第一章 现状分析与复杂工程问题归纳

1.1 背景与意义

在数字经济时代，随着大数据技术和人工智能的快速发展，数据作为最活跃的生产要素，已经深刻融入到经济活动中，推动着生产力的提升和生产关系的变化。尤其在金融保险、个性化推荐、智能物联网、医疗健康等数据密集型行业，数据的流通和融合已成为数字化转型的核心动力，各种系统的相互交叉使得数据的统一管理和共享变得愈加复杂。同时在数据共享过程中，未经授权的访问、数据泄露以及数据篡改等安全问题日益凸显，这使得数据隐私保护成为社会各界关注的焦点[1]。

随着《数据安全法》、《个人信息保护法》、《网络安全法》等法律法规的逐步实施，数据合规和隐私保护成为企业在数据流通中的首要任务。在数据流处理的实际场景中，如何在保证用户隐私的前提下实现高效计算，已经成为亟待解决的技术难题[2]。

在此背景下，隐私计算技术应运而生，并成为了确保数据在流通过程中隐私安全的关键手段，联邦学习作为一种分布式机器学习框架，能够在确保数据隐私的同时进行有效的模型训练，成为解决这一难题的重要手段。结合安全多方计算、同态加密和差分隐私等先进技术，能够进一步提升数据处理过程中的安全性和隐私保护能力。这些技术为建立一个面向数据流处理的隐私计算平台提供了有力的技术支持，使得各行业能够在保证合规性和隐私保护的基础上，实现高效的数据流处理与分析，具有重要的社会、经济和技术意义。

1.2 国内外研究分析

近年来，国内外在隐私计算领域取得了诸多进展。目前国内外实现隐私计算的主要技术方向有如下三类[3]：

1. 基于可信硬件的隐私计算技术。以可信执行环境为代表，通过硬件级别的安全机制，为数据计算提供受保护的执行环境，确保应用程序在受保护的容器内运行，实现数据和代码的加密与隔离，提高数据安全性和隐私保护。

2. 基于多方安全计算的隐私计算技术。多方安全计算会约定一个特殊的函数，每个人都可以知道计算的结果，但是无法通过计算结果直接推导出相应的参数。核心是实现多个参与方之间的数据计算并保护隐私，多方安全计算并不是某一个单独的技术，而是一系列 技术的结合体，包括隐私查询、差分隐私、比较运算、哈希函数、秘钥交换、同态加密、OPRF协议。

3. 基于联邦学习的隐私计算技术。联邦学习是一种分布式机器学习框架，适用于敏感数据不能集中训练的场景。参与方在不暴露自身数据的情况下，通过加密协议或其他隐私保护手段共同完成模型训练，根据数据维度不同分为横向联邦学习、纵向联邦学习和联邦迁移学习。横向联邦学习适用于参与者业态相同但触达客户不同的场景，基于样本联合；纵向联邦学习适用于样本相同但业态不同的场景，基于特征联合；联邦迁移学习适用于用户和样本模型都不同的情况，利用数据、任务或模型之间的相似性进行迁移学习。

目前，隐私计算平台构建和应用已经有了不少的案例：如浙江省绍兴市公安局提出构建的基于多方安全计算的广域大数据计算平台，结合云计算资源、分布式存储及安全通道技术，解决了跨域异构的数据共享问题。通过安全计算模块、安全协议和算法模块、计算任务管理模块等实现数据安全共享与隐私保护。在公安领域应用中，通过隐私集合求交等技术，在保障数据隐私的前提下，实现了数据的高效共享与协同，提升了业务处理实时性和隐私性[1]；山西移动与晋商银行合作，通过基于联邦学习的山西移动洞见数智联邦平台（INSIGHTONE）建立联合模型，在进行参数优化和模型评估后输出可供银行侧营销的数据。在保证数据价值安全释放的基础上，实现了数据价值的商业变现，达到了双方共赢的良性发展目标，助力数字经济快速发展[4]。

1.3 现有技术存在的主要问题

目前隐私计算正处于发展阶段，技术方面目前存在的主要问题如下：

1. 计算效率问题。联邦学习通信开销大：在联邦学习中，尤其是基于秘密共享的纵向联邦学习模型，存在大量的数据传输，通信量较大，影响了整体性能。虽然可以通过一些技术手段进行优化，如在带宽巨大的步骤中引入同态加密来减少通信量，但仍面临着如何进一步平衡计算和通信开销的挑战；同态加密计算资源开销大：全同态加密目前的实现路径主要基于自举，普遍存在效率低的问题，使得其在实际生产环境中的应用受限。虽然部分同态加密能够高效计算，但因其只支持加法或乘法单一计算，应用面受限，且在实现复杂算法时可能导致计算精度损失和效率降低。尽管学术界和产业界不断进行算法优化，同态加密的计算复杂度仍较高，与明文操作相比性能差异较大[5]。

