
安全多方计算	聚合过程、跨机构联合计算	不依赖单点可信，安全边界明确	多轮交互多，通信与时延开销较大
同态加密	参数上传与密文运算	聚合方无需见明文即可完成计算	密文膨胀明显，算力需求高
可信执行环境	本地训练、边缘聚合、密钥保护	开销相对较低，工程落地性较强	依赖特定硬件，需防范侧信道与供应链风险

3 边缘-云协同联邦学习系统的安全威胁分析

3.1 威胁模型构建

为了分析系统风险，结合现有研究与工程实践，可将攻击者分为三类：恶意边缘节点、半可信云协调方和外部网络攻击者。恶意边缘节点可能控制训练数据、修改本地参数、伪造身份、上传异常梯度，也可能利用系统的容错机制长期潜伏；半可信云协调方通常按协议完成任务，但可能出于商业利益、分析目的或被入侵后，尝试从聚合视角获取参与者信息；外部网络攻击者则重点利用通信链路、接口暴露、凭据泄漏和配置错误发起窃听、篡改、重放或拒绝服务攻击。三类攻击者的存在说明，系统设计不能把风险简单归因于“外部黑客”，而应承认内部与外部边界均可能失效。

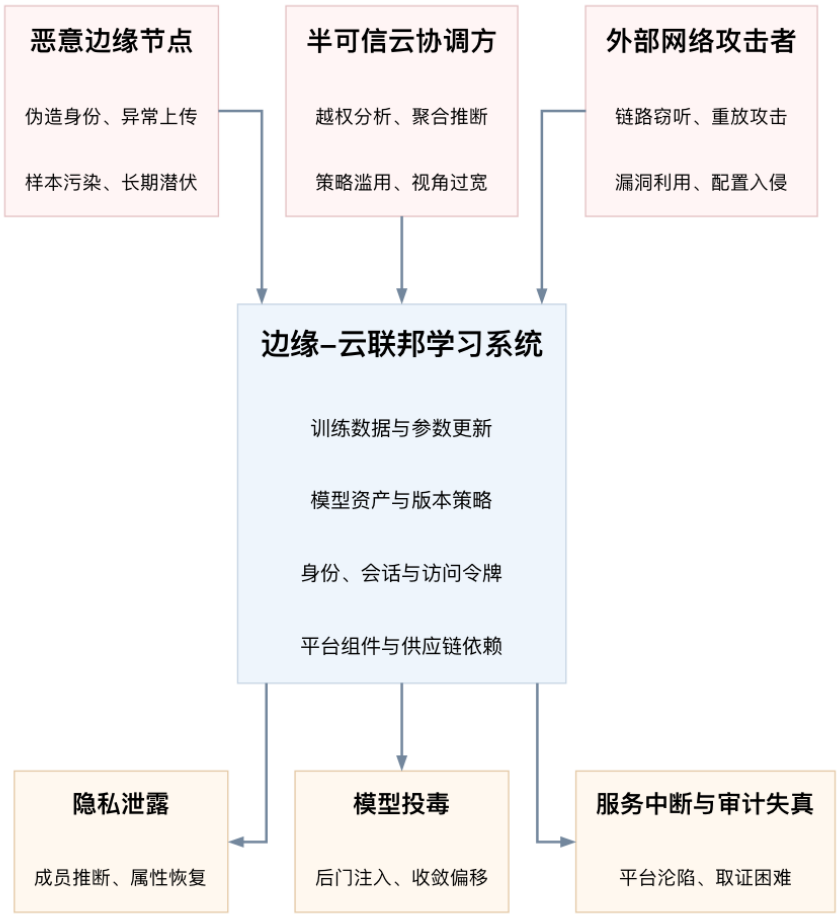

图 3-1 边缘-云联邦学习威胁模型

3.2 数据层与隐私层威胁

联邦学习虽然不直接交换原始数据，但上传的梯度、参数差分和统计量仍然携带与原始数据相关的信息。攻击者可利用梯度泄露、成员推断、属性推断等技术分析模型更新，从而推断某一条记录是否参与训练，甚至恢复部分输入特征[2,3]。这类风险在医疗、金融和车联网等高敏感场景尤为严重，因为即使只暴露少量样本特征，也可能造成身份识别、商业泄密或安全事故。

