面向隐私保护的非聚合式数据共享综述

李尤慧子[1]，殷昱煜[1]，高洪皓[2,3]，金一[4]，王新珩[5]

（1. 杭州电子科技大学计算机学院，浙江 杭州 310018；2. 上海大学计算机工程与科学学院，上海 200444；
3. 韩国嘉泉大学计算机工程系，城南市 461701；4. 北京交通大学计算机与信息技术学院，北京 100044；
5. 西交利物浦大学电气与电子工程系，江苏 苏州 215123）

摘 要：海量数据价值虽高但与用户隐私关联也十分密切，以高效安全地共享多方数据且避免隐私泄露为目标，介绍了非聚合式数据共享领域的研究发展。首先，简述安全多方计算及其相关技术，包括同态加密、不经意传输、秘密共享等；其次，分析联邦学习架构，从源数据节点和通信传输优化方面探讨现有研究；最后，整理对比面向隐私保护的非聚合式数据共享框架，为后续研究方案构建和运行提供支撑。此外，总结提出非聚合式数据共享领域的挑战和潜在的研究方向，如复杂多参与方场景、优化开销平衡、相关安全隐患等。

关键词：隐私保护；数据共享；联邦学习；安全多方计算

中图分类号：TN92

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021120

Survey on privacy protection in non-aggregated data sharing

LI Youhuizi[1], YIN Yuyu[1], GAO Honghao[2,3], JIN Yi4, WANG Xinheng[5]

1. School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China
2. School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
3. Department of Computer Engineering, Gachon University, Seongnam 461701, South Korea
4. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
5. Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Abstract: Although there is a great value hidden in the massive data, it can also easily expose user privacy. Aiming at efficiently and securely sharing data from multiple parties and avoiding leakage of user private information, the development of related research and technologies on the non-aggregated data sharing field was introduced. Firstly, secure multi-party computing and its technologies were briefly described, including homomorphic encryption, oblivious transfer, secret sharing, etc. Secondly, the federated learning architecture was analyzed from the aspects of source data nodes and transmission optimization. Finally, the existing non-aggregated data sharing frameworks were listed and compared. In addition, the challenges and future potential research directions were summarized, such as complex multi-party scenarios, the balance between optimization and cost, as well as related security risks.

Keywords: privacy protection, data sharing, federated learning, secure multi-party computation

1 引言

当今世界处在信息时代，并正快速进入全面的数字世界。随着 5G 的广泛应用，物联网爆发出了蓬勃的生命力，移动与物联网终端发展情况如图 1 所示[1]。海量数据隐藏着重要的价值，这也是近年来人工智能、深度学习等领域飞速发展的主要因素之一，然而，数据一旦非法泄露，会造成巨大的损失。2020 年中国网络安全报告[2]显示，病毒样本总量为 1.48 亿个，较 2019 年同期上涨 43.71%；超两亿条用户信息被

收稿日期：2020−12−16；**修回日期**：2021−03−10

基金项目：国家重点研发计划基金资助项目（No.2020YFB2103805）；国家自然科学基金资助项目（No.61802093, No.61972358）

Foundation Items: The National Key Research and Development Program of China (No.2020YFB2103805), The National Natural Science Foundation of China (No.61802093, No.61972358)

出售，造成数千万经济损失。国外数据隐私问题也十分严峻，2019 年英国航空公司因违反用户隐私条例被信息监管局罚款近 2 亿英镑（约合 15.8 亿元人民币）。各国为了推动数据隐私保护，颁布了一系列法律条文，如欧盟的 GDPR（General Data Protection Regulation）[3]、美国的 CCPA（California Consumer Privacy Act）以及我国的《中华人民共和国网络安全法》。由此可见，数据隐私保护十分重要。

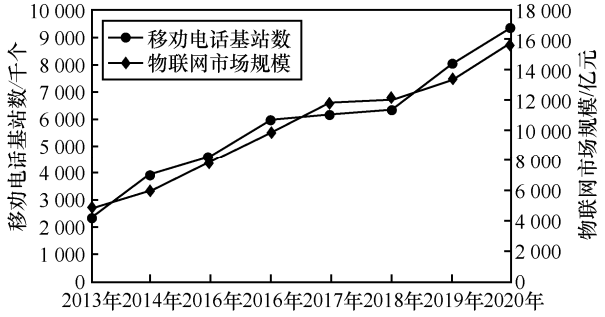


图 1 移动与物联网终端发展情况

加密是保护数据隐私的主要手段。在聚合式数据共享方法中，各数据生产者使用加密算法编码源数据，然后传输至数据处理中心；数据处理中心通过解密算法获取数据信息，聚合所有源数据进行数据挖掘等复杂的操作。加密保证了数据传输时的隐私安全，但无法确保数据处理中心的安全，如果数据处理中心被攻破，则会造成全部数据的泄露。相对地，非聚合式数据共享方法旨在不汇集所有源数据的情况下，同样达到数据共享要完成的最终目标。非聚合式数据共享主要包含两层意思：首先，数据不汇聚，避免了单点故障（中心点被攻破）造成的潜在隐私泄露危险；其次，数据共享不是指狭义的源数据分享，而是从广义角度来看，期望达到数据共享的最终目标，例如数据处理和数据挖掘。最优场景是在不分享源数据的情况下完成处理和挖掘操作，进一步避免数据隐私泄露的可能。

从非聚合式架构来看，1982 年提出的安全多方计算（SMC, secure multi-party computation）[4]是早期主要的非聚合式数据共享方法。安全多方计算继承了分布式的特点，计算参与方地位平等且互不信任，无中心节点。利用加密算法，如同态加密等，数据接收方只能处理加密后的数据，无法获知源数据信息。处理后的数据再传送回数据发送方，发送方经过解密后获得计算结果。安全多方计算利用密码学和底层数据交互协议，保证计算参与方在不获取源数据的情况下，完成数据处理操作，增强了数据的隐私性。早期，由于安全多方计算复杂度高，通常采用哈希映射[5]等方法进行传输，但其安全性不足。其他参与方可以通过枚举操作来确定对方数据集中存在的元素，从而获取对方的隐私信息。随着边缘计算[6]等新型计算架构的发展以及设备计算能力的增强，安全多方计算的实际需求和部署能力也相应提升，相关技术（如不经意传输拓展协议、秘密分享、隐私求交等）为数据隐私保护提供了可靠的指导方案和技术基础。

近年来，云计算、物联网和机器学习高速发展，广泛应用于智慧城市、智慧安防等领域。海量异构数据聚合在云平台上，通过各类机器学习和深度学习算法对数据进行分析，挖掘数据隐藏的知识信息。然而，随着数据和用户隐私的关联度越发密切，把全部源数据聚合到中心节点再进行模型训练的方法隐私泄露风险就越大。因此，谷歌于 2017 年提出联邦学习，旨在不需要通过中心化的数据训练就能获得机器学习模型。各数据提供方在本地进行数据训练，把参数等隐私无关信息发给参数服务器进行全局调优，再把优化后的模型应用到本地，以实现在不提供原始数据的情况下，获取全局的数据“知识”。联邦学习是典型的非聚合式数据共享方法，源数据不出本地，降低了隐私泄露的风险，同时完成了数据共享的目标，即获得优化训练模型。

数据的重要性日益增加，为了更好地保护数据隐私安全，本文针对非聚合式数据共享方法进行介绍和分析，主要综述了安全多方计算和联邦学习的相关研究。安全多方计算侧重于数据传输和计算外包，本文从原理、算法复杂度、适用场景等情况比较分析各类交互协议，增强数据交互的安全性。联邦学习侧重于数据挖掘，旨在从多个数据孤岛上全局分析数据潜在价值，其框架主要包括本地数据训练、参数信息传输交互、参数服务器全局调优。本文从与数据隐私关联性较强的数据源（本地数据）和通信传输两方面讨论对比现有的联邦学习优化方法。此外，本文总结整理了现有面向隐私保护的非聚合式数据共享框架，同时从复杂多参与方场景、优化开销平衡等方面提出了非聚合式数据共享隐私保护潜在的 4 个研究方向及建议。

2 安全多方计算

安全多方计算是密码学领域用于多个数据方

在无可信第三方的情况下，安全且保护隐私地协同计算某个或某些约定函数的方法。SMC 由 Yao[4]于 1982 年提出，用来解决著名的百万富翁问题，即 2 个百万富翁比较谁更加富有，而不能泄露具体的财富值。

安全多方计算广泛应用于敏感数据协同计算的场景。例如，根据个人的信用记录、购买记录、社交喜好等协同挖掘个性化推荐服务；广告转化率[7]收益计算，即利用属于第三方平台的广告点击数据和商品平台的购买数据分析观看特定广告的用户中有多大比例购买了该商品等。

安全多方计算利用数据交互协议，保证计算参与方在不知道源数据的情况下，完成数据处理操作，增强了数据的隐私性。不同的数据交互协议特点各异，适用场景也不同，主要的 MPC 数据交互协议如下。

2.1　同态加密

在基于加密的数据交互协议中，最常用的是同态加密技术，由 Rivest、Adleman 和 Dertouzos 提出[8]。该技术不仅支持加密数据传统的传输、存储操作，还支持用户直接对加密数据进行计算，其结果等价于原始数据计算后再加密。

同态性质是针对加密函数来说的，一般分为加（减）法同态、乘（除）法同态（也称单同态）和全同态。其简单定义如下。设存在映射 $g:G_1\to G_2$，则有

加法同态：$g(x+y)=g(x)+g(y),\ x,y\in G_1$

乘法同态：$g(x+y)=g(x)\times g(y),x,y\in G_1$

全同态：同时满足上述 2 个性质

若一个加密函数满足加法或乘法同态，则其支持在加密数据上做加法或乘法计算而不损害数据，若同时满足加法、乘法同态，即全同态，则几乎可以支持任何计算操作。广泛使用的 RSA 算法[9]满足乘法同态，Paillier 算法[10]满足加法同态。

同态加密常常应用在云计算环境中，保证数据传输过程和云中心节点计算时的数据隐私安全。例如，在云外包的场景中，Abadi 等[11]实现了 2 个协议，通过加法同态加密等操作实现安全对抗半诚实对手，其中 EO-PSI（efficient outsourced private set intersection）具有较好的大数据集拓展性。而文献[12-13]采用加法同态加密技术实现隐私集合交集（PSI, private set intersection）协议和 PSI-CA（private set intersection cardinality）协议。

针对两方数据集大小差异大的情况，Resende 等[14]使用同态加密和布谷鸟过滤器优化[15]中提出的协议，在半诚实对手模型中实现了单向隐私求交协议；文献[16-18]则利用全同态加密解决数据集差异大的问题。

同态加密的优点是可以在不泄露源数据的情况下，得到同样加密的计算结果，但是，合适的加密函数定义难，而且其最大的局限性在于复杂度过高，普遍应用还需要进一步的研究和发展。

