CAICT 中国信通院

集智 蓝皮报告

联邦学习应用安全研究报告

（2023 年）

中国信息通信研究院安全研究所

2023年12月

版权声明

前　言

“数据孤岛”，是数据为“王”的时代的一个不可被忽视的现象，各组织机构的数据如同大洋上的岛屿，隔海相望、孤立无援。这种现象来源于组织机构对敏感数据域外共享的数据安全担忧，随着数据安全法律法规日趋严格，各组织机构难以承担数据泄露所带来的严重后果，使数据既出不去，也进不来。

在追求数据要素高效高质流通的当下，“数据孤岛”现象无疑是数据要素市场化建设进程中的障碍，于是“原始数据不出域，数据可用不可见”的新范式被提出，联邦学习也作为能够实现该范式的代表技术之一，得到了快速的发展。联邦学习技术可避免原始数据流出本地，转而通过“本地存储 + 分布式学习”的联合机器学习建模方式完成多方数据价值的释放，很好地解决了数据流通与数据安全之间的矛盾。

联邦学习作为能够打破“数据孤岛”的有力技术工具，为实现保障数据安全流通的初衷，仍然需要确保其自身的各属性的安全可靠。本研究报告着眼于联邦学习技术产品、系统、平台等形式的应用的安全，介绍了联邦学习应用的安全现状，分析了联邦学习在应用中面临的安全问题，并针对以上痛点问题，提出了联邦学习应用的未来发展建议。本报告的编写得到了不少业界同仁的大力支持，希望本报告能为社会各界深入了解联邦学习应用安全的现状与发展提供有价值的参考。

目 录

图 目 录

表 目 录

一、联邦学习概述

（一）背景

在数据价值被充分重视的大数据时代，数据流通成为了数据价值释放的重要步骤。2022 年 1 月 6 日国务院办公厅印发的《要素市场化配置综合改革试点总体方案》提出了要探索“原始数据不出域、数据可用不可见”的数据交易范式。联邦学习技术是实现该交易范式的典型代表技术之一，具有巨大的发展潜力。近年来，联邦学习的应用实践正在不断落地，其实用性已经得到了反复印证。

联邦学习作为数据流通领域的重要技术应用，一旦其出现安全问题，则保护数据的初衷将无法实现。因此，联邦学习的使用者对其安全性要求普遍较高。目前，联邦学习多被用于金融、医疗、政务等行业[1]，这些行业对数据安全及个人隐私保护有着严格要求，一旦联邦学习应用的安全性存疑，数据系统将面临着数据泄露的风险，并可能对企业或组织机构造成巨大损失。

近年来，联邦学习安全已经得到了学界的高度重视。从研究热度上看，在 2016 年至 2022 年的区间内，联邦学习安全方面的论文数量持续增加[1]，整体研究热度呈现上升的趋势。从研究广度上看，联邦学习安全方面的研究主题已经涵盖了恶意攻击、网络安全、隐私泄漏、容错、以及与其他隐私保护技术融合应用等多个领域[1]。联邦学习应用的安全风险发现和防御理论持续得到更新。

[1] AMiner.org，2023 全球联邦学习研究与应用趋势报告，2023

（二）联邦学习技术体系

联邦学习的概念最初在 2016 年由谷歌提出，经过一段时间的发展，有了比较明确的定义——“联邦学习是一种机器学习的形式，这种形式中多个实体（客户端）在中央服务器或服务提供商的协调下协作解决机器学习问题。每个客户端的原始数据都存储在本地，不进行交换或传输，并以聚合更新的方式达成学习目标”[2]。它的出现打破了传统机器学习的集中式数据训练模式，各组织、机构的原始数据不必流出本地，各自使用本地原始数据参与模型训练，通过迭代、聚合等过程最终得到全局模型。同时，各组织、机构、设备间的交互被以保护隐私为目标而精心设计，使得联邦学习应用可以在保护隐私的前提下，完成多方数据联合建模的任务。

联邦学习可以从以下三个维度进行分类，如表 1 所示。一是，**根据参与方的性质，联邦学习可划分为跨机构（cross-silo）联邦学习和跨设备(cross-device)联邦学习。**跨机构联邦学习指不同组织、机构之间，或者地理分离的数据中心之间的联邦学习，其特点是参与方数量少，各方的数据规模、质量等方面相对一致，技术实现相对简单;跨设备联邦学习指大量移动通信设备或物联网终端、边缘计算设备等之间的多方数据建模模式，其特点是参与方数量规模巨大，且各方的数据质量以及所处的网络、硬件环境相差较大，因此需考虑数据不平衡、设备性能不平衡、网络性能差等问题，实现难度较大。二是，**根据多方训练数据样本和特征空间的异**

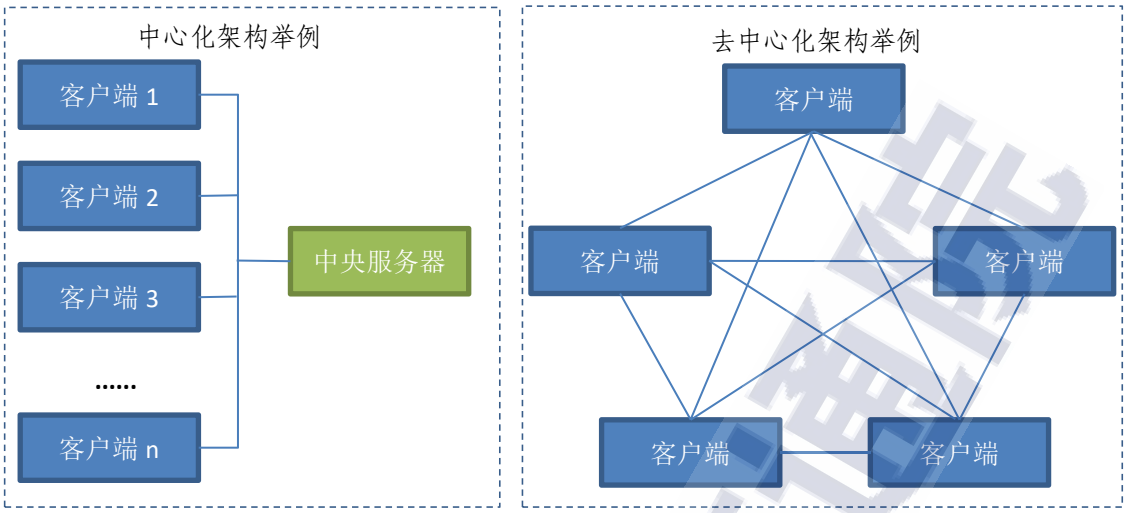
[2] Peter Kairouz，H. Brendan McMahan 等，Advances and Open Problems in Federated Learning，2019

