计算机科学
COMPUTER SCIENCE

一种面向多模态医疗数据的联邦学习隐私保护方法

张连福, 谭作文

引用本文

张连福, 谭作文. 一种面向多模态医疗数据的联邦学习隐私保护方法[J]. 计算机科学, 2023, 50(11A): 230800021-8.

ZHANG Lianfu, TAN Zuowen. Federated Learning Privacy-preserving Approach for Multimodal Medical Data [J]. Computer Science, 2023, 50(11A): 230800021-8.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230800021

一种面向多模态医疗数据的联邦学习隐私保护方法

张连福[1]　谭作文[2]
1 宜春学院数学与计算机科学学院　江西 宜春 336000
2 江西财经大学信息管理学院计算机科学与技术系　南昌 330032
(zlf_jx@163.com)

摘　要　电子健康记录(Electronic Health Records,EHRs)数据已成为生物医学研究的宝贵资源。通过学习隐藏在EHRs数据中的人类难以区分的多维特征,机器学习方法可以获得更好的结果。然而,现有的一些研究只考虑了模型训练过程中或模型训练后可能面临的一些隐私泄露,导致隐私防护措施单一,无法实现覆盖机器学习全生命周期。此外,现有的方案大多是针对单模态数据的联邦学习隐私保护方法的研究。因此,提出了一种面向多模态数据的联邦学习隐私保护方法。为防止敌手通过反向攻击窃取原始数据信息,对每个参与者上传的模型参数进行差分隐私扰动。为防止在模型训练过程中各参与方的局部模型信息泄露,利用Paillier密码系统对局部模型参数进行同态加密。从理论的角度对该方法进行了安全性分析,给出了安全模型定义,并证明了子协议的安全性。实验结果表明,该方法在几乎不损失性能的情况下,保护了训练数据和模型的隐私。

关键词:联邦学习;多模态数据;电子健康记录;安全聚合;隐私保护
中图法分类号　TP391

Federated Learning Privacy-preserving Approach for Multimodal Medical Data

ZHANG Lianfu[1] and TAN Zuowen[2]
1 College of Mathematics and Computational Science,Yichun University,Yichun,Jiangxi 336000,China
2 Department of Computer Science and Technology,School of Information Technology,Jiangxi University of Finance and Economics,Nanchang 330032,China

Abstract　Electronic health records(EHRs) data has become a valuable resource for biomedical research. By learning multi-dimensional features hidden in EHRs data that are difficult for humans to distinguish,machine learning methods can achieve better results. However,some existing studies only consider some privacy leaks that may be faced during or after model training,resulting in a single privacy preservation measure that cannot cover the whole life cycle of machine learning. In addition,most of the existing programs are focused on federated learning privacy preservationmethods for single-mode data. Therefore,a federated learningprivacy preservation approach for multimodal data is proposed. To prevent the adversaryfrom stealing the original data information through reverse attack,differential privacy perturbation is performed on the model parameters uploaded by each participant. To prevent the leakage of local model information of each participant in the process of model training,the Paillier cryptosystem is used for homomorphic encryption of local model parameters. The security of the method is analyzed from the theoretical point of view,the security model is defined,and the security of the subprotocol is proved. Experimental results show that this method can preserveprivacy of training data and model with almost no loss of performance.

Keywords　Federated learning,Multimodal data,EHRs,Secure aggregation,Privacy-preserving

1　引言

机器学习方法需要从大量数据中学习才能获得更好的结果。然而,由于网络带宽、存储等限制,导致将训练数据集中到单个数据中心不可能实现。另一方面,金融、医疗等数据的高敏感性带来的隐私顾虑,导致各大机构和平台不愿与外界共享训练数据。此外,越来越严格的法律法规逐步出台,也限制了大数据的共享共用。这就阻碍了机器学习在各领域的应用。

联邦学习(FL)[1-2]作为一种机器学习范式,它使用来自多方的数据共同学习一个模型,同时保持数据的隐私性。因此,FL对各领域的未来具有重要意义[3]。然而,尽管FL试图让训练数据不离开客户端设备,以保持用户数据的隐私性,但它本身并没有提供有意义的隐私保证[4],这就导致FL可能面临一系列潜在的安全和隐私问题。例如,敌手可能会根据模型输出推断训练数据集的成员信息,或者从参与者上传的权重/梯度更新中反推出原始数据[5-6],特别是当权重矩阵中的某些权重对特定特征敏感时。现有的一些研究只考虑了

基金项目:国家自然科学基金(62362036);江西省自然科学基金重点项目(20232ACB202012)
This work was supported by the National Natural Science Foundation of China(62362036) and Key Project of Jiangxi Provincial Natural Science Foundation(20232ACB202012).
通信作者:谭作文(tanzyw@163.com)

模型训练过程中或模型训练后可能面临的一些隐私泄露[4,7-14]，导致隐私防护措施单一，无法覆盖机器学习全生命周期。

此外，医疗领域大数据除了具备一般大数据“4V”特性之外，还具有自己的独有特性——多模态性和多任务类型。多模态性是指医疗数据既包含化验产生的检测、检验数据，也包含体检产生的图像数据，如腹部 CT、胸部 CT、皮肤 X 射线、视网膜 X 射线和心电图等信号图谱，脑部磁共振、心电超声、颈脑血管超声等数据，以及像心跳声、咳嗽声、哭声等音频数据和胎动影像等视频数据。多任务类型是指根据不同的业务领域，医疗大数据还可能涉及多分类、多标签、有序回归等多种处理类型。现有的联邦学习隐私保护方案一般是针对某种单一模态数据来训练特定的联邦学习模型[4,7-8,11-14]，其任务类型也大多为单一的。因此，如何设计一种既考虑到训练阶段，又考虑到最终训练的模型可能面临的各种推理攻击的面向多模态数据的、多任务类型的通用 FL 模型，是本研究的关键性问题之一。

面对上述挑战，本文提出了一种隐私保护的联邦学习模型 BioMed-PPFL（Privacy-Preserving Federated Learning Framework for BioMedical Data）。利用 Paillier 同态加密来降低模型训练过程中模型参数泄露的风险，利用差分隐私扰动来最大限度地降低最终模型包含或记忆相关原始训练数据的隐私的风险。本文的工作将 FL 和密码学技术相结合，以促进数据科学合作，而无需明确地共享原始数据，并生成一种面向多模态数据的、多任务类型的通用 FL 模型，为 FL 在医疗保健领域的更广泛应用奠定了基础。表 1 列出了提出的方法与其他模型的比较。

