2024 年 第 24 卷 第 19 期
2024, 24(19): 08166-10

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2303534
引用格式:龚颖, 许文韬, 赵策, 等. 基于零信任机制的联邦学习模型[J]. 科学技术与工程, 2024, 24(19): 8166-8175.
Gong Ying, Xu Wentao, Zhao Ce, et al. Federated learning model based on zero trust mechanism[J]. Science Technology and Engineering, 2024, 24(19): 8166-8175.

基于零信任机制的联邦学习模型

龚颖, 许文韬, 赵策, 王斌君*
(中国人民公安大学信息网络安全学院, 北京 100240)

摘 要 为使联邦学习能够满足更高的安全与效率需求,提出了一种采取双重加密与批处理加密方法的零信任模型。首先,利用双重加密防范来自服务器与其他参与方的多方威胁,且通过选取不同的加密方式并设置加密顺序,保证联邦学习模型在更安全的情况下正常运转;其次,在双重加密的基础上引入批处理模块,通过以密钥位数为依据的拆分再拼接操作,提升加密的效率,保证联邦学习模型在更高效的情况下正常运转。理论分析与实验结果表明:所提出的零信任机制的联邦学习模型能够防范来自多方的推理攻击,并维持与单层同态加密相近的开销。可见零信任机制在联邦学习中的应用具备相当程度的可行性,能够同时满足高安全性、高效率的需求。

关键词 联邦学习; 零信任; 双重加密; 批处理加密
中图法分类号 TP391; **文献标志码** A

Federated Learning Model Based on Zero Trust Mechanism

GONG Ying, XU Wen-tao, ZHAO Ce, WANG Bin-jun*
(Information Network Security College, People's Public Security University of China, Beijing 100240, China)

[**Abstract**] In order to enable federated learning to meet higher security and efficiency requirements, a zero trust model using double encryption and batch encryption was proposed. Firstly, double encryption was used to prevent multi-party threats from the server and other participants. By selecting different encryption methods and setting the encryption order, the federated learning model can be guaranteed to operate normally in a more secure environment. Secondly, the batch processing module was introduced on the basis of double encryption. Through splitting and splicing operations based on the number of key bits, the efficiency of encryption was improved to ensure the normal operation of the federated learning model in a more efficient manner. Theoretical analysis and experimental results show that the proposed federated learning model of zero trust mechanism can prevent inference attacks from multiple parties, and maintain the overhead similar to that of single-layer homomorphic encryption. It can be seen that the application of zero trust mechanism in federated learning has a certain degree of feasibility, and can meet the requirements of high security and high efficiency at the same time.

[**Keywords**] federated learning; zero trust; double encryption; batch encryption

随着5G、物联网与人工智能等技术快速发展及深度应用,各类行业内的数据呈现出井喷的状态,也推动了深度学习等数据挖掘技术的深入研究和广泛应用,人类由此进入了大数据时代。在大数据时代,数据已然成为人类社会生产、生活的要素,数据需要共享,才能充分地增益其价值,以数据为生产要素的共享经济方能形成。然而,海量数据自然状态下是分散在各采集点的本地应用系统、移动设备中;且大数据中包含敏感信息、个人信息等,使得大数据共享受到诸多的限制。以欧盟《通用数据保护条例》和中国的《数据安全法》为代表的各类数据处理和管理规定的问世说明了保护个人隐私和维护数据产权的重要性,对原始数据的共享进行了相关限制。因此,保护个人信息与促进数据安全共享并举的要求促使联邦学习这一概念孕育而出。

在联邦学习(federated learning, FL)[1]的场景下,原始数据被保持在原持有者的控制之下,而对其他参与方不可见。各参与方共享一个初始化状态的神经网络模型,并分别使用其本地数据进行模型训练,然后将模型参数的梯度值上传至聚合服务

收稿日期: 2023-05-15; **修订日期**: 2023-11-27
基金项目: 国家社会科学基金(20AZD114)
第一作者: 龚颖(1999—),女,汉族,湖北襄阳人,硕士研究生。研究方向:联邦学习、生成对抗网络。E-mail:931415696@qq.com。
* **通信作者**: 王斌君(1962—),男,汉族,河南偃师人,博士,教授。研究方向:神经网络、深度学习、网络安全与执法。E-mail:2335371229@qq.com。

器,聚合服务器负责将梯度求取平均并分发给各参与方,各参与方使用聚合梯度更新神经网络模型。在聚合服务器的协调下,通过本地训练、梯度上传、梯度聚合以及梯度更新,对共享的神经网络模型进行迭代优化,实现分布式机器学习。联邦学习在数据不流动的条件下,使用梯度信息的流动替代了原始数据的流动,构建了一种新的协同计算范式。这为多个数据持有者之间进行数据共享与数据挖掘、数据增益提供了新的解决方案,符合大数据时代的实际需求,在多个领域得到了广泛的应用[2-4]:互联网领域运用 FL 在不侵犯个体隐私的情况下收集群体行为模式信息以优化搜索引擎、输入法联想等;金融领域运用 FL 综合多方数据实现各企业单位之间的安全合作与互利共赢、并可实现风险评估与定向推广等;医疗卫生领域运用 FL 在不暴露患者病情信息的同时完成信息整合以进行更好的诊断;在其他诸如政务、物流、自动驾驶等领域,FL 也同样大放异彩。

近年来,随着联邦学习深入研究及实践应用,其隐私安全问题获得了广泛关注。研究者们发现联邦学习仍然存在隐私泄露的风险[5]。Zhu 等[6]提出了根据泄露的梯度进行推理的方法,诚实但好奇的聚合服务器或恶意参与者可以利用某参与者梯度信息,还原其原始数据,进而导致数据隐私泄露。造成泄露的根本原因在于梯度信息是由原始数据信息计算、抽象得到的,那么一定存在方法可以利用、分析梯度信息甚至还原原始数据信息。针对这一系列隐私泄露安全隐患,研究者们提出了一系列将隐私保护技术应用于梯度信息的方法和技术,主要包括:以差分隐私为代表的数据扰动法[7]、以 k-匿名为代表的数据泛化法[8]、以安全多方计算为代表的数据加密法[9]等。

在中国也同样有相当多的研究者对联邦学习攻防技术这一联邦学习系统安全的核心问题进行了相关研究[10],并从联邦学习系统有无目标性的防御措施出发,将防御措施分为通用性防御措施及针对性防御措施两类。前者如张海超等[11]提出的边缘计算下的轻量级联邦学习隐私保护方案,提出一种云-边-端分层的联邦学习框架;其次,对不同层进行隐私保护;最后,提出一种周期性更新策略,极大地提高了收敛速度。后者如许文韬等[12]提出的基于随机断层与梯度剪裁的横向联邦学习后门防御研究,中心服务器在收到参与方提交的梯度信息后,随机确定每个参与方的神经网络层,然后将各参与方的梯度贡献分层聚合,并使用梯度阈值对梯度参数进行裁剪,实现对后门攻击的防御。