同，联邦学习可划分为横向应用与纵向应用。在横向应用中，各参与方数据集的特征相同，而样本不同，其“横向”扩展了训练数据的样本空间。纵向应用则与横向应用相反，各参与方的数据拥有相同的样本空间，但在特征上各不相同，纵向应用实现了训练数据特征空间的“纵向”扩展。三是，**根据技术架构的不同，联邦学习可以划分成中心化架构和去中心化架构。**中心化架构中需要中央服务器作为协调方协助完成联邦学习过程，中央服务器及协调方程序通常部署于诚实的第三方中。去中心化架构中则没有处于中心地位、用以协调的第三方，如图 1 所示。

表 1 联邦学习应用分类

维度	参与方		样本和特征空间		技术架构	
分类	跨机构	跨设备	横向	纵向	中心化	去中心化架构

来源：中国信息通信研究院



来源：中国信息通信研究院

图 1 联邦学习架构

二、联邦学习应用概况

联邦学习应用已在我国多个行业落地实践，在此背景下，中国信息通信研究院安全研究所（以下简称安全所）于 2021 年至 2023 年间开展了联邦学习安全测评活动（以下简称“活动”），对 40 余款联邦学习产品进行了安全测评。同时期，中国信通院也针对 20 余项联邦学习产品或应用进行了安全性调研（以下简称“调研”）。本报告以本次“活动”与“调研”中积累的数据为基础，从联邦学习的应用情况、存在的安全风险、技术保障措施等方面分析了当下联邦学习技术的应用现状。

(一) 跨机构应用是国内联邦学习应用的主要形态

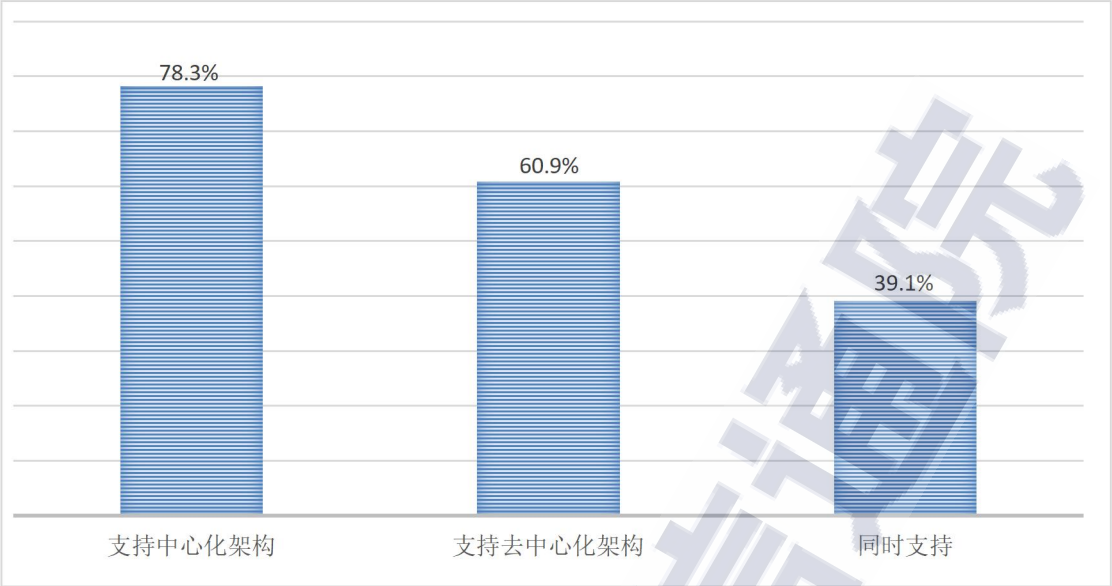
目前国内联邦学习应用需求主要来自金融、医疗、政务等行业[3]，实现的是跨“孤岛”的联合建模，即跨机构的联邦学习。跨机构联

[3] 中国信息通信研究院，数据价值释放与隐私保护计算应用研究报告，2021

邦学习应用的参与方数量少，相应的计算和通信压力相对较小，且成功的数据共享是各参与方的共同需求，在联合建模过程中各参与方为达成共同目标通常不会主动发起攻击行为，因而其对性能和安全性方面的技术要求相对较低，当下的联邦学习技术已可满足跨机构应用的大部分技术要求，这使得跨机构的联邦学习应用能够获得相对广泛的落地。相反，在跨设备的联邦学习应用中，参与方数量巨大，应用对计算效率、通信开销、安全防御等方面的要求更高。而现有技术在这些方面仍显不足，不能满足该类应用的高性能计算、低通信开销的要求，也难以应对恶意设备的投毒与攻击。技术上的不足使联邦学习实践难以向跨设备应用方面扩展。因此，跨机构应用成为了目前联邦学习应用最主要的应用形态。

