基因数据隐私问题及相关保护技术进展研究*

宋思嘉

(上海交通大学医学院附属新华医院
心血管内科 上海 200092)

单晨璐

(杭州诺崑信息科技有限公司 杭州 310023)

王爽

(1 杭州诺崑信息科技有限公司 杭州 310023
2 同济大学上海普陀区人民医院 上海 200040)

陈如梵

(杭州诺崑信息科技有限公司 杭州 310023)

张涛

(公安部第三研究所
上海 201204)

郑灏

(杭州诺崑信息科技有限公司
杭州 310023)

韩雅琴

(上海交通大学医学院附属新华
医院心血管内科 上海 200092)

〔摘要〕 分析基因组数据存在的隐私安全问题, 详细阐述基因数据隐私保护相关技术、研究进展、技术不足之处等方面, 为相关研究提供参考。

〔关键词〕 基因组数据隐私; 基因数据隐私计算; 密码学技术

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.06.001

Study on the Privacy Issues of Genetic Data and the Progress of Related Protection Technologies SONG Sijia, Department of Cardiovascular Medicine, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China; SHAN Chenlu, Hangzhou Nuowei Information Technology Co. Ltd., Hangzhou 310023, China; WANG Shuang, 1 Hangzhou Nuowei Information Technology Co. Ltd., Hangzhou 310023, 2 People's Hospital of Shanghai Putuo District, Shanghai 200040, China; CHEN Rufan, Hangzhou Nuowei Information Technology Co. Ltd., Hangzhou 310023, China; ZHANG Tao, The Third Research Institute of Ministry of Public Security, Shanghai 201204, China; ZHENG Hao, Hangzhou Nuowei Information Technology Co. Ltd., Hangzhou 310023, China; HAN Yaqin, Department of Cardiovascular Medicine, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

〔修回日期〕 2021-06-09

〔作者简介〕 宋思嘉, 主管护师; 通讯作者: 韩雅琴, 主管护师。

〔基金项目〕 信息网络安全公安部重点实验室开放课题项目 (项目编号: C20605)。

[Abstract] The paper analyzes the privacy security problems existing in genomic data, and elaborates the related technologies, study progress, technical shortcomings and other aspects of genetic data privacy protection in detail, so as to provide references for relevant study.

[Keywords] genomic data privacy; genetic data privacy computing; cryptography technology

1 引言

1.1 研究背景

1990年人类基因组计划启动,基因研究和探索进入新的发展阶段,基因组信息开始广泛被应用于医疗健康领域研究中。与此同时世界3大基因数据库即美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI),欧洲生物信息研究所(European Bioinformatics Institute, EBI)和日本DNA数据库(DNA Data Bank of Japan, DDBJ)开始建设,发展至今其所存储的基因数据已达到PB级,涵盖多种常用生物数据类型。近年来随着二代测序技术发展,基因组数据获取已解决复杂、耗时等问题^[1-3],为相关研究工作提供便利。我国在基因数据库搭建方面获得突破性进展,已建立综合性的国家基因库(China National GeneBank, CNGB),生命与健康大数据中心(BIG Data Center, BIGD),国家组学数据百科全书(The National Omics Data Encyclopedia, NODE),蛋白质组学整合资源库(Integrated Proteome Resources, iProX)等不同类型基因组数据库。但目前我国基因库相互独立,缺乏与样本的关联和数据之间的相互作用,尚未实现共有、共为、共享^[4]。

1.2 问题分析

基因组数据共享利用引发相关数据隐私安全问题。近年来基因测序技术发展使得基因检测更加便捷,催生建立一批基因检测公司和第3方机构,其检测过程涉及大量基因组数据且通常被储存在公开基因组数据库中,尽管经过匿名化处理,但数据隐私安全保护较难实现。基因组数据一旦泄露将会对数据主体造成无法预计的伤害。因此如何在实现数据存储和共享的同时避免个体基因数据隐私泄露成

为新的挑战。

1.3 解决方法

目前较广泛使用的方法之一是在基因组数据共享过程中实施隐私保护政策。然而由于各个国家和地区有关数据共享法律法规不同,该方法具有较大局限性^[5]。例如管理机构审查研究者访问基因组数据集请求需要一定时间,可能影响研究效率。同时由于无法预见对基因组数据的攻击,隐私保护策略可能过于笼统而无法覆盖所有类型攻击,形成漏洞,可能导致新的隐私安全问题。另一种更为有效的解决方法是将密码学应用到基因组数据共享过程中。实践证明在共享基因组数据时使用不同加密方法可以有效避免隐私信息泄露。因此利用密码学技术和加密方法保证基因组数据安全将是未来重要研究方向。本文从基因组数据隐私保护角度出发梳理、分析相关研究,重点关注基因组数据在安全性和隐私性方面的差异;针对当前不同环境下共享和计算基因组数据存在的隐私安全问题,解决基因组数据隐私安全问题的主要密码学方法及其进展,现有加密原语工具及其限制,针对密码学隐私保护方法相关研究进行探讨。

2 基因组数据隐私安全问题

2.1 共享方面

2.1.1 概述 全基因组关联分析(Genome Wide Association Study, GWAS)是将特定的遗传变异与特定疾病相关联的方法,其涉及从基因组数据库中寻找可用于预测疾病存在的遗传标记。这些分析广泛涉及基因组数据共享。随着研究发展基因组数据共享过程中可能会造成个体隐私泄露的问题,逐渐获得广泛关注^[6]。

2.1.2 相关研究 2000年Malin B和Sweeney L^[7]

