《计算机科学、工程与信息技术国际科学研究期刊》ISSN：2456-3307 在

OPEN ACCESS



云端隐私保护数据处理：从同态加密到联邦分析

Gaurav Sarraf、Vibhor Pal美国独立研究人员

文章信息
第8卷第2期 页码：735- 749

出版期：2022年3月至4月

文章历史
录用：2022年4月10日
发表：2022年4月30日

摘要

隐私保护数据处理是指能够在保证机密性的前提下，对敏感数据进行计算和分析的方法与模型。随着云计算以及依赖数据的应用不断发展，保护个人、金融和医疗保健信息的需求日益增加。传统的集中式数据处理方法会使敏感数据面临泄露风险，因此亟须采用去中心化且安全的数据处理方法。本文详细回顾了云平台中的隐私保护机制，例如差分隐私等统计方法，以及同态加密等密码学解决方案。本文还讨论了联邦分析和联邦学习这两种分布式学习框架，并回顾其原理、应用、优势和局限性，以及它们在医疗保健、金融、物联网和工业场景中的作用。通过对比分析，评估了安全性、效率、可扩展性和准确性之间的权衡，并探讨了新兴的混合框架，以实现更好的隐私保护。本文还讨论了计算开销、隐私-效用权衡、标准化、对抗性威胁和云集成等关键问题。本文详细考察了近期云计算中的隐私保护方法，并为学者和从业者提供了有关安全、有效的数据处理解决方案的重要信息。

关键词——隐私保护、同态加密、安全多方计算、差分隐私、联邦分析、云计算、数据安全、混合隐私框架。

引言

数据隐私保护已成为学术界和商业领域亟待解决的问题，因为数据驱动技术正呈指数级增长 [1]。每天都有大量敏感信息产生并接受处理，因此数据安全至关重要。传统的集中式存储和基于计算的数据处理方案通常会将数据汇集到中央服务器上。然而，通过这种集中式系统传输和存储数据，很可能会侵犯隐私。当集中式存储或传输过程遭到破坏或攻击，导致敏感信息面临泄露风险时，个人和组织都将面临极大的危险。

版权：© 作者、出版方及许可方 Technoscience Academy。本文为开放获取文章，依据知识共享署名-非商业性使用许可协议的条款发布。该协议允许在正确引用原作品的前提下，不受限制地进行非商业性使用、传播和复制。

联邦学习已成为解决这些隐私问题的一种日益流行的去中心化数据处理方法 [2][3]。联邦学习能够通过多个设备或实体协作来训练模型，而无需共享原始数据，从而在一定程度上降低隐私风险。尽管联邦学习能够解决集中式系统中的某些隐私问题，但它仍然存在自身的漏洞。在联邦系统中传输模型更新，可能会在无意间泄露有关本地数据集的信息。这促使研究人员致力于改进联邦学习系统的隐私保护。

正因如此，同态加密（HE）已被证明是保护联邦学习隐私的一种有效技术。在同态加密中，无需解密即可对加密数据进行计算[4]，，因此在整个数据处理链中都能保护敏感数据的安全。凭借这一特性，同态加密成为保护联邦学习隐私的首选方案。

论文结构

本文结构如下：第二节介绍背景与基础知识，第三节阐述隐私保护方法，第四节讨论协作框架与新兴框架，第五节探讨应用并进行比较分析，第六节回顾文献，第七节总结并提出主要挑战。

背景与基础知识

云计算提供按需且可扩展的计算资源和存储服务，使组织能够轻松处理大量数据。然而，这种集中式架构带来了显著的隐私风险，因为个人健康记录或财务信息等敏感数据很可能存储并处理在第三方服务器上。一些关键问题包括未经授权的访问、数据泄露、内部人员威胁以及个人信息滥用。法规遵从也是这一过程的另一项障碍，而 GDPR、HIPAA 和 CCPA 等框架对数据的处理、存储和共享方式提出了严格要求，这使问题更加复杂。此外，云多租户增加了客户之间数据泄露的风险，而将计算任务外包给第三方也让数据机密性和控制权受到质疑。云服务信任取决于能否解决这些隐私问题，而这也为实施加密、差分隐私和联邦分析等隐私保护措施铺平了道路。

云计算的基本特征

根据 NISO 的定义，下面的表I概述了云模型的五个主要组成部分 [5]。表I. 云计算的基本特征

技术	描述
同态加密（同态加密）	能够直接处理加密数据，解密后得到的结果与明文运算的结果完全相同。实现隐私保护分析、安全机器学习和外包计算。加密有两种类型：全同态和部分同态。
安全多方-参与方计算（安全多方计算）	通过秘密输入协作计算函数，可使多个参与方保持匿名。不经意传输、秘密共享和姚氏混淆电路是最常用的几种协议，有助于联邦计算以及协作分析。

