联邦学习隐私保护机制综述

摘要：随着数据孤岛的出现和隐私意识的增强，传统的中心化的机器学习模式遇到了一系列挑战。联邦学习作为一种新兴的隐私保护的分布式机器学习模型迅速成为一个热门的研究问题。有研究表明，机器学习模型的梯度会泄露用户数据集的隐私，能够被攻击者利用以获取非法的利益，因此，需要采用一些隐私保护的机制来保护这种敏感信息。研究了当前联邦学习系统中采用的隐私保护机制，并根据研究者们采用的隐私保护技术，将联邦学习中的隐私保护机制主要分为了五类，总结了不同的隐私保护机制的研究思路和研究进展。通过对当前联邦学习中使用的隐私保护机制的研究，联邦学习系统的设计人员提高联邦学习系统的安全性，更好的保护数据隐私。

关键词：联邦学习；隐私保护；

Abstract：With the emergence of data islands and the enhancement of privacy awareness, the traditional centralized machine learning model has encountered a series of challenges. Federated learning, as a privacy-protected distributed machine learning model, has quickly become a hot research topic. Studies have shown that the gradient of the machine learning model will leak the privacy of user data sets and can be used by attackers to obtain illegal benefits. Therefore, some privacy protection mechanisms are needed to protect this sensitive information. We study the privacy protection mechanisms used in the current federal learning system, and according to the privacy protection technology adopted by the researchers, divides the privacy protection mechanisms in the federal learning into five categories, and summarizes the ideas of different privacy protection mechanisms. We want to introduce the current privacy protection mechanisms used in federated learning to enable designers of federated systems to improve the security of federated learning systems and protect data privacy.

Keywords：Federated Learning；privacy protection

引言

近年来，机器学习算法在人工智能领域内迅猛发展，在计算机视觉、自然语言处理和推荐算法等领域都有良好的表现。这些机器学习技术的成功，都是通过大量数据的训练得到的，然而，一方面，随着人工智能在各个行业的不断落地，人们对于数据安全和隐私保护的程度在不断提高。另一方面，在法律层面，法律制定者和监管机构也出台了新的法律来规范数据的管理和使用，法律的实施进一步加剧了在不同组织之间收集和分享数据的困难。因此，各方面原因导致在许多应用领域，满足机器学习规模的数据量是难以达到的，人们不得不面对难以桥接的数据孤岛。

为了解决上述数据孤岛问题，一种可行的方案是谷歌的 McMahan 等人于 2016 年提出的联邦学习（federated learning）[1]。联邦学习是多个实体（客户端）协作解决机器学习数据孤岛问题的一种方案，它在一个中央服务器或服务提供商的协调下进行，每个客户端的原始数据存储在本地，无法交换或迁移，满足了隐私保护和数据安全的要求。谷歌的联邦学习系统能够通过智能手机的私人数据来更新其 Gboard 系统（一种虚拟键盘系统）的输入预测模型，所有使用 Gboard 的智能手机的数据都可以被用来优化模型，而这一过程并不需要将用户的隐私数据从智能手机发送到中央设备。

本文主要就联邦学习中的隐私保护机制进行了调查和研究，从隐私保护技术的角度对当前的研究联邦学习中的隐私保护的文献进行了分类，目前联邦学习中主要采用隐私保护技术有差分隐私、同态加密和安全多方计算，部分文献也混合使用了上述隐私保护技术的两种或三种，也有部分文献使用了其它隐私保护技术。

本文的主要内容如下：

第 1 节介绍了当前联邦学习面临的主要隐私保护问题，隐私需求部分解释了为什么需要保护联邦学习中的梯度，威胁模型部分主要介绍了联邦学习系统面临的隐私和安全攻击手段。

第 2 节主要介绍了现有的隐私保护及技术，主要有差分隐私、同态加密和安全多方计算三种。

第 3 节对现有联邦学习中隐私保护技术的研究进行了分类并且介绍了隐私保护技术的基本原理和研究现状，将主要从差分隐私、同态加密、安全多方计算、混合模式和其它方法，一共五个方面进行介绍。

第 4 节将总结本文的工作和未来联邦学习隐私保护问题的研究方向。

1　联邦学习的隐私保护问题

1.1 隐私需求

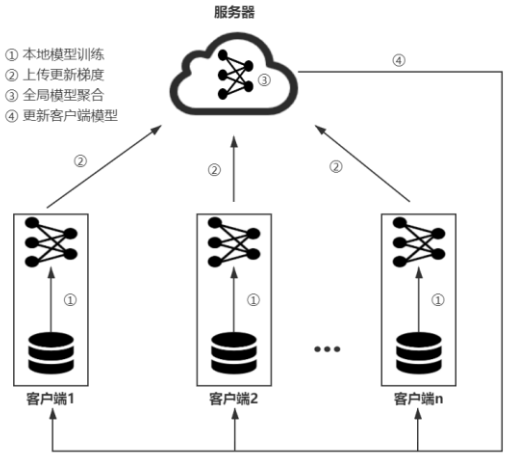

Fig.1　Client-Server Architecture

图 1 客户端-服务器架构

谷歌提出了一种基于客户端-服务器架构的典型的横向联邦学习过程，参与训练的移动设备数据具有类似的数据特征，但是每个移动设备的数据量是不同的，如图 1 所示。

客户端-服务器的联邦学习过程如下：客户端使用本地数据和收到的初始化模型进行模型训练，得到更新的模型梯度；客户端将模型梯度上传到聚合服务器；服务器进行模型聚合，可以以客户端数据集的规模为权

重进行模型平均；服务器将新的模型参数下发给客户端，客户端更新本地模型。

在客户端-服务器的联邦学习的联邦学习过程中，用户数据没有离开用户的设备，数据隐私得到了一定程度的保护，但是，Zhu 等人的研究结果表明分享梯度会泄露隐私数据[2]，服务端能够根据客户端上传的梯度恢复出客户端的训练数据。因此，在联邦学习过程中，客户端向服务器分享本地模型的更新梯度存在敏感数据泄露的隐患。联邦学习将需要传输的训练数据转化为了训练后得到的梯度，需要进行隐私保护的敏感数据也随之变成了梯度，客户端需要使用隐私保护技术对梯度进行保护，可以使用的技术有：同态加密、安全多方计算和差分隐私等。