边缘-云协同架构还会带来额外的数据暴露面。边缘节点通常需要缓存中间模型、训练日志和临时密钥，若设备丢失、系统被提权或容器逃逸，攻击者可能在本地直接获取这些敏感对象。云侧如果对策略库、模型仓库和日志中心缺乏细粒度隔离，也会使“原始数据未出域”这一保护承诺被间接削弱。因此，数据层安全应从“只保护明文样本”扩展到“保护所有可推断出样本的信息载体”。

3.3 模型层攻击

模型层攻击主要包括投毒攻击、后门攻击、模型替换和收敛操纵。恶意节点可以通过污染本地数据或直接构造恶意更新，使全局模型在特定输入上输出攻击者期望结果，例如在医疗诊断中故意降低某类病灶识别率，在交通系统中误导异常路况识别。与传统单机学习相比，联邦学习难点在于攻击行为被分散到多个轮次和多个节点中，恶意更新可能伪装成正常异构数据导致的偏差，从而逃避简单的异常检测[2,3]。

此外，模型本身也是高价值资产。参与方获取全局模型后，可能未经授权将其用于其他任务，或者通过模型抽取和蒸馏手段复制核心能力。若系统仅强调“保护参与方数据”，却忽视“保护联合训练形成的模型资产”，将导致知识产权和供应链安全问题。对课程研究报告而言，这一问题也体现了信息安全技术研究正在从“保数据”扩展到“保数据、保模型、保流程、保责任”的综合治理阶段。

3.4 系统、网络与供应链威胁

在系统与网络层面，典型威胁包括中间人攻击、重放攻击、会话劫持、接口滥用、拒绝服务、容器逃逸和镜像污染。边缘节点数量巨大，升级节奏不一致，极易出现老旧组件、默认口令、弱证书和补丁滞后问题。一旦攻击者取得某个边缘网关、编排节点或镜像仓库的控制权，便可能扩散至整个训练链路。特别是在以容器和微服务为基础的联邦学习平台中，供应链安全成为不能忽视的议题。

供应链风险体现在模型训练代码、依赖库、预训练权重、容器基础镜像和硬件固件等多个层面。攻击者不必直接破解加密协议，只要在依赖包中植入后门、替换训练镜像或伪造更新组件，就可能在系统内部实现“合法形式下的恶意执行”。这说明边缘-云协同联邦学习的安全不是某个单独模块的问题，而是贯穿开发、部署、运行和运维全过程的系统工程问题。

3.5 信任崩塌与传统模型失效原因

传统边界安全模式默认内网可信、认证一次长期有效，而边缘-云协同联邦学习的实际运行环境恰恰表现为跨组织、跨地域、跨设备和跨生命周期的动态连接。一方面，参与节点可能随时加入、退出、离线或恢复；另一方面，同一个节点在不同时间、不同任务中风险水平并不相同。如果仍然使用静态白名单、固定凭据和粗粒度角色权限，就会出现“初始可信后长期放权”的问题，使异常行为有充足时间扩散。

因此，这类系统最关键的挑战不是“怎样建立永久信任”，而是“怎样用最小成本持续验证并及时收缩信任”。这也是零信任与隐私计算必须结合的原因所在：零信任解决主体和行为的可信判定问题，隐私计算解决即使主体不完全可信也尽量不暴露敏感内容的问题，二者共同构成新的安全底座。