表 1 提出的方法与其他方法比较

Table 1 Comparison between the proposed method and other methods

指标	[15]	[4]	[8]	[16]	[17]	BioMed-PPFL
分布式系统	●	●	●	●	○	●
训练阶段隐私保护	○	○	●	●	○	●
最终模型隐私保护	○	●	○	○	○	●
面向多模态数据	○	○	○	○	●	●

注：●表示支持；○表示不支持。

本文的主要贡献概括如下：

1）提出了一种隐私保护联邦学习模型，用于医学图像分类的安全聚合。为了方便地测试模型在不同类型的医学数据上的通用性，使用了具有多数据集设计的 MedMNIST[17] 数据集作为基准。据了解，这是首个在面向多模态、多任务类型的生物医学数据集上研究联邦学习隐私保护的工作。

2）对参与者上传的模型参数进行差分隐私扰动，以防止敌手通过模型反向攻击或成员推断攻击窃取原始数据信息。利用 Paillier 密码系统对局部模型参数进行同态加密，以防止半诚实但好奇的参数服务器和敌手在模型训练过程中获取各参与方的局部模型信息。

3）对 BioMed-PPFL 模型进行了安全性分析，给出了安全模型定义，证明了子协议的安全性，并对全局信息隐私性进行了分析。

4）大量实验结果表明，提出的方案的性能与经典的 FedAvg 和其他最先进的解决方案的性能接近，但本方案提供了更高的安全保障。

2 相关研究

在医疗保健领域，FL 在疾病诊断、疾病预测或患者病情分析方面发挥着越来越重要的作用。为了避免模型训练过程中的隐私泄露，一些医疗联邦学习系统采用同态加密（HE）对局部模型参数进行加密，或者采用安全多方计算（SMC）协议对局部模型参数进行加法随机化。Lee 等[7] 试图通过分析 EHR 数据，在不共享患者原始信息的情况下，有效地搜索不同医疗机构中的相似患者，作者将同态加密应用于联邦设置中的患者相似性搜索。Wibawa 等[8] 基于同态加密和联邦学习等隐私保护技术提出了一种 COVID-19 检测模型。作者使用公钥对模型权重矩阵进行加密，以保护深度学习模型免受攻击。Hosseini 等[11] 采用 SMC 技术基于组织病理学肺癌数据集构建了一个保护隐私的联邦学习框架。实验结果表明，与基于差分隐私的算法相比，该算法具有更高的精度。Dong 等[16] 基于秘密共享协议与 Top-K 梯度选择算法提出了一个高效安全的联邦学习方案，以减少联邦学习中用户之间梯度同步的通信开销，同时确保用户的数据隐私和安全。此外，作者还通过构造一个高效的消息验证码，来验证服务器返回的聚合结果的有效性。实验结果表明该方案会引入少量额外的开销，但达到了和明文训练同一水平的模型准确率。但是，HE 和 SMC 无法防范各种推理攻击[18]。攻击者对完成训练的模型执行多次查询，可以窃取关于模型参数及其训练数据的一些信息，例如特定样本是否包含在训练集中。

为了避免对最终模型的推理攻击，一些研究工作利用差分隐私来扰动局部模型参数。Choudhury 等[12] 基于来自 100 万患者的真实电子健康数据训练了一个全局模型，而没有显式地共享原始数据。该方案采用差分隐私机制，进一步保护模型免受潜在的隐私攻击。Aziz 等[13] 基于联邦学习和差分隐私技术提出了一种方法来共享基因表达数据。Islam 等[14] 基于 FL 和患者的基因组数据建立了一个医疗保健实用程序，来预测某些癌症疾病。为了保护数据和模型的隐私性，研究人员在传输前采用差分隐私（DP）方法扰动局部模型参数。Adnan 等[4] 进行了一个案例研究，该研究基于差分隐私和联邦学习技术来分析组织病理学图像。结果表明，与传统集中式训练相比，分布式训练可以在较强的隐私保障下取得相似的性能。上述研究通过在模型参数或梯度中添加噪声，可以确保攻击者很难从模型结果中反向推导出训练样本信息。但是，由于差分隐私的目标是保护计算结果而不是计算过程，因此，原始数据的隐私仍然有可能泄露[19]。

还有一些方案利用可信执行环境来进行隐私保护。可信执行环境（Trusted Execution Environment，TEE）[20-22] 是一种基于内存隔离的安全计算技术，可以在保证计算效率的同时完成保护隐私的计算。但缺点是其安全性很大程度上依赖于硬件实现，难以给出具体的安全边界定义，更容易受到来自不同攻击面的侧信道攻击。

上述现有的联邦学习隐私保护方案一般是针对某种单一模态数据来训练特定的 FL 模型，其任务类型也大多为单一的。与上述解决方案不同，本文的工作主要是设计一种既考虑了训练阶段，又考虑到最终训练的模型可能面临的各种推理攻击，同时面向多模态数据、多任务类型的通用 FL 隐私保护模型。

3 问题描述

本节简要概述了提出的隐私保护医疗联邦学习模型BioMed-PPFL的系统模型和威胁模型，然后基于这两种模型，确定了BioMed-PPFL的设计目标。

3.1 系统模型

所提模型BioMed-PPFL的框架如图1所示。系统分为两层：上层是参数服务器，下层是参与节点。参数服务器负责参与节点的选择、模型参数的聚合和模型参数的共享。参与节点包括医院、医学研究所等。这些参与节点有自己的本地训练数据，负责数据处理、模型初始化、本地模型训练、本地模型更新等。在模型训练过程中，所有参与节点使用认证的SSL通道将其本地模型更新与参数服务器通信，这确保了与服务器的安全通信。但是，SSL通道并不能防止参数服务器看到每个参与者提交的原始模型参数，因而考虑利用Paillier同态加密来防止模型参数的直接泄漏。尽管如此，最终的模型可能仍然包含或记住有关训练数据的隐私信息。如果存在半诚实但好奇或恶意的服务器，则可能存在模型反转或推理攻击，因而考虑利用差分隐私作为对HE的有力补充，从而加强对机器学习模型全生命周期的隐私保护。

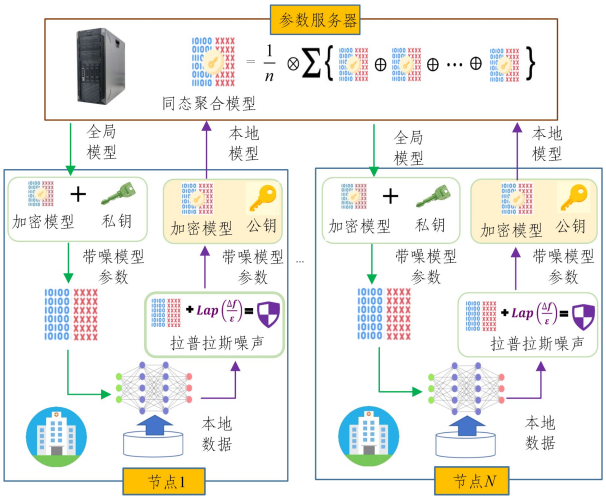

图1 BioMed-PPFL框架图

Fig. 1 Framework of BioMed-PPFL

3.2 威胁模型

系统潜在的威胁可能来自3个实体：参数服务器、参与节点和外部敌手。

1)参数服务器。参与模型训练的参数服务器一般被认为是诚实的，但同时也是好奇的(honest-but-curious)，即参数服务器会诚实地遵循预先约定的训练协议，但会试图窃取各参与者的局部模型参数，从而反向推断出训练数据的隐私信息。

2)参与节点。参与节点考虑以下两种类型：(1)参与节点为诚实实体，即参与节点诚实地遵循训练协议，但对其他参与节点的模型和训练数据的隐私感到好奇；(2)参与节点沦陷，它们的局部模型参数可作为攻击者的背景知识。

3)外部敌手。外部敌手考虑具有以下能力：(1)窃听参与节点与参数服务器之间通信信道中的密文；(2)沦陷参与节点。

4 BioMed-PPFL模型设计

本节阐述BioMed-PPFL模型的工作流程、设计思路和算法流程。BioMed-PPFL模型的设计主要包括构建联邦学习本地模型以及联邦学习全局模型聚合与更新两大任务。

4.1 本地模型训练协议

开始训练之前，每个参与节点首先初始化自己的本地模型。Paillier密码系统生成私钥key_{pri}和公钥key_{pub}，并分配给各参与节点。在第一轮全局模型聚合之后，各参与节点从参数服务器下载上一轮全局模型参数，并利用私钥key_{pri}解密全局模型参数，为下一轮的训练做准备，如式(1)所示：

$$w_{\text{global}}^{t-1}=Dec(\llbracket w_{\text{global}}^{t-1}\rrbracket,key_{\text{pri}}) \tag{1}$$

其中，$\llbracket w_{\text{global}}^{t-1}\rrbracket$为上一轮密文全局模型参数。

每个参与节点利用小批量梯度下降(Mini-Batch Gradient Descent，MBGD)方法基于其本地数据集对局部模型进行训练。设w^{t-1}为第$t-1$轮全局模型参数，$\chi_k^{t-1}\subseteq D_k$为参与者$k$的小批量训练数据，批大小为$B$，学习率为$\eta$，本地训练周期为$E$，则在第$t$轮迭代时，参与者$k$的模型参数$w_k^t$的更新规则如式(2)、式(3)所示：

$$w_k^t:=w^{t-1}-\sum_{i=1}^{E}\eta g_k(w_k^{(t-1,i)};\chi_k^{(t-1,i)}) \tag{2}$$

$$g_k(w_k^{(t-1,i)};\chi_k^{(t-1,i)}):=(1/B)\sum_{j=1}^{B}\nabla f_k(w_k^{(t-1,i)};x_k^{(j,i)},y_k^{(j,i)}) \tag{3}$$

其中，$k\in\{1,2,\cdots,K\}$，K为参与节点总数。

为避免攻击者在模型训练后通过模型参数$\{w_k^t\}$推断出训练数据信息，各参与节点在局部模型训练结束后在其模型参数中添加DP机制噪声，如式(4)所示：

$$w_{(k,dp)}^t:=\begin{cases}w_k^t+Lap\left(\dfrac{\Delta f}{\varepsilon}\right), & \text{Laplace noise}\\ w_k^t+\mathcal{N}(0,\sigma^2C^2), & \text{Gaussian noise}\end{cases} \tag{4}$$

其中，$w_{(k,dp)}^t$表示参与节点k在第t轮利用差分隐私加噪后的模型参数。

以高斯机制为例，当添加的高斯噪声的标准差满足$\sigma\geqslant c\Delta f/\varepsilon$，且$c^2>2\ln(1.25/\delta)$，则可保证算法中每一轮训练都满足$(\varepsilon,\delta)$-差分隐私。其中，噪声方差取值由隐私预算$(\varepsilon,\delta)$和查询敏感度$\Delta f$共同决定。在本实验中，结合具体的$\Delta f$和$\sigma$，我们将差分隐私预算$\varepsilon$设为10，$\delta$设为$1\times10^{5}$。

为防止半诚实但好奇的服务器和敌手在模型训练过程中获取本地模型信息，各参与节点在将本地模型参数上传到参数服务器之前使用Paillier密码系统对其进行加密，如式(5)所示：

$$\llbracket w_k^t\rrbracket=Enc(w_{(k,dp)}^t,key_{\text{pub}}) \tag{5}$$

其中，key_{pub}为参与节点用于本地模型参数加密的公钥。

本地模型训练协议(Local Model Training Protocol，LMTP)，如算法1所示。

算法1 参与者k本地模型训练协议(LMTP)

输入：第t−1轮全局模型$\llbracket w_{\text{global}}^{t-1}\rrbracket$，学习率$\eta$，本地迭代次数E，训练数据$D_k$

输出：参与者k第t轮模型参数$\llbracket w_k^t\rrbracket$

1. 下载全局模型参数$\llbracket w_{\text{global}}^{t-1}\rrbracket$

2. Paillier密码系统生成私钥key_{pri}和公钥key_{pub}，并分配给各参与节点。

// 参与节点k使用私钥key_{pri}解密上述密文

3. $w_{\text{global}}^{t-1}=Dec(\llbracket w_{\text{global}}^{t-1}\rrbracket,key_{\text{pri}})$

4. for i=1,2,…,E do

5. batches←将训练数据D_k分割成多个小批量数据$\chi_k^{(t-1,i)}$
6. for b in batches do
7. $w_k^t := w_{global}^{t-1} - \eta g(w_k^{(t-1,i)}; \chi_k^{(t-1,i)})$
8. end for
//利用差分隐私噪声扰动本地模型参数
9. $w_{(k,dp)}^t := w_k^t + Lap(\Delta f/\varepsilon)$或$w_{(k,dp)}^t := w_k^t + \mathcal{N}(0, \sigma^2 C^2)$
//参与节点 k 利用 Paillier 密码系统加密本地模型参数
10. $[\![w_k^t]\!] = Enc(w_{(k,dp)}^t, key_{pub})$
11. 参与者 k 将加密模型参数$[\![w_k^t]\!]$上传给参数服务器

4.2 全局模型聚合与更新协议

参数服务器接收到所有参与者第 t 轮加密的模型参数$([\![w_1^t]\!], [\![w_2^t]\!], \cdots, [\![w_K^t]\!])$后，如式(6)所示进行参数聚合操作。

$$[\![w_{global}^t]\!] = \frac{1}{n} \otimes ([\![w_1^t]\!] \oplus [\![w_2^t]\!] \oplus \cdots \oplus [\![w_K^t]\!]) = \frac{1}{n} \sum_{k=1}^{K} [\![w_k^t]\!] \tag{6}$$

其中，$k \in \{1, 2, \cdots, K\}$，$\otimes$表示同态乘法，$\oplus$表示同态加法，$n=K$。

这里，同态聚合操作可以确保参数服务器对全局模型进行更新时，每个本地模型参数都是私有的。这种增加的安全性是以增加服务器的计算开销为代价的，但它在医疗保健应用中发挥着至关重要的作用，可以确保每个医疗机构原始训练数据的机密性，同时仍然受益于与其他机构的协作学习。在聚合这些加密的模型参数之后，参数服务器将更新后的全局模型发送给所有的参与节点。

全局模型聚合和更新协议(Global Model Aggregation and Update Protocol，GMAP)伪代码如算法 2 所示。

算法 2 全局模型聚合与更新协议(GMAP)

输入：参与者 k 第 t 轮本地模型参数$[\![w_k^t]\!]$，全局模型训练次数 R，节点总数 M，选中的参与节点数 K

输出：第 t 轮全局模型参数$[\![w_{global}^t]\!]$

1. 初始化全局模型$[\![w_{global}^t]\!]$
2. for t=1,2,…,R do
3. 从 M 中随机选取 K 个参与节点
4. for k∈K in parallel do
5. 接收局部模型参数$[\![w_k^t]\!]$
//聚合各局部模型
6. $[\![w_{global}^t]\!] = [\![w_{global}^t]\!] \oplus [\![w_k^t]\!]$
7. end for
//全局模型同态乘法运算
8. $[\![w_{global}^t]\!] = \frac{1}{K} \otimes [\![w_{global}^t]\!]$
9. 将全局模型参数$[\![w_{global}^t]\!]$发送给各参与节点
10. end for

5 安全性分析

下面从信息安全理论的角度对提出的方法 BioMed-PPFL 进行安全性和隐私性分析。首先提出安全模型定义，然后基于安全模型定义证明提出的协议是安全的，最后对全局信息隐私性进行分析。

5.1 安全模型定义

假设 $\mathcal{P} = (P_a, P_b, P_c, S)$为协议参与方集合，其中$P_a$，$P_b$，$P_c$代表客户端集合，$S$ 代表服务器。考虑两种敌手($\mathcal{A}_P$，$\mathcal{A}_S$)，分别破坏客户端集合P_a，P_b，P_c和服务器 S。在实际执行过程中，分别输入 x，y 和 z 来运行P_a，P_b，P_c，而令 λ 为 S 的辅助输入。考虑每个 $P \in \mathcal{P}$，针对敌手($\mathcal{A}_P$，$\mathcal{A}_S$)，在协议 Π 实际执行过程中，P 的部分视图可定义为随机变量$REAL_{\Pi,\mathcal{A}}^P(\lambda, x, y, z)$。考虑每个 $P \in \mathcal{P}$，针对相互独立的模拟器$Sim = (Sim_P, Sim_S)$，在协议 Π 的理想执行过程中，P 的部分视图可定义为随机变量$IDEAL_{f,Sim}^P(\lambda, x, y, z)$，其中，$f$ 为理想世界中的理想的功能性函数。

非正式地，如果现实执行中的协议模拟理想世界中执行的 f，那么针对 PPT 敌手，协议 Π 是安全的。下面分别针对敌手$\mathcal{A}_P$和$\mathcal{A}_S$，给出安全模型的形式化定义。

定义 1 针对 PPT 敌手$\mathcal{A}_P$、独立的模拟器Sim_P、所有的输入 x, y, z 以及所有的参与方 $P \in \mathcal{P}$而言，如果 BioMed-PPFL 模型是安全的，则可以表示为：

$$IDEAL_{f,Sim,H}^P(x, y, z) \approx_c REAL_{\Pi,\mathcal{A},H}^P(x, y, z)$$

其中，f 为 $\mathcal{P}$中各参与方之间相互协作的功能性函数，Π 为实际执行中的协议，$H \subset \mathcal{P}$为 $\mathcal{P}$中诚实参与方集合，$\approx_c$表示计算上不可区分。

定义 2 针对 PPT 敌手$\mathcal{A}_S$、独立的模拟器Sim_S、辅助输入 λ 以及所有的参与方 $P \in \mathcal{P}$而言，如果 BioMed-PPFL 模型是安全的，则可以表示为：

$$IDEAL_{f,Sim,H}^P(\lambda) \approx_c REAL_{\Pi,\mathcal{A},H}^P(\lambda)$$

其中，f 为 $\mathcal{P}$中各参与方之间相互协作的功能性函数，Π 为实际执行中的协议，$H \subset \mathcal{P}$为 $\mathcal{P}$中诚实参与方集合，$\approx_c$表示计算上不可区分。

5.2 子协议的安全性

引理 1[23] 遵循 Paillier 同态密码系统的所有基础操作都被认为是安全的。

引理 2[24] 如果一个协议所有的子协议都是完美的模拟，即不可区分的，则该协议为不可区分的。

定理 1 在半诚实模型下，即使存在威胁客户端集合的 PPT 敌手$\mathcal{A}_P$，本地模型训练协议(LMTP)仍然可以安全地实现理想的功能，而不会泄露隐私。

证明：考虑存在威胁客户端集合的 PPT 敌手$\mathcal{A}_P$，构造模拟器 Sim_P，在理想世界中执行，其中Sim_P构造如下。对于Sim_P，协议理想执行中的视图为$View_P = [\![w_k^t]\!]$。LMTP 协议包括下载全局模型参数、局部模型训练和上传本地模型参数至服务器 3 个主要操作。其中本地模型训练在参与节点本地上进行计算，其操作是满足安全的。因此，只要保证下载全局模型参数和上传本地模型参数至服务器两个操作满足安全性即可。参与节点以密文形式从参数服务器下载全局模型参数$[\![w_k^t]\!]$的，敌手$\mathcal{A}_P$无法获得任何信息。由于每轮都执行重复的操作，因此可以根据一轮参数服务器和参与节点之间操作的安全性，将整个迭代过程视为安全的。根据以上分析，结合引理 1、引理 2 可知，模拟器 Sim_P将生成一种在计算上与实际执行无法区分的视图。因此，LMTP 协议在理想执行和实际执行中无法区分。

定理 2 在半诚实模型下，即使存在威胁参数服务器的 PPT 敌手$\mathcal{A}_S$，全局模型聚合和更新协议(GMAP)仍然可以安全地实现理想的功能，而不会泄露隐私。

证明：假定存在威胁参数服务器的 PPT 敌手$\mathcal{A}_S$，构造模拟器Sim_S，在理想世界中执行，其中Sim_S构造如下。对于

Sim_S，协议理想执行中的视图为$View_S=[\![w_k^t]\!]$。参数服务器接收到各参与节点发送过来的加密局部模型参数，然后按照$[\![w_{\text{global}}^t]\!]=\frac{1}{n}\otimes([\![w_1^t]\!]\oplus[\![w_2^t]\!]\oplus\cdots\oplus[\![w_K^t]\!])$进行参数聚合，根据引理1，这一操作是安全的。综上分析，模拟器 Sim_S将生成一种在计算上与实际执行中无法区分的视图。因此，GMAP 协议在实际执行和理想执行中是不可区分的。

5.3 全局信息隐私性分析

定理 3 假定训练总轮数为 T，选中参与节点总数为 K，每个数据子集规模相同，即$|D_i|=N$，C 为权重参数的界，那么，当添加的高斯噪音标准差满足 $\sigma\geqslant 2Cc/KN\varepsilon$，且 $c^2\geqslant 2\ln(1.25/\delta)$，则方案 BioMed-PPFL 满足$(T\varepsilon,T\delta)$-差分隐私。

证明：令算法 1 第 9 步中添加的高斯噪声的标准差满足 $\sigma\geqslant 2Cc/KN\varepsilon$，且$c^2>2\ln(1.25/\delta)$，根据差分隐私的并行组合特性，所有参与节点第一次聚合后的全局模型参数满足(ε,δ)-差分隐私。在整个训练过程中，参与节点对D_i进行了 T 次查询，由差分隐私的串行组合性可知，方案 BioMed-PPFL 至少满足$(T\varepsilon,T\delta)$-差分隐私。

6 实验结果及分析

本节将评估提出的 BioMed-PPFL 模型的性能。首先描述了数据集和实验设置，然后在 MedMNIST2D 数据集上运行 BioMed-PPFL，接着将所提出的方案与经典的 FedAvg[1]，DP-FedAvg[25]以及其他最先进的解决方案在 Auroc、准确率(Accuracy)、精度(Precision)、召回率(Recall)和 F1 评分(F1-Score)等方面进行了比较，最后研究了隐私预算对模型精度的影响。

6.1 模型与数据集

考虑到 Paillier 密码系统计算开销较大，对于 ResNet18，ResNet50 等深度神经网络模型训练收敛速度较慢，因此，本文所有实验都基于 4 层卷积神经网络模型(4Layer Convolutional Neural Network model，4Layer-CNN)。神经网络结构如图 2 所示。

```
model=Sequential()
model.add(Conv2d(in_channels=3,out_channels=16,kernel_size=4,stride=2,padding=2,activation='relu'))
model.add(MaxPool2d(2))
model.add(Conv2d(out_channels=32,kernel_size=3,stride=2,padding=1,activation='relu'))
model.add(MaxPool2d(2))
model.add(Conv2d(out_channels=64,kernel_size=3,stride=1,padding=1,activation='relu'))
model.add(MaxPool2d(2))
model.add(Conv2d(out_channels=128,kernel_size=3,stride=1,padding=1,activation='relu'))
model.add(MaxPool2d(2))
model.add(Flatten())
model.add(Dense(num_classes))
```

图 2 BioMed-PPFL 神经网络结构

Fig. 2 Neural network architecture of BioMed-PPFL

本文所有实验都基于 MedMNIST v2 数据集[17]。该数据集是二维和三维生物医学图像分类的大规模基线实验数据，包括 12 个 2D 数据集(28 * 28，224 * 224)和 6 个 3D 数据集(28 * 28 * 28)。数据集涵盖了生物医学图像中的主要数据模式(CT、X 射线、超声等)，支持多种任务类型(多分类、多标签、有序回归)，并提供各种数据大小(从数百到十万)。对于 2D 数据集，支持 ResNet18 和 ResNet50 分别在 28 * 28 和 224 * 224 分辨率上进行测试；对于 3D 数据集，支持 2.5D，3D，ACS 卷积的 ResNet18 和 ResNet50 进行测试。本文选择了 OrganAMNIST，OrganCMNIST 和 BloodMNIST 这 3 种 2D 数据集进行测试。数据集描述如表 2 所列。为了比较的公平性，所有实验均取测试集上 3 次结果的平均值作为平均性能指标。

表 2 数据集描述

Table 2 Description of datasets

数据集	数据种类	训练集	验证集	测试集
OrganAMNIST2D	多分类(11)	34581	6491	17778
OrganCMNIST2D	多分类(11)	13000	2392	8268
BloodMNIST2D	多分类(8)	11959	1712	3421

6.2 实验设置

1)实验环境

操作系统为 ubuntu20.04，CPU 为 IntelCore i9-10850K，显卡为 CUDA Toolkit 10.1，开发环境为 Anaconda3，Python3.7.0 和 PyCharm。联邦学习框架为基于 Pytorch 的客户端/服务器架构。

