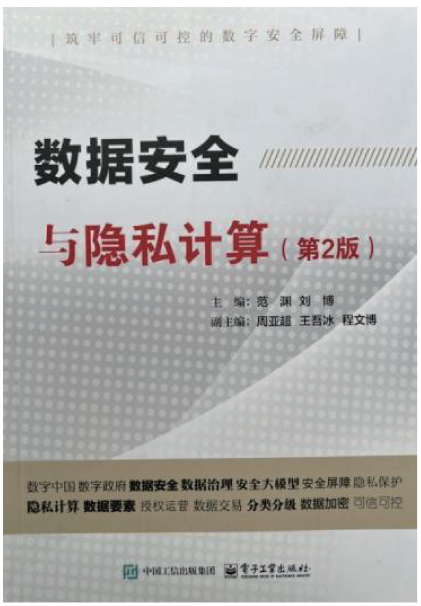
开放数据安全中的隐私保护策略
——评《数据安全与隐私计算(第2版)》

数据要素流通加速推进,开放数据已成为推动社会治理、科研创新与经济发展的关键资源。然而,数据开放过程中广泛存在的个人信息泄露风险,不仅威胁用户隐私权益,更动摇公众对数据共享的信任基础。逐步兴起的基于密码学、匿名化技术的隐私计算范式,为破解数据利用与隐私保护的矛盾提供了技术路径,对此开展相关研究,对完善数据安全治理体系、促进数据要素有序流通具有重大现实意义。



书名:数据安全与隐私计算(第2版)
编者:范渊,刘博
出版社:电子工业出版社
ISBN:9787121487378
出版时间:2024年9月
定价:88元

笔者在开展2024年度甘肃省科技计划项目:“甘肃省市州政府数据开放隐私保护成熟度评价研究:基于数智化转型视角(24JRZA109)”的研究过程中,认真地阅读了《数据安全与隐私计算(第2版)》一书。该书立足数字经济时代数据要素市场化配置需求,系统性构建“理论-技术-实践”三位一体的数据安全与隐私计算知识体系,为数据要素安全流通与价值释放提供了全流程解决方案。全书分3部分,形成从制度基础到技术实践的完整闭环,既可作为高校相关专业教材,也为政府、企业数据治理提供实践指南。第1部分聚焦数据要素市场培育机制,构建数据要素流通的制度基础设施。第1章阐释了数据产权、流通、分配、监管等基础制度框架,梳理了《数据安全法》《个人信息保护法》等核心法律法规。第2章剖析了数据要素市场化配置的现实困境,提出涵盖数据资产化、数字证书体系、会计处理规范等创新解决方案,重点构建“数据沙箱-可信存证-隐私计算”三位一体的基础设施架构,力求破解数据流通中的“不敢”“不愿”“不会”共享难题。第2部分围绕数字化转型中的安全需求,构建全生命周期防护体系。第3—5章建立了风险识别-理论框架-防护实践递进式研究路径;基于数据转接客户访问协议、数据安全能力成熟度模型等国际标准构建治理框架,系统梳理数据库配置缺陷、过度授权、结构化查询语言注入等15类典型风险,形成覆盖“事前评估-事中管控-事后审计”的防护闭环。第6—8章创新性地提出持续改进能力实践框架,构建涵盖资产扫描、分类分级、加密脱敏、行为审计等技术工具链,重点突破动态脱敏、应用程序编程接口防护、数字水印等关键技术。第9章研究了大模型训练中的数据投毒防护、模型泄露监测等新型安全挑战,提出“联邦学习+差分隐私”等技术融合路径。第3部分开创隐私计算技术落地新范式,构建数据要素流通的技术基础设施。第10章提出可信数据流通交易空间框架,构建“供给-交易-交付”全流程技术支撑体系。第11章深度解析了机密计算、安全多方计算、联邦学习3大技术路线,重点突破可信执行环境、秘密分享、同态加密等核心算法。第12章通过政务数据授权运营、金融风控、医疗研发等10大典型场景,展示隐私计算技术在跨域数据流通中的实践价值。第13章针对大模型训练中的数据隐私风险,提出基于机密计算的“数据不动模型动”解决方案,构建隐私保护大模型基础设施新形态。全书突出3大特色:理论体系上融合法学、管理学、计算机科学多学科视角;技术实践上覆盖工具研发、系统部署、场景应用全链条;案例选择涵盖数字政府、金融科技、智慧医疗等8大重点领域。

作者指出,数据脱敏与匿名化是通过抑制、泛化或扰动敏感信息的技术体系,构成数据开放的首道安全防线。其核心在于平衡“数据可用性”与“隐私暴露风险”,技术演进已呈现智能化与数学可证明的显著特征。静态脱敏采用数据掩码、格式保留加密等技术固化处理敏感字段。以医疗数据为例,患者身份证号通过哈希算法可转换为固定长度字符串,既保持数据关联性又阻断逆向解析。动态脱敏则基于访问上下文实时控制数据可见性,如金融机构根据用户权限动态隐藏信用卡号的4—8位数字。静态脱敏可能导致数据分布失真,影响后

