Computer Science and Application 计算机科学与应用, 2026, 16(4), 186-194
Published Online April 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.164121

Hans汉斯

面向弱信任环境的量子安全联邦学习协议

何芯雨，岳笑含

沈阳工业大学信息科学与工程学院，辽宁 沈阳

收稿日期：2026年3月9日；录用日期：2026年4月10日；发布日期：2026年4月20日

摘　要

联邦学习虽具“数据不出域”优势，但仍面临梯度泄露、身份关联、串通攻击及量子计算威胁，难以兼顾抗量子、抗串通与不可关联性。为此，本文提出了一种支持抗量子与抗串通的隐私联邦学习协议，旨在为跨机构弱信任环境下的模型训练与数据协作提供安全保障。该方案基于RLWE构建后量子安全的同态加密聚合机制，实现梯度机密性保护；通过加法秘密共享与双混洗服务器设计，实现抗串通安全；结合混洗与虚拟客户端机制，实现身份与梯度的不可关联性。该协议在保障模型效用的同时，实现了后量子安全、抗串通性与匿名性的统一，兼顾安全性与系统性能，提升了协议在复杂现实场景中的可部署性与长期安全稳定性，增强了系统整体的可信性与工程应用价值。

关键词

联邦学习，量子安全，抗串通，不可关联性

Post-Quantum Secure Federated Learning Protocol for Weak-Trust Environments

Xinyu He, Xiaohan Yue

School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning

Received: March 9, 2026; accepted: April 10, 2026; published: April 20, 2026

Abstract

Although federated learning has the advantage of “data not leaving the domain”, it still faces issues such as gradient leakage, identity association, collusion attacks, and quantum computing threats, making it difficult to balance anti-quantum, anti-collusion, and non-correlation. Therefore, this paper proposes a privacy federated learning protocol that supports anti-quantum and anti-collusion, aiming to provide security guarantees for model training and data collaboration in cross-institutional weak-trust environments. This scheme is based on RLWE to build a post-quantum secure

文章引用: 何芯雨, 岳笑含. 面向弱信任环境的量子安全联邦学习协议[J]. 计算机科学与应用, 2026, 16(4): 186-194.
DOI: 10.12677/csa.2026.164121

homomorphic encryption aggregation mechanism, achieving gradient confidentiality protection; through addition secret sharing and double mixing server design, achieving anti-collusion security; combining mixing and virtual client mechanisms, achieving non-correlation between identity and gradient. This protocol ensures the model utility while achieving the unification of post-quantum security, anti-collusion, and anonymity, balancing security and system performance. This improves the deployability and long-term security stability of the protocol in complex real-world scenarios, enhancing the overall reliability and engineering application value of the system.

Keywords

Federated Learning, Post-Quantum Security, Collusion Resistance, Non-Correlation

1. 引言

在数字经济飞速发展的当下，数据已成为驱动科技创新与产业升级的核心生产要素，而联邦学习凭借“数据不出域”的独特优势，成功破解了隐私保护与数据协同的矛盾，在医疗健康、金融风控、工业物联网等隐私敏感领域展现出广阔应用前景[1]。在医疗领域，它支持多机构跨区域协同构建医学影像诊断模型，显著提升疾病检出精度；在金融场景[2]中，其助力跨机构风险建模与欺诈检测，实现合规前提下的高效协作；在工业物联网环境[3]下，通过整合分布式终端数据，有效降低通信开销并提升系统实时响应能力。同时，在智慧政务、智能交通与个性化推荐等新兴场景中，联邦学习也为跨主体数据融合提供了安全可行的技术路径，逐步成为数据要素市场化配置的重要支撑技术。

然而，联邦学习的隐私安全防护体系仍面临多重严峻挑战。梯度作为模型更新的核心载体，隐含大量原始数据语义信息，易遭受重构攻击与属性推理攻击[4]，导致隐私泄露；多轮训练中，客户端的上传行为、通信模式与梯度特征形成可追溯关联，攻击者可借此实现跨轮次身份追踪；现有安全聚合方案多依赖单一可信实体，中心服务器与部分节点串通时极易引发隐私破坏[5]；同时，量子计算技术的快速发展使传统密码机制面临失效风险，而现有方案难以系统性融合抗量子安全性[6]、抗串通性与不可关联性，无法满足复杂场景的安全需求[7]。此外，在跨域协同环境下，各参与方之间缺乏完全信任关系，攻击面更加复杂，传统“诚实但好奇”假设难以覆盖现实威胁模型，进一步提升了系统设计难度。