2)比较基线

为了展示所提方案的性能，将其与经典的 FedAvg[1]，DP-FedAvg[25]方法以及最先进的解决方案 FedHE[9]进行了比较。在 DP-FedAvg 方法中，每个参与节点的模型参数都受到差分隐私噪声的干扰。在 FedHE 方法中，各参与节点的模型参数采用 Paillier 密码系统加密。

3)参数设置

本文实验使用自定义 4 层卷积神经网络模型(4Layer-CNN)分别在 28 * 28 和 224 * 224 分辨率上进行测试。这两种分辨率上的模型分别简称为 4Layer-CNN(28)和 4Layer-CNN(224)。其通用参数设置为：密钥长度 192，批大小 128，输入通道均为 3，本地训练迭代次数为 5，损失函数使用交叉熵，优化器使用 Adam，初始学习率设置为 0.01，隐私预算 ε 为 10，加噪方式为在本地模型参数中添加拉普拉斯噪音，在训练进行到全局 Epochs 的 50% 和 75% 时，学习率减半。其特别参数设置如表 3 所列。

表 3 特别参数设置

Table 3 Special parameter settings

数据集	分辨率	节点总数	选定节点数	全局 Epochs
OrganAMNIST2D	28	30	10	30
	224	30	10	10
OrganCMNIST2D	28	15	10	30
	224	15	10	10
BloodMNIST2D	28	10	10	30
	224	10	10	10

6.3 性能比较和结果分析

下面分别对 OrganAMNIST2D，OrganCMNIST2D 和 BloodMNIST2D 数据集上的实验结果进行讨论分析。

1)OrganAMNIST2D 数据集上实验结果及分析

表 4 和图 3－图 7 展示了 OrganAMNIST2D 数据集上的

实验结果。实验结果表明，FedHE 方法与 FedAvg 方法训练的模型性能相当，并且表现非常稳定，表明了 Paillier 同态密码技术对模型的精度、准确率等性能指标无不良影响。提出的方案与 DP-FedAvg 方法表现相当，与 FedHE 方法与 FedAvg 方法在性能上的差距主要是由于受到差分隐私噪音的影响。只要隐私预算取到一个合理值，就可以达到可用性与隐私保护之间的较好平衡。

表 4 OrganAMNIST2D 数据集上的性能对比

Table 4 Comparison of performance on OrganAMNIST2D dataset

分辨率	评价指标	BioMed-PPFL	DP-FedAvg	FedHE	Fed Avg
28	Auroc	0.979	0.980	0.982	0.983
	Accuracy	0.886	0.885	0.888	0.889
	Precision	0.889	0.887	0.892	0.890
	Recall	0.880	0.878	0.882	0.882
	F1-score	0.884	0.882	0.887	0.886
224	Auroc	0.981	0.981	0.982	0.983
	Accuracy	0.892	0.891	0.894	0.894
	Precision	0.896	0.895	0.898	0.900
	Recall	0.886	0.885	0.889	0.890
	F1-score	0.891	0.890	0.893	0.895

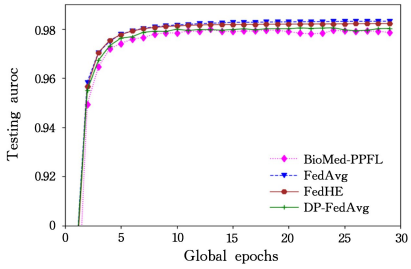

图 3 OrganAMNIST2D 数据集 4Layer-CNN(28)模型上的 Auroc

Fig. 3 Auroc of 4 Layer-CNN(28) on OrganAMNIST2D dataset

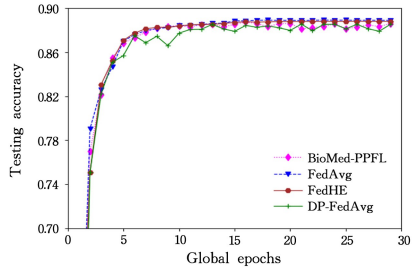

图 4 OrganAMNIST2D 数据集 4Layer-CNN(28)模型上的 Accuracy

Fig. 4 Accuracy of 4Layer-CNN(28) on OrganAMNIST2D dataset

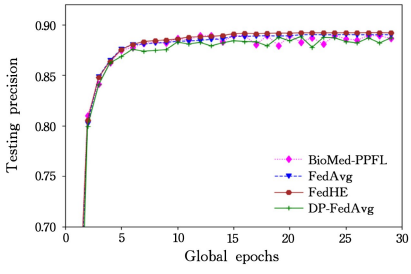

图 5 OrganAMNIST2D 数据集 4Layer-CNN(28)模型上的 Precision

Fig. 5 Precision of 4Layer-CNN(28) on OrganAMNIST2D dataset

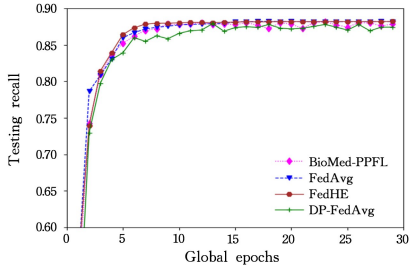

图 6 OrganAMNIST2D 数据集 4Layer-CNN(28)模型上的 Recall

Fig. 6 Recall of 4Layer-CNN(28) on OrganAMNIST2D dataset

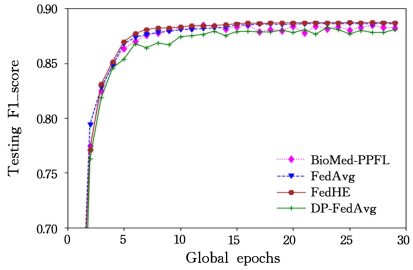

图 7 OrganAMNIST2D 数据集 4Layer-CNN(28)模型上的 F1-score

Fig. 7 F1-score of 4Layer-CNN(28) on OrganAMNIST2D dataset

2)OrganCMNIST2D 数据集上的实验结果及分析

表 5 列出了 OrganCMNIST2D 数据集上的实验结果。实验结果表明，FedHE 方法与 FedAvg 方法训练的模型性能相当，并且表现稳定，表明了 Paillier 密码技术对模型的精度等性能指标无不良影响。提出的方案与 DP-FedAvg 方法表现相当，与 FedHE 方法与 FedAvg 方法在性能上的差距主要受差分隐私噪音的影响。只要隐私预算取到一个合理值，就可以实现可用性与隐私保护之间的较好平衡。