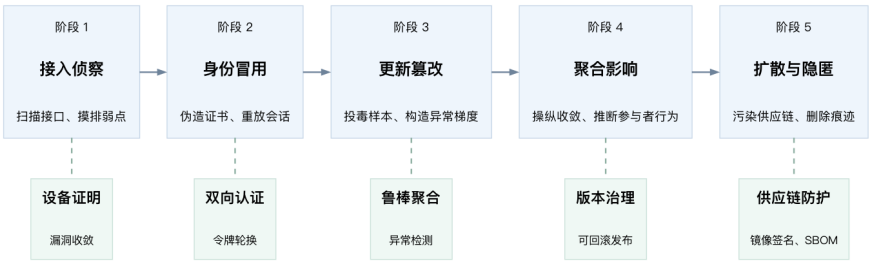

图 3-2 联邦学习攻击链与防御插入点

表 3-1 典型安全威胁与对应防护思路

威胁类型	典型表现	可能后果	优先防护思路
梯度泄露与推断攻击	通过参数更新推断样本属性或成员关系	个人隐私暴露、商业秘密泄漏	梯度裁剪、差分隐私、安全聚合、最小日志暴露
模型投毒与后门攻击	恶意节点上传异常更新或污染本地样本	全局模型失真、特定输入被定向操控	鲁棒聚合、异常检测、动态信任评分、节点隔离
链路窃听与会话劫持	传输过程被监听、篡改或重放	参数泄露、会话被冒用、训练轮次失真	双向认证、密钥轮换、传输加密、重放防护
供应链与平台越权	镜像污染、依赖包后门、管理员越权取数	平台整体沦陷、审计失真、模型资产丢失	镜像签名、最小权限、远程证明、不可抵赖审计

4 基于零信任与隐私计算的安全架构归纳

4.1 总体思路与架构概览

针对上述风险，现有研究中的典型做法可以归纳为一种“三层协同、内生安全、按需增强”的边缘-云联邦学习安全架构。该架构包括边缘训练层、云协调层和安全控制层。边缘训练层负责本地数据处理、模型训练与初步筛选；云协调层负责任务编排、安全聚合、全局模型分发和策略同步；安全控制层横跨前两层，提供身份认证、远程证明、动态授权、策略下发、密钥管理、审计留痕和风险处置等能力。其核心思想不是把安全机制附加在训练流程外部，而是让安全控制与训练过程同步发生，这与零信任架构和联邦学习系统化演进方向是一致的[3,10]。

在设计上，该架构遵循四项原则。第一，默认所有主体不可信，任何接入都需要身份与环境双重校验。第二，敏感数据和敏感参数尽量在源头完成最小暴露处理。第三，控制策略必须可动态调整，以适应节点风险、任务类型和法规要求变化。第四，所有关键动作都应具备可追溯性，便于合规审计、责任界定和事后复盘。

