arXiv:2602.01185v1 [cs.CR] 1 Feb 2026

FedBGS：一种在去中心化系统中实现分段Gossip学习的区块链方法

Fabio Turazza、Marcello Pietri、Marco Picone、
Marco Mamei摩德纳-雷焦艾米利亚大学工程科学与方法系，
意大利雷焦艾米利亚name.surname@unimore.it

摘要——**隐私保护联邦学习（PPFL）是一种去中心化机器学习范式，它通过整合密码学与隐私保护技术，增强整个系统的安全性，使多个参与方能够在不共享数据的情况下协同训练全局模型。这种面向隐私的方法使 PPFL 成为在数据隐私至关重要的领域训练共享模型的理想方案。在传统联邦学习中，边缘设备负责训练本地模型，且仅将模型更新发送到中央服务器，由中央服务器聚合这些更新，以改进全局模型。然而，尽管采用了上述隐私保护技术，经典的联邦架构仍存在服务器单点故障的问题，从而在安全性和可扩展性方面受到限制。本文介绍 FedBGS，这是一种完全去中心化的区块链框架，通过联邦分析实现分段式 Gossip 学习。该系统旨在优化区块链使用，同时提供全面的防攻击保护，确保联邦环境中的隐私、安全性及非独立同分布数据处理。**

索引词—**联邦学习，gossip 学习，区块链，可扩展性，数据隐私，分析**

一、引言

过去十年间，随着数字应用迅速普及，以及数据存储解决方案变得更加实惠、易于实施，数据获得了显著的法律和经济价值。从法律角度看，这些信息中有很大一部分属于个人信息，被视为每个人不可侵犯的资产，并受到欧盟GDPR（通用数据保护条例）等措施的保护；对于敏感医疗信息，美国HIPAA（健康保险流通与责任法案）也提供了保护。从经济角度看，企业将这些数据视为关键资产，往往不愿共享。这种不情愿造成了一种“囚徒困境”：数据交换能够通过改善共享模型的训练、提升其泛化能力并减少偏差，带来集体利益。

联邦学习由谷歌于 2016 年通过 FedAvg（联邦平均）算法引入， [1],旨在通过建立去中心化训练流程来解决这些问题，在该流程中，数据始终由其原始所有者持有。在 FedAvg 方法中，每位参与者都会训练一个本地模型，并仅将模型更新发送至中央服务器。服务器对这些更新求平均并进行聚合，以生成全局模型，然后将其重新分发给参与者。这个过程会重复进行预先设定的轮次直到模型收敛。尽管 FedAvg 确保数据始终保留在各所有者的领域内，但它在安全性方面仍存在漏洞，可能遭受多种攻击，例如重构攻击、投毒攻击和女巫攻击。此外，FedAvg 对来自客户端的非独立同分布异构数据较为敏感，这可能导致性能显著下降。

为应对这些挑战，隐私保护联邦学习（PPFL） [2] 应运而生，旨在整合额外的隐私和加密技术，以增强传统联邦学习固有的安全性。应用最广泛的隐私保护方法是差分隐私（DP）[3] [4], ，其通过向数据添加噪声（通常为高斯噪声）来减轻数据泄露可能造成的损害。其他策略则包括将FedAvg与安全多方计算（SMPC）[5]和同态加密（HE） [6] [7]等方法结合。FedAvg所凸显的另一个重要问题，是如何管理客户端之间的非独立同分布数据：当参与方之间的数据分布存在异质性时，全局模型的性能会呈指数级下降。近年来的多项研究致力于解决联邦场景中的非独立同分布数据问题：FedProx [8] 通过在局部损失函数中加入邻近项来处理非独立同分布数据问题，该项限制客户端模型偏离全局模型的程度，从而稳定训练。SCAFFOLD [9]通过利用梯度方差缩减来校正客户端漂移，确保客户端之间的更新更加一致，并加速收敛。另一方面，MOON [10], 利用对比学习使本地模型与全局模型对齐，缩小局部数据分布之间的差异并提升泛化能力。FedDyn[11]采用动态正则化来平衡客户端与中心模型之间的更新，从而补偿异构数据分布引入的偏差。

为缓解联邦场景中参与方之间异构数据分布的问题，同时确保系统保持可扩展性和透明性，我们提出 FedBGS：一种基于区块链的自分段流言学习框架，旨在实现可扩展、模型无关，并能够适应非独立同分布数据分布。FedBGS 兼容任何基于以太坊的区块链，并借助 IPFS 实现优化的数据存储；若依赖公有区块链，这种存储方式将难以实现。

本研究的主要贡献如下：

- 联邦 K 均值++自动分割：本研究提出了一种利用联邦分析的完全去中心化聚类机制，可根据数据分布自动对客户端进行分段，并减轻非独立同分布数据的影响。

- 基于区块链的分段式 Gossip 学习方法：提出一种基于区块链的分段式 gossip 学习方法，客户端在动态形成的分段内交换模型更新，从而确保可扩展性、稳健性和去中心化。
- 用于高效存储的混合区块链与 IPFS：该系统将链上聚合与通过 IPFS 实现的链下存储相结合，在保持去中心化学习透明性和安全性的同时，克服区块链的可扩展性问题。
- 基于非独立同分布数据的实证验证：在*cifar-10* [12] 和*mnist* [13] [14] [15][16] 数据集上进行的大量实验表明，与异构环境和资源受限区块链中的传统联邦学习及 gossip 学习方法相比，该方法提升了安全性和准确率。此外，还将对 FedBGS 的完全去中心化和可扩展性进行理论评估（表 I）；与最先进的方法相比，这两点正是其真正优势所在。

II. 相关工作

尽管训练具有分布式特性，传统联邦架构仍依赖中心化聚合，使服务器成为整个系统的关键薄弱环节。近期多项研究致力于消除服务器这一单点故障（SPOF），转而采用点对点架构，在客户端层面或通过利用区块链技术的智能合约进行聚合。FLchain [17], BFLC [18], 和 BlockFL 等框架利用区块链实现联邦学习的完全去中心化，使参与方能够全面了解聚合操作及整个系统的运行情况。然而，与传统联邦学习类似，这些框架仍会因非独立同分布数据而出现性能下降；使用公有区块链时也存在可扩展性问题，因此不适合存储大量模型参数。为解决区块链系统固有的存储限制， [20] 提出了一种结合链上和链下存储方案的混合架构。在此方法中，使用星际文件系统（IPFS） [21][22] [23] 存储中间模型参数，从而缓解区块链瓶颈带来的存储限制。

gossip 学习 [24] 是一种去中心化的联邦方法，参与方通过点对点机制进行边缘端聚合，直接与网络中的其他节点交换模型权重。这种方法克服了中央服务器的局限性，使系统更具容错能力，并有望实现更强的可扩展性。实证研究表明，gossip 学习能够取得与联邦学习相当、某些情况下甚至更优的性能，尤其是在训练数据均匀分布于各节点上的情况下。值得注意的是， [25]对 gossip 学习与联邦学习进行了深入的实证比较，指出了去中心化方法表现相当甚至超越集中式方法的情形。

拆分学习 [26] [27] 是一种通过在客户端和中央服务器之间划分神经网络来降低通信开销的技术。在这种方法中，客户端在本地计算前几层，并仅将中间激活值传输给服务器，由服务器完成前向传播和反向传播。这种策略通过将原始数据保留在客户端来最大限度地减少数据交换并增强隐私；然而，它依赖中央服务器，可能重新引发中心化问题，并可能限制可扩展性和安全性。