1.2 威胁模型

联邦学习不可能总是提供足够的隐私保证，也会遭受潜在的隐私攻击，危及模型和数据的完整性。Lyu 等人[3]详细的介绍和总结了联邦学习面对的威胁和攻击。

在联邦学习的威胁模型中，通常会受到来自内部或外部的攻击，其中内部攻击往往比外部攻击要强。对于联邦学习的内部攻击，可以采取单一攻击，拜占庭攻击和女巫攻击。

根据主动性可以将对手分为半诚实的对手和恶意的对手，半诚实的对手被认为是诚实但好奇的，在不违背联邦学习协议的情况下试图了解其他方的私密状态；而一个活跃或恶意的对手可以通过修改、重放或删除消息来随意违背联邦学习协议，还可以进行毁灭性的攻击。

联邦学习可以分为训练和推理两个阶段，所以每个阶段受到的攻击也不同。训练阶段的攻击企图学习、影响和破坏模型本身，对手可以用数据投毒攻击和模型投毒攻击，还可以对参与者的更新发动一系列推理攻击。推理阶段的攻击通常不会篡改目标模型，只会导致模型产生错误的输出，或收集关于模型特征的证据。

投毒攻击主要有本地数据采集时的数据投毒攻击和在局部模型训练过程中的模型投毒攻击。数据投毒可以分为 clean-label 和 dirty-label 攻击，clean-label 攻击假设对手无法改变任何训练数据的标签；相比 dirty-label 攻击中，对手可以将一些它希望用所需目标标签误分类的数据样本引入到训练集中。dirty-label 投毒攻击的常见方式是标签翻转攻击和后面攻击。模型投毒攻击的目标是在将本地模型更新发送到服务器之前进行投毒，或者在全局模型中插入隐藏的后门。

推理攻击则可以分为成员推理攻击和属性推理攻击。成员推理攻击的目的是确定某一数据点是否被用于训练模型，而属性推理攻击的用来推断其他参与者的训练数据的属性。

2 隐私保护技术

2.1 差分隐私

差分隐私是由 Dwork 等人[4]在 2006 年首次提出的，它提供了量化和限制个人信息泄露的一种输出隐私保护模型。差分隐私的中心思想是，当攻击者试图从数据集中查询个体信息时将其混淆，使得敌手无法从查询结果中辨别个体的敏感性，即函数的输出结果对于数据集中的任何特定记录都不敏感，因此，差分隐私能被用来抵抗成员推理攻击。

定义 1（(ϵ,δ) −差分隐私）一个随机化机制$\boldsymbol{\mathcal{M}}$，其定义域为$\mathbb{N}^{|\boldsymbol{x}|}$，如果$\boldsymbol{\mathcal{M}}$满足$(\epsilon,\delta)$ −差分隐私，那么对于任意的输出集合$\boldsymbol{S}\subseteq \mathrm{Range}(\boldsymbol{\mathcal{M}})$和两个最多只有一个元素不同的相邻数据集$\boldsymbol{x}$和$\boldsymbol{x'}$，有：

$$Pr[\boldsymbol{\mathcal{M}}(\boldsymbol{x})\in \boldsymbol{S}]\le e^{\varepsilon}Pr[\boldsymbol{\mathcal{M}}(\boldsymbol{x'})\in \boldsymbol{S}]+\delta \quad (1)$$

(1)式中，ϵ表示隐私预算；δ表示失败概率。一般而言，ϵ越小，隐私保护程度越高，噪声越大，数据可用性越差。

差分隐私的主要实现方式是向数据添加噪声，主要有两种方式，一种是根据数值型的输出函数的敏感度添加噪声，比如基于l_1−敏感度的拉普拉斯噪声（laplace noise）和l_2−敏感度高斯噪声（gaussian noise）；另一种是根据离散值的指数分布选择噪声。

2.2 同态加密

同态加密（homomorphic encryption，HE）是一种加密形式，此概念是 Rivest 等人在 20 世纪 70 年代首先提出的[5]，一个同态加密方法 H 通常由四部分组成，即

$$H = \{KeyGen, Enc, Dec, Eval\}$$

其中，$KeyGen$为密钥生成函数，在非对称加密中，输入密钥生成源$\boldsymbol{g}$，输出用于加密的密钥对$\{\boldsymbol{pk}, \boldsymbol{sk}\} = KenGen(\boldsymbol{g})$，$Enc$为加密算法将明文$\boldsymbol{m}$作为输入，输出密文$\boldsymbol{c} = Enc(\boldsymbol{m})$，$Dec$为解密算法，即输入密文$\boldsymbol{c}$，输出明文$\boldsymbol{m} = Dec(\boldsymbol{c})$，$Eval$为评估函数，将密文$\boldsymbol{c}$和公钥作为输入，输出与明文相对应的密文。

它允许数据的接收者能够对加密后的信息进行特定形式的代数运算得到的仍然是加密的结果，将其解密之后所得到的结果与对明文进行同样的运算所得到的结果一致。可以形式化的表示为：

$$Enc(\boldsymbol{m_1} \circ \boldsymbol{m_2}) = Enc(\boldsymbol{m_1}) \circ Enc(\boldsymbol{m_2}) \quad (2)$$

(2)式中∘为特定形式的代数运算。

换言之，这项技术令人们可以在无需对数据进行解密的情况下对加密的数据中进行诸如检索、比较等操作时得出正确的结果。其意义在于，真正从根本上解决将数据及其操作委托给第三方时的保密问题，例如对于各种云计算的应用。

同态加密一直是密码学领域的一个重要课题，以往人们只找到一些部分实现这种操作的方法。而 2009 年 9 月 Gentry 从数学上提出了“全同态加密”（fully homomorphic encryption，FHE）的可行方法[6]，即可以在未解密的情况下对加密数据进行任何可以在明文上进行的运算，使这项技术取得了决定性的突破。人们正在此基础上研究更完善的实用技术，这对信息技术产业具有重大价值。