差分 隐私（差分隐私）	通过添加受控噪声，隐藏特定数据点的方法 查询结果或模型。包括集中式差分隐私（集中式差分隐私）和本地差分隐私（本地差分隐私）。 应用包括基于云的分析、机器学习、物联网和医疗保健。
联邦 分析 / 联邦 学习（联邦分析/联邦学习）	所有数据都保留在本地系统中；仅更新后的模型或聚合结果 被传输。可与同态加密或差分隐私结合使用，以增强隐私保护。 在组织间的医疗保健、金融和协作式机器学习领域的应用。
可信执行 环境 （TEE）	内置于硬件中的安全飞地（如 Intel SGX 和 AMD SEV）会加密 信息在处理过程中。确保安全的云端处理，而不会暴露 明文给服务提供商。
Data 匿名化与 假名化	k-匿名性、l-多样性和 t-接近性等方法可以对敏感 数据进行去标识化。实施这些方法是为了支持公共数据集共享和满足监管 合规性。
访问控制与 身份 管理	基于角色或属性的访问控制旨在将访问权限限制为仅授予获准人员， 使其能够访问敏感信息。密码学和审计是 经常一起使用。
安全 Data 聚合	安全地整合多个数据源，常用于物联网和传感器网络。 在计算全局统计数据的同时，防止个体贡献泄露。

云架构

“云架构”是对云计算环境结构和布局的描述，包括计算资源、应用和服务在各个云服务提供商之间的分配。云架构对于云计算环境的可扩展性、可靠性和安全性至关重要。要理解并有效利用云计算，关键在于认识到它本质上是一种同时服务 [6]。对于数据、信息和知识等宝贵资源而言，安全存储、防范自然灾害，以及与用户之间便捷、安全地传输，都是至关重要的要求。

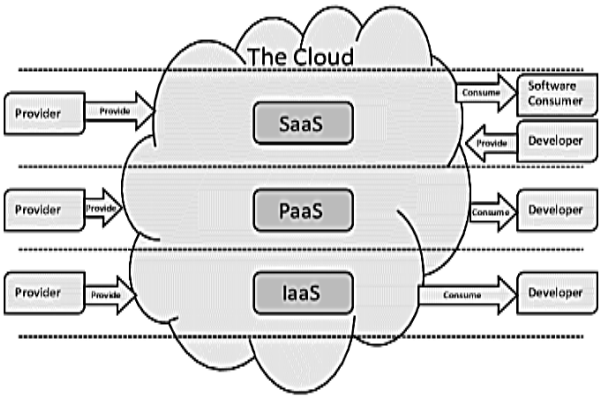


图 1. 云计算架构

计算、存储、数据库管理、应用服务和大规模计算，只是众多从云计算中受益的 IT 流程中的一部分。云计算方面的专业知识在很大程度上依赖于其架构（图 1），该架构描述了系统的组件、子系统和总体组织方式 [7]。云计算有许多不同的实现方法，每种方法都有其自身的一系列优点和

缺点。这些设计为如何构建和实现基于云的应用设计提供了指导。在大多数情况下，云架构由以下几个基本部分组成：

1) 云服务模型
   - 基础设施即服务：提供在线虚拟化服务，例如云计算、存储和网络。
   - 平台即服务：使用户能够专注于应用创建、发布和管理，而不必受底层基础设施限制。
   - 软件即服务：通过互联网分发软件程序，使用户免于亲自安装、维护和监控这些程序。

2) 部署模型
   - 公有云：所有希望使用这些设施的个人都可以通过开放的互联网访问它们。
   - 私有云：私有网络通常用于组织内部，在提供服务时能够提供更多控制权和个性化设置[8]。
   - 混合云：结合公有云和私有云的特性，促进应用和数据共享。

3) 云服务
   - 身份与访问管理（IAM）：有效管理用户身份和权限，可保护云资源免遭入侵。
   - 数据库服务： 托管数据库支持：支持托管数据库，并提供安全且可扩展的数据存储。

4) 编排与管理
   - 自动化： 提供用于自动执行日常任务的工具和资源，减少人工干预并提高产能。
   - 编排： 通过管理和维护多种云服务，确保沟通顺畅。

5) 可扩展性与弹性
   - 可扩展性：系统通过增加资源来处理不断增长的工作负载的能力，称为其可扩展性。
   - 弹性：自动调整资源以满足需求的能力。

云架构具有动态性和适应性，并不断发展以适应新技术。企业在设计云基础设施时，会根据自身在效率、效能、成本效益和法规遵从方面的需求进行考量 [9]。请注意，目前基于云的应用尚无通用架构，甚至也没有基本的设计原则。云计算现有的一项目标是建立受控且可扩展的云基础设施。这种设计必须能够适应服务级别的变化，并以市场为导向，从而应对云资源供需的波动。

了解法规遵从与云数据保护

法规遵从是云数据保护中至关重要的流程，因为它可以验证组织是否遵守了旨在保护数据隐私、安全性和完整性的法律与标准。人们对数据泄露、网络攻击和隐私侵犯的担忧日益加剧，监管压力也随之增加[10]。不合规可能会使未能遵守合规标准的组织遭受法律、财务和声誉损失。因此，鉴于云数据保护方面存在特定的监管要求，组织，尤其是从事处理敏感信息的行业的组织，必须了解并满足这些要求。

法规遵从是指遵守有关信息收集、存储、处理和传播的法律、道德及行业特定法规。在云计算领域，合规意味着所有存储在云端的数据都按照所有适用的规则和法规进行处理。这些法律可能因地区、行业性质以及数据类型而异。合规还取决于是否落实必要的安全控制、审计和文档，以证明符合这些法规。

6) 云数据保护的全球监管格局

《通用数据保护条例》（GDPR）

GDPR 旨在确保欧盟居民个人信息和数据的安全性。数据加密、数据泄露通知和跨境数据传输只是该条例对企业在收集、存储和处理消费者个人信息方面施加的众多限制中的几项。组织必须按照 GDPR 的要求采取必要的组织和技术措施，以保护个人数据。无论数据是否存储在云端，这些要求均适用。