分段式 gossip 学习是 gossip 学习的一项重要扩展，旨在应对非独立同分布数据分布以及节点间通信开销所带来的挑战。它不再于每一步共享和更新整个模型，而是仅在节点间交换模型参数的特定分段。 [28]探讨了如何在去中心化 gossip 学习框架中对模型进行分段，通常是将最终层划分到多个簇中的节点上。这种方法降低了通信开销，并缓解了异构数据场景下的梯度冲突，从而加快收敛并提升模型准确率，同时维持去中心化学习过程。在 GossipFL [29] ，中，没有明确进行分段，而是通过稀疏化仅传输最重要的模型更新来降低开销，从而减少节点间交换的数据量。此外，自适应通信会根据网络状况和收敛进程动态调整更新的频率和数量，实现高效的资源利用率并加快训练。

III. 所提出的框架

FedBGS （联邦区块链 gossip 分段）是一种基于 gossip 学习的框架，通过单轮联邦 K 均值进行区块链调度，从而自动对对等模型进行分段。这种分段有助于在客户端数据分布存在异质性时保护系统，并避免产生计算开销。该方法能够处理非独立同分布的数据分布，同时仍为每个组件提供对系统的整体视图。为确保相关各方之间的充分透明性和信任（这在跨孤岛场景中尤为重要），FedBGS 集成了一个基于以太坊智能合约的区块链系统。这样便能验证参与方，防止恶意客户端混入联邦过程，并通过对等节点层和区块链层的双重验证，使系统能够抵御所谓的“投毒攻击”。在聚类阶段，区块链还充当中央聚合器，使系统实现完全去中心化，并避免依赖可能成为单点故障的单一服务器。使用以太坊这一标准区块链所固有的局限性——尽管它高度安全，但可扩展性有限，

TABLE I：FedBGS 与联邦学习（FL）及区块链联邦学习方法的比较 s

方法	非独立同分布数据处理	混合存储（区块链 + IPFS）	隐私保护技术	可扩展性	完全去中心化
FedAvg [1]	有限，需要增强	×	基本隐私保护，易受推理攻击	高，但中心化	×
FLchain [17]	中等，继承 FedAvg 局限性	部分（仅链上）	侧重透明性，不具备广告差分隐私（DP、同态加密）	中等，共识瓶颈	✓
BFLC [18]	有限，无特殊处理	有限，链上 with 委员会共识	通过委员会验证实现鲁棒性验证	有所改进，但仍有限大规模地	✓
BlockFL [19]	无特殊处理，标准 FedAvg	仅链上，高存储开销	透明验证，无差分隐私或加密	有限，共识开销	✓
区块链图联邦学习 [30]	有限的显式非独立同分布处理理，侧重于八卦通信	仅限链上，不使用 IPFS	去中心化，不涉及明确的差分隐私或加密	通过八卦通信实现高	✓
FedBGS（所提出的方法）	**鲁棒（单轮 K-均值++ 聚类）**	**混合（区块链 + IPFS）**	**同态加密、差分隐私和截尾均值聚合**	**高、去中心化分段训练**	✓

注：粗体条目表示 FedBGS 相较于其他现有方法的显著优势。

不适合链上存储模型权重，且 Gas 费用高昂——通过一种混合方法得以解决：聚合和存储通过 IPFS 在链下进行，同时将已完成交易的引用以及联邦过程的完整历史记录保存到链上。

A. 阶段 *1*：通过联邦分析实现自动聚类 (1)

随着联邦学习的兴起，另一种去中心化、面向隐私的方法——联邦分析（FA） [31], 开始逐渐成形，而联邦学习是联邦分析的一个子类别。联邦分析不同于传统的联邦学习，因为它的目标并非训练全局人工智能模型，而是以联邦方式进行简单的统计计算。因此，它的结构要简单得多，也不依赖多个轮次。单轮联邦K均值是联邦分析的一种具体形式，它通过一轮通信从对等节点获取质心。这样便可以根据相似性将对等节点分组为不同的簇，同时不泄露任何参与者的内部数据。与传统的联邦架构一样，联邦分析也依赖中央聚合器。然而，这种方法的优势在于无需共享完整的深度学习模型。只需发送质心等统计参数（其存储需求远低于权重），便可利用智能合约进行链上存储和聚合，且不会产生过高费用。值得注意的是，聚类步骤采用了联邦分析中典型的单轮方法，专门用于避免区块链负载过重。另一方面，单轮方法往往会牺牲聚类准确率，通常只能得到次优解。然而，我们的实验表明，即使采用次优方法，这种轻微的性能下降通常也可以忽略不计。尽管如此，只要谨慎评估其区块链存储方面的权衡，属于联邦学习范畴的少轮迭代方法仍不失为一种可行方案。为提高聚类过程的整体精度，我们采用了 K 均值++ 算法 [32], : 该算法随机选取第一个质心，随后根据概率分布选择后续质心，该概率分布与其到最近已有质心的平方距离成正比。这种方法降低了随机性，并提高了聚类效率。

FedBGS 将使用初始阶段生成的簇来创建全局模型的分段；每个簇对应模型的一个分段，也就是说，每个簇负责训练模型的不同部分。分段通过将模型的最后一层等分来实现，确保每个簇都训练模型的特定部分。这种方法可减少梯度冲突，同时降低计算和通信开销。

检索过程通过访问链上映射来实现，该映射将每个簇与最后一个全连接层中的特定神经元范围关联起来。当对等节点请求其分段边界时，智能合约会从存储中获取分配的起始和结束索引并返回。这确保每个簇仅更新模型中指定的部分，从而防止未经授权的修改，并实现去中心化分段。

设 C_k 为第 k 个簇， $\mathcal{S}_k$ 表示分配给该簇的模型分段，该分段由神经元索引范围$[s_k, e_k]$定义。对于模型参数 w_i，智能合约强制执行的更新规则如下：

$$\mathcal{S}_k = \{w_i \mid s_k \le i \le e_k\} \tag{1}$$

$$\forall w_i \in \mathcal{W}, \quad w_i \text{ is updated only if } w_i \in \mathcal{S}_k \tag{2}$$

图例：
C_k：第 k 个簇。$\mathcal{S}_k$：分配给簇 C_k 的分段。s_k， e_k：定义 C_k 分段的起始和结束索引。w_i：最后一个全连接层中的一个模型参数。$\mathcal{W}$：最后一层中所有模型参数的集合。

B. 区块链集成

以太坊是一个广受欢迎的区块链平台，让人们能够在没有中央权威机构的情况下构建和运行智能合约及去中心化应用。区块链由区块组成，每个区块（单个区块的结构如图 1 所示）都包含一系列交易以及对前一区块的引用。区块之间的这种连接形成了一条安全的链，使过去的数据难以被篡改。

在以太坊中，交易由网络参与者验证，并由他们创建新区块。在最初的系统中，这些参与者被称为矿工，他们使用一种称为工作量证明（PoW）的方法，在添加区块前解决复杂的谜题。不过，以太坊现已转向一种称为权益证明（PoS）的系统。在 PoS 中，不再由矿工参与，而是由验证者根据其锁定的以太币数量（即“质押”）被选出创建区块。这种方法能耗更低，效率也更高。

