安全多方学习:从安全计算到安全学习

韩伟力 宋鲁杉 阮雯强 林国鹏 汪哲轩
(复旦大学计算机科学技术学院 上海 200438)

摘 要 如何在保护原始数据隐私的前提下,利用分散在多方的数据,高效且安全地完成高质量的机器学习模型训练和预测成为当前安全多方计算和机器学习两个研究方向的一个共同研究热点.本文在调研这一研究热点最新进展的基础上,提出安全多方学习这一概念.作为一个安全攸关软件工程领域的研究主题,安全多方学习是指基于安全多方计算实现隐私保护机器学习的方法、框架与平台.本文分析了安全多方学习中的安全模型、系统部署方式和功能场景,从底层安全多方计算原语和隐私保护技术入手,对现有安全多方学习框架进行了系统全面综述.首先,本文根据所使用的底层技术将安全多方学习框架分成了四类,并从计算复杂度、通信轮次、通信量、线性操作效率、非线性操作效率、支持的功能场景6个方面总结了不同安全多方学习框架的特点.进一步地,本文调研了38个典型的安全多方学习框架,根据支持的参与方数量、安全模型、功能场景,支持的机器学习模型,支持的激活函数,所实现的池化方式以及准确率等要素对它们进行对比,以展示它们的优势和局限.最后,本文分析了安全多方学习与其他隐私保护机器学习技术之间的区别,给出了安全多方学习提高安全性、可证明安全、提高性能和效率以及安全多方学习框架间的互联互通等方面的未来发展方向.

关键词 安全多方学习;机器学习;数据隐私;安全多方计算;隐私保护机器学习;访问控制
中图法分类号 TP309 **DOI号** 10.11897/SP.J.1016.2023.01494

Secure Multi-Party Learning: From Secure Computation to Secure Learning

HAN Wei-Li SONG Lu-Shan RUAN Wen-Qiang LIN Guo-Peng WANG Zhe-Xuan
(*School of Computer Science, Fudan University, Shanghai* 200438)

Abstract How to leverage the data distributed among/between multiple parties to efficiently and securely enforce high-performance machine learning training and inference with privacy preservation has become a hot spot of two research topics, i.e., secure multi-party computation and machine learning. This paper proposes the concept of secure multi-party learning based on the investigation of the latest developments in the hot spot. Secure multi-party learning, a research topic in (secure) software engineering rather than cryptography, hereby refers to the methods, frameworks, and platforms that enforce privacy-preserving machine learning based on secure multi-party computation. It enables multiple parties to perform secure training and secure inference of machine learning models without directly leveraging their plaintext data and any private information beyond the final result. Therefore, secure multi-party learning can be applied to several practical fields involving private data, such as risk control in the financial field and medical diagnosis. Researchers have proposed a dozen of secure multi-party learning frameworks recently. Considering the rapid development of secure multi-party learning, a comprehensive and systematic survey, which covers

收稿日期:2022-09-19;在线发布日期:2023-03-14.本课题得到国家自然科学基金项目"可编程安全多方学习机制及优化方法研究"(62172100)、上海市科委"科技创新行动计划"项目"基于大数据的电信网络诈骗受害人群评估与防范技术研究"(21DZ1201400)资助.
韩伟力(通信作者),博士,教授,中国计算机学会(CCF)杰出会员,主要研究领域为数据安全、隐私计算.E-mail:wlhan@fudan.edu.cn.
宋鲁杉,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为数据安全、隐私计算.E-mail:19110240022@fudan.edu.cn.
阮雯强,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为数据安全、安全多方学习、差分隐私.E-mail:20110240031@fudan.edu.cn.
林国鹏,博士研究生,主要研究方向为数据安全、隐私计算.E-mail:17302010022@fudan.edu.cn.
汪哲轩,硕士研究生,主要研究方向为数据安全、隐私计算.

the underlying technologies and classification of secure multi-party learning frameworks, is still absent so far. Therefore, this paper is motivated to conduct a literature review of the categories, characteristics, and frameworks of secure multi-party learning to help researchers choose suitable secure multi-party learning frameworks for various scenarios, further identify research gaps, and improve the weaknesses of secure multi-party learning frameworks. This paper analyzes the security models, system deployment methods, and functional scenarios in secure multi-party learning and starts with the underlying secure multi-party computation primitives and the privacy-preserving technologies to summarize secure multi-party learning frameworks systematically and comprehensively. The underlying technologies used in secure multi-party learning include holomorphic encryption, oblivious transfer, garbled circuit, and secret sharing. According to these underlying technologies, secure multi-party learning frameworks are classified into four categories: homomorphic encryption-based secure multi-party learning frameworks, garbled circuit-based secure multi-party learning frameworks, secret sharing-based secure multi-party learning frameworks, and mixed-protocol-based secure multi-party learning frameworks. Besides, this paper summarizes the characteristics of these four categories of secure multi-party learning frameworks from six aspects: computational complexity, communication rounds, communication size, linear operation efficiency, nonlinear operation efficiency, and functional scenarios supported. Further, this paper investigates 38 typical secure multi-party learning frameworks and compares them regarding the number of parties supported, security models, functional scenarios supported, machine learning models supported, activation functions supported, pooling implemented, and accuracy. Then, this paper analyzes the differences between secure multi-party learning and other privacy-preserving machine learning techniques, including federated learning and confidential computing based on trusted execution environment. Finally, this paper presents suggestions for future development of secure multi-party learning as follows: (1) to improve security, including support for a security model with stronger security guarantees, access control to the final model, and protection of the final model; (2) to prove the security of secure multi-party learning processes by the universally composable scheme; (3) to improve performance and efficiency by reducing the online communication overhead, accelerating the local computation with GPU and designing the machine learning models that adapt to the underlying technologies of secure multi-party learning; (4) to realize interoperability between secure multi-party learning frameworks.

Keywords　secure multi-party learning; machine learning; data privacy; secure multi-party computation; privacy-preserving machine learning; access control

1　引　言

随着互联网＋、云计算、大数据等技术的发展，数据爆炸式增长，几乎所有的活动都由数据所驱动，数据因而成为了当前最具时代特征的生产要素.数据、算力和算法相结合，逐渐演变为一种新型的社会生产力，对社会各方面发展的影响日益凸显.目前数据已经成为企业间竞争的关键，加强和创新社会治理的重要依据、重要资源和影响国家竞争力的重要因素.同时，机器学习成为图像处理[1]、语音识别[2]、自然语言处理[3]等研究领域的核心技术和工具，进而给金融风险预警[4]、医疗诊断[5]等多种社会生产和生活场景带来了技术突破.在这些场景中，海量的高质量训练数据通常是影响机器学习模型性能的主导因素.然而，机器学习所需要的数据通常又分散在多个数据所有者手中，并由于商业竞争或者不同的利益诉求而难以直接收集、合并或共享，从而形成了数据孤岛现象.

此外，由于这些数据通常蕴含较高的价值和具有较强的敏感性，数据安全问题甚至已成为事关国家安全与经济社会发展的重大问题.面对日益严峻

的数据安全形势，为了进一步保护个人隐私，世界各国和组织纷纷发布数据安全相关的法律法规，如欧盟的《通用数据保护条例》(General Data Protection Regulation，GDPR)[6]、我国的《数据安全法》和《个人信息保护法》等.这些法律法规的推出规范了数据的合规利用，但一定程度上加剧了数据孤岛现象.

因此，如何利用多个数据所有者持有的数据，并在保护这些数据隐私的前提下，高效地训练高性能机器学习模型，成为了当前机器学习领域的一个关键挑战[7-9].安全多方学习(Secure Multi-Party Learning，MPL)[10]，即基于安全多方计算(Secure Multi-Party Computation，MPC)的隐私保护机器学习技术[11-14]为此提供了一种可行方案.作为一个安全攸关软件工程(也即安全系统)领域的研究主题，安全多方学习通过安全的方法、框架和平台，使各个参与方在不泄露各自明文数据的前提下完成机器学习模型的安全训练或预测.

安全多方计算[15]最早由姚期智先生于 1982 年针对百万富翁问题提出，即两个百万富翁想比较谁更富有，但都不想让对方知道自己财富的具体数值，且不能借助于可信的第三方.随后，安全多方计算技术于 1986 年扩展为安全计算任意多项式可计算函数的一般定义[16]：安全多方计算技术可以使一组互不信任的参与方，在不依赖可信第三方的前提下安全地计算一个约定的函数，且不泄露除结果外的任何隐私信息.而基于安全多方计算的机器学习(也即安全多方学习)在无须依赖可信第三方的前提下，在保证各个参与方数据安全的同时，联合利用分散的各方数据训练机器学习模型，或者基于模型实现安全推理，从而更大程度发挥这些数据的价值，实现数据的可用不可得、可用不可见、可用不可拥.

1.1 相关综述

近年来，一些研究人员调研了隐私保护机器学习的相关技术.Hastings 等人[17]调研了 11 个安全多方计算的通用型编译器，包括 9 个端到端框架和 2 个电路编译器.他们从可用性(使用该编译器安装、运行和编写程序所需的工具和文档)、示例程序(三方乘法、内积、交叉制表)、功能性(数据类型、算子、语法)、实施标准(架构、计算模型、输入/输出)四个维度评估了这 11 个通用型编译器.Hastings 等人的调研重点在于分析编译器的可用性，而没有关注到编译器与机器学习的结合，且只分析了通用型编译器，对于安全多方学习框架缺少分析和调研.宋蕾等人[7]总结了机器学习安全及隐私保护研究的一些进展.首先根据机器学习的流程，对机器学习的敌手模型进行了描述，之后总结了机器学习常见的安全威胁以及应对的防御方法，以及常见的隐私威胁和相应的隐私保护技术，即同态加密和差分隐私.然而，宋蕾等人的综述没有调研并分析安全多方计算这一重要的隐私保护技术.谭作文和张连福[8]及魏立斐等人[18]也针对机器学习安全及隐私保护问题进行了综述.这两篇综述和宋蕾等人的研究思路相似，均以机器学习的敌手模型入手，总结了机器学习中常见的安全及隐私威胁，进而对机器学习隐私保护的主流技术进行了总结和比较.他们分析了同态加密、差分隐私以及其他安全多方计算技术.蒋瀚等人[9]针对隐私保护机器学习的密码学方法进行了综述.他们介绍了一些密码学技术，如安全多方计算、隐私保护集合运算、同态加密等，并对这些密码学技术在隐私保护机器学习中的各个阶段的应用进行了总结.综述[8-9,18]对最近这几年快速发展的安全多方学习框架的调研完备性不足，尤其是缺少对于树类模型安全训练和预测的框架分析.此外，所分析的安全多方学习框架特征较少，缺少框架所对应的底层安全多方计算技术，所支持的机器学习类型等特征.

1.2 本文贡献

近年来，在安全多方计算协议的支持下，研究人员设计了许多安全多方学习框架.然而，其底层协议昂贵的通信开销和较高的计算复杂性仍然是将这些框架应用于现实场景中亟需解决的两个重要问题.本文充分调研了现有的安全多方学习技术及框架，贡献如下：

(1)本文提出了安全多方学习概念，并从安全攸关的软件工程(也即安全系统)领域问题出发，全面分析了安全多方学习所采用的安全模型、系统部署和功能场景等.