《健康保险流通与责任法案》（HIPAA）

美国的 HIPAA 规定了受保护健康信息的隐私和安全规范。处理受保护健康信息的医疗保健服务提供者、保险公司和云服务提供商必须遵守 HIPAA 标准，例如必须对受保护健康信息进行加密，并在静态存储和传输过程中加以保护。云服务供应商应提供符合 HIPAA 要求的安全保障，医疗保健机构则应确保所使用的云服务符合 HIPAA 对处理受保护健康信息所要求的安全和隐私标准。

《加利福尼亚州消费者隐私法》（CCPA）

CCPA 是一项州级隐私立法，赋予加利福尼亚州居民更多掌控其个人信息的权利。根据这些保护措施，个人有权访问、更正或删除自己的个人数据，也有权拒绝出售这些数据。CCPA 规定，处理加利福尼亚州居民数据的云服务提供商必须遵守该法案的要求，包括确保数据收集实践透明、保障数据安全，并建立适当的同意管理流程。

云计算中的隐私保护技术

在云计算环境中，下一节将介绍一些保障隐私保护数据处理的方法：

同态加密（HE）

一种称为同态加密的加密过程，可以对加密数据进行计算，所得结果与对明文执行相同操作后得到的结果相同。利用这一特性，分析人员能够保护敏感信息的隐私。尽管这一构想最初由 Rivest、Adleman 和 Detrusors 于1978年提出，但直到21世纪初，切实可行的实现方案才开始出现。该领域最近的一项进展是 Gentry 于2009年提出的全同态加密方法，为后续研究奠定了基础。部分同态加密的例子包括 RSA 和 El Gamal；有限同态加密支持有限次数的运算，而全同态加密则支持不受限制的计算。RSA 和 El Gamal 也是有限同态加密的例子。尽管全同态加密安全性高，但计算开销巨大，因此无法应用于大规模的实际场景。

7) 同态加密在云计算中的应用

同态加密用于云计算的原因如下：

- 安全数据分析： 同态加密可在不泄露数据本身的情况下，对机密数据执行基于云的计算，从而保护组织的隐私 [11][12]。

- 医疗数据保护： 云计算支持安全存储和分析患者数据，并能以完全符合所有适用标准的方式执行关键计算。

- 隐私保护机器学习：使用同态加密对加密数据集进行训练并推断机器学习模型（如支持向量机和线性回归），从而实现基于安全数据的协作学习。

- 安全多方计算（SMPC）：多个参与方在不泄露私有输入的情况下，对这些输入计算某个函数，从而确保数据保密性。
- 跨组织数据共享：同态加密促进了对分布于不同实体的敏感数据进行安全共享和协作计算。

8) 同态加密（HE）的挑战

下面讨论同态加密面临的一些挑战和局限：

- 高计算开销：加密、解密以及对密文执行操作都需要大量计算，导致性能瓶颈。

- 在时间敏感型应用中的效率低下：由于处理延迟，同态加密无法有效应用于实时或其他对延迟敏感的应用场景[13]。
- 实现要求复杂：实现同态加密需要深入掌握该技术的数学原理，并选择合适的参数。

- 普及程度有限：技术复杂性使其尚未在行业和研究领域得到广泛应用。
- 缺乏标准化工具和框架：目前用于实际实现同态加密的框架和实用工具有限。

- 可扩展性问题：对于数据密集型或大规模应用而言，计算和存储开销带来了巨大挑战。

安全多方计算 (SMPC)：

SMPC 协议用于保障由多个参与方共同执行的计算的隐私和完整性。这些协议还应能够抵御试图窥探参与方隐私或干扰计算过程的对手 [14][15]。因此，安全模型对于确定潜在对手的能力和意图至关重要。本节将讨论 SMPC 的技术层面，包括其数学基础、算法实现和密码协议。

9) SMPC 中的密码协议

SMPC 采用多种密码学技术，以确保在整个计算过程中每项数据输入的机密性。机密信息共享、同态加密和隐匿传输都是此类协议的例子。

- 秘密共享：将输入拆分为多个份额，并分发给参与方。每个份额单独来看可能毫无价值，但将它们合并后便可还原原始输入，同时不向任何单个参与方泄露该输入。

- 同态加密：使用加密数据进行计算，得到加密结果；解密后，该结果与使用明文计算所得的结果一致。[16]此特性可确保信息在处理过程中得到安全保护和加密。

- 不经意传输：确保接收方享有隐私权：一方能够向另一方发送消息，而无需确定对方接收的是谁的信息。

10) 安全多方计算（SMPC）中的隐私与安全

安全多方计算（SMPC）是一种框架，可在不泄露参与方输入的情况下，对分布式数据进行函数计算。本部分将讨论安全多方计算所提供的隐私和安全保障、其能够抵御的威胁类型，以及其在数据保护方面的局限性。安全多方计算保障各方的隐私，使任何一方都无法获得超出了解计算结果所需的信息。这可以通过密码协议实现，从而无需知晓具体输入即可完成计算 [17]。只要参与方中的大多数没有串通，安全多方计算就能保护任何具体输入的隐私；通常可以通过假设一定数量的协作方会保持诚实来确保这一点。