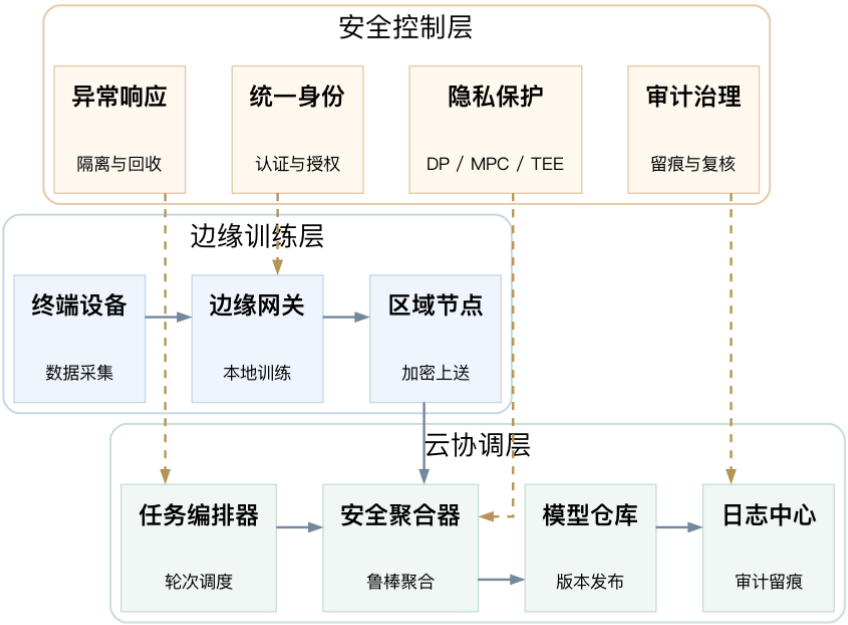

图 4-1 基于零信任与隐私计算的边缘-云联邦学习安全架构

4.2 基于零信任的身份认证与动态授权

在接入控制方面，系统为用户、设备、服务和任务建立统一身份体系。边缘节点在参与训练前，不仅需要提交传统证书或密钥材料，还应提供运行环境度

量、容器镜像摘要、硬件证明结果等上下文信息。云协调层根据这些信息形成初始信任评分，并结合节点的历史贡献质量、在线稳定性、异常行为记录和地理位置等因素进行持续更新。换言之，节点并非“通过一次认证就永久可信”，而是在每轮训练、每次上传和每次敏感操作前都处于被重新审视的状态。

动态授权机制可进一步将安全与效率结合起来。高信任节点可以参与完整训练流程，低信任节点则被限制访问模型片段、降低更新权重，甚至仅允许参与本地推理而不再上传参数。若某节点在短时间内出现更新分布异常、远程证明失败或会话行为异常，系统可以立即触发二次验证、密钥轮换和网络微隔离。这种以风险为中心的权限收缩能力，使联邦学习平台能够在不停止整体业务的情况下实现“局部隔离、全局连续”。

4.3 联邦学习中的分层隐私保护机制

为了应对数据推断和聚合分析风险，现有研究与工程实践通常采用分层式隐私保护策略。首先，在终端和边缘设备侧，对本地梯度进行裁剪和差分隐私噪声注入，限制单个样本对最终更新的影响；其次，在区域边缘或云侧聚合阶段，通过安全多方计算或安全聚合协议确保聚合方只能得到统计结果而非单个节点明文更新；再次，在高敏感任务中，对关键参数采用同态加密或可信执行环境保护，从而避免聚合过程中出现“中间明文暴露”[4,5,6,8]。这三层机制并不要求在所有场景同时满配，而是可根据任务敏感度进行组合编排。

分层保护的另一个重要意义在于平衡性能。若对全部训练流程使用重型密码学技术，边缘节点往往难以承受算力和时延开销，系统也难以扩展到大规模场景。因此，设计重点应放在“在哪一层使用何种强度的保护最划算”。例如，普通环境感知任务可采用传输加密加差分隐私，医疗协同诊断可进一步启用安全聚合和可信执行环境，涉及高价值模型资产或跨境协作时再引入更强的密文计算与审计约束。安全架构只有体现出可分层、可配置、可量化的能力，才可能真正落地。



图 4-2 单轮安全训练流程

4.4 密文检索、模型治理与安全审计

联邦学习平台通常需要维护模型仓库、样本标签、任务策略、风险规则和审计记录。若这些对象在检索时以明文索引和明文日志暴露，就可能形成新的侧信道。因此，在这类安全架构中也应将密文检索纳入能力范围，用于保护模型片段查询、策略模板检索和审计条件检索过程。通过将关键词、标签或向量索引进行加密处理，平台可以在不泄露检索意图的情况下返回授权结果，从而减少攻击者基于访问模式推测业务敏感信息的可能性[7]。

与此同时，模型治理与审计必须成为架构的常规能力，而非事后补丁。每一次训练任务的发起、每一轮模型上传、每一次异常剔除、每一版全局模型发布，都应生成可校验、不可随意篡改的日志记录。结合时间戳、签名、策略版本和节点标识，管理者可以回溯模型为何产生偏差、哪一类节点贡献了异常更新、某次权限提升是否符合审批流程。对于课程研究报告而言，这一部分也是体现“工程应用”和“合规治理”融合的重要抓手。