同时有一些主要用于防范其他类型攻击的技术,也可以对模型的隐私保护产生助益。如杨宇等提出的入侵检测方法[13]、江欣俞等[14]提出的基于兴趣点推荐的方法等框架。在针对联邦学习以及其所依托的神经网络框架所采取的多样保护中,以同态加密(homomorphic encryption,HE)为当前联邦学习系统中所最常用。

然而,同态加密算法仅防范诚实但好奇的聚合服务器的隐私推理行为。由于参与方共用一组同态加密密钥,好奇的参与方可截获其他参与方的梯度信息,使用共有的私钥对信息解密,获取该参与方的原始梯度信息,进而推理还原其原始数据。故而仅采用同态加密隐私保护机制仍未完全解决联邦学习中的隐私安全问题。同时,同态加密算法在应用过程中还存在着加密时间长,数据传输量大等问题。

有鉴于此,基于梯度拼接与量化技术[15]、Paillier 同态加密技术[16]和 AES(advanced encryption standard)加密通信技术[17]提出了一种基于零信任机制的联邦学习模型。该方法通过梯度拼接与量化技术将模型的参数进行压缩,降低了后续同态加密和参数传输中的数据量,通过同态加密保证了各参与方的梯度参数对于聚合服务器的匿名性,在交换梯度参数的通信过程中使用会话密钥的 AES 加密使得保证了参与方之间无法截获其他参与方的梯度信息,由此在保证了梯度参数在各参与方(包括聚合服务器)之间的匿名性的同时实现了梯度参数的高效计算和分发。下面将分别予以阐述。

1　相关工作

针对联邦学习隐私安全问题的相关讨论一直在进行。Paillier 等[18]提出的同态加密算法被应用到联邦学习模型后,由于参与方与服务器之间的梯度信息传输得到了加密保护,较好地防范了不可信服务器的各种隐私攻击。然而,在实际中,由身份更为隐蔽、数量更为庞大的参与者所主导的攻击同样层出不穷,这些攻击难以被同态加密防御。

为更清晰阐述本文对联邦学习模型防御恶意参与者攻击能力的改进,下面将从模型容易遭受的主要推理攻击方式及相应防御措施方面,对已有联邦学习模型安全性进行梳理分析。

1.1　针对模型训练数据的推理攻击

常见的针对机器学习模型进行隐私推理攻击的方式,根据其想要获取的隐私信息类别可大致分为:①针对模型参数数据的推理攻击(即模型窃取攻击[19]);②针对模型训练数据的推理攻击。联邦

学习中各参与者天然持有模型参数数据,无需实施模型窃取攻击;而与训练无关的非参与方在同态加密的传输环境下,难以获取到有效的模型参数信息,无法实施模型窃取攻击。有鉴于此,本节将主要围绕训练数据集的隐私问题,对各种以参与者为主导的攻击按照被动和主动分类进行归纳、梳理。

1.1.1 被动推理攻击

被动推理攻击的攻击者以消极模式(passive mode)进行观察,而并不主动影响模型以获取额外的隐私信息。该类攻击在传统机器学习与联邦学习等模型中均可实施,且攻击行为较为隐蔽,难以察觉。

Shokri 等[20]于 2017 年首次提出了成员推理攻击[21],将隐私推理视为一个二分类的问题,即针对给定数据,判断其是否属于目标模型的训练数据集。其通过多次影子训练(shadow training)区分目标模型对属于训练集中的数据输入与不属于训练集中的数据输入的行为差异,利用该差异将给定的目标数据点进行二分类,从而得到其与隐私训练数据集的隶属关系。然而 Shokri 等[20]的方法需要通过多次影子训练得到多个影子模型(shadow models),攻击成本较高,且对训练数据的分布、模型结构等都提出了较高的要求,攻击的落地性不强。Salem 等[22]在其方法的基础上进行了数次改进,放宽了该攻击方法对数据分布与模型结构的限制,适用于任何类型的训练数据集,并提升了攻击的准确率与召回率。

Nasr 等[23]综合前人研究,利用随机梯度下降算法的隐私漏洞设计了一种针对联邦学习模型的白盒隐私推理攻击。其将从目标模型中所获得的梯度信息作为攻击模型的输入,分别处理从目标模型不同层提取的梯度特征,并将其结合来计算目标数据点的隶属度概率,该方法对泛化良好且具有大量参数的神经网络十分有效。同时,Nasr 等[23]还考虑到了攻击者是否具有先验知识这一问题,对于知晓目标训练集某个子集的攻击者,采取有监督的方式训练攻击模型;而对缺乏相应先验知识的攻击者,则使用无监督的方式训练攻击模型。这一区分可帮助不同类型的攻击者在实践中取得了更好的效果。

Melis 等[24]也针对联邦学习、协作学习(collaborative learning)等联合模型提出了成员推理攻击,并罕见地在自然语言文本上也进行了实践。自然语言文本模型会先用嵌入层将输入转化为低维向量表示,而其中给定文本嵌入矩阵的更新仅仅与单词是否出现在文本中相关联,如果单词未出现在文本中,则其梯度为零。Melis 等[24]正是利用这一特点,在一轮更新中得到关于此轮更新里参与训练的词袋(bag of words),对于给定的文本记录,只需判断其是否在词袋中,即可知晓其是否属于本轮训练的训练集。该方法同样可以用于推知特定的数据何时首次出现、何时消失在训练集中。这同样对隐私问题造成严重影响,比如可以知晓某位参与者何时开始拜访某种类型的医生。

1.1.2 主动推理攻击

主动推理攻击的攻击者以积极模式(active mode)参与到目标模型的训练中来,通过有意改变其上传的更新参数等主动行为,引导目标模型透露出更多隐私信息,其攻击性更强。

Nasr 等[23]设计了一种主动攻击方式,恶意参与方在上传和更新全局参数之前,先对一组目标数据点执行梯度上升。这一行为放大了其他人训练集中该数据点的存在,如果该目标数据点属于训练集,随机梯度下降就会突然降低目标数据点上的梯度,从而导致隐私泄露。

Melis 等[24]提出一个主动的攻击者可以使用多任务学习来执行更强大的攻击,通过一个连接到最后一层的增强属性分类器来扩展其协作训练模型的本地副本,在处理训练任务的同时识别此属性,训练数据含有一个主标签 y 和一个属性标签 p,通过计算联合损失,使模型学习到有该属性的数据和无该属性的数据的可分离表示,并对梯度信息进行分离,从而攻击者即可判断判断训练数据是否具有该属性,进而泄露隐私。值得注意的是,在密码学的角度上,该类攻击者仍然是“诚实但好奇的”,忠实遵循学习协议且不提交错误信息。该类攻击与被动攻击的区别是执行了额外的本地计算。