这些问题的存在，严重制约了联邦学习在跨机构、弱信任环境中的规模化应用。因此，构建一套兼具抗量子安全、抗串通能力与不可关联性的隐私联邦学习体系，成为当前隐私计算领域的迫切需求。本文聚焦上述核心挑战，提出一种支持量子安全及抗串通的隐私联邦学习协议，通过融合先进密码学技术与优化聚合机制，在保障模型训练效果的同时，实现梯度隐私、身份安全与量子抗性的全方位防护，为联邦学习在高安全需求场景的落地提供技术支撑，推动隐私计算技术的创新发展与产业应用。

2. 基础知识

2.1. 环学习带误差问题

环学习带误差(Ring Learning With Errors, RLWE)问题[8]是格密码体系中的核心计算困难问题之一，其安全性可归约至理想格上若干经典最坏情况问题的困难性，被广泛认为在经典与量子计算模型下均具有较强的安全性保证。因此，RLWE 判定性问题已成为构建后量子安全密码方案(如同态加密、安全聚合

协议等)的重要理论基础。

设 $R_q=\mathbb{Z}_q[x]/\langle f(x)\rangle$，其中 $f(x)$ 为次数为 n 的不可约多项式，q 为模数。令 χ 表示定义在 R_q 上的误差分布(通常选取离散高斯分布或其变体)。随机采样 $\boldsymbol{s}\xleftarrow{\$}R_q,\boldsymbol{s}\xleftarrow{\$}R_q,\boldsymbol{e}\xleftarrow{\$}\chi$，并定义 RLWE 样本对为 $(\boldsymbol{a},\boldsymbol{b}=\boldsymbol{a}\cdot\boldsymbol{s}+\boldsymbol{e}\bmod q)$。RLWE 判定性问题要求攻击者在给定若干样本对的情况下，判断其是否来自如下两种分布之一：RLWE 分布 $\mathcal{D}_{\mathrm{RLWE}}=\left\{(\boldsymbol{a},\boldsymbol{a}\cdot\boldsymbol{s}+\boldsymbol{e}\bmod q)|\boldsymbol{s}\xleftarrow{\$}R_q,\boldsymbol{s}\xleftarrow{\$}R_q,\boldsymbol{e}\xleftarrow{\$}\chi\right\}$；均匀随机分布 $\mathcal{D}_U=\left\{(\boldsymbol{a},\boldsymbol{u})|\boldsymbol{a},\boldsymbol{u}\xleftarrow{\$}R_q\right\}$。攻击者的目标是区分给定样本对来自 $\mathcal{D}_{\mathrm{RLWE}}$ 还是 $\mathcal{D}_U$。

2.2. 加法秘密共享

加法秘密共享[9] (Additive Secret Sharing, ASS)是一类经典的秘密共享方法，广泛应用于安全多方计算、隐私保护协议以及分布式密码系统中。其基本思想是将一个秘密值拆分为若干个随机份额，使得单个或不足阈值数量的参与方无法获得关于秘密的任何有效信息，而所有份额在加法意义下可以唯一恢复原始秘密。设秘密 $\boldsymbol{s}\in\mathbb{Z}_q$，加法秘密共享方案 Σ_{ASS} 由以下两个算法组成：

$\Sigma_{\mathrm{ASS}}=\left(\Sigma_{\mathrm{ASS}}.\mathrm{Share}(\cdot),\Sigma_{\mathrm{ASS}}.\mathrm{Reconstruct}(\cdot)\right)$。

$\Sigma_{\mathrm{Ass}}.\mathrm{Share}(\boldsymbol{s})\to(\boldsymbol{s}_1,\boldsymbol{s}_2)$：秘密共享算法以秘密值 $\boldsymbol{s}\in\mathbb{Z}_q$ 为输入，随机选取 $\boldsymbol{s}_1\xleftarrow{\$}\mathbb{Z}_q$，并计算 $\boldsymbol{s}_2=(\boldsymbol{s}-\boldsymbol{s}_1)\bmod q$，输出秘密份额对 $(\boldsymbol{s}_1,\boldsymbol{s}_2)$。

$\Sigma_{\mathrm{ASS}.}.\mathrm{Reconstruct}(\boldsymbol{s}_1,\boldsymbol{s}_2)\to\boldsymbol{s}$ 秘密重构算法以秘密份额 $(\boldsymbol{s}_1,\boldsymbol{s}_2)$ 为输入，计算 $\boldsymbol{s}=(\boldsymbol{s}_1+\boldsymbol{s}_2)\bmod q$，并输出秘密 $\boldsymbol{s}$。