2.3 安全多方计算

安全多方计算（secure multi-party computation，SMPC）最早是由图灵奖获得者、中国科学院院士姚期智于 1982 年正式提出的[7]， 其目的是从每一方的隐私输入中协作计算一个函数的结果，而不用将这些输入展示给其它方。安全多方计算保证了参与方获得正确结果的同时，无法获得计算结果之外的任何信息。

安全多方计算可以通过三种不同的框架来实现：不经意传输（oblivious transfer，OT）、秘密共享（secret sharing，SS）和阈值同态加密。本文将在下面内容中简单介绍不经意传输和秘密共享。

不经意传输是一种由 Rabin 在 1918 年提出的两方安全协议[8]，在不经意传输中，发送方拥有一个“索引-消息”对，每次传输时，接收方选择一个索引，并接受对应的消息。接收方不能得知数据库的任何其它消息，发送方也不了解接受方选择的索引的任何信息。

秘密共享是指将秘密信息分割成若干份，然后将这些秘密份额分给不同的人保管，以达到隐藏秘密和风险分散的目的。一般来说，一个秘密分享方案包含了一个秘密分割算法和一个秘密重建算法，包含了分发者、持有者和接受者三种角色。分发者负责将秘密信息通过秘密分割算法进行分割，并发送给持有者。接收者在需要秘密的时候对秘密进行重建，收集一组持有者的秘密份额，并执行秘密重建算法来恢复秘密，当有足够的秘密份额时就可以得到秘密信息。常用的一种秘密共享的方法是 shamir 秘密共享[9]（shamir’s secret sharing），是基于多项式方程构建的。

3 联邦学习隐私保护机制

联邦学习中目前主要使用的隐私保护机制有三种：差分隐私、同态加密和安全多方计算，部分研究人员也混合使用了两种或者三种来实现更好的隐私保护，部分研究人员也采用了一些其它的隐私保护方式。本节将从这个五个方面来介绍联邦学习中的隐私保护机制。

3.1 差分隐私

差分隐私机制通常被应用到联邦平均算法 FedAvg[10]中，算法的架构如图 1 所示，客户端直接将梯度发送给服务器会向服务器泄露隐私信息，服务器可以通过梯度来推断客户端的数据集中有无某个样本，差分隐私算法可以应用到联邦平均算法的过程中，通过添加噪声的方式以抵抗推理攻击。差分隐私算法可以根据噪声添加的位置分为用户级和样本级，样本级差分隐私算法也可以称为本地差分隐私（local different privacy，LDP）。

McMahan 等人[11]率先将差分隐私技术应用到联邦平均算法中。在本地客户端更新过程中，使用参数裁剪来限制样本梯度的大小，在聚合全局模型时，使用差分隐私的高斯机制对缩放后的模型更新添加噪声，使用 Moment accountant 算法[12]计算总体隐私损失，实现了用户级别的隐私保护，其实验结果表明，在一个具有足够多用户的数据集上，实现差分隐私是以增加计算量为代价的，在大型数据集上进行训练时，私有 LSTM 语言模型在数量和质量上都与无噪声模型相似。Geyer 等人[13]也使用了类似的方法在客户端级别使用差分隐私来隐藏客户端的贡献，有所不同的是，前者在本地客户端进行训练过程中进行参数裁剪，而后者在中央服务器进行参数裁剪，二者参数裁剪的阈值也不相同。Geyer 等人还指出在联邦学习训练过程中动态调整差分隐私机制可以提高模型性能，并通过实验表明，在有足够数量的参与客户的情况下，其程序可以在模型性能上以较小的成本维持客户端级别的差异隐私。

Thakkar 等人[14]进一步改进了参数裁剪过程，提出了自适应分位数裁剪策略，该策略是为迭代隐私机制设计的，用于训练具有用户级别的差分隐私的联邦学习模型，从而无需进行大量参数调整，然后描述了特定层的噪声添加策略，后者比基本策略具有更高的实用性。

Agarwal 等人[15]对二项机制进行了改进和扩展，应用到了分布式向量平均估计（distributed mean estimation，DEM）问题中，在联邦学习中，分布式向量可以是模型梯度或参数），Agarwal 等人的主要思想是：对于每个客户端，在给服务器发送梯度之前，将经过适当参数化的二项式分布得出的噪声添加到每个量化值，服务器进一步减去噪声引入的偏差，以实现无偏差的均值估计器。还进一步表明，随机轮换有助于减少因差分隐私导致的附加错误。

Bhowmick 等人[16]针对大规模分布式学习和联邦学习，提出了一种针对不同隐私要求的统计学习问题的最小最大（minimax）差分隐私机制，提出了一个新的联邦学习隐私保护框架：在客户端本地计算模型参数更新的过程中，使用自身提出的本地隐私保护机制保护本地数据，在中央服务器执行聚合过程中，使用差分隐私保证模型参数的通信过程是全局私有的。整个反馈环路提供了有效的隐私保护，用户的本地数据不会传输到中央服务器，而集中式隐私保护能够保证过程和最终参数都不存在敏感性披露的风险。

Li 等人[17]研究了元学习背景下的隐私问题，提出了一种基于梯度元学习的差分隐私算法，该算法在每次任务内迭代时使用了高斯机制的差分隐私梯度下降，保护了任务中的单个样本的隐私，在凸环境下具有学习保证，并证明了在联邦语言模型和小样本图

像分类任务中具有出色的性能。

Triastcyn 等人[18]提出了一种新的贝叶斯差分隐私框架和一种实用的隐私损失计算方法，贝叶斯差分隐私是对类似分布数据的差分隐私的放松版本。文献将差分隐私框架应用到联邦环境中，并使用提出的隐私损失计算方法估算客户的隐私保证；文献强调了样本级差分隐私的重要性，并提出了估计样本级隐私损失的两种变体；最后，文献介绍了一种新的联合计算隐私损失的方法，可以同时估计客户端级别和样本级别的隐私。

Liu 等人[19]针对联邦梯度下降算法提出了一个两阶段的本地差分隐私框架 FedSel，其关键思想是首次尝试通过延迟不重要的梯度减轻了注入噪声的维度问题。FedSel 有两个阶段组成，分别是“维度选择”和“值扰动”，文献列举了三种“维度选择”的方法，并提供了对应的隐私保证。从理论上分析了 FedSel 的隐私、准确度和时间复杂度，其性能优于最新的解决方案，在现实世界和合成数据集上进行实验测试了 FedSel 框架的有效性和效率。

Truex 等人[20]提出了 LDP-Fed，使用本地差分隐私（local differential privacy,LDP）提供正式的隐私保证。LDP-Fed 主要有两个新的组件：在每个客户端运行的本地差分隐私模块和 k-客户端选择模块，前者使用指数机制和 CLDP（本地差分隐私的变体）提供隐私保护，后者用于在迭代 LDP-Fed 训练过程的各个回合中选择性共享模型参数更新。

Liang 等人[21]将基于拉普拉斯平滑（laplaian smoothing）的训练效果增强方案应用到差分隐私联邦学习过程中，在梯度聚合时注入了高斯噪声，在模型精度上得到了改善，并为均匀二次抽样和泊松二次抽样提供了严格的封闭形式的隐私边界，得出了差分私有联邦学习的差分隐私保证。

Sabater 等人[22]提出了一种新颖的差分私有的平均协议——GOPA（gossip noise for private averaging, GOPA）。该协议可以匹配可信管理者设置的准确性，同时自然的扩展到大量用户。客户端需要将两种噪声加到了私有值上，第一种噪声是客户端与相邻客户端交换的高斯噪声，该噪声在聚合时可以相互抵消；第二种噪声是独立的服从高斯分布的噪声，无法消除，是对一种噪声的补充，文献分析了 GOPA 协议的差分隐私保证和在面对恶意行为时的正确性。另外，文献通过承诺方案和零知识证明，保证了用户能够证明他们计算的正确性，而又不影响协议的效率和保密性。

3.2 同态加密

同态加密作为计算机安全领域和密码学领域一项热门的隐私保护手段被人们广泛应用，但由于加密算法需要大量的计算成本，在过去只有少量应用于机器学习领域的尝试，但由于技术的不断进步，以及社会的需求，人们对于隐私保护的需求远远大于对计算资源的需求，进而也逐步应用到联邦学习之中。在联邦学习中，同态加密大多运用在对客户端所上传的梯度进行保护，具体流程基本如图 2 所示。

Hardy S 等人[23]就提出了基于纵向联邦学习的隐私保护两方的逻辑回归算法，该算法指定了一个协调方，以及两个数据提供者，利用 Paillier 算法[24]进行安全梯度下降，来训练逻辑回归模型，其中利用 Paillier 算法的加密的掩码和各方计算得到的中间数据进行加法运算以及常数乘法运算。而在安全梯度下降算法中，双方交换加密后的中间结果掩码。最后则是将加密的梯度信息发送给协调方进行解密以及模型的更新。

Zhang 等人[25]针对跨组织的场景下，提出加速基于同态加密的安全聚合方案 BatchCrypt，并减少一定量的通信开销。

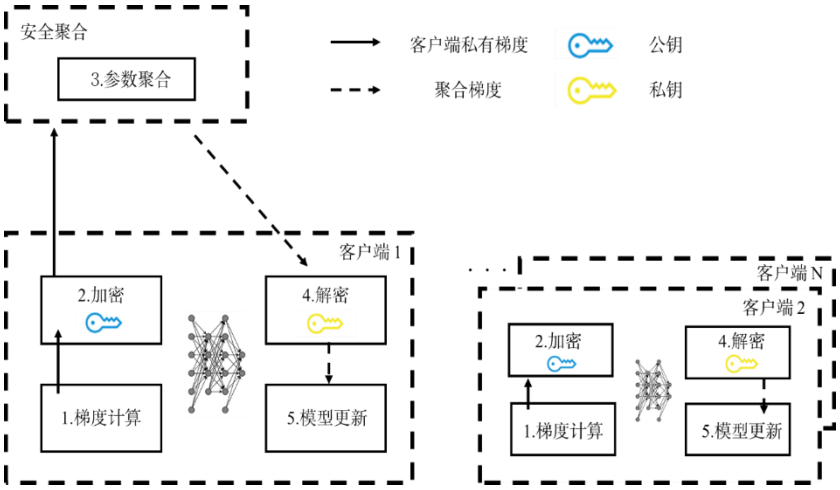

Fig.2 Training Process Of Federated Learning Model Based On Homomorphic Encryption

图 2. 基于同态加密的联邦学习模型训练过程

Zhang 等人设计了一种梯度量化的方法，其加密方案是在量化梯度的基础上，将多个量化梯度编码到一个整数，对整数进行进一步的同态加密操作。BatchCrypt 对比微众银行所提出的 FATE（federated AI technology enabler,FATE）架构中的 stock 方法有很大的性能提升。但是在和明文方案对比，性能还是有些差距，尤其是当模型规模比较大时，性能差距则更加明显。

Liu 等人[26]提出了一个基于分类与回归（classification and regression Trees, CART）树和 bagging 的联邦森林框架，它既能处理分类问题，又能处理回归问题，称为联邦森林（federated forest，FF）。FF 利用分布式模型存储策略的预测算法，每次预测只需对整个森林的每棵树进行一轮集体通信，可以大大减少通信开销，从而提高预测效率。但 FF 在处理分类问题的过程中，即使标签已编码，仍然可以猜测出真实的标签值，尤其是对于二叉树分类；而对于回归问题，即使可以使用同态加密对标签进行加密，建模也将非常耗时。与此类似利用决策树的还有 Cheng 等人[27]提出的 SecureBoost，这也是一种基于联邦学习的无损隐私保护树增强系统。SecureBoost 应用于纵向联邦学习场景，主要包括数据对齐和训练构造 Boost 树两方面，对于数据对齐方面，提出了一种在不泄露数据标识的情况下寻找数据交集的方法；对于训练构造 Boost 树方面，特征分割的选择和叶子的权重的计算仅依赖于标签，这是由主动方所提供的，为了保证数据安全，使用同态加密技术进行处理，且为了能够使用同态加密，对数据进行处理以除去无法进行同态加密的运算。

Mandal K 等人[28] 设计了一种称之为 PrivFL 的用于联邦环境下对预测模型(如线性和逻辑回归)进行训练和遗忘预测的隐私保护系统，该方案能够在用户退出的情况下保证系统的鲁棒性。该系统基于一个加法同态加密方案和一个聚合协议，设计了分别用于训练线性和逻辑回归模型的隐私保护协议，同时利用联合学习的训练算法，设计出对活跃用户数据的安全多方全局梯度计算。该方案考虑了三种不同的威胁模型:只针对用户的对手；纯服务器对手；和用户-服务器对手，故使用加密的方式保护用户隐私，但由于是采用的加密方式，相较于其他的隐私保护手段可能会占用更大的通信开销。并且，Mandal 等人特意考虑到了在移动端的联邦学习，通过迭代执行使用适合移动应用的轻量级加密原语构建的安全多方全局梯度协议，使得 PrivFL 实现了健壮和安全的训练过程。

3.3 安全多方计算

在联邦学习过程中，很多情况下的模型聚合过程可以视为在不泄露任何一个参与方的私有信息的情况下，安全计算一个函数的值，这与安全多方计算的思想不谋而合。

Bonawitz 等人[29][30]使用安全多方计算在横向联邦学习过程中进行安全聚合，并支持客户端中途退出，每个客户端对要传输的私有向量加上两个掩码，得到掩饰值，将两个掩码的生成种子使用秘密分享的方式分享给其它用户，将掩饰值传输到服务器端进行聚合，服务器端根据仍然存在的客户端重建掉线客户端的生成种子，消除掉线客户端的掩码对聚合值的影响，得到完美的聚合值。掩饰值保证了数据的私密性，秘密分享提高了协议的鲁棒性，消除掉线用户的掩码对聚合值的影响。

Sanil 等人[31]提出了一种在跨组织的纵向联邦学习的场景下，训练回归模型的方法，每个参与方拥有部分属性值，在每个参与方不泄露自身数据的情况下，共同建立一个回归模型。在训练回归模型的过程中，主要使用了 Powell 优化算法和一个简单的安全求和协议，安全求和协议指定第一个参与方作为负责人，生成一个随机数，然后每个参与方把自己的私有值加到随机数上，之后发送给下一个参与方，最后的结果将被重新发送给第一个参与方，第一个参与方将生成的随机数从结果中减去，得到完美的聚合值。

Zhao 等人[32]提出一种安全的成员选择策略（Secure Member Selection Strategy，SMSS），在训练之前评估成员的数据质量。SMSS 利用 Shamir 的方案共享对称密钥，这可以避免恶意客户端或不正确的数据集在不进行数据交换的情况下对模型进行训练；SMSS 还使用 PSI（Private Set Intersection）解决了实时秘密块分发问题，并引入了 RANSAC 算法来从并非所有正确的机密块中恢复正确的密钥。Zhao 等人通过严格的分析显示了 SMSS 的可行性和安全行，通过实验证明了 SMSS 的高效性和鲁棒性。

Xu 等人[33]提出了一种使用基于函数加密的 SMC 协议保护隐私的联合学习的方法——HybridAlpha。HybridAlpha 引入了一个可信的第三方验证结构（Third Party Authority，TPA），TPA 负责生成公钥和私钥，并向每个参与聚合的客户端发送不同的公钥，客户端完成本地的训练后，使用差分隐私机制为本地模型参数添加噪声，然后使用得到的公钥加密带有噪声的参数并发给聚合服务器，服务器收到来自客户端的一定数目的向量后，生成一个权重向量并发送给 TPA，TPA 根据权重向量生成私钥返回给聚合服务器，聚合服务器可以根据函数加密的解密算法计算聚合值，而不会得到每个服务器的具体信息。HybridAlpha 为了防止好奇的服务利用权重向量进行推理攻击，在 TPA 中还附加了一个推理预防模块，对权重向量进行审查。Xu 等人通过实验验证 HybridAlpha 可以提供与现有方案相同的模型性能和隐私保证，并可以减少训练时间和通信量。

He 等人[34]提出了一种安全的两服务器协议，该协议可提供输入隐私和拜占庭的鲁棒性，具有通信效率高，容错能力强并且可以保证本地差分隐私。框架使用秘密分享来保护隐私，具体过程是客户端将私有梯度向量随机划分为两份，分别发送给两个服务器 A 和 B，服务器 B 使用拜占庭鲁棒性协议选择部分客户端并生成客户端索引的二进制向量，将向量秘密分享给服务器 A，两个服务器根据二进制向量聚合客户端的秘密份额，然后服务器 B 将结果发送给服务器 A 得到完整的聚合值，可以在秘密分享过程中添加噪声以获得本地差分隐私。

3.4 混合模式

由于使用单一的隐私保护手段都存在着一些不足，或无法保证无损训练，或占用大量的计算通信资源，为了设计出适用性更好的联邦学习架构，研究者们尝试融合多种隐私保护手段来达到这一目的。

Truex 等人[35]提出了一个新的联邦学习系统，在保证数据隐私的基础上有着比普通的联邦学习系统更高的准确率，同时该联邦学习系统中包含一个可调的参数，并表示可以通过这个参数在系统的准确性和隐私性做一个权衡。Truex 等人组合使用了差分隐私技术、同态加密技术和安全多方计算技术，使用差分隐私保护客户端数据集的隐私，使用同态加密进行安全聚合，安全多方计算技术被用来减少差分隐私中的噪声以提高训练出来的模型的准确率。

Liu 等人[36]提出了一种用于移动人群感知的保护隐私的联邦极端梯度增强方案 FedXGB（federated extreme gradient boosting scheme，FedXGB），该方案通过结合同态加密和秘密共享，来防止恶意参与方的攻击，即每个用户先通过密钥协商函数计算出与其他用户共享的掩码密钥，并秘密共享它的私有掩码密钥，随机选择一个随机值r_u通过$SecMask$函数计算掩码值$[[x_u]]$。服务器记录接收到的掩码值和发送者，然后发布发送者的列表。列表中的活跃用户返还g^{r_u}和退出用户的私有掩码密钥的共享；服务器通过共享恢复出退出用户的私有掩码密钥，计算退出用户与其他用户之间的共享掩码密钥；最后，服务器计算得到聚合结果。

Nikolaenko 等人[37] 提出一种结合了同态加密和姚氏混淆电路的混合岭回归方法来保护隐私，其中包含参与方与评估方。该回归算法可以分为两个阶段，第一阶段利用同态加密来处理计算的线性部分，第二阶段利用混乱电路处理非线性部分，执行回归算法的其余操作。该系统利用 Paillier 作为加性同态系统，使用 FastGC[38]作为底层的姚氏混淆电路的框架，构建了一个真实的系统。

Xu 等人[39]提出了名为 VerifyNet 的框架，设计了一种基于同态哈希函数和伪随机技术的方案来支持每个用户验证服务器返回结果的正确性；然后采用一种双掩码协议保证联邦学习过程中用户局部梯度的机密性，同时能够允许用户中途退出，并且这些退出的用户的隐私仍然受到保护，但对于云服务器，需要对所有用户的加密梯度进行聚合，并在屏蔽轮中还原所有在线用户的秘密，导致计算和通信开销较大。

Hao 等人[40] 提出了一种框架称之为隐私增强联邦学习（Privacy-Enhanced Federated Learning，PEFL）。PEFL 在每聚合中都是非交互式的，将私有梯度的同态密文（同态加密使用的是 BGV 方案，消除了密钥交换操作并增加了明文空间）嵌入到 A-LWE 中，实现安全的聚合协议。同时该框架为了进一步防止隐私从局部梯度和共享参数中泄露，还使用了分布式高斯机制实现的差分隐私技术对梯度进行扰乱。PEFL 是一种后量子安全且非交互式的协议，即使多个实体相互串通，也可以防止私有数据泄露。可以抵御推理攻击，模型反转攻击，适合于大规模用户场景。

3.5 其它

3.1-3.4 节讨论了使用差分隐私技术、安全多方计算技术、同态加密技术在联邦学习中进行隐私保护，在联邦学习系统中，除了这些隐私保护方法，还有一些其它的隐私保护技术，比如使用可信执行环境、哈希函数等等。

Chen 等人[41]提出使用可信执行环境（trusted execution environment，TEE）来抵抗模型投毒攻击。TEE 可以依靠公共超参数，强制每个不诚实的参与者在本地运行标准的 SGD 算法；另外，还可以提供证明和密封的能力，安全区域内的代码可以获取使

用每个处理器私钥签名的消息以及安全区域的摘要。类似的，Lie 等人[42]提出在可信硬件 SGX 上实现 Glimmers，它可以在联邦学习中提供用户贡献的数据的可信性保证。同时在客户端上进行部署，通过验证用户贡献的数据是否可信来解决客户端隐私与服务端信任之间的冲突。

Feng 等人[43]提出了一种用于联邦深度学习的隐私保护方法，该方法可以在加密的情况下支持非线性激活函数和广泛使用的损失函数的操作，从而可以支持半诚实的客户端使用本地训练数据在加密的模型进行迭代训练，即保证服务器端模型的保密性，并且使用秘密分享技术来保障服务器端不能获得每个客户端的本地梯度从而保护客户端的隐私，其主要思想是服务器用一次性随机数对全局模型进行加密，然后将其发送给各个客户端，客户端根据本地训练数据和来自服务器端的加密模型得到本地梯度，使用一次性随机数扰动局部梯度，然后将经过扰动的局部梯度返回给服务器，保证所有客户端的一次性随机扰动值的和为 0，服务器就可以得到没有噪声的全局梯度用于更新全局模型。

Triastcyn 等人[44]提出了 FedGP，在客户端上训练生成式对抗网络（generative adversarial networks，GAN），以生成可替代客户真实数据的人工数据。使用差分平均案例隐私来估计并限制普通客户的预期隐私损失，从而增强传统联合学习的隐私。并且证明了在人工数据上训练的下游模型在保持良好的平均案例隐私性和对模型反演攻击的抵抗力的同时，具有较高的学习性能和准确性。

Liu 等人呢[45]将 sketching 算法应用于联邦学习中，当用户计算本地梯度后，利用 sketching 算法的独立哈希函数混淆原始数据，它主要通过用户拥有的秘密哈希索引和种子来隐藏每一轮模型更新的身份，并基于精度和空间的权衡，选择合适的压缩比，得到压缩后的梯度，然后再将该梯度发送给服务器进行聚合，聚合后的梯度返还给用户。用户查询已知的哈希索引获取更新的梯度并在本地更新。该方案不仅能够实现联邦学习任务的隐私性，还能同时保持或提高准确性和性能。

Choudhury 等人[46]提出了一种在联邦学习环境中保证隐私的语法方法（syntactic approach）。训练阶段中，采用基于(k, k^m)匿名的方法来对每个客户端上的本地数据进行匿名化，使用匿名的局部数据集训练模型，然后将参数更新合并到全局模型中，每个客户端只共享由匿名数据训练的模型参数。预测过程中，则利用全局匿名映射语法将样本映射到适当的等价类进行联邦学习预测。

Liu 等人[47]提出了一种名为 Forsaken 的联合学习框架，其为用户提供了消除记忆（k消除法）的服务，具体来说就是每个用户都部署有可训练的虚拟梯度生成器，经过训练步骤后，生成器可以产生虚拟梯度来刺激机器学习模型的神经元，从而消除特定数据的记忆，并用遗忘率（FR）来评估记忆消除的性能，且不需要重新训练机器学习模型，也不会破坏联合学习的通用过程。

Wainakh 等人[48]提出分层联邦学习（hierarchical federated learning，HFL），就是一台根服务器连接到多台组服务器，这些服务器以树结构进行组织。组服务器的最低层连接到用户，这些用户聚集在用户组中。层次结构可以包含多层组服务器，并且可以不平衡，以便不同分支的层数有所不同，HFL 的架构介于集中式 FL 和完全分散式学习之间，并具有一系列隐私优势。

Chamikara 等人[49]提出名为 DISTPAB（distributed privacy-preserving approach，DISTPAB）的分布式扰动算法来保护水平分区数据的隐私，DISTPAB 通过利用分布式环境中资源的不对称性来分配隐私保护任务从而缓解计算瓶颈，分布式环境可以具有资源受限的设备以及高性能计算机。其主要