以太坊中的现代验证技术着重确保验证者诚实行事。例如，如果验证者行为不端，系统可以罚没（或惩罚）其质押。这有助于保持网络的安全与可靠。

我们选择在联邦 Gossip 学习项目以及作为联邦单次 K-means 的服务器中使用以太坊，因为它提供了一种去中心化且透明的方式来验证更新和客户端注册。使用以太坊，我们可以确保每位参与者的贡献都记录在安全且不可篡改的账本中。

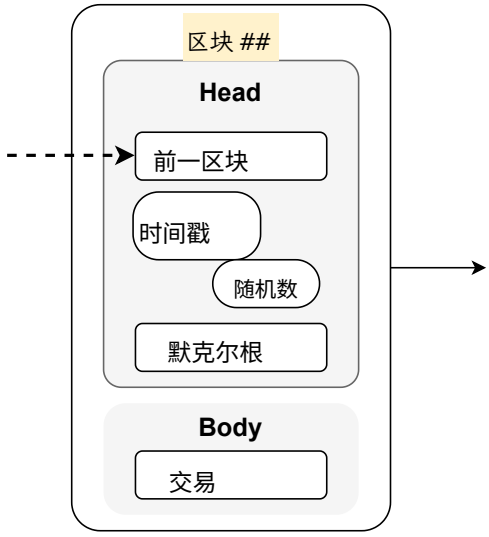


图 1：以太坊区块示例：在以太坊区块中，默克尔根 [33] 是一个汇总并验证区块中所有交易的单一哈希，从而确保数据完整性。随机数是在共识过程中用于获得有效区块哈希值的数字。前一区块的哈希值将当前区块与前一区块相连，从而形成一条不可篡改链。

在去中心化学习中，区块链通常用于以下目的：

- **去中心化：**通过智能合约进行链上聚合，使系统实现去中心化，从而消除服务器这一单点故障。
- **不可变客户端注册表：**维护获准参与的客户端的不可变账本，有助于防范女巫攻击。
- **安全交易记录：**确保以分布式且不可更改的方式记录交易（即节点发送的更新），从而降低遭受篡改（投毒攻击）的风险。
- **双重花费防范：**防止双重花费攻击：节点可能试图重复使用旧更新，或将其复制为新更新，以操纵联邦模型；由于每笔交易都会被唯一追踪，且无法再次“花费”，因此可防范此类攻击。
- **激励机制：**通过基于代币的智能合约实施激励系统。

C. IPFS 与区块链

IPFS（星际文件系统）是一种以分布式方式存储和共享文件的系统，无需依赖中心化服务器。IPFS 的基本原理如下：

- **内容寻址：**
  在 IPFS 中，文件并非通过其位置（例如 URL）来识别，而是通过其内容来识别，如图 2 所示。添加文件时，系统会将其转换为唯一的密码学哈希。这个哈希会成为文件的地址，确保每个文件都能被安全且不可篡改地识别。
- **区块和默克尔有向无环图：**
  IPFS 中的文件会被拆分成称为区块的小块。每个区块包含：

  – **数据：** 文件的一部分。– **链接：** 指向其他区块的引用（如果文件被拆分成多个部分）。

  这些链接构成一种称为默克尔有向无环图（有向无环图）的结构，可用于验证每个区块是否未被篡改。即使单个区块发生变化，整体哈希也会改变，从而表明内容已被修改。
- **分发：**
  当你通过哈希请求文件时，IPFS 会在整个网络中搜索与该哈希对应的区块（或多个区块），并将它们重新组装以还原文件。

IPFS 和以太坊都采用默克尔图结构，以高效确保数据完整性，并通过最少的密码学证明验证内容。在以太坊中，这种设计（帕特里夏树）能够快速、去中心化地验证交易和状态变化，从而降低燃料费用并提高可扩展性。同样，IPFS 使用默克尔图实现不可篡改的内容寻址存储，形成了天然的协同效应，使两个系统的集成既顺畅又稳健。

在 FedBGS 中，区块链的使用将分为两个阶段：

阶段 1：

- **任务 1：** 区块链将负责通过唯一凭证管理每位参与者的注册；未注册的客户端无权参与共享训练。
- **任务2：**我们将通过智能合约对联邦 K 均值的质心进行聚合，并将结果直接存储在链上。

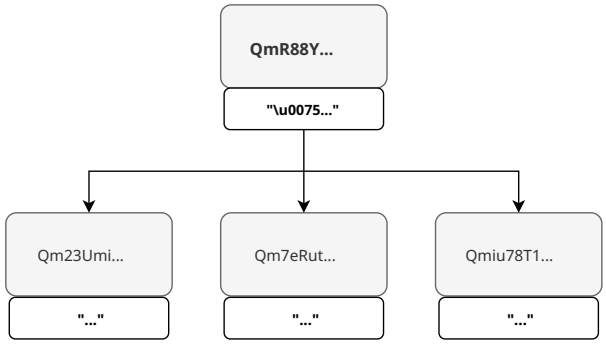


图2：IPFS 中的区块大小可能因存储的数据而异。得益于基于哈希链接的层次结构，每个区块都会指向其他相连的区块，从而促进节点间的冗余与共享。这种可扩展的方法确保能够从不同位置快速、可靠地检索内容。

- **任务3：**智能合约使用内部策略算法，确定模型的哪一部分将被分配给每个簇。在流言传播周期开始时，客户端查询智能合约，以了解其指定的簇以及应训练模型的哪个“分段”。

第2阶段：

- **任务1：**智能合约会定期随机选出一位领导者，负责处理跨分段的全局聚合，重新生成原始的“未分段”模型，并将其存储在 IPFS 上。
- **任务2：** 每个对等节点获取全局模型时，都会执行验证流程，以确认模型的完整性（对等节点提取的模型哈希必须与区块链上存储的不可变哈希一致）。
- **任务3：**基于代币的激励机制会奖励那些向后续客户端提供更优模型的客户端，并惩罚与当前损失偏差过大的客户端。

D. 隐私保护的去中心化学习

如引言中所述，gossip 学习本身提供的安全性不足以确保足够的隐私保护，尤其是在跨孤岛场景中。将区块链与 IPFS 引入系统，旨在防止某些可能的攻击，例如女巫攻击或投毒攻击；然而，数据泄露或“好奇”的参与方可能进行的间谍活动，仍是此前所述技术未能解决的问题。尤其是在聚类阶段，系统传输的是质心等统计数据，而非深度模型的参数，因此特别容易受到多种攻击，包括：

- **成员推理攻击：**对手试图确定某个特定样本（例如，用户数据）是否曾出现在训练集中。
- **梯度反演（或重构）攻击：**攻击者试图从客户端在训练期间共享的梯度中重构敏感原始数据（图像、文本等）。
- **模型反演攻击：**其目的是利用模型返回的信息，推断训练样本中的敏感属性。
- **窃听／中间人攻击：**攻击者会在参数（或梯度）于参与方与中央服务器之间（或对等节点之间）传输时进行拦截或篡改。
- **模型提取攻击：**对手通过发起查询或分析系统的响应，试图提取或复制模型的架构和参数。

在 FedBGS 中，我们集成了多种先进技术，以增强采用区块链和 IPFS 的去中心化 gossip 学习系统的安全性和鲁棒性。这些技术包括差分隐私、同态加密以及通过截尾均值实现的鲁棒聚合，它们协同工作以保护敏感信息、确保安全聚合并抵御对抗攻击。