(2)本文根据安全多方学习框架底层所依赖的原语和隐私保护技术，对最近几年提出的 38 个框架进行分类.此外，本文根据支持的参与方数量、安全模型、功能场景，支持的机器学习模型，支持的激活函数，所实现的池化方式以及准确率等要素比较所调研的安全多方学习框架，以展示它们各自的优势和局限性.

(3)本文分析了安全多方学习与其他隐私保护机器学习技术(如联邦学习、机密计算)的区别.并从安全性强化、安全性证明机制、效率和性能优化方法、不同安全多方学习框架的互联互通等方面给出

了安全多方学习的未来发展建议.

2　问题定义

从技术栈的角度，安全多方学习是指基于安全多方计算实现隐私保护机器学习的技术、框架与平台. 它基于一种或多种安全多方计算方法，联合并利用分布于多方的受保护数据，实现安全的分布式机器学习，是安全攸关软件工程（也即安全系统）领域的一个研究主题. 安全多方学习的技术栈如图 1 所示，底层是一组参与方和数据所有者，中间两层是安全多方学习中会用到的一些底层原语和现有的一些安全多方学习框架，最上层是一系列安全多方学习所支持的机器学习模型.

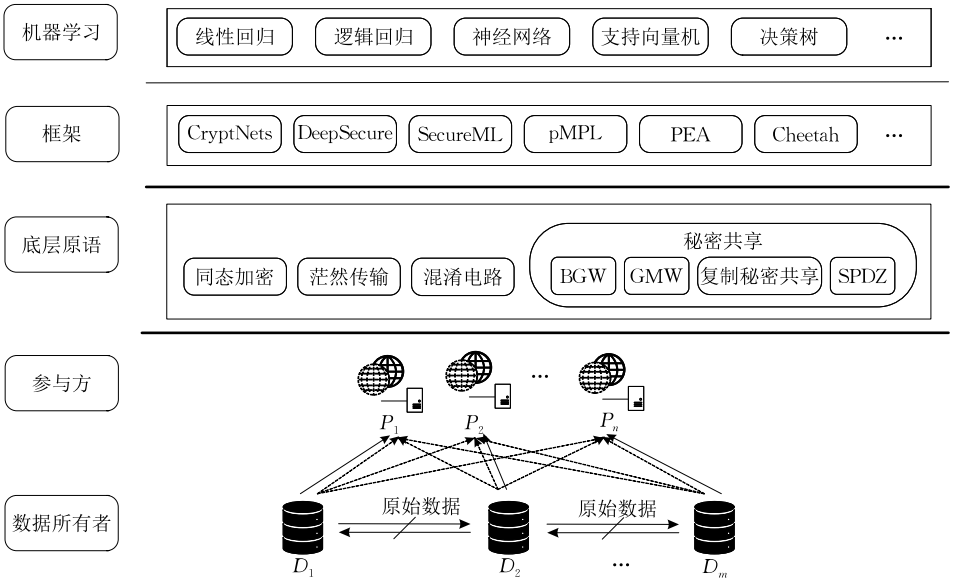

图 1　安全多方学习技术栈

就安全模型而言，安全多方学习的安全模型可以从两个方面来考虑：参与方的可信程度和参与方可以合谋的数量. 根据参与方的可信程度，可以分为两种安全模型，即半诚实模型和恶意模型. 它们之间的区别在于参与方是否遵循协议. 半诚实（也称为诚实但好奇或被动）的参与方严格遵循协议的要求，但试图从所获得的中间结果中获取尽可能多的其他参与方的隐私信息. 恶意（也称为主动）的参与方会任意破坏协议的执行，例如向其他参与方发送不正确的消息. 同时，根据可能合谋的参与方数量，可以分为两种安全模型，即诚实大多数模型和不诚实大多数模型. 他们之间的区别在于可能合谋的参与方数量是否少于总参与方的一半. 结合以上两个方面的安全假设形式，有四种组合. 最强的安全假设是恶意的不诚实大多数模型，最弱的安全假设则是半诚实的诚实大多数模型.

就系统部署而言，安全多方学习中通常包含两种角色：数据所有者（拥有原始数据的角色）和参与方（实际进行安全学习的角色）. 两者可以融合在一起，即数据所有者作为参与方直接进行安全学习；两者也可以分离，即数据所有者通过安全的方式把数据发送给多个参与方，由参与方完成安全学习，也即安全外包计算. 此外，在安全预测的场景中，还存在一个模型所有者，即拥有模型的角色.

就功能场景而言，安全多方学习的功能场景可以根据机器学习的阶段分为两种：安全训练和安全预测. 安全训练即多个参与方互相合作来安全地训练一个机器学习模型. 安全预测即多个参与方使用训练好的模型对数据集进行预测，或者数据所有者将自己的数据加密后发给模型所有者，由模型所有者在加密的数据上进行预测. 安全训练得到的输出是一个机器学习模型，安全预测得到的输出是预测结果. 安全训练需进行正向及反向传播的多轮迭代，因而需要大量的加法、乘法以及比较等运算，而安全预测只需要进行一次正向传播，所需的运算量较少. 所以一个高效的安全多方训练方法和框架从技术实现上来说更具难度. 特别地，在安全外包计算场景中，对于安全训练，一个或者多个（这里可以远远超过参与方数目）数据所有者将加密（通常是秘密共享）后的数据发送给多个参与方，由参与方利用这些加密后的数据安全地训练出一个机器学习模型；对于安全预测，模型所有者和数据所有者分

别将密态(通常是秘密共享或同态加密)的模型和数据发送给多个参与方,由参与方对这些数据在模型上进行预测.

3 底层原语和隐私保护技术

本节介绍安全多方学习中的底层原语以及隐私保护技术,包括同态加密、茫然传输、混淆电路以及秘密共享.

3.1 同态加密

同态加密(Homomorphic Encryption,HE)[19]是一种加密方式,将原始数据进行同态加密后,可以直接对密文执行特定的代数运算,而不必解密密文,也不需要知道有关密钥的任何信息.且对得到的密文计算结果进行解密后得到的结果,与对原始数据执行相同的代数运算得到的结果完全相同.当前,同态加密通常可以分为三种类型[20]:部分同态加密(Partially Homomorphic Encryption,PHE)[21-23]、特定同态加密(Somewhat Homomorphic Encryption,SWHE)[24]和全同态加密(Fully Homomorphic Encryption,FHE)[25-28].部分同态加密只支持一种操作类型(加法或乘法),可执行的次数不受限制.常用的加法同态加密有 Paillier 算法[23],乘法同态加密有 RSA 算法[21].为了对密文既进行加法运算又进行乘法运算,可以使用特定同态加密和全同态加密.特定同态加密可以在有限的次数内执行某些类型(如加法和乘法)的代数运算,而全同态加密可以不受次数限制地执行所有代数运算.但是,全同态加密的计算复杂度通常要远高于部分同态加密和特定同态加密.参与方利用同态加密进行安全多方计算时,通信的轮次通常是恒定的.

3.2 茫然传输

茫然传输(Oblivious Transfer,OT)[29],也叫不经意传输,是一种基础的密码学原语,是安全多方计算的重要构建模块,可以在多个参与方之间安全地传输信息.通过茫然传输协议,发送方从一些待发送的消息中发送一条给接收方,接收方可以获得自己想要的消息,而不能获得发送方的其他消息,同时发送方不知道接收方获得的是哪条消息.本文在此介绍一种基于 RSA 的半诚实 1-out-of-2 茫然传输协议.发送方$\mathcal{S}$拥有两条信息 m_0,m_1,接收方$\mathcal{R}$拥有一个选择比特 $\theta\in\{0,1\}$.在协议结束后,接收方$\mathcal{R}$得到自身选择比特所对应的消息 m_θ,但是得不到 $m_{1-\theta}$的任何信息,发送方$\mathcal{S}$也无法得到关于选择比特 θ 的任何信息.具体步骤如下:(1)发送方$\mathcal{S}$生成两对公私钥(s_{k0},p_{k0})和(s_{k1},p_{k1}),并将两个公钥 p_{k0}和 p_{k1}发送给接收方$\mathcal{R}$;(2)接收方$\mathcal{R}$生成一个随机数 r,根据选择比特 θ 所相应的公钥 $p_{k\theta}$对随机数 r 加密,从而得到 $v=Enc_{p_{k\theta}}(r)$,并将 v 发送给发送方$\mathcal{S}$;(3)发送方$\mathcal{S}$分别用两个私钥对 v 进行解密得到 $k_0=Dec_{s_{k0}}(v)$和 $k_1=Dec_{s_{k1}}(v)$.随后,发送方$\mathcal{S}$将解密得到的结果与自己拥有的两条消息 m_0,m_1进行异或得到 $e_0=m_0\oplus k_0$和 $e_1=m_1\oplus k_1$,并将 e_0和 e_1发送给接收方$\mathcal{R}$;(4)接收方$\mathcal{R}$计算 $m_\theta=e_\theta\oplus r$,即可得到自己的选择比特所对应的信息.

为了提高茫然传输的效率,研究人员提出了一系列茫然传输扩展协议[30-33].Ishai 等人[31]提出了在半诚实安全模型下的茫然传输扩展协议 IKNP,使用 λ 个基础的茫然传输协议来实现 n 个茫然传输,其中 $n\gg\lambda$.KOS 协议[32]对 IKNP 协议进行改进,使其支持恶意安全模型.上述两个协议都是 1-out-of-2 茫然传输协议.之后 Kolesnikov 等人[33]将 IKNP 协议推广至了 1-out-of-k 茫然传输.

3.3 混淆电路

混淆电路(Garbled Circuit,GC)[16,34]是一种密码学协议,使两个参与方(通常分别被称为混淆者和评估者)在不知道对方数据的情况下,安全地计算任何可以表示为布尔电路的函数.典型的混淆电路协议有三个过程:(1)首先,对于电路的每路导线 i,混淆者生成两个随机字符串 k_i^0和 k_i^1作为标签,分别代表导线的两个可能的位值"0"和"1".对于电路中的每个门,混淆者都会创建一个真值表.真值表的每个输出都使用与其输入相对应的两个标签进行加密.之后混淆者将加密后的真值表进行打乱得到乱码表;(2)混淆者将乱码表连同与他的输入位所对应的导线标签一起发送给评估者;(3)评估者通过茫然传输协议安全地获取与其输入相对应的标签.然后评估者利用乱码表和收到的混淆者的输入所对应的标签以及自己的输入所对应的标签,逐门对乱码表进行解密,直到获得函数的最终结果为止.

