大数据与决策研究

（政策与技术跟踪专题）

2021 年第 35 期（总第 78 期）

广西壮族自治区信息中心
广西壮族自治区大数据研究院　　2021 年 7 月 31 日

编者按：数据流通是释放数据价值的关键环节，隐私计算技术为数据流通提供了解决方案。2021 年 5 月印发的《全国一体化大数据中心协同创新体系算力枢纽实施方案》明确提出建设数据共享开放、政企数据融合应用等数据流通共性设施平台，试验多方安全计算、区块链、隐私计算、数据沙箱等技术模式，构建数据可信流通环境。本期主要介绍隐私计算相关技术情况。

本期要目

隐私计算的概念和技术原理

一、隐私计算的概念

2016 年发布的《隐私计算研究范畴及发展趋势》正式提出“隐私计算”一词，并将隐私计算定义为：“面向隐私信息全生命周期保护的计算理论和方法，是隐私信息的所有权、管理权和使用权分离时隐私度量、隐私泄漏代价、隐私保护与隐私分析复杂性的可计算模型与公理化系统。”

通俗地说，隐私计算是指在保护数据本身不对外泄露的前提下实现数据分析计算的一类信息技术，包含了数据科学、密码学、人工智能等众多技术体系的交叉融合。

二、隐私计算的技术原理

从技术实现原理看，隐私计算主要分为密码学和可信硬件两大领域。密码学技术目前以多方安全计算为代表；可信硬件领域主要指可信执行环境；此外，还有基于以上两种技术路径衍生出的联邦学习等相关技术。

（一）多方安全计算

多方安全计算技术的核心思想是设计特殊的加密算法和协议，基于密码学原理实现在无可信第三方的情况下，在多个参与方输入的加密数据之上直接进行计算。

多方安全计算由姚期智等人于 1982 年提出，以交互不可逆的密文数据的方式实现了对数据的安全保护，每个参与方不能得到其他参与方的任何输入信息，只能得到计算结果。

多方安全计算的性能相对较低且技术开发难度较大，但近几年产业界的高度关注使得其性能得以迅速提升、技术可用性得到很大改善。多方安全计算技术框架见图 1。

图 1 多方安全计算技术框架

多方安全计算的实现包含多个关键的底层密码学协议或框架，主要包括不经意传输、混淆电路、秘密分享、同态加密等。

（1）不经意传输，提出了一种在数据传输与交互过程中保护隐私的思路。在不经意传输协议中，数据发送方同时发送多个消息，而接收方仅获取其中之一。发送方无法判断接收方获取了具体哪个消息，接收方也对其他消息的内容一无所知。

（2）混淆电路，是一种将计算任务转化为布尔电路并对真值表进行加密打乱等混淆操作以保护输入隐私的思路。利用计算机编程将目标函数转化为布尔电路后，对每一个门输出的真值进行加密，参与方之间在互相不掌握对方私有数据的情况下共同完成计算。

（3）秘密分享，也称秘密分割或秘密共享，给出了一种分而治之的秘密信息管理方案。秘密分享的原理是将秘密拆分成多个分片，每个分片交由不同的参与方管理。只有超

过一定门限数量的若干个参与方共同协作才能还原秘密信息，仅通过单一分片无法破解秘密。

（4）同态加密，是一类实现在基础的加密操作之上直接完成密文数据间运算的加密算法。数据经过同态加密后进行计算得到的结果与用同一方法在明文计算下得到的结果保持一致，即先计算后解密等价于先解密后计算。

在经典的多方安全计算中，两方计算主要采用不经意传输与混淆电路结合的方案，三方及以上的计算则进一步结合了秘密分享，也有观点将同态加密视作一套基于密码学理论但独立于多方安全计算的隐私计算技术。

（二）可信执行环境

可信执行环境的核心思想是构建一个独立于操作系统而存在的可信的、隔离的机密空间，数据计算仅在该安全环境内进行，通过依赖可信硬件来保障其安全。

可信执行环境最本质的属性是隔离，通过芯片等硬件技术并与上层软件协同对数据进行保护，且同时保留与系统运行环境之间的算力共享。目前，可信执行环境的代表性硬件产品主要有 Intel 的 SFX、ARM 的 TrustZone 等。可信执行环境技术体系见图 2。

图 2 可信执行环境技术体系

严格地讲，可信执行环境并不属于“数据可用不可见”，但其通用性高、开发难度低，在通用计算、复杂算法的实现上更为灵活，使得其在数据保护要求不是特别严苛的场景下仍有很多发挥价值的空间。

（三）联邦学习

联邦学习的本质是分布式的机器学习，在保证数据隐私安全的基础上，实现共同建模，提升模型的效果。对于基于数据隐私保护的分布式机器学习，早在 2012 年即有学者发表了相关研究成果，直到 2016 年谷歌率先提出联邦学习的概念，才逐步受到更广泛的关注。

联邦学习的目标是在不聚合参与方原始数据的前提下，实现保护终端数据隐私的联合建模。根据数据集的不同类型，联邦学习分为横向联邦学习、纵向联邦学习与联邦迁移学习。

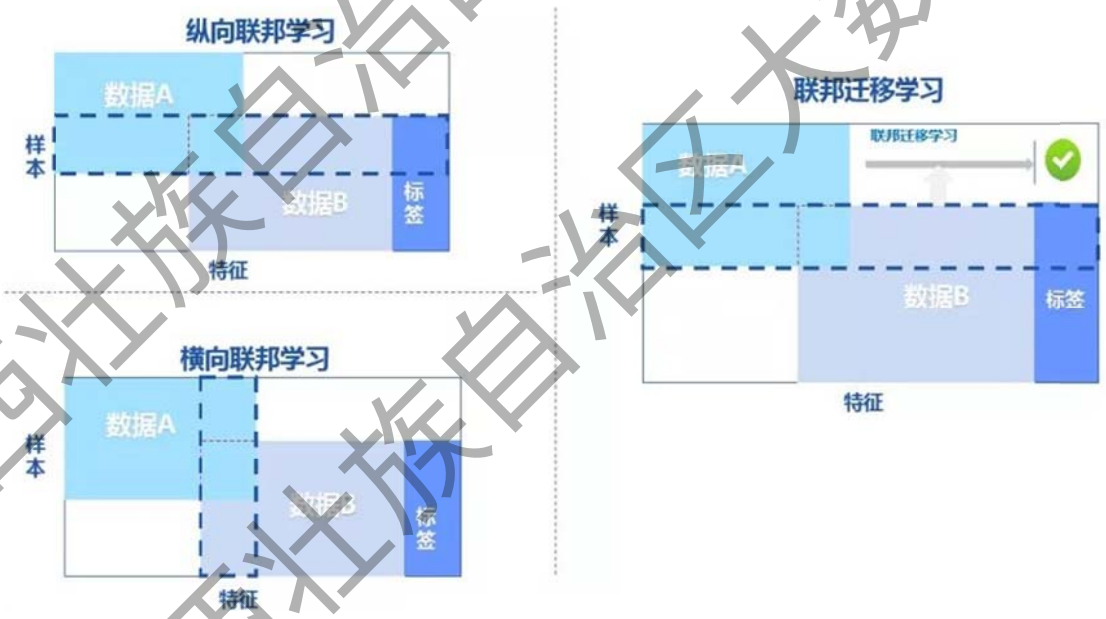

图 3 联邦学习分类示意图

（1）横向联邦学习。横向联邦学习更适用于在特征重合较多，而样本重合较少的数据集间进行联合计算的场景。以样本维度（即横向）对数据集进行切分，以特征相同而样

本不完全相同的数据部分为对象进行训练。谷歌在 2016 年提出的安卓手机模型更新数据联合建模方案是利用单个用户使用安卓手机时，不断在本地更新模型参数并上传到安卓云上，从而使特征维度相同的各数据拥有方联合建模。

（2）纵向联邦学习。纵向联邦学习更适用于样本重合较多，而特征重合较少的数据集间联合计算的场景。以特征维度（即纵向）对数据集进行切分，以样本相同而特征不完全相同的数据部分为对象进行训练。以同一地区的银行和电商为例，由于两机构在特定地区的用户群体交集较大，因此可以对两机构的不同维度的用户特征进行聚合以增强模型能力。

（3）联邦迁移学习。联邦迁移学习则适用于数据集间样本和特征重合均较少的场景。在这样的场景中，不再对数据进行切分，而是利用迁移学习来弥补数据或标签的不足。以不同地区、不同行业机构之间进行联合建模为例，用户群体和特征维度的交集都很小，联邦迁移学习即用来针对性解决单边数据规模小、标签样本少的问题。

此外，许多企业推出的共享学习、知识联邦、联邦智能等一系列技术大多以联邦学习为基础进行改进，目的仍然是实现多方数据的机器学习。

（四）各类技术对比

由于技术路径的不同，各类隐私计算技术均有其更加适用的场景：多方安全计算技术不依赖硬件且具备较高的安全

性，但是仅支持一些相对简单的运算逻辑；可信执行环境技术具备更好的性能和算法适用性，但是对硬件有一定依赖；联邦学习技术则可以解决复杂的算法建模问题，但是性能存在一定瓶颈。相关技术的主要对比如表 1 所示。

表 1 隐私计算相关技术主要对比

技术	多方安全计算	可信执行环境	联邦学习
安全机制	基于密码学原理对数据加密	引入可信硬件	数据不动模型动
性能	低~中	高	高
通用性	高	中	低
高效性	中	中	低
准确性	高	高	中~高
可控性	高	中	高
保密性	高	中~高	中
可信方	不需要	需要	不需要
整体描述	开发难度大、关注度高使得性能提升迅速	易开发、性能佳，但需信任芯片厂商	综合运用各类密码学方法，主要针对机器学习

（信息通信技术与政策《隐私计算发展综述》等）

隐私计算主要应用场景

一、联合营销

随着营销业务进入智能时代，应用于营销的数据维度不断丰富，应用场景也不断增加。用户画像的数据往往是相互割裂的，只有通过整合多机构间、多维度的数据才能构建更立体的用户画像。利用隐私计算可以帮助机构在不输出原始数据的基础上共享各自的用户数据进行营销模型计算，根据建模结果制订营销策略，实现双赢的联合营销目的。

在构建营销模型中，通过隐私计算技术，对交互的标签、特征、梯度等数据进行等密码学处理，保证密文接收方或外部第三方无法恢复明文，直接基于密文进行计算并获得正确的计算结果，从而达到各参与方无需共享数据资源即可实现联合构建营销模型，从而进行精准营销。

在高价值用户识别中，可以利用隐私计算技术，通过联合统计、隐匿查询等方式将内部和外部数据进行安全融合，打通多方数据孤岛，利用外部数据更精准的对用户客群进行分类，识别高价值用户，制定更精准的营销策略。

银行机构利用隐私计算技术，可对运营商、政务、征信等数据实现应用场景所需的价值融合，从而为用户提供聚合金融服务。**保险公司**将用户基本信息、购买保险、出险赔付和电商、航旅等其他合作方的消费、出行、行为偏好等数据

进行安全融合。通过匿踪查询技术可信地获取客户的黑名单、消费能力、画像标签等信息，用于识别消费者的潜在风险等应用。**电信运营商**通过融合金融机构数据在共有的用户群中找到对理财产品、保险产品感兴趣的用户群，筛选找到更精准的目标用户进行营销，提升交叉销售效果，获取更多的新客。**互联网公司**利用自身拥有的大量用户行为信息和基础画像数据，与广告数据方拥有的深度转化链路数据（如付费信息）进行安全求交，并通过多方安全计算或联邦学习技术联合训练、建模、优化广告模型效果。在游戏、金融、教育、电商行业的广告应用案例中都能提升广告投放效果和用户体验。

二、联合风控

联合风控是隐私计算在金融领域的一个重要应用场景。一般而言，用户在本机构的金融业务数据难以满足金融风控的需求，但由于不同机构间数据分散、数据保护等原因，金融机构之间、金融机构与其他行业机构之间的数据融合壁垒较高，“数据孤岛”现象严重，提升了金融机构的风险识别难度，难以降低融资成本。

利用隐私计算技术，可以实现跨机构间数据价值的联合挖掘，更好地分析客户的综合情况，交叉验证交易真实性等业务背景，降低欺诈风险，从而综合提升风控能力。

在构建风控模型时，一方面可通过融合多个金融机构数据，解决单个金融机构样本量有限的问题，形成在相关场景

中的全局认知，提升模型精准度；另一方面，可以综合利用金融机构同其他行业数据，在各方原始特征不出域的前提下建立风控模型，形成对业务的多维度认识，提升风控质量。

在信息核验时，可通过隐私计算实现多方黑名单数据共享，对电诈、洗钱、骗贷等行为的黑名单用户进行匿踪识别，数据方不能获知查询的具体内容，提升客户背景调查的安全可信程度。

三、智慧医疗

在智慧医疗领域，医学研究、基因分析等工作非常依赖大量数据的积累，然而，医疗相关机构的这些数据割裂，分散在不同机构及业务系统内，机构间的数据难以互通互联，严重制约了临床科研成果的产出。

利用隐私计算技术，可实现在数据隐私保护下医学数据安全统计分析和医学模拟仿真和预判，从而进行跨机构的精准防疫、基因分析、临床医学研究等应用。

在疫情防控中，通过隐私计算保障个人数据的安全性，对高危人群进行筛选疫情传播仿真分析，通过防控筛查模型精准筛查高风险易感人群，构建潜在传染的关系网，结合病患信息，快速追溯传染路径方向和传播源。

在基因分析中，要依赖大量隐私数据，可利用隐私计算技术在原始基因数据不出库的基础上，实现基因数据的安全共享，进行全基因组的联合计算及关联分析等，在隐私保护的前提下挖掘多样的基因资源。

在临床医学研究中，在数据不出本地的情况下，可以实现分布式的统计分析算法，对数据进行联合建模、分析，从而获得临床科研的成果，例如临床研究可行性分析、大样本量队列研究、疾病预测模型、药物市场洞察等。应用隐私计算，可大幅提升医疗研究效率，加速科研成果转化。

四、电子政务

隐私计算技术为政务数据的开放提供了有效解决方案。在企业自有数据、第三方数据或政府共享数据都需要保护且不能离开本地节点的场景下，基于隐私计算进行数据安全利用。

在政务数据共享上，政务公共数据分布在各部门，通过隐私计算技术搭建政务公共数据密文开放共享交换平台，打通跨域数据的应用价值链，使得数据基于业务应用需要在各业务条线之间，安全地共享和流通，实现数据安全共享融合而不泄密。

在政务数据开放上，政府机构建设保护各方隐私安全的公共数据开放平台，使用隐私计算技术融合政府数据和社会、企业数据进行安全计算，联合统计，联合建模，实现数据融合价值，可以广泛应用在信用评估、服务选址、健康医疗、家政服务、旅游投资、营销设计等众多领域，让政府部门掌握的数据在安全保护前提下，最大限度造福社会。（隐私计算联盟《隐私计算白皮书（2021 年）》）

隐私计算的产业现状与发展趋势

一、隐私计算的产业现状

（一）国外隐私计算技术研究活跃，但商业化形态较为局限

从隐私计算本身的发展历程来看，国外企业布局隐私计算较早。早在 2008 年第一家专攻多方安全计算解决方案的技术厂商 Partisia 就已在丹麦成立，为商务合同、加密拍卖等场景提供安全方案。

科技巨头中，微软从 2011 年开始深入研究多方安全计算、谷歌在全球率先提出联邦学习的概念、Intel 打造 SGX 成为绝大部分可信执行环境实现方案的底座，均已成为各条技术路线主要的领路人。其他如 IBM 致力于将同态加密与云服务结合，帮助用户数据安全上云；Facebook 则是专攻基于隐私计算的机器学习。

创业公司中，Sharemind、Privitar 致力于搭建自研的多方安全计算平台；Duality 基于密码学开发的 SecurePlus 平台在新冠疫情中支撑了医学机构进行病毒基因分析。此外，AI 公司 Zama、区块链公司 Enigma 等均在推进多方安全计算、同态加密等方向的技术研发。

但从总体的应用场景来看，目前国外隐私计算项目中的很大一部分是面向区块链和加密虚拟货币的场景。如美国的

Unbound Tech和丹麦的Sepior均集中于将多方安全计算应用于分布式密钥管理领域。

（二）国内隐私计算技术产品蓬勃发展，形成一定优势

我国的隐私计算技术产业化在2018年后开始进入快速启动阶段，形成了互联网大厂、大数据公司、运营商、金融机构和金融科技企业、隐私计算初创企业为代表的五大类市场主要参与者。

互联网大厂方面，阿里巴巴、百度、腾讯、京东、蚂蚁等凭借自己在技术领域的积累，自2019年开始就纷纷推出了各自的隐私计算产品，形成了跨业务、多团队、强支撑的发展态势。**运营商方面**，三家电信运营商不仅在集团层面开始了隐私计算技术的选型与应用，天翼支付、电信云等子公司还自建平台服务于内部或其他机构的数据流通业务。**金融机构和金融科技企业方面**，国有银行的研究院或是事业部也均开始了隐私计算技术的研究工作，新心数科、神谱科技、平安科技、百融云创、度小满等金融科技类企业也将传统的数据建模、数据分析等业务拓展到基于联邦学习平台等的隐私计算服务中。**大数据公司方面**，同盾科技、星环科技、Talking Data、京信数科等代表性的大数据公司也快速布局基于隐私计算的数据流通产品或平台。**初创企业方面**，富数科技、华控清交、矩阵元、翼方健数、数牍科技、锘崴科技、光之树科技、零知识科技等一批专注于隐私计算产品化的初创企业不断涌现。