(二) 中心化架构在联邦学习产品中占比最多

在技术架构方面，从总体上看中心化架构在各类联邦学习产品中占比最高，如图 2 所示。同时，联邦学习产品采用何种技术架构与其内置算法有关，部分联邦学习产品内置了多种算法以适应不同场景，因此出现了可同时支持中心化架构与去中心化架构的情况，如图 2。

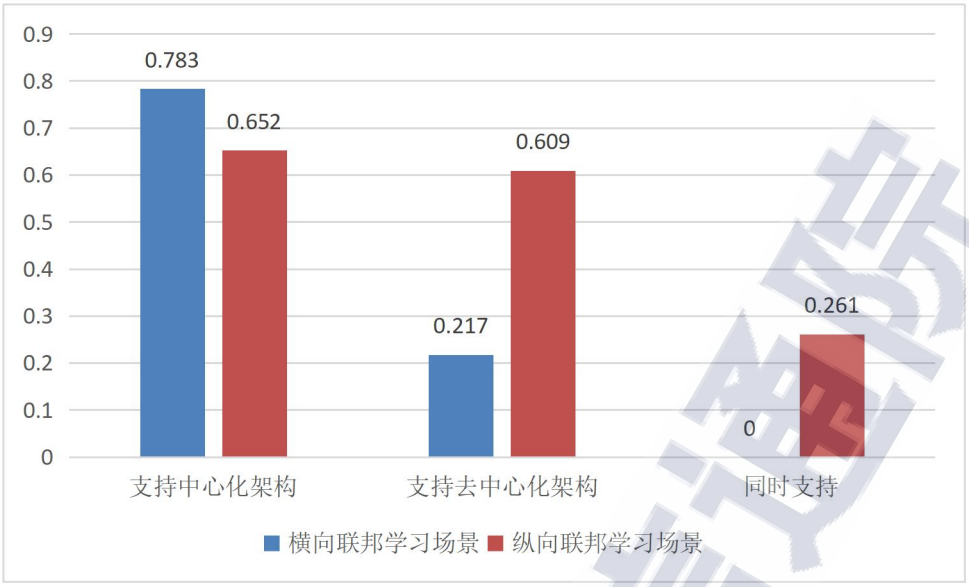


来源：中国信息通信研究院

图 2 联邦学习产品架构总体分布统计

在横向与纵向的联邦学习中，技术架构的分布有所不同。绝大部分横向联邦学习产品采用了中心化架构，如图 3。其原因在于大部分横向联邦学习算法需要协调方（中央服务器）的加入，以完成参数的聚合与分发操作,如 fedAvg 算法[4]。在纵向联邦学习方面，中心化架构与去中心化架构的分布占比近似，中心化架构占据微弱优势，如图 3。其原因是，纵向联邦学习中，协调方（中央服务器）所负责的聚合、分发等任务在部分算法中是不必要的（如 secureboost 算法[5]），同时一些产品采用安全多方计算技术保护参数交互，无需协调方的介入。如此导致了两种架构分布的平分秋色。综合以上数据，中心化架构在联邦学习产品各应用场景中均占比最多。

[4] H. B. McMahan 等，Federated learning of deep networks using model averaging，2016
[5] Kewei Cheng 等，Secureboost:A lossless federated learning framework，2019



来源：中国信息通信研究院

图 3 不同场景类别中的联邦学习产品架构分布统计

(三) 半诚实敌手环境是当下联邦学习主要的应用环境

从参与方对联邦学习协议的遵守程度看，联邦学习的参与方可被划分为诚实参与方、半诚实敌手、恶意敌手，相应地，根据参与方的诚实程度，将联邦学习的应用环境划分为诚实环境、半诚实敌手环境和恶意敌手环境，如表 2 所示。其中，诚实环境中的所有参与方均为诚实的；半诚实敌手环境中的参与方会诚实地遵守协议，但也有可能会被动接收或推测其他参与方的隐私信息，即该环境中存在半诚实的参与方；恶意敌手环境中的部分参与方会不遵守协议，而主动发起攻击，即该环境中存在恶意的参与方。在安全风险的应对措施上，恶意敌手环境下的应用对安全措施要求最高，半诚实敌手环境其次。“活动”与“调研”显示，所有统计对象均支持半诚实敌手环境，同时所有调研对象均不支持恶意敌手环境。造成这种