2.2　逻辑电路和不经意传输

安全多方计算主要包括计算和数据传输 2 个方面。基于逻辑电路方案被认为是通用的计算设计方法，因为任何函数都可以转化成对应的逻辑电路，借助对电路真值表的替换、加密和打乱，形成 Garbled Table。参与方间的数据传输可以通过不经意传输技术交换必要的消息，最后某一方计算出最终的结果。图 2 展示了以两方为例的情景。首先，由 Alice 根据需求构造电路，确定真值表内容。为门中的每条线秘密地生成 2 个密钥 X_i^0，X_i^1，分别对应输入 0、1，替换该门对应真值表中的值。用前两项加密第三项，然后随机打乱得到 Garbled Table。Bob 通过不经意传输从 Alice 那里获得其输入对应的密钥、Garble Table 和自己输入对应的密钥等必要信息，最后计算该门对应的输出 C。在这一过程中，Bob 不用泄露自己的输入，也不知道 Alice 的任何信息。

基于逻辑电路的计算方法有一定可行性，文献[19]在智能手机上成功配置了基于电路的协议，并使用了 Wi-Fi 通信。但该方法需要的逻辑门较多，实现复杂度较高。例如，计算编辑距离需要 30 000 个门电路。为了降低复杂度，Pinkas 等[20]于 2019 年提出了具有线性渐进通信复杂度的基于电路的协议，在集合大小为 2^{20} 时，该协议的通信复杂度不到文献[21]协议的十一分之一，但后者可以拓展到多方环境。

不经意传输技术保证接收者只能从发送者的多个数据中获取自己想要的数据，而发送者却不知道这个具体的数据。不经意传输的构造方法有许多，如基于 RSA（Rivest, Shamir, Adleman）构造[22]、基于椭圆曲线[23]等。近年来，高效的 SMC 数据交互协议大部分是基于不经意传输拓展的[24-25]，其性能较好，能够仅通过少数公钥操作和位操作完成大量的数据传输。例如，文献[26]实现了在普通带宽环境中（30～100Mbit/s）最快的 SMC 协议，在半

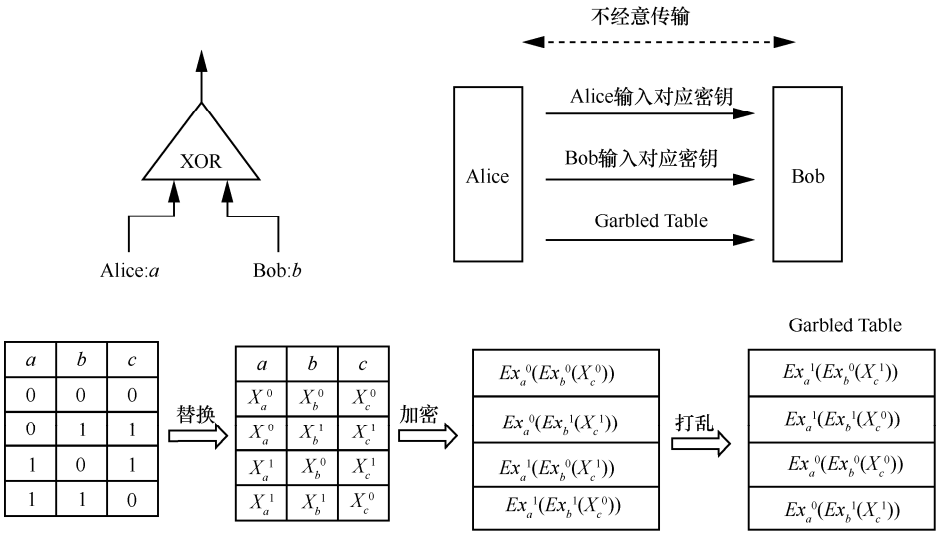


图 2　混淆电路示意

诚实对手环境和恶意对手环境中都能保证一定的安全性。它的底层协议是基于不经意拓展的轻量级多点不经意伪随机函数。由于在某些环境下，协议计算的结果可能只是一个中间结果，因此，Ciampi 等[27]基于不经意传输拓展实现能够输出“加密”的协议，类似的工作还有文献[28-35]等。此外，文献[36]利用差分隐私技术，选择性地泄露部分不重要的信息，相比于文献[32]提出的对抗恶意对手的协议，性能提升近 63%。

2.3　秘密共享

秘密共享（secret-sharing）也是 SMC 领域经常使用的技术，由 Shamir[37]提出。秘密共享的关键是保证数据共享过程使用的信道是安全的。传统方法采用的是加密技术，然而，密钥的分享也面临同样的问题，即如何保证密钥分享信道安全。非对称加密技术的发展在很大程度上解决了该问题，但它的安全性也依赖于信道的诚实性，存在被中间人攻击的可能。为了解决这一问题，秘密共享技术将秘密分成多份，由不同的信道传输，就算有恶意信道存在，也不能轻易还原完整的秘密。其形式化定义为

$$S(s,t,n)\to\{(s_0),(s_1),\cdots,(s_n)\}$$

其中，s 表示需要分享的秘密，t 表示需要还原秘密的阈值，n 表示将秘密拆分的数目。在一组参与者中，秘密由大家共同保管，只有当超过阈值的参与者相互合作时才能恢复完整的秘密数据。

秘密共享技术在安全多方计算场景中应用广泛[38-39]，数据方将数据和模型用秘密共享的方法加密拆分，发送至各参与方。在解决分布式集群一致性问题的拜占庭协议[40]中，也采用秘密共享协议将秘密分成多份传送给各个假设的将军，并在需要的时候恢复秘密。

2.4　隐私集合交集协议

隐私集合交集协议是 MPC 领域使用较为广泛的应用层协议，旨在多个数据方间计算数据交集。其具体定义如下。

发送方 S 和接收方 R 分别持有一个私有数据集 X 和 Y，协议的目标是计算它们各自隐私集合的交集，即以 X 和 Y 为输入，计算 $f(X,Y)=X\bigcap Y$，而不能泄露除交集之外的数据信息。

PSI 作为具体的应用层协议，它底层的数据传输和计算功能可以由同态加密、不经意传输、秘密分享等技术实现。此外，PSI 还可以根据具体的应用要求拓展出各种各样的变体，例如，文献[41]受可交换加密技术启发，采用布隆过滤器优化，接收者可以更轻松地计算输出，确保只有一方得知交集内容；只公布交集大小的应用场景[42-44]，其中文献[42]借助不经意评估方法实现恶意环境安全，结果表明其在云环境时的开销比半诚实协议高 25 倍，而且允许动态调节通信和计算之间的开销比；隐藏一方输入大小的应用场景[45-46]，文献[12]提出了一种适用于云计算环境的新的集合表示方法，基于加法同态公钥密码（PKC，private key cryptography）实现了 2 个实用的协议，即 PSI 和 PSI-CA。

主要的 SMC 数据交互协议比较如表 1 所示。

3　联邦学习

随着云计算、人工智能、大数据等技术的发展，智能手机、可穿戴设备等产生了超大规模的数据，深度学习也随着数据量的增长而得到更好的发展。由于设备的计算能力不断增强，端设备数据与用户隐私关联度较高，非聚合式数据共享方法更受青睐，训练数据保留在本地的联邦学习算法得到广泛研究。

联邦学习是一种关注隐私保护的机器学习技术[47]，源数据不离开本地设备，在多参与方或多计算节点之间开展高效的机器学习。联邦学习是一个统称，各数据提供方构成联邦，共同训练模型，其可使用的机器学习算法不局限于神经网络，还包括随机森林等。其框架主要由 3 个部分构成：提供数据的多个本地节点、负责参数调优的中心参数服务器、参数传输链路（传输本地数据训练后的模型参数至服务器，传输全局调优后的参数至各本地节点）。根据具体实现的深度学习模型和算法的不同，联邦学习框架可以适用于各类数据分散的学习场景中。

本文侧重于从源数据节点和通信传输 2 个方面分析联邦学习的优化方法及其相关的数据隐私保护技术。

3.1　源数据节点

联邦学习中各源数据节点在本地进行训练，仅上传参数信息，避免数据隐私信息泄露。针对源数据节点的工作主要可以分为数据获取和非平衡数据优化 2 个方面。

3.1.1　数据获取

联邦学习拥有一定规模的本地数据节点提供源数据，但本地节点的管理较为松散，是自主构建联邦模式，可以任意加入和离开。此外，本地节点的状态是不可控的，如在线情况、诚实度、参与度、贡献程度等。如何选择适合的本地数据节点、如何激励优质的节点参与并提供高质量的数据都是数据获取阶段需要考虑的问题。

1) 参加者选择

本地数据节点的质量会影响总体的训练过程，联邦学习可通过选择部分高质量参与者，提升算法效率。例如，FedAvg[48]从符合要求的本地节点（如

表 1　　**主要的 SMC 数据交互协议比较**

类型	文献及协议		应用场景	安全	支持多方	输出	运行时间/s（输入元素个数为 2^{16}）	输入上限 k（输入元素个数为 2^k）	备注
同态加密	文献[11]	O-PSI	云外包	半诚实	√	交集	—	20	—
		EO-PSI					415.8		
	文献[12]	PSI	云外包	半诚实	×	交集大小	1 670	20	—
	文献[18]	PSI	本地	半诚实	√	交集	32.78	19	—
逻辑电路	文献[20]	Basic	本地	半诚实	×	交集	13.767	20	LAN
		Advanced					9.763		
	文献[21]	Separate	本地	半诚实	×	交集	11.251	20	LAN
		Combined					9.076		1 Thread
不经意传输	文献[26]	PSI	本地	半诚实 恶意	×	交集	0.63	24	LAN
	文献[29]	Spot-low	本地	半诚实	×	交集	13.7	24	100 Mbit/s
		Spot-fast		恶意			2.91		1 Thread
	文献[30]	PSI	本地	半诚实	×	交集	0.702	24	LAN 1 Thread
	文献[31]	PSI	本地	恶意	×	交集	9.7	20	LAN
	文献[32]	EC-ROM	本地	恶意	×	交集	0.94	24	LAN
		DE-ROM					1.3		1 Thread

注：O-PSI 在输入为 2^{15} 时，运行时间达到 57 357 s，时间过大，故未标出。

在线且空闲的设备）中选取一定量的客户端进行聚合；Goetz 等[49]提出主动联合学习框架 AFL（active federated learning），每个通信回合不是随机均匀地选择客户，而是以模型和客户数据为条件进行概率选择；FedCS[50]根据资源的状况主动管理本地节点，使其产生更多的聚合更新。