作为促进数据流通的关键技术，我国隐私计算技术产品日渐成熟，各领域应用场景加速落地，产业快速发展。面对隐私计算技术领域的国际竞争，我国已初具竞争优势。

从技术路线上来看，多方安全计算的复杂度高、开发难度大，以华控清交、富数科技、矩阵元等为代表的隐私计算初创企业多致力于此，专注于打造以底层多方安全计算技术为基础的数据流通基础设施。可信执行环境对于硬件的局限及国外芯片的强依赖，使得其在国内的产品选型相对较少，较集中于百度、阿里巴巴等互联网大厂和冲量在线、隔镜科技等初创企业。对于联邦学习，由于机器学习类应用需求突出，且有较成熟的开源社区为基础，开发难度相对轻松，因而运营商、金融科技公司等业务需求方大多专注于基于联邦学习的隐私计算产品化。

二、隐私计算未来的发展趋势

（一）从技术角度来看，隐私计算仍在加速成熟

软硬件协同优化提升隐私计算性能。硬件加速在隐私计算性能提升方面正在发挥越来越关键的作用，在算法不断优化的基础上，一些专用芯片和控件的使用将进一步提升隐私计算的性能。

逐步向大规模分布式计算迈进。2020 年以来，隐私计算逐渐成熟的一个表现就是分布式隐私计算的逐渐应用，为解决隐私计算在计算量方面的瓶颈提供了优秀实践。

提供工具化、模块化的服务能力。如何满足用户的个性

化与定制化需求、提升用户使用效率将成为产品形态趋同之下，技术提供者提升竞争力的关键。低代码甚至零代码开发、图形化拉拖拽替代编码和多版本轻量化部署等将成为产品升级优化的关键之一。

（二）从应用角度来看，隐私计算将加速与其他技术的协同以推进大规模落地应用

增强隐私计算与区块链等其他技术的不断协同。区块链与隐私计算的功能是天然互补的，借助区块链去中心化、不可篡改、公开透明的特性，将增强隐私计算任务的可验证性、可审计性，目前已成为诸多厂商的技术融合方向。此外，隐私计算与云计算的协同，将在支持云端数据存储、处理的同时加强任务过程中的安全与隐私控制；而隐私计算与人工智能的协同，将有力推进数据智能的应用和发展。

促进跨技术平台间的互联互通。隐私计算的目标在于促进多方数据之间的互联互通，但从应用现状看，不同技术路径之间的差异明显，而同一路径下不同产品的实现方案也相互独立，数据资源的互联互通只能基于不同的技术平台分块实现，增加了应用侧的使用成本。从长期发展来看，跨技术路径、跨系统平台之间的隐私计算技术工具的互联互通将成为广泛需求。

（三）从产业角度来看，隐私计算需与数据治理相互配合，共同促进数据要素的共享利用和价值释放

隐私计算有望成为数据流通的关键基础设施。随着近两

年国内隐私计算的技术产品快速落地，越来越多的行业客户开始在数据流通活动中实施部署相应的技术解决方案，隐私计算逐步推动着传统数据流通模式和流程的变革，当技术能力和应用模式越发成熟之时，隐私计算有望成为全社会数据流通网络的支撑型基础设施。

技术发展将随着数据合规要求的不断变化而演变。随着《数据安全法》《个人信息保护法》等法规逐步出台，金融、电信、互联网等各个行业将陆续出台数据流通相关的监管规定，虽然监管层面逐步认可并鼓励应用隐私计算以促进数据流通与权益保护之间的平衡，但法律法规一般不会明确给出技术应用的具体路径。随着攻击手段和破解技术的不断发展，数据合规要求将会不断更新，隐私计算技术发展也将随之动态演变。(《隐私计算白皮书(2021 年)》《隐私计算发展综述》等)

编辑部地址：南宁市体强路 18 号广西信息中心 1412 号房
联系电话：0771-6113592
电子邮箱：dsjyjs@gxi.gov.cn
网　　址：http://gxxxzx.gxzf.gov.cn/

扫描二维码获取
更多决策参考信息