现状的原因，可从以下两个方面进行分析。

表 2 联邦学习应用的安全假设

	协议执行	行为特征	防御难度
恶意敌手环境	不遵守协议	主动攻击	高
半诚实敌手环境	遵守协议	隐私推理	中等
诚实环境	遵守协议	诚实	无需防御

来源：中国信息通信研究院

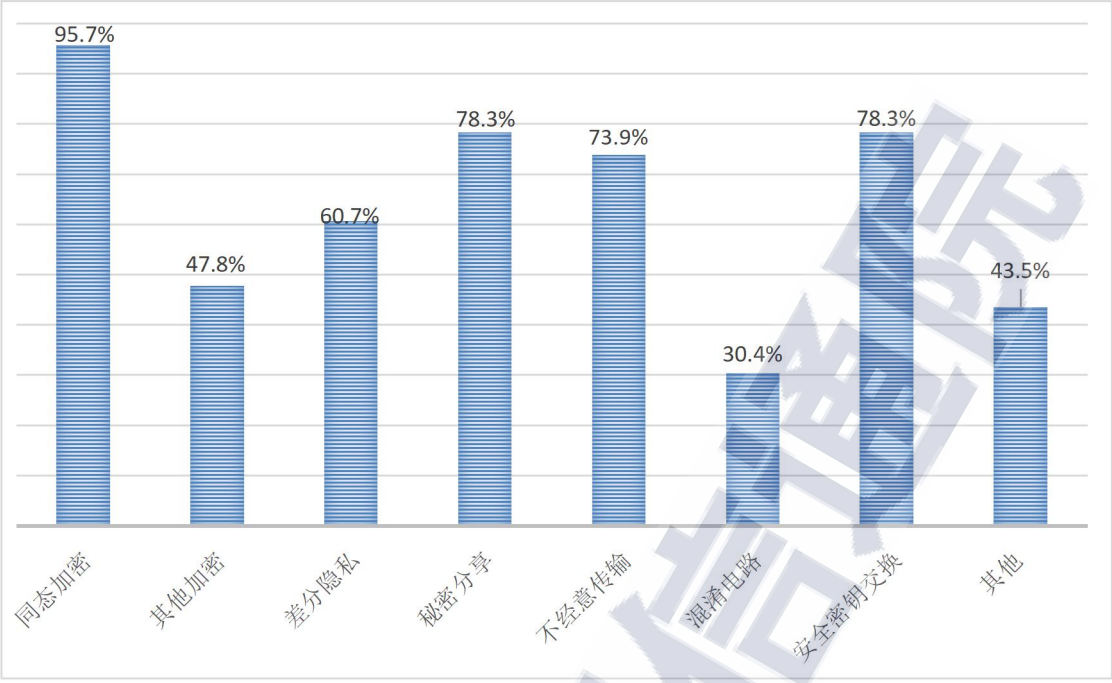
一方面，国内联邦学习应用的参与方之间多存在一定的信任基础，基于半诚实敌手假设的安全设计可以满足大部分联邦学习应用的安全需求。目前国内联邦学习应用多为跨机构的应用，各参与方之间为合作共赢的关系，顺利且准确地训练出最终结果符合联邦学习应用各参与方的利益，因此各参与方没有必要针对联邦学习过程发起攻击。然而，不排除各组织或机构在不影响联合训练结果的情况下“窥伺”其他参与方敏感数据的情况。多方交互中安全设计缺陷为这种“窥伺”行为提供了发生的可能，这使得“好奇”的参与方可直接获取或间接推理出其他参与方的敏感数据。这种“好奇”的行为虽然造成了数据泄露，但并不影响联邦学习过程的正常进行，因此其对应的环境安全假设为半诚实敌手环境。可见，参与方间的“合作”与“窥伺”是造成绝大多数联邦学习应用基于半诚实敌手环境进行设计开发的主要原因之一。

另一方面，恶意敌手模型下的安全研究成果尚不足以支撑联邦学习实际应用。一是学界尚未探明恶意敌手模型中的安全风险。在

恶意攻击方面的学术研究近些年呈现出上升趋势[1]，不断有新的安全风险被发现。在仍存在未知安全风险的背景下，很少会有用户愿意在存在恶意参与方的环境下部署联邦学习应用。二是恶意敌手环境下的已有安全研究成果仍有待验证。生产环境与实验室环境存在较大区别，实验室环境难以模拟大规模的参与方，各类的防御手段往往在计算性能、通信效率等方面具有局限性。因此，恶意敌手环境下的安全研究成果转化缓慢，当下联邦学习应用的设计开发仍以应对半诚实敌手为重点。

(四) 密码技术是当下联邦学习产品的主要安全保护技术

当前的联邦学习产品采取了多种安全技术手段来应对安全风险。图 4 展示了各类安全保护技术的使用占比情况，其中，绝大多数联邦学习产品使用了同态加密，使用率达到了 95%以上，在各类安全保护技术中处于第一梯队；秘密分享、不经意传输、安全密钥交换等安全多方计算技术处于第二梯队，使用率均达到 70%以上；差分隐私技术、其他加密（对称密码、非对称密码、Hash 函数）和混淆电路技术的使用率分别为 60.7%、47.8%、30.4%，其他安全保护技术，如布隆过滤器、梯度扰动等，合在一起使用率达到 43.5%。上述技术除差分隐私与“其他”技术外同属密码技术，可见密码技术是当前联邦学习产品的主流安全保护技术。



来源：中国信息通信研究院

图 4 联邦学习产品安全保护技术使用占比统计

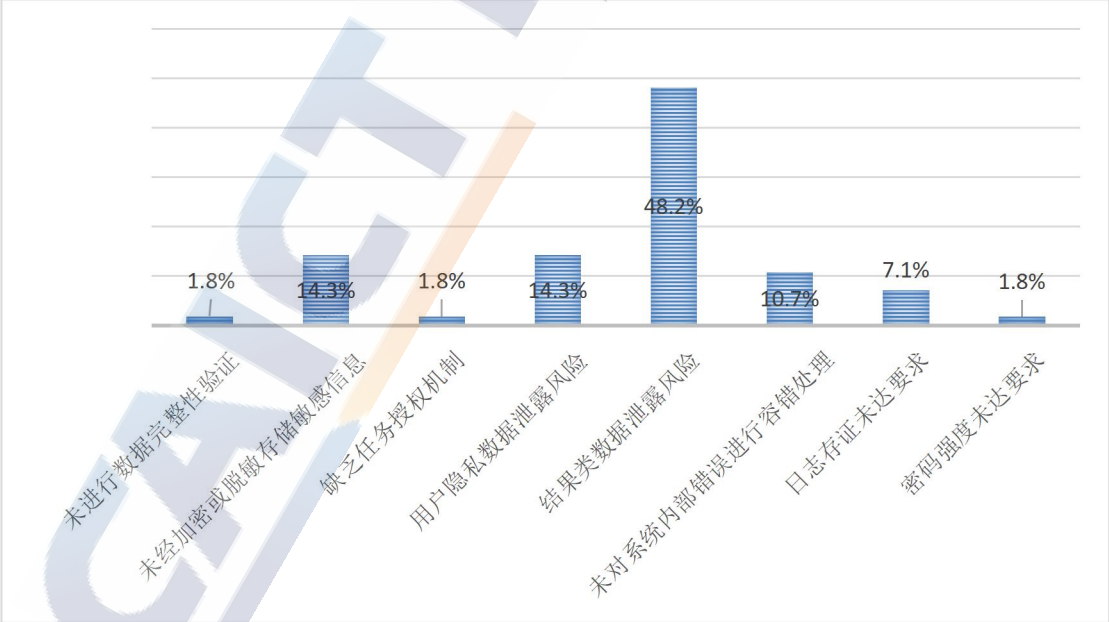
使用同态加密保护技术与使用安全多方计算技术保护是联邦学习安全保护的两条主要技术路线。一方面，同态加密主要应用于联邦学习的多方交互算法中，它的作用主要是使各方参数的运算在密态空间中进行，避免参数明文被他方获知，从而达到保护敏感数据的目的。另一方面，安全多方计算技术亦可实现对联邦学习过程的保护。除了负责完成匿踪查询、联合运算等隐私场景的需求外，安全多方计算技术也作为底层库在同一平台内为联邦学习的上层应用提供安全支撑。与同态加密类似，安全多方计算也保护了联邦学习过程的参数交互。不同之处在于，以安全多方计算作为主要安全保护技术的联邦学习产品通常不需要协调方的参与，适用于无可信第三方的场景下的多方数据安全流通共享。图 4 中，秘密分享、不经意传输、混淆电路同属安全多方计算技术，它们常与联邦学习集成