混淆电路协议的总通信开销与电路的大小成正比.但混淆电路协议交互的轮次与电路的深度及所计算的函数无关.近年来,研究人员提出了几种混淆电路协议的变体,以提高混淆电路协议的性能.Beaver 等人[35]提出的 P&P(Point and Permute)机制减少了加密运算的数量.Naor 等人[36]提出的 GRR3(Garbled Row Reduction 3 Ciphertexts)方法,将乱码表的行数减少为 3 个.Kolesnikov 和 Schneider

提出的 Free XOR 协议[37]使得参与方在本地进行异或运算而无需交互，从而提高了异或门的计算效率. Pinkas 等人[38]提出了 GRR2(Garbled Row Reduction 2 Ciphertexts)方法，将乱码表的行数从四个减少到两个. 然而 GRR2 与 Free XOR 不能兼容. Zahur 等人[39]提出了 Half Gates 技术将密文的数量从四个减少到两个，并兼容了 Free XOR. 虽然混淆电路协议一直被优化，然而，这些协议依旧需要按比特逐位去计算一个函数，这会导致很多操作(尤其是乘法)的通信成本显著增加.

3.4　秘密共享

秘密共享(Secret Sharing，SS)的主要思想是将一个秘密值 x 分成多个份额，每个份额由一方持有，单方无法恢复秘密值，只有若干个参与方一起合作才能恢复秘密值.

常用的秘密共享机制包括加法秘密共享[40]和 Shamir 秘密共享[41]. 加法秘密共享是指份额的总和是秘密值. 因此，可以通过简单地将所有份额相加来恢复秘密值. 在 Shamir 秘密共享中，份额是根据多项式构造的，并且通过拉格朗日插值法求解该多项式来恢复秘密值. 作为一个阈值协议，Shamir 秘密共享确保任意 t 个份额可以恢复秘密值，其中 $t<n$，t 是阈值，n 是参与方的数量. 即如果获得的份额数量少于 t 个，则无法推断出秘密值 s. 典型的基于秘密共享的安全多方计算协议有：GMW(Goldreich-Micali-Wigderson)协议[42]、BGW(Ben-Or-Goldwasser-Wigderson)协议[43]、复制秘密共享协议[44]和 SPDZ 协议[45]，这些协议均适用于多个参与方参与计算的场景.

GMW 协议允许任意数量的参与方安全地计算一个可以表示为布尔电路或算术电路的函数. 在计算之前所有参与方使用加法秘密共享交互地共享他们的输入数据，并逐门进行评估. 电路中的加法(异或)门可以通过在本地将份额相加(异或)完成评估，而无需任何通信. 且各方的本地计算成本可以忽略不计. 至于乘法(与)门，各方需要使用 1-out-of-4 茫然传输或乘法三元组[46]来运行交互计算，以进行安全的评估. 此外，电路中同一层的乘法(与)门可以并行计算. GMW 协议的性能主要取决于电路的深度. 与混淆电路不同，GMW 协议不需要生成真值表，计算只需要加法(异或)和乘法(与)操作，并允许在离线阶段预先计算所有对称密码操作. 因此，GMW 协议通常可以在低延迟网络和低深度电路中取得良好的性能[47].

BGW 协议是一种支持三方及以上的安全多方计算协议. 在计算之前所有参与方使用 Shamir 秘密共享交互地共享他们的输入数据，然后逐门计算结果. 与 GMW 协议类似，对于电路中的加法门，各方本地计算，不需要进行通信，且计算开销可以忽略不计. 而对于乘法门，各方需要交互. 然而，BGW 协议和 GMW 协议的交互方法是不同的，BGW 协议依赖于多项式乘法和降阶操作来完成乘法计算. 由于 Shamir 秘密共享是一个阈值协议，所以 BGW 协议能够很好的解决单点失效问题，但是 BGW 协议只适用于诚实大多数安全模型. BGW 协议可以容忍 $k<n/2$ 个半诚实的参与方合谋或 $k<n/3$ 个恶意的参与方合谋，其中 n 是参与方的数量.

复制秘密共享协议是一种加法秘密共享的变体. 它源于加法秘密共享，且遵循加法秘密共享的性质，但是一个参与方持有多份份额. 以三方为例，一个秘密值 x 利用加法秘密共享生成三个份额 x_1，x_2，x_3，满足 $x=x_1+x_2+x_3$，这些份额以 (x_1,x_2)，(x_2,x_3)，(x_3,x_1)的形式分发给三个参与方，每一个参与方持有其中的一对份额. 由上述份额分配方式可知，任意两个参与方即可恢复出来秘密值 x，则当一个恶意参与方发送错误份额时，其他参与方可以根据收到的份额进行验证，如果不一致说明存在恶意的参与方. 如恶意参与方 P_1 向 P_3 发送错误份额 $(x_1,\tilde{x}_2)$，P_2 向 P_3 发送正确份额 (x_2,x_3). P_3 根据收到来自 P_1 和 P_2 的份额分别计算 $\tilde{x}=x_1+\tilde{x}_2+x_3$ 和 $x=x_1+x_2+x_3$，由于 $\tilde{x}\neq x$，则说明 P_0 和 P_1 中有一个是恶意参与方. Araki 等人[48]提出了一种半诚实安全模型下的三方复制秘密共享协议. 在此协议中，每个参与方分别持有份额对 (a_1,a_3-x)，(a_2,a_1-x)，(a_3,a_2-x)，其中 a_i 是 0 的加法秘密共享份额，即 $0=a_1+a_2+a_3$. 零共享可以使用伪随机数生成来执行，其中每对参与方共享一组随机密钥，在此设置后无需任何通信即可生成零的随机份额. 此外，零的随机份额是 Araki 等人设计的乘法算子协议所需的唯一随机值. 在该乘法算子协议中，每一方只需要发送一个环元素. 因此，该协议的通信量非常低，且每个乘法门只需进行单轮通信，而加法门不需要通信.

SPDZ 协议由 Damgård 等人[45]2012 年提出，用于支持恶意的不诚实大多数安全模型，并支持任意多方进行安全计算. 它是一种可验证的秘密共享机制[49]，分为离线阶段和在线阶段. SPDZ 离线阶段用于生成乘法三元组并执行昂贵的公钥加密计算，而在线阶段只使用廉价的信息-理论上安全的原语.

SPDZ 的主要思想是使用理论安全的消息认证码(MAC)来保护秘密值不被恶意参与方破坏. 具体来说,每个参与方 P_i 都持有份额对 $(x_i,\gamma(x)_i)$,其中 x_i 和 $\gamma(x)_i$ 分别是 x 和 $\gamma=\alpha(x+\eta)$ 的加法秘密共享份额值,η 是公开的常数值. 此外,γ 是全局密钥 α 下认证 x 的 MAC,并且 α 也以加法秘密共享的形式分布在各个参与方手中. SPDZ 协议可以容忍多达 $k<n$ 个恶意参与方,其中 n 是参与方的数量. Damgård 等人[45]提出的 SPDZ 协议执行在域上,之后 Cramer 等人[50]提出了执行在环上的协议 SPDZ_{2^k},支持恶意的不诚实大多数安全模型. 此协议的通信开销高于执行在域上的协议,但是由于可以使用 32 位或 64 位 CPU 进行自动取模来进行计算,而不需要在每一次计算完成后对一个大素数进行取模操作,所以增加的通信量带来的时间开销可以由减少的本地取模操作弥补.

4 安全多方学习框架

如图 2 所示,本节根据安全多方学习所使用的底层原语以及隐私保护技术,将其分成了四种,即基于同态加密的安全多方学习框架,基于混淆电路的安全多方学习框架,基于秘密共享的安全多方学习框架和基于混合协议的安全多方学习框架. 此外,我们分析了 38 个安全多方学习框架的优缺点.

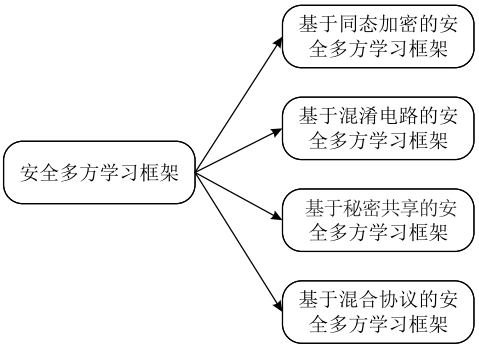

图 2 安全多方学习框架分类

4.1 基于同态加密的安全多方学习框架

通过直接对加密数据进行计算,同态加密可以保证计算过程的安全性. 在进行机器学习前,参与方采用同态加密技术对数据进行加密,后续计算直接在密文上进行,从而保证运算过程中不会泄漏任何原始数据的信息. 如图3所示,以安全两方计算为例,P_0 生成一对公私钥 (s_k,p_k),利用公钥 p_k 对其数据 m_0 加密得到密文 $M_0=Enc(p_k,m_0)$,并将加密数据连同公钥 p_k 发送给 P_1. P_1 利用公钥 p_k 对其数据 m_1 加密得到密文 $M_1=Enc(p_k,m_1)$. 之后 P_1 利用密码系统的同态特性在密文 M_0 和 M_1 上进行计算得到结果的密文形式 $R=F(M_0,M_1)$,其中 F 为所要计算的电路. 计算完成后,P_1 将 R 发送给 P_0,P_0 使用私钥 s_k 对其解密得到最终结果 $r=Dec(s_k,R)$. 在整个过程中,所有的运算均在密文上进行,没有泄漏原始数据的信息.

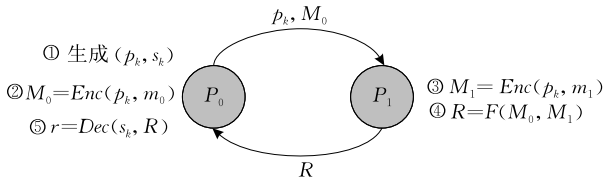

图 3 两个参与方使用同态加密进行安全计算(其中 (s_k,p_k) 为 P_0 生成的同态加密公私钥对,m_0 和 M_0 分别为 P_0 的原始数据和其密文,m_1 和 M_1 分别为 P_1 的原始数据和其密文,F 为所要计算的电路,r 和 R 分别为最终结果及其密文形式)

Vaidya 等人[51]利用同态加密设计了一个决策树安全训练和安全预测框架. 该框架支持任意数量的参与方在半诚实的诚实大多数安全模型下进行安全计算. 此外该框架针对不同的数据分布以及不同的隐私保护等级给出了不同的协议实现. 在 Gilad-Bachrach 等人[52]提出的 CryptNets 框架中,一个数据所有者将数据利用同态加密进行加密后发送给云服务器,由云服务器在这些加密的数据上对神经网络模型进行安全预测. CryptoDL[53]在 CryptoNets 的基础上进行了改进,利用低阶多项式对非线性函数进行了更有效的逼近,并使用平均池化替换最大池化层,提高整体框架的运行效率. Li 等人[54]使用多密钥同态加密设计了云辅助决策树训练框架. 在该框架中,数据所有者将自己的原始数据利用同态加密加密后传输给两个参与方(云存储服务器和云计算服务器)进行计算. 其中这两个参与方被假设为半诚实的. 许心炜等人[55]针对多分类逻辑回归模型提出了一个基于同态加密的安全多方学习框架. 数据所有者将自己持有的数据利用同态加密技术加密后发送给一个服务器,之后此服务器在加密数据上进行训练,并将训练完成的结果返回给数据所有者,数据所有者将训练结果解密即可进行对逻辑回归模型进行多分类预测.

特点. 基于同态加密的安全多方学习框架具有较低的通信开销和常数通信轮次. 然而,这些框架通常具有巨大的计算复杂度和较高的内存占用. 利用同态加密计算线性函数如矩阵乘法等效率较高,但安全计算机器学习中常用的非线性函数如 ReLU、Sigmoid 时,其效率较低. 此外,基于同态加密的安全

多方学习框架容易出现密文爆炸的问题，而机器学习模型的训练，尤其是层数较深的神经网络模型所要求的计算量较大，因此当前基于同态加密的安全多方学习框架不适用于深度神经网络的安全训练，一般只用于机器学习模型的安全预测. 通常基于同态加密的安全多方学习框架用于两方安全预测的场景，其中数据持有者将数据利用同态加密进行加密，并发送给服务器(模型所有者)，服务器在加密的数据上进行模型的预测，预测完成后将加密的预测结果返回给数据所有者进行解密，从而完成机器学习模型的安全预测.

4.2　基于混淆电路的安全多方学习框架

Lindell 和 Pinkas[56]使用混淆电路实现了两方场景下的决策树训练算法 ID3，并使用截断的泰勒级数来近似对数函数. 该实现考虑了半诚实的安全模型. 在训练过程中，参与的双方不会获得除了训练结果以外的任何信息，训练结束后，两个参与方将获得明文的决策树模型. 在此基础上，Lory[57]使用切比雪夫多项式展开改进了对数函数的近似方法，进一步提升了训练的效率. Fairplay[58]是一个应用混淆电路来安全地评估函数的安全两方计算系统，其安全模型为半诚实模型. Fairplay 允许开发人员使用高级语言 SFDL 指定要安全计算的函数，该语言被编译和优化为存储在文件中的布尔电路. FairplayMP[59]将原来的 Fairplay 框架扩展到多方的场景. Mohassel 等人[60]提出了一种基于混淆电路的安全三方计算方法，该方法针对单个恶意参与方提供安全性，即支持恶意的安全大多数安全模型. Gascón 等人[61]研究设计了一个基于混淆电路的安全多方学习框架，用于在分布于任意多方的垂直分割数据上训练线性回归模型. Rouhani 等人[62]提出了基于混淆电路的安全两方深度学习预测框架 DeepSecure，考虑了半诚实的安全模型，并通过在安全计算开始之前对数据和网络结构进行预处理加快在线阶段的运算速度.

特点. 基于混淆电路的安全多方学习框架大多用于两方场景. 这些框架通常具有常数通信轮次，但通信开销与电路的大小成正比，通信量较大. 由于混淆电路是按比特进行计算的，在计算线性操作如矩阵乘法时计算复杂度和通信开销较大，在计算非线性函数如比较时较为高效，因此基于混淆电路的安全多方学习框架通常只适用于简单的机器学习模型训练，如逻辑回归. 对于神经网络等模型，基于混淆电路的安全多方学习框架通常只适用于安全预测.

4.3　基于秘密共享的安全多方学习框架

在基于秘密共享的安全多方学习框架中，各个参与方在机器学习开始前使用秘密共享方案共享他们的数据. 并且在整个机器学习过程中，各方持有的所有数据都是以份额的形式存在.

Choi 等人[63]在半诚实安全模型中为任意数量的参与方实现了 GMW 协议. Hoogh 等人[64]使用 Shamir 秘密共享实现了支持大于等于 3 个参与方的决策树训练及预测框架，其安全模型为半诚实的诚实大多数. 该框架采用的训练算法为 ID3 算法，但是使用了基尼系数替代了信息增益，从而避免了秘密共享下的对数计算. 此外，该框架只支持离散型的特征. Abspoel 等人[65]在此基础上进一步扩展，使用 MP-SPDZ[66]实现了一个 3 方下的支持连续型特征的决策树训练及预测框架，同样考虑了半诚实下的诚实大多数安全模型. 为了减少连续型特征所需要的秘密共享下的比较操作次数，该框架在训练开始前对连续型特征进行排序操作. Damgård 等人[67]应用 $SPDZ_{2^k}$ 并进一步设计了基础的计算协议，如比较、截断、位分解等. 此外他们将在线阶段集成到 FRESCO 框架[68]中，以支持决策树和支持向量机模型的安全预测. Wagh 等人[69]提出了基于秘密共享的 SecureNN 框架，构建了三方的安全协议来支持多种神经网络模型的安全训练和预测. SecureNN 支持半诚实的诚实大多数模型，并且确保隐私不会受到恶意行为的损害，但是不能保证存在恶意行为时结果的正确性. CrypTFlow[70]在 SecureNN 基础之上，利用秘密共享技术对将 TensorFlow[71]代码转换成一种可以三方安全运行的表示形式，以支持对 RseNet50 和 DenseNet121 的安全预测. CrypTFlow 可以支持半诚实/恶意的诚实大多数安全模型，但是恶意安全模型需要利用可信硬件来实现. Wagh 等人[72]基于秘密共享提出了一个高效的安全三方学习框架 Falcon，可以支持多种复杂的神经网络模型安全训练和预测，如 LeNet、AlexNet、VGG-16 等. Falcon 既可以支持半诚实的诚实大多数安全模型，又可以支持恶意的诚实大多数安全模型. Byali 等人[73]设计了一个支持四方的 FLASH 框架以实现多种机器学习模型的安全训练和预测，如线性回归、逻辑回归、神经网络等. 在诚实大多数的设置下，FLASH 框架既可以支持半诚实安全模型，又可以支持恶意安全模型. Koti 等人[74]提出了一个安全多方学习框架 SWIFT，其核心是在诚实大多数设置上的恶意安全的三方计算协议，并将其扩展

到了四方. SWIFT 支持逻辑回归的安全训练和安全预测,但仅支持如 LeNet 和 VGG-16 等神经网络模型的安全预测. Dalskov 等人[75]提供了一种在恶意的诚实大多数安全模型下的四方安全多方学习框架 Fantastic Four,它可以支持逻辑回归和神经网络的安全训练和安全预测. Fantastic Four 在 MP-SPDZ 中提供了实现. SWIFT、FLASH 和 Fantastic Four 均提供一种鲁棒性,即可以保证不管参与方做出怎样的行为,都能输出正确的结果. 但是在 SWIFT 和 FLASH 的设置中,只有一个参与方为恶意方,其他参与方都是诚实的,而在 Fantastic Four 的设置中,一个参与方是恶意的,其他参与方是半诚实的. SWIFT 和 FLASH 通过找到诚实方,并且把后续的计算全部交给诚实方完成来实现鲁棒性,而 Fantastic Four 通过排除恶意方,让其他的参与方继续安全协作训练来实现鲁棒性,因此 Fantastic Four 相较于 SWIFT 和 FLASH 具有更高的安全性. Hamada 等人[76]通过设计新的秘密共享下的数据结构-组,进一步降低了决策树训练框架的通信复杂度. 此框架支持 3 个参与方,且考虑了半诚实的诚实大多数安全模型. Song 等人[77]设计了一个基于向量空间秘密共享的鲁棒性三方学习框架 pMPL,支持线性回归、逻辑回归和 BP 神经网络等常见模型的安全训练. pMPL 保证其中一方可以在机器学习之前被设置为特权方. 即使两个协助方互相勾结,也只有特权方可以获得最终的模型. 同时,pMPL 可以容忍任一协助方在训练期间退出. pMPL 中引入的特权方设置突破了当前安全多方学习中所采用的对等结构不符合常见商业场景这一缺陷.

特点. 基于秘密共享的安全多方学习框架通常具有较低的计算复杂度,其通信量和通信轮次和电路深度成正比. 这些框架在计算线性操作时通常具有较高的效率,但是对于非线性操作,如比较等效率较低. 基于秘密共享的安全多方学习框架既可以用于机器学习模型的安全预测,也可以用于机器学习模型的安全训练. 基于秘密共享的安全多方学习框架通常用于安全外包计算场景,将数据持有方和参与方进行分离,以支持任意数量的数据持有方. 数据持有方将自己的原始数据通过秘密共享协议共享给其他参与方,之后多个参与方在这些共享值上进行机器学习模型的训练或预测. 在基于秘密共享的安全多方学习中,各个参与方基于秘密共享份额的本地计算和明文下的本地计算(如本地的矩阵乘法),计算复杂度一致. 此外,基于秘密共享的安全多方学习根据所使用的秘密共享协议不同,可以适用于多种场景. 如基于复制秘密共享的安全多方学习框架适用于半诚实/恶意的诚实大多数安全模型,基于 SPDZ 的安全多方学习框架适用于恶意的不诚实大多数安全模型.

4.4 基于混合协议的安全多方学习框架

除了前面所述的基于单个协议的安全多方学习框架,还有一些安全多方学习框架,不仅使用一种技术,而是将两种及以上技术结合起来,发挥各自的优点来实现高效的安全多方学习. 例如,将同态加密和混淆电路结合起来的混合协议背后的基本思想是使用同态加密计算可以有效表示为算术电路的操作(例如加法和乘法)和使用混淆电路计算可以有效表示为布尔电路的操作(例如比较). 但是,不同协议之间的转换也比较昂贵.

Mohassel 等人[11]提出的 SecureML 框架,在半诚实安全模型下,在现阶段结合了加法秘密共享和混淆电路等技术. SecureML 支持两方场景下的线性回归、逻辑回归和神经网络等机器学习模型的安全训练和安全预测过程. Liu 等人[78]所提出的安全两方学习框架 MiniONN 结合了同态加密、秘密共享、混淆电路等技术,在半诚实安全模型下,对神经网络模型进行安全预测. MiniONN 利用基于格的加法同态加密为 GMW 协议生成乘法三元组,加快在线阶段乘法的计算. Chameleon[79]是一个结合了秘密共享和混淆电路的混合协议框架,其中利用加法秘密共享执行线性操作,利用 GMW 协议计算电路深度较浅的非线性函数,混淆电路计算电路深度较深的非线性函数. Chameleon 使两个参与方在半诚实安全模型下对支持向量机和神经网络进行安全预测. Juvekar 等人[80]提出了 GAZELLE 框架. GAZELLE 结合同态加密、混淆电路、秘密共享等技术在半诚实安全模型下实现了两方卷积神经网络的安全预测. 他们利用混淆电路计算非线性的函数,并设计了同态加密库,通过使用打包同态加密技术并行化地实现密文上的矩阵加法和矩阵乘法,从而加速了矩阵的运算. Mohassel 等人[12]基于复制秘密共享和混淆电路设计了一个用于三方安全计算的框架 ABY^3. 它支持在算术共享、布尔共享和姚氏共享之间来回切换,并且支持对线性回归、逻辑回归和神经网络等机器学习模型进行安全训练和预测. ABY^3 为半诚实的诚实大多数安全模型协议提供了具体的实现和实验结果,并从理论上描述了如何使他们的协议支持恶意的诚实大多数安全模型. Trident[81]通

过引入一个额外的诚实方将 ABY³ 扩展至四方场景，并且在通信轮次和通信复杂度方面均优于 ABY³. QUOTIENT[82] 利用秘密共享、混淆电路以及茫然传输等安全多方计算技术在两方场景中对多种神经网络模型进行安全训练和预测. 此框架在半诚实的安全模型下为机器学习模型定制了新的安全计算方法，并且实现了残差层的安全训练. Riazi 等人[83] 提出了 XONN 框架，在两方场景中结合了混淆电路和秘密共享对二值化的神经网络进行安全预测，并利用了 XOR 可以在混淆电路协议中免费计算的特点来提高其效率. XONN 既可以支持半诚实的安全模型，又可以支持恶意的安全模型. Astra[84] 是一个支持三方的安全多方学习框架，它结合了混淆电路和秘密共享技术，在半诚实的诚实大多数安全模型和恶意的诚实大多数安全模型下，对线性回归、逻辑回归、支持向量机等机器学习模型进行安全预测. BLAZE[85] 是一个基于 Astra 的安全三方学习框架，最多可以容忍一个恶意的参与方，可以支持半诚实的诚实大多数安全模型和恶意的诚实大多数安全模型. 它利用混淆电路和秘密共享技术对线性回归和逻辑回归模型进行安全训练和安全预测，并支持神经网络的安全预测. Delphi[86] 建立在 GAZELLE 的基础上，同样在半诚实安全模型下对神经网络进行安全预测. Delphi 结合同态加密和加法秘密共享进行线性计算，利用混淆电路实现非线性计算. 此外，Delphi 提出了一种规划器，可以自动生成神经网络架构配置，以在性能和准确率之间进行权衡. CrypTFlow2[87] 针对半诚实安全模型，结合加法秘密共享、茫然传输和同态加密，对 ResNet50 和 DenseNet121 等深度神经网络模型进行安全预测. Tan 等人[88] 提出了一个三方的安全多方学习框架 CRYPTGPU，它基于半诚实的诚实大多数安全模型在 GPU 上实现所有的密文计算操作. CRYPTGPU 构建在 PyTorch[89] 和 CRYPTEN[90] 之上，并支持 LeNet、AlexNet、VGG-16 等神经网络模型的安全训练和预测，以及 ResNet 的安全预测. CAESAR[91] 是一个基于同态加密和秘密共享技术的半诚实安全两方学习框架，它支持数据纵向分割的逻辑回归安全训练和预测. Patra 等人[13] 构造了半诚实安全两方学习框架 ABY2.0. 他们设计了多输入的乘法门，提高了在线阶段的效率，并且支持逻辑回归和神经网络的安全训练和预测. MUSE[92] 针对在半诚实安全模型下基于加法秘密共享的安全预测协议设计了一种新的模型提取攻击，并且设计了一个在两方恶意安全模型下高效的神经网络安全预测协议. Zhang 等人[93] 在 GAZELLE 基础上提出了 GALA，利用同态加密进行线性计算，利用混淆电路完成非线性计算. 他们发现基于同态加密的线性计算主导了总计算时间，因此对基于同态加密的线性计算进行了深度优化，从而大大减少总体计算时间. GALA 在半诚实安全模型下，支持 AlexNet、VGG 和 ResNet 等神经网络的安全两方预测. Huang 等人[14] 提出了 Cheetah，结合了同态加密、加法秘密共享、茫然传输等技术，在两方半诚实的安全模型下进行设计. Cheetah 从两个方面进行了优化，首先通过为神经网络模型中的线性层（卷积、全连接、批量归一化）设计基于同态加密的优化协议，消除同态 rotation 操作，从而加快同态运算的效率. 此外，使用基于 VOLE 类型的茫然传输扩展协议对非线性操作（ReLU、截断）进行优化，极大地降低了深度神经网络模型安全预测的计算和通信开销.

特点. 基于混合协议的安全多方学习框架充分利用了不同协议的优点，如利用混淆电路计算非线性函数，利用同态加密或秘密共享计算线性函数，在计算复杂度和通信开销之间做了权衡，广泛应用于机器学习模型的安全训练和安全预测. 但是不同协议之间转换所带来的开销也是十分昂贵的，因此需要在混合协议带来的性能提升与协议转换的开销之间进行权衡. 常用的混合协议组合有：同态加密＋混淆电路、秘密共享＋混淆电路、同态加密＋混淆电路＋秘密共享等.

4.5　比较与分析

根据 4.4 节的描述，我们从计算复杂度、通信轮次、通信量、线性操作效率、非线性操作效率、支持的功能场景 6 个方面总结了不同类型的安全多方学习框架特点，如表 1 所示.

表 1　不同技术路线的安全多方学习对比

技术路线	计算复杂度	通信轮次	通信量	线性操作效率	非线性操作效率	功能场景
基于同态加密的安全多方学习	较高	常数	较小	较高	较低	通常用于安全预测
基于混淆电路的安全多方学习	较高	常数	较大	较低	较高	通常用于安全预测
基于秘密共享的安全多方学习	较低	较高	适中	较高	较低	既可以用于安全预测 也可以用于安全训练
基于混合协议的安全多方学习	较低	较高	适中	较高	较高	既可以用于安全预测 也可以用于安全训练

此外，如表 2 所示，我们调研了 38 个安全多方学习框架. 并且我们将这些安全多方学习框架根据 4.1 节～4.4 节的四个技术路线进行分类. 我们在表 2 中列出了这些框架的一些信息，包括支持的参与方数量、安全模型、功能场景，支持的机器学习模型，支持的激活函数，所实现的池化方式以及准确率等. 由于大部分框架没有开源，所以我们无法对这些框架进行统一的效率评估. 而且文献中评估这些框架所用的硬件运行环境、数据集、模型结构及功能场景都有所差异，因此这些框架所对应的文献中的实验数据并不能直接用来展现这些框架的效率差异. 我们根据表格中的信息总结如下：

表 2 38 个安全多方学习框架的对比

框架	技术路线	参与方数量	安全模型	功能场景	模型	激活函数	池化方式	准确率(模型，数据集)	年份
文献[51]	基于同态加密的安全多方学习	2^+	SH	安全训练 安全预测	DT	\	\	85.55%(ID3，Car) 71.6%(RDT，Car) 89%(ID3，Nursery) 83.2%(RDT，Nursery) 98.06%(ID3，Mushroom) 89.4%(RDT，Mushroom)	2013
CryptoNets[52]		2^*	\	安全预测	NN	ReLU Sigmoid	Mean	99%(CNN，MNIST)	2016
文献[54]		2	SH，DM	安全训练	DT	\	\	\	2017
CryptoDL[53]		2^*	\	安全预测	NN	ReLU Sigmoid Tanh	Mean	99.52%(CNN，MNIST) 91.5%(CNN，CIFAR-10)	2017
文献[55]		2^*	SH	安全训练	LOR	\	\	\	2019
文献[56]	基于混淆电路的安全多方学习	2	SH，DM	安全训练	DT	\	\	\	2002
文献[57]		2	SH，DM	安全训练	DT	\	\	\	2012
文献[61]		2^+	SH，HM	安全训练 安全预测	LOR	\	\	\	2017
DeepSecure[62]		2	SH，DM	安全预测	NN	ReLU Sigmoid Tanh Softmax	Mean Max	\	2018
文献[64]	基于秘密共享的安全多方学习	3^+	SH，HM	安全训练 安全预测	DT	\	\	\	2012
文献[67]		2^+	Mal，DM	安全预测	DT SVM	\	\	\	2019
SecureNN[69]		3	SH/Mal，HM	安全训练 安全预测	NN	ReLU	Max	93.4%(3 层 DNN，MNIST) 98.77%(4 层 CNN，MNIST) 99.15%(LeNet，MNIST)	2019
CryptFlow[70]		3	SH/Mal，HM	安全预测	NN	ReLU	Max	76.45%(ResNet50，ImageNet)	2019
Falcon[72]		3	SH/Mal，HM	安全训练 安全预测	NN	ReLU	Max	97.42%(3 层 DNN，MNIST) 97.81%(3 层 CNN，MNIST) 98.64%(4 层 CNN，MNIST) 99.15%(LeNet，MNIST)	2020
FLASH[73]		4	SH/Mal，HM	安全训练 安全预测	LR LOR NN	ReLU Sigmoid	\	\	2020
文献[65]		3	SH/Mal，HM	安全训练 安全预测	DT	\	\	\	2020
SWIFT[74]		3，4	Mal，HM	安全训练 安全预测	LOR NN	ReLU Sigmoid	Max	\	2021
Fantastic Four[75]		2～4	SH/Mal，HM	安全训练 安全预测	LOR NN	ReLU Sigmoid Softmax	\	92.3%(1 层 DNN，MNIST) 95.0%(2 层 DNN，MNIST) 92.9%(3 层 DNN，MNIST)	2021
文献[76]		3	SH，HM	安全训练 安全预测	DT	\	\	\	2021
pMPL[77]		2，3	SH，DM	安全训练	LR LOR NN	ReLU Sigmoid	\	96.41%(DNN，MNIST) 86.47%(DNN，Fashion-MNIST) 73.31%(DNN，SVHN)	2022

（续　表）

框架	技术路线	参与方数量	安全模型	功能场景	模型	激活函数	池化方式	准确率(模型，数据集)	年份
SecureML[11]	基于混合协议的安全多方学习	2	SH，DM	安全训练 安全预测	LR LOR NN	ReLU Sigmoid Softmax	\	93.4%（3 层 DNN，MNIST）	2017
MiniONN[78]		2	SH，DM	安全预测	NN	ReLU Sigmoid Tanh SoftMax	Mean Max	99.0%（CNN，MNIST） 81.61%（CNN，CIFAR-10）	2017
Chameleon[79]		2	SH，DM	安全预测	SVM NN	ReLU	Mean Max	99%（CNN，MNIST） 81.61%（CNN，CIFAR-10）	2018
GAZELLE[80]		2	SH，DM	安全预测	NN	ReLU	Max	\	2018
ABY^3[12]		3	SH/Mal，HM	安全训练 安全预测	LR LOR NN	ReLU Sigmoid Softmax	Max	98.3%（3 层 DNN，MNIST） 99%（3 层 CNN，MNIST）	2018
QUOTIENT[82]		2	SH，DM	安全训练 安全预测	NN	ReLU Sigmoid Softmax	Mean Max	99.48%（DNN，MNIST） 95.65%（CNN，MotionSense） 98.3%（DNN，Thyroid） 80.0%（CNN，Breast cancer） 80.5%（ResNet-20，Skin Cancer MNIST）	2019
XONN[83]		2	SH/Mal	安全预测	NN	Softmax	Max	97.6%（DNN，MNIST） 99.0%（CNN，MNIST） 81.85%（CNN，CIFAR-10）	2019
Astra[84]		3	SH/Mal，HM	安全预测	LR LOR SVM	\	\	\	2019
BLAZE[85]		3	SH/Mal，HM	安全训练 安全预测	LR LOR NN	ReLU Sigmoid	\	97.8%（DNN，MNIST）	2020
Delphi[86]		2	SH，DM	安全预测	NN	ReLU	Mean	\	2020
CrypTFlow2[87]		2	SH，DM	安全预测	NN	ReLU	Max	55.86%（SqueezeNet，ImageNet） 76.47%（ResNet50，ImageNet） 74.25%（DenseNet121，ImageNet）	2020
Trident[81]		4	SH/Mal，HM	安全训练 安全预测	LR LOR NN	ReLU Sigmoid Softmax	\	98.3%（CNN，MNIST）	2020
CRYPTGPU[88]		2	SH，DM	安全训练 安全预测	NN	ReLU Softmax	Mean	93.97%（LeNet，Mnist） 59.60%（AlexNet，CIFAR-10） 17.82%（AlexNet，Tiny ImageNet）	2021
CAESAR[91]		2	SH，DM	安全训练 安全预测	LOR	\	\	\	2021
ABY2.0[13]		2	SH，DM	安全训练 安全预测	LR SVM NN	ReLU Sigmoid	\	\	2021
MUSE[92]		2	SH，DM	安全预测	NN	ReLU	Mean	\	2021
GALA[93]		2	SH，DM	安全预测	NN	ReLU Sigmoid Tanh SoftMax	Mean	78.43%（AlexNet，ImageNet） 92.05%（VGG，ImageNet） 93.21%（ResNet-18，ImageNet） 93.86%（ResNet-50，ImageNet） 94.12%（ResNet-101，ImageNet） 94.15%（ResNet-152，ImageNet）	2021
Cheetah[14]		2	SH，DM	安全预测	NN	ReLU	Mean Max	\	2022

注：(1) 支持的参与方数量中带“*”的表示框架中只存在一个数据所有者和一个参与方，数据所有者将加密后的数据发送给参与方，由参与方在加密后的数据上进行计算，并将计算后的结果交给数据所有者进行解密，此处将支持的参与方数量表示为 2；(2) SH、Mal、HM、DM 分别代表半诚实、恶意的、诚实大多数和非诚实大多数安全模型；(3) LR、LOR、NN、DT、SVM 分别代表线性回归、逻辑回归、神经网络、决策树、支持向量机模型；(4) Mean 和 Max 分别代表平均池化和最大池化；(5)“\”代表文献中没有提供相关信息.

(1)大部分安全多方学习框架是基于秘密共享和基于混合协议的.基于秘密共享的安全多方学习框架和基于混合协议的安全多方学习框架在通信开销和计算复杂度之间进行了权衡,通常既可以支持安全训练又可以支持安全预测功能场景.而同态加密技术本身存在计算复杂度高、密钥及密文膨胀、高内存占用、非线性函数计算困难等不足.但是同态加密技术可在两个参与方只交互一次的情况下完成学习任务,因此基于同态加密的安全多方学习框架更适合应用于安全预测的功能场景.此外混淆电路技术虽然通信轮次恒定,但是通常具有巨大的通信开销,适合用于计算非线性的函数,因此更适合与其他技术混合起来使用,同样基于混淆电路的安全多方学习框架更适合应用于安全预测的功能场景.

(2)当前的安全多方学习框架通常使用定制的协议来支持特定数量的参与方,导致可扩展性较差.此外,目前框架所涉及的参与方数量一般为两方或三方,无法做到在计算过程加入新的参与方.

(3)目前大部分安全多方学习框架只能支持诚实大多数/不诚实大多数的半诚实安全模型,以及诚实大多数的恶意安全模型,而不诚实大多数的恶意安全模型只有基于 SPDZ 协议的安全多方学习框架才可以支持.

(4)大部分支持安全训练功能场景的安全多方学习框架所能高效支持的模型较为简单,如线性回归、逻辑回归和网络结构较为简单的神经网络,更为复杂的模型,如 ResNet 等,目前只有安全预测功能场景的安全多方学习框架可以高效支持.

5 讨论与发展建议

本节讨论了安全多方学习与其他隐私保护机器学习技术,如联邦学习,机密计算的区别,并提出了安全多方学习未来的发展建议.

5.1 与其他隐私保护机器学习技术的区别

5.1.1 联邦学习

联邦学习(Federated Learning,FL)[94-98]是一类机器学习框架.它通常需要一个中心服务器,使多个参与方在原始数据不出本地的情况下进行联合机器学习建模.在联邦学习中,每个参与方都使用本地数据在其本地计算环境中训练机器学习模型,并将中间的训练结果(而不是原始数据)上传给中心服务器.如图 4 所示,在一个典型的联邦学习模型中,首先中心服务器初始化一个全局模型,之后迭代执行以下三个步骤:(1)中心服务器将当前全局模型发送给每个参与方或部分参与方;(2)每个参与方使用自己的本地数据训练从中心服务器所接收到的全局模型得到本地模型,并将此本地模型上传到中心服务器;(3)中心服务器通过聚合各参与方上传的本地模型得到一个新的全局模型.

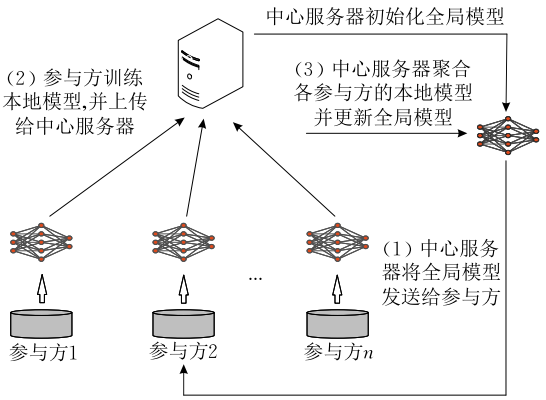

图 4 典型的联邦学习模型

安全多方学习与联邦学习相比,通常具有较高安全性保障,且不需要考虑数据在参与方之间如何分布的,但是其效率较低.

(1)安全性.安全多方学习提供了严格的安全保证,除了最终结果外不泄露任何其他隐私信息;联邦学习虽然保证原始数据不出域,只交换中间梯度信息,但是这些中间梯度信息中同样包含了大量的隐私信息,可能被攻击者利用从而推出原始数据[99-100].当前部分联邦学习框架在交换中间梯度信息时,利用安全多方计算技术(包括同态加密)进行加密,虽然这种方式可以使中心服务器无法窃取明文的中间梯度信息,但是并不能防止参与方获得中间的聚合梯度信息.每个参与方在得到中心服务器聚合后的梯度信息时,通过解密,依旧能够得到明文的聚合后梯度信息.因此在联邦学习中即使利用安全多方计算技术(包括同态加密)对中间梯度进行加密,其安全性依旧不可证明与量化.此外,有些联邦学习框架利用差分隐私技术对中间梯度信息添加噪声来达到保护中间梯度信息的目的,但是添加的随机噪声会影响最终模型的准确率.

(2)数据分布和模型准确率.在安全多方学习框架中,每个参与方都可以在进行机器学习前将自己的数据进行加密(通常利用秘密共享),并将加密后的数据发送给其他参与方,之后每个参与方均拥有全局数据集的一个加密数据集,因此无论原始数据在参与方之间是如何分割的(横向或纵向分布)或

者数据是否独立同分布都不会影响后续的学习.此外,安全多方学习所训练的最终模型在理论上与数据集中训练所得到的机器学习模型准确率一致.然而在联邦学习中,对于非独立同分布的数据和数据纵向分割场景的处理会复杂很多.同时,倘若原始数据是非独立同分布的,通常会对所训练得到的最终模型的准确率产生一定的影响.

(3)效率.在安全多方学习中,参与方交互的方式通常是 Peer-to-Peer 或广播形式的,不需要第三方的参与,且依赖于底层复杂的密码学协议来确保安全性.而在联邦学习中参与方只需要本地训练模型,和第三方交互中间梯度信息,因此联邦学习的计算复杂度和通信开销通常要低于安全多方学习.

5.1.2　机密计算

可信执行环境(Trusted Execution Environment,TEE)通过硬件上的一个独立安全区域,在参与方构建计算平台,保护系统代码和用户数据的机密性和完整性.基于TEE的机密计算(Confidential Computing,CC)[101]可以使得分布式的参与方在可信执行环境中训练模型.参与方部署的数据处理环境的可信性保障了数据隐私和安全性,机密计算的发起者只能获得最终结果,而无法知道参与方的原始敏感数据.然而,基于 TEE 的机密计算需要额外的硬件支持,且用户需要信任此硬件.多个参与方在可信执行环境中训练机器学习模型,其隐私是由数据处理环境的可信赖性来保护的,且参与方通常只能获得最终结果,无法知道原始数据的信息.而安全多方学习利用密码学技术保证机器学习过程的安全,可以不需信任其硬件环境.

5.2　未来发展建议

5.2.1　提高安全性

提高安全多方学习框架的安全性在未来可以从三个方面来考虑:支持安全性保障更强的安全模型,对最终模型进行访问控制,对最终模型的保护.

(1)支持安全性保障更强的安全模型.由 4.5 节的比较分析可知,目前,大多数安全多方学习框架侧重于半诚实或恶意的诚实大多数安全模型.为了保护各种场景下的数据安全,我们应该支持安全性保障更强的安全模型,如恶意的不诚实多数安全模型.利用 SPDZ 协议来支持这一安全模型目前是有效的可行方案.

(2)对最终模型的访问控制.在安全多方学习框架中,模型可以以密文的形式存储在各方中,但是任何 t 个参与方的合作都可以恢复全局模型,其中 t 是阈值.且每个参与方对于获得最终模型的权利是平等的.然而,在现实世界的场景中,通常只有在特权方(即能够为其他方提供的数据付费的参与方)的参与下才能获得最终模型[77].在这种具有特权方的场景下,可以利用空间向量秘密共享[102]技术,通过设计不同参与方所持有的共享值,来实现对最终模型的访问控制.

(3)对最终模型的保护.现有的大多数安全多方学习框架大多数只能保证中间结果的安全,即每个参与方均不能从获得的中间结果推导出其他参与方的原始数据,但是不能保证最终模型不会泄露原始数据的隐私.Ruan 等人[103]提出 PEA 方法,通过在模型训练过程中引入差分隐私技术,添加随机噪声扰动保护最终模型,进一步加强了安全多方学习的安全性保障,实现多方数据共享场景下的训练过程的可证明安全和训练结果的可度量安全.

5.2.2　可证明安全

尽管安全多方学习所使用的底层安全多方计算技术拥有良好的安全性证明机制,但是一个安全多方学习框架包含各种机器学习的步骤,并可能结合多种安全多方计算技术(基于混合协议的安全多方学习框架).因此难以证明安全多方学习整体框架的安全性.安全多方学习的可证明安全可以利用通用可组合(Universally Composable,UC)[104]方案来实现.安全多方学习框架可以从底层模块设计出发,设计满足通用可组合安全定义的子协议或安全组件.通过模块化的组合形成完整的安全多方学习方法,保证整个框架的安全性.

5.2.3　提高性能和效率

提高性能和效率主要可以通过两个方面来实现:减少通信开销和利用 GPU 加速计算.

(1)减少通信开销.在安全多方学习框架中,乘法、比较等计算均涉及到多方之间的通信,其中总的通信开销与通信轮次和单次通信量成正比,因此减少通信开销可以从两个方面进行,即减少通信轮次以及减少单次通信量.其中减少通信轮次可以通过以下两个方式实现:① 将在线阶段的运算转换为本地计算和离线计算,从而减少在线计算的通信轮次;② 设计并行机制,将不相互依赖的算子进行并行计算,从而减少通信轮次.减少单次通信量同样可以通过两个方式去实现:① 优化所设计的底层算子,以减少计算中单次交互所需的通信量;② 针对不同的计算使用不同大小的域或环表示中间数据以减少通信量.

(2)利用 GPU 加速计算. 由于在安全多方学习中涉及大量的矩阵运算,并且在实际的计算过程中,矩阵运算是除了网络通信之外最为耗时的部分. 因此令安全多方学习框架中的算子兼容 GPU 计算,利用 GPU 并行计算的特性来加速计算,可以进一步提高整体框架的效率,为其应用到实际的场景中提供进一步的优化[88,105].

(3)适配安全多方学习底层原语的机器学习模型. 现有的机器学习模型基于的是高性能的服务器或者联网服务器集群设计,大量的计算并不考虑安全多方学习所需要适应的分布式计算场景. Ruan 等人[103]提出 PEA 方法探索了一种利用简化机器学习模型来实现高效模型训练的技术路径,是安全多方学习技术发展的方向之一.

5.2.4 安全多方学习框架的互联互通

由于不同技术平台所使用的安全模型、技术方案和平台设计各有差异,为保证各安全多方学习框架的独立性、完整性和安全性,安全多方学习框架跨平台互联互通应在各平台内部自治的前提下实现,即不暴露平台内部的设计细节且进行任务协同. 为此,首先可以对安全多方学习框架进行模块拆分,如乘法模块,非线性函数计算模块等,对每一个模块可以设计并实现统一标准接口,使得不同安全多方学习框架间通过同一套标准接口实现互联互通.

作者贡献声明 韩伟力、宋鲁杉、阮雯强、林国鹏对本文贡献相同,为共同第一作者.

致 谢 我们向对本文的工作给予支持和宝贵建议的评审老师和同行表示衷心的感谢!

参 考 文 献

[1] Rastegari M, Ordonez V, Redmon J, et al. XNOR-Net: ImageNet classification using binary convolutional neural networks//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 525-542

[2] Povey D, Ghoshal A, Boulianne G, et al. The Kaldi speech recognition toolkit//Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society, 2011: 6465-6469

[3] Gu Y, Tinn R, Cheng H, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH), 2021, 3(1): 1-23

[4] Kou G, Chao X, Peng Y, et al. Machine learning methods for systemic risk analysis in financial sectors. Technological and Economic Development of Economy, 2019, 25(5): 716-742

[5] Fakoor R, Ladhak F, Nazi A, et al. Using deep learning to enhance cancer diagnosis and classification//Proceedings of the International Conference on Machine Learning. New York, USA, 2013, 28: 3937-3949

[6] Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. 1st Edition. New York, USA: Springer International Publishing, 2017

[7] Song Lei, Ma Chun-Guang, Duan Guang-Han. Machine learning security and privacy: A survey. Chinese Journal of Network and Information Security, 2018, 4(8): 1-11 (in Chinese)
(宋蕾, 马春光, 段广晗. 机器学习安全及隐私保护研究进展. 网络与信息安全学报, 2018, 4(8): 1-11)

[8] Tan Zuo-Wen, Zhang Lian-Fu. Survey on privacy preserving techniques for machine learning. Journal of Software, 2020, 31(7): 2127-2156(in Chinese)
(谭作文, 张连福. 机器学习隐私保护研究综述. 软件学报, 2020, 31(7): 2127-2156)

[9] Jiang Han, Liu Yi-Ran, Song Xiang-Fu, et al. Cryptographic approaches for privacy-preserving machine learning. Journal of Electronics & Information Technology, 2020, 42(5): 1068-1078(in Chinese)
(蒋瀚, 刘怡然, 宋祥福等. 隐私保护机器学习的密码学方法. 电子与信息学报, 2020, 42(5): 1068-1078)

[10] Song L, Wu H, Ruan W, et al. SoK: Training machine learning models over multiple sources with privacy preservation. arXiv preprint arXiv:2012.03386, 2020

[11] Mohassel P, Zhang Y. SecureML: A system for scalable privacy-preserving machine learning//Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP). San Jose, USA, 2017: 19-38

[12] Mohassel P, Rindal P. ABY^3: A mixed protocol framework for machine learning//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. Toronto, Canada, 2018: 35-52

[13] Patra A, Schneider T, Suresh A, et al. ABY2.0: Improved mixed-protocol secure two-party computation//Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). 2021: 2165-2182

[14] Huang Z, Lu W, Hong C, et al. Cheetah: Lean and fast secure two-party deep neural network inference. IACR Cryptology ePrint Archive, 2022, 2022: 207

[15] Yao A C. Protocols for secure computations//Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). Chicago, USA, 1982: 160-164

[16] Yao A C C. How to generate and exchange secrets//Proceedings of the 27th Annual Symposium on Foundations of Computer Science (SFCS 1986). Toronto, Canada, 1986: 162-167

[17] Hastings M, Hemenway B, Noble D, et al. SoK: General purpose compilers for secure multi-party computation//Proceedings of the 2019 IEEE Symposium on Security and Privacy (S&P). San Francisco, USA, 2019: 1220-1237

[18] Wei Li-Fei, Chen Cong-Cong, Zhang Lei, et al. Security issues and privacy preserving in machine learning. Journal of Computer Research and Development, 2020, 57(10): 2066-2085(in Chinese)
(魏立斐，陈聪聪，张蕾等. 机器学习的安全问题及隐私保护. 计算机研究与发展，2020，57(10)：2066-2085)

[19] Rivest R L, Adleman L, Dertouzos M L. On data banks and privacy homomorphisms. Foundations of Secure Computation, 1978, 4(11): 169-180

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

[21] Rivest R L, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, 1978, 21(2): 120-126

[22] ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory, 1985, 31(4): 469-472

[23] Paillier P. Public-key cryptosystems based on composite degree residuosity classes//Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques. Prague, Czech Republic, 1999: 223-238

[24] Boneh D, Goh E J, Nissim K. Evaluating 2-DNF formulas on ciphertexts//Proceedings of the Theory of Cryptography Conference. Cambridge, USA, 2005: 325-341

[25] Dijk M, Gentry C, Halevi S, et al. Fully homomorphic encryption over the integers//Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques. Monaco, 2010: 24-43

[26] Gentry C. Fully homomorphic encryption using ideal lattices//Proceedings of the 41st Annual ACM Symposium on Theory of Computing. Bethesda, USA, 2009: 169-178

[27] Smart N P, Vercauteren F. Fully homomorphic encryption with relatively small key and ciphertext sizes//Proceedings of the International Workshop on Public Key Cryptography. Paris, France, 2010: 420-443

[28] Coron J S, Mandal A, Naccache D, et al. Fully homomorphic encryption over the integers with shorter public keys//Proceedings of the Annual Cryptology Conference. Santa Barbara, USA, 2011: 487-504

[29] Rabin M O. How to exchange secrets by oblivious transfer. Harvard Aiken Computation Laboratory: Technical Report TR-81, 1981

[30] Beaver D. Correlated pseudorandomness and the complexity of private computations//Proceedings of the 28th Annual ACM Symposium on Theory of Computing. Philadelphia, USA, 1996: 479-488

[31] Ishai Y, Kilian J, Nissim K, et al. Extending oblivious transfers efficiently//Proceedings of the Annual International Cryptology Conference. Santa Barbara, USA, 2003: 145-161

[32] Keller M, Orsini E, Scholl P. Actively secure OT extension with optimal overhead//Proceedings of the Annual Cryptology Conference. Santa Barbara, USA, 2015: 724-741

[33] Kolesnikov V, Kumaresan R. Improved OT extension for transferring short secrets//Proceedings of the Annual Cryptology Conference. Santa Barbara, USA, 2013: 54-70

[34] Huang Y, Evans D, Katz J, et al. Faster secure two-party computation using garbled circuits//Proceedings of the 20th USENIX Security Symposium (USENIX Security 11). San Francisco, USA, 2011: 331-335

[35] Beaver D, Micali S, Rogaway P. The round complexity of secure protocols//Proceedings of the 22nd Annual ACM Symposium on Theory of Computing. Baltimore, USA, 1990: 503-513

[36] Naor M, Pinkas B, Sumner R. Privacy preserving auctions and mechanism design//Proceedings of the 1st ACM Conference on Electronic Commerce. Denver, USA, 1999: 129-139

[37] Kolesnikov V, Schneider T. Improved garbled circuit: Free XOR gates and applications//Proceedings of the International Colloquium on Automata, Languages, and Programming. Reykjavik, Iceland, 2008: 486-498

[38] Pinkas B, Schneider T, Smart N P, et al. Secure two-party computation is practical//Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security. Tokyo, Japan, 2009: 250-267

[39] Zahur S, Rosulek M, Evans D. Two halves make a whole-reducing data transfer in garbled circuits using half gates//Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques. Sofia, Bulgaria, 2015: 220-250

[40] Bogdanov D, Laur S, Willemson J. Sharemind: A framework for fast privacy-preserving computations//Proceedings of the European Symposium on Research in Computer Security. Málaga, Spain, 2008: 192-206

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

[42] Goldreich O, Micali S, Wigderson A. How to play any mental game, or a completeness theorem for protocols with honest majority//Proceedings of the 19th Annual ACM Symposium on Theory of Computing. New York, USA, 1987: 218-229

[43] Ben-Or M, Goldwasser S, Wigderson A. Completeness theorems for non-cryptographic fault-tolerant distributed computation//Proceedings of the 20th Annual ACM Symposium on Theory of Computing. Chicago, USA, 1988: 1-10

[44] Benaloh J, Leichter J. Generalized secret sharing and monotone functions//Proceedings of the Conference on the Theory and Application of Cryptography. New York, USA, 1988: 27-35

[45] Damgård I, Pastro V, Smart N, et al. Multiparty computation from somewhat homomorphic encryption//Proceedings of the Annual Cryptology Conference. Santa Barbara, USA, 2012: 643-662

[46] Beaver D. Efficient multiparty protocols using circuit randomization//Proceedings of the Annual International Cryptology Conference. Santa Barbara, USA, 1991: 420-432

[47] Schneider T, Zohner M. GMW vs. Yao? Efficient secure two-party computation with low depth circuits//Proceedings of the International Conference on Financial Cryptography and Data Security. Okinawa, Japan, 2013: 275-292

[48] Araki T, Furukawa J, Lindell Y, et al. High-throughput semi-honest secure three-party computation with an honest majority//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna, Austria, 2016: 805-817

[49] Evans D, Kolesnikov V, Rosulek M. A pragmatic introduction to secure multi-party computation. Foundations and Trends in Privacy and Security, 2018, 2(2-3): 70-246