表 5 OrganCMNIST2D 数据集上的性能对比

Table 5 Comparison of performance on OrganCMNIST2D dataset

分辨率	评价指标	BioMed-PPFL	DP-FedAvg	FedHE	Fed Avg
28	Auroc	0.976	0.978	0.980	0.981
	Accuracy	0.877	0.879	0.880	0.881
	Precision	0.867	0.869	0.870	0.872
	Recall	0.863	0.865	0.868	0.866
	F1-score	0.865	0.867	0.869	0.869
224	Auroc	0.976	0.976	0.977	0.978
	Accuracy	0.873	0.874	0.875	0.876
	Precision	0.870	0.872	0.873	0.874
	Recall	0.852	0.853	0.855	0.855
	F1-score	0.861	0.862	0.864	0.864

3)BloodMNIST2D 数据集上的实验结果及分析

表 6 列出了 BloodMNIST2D 数据集上的实验结果。实验结果表明，FedHE 方法与 FedAvg 方法训练的模型性能相当，并且表现非常稳定，表明了 Paillier 同态密码技术对模型的精度、准确率等性能指标无不良影响。提出的方案与 DP-FedAvg 方法表现相当，与 FedHE 方法与 FedAvg 方法在性能上的差距主要受差分隐私噪音的影响。只要隐私预算取到一个合理值，就可以达到可用性与隐私保护之间的较好平衡。

表 6　BloodMNIST2D 数据集上的性能对比

Table 6　Comparison of performance on BloodMNIST2D dataset

分辨率	评价指标	BioMed-PPFL	DP-FedAvg	FedHE	Fed Avg
28	Auroc	0.985	0.984	0.987	0.986
	Accuracy	0.918	0.917	0.920	0.919
	Precision	0.912	0.910	0.916	0.914
	Recall	0.912	0.910	0.914	0.913
	F1-score	0.912	0.910	0.915	0.913
224	Auroc	0.987	0.986	0.988	0.989
	Accuracy	0.924	0.924	0.925	0.926
	Precision	0.922	0.921	0.923	0.925
	Recall	0.902	0.900	0.903	0.904
	F1-score	0.912	0.910	0.913	0.914

4）参与节点的数量对模型精度的影响

图 8 展示了 OrganAMNIST2D 数据集上参与节点的数量对模型精度的影响。在 OrganAMNIST2D 数据集上，创建了 30 个节点。为了简单起见，设定每个节点上的数据分布均匀。从图 8 可以看出，选中参与训练的客户端节点越多，模型精度越高，这是由于训练数据增多导致的，这与理论上是一致的。在 OrganCMNIST2D 数据集和 BloodMNIST2D 数据集上也有相同的结论。

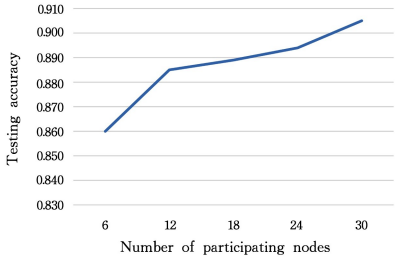

图 8　OrganAMnist2D 数据集上参与节点的数量对模型精度的影响

Fig. 8　Impact of the number of participating nodes on model accuracy on OrganAMnist2D dataset

5）参与节点的数量对训练时间的影响

图 9 展示了 OrganAMnist2D 数据集上参与节点的数量对训练时间的影响。设定每个节点上的数据分布均匀。从图 9 可以看出，选中参与训练的客户端节点越多，模型训练耗时越长，这是由于训练数据增多导致的。实验结果表明，Paillier 同态加解密还是比较耗时的，这也是同态加密技术的不足之处。在 OrganCMNIST2D 数据集和 BloodMNIST2D 数据集上也有类似的结论。

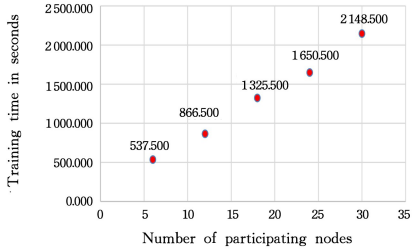

图 9　OrganAMnist2D 数据集上参与节点的数量对训练时间的影响

Fig. 9　Effect of the number of participating nodes on training time on OrganAMnist2D dataset

6）不同方法的效率对比

图 10 展示了 OrganAMnist2D 数据集上不同方法的效率对比。实验展示了 ResNet28 模型上的结果，假定参与节点总数为 30，选定参与节点为 10，全局迭代次数为 30。

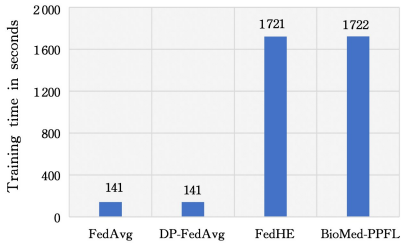

图 10　OrganAMnist2D 数据集上不同方法的效率对比

Fig. 10　Efficiency comparison of different methods on OrganAMnist2D dataset

从图 10 可以看出，FedAvg 方法与 DP-FedAvg 方法的效率相同，BioMed-PPFL 方法与 FedHE 方法的效率接近。实验结果表明，差分隐私对实验效率几乎没有影响，同态加密是影响实验效率的主要因素，但也在可接受的范围之内。

7）不同隐私预算对模型精度的影响

图 11 展示了不同隐私预算对模型精度的影响。在本文实验中，各参与节点在本地模型训练结束后向服务器提交参数前，为本地模型参数注入拉普拉斯机制噪音。从图 11 可以看出，隐私预算 ε 越小，表示添加的噪音越大，模型精度越低；反之，隐私预算 ε 越大，则表示添加的噪音越小，模型精度越高。