2) 激励机制

通常的研究工作都假设设备无条件贡献资源，但实际上设备是自私的，很难组建联邦。因此，部分学者研究如何使本地数据节点积极参与到联邦系统中，并制定高效的激励机制[51-53]。首先，参与的节点之间可能会存在敌对竞争关系，双方正面积极参与并均获利十分关键。其次，参与者的贡献程度不同，产生的效益也不一致，如何合理公平地分配收益也是值得研究的问题。此外，部分工作将激励机制与防御恶意参与者结合，例如，Fmore[54]是移动边缘场景的激励机制框架，实验证明提出的激励机制能提高学习算法的性能。

3.1.2 非独立同分布及不平衡数据的优化

数据分布及其质量对模型训练有重要影响，尤其是在非聚合数据共享方法中。2017 年谷歌在 FedSGD 算法的基础上提出了 FedAvg 算法[48]，该算法对非独立和不平衡数据都具有稳健性，为后续联邦学习优化奠定了基础。

1) 数据及模型架构优化

首先，从数据入手，部分解决方案利用共享数据子集[55-56]方法，在保护用户隐私的前提下，提高模型准确率。其次，从模型本身架构入手，使之更加适应联邦学习，例如，基于随机梯度下降（SGD, stochastic gradient descent）的优化算法。从本地数据节点角度考虑，优化方法包括：对数据节点的要素进行插值更新[57]、对节点的目标函数添加一个近端项（proximal term）[58]、局部梯度添加可控的噪声干扰并控制变量[59]、矫正本地数据的用户漂移[60]等。

2) 模型个性化

模型个性化的联邦学习也解决了数据非独立同分布影响模型精度的问题。传统联邦学习下，中心参数服务器提供统一的本地模型，但在某些联合学习场景下，参与方希望训练的模型对自身利益最大化，而不是为了达成全局模型的共识而进行妥协。此外，多方数据的非独立同分布特性可能会降低全局模型的精度，导致用户参与联邦学习所得的模型精度不理想[61]。模型个性化指参与方可自定义所训练的本地联邦学习模型，在利用本地数据集训练本地私有模型的基础上，参与全局公共模型的训练，优化私有模型。

现有的模型个性化算法大多基于知识蒸馏实现，以完成不同模型间交换知识达成共识。表 2 展示了支持模型个性化的联邦学习框架对比分析，梳理了框架的异同点。

Liu 等[62]提出了一种允许客户独立定制模型和设计训练的联邦学习框架——联邦相互学习（FML, federated meta learning），其实现方式如图 3 所示。首先，训练 3 个本地模型（私有模型和中间模型）；其次，通过深度相互学习与中间模型交换知识（不同于师生模型，模型间不存在强弱关系，交换的方向也是双向的）；随后，通过合并中间模型获取全局模型；最后，借助中间模型向私有模型传达联邦学习的知识。该方法可以让本地数据节点完全独立的设计不同于全局模型的私有模型，但同时会导致全局模型不再能开箱即用。

Ramanan 等[63]使用迁移学习和知识蒸馏开发一个通用的支持本地节点自主设计模型的联邦学习框架。其假定参与方拥有较少的标准数据以及足量的公共数据集，借此联合学习一个分类任务，并提出 FedMD 算法：每个模型首先在公共数据方面进行充分的训练；然后，借助迁移学习，对自己少量的私有数据进行训练；随后，借助知识蒸馏将私有模型的知识转化为一个统一形式；最后，中央参数服务器收集这些转化后的知识，计算出共识。相比于传统联邦学习，不存在所谓的全局模型，只存在一个知识蒸馏结果的共识，也不支持新的参与方，因为新的参与方可能会破坏现有的模型。

表 2 模型个性化下的联邦学习框架对比

模型个性化程度	文献	对个性化模型的保护	全局模型直接可用	接入设备动态改变	局限性	模型个性化方法
完全自主	文献[62]	支持	不支持	自适应	对持有更多样本数据的参与者不公平	知识蒸馏
模型体系结构由中央服务器控制	文献[63]	无	支持	自适应	模型个性化程度不高	模型不可知元学习
完全自主	文献[64]	无	不支持	不支持	需要公共数据集	知识蒸馏

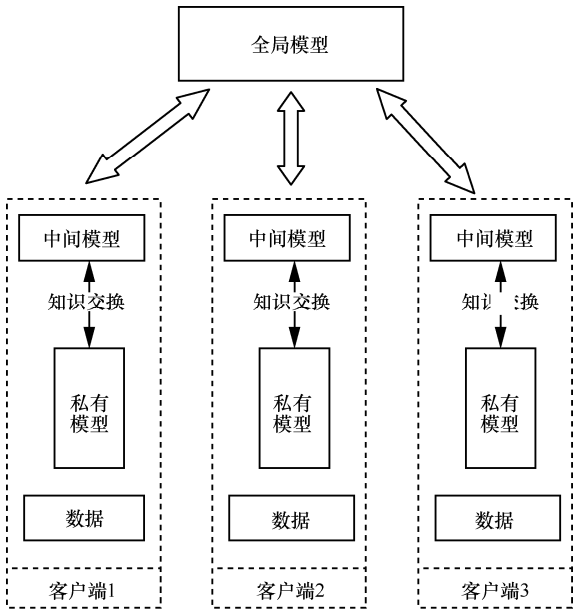


图 3　FML 实现方式

Roy 等[64]论证了 FedAvg 也是一种元学习，并提出一种平均后悔上限分析框架，使联邦学习与元学习构建联系，让模型个性化成为可能。

表 3 对现有的联邦学习数据源优化策略进行了对比总结。

3.2　通信传输优化

联邦学习中有 2 个必要的通信传输过程：数据节点传输本地数据训练后的模型参数至服务器，中心参数服务器传输全局调优后的参数/模型至各本地节点。在双向通信过程中，本地数据节点潜在的不可控网络状态（网络时延、通信费用等）容易导致通信成为联邦学习的瓶颈。

3.2.1　模型压缩与降低通信频次

针对通信这一任务本身，可以从降低通信频次[48-49, 65-68]和通信量方面进行优化。例如，AdaComm 自适应通信策略[69]从低通信频率求平均值开始，以节省通信时延并提高收敛速度；增加通信频率以实现较低的错误率，实验证明其总体花费时间可减少到 1/4。

从降低通信量考虑，可以进行通信数据压缩，包括上行压缩和下行压缩。因为上行的速率总是低于下行的速率，所以更多的研究关注上行压缩[70-73]，部分研究同时压缩上行和下行[74-75]。压缩可分为有损压缩和无损压缩，相较于无损压缩，有损压缩能大大提升压缩率，达到很好的压缩效果，但是有一定损耗。在保证收敛率和准确率的前提下，更多的研究偏向于有损压缩。在表 4 中，本文对各个压缩算法是否为有损压缩、是否有下行压缩、对非独立同分布（Non-IID, non independent and identically distributed）数据是否有稳健性以及压缩倍率等进行了分析和比较。

3.2.2　改变分布式体系结构

改变分布式体系结构也是优化计算节点间通信的有效方式之一。分布式体系结构包括对等、循环、服务器−客户端等。单中央服务器−多客户端的结构存在单点故障、主干网络开销过大等问题，可以利用对等架构解决。例如，对等网络结构下的联邦学习，其计算节点在有限通信的条件下，可以通过与相邻节点联合学习的方式解决通信瓶颈问题。相关研究主要集中在达成共识、协调构造全局模型、无中心模型聚合等方面。本节将去中心化的联邦学习框架根据其训练网络的结构划分为网型拓扑、树型拓扑、抽象总线拓扑，并总结讨论其模型聚合的方式及特点。

表 3　　**联邦学习数据源优化策略**

解决方法	文献	核心方法/特点	局限性
参与者选择	文献[49]	根据值函数选择一个优化（有用程度较高）的数据子集	客户不断收集（并可能获取）数据，并且在许多情况下，分发可能是不稳定的
	文献[61]	主动管理客户端，FedCS 为客户端设置了在 FL（federated learning）协议中下载、更新和上载 ML 模型的最后期限	在动态的场景中工作平均资源量以及更新和上载所需的时间会动态波动
	文献[76]	只有极少数客户端（例如，少于 1%）允许其数据上传到 FL 服务器。通过考虑每个客户端的数据分布和通道条件来计划数据上载客户端和模型上载客户端	考虑客户的能源消耗
	文献[77]	采用 deepQ 学习算法，该算法允许服务器学习和找到最佳决策	未考虑 Non-IID 数据
奖励机制	文献[62]	契约理论	确保本地模型更新的可靠性
	文献[63]	将声誉与合同理论相结合的有效激励机制	只考虑了一个联邦服务器
	文献[64]	斯塔克伯格博弈的平衡解	服务器获利较低，只考虑了一个联邦服务器

表 4　不同联邦学习压缩算法对比分析

名称	有损压缩	特点	下行压缩	对 Non-IID 数据有稳健性	压缩的倍数/通信时间缩小倍数（选取最好情况）	测试数据集
深度梯度压缩[70]	×	动量修正、局部梯度剪裁、动量因子映射	×	√	压缩了 270～600 倍	CIFAR10、ImageNet、Penn Treebank、AN4、Librispeech
量化 SGD[71]	√	量化、黑盒压缩	×	×	训练减少了 2.5～6.8 倍	ImageNet、CIFAR10、MNIST、CMU AN4、AlexNet、VGG、ResNet
PowerSGD[72]	√	功率迭代、快速	×	×	压缩了 24～128 倍	CIFAR10
SignSGD[73]	√	top-k 梯度稀疏化，实现下游压缩以及权重更新的分层和最佳 Golomb 编码	√	√	压缩了 1 050 倍	CIFARA、KWSA、F-MNIST
稀疏三进制压缩[75]	√	多数投票的方式	√	×	速度提高 25%	ImageNet、RESNET50

Savazzi 等[76]为降低主干网络和中央服务器的通信开销，避免边缘设备长距离通信而导致训练时延，提出基于云边协同的联邦学习框架，其网络拓扑结构由传统的星形拓扑结构转化为树形结构，子树节点的层层聚合减少了主干网络的数据传输量。边缘云辅助的联邦学习架构如图 4 所示。边缘云聚合边缘网络下的本地节点的模型权重，中央云聚合边缘云的权重，其聚合函数是专门设置的 HierFAVG 函数。此方案仍存在中央服务器，所有边缘云聚合后才能进行云聚合，无法异步更新，降低了每轮聚合收敛的速度。

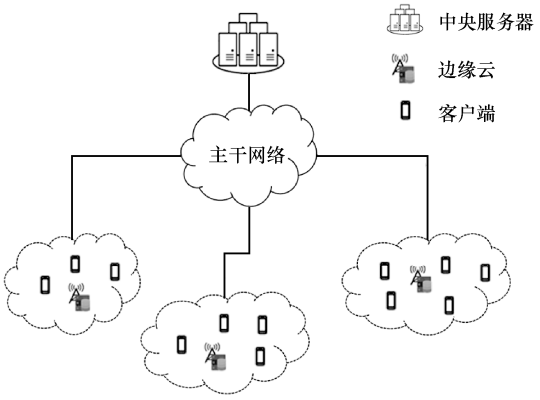


图 4　边缘云辅助的联邦学习架构

Sattler 等[77]提出与区块链结合的去中心化联邦学习框架。利用抽象总线拓扑结构，各参与方构建一个区块链，基于区块链存取全局模型。区块链智能合约帮助训练网络的计算节点达成共识不借助中央服务器就完成全局模型聚合，具体包括：将全局模型权重向量划分为数个块数据，基于智能合约达成全局模型副本的共识，每个参与者投标训练区块上的块。该框架还有一个显著优势，就是借助区块链的记账特性，在区块链上推动计算，用户可以独立评估自己的成本效益比，并决定他们希望更新的块数，量化多方参与计算的绩效。但是，区块链块的数据大小限制会影响全局模型的划分，从而影响该方案性能。

Yoshida 等[51]提出了对等网络的去中心化联邦学习框架，解决参与方分担中央服务器聚合模型工作的问题。首先，随机选择一个节点开始广播，获取其他节点的权重、样本数和版本号，仅记录版本号大于自己的权重；其次，对权重进行加权平均，更新权重并增加版本号。此方案利用参与方充当某一轮的中央服务器角色，适合相互信任但又无法找出共识可靠第三方的场景。但是，该方案对节点的可靠性要求高，因为节点不仅承担一个轮次的权重更新工作，还会广播获取其他节点的样本数量和权重值，若存在恶意参与方，或者没有可靠的安全协议支撑，就会造成数据不可用甚至出现模型安全问题。

Kang 等[78]提出了基于支持设备到设备（D2D，device-to-device）通信的 5G 网络进行去中心化联邦学习的框架及去中心化的模型聚合算法 CFA-GE（consensus based federated averaging with gradients exchange）。利用 5G 网络环境，该方法更适合大规模密集和完全分散的网络，与以 Mcmanhan 等[79]所提方法为代表的集中式深度学习方法形成鲜明对比。

表 5 从算法的特点、隐私保护情况、适用场景等方面对部分联邦学习优化算法进行了对比分析。

4　面向隐私保护的非聚合式数据共享框架

本节主要介绍支持隐私保护的非聚合式数据共

享框架，包括安全多方计算平台 JUGO 及编译器 Tasty、百度共享数据框架 PaddleFL、微众银行 FATE（federated AI technology enabler）、谷歌的 TFF（TensorFlow federated）以及英伟达 Clara。表 6 对这些框架进行了总结对比，分析了其各自特性及适用的场景。

4.1 安全多方计算平台

2018 年计算架构服务提供商矩阵元发布了通用的安全多方计算平台 JUGO，帮助用户快速开发通用半诚实的两方安全计算算法。JUGO 架构如图 5 所示，算法模块主要集成了混淆电路、同态加密等底层协议，供 MPC-SDK 模块调用。当参与方间协同商定计算逻辑后，借助矩阵元开发的 Frutta 高级编程语言在 MPC-IDE 集成开发环境上编写实现，为上层的应用提供安全多方计算服务。最后，电路编译器把电路逻辑编译成电路文件。这些操作都可在 GPU、FPGA 等硬件加速下实现，使协同计算过程更快地完成。

表 5　　联邦学习的优化算法对比分析

算法—框架名称	Baseline	特点	通信情况	隐私保护	适用场景
FedAvg[48]	FeedSGD	迭代平均	与同步随机梯度下降相比，通信效率提高 10~100 倍	DP、MPC	非平衡数据 non-IID 数据
DSVGR[57]	SVRG DANE	稀疏数据 (结构)	—	—	不平衡数据 可用节点少
Parallel Restarted SGD[65]	Parallel mini-batch SGD parallel SGD	并行、重启	与 parallel mini-batch SGD 相比减少了 $O(T^{1/4})$的通信交流次数	—	非凸优化
PASGD[69]	Fully synchronous SGD	周期平均	VGG-16 中，ADACOMIS 比全同步 SGD 快 3.5 倍；ResNet-50 中，ADACOMM 比完全同步 SGD 快 2 倍	—	—
AFL[49]	Active Learning	主动、选择性抽样迭代	减少 20%～70%的迭代	有	—
AFL [61]	FedAvg	分布、公平	—	—	集中式模型
Fmore[54]	—	激励机制，迁移学习	减少近 51.3%的训练轮数，真实系统，时间减少了 38.4%	—	移动边缘计算
FedBCD[66]	—	基于块坐标梯度下降	可以显著降低通信成本	有	共享样本
MMD[67]	单流	双流模型代替单一模型	非 IID 数据分布所需的通信轮数减少了 20%以上	—	non-IID 数据
Mini-batch SGD[68]	SGDZ	小型批量	通信回合的数量最多可以减少 $T/2$（其中 T 表示总步数）	—	高网络时延 带宽受限

表 6　　面向隐私保护的非聚合式数据共享框架

框架	支持类型	安全协议	主要支持场景	软件	Kubernetes	开发机构
JUGO	安全两方计算	同态加密 混淆电路	商业应用	不开源	不支持	矩阵元
Tasty	安全多方计算	PSI 同态加密 混淆电路	学术研究	开源	不支持	—
PaddleFL	横向联邦学习 纵向联邦学习	PSI 秘密分享 同态加密	学术研究	开源	仅 Data Parallel 支持	百度
FATE	横向联邦学习 纵向联邦学习 联邦迁移学习	PSI 秘密分享 同态加密	商用数据共享	开源	支持	微众银行
TTF	横向联邦学习	DP GAN	学术研究	开源	不支持	谷歌
Clara	横向联邦学习	DP	商业应用	不开源	支持	英伟达

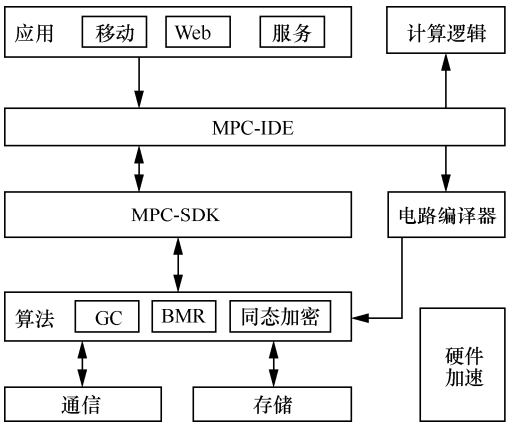


图 5　矩阵元 JUGO 架构

此外，编译器 Tasty[80]能自动生成高效的基于同态加密和混淆电路技术的组合协议，用户可以使用高级语言方便快速地描述该协议。Tasty 用于许多隐私保护协议，如 PSI 相关应用、人脸识别等。图 6 展示了其主要工作流程，包括：1) 分析阶段，运行时环境首先检查协议描述是否存在语法错误，协议两方是否在执行同一个协议，通过分析该协议自动确定哪方可以进行预计算；2) 设置阶段，可预计算的参与方提前计算协议中独立于它们输入的部分，如混淆电路的生成和发送等；3) 在线/执行阶段，各参与方提供自己的输入，协议的在线部分（加密、解密、电路评估等）开始执行，直到计算出各参与方相应的输出为止。

4.2　百度数据共享框架

PaddleFL 是百度于 2019 年开源的联邦学习框架，主要提供 2 种联邦学习解决方案：Data Parallel 和 PFM（parallel federated learning with MPC）[81]。

4.2.1　数据并行化

各数据方通过 Data Parallel 可以基于经典的横向联邦学习算法 FedAvg[81-82]、DPSGD[83-84]等完成模型训练。其服务架构如图 7 所示，分为编译阶段和运行阶段。

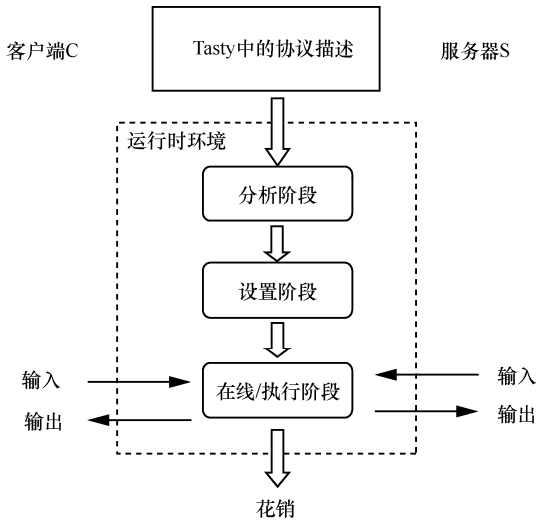


图 6　Tasty 工作流程

编译阶段包含以下 4 个主要部分。1) FL-Strategy。用户可以使用 FL-Strategy 定义联邦学习策略，如 Fed-Avg。2) UDP（user-defined-program），为 PaddleFL 程序，定义机器学习模型结构和训练策略，如多任务学习。3) Distributed-Config。联邦学习系统会部署在分布式环境中，需要对分布式训练进行配置并定义分布式节点信息。4) FL-Job-Generator。给定 FL-Strategy、UDP 和 Distributed Training Config，生成联邦参数的服务器端和客户端的 FL-Job。

运行阶段包含以下 3 个组件。1) FL-Server，在云或第三方集群中运行的联邦参数服务器。2) FL-Trainer，参与联邦学习的每个组织都将有一个或多个与参数服务器通信的客户端。3) FL-Scheduler，训练过程中调度客户端，在每个更新周期前，决定哪些客户端可以参与训练。