差分隐私（DP）：即使共享模型更新，差分隐私也能确保每个客户端的贡献保持私密。在联邦设置中，本地差分隐私是在发送更新之前于客户端层面添加噪声，而全局差分隐私则是在聚合过程中添加噪声。在我们的 gossip 学习框架中，差分隐私用于训练阶段，对梯度进行裁剪和扰动，从而保护个体数据贡献，同时仍能实现有效的协作学习。

在分段式 gossip 学习的训练阶段，通过裁剪并添加高斯噪声，使每个客户端的梯度 g 满足差分隐私。其表示如下:

$$\tilde{g} = g \cdot \min\left(1, \frac{C}{\|g\|_2}\right) + \mathcal{N}\left(0, \sigma^2 C^2\right) \tag{3}$$

图例：
g：客户端计算的原始梯度。$\tilde{g}$：差分隐私梯度。C：裁剪阈值。$\|g\|_2$：梯度 g 的 L_2-范数。$\mathcal{N}(0, \sigma^2 C^2)$：均值为 0、方差为 $\sigma^2 C^2$ 的高斯噪声。σ：控制所添加噪声量的噪声乘数。

FedBGS实现了一个内部调度器，允许差分隐私动态变化。我们的调度器随时间对差分隐私应用线性递减，并在最终迭代达到最低水平。早期轮次尤为敏感，因为模型直接受到每个节点初始数据的影响。由于模型尚未与许多其他参与方的信息“混合”，所发送的更新可能包含更多原始数据的可识别痕迹。换言之，隐私可能面临更大风险，因为“噪声”（即其他节点引入的变异性以及任何人工差分隐私噪声）尚未积累到足以掩盖本地数据集个体特征的程度。

此外，在初始轮次中，如果更新更为显著，则需要引入更多噪声以保持其受保护状态，这可能会迅速消耗隐私预算。如果该预算在极早的轮次中耗尽或大幅减少，则存在无法在后续迭代中保证同等保护水平的风险。

同态加密（HE）： HE 允许在不泄露底层信息的情况下对加密数据进行计算。虽然全同态加密支持对密文进行任意计算，但其计算成本很高，因此实用性较低。为此，我们采用部分同态加密，具体来说，使用加法同态 Paillier 方案安全地聚合标签分布和模型更新。这使我们能够合并加密的客户端更新，而不暴露其原始值，从而在聚合过程中同时维护隐私和完整性。

为在聚类阶段保护质心，我们使用 Paillier 方案 [34] 聚合加密的标签分布。设 $x_{i,\ j}$ 为客户端 i的标签分布（质心）的第 j 个分量。令 $E(\cdot)$和 $D(\cdot)$分别表示加密和解密函数，并使用同态加法算子 $\bigoplus$，聚合（并解密后）的质心分量计算如下：

$$\hat{c}_j = \frac{D(\bigoplus_{i=1}^{n} E\left(x_{i,j}\right))}{n} \tag{4}$$

图例：
$E(x)$：使用 Paillier 方案对 x进行加密。$D(y)$：对密文 y进行解密。$\bigoplus$：同态加法算子。n：客户端（或标签分布）的总数。$x_{i,\ j}$：客户端 i 的标签分布的第 j 个分量。

在 **FedBGS** 中，部分同态加密仅在聚类阶段使用；我们认为从安全性角度来看，该阶段最为脆弱。它并未应用于整个过程，以避免八卦通信带来的计算开销（需要加密更多数据，且迭代次数更多）。

截尾均值聚合： 为应对拜占庭攻击 [35]，（某些节点可能提交恶意更新），聚合采用截尾均值进行计算。这种鲁棒统计方法会在求平均之前丢弃固定比例的最极端值，从而降低离群值的影响。在我们的八卦学习系统中，截尾均值通过减轻对抗性贡献的影响，有助于维持全局模型的可靠性。

这些技术共同使我们的系统能够在联邦环境中执行安全、保护隐私且鲁棒的模型聚合。通过将差分隐私用于隐私保护、将部分同态加密用于安全计算，并将截尾均值用于鲁棒聚合，我们的方法非常适合应对去中心化且可能存在对抗性的环境所带来的独特挑战。

设 $\{\theta_i\}_{i=1}^{n}$ 为模型更新的集合，$\theta_{(i)}$ 表示排序后的第 i 个更新。在截尾比例 r 下，稳健聚合

算法 1 FedBGS 阶段 1：链上聚类与客户端注册**需要：** 对等节点集合 P 及其本地数据 D_p**确保：** 每个对等节点 $p \in P$ 均已在链上注册并被分配到分段 S_i1: **对于** 每个对等节点 $p \in P$ **执行**2: 使用唯一凭证在区块链上注册 p 3: 根据 D_p计算本地标签分布 x_p 4: **结束循环**5: 通过对 $\{x_p\}_{p\in P}$ 执行联邦 K 均值++形成分段 $S_1,\ S_2,\ \ldots\ldots,\ S_S$ 6: **对于**每个分段 S_i **执行**7: **安全聚合：** 使用部分同态加密聚合标签分布▷ 参见式 (4) 中的同态加密表述 8: 计算分段 S_i 的安全质心$\hat{c}_i$ 9: 在链上更新$\hat{c}_i$ （充当聚合服务器）10: **结束循环** 11: **簇发现：** 每个对等节点 $p \in P$ 通过智能合约查询获取其被分配的分段S_i ，并受式 (1) 约束

更新 θ^* 的计算方式如下：

$$\theta^* = \frac{1}{n(1-2r)} \sum_{i=\lceil rn \rceil}^{\lfloor (1-r)n \rfloor} \theta_{(i)} \tag{5}$$

图例：
θ_i：第 i个个体模型更新。$\theta_{(i)}$：第 i个顺序统计量（排序后的更新）。n：更新总数。r：截尾比例（从两端截去的数值占比）。θ^*：稳健聚合模型更新。

E. 第*2*阶段：分段式 *Gossip* 学习(2)

如前所述，gossip 学习（GL）是一种完全去中心化地训练共享模型的策略，无需参与方显式共享其数据。gossip 学习不同于联邦学习（后者是半去中心化的），它不依赖中央服务器进行聚合；相反，聚合在客户端层面进行，使系统不受恶意或故障服务器的影响。

gossip 学习可以采用两种不同的范式：

同步逻辑：这种方法保留了联邦学习中的轮次概念，即所有对等节点都要等待每个参与者完成训练阶段，然后再发送各自的模型更新。尽管从分析角度看，这种方法更简单、更有条理，但当客户端存在异构性时，会造成显著的瓶颈；在参与者众多且计算能力和数据各异的场景中，系统的可扩展性会因此降低。异步逻辑：这种方法没有轮次；每个客户端会根据其内部策略，在不同时间聚合模型更新。