4.5 操作系统层与可信执行支撑

从底层实现看，容器隔离、虚拟化、可信执行环境、安全启动和镜像签名构成了联邦学习安全架构的根部支撑。边缘节点可通过容器编排实现训练任务、密钥服务、监控代理彼此隔离，防止单一进程被攻陷后横向移动；关键聚合逻辑和密钥处理可放入可信执行环境中运行，并借助远程证明向控制平面证明自身处于未被篡改的状态；设备启动链若加入固件校验和完整性度量，则可从更早阶段阻断恶意代码进入训练环境[8,10]。

需要强调的是，可信执行环境也不是“绝对安全”的银弹。它仍可能面临侧信道、微架构漏洞和供应链风险，因此必须与补丁管理、运行时监测和策略审计共同使用。只有当底层运行环境能够被持续验证，上层零信任决策与隐私计算协议才不会建立在脆弱基础之上。

4.6 架构优势与现实局限

该架构的主要优势在于：其一，能够把身份可信、数据保护、模型防护和审计追踪贯通起来，避免“单点安全、整体失守”的问题；其二，能够根据场景差异按需启用保护能力，使安全策略与业务敏感度匹配；其三，能够较好地适应边缘节点动态加入和退出的现实特点；其四，便于与现行法律法规和行业标准对接，形成可解释的治理闭环。

当然，该架构也存在现实局限。首先，多层安全机制会引入额外通信、加密、验证和日志成本，特别是在资源受限终端中可能影响训练效率。其次，动态信任评分与异常检测本身也需要较高质量的数据支持，若策略设计不当，可能对弱网环境或异构节点产生误判。再次，跨机构场景下的责任划分、策略共享和密钥

托管仍具有组织协调难度。因此，安全架构设计既要追求技术先进性，也必须尊重工程复杂性和治理边界。

表 4-1 安全架构中各层能力映射

架构层次	核心对象	关键能力	主要安全目标
边缘训练层	终端、网关、边缘服务器	本地训练、梯度裁剪、可信启动、容器隔离	保证数据不出域，降低源头泄露和设备入侵风险
云协调层	聚合器、模型仓库、编排服务	安全聚合、模型版本管理、策略同步、异常剔除	保证全局模型完整性与聚合可验证性
安全控制层	身份、策略、审计、密钥服务	动态授权、远程证明、密钥轮换、审计追踪	保证主体可信、过程可控、责任可追溯

5 工程应用与实践分析

5.1 医疗数据共享场景

医疗场景是联邦学习最具代表性的高敏感应用之一。不同医院之间往往拥有相同类型但彼此隔离的诊疗数据，既存在联合训练价值，又面临严格的隐私和伦理约束。采用边缘-云协同联邦学习后，各医院可在本地服务器或院内边缘节点完成模型训练，仅将处理后的更新上传至区域医疗平台或云端协调中心。通过差分隐私、安全聚合和可信执行环境的组合，可以在较大程度上降低病历、影像和检验数据泄露风险。

这一场景的工程难点主要体现在数据标准不统一、节点网络环境复杂和模型收敛受非独立同分布影响明显。安全架构在这里不仅要“防攻击”，还要通过统一身份、模型版本控制和细粒度审计保证协作过程可持续。换言之，医疗联邦学习平台若不能让参与医院清楚知道“谁在何时以何种权限使用了什么模型与策略”，就很难形成长期稳定的数据合作机制。

5.2 车联网与智能交通场景

车联网与智能交通系统需要处理道路状态、车辆运行、传感器感知和行为预测等连续数据，对低时延和高可靠性要求极高。边缘节点通常部署在路侧基础设施、车载网关和区域交通控制中心，云端负责全局模型优化与协同调度。在此场景中，联邦学习能够减少大规模原始视频和传感器数据的集中回传压力，但同时也暴露出更多身份伪造、通信干扰、设备被劫持和恶意模型注入风险。

因此，车联网场景尤需强化零信任接入和任务级授权。一辆车、一个路侧单元、一个交通分中心都不能因为属于“内部交通网络”就被天然信任，而应通过设备证书、位置上下文、软件版本和运行状态综合判定。对于时延敏感任务，可

