计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20220800
59(10):2130-2163,2022

基于通用数据保护条例的数据隐私安全综述

赵景欣[1,5] 岳星辉[1,5] 冯崇朋[2,5] 张 静[2,5] 李 印[3,5] 王 娜[4,5] 任家东[1] 张昊星[6]
伍高飞[4,5] 朱笑岩[7] 张玉清[1,2,3,4,5]
[1](燕山大学信息科学与工程学院 河北秦皇岛 066004)
[2](西安邮电大学网络空间安全学院 西安 710121)
[3](海南大学网络空间与安全学院(密码学院) 海口 570228)
[4](西安电子科技大学广州研究院 广州 510555)
[5](国家计算机网络入侵防范中心(中国科学院大学) 北京 101408)
[6]中国信息通信研究院安全研究所 北京 100191)
[7](西安电子科技大学通信工程学院 西安 710071)
(zhao104825@stumail.ysu.edu.cn)

Survey of Data Privacy Security Based on General Data Protection Regulation

Zhao Jingxin[1,5], Yue Xinghui[1,5], Feng Chongpeng[2,5], Zhang Jing[2,5], Li Yin[3,5], Wang Na[4,5], Ren Jiadong[1], Zhang Haoxing[6], Wu Gaofei[4,5], Zhu Xiaoyan[7], and Zhang Yuqing[1,2,3,4,5]
[1](*School of Information Science and Engineering, Yanshan University, Qindhuangdao, Hebei* 066004)
[2](*School of Cyberspace Security, Xi'an University of Posts & Telecommunications, Xi'an* 710121)
[3](*School of Cyberspace Security(School of Cryptology), Hainan University, Haikou* 570228)
[4](*Guangzhou Institute of Technology, Xidian University, Guangzhou* 510555)
[5](*National Computer Network Intrusion Protection Center (University of Chinese Academy of Sciences), Beijing* 101408)
[6](*Security Research Institute of China Academy of Information and Communications Technology, Beijing* 100191)
[7](*School of Telecommunication Engineering, Xidian University, Xi'an* 710071)

Abstract With the gradual acceleration of the global digitalization process, data has now become an important factor of production in society. The flow of data has created infinite value for the society, but it also has huge hidden privacy risks. With the introduction of the EU General Data Protection Regulation (GDPR), personal data security has become a sensitive topic in the era of big data, and it has been paid more and more attention by researchers. Firstly, we review the development of data privacy security, introduce the EU General Data Protection Regulation (GDPR) and analyze its applications and influences. Secondly, we summarize the domestical and abroad related research literatures in recent years. We divide the GDPR compliance into three aspects: GDPR violation analysis, privacy policy analysis and GDPR model framework, and analyze the research status of these three aspects. We analyze the data technology based on GDPR, and discuss the application of GDPR in

收稿日期:2022-09-14;**修回日期**:2022-09-28
基金项目:国家自然科学基金项目(U1836210,61941105,61772406);海南省重点研发计划项目(ZDYF202012,GHYF2022010);河北软件工程重点实验室项目(22567637H)
This work was supported by the National Natural Science Foundation of China (U1836210, 61941105, 61772406), the Key Research and Development Science and Technology of Hainan Province (ZDYF202012, GHYF2022010), and the Project of Hebei Key Laboratory of Software Engineering (22567637H).
通信作者:张玉清(zhangyq@nipc.org.cn)

specific fields such as blockchain and Internet of Things respectively. Finally, according to the limitations of the existing research work, we summarize the main challenges and opportunities of data privacy security research based on GDPR, and put forward some inspirations for data privacy protection in China.

Key words data privacy protection; General Data Protection Regulation (GDPR); privacy policy; compliance; cross border data flow; data protection impact assessment

摘 要 随着全球数字化进程逐渐加快,数据已经成为当今社会重要的生产要素.数据的流动为社会创造了无穷的价值,但也潜藏着巨大的隐私风险.随着欧盟通用数据保护条例(GDPR)的出台,个人数据安全成为了大数据时代下的敏感话题,也越来越受到研究人员的重视.首先,对数据隐私安全发展历程进行了回顾,介绍了欧盟数据保护条例 GDPR 及其应用领域和影响;其次归纳分析了近几年国内外相关研究文献,将 GDPR 合规问题划分为 3 个方面:GDPR 违规行为分析、隐私政策分析、GDPR 模型框架,并分析了这 3 个方面的研究现状.总结分析了基于 GDPR 的数据技术,并分别探讨了 GDPR 在区块链、物联网等具体领域的应用;最后,根据现有研究工作存在的不足与问题,指出了基于 GDPR 的数据隐私安全研究面临的主要挑战和机遇,并针对中国数据隐私保护提出了一些启示.

关键词 数据隐私保护;通用数据保护条例(GDPR);隐私政策;合规性;跨境数据流动;数据保护影响评估

中图法分类号 TP391

随着大数据时代的飞速发展,数据成为了当今世界最宝贵的资源之一.企业也纷纷进行数字化转型,在数字经济时代下,数据的社会价值和经济价值不断凸显.然而数据的共享、加工、使用的过程又给数据的隐私安全带来了极大的风险,数据可以产生无数的副本,且形态多样化,如何实现数据的隐私保护成为了现今亟待解决的难题.

由于欠缺有力的监管机制,个人数据隐私遭到侵犯的事件屡屡发生,个人数据隐私时刻面临被泄露的风险.诸如,Uber 公司为掩盖 2016 年 60 万司机和 5 700 万用户信息失窃事件,私下向作恶者支付封口费,这项隐瞒行为也为公司带来了巨额的罚款[1].美国互联网公司雅虎在 2017 年承认公司曾在 2013 年受黑客袭击并泄露了所有用户信息(约 30 亿用户)[2].安全研究人员阿隆·加尔在2021 年 1 月发现由于入侵者利用了 Facebook 在 2019 年 8 月修复的漏洞,来自 106 个国家的超过 5.33 亿 Facebook 用户的个人信息已被免费在线泄露,涉及了不少知名人士和公众人物,还包括 67 万的国内用户[3].v.pnMentor 的研究团队在 2021 年 8 月份发现,B2B 营销公司 OneMoreLead 将至少 6 300 万美国人的私人数据存储在一个不安全数据库中,该公司任由此数据库完全敞开[4].2021 年 8 月,美国电信巨头 T-Mobile 官方确认服务器被黑客入侵,本次入侵大规模影响了大约 780 万 T-Mobile 后付费用户、850 000 名 T-Mobile 预付费用户以及大约 4 000 万以前或潜在用户,导致 T-Mobile 支付了 3.5 亿美元的索赔[5].

为了更好地保障个人权利,堪称史上最严格的数据隐私保护法案——《通用数据保护条例》(General Data Protection Regulation, GDPR),于 2016 年 4 月由欧盟议会通过,并于 2018 年 5 月 25 日起生效.GDPR 的出台使欧盟对个人信息的保护及监管达到了前所未有的高度,并统一了欧盟成员国有关数据保护的法律法规.虽然 GDPR 的保护范围只限于欧洲生活的民众,但由于互联网的全球性和开放性,几乎所有的服务都会受到隐私政策的限制,所以 GDPR 也通过各种机制对欧盟以外的国家产生了广泛的影响.

本文主要侧重于基于 GDPR 的数据隐私安全工作研究,为此对自 2016 年到 2022 年 6 月期间的网络与信息安全领域的四大顶级会议 USENIX Security (USENIX Security Symposium),NDSS(Network and Distributed System Security Symposium),CCS (ACM Conference on Computer and Communications Security),IEEE S&P(IEEE Symposium on Security and Privacy)的论文,及来自 Web of Science 核心数据库、EI 数据库、arXiv、中国知网(CNKI)等国内外数据库收录的相关论文进行了深入调研分析,如图 1 所示,相关的文献数量正在逐年增加.同时对基于 GDPR 的数据隐私安全领域的现有研究成果进行了

总结归纳,指出了现有研究工作不足和基于 GDPR 的数据隐私安全面临的挑战和机遇,为未来的安全研究工作指出了方向,并探讨了 GDPR 为中国的数据隐私安全工作带来的启示.

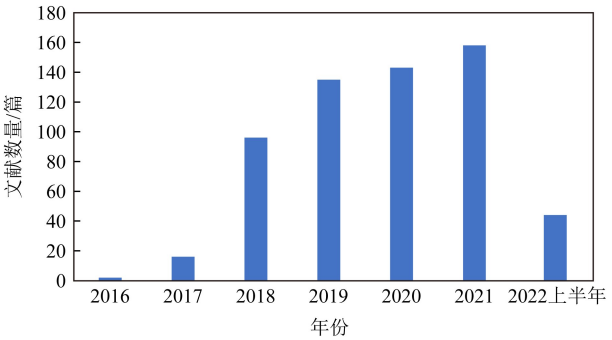

Fig. 1 Literature number of data privacy security based on GDPR

图 1 基于 GDPR 的数据隐私安全文献数量

自 GDPR 出台以来,国内外已围绕 GDPR 展开了许多相关的研究工作,其中不乏针对 GDPR 的隐私保护综述研究,但主要都是针对某一特定领域,如区块链[6]、物联网[7-9]等领域,或针对条例中的某部分规定[10]展开讨论.本文在关注 GDPR 在特定领域应用的同时,还聚焦于 GDPR 政策本身的可读性与完整性,不仅分析了更正权与被遗忘权的争议,还对知情权、访问权、数据保护影响评估等规定的合规方法进行了探讨.同时从国情出发分析了 GDPR 对中国产生的启示,为中国的数据隐私安全工作提出建议.

本文的主要贡献包括 5 个方面:

1) 分析了数据隐私安全的发展历程与现状,介绍了欧盟出台的影响广泛的数据保护法案《通用数据保护条例》(GDPR),分析了 GDPR 的应用领域及其带来的影响.

2) 深入调研了近几年国内外 GDPR 合规性相关的研究文献,从 GDPR 合规检测、隐私政策分析、GDPR 模型框架 3 个方面总结了 GDPR 合规性研究现状.

3) 总结分析了基于 GDPR 的数据技术,包括数据保护影响评估和数据跨境流动 2 个方面,并分别探讨了 GDPR 在区块链、物联网等具体领域的应用.

4) 通过分析 GDPR 合规的潜在安全问题以及现有研究工作的不足,指出了基于 GDPR 的数据隐私安全研究中面临的挑战与机遇,为相关的隐私安全研究指出了未来的研究方向.

5) 结合 GDPR 出台后的实施情况,从 6 个方面探讨了 GDPR 给中国的数据保护工作带来的启示.

1 相关背景介绍

1.1 数据隐私安全发展历程

为了保护个人数据隐私,需要有一定的法律制度为处理个人信息时提供保障.早在 1970 年德国联邦黑森州就通过数据保护法来保护数据隐私,瑞典在 1973 年通过了数据保护法,美国政府也在 1973 年制定了的公平信息惯例(FIPs).经合组织《1980 年 9 月隐私保护准则》列出了 8 项数据处理原则:收集限制原则、数据质量原则、目的规范原则、使用限制原则、安全保障原则、开放性原则、个人参与原则和问责原则,为各国制定个人数据处理法律提供了依据[11].随着数字经济社会的兴起,个人数据隐私问题也越来越多,于是越来越多的数据保护法条例在世界各国涌现.据美国法学教授 Bertil Cottier 统计,截至 2020 年共有 142 个国家发布了数据隐私立法,例如《1998 英国数据保护法案》《2008 年阿尔巴尼亚数据保护法》《2012 年加纳数据保护法》《2012 美国消费者隐私权利法案》、欧盟《通用数据保护条例》(GDPR)、《加州消费者隐私法案》(CCPA)、《巴西通用数据保护法》(LGPD)、《2019 年肯尼亚数据保护法》、《新加坡个人信息保护法例》(PDPA)、《中国个人信息保护法》等[12]. 其中欧盟出台的 GDPR 影响力最大,我们在这篇论文中主要对 GDPR 的相关研究进行了讨论. 也有研究者对各国的数据保护法进行了对比和评估,文献[11]根据经合组织指南对欧盟 GDPR、《2012 年加纳数据保护法》和《2019 年肯尼亚数据保护法》进行了比较,三者在一些原则的应用方面略有不同,加纳保护法缺乏数据可移植性的权利和记录个人数据泄露的义务,肯尼亚保护法包含了所有与个人和数据主体的权利和义务有关的经合组织修订原则,在很大程度上复制了 GDPR. 本文从适用范围、数据主体权利、数据处理者责任等要点出发,对各国出台的数据保护法进行了对比,具体内容如表 1 所示.

GDPR 的出台推动了许多国家对于数据保护方面的立法进程,但由于国情与隐私文化的差异,各国相应的应对措施不尽相同.如欧洲强调个人权利,以保护人权为出发点,因此 GDPR 法规要求严格,内容全面;印度、澳大利亚纷纷依照 GDPR 对自己的隐私法规进行了审查和修改,增强了监管机构的权利并加大了处罚力度;巴西借鉴了 GDPR 的主要结构出台了第一部综合性的数据保护法,但在处罚方面

Table 1 Comparison of Data Protection Laws in Different Countries[13]

表 1 各国数据保护法对比[13]

各国法规	适用范围	数据主体权利	数据处理者责任
2021 年中国《个人信息保护法》	境内处理个人信息及境外为境内个人提供服务方	知情权、决定权、查阅权、复制权、更正权、删除权及获得解释权等	管理制度和操作规程、分级分类管理、加密、去标识化等安全技术措施等
2019 年印度《个人数据保护法》	境内数据及境外处理境内数据方	确认和访问权、被遗忘权、数据可移植权、纠正权等	当重要数据受托人满足一定条件或情况时,需附加对应的责任和义务
2018 年美国《加州消费者隐私法案》(CCPA)	加州开展业务的公司	访问权、知情权、删除权、拒绝权、不受歧视权、儿童的选择加入权等	在官网首页作出“不要售卖我的个人信息”链接;在隐私政策或 CCPA 特殊页中给出消费者权利等
2018 年欧盟《通用数据保护条例》(GDPR)	所有向欧盟民众提供商品和服务或收集并分析欧盟居民相关数据的组织	访问权、删除权、知情权、拒绝权、可携带权等	记录数据处理活动;进行数据保护影响评估;任命数据保护官;实施技术安全措施;数据泄露时进行通知;问责制
2018 年巴西《巴西通用数据保护法》(LGPD)	任何境内的数据处理操作	访问权、删除权、知情权、拒绝权、可携带权、匿名权、拦截权、消除权等	记录数据处理活动;根据数据控制者的指示处理个人数据;从控制者处接收并支持实现数据主体对数据的更正、消除、匿名化或阻塞的要求
2020 年韩国《个人信息保护法》(PIPA)	境内和本国跨境公司处理的境内个人数据	访问权、更正权、暂停或删除权、拒绝权等	最小化原则;匿名化;建立内容管理个人数据的规制,包括保留访问日志;在处理个人数据时进行通知;个人信息和敏感个人信息分别获得同意
2020 年日本《个人信息保护法(修订)》(APPI)	境内处理个人数据的所有经营者	访问权、公开权、知情权、更正权、删除权、拒绝权、要求停止处理权	要求数据控制者对第三方进行管控和监督,包括执行数据控制者与服务提供商之间的协议;向服务提供商提供安全措施;围绕数据处理行为指示和调查服务提供商
2020 年新加坡《个人信息保护法(修订)》(PDPA)	所有向境内民众提供商品和服务或收集并分析境内居民相关数据的组织	访问权、删除权、知情权、拒绝权、可携带权	任命数据保护官、实施技术安全措施、问责制

较 GDPR 宽松很多;美国注重企业发展,强调数据利用,更偏向从消费者的角度对数据进行监管;我国在数据立法方面也并没有照抄照搬欧洲立法,而是兼顾个人权利与经济发展,探索出一条适合自己的发展道路.在数据的跨境流动中,面对更多的网络安全威胁和不同国家数据保护法的不同要求,特别是面对非常严格的 GDPR,世界各国也在不断地修改并完善数据保护法.2021 年 1 月,韩国个人信息保护委员会向社会公布了《个人信息保护法(修正案草案)》(PIPA),并在一年内修订了 3 次.2021 年 5 月

Fig. 2 Global data security protection legislation (partial)

图 2 全球数据安全保护立法情况(部分)

12 日,日本国会通过了包括《为形成数字化社会完善相关法律的法案》在内的 6 部数字化改革法律案,《个人信息保护法》(APPI)修正案也作为完善法案的一部分同时生效.2021 年 6 月,中国通过了《中华人民共和国数据安全法》.2021 年 8 月 20 日,《中华人民共和国个人信息保护法》历经三审正式通过.图 2 展示了全球一些具有代表性的数据保护法案[14].

1.2 GDPR 介绍

欧盟通用数据保护条例(GDPR)是关于欧盟(EU)和欧洲经济区(EEA)数据保护和数据隐私的法律条例,是欧盟隐私法和人权法的重要组成部分.

GDPR 规定了与个人数据处理以及个人数据自由流动相关的自然人保护法规,旨在尊重自然人的基本权利和自由,并重点强调其保护个人数据的权利.GDPR 被誉为是最严格的个人数据保护和数据监管条例,适用于欧洲经济区内数据主体产生的所有数据,无论收集相关数据的企业是否位于欧盟境内都要遵守 GDPR.GDPR 于 2016 年 4 月 14 日获欧洲议会和欧盟理事会通过,并于 2018 年 5 月 25 日开始强制实施.该法规取代了 1995 年的《数据保护指令》(95/46/EC)(简称 95/46/EC 指令),解决了 95/46/EC 指令成员国在处理个人数据时对保护自然人权利和自由水平之间的差异,具有直接的约束力和适用性[15].同时,《电子隐私指令》(2002/58/EC)旨在补充 GDPR 并完成协调过程,目前该法案正在通过欧盟的立法程序[16].图 3 展示了 GDPR 从立项到实施过程中的关键日期和事件.继 GDPR 之后,欧盟《数据法案》《数据治理法案》《数据市场法案》等一系列数据治理法规的出台也展现了构建未来数字驱动创新生态的欧洲方案.

Fig. 3 Key time points for GDPR legislation

图 3 GDPR 立法的关键时间点

本节主要从 GDPR 框架、处理个人数据相关原则、数据主体的基本权利以及违规行为的补救措施、责任和处罚 4 个方面来具体介绍 GDPR.

1.2.1 GDPR 框架

GDPR 框架如图 4 所示.通用数据保护条例共包含十一章内容,涉及一般规定、原则、数据主体权利、

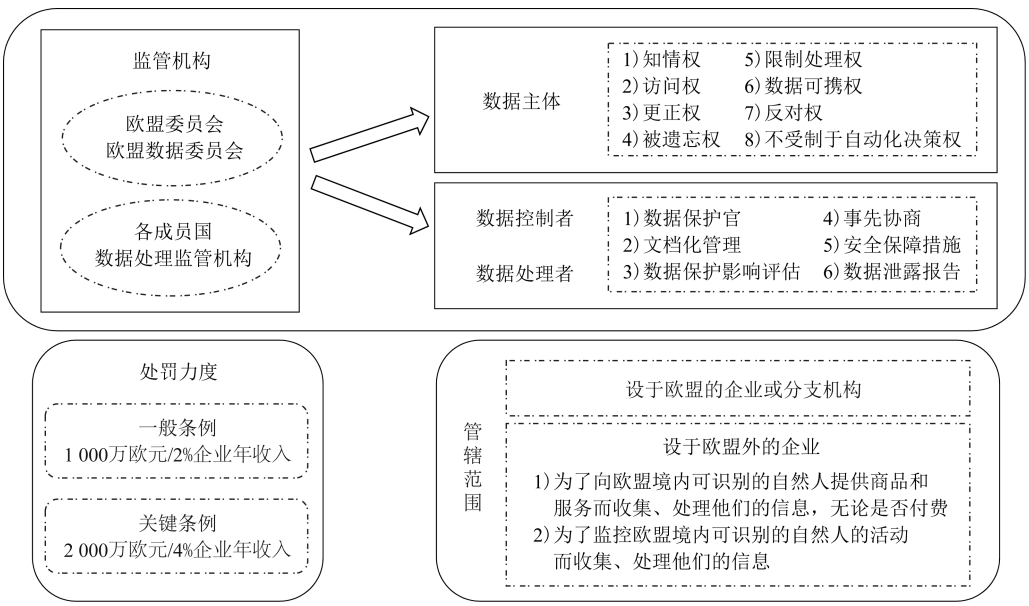

Fig. 4 The framework of GDPR

图 4 通用数据保护条例框架

数据控制者和处理者义务、向第三国或国际组织传输个人数据、独立监管机构、成员国之间的合作与一致性、补救措施,责任和处罚、有关特定处理情况的规定、授权和实施法案以及最终条款[15].其中定义了与个人数据相关的3种不同实体:1)数据主体,即个人数据所有者;2)数据控制者,即收集和使用个人数据的个人或组织;3)数据处理者,即为控制者处理个人数据的个人或组织.同时,任命具有数据保护法和实践专业知识数据保护官(DPO),以协助数据控制者和处理者监控其法规的遵守情况.

