数据管理与隐私计算平台的设计案例分析

陈正伟
（云赛智联股份有限公司，上海 201108）

摘要：阐述主流的隐私计算技术，构建针对疾控中心的数据管理和隐私计算平台，探讨基于隐私计算的传染病多点触发预警和监测平台，以及基于隐私计算的食品安全风险监测平台应用案例。
关键词：隐私计算，数据管理，数据监测系统。

中图分类号：TP311.13　文章编号：1674-2583(2024)06-0220-02
DOI：10.19339/j.issn.1674-2583.2024.06.099
文献引用格式：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.

Design and Case Analysis of Data Management and Privacy Computing Platform

CHEN Zhengwei
(INESA Intelligent Tech Inc., Shanghai 201108, China.)

Abstract — This paper describes mainstream privacy computing technologies, constructs a data management and privacy computing platform for disease control centers, explores a multi-point trigger warning and monitoring platform for infectious diseases based on privacy computing, and presents application cases of a food safety risk monitoring platform based on privacy computing.
Index Terms — privacy computing, data management, data monitoring system.

0 引言

目前常规的数据共享和数据使用方式存在较大的安全隐患，安全的共享方式是要求数据使用方在无法知晓原始数据的情况下进行数据开发和利用。

1 研究背景

隐私计算技术已经应用于社会的很多方面，但是目前对于隐私计算技术应用于公共卫生和医疗领域的研究相对较少。邓维维[1]提出基于密文策略和隐私计算的方案，所有的卷积、激活等过程都是出于非明文状态，不会泄露预测的成果，只有用户自己通过私钥获得真实预测结果。可以保护用户隐私、同时提升访问的效率，也可以对恶意行为进行逆向追踪。张莺耀[2]提出通过搭建医疗隐私计算平台，可以使数据出于可用不可得的状态中，通过传染病智慧防控、AI建模以及生物信息平台的应用案例，讨论平台的隐私计算能力。华佳烽提出利用skyline查询构造了医学知识模型，应用包括三方在内的交互数据应用于医疗预诊隐私保护方法用以保障医学知识模型中查询数据时的隐私性能，并利用决策树分类提高预测模型的准确性。魏晋提出根据医疗数据的特点，数据共享、数据资源查询、隐私数据传输等流程进入，通过量子信息转换、索引搜索等方式实现医疗隐私数据的共享服务。赵伟提出为了确保数据安全性、完整性和透明度，采用区块链和隐私计算技术为医保数据共享提供技术支撑。

2 隐私计算技术

常用的隐私计算技术包括以下三类。（1）可信计算指借助硬件CPU芯片实现可信执行环境（TEE），构建一个受保护的可信执行环境模块，对于应用程序来说，它的可信执行环境模块是一个安全的容器，用于存放应用程序的敏感数据与代码，并保证它们的机密性与完整性。可信执行环境模块的内存区域是由CPU默认加密的，且只能被同一个可信执行环境模块中的代码所访问，即便是外部高权限实体（VMM、BIOS、SMM）也无法访问。目前，TEE的实现也包括ARM平台的TrustZone、AMD下的SEV等，但在隐私计算领域，可信计算（TEE）是基于硬件和密码学原理的隐私计算方案，相比于纯软件解决方案，具有较高的通用性、易用性和较优的性能。其缺点是需要引入可信方，即信任芯片厂商。但其侧信道攻击也成为不可忽视的攻击向量，需要关注相关漏洞和研究进展。（2）多方安全计算1982年时提出，提出的问题是“两个百万富翁都不愿意对方知道自己具体有多少钱，但是他们又想要知道谁更加有钱？”所以，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。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

作者简介：陈正伟，云赛智联股份有限公司，高级工程师；研究方向：智慧城市、区块链、隐私计算、公共数据。
收稿日期：2024-04-15；修回日期：2024-05-22。

等一系列技术。（3）联邦学习是将互不信任的各方的不同的样本汇聚到一起，共同获得更好的预测模型，因为就像大家联手学习，因此取名为联邦学习。因为各方都掌握不同的数据，而各方都希望汇聚更多的数据从而得到预测模型，但是不同的公司之间又不希望将自己的数据直接交予他人手中。联邦学习一般包含横向联邦学习、纵向联邦学习、联邦迁移学习。横向联邦学习是指业务类似的几家架构，要学习、预测的特征都是类似的，但是由于样本不同，他们之间的合作就叫“横向联邦学习”，纵向联邦学习是指组织的客户的近似的，但是其特征不尽相同，例如银行和保险公司在同一个区域，样本相同，但是模型不相同，因此称为“纵向联邦学习”。联邦迁移学习，是指用户和样本模型都不相同的情况，特征发生迁移，而产生的共同学习，称为“联邦迁移学习”。