算法 2 FedBGS 第2阶段：采用混合存储和全局聚合的分段式 Gossip 学习**要求：** 对等节点集合 P 及其分配的分段 S_i，差分隐私参数 C，σ，截尾比例 r**确保：** 全局聚合后的模型参数 θ^*1：**对于** 每个对等节点 $p \in P$ 在其分配的分段 S_i 内**执行**2：训练本地模型并计算梯度 g_p3：**应用差分隐私：** 裁剪并扰动 g_p 以获得 $\tilde{g}_p$ ▷，参见公式（3）中的差分隐私表述 4：通过混合存储共享 $\tilde{g}_p$ ——将模型更新存储在 IPFS 上——将更新哈希记录在链上，以进行完整性验证 5：将 IPFS 哈希与区块链记录匹配，验证更新 6：**如果**验证失败**则**7：通过智能合约强制扣除代币，对 p 进行惩罚 8：**结束如果**9：**结束循环**10：**全局聚合：** 11：通过智能合约选举一名领导者 p^* （例如，使用链上随机性）12：**对于**每个分段 S_i **执行**13：领导者 p^* 收集更新 $\{\tilde{g}_p\}_{p \in S_i}$14：使用截尾均值聚合更新（公式（5））15：获得分段更新 $\Delta\theta_i$ 16：使用 $\Delta\theta_i$更新全局模型分段 i17：**结束循环**18：领导者 p^* 重构完整的全局模型 θ^*19：将 θ^* 存储在 IPFS 上，并将其 CID 记录在链上 20：每个对等节点 $p \in P$ 获取最新 CID 并更新其本地模型

虽然这种策略不如同步方法有序，但性能更佳。

如图 4 所示的整体架构所述，FedBGL 分两个阶段运行：

阶段 1（图 3a）：链上聚类与客户端注册：

- **客户端注册：** 每个对等节点使用唯一凭证在区块链上注册。
- **本地计算：** 每个对等节点根据自己的本地数据集计算本地标签分布。
- **聚类：** 对这些分布应用联邦 K 均值++算法，以形成同质分段。
- **安全聚合：** 对于每个分段，标签分布使用部分同态加密进行安全聚合，以计算安全质心。
- **链上更新：** 安全质心直接在链上更新，充当聚合服务器。
- **分段分配：** 每个对等节点通过智能合约查询区块链，检索分配给自己的分段 S_i 。

第2阶段（图 3b）：采用混合存储与全局聚合的分段式 Gossip 学习：

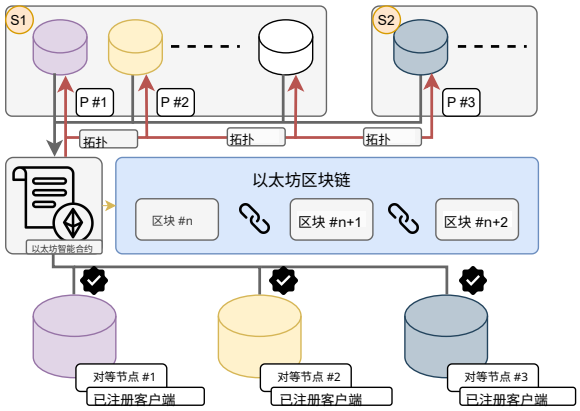
- **本地训练：** 在每个分段内，对等节点训练其局部模型并计算梯度。
- **差分隐私：** 每个梯度都经过用于保护敏感信息的差分隐私机制信息 .
- **混合存储：** 隐私保护更新通过以下方式共享：
  - 将模型更新存储在 IPFS 上。
  - 在以下位置记录相应的更新哈希（内容标识符）：区块链以供验证。
- **更新验证：** 每个对等节点都会验证更新哈希方法是将 IPFS 哈希与区块链记录进行比较。若验证失败，对等节点将受到区块链代币扣除 .
- **领导者选举：** 通过以下方式在链上选举领导对等节点一个智能合约（例如，使用链上随机性）。
- **鲁棒聚合：** 当选的领导者收集每个分段中对等节点的隐私保护更新并使用截尾均值方法对其进行聚合，以计算分段更新。
- **全局模型重构：** 领导者重构来自聚合后的分段更新。
- **全局存储与传播：**
  - 领导者将全局模型存储在 IPFS 上，并将相应的内容标识符记录在链上。– 每个对等节点通过记录的内容标识符检索最新的全局模型，并据此更新自己的本地模型分段。

在 FedBGS 的第二阶段，我们实现了一个异步 gossip 学习系统，其中对等节点各自独立运行（每个节点都在自己的线程中），并进行本地模型训练。训练、共享和测试操作均异步进行，时间间隔则随机确定。

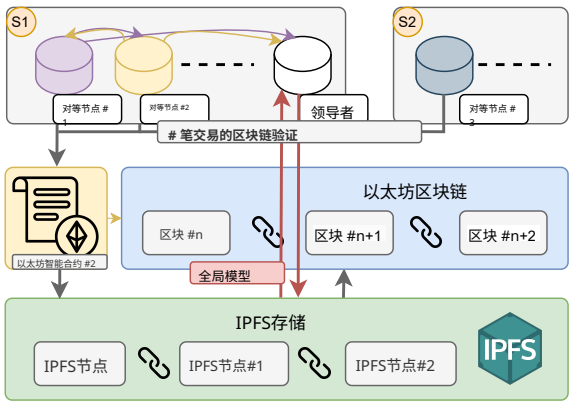
GOSSIP 学习的全局模型推导与迭代

在所提出的协议中，每个对等节点首先使用其私有数据集训练本地模型并计算梯度更新，随后通过差分隐私机制处理，以保护数据机密性。在通过联邦 K 均值聚类算法++得到的每个同质分段内，对等节点与指定邻居交换隐私保护更新。分段内聚合采用截尾均值等鲁棒方法进行，从而优化该分段的本地模型。

系统定期通过采用随机化机制的智能合约，在链上选举一名领导对等节点。选出的领导者收集所有分段的聚合更新，并将其融合为统一的全局模型。随后，领导者将全局模型存储在 IPFS 上，并在区块链上记录相应的内容标识符。最后，每个



(a) **阶段 1**：每个客户端都在区块链上注册，并根据数据相似性进行聚类。区块链充当聚合器，采用联邦式K均值聚类方法更新质心。FedClustering 以去中心化方式采用单次方法运行，质心存储在链上。簇分配直接影响模型分割，因为每个簇负责模型参数的一个子集。



(b)**第2阶段**：客户端执行差分隐私簇式分段 gossip 学习，各簇仅更新最后一个全连接层中的特定分段。模型更新存储并从 IPFS 中检索，以确保高效验证和学习透明性。客户端侧的聚合采用截尾均值聚合，对等节点训练的激励系统则通过专用智能合约进行管理。

图 3：FedBGS 的整体架构，采用基于簇的分段式 gossip 学习。得益于智能合约和以太坊的执行逻辑，委托给区块链的每项操作都会在多个区块上执行并得到验证。分段策略通过将更新限制在最终层中指定的神经元块内，确保在非独立同分布数据上高效训练。

对等节点使用记录的内容标识符获取更新后的全局模型，并据此更新其本地模型。

这一迭代过程包括本地训练、分段内聚合、领导者驱动的全局聚合以及模型传播，确保全局模型持续改进并收敛。

IV. 实验评估

本实验评估旨在通过图表展示 FedBGS 在各种数据集及不同初始数据分布条件下的收敛情况。我们将通过折线图和柱状图展示各参与方对训练的贡献，并以表格总结 FedBGS 在以太坊区块链上的成本影响。有关系统安全性和可扩展性的其他评估将在理论层面进行。为验证 FedBGS 的性能，我们在 6 个不同的标准数据集上进行了测试。本文概述了用于验证 FedBGS 性能的标准数据集。这些数据集分为两大类：基于灰度图像的数据集（28×28）和使用彩色图像的 CIFAR-10 数据集（32×32）。以下实验均在配备 NVIDIA RTX 4080 的笔记本电脑 GPU 上本地运行，并可按照 GitHub 上的部署说明完整复现[1]。

灰度图像数据集（*28×28*）

这些数据集的特点是尺寸统一且为灰度格式。它们被广泛用作基准因其简单性、标准化结构以及便于跨模型比较。

- **MNIST:** 包含以 28×28灰度图像表示的手写数字（0–9）。其中包括 60,000 张训练图像和 10,000 张测试图像。
- **EMNIST:** 这是 MNIST 的扩展版本，包含额外的字母类别。“digits”划分与 MNIST 相同，提供了一个更大的数据集，总计约 280,000 张图像。
- **Fashion-MNIST (FMNIST):** 由 28×28 灰度图像组成，内容为服装物品，其中包含 60,000 张训练图像和 10,000 张测试图像。它为数字识别提供了更具挑战性的替代方案。
- **KMNIST:** 由 28×28 日文草书（Kuzushiji）字符的灰度图像组成。它通常采用与 MNIST 类似的数据划分（约 60,000 张训练图像和 10,000 张测试图像），带来独特的视觉挑战。

CIFAR-10

- **CIFAR-10:** 由涵盖 10 个类别的 32×32 RGB 彩色图像组成（例如飞机、汽车和鸟类）。其中包含 50,000 张训练图像和 10,000 张测试图像，其自然复杂性和多样性使其成为图像分类任务中颇具挑战性的基准数据集。

A. 客户端模型

为对结果进行实验验证，我们实现了两种不同的卷积网络。第一种（NetM-

[1]https://github.com/FabioTur-dev/gossip_bc_full

NIST）通用于所有源自 MNIST 的灰度数据集；而第二种（NetCIFAR）结构更复杂，用于在 CIFAR-10 等 RGB 数据集上实现所需的性能。由于硬件条件有限，我们未对 STL-10 和 ImageNet 等更大型的数据集进行进一步实验。

NetMNIST

NetMNIST 网络是一种专为 MNIST 数据集设计的卷积神经网络，该数据集包含 28×28 表示手写数字的灰度图像。其架构包括：

- **卷积层：**使用两个卷积层。第一层采用 6 个滤波器，卷积核大小为 5；第二层采用 16 个滤波器（卷积核大小同样为 5）。这些层后接 ReLU 激活和最大池化操作，从图像中提取显著特征。
- **展平：**将卷积层的输出展平，以便输入后续层。
- **全连接层：**使用由三个全连接层组成的序列，前两层分别包含 120 个和 84 个神经元，随后将数据映射到 10 个类别（从 0 到 9）。

这种简单轻量的架构非常适合数字识别，因为 MNIST 数据集具有标准化且相对简单的特点。

NetCIFAR（适用于 *CIFAR-10*）

用于 CIFAR-10 数据集的 **NetCIFAR**网络是专为该数据集开发的，该数据集包含分布在 10 个类别中的 32×32 RGB 彩色图像。其架构明显更为复杂，具有以下特点：

- **卷积结构：**采用一系列带有 3×3 内核的卷积层。第一个区块使用 32 个滤波器，第二个区块使用 64 个滤波器，第三个区块则使用 128 个滤波器。每个卷积层后都接批量归一化和 ReLU 激活，以增强训练稳定性。
- **池化：**应用两次最大池化操作，将特征图的空间维度缩减至8×8，从而便于处理更复杂的图像。
- **分类器：**展平特征图后，分类器包含一个具有 256 个神经元的全连接层，随后通过随机失活缓解过拟合，最后使用一个将特征映射至 10 个类别的线性层。

滤波器和神经元数量的增加，体现了 CIFAR-10 数据集中自然图像的复杂性和多样性有所提升，从而确保泛化能力与训练稳定性之间取得良好平衡。

狄利克雷划分

在我们的设置中，我们使用狄利克雷分布 [36] [37] 将数据集划分给客户端。这种方法通过为每个类别分配不同的比例来模拟非独立同分布数据，分配对象为不同的客户端。对于包含 K 个类别和 N 个客户端的数据集，我们为每个类别 k 生成一个概率向量

$$\mathbf{p}_k \sim \text{Dirichlet}(\beta, \beta, \ldots, \beta), \tag{6}$$

其中，β 是控制异质性程度的浓度参数。较小的 β 会产生更加不均衡的分布，而较大的 β 则会得到更加均匀、近似独立同分布的划分。

对于每个类别 k，假设共有 n_k 个样本，则分配给客户端 i 的样本数量计算如下：

$$n_{k,i} = \lfloor p_{k,i} \cdot n_k \rfloor, \tag{7}$$

其中，$p_{k,\ i}$ 是$\mathbf{p}_k$的第 i个分量。其余样本随后会随机分配给各客户端。

这种划分策略用于模拟客户端之间真实的数据异质性，这对于评估我们的联邦学习方法在非独立同分布条件下的鲁棒性和性能至关重要。

图4中给出的三幅准确率图表显示，随着 β 参数变化（β的值越小，对等节点之间的分布越异质），整体训练质量几乎保持不变。例如，FedAvg 则并非如此，其性能会随着 β 降低而显著下降。我们希望通过这些结果表明，借助完全去中心化的自动化系统，可以实现与最先进方法相当的训练性能，同时不必在可扩展性、成本或安全性方面作出取舍。使用基于以太坊的公有区块链是实现去中心化的关键要素之一；私有区块链仍由单一实体管理，因此无法实现这一特性。柱状图更清晰地展示了各对等节点在准确率方面的一致性；在本实验中，准确率本身并不重要，因为它严格受到可用硬件资源和使用场景的影响，也就是说，它取决于所用网络及其相应的超参数，以及为确保足够的系统安全性而需要的隐私预算；而隐私预算又与数据集和可用数据量密切相关。除了直接从理论上比较 FedBGS 在关键方面与最先进系统的差异外，我们还将以准确率增长量作为评估系统有效性的指标。柱状图还突显了另一个重要方面，即各个客户端对齐的收敛速度：每幅图中的黑色标记表示每次八卦迭代的准确率历史记录；在 β 值较低的情况下，标记较为稀疏，表明收敛速度较慢，且总体准确率也较低。在 FedBGS 的两个设计阶段中，根据所涉及数据集的规模和类型，解决隐私、聚类期间的可扩展性与准确率之间的权衡将至关重要。另一方面，图5展示了 CIFAR-10 数据集上各个对等节点记录的准确率值。尽管与 MNIST 变体相比，RGB 数据集带来了分类方面的挑战，但模型在收敛过程中表现出很强的鲁棒性，即使存在非独立同分布数据也是如此。