将大部分隐私保护能力前置到边缘侧，减少云侧重加密计算造成的实时性损失；对于安全事故追溯，又必须保留足够的审计证据，以支持责任判定和应急处置。

5.3 工业互联网与智慧城市场景

工业互联网中的设备预测维护、质量控制与能耗优化，是边缘-云联邦学习的另一重要应用方向。工厂现场存在大量异构传感器、控制终端和工业协议，数据具有较强业务私密性。企业通常不愿直接共享产线数据，但希望通过协同建模提高故障预测与质量检测效果。此时，联邦学习可在保障商业边界的前提下实现经验共享，而零信任与隐私计算则用于限制设备越权、保护工艺参数和防止模型被竞争对手推断。

在智慧城市层面，边缘-云协同还会涉及安防感知、环境监测、能源调度和公共服务优化等任务。系统连接对象从单一机构扩展到政府、企业与公众终端，其社会影响面更广，对透明度、公平性和审计性的要求更高。因此，工程实践不仅要关注“模型是否更准”，还要关注“系统是否可解释、是否可申诉、是否避免过度监控”。这也说明信息安全新技术研究需要走出单纯的技术性能比较，进入更完整的治理框架。

如果从部署结构抽象来看，上述场景虽然业务边界不同，但大体都可以归纳为“数据源位于现场或机构内部、训练能力尽量下沉到边缘、全局协调与审计能力保留在云侧”这一三层形态。不同之处主要在于哪一层承担更高的实时性压力、哪一层承担更强的合规与治理要求，以及节点之间的信任边界如何划分。

5.4 性能、安全与合规的权衡

在所有工程场景中，性能、安全和合规三者之间都存在显著张力。安全强度越高，往往意味着更多加密、验证和日志开销；合规要求越严格，系统在数据使用、跨域流转和保留周期上的自由度就越低；性能指标越苛刻，越需要对安全策略做场景化裁剪。真正可用的工程方案并不是“把所有最强的安全技术叠加”，而是建立一套能够根据任务等级、数据敏感度和节点能力自动调整的策略体系。

例如，可以把训练任务划分为普通、重要和关键三级，对应不同的接入门槛、隐私预算、聚合保护强度和审计保留要求。这样既能避免资源浪费，也能使安全投入更符合风险实际。课程研究报告若要体现工程思维，就必须展示这种权衡，而不是停留在“技术越多越安全”的表面判断。

若进一步将工程决策抽象为“数据敏感度”和“时延容忍度”两个维度，就可以更清楚地解释为什么不同场景会选择不同的保护组合。高敏感且低时延容忍的任务更适合在边缘侧采用可信执行环境、严格准入和快速聚合；高敏感但对时延要求相对宽松的任务则可以引入更强的密文计算和长周期审计；低敏感场景则可将重点放在链路保护、轻量认证和稳定运行上。这种分级配置思路，比简单堆叠全部安全技术更符合学术研究和工程实践的共同逻辑。

表 5-1 典型应用场景的安全需求对比

场景	主要数据类型	首要约束	更关注的安全能力	工程侧重点
医疗协同诊断	病历、影像、检验结果	隐私与伦理合规	差分隐私、安全聚合、细粒度审计	数据标准统一与长期协作机制
车联网与交通控制	感知数据、轨迹、道路状态	低时延与高可靠	动态身份认证、链路保护、异常节点隔离	实时性和事故可追溯性并重
工业互联网	工艺参数、设备日志、质检结果	商业保密与稳定运行	容器隔离、可信执行、供应链防护	异构设备接入与生产连续性保障

6 技术标准、法律法规与治理要求

6.1 国际标准与技术规范

从国际标准看，NIST SP 800-207 将零信任架构明确为持续认证、最小权限和策略决策驱动的安全模型，为联邦学习这类多主体分布式系统提供了重要设计依据[10]。ISO/IEC 27001:2022 则从信息安全管理体系角度提出组织层面的风险管理、控制措施和持续改进要求，说明仅有技术协议并不足以支撑长期安全运行，仍需配套制度、流程和审计机制[11]。对于跨国数据协作场景，GDPR 所强调的合法性、最小必要、目的限定和数据主体权利，也对联邦学习平台的架构设计形成直接约束[15]。