2. 数据安全与隐私问题。隐私泄露风险：在一些隐私计算技术应用中，如基于同态加密的纵向 LR 模型，每次迭代可能会有部分明文的中间信息暴露给其他方，即使在参与者都是半诚实的假设下，模型仍存在一定风险。此外，传统的多方安全计算和联邦学习在面对复杂攻击场景时，数据隐私保护的安全性仍有待进一步提高；安全假设与实际情况差异：现有隐私计算技术通常基于一些安全假设，如半诚实模型假设，但在实际应用中，可能面临恶意攻击等更复杂的情况，这就要求技术方案能够更好地适应实际安全需求，提供更强的安全保障[5]。

3. 技术复杂性与易用性问题.技术实现复杂：隐私计算涉及多种技术的融合应用，如密码学、分布式计算、机器学习等，技术实现难度较大，对开发人员和使用者的技术要求较高，这在一定程度上限制了其广泛应用。例如，同态加密算法的实现需要深入理解密码学原理和复杂的数学计算，开发和部署成本较高；易用性有待提高：目前隐私计算平台和技术的使用门槛较高，普通用户和企业在应用时可能面临操作复杂、理解困难等问题，需要进一步优化用户体验，提供更便捷、易用的解决方案，以促进其在更多领域的普及和应用[3]。

金融保险行业，传统个性化推荐行业，大数据交易行业，智能物联网行业和医疗健康应用行业作为数据密集型行业，具有丰富的业务场景。数据要素的流通为金融保险，传统个性化推荐，大数据交易，智能物联网和医疗健康应用数字化转型带来了全新的机遇。在数据融合计算需求激增和数据安全保护趋严的背景下，隐私计算技术作为保障数据融合使用过程中隐私安全的有效手段，成为了金融数据流通领域的主要探索方向。针对金融保险领域，传统个性化推荐领域，大数据交易领域，智能物联网领域和医疗健康应用领域数据流通的安全和隐私保护问题，我们提供了完备的隐私计算技术解决方案，能够帮助这些领域的机构构建了“数据可用不可见，用途可控可计量”的数据可信流通范式。

1.4 解决方案

针对目前隐私计算平台存在的问题，我们基于联邦学习构建了banana隐私计算数据安全平台。Banana隐私计算数据安全平台集成了安全多方计算、可信硬件执行环境、联邦学习等主流的隐私计算技术优点，能够在保证敏感信息安全的同时，实现有价值的数据处理和分析。以下是我平台的关键价值点：

1. 数据隐私保护：为确保数据的隐私性和安全性，隐私计算平台采用了安全多方计算、同态加密、联邦学习、可信硬件执行环境及零知识证明等先进的密码学协议，增强了数据的隐私安全处理能力。在计算和分析过程中，原始数据不出域，从而降低了数据泄露的风险。

2. 满足合规性要求：为确保数据共享和交换的基本前提，平台在数据的采集、加工、共享和使用等多个环节均严格遵守三法一例、《数据出境安全评估办法》及相关标准和规范。该平台为不同行业提供了对应的技术支持，以满足他们在数据流通方面的合规性要求。

3. 数据跨机构共享协作：banana平台为机构间数据共享和开放提供了技术支撑，使得机构间在不共享原始数据的情况下，可进行协同计算，包括联合统计，联合建模等场景。在保护数据隐私的前提下，最大化发挥数据价值，实现有效共享和合作。

4. 降低隐私计算使用门槛：平台对安全多方计算、同态加密、联邦学习、可信执行环境及其他密码学协议进行了技术封装，为用户提供了可视化操作界面。用户无需深入了解底层技术实现细节，采用作业模板、可视化拖拽式DAG图配置等便捷操作实现隐私数据分析及机器学习等场景应用。

5. 安全性与可靠性保障：平台采用多种安全措施和技术手段来确保数据的机密性、完整性和可用性。平台采用随机响应分布式安全多方计算，和Paillier半同态加密等底层安全计算协议，为用户提供了一个安全可靠的数据处理环境。综上所述，我们的隐私计算数据安全平台在保护数据隐私的同时，实现了跨机构之间的数据共享和业务协作。借助降低使用门槛、提供可信计算等多种优势，隐私计算平台为各类业务场景提供了有力支持，实现了数据要素的可信、安全、合规流通，进一步推动了数据要素市场的发展。

第二章 需求分析与总体方案概要设计

2.1 需求分析与建模

（一） 数据接入与隐私保护处理需求

Banana隐私计算平台需支持多种数据源的接入，包括数据库、实时数据流等。同时提供数据采集工具和接口，方便用户将数据快速导入平台；支持数据的压缩和加密存储；提供数据存储管理工具，方便用户对数据进行查询、检索和管理；提供数据清洗功能，能够自动识别和处理数据中的噪声、缺失值、异常值等问题。支持数据去重、数据标准化、数据归一化等预处理操作，提高数据质量。

Banana隐私计算平台需支持对各类敏感数据（如个人身份信息、财务数据、医疗记录等）进行加密，并对不同加密需求提供多种加密算法选择，包括同态加密，差分隐私等，以适应不同类型数据和应用场景的加密需求。用户可以根据数据的敏感程度和安全要求，选择合适的加密算法，从而确保数据在存储、传输和计算过程中的保密性，防止数据被未经授权的访问和窃取。

Banana隐私计算平台结合多种方案优点，需支持以下隐私计算功能：

1. 联邦学习功能

Banana隐私计算平台需支持横向联邦学习、纵向联邦学习和联邦迁移学习等不同类型的联邦学习算法。在横向联邦学习中，实现不同机构间相同特征但不同样本的数据联合建模；在纵向联邦学习中，实现不同机构间相同样本但不同特征的数据联合建模；在联邦迁移学习中，实现不同领域数据间的知识迁移和模型共享；提供联邦学习模型训练、模型评估、模型部署和模型更新等功能，方便用户进行联邦学习模型的全生命周期管理。

2. 同态加密计算功能

Banana隐私计算平台需支持同态加密功能，如半同态加密、全同态加密等分别适应不同数据隐私保密性需求；提供同态加密密钥管理功能，包括密钥生成、密钥分发、密钥存储和密钥更新等，确保密钥的安全性和可靠性。

3. 安全多方计算功能

Banana隐私计算平台实现安全多方计算协议，支持多方之间的联合计算，如多方求和、多方求平均、多方乘法等基本计算操作；需确保在计算过程中各方数据的隐私性，计算结果只能被授权方获取。

（二） 合法合规性需求

banana隐私计算平台需遵守国家相关法律法规，建立合规管理流程，定期对平台的合规性进行评估和审查，及时发现并纠正可能存在的违规行为，杜绝泄露用户隐私，同时根据不同行业的特点和要求，调整数据处理流程和安全策略，确保在满足行业标准的同时，提供最佳的隐私保护和数据利用方案。

（三） 数据共享需求

Banana隐私计算平台需基于联邦学习的技术原理，支持机构间在不泄露原始数据的情况下，进行联合统计分析、联合建模等复杂的数据处理任务，同时结合安全多方计算技术对共享数据进行加密和签名，确保数据在共享过程中的隐私性，只有授权的接收方才能解密和使用共享数据，防止数据在传输过程中被篡改和窃取。例如针对金融保险领域，banana平台需支持多家银行可以在平台上联合进行风险评估模型的训练，而无需共享各自的客户数据，同时banana平台需提供高效的数据交互接口和通信协议，确保机构间数据传输的安全和高效，支持数据格式的标准化转换，方便不同机构的数据进行融合和处理。

（四） 降低使用门槛需求

Banana隐私计算平台需打造简洁、直观的可视化用户操作界面，采用图形化界面元素（如图表、流程图等）展示数据处理流程和结果。用户可以通过拖拽、点击等简单操作完成复杂的数据隐私计算任务，无需编写代码或深入了解底层技术原理。

同时banana隐私计算平台将内置一系列预定义的系统模板，涵盖常见的数据隐私计算场景（如金融保险行业，传统个性化推荐行业，大数据交易行业，智能物联网行业和医疗健康应用行业等）。用户可以根据自己的需求选择合适的模板，并进行简单的参数配置，即可快速启动任务。

（五） 安全性与可靠性需求

Banana隐私计算平台需具备系统配置管理功能，管理员可以对系统的参数、配置文件等进行管理和调整。Banana隐私计算平台需部署防火墙防止外部网络攻击和恶意入侵，保护用户数据安全，确保系统的网络安全。同时平台将定期进行系统漏洞扫描和安全评估，及时发现并修复潜在的安全漏洞；建立完善的数据备份策略，定期对数据进行全量和增量备份，备份数据存储在异地的安全存储设施中，以防止本地数据丢失或损坏。Banana隐私计算平台需对系统的安全配置进行定期审查和更新，确保系统始终处于最佳的安全状态。

2.2 系统总体方案设计

2.3 系统流程设计

参考文献

[1] 汤文全,孙位栋,赵建伟,等.基于隐私计算的广域大数据计算平台构建研究[J].警察技术,2024,(05):38-41.

[2] 杭州云象-隐私计算数据安全平台YunPCDS产品白皮书

[3] 陈正伟.数据管理与隐私计算平台的设计案例分析[J].集成电路应用,2024,41(06):220-221.DOI:10.19339/j.issn.1674-2583.2024.06.099.

[4] 秦鹏,上官丽丽.联邦学习法助力隐私计算新应用[J].中国电信业,2024,(02):77-80.

[5] 邵航,高思琪,钟离,等.同态加密在隐私计算中的应用综述[J].信息通信技术与政策,2022,(08):75-88.