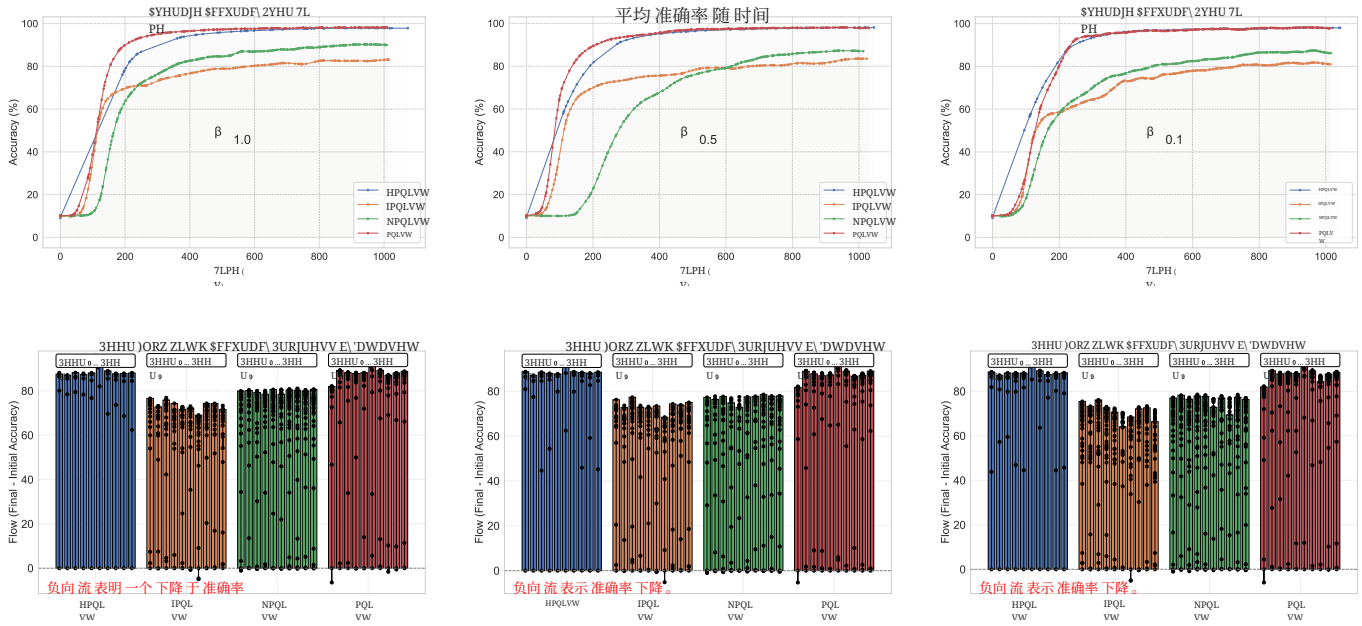


图 4：上方三张折线图报告了所有灰度数据集（MNIST 及其变体）上的平均准确率；从左到右，结果分别对应于 β=1.0、 β=0.5 和 β=0.1（后者是在对等节点之间具有高度异质性的条件下获得的）。下方显示了相应的柱状图（每张柱状图对应其上方的折线图），表示每个独立对等节点的最终值和中间值。

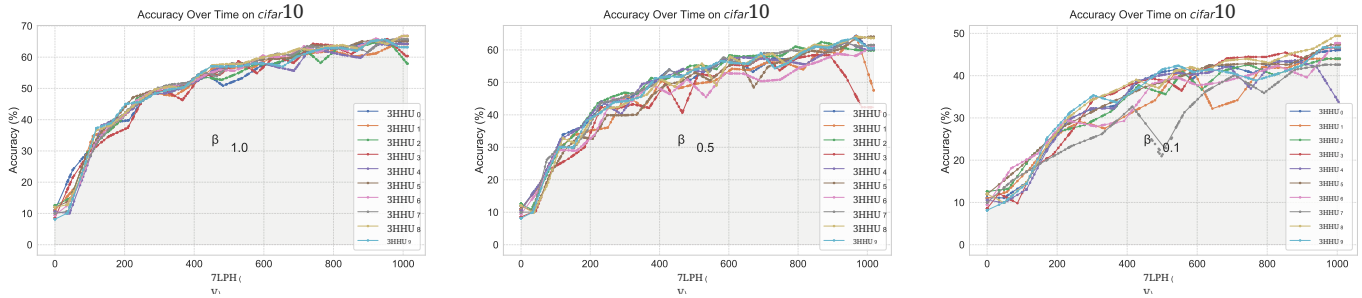


图 5：*CIFAR-10* 数据集上每个对等节点的准确率趋势。与之前的折线图一样，图中突出显示了三种不同的β 值（1.0、0.5、0.1）下的性能，以展示 FedBGS 在使用更复杂的 RGB 数据集时的鲁棒性。

表 II：采用分段式 Gossip 学习的 FedBGS 区块链使用

操作 Type	Gas 用量（Gas 单位）[a]		
	交互成本	次数	智能合约类型
智能合约部署	1.418.084	Once	#1
智能合约#2 部署	1.566.634	Once	#2
注册	100.340	适用于每个客户端	#1
重置余额（可选）	257.032	对于每个客户端	#2
令牌惩罚	77.102	变量	#2
保存哈希	50.527	对于每个客户端	#2
保存聚类中心	257.000	每次更新时	#1
将分段分配给对等节点	120.450	对每个客户端	#1
获取分段边界	35.210	每个训练轮次	#1
验证分段更新	65.800	变量（每次更新）	#2

[a]请注意，Gas 用量也取决于智能合约的优化程度。

表II中，我们展示了 FedBGS 两个阶段所涉及的各个函数消耗的以太坊 Gas。各函数

如图4所示，这些智能合约分为两种类型：#1 专用于注册和聚类功能，而 #2 用于八卦通信和惩罚阶段。虽然我们最初为 FedBGS 选择以太坊，是因为它作为一种标准，能够提供较高的安全性和丰富的功能，但我们的八卦通信系统旨在支持任何具备去中心化应用高级功能的区块链，例如卡尔达诺、索拉纳或波卡。因此，这类区块链均可用于实现 FedBGS，具体选择取决于系统的特定需求。

V. 结论

总而言之，FedBGS 通过提出一种稳健的

去中心化学习方法，在收敛方面取得了出色的成果，同时应对了联邦学习中的若干开放挑战，包括实现完全去中心化，以及管理参与方之间的异构分布数据。此外，鉴于用户不愿信任中心化结构，实现完全去中心化至关重要；而许多提出“无服务器”架构的最先进方法，在可扩展性方面都存在显著的权衡（这是许多基于区块链的方法面临的问题）。FedBGS 的未来发展将着重于减轻本地差分隐私的影响，因为在数据量较少时，即使采用调度，其也会显著降低整体性能；此外，还将研究如何使 FedBGS 适用于受限架构。这一挑战在 gossip 学习场景下尤为严峻，因为该场景不存在服务器。