这些标准与规范共同表明，未来的信息安全技术不再只是“把系统守住”，更是要“证明系统在持续受控状态下运行”。对边缘-云联邦学习而言，这意味着模型训练、参数传输、权限分配、日志审计都应具有被验证、被记录和被解释的能力。

6.2 国内法律法规要求

我国在数据安全与网络治理领域已形成较为完整的法规体系。《中华人民共和国网络安全法》强调网络运行安全、关键信息基础设施保护和个人信息保护义务；《中华人民共和国数据安全法》从数据分类分级、风险监测和数据活动安全义务等方面提出要求；《中华人民共和国个人信息保护法》则围绕个人信息处理的合法性、必要性、透明性和个人权利保护建立了基础规则[12,13,14]。对于涉及人工智能服务的场景，《生成式人工智能服务管理暂行办法》进一步体现出算法安全、内容治理和主体责任的监管趋势[17]。

如果联邦学习平台面向医疗、交通、金融等重点行业，还需考虑行业主管部门的专项规范和本地化监管要求。也就是说，联邦学习虽然减少了原始数据集

中汇聚，但并不意味着天然满足所有合规要求。平台仍需证明其数据处理目的明确、权限边界清晰、审计记录完整、异常处置及时，并能够支持监管检查与责任追踪[12,13,14,16,17]。

6.3 合规落地建议

从工程实践角度，合规落地至少应包含四个方面。第一，建立数据和模型的分类分级体系，明确哪些任务允许跨机构协作、哪些参数必须加密保存、哪些日志需要长期留存。第二，将统一身份、策略审批、权限变更和审计分析纳入日常运维流程，而不是只在上线前做一次检查。第三，对跨域协作建立合同、责任和应急处置机制，避免技术平台运行后出现“安全责任悬空”。第四，对平台引入周期性评估，包括算法偏差、隐私预算消耗、异常节点处置效果和法规符合性复核。

若从要求来源进一步归纳，可以发现国际标准更强调持续认证、风险管理和可验证审计，国内法律法规更强调分类分级、最小必要、用途限定和责任追踪，而行业规范则要求把这些原则转化为具体任务模板、审批链路和留痕机制。换言之，合规并不是在系统上线前准备一套文档，而是要把“标准要求”持续翻译成“平台动作”和“运维制度”，使训练发起、节点接入、参数上传、模型发布和异常处置都处于可解释、可检查、可回溯的状态。

也正因为如此，联邦学习平台的治理不应停留在静态制度层面，而应形成覆盖训练前、训练中和训练后的连续闭环。只有把分类分级、策略审批、安全训练、合规评估与审计复盘串联起来，平台才能真正把标准和法规要求落实到日常运行过程之中。



图 6-2 联邦学习安全治理闭环

7 对社会可持续发展的影响

7.1 正面影响

从积极面看，边缘-云协同联邦学习安全架构有助于在保护隐私和商业边界的同时释放数据价值，推动医疗资源共享、城市治理优化、工业节能增效和公共服务智能化。相比粗放式数据汇聚模式，这种技术路径更符合“数据最小化使用”和“按需共享”的可持续发展理念，也有助于减少因过度集中存储带来的系统性泄露风险。

7.2 潜在风险

但技术进步也可能带来新的社会风险。第一，若平台治理不足，联邦学习可能被包装成“天然安全”，从而掩盖模型推断、身份滥用和算法歧视问题。第二，零信任和持续审计若被过度扩张，可能演化为过度监控，压缩个人和组织的合理自治空间。第三，复杂安全架构提高了中小机构的技术门槛，若缺乏公共基础设施支持，可能进一步扩大数字能力差距。