SMPC 的安全保障：尽管存在攻击性因素，仍能正确且一致地执行计算的能力，被称为 SMPC 中的安全性。SMPC 协议应能够抵御多种恶意攻击，包括篡改和窃听。它们还假设部分参与方可能受到污染或实施恶意行为，并确保计算结果正确。

差分隐私（DP）

差分隐私是一种数学概念，有助于保护个人隐私，同时允许对数据进行分析。它通过向数据或针对数据的查询响应中添加适量噪声来实现这一点，从而使单个数据点的存在（或缺失）不会显著影响答案。这种方法可以确保人们几乎无法根据汇总结果推断某个特定个人的数据，从而提供较高的隐私保护水平。当需要分析医疗记录或金融交易等私人信息以获取更多信息，同时又不泄露相关人员的身份时，差分隐私十分重要 [18]。图2展示了如何在保护用户数据安全的同时进行数据分析。收集来自各种地点（包括医院和学校）的敏感用户信息。考虑数据的敏感性，并对原始数据应用特定函数——噪声函数，在此过程中对其进行转换。这样做是希望在保护个人隐私的同时保留数据效用。采用深度学习等方法处理添加噪声后的数据，从中提取有用见解，同时不侵犯任何人的隐私。

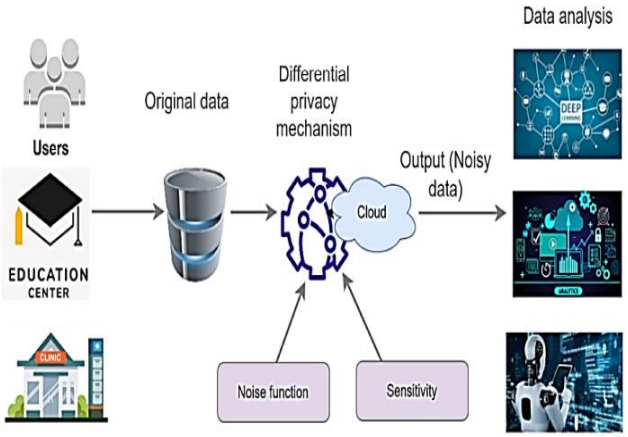


图2。 差分隐私概念

差分隐私之所以重要，是因为它提供了强大且可量化的隐私保护，而在大数据和高级分析时代，这类保护正变得愈发重要。随着组织越来越依赖大型数据集来推动创新和决策，保护这些数据库中个人的隐私至关重要 [19]。差分隐私通过降低重新识别或

即使对手掌握了额外信息，也能避免敏感信息泄露。这对于确保符合严格的隐私要求（包括《通用数据保护条例》）以及维护数据提供方与数据使用方之间的信任，都至关重要。

协作式与新兴隐私框架

本节介绍联邦分析/联邦学习（FA/FL）和新兴趋势，重点阐述用于隐私保护云计算的协作方法与新一代框架。

联邦分析与联邦学习（联邦分析/联邦学习）

联邦分析与联邦学习是隐私保护计算领域的一种范式转变：原始数据不会发送到中央服务器，而是将中间结果或模型更新发送到中央服务器。该模型解决了集中式数据聚合带来的隐私问题，尤其适用于医疗保健和金融等高度敏感的领域。

11) 范式

联邦分析（FA）侧重于协作式统计分析，而联邦学习（FL）则将其扩展到机器学习（ML），无需直接共享数据即可进行联合模型训练。

12) 安全聚合

可利用密码协议（例如安全多方计算（SMPC）和同态加密（HE）[20], ，确保个体贡献不被暴露，且服务器只能看到聚合结果。

13) 与差分隐私及同态加密的集成

差分隐私（DP）常被添加到联邦系统中，以降低模型反演和成员隐私受到威胁的风险；同态加密（HE）则可用于对加密更新进行计算。这种叠加方法可提供本地与全局隐私保障。

14) 现实世界部署

谷歌的 Board 是最早的大规模联邦学习应用之一，因为该公司会将用户的输入数据保存在本地，仅分享模型改进成果 [21]。在联邦学习中，医疗联盟可以跨多家医院训练模型来预测疾病，而无需提供任何患者信息。金融机构也已开始在信用评分和欺诈检测中采用联邦学习。

15) 挑战

尽管联邦分析/联邦学习潜力巨大，但仍面临通信开销、系统高度异构、数据不平衡（非独立同分布）以及容易遭受模型投毒或后门攻击等挑战。解决这些问题对于实现现实世界中的可扩展性至关重要。

新兴趋势

除了联邦分析/联邦学习之外，目前还在开发各种新模型和技术，以增强基于云的数据处理中的隐私保障。

16) 可信执行环境（TEE）

可信执行环境（TEE），包括 Intel SGX 和 AMD SEV，可在隔离的硬件飞地中对明文数据执行计算，并提供近乎原生的效率 [22]。尽管可信执行环境（TEE）降低了密码学开销，但仍容易受到硬件攻击和侧信道攻击。

17) 混合方法（同态加密 + 差分隐私 + 联邦学习）

近年来的研究越来越多地采用这种方法组合，以在安全性、可扩展性和准确性之间取得平衡。例如，通过 HE 实现安全聚合（属于同一方），通过 FL 实现去中心化训练，以及通过差分隐私应对噪声推断攻击。这类混合框架可针对多种威胁提供分层保护。