Hitaj 等[25]在其所提出的基于生成对抗网络(generative adversarial networks,GAN)隐私推理攻击方法中,则是通过在联邦学习中提交错误的信息引导被攻击者泄露更多隐私。该攻击并不在意攻击者是否拥有先验知识,任何精确的深度学习机器,无论它如何被训练,都可能泄露关于它可以区分的不同类的信息。Hitaj 等[25]利用了这一点,将联邦学习中每轮更新的联合模型作为判别器,攻击者训练生成器与之对抗,只要联邦学习的模型能够正确对输入进行分类,攻击者所持有的生成器就可以从中恢复出训练数据,侵犯其他参与者的隐私。

1.2 防御措施

隐私推理攻击能够奏效的原因主要在于以下两点:①过拟合(overfitting)。过拟合是一种机器学

习行为,是指机器学习模型为训练数据而非新数据提供准确的预测,模型的过拟合会导致训练数据与非训练数据易于区分。②训练集数据不具有代表性。当训练集数据不具有代表性时,与测试集数据分布显著不同,也会致使隐私推理攻击的成功。

为了抵御层不出穷的攻击方式,研究者们也相应提出了诸多防御措施,主要集中在①模型修改;②扰动添加;③数据加密。

1.2.1　模型修改

模型修改即通过对模型本身的结构、参数等进行调整与改进,使得模型尽量少地学到与主任务无关的训练集额外信息(隐私信息),减少模型自身过拟合程度。在完全理想的情况下,模型自身无法习得与主任务无关的信息,其输出特征向量、梯度等信息同样无法泄露隐私,从而在根本上规避了隐私风险。

模型修改的常用方法为随机失活(dropout)与模型堆叠(model stacking)。随机失活由 Geoffrey 等[26]于 2012 年提出,即在训练过程中每次随机使部分隐藏神经元失活,在实践中这一方法被证明能够大大减少模型的过拟合程度,并在之后被广泛沿用[27-28]。模型堆叠的考量则在于使用不同的数据子集训练模型的不同部分,组合起来的完整模型同样被证明不易发生过拟合。

在实际中,机器学习模型特别是深度学习模型的神经元,在训练中或多或少总会习得一些与主任务无关的训练数据集隐私信息。通过模型修改进行防御的方法,在抵御部分基于输出特征向量进行隐私推理的黑盒攻击方案上或有效果,而在联邦学习的环境下作用较为有限。

1.2.2　扰动添加

扰动添加是指通过添加扰动的方式,使可能会被窥探到的梯度信息等无效化,攻击者难以从中推理出真正的隐私。该类防御的代表方法是差分隐私(differential privacy,DP),根据添加扰动位置的不同,可大致分为本地差分隐私(由客户端本地添加扰动)、分布式差分隐私(由可信中间节点添加扰动)、中心化差分隐私(由服务器添加扰动)。其中,联邦学习模型下所采用的基本为本地差分隐私(local differential privacy,LDP)。

Bhowmick 等[29]在数据层面上提出并实现了差分隐私技术,并设计了新的最优隐私机制。即先对数据进行差分隐私扰动后,再将扰动后的数据投入模型进行训练。Heikkilä 等[30]与 Kerkouche 等[31]则聚焦于目标函数扰动,对本地模型的目标函数添加差分隐私噪声项,通过该目标函数训练模型,从而得到差分隐私保护的本地模型。相较而言,更为常见的是输出扰动这一类 LDP[32-34],对数据参与方上传的梯度加入差分隐私噪声项,使得聚合后的全局模型是差分隐私保护的。

无论是哪一种扰动添加,往往会导致模型的可用性降低,参数选取上均需要在模型主任务精确度与隐私保护之间小心平衡,对准确度较高的场合,如人脸识别、金融风险计算等,难以应用该技术。同时差分隐私的应用也仅仅是以模型可用性为代价换来隐私泄露的量化可控,并非完全安全,所泄露的隐私仍会带来一定程度的风险。

1.2.3　数据加密

数据加密通过密码学的方法进行加密保护,使梯度等关键信息不会被泄露,从而攻击者无法窥探隐私信息,该类代表防御方法是同态加密。目前,主流的联邦学习隐私保护技术是采用各参与方的梯度参数对聚合服务器可用不可见的同态加密技术实现。

同态加密是一种允许在密文上进行数学运算的加密方法。同态加密的关键特征就是解密经数学运算的密文获得的输出和仅在明文上进行数学运算操作获得的输出一致。现阶段同态加密方案主要有 3 类:近似同态加密(somewhat homomorphic encryption,SHE)、部分同态加密(partially homomorphic encryption,PHE)和全同态加密(fully homomorphic encryption,FHE)。近似同态加密支持加法运算和乘法运算,但仅限于一定数量的操作。由于近似同态加密中每次操作都会增加噪声,在噪声达到一定量之后,便无法再解密数据。部分同态加密支持任意数量的操作,但仅限于一种类型的操作。全同态加密支持对密文进行任意次数的加法操作和乘法操作。由于在联邦学习的过程中聚合服务器上基本只涉及加法,因此加解密开销较小的 Paillier 部分同态加密系统是最为常见的选择。不过即使选择 Paillier 部分同态加密系统,同态加密技术仍然存在加密小数慢、安全密钥位数长、计算复杂度高、开销成本大等问题。有鉴于此,研究者们采取了多种手段进行干预,比如对明文进行拼接后整体加密以减小密文总长度[35]、在同态加密的同时引入安全多方计算[36]技术来高效保护隐私[37-38]等。

虽然使用同态加密方法能够较好地保护隐私,但如前文所述,当前主流联邦学习框架中的同态加密算法只考虑了服务器不可信的情况,却对同样可能窃取梯度信息的其他参与方几乎不设防,恶意参与方的威胁普遍存在,相关的隐私保护技术亟待被提出。

2 零信任机制的联邦学习安全模型

2.1 一般隐私原则

安全且理想的联邦学习模型应当在保护各方隐私的前提下进行高效学习。根据 Nasr 等[23]的描述,本文中将模型的训练数据隐私泄露定义为“攻击者可以从模型中学习到一些信息,该信息无法从其他基于相同分布的训练数据所训练的其他模型中推断出来”。这就区分了人们可以从模型中学习到的关于数据总体的信息和模型泄漏的关于在其训练集中的样本特定数据的信息。

如第 1.2 节中所述,现行的诸多联邦学习模型在防范隐私攻击时仅考量了特定身份的攻击者,而防范措施无法覆盖所有类型的攻击者会在实际应用中带来巨大的安全隐患,提出的零信任机制联邦学习模型在这方面作了尝试。

2.2 模型工作流程

在本节对所提出的零信任联邦学习模型之工作流程予以描述,阐释了在参与方(数据所有者)与服务器(模型所有者)这两个联邦学习系统必备组件[39]中所执行的操作,探讨了如何利用双重加密与量化拼接等手段实现安全、高效的联邦学习。