续分析准确性,解决方案是通过领域知识库构建语义感知的脱敏规则。传统 k-匿名模型要求每条记录至少与 k-1 条其他记录在准标识符上不可区分,但难以防御背景知识攻击。现代研究引入差分隐私机制,在等价类分组后注入拉普拉斯噪声,实现差分隐私保护。美国人口普查局在发布通勤数据时,采用此混合方法将重识别风险降低至非常低的数量级。高维数据场景下,同时满足 k-匿名和差分隐私会导致数据过度泛化。应对策略是采用基于信息熵的维度约简算法,优先保留关键分析维度。本地差分隐私将噪声注入环节前移至用户终端,彻底阻断原始数据收集。苹果公司 iOS 系统中的“隐私保护广告”功能,通过随机响应机制扰动用户行为数据,确保单个数据点不可解析而群体特征有效保留。开发上下文感知的链接数据发布机制,根据数据类型动态调整隐私预算分配。如对位置数据采用更高强度的扰动,对搜索记录采用较低强度。保障层面则需建立脱敏效果量化评估体系,采用信息损失度和重识别风险双维度指标。同时,构建领域自适应的脱敏模板库,如金融、医疗、教育等不同场景的标准化处理流程。

笔者认为,加密访问控制技术通过密码学工具构建数据流通的“动态防护网”,可实现“最小权限原则”下的细粒度管控。其核心在于将访问控制策略与加密机制深度融合,形成“数据-权限-用户”的三元安全映射。密文策略属性基加密将访问控制树嵌入加密密钥,数据请求方需满足预设属性集合(如机构资质、研究目的)才能解密。如在智慧医疗场景中,只有同时具备“三甲医院”和“肿瘤学研究”属性的机构才能访问特定患者队列数据。采用基于属性的加密方案降低计算开销,结合代理重加密技术实现密钥分发效率提升。代理重加密允许数据提供方将密文转换为目标接收方的可解密格式,无需暴露原始密钥。结合多方安全计算,可在不解密状态下完成联合统计分析。又如在跨境金融反诈骗场景中,多家银行通过多方安全计算技术实现可疑交易模式识别,单个机构仅掌握局部计算结果。引入零知识证明技术,确保计算过程可验证但不可窥探。利用区块链的不可篡改特性记录数据访问全链路信息,构建可审计的权限管理系统。北京市政务数据开放平台采用超级账本框架,实现数据申请、审批、使用全流程上链存证,有效震慑越权访问行为。开发智能合约自动执行访问控制策略,如当数据请求方属性变更时,自动触发权限更新流程。

另外,面向数据分析场景的隐私保护技术,通过算法设计实现“数据可用不可见”的目标,破解数据利用与隐私保护的固有矛盾。其核心在于将安全计算机制内嵌于分析流程,形成“计算过程安全”的新型范式。联邦学习技术使多方在本地训练模型仅交换梯度更新而非原始数据。如微众银行在风控模型训练中,联合多家金融机构采用联邦学习技术框架,使模型安全风控水平提升的同时确保数据不出本地。横向联邦学习适用于同构数据场景,纵向联邦学习解决异构数据联合建模问题,迁移联邦学习处理非独立同分布数据。安全多方计算允许参与方在不暴露输入数据的前提下完成协同计算。在基因组学研究中,多家医院利用安全多方计算技术联合分析患者基因序列,成功发现罕见病致病基因关联位点,而单个机构无法获取其他医院的数据明文。此种情况下,采用混淆电路与秘密共享的混合协议,将计算效率提升 3 个数量级。同态加密支持在加密数据上直接进行计算,结果解密后与明文计算一致。云服务商采用“切比雪夫-卡尔达诺同态加密”方案,可为企业提供外包数据分析服务,如加密文本的情感分析准确率可达 90%以上,而云服务提供商无法获取原始文本内容。后量子同态加密方案(如基于格理论的加密)正在研发中,以应对量子计算威胁。需建立隐私预算动态分配机制,根据数据敏感度调整差分隐私参数。开发专用硬件加速库提升计算性能,同时构建跨平台协议转换层,实现不同隐私计算框架的互联互通。

综上,开放环境下的隐私保护策略已形成“预处理-流通-计算”的全链路技术体系,其有效性取决于具体场景的隐私需求与数据特性。未来研究需重点关注跨策略协同机制,如将差分隐私与联邦学习结合构建混合保护框架,同时,加强技术标准的国际化对接。数据安全与隐私保护是一个系统工程,需要政府、企业、科研机构及社会各界的共同努力,通过制度创新、技术创新和人才培养等多维度推进,以构建更加完善的数据安全治理体系,实现数据要素的安全、高效、可持续利用。随着隐私计算技术的持续演进,有望构建既满足数据开放需求又符合伦理规范的新型数据治理体系,切实引导信息数据产业健康可持续发展。

(闾燕/甘肃政法大学网络空间安全学院/教授)