4.2.2　基于安全多方计算的联邦学习

作为 PaddleFL 的一个重要组成部分，PFM 是

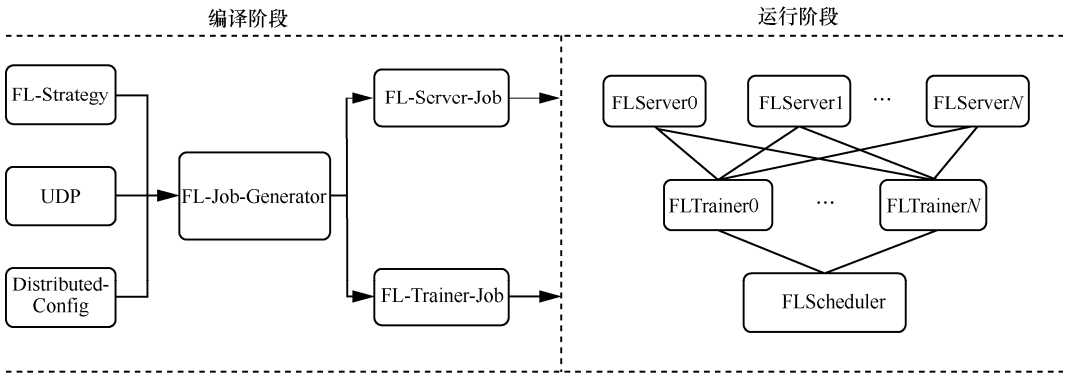


图 7　数据并行化下的横向联邦学习服务架构

基于多方安全计算实现的联邦学习方案。PFM 可以很好地支持横向、纵向及联邦迁移学习等多个场景，既可提供可靠的安全性，也具有可观的性能。

PFM 中的安全训练和推理任务是基于高效的多方计算协议，如三方安全计算协议 ABY3（three-party arithmetic-binary-Yao）[38]。在 ABY3 中，参与方可分为输入方、计算方和结果方。输入方为训练数据及模型的持有方，负责加密数据和模型，并将其发送到计算方。计算方为训练的执行方，基于特定的多方安全计算协议完成训练任务，只能得到加密后的数据及模型。计算结束后，结果方会拿到计算结果并恢复出明文数据。每个参与方可充当多个角色，如一个数据拥有方也可以作为计算方参与训练。PFM 的整个训练及推理过程如图 8 所示。其主要由 3 个阶段组成：数据准备、安全训练/推理、结果解析。数据准备阶段包括私有数据对齐和数据加密及分发。首先，PFM 通过 PSI 协议允许数据拥有方在不泄露自己数据的情况下，找出多方共有的样本集合。此功能主要支持纵向联邦学习，因为其要求多个数据方在训练前进行数据对齐，同时保护用户的数据隐私。其次，数据方将数据和模型用秘密共享的方法加密，然后用直接传输或者数据库存储的方式传到计算方。每个计算方只会拿到数据的一部分，因此计算方无法还原真实数据。

安全训练/推理阶段。PFM 拥有与 PaddleFL 相同的运行模式。在训练前，用户需要定义 SMC 协议、训练模型以及训练策略。PaddleFL 的多方安全计算模块提供了可以操作加密数据的算子，在运行时算子的实例会被创建并被执行器依次运行。

结果解析阶段。安全训练和推理工作完成后，模型（或预测结果）将由计算方以加密形式输出。结果方可以收集加密的结果，使用 PFM 中的工具对其进行解密，并将明文结果传递给用户。

4.3　微众银行数据共享框架

4.3.1　概述

FATE 是由微众银行开源的一款工业级的联邦学习框架，为用户提供保护隐私的分布式机器学习服务[85]。FATE 涵盖联邦特征工程、联邦机器学习模型训练（FATE FederatedML）、联邦模型评估、联邦在线推理等。其中，FATE FederatedML 是联邦学习算法功能组件，提供许多常见机器学习算法联邦化实现。其主要功能如图 9 所示，具体如下：联邦样本对齐，包括纵向样本 ID 对齐、基于 RSA+哈希等对齐方式；联邦特征工程，包括联邦采样、联邦特征分箱、联邦特征选择、联邦相关性、联邦统计等；联邦机器学习，包括联邦逻辑回归、线性回归、泊松回归、联邦 SecureBost、联邦 DN、联邦迁移学习等[86]；多方安全计算协议，包括同态加密、秘密分享、RSA、Diffie-Hellman 交换算法等。

4.3.2　无损隐私保护系统

FATE 在实现纵向联邦学习时，为联邦学习环境下的模型训练提供一种称为 SecureBoost 的无损隐私保护 tree-boosting 系统。如图 10 所示，SecureBoost 在隐私约束下对数据进行对齐，协同学习共享 gradient-tree boosting 模型[87]，同时对多个私有方的所有训练数据保密。FATE 利用基于隐私保护的样本 ID 匹配进行数据对齐。当数据垂直划分在多个参与方上时，不同的参与方持有不同但部分重叠的用户，这些用户可以通过其唯一的用户 ID 来识别。为了在没有仲裁的情况下兼顾隐私并找到各方的共享用户或数据样本用户集的非共享部分，FATE 使用文献[88-89]所提的隐私保护协议，在加密方案下寻找数据样本的用户交集。

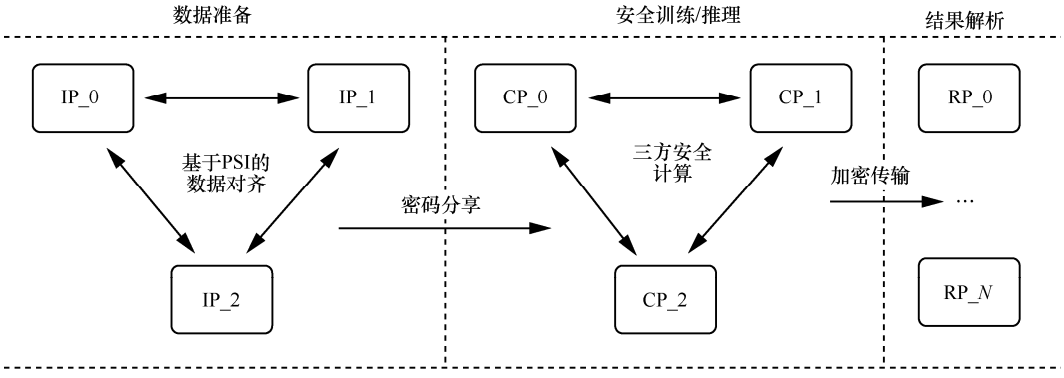


图 8　PFM 训练及推理过程（IP_i、CP_i 以及 RP_i 分别表示数据或模型的拥有方、计算方以及结果获取方）

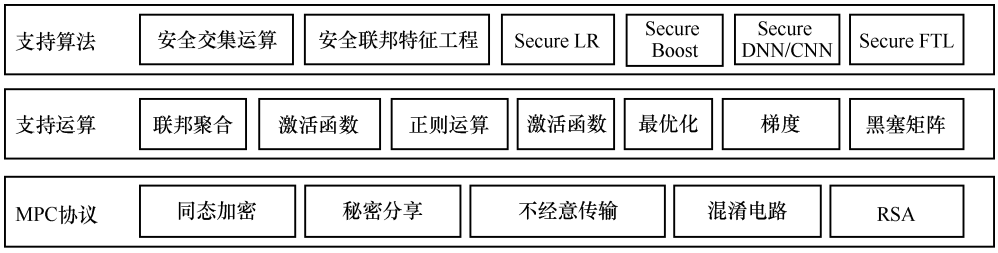


图 9　FATE FederatedML 主要功能

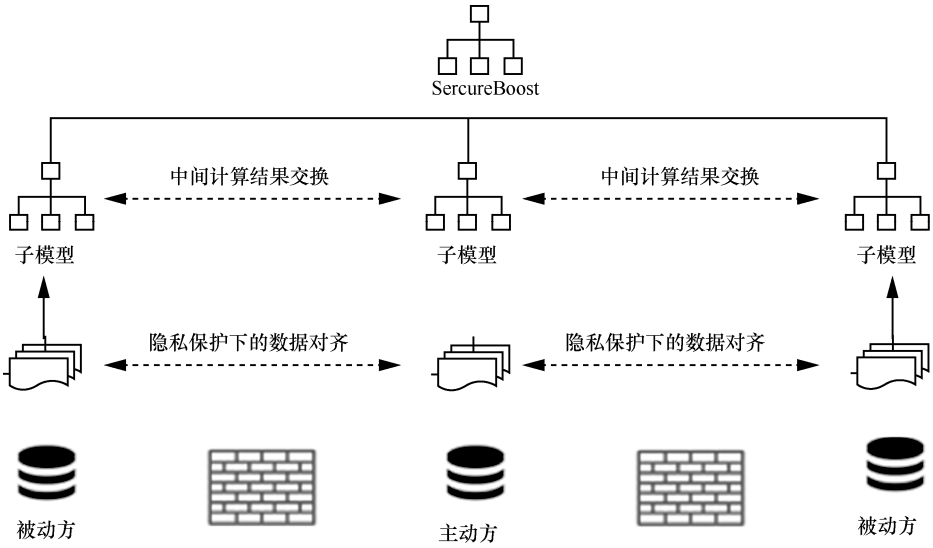


图 10　SecureBoost 框架

图 11 描述了在纵向联邦学习时，隐私保护约束下的数据对齐流程，具体如下。

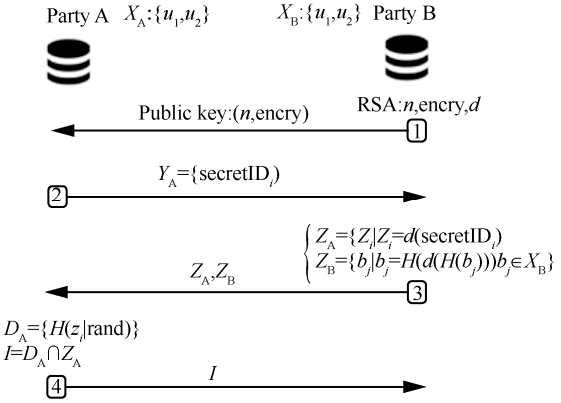


图 11　FATE 纵向联邦学习的数据对齐实现流程

1) B 将公钥（n, encry）加密后传给 A，建立加密通道。其中，n 是公有密钥，encry 是加密算法。

2) A 通过哈希函数 H 逐个映射 u_i，并乘以加密后的随机噪声，然后将结果 Y_A 回传给 B。

3) B 解密后得到 Z_A，计算 Z_B，并将 Z_A 和 Z_B 回传给 A。

4) A 消除 Z_A 中的随机噪声，然后进行一次哈希运算生成 D_A，再求 Z_A 与 D_A 的交集，最后 A 回传交集结果给 B。

在多方安全训练阶段，FATE 利用 SecureBoost 的模型，实质上是将梯度提升树学习算法 XGBoost[90] 进行转换，使其适应联邦学习环境。分裂节点的选择和叶的最优权重计算仅取决于叶的 g 和 h。其中，g 和 h 分别是 XGBoost 损失函数的一阶导数和二阶导数。而 g、h 与分类标签存在关联，攻击方在一定条件下可以通过 g 和 h 恢复分类标签。由 XGBoost 特点可知，每个被动方（无标签数据的参与方）一旦获得 g 和 h，仅用其本地数据就能够独立地确定局部最优分裂。因此，非联邦学习下活动方将 g 和 h 发送到每个被动方是可行的。但由于 g 和 h 可以用来获取分类标签信息，为了确保安全，联邦学习要求各被动方无法直接访问 g 和 h，主动方（有标签数据的参与方）在将 g 和 h 发送给被动方之前要进行加密。随后，每个被动方使用加密的 g 和 h 确定局部最优分裂。被动方 A 使用由主动方 B 加密的 g 和 h 进行计算。其中，g 和 h 在主动方 B 侧本地

计算，B 侧没有泄露样本分类标签；被动方 A 本地计算经加法同态加密后的梯度直方图，B 解密梯度直方图，但是不知道具体对应的 ID 集合，保护了 A 侧 ID 集合隐私信息。