零信任联邦学习模型的核心思想在于双重加密,即参与方每轮在将梯度提交予服务器前,先进行两次加密:第一次同态加密旨在防范不可信服务器,其非对称加密密钥由各参与方之间彼此共享;第二次加密旨在防范不可信的其他参与方,其通过会话密钥过程获得的对称密钥仅由该参与方自身与服务器共享。服务器在接收到各参与方二次加密的梯度后,先使用其与各参与方分别共享的会话密钥进行一次解密,并将其结果(实际仍为密文)进行聚合得到全局更新,然后分别返回给各参与方。在整个流程中,无论是服务器抑或是各参与方,所能直接接触到的非己方梯度信息均以密文形式呈现,因而有效保障了联邦学习的隐私安全。

为描述方便,假设:① $N=\{1,2,\cdots,N\}$ 表示 N 个参与方的集合:其中每个参与方 k 都拥有自己的隐私数据集 $D_{k\in N}$,各参与方分别使用自己的数据集训练本地模型,将且仅将模型更新梯度共享到服务器。② β、ε 分别表示参与方两次加密所使用的密钥:其中,β 表示第一次加密(同态加密、非对称加密)所使用的密钥,β_{pub}表示公钥,β_{pri}表示私钥,均仅由各参与方持有,服务器在同态加密特殊性质的保障下直接对密文进行操作;ε 表示第二次加密(对称加密)所使用的密钥,ε_k表示服务器与第 k 个参与方的会话密钥[40],ε_k仅由第 k 个参与方与服务器在联邦学习前产生并共同持有。

零信任联邦学习算法流程如下。

算法 1 零信任联邦学习算法

secret_sharing(client_1,client_2,…,client_N,β_{pub},β_{pri})
//在各参与方之间产生并共享非对称密(β_{pub},β_{pri}).
secret_sharing(client_k,server,ε_k)
//分别在参与方 k 与服务器之间产生并共享会话密钥 ε_k.
[服务器]
initializeW_{G}^0
model_sharing(W_{G}^0)
//将 W_{G}^0 分发给各参与方.
for t in round$\{1,2,\cdots,r\}$ do:
[参与方]
$\Delta W_k^t \leftarrow$local_training(t,μ,W_{G}^{t-1},D_k)
//μ 为学习率,ΔW_k^t 为参与方 k 第 t 轮的梯度变化.
$\Delta W_{ke1}^t \leftarrow$encryption($\beta_{\text{pub}}$,$\Delta W_k^t$)
$\Delta W_{ke2}^t \leftarrow$encryption($\varepsilon_k$,$\Delta W_{ke1}^t$)
[服务器]
$\Delta W_{ke1,k=1,2,\cdots,N}^t \leftarrow$decryption($\varepsilon_k$,$\Delta W_{ke2,k=1,2,\cdots,N}^t$)
$\Delta W_{\text{Ge1}}^t \leftarrow$aggregation($\Delta W_{1e1}^t$,$\Delta W_{2e1}^t$,…,$\Delta W_{Ne1}^t$)
model_sharing(ΔW_{Ge1}^t)
//将 ΔW_{Ge1}^t 分发给各参与方.
[参与方]
$W_{\text{G}}^t \leftarrow$decryption($\Delta W_{\text{Ge1}}^t$,$\beta_{\text{pri}}$,$\varepsilon_k$)
$W_{\text{G}}^t \leftarrow$model_update($\Delta W_{\text{G}}^t$)
//更新本地模型.
end for.

(1)参与方所执行的操作。参与方进行本地训练并更新参数:各参与方从服务器处获得初始参数 W_{G}^0,并以己方所拥有数据为基础进行学习,最小化损失函数 $L(\Delta W_k^t)$[41],并搜索最优超参数。当参与方完成训练,依次使用密钥 β_{pub}、ε_k对其梯度更新进行加密,再将其传至服务器。这些步骤有效防止了隐私的泄露,使得无论身份是服务器还是其他参与方的攻击者均难以获知有效信息。

(2)服务器所执行的操作。服务器首先进行权重初始化:全局模型 W_{G}^0 以及相关超参数从服务器端开始传播。

服务器而后进行聚合与全局更新:服务器利用其与各参与方 k 所分别共享的密钥 ε_k对加密更新进行一次解密,然后利用同态加密的同态性对一次解密后(仍为密文)的信息进行聚合。ΔW_{G}^t 来自各参与方的局部梯度更新,服务器最小化全局损失函数 $L(\Delta W_{\text{G}}^t)$[39]。服务器聚合并更新全局参数后,再发送给各参与方。全局损失函数表达式为

$$L(\Delta W_G^t) = \frac{\sigma}{N}\sum_{k=1}^{N} L(\Delta W_k^t) \tag{1}$$

式(1)中:σ 为全局学习率。

2.3　安全性与性能分析

2.3.1　同态加密

零信任联邦学习模型的安全性首先依赖于同态加密的实现。同态加密对密文直接进行处理,得到与先处理明文再加密相同的结果。从抽象代数的角度讲,保持了同态性,可以实现令处理者无法访问到数据自身的信息。即:定义一个运算符⊙,对于加密算法 E,满足

$$E(X\odot Y)=E(X)\odot E(Y) \tag{2}$$

则意味着该运对于算满足同态性[42]。

同态加密的这种特殊性质使得其再联邦学习框架中发挥出巨大的作用,令数据处理者(服务器)无法直接访问数据本身,保证了各参与方与服务器联络的隐私安全性。

2.3.2　零信任模式

在同态加密算法的有效加持之下,来自不可信服务器的威胁被有效遏制,然而联邦学习模型依然暴露在来自其他参与方的可能攻击之下。我们使用零信任模式对此给出的解答是,在进行敏感数据(如梯度信息等)传输之前,各参与方先使用同态加密密钥进行一次加密,防范不可信服务器;而后再使用各参与方分别与服务器共享的密钥进行二次加密,以防范其他参与方。经过如是双层加密,敏感数据的每一步传输都以密文形式进行,且不被除数据所有者以外的任何处理方、窃听者所直接接触,有效保障了联邦学习模型整体的安全性。

2.3.3　批处理加密

虽然在 2.3.1、2.3.2 两节之中,阐述了所提出零信任模型的安全性,然而众所周知,加密操作会带来比明文更大的数据传输量,其中以同态加密尤甚,甚至在实际操作中能够将所需传输的数据数量增加两个数量级[43]。以 Pailliar 算法为例,若对每个梯度参数逐一加密,该算法要求密钥位数与明文位数相同,而通常密钥位数远大于梯度参数的位数长度,这就需要将每个梯度信息均膨胀至密钥位数的长度,从而势必造成更大的时间开销和通信开销。同时,本文所提出的零信任模式由于涉及两次加密,更是令计算与通信成本只增不减。WeBank[44] 等也曾表示其大多数 FL 应用程序无法负担加密的梯度,因而在许多情形下会弃用这一安全的学习方式。