思想就是使用多维变换和随机扩展，在数据离开局部边缘和雾层之前将扰动转移到分布式分支，而仅将全局扰动参数生成留给中心实体，DISTPAB 使用Φ −分离作为其底层隐私模型，该模型允许对给定实例进行最佳数据扰动。

除了使用传统的隐私

Li 等人[50]提出了在不交换数据的情况下联合训练一个梯度提升决策树（Gradient Boosting Decision Trees，GBDT）的框架 SimFL。整个联邦学习框架分为两个阶段：预处理阶段和训练阶段，预处理阶段的目标是收集相似信息，此处用到的是局部敏感哈希函数（Locality sensitive hash function，LSH），作者采用了多个 LSH 函数，每个组织首先计算自己样本对应的哈希值，经过广播后，所有的组织都可以构建一个哈希表，里面存储着样本序号和对应的哈希值。然后，每个组织都可以通过这个哈希表来计算相似信息。预处理阶段过后，每个组织对于自己的样本，都能在其他各个组织中找到一个相似样本。收集完相似信息后，进入训练阶段，每个组织轮流训练一些树，最终的模型为各个组织训练的树之和。

4 未来挑战

联邦学习对现有的隐私保护算法提出了新的挑战，除了提供严格的隐私保证外，还必须考虑使用隐私保护技术增加的计算量、通信量和模型损失。尽管当前的联邦学习隐私机制的研究已经取得了不错的效果，但是仍然存在一些挑战。

差分隐私技术可以通过添加噪声的方式保护用户级或者样本级的隐私，能提供严格的隐私保证，可以抵抗推理攻击，并且不需要花费太多的计算量和通信量。然而，虽然添加的噪声是在可预估的范围之内，但是噪声的添加仍然对模型的训练带来了影响，导致最终模型的性能下降，这相当于通过牺牲模型准确度来保护数据隐私，在一些要求高精度的模型的情况下，差分隐私可能并不合适。

同态加密技术通常是供隐私保护模型训练和预测使用的，一般用于加密梯度信息，避免泄露关于训练数据的额外信息，同时又能保证服务器可以对加密后的梯度进行聚合，可以抵御重构攻击，部分还可以抵御推理攻击。可是，虽然同态加密能够严格保护隐私，但由于是对密文进行操作，这就要求无论是客户端还是服务器端又要有一定的计算能力，同时又会占据较高的通信资源，在需要频繁交互或数据量大的环境下会暴露一些问题。如何选择合适的同态加密方案以适用于不同场景的数据处理是目前所面临的问题，同时研究者们也不断尝试与其他隐私保护手段相结合，比如安全多方计算，所以设计出支持复杂运算且高效的协议也是今后努力的方向。

安全多方计算技术可以安全的聚合来个每个客户端的局部更新，是针对联邦学习的设置而量身定制的，并且对于客户端在执行过程中退出具有鲁棒性，但是这种方法也有一定的局限性。使用安全多方计算技术往往会带来额外的通信成本，这在某些通信昂贵的场景下可能是不适用的；服务器通常被假定为半诚实的服务器，或者需要引入一个可信赖的第三方机构；允许服务器查看聚合结果可能仍然会泄露信息（比如说推理攻击）；缺乏强制客户端输入格式正确的能力。如何在解决上述问题的情况下构建一个强大的安全聚合协议仍然是个巨大的挑战。

使用混合的方案可以在一定程度上结合不同技术的优缺点，提供更加强力的隐私保证。安全多方计算中的秘密共享技术通常可以与其它技术一起使用，基于它的门限特性，可以保证用户中途退出，但通常需要较大的通信开销。另外，掩码协议也可以作为一种保护共享数据的手段与其他隐私保护方案联合使用。在混合方案中，也有许多方法是在共享数据中添加噪声，以获得一些差分

隐私的优点。但是在使用混合方案后，如何放大各个技术的优点，缩小各种技术的缺点，在计算量、通信量、模型性能和数据隐私性之间达到权衡，在未来需要进行进一步的研究。

5 结束语

联邦学习是一种分布式的架构，以打破数据孤岛为目的，在保证不泄露数据隐私的前提下，各个参与方合作构建模型。为了保护参与方的数据隐私，联邦学习使用了各种隐私保护技术来保障通信过程中的隐私安全。本文对联邦学习中使用的隐私保护机制进行了深入的研究和调查，根据使用的技术，对目前的联邦学习隐私保护机制进行了分类，并且总结了每种隐私保护机制的优点和不足。

目前，联邦学习正逐渐发展为一个综合性的研究项目，但是隐私保护永远是联邦学习的重点，当前的隐私保护技术能够在一定程度上保障联邦学习系统的隐私安全，但是面对恶意对手的攻击还缺乏防御手段，对联邦学习系统的鲁棒性的研究是很有意义的；联邦学习系统也缺乏识别恶意参与节点的方式，部分节点对于全局模型的贡献可能不是是有益的，这方面也是联邦学习的研究重点；另外，目前，联邦学习对于收益的分配是相同的，缺乏对收益公平性的研究，这方面可以考虑通过激励机制来提高参与方的动力，这将对联邦学习系统的落地有巨大帮助。最后，在保障参与方隐私的的情况下，提高模型的训练速度，提高模型的性能，也是一个值得研究的方向。

参考文献

[1] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.

[2] Zhu L, Han S. Deep leakage from gradients[M]//Federated learning. Springer, Cham, 2020: 17-31.

[3] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020.

[4] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of cryptography conference. Springer, Berlin, Heidelberg, 2006: 265-284.

[5] Rivest R L, Adleman L, Dertouzos M L. On data banks and privacy homomorphisms[J]. Foundations of secure computation, 1978, 4(11): 169-180.

[6] Gentry C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the forty-first annual ACM symposium on Theory of computing. 2009: 169-178.

[7] Yao A C. Protocols for secure computations[C]//23rd annual symposium on foundations of computer science (sfcs 1982). IEEE, 1982: 160-164.

[8] Rabin M O. How To Exchange Secrets with Oblivious Transfer[J]. IACR Cryptol. ePrint Arch., 2005, 2005(187).

[9] Shamir A. How to share a secret[J]. Communications of the ACM, 1979, 22(11): 612-613.

[10] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.