3 疾控中心数据管理和隐私计算平台

为了解决公共卫生和疾控控制领域多个机构之间进行数据协作过程中存在的数据权责不清、数据安全无保障、数据获取门槛高、数据协作流程复杂等问题，通过数据生命周期管理和隐私计算平台可以对数据协作方式进行精细化的安全策略定制，确保数据使用方在限定的方式下使用数据，通过多种安全组件提供全生命周期的数据安全与隐私保障，对链上所有可用数据资源及其可能的应用方式进行筛选查看，便于协作前寻找适用数据，降低数据获取门槛，实现低技术门槛、多场景、可自定义的数据协作，见图1。基于多样的数据安全组件、隐私计算引擎以及易编辑、可扩展的协作应用，数据隐私计算协作平台为多机构间频繁、复杂的数据协作提供了安全、便捷、易懂的解决方案。平台包括应用层、安全流转接口层、隐私计算引擎层、数据管理层和基础资源层。

图1 疾控中心数据管理及隐私计算平台功能架构

4 隐私计算平台应用

（1）案例1：基于隐私计算的传染病多点触发预警和监测平台。基于隐私保护计算的传染病多点触发预警和监测平台，建立多点触发、灵敏可靠的智能化传染病早期预警手段。核心模型为传染病智慧化模型，通过构建重点人群监测、动力学模型、个体网络模型计算、预警、分类和研判传染病的发病情况。全市发热门诊、肠道门诊、住院病例监测、哨点监测、公共卫生事件监测、健康危害因素监测等疾控业务数据和外部业务数据等多个数据来源的监测数据通过隐私计算平台进行传染病预警模型的计算，实现传染病病监测的“一网统管”，为疾病防控提供在线实时监测监控，形成多点触发、动态灵敏的预警研判模式，实现科学、可视化的早期预警和发病趋势预测，通过可视化展示平台，实时监控传染病疫情动态，并进行发病趋势预测与疾病预报，建设科学高效、协同联动的传染病应急处置管理体系，为城市传染病应急处置和指挥决策提供信息化支撑。

（2）案例2：基于隐私计算的食品安全风险监测平台。由于大规模的食品安全的爆发属于突发事件，患者可能通过医院就医、社区问诊以及自行去药店购买腹泻药物，而这些数据一般未直接归集，而且由于数据量较大，也不需要进行全量归集，由于此数据属于个人隐私数据，因此可以采用隐私计算的方式进行食品安全风险监测。借助隐私计算平台，对食品安全进行风险监测评估工作，建设功能齐全、适合本市实际需求的信息化平台，整合数据录入、交换、上报以及统计利用和信息发布沟通等各项功能，保障监测和评估工作的顺利运行。构建基于隐私计算平台的食品风险安全风险监测平台，包括构建基于隐私计算平台的医院食源性病例检测、社区卫生服务中心急性肠胃炎监测、腹泻药物监测等模型。如腹泻药物检测模型，通过隐私计算平台导入药店每日销售记录，每日监测10类腹泻药物分区域销售量，以此监测大规模食品安全事件爆发的迹象。医院食源性病例检测通过在隐私计算平台中录入样本采集的信息，如样本种类、采集时间、采集量、检测项目等。基于隐私计算的社区病例对照监测，借助隐私计算平台的可信执行环境，充分保护患者隐私数据安全。

5 结语

通过构建疾控中心的数据管理和隐私计算平台，可以解决疾控中心各个机构之间进行数据协作过程中存在的数据权责不清、数据安全无保障、数据获取门槛高、数据协作流程复杂等问题，为多机构间频繁、复杂的数据协作提供了安全、便捷、易懂的解决方案，在实现的过程中，必须找到数据的私密性、计算的准确性以及计算的时效性之间的均衡点。通过隐私计算推动公共卫生和疾病控制的跨越式发展。

参考文献

[1] 邓维维. 面向智慧医疗的访问控制方案与隐私计算研究[D]. 重庆：重庆大学. 2022.

[2] 张莺耀, 贺成飞, 罗震. 医疗数据隐私计算平台设计与应用案例[J]. 集成电路应用, 2022, 39(02):56-59.