为了解决加密所带来的效率低下问题,使其更能适用于实际需要,提出批处理加密方法。之前的部分研究者已经做出了类似设想[45-46],并在同态加密的联邦学习任务中表现出良好的效果。将其改进并引入到零信任机制中,大大减小了双层加密所带来的开销。

批处理加密部署在联邦学习的客户端,在联邦学习开始前,各参与方即进行批处理加密参数(包括编码基数 t,梯度范围 $\Delta_{\max}$ 与 $\Delta_{\min}$,编码位数 r 等)协商,并在之后每轮得到梯度参数信息时,执行编码与拼接两个阶段。

编码阶段所执行的操作:将通常以小数形态存在的梯度参数转化为 r 位正整数,把梯度参数 Δ 转化为范围在 $[0,2^t]$ 的 batch_code(Δ),从而避免了 Pailliar 算法加密小数较慢的问题。

$$\Delta_c=\text{batch\_code}(\Delta)=\lfloor 2^t\frac{\Delta-\Delta_{\min}}{\Delta_{\max}-\Delta_{\min}}\rfloor \tag{3}$$

式(3)中:t 为编码基数,当时,编码所造成的误差小于 e^{-5},可实现无损编码[43];Δ 为原始梯度值,且应介于 $\Delta_{\max}$ 与 $\Delta_{\min}$ 之间,如果参与者本地训练所得 Δ 无法满足既定梯度范围,则可先执行梯度裁剪。

拼接阶段所执行的操作:将若干个表达梯度参数的整数拼接为一个长整数。以密钥长度 $k=2\,048$ 为例,2 048 位二进制最多可表示($\lfloor \lg(2^k)\rfloor+1$)=617 位十进制整数,根据协商所得的编码位数 r,则可知每次最多可拼接 $\lfloor 617/r\rfloor$ 个整数。参与方按照零信任机制的要求,对拼接后的若干长整数执行两次加密后,提交至服务器。由此,同态加密效率可提高 $\lfloor 617/r\rfloor$ 倍。

在联邦学习进行过程中,服务器仍先按照算法 1 中的机制对梯度参数进行对称解密、聚合、分发等操作。参与方则在收到梯度更新值 $\Delta W_{\text{Gel}}^{t'}$($t'$为联邦学习进行轮数)并执行同态解密后,若干长整数中提取出原始梯度信息公式为

$$\Delta=\text{batch\_decode}(\Delta_c)=\Delta_c\frac{\Delta_{\max}-\Delta_{\min}}{2^t}+N\Delta_{\min} \tag{4}$$

式(4)中:N 为参与方总数,参与方在对长整数解码后,得到聚合梯度信息,更新本地模型,并等待进行下一轮联邦学习。

除此之外,为避免服务器对长整数执行聚合操作时出现溢出问题,编码位数 r 的选取至关重要。在实际应用中,可以根据参与方总数 N、采样率 s、编码基数 t、梯度范围等来进行选取,设置编码位数 r,以便为梯度聚合时留出进位空间。方法为

$$r=(\lfloor \lg(2^tNs)+1\rfloor) \tag{5}$$

式(5)中:$\lfloor\cdot\rfloor$ 表示向下取整。

以本文的实验环境为例,选取 $t=15$ 作为编码基数,参与方总数 $N=50$,采样率 $s=0.2$,梯度范围 $\Delta\in[-1,1]$,则编码后的梯度信息 batch_code(Δ)$\in[0,2^t=32\,768]$,需占用($\lfloor \lg(2^t)\rfloor+1$)=5 个十进制位,考虑到每轮将有 $N\times s=10$ 位参与方投入到联

邦学习中,则根据式(5)、式(6)可知聚合梯度的取值范围为[0,327 680],最佳编码位数 $r=6$。

$$\Delta W_{\mathrm{Ge1}}^{t'} \in [0,2^{t}Ns] \tag{6}$$

综上所述,采用批处理加密技术的零信任联邦学习算法可总结如算法 2 所示。

算法 2 应用批处理加密的零信任联邦学习算法

1. secret_sharing(client_1,client_2,…,client_N,β_{pub},β_{pri})
//在各参与方之间产生并共享非对称密(β_{pub},β_{pri}).
2. secret_sharing(client_k,server,ε_k)
//分别在参与方 k 与服务器之间产生并共享会话密钥 ε_k.
3. batch_code_consult(t,$\Delta_{\max}$,$\Delta_{\min}$,r)
//参与方协商批处理加密相关参数.
//其中 t 为编码基数,$\Delta_{\max}$ 与 $\Delta_{\min}$ 为梯度范围,r 为编码位数.
[服务器]
4. initialize W_{G}^{0}
5. model_sharing(W_{G}^{0})
//将 W_{G}^{0} 分发给各参与方.
6. for t' in round{1,2,…,r'} do:
[参与方]
7. $\Delta W_k^{t'}$←local_training(t',μ,$W^{t'}\_1G$,D_k)
//μ 为学习率,$\Delta W_k^{t'}$ 为参与方 k 第 t'轮的梯度变化.
8. $\Delta W_k^{t'}$←batch_code($\Delta W_k^{t'}$,t,$\Delta_{\max}$,$\Delta_{\min}$)
//将梯度编码为整数
9. $\Delta W_k^{t'}$←joint_gradient($\Delta W_k^{t'}$,r)
//拼接成若干长整数
10. $\Delta W_{ke1}^{t'}$←encryption(β_{pub},$\Delta W_k^{t'}$)
11. $\Delta W_{ke2}^{t'}$←encryption(ε_k,$\Delta W_{ke1}^{t'}$)
[服务器]
12. $\Delta W_{ke1,k=1,2,\cdots,N}^{t'}$←decryption($\varepsilon_k$,$\Delta W_{ke2,k=1,2,\cdots,N}^{t'}$)
13. $\Delta W_{\mathrm{Ge1}}^{t'}$←aggregation($\Delta W_{1e1}^{t'}$,$\Delta W_{2e1}^{t'}$,…,$\Delta W_{Ne1}^{t'}$)
14. model_sharing($\Delta W_{\mathrm{Ge1}}^{t'}$)
[参与方]
15. $\Delta W_{\mathrm{G}}^{t'}$←decryption($\Delta W_{\mathrm{Ge1}}^{t'}$,$\beta_{\mathrm{pri}}$)
16. $\Delta W_{\mathrm{G}}^{t'}$←batch_decode($\Delta W_{\mathrm{G}}^{t'}$,$t$,$\Delta_{\max}$,$\Delta_{\min}$)
//将梯度解码为长整数
17. $\Delta W_{\mathrm{G}}^{t'}$←return_gradient($\Delta W_{\mathrm{G}}^{t'}$,$r$)
//将长整数还原为聚合梯度
18. $W_{\mathrm{G}}^{t'}$←model_update($\Delta W_{\mathrm{G}}^{t'}$)
19. end for.