于同一大平台。

三、联邦学习应用安全现状与问题分析

(一) 数据泄露类风险是联邦学习产品最易出现的安全风险

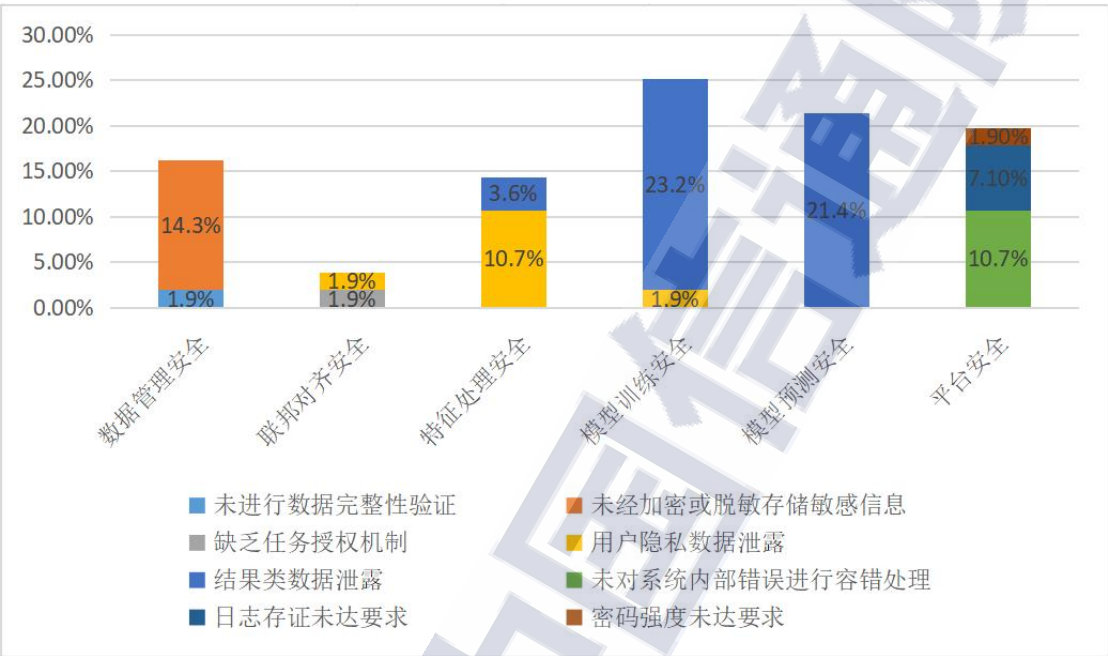
“活动”针对半诚实敌手环境下的安全风险进行了检测，发现目前的联邦学习产品主要存在未进行数据完整性验证、存储敏感信息未经加密或脱敏、缺乏任务授权机制、泄露用户隐私数据、泄露结果类数据、未对系统内部错误进行容错处理、日志存储未达要求、密码强度未达要求等安全风险。图 5 展示了联邦学习产品各类别安全风险的占比，存储敏感信息未经加密或脱敏、用户隐私数据泄露、结果类数据泄露等数据泄露类风险在各类安全风险中占比最高，其中结果类数据泄露风险占比最为突出，达到了 48.2%。



来源：中国信息通信研究院

图 5 联邦学习产品安全风险占比统计（半诚实环境）

图 6 从联邦学习流程的维度，统计了全流程中每个环节的安全风险分布，数据泄露类风险在联邦学习流程中的模型训练环节与模型预测环节占比最高。



来源：中国信息通信研究院

图 6 联邦学习产品安全风险分布统计（半诚实环境）

从联邦学习的运算过程看，多方交互协议的设计不当是导致数据泄漏风险多发的原因之一。在联邦学习的运算过程中，各参与方仅拥有联合建模所需的“部分”训练数据，这些原始的训练数据在经过本地计算后产生了中间参数，各参与方以交换中间参数的方式进行协作，最终完成整个建模过程。其中，多方交互协议约定了各方参数交互的内容和时序。在内容上，协议规定了参数在各参与方本地的形成过程，形成的参数可能是密钥、随机数，甚至是经过加密的原始数据。若这些参数没有经过一定程度的安全处理，半诚实

的参与方往往会借此推导出其他参与方的敏感数据，从而造成数据泄露。在时序上，协议规定了各类参数在各参与方之间交互的时机和顺序。各类参数之间存在着一定的依赖关系，前一个参数往往是后一个参数形成的计算条件，时序的错误一方面可能会造成联邦协作过程无法继续，另一方面，混乱的密钥分配、滞后的安全保护措施等都会带来安全风险。因此，多方交互过程是出现数据泄露安全风险的重灾区，严谨的多方交互协议对确保联邦学习安全起着至关重要的作用。