18) 抗量子方法

随着量子计算不断发展，传统加密方案可能会逐渐过时。如今，云端隐私保护方法开始采用基于格问题（例如环上学习带误差问题）的后量子密码学，以确保后量子时代 HE 和 SMPC 协议的安全性。

19) 边缘云协同框架

新兴架构将边缘计算与云分析相结合，使敏感数据能够在边缘（例如物联网设备、手机）进行预处理和匿名化，然后再汇总到云端[23]。智慧城市、自动驾驶汽车和工业物联网等涉及实时数据的应用，能够从这种设计所带来的低延迟和避免暴露原始数据中获益。

应用与未来方向

医疗保健、金融、物联网和工业领域均可从隐私保护数据处理中受益，该技术能够安全、高效地交换数据。多种方法可提供不同程度的安全性、可扩展性、准确性和效率，包括混合方法、联邦学习、差分隐私、安全多方计算和同态加密等 [24]。这些方法支持安全电子健康记录交换、隐私感知金融分析和预测性维护等应用，但在可扩展性、隐私与效用权衡、标准化、对抗性威胁以及云集成方面仍面临挑战。以下部分将讨论隐私保护数据处理技术的应用、比较分析及相关挑战：

医疗保健、金融、物联网和工业领域的应用

下面将讨论医疗保健、金融、物联网和工业领域的一些相关应用：

- 医疗保健： 在保障患者保密性的同时，支持各机构之间安全交换电子健康记录（EHR）[25]。它还支持用于精准医疗的隐私保护基因组分析，以及通过可穿戴设备和基于传感器的设备进行安全远程监测。
- 金融： 在利用加密交易数据识别欺诈和评估信用风险的同时，满足数据隐私要求。支持注重隐私的客户画像、个性化金融服务以及安全的基于区块链的支付系统。

- 物联网 (IoT)： 保护智能家居、联网汽车和可穿戴设备中的敏感信息。支持医疗物联网中的安全信息传输，并保障工业传感器系统中的隐私，还可支持预测性维护等多种应用。

- 行业案例研究： 联邦学习已应用于制造业，无需共享原始数据即可预测产品维护需求和质量。供应链中采用隐私敏感分析来优化物流、跟踪和需求预测。这些方法有助于安全管理智能电网以及能源行业中注重隐私的需求响应流程。

比较分析：技术分类体系

处理个人数据的方法多种多样，各有优缺点。这些技术的分类体系大致可分为密码学方法、统计隐私方法、分布式学习方法和混合解决方案 [26]。安全多方计算和同态加密是两种密码学算法，能够提供高水平的安全性，但计算成本高昂。包括差分隐私在内的统计方法，在效率相对较高的同时，能够提供可行的隐私保护水平，但可能会影响数据准确性。分布式学习方法，尤其是联邦学习，通过保持原始数据分散存储来实现可扩展的分析，但可能容易受到对抗性攻击。混合系统通过结合多种技术，实现了可扩展性、精确度、效率和安全性，

因此非常适合复杂的现实世界应用。这些方法的主要特点和缺点总结于表II。

表II. 比较分析：技术分类体系

技术	安全性	效率	可扩展性	准确性	关键协同效应
同态加密 (HE)	非常高	Low（计算上昂贵）	Low–中	High	有效集成 when with 联邦学习安全训练
安全多方计算（SMC）	非常高	低–中	中	High	支持协作金融和医疗保健
差分隐私（差分隐私）	高（可调通过 ε）	High	High	中–高（取决于 on 隐私预算）	适用于大规模机器学习任务
联邦学习（联邦学习）	中–High	High	Very High	High	非常适用于物联网和 edge 计算；通过差分隐私/同态加密增强
混合方法	非常高	中	中–High	High	结合密码学与机器学习实现平衡的隐私-效用权衡

挑战与开放性研究方向

下面讨论一些挑战和开放的研究方向；

- 可扩展性：许多密码学和隐私保护技术仍然需要大量计算资源，因此难以高效处理大规模数据[27]。将这些方法扩展到实时或大数据应用仍然是一个活跃的研究问题。

- 隐私–效用权衡：降低数据效用或准确性是实现强隐私保护时常见的权衡。在医疗保健、金融等不同领域，如何制定策略，在这些相互对立的目标之间取得平衡，仍然是备受关注的问题。

- 标准化和互操作性：缺乏一套普遍认可的隐私保护系统标准，限制了其在异构平台上的应用 [28]。不同框架之间还存在互操作性问题，使其在多组织环境中的集成更加困难。

- 对抗性威胁：隐私系统可能遭受模型反演、成员推断和数据投毒等攻击。研究人员关注的问题之一，是设计强大的防御措施，使这些系统可靠且值得信赖。

与云环境的集成：部署在云基础设施中的隐私保护方法面临安全性、效率和透明度方面的挑战。随着越来越多的数据处理任务转移到云系统，确保集成协调一致且安全至关重要。

文献综述

本节比较了联邦学习和协作学习中隐私保护方法的相关文献。这些方法提供了多种机制，包括梯度稀疏化、同态加密、Blockchain、迁移学习和异常值检测，以优化隐私、可扩展性和效率，如表III所示。

Feng 和 Du（2021）提出的 FLZip 是一个新框架，有助于降低联邦学习中的同态加密开销。根据各层梯度的大小，客户端会筛除不显著的梯度值，而不是逐个加密梯度值。FLZip 采用键值对编码来聚合经筛减梯度的密文。此外，该框架还集成了误差累积机制，以确保即使进行筛减，模型准确性也能保持不变。在不损失模型准确性的前提下，FLZip 相较于先进的基于同态加密的联邦学习框架 Batch Crypt，性能提升显著：加密和解密操作数量减少 6.4 倍，服务器收发数据量减少 13.1 倍 [29]。

Itokazu、Wang 和 Ozawa（2021）利用一种基于所谓联邦学习方案的决策树集成，提出了一种新方法，用于检测由多个组织评估的数据中的异常值。通过将一种名为孤立森林的现有异常值检测模式纳入联邦学习范式，这项研究能够采用加法同态加密来保护各实体的数据。多个基准数据集的实验结果表明，无论组织数量如何增加，所提出的隐私保护型孤立森林（pp-forest）都能保持稳定的分类性能。其性能与单一组织条件下的表现基本相同 [30]。

Passerat-Palmbach 等人（2020）为一种先进的区块链编排机器学习系统奠定了基础。该系统能够在保护患者隐私的同时，在医疗保健领域开展联邦学习，为与健康相关的应用开辟了新的可能性。凭借以下六大支柱，这项计划有望在未来几年取得成功：(a) 在兼具安全性与透明度的公有区块链上匿名存储数据和分析流程。(b) 通过数据与计算匹配，释放过去被认为不合乎伦理、违法或无法实现的价值。(c) 将联邦学习与高级密码学结合，以获得保障。(d) 同态加密、安全多方计算和 AMD SEV-SNP 加密，都是软件和硬件安全能力的例子。(e) 利用代币化信誉系统，激励提升数据质量。(f) 通过降低模型投毒风险，消除错误数据 [31]。

Zhang 等人（2020）提出的方法通过仅与服务器交换梯度，防止训练集暴露，从而解决了联邦学习中的隐私问题。该方法解决了现有的两种漏洞：从梯度中提取信息，以及服务器端可能伪造结果。该方法利用 Parlier同态加密和中国剩余定理，以较低的通信和计算成本实现隐私保护的联邦学习。进一步集成双线性聚合签名技术，可确保聚合梯度有效，同时保持较高的效率和准确性，即使启用了验证机制也是如此 [32]。

Gao 等人（2019）提出了一种迁移学习策略，用于处理重叠的同质特征空间中的协变量偏移，连接不同数据拥有者所拥有的多个特征空间，并确保联邦学习中的隐私。他们提出了 HFTL 框架，该框架有三个可能的版本：两个基于秘密共享方法，另一个基于端到端隐私保护多方学习机制。他们的研究最终通过在五个基准数据集上验证 HFTL 的有效性、可扩展性和安全性，取得了成果。

随后，他们利用 MIMIC-III数据集预测院内死亡率，将其付诸实践。该数据集包含高度敏感的患者信息[33]。

Zhang 等人（2018）提出的机器学习框架可通过共享用户数据提升学习效率，近来备受关注，并已开始应用于实际问题。处理高度敏感的用户数据时，由于协作式深度学习涉及多个用户交换和交互数据，因此数据泄露确实有可能发生。由此产生一个重大问题：如何在协作式深度学习处理过程中保护数据隐私。本研究的作者以协作式深度学习的两个阶段为基础，总结了隐私保护技术的应用，并考察了该领域的当前研究状况。最后，文章讨论了这一问题的未来发展方向和趋势 [34]。

表III. 隐私保护方法的比较分析

作者 / Year	所提框架 / 方法	技术 Used	关注领域 / 问题 应对	Key 贡献	结果 / 优势
Feng & Du（2021）	FLZip 框架 to 减少 同态加密开销 在联邦学习中	同态加密（HE），梯度稀疏化，键值编码，误差累积	High 计算开销 of HE in 联邦式学习	高效加密，通过梯度稀疏化 and 误差处理	减少的加密/解密 by 6.4×/13.1×，网络占用减少了 5.9×/12.5×，准确性得以保持
Itokazu, Wang 与 Ozawa 2021）	隐私保护的孤立森林 (pp-iForest)	决策 Tree 集成，联邦学习，加法同态加密	异常值检测 跨越多个组织 with 隐私保障	扩展孤立森林转入联邦环境 使用 HE	稳定的 分类跨越 数据集，准确性 关闭 to 集中式场景
Passerat-Palmbach et al.（2020 年）	用于联邦学习已开启区块链技术	Blockchain、联邦学习，HE, SMC, 硬件密码学（Intel SGX, AMD SEV-SNP）	安全编排 of FL in 敏感的医疗 data 共享	整合 Blockchain，密码学，和硬件 用于联邦学习	强大的 隐私 and 计算 保障，激励的 data 质量，模型模型投毒防御
Zhang 等人（2020）	隐私保护 FL with 梯度共享 and 验证	Paillier HE, 中国剩余定理，双线性聚合签名	梯度泄露 and 伪造 联邦学习中的威胁	保护梯度 and 确保正确性通过密码学验证	Low 成本，high 准确性，安全聚合 with 验证

Gao等人（2019年）	异构联邦迁移学习（HFTL）	同态加密，秘密共享，迁移学习	处理异构特征空间跨参与方在联邦学习中	端到端安全联邦的迁移学习框架	安全，有效，可扩展；已在5数据集+real医院死亡率预测
Zhang 等人（2018）	协作隐私保护智能机器	协作DL，隐私-保护方法（各种）	数据泄露风险泄露in协作深度学习	综述方法in协作深度学习阶段和指出不足	描述当前实践，指出潜在危险，并提出.....方向未来研究

结论与未来方向

在云计算和分布式计算场景中，众多实体处理和共享海量数据，使敏感信息保护成为一项重大难题。安全多方计算、联邦学习、差分隐私、同态加密和混合框架等隐私保护数据处理方法，使分析、协作学习和计算成为可能。这些方法已应用于医疗保健、金融、物联网和工业领域，在安全性、可扩展性、效率和准确性之间取得平衡。然而，仍存在许多障碍。云平台难以无缝集成、通信和计算开销庞大、隐私与数据效用权衡、缺乏标准化以及易受对抗性攻击等问题，导致这些方法尚未得到广泛应用。要确保隐私敏感型系统兑现其承诺，就必须解决这些局限。未来研究方向包括：创建适用于大规模和资源受限场景的轻量级、可扩展密码协议；设计能够互补运用隐私保护技术的混合框架；以及建立可根据具体情境灵活调整隐私与效用平衡的机制。此外，还需要标准化的评估体系和规范，以确定这些解决方案的有效性、安全性和互操作性。要在现实世界中有效、安全且切实可行地实施隐私保护数据处理，这些领域至关重要。

参考文献

1. H. Fang 和 Q. Qian，《基于同态加密与联邦学习的隐私保护机器学习》，Futur. Internet，第 13 卷，第 4 期，第 94 页，2021年4月，doi：10.3390/fi13040094。 2. Y. Bai 和 M. Fan，《一种提升联邦学习隐私与安全性的方法》，2021 IEEE第六届计算机与通信系统国际会议，第 704–708 页，2021，doi：10.1109/ICCCS52626.2021.9449214。 3. Abhishek 和 P. Khare，《云安全挑战：实施安全SaaS应用开发最佳实践》，Int. J. Curr. Eng. Technol.，第 11 卷，第 06 期，第 669–676 页，2021年11月，doi：10.14741/ijcet/v.11.6.11。 4. M. Hao、H. Li、G. Xu、S. Liu 和 H. Yang，《迈向高效且保护隐私的联邦深度学习》，载于 ICC 2019 - 2019年IEEE国际通信会议（ICC），IEEE，2019年5月，第 1–6 页。doi：10.1109/ICC.2019.8761267。

5. P. Mell 和 T. Grance，《美国国家标准与技术研究院云计算定义》，载于《数字企业中的应用性能管理（APM）》，爱思唯尔，2017 年，第 267–269 页。doi：10.1016/B978-0-12- 804018-8.15003-X。 6. A. Papaioannou 和 K. Magoutis，《一种评估多云中分布式应用部署的架构》，载于《 2013 年 IEEE 第五届云计算技术与科学国际会议》，IEEE，2013 年 12 月，第 547–554 页。doi：10.1109/CloudCom.2013.79。 7. B. Soltani、A. Ghenai 和 N. Zeghib，《迈向多云环境中的分布式容器化无服务器架构》，Procedia Comput. Sci.，第 134 卷，第 121–128 页，2018，doi：10.1016/j.procs.2018.07.152。 8. V. Singh，《从大规模 Oracle Fusion Cloud 数据迁移中汲取的经验》，Int. J. Sci. Res.，第 10 卷，第 10 期，第 1662–1666 页，2021。 9. Z. S. Ageed、S. R. M. Zeebaree、M. A. M. Sadeeq、R. K. Ibrahim、H. M. Shukur 和 A. Alkhayyat，《从网格与云计算经雾计算和边缘计算迈向露计算的综合研究》，载于第四届伊拉克工程技术及其应用国际会议 IICETA 2021，2021。doi：10.1109/IICETA51758.2021.9717894。 10. G. Sarraf，《工业物联网的弹性通信协议：大规模保护信息物理系统》，Int. J. Curr. Eng. Technol.，第 11 卷，第 6 期，2021。 11. D. Liu、Z. Yan、W. Ding 和 M. Atiquzzaman，《边缘计算中的安全数据分析综述》，IEEE Internet Things J.，第 6 卷，第 3 期，第 4946–4967 页，2019 年 6 月，doi：10.1109/JIOT.2019.2897619。 12. S. B. V. Naga 、 K. C. Sunkara 、S. Thangavel 和 R. Sundaram，《分布式存储网络中的安全且可扩展的数据复制策略》，Int. J. AI, BigData, Comput. Manag. Stud.，第 2 卷，第 2 期，第 18–27 页，2021，doi：10.63282/3050-9416.IJAIBDCMS-V2I2P103。 13. A. Acar、H. Aksu、A. S. Uluagac 和 M. Conti，《同态加密方案综述：理论与实现》，ACM Comput. Surv.，第 51 卷，第 4 期，2018，doi：10.1145/3214303。 14. D. Quang、B. Martini 和 C. K.-K. Raymond，《对手模型在应用安全研究中的作用》，Comput. Secur.，第 81 卷，第 156–181 页，2019年3月，doi：10.1016/j.cose.2018.12.002。 15. V. M. L. G. Nerella，《面向可扩展合规性的安全、自动化多云数据库平台架构策略》，Int. J. Intell. Syst. Appl. Eng.，第 9 卷，第 1 期，第 128–138 页，2021。 16. F. Wang 等，《电子医疗保健的隐私保护协同模型学习方案》，IEEE Access，第 7 卷，第 166054–166065 页，2019，doi：10.1109/ACCESS.2019.2953495。 17. J. Zhou、Y. Feng、Z. Wang 和 D. Guo，《利用安全多方计算保护许可型区块链上的隐私》，Sensors，第 21 卷，第 4 期，第 1540 页，2021年2月，doi：10.3390/s21041540。 18. T. Wang、X. Zhang、J. Feng 和 X. Yang，《面向数据统计与分析的局部差分隐私综合综述》，Sensors，第 20 卷，第 24 期，第 7030 页，2020年12月，doi：10.3390/s20247030。 19. P. C. M. Arachchige、P. Bertok、I. Khalil、D. Liu、S. Camtepe 和 M. Atiquzzaman，《深度学习的局部差分隐私》，IEEE Internet Things J.，第 7 卷，第 7 期，第 5827–5842 页，2020年7月，doi：10.1109/JIOT.2019.2952146。 20. Q. Wu、K. He 和 X. Chen，《面向智能物联网应用的个性化联邦学习：基于云和边缘的框架》，IEEE Open J. Comput. Soc.，第 1 卷，第 35–44 页，2020，doi：10.1109/OJCS.2020.2993259。 21. Y. Chen、X. Qin、J. Wang、C. Yu 和 W. Gao，《FedHealth：面向可穿戴医疗保健的联邦迁移学习框架》，IEEE Intell. Syst.，第 35 卷，第 4 期，第 83–93 页，2020年7月，doi：10.1109/MIS.2020.2988604。