与算法 1 相比,算法 2 要求参与方在联邦学习开始前,协商批处理加密参数,并在本地训练结束后、两次加密前,将梯度参数编码、拼接为若干长整数;在参与方解密聚合梯度后,将梯度参数解码、还原为真正的聚合梯度。在同态加密时应用批处理加密方法,本文的实验环境下,可将用时压缩到了原所须时间的 1%(见 3.2.1 节);并且,在第二次加密时选取了 AES 加密算法,速度较快,使得双重加密并不会带来训练时间的显著延长,大大提升了其实际应用的可行性,提高了同态加密的效率。与此同时,批处理加密的方法即使在第二次 AES 加密中,仍然能够有效减小加密的数据量,进一步缩短 AES 算法的加密时长,更进一步提升了模型的高效性。改进后的模型在不影响原联邦学习模型的效率的前提下,有效确保了模型的安全性。

3 实验与分析

从图片分类与文本分类两个任务,将零信任机制部署于联邦学习环境中,统计通信和计算代价,并与仅采用同态加密的联邦学习通信和计算代价对比分析,验证了为联邦学习部署零信任机制的可行性。

3.1 实验设置

3.1.1 数据集

本文使用 FEMNIST[47]、SENTI140[48] 数据集的子集进行实验验证,如表 1 所示。FEMNIST 数据集属图片分类任务,包含了数字和大小写字母共 62 种数据类别。SENTI140 数据集属文本分类任务,数据来源于 Twitter,用于情感分析。数据集均属于 non-i.i.d[49-50] 性质。

3.1.2 模型与参数设置

本文使用 Lenet-5 模型用于图像分类、Bi-LSTM 模型用于文本分类。模型基本信息如表 2 所示。

其中,Lenet-5 模型包括 2 个卷积层、2 个最大池化层、3 个全连接层;Bi-LSTM 模型包括 1 个嵌入层、1 个双向 LSTM 层、1 个全连接层,词典容量为 20 000,嵌入维度为 4、隐藏层特征维度为 32。

联邦学习运行参数如表 3 所示。

联邦学习实验设置了 $N=50$ 名参与者,采样率 $s=0.2$,采取批处理加密技术的编码基数 $t=15$,梯度范围选取[-1,1],即 $\Delta_{\max}=1$,$\Delta_{\min}=-1$ 联邦学习将运行 20 轮,统计参与方与服务器端各项用时情况。

表 1 数据集基本信息

Table 1 Basic information of dataset

数据集	训练样本数/个	数据属性	类别/个	测试样本数/个
FEMNIST	156 000	图片(1×28×28)	62	18 124
SENTI140	455 100	文本	2	142 913

表 2 模型基本信息

Table 2 Basic information of model

模型名称	数据集	模型参数量/个	训练任务
Lenet-5	FEMNIST	48 558	图像分类
Bi-LSTM	SENTI140	89 801	文本分类

表 3 联邦学习参数信息

Table 3 Parameter information of federated learning

参与者总数/名	采样率/%	全局学习率/%	运行轮次/轮
50	20	80	20

3.1.3　评价指标

本文采用的实验效果主要指标如下:

(1)参与方每轮平均用时(participate average time per round,PAT):全部参与方每轮训练平均用时。

(2)服务器每轮平均总用时(server average time per round,SAT):服务器每轮平均用时。

(3)服务器每轮等待平均用时(server average waiting time per round,SAWT):服务器每轮等待全部参与方均提交参数平均用时。

(4)服务器每轮聚合平均用时(server average aggregation time per round,SAAT):服务器每轮聚合梯度平均用时。

(5)服务器每轮解密平均用时(server decode time per round,SDT):零信任机制服务器AES解密平均用时。

3.2　实验对比分析

3.2.1　批处理加密技术可行性分析

本小节测试了批处理加密技术的可行性。分别在联邦学习环境下执行同态加密方案,对照组要求参与方对梯度参数逐一同态加密(执行算法1),实验组要求参与方采取批处理加密技术。图1显示了二者参与方每轮平均用时差异。

由图1可知,是否采取批处理加密技术会显著影响参与方每轮参与联邦学习所用时间。由于同态加密密钥长度长,导致加密时间长、易造成数据量膨胀。若参与方使用同态加密对梯度参数逐一加密,将导致参与方进程耗费大量时间对数据进行加密、服务器进程长期处于等待状态,造成资源的浪费。因此,参与方采取批处理加密技术,对数据进行预处理,可以大大缩短同态加密所需时间,提高同态加密于联邦学习的应用性。

3.2.2　批处理加密的零信任机制模型可行性分析

本小节测试了零信任机制联邦学习模型的可行性,统计其通讯开销,并与仅采用同态加密的联邦学习通讯开销比较。二者均采用批处理加密技术减小通信代价。具体时间开销情况如图2~图4所示。

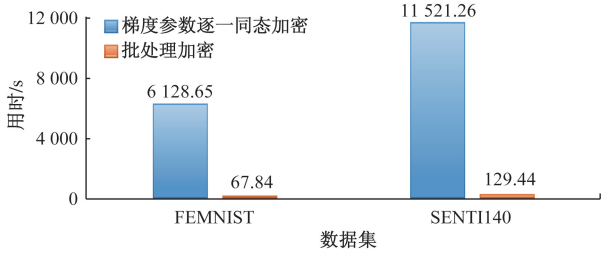

图1　量化与拼接参与方每轮平均用时

Fig. 1　The average time spent per round of quantification and splicing participants

其中,图2显示了采用不同数据集、不同训练模型时,仅采用同态加密与采用零信任机制联邦学习的参与方每轮平均用时。与仅采用同态加密方案相比,零信任机制要求每个参与方使用预先协商的AES对称密钥,对同态加密后的密文进行二次加密。AES加密算法加密速度快,并不会显著延长训练所用时间。因此,参与方每轮平均用时无显著差异。值得一提的是,采用批处理加密技术也可以减少待加密的数据量,缩短AES算法的加密时长。

图3、图4显示了采用不同数据集、不同训练模型时,仅采用同态加密与采用零信任机制联邦学习的服务器端每轮平均总用时及各项时间指标。服务器端总用时由等待全部参与方完成训练用时、聚合用时、AES解密用时(仅采取同态加密方案的联