从联邦学习的结果保护方面看，对结果聚合的疏于防护也是引发数据泄露风险的重要原因。在部分联邦学习应用中，需要由协调方对各参与方在本地产生的中间结果进行聚合运算，中间结果包括每轮迭代的模型参数、各参与方的本地预测结果参数等。这些中间结果属于敏感数据，倘若没有一定措施对其进行保护，中间结果与最终结果等敏感数据会直接暴露给部署了协调方程序的第三方，半诚实的第三方可能会利用中间结果，通过推导的方式获知用户的原始数据。在模型训练和模型预测环节存在着较多的结果聚合场景，这使得结果类数据泄露的风险集中出现在这两个环节中。例如，在 FedAvg 算法[4]的横向联邦学习的训练环节中，各参与方会将每一轮迭代的梯度或模型参数结果上传到第三方进行加和平均；再如，在纵向联邦学习的模型预测环节中，部分应用将各参与方本地的预测结果汇总至第三方聚合以获得最终预测结果。上述结果聚合过程中，若未采取同态加密等措施，那么便可能出现结果类数据暴露给第三

方的问题。因此如何安全地进行结果聚合应是联邦学习应用关注的重点问题之一。

(二) 联邦学习应用安全风险的隐蔽性高

联邦学习应用的安全风险不仅类型多、分布广，而且这些安全风险也具备隐蔽性高的特点，其原因可以从下述两方面分析。

一方面，联邦学习应用中多方交互的高复杂性造成了其安全风险不易被发现。联邦学习比集中式机器学习增加了分布式过程，其针对集中式机器学习的各类算法进行了拆分改造，并且将拆分后的计算参数分布到各参与方进行计算，同时通过多方交互的方式来满足各拆分后参数的计算所需条件，从而完成计算。拆分后的计算参数在各参与方之间进行流转，参与方的当前状态亦随时序的变化而变化，在特定时刻下，不易确定参与方是否可根据其自身状态推导出其他参与方的敏感数据，因此其中的安全风险具有很强的隐蔽性。例如，在联邦学习的特征处理过程中，非标签拥有方可通过看似“无害”的 woe/iv 明文结合自身已有数据，推导出标签拥有方的敏感数据，该风险隐藏在复杂的算法逻辑和数学推导之下，使得其难以被发现。因此联邦交互过程的复杂性为安全风险提供了“掩护”。

另一方面，联邦学习应用不易进行安全审计，也是其安全风险隐蔽性高的重要原因。联邦学习应用中大量采用了密码技术，这些加密措施保护了联邦学习应用的多方交互过程的同时，也使流转的参数经历了复杂的数学变形，改变了其原有形态。在这种情况下，

使用常规审计手段难以复原联邦学习过程的原貌，从而造成了联邦学习应用安全审计的困难。同时，联邦学习是跨分布式计算、机器学习、密码学等多领域的融合应用，对安全审计人员的知识背景要求极高。目前企业安全部门中，兼具这些知识背景的安全审计人员较少，开展专业的联邦学习审计活动比较困难。因此，联邦学习过程是否按照“原始设计”无误进行、操作人员是否执行了不安全的操作、参与方是否存在恶意攻击的行为等等安全问题，变得不易发现和评估。

(三) 协调方的存在为联邦学习应用带来了安全方面的不确定因素

中心化架构在当下的众多联邦学习产品中占据较大比例，采用该架构的应用使用协调方（中央服务器）协助完成联邦学习的全过程。协调方的引入为联邦学习应用带来了安全方面的不确定性，体现在如下三个方面。

一是将敏感数据暴露给第三方是潜在的安全隐患。联邦学习过程中，原则上由第三方担任协调方角色。如此，作为协调方的第三方便承担着密钥分发、参数聚合等任务，其在服务的过程中可能会接触到密钥、模型参数等敏感数据。一旦多方交互协议设计不当或第三方存在不诚实行为，极易出现第三方获取敏感数据的情况。同时，也存在着不诚实第三方与某些参与方利用自身参数合谋获取隐私数据的情况。因此，让第三方接触敏感数据是潜在的安全风险。

二是目前现实中难以找到诚实可信的第三方担任协调方角色。绝对诚实的第三方，不会窃取隐私数据，是理想状态下协调方角色的担当。然而，现实中尚无诚实第三方的认定标准，难以对第三方进行有效的可信认定。因此，盲目的引入第三方，一方面可能会出现数据泄漏等安全风险，另一方面也可能会带来安全合规方面的问题。

三是协调方程序的部署易受到机构间合作关系的影响而导致潜在安全风险。在机构间的联邦学习合作中，可能会出现话语权不对等的问题。譬如，机构体量、数据量、数据价值等方面的差异，导致在联邦学习合作中出现优势方。部分优势方为避免引入第三方而产生的安全隐患，而选择将协调方程序部署到自身一侧。如此，优势方同时拥有了参与方程序与协调方程序，若优势方是不诚实或半诚实的，则其可能会结合两种角色程序所产生的中间参数，利用数据推导等手段，窃取其他方的敏感数据。这种部署方式保证了优势方在一定程度上的数据安全，却使其他参与方存在敏感数据泄露的安全风险。因此，机构间的合作关系不对等可能会成为联邦学习应用安全风险出现的诱因。