7.3 可持续发展路径

因此，可持续发展的关键不在于简单追求“更强的监控”或“更多的数据”，而在于推动技术、制度和伦理协同演进。一方面，要持续提升隐私计算效率、可信硬件可用性和安全策略自动化水平；另一方面，要建立透明的治理机制、明确的数据权责边界和可申诉的审计制度。只有当技术进步与社会信任同步增长，边缘-云联邦学习的安全架构才可能真正成为可持续的信息基础设施。

8 总结与展望

本报告围绕边缘-云协同的联邦学习安全架构进行了系统分析，从边缘计算、联邦学习、零信任、隐私计算、密文检索和可信执行等关键技术出发，梳理了数据层、模型层、系统层和治理层的主要风险，并归纳了三层协同、按需增强、持续验证的融合安全架构，同时结合医疗、车联网和工业互联网场景分析了工程可行性与现实约束。

总体来看，联邦学习的未来发展不应停留在“把数据留在本地”的单一目标上，而应迈向“主体可信、过程可控、结果可审、责任可追”的综合安全体系。随着大模型、边缘智能、机密计算和数据治理制度进一步成熟，边缘-云协同联邦学习将成为信息安全技术与人工智能交叉领域的重要研究方向。后续研究仍需继续解决高强度保护下的性能优化、跨域治理协同和可信评估自动化问题，以推动这一新技术真正走向大规模、长期、稳定的工程应用。

参考文献

[1] McMahan B, Moore E, Ramage D, Hampson S, Aguera y Arcas B. Communication-Efficient Learning of Deep Networks from Decentralized Data[C]. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017: 1273–1282.
[2] Li T, Sahu A K, Talwalkar A, Smith V. Federated Learning: Challenges, Methods, and Future Directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50–60.
[3] Kairouz P, McMahan H B, Avent B, et al. Advances and Open Problems in Federated Learning[J]. Foundations and Trends in Machine Learning, 2021, 14(1–2): 1–210.
[4] Bonawitz K, Ivanov V, Kreuter B, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning[C]. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017: 1175–1191.
[5] Dwork C, McSherry F, Nissim K, Smith A. Calibrating Noise to Sensitivity in Private Data Analysis[C]. Theory of Cryptography Conference, 2006: 265–284.
[6] Gentry C. Fully Homomorphic Encryption Using Ideal Lattices[C]. Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009: 169–178.
[7] Song D X, Wagner D, Perrig A. Practical Techniques for Searches on Encrypted Data[C]. Proceedings of the 2000 IEEE Symposium on Security and Privacy, 2000: 44–55.
[8] Costan V, Devadas S. Intel SGX Explained[EB/OL]. https://eprint.iacr.org/2016/086
[9] Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge Computing: Vision and Challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637–646.
[10] NIST. Zero Trust Architecture: NIST Special Publication 800-207[EB/OL]. https://csrc.nist.gov/pubs/sp/800/207/final
[11] ISO. ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection – Information Security Management Systems – Requirements[EB/OL]. https://www.iso.org/standard/27001
[12] 全国人民代表大会常务委员会. 中华人民共和国网络安全法 [EB/OL]. https://flk.npc.gov.cn/detail?id=021e7d7684474107b8f3febbb1c4f8b5
[13] 国家互联网信息办公室. 中华人民共和国个人信息保护法 [EB/OL]. https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm
[14] 国家互联网信息办公室. 中华人民共和国数据安全法 [EB/OL]. https://www.cac.gov.cn/2021-06/11/c_1624994566919140.htm
[15] European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016[EB/OL]. https://eur-lex.europa.eu/eli/reg/2016/679/oj
[16] 工业和信息化部, 中央网信办, 教育部, 国家卫生健康委, 中国人民银行, 国务院国资委. 工业和信息化部等六部门关于印发《算力基础设施高质量发展行动计划》的通知 [EB/OL]. https://www.gov.cn/zhengce/zhengceku/202310/content_6907900.htm
[17] 国家互联网信息办公室, 国家发展改革委, 教育部, 科技部, 工业和信息化部, 公安部, 广播电视总局. 生成式人工智能服务管理暂行办法 [EB/OL]. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm