基于区块链和隐私计算的工业数据安全方案应用探索

刘冬 谢文刚 张浩 王俊豪 何聪

中冶赛迪信息技术（重庆）有限公司，重庆，401122；

摘要：随着工业互联网的迅猛发展，工业数据安全已经成为亟待解决的重要问题。工业数据包含大量敏感信息，如生产流程数据、用户个人信息、知识产权等，一旦泄露或被篡改，将给企业和用户带来巨大损失。本文旨在探讨区块链和隐私计算技术在工业数据安全方案中的应用，通过构建可信、安全的数据流通机制，提升工业数据的安全性和可信度。本文还对比当前区块链加隐私计算的设计方案，结合不同业务需求和安全需求，选择不同的设计方案，开发工业安全互通平台，并在数据采集和数据分享场景下对平台进行安全性和可信验证。本研究为工业行业的数据安全与隐私保护给出了新的思路和方法，具有重要的实际应用价值和深远的社会意义。

关键词：区块链；隐私技术；数据安全；工业数据

DOI:10.69979/3041-0673.25.02.014

引言

工业互联网作为新一代信息技术与制造业深度融合的产物，已成为推动制造业高质量发展的主要力量。工业数据的安全性和隐私保护问题日益凸显，传统的数据安全手段已难以满足工业互联网环境下数据复杂、多源、异构的特点。因此，探索新型的数据安全方案，成为工业互联网发展的重要课题。

区块链作为结合密码学技术，采用链式结构存储数据的分布式数据管理技术，区块链系统内的数据安全管理和隐私保护技术至关重要。区块链摒弃了传统数据库中心化、易发生单点故障且依赖中央机构正确管理和保护数据的缺点，而通过去中心化，允许未知节点之间进行点对点的可信价值转移，无需第三方信任机构，从而降低交易成本并提高交互效率。

但是在隐私保护领域，为了在分散的区块链节点中达成共识，区块链中所有的交易记录必须公开给所有节点，这显著增加了隐私泄露的风险。例如，在数字货币应用中，可以通过分析子节点的交易记录以获得用户的交易规律，甚至能够获取到用户的身份信息和位置信息。所以利用隐私保护算法保护防止这些信息被窃取是当前一个重要课题。

区块链和隐私计算作为新兴技术，为工业数据安全提供了新的解决方案。本文将从区块链和隐私计算的技术原理出发，探讨其在工业数据安全方案中的应用。

1 区块链技术原理与特点

区块链的核心在于利用去中心化和加密算法确保数据的安全性与不可篡改性。区块链的应用范围非常广泛，从最初的比特币等加密货币到智能合约、供应链管理、身份验证等多个领域，展现出极大的潜力和多样性[1]。区块链在工业数据安全领域主要有以下几个方面的应用：

1.1 数据确权与追溯

区块链的分布式账本和不可篡改性，为工业数据的确权和追溯提供了有力支持。通过区块链技术，可以记录数据的生成、传输、存储、使用等全生命周期信息，实现数据的可溯源和可审计。

1.2 数据共享与流通

区块链的智能合约技术，可以实现数据的自动化共享和流通。通过智能合约，可以实现数据的安全共享和流通，同时保证数据的隐私性和安全性。

1.3 数据防篡改与保护

区块链的哈希函数和非对称加密技术，可以确保数据的完整性和安全性。哈希函数可以实现数据的唯一标识和防篡改。非对称加密技术则可以实现数据的加密传输和存储，防止数据在传输和存储过程中被窃取或篡改。

2 隐私计算技术原理与特点

隐私计算是一种在保护数据隐私的前提下进行计算和分析的技术。它允许数据在不暴露原始数据的情况下，进行加密计算、多方联合计算等操作，从而实现数据的隐私保护和安全共享。隐私计算的核心技术包括同

态加密、多方安全计算、联邦学习等[2]。在工业数据安全领域，隐私计算主要有以下几方面的应用：

2.1 数据加密与脱敏

隐私计算技术可以通过同态加密等技术，对工业数据进行加密处理，保证数据在传输和存储过程中的安全性。同时，脱敏技术，对敏感数据进行匿名处理，保护关键信息。

2.2 多方联合计算

在工业互联网环境下，多个企业可能需要共同使用数据进行联合分析或决策。隐私计算技术可以实现多方联合计算，即在不暴露原始数据的前提下，进行数据的联合分析和计算。这有助于打破数据孤岛，促进数据资源的共享和利用。

2.3 联邦学习

联邦学习是一种分布式机器学习技术，可以在多个数据节点上进行模型的训练和优化，而无需将数据集中到一个中心节点。通过联邦学习，可以在保护数据隐私的前提下，实现模型的训练和预测。

3 工业安全互通平台技术实现

隐私计算与区块链技术的结合，为工业数据的安全共享与利用提供了新方向。本节围绕技术选型、系统架构以及功能模块三方面，深入探讨工业数据安全的业务场景，阐述工业数据共享协同平台的设计理念与实现过程，描绘功能架构中各模块的职责与交互方式。

3.1 技术选型

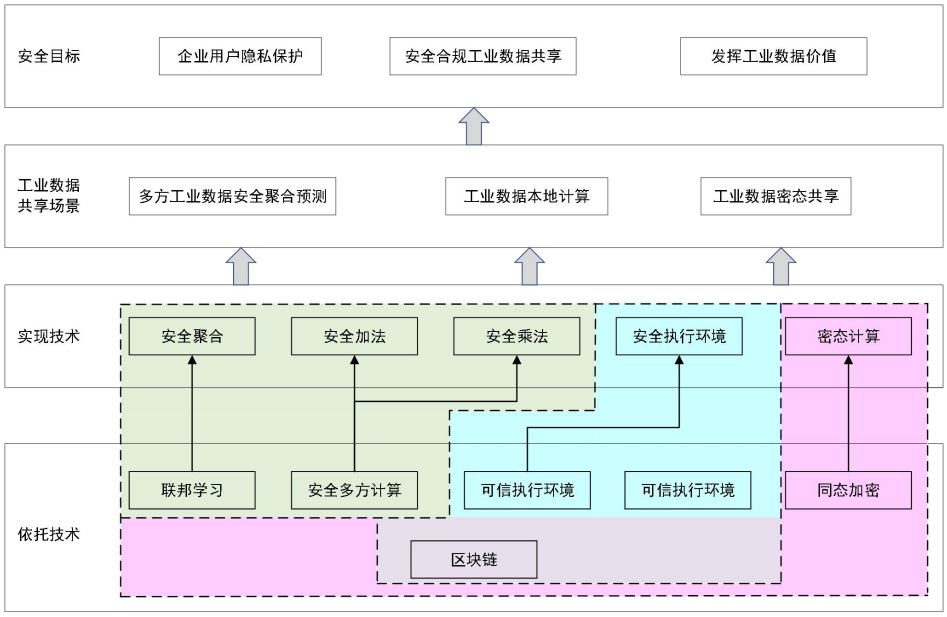

图 1 工业互通技术体系框架

如图 1 所示，结合区块链技术和隐私计算技术，在工业互通领域中主要有三大方向：

3.1.1 联邦学习（Federated Learning，简称 FL）结合多方安全计算（Secure Multi-Party Computation，简称 MPC），将 FL 的本地数据与 MPC 的安全计算结合，实现多方数据的安全聚合。

3.1.2 区块链、基于硬件的可信执行环境（TEE）、和加密算法结合各企业工业数据进行本地计算，然后利用区块链技术，共享最后的计算结果数据，减少原始数据的泄露风险。

3.1.3 区块链技术结合同态加密技术（Homomorphic Encryption，HE，是一种允许在加密数据上直接进行计算的加密方式）实现共享密态数据，数据接收方可将密态数据用于计算，最终得到正确计算结果，整个过程数据一直处于密态，实现了安全的数据共享。

3.2 系统架构

工业互通平台的系统架构如图 2 所示。

图 2 工业互通平台系统架构

在安全加密层次，架构使用国家密码算法，例如 SM2 和 SM4 等算法，同时支持 RSA 和 ECDSA 等标准加密算法。在区块链层面上，兼容超级账本 fabric 联盟链[3]。在隐私计算层面，依托开源框架封装改造，引入多种多方安全计算和联邦学习的算法，如同态加密(Homomorphic Encryption)、零知识证明(Zero-Knowledge Proof)和 XGBoost (eXtreme Gradient Boosting)[4]。在功能层，包含数据识别、数据脱敏、数据共享等核心业务模块，采用链上链下协同方式，关键数据非必要不上链，上链必须脱敏加密，保证工业数据的可信性和隐私安全。

3.3 功能架构

工业互通平台在主要功能上的设计主要考虑敏感数据识别、敏感数据脱敏和数据安全共享功能。工业数据上链共享前，必须进行脱敏处理，为实现数据脱敏，首先需要将敏感数据识别出来。如图 3 工业敏感数据识别流程所示，工业数据分为结构化数据和非结构化数据，对于非结构化数据，处理前需要将这些数据中提取出来主要内容，并对内容去噪，再根据各个行业的敏感数据特征进行分词，引入词性选择、词频统计、词长选择、词频选择的四维特征，实现非结构化的敏感数据的自动识别；对于结构化数据，可以直接去噪，预处理后进行特征提取，实现敏感数据的自动识别。

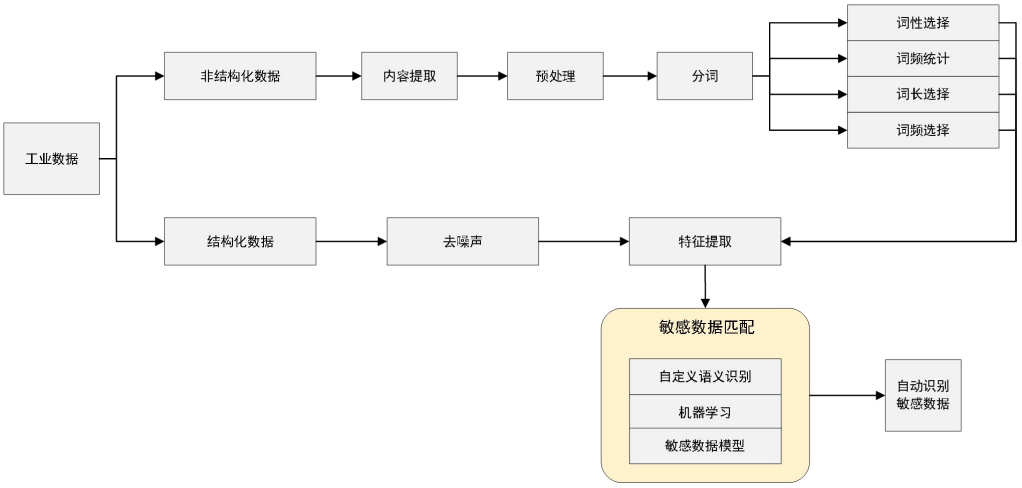

图 3 工业敏感数据识别流程

敏感数据识别完以后，需要对工业数据进行脱敏处理：首先对敏感数据使用 k-means 聚类、机器学习或者 XGBoost 等算法进行分类处理，再根据各行业的自定义的脱敏规则和脱敏算法形成工业数据脱敏策略，实现敏感数据脱敏。

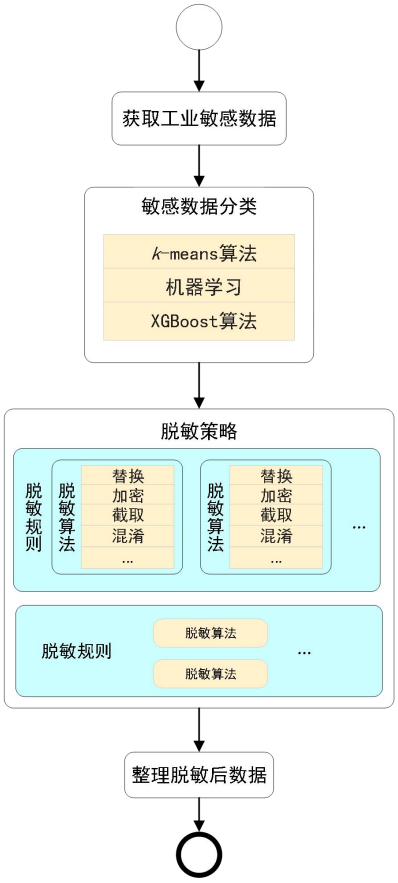

图 4 敏感数据脱敏流程

敏感数据脱敏后可以进入数据共享阶段，在数据共享阶段中，使用区块链技术实现工业数据安全共享，工业领域采用联盟链作为功能区块链最为合适。基于区块链的工业数据安全共享模型如图 5 所示。由于企业对于数据有不同的安全标准，平台采用两种基于区块链加隐私计算技术的数据共享方案。第一种方案如图中 a1 到 a3 的路径，使用单一的联盟链，采用密文策略属性基加密(CP-ABE)、哈希算法(Hash)、Paillier 同态加密算法、BLS 签名(Boneh-Lynn-Shacham)等算法对工业数据进行加密，加密后将数据共享到数据需求方；第二种方案如图中 b1 到 b5 所示，企业对数据使用上述算法对数据进行加密，然后并不将数据上传到联盟链，而是将加密后的数据上传到本地私有链上面，本地私有链会对上传的数据进行进一步融合计算，然后将融合计算结果报告上传到联盟链，数据需求方可通过联盟链获取到最后的计算结果报告，并对报告进行加工分析。