致谢

本研究获得项目“数据空间、高性能计算与边缘计算时代欧洲云服务催化剂（NOUS）”的资助，资助协议编号 101135927。

参考文献

[1] H. B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y Arcas，“利用去中心化数据进行深度网络的高效通信学习”，2023 [ ] 1602 05629 。在线。可访问：https://arxiv.org/abs/。[2] X. Yin、Y. Zhu 和 J. Hu，“隐私保护联邦学习综合综述”，*ACM*计算调查 *(CSUR)*，第 54 卷，第 1–36 页，2021年。[在线]。可访问：https://api.semanticscholar.org/CorpusID:236898578。[3] Z. Ji、Z. C. Lipton 和 C. Elkan，“差分隐私与机器学习：调查与综述”，2014年。 [在线]。可访问：https://arxiv.org/abs/ 1412 7584 。[4] A. E. Ouadrhiri 和 A. Abdelhadi，“深度学习和联邦学习中的差分隐私：综述”，*IEEE Access*，第 10 卷，2022年。[5] S. M. Hosseini、M. Sikaroudi、M. Babaei 和 H. R. Tizhoosh，“用于组织病理学图像联邦学习的基于簇的安全多方计算”，2022年。 [在线]。可访问：https://arxiv.org/abs/2208.10919。[6] W. Jin、Y. Yao、S. Han、J. Gu、C. Joe-Wong、S. Ravi、S. Avestimehr 和 C. He，“Fedml-he：一种基于同态加密的高效隐私保护联邦学习系统”，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2303 10837 。[7] A. Viand、C. Knabenhans 和 A. Hithnawi，“可验证全同态加密”，2023年。 [在线]。可访问：https://arxiv.org/abs/2301.07041。[8] T. Li、A. K. Sahu、M. Zaheer、M. Sanjabi、A. Talwalkar 和 V. Smith，“异构网络中的联邦优化”，2020年。 [在线]。可访问：https://arxiv.org/abs/ 1812 06127 。[9] S. P. Karimireddy、S. Kale、M. Mohri、S. J. Reddi、S. U. Stich 和 A. T. Suresh，“SCAFFOLD：用于联邦学习的随机控制平均”，2021 [ ] 1910 06378 。在线。可访问：https://arxiv.org/abs/。[10] Q. Li、B. He 和 D. Song，“模型对比联邦学习”，2021 [ ] 2103 16257 。在线。可访问：https://arxiv.org/abs/。[11] D. A. E. Acar、Y. Zhao、R. M. Navarro、M. Mattina、P. N. Whatmough 和 V. Saligrama，“基于动态正则化的联邦学习”，2021 [ ] 2111 04263 。在线。可访问：https://arxiv.org/abs/。[12] Y. LeCun、L. Bottou、Y. Bengio 和 P. Haffner，“应用于文档识别的基于梯度的学习”，*IEEE*会议录，第 86 卷，第 11 期，第 2278–2324 页，1998年。[13] H. Xiao、K. Rasul 和 R. Vollgraf，“Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集”，*arXiv* 预印本 *arXiv:1708.07747*，2017年。[14] A. Krizhevsky，“从微小图像中学习多层特征”，多伦多大学，技术报告，2009年。[15] G. Clanuwat、H. Kuratani 和 Y. Wada，“Kuzushiji-MNIST：用于现代日语（くずし字）识别的新型基准数据集”，2018年。[16] G. Cohen、S. Afshar、J. Tapson 和 A. van Schaik，“EMNIST：将 MNIST 扩展至手写字母”，载于2017年国际神经网络联合会议（IJCNN）。IEEE，2017年，第 2921–2926 页。[17] X. Bao、C. Su、Y. Xiong、W. Huang 和 Y. Hu，“FLchain：一种用于可审计联邦学习、具备信任和激励机制的区块链”，载于2019年第五届大数据计算与通信国际会议（BIGCOM），2019年，第 151–159 页。[18] Y. Li、C. Chen、N. Liu、H. Huang、Z. Zheng 和 Q. Yan，“一种基于委员会共识的区块链去中心化联邦学习框架”，IEEE Network，第 35 卷，第 1 期，第 234–241 页，1月。2021 [ ] 10 1109 011 2000263 。在线。可访问：http://dx.doi.org/。/MNET.。[19] H. Kim、J. Park、M. Bennis 和 S.-L. Kim，“基于区块链的设备端联邦学习”，2019年。 [在线]。可访问：https://arxiv.org/abs/ 1808 03949 。[20] Q. Zheng、Y. Li、P. Chen 和 X. Dong，“一种用于区块链的创新型基于IPFS的存储模型”，载于2018年IEEE/WIC/ACM国际网络智能会议（WI），2018年，第 704–708 页。[21] J. Benet，“IPFS——内容寻址、版本化的点对点文件系统”，2014年。 [ ] 1407 3561 在线。可访问：https://arxiv.org/abs/。[22] X. Wu、Z. Wang、J. Zhao、Y. Zhang 和 Y. Wu，“FedBC：基于区块链的去中心化联邦学习”，载于2020年IEEE人工智能与计算机应用国际会议（ICAICA），2020年，第 217–221 页。[23] H. Zhang、S. Jiang 和 S. Xuan，“基于区块链的去中心化联邦学习：概念、框架与挑战”，Computer Communications，第 216 卷，第 140–150 页，2024年。 [在线]。可访问：https://doi.org/ 10 1016 2023 12 042 。/j.comcom.。。[24] I. Hegedus、G. Danner 和 M. Jelasity，“作为联邦学习去中心化替代方案的 gossip 学习”，载于IFIP分布式应用与互操作系统国际会议，2019年。 [在线]。可访问：https://api.semanticscholar.org/CorpusID:174800884。[25] I. Heged˝us、G. Danner 和 M. Jelasity，“去中心化学习行之有效：gossip 学习与联邦学习的实证比较”，Journal of Parallel and Distributed Computing，第 148 卷，第 109–124 页，2021年。 [在线]。可访问：https://www.sciencedirect.com/science/article/pii/S0743731520303890。[26] C. Thapa、M. A. P. Chamikara、S. Camtepe 和 L. Sun，“SplitFed：联邦学习与拆分学习相遇”，2022年。 [在线]。可访问：https://arxiv.org/abs/ 2004 12088 。[27] P. Vepakomma、O. Gupta、T. Swedish 和 R. Raskar，“面向医疗健康的拆分学习：无需共享原始患者数据的分布式深度学习”，2018 [ ] 1812 00564 。在线。可访问：https://arxiv.org/abs/。[28] C. Hu、J. Jiang 和 Z. Wang，“去中心化联邦学习：一种分段式八卦方法”，2019年。 [在线]。可访问：https://arxiv.org/abs/ 1908 07782 。[29] Z. Tang、S. Shi、B. Li 和 X. Chu，“GossipFL：一种采用稀疏化自适应通信的去中心化联邦学习框架”，IEEE Transactions on Parallel and Distributed Systems，第 34 卷，第 3 期，第 909–922 页，2023年。[30] A. Janjua、S. Dhalla、S. Gupta 和 S. Singh，“一种区块链赋能的去中心化 gossip 联邦学习框架”，载于2023年国际网络与通信会议（ICNWC），2023年，第 1–7 页。[31] A. R. Elkordy、Y. H. Ezzeldin、S. Han、S. Sharma、C. He、S. Mehrotra 和 S. Avestimehr，“联邦分析：综述”，2023年。 [在线]。可访问：https://arxiv.org/abs/ 2302 01326 。[32] D. Arthur 和 S. Vassilvitskii，“k-means++：谨慎初始化的优势”，载于第十八届 ACM-SIAM 离散算法年度研讨会论文集。SIAM，2007年，第 1027–1035 页。[33] 维基百科贡献者，“默克尔树——维基百科，自由的百科全书”，https://en.wikipedia.org/wiki/Merkle tree#， [在线；访问于25-May-2025]。[34] Z. Cao 和 L. Liu，“重新审视 Paillier 密码系统及其若干变体”，2015 [ ] 1511 05787 。在线。可访问：https://arxiv.org/abs/。[35] D. Cajaraville-Aboy、A. Fernandez-Vilas、R. P. Dıaz-Redondo 和 M. Fern´andez-Veiga，“用于保障去中心化联邦学习安全的拜占庭鲁棒聚合”，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2409 17754 。[36] B. Osting 和 T. H. Reeb，“狄利克雷分区的一致性”，2017年。 [ ] 1708 05472 在线。可访问：https://arxiv.org/abs/。[37] Q. Li、Y. Diao、Q. Chen 和 B. He，“非独立同分布数据孤岛上的联邦学习：一项实验研究”，2021年。 [在线]。可访问：https://arxiv.org/abs/2102.02079