4.4　谷歌数据共享框架

4.4.1　概述

TFE 是由谷歌开源的联邦学习框架，可用于对分散式数据进行机器学习和计算[91]。开发者可基于其模型和数据来模拟联邦学习算法并实验新算法。TFF 提供的构建块也可用于实现非学习计算，例如对分散式数据进行聚合分析。TFF 的接口可以分成两层：1) FL API，提供了一组高阶接口，开发者能够利用其联合训练和评估实现 TensorFlow 模型。2) FC（federated core）API，可以通过在强类型函数式编程环境中结合使用 TensorFlow 与分布式通信运算符，简洁地表达新的联合算法。这一层也是构建联合学习的基础。

TFF 可用于模拟对联合学习系统的目标攻击和基于隐私的差异防御[92]，使用潜在的恶意客户端构建一个迭代进程。同时，TFF 还支持自定义的攻击方式，通过编写一个客户端更新函数来实现新的攻击算法。此外，新的防御方案可以通过定制状态聚合函数及聚合客户端输出以获得自定义安全全局更新策略。

4.4.2　隐私保护库

谷歌开源的 TensorFlow Privacy 是将差分隐私技术集成到诸如随机梯度下降的迭代训练过程中，提供了隐私保护 Python 库，以训练具有差异隐私的机器学习模型[93]。引入模块化方法最大限度地减少对训练算法的更改，为隐私机制提供各种配置策略；隔离和简化关键逻辑解决了在隐私敏感数据集上训练机器学习模型的实际挑战[94]。而 TFF 与 TensorFlow Privacy 隐私库是互操作的，可以支撑联邦训练算法的不同隐私模型，例如，支持基础的 DP-FedAvg 算法进行差分隐私训练。此外，TFF 还提供可扩展的隐私保护接口，可以实现自定义差分隐私算法并将其应用于联邦平均的参数更新。

TFF 实现本地数据隐私保护的另一个重要方式是与生成式对抗网络（GAN, generative adversarial network）结合，如 DP-FedAvg-GAN 算法。Rangan 等[95]展示了 DP-FedAvg-GAN 算法下联邦学习、生成模型和差分隐私相结合的有效性。

4.5　英伟达医疗数据共享框架

NVIDIA Clara 是一个医疗保健应用框架，用于人工智能成像、基因组学以及智能传感器的开发和部署[47]。以服务器−客户端结构的联邦学习为特色，各数据持有方和中央服务器通过边缘 AI 计算平台 NVIDIA EGX 构建训练网络，实现支持隐私保护的智能计算。NVIDIA Clara 作为一个商业应用产品，需基于英伟达的 GPU 硬件来获取服务。

联邦学习的中央服务器虽然通过适当地聚合客户端本地模型更新可以获得一个高精度的全局模型，但是共享的模型可能间接地泄露本地的训练示例。NVIDIA Clara 云边协同的架构，可以通过控制站点的全局模型只共享部分模型权重，从而保护隐私，并且数据较少暴露在模型反转中。Paillier 等[10]探讨了在联邦学习系统中应用差分隐私技术来保护病人资料的可行性。其实验结果表明，模型性能与隐私保护代价之间存在折中关系。

此外，为确保在客户机−服务器通信时数据和模型的安全性，Clara 使用联合学习令牌来建立客户端和服务器之间的信任[22]。联邦学习令牌会在整个联邦训练会话生命周期中使用。客户端需要验证服务器标识，服务器也需要验证客户端。客户端−服务器数据交换基于 HTTPS 协议进行安全通信。

5　挑战与展望

随着海量异构数据的日益增长，数据隐私保护问题迫在眉睫。非聚合式数据共享方法在数据分享模式基础上增强了隐私保护，降低数据泄露的风险。通过总结分析安全多方计算、联邦学习及非聚合式数据共享框架的发展，本文进一步提出了非聚合式数据共享领域在未来更为复杂的信息世界中面临的挑战和机遇。

5.1　复杂的多参与方场景

随着物联网和边缘计算的发展，智能设备能力不断提升，数据往往分布在多个节点上。而且，节点动态性强，可以自由加入和离开，导致多参与方情况更加复杂和不稳定。

由于安全多方计算协议复杂度高，传统的方法侧重研究两方参与场景，无法很好地拓展到多方环境。不能简单地将两方协议执行多次来达到多方计算，因为在简单重复两方协议的过程中，参与方会得到一些中间结果，而它本身是不能获知这些中间结果的。虽然有部分研究者针对这一问题展开研

究，但还未能完全解决。例如，Wang 等[96]提出让某参与方充当领导者，组织协议在多方之间执行，但无法很好地应对领导者是恶意节点的情况，而且领导者的选择也影响协议效率；Kolesnikov 等[97]借助不经意可编程的伪随机函数（OPPRF, oblivious evaluation of a programmable pseudorandom function）的特性来完成多方计算，5 个各有 2^{20} 个元素的参与者执行协议仅需 72 s，但其实验选取的是较小数字或较短字符串等简单元素，不能很好地应对复杂场景。云计算给多参与方场景的安全计算带来了机遇[98]，其强大的计算和存储能力可以用于支撑复杂的协议，但也存在一定挑战。首先，传统方法是针对数据存储在本地、计算执行在本地的特点而设计的，不能照搬到云环境中。其次，云中心节点一般被认为是半诚实的，需要对传输数据进行加密，而这个操作会导致数据拥有方难以在协议执行过程中对外包的数据集进行访问控制。因此，需要研究适应动态多参与方的安全计算协议，可以利用现有的一些高级密码学成果和边缘计算环境中的智能节点等，达到高效合理计算的目标。

参与方的增多也会影响联邦学习算法的准确性和性能。首先，管理和筛选合格的参与方需要额外的开销；其次，本地数据节点增多，其不同的数据分布会影响全局优化模型的效果。在最坏的情况下生成的联合模型并不比在单个节点上训练生成的模型好[99]，在非独立同分布的数据情况下，FedAvg 训练卷积神经网络的准确率会显著下降[79]。改进的潜在方法包括增加并行性、增加本地节点的计算量、参与方聚类择优、模型自适应调整等。针对参与方的动态变化及应对短时间内参与方数量弹性增加/降低等情况，如何设计一个快速收敛且保证准确率的算法模型还需要进一步的研究。

5.2 性能优化与开销代价的平衡

为了增强数据隐私保护，需要复杂的协议和算法保证最低程度的数据泄露，但在实现过程中，则会造成系统开销增大。因此，如何平衡所需的优化性能和执行开销是十分重要的问题，关系到方案的实际应用拓展情况。

在安全多方计算领域，Pinkas 等[29]提出货币成本衡量标准，其表示 PSI 协议在云计算平台执行所带来的计算和通信开销，并根据该标准设计了 2 个半诚实对手安全协议，其中一个具有非常低的通信开销，另一个则在计算开销上表现出色，然而这 2 种特性并未出现在同一个协议上。如何权衡计算和通信开销，在满足应用场景和用户隐私需求的情况下达到平衡，能否借助货币成本标准设计出更加折中的方案等，都是亟待解决的问题。

在联邦学习领域，其分布式架构存在多节点的数据交换过程，需要较好的通信带宽，多节点训练的拓展性也与其正相关。然而，在大型异构环境分布式训练中，本地数据节点常常会受到可用通信带宽和资源的限制。梯度压缩是解决这类问题的一种潜在的有效方法，但大多压缩算法采用近似编码表示内容，存在一定的信息损失，无法被广泛采用。其次，运行速度会受到影响，而且可能造成优化后的所有梯度聚集或无法达到相同的测试性能。因此，如何在保证算法准确度的前提下减少通信成本仍然需要进一步研究。

5.3 潜在安全问题

非聚合式数据共享虽然能保证源数据不在单点聚合，但仍面临一些潜在的安全问题。

虚假数据。攻击者可以冒充虚假参与方，提交模拟的数据，造成数据中毒的情况。在安全多方计算场景中，此方案可以用来攻击外包计算节点，消耗其资源，使真正参与方得不到公平的资源使用机会。在联邦学习场景中，虚假数据会严重影响全局优化模型的准确性，造成严重后果。针对这一问题，可以尝试从参与方认证、可靠性激励等方面提出解决方案。

架构安全性。安全多方计算中，由于加密后的数据需要传输交互，若一个参与方被攻破，则信息可能被泄露。如果攻击者获悉加密方法及部分真实数据信息，就能在一定程度上破解密文。安全多方计算的数据安全性取决于使用的加密方法和数据传输通道的安全性。联邦学习中参数服务器对本地数据及其训练过程是不可见的，攻击者可利用缺乏透明性对系统进行攻击。通过有目的或无目的的模型攻击，使训练过程中数据样本偏差不可察觉，影响全局模型的性能或操控模型偏向。如何确保本地数据节点提供诚实可信的训练，保证整个联邦学习流程的安全性仍然是一个难点。

5.4 隐私保护技术结合

随着学术界和工业界对数据隐私保护问题的重视，研究者致力于开发各种增强数据隐私保护的技术，如何结合现有的技术，针对不同的应用场景，提升整体系统的数据隐私保护能力，也是未来值得

研究的方向。

生成对抗网络是一种深度神经网络结构，由 Goodfellow 等[100]在 2014 年提出，它可以学习数据集分布，生成与数据集相似的逼真数据。利用 GAN 可以把同等特征的模拟数据发送给对方，而不泄露真实数据内容，保护源数据隐私。非聚合式数据共享方法主要强调数据共享的作用及其目标，若数据特征保留，则同样可以训练准确的模型。但现有的 GAN 相关研究缺乏对 GAN 模型的性能、可用性、生成样本的质量等方面的客观评价，导致模型的判定受一定主观因素的影响。其次，仍然存在训练过程不稳定、训练结果难以收敛、模式崩溃等问题，PPGAN[101]、DPGAN[102]虽在一定程度上有所改善，但仍有很大的优化空间。

本地差分隐私（LDP, local differential privacy）[103]技术让数据所有者在发布数据之前对数据进行扰动，避免需要信任的第三方进行数据收集和处理，保护源数据隐私。该技术能够消除所增加的正、负噪声，并且基于预定义的查询来设计扰动机制，对指定的查询能得到准确的估计。但目前的查询类型比较单一，包括离散型数据的计数查询和连续型数据的均值查询，不支持其他类型的查询，如范围查询、最值查询等。其次，本地差分隐私对复杂数据类型研究的工作还不足，目前仅有文献[104-105]等对图等复杂数据进行了研究，不能很好地处理具有边权或边属性的图，以及特定图的挖掘任务研究，如三角形计数和频繁的子图结构挖掘等。

联邦学习框架需要上传本地训练模型的参数至中央参数服务器，中央服务器结合全局信息进行调优操作后，再把优化后的参数信息返回给各本地节点。参数数据传输链路及中央参数调优计算可结合安全多方计算技术，保证参数信息交互过程中的隐私安全性。具体的技术选型、计算操作实现还需要结合联邦学习实际应用场景进行进一步的分析和优化。

6　结束语

随着数据量的迅速增多以及用户隐私保护需求的提高，传统的集中式获取全部数据的方式已不能很好地应对多方数据共享的场景，非聚合式数据共享有望成为主流。非聚合式数据共享在有效降低源数据隐私泄露风险的情况下，完成数据处理和挖掘的目标。本文简要介绍了主要的非聚合式数据共享方法的研究现状，首先，阐述早期非聚合式数据共享方法安全多方计算的相关技术，包括同态加密、不经意传输、秘密共享、隐私集合交集协议等。其次，从源数据节点和通信传输优化 2 个方面介绍近期的非聚合式数据共享研究热点联邦学习技术。此外，本文还对比分析了非聚合式数据共享框架，如百度 PaddleFL、微众银行 FATE、谷歌 TFF 等，给未来研究方案的实际构建和运行提供支撑。最后，本文提出了 4 个非聚合式数据共享领域面临的挑战和潜在研究方向。期望本文可以为研究人员快速全面地了解和掌握非聚合式数据共享领域的基本现状和研究发展提供参考和帮助。

参考文献：

[1] 于文平. 中国物联网市场发展现状分析及未来五年发展趋势预测[J]. 物联网技术, 2018, 8(3): 9-11.
YU W P. Analysis of the development status of China's Internet of things market and forecast of its development trend in the next five years[J]. Internet of Things Technologies, 2018, 8(3): 9-11.