[50] Cramer R, Damgård I, Escudero D, et al. SPDZ$_{2^k}$: Efficient MPC mod 2^k for dishonest majority//Proceedings of the Annual International Cryptology Conference. Santa Barbara, USA, 2018: 769-798

[51] Vaidya J, Shafiq B, Fan W, et al. A random decision tree framework for privacy-preserving data mining. IEEE Transactions on Dependable and Secure Computing, 2013, 11(5): 399-411

[52] Gilad-Bachrach R, Dowlin N, Laine K, et al. CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy//Proceedings of the International Conference on Machine Learning. New York, USA, 2016: 201-210

[53] Hesamifard E, Takabi H, Ghasemi M. CryptoDL: Deep neural networks over encrypted data. arXiv preprint arXiv: 1711.05189, 2017

[54] Li Y, Jiang Z L, Wang X, et al. Privacy-preserving ID3 data mining over encrypted data in outsourced environments with multiple keys//Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). Guangzhou, China, 2017, 1: 548-555

[55] Xu Xin-Wei, Cai Bin, Xiang Hong, Sang Jun. Multinomial Logistic regression model based on homomorphic encryption. Journal of Cryptologic Research, 2020, 7(2): 179-186 (in Chinese)
(许心炜, 蔡斌, 向宏, 桑军. 基于同态加密的多分类 Logistic 回归模型. 密码学报, 2019, 7(2): 179-186)

[56] Lindell Y, Pinkas B. Privacy preserving data mining. Journal of Cryptology, 2002, 15(3): 177-206

[57] Lory P. Enhancing the efficiency in privacy preserving learning of decision trees in partitioned databases//Proceedings of the International Conference on Privacy in Statistical Databases. Palermo, Italy, 2012: 322-335

[58] Malkhi D, Nisan N, Pinkas B, et al. Fairplay—A secure two-party computation system//Proceedings of the USENIX Security Symposium. San Diego, USA, 2004: 287-302

[59] Ben-David A, Nisan N, Pinkas B. FairplayMP: A system for secure multi-party computation//Proceedings of the 15th ACM Conference on Computer and Communications Security. Alexandria, USA, 2008: 257-266

[60] Mohassel P, Rosulek M, Zhang Y. Fast and secure three-party computation: The garbled circuit approach//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, USA, 2015: 591-602

[61] Gascón A, Schoppmann P, Balle B, et al. Privacy-preserving distributed linear regression on high-dimensional data. Privacy Enhancing Technologies, 2017, 2017(4): 345-364

[62] Rouhani B D, Riazi M S, Koushanfar F. DeepSecure: Scalable provably-secure deep learning//Proceedings of the 55th Annual Design Automation Conference. San Francisco, USA, 2018: 1-6

[63] Choi S G, Hwang K W, Katz J, et al. Secure multi-party computation of Boolean circuits with applications to privacy in on-line marketplaces//Proceedings of the Cryptographers' Track at the RSA Conference. San Francisco, USA, 2012: 416-432

[64] Hoogh S, Schoenmakers B, Chen P. Practical secure decision tree learning in a teletreatment application//Proceedings of the International Conference on Financial Cryptography and Data Security. Christ Church, Barbados, 2014: 179-194

[65] Abspoel M, Escudero D, Volgushev N. Secure training of decision trees with continuous attributes. Privacy Enhancing Technologies, 2021(1): 167-187

[66] Keller M. MP-SPDZ: A versatile framework for multi-party computation//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. 2020: 1575-1590

[67] Damgård I, Escudero D, Frederiksen T, et al. New primitives for actively-secure MPC over rings with applications to private machine learning//Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2019: 1102-1120

[68] Institute A. FRESCO—A framework for efficient secure computation. https://github.com/aicis/fresco, 2018

[69] Wagh S, Gupta D, Chandran N. SecureNN: 3-party secure computation for neural network training. Privacy Enhancing Technologies, 2019, 2019(3): 26-49

[70] Kumar N, Rathee M, Chandran N, et al. CrypTFlow: Secure TensorFlow inference//Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2020: 336-353

[71] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016

[72] Wagh S, Tople S, Benhamouda F, et al. Falcon: Honest-majority maliciously secure framework for private deep learning. Privacy Enhancing Technologies, 2021, 2021(1): 188-208

[73] Byali M, Chaudhari H, Patra A, et al. FLASH: Fast and robust framework for privacy-preserving machine learning. Privacy Enhancing Technologies, 2020, 2020(2): 459-480

[74] Koti N, Pancholi M, Patra A, et al. SWIFT: Super-fast and robust privacy-preserving machine learning//Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). 2021: 2651-2668

[75] Dalskov A, Escudero D, Keller M. Fantastic four: Honest-majority four-party secure computation with malicious security //Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). 2021: 2183-2200

[76] Hamada K, Ikarashi D, Kikuchi R, et al. Efficient decision tree training with new data structure for secure multi-party computation. arXiv preprint arXiv:2112.12906, 2021

[77] Song L, Wang J, Wang Z, et al. pMPL: A robust multi-party learning framework with a privileged party//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security(CCS'22). New York, USA, 2022: 2689-2703

[78] Liu J, Juuti M, Lu Y, et al. Oblivious neural network predictions via MiniONN transformations//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas, USA, 2017: 619-631

[79] Riazi M S, Weinert C, Tkachenko O, et al. Chameleon: A hybrid secure computation framework for machine learning applications//Proceedings of the 2018 on Asia Conference on Computer and Communications Security. Incheon, Korea, 2018: 707-721

[80] Juvekar C, Vaikuntanathan V, Chandrakasan A. GAZELLE: A low latency framework for secure neural network inference//Proceedings of the 27th USENIX Security Symposium(USENIX Security 18). Baltimore, USA, 2018: 1651-1669

[81] Chaudhari H, Rachuri R, Suresh A. Trident: Efficient 4PC framework for privacy preserving machine learning//Proceedings of the 27th Annual Network and Distributed System Security Symposium. San Diego, USA, 2020

[82] Agrawal N, Shamsabadi A S, Kusner M J, et al. QUOTIENT: Two-party secure neural network training and prediction//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. London, UK, 2019: 1231-1247

[83] Riazi M S, Samragh M, Chen H, et al. XONN: XNOR-based oblivious deep neural network inference//Proceedings of the 28th USENIX Security Symposium(USENIX Security 19). Santa Clara, USA, 2019: 1501-1518

[84] Chaudhari H, Choudhury A, Patra A, et al. ASTRA: High throughput 3PC over rings with application to secure prediction //Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop. London, UK, 2019: 81-92

[85] Patra A, Suresh A. BLAZE: Blazing fast privacy-preserving machine learning//Proceedings of the 27th Annual Network and Distributed System Security Symposium. San Diego, USA, 2020

[86] Mishra P, Lehmkuhl R, Srinivasan A, et al. Delphi: A cryptographic inference service for neural networks//Proceedings of the 29th USENIX Security Symposium(USENIX Security 20). 2020: 2505-2522

[87] Rathee D, Rathee M, Kumar N, et al. CrypTFlow2: Practical 2-party secure inference//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. USA, 2020: 325-342

[88] Tan S, Knott B, Tian Y, et al. CryptGPU: Fast privacy-preserving machine learning on the GPU//Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2021: 1021-1038

[89] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 8026-8037

[90] Knott B, Venkataraman S, Hannun A, et al. CrypTen: Secure multi-party computation meets machine learning//Proceedings of the Advances in Neural Information Processing Systems. 2021, 34: 4961-4973

[91] Chen C, Zhou J, Wang L, et al. When homomorphic encryption marries secret sharing: Secure large-scale sparse logistic regression and applications in risk control//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Singapore, 2021: 2652-2662

[92] Lehmkuhl R, Mishra P, Srinivasan A, et al. MUSE: Secure inference resilient to malicious clients//Proceedings of the 30th USENIX Security Symposium(USENIX Security 21). 2021: 2201-2218

[93] Zhang Q, Xin C, Wu H. GALA: Greedy ComputAtion for linear algebra in privacy-preserved neural networks//Proceedings of the 28th Annual Network and Distributed System Security Symposium. 2021

[94] Konečný J, McMahan H B, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527, 2016

[95] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492, 2016

[96] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data//Proceedings of the Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2017: 1273-1282

[97] Liang P P, Liu T, Ziyin L, et al. Think locally, act globally: Federated learning with local and global representations. arXiv preprint arXiv:2001.01523, 2020

[98] Xu G, Li H, Liu S, et al. VerifyNet: Secure and verifiable federated learning. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926

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

[100] Zhu L, Liu Z, Han S. Deep leakage from gradients//Proceedings of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 14774-14784

[101] Ohrimenko O, Schuster F, Fournet C, et al. Oblivious multi-party machine learning on trusted processors//Proceedings of the 25th USENIX Security Symposium (USENIX Security 16). Austin, USA, 2016: 619-636

[102] Brickell E F. Some ideal secret sharing schemes//Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques. Houthalen, Belgium, 1989: 468-475

[103] Ruan W, Xu M, Fang W, et al. Private, efficient, and accurate: Protecting models trained by multi-party learning with differential privacy//Proceedings of the 44th IEEE Symposium on Security and Privacy. San Francisco, USA, 2023: 22-26

[104] Canetti R. Universally composable security: A new paradigm for cryptographic protocols//Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science. Las Vegas, USA, 2001: 136-145

[105] Watson J L, Wagh S, Popa R A. Piranha: A GPU platform for secure computation//Proceedings of the 31st USENIX Security Symposium. Boston, USA, 2022: 827-844

HAN Wei-Li, Ph. D., professor. His research interests include data security and privacy computing.

SONG Lu-Shan, Ph. D. candidate. Her research interests include data security and privacy computing.

RUAN Wen-Qiang, Ph. D. candidate. His research interests include data security, secure multi-party learning and differential privacy.

LIN Guo-Peng, Ph. D. candidate. His research interests include data security and privacy computing.

WANG Zhe-Xuan, M. S. candidate. His research interests include data security and privacy computing.

Background

Privacy-preserving machine learning based on secure multi-party computation (Secure Multi-party Learning, MPL for short), which allows multiple parties to jointly perform machine learning over their private data while protecting the privacy of their raw data, has been a hot spot in recent. MPL breaks the barriers that different organizations or companies cannot directly share their private raw data mainly due to released privacy protection regulations and laws (e. g. GDPR). Therefore, MPL can be applied to several practical fields involving private data, such as risk control in the financial field and medical diagnosis.

Researchers have proposed a doze of MPL frameworks recently. When we consider the rapid development of MPL, however, a comprehensive and systematic survey, which covers the underlying privacy technologies and classification of MPL frameworks, is still absent so far. Thus, we are motivated to conduct such a literature review of the technical routes, and frameworks of MPL. Our study would help researchers choose suitable MPL frameworks for various scenarios, further identify research gaps, and strengthen the weaknesses of the approaches.

This paper is supported in part by the National Natural Science Foundation of China (No. 62172100), and the project "Science and Technology Innovation Action Plan" of the Science and Technology Commission of Shanghai Municipality (No. 21DZ1201400).