2.3. 多重互斥综合译码问题

多重互斥综合译码问题[10] (Multi-Disjoint Syndrome Decoding，简称 MDSD)是综合译码问题(Syndrome Decoding, SD)的计算困难推广形式，是描述“拆分–混合”结构计算安全性的核心假设。该问题在本方案的联邦学习与不可关联性聚合中为基于混洗与份额重排的安全聚合协议提供计算安全保证。

定义 1 (MDSD 分布)：设 $\mathbb{F}$ 为有限域，n,m,c,w 为正整数，定义 (n,m,c,w) -多不相交综合征解码(MDSD)分布为 $\mathrm{MDSD}_{n,m,c,w}=\left\{(\mathbf{H},\mathbf{H}\cdot\mathbf{E})\,|\,\mathbf{H}\xleftarrow{\$}\mathbb{F}_q^{n\times m},\mathbf{E}\xleftarrow{\$}\mathrm{DisError}_{n,m,c,w}\right\}$，其中 $\mathbf{H}$ 均匀随机选自所有 $n\times m$ 矩阵，$\mathbf{E}$ 均匀随机选自不相交错误集 $\mathrm{DisError}_{m,c,w}$，且 $\mathbf{H}\cdot\mathbf{E}$ 表示矩阵乘法。给定参数 n,m,c,w，MDSD 问题要求区分 $\mathrm{MDSD}_{n,m,c,w}$ 与均匀随机分布 $\mathbb{F}^{n\times m}\times\mathbb{F}^{n\times c}$。即对于任何多项式时间敌手，其区分优势为关于安全参数 λ 的可忽略函数。

定义 2 (判定型 MDSD)给定参数 n,m,c,w，MDSD 问题要求区分 $\mathrm{MDSD}_{n,m,c,w}$ 分布与均匀随机分布 $\mathbb{F}^{n\times m}\times\mathbb{F}^{n\times c}$。即对于任何多项式时间敌手，$\mathbf{H}\leftarrow\mathbb{F}^{n\times m},\mathbf{E}\leftarrow\mathrm{DisError}_{n,m,c,w},\boldsymbol{Y}\leftarrow\mathbb{F}^{n\times c}$ 其区分优势 $\left|Pr\left[\mathcal{A}(\mathbf{H},\mathbf{H}\cdot\mathbf{E})=1\right]-Pr\left[\mathcal{A}(\mathbf{H},\mathbf{Y})=1\right]\right|\le\mathrm{negl}(\lambda)$，其区分优势定义为其中 $\mathrm{negl}(\lambda)$ 为关于安全参数 λ 的可忽略函数。

3. 威胁模型

在跨机构弱信任环境下，各参与方之间缺乏完全信任关系，因此有必要对系统的威胁模型进行形式化定义。本文采用半诚实(honest-but-curious)敌手模型并考虑有限串通场景。系统参与实体包括客户端、中心服务器以及两台独立部署的混洗服务器。各参与方在协议执行过程中按照既定流程完成计算与通信，但可能通过分析接收到的中间数据或通信信息以推断额外隐私。攻击者能够监听客户端与服务器之间的通信信道，并获取加密的模型更新信息，通过梯度分析、统计推断或行为模式分析等方式尝试恢复客户端训练数据或关联客户端身份。此外，本文允许存在有限的实体串通，例如中心服务器可能与至多一台混洗服务器发生合谋，但假设两台混洗服务器不会同时与服务器串通。

在该威胁模型下，攻击者的主要目标包括推断客户端本地训练数据、恢复单个客户端的梯度更新以及通过多轮训练过程建立客户端身份关联关系。需要指出的是，当系统面对更强的恶意敌手(malicious adversary)时，即部分参与方可能主动偏离协议流程或发送伪造数据，协议仍可能存在潜在风险。例如，恶意客户端可能上传构造的梯度更新以影响全局模型训练结果，服务器或混洗服务器也可能通过篡改、丢弃或重排消息破坏协议执行过程。为进一步增强系统鲁棒性，可以在协议扩展中引入可验证机制，例如利用零知识证明验证客户端梯度计算的正确性，或采用可验证计算与可验证混洗技术保证服务器执行过程的合法性，从而在更强敌手模型下提升协议的安全性与可靠性。