(四) 联邦学习应用的安全保护强度与性能要求在一定程度上相互制约

联邦学习在应用中通常会使用密码等技术手段对多方交互的过程进行保护，而这些保护措施的强度往往与实际中的性能要求存在

着相互牵制的关系。

一方面，高强度的安全保护往往制约了联邦学习应用的性能。在通信开销上，大部分联邦学习应用对多方交互中产生的中间参数进行了加密，这使多方交互过程相对于明文交互，增加了密钥分发、加密、解密等过程，从而使联邦学习整体的通信成本增加；同时，在密码体系中，安全参数代表着加密的安全保护强度，安全参数越大，安全保护强度越强，在通常情况下，密文的体量不仅较明文会有明显增加，而且会随着安全参数的增大而增大。以训练 DNN（深度神经网络）模型为例，生产环境的 DNN 模型往往拥有数百万个参数，且该类模型在联邦学习中的通信成本会随着迭代伦次的增加而持续增大，因而在此基础上的密文通信将会是巨大的开销。在计算效率上，密码技术带来了加密、解密、密态运算等过程，而这些过程包含了大量的取模和幂等高复杂度计算，开销相对较大，并且安全参数的增大会使此开销进一步增加。因此安全保护强度会对联邦学习应用的性能有一定的影响。

另一方面，可用性要求迫使联邦学习应用在实际应用中降低部分安全配置。某些情况下，高级别的安全配置将使联邦学习应用的性能下降，甚至陷入“不可用”的状态，设计者不得不降低安全配置，以获得安全性与性能的平衡。以纵向联邦学习中的联邦数据对齐环节为例，绝大多数联邦学习应用为避免重复运算，在整体设计上进行了解耦，将数据对齐环节设计成独立的功能模块，使其能够进行独立运算、独立存储。而纵向场景下的联邦数据对齐实质上是

对各参与方数据集中样本 ID 的联合求交。如此，在运算结束后，各参与方均可获得样本 ID 的交集，并可根据该 ID 交集及自身的已有数据推导出对方的用户群体数据，从而造成用户群体数据的暴露。反之，若将数据对齐环节与模型训练环节集成起来，样本交集以密文中间参数的形式继续加入运算，即可避免上述的数据泄露风险。然而，此举一方面使数据对齐的结果难以被重复利用，另一方面对规模巨大的样本交集进行密态计算，将极大增加计算性能开销，甚至可能使功能陷于瘫痪。因此，在实际应用中，当技术手段无法解决安全与性能的矛盾时，牺牲部分安全性以换取性能上的“可用”便成了无奈之举。安全与性能呈现出相互制约的关系，如何解决二者的矛盾仍是当下联邦学习应用面临的严峻问题。

(五) 联邦学习应用安全相关标准尚未健全

联邦学习技术多应用于数据生命周期的共享阶段，目前联邦学习应用安全相关的国家/行业标准较少，行业亟需标准化文件的规范和指导。

当下各类联邦学习应用对安全尺度的把控不统一。对于联邦学习应用的设计者和使用者而言，是否可以降低安全尺度以提高性能、在特定场景下该采取何种安全强度，这些均是联邦学习设计和使用过程中的关键问题，这些问题答案的不统一，使目前各类联邦学习产品、系统、平台等形式的应用各自为战，为数据应用带来极大的安全隐患。因此，需要有标准化文件对联邦学习应用的各环节的安

全尺度进行规范指导。

同时，**当下已发布的安全标准主要集中于产品的安全要求上，联邦学习应用部署、运营阶段的安全标准较少。**联邦学习应用的部署、运营不当，会使数据安全风险出现于管理环节。例如，在无身份、权限的管理机制的情况下，操作人员可能会接触到其不应访问到的数据。因此，健全、加速联邦学习应用安全的标准化建设十分必要。

四、联邦学习应用安全的学界研究现状

(一) 偏重于恶意安全环境下的安全研究

相较于联邦学习在工业界的应用，学界的研究更具有前瞻性。在研究对象上，工业界的实践主要以 toB 的联邦学习应用为主，学界则更加关注性能、安全等方面技术要求更高的 toC 联邦学习应用。相应的，在安全方面，研究也更偏重于与 toC 场景安全要求相适配的恶意环境下的安全攻防。目前，安全研究可大致分为三大类：投毒攻击、对抗攻击与合谋攻击的相关研究。

投毒攻击是指联邦学习中的恶意的参与方对训练集进行恶意操纵，例如向训练集中插入精心制作的恶意样本，从而改变训练数据的分布，以达到破坏训练过程或结果的目的，其种类包括了数据投毒、模型投毒、后门投毒等。在存在多方训练集参与的联邦学习中，相较集中式机器学习，攻击面更大，威胁更甚。学界提出了基于数据清洗、异常性探测等防御方法，以针对投毒攻击进行防御。这些

基于攻击检测的方法存在一定的误报或漏报，提高攻击发现的准确率是未来研究的重点方向。

对抗攻击是指联邦学习中的恶意参与方对输入样本添加一些难以察觉的细微干扰，导致全局模型以高置信度给出一个错误的输出，通过这种干扰，恶意参与方可在联邦学习中发动无形的对抗攻击，其他参与方无法提前预判对抗攻击的时机和方式。对抗攻击具有较强的隐蔽性，对此，学界提出了防御蒸馏、梯度正则化等防御方法。这些方法旨在提高模型的泛化能力，取得了良好的防御效果。

合谋攻击指的是多个联邦学习的恶意参与方相互掩饰联合破坏联邦学习的过程与结果。恶意参与方可利用自身资源为其他恶意参与方提供攻击便利，从而对联邦学习计算、通信等多个层面进行破坏，具有较强的破化性。为应对该攻击，学界提出了收敛异常检测、指纹嵌入等防御手段。这些防御手段多偏重于对合谋攻击的发现和攻击者的识别，如何提高识别的准确率并实时阻断依然是学界面临的严峻挑战。