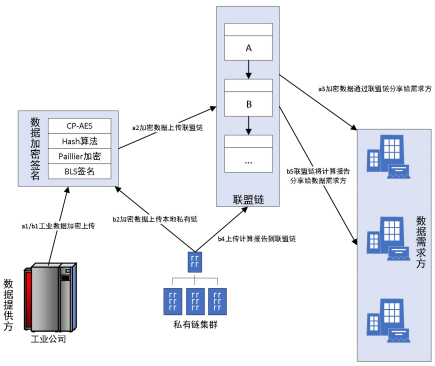

图 5 工业数据安全共享模型

4 安全与可信验证

4.1 工业数据采集存储场景下的安全与可信验证

在普通的数据采集场景下，设备制造商一般会采用

通用协议或者定制化协议进行数据传输和采集，第三方可以使用相同协议进行数据采集并完成进一步加工。整个过程中涉及的厂商较多，数据泄露的风险成倍增加。另外，被三方厂商加工后或逻辑处理后的数据也会增加数据的不可信程度。

在实验场景中，分别模拟两种情况，一种是中间厂家篡改数据，另一种场景下不篡改数据，然后对设备数据进行采集。一般采集技术无法识别数据是否被篡改，但基于区块链和隐私计算技术的工业数据采集技术，通过搭载可信芯片的物联网设备进行实时数据采集，能够确保所采集数据的真实性和完整性，利用隐私计算技术如同态加密、数据脱敏、差分隐私等，可以在保护数据隐私的同时，将密文数据安全地上链存储，避免原始数据泄露，实现数据的“可用不可见”。

4.2 工业数据分享场景下的安全与可信验证

在工业数据分享场景中，需要在确保数据安全和隐私的前提下，实现数据的流通与共享。在数据分享中，工业数据存在隐私泄露等风险，基于区块链和隐私计算的工业数据需要对数据分享场景进行安全性验证和可信性验证。

在实验场景中，通过利用加权阈值秘密共享方案实现的工业数据共享，该方案改进了通用数据共享流程，为工业物联网的终端成员提供了一个既灵活又安全的数据共享访问控制机制。在该方案中，参与数据共享的实体包括：证书颁发机构(CA)、属性权威(AA)、监管节点(RN)、终端成员以及星际文件系统(IPFS)。方案通过身份认证、加密存储和访问控制等方面来确保数据信息的隐私保护和安全共享。该方案包括五个部分：初始化阶段、注册阶段、加密阶段、认证阶段和解密阶段，由五个参与实体执行：CA、AA、RN、终端成员和 IPFS。这个过程确保了只有经过验证和授权的实体才能访问敏感数据，同时保护了数据的完整性和机密性，防止了未经授权的访问和数据泄露。

5 结语

在工业领域内，工业数据是智能工业的神经和血液，针对工业数据全流程场景，现有的工业数据在采集、检索、协同中存在隐私保护不足、效率低下等问题，基于区块链和隐私计算技术的全新方案能有效解决这些问题，帮助整合全流程的工业数据，并保证工业数据在流转中的安全性和可靠性，进一步提高生产实际中的安全管理和效率管理。当然，基于区块链和隐私计算的方案仍然存在不足和挑战，需要从区块链数据结构、存储优化、加密优化等多方面进一步在工业场景下展开研究。随着区块链技术和隐私计算技术的不断发展，基于区块链和隐私计算的工业数据安全方案将会体现出更大的应用和推广价值。

参考文献

[1]滕亮,陈兵,赵开斌,等.基于区块链的医疗数据安全共享模型研究与应用 J1.信息安全研究,2023,9(9):884-891.
[2]沈传年,徐彦婷,陈滢霞.隐私计算关键技术及研究展望[J].信息安全研究,2023,9(8):714-721.
[3]隐私计算联盟.隐私计算与区块链技术融合研究报告[EB/OL].(2021-08-13)[2024-04-07].
[4]程亚玲.区块链探析[J].发明与创新:职业教育，2020(7):131.

作者简介：刘冬（1995-），男，汉族，重庆人，工程师，硕士，单位：中冶赛迪信息技术（重庆）有限公司，研究方向：计算机