4. 方案构建

4.1. 方案系统模型

本文的系统模型如图 1 所示，核心由三类参与实体构成，协同完成隐私联邦学习，整体采用分层保护与安全聚合相结合的设计思路，流程如下。

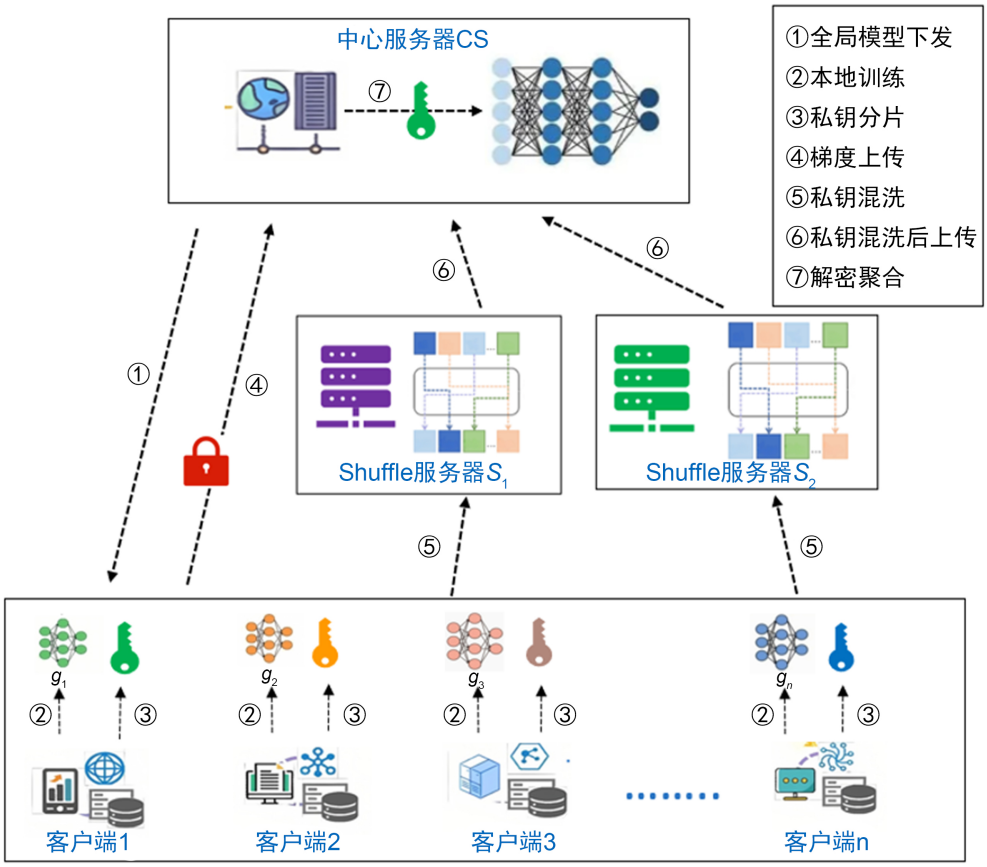

Figure 1. System model
图 1. 系统模型图

客户端：持有本地数据集，完成本地模型训练生成梯度；对梯度执行“量化-IntCRT-PolySubR”打包与 RLWE 加密后上传至中心服务器；将自身私钥通过加法秘密共享拆分为两份份额，分别发送至两台混洗服务器，从源头上避免单点泄露风险。

中心服务器：初始化系统公共参数与全局模型一并下发；接收客户端加密梯度与混洗服务器的混洗

结果；重构聚合私钥，对密文执行聚合与解密，恢复全局聚合梯度以更新模型，并进入下一轮迭代训练。

双混洗服务器(S_1、S_2)：独立接收客户端私钥份额；拆分份额并与虚拟客户端份额混合，通过 MDSD 机制随机重排；将混洗后的份额矩阵发送至中心服务器，切断身份与梯度的关联，确保多轮训练过程中的不可关联性。

4.2. 方案形式化定义

下面给出本文所提支持量子安全及抗串通的隐私联邦学习协议的形式化定义。设安全参数为 λ，客户端集合为 $\mathcal{C}=\{C_1,\cdots,C_c\}$，中心服务器为 CS，两台混洗服务器为 S_1,S_2。