(二) 如何优化性能是热门研究方向

联邦学习应用的性能与安全存在着相互制约的关系，性能优化技术的进步可以为更强的安全措施提供计算资源上的冗余，因此如何优化联邦学习的性能一直是学界的研究热点。

影响联邦学习性能的主要因素包括通信性能和运算性能，其中，通信效率低是现阶段制约联邦学习发展最严重的问题。学界主要以

如下几个思路开展研究，一是通过减少全局模型训练的迭代轮数，从而优化通信的效率。例如，进行更多的本地模型参数更新等，这种方式减少了各客户端节点向中央服务器通信的次数，但未解决单次迭代导致的信道拥堵问题。二是使用参数的压缩技术，例如对梯度、模型参数进行量化与稀疏化等，这种方式减少交互信息的长度，但可能会损失一定的训练精度。三是采用分散式的拓扑结构，例如，去中心化拓扑、分层式拓扑等，这种方式能够减少客户端节点与中央服务器的高交互开销，解决了各节点与中央服务器的信道拥堵问题。

五、联邦学习应用安全发展建议

(一) 加速联邦学习应用安全的标准化建设

在联邦学习应用于各行业已有初步实践的背景下，尽快健全联邦学习应用安全相关标准。

一是出台基于行业细分的联邦学习应用安全技术类标准。当下联邦学习技术已在金融、电信、互联网、能源等多个行业得到了应用，一方面针对联邦学习应用相对较成熟的行业，出台本行业的技术安全应用的要求和规范类标准，帮助行业内企业把控联邦学习应用的安全尺度；另一方面，针对尚处于联邦学习应用探索阶段的行业，出台本行业的联邦学习安全方面的指南和方法类标准文件，以指导行业内企业搭建具备该行业特征的安全联邦学习应用。

二是出台联邦学习应用安全部署、运营的规范类标准。以标准

化的方式来解决联邦学习应用在部署、运营环节所面临的诸多问题，例如，联邦学习的协调方程序应在何处部署、如何确保担任协调方角色的第三方诚实可信、系统平台的用户域与运维域该如何进行隔离、产品应用在物理机环境或云环境下部署的安全要求是否相同等等。为进一步推动联邦学习技术的深化落地提供有力保障。

三是建立联邦学习应用的安全评估制度。一方面，针对联邦学习产品，建立产品安全性评估制度，检验其技术实现是否合乎产品的安全性标准。另一方面，针对联邦学习应用方的部署环境安全、运营管理安全等方面，建立相应的评估制度，检验联邦学习在部署、运营环节是否存在安全风险，评估结果存档并作为监督检查的依据。

(二) 加强联邦学习应用安全的研究

联邦学习是新兴技术，其应用虽已得到初步实践，但安全方面仍存在待优化的问题，相关研究仍然具有高价值。

一是加强联邦学习应用的审计技术研究。一方面是针对联邦学习应用的自动化审计技术进行探索，从而解决现实中审计相关工作严重依赖人工，审计效率低下，且审计过程受人的主观性影响较大的问题。另一方面，要研制通用性的联邦学习应用审计工具，解决因各类联邦学习应用的技术异构性导致普通审计手段不能适配各类异构应用的问题。审计技术的突破将为企业自查、第三方检验、政府监管扫除技术障碍，从而推动联邦学习应用的健康发展。

二是探索联邦学习与新技术的融合应用。推进联邦学习与新型

硬件、算力网络等新型算力技术的结合，有利于提高联邦学习的性能表现，从而缓解联邦学习应用中性能与安全的矛盾，进一步提高联邦学习的实用性。因此研究联邦学习与新技术的融合应用具有重大现实意义。

三是加强恶意环境下联邦学习应用的安全攻防的研究。恶意环境下安全研究尚不足以支撑当下的现实应用，是跨设备联邦学习应用难以落地的原因之一。在此方面的突破将有助于拓展联邦学习的应用场景，使联邦学习应用下沉到对安全要求更高的 toC 场景，对隐私计算的大规模应用有着促进的作用。

(三) 推动联邦学习应用安全的基础设施建设

加强基础设施建设对联邦学习应用的夯实地基、稳固发展有着重要意义。

一是推动区块链基础设施在联邦学习中的应用。区块链作为一种去中心化信息基础设施，可为联邦学习应用提供安全的身份管理、多方数据的一致性保障，其内容不可篡改特性也可为联邦学习应用提供优良的安全存证、审计平台。

二是推动建立可靠的“第三方”设施。多数联邦学习应用依赖协调方程序，这些程序需要部署于第三方。但在很多情况下各机构出于对数据安全的担忧，而找不到合适的第三方。此时由行业协会、高校、科研院所、数交所等中立或权威机构按照法律法规建立可靠的“第三方”设施，则在一定程度上可缓解联邦学习参与方的担忧，

减少联邦学习应用在发展上的障碍。

编制说明

在本报告的研制过程中，得到了以下单位的支持协助，在此表示感谢（以企业名称首字母先后为序）：北京百度网讯科技有限公司、北京数牍科技有限公司、广州大学网络空间先进技术研究院、恒安嘉新（北京）科技股份公司、哈工大(深圳)-奇安信数据安全研究院、华控清交信息科技(北京)有限公司、黑龙江大学数据科学与技术学院、华中科技大学、杭州金智塔科技有限公司、京东科技集团有限公司、上海富数科技有限公司、陕西数盾慧安数据科技有限公司、苏州黑云智能科技有限公司、深圳市洞见智慧科技有限公司、深圳致星科技有限公司、平安银行股份有限公司、同盾科技有限公司、天翼云科技有限公司、星环信息科技（上海）股份有限公司、中国电信股份有限公司研究院、中国工商银行业务研发中心、中国联合网络通信有限公司研究院。

本报告提供给媒体、公众和相关政府及行业机构作为联邦学习应用安全的研究资料，请相关单位酌情使用。如若本报告阐述之状况、数据与其他机构研究结果有差异，请使用方自行辨别，中国信息通信研究院安全研究所不承担与此相关的一切法律责任。因研究团队能力有限，报告仅作为参考研究，多有纰漏与不足之处，欢迎各领导专家批评指正，我们持续改进。

中国信息通信研究院 安全研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62308087

传真：010-62300264

网址：www.caict.ac.cn