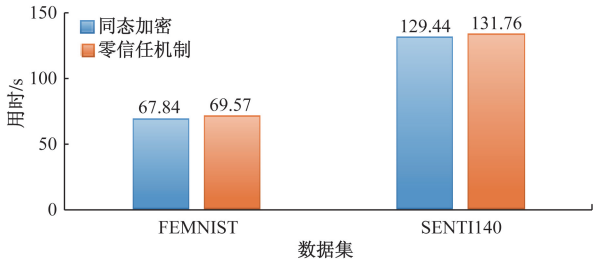

图2　不同方案下客户端每轮平均用时

Fig. 2　The average time spent on each round of the client under different schemes

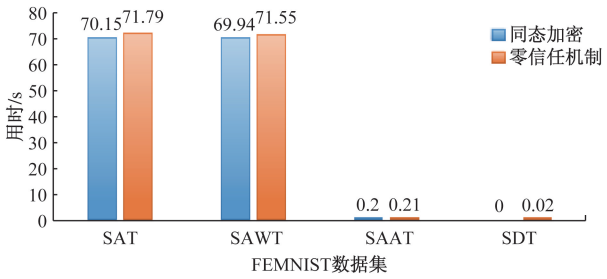

图3　FEMNIST不同方案下服务器端用时情况

Fig. 3　The time spent on the server side under different schemes of FEMNIST

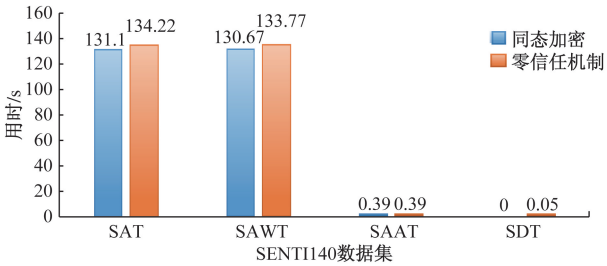

图4　SENTI140不同方案下服务器端用时情况

Fig. 4　The time spent on the server side under different schemes of SENTI140

邦学习无 AES 解密用时)组成。由图可知,采用零信任机制的联邦学习各项时间指标与仅采用同态加密方案的联邦学习无明显差异。

参与方、服务器各项时间差异均在 2 s 以内。因此,零信任机制的应用不会为现有采取同态加密的联邦学习增加过多的时间开销,证明了零信任机制在联邦学习环境下的可行性。

4 结论与展望

联邦学习为解决数据共享与隐私保护的矛盾提供了新方法,也已成为了由多方参与的分布式机器学习新范式。然而,联邦学习仍存在着隐私安全问题,且现有的隐私保护方法无法完全解决其中的问题。因此,本文提出了一种基于零信任机制的联邦学习模型。

(1)模型每个参与方均采用同态加密与 AES 加密结合的方法,保障每个参与方的梯度信息对服务器和其它参与方均不可见,从而实现零信任基础上的安全联邦学习。

(2)针对多重加密导致模型的性能降低问题,本文也探讨了批量加密处理的解决方案,基本能在不降低原模型效率的前提下保证模型的安全性。

基于零信任机制的联邦学习模型虽然防御了诚实但好奇的中心服务器和其他恶意参与方针对梯度信息泄露的隐私推理攻击,但仍无法防御基于 GAN 的隐私推理攻击。因此,联邦学习的隐私安全问题仍需要进一步研究。

参 考 文 献

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

[2] Xu J, Glicksberg B S, Su C, et al. Federated learning for healthcare informatics[J]. Journal of Healthcare Informatics Research, 2021, 5(1): 1-19.

[3] Lin B Y, He C, Zeng Z, et al. FEDNLP: Benchmarking federated learning methods for natural language processing tasks[C]//The Proceedings of Findings of the Association for Computational Linguistics: NAACL 2022. Stroudsburg: ACL, 2022: 157-175.

[4] Byrd D, Polychroniadou A. Differentially private secure multi-party computation for federated learning in financial applications[C]// The Proceedings of the First ACM International Conference on AI in Finance. New York: ACM, 2020: 1-9.

[5] 刘俊旭, 孟小峰. 机器学习的隐私保护研究综述[J]. 计算机研究与发展, 2020, 57(2): 346-362.
Liu Junxu, Meng Xiaofeng. Survey on privacy-preserving machine learning[J]. Journal of Computer Research and Development, 2020, 57(2): 346-362.

[6] Zhu L, Han S. Deep leakage from gradients[C]//The Proceedings of Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2019: 14747-14756.

[7] Dwork C. Differential privacy[C]// The Proceedings of ICALP 2006: Automata, Languages and Programming, 33rd International Colloquium. Venice: Springer, 2006: 1-12.

[8] Sweeney L. K-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge Based Systems, 2002, 10(5): 557-570.

[9] Bogetoft P, Christensen L D, Damgard I, et al. Secure multiparty computation goes live[C]// The Proceedings of FC 2009: International Conference on Financial Cryptography and Data Security. Accra Beach: Springer, 2009: 325-343.

[10] 张世文, 陈双, 梁伟, 等. 联邦学习中的攻击手段与防御机制研究综述[J]. 计算机工程与应用, 2024, 60(5): 1-16.
Zhang Shiwen, Chen Shuang, Liang Wei, et al. Survey on attack methods and defense mechanisms in federated learning[J]. Computer Engineering and Applications, 2024, 60(5): 1-16.

[11] 张海超, 赖金山, 刘东, 等. 边缘计算下的轻量级联邦学习隐私保护方案[J]. 计算机技术与发展, 2023, 33(9): 161-167.
Zhang Haichao, Lai Jinshan, Liu Dong, et al. Lightweight federated learning privacy protectionscheme under edge computing[J]. Computer Technology and Development, 2023, 33(9): 161-167.

[12] 许文韬, 王斌君. 基于随机断层与梯度剪裁的横向联邦学习后门防御研究[J]. 计算机科学, 2023, 50(11): 356-363.
Xu Wentao, Wang Binjun. Backdoor defense of horizontal federated learning based on random cutting and gradient clipping[J]. Computer Science, 2023, 50(11): 356-363.

[13] 杨宇, 闫钰, 申芳, 等. 基于机器和深度学习的入侵检测综述[J]. 科学技术与工程, 2023, 23(18): 7607-7621.
Yang Yu, Yan Yu, Shen Fang, et al. Review of intrusion detection based on machine and deep learning[J]. Science Technology and Engineering, 2023, 23(18): 7607-7621.

[14] 江欣俞, 李晓会, 秦若婷, 等. 基于图神经网络的兴趣点推荐的隐私保护框架[J]. 科学技术与工程, 2023, 23(17): 7407-7419.
Jiang Xinyu, Li Xiaohui, Qin Ruoting, et al. Privac-preserving framework for point of interest recommendation based on graph neural network[J]. Science Technology and Engineering, 2023, 23(17): 7407-7419.

[15] Zhang C, Li S, Xia J, et al. Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning[C]//The Proceedings of the 2020 USENIX Annual Technical Conference(USENIX ATC 2020). Boston: USENIX ATC'20, 2020: 493-506.