1.2.2 处理个人数据相关原则

GDPR规定了7项处理个人数据相关原则(第5条)包括"合法性、公平性和透明度""目的限制""数据最小化""准确性""存储限制""完整性和保密性"的数据处理原则以及控制者责任与义务的"问责制",具体原则内容如表2所示.并对处理的合法性进行说明(第6条),且只有满足至少一项原则才认为该处理是合法的:1)数据主体已经同意处理其个人数据;2)履行与数据主体的合同义务,或在签订合同时采取相应措施满足数据主体要求;3)遵守数据控制者的法律义务;4)保护数据主体或其他个人的利益;5)为公共利益或数据控制者官方权力执行任务;6)在利益与被保护数据主体的利益、基本权利和自由不相冲突的情况下,保护数据控制者或第三方的合法利益.

Table 2 Principles Related to the Processing of Personal Data

表2 处理个人数据相关原则

原则	具体要求
合法性、公平性和透明度	必须清楚告知数据主体其数据将如何使用,且满足合法、公平和透明的处理方式
目的限制	依据特定、明确且合法的目的收集数据,且不会以其他目的进一步处理数据
数据最小化	充分相关且仅限于特定目的的必要数据量
准确性	确保数据控制者在收集个人数据时的准确定性并在必要时进行更新或删除
存储限制	存储时间不超过处理个人数据的目的所必需的时间
完整性和保密性	采取适当安全措施保护个人数据,以防止未经授权或非法处理
问责制	数据控制者及处理者应能证明遵守以上原则,并对其行为负责

1.2.3 数据主体的基本权利

GDPR详细阐明了数据主体的基本权利,共涉及8项权利:1)知情权(第12,13,14条).数据控制者以简洁、透明、可理解和易于访问的形式向数据主体提供信息;2)访问权(第15条).数据主体有权要求数据控制者告知其个人数据是否正在被处理.数据控制者必须根据要求提供正在处理的数据的目的(用途)、数据类别、存储期限或标准并为数据主体提供一份实际数据的副本;3)更正权(第16条).数据主体有权更正错误的个人数据;4)删除权(第17条).数据主体有权要求控制者及时删除有关的个人数据;5)限制处理权(第18条).在特定场景下,数据主体有权要求数据控制者限制对他的个人数据的使用;6)可携带权(第20条).数据主体有权要求将自己的数据转移到另一家数据控制者,数据控制者应当配合;7)反对权(第21条).允许个人反对出于营销或非服务相关目的处理个人信息;8)不受制于自动化决策(第22条).数据主体有权不受基于自动化决策所做决定的影响.

1.2.4 违规行为的补救措施、责任和处罚

针对违规行为的补救措施、责任和处罚,GDPR也做出了相应的限定,要求数据控制者必须在违规行为发生后72 h内通知监管机构,依据违规的严重程度、违规的持续时间、受违规影响的数据主体数量以及违规造成的损害程度来处罚违规行为责任方.GDPR的出台对其他国家及地区的个人数据相关法律产生较大的影响,成为全球个人数据保护法的典范.

1.3 GDPR的应用和影响

GDPR的实施影响了各行各业,对数据隐私的立法极大提高了公民的隐私权,在不同的领域内产生了积极影响,本节以医疗健康和物联网为例阐述了GDPR产生的积极影响,同时也讨论了GDPR的潜在风险.

现阶段的医疗正在经历数字化转型,向个性化、预防性和精准医疗进行转变,由于个人的健康状态、条件和背景都是高度动态的,导致了分布式、高复杂度的业务流程,因此不可能以静态的方式进行全局管理.随着个人可穿戴设备数量的指数级增长,利用个人健康数据进行分析有很多的益处,比如识别医疗服务中的风险和成本、提高服务效率、疾病预防等,但同样也带来了更多的用户隐私泄露风险.

GDPR扩展了个人数据的定义范围,包括自然人身体、生理、遗传、经济、文化或社会身份的特定因素;进一步的,GDPR定义了对个人数据进行处理的要求,以确保在处理用户数据过程中的合理合规,此外,由于系统环境的变化,GDPR要求处理用户数据时进行动态的管理.在GDPR的规范下,数据的保护者变成了风险的管理者,必须积极主动地动态管理

系统,这对于医疗健康类个人敏感信息的处理具有指导意义.在GDPR的驱动下,未来医疗系统对个人数据的处理应该是一个政策驱动的多领域自动化业务系统,将政策和业务流程中的数据对象进行绑定,在这个过程中保证数据处理的高透明度以保证用户的知情权,GDPR很好地适应了医疗的数字化转型,保护了个人健康敏感数据的隐私.

近年来物联网设备数量呈现井喷式增长,同时也意味着设备厂商针对个人数据进行大量的存储、分发和利用,从厂商的角度来看,分析这些数据可以更好地理解用户的行为,及时发现消费者的行为模式和使用某类设备的关系,能够帮助厂商对产品进行进一步的改进以提高用户体验,但是也存在厂商在用户不知情的情况下将这些数据出售给第三方,或者从同一用户的不同设备同时收集数据建立用户画像的情况,这进一步增加了用户的隐私风险.

在此情况下,GDPR的实施使得信息的控制权大大地转向了个人.首先,这些收集的用户数据在GDPR的扩展定义中都属于个人隐私数据,其次,GDPR要求对个人数据的收集和处理必须基于明确的用户同意,并且用户有权在任何时候撤销自己的同意,否则将面临严重的罚款.根据GDPR官方的处罚规定[17],在处罚方面将有一个两级的制裁制度,若是情节较轻的违规行为,可导致1 000万或公司全球营业额的2%的罚款(以较高者为准),最严重的违规行为可能导致2 000万或公司全球营业额的4%的罚款(以较高者为准).因此GDPR的实施是对物联网设备厂商极大的警告,迫使他们按照GDPR的要求重新设计隐私政策和收集用户数据的范围,以及必须取得用户的知情同意.

总体来看,GDPR改善了网络安全,网络、服务器和其他基础设施的安全升级是网络安全的保障,GDPR直接影响了数据隐私的安全,鼓励企业制定政策和升级设备来预防潜在的安全风险;其次,GDPR将数据保护进行标准化,在欧盟国家直接实施,建立了区域统一数据保护标准,而无需建立每个国家的个人数据保护法.

GDPR也带来了一些负面的影响,比如极其严厉的两级处罚措施,一旦企业因为各种原因未能保护好用户数据导致泄露将会付出巨大代价,而中小型企业对数据的保护能力和抗风险能力本身就较弱,一旦遭遇此类事件将对企业造成很大的打击,中小型企业的隐私保护意识也相对薄弱,根据网站superoffice[18]的统计,截止2021年5月,超过四分之一的企业尚未根据GDPR进行整改,由此可见部分企业并未意识到GDPR的重要性.严格的GDPR给新兴的物联网企业带来了繁重的负担,合规工作消耗了大量的资源,使得企业的业务运营变得更加艰难.此外,企业必须进行合规性的审计,需要招聘更多专业的隐私保护方面的人才,因此,带来了更多成本的负担,相对应的也给执法机构带来了新的挑战.而对于需要留存大量患者信息的医疗企业来说,如何实现高度敏感数据的安全存储仍是一个十分严峻的挑战.随着全球数据隐私安全意识的增强,医学实验的开展也受到重重阻隔,影响了医学科技的高效发展.

2 GDPR合规性研究现状

通过对现有的基于GDPR的数据隐私安全研究工作进行梳理和分析,发现目前相关研究方向主要集中在GDPR合规检测、隐私政策分析、GDPR模型框架3个方面.图5中给出了现有研究工作文献数量的占比情况,以便读者有一个直观的认识.本节将通过这3个研究方向分类阐述现有具有代表性的GDPR相关的研究工作,同时在现有研究工作基础上,本文将对每个研究方向的具体工作进行对比分析和讨论,并给出观点,供感兴趣的研究人员对该领域进行进一步研究.

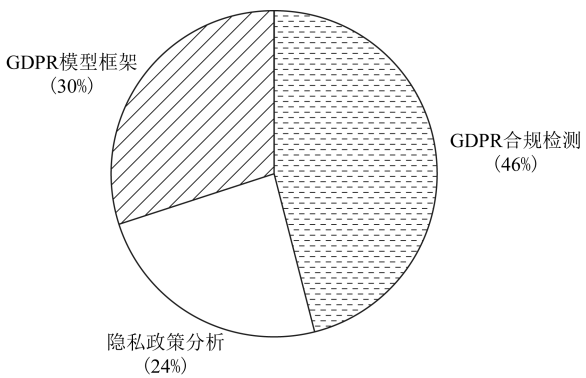

Fig. 5 Proportion of GDPR compliance related studies

图5 GDPR合规性相关研究占比

2.1 GDPR合规检测

自2018年5月25日起,欧盟开始实施《通用数据保护条例》(GDPR),条例涉及个人数据处理和数据隐私保护,直接适用于所有成员国.GDPR旨在保护欧盟成员国内所有公民的个人数据隐私,并对违规行为实施严厉制裁.数据隐私保护网站DataPrivacyManager披露了2020—2022年欧盟国

家根据 GDPR 对企业的数笔大额罚款[19],如图 6 所示,罚款金额从数百万欧元到数亿欧元不等,拥有越多用户数据的企业遭受处罚的风险也更大.

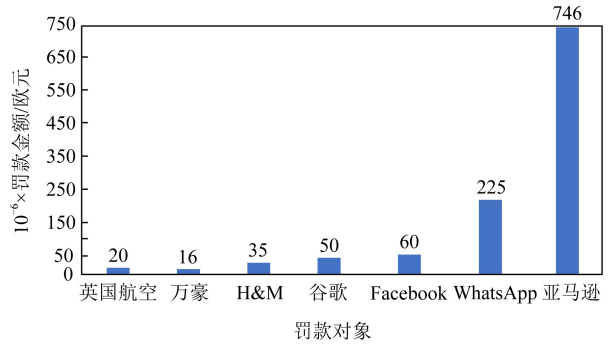

Fig. 6 Large GDPR fines from 2020 to 2022

图 6 2020—2022 年 GDPR 大额处罚情况

2.1.1 违规案例分析

根据文献[20]分析的 277 项制裁案例,违规行为主要分为 4 种类型:非法处理个人信息、披露个人信息、未保护个人信息和与监管机构合作不足.这些处罚主要针对违反 5 项主要与用户隐私保护相关的特定条款.由此可见,违规的主要原因是企业未能充分向用户披露他们的个人信息是如何被收集的,以及没有明确告知用户收集这些信息的用途,并且在利用这些数据时未适当征得用户的同意,具体表现形式分为违规采集和传输、隐私政策不规范、数据滥用、网站 cookie 跟踪 4 个方面.

1) 违规采集和传输

文献[21]对一些流行的健康软件进行安全审计,结果表明被分析的应用程序大多数都没有遵守 GDPR 要求的法律限制,在没有征求用户同意的情况下,通过不同的方式违规收集用户的个人敏感信息,从而威胁了数百万用户的隐私.根据 GDPR 的要求,在线服务必须获得用户的同意才能与第三方共享用户数据,文献[22]通过检测 Android 应用程序中事先未经用户同意发送到互联网的数据表明 30%应用程序在未经用户事先明确同意的情况下将个人数据发送给第三方,由此可见对于用户数据的滥用目前在业内非常普遍.

2) 隐私政策不规范

近年来随着物联网设备(IoT)的兴起,因其涉及大量的个人隐私数据,应该具有相应的数据收集规范.文献[23]通过捕获物联网设备与云之间、物联网设备与其对应的在智能手机上的应用程序之间的数据流量,对 11 家物联网制造商进行分析测试,结果显示其中一半的物联网制造商没有专门针对其物联网设备制定对应的隐私政策,对于目前大量不同类型的物联网设备,只有大致的隐私政策框架是不够的,需要按照不同的应用场景对隐私政策进行细分.

3) 数据滥用

文献[24]披露了在 GDPR 实施之前,科技企业 Facebook 为 73%的欧盟用户贴上了潜在敏感兴趣的标签,从而针对性地推送个性化广告,并且恶意第三方可以以极低的成本获取已被分配潜在敏感兴趣的 Facebook 用户的身份.进一步的,文献[25]通过检测第三方广告和跟踪服务发现,广告商会在未经用户同意的情况下,与第三方关联公司共享收集到的数据,并且揭示了这类做法在业界已成为常态化,此类违规行为利用用户个人数据达到其商业目的,并未征求用户的同意.

4) 网站 cookie 跟踪

由于 GDPR 的实施,欧洲用户几乎在每个网站上都会遇到 cookie 同意选择框.通过 cookie,网站可以经用户同意收集数据并将其传播给第三方.文献[26]的研究表明,即使用户没有做出选择,部分网站也会默认用户已经同意 cookie 跟踪,或者用户已明确选择退出,部分网站也会存储用户同意.在测试中一半以上的网站至少存在一项可疑违规收集行为.并且,cookie 跟踪的范围非常广泛,文献[27]对欧盟境内 2 000 多个高流量网站通过 cookie 实现的跟踪进行评估,发现 cookie 可以在访问数据集中 90%以上的网站识别用户,因此用户很难避免被跟踪.此外,根据文献[28]对在线广告业务的研究,出于研究用户行为的目的,即使用户已经选择退出广告,网站也会继续追踪用户的浏览器.因此在现有的框架下,用户缺乏可行的机制来同意或者拒绝其在互联网上的行为被跟踪分析.

2.1.2 违规检测方法

若应用程序或者网站需要处理用户的个人数据,就要符合 GDPR 的要求,然而 GDPR 的隐私处理条款针对软件的开发过程只给出了一般性原则,而非详细的操作指南,因此现有软件和网站中可能存在大量违反 GDPR 法规的情况.但是先分析隐私泄露事件,而后确定后果和罚款需要执法机构付出大量的时间和精力来评估数据收集和处理机制是否符合 GDPR 条款,因此使用技术手段对违规行为进行规模性的检测和评估是十分有必要的.

移动应用程序经常访问用户的个人信息以满足

业务需求,由于此类信息通常很敏感,因此监管机构要求移动应用程序开发人员发布详细的隐私政策,文献[29]提出了一个半自动框架,建立隐私政策术语到 API 方法的映射集合,用于检测隐私政策和应用程序代码不一致的违规行为.文献[30]在文献[29]的基础上,将 GDPR 要求的隐私政策规范进行量化分析,通过自然语言处理(natural language processing, NLP)与机器学习算法,生成 6 个通知分类器来检测应用程序的隐私政策是否完整,并通过实验证明了自动化分析隐私政策的有效性.同样基于文本分析,文献[31]使用语义相似性来识别不同文章对应的和特定违规类型相关的主题,用识别的特征来训练一个长短期记忆(long short term memory, LSTM)深度学习器,可以有效识别给定文本描述的潜在违规行为.

进一步的,隐私政策的合规只是基本要求,在实际的应用场景中,还需要知道应用程序是否会收集隐私政策之外、没有取得用户同意的数据.文献[22]基于字符串匹配,利用半自动化工具来检测未经事先同意而发送到互联网上的数据流量,并检测了 86 163 个应用程序,发现有三分之一应用程序在未经用户事先明确同意的情况下向第三方发送个人数据.

此外,许多移动应用开发者为了各种目的在他们的应用中整合第三方服务,包括应用维护分析服务、用户参与、社交网络整合和广告.第三方服务访问大量有价值的用户数据,这些数据往往超出了它们向应用开发者或用户提供服务的需要,并且在用户不知情的情况下进行跟踪.文献[32]根据应用程序无限制收集个人信息的情况,提出了一种基于关联挖掘的个人身份信息泄漏检测方法,设计并实现了一个自动化系统,用于检测 APP 发送的流量数据是否暴露了用户的个人身份信息,有助于在流量数据中发现隐藏的隐私泄露.文献[25]使用 URL 分类器来自动检测流量级别的第三方广告和跟踪服务,使用这种技术识别出 2 121 项此类第三方服务,其中 233 项不为其他流行的广告和跟踪黑名单所知.第三方跟踪服务的隐秘性和高权限使得该类行为具有较大的违规风险,一种解决隐秘性的可行方案是利用区块链技术增强数据使用的透明度,文献[33]提出了一种利用区块链和智能合约技术开发符合 GDPR 的个人数据管理平台的方案,该方案为服务提供商和数据所有者提供去中心化机制用于处理个人数据,确保只有指定方可以处理个人数据,并使用智能合约和加密技术将所有数据活动记录在不可变的分布式账本中,任何违规行为都会被永久记录下来并自动报告,该方法也能有效地解决数据所有者无法感知服务提供商是否遵守 GDPR 并且有效保护了其个人数据的问题.表 3 整理了这 5 种典型的违规检测手段,分别从违规检测方法、分析对象以及违规行为 3 方面进行展示.

Table 3 Analysis and Comparison of Violation Detection Methods Based on GDPR

表 3 基于 GDPR 的违规检测方法分析对比

文献	违规检测方法	分析对象	违规行为
文献[22]	字符串匹配	APP 流量	未经同意的数据收集
文献[25]	URL 分类器	APP 流量	第三方跟踪
文献[30]	NLP+随机森林	隐私政策	隐私政策不完整
文献[31]	LSTM	任意文本	给定文本潜在违规
文献[33]	以太坊智能合约	个人敏感数据	未经同意的数据收集与处理

2.1.3 规避违规

企业使用包括企业网站、社交媒体资料、在线商店等媒介和用户进行交互,特别是当前社交媒体已成为重要的企业平台.由于这些媒介处理着大量用户数据,因此企业必须考虑出台新的数据保护和隐私保护政策以适应 GDPR 条例,履行 GDPR 所规定的义务,确保其软件系统达到 GDPR 的要求.本节总结了增加隐私政策可读性、增加数据访问透明度及同意管理 3 种规避违规的手段.

1) 增加隐私政策可读性

隐私政策是用户了解哪些个人信息被收集和使用的重要媒介,但是隐私政策的可读性普遍较差,结合其他复杂性使其无法达到预期目的.文献[34]引入了一种基于自然语言处理技术的隐私政策摘要工具,该方案能够以高准确度将隐私政策进行分类以及阐明相关的风险级别.进一步的,文献[35]提出的方法可以将相当长的隐私政策总结为简短而浓缩的注释,从而让用户更准确的辨别数据收集的范围.

2) 增加数据访问透明度

增加数据收集中的透明度有利于数据收集安全.文献[36]调研了谷歌工具“我的活动”,尽管大多数参与者并不关心数据收集,但是可以随时查看数据收集情况使得大部分用户增加了对产品的信任度.同样的,文献[37]介绍了一种用 Web 界面从不同在线服务导出数据收集和处理情况进行可视化展示,有效提高了用户对在线服务收集数据的行为认识.

在某些情况下,服务提供商并非真正需要采集到个人敏感数据,采集的目的可能只是为了收集丰富的某项数据以满足其分析的需求,文献[38]提出的区块链系统采用类似零知识证明的机制,允许用户在不透露其身份的情况下证明拥有某些属性,最大限度地在满足服务提供商需求的情况下使得用户提供的信息最少,因此 GDPR 所强调的数据最小化原则在一定程度上可以增加数据采集的安全性.

3) 同意管理

由于同意通知的复杂性和动态性,必须执行合规性验证或审计来保证数据采集的合规.采用工具进行验证是一种有效的手段.文献[39]提出了隐私政策和同意管理需要的机器策略语言,使用推理器进行语义合规性检查.基于此项研究,文献[40]提出了一种数据保护设计工具,将 GDPR 法规转换为软件代码,从而实现自动化合规性验证.文献[41]认为通过确保数据处理中使用的数据集从一开始就符合同意,使用给定同意的结构化表示来“实时”生成数据集,可以增加透明度,方便用户给予、撤回他们对系统数据处理的同意许可,减少了事后进行遵守情况分析的需要.

根据 GDPR 的要求,服务提供商需要告知用户他们的数据收集情况,经过用户允许才能收集特定的数据.文献[42]通过对数千名参与者的调研得出,用户普遍在服务提供商数据收集和数据使用的解释上没有仔细地阅读,削弱了同意通知的作用,这表明了用户体验需要进一步改进.此外,对于应用程序和网站的开发人员,根据文献[43]的研究,只有不到四分之一的专业人员能够接触到安全专家,而且很少有技术人员因为欧洲 GDPR 立法而对其软件进行针对性的优化,因此规避违规行为还需要很长一段路由走,用户既要提高隐私保护意识,明确 GDPR 赋予的权利,企业也要采取积极手段响应政策,避免受到处罚.

2.1.4 小结

2.1 节主要从典型的 GDPR 违规案例出发,对基于 GDPR 的合规检测研究工作进行了分析和讨论,并给出了 GDPR 合规检测研究领域的一些观点.

讨论 1. 本节从不同角度分析和归纳了几种违规检测方法,对应着不同的违规行为.从分析对象来看,违规检测方法针对的是不同场景、不同阶段下的特定行为,基本覆盖了数据收集、处理的各个阶段;从分析方法来看,目前大多数文献主要集中于使用自动化工具进行违规识别,此类问题的解决方法大都是以机器学习技术为基础构建的自动化工具进行分析,在各自的实验场景中表现出了出色的检测效果,此外,使用区块链技术进行违规检测的研究目前较少,且都是以理论框架构建为主,如何将区块链和智能合约技术进行有效地应用仍然需要研究人员进一步探索.

观点 1. 目前的自动化检测的手段大都围绕机器学习算法,实验对象也是特定场景下的应用程序或网站,有一定的局限性.并且由于分析的对象并不相同,数据集的训练性能有一定的针对性,能否在跨平台跨类别的应用程序上达到相近的性能值得进一步的探讨.进一步的,违规行为分布在数据收集、处理、共享、流动的各个阶段,每个阶段所涉及到的场景都是非常多样的,这给违规行为检测带来了很大的挑战,如何找到违规检测在不同阶段的普适性方法以及如何针对不同场景进行优化是未来的一大难点.

2.2 隐私政策分析

隐私政策是一份声明或法律文件,它向用户披露数据收集、使用、存储和共享的部分或全部方式,使用户能够在注册任何服务或决定是否继续使用服务时做出明智的决定,是数据控制者和用户之间信息传播的主要媒介.随着数据隐私保护成为一个重要的社会问题,不同国家和地区都制定了相应的法律法规来保证用户数据的安全性和隐私性,其中最具有代表性的就是 GDPR.但是检测收集、处理或存储用户个人数据服务商的合规性是法律执行的一大困难挑战.这个困难主要来自于 2 个方面:1)GDPR 等法律法规是用自然语言编写的,包含了大量的法律术语,没有法律知识的用户很难读懂.2)隐私政策通常用冗长而复杂的文档展示,用户阅读起来非常耗时.文献[44]在 2008 年就指出,如果一个用户要阅读在互联网上访问的每一项服务的隐私政策,平均每年需要 244 h.因此当前研究的主要方向是自动地发现法规与隐私政策之间的合规性问题,并为数据主体(即用户)、数据收集方(即服务提供商)和监管当局提供直观的结果[45].

在 GDPR 出现之前,已有很多对隐私政策的分类研究,大多数方法都是利用自然语言处理技术对隐私政策进行分析[46-48],但使用的方法欠缺迁移性,在 GDPR 相关的隐私分析中并不适用.新兴的机器学习技术越来越多地被用于辅助隐私保护,通过对隐私政策的评估与分析,使政策更具可读性,并检测隐私政策中的模糊内容.文献[30]提出了一个自动系统 HPDROID,通过识别应用隐私政策中声明的数据实践和应用代码中的数据相关行为来弥合 GDPR 的一般规则和应用实现之间的语义鸿沟.该系统根据 GDPR 第 5 条相应的 3 个基本要求,即透明度、数据最小化、保密性,将自然语言处理技术与机器学习相结合,对 796 个移动健康应用程序隐私政策进行了检测,发现其中 189 个没有提供完整的隐私政策,HPDROID 提高了应用程序用户和开发者的隐私保护意识.

文献[35]在 2018 年受到 GDPR 和机器学习技术的影响,根据 GDPR 第 12,13 条的规定提出了风险指标,并使用了朴素贝叶斯、支持向量机(support vector machine, SVM)、决策树和随机森林 4 种有监督的机器学习技术.基于风险指标对冗长的隐私政策进行了分类,简化了隐私政策的解释,并提醒用户注意建议的风险指标.文献[49]在文献[35]的基础上,增加了数据集的范围,从网上爬取了 1200 个隐私政策,按照 5 项 GDPR 隐私政策核心要求进行标记,并增加了单词嵌入技术与监督学习相结合,对隐私政策进行了分类,发现超过 76%的隐私政策不满足 5 项基本要求,因此可能不完全符合 GDPR.文献[35,49]提出的各种基于机器学习的方法在一定程度上解决了隐私政策总结问题,但是他们使用的都是美国或者欧盟网站的数据集,对其他国家的网站效率并不高.文献[50]从 GDPR 和《巴基斯坦数保护法》中提取了 10 个隐私惯例,定义了 27 个类别标签,从 5 个部门的巴基斯坦网站编译了 120 条隐私政策的标记数据集,使用了 SVM、Logistic 回归、KNN 和朴素贝叶斯 4 个机器学习分类器对数据集进行了训练和测试,实现了对巴基斯坦网站隐私政策的合规性检查.

对隐私政策的大量研究都依赖于有监督的机器学习方法,这些方法需要标注隐私政策的数据集,但是这种公开的数据集很少,因此隐私政策语料库的建立极其重要.文献[51]基于众包创建了一个名为 OPP-115 的网站隐私政策语料库,其中包含 23 000 细粒度的数据实践.文献[52]扩展了 OPP-115 语料库,增加了标记“退出选择”的细粒度信息,该文献专注于自动识别隐私政策文本中的用户选择的任务.文献[53]引入了从 GDPR 条款到 OPP-115 注释方案的映射,证明了 OPP-115 的广泛适用性.文献[54]建立了一个包含 350 条移动应用隐私政策的语料库,并提供了一个可扩展的管道来分析带有隐私政策的 APP 可执行文件的潜在合规性问题.文献[55]提出了一种自动检测隐私政策中模糊词和句子的方法,通过众包创建了一个模糊词语料库.文献[35]向前迈出了一步,创建了一个包含 45 个手动标记的隐私政策的语料库,专注于由专家定义的隐私政策的风险级别.文献[45]根据 GDPR 第 13 条对隐私政策进行合规性分析,设计了一种基于 GDPR 的分类方案,并为此手动策划了 304 个隐私政策的语料库.对于语料库的扩大和填充,还需要研究人员进一步努力.

除了文献[35,49-50]对网站的隐私政策进行分析,还有许多研究对其他领域的隐私政策的分析.文献[56]根据 GDPR 一般规则,采用有监督的 NLP 技术对基金行业的 234 个隐私政策进行了检测.文献[45]从 Google Play6(应用程序商店之一)收集应用程序的隐私政策,涵盖了 22 个应用程序类别,并基于 GDPR 第 13 条的分类方案注释了一个包含 304 个隐私政策的语料库.算法采用了 SVM,以及基于嵌入的双向长短期记忆网络((bi-directional long short-term memory, BiLSTM)和基于上下文 Bert 网络 2 种具有代表的神经网络模型.文献[57]采用了文本模糊解释结构建模(textual fuzzy interpretive structural modeling, TFISM)确定了 GDPR 中的关键因素,并将它们与各种云服务隐私政策进行了比较,检测了 GDPR 与服务隐私政策之间对于不同关键术语或因素的优先级设置的相似性.文献[58]开发了一个集成的、语义丰富的知识图谱来表示 GDPR 所规定的规则,并将其应用于云隐私政策中对比语义相似性,大数据从业者可以利用该方法根据授权文件定期更新其参考文件.

小结:2.2 节从现有的基于 GDPR 的隐私政策合规性研究工作中,挑选和总结了 5 项具有代表性的研究工作,并给出了基于 GDPR 的隐私政策合规性研究领域的一些观点.表 4 分别从分析依据、数据集、算法以及最优算法多个角度进行分析和讨论.

讨论 2. 由表 4 可知,1)从分析依据而言,基于 GDPR 第 13 条的要求进行合规性分析占研究的多数.2)从数据集来看,大多采用的是英文的隐私政

策,数据集的范围领域在不断的扩大.3)从算法来看,文献[35,45,49-50]都采用了3个及以上的算法进行对比分析,结果都得出SVM算法对隐私政策的分析领域适用性最好.

观点2. 基于GDPR的隐私政策的合规性研究能够将隐私政策中数据收集、使用、存储和共享的部分或全部方式直观地展现给用户和服务提供商,促进了数据隐私保护领域的发展,其研究意义重大.通过对比和归纳现有工作,本文发现:1)相比于深度学习算法,SVM算法在隐私政策分类上有更好的结果,这或许是因为深度学习算法欠缺专业标注的数据集,同时也缺少大量的正样本来训练神经网络.2)隐私政策语料库目前大部分涉及的是英文,多语言融合的语料库有待研究人员进一步开发.

Table 4 Comparison of Privacy Policy Analysis Based on GDPR

表4 基于GDPR的隐私政策分析工作对比

文献	算法	最优算法	分析依据	数据集
文献[35]	朴素贝叶斯、SVM、决策树、随机森林	SVM	GDPR12、13条	45个欧洲网站的隐私政策
文献[49]	SVM、Logistic回归、KNN	SVM	GDPR五项核心要求	1200个网站隐私政策
文献[50]	SVM、Logistic回归、KNN、朴素贝叶斯	SVM	GDPR和《巴基斯坦数据保护法》	120条巴基斯坦网站隐私政策
文献[45]	SVM、BILSTM、Bert	SVM	GDPR13条	304个应用程序隐私政策
文献[57]	TFISM		GDPR中关键因素	224份美国公司隐私政策

2.3 GDPR模型框架

通过调研现有研究工作,目前常见的GDPR模型主要基于合规性检测、隐私政策分析以及系统模型设计来遵循GDPR基本原则.因此,本节从这3种不同的技术角度,分别阐述了基于合规性检测的框架模型、隐私设计的框架模型以及系统设计的框架模型的研究进展.同时,在现有研究工作的基础上,对每种模型框架进行分析讨论,并给出观点.

2.3.1 合规性检测框架模型

通用数据保护条例(GDPR)的合规性对组织在个人数据隐私保护上提出了更高的要求,每个组织都必须考虑适用其组织架构的框架模型,然而庞大且复杂的法律合规需求极大地限制了组织的效率,如何为组织提供良好语义化的GDPR框架仍是一项重要的挑战.现有的研究工作主要通过合规检查表、合规评估工具以及法律模型来实现GDPR的合规性检测.

公共机构和公司开发构建的合规检查表[59-61],能够有效支持组织检查其对GDPR的遵守情况.文献[59]提出了GDPR文本扩展(GDPRtEXT),使用欧洲立法标识符(European legislation identifier, ELI)本体将GDPR公开为链接数据,将概念与GDPR相关文本链接起来.组织可引用查询结果并链接至相关文本,从而记录和衡量对GDPR的遵守情况.处理活动登记册(record of processing activities, ROPA)是组织个人数据处理活动的综合记录,创建和维护ROPA是实现问责制并帮助监管机构实施GDPR合规监管的重要过程.然而,传统的通过电子表格维护的ROPA缺乏适合构建自动化工具链的数据结构及语义.文献[60]通过语义网络将不同监管机构发布的模板合并为良好交互性的ROPA通用语义模型(common semantic model for ROPAs, CSM-ROPA),并基于扩展数据隐私词汇(data privacy vocabulary, DPV)为跨域法管辖合规性提供统一数据模型.文献[61]在文献[60]的基础上构建使用DPV审计个人数据国际传输的GDPR合规性工具,并在识别数据转移、合规性以及问责制方面有积极反馈.但受限于测试规模,该模型性能还需要进一步考量.

在合规评估方面,工具的实现往往需要基于具体的数据保护技术(例如,区块链、数据挖掘技术)或集成定制满足GDPR原则的工具实现.GDPR强调必须确保组织在用户同意情况下使用数据,用户同意也是执行同意机制的互操作性、正确性和完整性的基础.文献[62]提出了一种基于区块链的合规验证模型,确保只有获得用户授权的实体才能访问用户数据,且所有数据交互都记录在区块链上,但是该方案仅保障GDPR同意机制的实施,无法满足GDPR整体合规验证.此外,数据驱动型组织严重依赖于数据处理,存在数据交互的业务很容易违反GDPR.文献[63]提出一种基于事件日志行为的在线流程挖掘框架以实现支持业务流程的GDPR合规性.通过前向合规技术检测业务流程的合规性,由流程挖掘技术从事件日志中发现组织的违规行为来为流程提供用例.一致性检查技术将观察到的行为

与业务流程期望的行为进行对比,以评估它们的偏差值,然后通过向后合规检查技术发现不合规方面并相应地调整模型,但该框架在复杂度高、跨越组织的业务流程中存在一定的局限性.

从学术届和产业届方面的工作来看,许多工具和模型仅满足特定或孤立的 GDPR 需求,例如透明度、问责制或数据最小化,较少存在全面支持 GDPR 原则的模型.文献[64]设计了一个支持 GDPR 数据治理的 DEFeND 平台框架,能够有效复用和集成满足特定或孤立 GDPR 异构的工具,围绕隐私保护的设计、同意机制管理和隐私影响评估管理 3 个概念,帮助组织模块化实现 GDPR 合规性.该方案能满足 GDPR 多方面的要求,对 GDPR 的实施提供了完整的参考实例.然而,目前仍缺乏应用于大数据场景中多源数据、不同目的和密集型数据处理的 GDPR 合规性解决方案.文献[65]提出了一个组件化框架来实现大数据场景中的 GDPR 应用,该框架允许对与 GDPR 相关的工作进行分类并集成在框架组件中,解决了大数据系统中的异构性和多源数据分析的需求,但还需要大量的测试来平衡安全解决方案和性能开销.

除此之外,法律建模方法[66-67]建议对监管概念进行建模以实现 GDPR 合规性,法律文本通常包含特定领域术语的定义、交叉引用和歧义,其可解释性对于开发人员可能具有挑战性,公司通常使用法规评估工具来提升法律文本的可读性,帮助组织了解其法律义务.文献[66]提出了一个描述 GDPR 原则的企业架构模型(enterprise architecture models, EAM),将 GDPR 法规形式化为遵循合规性原则的 EAM 片段并强调 GDPR 原则和义务之间的联系,帮助组织积极履行法规义务.同时,该方案在企业架构的不同层次上对 GDPR 法规建模,解决单一方面建模的局限性.然而,现有的法律建模倾向于考虑特定法规,但在实际环境中企业将面临诸多法规制约.文献[67]提出了一个灵活的模块化立法合规评估框架,该框架旨在支持多项立法,此外,该框架还扩展了开放数字版权语言(open digital rights language, ODRL)用于表达立法义务,这两者都是迈向上下文内容相关合规系统的重要一步,使系统可以轻松适应不同的监管领域.

合规检查表、合规评估工具以及法律建模方法都能够在不同程度上实现 GDPR 合规原则检测,为检查合规性和理解 GDPR 的影响提供指导.

2.3.2 隐私设计框架模型

隐私设计是指在设计系统时需考虑到隐私问题,即在处理方法的设计阶段必须已经考虑到所需的隐私保护问题,与隐私相关的问题应该在设计层面解决,而不是在实施之后.这种方法通常被称为隐私设计.在设计隐私框架时通常需要满足 GDPR 的相关原则,如主体同意原则、透明度原则、真实性和准确原则以及问责原则等,以有效保障合规性.

文献[68-69]基于 GDPR 基本隐私原则设计了满足 GDPR 要求的隐私设计框架,以实现在源头满足 GDPR 合规性的方法.GDPR 定义了问责机制以保障个人数据的隐私,通过赋予个人对隐私数据的控制权来提升个人数据隐私权限.文献[68]提出了一种满足 GDPR 数据处理要求的隐私设计框架(privacyTracker),该框架支持包括数据可追溯性在内的 GDPR 基本原则,允许用户从任意节点以不同索引方式遍历引用来构建跟踪树,即所有接收数据实体的树状记录,跟踪收集数据的披露情况,实现个人数据泄露问责的同时评估数据完整性.这些隐私设计框架虽能有效保存和处理个人数据,但仅关注了局部的隐私原则,缺乏对隐私设计问题的整体认识.文献[69]依据 GDPR 基本原则为企业架构提供模式库、集成用例来实现 GDPR 合规性;通过对来源的检索、识别实体对象并分析所需的业务流程来定义用例;选择模式对应的 GDPR 原则或创建新模式来确保信息系统符合 GDPR.该方案能够依据检索模式实现满足 GDPR 的隐私设计,融合多模式解决整体隐私设计问题,具有良好的泛化能力.但同时,需要不断更新模式库,以满足不断出现的隐私设计问题.

2.3.3 系统设计框架模型

GDPR 的提出使得组织需要设计同时兼顾功能和隐私原则的系统模型.文献[70-71]针对不同应用场景设计系统框架以遵循 GDPR 原则.

社会技术安全(science,technologhy and society, STS)是一种设计安全复杂系统的方法,其中自主参与者和机器之间建立相互依赖关系通过交互和共享数据实现目标.文献[70]提出了一种由建模语言和推理框架构成的社会技术系统设计方法,通过建模识别参与者之间的依赖关系实现满足 GDPR 的社会层面建模,并由推理框架自动验证隐私政策合规性.

工业领域中识别或分析人类行为的算法,有助于实现和增强人机协作,但数据主体隐私与工作流程有效性之间存在冲突.文献[71]基于自动化工业

生产场景提出了符合 GDPR 自动化服务的分布式隐私感知软件架构,在保证个人数据(personal data, PD)自动化感知服务隐私性的同时,规定自动化服务公司的义务和职责.但适用范围较为局限,需要与企业资源规划和信息安全管理系统的协同作用.

在医疗行业这种依赖个人敏感信息(病例)的系统中,需要着重考虑数据处理的安全性.文献[72]为电子健康记录(electronic health records, EHR)提出一种可互操作的 openEHR 系统架构,允许用户实时接收数据并在同意的情况下共享数据.该模型实现系统功能层和数据可追溯性、完整性和机密性相关需求,提供了开发兼容卫生系统的完整方法.同时,文献[73]提出一个面向患者基于区块链和快速医疗互操作性资源(fast healthcare interoperability resources, FHIR)的电子健康钱包(electronic health wallet, EHW)系统,和一个兼容 GDPR 法规的基于健康物联网系统数据的 PHR 系统框架.PHR 系统可以兼顾数据隐私保护以及数据互操作性,鼓励患者选择性的共享数据,并以保护隐私的方式对物联网健康数据进行分析,进一步解决了医疗系统设计中的互操作性及隐私保护问题.但是,基于系统的设计框架需要平衡系统功能的可用性和数据主体隐私之间的关系,进而有效遵循 GDPR 原则,针对不同系统实现模型设计的统一方案还未实现.

2.3.4 小结

2.3 节从现有的基于 GDPR 的模型框架的研究工作中,总结了 4 类具有代表性的合规方法并针对每类合规方法挑选出 2 种及以上代表性的研究工作,具体如表 5 所示.表 5 分别从合规方法、具体方式、分析对象以及使用领域多个角度进行分析和讨论.

Table 5 Comparison of Compliance Methods Based on the GDPR Model Framework

表 5 基于 GDPR 模型框架的合规方法工作对比

文献	合规方法	具体方式	分析对象	使用领域
文献[59]	合规检查表	GDPRtEXT 文本链接	GDPR 文本	公共领域
文献[60]	合规检查表	CSM-ROPA 通用语义模型	处理活动登记册(POPA)	监管机构
文献[62]	合规评估工具	区块链	GDPR 原则	公共领域
文献[63]	合规评估工具	流程挖掘框架	事件日志	企业
文献[64]	合规评估工具	数据治理 DEFeND 平台框架	合规工具	企业
文献[65]	合规评估工具	组件框架	合规工具	大数据领域
文献[66]	法律建模	形式化法规、不同层次架构建模	GDPR 文本	企业
文献[67]	法律建模	模块化立法评估框架	法律文本	公共领域
文献[68]	隐私设计	privacyTracker 隐私设计框架	GDPR 原则	用户、企业
文献[69]	隐私设计	架构模式库	GDPR 原则	企业
文献[71]	系统设计	分布式隐私感知软件架构	自动化服务	企业
文献[72]	系统设计	个人健康记录系统架构	物联网医疗数据	医疗

讨论 3. 通过研究大量文献,本节将 GDPR 模型框架分为合规性检测框架、隐私设计框架以及系统设计框架 3 部分.其中,合规性检测框架通过合规检查表[59-61]、合规评估工具[62-65]以及法律建模[66-67]检测 GDPR 的合规性.由文本扩展[59]、通用语义模型[60]和数据隐私词汇[61]构成的合规检查表通过建立概念与 GDPR 法规的映射关系,在一定程度上帮助组织实现合规性检测,但其多依赖于人工实现,在实现效率和灵活度上有待考量.合规性评估工具基于数据保护技术[62-63]或满足特定或孤立 GDPR 要求的工具集合[64-65],评估 GDPR 的遵循情况进而保障组织合规性.法律建模方法[66-67]对监管概念建模,提高法律文本的可解释性,以减轻组织 GDPR 合规性挑战.隐私设计框架[68-69]在设计层面基于 GDPR 基本原则设计隐私框架以保障隐私合规.系统设计框架[70-73]针对不同场景设计兼顾系统功能可用性、效率和数据隐私的系统框架.

观点 3. 针对合规性检测框架多层架构、多源数据和不同目的数据处理的需求,隐私设计框架构建整体隐私设计框架的要求,以及系统设计框架平衡系统功能和数据隐私保护之间关系的问题,本文通过对比和归纳现有研究工作发现:1)使用集成性和自动化合规性检测工具并通过具有明确组件的模块化框架能够快速解决概念合规、数据合规以及流程合规的挑战,有效实现 GDPR 合规性;2)集成多种隐私设计模式方案的模式库能够实现整体隐私设计需求,但需要

与自动化工具结合以实现高效的隐私设计;3)异构性及其功能和隐私保护等级需求不同,使得现有研究工作并没有提出满足异构系统的系统设计框架.

2.4 小 结

建立合理的合规性检测框架,进一步规范现有的隐私政策,并且采用普适性的高效检测方法,才能达到 GDPR 的政策预期,保护公民的数据安全和隐私.本节分别从违规检测手段、隐私政策设计和 GDPR 合规性模型框架等方面分析了推动 GDPR 政策落实的技术手段,并加以概括总结,同时指出了进一步的研究方向,图 7 选择部分代表性文献展示了 GDPR 的合规性研究发展历程.

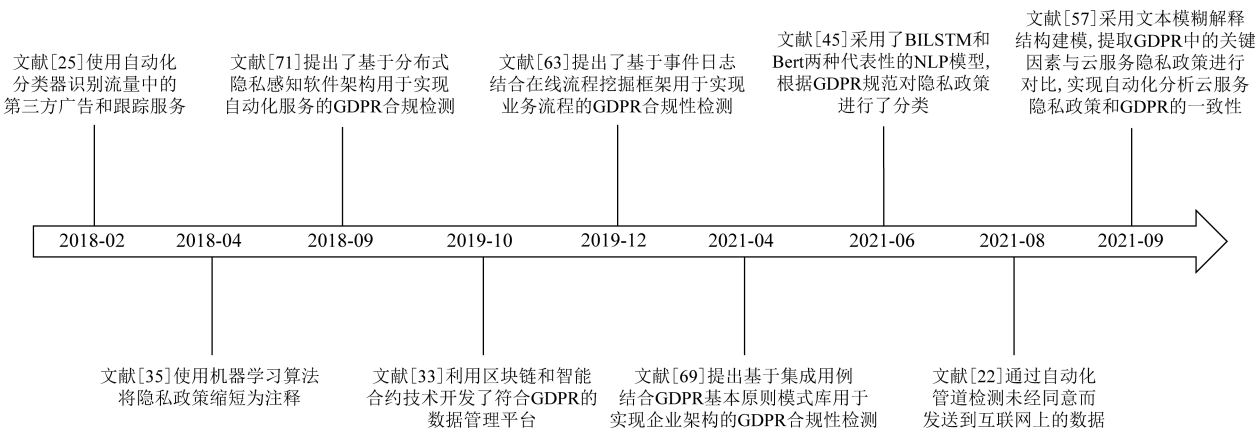

Fig. 7 The development of GDPR compliance testing

图 7 GDPR 合规性检测发展历程

3 GDPR 相关的技术应用

3.1 基于 GDPR 的数据技术

GDPR 指导了数据控制者和数据处理者如何对数据进行合理的处置,但数据处理必然存在着一定的风险,为了降低数据处理的安全风险,需要制定相应的保护措施.逐渐增多的跨境业务也为个人数据隐私增添了一份安全隐患,跨境流动数据需要得到更有力的保护.本节将从数据保护影响评估和数据跨境流动 2 个方面探讨 GDPR 相关的数据技术研究进展.

3.1.1 数据保护影响评估

GDPR 第 35 条规定当个人数据处理的过程中可能会对个人权利和自由产生高风险时,数据控制者应提前做好数据保护影响评估(data protection impact assessment, DPIA).DPIA 建立在隐私影响评估(PIAs)的基础上,是组织和企业必须履行的一项有关 GDPR 数据问责制的关键义务,它可以帮助企业实现风险最小化,并帮助企业证明合规性.如果企业未能履行这项义务,则将面对极其严厉的处罚,包括高达 1000 万欧元的罚款,或者高达到 2%的全球年营业额.

虽然 GDPR 对数据保护影响评估提出了相关要求,但是 GDPR 只规定了实施 DPIA 的最低标准,并没有涉及明确的执行方法[74].文献[74-75]结合了德国数据保护机构采用的标准数据保护模型(standard data protection model, SDM)方法,文献[74]设计了一种跨学科的风险评估方法,将 DPIA 过程分为了准备、评估以及报告和保障 3 个阶段,并提出了可用性、完整性、机密性、不可链接性、透明性和可干预性 6 个评估要素.文献[75]就有关如何实施 DPIA 框架的问题展开了讨论,并通过 2 个案例的分析总结,实现利用 SDM 的数据保护目标来对风险进行结构化分析,未来也可以将这项工作纳入 SDM.文献[76]在文献[74]的基础上对方法进行了实践,使用文献[74]的方法实施 DPIA,与 12 个组织展开了合作,并分享了与公司合作实施以来积累的经验,以及不同的利益相关者在实施 DPIA 时需要注意的事项.

文献[77-78]针对特定的 DPIA 实施环境进行了分析.文献[77]专门对慈善机构及中小型企业展开了研究,与其他组织不同的是,慈善机构和中小型企业通常在财务和资源方面能力有限,因此在处理特殊类型数据和个人身份数据的工作上缺乏专业性.文章展示了实施 DPIA 的示范过程及设计框架,并通过一家实际的慈善机构进行了验证,该框架同样可以应用于其他需要实施 DPIA 的组织.文献

[78]则主要针对 IT 系统,在系统开发早期通过基于模型的隐私安全分析来实现 DPIA,并通过 3 个工业案例研究对该方法进行验证和评估.

因为现有的 DPIA 方法主要由分析师来进行评估,所以很容易受到分析师的主观影响,为了解决这个问题,文献[79]提出了一套有明确定义的标准用来帮助分析师评估隐私风险的影响和可能性,同时使用模糊多准则决策方法来系统评估隐私威胁的严重程度并进行建模.文献[80]结合数据保护影响评估(DPIA)和信息安全风险评估(information security risk assessment, ISRA)提出了一个信息安全风险评估模型 pISRA,该模型为评估者提供了一个可以进行比较和重复的评估方法,但是还没有得到具体实现.

小结:本节主要针对现有的数据保护影响评估(DPIA)方法进行了讨论和分析,并给出了基于数据影响保护评估领域的一些观点.

讨论 4. 目前大多数研究工作都集中在构建 DPIA 实施框架及流程示范上,但有关 DPIA 框架的具体实现较少.部分研究工作针对一些特定的组织分析了 DPIA 实施环境,并对 DPIA 框架进行了验证.除此之外分析师的主观想法也会影响到 DPIA 流程,明确的标准和系统的评估能够帮助 DPIA 的顺利实施.

观点 4. 由于 GDPR 没有对 DPIA 的具体实施方法进行详细说明,因此目前仍缺乏标准化的 DPIA 实施流程,同时每个领域需要解决的问题不同,对于 DPIA 流程设计的需求也不尽相同,这给 DPIA 的实施带来了很大的挑战.如何针对特定的领域设计专门的标准化 DPIA 程序仍需要进一步研究.

3.1.2 数据跨境流动安全

由于各国之间对于个人数据的相关法律要求不尽相同,比如欧盟的 GDPR 标准对于个人数据的保护十分严格,而中国的《数据安全法》出台还没多久,在个人数据的保护方面还较为薄弱,因此个人数据的跨境流动会带来较大的安全隐患.数据跨境流动要注意数据合规性问题,合规性验证对于保证业务流程的整个生命周期的安全性至关重要.

目前针对数据跨境安全领域的研究主要集中在政策解读和模型架构、技术支持方案以及医疗健康数据方面.

1) 政策解读和模型架构

在数据跨境政策解读和现有模型架构方面,文献[81]认为 GDPR 不仅保护了数据基本权利,也促进了个人数据自由流动,这样形成的分层数据保护制度的架构保障了包括研究在内的以不同的公共或经济利益为基础进行的数据处理活动;而文献[82]提出了 GDPR 标准在如何评估第三国制度数据保护水平等问题上的缺失和不足,认为可以确定一套实质性要求以及第三国必须提供的支持性程序和执行机制来确保其数据保护水平符合欧盟标准.另外,在国内也有大量法律、金融等领域的学者和专家对 GDPR 进行了分析和研究,比如,文献[83]指出 GDPR 客观上对国际服务贸易规则产生了深远而广泛的影响,其中包含的数据跨境转移规则造成了数字封锁,构成了贸易障碍,企业和第三方满足 GDPR 标准的数据合规要求困难重重,最终导致数据本地化是满足 GDPR 合规要求的最佳选择.文献[84]将 GDPR 与当前全球其他主要经济体的跨境数据流动政策及其实践进行比较,分析了"数字主权"下全球跨境数据流动政策的新动向,从贸易框架探寻国际合作机制并由此提出对中国相关体系建设的建议.

不同国家和组织之间的差异是目前跨境数据流动面临的最大问题,不同的政策措施也意味着不同的数据保护水平,因此,文献[85]对各国及各组织在数据跨境流动问题上的现有政策和态度进行了对比解读,分析了在 APEC、CBPR 以及 GDPR 等多个标准协定之间建立互操作系统的潜在挑战和影响,以及站在美国角度,提出了目前在数据跨境流动问题上还需要考虑和解决的一系列事项.文献[86]通过分析欧盟和美国的跨境数据保护政策,提出中国应该继续保护当地居民的个人数据,同时,在考虑到互联网带来的巨大价值,中国应该开放非个人数据传输市场的建议.针对这些争议和讨论,文献[87]提出一种通用交换数据模型(EDM),该模型利用现有的开放式欧洲标准和技术规范作为构建块,以更加内聚和统一的方式描述一次性跨境消息交易.文献[88]从中外数据本地化实践中,抽象出描述数据本地化存储的严苛度模型,并以目的和手段之间的适当性和必要性为指针,构建出一套"数据本地化存储合理界限"理论,并从该理论出发,检视中国《网络安全法》相关规定,给出基本评价并提出了数据跨境安全评估办法的总体框架.

2) 技术支持方案

跨境数据流动亟待解决的问题和需求已经开始催生新的技术支持方案.例如,文献[89]引入隐私证书颁发机构(certificate authority, CA)的概念,

设计了一个多个隐私 CA 的访问控制层次模型,这些 CA 负责管理不同领域的法规,不仅可以控制不同国家的数据传输,还可以控制不同经济或政治集团或城市的数据传输.另外,他们还将城镇管理应用程序作为 iKaaS 平台的一个用例,介绍了该访问控制机制的工作原理.文献[90]在 GEO-TRUST 项目中提出了一种称为偏移量证明(proof of offset, POO)的创新协议,以通过地理位置、责任、数据公开最小化、数据语义注释实现更高的控制和数据访问限制,从而保证跨域数据重用,并提高数据保护意识,以此来促进数据交换、可信任性、同意管理、声誉和安全的监管.

随着时代进步和科学技术的不断发展,区块链系统提供了一个分散、不变和透明的架构,可以将数据的所有权和控制权交还给用户,实现可信和负责的数据共享,但对于数据共享等领域,区块链网络中仍存在不同的可扩展性、安全性和潜在的隐私问题,如链上数据隐私、数据源身份验证或遵守隐私法规,因此,文献[91]提出了一种基于区跨链系统和密文策略属性加密的隐私保护和用户控制的数据共享架构 ThemisABE,该方案具有一对多数据加密和细粒度访问控制等特性,能解决数据共享的隐私安全和本地化问题.针对跨境数据共享,文献[92]提出了一个使用区块链的跨订单可问责数据共享平台,其中全球云构建在不同国家设置的多个安全网关之上,分别使用包括 5 种算法来处理数据访问请求、数据共享、区块链交易、检测和惩罚行为不端的实体等问题.文献[93]针对 GDPR 下共享数据的安全性问题部署了一种基于风险的评估方法来确定如何评估现有的数据匿名化技术,以此来与 GDPR 中的新数据类型相协调;还进一步开发了一个基于机器学习的隐私风险挖掘框架,该框架由两阶段聚类算法和隐私风险树模型组成,可以用于检测发布新净化数据集的记录链接风险;此外,文献[93]不仅为数据控制者提出了一个隐私管理框架以提高区块链技术差异私有数据共享的效用和安全性,还提出了另一个结合区块链和同态加密的框架,以外包集中式匿名服务帮助数据所有者与多个数据控制者之间共享数据.除了区块链技术,文献[94]为了让任何非结构化数据云存储系统都必须满足跨境数据流法规遵从性的要求,还使用深度学习模型将驻留在统一文件和对象存储中的数据分类为个人信息,以及在集群文件系统级别实现地理围栏功能,以此来规范分类个人信息的跨境数据流.另外,在 Web 服务方面,文献[95]设计了一种测量方法,用来量化跨境的大规模跟踪流量,测量结果显示,大部分的跟踪流量都会在欧盟境内终止,也就是跟踪流量还在 GDPR 规则的管辖之下.文献[96]全面总结了有关第三方网站跟踪的政策及技术研究,来帮助决策者制定更加安全的解决方案.在移动应用方面,文献[97]针对安卓应用程序的数据跨境传输制定了合规评估标准,并设计了合规性评检测方法,并用此方法对 100 个常用的安卓应用程序进行了评估,发现有高达 66%的应用程序存在着跨境合规问题.

3) 医疗健康数据

在医疗数据方面,为了实现更好的医疗服务,患者的跨境移动、远程医疗和医疗研究的交流都给数据安全带来了极大的挑战,为此文献[98]借助私人区块链搭建了用于评估的平台,通过推荐最佳的安全策略来为业务和应用系统量身定制防御措施.文献[99]介绍了可自动识别风险的系统安全建模器(system security modeller, SSM),并以欧盟内部的跨国医疗数据交换为场景进行了讲解,该工具可以在系统设计的同时检测合规性,当出现不符合合规性的情况时还会计算出对整个体系结构的影响.文献[100]针对跨境电子身份认证保护进行了讨论,并建议通过假名化和选择性披露的方法使电子身份识别的互操作性框架达到要求的数据保护级别.为实现有效的跨订单医疗保健供应,欧盟发布了 OpenNCP 平台来解决国家间卫生信息交换中的互操作性问题,针对其中存在的一些安全问题,文献[101]在 OpenNCP 的基础上进行扩展并详细描述了 KONFIDON 项目方法以及如何通过结合互补的安全增强技术来部署该方法,以达到最终提高电子健康数据交换的信任和安全性.文献[102]提出了一种实现破坏性日志记录的新方法,即一种用于在 OpenNCP 上跨境交换电子健康数据的审计机制,在 OpenNCP 基础设施内提供可追溯性和责任支持.文献[103]提出一种访问控制方案,该方案允许请求数据和服务的消息在发送方和接收方验证安全问题后跨不同的区域或国家节点,它可以拒绝那些被检测为恶意的访问请求;并通过放置在发送方和接收方的威胁检测软件的明确反馈来抑制许可消息流,以此来提高在分布式系统如 OpenNCP 下运行的跨境健康数据访问的安全性;并使用一个分析模型来评估了安全系统造成的开销.考虑到医院中数据和软件使用的异构性和高度敏感性所带来的具体限制和批评,文献[104]提出了一种为医疗信息系统执行 DPIA 的方法,通过支持风险评估和管理,该方法可以应用于在医疗环境中

执行 DPIA 以维护医疗保健信息系统的安全性.针对系统的互操作性问题,文献[105]也给出了解决方案,它设计了用于医疗保健行业的工业 4.0 模型,并集成了不同的工具,如同意管理器和数据隐藏工具,来确保医疗体系的隐私性.

4) 小结

3.1.2 节针对现有的数据跨境流动安全领域的研究工作进行了总结和讨论,并给出了数据跨境流动研究领域的一些观点.

讨论 5. 3.1.2 节从数据跨境领域出发,分别介绍了国内外专家学者对于当前多个经济体的政策的解读和模型架构的分析、以及应运而生的新技术、新方法,另外也单独从医疗数据角度出发介绍其跨境安全和现有方案.不管是 GDPR 对欧盟数据保护起到的积极作用,还是其掣肘发展和交流的消极影响,都说明目前各个经济体针对数据跨境的制度和政策都有一定的局限性.此外,不管是框架还是技术方案,目前都还处于研究阶段,而数据跨境安全体系建设势必要落实到实践中去,既要考虑其适配性和合理性,也要不断从实践和反馈结果中发现问题并提出解决和提升的方案.此外,目前的有效技术方案较少,层次相较于普通数据共享方案也没有明显的融入数据跨境需求,缺少针对性探讨和研究.个人健康数据在跨境过程中的隐私性和安全性确实需要得到重视,但其他领域的数据也需要相应的研究和评估,这是目前研究领域存在的短板和不足.

观点 5. 目前不同国家的数据跨境政策之间的差异较大,且安全和发展侧重点不同,以至于短期内很难在全球范围内形成统一且高效的跨境数据治理监管体系,也就无法应对未来发展带来的大规模数据跨境安全需求问题.目前世界各大经济体都在致力于探寻符合自身利益的数据跨境方案和界限,但缺少交流协商寻求全球共识的契机.技术工具和框架建设不应止步于个人健康数据,经济、政治、科技等领域也是国家有序健康发展的重要动力,不同的数据拥有不同的敏感性和安全级别,相应的就会在跨境的各个环节产生不同等级保护措施的要求,中国现如今已经逐渐形成统一的数据分类分级制度,相关技术方案可以以此为研究角度进行设计、改进和升级.

3.2 GDPR 合规应用场景

GDPR 的出台确保了数据主体的数据隐私安全,为数据主体、数据控制者和数据处理者之间搭建起了一座信任的桥梁,特别是数据流动频繁、数据敏感度高的应用场景,GDPR 的合规性显得尤为重要.本节分别针对区块链、物联网、电子健康及其他领域(教育、生物特征识别)等不同的应用场景对 GDPR 合规性进行探讨.

本文选择区块链、物联网等应用领域进行分析和讨论,主要有 3 点原因:1)通过对现有研究工作的梳理发现,现有的基于 GDPR 的数据隐私安全研究成果主要集中在这几个领域,有必要对其进行单独调研分析.2)目前大多数的区块链应用都不符合 GDPR 标准,区块链的永久存储不可更改的特性使得区块链的合规性变得困难.同时,物联网设备之间传输的数据量大且类型复杂,其中不乏大量的用户个人敏感信息,一旦设备遭到攻击将可能造成十分严重的数据泄露事故.3)生物特征数据属于 GDPR 规定的特殊类别数据,非特殊情况不得处理,因此作为当今社会重要的生产要素,生物特征数据的隐私安全不可轻视.学术研究需要用到大量的研究数据,如何确保这些数据的合规性,将在很大程度上关系到学术研究能否顺利开展.但目前有关生物特征数据和学术研究领域的研究工作较少,因此本文将其归纳到其他领域进行探讨.

3.2.1 区块链合规领域

区块链技术具有分散性、透明性、可追溯性、不变性的特性,消除了个人数据的集中化,为数据的管理和存储提供了很大的帮助.但 GDPR 的出台也为区块链技术带来了新的挑战,为了了解区块链领域是否能够有效应对 GDPR 带来的合规问题,文献[106]对区块链系统做了一项分析调查,调查包含了区块链系统的开发商和服务提供商公开发布的法律文件及官方的 Twitter 账户推文.然而调查结果不容乐观,虽然 GDPR 已经颁布了 3 年并实施了一年,但在区块链领域仍然存在着如何解决 GDPR 合规性的严峻挑战.调查显示,在 314 个区块链系统中只有 86 个(27.5%)系统涉及到了 GDPR,且仅有 27 个(8.6%)系统有关于 GDPR 合规性的确切的法律文件.因此,要解决区块链技术与 GDPR 合规性之间的问题仍然任重道远.

本节将从数据责任和来源追踪、数据管理和数据擦除 3 方面来讨论区块链技术为 GDPR 的合规性提供的助力以及其产生的阻碍.

1) 数据责任和来源追踪

虽然 GDPR 出台后对拥有信息的服务提供商提出了更严格的要求,但服务提供商能否一直坚守高要求还是一个变数,数据的收集和处理过程仍缺乏透明度,用户无法了解自己的数据流向了哪里,被用在何处.区块链技术为此提供了合适的解决方案[33,107-110],

通过分布式账本来记录服务提供商的所有数据活动,这样一旦服务提供商违反 GDPR 标准,他们的行为将会被记录在案.通过区块链技术可以实现数据流动的透明度,增进个人数据利益相关方之间的信任.

文献[107]为云存储应用设计了云数据溯源架构 Provchain,该架构将数据操作的历史记录散列到 Merkle 树节点中,并链接到区块链上,生成防篡改的数据记录以供验证,实现云数据的透明性.文献[33,108-110]则利用了基于区块链的智能合约技术实现了数据来源的追踪和记录,通过智能合约捕获服务提供商和用户之间的交易条件,而无需第三方的参与,既实现了去中心化,又能够降低成本.文献[108]设计实现了 2 个具有不同粒度和可伸缩性的模型,其中第一个由数据主体为每个接受数据的控制器部署访问控制策略,第二个则由数据控制器部署策略来让数据主体加入.但文献[108-110]只提出了相应的概念框架,并没有涉及更详细的技术细节.文献[33]为合规的基于区块链的个人数据管理平台提供了详细的技术机制,他们在 Hyperledger Fabric 区块链框架之上开发了基于业务连续性的个人数据管理系统,证明了概念的可行性.

文献[109]具体说明了如何将一组 GDPR 规则转换为智能合约中的操作代码,使物联网设备实现对个人数据的自动验证.该方法不仅可以应用于物联网场景,还可应用于云系统或其他的服务场景[110].未来还可以在公共许可区块链或私有区块链上实现设计的抽象模型[109].

2) 数据管理

GDPR 规定数据主体要对个人数据的流向知情并予以同意,还要以易于理解的方式对个人数据进行控制.基于区块链技术的同意管理平台[111-118]可以帮助用户理解同意申请并轻松地管理同意许可,确保了用户对于其个人数据的控制权.在 GDPR 出台之前,文献[111]就针对个人数据隐私问题,将区块链作为自动化访问控制管理器,设计了基于区块链的个人数据管理系统.GDPR 出台后,文献[112-122]也利用区块链技术提出了各自的解决方法.

文献[112]针对在线社交网络现有的同意管理机制与 GDPR 的规定进行了比较分析,并确定了其中存在的风险,作者建议设计基于区块链的同意管理模型为在线社交网络用户提供所需的透明度.文献[113]设计了一个个人数据管理系统 BPDIMS,该系统以用户为中心,最大限度地实现了用户对个人数据的控制,并通过个人数据的货币化提升了用户对于个人数据价值的认知,使用户能够在分享个人数据的同时获取金钱收益.文献[114]利用区块链技术为用户提供了一个轻量级管理系统,该系统可以显示服务提供商有关个人数据的协议.文献[114]通过对控制器和处理器进行识别来区分 2 种同意许可,解决了其他文献并没有将数据收集和数据处理 2 方面的许可区分开来的问题,未来还可以为系统增加可视化图形界面来方便用户的管理.

文献[115]结合了加密技术,保证了同意管理系统的隐私性,并且为公司设计了代理应用程序,该程序会定期查询区块链,更新有关的同意状态,并以发布-订阅的形式告知相关的服务,使其能够及时做出反应.该文献首次实现了使公司服务与数据主体的动态同意许可之间保持实时同步.

文献[112-115]仅进行了基于区块链的概念设计,并没有进行概念验证,文献[116-118]则分别在不同的区块链上开发了相应的系统.文献[116]借助语义网和以太坊区块链构建了自动验证数据合规性的系统,当数据分享给第三方时,该系统能够强制执行 GDPR 规则.但该系统仅使用了以太坊区块链,未来还可以在更多的区块链框架上进行探索.文献[117]则通过 Hyperledger Fabric 框架实现了概念验证,设计了一个同意管理模型,利用区块链技术为数据主体、数据控制者和数据处理者提供了交互的工具,并维护了数据主体的权力.文献[118]提出了一个数据安全共享方案,将智能合约设置为访问控制列表,并为不同的对象设计了 4 种智能合约,文中探讨了哪些数据是不可变类型且可以存储在区块链的数据,并对该方案在不同区块链平台下的性能进行了测试.

不同于其他系统的单链结构,文献[119]设计了一种新颖的双层区块链结构,开发了用户权限管理系统 Soteria,该系统可以同时满足分布式系统 CAP 定理中一致性(C)、可用性(A)和分区容忍性(P)3 个属性,其中主链满足了可用性和分区容忍性,侧链满足了一致性和可用性,保证了系统的透明性、可证明性和可扩展性.除了双层区块链的分布式账本模块,该系统还包括用户权限管理模块 URM 和审计跟踪模块 ATS.但由于侧链将块散列到主链上的频率会影响到整个系统的延迟和吞吐量,因此 Soteria 的链间管理策略还需要进一步的调整优化.

基于区块链技术的自我主权身份(self-sovereign identity, SSI)[120]也是实现数据的完全控制的一种途

径.区块链技术使得身份管理(identity management,IdM)系统由传统的集中化的方法逐渐向开放、分散的自我主权身份转变.自我主权身份系统通过结合分布式分类账本技术和加密技术来创建不可篡改的身份记录,实现了用户对个人数据的完全控制权[121].文献[122-124]研究了现有的自我主权身份技术方案,并对 SSI 系统与 GDPR 原则的兼容性进行了分析.

文献[122]对现有的 3 种区块链身份管理系统 uPort,Sovrin 和 ShoCard 进行了分析,并指出了它们存在的缺陷,提出了新型身份管理系统 DNS-IdM,该系统可以通过自主身份管理实现去中心化.文献[123]对基于公共无许可的 uPort 和基于公共许可的 Sovrin 两种不同类型的身份管理系统进行了比较,发现 Sovrin 区块链系统更加符合 GDPR 的大部分要求,因为 Sovrin 生态系统包含一个治理模型,且由可信组织联盟管理.除了 uPort 和 Sovrin 系统之外,文献[124]还分析了在公共无许可的以太坊区块链上应用的 Jolocom 框架,并讨论了 SSI 与 GDPR 标准之间的一致性.

3) 数据擦除

GDPR 第 17 条规定了数据主体的被遗忘权,即当满足一定的条件时,数据主体有权要求删除自己的个人数据.用户需要合适的机制确保他们能够选择自己想要的服务,当他们不需要这种服务时也能够完美地退出,例如当用户想要退出某种服务时,服务提供商需要删除用户使用该服务的所有历史记录[113].但是区块链的不变性意味着数据一旦存储在区块链上就不能再被删除或者改变,因此如何实现区块链数据的擦除成为了一项亟待解决的挑战.在先前有关区块链的 GDPR 合规性问题的文章中,讨论的最多的问题也是有关数据删除和修改的规定[125].文献[125]综合研究了有关使用区块链技术进行身份管理的文献,探讨了区块链在遵守 GDPR 的要求方面存在的优点及产生的矛盾,尤其是区块链的不变性与 GDPR 的被遗忘权之间存在的冲突.

比较常见的方法有针对区块链的离线数据存储解决方案[126-130].离线存储即构建链外数据库用来存储个人数据,区块链上则仅保存指向对应的个人数据存储位置的散列数据指针.文献[126]将个人身份信息与非个人身份信息分开存储,个人身份信息存储在本地数据库中,而非个人身份信息以及个人身份信息的哈希则存储在区块链中.文献[127]详细讨论了有关区块链的链外功能集成的方法,并提出了一个概念框架实现链外结构与传统区块链技术的结合.

由于区块链上的数据会在许多节点被复制,导致了数据的大量冗余,因此在区块链存储个人数据是不现实的.如今分布式文件系统(distributed file system, DFS)越来越多地应用于区块链技术,用来解决区块链技术与 GDPR 中的被遗忘权之间的冲突,优异的可扩展性及内容寻址能力使 DFS 系统成为替代传统区块链存储的新方向[128].文献[129]提出了一个在星际文件系统(inter planetary file system,IPFS)中应用的匿名委托擦除协议,该协议可以轻松地集成到 IPFS 中,使 IPFS 符合被遗忘权的要求并被认可其合规性.协议规定只有原始数据的提供者或其代表才能对数据进行擦除,发出的擦除请求会传至所有的 IPFS 节点,且所需的开销并不会影响系统的性能.文献[130]对 IPFS,Sia 和一种专有服务 3 种不同的 DFS 方法进行了评估,发现 3 种方法展现了不同的性能,当出现一定的过载情况时,专有服务的响应和可靠性会优于另外 2 种方法.虽然离线存储有效地解决了区块链的数据存储问题,但此种方法实际上破坏了区块链的分散性,同时也需要可信的数据管理机构[131].

文献[132]开创了另一种可行的解决方法,利用变色龙哈希函数(Chameleon Hash)构建可编辑区块链,传统哈希函数的抗碰撞性保证了区块链的不变性,变色龙哈希利用陷门可以轻松地找到哈希碰撞,从而对区块链任意块中的内容进行重写.该系统扩展了变色龙哈希函数与区块链的兼容性,可以与所有流行的区块链兼容.文献[133]在文献[132]的基础上结合基于密文策略属性的加密(CP-ABE)方法,提出了新的基于政策的变色龙哈希(PCH)的概念,实现了对区块链事务级重写的细粒度控制.为了解决文献[133]的方法可能面临恶意攻击的问题,文献[134]限制了修改者重写特权,修改者最多只能修改 k 次,次数由中央机构定义,除此之外加入了恶意行为惩罚机制,修改者在授权期间需要在链中存入押金,一旦发生任何恶意行为,中央机构可以提取押金.由于 PCH 机制需要一个完全可信的中央机构,文献[135]针对这一弱点提出了去中心化的解决方案 DPCH,并通过基于 RSA 加密算法的变色龙散列和 BLS 短签名进行了实例化.

除了离线存储和变色龙哈希的方法之外,文献[136]提出了一种不同于侧链的解决方法,他们采用树的结构构建区块链,根据业务上下文将交易分到线性子链中,这种方法的优点在于当其中一个线性子链被删除时不会影响到其他子链.文献[137]运用

设计科学研究(design science research, DSR)的方法设计了一个概念原型解决了删除区块链数据的问题,建议在一定的时间过后自动删除区块链中的数据,来实现区块链与GDPR的兼容性.但该方法的前提是需要区块链所有的节点都能有足够的诚信,而且因为删除的时间是预定的,所以该方案并不能满足用户能够随时删除数据的要求.相比于文献[137]的方法,文献[131]的方法则完全不需要修改区块链,文章利用了假名数据的法律属性,即只有当假名数据能够与个人身份联系起来时才能被当作个人数据.文献[131]通过假名生成算法为安全使用日志设计了假名供应系统,该系统会为每一个新块提供一个一次性的交易假名来保证GDPR的合规性.

4) 小结

3.2.1节阐述了现有的区块链GDPR合规性的研究工作进展,并对3类具有代表性的合规性问题以及相应的合规性方法进行了总结和讨论,具体如表6所示.

Table 6 Compliance Issues and Approaches of Blockchain

表6 区块链合规性问题及解决方案

文献	合规性问题	提出的解决方法	涉及的GDPR主要条款
文献[107]	数据问责	分布式账本	GDPR第5条
文献[108]	数据问责	智能合约	GDPR第5条
文献[113]	数据管理	智能合约	GDPR第5~7条
文献[122]	数据管理	自我主权身份	GDPR第5~7条
文献[129]	数据删除和修改	离线数据存储	GDPR第16~17条
文献[133]	数据删除和修改	变色龙哈希	GDPR第16~17条
文献[136]	数据删除和修改	上下文链	GDPR第16~17条
文献[137]	数据删除和修改	遗忘区块链	GDPR第16~17条
文献[131]	数据删除和修改	假名生成算法	GDPR第16~17条

讨论6. 目前有关区块链合规性问题的文献主要集中在数据问责、数据管理以及数据的删除和修改上,其中有关数据的删除和修改的讨论最多.区块链提供的智能合约、自我主权身份等技术,能够帮助企业更好地实现GDPR的合规性.而针对区块链如何进行数据删除和修改的问题,较为广泛的方法是离线数据存储,将数据存储在链外数据库中,区块链上保存数据的散列指针.除此之外,上下文链、遗忘区块链、假名数据等方法也可以用来实现数据的删除和修改.

观点6. 区块链为个人数据隐私安全提供助力的同时也带来了相应的安全风险.区块链的不变性成为GDPR的被遗忘权与区块链之间难以调和的矛盾,如何解决区块链与GDPR之间的冲突是一个值得探索的方向.

3.2.2 物联网平台合规领域

物联网中设备繁多,数据量大,数据流动频繁,个人数据隐私时刻都有遭受侵犯的风险,因此如何实现物联网的GDPR合规性是一个亟待解决的难题.GDPR标准在涉及较多用户的应用领域的影响更为明显,尤其是基于服务的物联网场景如智能医疗、智慧城市等.文献[138-143]致力于为用户提供数据同意管理平台以实现数据隐私保护.文献[138]开发了物联网管理平台ADVOCATE,该平台以用户为中心,帮助用户轻松管理物联网系统中有关个人数据访问的同意请求,同时也帮助数据控制者能够遵循GDPR的原则进行活动.文献[139]提出了Privysharing框架,将区块链技术应用到智慧城市场景中,将数据分成不同的类型,并通过不同的通道处理数据,实现了物联网数据的安全共享,实验证明多通道系统比单通道系统的可扩展性更好,文章还设计了奖励机制以激励用户分享个人数据.文献[140]为物联网智能家居平台提供了一个同意管理器,管理器将复杂事件处理(comples event processing, CEP)与边缘计算结合在一起,复杂事件处理负责数据流动的控制,边缘计算则负责为复杂事件提供安全策略.

在智能医疗领域,文献[144-145]探讨了新实施的GDPR法规给医疗领域带来的变化.文献[144]针对移动医疗应用方面,提出了将GDPR关键规则集成到移动应用程序中的可视化方法,但该研究还未经过真实的场景测试.文献[145]通过文献计量学和科学计量学的方法对医疗领域有关GDPR研究的热点进行了可视化分析,分析揭示了目前的研究热词

是数据保护、隐私和大数据,区块链和机器学习成为了 GDPR 研究的新方向.

对于更为敏感、数据交换频率也更低的医疗数据,可以采用粒度更细的解决方案[141-143].文献[141]针对用户的动态健康数据设计了一个数据共享系统,该系统结合了区块链技术和云存储技术,为大型数据集提供了离线存储的方法,解决了区块链无法存储大量数据的问题,并添加了数据质量验证模块来控制数据的质量.文献[142]设计了一种用于物联网电子健康系统的 GDPR 控制器,能够让用户通过细粒度的访问控制策略完全控制自己的个人数据,当非法访问的情况发生时还能及时收到通知.文献[143]提出了数据安全共享方案 MedSBA,利用私有区块链来实现云存储医疗数据的访问控制策略,提供对医疗数据的细粒度访问和共享过程中的安全保障.文献[146]在容器的虚拟化技术和分布式账本技术的基础上搭建了一个云服务架构,容器技术用于数据的监控;分布式账本如区块链、智能合约等则用来记录对数据的操作,该架构在网上药店的场景中进行了验证,并可以推广到更多的医疗场景.

文献[147]分析了物联网电子健康领域面临的安全挑战,并设计了一个完整的架构来为中老年人提供更加安全的医疗服务;介绍了有关环境辅助生活(ambient assisted living, AAL)和移动医疗 2 种应用程序的设计和实现.文献[148]建议对医疗数据处理进行系统的风险管理和错误管理,以防止医疗项目因合作者没能正确处理数据导致的人为失误.

小结:3.2.2 节针对现有的物联网领域的 GDPR 合规性研究工作进行了总结和讨论,并给出了物联网 GDPR 合规领域的一些观点.

讨论 7. 3.2.2 节主要从智慧城市[138-140]和智能医疗[141-148]2 个应用场景分析了物联网领域在 GDPR 合规性方面的研究进展.数据同意管理平台的建立保证了物联网系统的合规性,高效的身份验证和细粒度的访问控制也进一步为物联网数据共享提供了隐私保护.一些研究工作聚焦在了将区块链技术应用于物联网的课题上,并结合数据加密、云存储、容器虚拟化等技术为物联网用户数据提供安全保障.

观点 7. 目前主要的研究方向主要是为物联网开发实现数据的安全共享.通过对现有研究工作的归纳分析,本文发现:1)对于用户众多数据庞大的物联网应用场景,开发一个保护用户隐私的数据管理控制平台是很有必要的,考虑到物联网资源受限的设备,平台最好能够实现轻量化.2)区块链技术为物联网系统的合规性提供了很大的助力,未来还可以将边缘计算引入区块链系统,以减轻物联网终端节点的维护压力.3)目前有关物联网合规性的研究工作大多都集中于概念架构的设计,还未能投入物联网系统,且应用场景较为单一,如何将合规方法推广到更多的应用场景还有待进一步探索.

3.2.3 其他合规领域

1) 生物特征识别领域

GDPR 引入了一种新的个人数据类别——生物数据,即通过与自然人的身体、生理或行为特征相关的特定技术处理产生的个人数据,这些数据可以确认自然人独一无二的身份,如面部图像或指纹.这种生物特征数据被广泛用于考勤或门禁系统.

对于生物特征数据的立法是很有必要的,但即使在欧盟内部,成员国之间也未能在生物特征数据的使用方面达成一致意见,各国对此的法律要求各不相同,导致 GDPR 在生物数据方面的要求无法实现[149],因此仍需要从法律和技术方面继续分析这一问题.文献[150]总结了法律界和技术界的专家们对于 GDPR 对语音数据影响的看法,由于目前法律界和技术界还无法达成共识,因此作者提出了分类法的方案以实现语音技术与隐私立法之间的协调.文献[151]对智能语音设备的隐私问题进行了详细的研究,作者对市场上流行的亚马逊 Echo 设备进行了测试,发现设备存在着很大的安全风险,在没有安全措施的情况下,用户的个人数据很容易遭到泄露.作者在文中提出了一系列降低安全风险的建议,并指出通过语音识别的生物特征控制可以成功阻止未授权的人访问设备数据.

除此之外,某些类型的软生物特征如情绪反应等,同样会带来数据隐私方面的威胁,甚至不亚于用于识别的生物特征的威胁,但这样的特征并不受 GDPR 规则的保护[152].因此关于 GDPR 生物数据相关的内容仍需要更加系统化的定义.

2) 学术研究领域

在学术研究领域,由于 GDPR 的合规性引起的有关受试者的数据隐私问题,使研究人员而不得不望而却步,甚至直接放弃有涉及到欧盟受试者的研究.尤其是数据密集型研究离不开物联网的支持,但 GDPR 的出台为研究带来了风险,因此文献[153]讨论了如何使学术环境下的物联网数据研究符合数据隐私标准的问题,确定了 3 个信任原则,并实现了一种物联网数据研究的可信架构.教育研究领域也同样受到了来自 GDPR 的影响,例如招收欧盟学生的 at-scale 教育项目在研究中就遇到了 GDPR 合规性

带来的困难[154].因此文献[154]对他们面临的挑战进行了总结,并提出了一些解决方案,如了解 GDPR 的法律细节、及时与法律团队合作、提前征求潜在受试者的同意等.

4 挑战与机遇

在深入调研现阶段基于 GDPR 合规性研究现状,以及总结 GDPR 相关的技术应用研究现状的基础上,指出了基于 GDPR 的数据隐私安全面临的十大挑战,并给出了可用于应对这些挑战的潜在安全技术研究方向,其对应关系如表 7 所示:

Table 7 Challenges and Opportunities of Data Privacy Security Based on GDPR

表 7 基于 GDPR 的数据隐私安全的挑战与机遇

序号	挑战	机遇
1	软件合规性检测	基于 GDPR 的软件开发规则
2	合规性审计	区块链、智能合约
3	第三方追踪	细粒度更高的流量检测
4	隐私政策语料库的扩建	多语言融合的语料库
5	异构系统设计	分布式、多层次系统框架组件
6	隐私设计框架模型	集成化隐私设计框架
7	DPIA 程序设计	标准化 DPI1A 程序设计指南
8	数据跨境国际规制和标准统一	完善国内跨境数据管控体系,积极参与国际规则制定
9	数据跨境安全技术支持缺乏针对性	融合跨境数据特点进行技术和研发
10	区块链数据擦除	去掉可信第三方的可编辑区块链

4.1 软件合规性检测

对于资源相对匮乏的中小企业来说,无论是遵守 GDPR 还是对已开发的应用软件进行 GDPR 合规性检测都是一个较大的挑战.但如果有一套在软件开发之初就能实现 GDPR 合规性的开发规则,以设计和默认来实现数据保护,就可以大大减少资源的浪费.文献[155]曾提出在需求工程期间解决这个问题,并打算基于 NLP 的自动化方法来实现.目前,对于这方面的研究才刚刚开始,实现这种挑战仍待安全研究人员进一步探索.

4.2 GDPR 合规性审计

由于监管机构的合规性审计是不定期进行的,并且个人用户也无法感知服务商是否有效保护了他们的个人数据,更无法感知服务商何时何处对他们同意的数据进行处理和利用.基于此类问题,一个潜在的研究方向是使用技术手段来提高企业访问用户数据的透明度使得用户可以感知,并且让违规行为不可篡改以便于监管机构进行执法.区块链和智能合约技术可以在一定程度上解决这个问题,分布式账本保证了所有的数据活动不可篡改,而智能合约可以保证触发违规行为之后不可撤销,违反 GDPR 规则的行为会被自动报告.文献[33,156-157]将区块链和智能合约技术应用到 GDPR 的规范中,目前智能合约技术在 GDPR 合规性检测方面的应用尚处于起步阶段,值得进一步的研究.

4.3 针对第三方跟踪服务的流量分析

移动应用系统中开发者会出于盈利目的整合具有强隐蔽性的第三方服务,用户往往无法察觉这类服务的存在,更不用说知道这些服务能够在多大程度上收集、关联和汇总他们的个人数据.尽管此类情况因 GDPR 的出台加以改善,但是由于应用市场包含了数以百万计的应用,很难大规模地执行这些法规,并且由于第三方跟踪服务的不透明性和开发者的授权,很难发现和追踪第三方服务的行为.更进一步的,GDPR 只是规定了对用户数据的收集和处理必须基于明确的用户同意,并没有限制这些第三方跟踪机构对数据的共享和销售.基于这种情况,现阶段在流量层面上对应用程序进行分析,研究应用程序和第三方跟踪服务之间的交互过程依然是非常有必要的,如何高效地对大量跟踪流量进行精准的识别和分析依然是未来的一大挑战.

4.4 隐私政策语料库的扩建

对隐私政策的分析研究多采用监督学习技术,这类技术的准确度都是大量可靠的数据集训练得来的.数据集的标注又是一个耗时耗力的工作过程,需要大量具备专业知识的人员耐心整理.业内工作者对数据集宽度与深度的持续要求,意味着要不断投入大量人力资源.那么是否可以利用对比学习、自注意力机制等无监督学习技术降低人力的投入,达到合理有效的利用社会资源目的.这样只需要一部分数据科学家对网上收集的大量的法律法规文件进行清理,再将这些文件用于预训练.此外,因预训练时的文件资料可包括多国语言信息,使用这类方法获得的模型具备良好的多模态基础,可通过巧妙的设置下游任务,实现多语言合规性的并行检测.

4.5 异构系统设计框架

开发人员在设计系统模型框架时考虑 GDPR 原则有助于帮助企业更好的处理个人数据并保障个人数据隐私.但当前的系统框架多基于不同应用场景、基于部分个人数据保护原则设计,存在一定的局

限性,并且如何平衡系统功能的有效性和 GDPR 隐私保护的合规性仍然是一个问题.分布式多层次的系统框架能够实现分化隐私保护等级并兼顾多项 GDPR 原则,尽管现有研究已经实现了基于部分 GDPR 原则的分布式框架设计,但面对大数据环境下系统的异构性和多源数据,如何设计分布式多层次的系统框架组件,实现系统性能的提升以及数据隐私的保护,还需要进一步加以研究.

4.6 隐私设计框架模型

在设计隐私框架时需尽可能多的考虑 GDPR 相关法规原则,以实现从源头保障组织的合规性,进而减少企业的经济损失.然而,现有隐私设计框架虽能有效保存和处理个人数据,但仅关注部分 GDPR 原则问题,缺乏对隐私设计整体的认识.尽管采用构建模式库的方法检索模式实现满足 GDPR 的隐私设计,解决了整体隐私设计问题,但需要不断更新模式库满足不断变化的设计要求.因此设计集成化隐私框架,仍然需要研究人员进一步探索.

4.7 DPIA 程序设计指南

有效的 DPIA 方法能够帮助企业在早期阶段识别并解决问题,使企业的安全风险最小化.但 GDPR 只提供了 DPIA 的相关标准,在如何实施 DPIA 方面并没有给出明确的 DPIA 模板.如何为每家企业提供可行的 DPIA 方法成为了当今的一大难题,尤其对于资源有限的企业来说,专业指导的缺失会使企业难以设计适合自身的 DPIA 流程,因此需要针对不同领域设计专门的标准化的 DPIA 程序,帮助企业建立自己的 DPIA 模板.

4.8 数据跨境国内外制度体系建设

我国目前在跨境数据领域的管理体系还在起步阶段,相关指南和标准尚处于起草和征询意见的阶段,这不仅需要不同领域的专家和研究人员针对不同敏感程度的数据制定相应的保护等级分划方案和具体说明来指导不同的数据操作,还需要不断完善整个数据管理体系.此外,不同国家或组织拥有的不同的法律规制意味着不同水平的跨境数据保护水平,这在很大程度上阻碍了数据的跨境流动,因此,如何最大限度降低国家之间政策差异导致的影响以及如何在数据跨境领域形成统一的国际规制,在保障我国重要数据安全性的同时更好地促进以数据为载体的国际交流和合作是目前亟待解决的问题.

4.9 数据跨境安全技术完善升级

数据跨境相比传统的数据操作具有步骤更加繁琐,风险因素更加多变和复杂,安全问题影响更深和代价更大等特点,因此,传统安全技术方案也需要得到相应的升级;另外,我国在数据跨境领域的风险评估体系也在积极建设,急需通过研究分析数据跨境流动潜在的风险因素并提前部署相关措施以便在支持数据健康流动的同时更好地保障跨境数据在整个周期的安全性.

4.10 实现区块链数据擦除

区块链的不变性可以为数据处理提供防篡改的记录,增强数据处理的透明度,但区块链的不变性意味着区块链不允许进行任何的修改,这一点并不符合 GDPR 有关数据修改和删除的规定,尤其是第 17 条规定的被遗忘权.现有的解决方法有离线数据存储、遗忘区块链、变色龙哈希等,但这些方法仍然需要借助可信的第三方来实现区块链的合规性,无法提供完全的隐私安全保证.若想要区块链技术在 GDPR 合规方面发挥更大的效用,则需要解决这一难题.因此如何设计更有效的方法解决区块链的不变性与 GDPR 被遗忘权之间的冲突将会是未来的一个研究热点.

5 对中国的启示

GDPR 作为个人数据保护领域的一部重要的法律规定,有着非常典型的示范意义.受其域外适用效力的影响,全球范围内的众多跨国企业的数据安全都面临了很大的挑战.中国在数据安全领域也同样出台了《个人数据保护法》等相关的法律法规及行业规则.中国出台的法律法规与 GDPR 的要求具有某种程度的一致性,但也存在一定的差异.在这样的背景下,中国应该如何更好的改进是一个值得探讨的问题.本节从 6 个方面探讨了 GDPR 给中国带来的启示.

5.1 跨境数据管控体系建设

GDPR 中关于数据跨境的具体要求对欧盟来说,主要是针对从境外流向欧盟境内的数据,而关于对我国个人数据跨境流动立法,我们必须明确我国在个人数据跨境流动中的地位和立场,理清中国作为数据输入国和输出国所需要的不同制度要求,同时做到维护国内用户信息数据跨境安全性以及与第三方国家进行交流和贸易的合规要求.如今我国相关立法体系还未完善,虽相较于完全的数据本地化态度和政策,当前所采用的“知情-同意”原则已经有一定的进步意义,但仍旧无法与中国互联网企业和数字经济大步向前迈进的趋势相适应,目前我们仍然

需要从其他经济体的数据跨境制度和实践中获取经验,探索属于中国的高适应性数据跨境方案.

除了制度,技术也要齐头并进,做到和跨境数据规制相互衔接.新的技术形式可能会给数据管理带来新的潜在风险,针对其中可能出现的漏洞,不仅需要及时了解技术发展新动态,将数据跨境需求融入技术更新,还要紧密联系技术和制度,为跨境数据管控筑牢保护屏障,从技术角度深入分析来辅助制度体系建设,以便更好地迎接挑战.

对于我国个人数据流动的监管,不仅需要建立相应的数据监管机构和数据评估机构对其进行职能划分,使其每个环节中的部门都明确相应的职责,更全面地对跨境数据进行评估,更好地监管数据在跨境前后以及整个生命周期中各个流程的安全性;还要积极引导企业和相关机构进行自评,因此,国家目前正在积极准备出台的《评估办法》和《评估指南》就需要细致化,衡量标准不宜过于笼统、模糊以及主观随意性太强,降低评估流程执行难度的同时提升数据保护强度;其次,各个行业应积极参与评估体系的建设,使其符合实际需求和落实条件,倡导行业自律,帮助建立可操行性强的数据跨境行业体系.

5.2 数据分类分级管理

为了应对 GDPR 以及各国的数据保护法,确保数据在流动过程中的安全性,对数据进行分类分级存储,建立分类分级跨境数据流动管理体系极其重要,《中华人民共和国数据安全法》对数据的分类分级保护作出了明确的要求.数据的分类分级管理是对数据全流程、全过程进行保障的基础,边界防护、入侵防范、身份鉴别、访问控制、数据加密等数据隐私防护方法如果建立在数据分类分级的基础上,可以达到事半功倍的效果[158].2022 年 9 月,全国信安标委完成了国家标准《信息安全技术 网络数据分类分级要求》征求意见稿,健全了《数据安全法》的数据分类分级保护规则[159].

数据分类重点在于理解数据的本质、属性、权属及其相关关系,清晰了解各个数据是如何被使用的,明确哪些数据属于哪个业务范畴,分类不能太细也不能太粗犷[160].可根据监管与合规、业务体系、功能单元、项目等维度进行分类.不同的企业分类的方法和标准也可能不同,例如烟草商业行业会根据数据的来源、敏感度等进行分类,按照业务类别将数据分为营销数据、专卖数据、财务数据、人事数据、供应链数据、考核数据、个人数据 7 个大类,然后再在大类下面细分小类,层级划分逐步扩大[161].

数据分级主要是根据数据泄露或被破坏所造成的影响范围、影响对象、影响程度来进行划分.还需要依据数据的关键性、数据对业务的重要性、以及国内外相关法律的要求进行划分,例如 GDPR 对于任何收集、传输、保留或处理涉及到欧盟所有成员国内的个人信息的机构组织均提出了规范要求[160].常用的数据分级步骤为首先确定分级对象,然后根据数据破坏对国家安全、社会秩序、公共利益造成的影响,数据破坏对企业利益造成的影响,数据破坏对用户利益造成的影响,3 个层面综合评定对客体的侵害程度,最后决定数据对象的安全等级[162].数据安全法把数据分为涉密数据和非涉密数据,涉密数据分为绝密、机密、秘密 3 个级别;非涉密数据根据对国家安全、社会秩序、公共利益以及相关公民、法人造成的危害程度依次分为了 5 个级别[163].

由于数据的海量、多元、非结构化成常态,数据的分类分级难度很大,我国目前在数据分类分级准则方面还有很多欠缺.目前主要努力的方向就是在遵守安全性、可执行性、时效性、就高不就低等分类分级原则的前提下健全数据分类分级管理制度,根据各行业各领域数据资源特点、流通场景,加快制定适应本行业本领域数据流通和开发利用需求的数据分类分级标准.表 8 列出了中国发布和在研的数据分类分级标准.

Table 8 Classification and Gradation Standards for Published and Developing Data[162] in China

表 8 中国发布和在研的数据分类分级标准[162]

标号	编号	名称
1	GB/Z 20986—2007	信息安全技术信息安全事件分类分级指南
2	YD/T 3813—2020	基础电信企业数据分类分级方法
3	JR/T 0197—2020	金融数据安全 数据安全分级指南
4	JR/T 0158—2018	证券期货业数据分类分级指引
5	T/ZAIF 1002—2020	互联网金融组织数据分类分级指南
6	JR/T 0171—2020	个人金融信息保护技术规范
7		工业数据分类分级指南(试行)
8	DB 52/T 1123—2016	政府数据 数据分类分级指南(贵州省)
9	DB 33/T 2351—2021	数字化改革 公共数据分类分级指南(浙江省)
10		信息安全技术 网络数据分类分级要求(征求意见稿)

5.3 重要数据识别与保护

作为数据安全中的重点保护对象,重要数据在中国的数据安全管理制度中一直占据着极其重要的地位.2017 年我国出台的《网络安全法》第一次提出

了“重要数据”的概念,2021 年出台的《数据安全法》再次在数据分类分级保护制度中提到了对“重要数据”的保护义务,但这 2 部法律均未对“重要数据”作出具体定义,重要数据的定义范围及其识别方法成为了一个关键的问题.在 2022 年发布的《信息安全技术 重要数据识别规则(征求意见稿)》中,“重要数据”被定义为“特定领域、特定群体、特定区域或达到一定精度和规模的数据,一旦被泄露或篡改、损毁,可能直接危害国家安全、经济运行、社会稳定、公共健康和安全”[164].

重要数据识别是数据安全管理工作的基石,一个企业对于重要数据的收集处理直接影响着企业数据的安全合规性.重要数据识别工作主要分为 3 步:1)通过扫描发现和流量检测的方式对企业数据进行初步识别,形成企业数据资产梳理清单;2)根据行业要求对企业数据进行分类分级;3)依据重要数据识别规则对重要数据进行判定并标识,并根据重要数据的基本信息、分类、重要性及用途等信息汇总出企业重要数据清单.重要数据的识别主要包括聚焦安全影响、突出保护重点、衔接既有规定、考虑风险、定量定性结合、动态识别复评六大原则.除此之外还要针对重要数据的收集、存储和使用采取重点保护措施,对于重要数据的数据处理者要提出更高的合规要求,这样才能保证数据流通的合规有序,充分发挥数据要素的价值.

目前我国有关重要数据识别相关规则的建立仍在起步阶段,重要数据识别总体要求《信息安全技术 重要数据识别规则》仍在不断修改中,各行业也依据标准制定行业内重要数据安全管理的相关细则,例如电信领域出台的《基础电信企业重要数据识别指南》及汽车领域出台的《汽车数据安全管理若干规定(试行)》等.中国亟待健全相关的重要数据识别与保护细则,走好重要数据安全防护体系建立的第一步.

5.4 关注不同规模企业的合规义务

虽然 GDPR 实现了对个人数据隐私的严格保护,但是对于市场经济的发展有时却会起到适得其反的效果.尤其在市场竞争方面,由于大型企业拥有充足的资金和研发能力,能够很好地应对 GDPR 带来的一系列合规性问题,而对于中小企业来说,过高的合规成本阻碍了企业发展的脚步,因此大型企业的竞争力大大增强,市场份额不断增加,而中小企业在这场浪潮中却步履维艰.虽然 GDPR 对于中小企业有相关的特殊豁免政策,然而实际执行的过程中并未能落到实处.

因此在中国相关数据保护政策实施过程中要重点关注中小企业的发展,平衡不同规模企业之间的市场竞争利益,这样有利于市场竞争的公平性,激发市场创新活力.对于合规性监管的过程中要避免进行一刀切管理,应对不同规模的企业赋予相应的合规责任,适当减轻中小企业的合规义务,使中小企业的特殊政策能够落到实处.

5.5 保护个人数据的同时兼顾社会经济发展

GDPR 基于个人控制论强化了数据主体对个人信息的控制,使得主体权利凌驾于社会利益、公共利益之上,并没有考虑个人信息的社会属性,造成了 GDPR 存在巨大的内在缺陷.数据控制者及处理者针对个人数据处理以及数据再利用或初始目的之外的使用需要通过大量设置同意实现,最终导致同意的滥用.同时,使用同意的预防保护机制处理泛在个人信息将会导致社会运行成本过高.并且泛在的个人信息及数据处理导致 GDPR 的适用范围无限扩大,进而引发侵害个人权利的风险,数据主体也可借助 GDPR 与其他众多法律的重叠现象来选择有利于自身的权限基础.

因此,面对 GDPR 确立的个人信息保护准则正在成为全球化标杆,我们应当从我国社会实际问题及需求出发,建立符合中国特色的数据经济制度需求.明确 GDPR 根植欧洲的政治和社会文化背景与我国社会经济文化的差异性,兼顾数字化时代个人数据控制困难问题,以及缓解泛在的个人信息处理同社会运行成本间的冲突,以促进我国个人数据保护法案的进一步升级,保障个人数据权益与数字经济的协同发展.

5.6 在数据保护的同时促进数据流通

在市场经济中要发挥好数据这一生产要素的作用,不仅要严格的数据保护,还要保证数据的流通,创造数据资源的价值,不能一味地强调数据权属,对数据进行僵化管理,让数据失去流动性.为支持数字经济发展,继《通用数据保护条例(GDPR)》之后,欧盟的《数据治理法案(DGA)》《数据法案(DA)》《数据市场法案(DMA)》《数据服务法案(DSA)》等相关法规也相继出台,为数据流动营造开放的环境.2021 年中国施行了《个人信息保护法》,它是我国的第一部个人信息保护方面的法律文件,在这之后又发布了许多数据相关立法,但主要焦点仍在数据保护监管方面,在促进数据流动和创造数据价值上中国仍需要更多的政策支持,对现有政策也需要不断调整和优化,保证数字经济的良好发展态势.

6 结 语

关于 GDPR 的数据隐私安全研究逐年的增加使得企业以及个人对数据隐私保护意识得到了很大的加强,但因其涉及领域较广,且随着各国数据法的不断更新、应用场景的不断变化,其整体还处于起步阶段.本文在调研大量基于 GDPR 的数据隐私安全相关论文及其研究成果后,首先介绍了数据隐私安全发展历程及欧盟 GDPR 法规主要内容,并将其与各国数据法进行了详细的对比;然后通过梳理总结现有的基于 GDPR 的数据隐私安全的研究工作,从 GDPR 违规行为分析、隐私政策分析、GDPR 模型框架 3 个方面阐述了 GDPR 合规性的研究现状;之后总结 GDPR 相关的数据技术应用以及各种合规应用场景.通过深入分析数据隐私安全问题以及现有研究工作的不足,指出了基于 GDPR 的数据隐私安全面临的十大安全技术挑战和机遇;最后指出了跨境数据管控体系建设、数据分类分级管理、重要数据识别与保护、不同规模企业的合规义务、兼顾社会经济发展、促进数据流通等 GDPR 相关研究对中国的启示.

作者贡献声明:赵景欣负责设计研究方案及论文撰写和修订;岳星辉负责调研分析、数据统计及论文部分撰写;冯崇朋、张静负责论文部分撰写及画图;李印负责最终版本修订;王娜负责论文部分撰写;任家东、张昊星、伍高飞、朱笑岩负责论文整体修订;张玉清提出论文整体研究思路,及最终论文的审核与修订.

参 考 文 献

[1] Larson S. Uber paid hackers $100,000 after they stole data on 57 million users [EB/OL]. [2022-05-10]. https://money.cnn.com/2017/11/21/technology/uber-hacked-2016/index.html

[2] Stempel J, Finkle J. Yahoo says all three billion accounts hacked in 2013 data theft. REUTERS [EB/OL]. [2022-05-10]. https://www.reuters.com/article/us-yahoo-cyber-idUSKCN1C82O1

[3] Bankhurst A. Over 533 million Facebook users' phone numbers and personal data has been leaked online [EB/OL]. [2022-08-05]. https://www.ign.com/articles/over-533-million-facebook-users-phone-numbers-and-personal-data-has-been-leaked-online

[4] Hernandez A. VpnMentor report discovers 63 million users information in data leak [EB/OL]. [2022-08-05]. https://techaeris.com/2021/08/04/vpnmentor-report-discovers-63-million-users-information-in-data-leak/

[5] Fung B. T-Mobile agrees to pay customers $350 million in settlement over massive data breach [EB/OL]. [2022-08-05]. www.cnn.com/2022/07/25/tech/tmobile-data-breach-settlement/index.html

[6] Haque A K M B, Islam A K M N, Hyrynsalmi S, et al. GDPR compliant blockchains—A systematic literature review [J]. IEEE Access, 2021, 9: 50593-50606

[7] Kounoudes A D, Kapitsaki G M. A mapping of IoT user-centric privacy preserving approaches to the GDPR [J]. Internet of Things, 2020, 11: 100179

[8] Akil M, Islami L, Fischer-Hubner S, et al. Privacy-Preserving Identifiers for IoT: A Systematic Literature Review [J]. IEEE Access, 2020, 8: 168470-168485

[9] Acosta L H, Reinhardt D. A survey on privacy issues and solutions for voice-controlled digital assistants [J]. Pervasive and Mobile Computing, 2022, 80: 101523

[10] Eugenia P, Efthimios A, Constantinos P. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions [J]. Journal of Cybersecurity, 2018, 4(1): 1-20

[11] Alunge R. Consolidating the right to data protection in the information age: A Comparative appraisal of the adoption of the OECD (Revised) guidelines into the EU GDPR, the Ghanaian Data Protection Act 2012 and the Kenyan Data Protection Act 2019 [C] //Proc of the Int Conf on Innovations and Interdisciplinary Solutions for Underserved Areas. Berlin: Springer, 2020: 192-207

[12] Hexun. A record decade: data privacy legislation in the 2010s [EB/OL]. [2022-05-10]. http://news.hexun.com/2020-07-06/201663668.html (in Chinese)
(和讯名家. 创纪录十年: 2010 年代的数据隐私立法 [EB/OL]. [2022-05-10]. http://news.hexun.com/2020-07-06/201663668.html)

[13] Wang Jie, Wei Tong. Comparison of the main points of the personal information protection law (draft) and the data protection laws of many countries [EB/OL]. [2022-05-10]. http://www.baijingapp.com/article/id-30896 (in Chinese)
(王捷, 魏彤. 个人信息保护法(草案)与多国数据保护法要点对比 [EB/OL]. [2022-05-10]. http://www.baijingapp.com/article/id-30896)

[14] San Zheng Science and Technology. Trends of data protection legislation in various countries in 2021. [EB/OL]. [2022-05-10]. https://www.163.com/dy/article/GQHH3GMP0552MROB.html (in Chinese)
(三正科技. 2021 年各国数据保护法立法动向 [EB/OL]. [2022-05-10]. https://www.163.com/dy/article/GQHH3GMP0552MROB.html)

[15] The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 (General Data Protection Regulation) [EB/OL]. [2022-04-10]. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[16] Degeling M, Utz C, Lentzsch C, et al. We value your privacy... now take some cookies: Measuring the GDPR's impact on Web privacy [J]. arXiv preprint, arXiv: 1808.05096, 2018

[17] Intersoft consulting. GDPR Fines/Penalties [EB/OL]. [2022-05-10]. https://gdpr-info.eu/issues/fines-penalties/

[18] Jennifer Lund. What is GDPR and how does it impact your business [EB/OL]. [2022-05-10]. https://www.superoffice.com/blog/gdpr/

[19] Blog GDPR. 20 biggest GDPR fines so far[2019, 2020, 2021 & 2022] [EB/OL]. [2022-05-10]. https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/

[20] Presthus W, Sønslien K F. An analysis of violations and sanctions following the GDPR [J]. International Journal of Information Systems and Project Management, 2021, 9(1): 38-53

[21] Papageorgiou A, Strigkos M, Politou E, et al. Security and privacy analysis of mobile health applications: the alarming state of practice [J]. IEEE Access, 2018, 6: 9390-9403

[22] Nguyen T T, Backes M, Marnau N, et al. Share First, Ask Later (or Never?) Studying Violations of {GDPR's} Explicit Consent in Android Apps [C] //Proc of the 30th USENIX Security Symp. Berkeley, CA: USENIX Association, 2021: 3667-3684

[23] Subahi A, Theodorakopoulos G. Ensuring compliance of IoT devices with their privacy policy agreement [C] //Proc of the 6th IEEE Int Conf on Future Internet of Things and Cloud. Piscataway, NJ: IEEE, 2018: 100-107

[24] Cabañas J G, Cuevas Á, Cuevas R. Unveiling and quantifying facebook exploitation of sensitive personal data for advertising purposes [C] //Proc of the 27th USENIX Security Symp. Berkeley, CA: USENIX Association, 2018: 479-495

[25] Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, et al. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem [C/OL] //Proc of the 25th Annual Network and Distributed System Security Symp (NDSS). San Diego, CA: Internet Society, 2018[2022-04-30]. https://dx.doi.org/10.14722/ndss.2018.23009

[26] Matte C, Bielova N, Santos C. Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe's transparency and consent framework [C] //Proc of the 2020 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2020: 791-809

[27] Sanchez-Rola I, Dell'Amico M, Kotzias P, et al. Can I opt out yet? GDPR and the global illusion of cookie control [C] //Proc of the 2019 ACM Asia Conf on Computer and Communications Security. New York: ACM, 2019: 340-351

[28] Sakamoto T, Matsunaga M. After GDPR, still tracking or not? Understanding opt-out states for online behavioral advertising [C] //Proc of the 2019 IEEE Security and Privacy Workshops (SPW). Piscataway, NJ: IEEE, 2019: 92-99

[29] Slavin R, Wang Xiaoyin, Hosseini M B, et al. Toward a framework for detecting privacy policy violations in android application code [C] //Proc of the 38th Int Conf on Software Engineering. New York: ACM, 2016: 25-36

[30] Fan Ming, Yu Le, Chen Sen, et al. An empirical evaluation of GDPR compliance violations in Android mHealth Apps [C] //Proc of the 31st IEEE Int Symp on Software Reliability Engineering (ISSRE). Piscataway, NJ: IEEE, 2020: 253-264

[31] AlEroud A, Masalha F, Saifan A A. Identifying GDPR privacy violations using an augmented LSTM: Toward an AI-based violation alert systems [C] //Proc of the 2021 IEEE Int Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). Piscataway, NJ: IEEE, 2021: 1617-1624

[32] Jia Qiwei, Zhou Lu, Li Huaxin, et al. Who leaks my privacy: Towards automatic and association detection with GDPR compliance [C] //Proc of the Int Conf on Wireless Algorithms, Systems, and Applications. Berlin: Springer, 2019: 137-148

[33] Truong N B, Sun Kai, Lee G M, et al. Gdpr-compliant personal data management: A blockchain-based solution [J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 1746-1761

[34] Tesfay W B, Hofmann P, Nakamura T, et al. Privacy Guide: Towards an implementation of the EU GDPR on Internet privacy policy evaluation [C] //Proc of the 4th ACM Int Workshop on Security and Privacy Analytics. New York: ACM, 2018: 15-21

[35] Tesfay W B, Hofmann P, Nakamura T, et al. I read but don't agree: Privacy policy benchmarking using machine learning and the EU GDPR [C] //Companion Proc of the The Web Conf 2018. New York: ACM, 2018: 163-166

[36] Farke F M, Balash D G, Golla M, et al. Are privacy dashboards good for end users? Evaluating user perceptions and reactions to Google's my activity [C] //Proc of the 30th USENIX Security Symp. Berkeley, CA: USENIX Association, 2021: 483-500

[37] Schufrin M, Reynolds S L, Kuijper A, et al. A visualization interface to improve the transparency of collected personal data on the Internet [J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 27(2): 1840-1849

[38] Russo A, Lax G, Dromard B, et al. A system to access online services with minimal personal information disclosure [J]. Information Systems Frontiers, 2021 [2022-06-30]. https://doi.org/10.1007/s10796-021-10150-8

[39] Bonatti P A, Sauro L, Langens J. Representing consent and policies for compliance [C] //Proc of the 2021 IEEE European Symp on Security and Privacy Workshops (EuroS&PW). Piscataway, NJ: IEEE, 2021: 283-291

[40] Chhetri T R, Kurteva A, DeLong R J, et al. Data protection by design tool for automated GDPR compliance verification based on semantically modeled informed consent [J]. Sensors, 2022, 22(7): No.2763

[41] Debruyne C, Pandit H J, Lewis D, et al. "Just-in-time" generation of datasets by considering structured representations of given consent for GDPR compliance [J]. Knowledge and Information Systems, 2020, 62(9): 3615-3640

[42] Kreuter F, Haas G C, Keusch F, et al. Collecting survey and smartphone sensor data with an App: Opportunities and challenges around privacy and informed consent [J]. Social Science Computer Review, 2020, 38(5): 533-549

[43] Weir C, Hermann B, Fahl S. From needs to actions to secure apps? the effect of requirements and developer practices on app security [C] //Proc of the 29th USENIX Security Symp. Berkeley, CA: USENIX Association, 2020: 289-305

[44] McDonald A M, Cranor L F. The cost of reading privacy policies [J]. A Journal of Law and Policy for the Information Society, 2008, 4(3): 540-565

[45] Liu Shuang, Zhao Baiyang, Guo Renjie, et al. Have You been properly notified? Automatic compliance analysis of privacy policy text with GDPR article 13 [C] //Proc of the Web Conf 2021. New York: ACM, 2021: 2154-2164

[46] Xiao Xusheng, Paradkar A, Thummalapenta S, et al. Automated extraction of security policies from natural-language software documents [C] //Proc of the 20th ACM SIGSOFT Int Symp on the Foundations of Software Engineering. New York: ACM, 2012: 1-11

[47] Costante E, Hartog J, Petković M. What websites know about you [M] //Data Privacy Management and Autonomous Spontaneous Security. Berlin: Springer, 2012: 146-159

[48] Brodie C A, Karat C M, Karat J. An empirical study of natural language parsing of privacy policy rules using the SPARCLE policy workbench [C] //Proc of the 2nd Symp on Usable Privacy and Security. New York: ACM, 2006: 8-19

[49] Müller N M, Kowatsch D, Debus P, et al. On GDPR compliance of companies' privacy policies [C] //Proc of the Int Conf on Text, Speech, and Dialogue. Berlin: Springer, 2019: 151-159

[50] Asif M, Javed Y, Hussain M. Automated analysis of Pakistani websites' compliance with GDPR and Pakistan Data Protection Act [C] //Proc of the 2021 Int Conf on Frontiers of Information Technology (FIT). Piscataway, NJ: IEEE, 2021: 234-239

[51] Wilson S, Schaub F, Dara A A, et al. The creation and analysis of a website privacy policy corpus [C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1330-1340

[52] Sathyendra K M, Wilson S, Schaub F, et al. Identifying the provision of choices in privacy policy text [C] //Proc of the 2017 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 2774-2779

[53] Poplavska E, Norton T B, Wilson S, et al. From prescription to description: Mapping the GDPR to a privacy policy corpus annotation scheme [C] //Legal Knowledge and Information Systems-JURIX 2020: 33rd Annual Conf. Amsterdam: IOS Press, 2020: 243-246

[54] Zimmeck S, Story P, Smullen D, et al. Maps: Scaling privacy compliance analysis to a million apps [J]. Proceedings on Privacy Enhancing Technologies, 2019, 2019(3): 66-86

[55] Lebanoff L, Liu Fei. Automatic detection of vague words and sentences in privacy policies [C] //Proc of the 2018 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 3508-3517

[56] Torre D, Abualhaija S, Sabetzadeh M, et al. An AI-assisted approach for checking the completeness of privacy policies against GDPR [C] //Proc of the 2020 IEEE 28th Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2020: 136-146

[57] Razavisousan R, Joshi K P. Analyzing GDPR compliance in cloud services' privacy policies using textual Fuzzy interpretive structural modeling (TFISM) [C] //Proc of the 2021 IEEE Int Conf on Services Computing (SCC). Piscataway, NJ: IEEE, 2021: 89-98

[58] Elluri L, Joshi K P, Kotal A. Measuring semantic similaity across EU GDPR regulation and cloud privacy policies [C] //Proc of the 2020 IEEE Int Conf on Big Data (Big Data). Piscataway, NJ: IEEE, 2020: 3963-3978

[59] Pandit H J, Fatema K, O'Sullivan D, et al. GDPRtEXT-GDPR as a linked data resource [C] //Proc of the European Semantic Web Conf. Berlin: Springer, 2018: 481-495

[60] Ryan P, Pandit H J, Brennan R. A common semantic model of the GDPR register of processing activities [J]. arXiv preprint, arXiv:2102.00980, 2021

[61] Hickey D, Brennan R. A GDPR International Transfer Compliance Framework Based on an Extended Data Privacy Vocabulary (DPV) [M] //Legal Knowledge and Information Systems. Amsterdam: IOS Press, 2021: 161-170

[62] Davari M, Bertino E. Access control model extensions to support data privacy protection based on GDPR [C] //Proc of the 2019 IEEE Int Conf on Big Data (Big Data). Piscataway, NJ: IEEE, 2019: 4017-4024

[63] Zaman R, Cuzzocrea A, Hassani M. An innovative online process mining framework for supporting incremental GDPR compliance of business processes [C] //Proc of the 2019 IEEE Int Conf on Big Data (Big Data). Piscataway, NJ: IEEE, 2019: 2982-2991

[64] Piras L, Al-Obeidallah M G, Praitano A, et al. DEFeND architecture: A privacy by design platform for GDPR compliance [C] //Proc of the Int Conf on Trust and Privacy in Digital Business. Berlin: Springer, 2019: 78-93

[65] Rhahla M, Allegue S, Abdellatif T. A framework for GDPR compliance in big data systems [C] //Proc of the Int Conf on Risks and Security of Internet and Systems. Berlin: Springer, 2019: 211-226

[66] Blanco-Lainé G, Sottet J S, Dupuy-Chessa S. Using an enterprise architecture model for GDPR compliance principles [C] //Proc of the IFIP Working Conf on The Practice of Enterprise Modeling. Berlin: Springer, 2019: 199-214

[67] Agarwal S, Steyskal S, Antunovic F, et al. Legislative compliance assessment: framework, model and GDPR instantiation [C] //Proc of the Annual Privacy Forum. Berlin: Springer, 2018: 131-149

[68] Gjermundrød H, Dionysiou I, Costa K. PrivacyTracker: A privacy-by-design GDPR-compliant framework with verifiable data traceability controls [C] //Proc of the Int Conf on Web Engineering. Berlin: Springer, 2016: 3-15

[69] Teixeira C, Vasconcelos A, Sousa P, et al. Enterprise architecture patterns for GDPR compliance [C] //Proc of the 23rd Int Conf on Enterprise Information Systems—Volume 2: ICEIS. Setúbal: SCITEPRESS, 2021: 715-725

[70] Robol M, Salnitri M, Giorgini P. Toward GDPR-compliant socio-technical systems: Modeling language and reasoning framework [C] //Proc of the IFIP Working Conf on the Practice of Enterprise Modeling. Berlin: Springer, 2017: 236-250

[71] Kittmann T, Lambrecht J, Horn C. A privacy-aware distributed software architecture for automation services in compliance with GDPR [C] //Proc of the 23rd IEEE Int Conf on Emerging Technologies and Factory Automation (ETFA). Piscataway, NJ: IEEE, 2018: 1067-1070

[72] Sousa M, Ferreira D N G, Pereira C S, et al. OpenEHR based systems and the general data protection regulation (GDPR)[J]. Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth, 2018, 247: 91-95

[73] Alamri B, Javed I T, Margaria T. A GDPR-compliant framework for IoT-based personal health records using blockchain [C] //Proc of the 11th IFIP Int Conf on New Technologies, Mobility and Security (NTMS). Piscataway, NJ: IEEE, 2021: 1-5

[74] Bieker F, Friedewald M, Hansen M, et al. A process for data protection impact assessment under the european general data protection regulation [C] //Proc of the Annual Privacy Forum. Berlin: Springer, 2016: 21-37

[75] Bieker F, Martin N, Friedewald M, et al. Data protection impact assessment: A hands-on tour of the GDPR's most practical tool [C] //Proc of the IFIP Int Summer School on Privacy and Identity Management. Berlin: Springer, 2017: 207-220

[76] Friedewald M, Schiering I, Martin N, et al. Data protection impact assessments in practice [C] //Proc of the European Symp on Research in Computer Security. Berlin: Springer, 2021: 424-443

[77] Henriksen-Bulmer J, Faily S, Jeary S. Implementing GDPR in the charity sector: A case study [C] //Proc of the IFIP Int Summer School on Privacy and Identity Management. Berlin: Springer, 2018: 173-188

[78] Ahmadian A S, Strüber D, Riediger V, et al. Supporting privacy impact assessment by model-based privacy analysis [C] //Proc of the 33rd Annual ACM Symp on Applied Computing. New York: ACM, 2018: 1467-1474

[79] Hart S, Ferrara A L, Paci F. Fuzzy-based approach to assess and prioritize privacy risks [J]. Soft Computing, 2020, 24 (3): 1553-1563

[80] Wei Yuzhi, Wu Weichen, Lai Guxin, et al. pISRA: privacy considered information security risk assessment model [J]. The Journal of Supercomputing, 2020, 76(3): 1468-1481

[81] Comandè G, Schneider G. Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for data protection-sound research activities [J]. Computer Law & Security Review, 2021, 41: 1-5

[82] Wagner J. The transfer of personal data to third countries under the GDPR: When does a recipient country provide an adequate level of protection? [J]. International Data Privacy Law, 2018, 8(4): 318-337

[83] Chen Dingzhuang. The influence of EU general data protection regulation on international service trade rules [J]. China Business and Market, 2021, 35(4): 93-102 (in Chinese)
(陈鼎庄. 欧盟《一般数据保护条例》对国际服务贸易规则的影响[J]. 中国流通经济, 2021, 35(4): 93-102)

[84] Zhang Monan. Cross-border data flow: Global situation and the countermeasures for China [J]. China Opening Journal, 2020, 29(2): 44-50 (in Chinese)
(张茉楠. 跨境数据流动:全球态势与中国对策[J]. 开放导报, 2020, 29(2): 44-50)

[85] Fefer R F. Data flows, online privacy, and trade policy [J/OL]. [2022-04-30]. https://sgp.fas.org/crs/row/R45584.pdf

[86] Feng Ran. Enlightenment of cross-border data security and personal privacy protection measures of Europe and the United States for China [C] //Proc of the 2019 Int Conf on Artificial Intelligence and Computer Science. New York: ACM, 2019: 793-796

[87] Siapera M, Douloudis K, Prentza A. A common data model for once-only cross-border data exchanges in Europe [C] //Proc of the 14th Int Conf on Theory and Practice of Electronic Governance. New York: ACM, 2021: 223-230

[88] Hong Yanqing. Building a framework for security review of the cross-border data flow [J/OL]. [2022-04-30]. http://dx.doi.org/10.2139/ssrn.3401615

[89] Hidano S, Kiyomoto S, Biswas A R, et al. Access control for cross-border transfer of sensor data [C] //Proc of the Int Symp on Mobile Internet Security. Berlin: Springer, 2016: 143-153

[90] Jara A J, Bocchi Y. GEO-Trust: Geo-aware security protocol for enabling cross-border trustable operations and data exchange in a global digital economy [C] //Proc of the 1st IEEE Sustainable Cities Latin America Conf (SCLA). Piscataway, NJ: IEEE, 2019: 1-6

[91] Zhang Xiaodong, Chen Taowei, Feng Yan, et al. A data sharing scheme based on blockchain system and attribute-based encryption [C] //Proc of the 3rd Int Conf on Blockchain Technology. New York: ACM, 2021: 195-202