(1) 初始化阶段

$\mathrm{Setup}\left(1^{\lambda}\right)\to pp$：由中心服务器执行。输入安全参数 λ，生成系统公共参数 pp，包括 RLWE 参数 (q,n,χ)、多项式环 R_q、公共随机多项式 $\boldsymbol{a}$、梯度量化界 B、缩放因子 Δ、CRT 模数组及最大客户端数量上界等。输出公共参数并广播给所有参与方。

(2) 全局模型下发阶段

$\mathrm{GlobalModelInit}\left(pp\right)\to \mathbf{w}^{(t)}$：由中心服务器执行。根据任务初始化或更新全局模型参数 $\mathbf{w}^{(t)}$，并将其广播给被选中的客户端。

(3) 客户端本地训练阶段

$\mathrm{LocalTrain}\left(\mathbf{w}^{(t)},D_i\right)\to \boldsymbol{g}_i^{(t)}$：客户端 C_i 在本地数据集 D_i 上进行若干轮梯度下降训练，输出本地梯度向量 $\boldsymbol{g}_i^{(t)}$。

(4) 梯度打包阶段

$\mathrm{Encode}\left(\boldsymbol{g}_i^{(t)},pp\right)\to \boldsymbol{m}_i$：客户端首先对梯度 $\boldsymbol{g}_i^{(t)}$ 执行量化操作(Quant)，将浮点梯度映射为有界整数；随后利用 CRT 进行整数层打包(IntCRT)；最后基于环同构结构执行多项式层嵌入(PolySubR)，得到 RLWE 明文多项式 $\boldsymbol{m}_i$。

(5) 加密与私钥拆分阶段

$\mathrm{KeyGen}\left(\mathrm{pp}\right)\to \boldsymbol{sk}_i$：客户端生成短多项式私钥 $\boldsymbol{sk}_i$。$\mathrm{Enc}\left(\boldsymbol{m}_i,\boldsymbol{sk}_i,\boldsymbol{a}\right)\to \boldsymbol{ct}_i$ 基于 RLWE 机制生成密文 $\boldsymbol{ct}_i$ 并发送至中心服务器。同时调用第 2.2 节介绍的加法秘密共享算法，客户端执行秘密共享算法 $\Sigma_{\mathrm{ASS}}.\mathrm{Share}\left(\boldsymbol{sk}_i\right)\to\left(\boldsymbol{s}_{i,1},\boldsymbol{s}_{i,2}\right)$，即随机选取 $\boldsymbol{s}_{i,1}\xleftarrow{\$}\mathbb{Z}_q$ 并计算 $\boldsymbol{s}_{i,2}=\left(\boldsymbol{sk}_i-\boldsymbol{s}_{i,1}\right)\bmod q$ 从而得到两个私钥份额 $\boldsymbol{sk}_i=\left(\boldsymbol{s}_{i,1}+\boldsymbol{s}_{i,2}\right)\bmod q$ 随后客户端将份额 $\boldsymbol{s}_{i,1}$ 发送至混洗服务器 S_1，将份额 $\boldsymbol{s}_{i,2}$ 发送至混洗服务器 S_2。由于任一服务器仅持有随机份额，因此无法恢复完整私钥。只有在秘密重构算法 $\Sigma_{\mathrm{ASS.}}.\mathrm{Reconstruct}$ 被执行时，原始私钥才可以恢复，从而保证系统在单点服务器被攻破时仍能保持私钥安全。

(6) 混洗阶段

$\mathrm{Shuffle}\left(\left\{s_{i,j}\right\}\right)\to H$：$S_1$ 接收各客户端私钥的第一份额 $\boldsymbol{s}_{i,1}$，S_2 接收第二份额 $\boldsymbol{s}_{i,2}$，两台服务器按相同逻辑处理，将每个真实份额拆分为 k 个子份额，前 $k-1$ 个子份额随机采样生成，最后一个子份额通过加法补偿确保总和与原份额一致，实现份额拆分的完整性。为增强不可区分性，服务器引入输入为零向量的虚拟客户端，将其拆分为 k_0 个子份额，前 k_0-1 个随机生成，最后一个份额通过补偿机制保证虚拟份额总和为零，避免对聚合结果产生干扰。最后，为了保证混洗后的份额分布在计算上不可区分，该过程调用第 2.3 节介绍的多重互斥综合译码问题作为安全基础，混洗服务器将真实客户端子份额与虚拟客户端子份额混合，通过均匀随机置换打乱顺序，生成混洗矩阵。S_1 输出包含混合份额矩阵、真实客户端最后一份额列及虚拟客户端补偿份额的集合 $\boldsymbol{H}_1$，S_2 输出对应集合 $\boldsymbol{H}_2$，最终发送至中心服务器，确保攻击者无法关联份额与客户端身份，为抗串通性提供支撑。