[2] 瑞星. 2020 年中国网络安全报告[J]. 信息安全研究, 2021, 7(2): 102-109.
RISING. China’s cybersecurity report in 2020[J]. Journal of Information Security Research, 2021, 7(2):102-109.

[3] REGULATION G D P. Regulation (EU) 2016/679 of the European Parlia ment and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46[J]. Official Journal of the European Union (OJ). 2016, 59(1-88): 294.

[4] YAO A C. Protocols for secure computation[C]//23rd IEEE Annual Symposium on Foundations of Computer Science. Piscataway: IEEE Press, 1982: 160-164.

[5] WINTERNITZ R S. A secure one-way hash function built from DES[C]//1984 IEEE Symposium on Security and Privacy. Piscataway: IEEE Press, 1984: 88.

[6] SHI W S, CAO J, ZHANG Q, et al. Edge computing: vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.

[7] ION M, KREUTER B, NERGIZ E, et al. Private intersection-sum protocol with applications to attributing aggregate AD conversions[J]. IACR Cryptology ePrint Archive, 2017, 2017: 738.

[8] RIVEST R L, ADLEMAN L, DERTOUZOS M L. On data banks and privacy homomorphisms[J]. Foundations of Secure Computation, 1978, 4(11): 169-180.

[9] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems[J]. Communications of the ACM, 1978, 21(2): 120-126.

[10] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes [C]//International Conference on the Theory and Application of Cryptographic Techniques. Berlin: Springer, 1999: 223-238.

[11] ABADI A, TERZIS S, METERE R, et al. Efficient delegated private

set intersection on outsourced private datasets[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 16(4): 608-624.

[12] RUAN O, WANG Z H, MI J, et al. New approach to set representation and practical private set-intersection protocols[J]. IEEE Access, 2019, 7: 64897-64906.

[13] KAVOUSI A, MOHAJERI J, SALMASIZADEH M. Improved secure efficient delegated private set intersection[C]//2020 28th Iranian Conference on Electrical Engineering. Piscataway: IEEE Press, 2020: 1-6.

[14] RESENDE A C D, ARANHA D F. Faster unbalanced private set intersection[M]. Berlin: Springer, 2018.

[15] BALDI P, BARONIO R, DE CRISTOFARO E, et al. Countering GATTACA: efficient and secure testing of fully-sequenced human genomes[C]//The 18th ACM Conference on Computer and communications Security. New York: ACM Press, 2011: 691-702.

[16] CHEN H, LAINE K, RINDAL P. Fast private set intersection from homomorphic encryption[C]//The 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 299.

[17] FREEDMAN M J, NISSIM K, PINKAS B. Efficient private matching and set intersection[C]//Advances in Cryptology-EUROCRYPT. Berlin: Springer, 2004: 1-9.

[18] INBAR R, OMRI E, PINKAS B. Efficient scalable multiparty private set-intersection via garbled bloom filters[C]//Security and Cryptography for Networks. Berlin: Springer, 2018:235-252.

[19] HUANG Y, CHAPMAN P, EVANS D. Privacy-preserving applications on Smartphones[C]//6th USENIX Workshop on Hot Topics in Security. Berkeley: USENIX Association, 2011: 1-6.

[20] PINKAS B, SCHNEIDER T, TKACHENKO O, et al. Efficient circuit-based PSI with linear communication[C]//Advances in Cryptology - EUROCRYPT 2019. Berlin: Springer, 2019: 122-153.

[21] PINKAS B, SCHNEIDER T, WEINERT C, et al. Efficient circuit-based PSI via Cuckoo Hashing[C]//EUROCRYPT 2018. Berlin: Springer, 2018: 125-157.

[22] HUANG H F, CHANG C C. A new design for efficent t-out-n oblivious transfer scheme[C]//19th International Conference on Advanced Information Networking and Applications. Los Alamitos: IEEE Computer Society, 2005: 28-30.

[23] 徐彦蛟, 李顺东, 王道顺, 等. 基于椭圆曲线公钥系统的不经意传输协议[J]. 计算机科学, 2013, 40(12): 186-191.
XU Y J, LI S D, WANG D S, et al. Oblivious transfer based on elliptic curve public key cryptosystems[J]. Computer Science, 2013, 40(12): 186-191.

[24] ISHAI Y, KILIAN J, NISSIM K, et al. Extending oblivious transfers efficiently[C]//Advances in Cryptology - CRYPTO 2003. Berlin: Springer, 2003: 145-161.

[25] ASHAROV G, LINDELL Y, SCHNEIDER T, et al. More efficient oblivious transfer and extensions for faster secure computation[C]//The 2013 ACM SIGSAC Conference on Computer & communications Security. New York: ACM Press, 2013: 673-701.

[26] CHASE M, MIAO P H. Private set intersection in the Internet setting from lightweight oblivious PRF[C]//Advances in Cryptology - CRYPTO 2020. Berlin: Springer, 2020: 34-63.

[27] CIAMPI M, ORLANDI C. Combining private set-intersection with secure two-party computation[C]//Security and Cryptography for Networks. Berlin: Springer, 2018: 105.

[28] KARAKOÇ F, KÜPÇÜ A. Linear complexity private set intersection for secure two-party protocols[C]//Cryptology and Network Security. Cham: Springer International Publishing, 2020: 409-429.

[29] PINKAS B, ROSULEK M, TRIEU N, et al. SpOT-light: lightweight private set intersection from sparse OT extension[C]//Advances in Cryptology - CRYPTO 2019. Berlin: Springer, 2019: 401-431.

[30] PINKAS B, SCHNEIDER T, ZOHNER M. Scalable private set intersection based on OT extension[J]. ACM Transactions on Privacy and Security, 2018, 21(2): 1-35.

[31] RINDAL P, ROSULEK M. Improved private set intersection against malicious adversaries[C]//Advances in Cryptology - EUROCRYPT 2017. Berlin: Springer, 2017: 235-259.

[32] RINDAL P, ROSULEK M. Malicious-secure private set intersection via dual execution[C]//The 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 1229-1242.

[33] GHOSH S, NILGES T. An algebraic approach to maliciously secure private set intersection[C]//Advances in Cryptology - EUROCRYPT 2019. Berlin: Springer, 2019:1064.

[34] KOLESNIKOV V, KUMARESAN R, ROSULEK M, et al. Efficient batched oblivious PRF with applications to private set intersection[C]//The 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 818-829.

[35] DONG C Y, CHEN L Q, WEN Z K. When private set intersection meets big data: an efficient and scalable protocol[C]//The 2013 ACM SIGSAC Conference On Computer & Communications Security. New York: ACM Press, 2013: 789-800.

[36] GROCE A, RINDAL P, ROSULEK M. Cheaper private set intersection via differentially private leakage[J]. Proceedings on Privacy Enhancing Technologies, 2019, 2019(3): 6-25.

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

[38] CHENG K W, FAN T, JIN Y L, et al. SecureBoost: a lossless federated learning framework[J]. IEEE Intelligent Systems, 2561, PP(99): 1.

[39] CRAMER R, DAMGÅRD I, MAURER U. General secure multi-party computation from any linear secret-sharing scheme[C]//Advances in Cryptology-EUROCRYPT 2000. Berlin: Springer, 2000: 316-334.

[40] RABIN M O. Randomized Byzantine generals[C]//24th Annual Symposium on Foundations of Computer Science. Piscataway: IEEE Press, 1983: 403-409.

[41] LYU S, YE J H, YIN S J, et al. Unbalanced private set intersection cardinality protocol with low communication cost[J]. Future Generation Computer Systems, 2020, 102: 1054-1061.

[42] MIAO P H, PATEL S, RAYKOVA M, et al. Two-sided malicious security for private intersection-sum with cardinality[C]//Advances in Cryptology - CRYPTO 2020. Berlin: Springer, 2020: 3-33.

[43] SHI R H, ZHANG M W. A feasible quantum protocol for private set intersection cardinality[J]. IEEE Access, 2019, 7: 105-112.

[44] DONG C Y, LOUKIDES G. Approximating private set union/intersection cardinality with logarithmic complexity[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(11): 2792-2806.

[45] CERULLI A, CRISTOFARO E, SORIENTE C. Nothing refreshes like a RePSI: reactive private set intersection[C]//Applied Cryptography and Network Security. Berlin: Springer, 2018: 280-300

[46] ATENIESE G, CRISTOFARO E, TSUDIK G. (if) size matters: size-hiding private set intersection[C]//Public Key Cryptography. Ber-

lin: Springer, 2011: 156-173.

[47] YANG Q, LIU Y, CHEN T J, et al. Federated machine learning[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

[48] ZHAO Y, LI M, LAI L, et al. Federated learning with non-IID data[J]. arXiv Preprint, arXiv:1806.00582, 2018.

[49] GOETZ J, MALIK K, BUI D, et al. Active federated learning[J]. arXiv Preprint, arXiv:1909.12641, 2019.

[50] KANG J W, XIONG Z H, NIYATO D, et al. Incentive mechanism for reliable federated learning: a joint optimization approach to combining reputation and contract theory[J]. IEEE Internet of Things Journal, 2019, 6(6): 10700-10714.

[51] YOSHIDA N, NISHIO T, MORIKURA M, et al. Hybrid-FL for wireless networks: cooperative learning mechanism using non-IID data[C]//IEEE International Conference on Communications. Piscataway: IEEE Press, 2020: 1-7.

[52] SARIKAYA Y, ERCETIN O. Motivating workers in federated learning: a stackelberg game perspective[J]. IEEE Networking Letters, 2020, 2(1): 23-27.