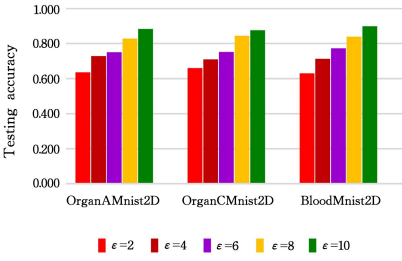

图 11　不同隐私预算对模型精度的影响

Fig. 11　Effects of different privacy budgets on model accuracy

结束语　FL 有助于利用现有的 EHR 数据来训练更一般化的模型，并直接影响患者的诊断、治疗和护理，为医疗保健行业的开放创新提供了巨大的机会。本文基于差分隐私和同态加密设计了一种联邦学习隐私保护方法 BioMed-PPFL。该方法可以在模型训练期间和模型训练后保护原始训练数据的隐私，从而消除了数据共享的许多障碍。为了方便地测试模型在不同类型的医学数据上的通用性，本文实验采用了基于多数据集设计的 MedMNIST 数据集。本文对 BioMed-PPFL 模型进行了安全性分析，给出了安全模型定义，并证明了子协议的安全性。本文使用 Auroc、准确度、精密度、召回率和 F1-评分作为性能指标，并将其与经典的 FedAvg 以及其他最先进的解决方案进行比较。大量实验结果表明，BioMed-PPFL 模型的性能与经典的 FedAvg 和其他最先进的解决方案的性能接近，但提供了更高的安全保障。但该方法中服务器与参与节点之间可能存在合谋，从而导致诚实节点的模型参数和训练数据存在隐私泄露的风险，并且，现有的联邦学习隐私保护方案大多是针对 2D 数据，很少有针对 3D 数据的联邦学习隐私保护方法研究。因此，研究提出抗合谋攻击的、支持 3D 数据的联邦学习隐私保护方法是下一步努力的方向。

参 考 文 献

[1]　MCMAHAN B, MOORE E, RAMAGE D, et al. Communica-

tion-efficient learning of deep networks from decentralized data [C]// Proceedings of the Artificial Intelligence and Statistics. 2017:1273-1282.

[2] TAN Z W, ZHANG L F. Survey on privacy preserving techniques for machine learning[J]. J Software, 2020, 31(7): 2127-2156.

[3] LONG G, SHEN T, TAN Y, et al. Federated learning for privacy-preserving open innovation future on digital health [M]. Springer, 2022: 113-133.

[4] ADNAN M, KALRA S, CRESSWELL J C, et al. Federated learning and differential privacy for medical image analysis[J]. Scientific Reports, 2022, 12(1): 1-10.

[5] SONG C, RISTENPART T, SHMATIKOV V. Machine learning models that remember too much[C]// Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017: 587-601.

[6] ATENIESE G, MANCINI LV, SPOGNARDI A, et al. Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers[J]. International Journal of Security and Networks, 2015, 10(3): 137-150.

[7] LEE J, SUN J, WANG F, et al. Privacy-preserving patient similarity learning in a federated environment: development and analysis[J]. JMIR Medical Informatics, 2018, 6(2): e7744.

[8] WIBAWA F, CATAK FO, KUZLU M, et al. Homomorphic encryption and federated learning based privacy-preserving cnn training: COVID-19 detection use-case[C]// Proceedings of the 2022 European Interdisciplinary Cybersecurity Conference. 2022: 85-90.

[9] ZHANG L F, XU J, VIJAYAKUMAR P, et al. Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system [J]. IEEE Transactions on Network Science and Engineering, 2022, 2022(1): 1-17.

[10] KALAPAAKING A P, STEPHANIE V, KHALIL I, et al. SMPC-Based federated learning for 6G-enabled internet of medical things[J]. IEEE Network, 2022, 36(4): 182-189.

[11] HOSSEINI S M, SIKAROUDI M, BABAEI M, et al. Cluster based secure multi-party computation in federated learning for histopathology images [C] // Proceedings of the International Workshop on Distributed, Collaborative, and Federated Learning, Workshop on Affordable Healthcare and AI for Resource Diverse Global Health. Springer, 2022: 110-118.

[12] CHOUDHURY O, GKOULALAS-DIVANIS A, SALONIDIS T, et al. Differential privacy-enabled federated learning for sensitive health data[J]. arXiv: 1910. 02578, 2019.

[13] AL AZIZ M M, ANJUM M M, MOHAMMED N, et al. Generalized genomic data sharing for differentially private federated learning [J]. Journal of Biomedical Informatics, 2022, 132: 104113.

[14] ISLAM T U, GHASEMI R, MOHAMMED N. Privacy-preserving federated learning model for healthcare data[C]// Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2022: 281-287.

[15] ZHANG W, ZHOU T, LU Q, et al. Dynamic-fusion-based federated learning for COVID-19 detection [J]. IEEE Internet of Things Journal, 2021, 8(21): 15884-15891.

[16] DONG Y, HOU W, CHEN X J, et al. Efficient and Secure Federated Learning Based on Secret Sharing and Gradients Selection [J]. Journal of Computer Research and Development, 2020, 57(10): 2241-2250.

[17] YANG J, SHI R, WEI D, et al. MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification[J]. Scientific Data, 2023, 10(1): 41.

[18] SABRY F, ELTARAS T, LABDA W, et al. Machine learning for healthcare wearable devices: the big picture [J/OL]. https://doi. org/10. 1155/2022/4653923.

[19] AONO Y, HAYASHI T, WANG L, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

[20] LU D, SHI M, MA X, et al. Smaug: A TEE-Assisted secured SQLite for embedded systems[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(5): 3617-3635.

[21] STEPHANIE V, KHALIL I, ATIQUZZAMAN M, et al. Trustworthy privacy-preserving hierarchical ensemble and federated learning in healthcare 4. 0 with blockchain[J]. IEEE Transactions on Industrial Informatics, 2023, 19(7): 7936-7945.

[22] MENG X, YANG Y, LIU X, et al. Active forgetting via influence estimation for neural networks[J]. International Journal of Intelligent Systems, 2022, 37(11): 9080-9107.

[23] MOHAMMED S J, TAHA D B. Performance evaluation of RSA, ElGamal, and Paillier partial homomorphic encryption algorithms[C]// Proceedings of the 2022 International Conference on Computer Science and Software Engineering (CSASE) IEEE, 2022: 89-94.

[24] BOGDANOV D, LAUR S, WILLEMSON J. Sharemind: A framework for fast privacy-preserving computations[C]// Proceedings of the Computer Security-ESORICS 2008: 13th European Symposium on Research in Computer Security. Málaga, Spain, Springer, 2008: 192-206.

[25] TRIASTCYN A, FALTINGS B. Federated learning with bayesian differential privacy[C]// Proceedings of the 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 2587-2596.

ZHANG Lianfu, born in 1978, Ph.D, lecturer, is a member of China Computer Federation. His main research interests include information security and privacy-preserving machine learning.

TAN Zuowen, born in 1967, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include cryptography, blockchain and privacy-preserving machine learning.