(7) 聚合与解密阶段

$\text{Aggregate}(\boldsymbol{ct}_i) \rightarrow \boldsymbol{ct}_{\text{agg}}$：中心服务器对所有密文执行加法同态聚合。$\text{Reconstruct}(\boldsymbol{H}_1, \boldsymbol{H}_2) \rightarrow \boldsymbol{s}_{\text{agg}}$ 根据混洗结果重构聚合私钥 $\boldsymbol{s}_{\text{agg}}$。$\text{Dec}(\boldsymbol{ct}_{agg}, \boldsymbol{s}_{agg}) \rightarrow \hat{\boldsymbol{g}}^{(t)}$ 执行 RLWE 解密，并依次将编码多项式分解为各分组的 CRT 整数、CRT 解码与反量化操作，得到聚合梯度 $\hat{\boldsymbol{g}}^{(t)}$，其在正确性条件下等价于明文联邦平均结果。

4.3. 安全及隐私需求

(1) 梯度隐私性

对于任意两个等长梯度向量 $\boldsymbol{g}_0, \boldsymbol{g}_1$，其对应密文在计算上不可区分，即不存在多项式时间攻击者能够以非忽略优势区分二者。该性质建立在 Ring Learning with Errors 困难假设之上。在本方案中，梯度经过量化与多项式嵌入后，通过 RLWE 加密生成密文对。由于密文中引入离散高斯噪声项，并满足标准 RLWE 分布特性，因此在判定 RLWE 困难假设成立的前提下，攻击者无法从密文中恢复明文多项式，从而无法推断单个客户端的原始梯度信息。即便中心服务器掌握所有密文，也只能得到同态聚合结果，而无法反推出任意个体梯度，实现计算意义下的语义安全。

(2) 身份不可关联性

在多轮训练过程中，中心服务器无法建立“客户端身份–上传密文”之间的稳定映射关系。即对任意两轮上传集合，其真实客户端排列与虚拟客户端混合后的输出在计算上不可区分。该性质依赖基于 Multi-Disjoint Syndrome Decoding 困难假设构造的混洗机制。双混洗服务器对私钥份额进行再次拆分、扩展与随机置换，并引入虚拟客户端份额，使得最终发送至中心服务器的份额矩阵满足 MDSD 型随机分布。在该假设下，攻击者无法判断哪些份额源自同一客户端，也无法区分真实与虚拟节点，从而切断跨轮次身份关联路径，实现计算不可关联性。

(3) 抗串通性

抗串通性要求在部分参与方合谋的情况下系统仍保持安全。具体假设为：中心服务器至多与一台混洗服务器串通。在协议中，客户端私钥采用加法秘密共享方式拆分为两份，分别发送至两台独立混洗服务器。任意单方仅持有随机份额，无法恢复完整私钥。当中心服务器与其中一台混洗服务器串通时，其所掌握的信息仍不足以重构客户端私钥或解密单个密文。只有在两台混洗服务器同时泄露全部份额时，私钥才可能被恢复。因此在既定威胁模型下，系统能够抵抗单点合谋攻击，消除传统安全聚合对单一可信实体的依赖。

(4) 正确性与完整性

正确性要求在所有参与方诚实执行协议时，解密得到的聚合梯度等价于明文联邦平均结果。该性质来源于 RLWE 加密方案的加法同态性。服务器对密文执行逐项加法后，得到的聚合密文等价于明文梯度之和的加密形式。只要噪声增长未超过参数界限，解密后即可正确恢复聚合明文。结合量化与 CRT 逆映射步骤，可保证数值误差处于可控范围内，不影响模型收敛与精度表现，从而实现功能正确性与训练完整性。

(5) 抗量子安全性

抗量子安全性要求在量子计算模型下，协议核心安全性仍成立。本方案的安全基础建立在 RLWE 与 MDSD 等格困难问题之上。现有研究表明，这类问题在经典与已知量子算法模型下均不存在多项式时间有效解法，且不受 Shor 算法等针对整数分解与离散对数问题的量子攻击影响。因此，相较于基于 RSA 或 ECC 的传统安全聚合机制，本方案具备后量子安全特性。即便未来量子计算能力提升，攻击者仍难以通过量子算法高效恢复私钥或破解密文分布，从而保障系统长期运行安全。