22. I. Kholod 等，“物联网开源联邦学习框架：比较综述与分析”，《传感器》，第21卷，第1期，第167页，2020年12月，doi：10.3390/s21010167。23. Q. Yang、Y. Liu、T. Chen 和 Y. Tong，“联邦机器学习”，《ACM智能系统技术汇刊》，第10卷，第2期，第1–19页，2019年3月，doi：10.1145/3298981。24. X. Gu、M. Li 和 L. Xiong，“PRECAD：基于密码学辅助差分隐私的隐私保护且鲁棒的联邦学习”，arXiv，2021。25. A. Haleem、M. Javaid、R. P. Singh、R. Suman 和 S. Rab，“区块链技术在医疗保健中的应用：概述”，《国际智能网络期刊》，第2卷，第130–139页，2021，doi：10.1016/j.ijin.2021.09.005。26. U. Sheikh、A. Khan、B. Ahmed、A. Waheed 和 A. Hameed，“溯源推断技术：分类体系、比较分析与设计挑战”，《网络与计算机应用期刊》，第110卷，第11–26页，2018年5月，doi：10.1016/j.jnca.2018.03.004。27. M. A. Ferrag、L. Shu、X. Yang、A. Derhab 和 L. Maglaras，“基于绿色物联网的农业的安全与隐私：综述、区块链解决方案与挑战”，《IEEE Access》，第8卷，第32031–32053页，2020年，doi：10.1109/ACCESS.2020.2973178。28. M. Gupta、M. Abdelsalam、S. Khorsandroo 和 S. Mittal，“智慧农业中的安全与隐私：挑战与机遇”，《IEEE Access》，第8卷，第34564–34584页，2020年，doi：10.1109/ACCESS.2020.2975142。29. X. Feng 和 H. Du，“FLZip：用于跨机构联邦学习的高效隐私保护框架”，收录于《2021年IEEE国际物联网会议（iThings）、IEEE绿色计算与通信会议（GreenCom）、IEEE网络、物理与社会计算会议（CPSCom）、IEEE智能数据会议（SmartData）及IEEE网络科学大会（Cybermatics）论文集》，2021年12月，第209–216页。doi：10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics53846.2021.00044。30. K. Itokazu、L. Wang 和 S. Ozawa，“使用同态加密的隐私保护集成决策树异常值检测”，收录于《国际神经网络联合会议论文集》，2021年。doi：10.1109/IJCNN52387.2021.9534464。31. J. Passerat-Palmbach 等，“区块链编排的机器学习：用于电子健康数据隐私保护联邦学习”，收录于《2020年IEEE国际区块链会议（Blockchain）》，IEEE，2020年11月，第550–555页。doi：10.1109/Blockchain50366.2020.00080。32. X. Zhang、A. Fu、H. Wang、C. Zhou 和 Z. Chen，“一种隐私保护且可验证的联邦学习方案”，收录于《ICC 2020——2020年IEEE国际通信会议（ICC）》，IEEE，2020年6月，第1–6页。doi：10.1109/ICC40277.2020.9148628。33. D. Gao、Y. Liu、A. Huang、C. Ju、H. Yu 和 Q. Yang，“隐私保护异构联邦迁移学习”，收录于《2019年IEEE国际大数据会议论文集（Big Data 2019）》，2019年。doi：10.1109/BigData47090.2019.9005992。34. D. Zhang、X. Chen、D. Wang 和 J. Shi，“协作式深度学习与隐私保护综述”，收录于《2018年IEEE第三届网络空间数据科学国际会议论文集（DSC 2018）》，2018年。doi：10.1109/DSC.2018.00104。