[16] Paillier P. Public-key cryptosystems based on composite degree residuosity classes[C]//The Proceedings of Advances in Cryptology (EUROCRYPT'99). Prague: Czech Republic, 1999: 223-238.

[17] Daemen J, Rijmen V. AES proposal: rijndael[C]//The Proceedings of First AES Candidate Conference(AES1). Ventura: NIST Journal of Research, 1999: 343-348.

[18] Paillier P. Public-key cryptosystems based on composite degree residuosity classes[C]//The Proceedings of International Conference on the Theory and Application of Cryptographic Techniques. Prague: Advances in Cryptology -EUROCRYPT'99, 1999: 223-238.

[19] F Tramèr, Fan Z, Juels A, et al. Stealing machine learning models via prediction APIs[C]//The Proceedings of 25th USENIX Security Symposium. Austin: USENIX Security, 2016: 601-618.

[20] Shokri R, Stronati M, Song C, et al. Membership inference at-

tacks against machine learning models[J]. IEEE Symposium on Security and Privacy(SP), 2017, 5(22): 3-18.

[21] 王璐璐, 张鹏, 闫峥, 等. 机器学习训练数据集的成员推理综述[J]. 网络空间安全, 2019, 10(10): 1-7.
Wang Lulu, Zhang Peng, Yan Zheng, et al. A survey of membership inference on machine learning training datasets[J]. Cyberspace Security, 2019, 10(10): 1-7.

[22] Salem A, Zhang Y, Humbert M, et al. Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models[C]//The Proceedings of 26th Annual Network and Distributed System Security Symposium. Rosten: The Internet Society, 2019: 01246.

[23] Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning[C]//The Proceedings of IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2019: 739-753.

[24] Melis L, Song C, De Cristofaro E, et al. Exploiting unintended feature leakage in collaborative learning[C]//The Proceedings of 2019 IEEE Symposium on Security and Privacy(SP). Piscataway: IEEE, 2019: 691-706.

[25] Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning[C]//The Proceedings of the 2017 ACM SIGSAC Conference. New York: ACM, 2017: 603-618.

[26] Geoffrey E H, Nitish S, Alex K, et al. Improving neural networks by preventing co-adaptation of feature detectors[EB/OL]. [2021-11-09]. https://arxiv.org/pdf/1207.0580.pdf.

[27] Wu L, Li J, Wang Y, et al. R-drop: Regularized dropout for neural networks[J]. Advances in Neural Information Processing Systems, 2021, 34: 10890-10905.

[28] Nguyen A T, Lu F, Munoz G L, et al. Out of distribution data detection using dropout bayesian neural networks[C]//The Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: Association for the Advancement of Artificial Intelligence. Reaton: AAAI, 2022: 7877-7885.

[29] Bhowmick A, Duchi J, Freudiger J, et al. Protection against reconstruction and its applications in private federated learning[EB/OL]. [2019-06-03]. https://arxiv.org/pdf/1812.00984.pdf.

[30] Heikkilä M A, Koskela A, Shimizu K, et al. Differentially private cross-silo federated learning[EB/OL]. [2020-07-10]. https://arxiv.org/pdf/2007.05553.pdf.

[31] Kerkouche R, Ács G, Castelluccia C, et al. Compression boosts differentially private federated learning[C]//Proceedings of the 2021 IEEE European Symposium on Security and Privacy (EuroS&P). Piscataway: IEEE, 2021: 304-318.

[32] Geyer R C, Klein T, Nabi M. Differentially private federated learning: A client level perspective[EB/OL]. [2018-03-01]. https://arxiv.org/pdf/1712.07557.pdf.

[33] Mcmahan H B, Ramage D, Talwar K, et al. Learning differentially private recurrent language models[EB/OL]. [2018-02-23]. https://arxiv.org/pdf/1710.06963.pdf.

[34] 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. New York: Association for Computing Machinery, 2019: 1-11.

[35] Zhang C, Li S, Xia J, et al. Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning[C]//Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC 2020). Boston: USENIX ATC'20, 2020: 493-506.

[36] Goldreich O. Secure multi-party computation[J]. Manuscript Preliminary Version, 1998, 78(110): 100-108.

[37] 张泽辉, 富瑶, 高铁杠. 支持数据隐私保护的联邦深度神经网络模型研究[J]. 自动化学报, 2022, 48(5): 1273-1284.
Zhang Zehui, Fu Yao, Gao Tiegang. Research on federated deep neural network model supporting data privacy protection[J]. Journal of Automation, 2022, 48(5): 1273-1284.

[38] 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.

[39] Pretom R O, Emon D, Nirmalya R, et al. Mixed precision quantization to tackle gradient leakage attacks in federated learning[EB/OL]. [2022-10-22]. https://arxiv.org/pdf/2210.13457.pdf.

[40] 王斌君, 佟晖. 信息安全技术体系[M]. 北京: 中国人民公安大学出版社, 2014: 48-56.
Wang Binjun, Tong Hui. Information security technology system [M]. Beijing: People's Public Security University Press, 2014: 48-56.

[41] Lim W Y B, Luong N C, Hoang D T, et al. Federated learning in mobile edge networks: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2031-2063.

[42] Acar A, Aksu H, Uluagac A S, et al. A survey on homomorphic encryption schemes: theory and implementation[J]. ACM Computing Surveys(Csur), 2018, 51(4): 1-35.

[43] Zhang C L, Li S Y, Xia J Z, et al. BatchCrypt: effificient homomorphic encryption for cross-silo federated learning[C]//Proceedings of the 2020 USENIX Annual Technical Conference. Boston: USENIX ATC, 2020: 493-505.

[44] Liu J, He X, Sun R, et al. Privacy-preserving data sharing scheme with FL *via* MPC in financial permissioned blockchain [C]//Proceedings of the 2021 International Conference on Communications. Xiamen: IEEE, 2021: 1-6.

[45] Liu C, Chakraborty S, Verma D. Secure model fusion for distributed learning using partial homomorphic encryption[J]. Policy-Based Autonomic Data Governance, 2019(1): 154-179.

[46] 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.

[47] Caldas S, Duddu S M K, Wu P, et al. Leaf: a benchmark for federated settings[EB/OL]. [2019-12-09]. https://arxiv.org/pdf/1812.01097.pdf.

[48] Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: The good the bad and the omg[C]//Proceedings of the International AAAI Conference On Web And Social Media. Menlo Park: Fifth International AAA Conference on Weblogs and Social Media, 2011: 538-541.

[49] Li Q, Diao Y, Chen Q, et al. Federated learning on non-iid data silos: An experimental study[C]//The Proceedings of 2022 IEEE 38th International Conference on Data Engineering(ICDE). New York: IEEE, 2022: 965-978.

[50] Kairouz P, McMahan H B, Avent B, et al. Advances and open problems in federated learning[J]. Foundations and Trends in Machine Learning, 2021, 14(1-2): 04977.