[11] McMahan H B, Ramage D, Talwar K, et al. Learning differentially private recurrent language models[J]. arXiv

preprint arXiv:1710.06963, 2017.
[12] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016: 308-318.
[13] Geyer R C, Klein T, Nabi M. Differentially private federated learning: A client level perspective[J]. arXiv preprint arXiv:1712.07557, 2017.
[14] Thakkar O, Andrew G, McMahan H B. Differentially private learning with adaptive clipping[J]. arXiv preprint arXiv:1905.03871, 2019.
[15] Agarwal N, Suresh A T, Yu F X X, et al. cpsgd: Communication-efficient and differentially-private distributed sgd[C]//Advances in Neural Information Processing Systems. 2018: 7564-7575.
[16] Bhowmick A, Duchi J, Freudiger J, et al. Protection against reconstruction and its applications in private federated learning[J]. arXiv preprint arXiv:1812.00984, 2018.
[17] Li J, Khodak M, Caldas S, et al. Differentially private meta-learning[J]. arXiv preprint arXiv:1909.05830, 2019.
[18] Triastcyn A, Faltings B. Federated learning with Bayesian differential privacy[C]//2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 2587-2596.
[19] Liu R, Cao Y, Yoshikawa M, et al. FedSel: Federated SGD under Local Differential Privacy with Top-k Dimension Selection[J]. arXiv preprint arXiv:2003.10637, 2020.
[20] Truex S, Liu L, Chow K H, et al. LDP-Fed: federated learning with local differential privacy[C]//Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking. 2020: 61-66.
[21] Liang Z, Wang B, Gu Q, et al. Exploring Private Federated Learning with Laplacian Smoothing[J]. arXiv preprint arXiv:2005.00218, 2020.
[22] Sabater C, Bellet A, Ramon J. Distributed differentially private averaging with improved utility and robustness to malicious parties[J]. arXiv preprint arXiv:2006.07218, 2020.
[23] Hardy S, Henecka W, Ivey-Law H, et al. Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption[J]. arXiv preprint arXiv:1711.10677, 2017.
[24] Paillier P. Public-key cryptosystems based on composite degree residuosity classes[C]//International conference on the theory and applications of cryptographic techniques. Springer, Berlin, Heidelberg, 1999: 223-238.
[25] Zhang C, Li S, Xia J, et al. Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning[C]//2020 {USENIX} Annual Technical Conference ({USENIX}{ATC} 20). 2020: 493-506.
[26] Liu Y, Liu Y, Liu Z, et al. Federated forest[J]. IEEE Transactions on Big Data, 2020.
[27] Cheng K, Fan T, Jin Y, et al. Secureboost: A lossless federated learning framework[J]. arXiv preprint arXiv:1901.08755, 2019.
[28] Mandal K, Gong G. PrivFL: Practical privacy-preserving federated regressions on high-dimensional data over

mobile networks[C]//Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop. 2019: 57-68.
[29] Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for federated learning on user-held data[J]. arXiv preprint arXiv:1611.04482, 2016.
[30] 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.
[31] Sanil A P, Karr A F, Lin X, et al. Privacy preserving regression modelling via distributed computation[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004: 677-682.
[32] Zhao K, Xi W, Wang Z, et al. SMSS: Secure Member Selection Strategy in Federated Learning[J]. IEEE Intelligent Systems, 2020, 35(4): 37-49.
[33] Xu R, Baracaldo N, Zhou Y, et al. Hybridalpha: An efficient approach for privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 13-23.
[34] He L, Karimireddy S P, Jaggi M. Secure Byzantine-Robust Machine Learning[J]. arXiv preprint arXiv:2006.04747, 2020.
[35] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11.
[36] Liu Y, Ma Z, Liu X, et al. Boosting privately: Privacy-preserving federated extreme boosting for mobile crowdsensing[J]. arXiv preprint arXiv:1907.10218, 2019.
[37] Nikolaenko V, Weinsberg U, Ioannidis S, et al. Privacy-preserving ridge regression on hundreds of millions of records[C]//2013 IEEE Symposium on Security and Privacy. IEEE, 2013: 334-348.
[38] Huang Y, Evans D, Katz J, et al. Faster secure two-party computation using garbled circuits[C]//USENIX Security Symposium. 2011, 201(1): 331-335.
[39] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.
[40] Hao M, Li H, Luo X, et al. Efficient and privacy-enhanced federated learning for industrial artificial intelligence[J]. IEEE Transactions on Industrial Informatics, 2019.
[41] Chen Y, Luo F, Li T, et al. A training-integrity privacy-preserving federated learning scheme with trusted execution environment[J]. Information Sciences, 2020, 522: 69-79.
[42] Lie D, Maniatis P. Glimmers: Resolving the privacy/trust quagmire[C]//Proceedings of the 16th Workshop on Hot Topics in Operating Systems. 2017: 94-99.
[43] Feng Y , Yang X , Fang W , et al. A Practical Privacy-preserving Method in

Federated Deep Learning[J]. 2020.
[44] Triastcyn A, Faltings B. Federated generative privacy[J]. IEEE Intelligent Systems, 2020.
[45] Liu Z, Li T, Smith V, et al. Enhancing the Privacy of Federated Learning with Sketching[J]. arXiv preprint arXiv:1911.01812, 2019.
[46] Choudhury O, Gkoulalas-Divanis A, Salonidis T, et al. Anonymizing Data for Privacy-Preserving Federated Learning[J]. arXiv preprint arXiv:2002.09096, 2020.
[47] Liu Y, Ma Z, Liu X, et al. Learn to Forget: User-Level Memorization Elimination in Federated Learning[J]. arXiv preprint arXiv:2003.10933, 2020.
[48] Wainakh A, Guinea A S, Grube T, et al. Enhancing Privacy via Hierarchical Federated Learning[J]. arXiv preprint arXiv:2004.11361, 2020.
[49] Chamikara M A P, Bertok P, Khalil I, et al. Privacy Preserving Distributed Machine Learning with Federated Learning[J]. arXiv preprint arXiv:2004.12108, 2020.
[50] Li Q, Wu Z, Wen Z, et al. Privacy-Preserving Gradient Boosting Decision Trees[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(01): 784-791.