5. 性能分析

实验在搭载 2.2GHz Intel Core i7 处理器与 8GB RAM Linux 平台上完成，实现基于 Ring Learning with

Errors (RLWE)的同态加密与秘密共享模块，安全参数设置为 $\lambda=128$ 。为验证方案的效率与可扩展性，实验分别从客户端计算时间、服务器聚合时间以及系统整体通信开销三个方面进行对比分析，并与明文 FedAvg 方案进行性能比较。

在计算开销方面，实验固定模型维度，逐步增加客户端数量，从 10 扩展至 100，统计单轮训练中客户端加密时间与服务器端聚合解密时间。结果表明，客户端侧的主要开销集中在梯度量化与加密阶段，计算时间随客户端数量变化基本保持稳定；服务器端的聚合时间随客户端规模线性增长，但增长幅度平缓，未出现指数级上升趋势。当客户端数量达到 100 时，单轮聚合与解密仍可在数十秒内完成，验证了协议在大规模场景下的可行性。

在通信开销方面，由于采用“量化-CRT-多项式打包”结构，每个客户端仅需上传一个密文对及少量私钥份额，相较于未打包加密方案显著减少通信数据量。实验曲线显示，系统总通信量随客户端数量呈线性增长，但单位客户端通信负担保持不变，体现出良好的扩展性能。

为分析客户端规模变化对联邦学习模型收敛性能的影响，在保持模型结构、优化器参数及本地训练轮次不变的条件下，分别设置不同数量的客户端参与联邦训练过程，并对模型在测试集上的准确率随通信轮次的变化情况进行统计分析。从图 2 和图 3 可以观察到，在 MNIST 与 CIFAR-10 两个数据集上，模型测试准确率均随着通信轮次的增加呈现出单调上升并逐步趋于稳定的变化趋势，表明在引入隐私保护机制后，模型仍能够实现稳定收敛。当客户端数量较少时，模型收敛速度相对较快，在较少的通信轮次内即可达到较高的测试准确率；随着客户端数量的逐步增加，模型的收敛速度出现一定程度的放缓，但整体收敛趋势保持一致，未出现明显的震荡或发散现象。

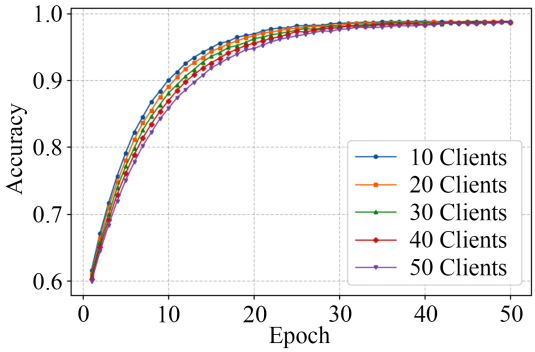

Figure 2. Accuracy rate under the MNIST dataset
图 2. MNIST 数据集下的准确率

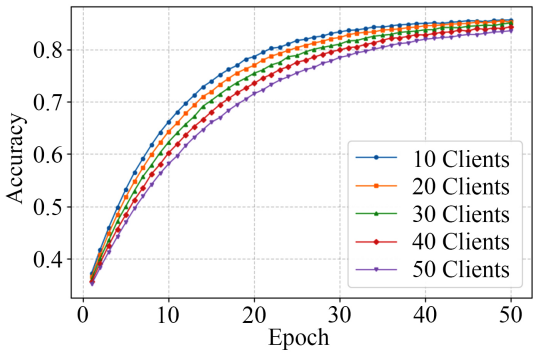

Figure 3. Accuracy rate under the CIFAR-10 dataset
图 3. CIFAR-10 数据集下的准确率

进一步分析可见，在不同客户端规模条件下，模型最终收敛时的测试准确率差异较小，说明客户端数量的增加并未对模型的最终性能造成显著影响。这主要是由于联邦学习过程中各客户端的本地更新在聚合阶段通过同态加密与安全聚合机制被有效整合，从而保证了全局梯度方向的整体一致性。尽管在客户端数量较多时，梯度量化误差与加密噪声在聚合过程中存在叠加效应，但该影响被控制在合理范围内，并未破坏模型参数更新的有效性。