[92] Rahman M S, Al Omar A, Bhuiyan M Z A, et al. Accountable cross-border data sharing using blockchain under relaxed trust assumption [J]. IEEE Transactions on Engineering Management, 2020, 67(4): 1476-1486

[93] Hu Runshan. Dealing with privacy risk: Solutions to data sharing under the GDPR for data controllers [D]. United Kingdom: University of Southampton, 2020

[94] Gangopadhyay B, Jetla V, Patil S R, et al. Cross border data flow governance in storage cloud leveraging deep learning techniques [C] //Proc of the 2018 IEEE Int Conf on Cloud Computing in Emerging Markets (CCEM). Piscataway, NJ: IEEE, 2018: 17-22

[95] Iordanou C, Smaragdakis G, Poese I, et al. Tracing cross border Web tracking [C] //Proc of the Internet Measurement Conf 2018. New York: ACM, 2018: 329-342

[96] Mayer J R, Mitchell J C. Third-party Web tracking: Policy and technology [C] //Proc of the 2012 IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2012: 413-427

[97] Guamán D S, Del Alamo J M, Caiza J C. GDPR compliance assessment for cross-border personal data transfers in Android Apps [J]. IEEE Access, 2021, 9: 15961-15982

[98] Mohammadi F, Panou A, Ntantogian C, et al. CUREX: secure and private health data exchange [C] //Proc of the IEEE/WIC/ACM Int Conf on Web Intelligence-Companion Volume. New York: ACM, 2019: 263-268

[99] Surridge M, Meacham K, Papay J, et al. Modelling compliance threats and security analysis of cross border health data exchange [C] //Proc of the Int Conf on Model and Data Engineering. Berlin: Springer, 2019: 180-189

[100] Tsakalakis N, Stalla-Bourdillon S, O'hara K. Data protection by design for cross-border electronic identification: Does the eIDAS interoperability framework need to be modernised? [C] //Proc of the IFIP Int Summer School on Privacy and Identity Management. Berlin: Springer, 2018: 255-274

[101] Staffa M, Sgaglione L, Mazzeo G, et al. An OpenNCP-based solution for secure eHealth data exchange [J]. Journal of Network and Computer Applications, 2018, 116: 65-85

[102] Castaldo L, Cinque V. Blockchain-based logging for the cross-border exchange of eHealth data in Europe [C] //Proc of the Int ISCIS Security Workshop. Berlin: Springer, 2018: 46-56

[103] Gelenbe E, Pavloski M. Performance of a security control scheme for a health data exchange system [C] //Proc of the 2020 IEEE Int Black Sea Conf on Communications and Networking (BlackSeaCom). Piscataway, NJ: IEEE, 2020: 1-6

[104] Todde M, Beltrame M, Marceglia S, et al. Methodology and workflow to perform the data protection impact assessment in healthcare information systems [J]. Informatics in Medicine Unlocked, 2020, 19: 100361

[105] Larrucea X, Moffie M, Asaf S, et al. Towards a GDPR compliant way to secure European cross border healthcare industry 4.0 [J]. Computer Standards & Interfaces, 2020, 69: 103408

[106] Sağlam R B, Aslan Ç B, Li Shujun, et al. A data-driven analysis of blockchain systems' public online communications on GDPR [C] //Proc of the 2020 IEEE Int Conf on Decentralized Applications and Infrastructures (DAPPS). Piscataway, NJ: IEEE, 2020: 22-31

[107] Liang Xueping, et al. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability [C] //Proc of the 17th IEEE/ACM Int Symp on Cluster, Cloud and Grid Computing (CCGRID). Piscataway, NJ: IEEE, 2017: 468-477

[108] Neisse R, Steri G, Nai-Fovino I. A blockchain-based approach for data accountability and provenance tracking [C] //Proc of the 12th Int Conf on Availability, Reliability and Security. New York: ACM, 2017: 1-10

[109] Barati M, Petri I, Rana O F. Developing GDPR compliant user data policies for Internet of things [C] //Proc of the 12th IEEE/ACM Int Conf on Utility and Cloud Computing. New York: ACM, 2019: 133-141

[110] Barati M, Rana O. Privacy-aware cloud ecosystems: Architecture and performance [J]. Concurrency and Computation: Practice and Experience, 2021, 33: No.e5852

[111] Zyskind G, Nathan O. Decentralizing privacy: Using blockchain to protect personal data [C] //Proc of the 2015 IEEE Security and Privacy Workshops. Piscataway, NJ: IEEE, 2015: 180-184

[112] Ahmed J, Yildirim S, Nowostaki M, et al. GDPR compliant consent driven data protection in online social networks: A blockchain-based approach [C] //Proc of the 3rd Int Conf on Information and Computer Technologies (ICICT). Piscataway, NJ: IEEE, 2020: 307-312

[113] Faber B, Michelet G C, Weidmann N, et al. BPDIMS: A blockchain-based personal data and identity management system [C] //Proc of the 52nd Hawaii Int Conf on System Sciences. Hawaii: HICSS, 2019: 6855-6864

[114] Daudén-Esmel C, Castellà-Roca J, Viejo A, et al. Lightweight blockchain-based platform for GDPR-compliant personal data management [C] //Proc of the 5th IEEE Int Conf on Cryptography, Security and Privacy (CSP). Piscataway, NJ: IEEE, 2021: 68-73

[115] Calani M, Denaro G, Leporati A. Exploiting the blockchain to guarantee GDPR compliance while consents evolve under data owners' control [C/OL] //Proc of the Italian Conf on Cybersecurity(ITASEC). [2022-04-30]. http://ceur-ws.org/Vol-2940/paper28.pdf

[116] Mahindrakar A, Joshi K P. Automating GDPR compliance using policy integrated blockchain [C] //Proc of the 6th IEEE Int Conf on Big Data Security on Cloud (Big Data Security), IEEE Int Conf on High Performance and Smart Computing (HPSC) and IEEE Int Conf on Intelligent Data and Security (IDS). Piscataway, NJ: IEEE, 2020: 86-93

[117] Camilo J. Blockchain-based consent manager for GDPR compliance [C/OL] //Proc of the Open Identity Summit 2019 [2022-04-30]. https://dl.gi.de/bitstream/handle/20.500.12116/20985/proceedings-14.pdf

[118] Qiu Weiyang, Meng Weizhi, Jensen C D. My data, my control: A secure data sharing and access scheme over blockchain [J]. Journal of Information Security and Applications, 2021, 63: 103020

[119] Fu Weikang, Lin Yishan, Campagna G, et al. Soteria: A provably compliant user right manager using a novel two-layer blockchain technology [C] //Proc of the 2020 IEEE Infrastructure Conf. Piscataway, NJ: IEEE, 2020: 1-10

[120] Tobin A, Reed D. The inevitable rise of self-sovereign identity [R/OL]. Provo: Sovrin Foundation, 2016 [2022-06-15]. https://www.evernym.com/wp-content/uploads/2017/07/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf

[121] Bernabe J B, Canovas J L, Hernandez-Ramos J L, et al. Privacy-preserving solutions for blockchain: Review and challenges [J]. IEEE Access, 2019, 7: 164908-164940

[122] Alsayed Kassem J, Sayeed S, Marco-Gisbert H, et al. DNS-IdM: A blockchain identity management system to secure personal data sharing in a network [J]. Applied Sciences, 2019, 9(15): No.2953

[123] Naik N, Jenkins P. Your identity is yours: Take back control of your identity using GDPR compatible self-sovereign identity [C] //Proc of the 7th Int Conf on Behavioural and Social Computing (BESC). Piscataway, NJ: IEEE, 2020: 1-6

[124] Kondova G, Erbguth J. Self-sovereign identity on public blockchains and the GDPR [C] //Proc of the 35th Annual ACM Symp on Applied Computing. New York: ACM, 2020: 342-345

[125] Sim W L, Chua H N, Tahir M. Blockchain for identity management: The implications to personal data protection [C] //Proc of the 2019 IEEE Conf on Application, Information and Network Security (AINS). Piscataway, NJ: IEEE, 2019: 30-35

[126] Al-Zaben N, Onik M M H, Yang Jinhong, et al. General data protection regulation complied blockchain architecture for personally identifiable information management [C] //Proc of the 2018 Int Conf on Computing, Electronics & Communications Engineering (iCCECE). Piscataway, NJ: IEEE, 2018: 77-82

[127] Molina F, Betarte G, Luna C. Design principles for constructing GDPR-compliant blockchain solutions [C] //Proc of the 4th IEEE/ACM Int Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB). Piscataway, NJ: IEEE, 2021: 1-8

[128] Casino F, Politou E, Alepis E, et al. Immutability and decentralized storage: An analysis of emerging threats [J]. IEEE Access, 2019, 8: 4737-4744

[129] Politou E, Alepis E, Patsakis C, et al. Delegated content erasure in IPFS [J]. Future Generation Computer Systems, 2020, 112: 956-964

[130] Zichichi M, Ferretti S, D'Angelo G. On the efficiency of decentralized file storage for personal information management systems [C] //Proc of the 2020 IEEE Symp on Computers and Communications (ISCC). Piscataway, NJ: IEEE, 2020: 1-6

[131] Zieglmeier V, Daiqui G L. GDPR-compliant use of blockchain for secure usage logs [C] //Proc of the Evaluation and Assessment in Software Engineering. New York: ACM, 2021: 313-320

[132] Ateniese G, Magri B, Venturi D, et al. Redactable blockchain-or-rewriting history in bitcoin and friends [C] //Proc of the 2017 IEEE European Symp on Security and Privacy (EuroS&P). Piscataway, NJ: IEEE, 2017: 111-126

[133] Derler D, Samelin K, Slamanig D, et al. Fine-grained and controlled rewriting in blockchains: Chameleon-hashing gone attribute-based [C/OL] //Proc of the 26th Annual Network and Distributed System Security Symp (NDSS). San Diego, CA: Internet Society, 2019 [2022-04-20]. https://dx.doi.org/10.14722/ndss.2019.23066

[134] Xu Shengmin, Ning Jianting, Ma Jinhua, et al. K-time modifiable and epoch-based redactable blockchain [J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4507-4520

[135] Ma Jinhua, Xu Shengmin, Ning Jianting, et al. Redactable blockchain in decentralized setting [J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1227-1242

[136] Kuperberg M. Towards enabling deletion in append-only blockchains to support data growth management and GDPR Compliance [C] //Proc of the 2020 IEEE Int Conf on Blockchain (Blockchain). Piscataway, NJ: IEEE, 2020: 393-400

[137] Farshid S, Reitz A, Roßbach P. Design of a forgetting blockchain: A possible way to accomplish GDPR compatibility [C] //Proc of the 52nd Hawaii Int Conf on System Sciences. Hawaii: HICSS, 2019: 7087-7095

[138] Rantos K, Drosatos G, Kritsas A, et al. A blockchain-based platform for consent management of personal data processing in the IoT ecosystem [J]. Security and Communication Networks, 2019: No.1431578

[139] Makhdoom I, Zhou Lan, Abolhasan M, et al. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities [J]. Computers & Security, 2020, 88: No.101653

[140] Allegue S, Rhahla M, Abdellatif T. Toward GDPR compliance in IoT systems [C] //Proc of the Int Conf on Service-Oriented Computing. Berlin: Springer, 2019: 130-141

[141] Zheng Xiaochen, Mukkamala R R, Vatrapu R, et al. Blockchain-based personal health data sharing system using cloud storage [C] //Proc of the 2018 IEEE 20th Int Conf on e-Health Networking, Applications and Services (Healthcom). Piscataway, NJ: IEEE, 2018: 1-6

[142] Rhahla M, Abdellatif T, Attia R, et al. A GDPR controller for IoT systems: Application to e-health [C] //Proc of the 2019 IEEE 28th Int Conf on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). Piscataway, NJ: IEEE, 2019: 170-173

[143] Pournaghi S M, Bayat M, Farjami Y. MedSBA: A novel and secure scheme to share medical data based on blockchain technology and attribute-based encryption [J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11 (11): 4613-4641

[144] Muchagata J, Ferreira A. Translating GDPR into the mHealth Practice [C] //Proc of the 2018 Int Carnahan Conf on Security Technology (ICCST). Piscataway, NJ: IEEE, 2018: 1-5

[145] Fatehi F, Hassandoust F, Ko R K L, et al. General data protection regulation (GDPR) in healthcare: Hot topics and research fronts [M] //Digital Personalized Health and Medicine. Amsterdam: IOS Press, 2020: 1118-1122

[146] Barati M, Aujla G S, Llanos J T, et al. Privacy-Aware cloud auditing for GDPR compliance verification in online healthcare [J]. IEEE Transactions on Industrial Informatics, 2021, 18(7): 4808-4819

[147] Koutli M, Theologou N, Tryferidis A, et al. Secure IoT e-Health applications using VICINITY framework and GDPR guidelines [C] //Proc of the 15th Int Conf on Distributed Computing in Sensor Systems (DCOSS). Piscataway, NJ: IEEE, 2019: 263-270

[148] Bienzeisler J, Fischer H, Thiemann V S, et al. Human-Induced errors in networked healthcare research: Risk management under the GDPR [J]. Studies in Health Technology and Informatics, 2020, 270: 1128-1132

[149] Vojković G, Milenković M. GDPR in access control and time and attendance systems using biometric data [C] //Proc of the 41st Int Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). Piscataway, NJ: IEEE, 2018: 1138-1142

[150] Nautsch A, Jasserand C, Kindt E, et al. The GDPR & speech data: Reflections of legal and technology communities, first steps towards a common understanding [J]. arXiv preprint, arXiv:1907.03458, 2019

[151] Furey E, Blue J. Can I trust her? Intelligent personal assistants and GDPR [C] //Proc of the 2019 Int Symp on Networks, Computers and Communications (ISNCC). Piscataway, NJ: IEEE, 2019: 1-6

[152] Bisztray T, Gruschka N, Bourlai T, et al. Emerging biometric modalities and their use: Loopholes in the terminology of the GDPR and resulting privacy risks [C] // Proc of the 2021 Int Conf of the Biometrics Special Interest Group (BIOSIG). Piscataway, NJ: IEEE, 2021: 1-5

[153] Bourgeois J, Kortuem G, Kawsar F. Trusted and GDPR-compliant research with the Internet of things [C] //Proc of the 8th Int Conf on the Internet of Things. New York: ACM, 2018: 1-8

[154] Duncan A, Joyner D A. With or without EU: Navigating GDPR constraints in human subjects research in an education environment [C] //Proc of the 8th ACM Conf on Learning @ Scale. New York: ACM, 2021: 343-346

[155] Aberkane A J. Automated GDPR-compliance in requirements engineering [C/OL] //Proc of the 33rd Int Conf on Advanced Information Systems Engineering (CAiSE 2021). [2022-05-10]. http://ceur-ws.org/Vol-2906/paper3.pdf

[156] Westerlund M, Jaatun M G. Tackling the cloud forensic problem while keeping your eye on the GDPR [C] //Proc of the 2019 IEEE Int Conf on Cloud Computing Technology and Science (CloudCom). Piscataway, NJ: IEEE, 2019: 418-423

[157] Barati M, Rana O, Theodorakopoulos G, et al. Privacy-aware cloud ecosystems and GDPR compliance [C] //Proc of the 7th Int Conf on Future Internet of Things and Cloud (FiCloud). Piscataway, NJ: IEEE, 2019: 117-124

[158] Zhang Xueying, Yang Shuaifeng, Wang Chonghua, et al. Research on industrial Internet data security classification and grading protection framework [J]. Information Technology and Network Security, 2021, 40(1): 2-9 (in Chinese)
(张雪莹, 杨帅锋, 王冲华, 等. 工业互联网数据安全分类分级防护框架研究[J]. 信息技术与网络安全, 2021, 40(1): 2-9)

[159] National Information Security Standardization Technical Committee. Notice on soliciting opinions on the national standard "Information security technology-Requirements for classification and grading of network data" (Draft for comment)[EB/OL]. (2022-09-14) [2022-09-15]. https://www.tc260.org.cn/front/bzzqyjDetail.html?id=20220914180530&norm_id=20211108000024&recode_id=48416 (in Chinese)
(全国信息安全标准化技术委员会. 关于国家标准《信息安全技术 网络数据分类分级要求》征求意见稿征求意见的通知[EB/OL]. (2022-09-14) [2022-09-15]. https://www.tc260.org.cn/front/bzzqyjDetail.html?id=20220914180530&norm_id=20211108000024&recode_id=48416)

[160] Chen Xingyue. "Data Security Law of the People's Republic of China (Draft)" is open for comments: Data classification is officially entered into the law [J]. Informatization in China, 2020, (7): 9-10, 8 (in Chinese)
(陈兴跃.《中华人民共和国数据安全法(草案)》公开征求意见:数据分级分类正式入法[J]. 中国信息化, 2020, (7): 9-10, 8)

[161] Zhang Fen. Classification management and security protection of data in the era of big data [J]. Computer Products and Circulation, 2019, (1): 129-129 (in Chinese)
(张芬. 大数据时代数据的分类分级管理及安全防护[J].计算机产品与流通, 2019, (1): 129-129)

[162] Zhang Feng, Yu Le, Ma Yusheng, et al. Research and practice of data security classification and grading [J]. Information and Communications Technology and Policy, 2021, 47(8): 45-50 (in Chinese)
(张峰, 于乐, 马禹昇, 等. 数据安全分类分级研究与实践[J].信息通信技术与政策, 2021, 47(8): 45-50)

[163] Jin Tao. Data security grading [J]. Journal of Information Security Research, 2021, 7(10): 969-972 (in Chinese)
(金涛. 数据安全分级划分[J]. 信息安全研究, 2021, 7(10): 969-972)

[164] National Information Security Standardization Technical Committee. Notice on soliciting opinions on the national standard "Information security technology-Guideline for identification of critical data" (Draft for comment) [EB/OL]. (2022-01-13) [2022-08-20]. https://www.tc260.org.cn/front/bzzqyjDetail.html?id=20220113195354&norm_id=20201104200036&recode_id=45625 (in Chiese)
(全国信息安全标准化技术委员会. 关于国家标准《信息安全技术 重要数据识别指南》征求意见稿征求意见的通知[EB/OL]. (2022-01-13) [2022-08-20]. https://www.tc260.org.cn/front/bzzqyjDetail.html?id=20220113195354&norm_id=20201104200036&recode_id=45625)

Zhao Jingxin, born in 1998. Master candidate. Her main research interests include network and information system security.
赵景欣,1998 年生.硕士研究生.主要研究方向为网络和信息系统安全.

Yue Xinghui, born in 1999. Master candidate. His main research interests include network and information system security.
岳星辉,1999 年生.硕士研究生.主要研究方向为网络和信息系统安全.

Feng Chongpeng, born in 1998. Master candidate. His main research interests include network and information system security.
冯崇朋,1998 年生.硕士研究生.主要研究方向为网络与信息系统安全.

Zhang Jing, born in 1999. Master candidate. Her main research interests include network and information system security.
张　静,1999 年生.硕士研究生.主要研究方向为网络与信息系统安全.

Li Yin, born in 1984. PhD candidate. His main research interests include network and information system security.
李　印,1984 年生.博士研究生.主要研究方向为网络和信息系统安全.

Wang Na, born in 1998. Master candidate. Her main research interests include network and information system security.
王　娜,1998 年生.硕士研究生.主要研究方向为网络与信息系统安全.

Ren Jiadong, born in 1967. Professor. Senior member of CCF, member of IEEE and ACM. His main research interests include data mining, temporal data modeling, and software security.
任家东,1967 年生.教授.CCF 高级会员,IEEE 和 ACM 会员.主要研究方向为数据挖掘、时态数据建模和软件安全.

Zhang Haoxing, born in 1987. Engineer. His main research interests include data security, information security and Internet security regulatory governance.
张昊星,1987 年生.工程师.主要研究方向为数据安全、信息安全以及互联网安全监管治理.

Wu Gaofei, born in 1987, PhD, lecturer, master supervisor. His main research interest is cryptography.
伍高飞,1987 年生.博士,讲师,硕士生导师.主要研究方向为密码学.

Zhu Xiaoyan, born in 1979. Professor. Her research interests include network and information security.
朱笑岩,1979 年生.教授.主要研究方向为网络与信息安全.

Zhang Yuqing, born in 1966. PhD, professor, PhD supervisor. His main research interest is information security.
张玉清,1966 年生.博士,教授,博士生导师.主要研究方向为信息安全.