[53] WANG S Q, TUOR T, SALONIDIS T, et al. Adaptive federated learning in resource constrained edge computing systems[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1205-1221.

[54] LIU L M, ZHANG J, SONG S H, et al. Client-edge-cloud hierarchical federated learning[C]//IEEE International Conference on Communications. Piscataway: IEEE Press, 2020: 1-6.

[55] SHOKRI R, SHMATIKOV V. Privacy-preserving deep learning[C]//The 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2015: 1310-1321.

[56] KONEČNÝ J, MCMAHAN B, RAMAGE D. Federated optimization: Distributed optimization beyond the datacenter[J]. arXiv Preprint, arXiv:1511.03575, 2015.

[57] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. arXiv Preprint, arXiv:1812.06127, 2018.

[58] CHEN X, CHEN T, SUN H, et al. Distributed training with heterogeneous data: bridging median-and mean-based algorithms[J]. arXiv Preprint, arXiv:1906.01736, 2019.

[59] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for federated learning[J]. arXiv Preprint, arXiv:1910.06378, 2019.

[60] XIE C, KOYEJO S, GUPTA I. Asynchronous federated optimization[J]. arXiv Preprint, arXiv:1903.03934, 2019.

[61] LI D, WANG J. FedMD: Heterogenous federated learning via model distillation[J]. arXiv Preprint, arXiv:1910.03581, 2019.

[62] LIU L M, ZHANG J, SONG S H, et al. Edge-assisted hierarchical federated learning with non-IID data[J]. arXiv Preprint, arXiv: 1905.06641, 2019.

[63] RAMANAN P, NAKAYAMA K. BAFFLE: blockchain based aggregator free federated learning[C]//2020 IEEE International Conference on Blockchain. Piscataway: IEEE Press, 2020: 72-81.

[64] ROY A G, SIDDIQUI S, PÖLSTERL S, et al. BrainTorrent: a peer-to-peer environment for decentralized federated learning[J]. arXiv Preprint, arXiv: 1905.06731, 2019.

[65] LIU Y, KANG Y, ZHANG X W, et al. A communication efficient collaborative learning framework for distributed features[J]. arXiv Preprint, arXiv: 1912.11187, 2019.

[66] YAO X, HUANG C F, SUN L F. Two-stream federated learning: reduce the communication costs[C]//2018 IEEE Visual Communications and Image Processing. Piscataway: IEEE Press, 2018: 1-4.

[67] STICH S U. Local SGD converges fast and communicates little[J]. arXiv Preprint, arXiv:1805.09767, 2018.

[68] WANG J Y, JOSHI G. Adaptive communication strategies to achieve the best error-runtime trade-off in local-update SGD[J]. arXiv Preprint, arXiv:1810.08313, 2018.

[69] MOHRI M, SIVEK G, SURESH A T. Agnostic federated learning[J]. arXiv Preprint, arXiv:1902.00146, 2019.

[70] ALISTARH D, GRUBIC D, LI J, et al. QSGD: communication-efficient SGD via gradient quantization and encoding[C]//Advances in Neural Information Processing Systems. Massachusetts: MIT Press, 2017: 1709-1720.

[71] KHALED A, RICHTÁRIK P. Gradient descent with compressed iterates[J]. arXiv Preprint, arXiv:1909.04716, 2019.

[72] VOGELS T, KARIMIREDDY S P, JAGGI M. PowerSGD: practical low-rank gradient compression for distributed optimization[C]//Advances in Neural Information Processing Systems. Massachusetts: MIT Press, 2019: 14259-14268.

[73] BERNSTEIN J, ZHAO J, AZIZZADENESHELI K, et al. SignSGD with majority vote is communication efficient and fault tolerant[J]. arXiv Preprint, arXiv:1810.05291, 2018.

[74] JIANG Y H, KONEČNÝ J, RUSH K, et al. Improving federated learning personalization via model agnostic meta learning[J]. arXiv Preprint, arXiv: 1909.12488,2019.

[75] SHEN T, ZHANG J, JIA X K, et al. Federated mutual learning[J]. arXiv Preprint, arXiv: 2006.16765, 2020.

[76] SAVAZZI S, NICOLI M, RAMPA V. Federated learning with cooperating devices: a consensus approach for massive IoT networks[J]. IEEE Internet of Things Journal, 2020, 7(5): 4641-4654.

[77] SATTLER F, WIEDEMANN S, MÜLLER K R, et al. Robust and communication-efficient federated learning from non-I.I.D. data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(9): 3400-3413.

[78] KANG J W, XIONG Z H, NIYATO D, et al. Incentive design for efficient federated learning in mobile networks: a contract theory approach[C]//2019 IEEE VTS Asia Pacific Wireless Communications Symposium. Piscataway: IEEE Press, 2019: 1-5.

[79] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. New York: ACM Press, 2017: 1273-1282.

[80] KISS Á, LIU J, SCHNEIDER T, et al. Private set intersection for unequal set sizes with mobile applications[J]. Proceedings on Privacy Enhancing Technologies, 2017, 2017(4): 177-197.

[81] NASR M, SHOKRI R, HOUMANSADR A. Improving deep learning with differential privacy using gradient encoding and denoising[J]. arXiv Preprint, arXiv: 2007.11524, 2020.

[82] MOHASSEL P, RINDAL P. ABY3: a mixed protocol framework for machine learning[C]//The 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2018: 35-52.

[83] 何雯, 白翰茹, 李超. 基于联邦学习的企业数据共享探讨[J]. 信息与电脑(理论版), 2020, 32(8): 173-176.
HE W, BAI H R, LI C. Research on enterprise data sharing based on federated learning[J]. China Computer & Communication, 2020, 32(8): 173-176.

[84] ZHANG C, LI S, XIA J, et al. Batchcrypt: efficient homomorphic

encryption for cross-silo federated learning[C]//USENIX Annual Technical Conference, Berkeley: USENIX Association, 2020: 493-506.

[85] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)[J]. The Annals of Statistics, 2000, 28(2): 337-407.

[86] GANG L, CHAWATHE S S. Privacy-preserving inter-database operations[C]//2nd Symposium on Intelligence and Security Informatics. Berlin: Springer, 2004: 66-82.

[87] ALEDHARI M, RAZZAK R, PARIZI R M, et al. Federated learning: a survey on enabling technologies, protocols, and applications[J]. IEEE Access, 2020, 8: 699-725.

[88] LE H Q, NGUYEN M N, HONGF C S. A federated learning for+ mage classification with heterogeneous data[J]. Journal of Korea Information Science Society, 2020: 352-354.

[89] WANG H Y, SREENIVASAN K, RAJPUT S, et al. Attack of the tails: yes, you really can backdoor federated learning[J]. arXiv Preprint, arXiv: 2007.05084, 2020.

[90] YU B, QIU W Y, CHEN C, et al. SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting[J]. Bioinformatics, 2020, 36(4): 1074-1081.

[91] MCMAHAN H B, ANDREW G. A general approach to adding differential privacy to iterative training procedures[J]. arXiv Preprint, arXiv: 1812.06210, 2018.

[92] AUGENSTEIN S, MCMAHAN H B, RAMAGE D, et al. Generative models for effective ML on private, decentralized datasets[J]. arXiv Preprint, arXiv: 1911.06679, 2019.

[93] GONZALEZ E, LEDIJU BELL M A. A GPU approach to real-time coherence-based photoacoustic imaging and its application to photoacoustic visual servoing[J]. Photons Plus Ultrasound: Imaging and Sensing, 2020, 1124: 1124054.

[94] LI W Q, MILLETARÌ F, XU D G, et al. Privacy-preserving federated brain tumour segmentation[C]//Machine Learning in Medical Imaging. Berlin: Springer, 2019: 133-141.

[95] RANGAN R, TURAKHIA N, JOLY A. Corrections to “countering load-to-use stalls in the NVIDIA Turing GPU”[J]. IEEE Micro, 2021, 41(1): 83.

[96] WANG Z S, BANAWAN K, ULUKUS S. Multi-party private set intersection: an information-theoretic approach[J]. IEEE Journal on Selected Areas in Information Theory, 2021, 2(1): 366-379.

[97] KOLESNIKOV V, MATANIA N, PINKAS B, et al. Practical multi-party private set intersection from symmetric-key techniques[C]//The 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 257-272.

[98] ALI M, MOHAJERI J, SADEGHI M R, et al. Attribute-based fine-grained access control for outscored private set intersection computation[J]. Information Sciences, 2020, 536: 222-243.

[99] ZHANG Y, DUCHI J C, WAINWRIGHT M J. Communication- efficient algorithms for statistical optimization[J]. The Journal of Machine Learning Research, 2013, 14(1): 3321-3363.

[100] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems. Berlin: Springer, 2014: 2672-2680.

[101] LIU Y, PENG J L, YU J J Q, et al. PPGAN: privacy-preserving generative adversarial network[C]//2019 IEEE 25th International Conference on Parallel and Distributed Systems. Piscataway: IEEE Prss, 2019: 985-989.

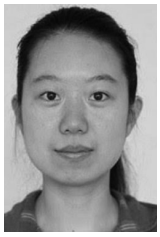
[102] XIE L Y, LIN K X, WANG S, et al. Differentially private generative adversarial network[J]. arXiv Preprint, arXiv: 1802.06739, 2018.

[103] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Local privacy, data processing inequalities, and statistical minimax rates[J]. arXiv Preprint, arXiv: 1302.3203, 2013.

[104] QIN Z, YU T, YANG Y, et al. Generating synthetic decentralized social graphs with local differential privacy[C]//The 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 425-438.

[105] 张佳程, 彭佳, 王雷. 大数据环境下的本地差分隐私图信息收集方法[J]. 信息网络安全, 2020, 20(6): 44-56.
ZHANG J C, PENG J, WANG L. A graph information collection method based on local differential privacy in big data environment[J]. Netinfo Security, 2020, 20(6): 44-56.

[作者简介]



李尤慧子（1989− ），女，河南新蔡人，博士，杭州电子科技大学副教授，主要研究方向为边缘计算、隐私安全、移动互联网计算、高能效系统。

殷昱煜（1980− ），男，重庆人，博士，杭州电子科技大学教授，主要研究方向为边缘计算、服务计算、大数据分析、软件形式化方法等。

高洪皓（1985− ），男，浙江临海人，博士，上海大学副教授、韩国嘉泉大学教授，主要研究方向为软件形式化验证、服务协同计算、无线网络和工业物联网、智能医学影像处理等。

金一（1982− ），女，河北石家庄人，博士，北京交通大学教授、博士生导师，主要研究方向为机器学习与认知计算、人工智能及应用、图像感知与识别。

王新珩（1968− ），男，山东平度人，博士，西交利物浦智能工程学院教授、博士生导师，主要研究方向为物联网、室内定位、智能化服务和智慧城市。