此外，在模型性能方面，对比明文 FedAvg 与本文方案在 MNIST 与 CIFAR-10 数据集上的收敛曲线可以看出，两者测试准确率基本一致，未因加密与量化操作引入明显精度损失。综合来看，所提方案在增强梯度隐私与抗串通能力的同时，仍能保持较高的计算效率与良好的实际应用价值。

6. 总结与展望

6.1. 总结

针对跨机构弱信任环境下联邦学习面临的梯度泄露、身份关联、串通攻击及量子计算威胁，本文提出一种量子安全联邦学习协议，实现了后量子安全、抗串通性与身份梯度不可关联性的协同保障。该协议以 RLWE 格困难问题为基础构建后量子安全的同态加密聚合机制，结合加法秘密共享与双混洗服务器设计抵御串通攻击，通过虚拟客户端与 MDSD 混洗机制切断身份与梯度的关联路径，同时采用“量化-IntCRT-PolySubR”打包策略优化计算与通信效率。实验结果表明，该协议在安全参数 $\lambda = 128$ 下，客户端计算开销基本稳定，服务器聚合开销呈平缓线性增长，通信扩展性良好，且在 MNIST、CIFAR-10 数据集上的训练精度与明文 FedAvg 基本一致，无明显精度损失，实现了安全性、性能与模型效用的平衡，为弱信任环境下的跨机构隐私保护数据协作提供了可行的技术方案。

6.2. 展望

本文提出的协议虽实现了核心安全目标与性能平衡，但仍有进一步优化和拓展的空间。后续可从三方面开展研究：一是轻量化优化，结合硬件加速技术降低加密、混洗等核心模块的计算开销，提升协议在边缘终端、物联网设备中的适配性；二是多维度安全增强，融合差分隐私、访问控制等技术，针对投毒攻击、模型窃取攻击设计防御机制，提升复杂攻击场景下的鲁棒性；三是实际场景落地，将协议部署于医疗、金融等隐私敏感领域，开展大规模跨机构联合训练验证，结合行业合规要求优化协议设计，推动量子安全联邦学习的产业化应用。同时，可探索新一代后量子密码算法与联邦学习的深度融合，构建全链路后量子安全防护体系，为量子计算时代的分布式机器学习提供更全面的安全保障。

参考文献

[1] 罗姚, 魏苏璟, 杨晶, 等. 公立医院数据安全风险治理路径研究[J]. 卫生经济研究, 2025, 42(3): 71-74, 78.

[2] 李莉莎, 谭镇锋. 金融数据共享: 理论、挑战与机制构建[J]. 经济与社会发展, 2024, 22(5): 59-74.

[3] Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y., Yang, Q., *et al*. (2020) Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, **22**, 2031-2063. https://doi.org/10.1109/comst.2020.2986024

[4] Geiping, J., Bauermeister, H., Dröge, H., *et al*. (2020) Inverting Gradients-How Easy Is It to Break Privacy in Federated Learning? *Advances in Neural Information Processing Systems*, **33**, 16937-16947.

[5] Mansouri, M., Önen, M., Ben Jaballah, W. and Conti, M. (2023) Sok: Secure Aggregation Based on Cryptographic Schemes for Federated Learning. *Proceedings on Privacy Enhancing Technologies*, **2023**, 140-157. https://doi.org/10.56553/popets-2023-0009

[6] Gentry, C. (2009) Fully Homomorphic Encryption Using Ideal Lattices. *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, Bethesda, 31 May-2 June 2009, 169-178. https://doi.org/10.1145/1536414.1536440

[7] Behera, S. and Prathuri, J.R. (2024) FPGA-Based Acceleration of K-Nearest Neighbor Algorithm on Fully Homomorphic Encrypted Data. *Cryptography*, **8**, Article 8. https://doi.org/10.3390/cryptography8010008

[8] Regev, O. (2009) On Lattices, Learning with Errors, Random Linear Codes, and Cryptography. *Journal of the ACM*, **56**, 1-40. https://doi.org/10.1145/1568318.1568324

[9] Shamir, A. (1979) How to Share a Secret. *Communications of the ACM*, **22**, 612-613. https://doi.org/10.1145/359168.359176

[10] Gascón, A., Ishai, Y., Kelkar, M., Li, B., Ma, Y. and Raykova, M. (2024) Computationally Secure Aggregation and Private Information Retrieval in the Shuffle Model. *Proceedings of the* 2024 *on ACM SIGSAC Conference on Computer and Communications Security*, Salt Lake, 14-18 October 2024, 4122-4136. https://doi.org/10.1145/3658644.3670391