根据可用健康记录和疾病背景知识识别 DNA 序列。2004 年 Gottlieb S^[8] 提出通过 75 个独立的单核苷酸多态性位点 (Single Nucleotide Polymorphisms, SNPs) 来识别对应数据主体。2008 年 Homer N、Szelinger S 和 Redman M 等^[9] 建立可从混合 DNA 池中识别已知基因型个体存在的理论框架,证明了在复杂混合的基因组池中识别特定个体的可能性,证实了在基因组数据共享过程中个体可以被识别从而导致隐私泄露。2009 年 Wang R、Li Y F 和 Wang X F 等^[10] 提出可以从 GWAS 相关研究文献中常规发布的相对较小的统计信息中心识别个体,同时提出存在两种隐私攻击,一是基于不同 SNPs 之间的关联对 Homer N^[9] 隐私攻击方法进行扩展,二是隐私攻击能够导致数百名参与者的 SNPs 信息完全泄露。在统计数据精度较低且部分数据丢失的情况下上述攻击方法仍被证实可能性,表明基因组数据隐私泄露的威胁被低估。2015 年 Shringarpure S 和 Bustamante C^[11] 提出可以通过查询有限数量的基因组数据库识别参与者。目前如何在实现基因组数据共享的同时加强个体隐私安全防护成为主要关注点,而如何在密码学基础上实现基因组数据的隐私保护成为主要研究方向。

2.2 存储或计算过程中

基因组数据在存储或计算中存在被泄露风险。一方面,由于预算限制部分研究人员选择使用公共云服务存储数据或执行计算任务,此类云服务具有公共特性而难以保证数据安全;另一方面,如将科研项目基因组数据交由第 3 方机构进行存储、处理或者计算同样存在数据泄露风险。因此应当使用更先进、完善的隐私计算技术。

2.3 查询或输出

基因组数据查询或输出过程中涉及隐私安全问题。即使在基因组数据共享、存储和计算中使用隐私保护技术,研究者在利用医疗健康数据、蛋白质结构以及基因组信息的公共数据库时会不可避免地泄漏研究目的及数据特征^[5]。Shringarpure S 和 Bustamante C^[11], Raisaro J L、Choi G 和 Pradervand

S^[12] 的研究提出在基因组数据查询或输出时隐私保护具有必要性。

3 隐私保护技术

3.1 基因信标

3.1.1 概述 经过脱敏或是匿名化处理的基因数据在共享或是公开披露时仍有可能遭到重识别攻击而造成隐私泄露。基因组数据共享信标被广泛应用于数据共享中从而规避一定重识别风险。通过基因信标可以查询某个数据集中是否存在特定对象。其目的是为数据共享提供安全、易于实现的标准化接口,即任何科研人员都能够在不提交申请的情况下使用访问受限的高敏感数据,为基因数据共享提供便利。

3.1.2 有效保护隐私安全 因为使用基因信标共享的数据和信息较有限,可较好地避免基因数据泄露风险,保证数据提供者隐私安全。首先,通过基因信标可以查询某个基因组位置上是否具有特定核苷酸,基因信标只为数据访问者提供较为有限的信息。数据访问者通过单次查询仅能获得“是”或“否”的反馈,即在特定的基因组位置上是否存在其询问的核苷酸,而无法获得更多信息。即使得到肯定反馈,由于查询结果可能同时关联到多个个体,无法锁定具体个体,可以避免重识别攻击。

3.1.3 相应问题及未来发展方向 有研究显示尽管基因数据信标所提供信息量较有限,但仍存在隐私泄露风险,即通过重复查询信标可以确定某个个体的基因组数据是否存在特定数据集中^[13]。另有研究提出即使所共享数据仅包括某个信标是否存在特定等位基因相关信息,其所对应的个体也将面临被重新识别的风险。例如当询问“是否在特定基因组位置(例如染色体 1 上的 11 272 位)具有特定核苷酸”时,无论信标反馈“是”或“否”都有可能泄露数据个体基因信息,导致个体身份被识别、隐私被暴露^[5]。上述研究说明过去所依赖的隐私保护手段不仅不能保证基因数据隐私安全反而会产生新漏洞,亟需更有效的隐私保护技术以推动基因数据共享^[10]。

3.2 同态加密

3.2.1 概述 同态加密 (Homomorphic Encryption, HE) 是一种基于计算复杂性的密码学技术。经过同态加密处理的数据可以直接进行计算而无需解密, 且其结果等同于直接使用明文数据计算。这一技术适用于云计算应用场景, 数据持有方可以将数据加密后传输至公有云服务器并进行加密计算, 生成加密计算结果。解密方只能获得最后结果而无法获得明文数据, 可以保证数据隐私安全, 同态加密效率较高且通信成本较低, 可用于基因领域数据隐私保护。

3.2.2 在半诚实云服务器上使用附加 HE (Paillier 加密)^[14] 有研究者根据二进制表示对云服务器中储存的基因组序列进行编码, 同时通过 Paillier 加密进行编码以实现基因数据共享过程中的隐私保护目的^[15]。另有研究者将 Paillier 加密使用在联邦^[16]和中心化环境中^[17]。上述研究表明使用该技术加密基因数据可以有效提高数据共享时的隐私安全性。Ghasemi R、Alaziz M M 和 Mohammed N 等^[17] 提出利用该技术建立的模型能够在数据持有方使用云服务时保护基因组数据隐私。Ayday E、Raisaro J L 和 McLaren P J 等^[18] 提出一种基于 DGK 密码系统的方法, 用于实现在联邦或中心化环境下存储和计算基因组数据的隐私保护, 应用于遗传风险计算过程中, 实验结果证明其对于隐私保护的有效性。

3.2.3 用于基因组研究的统计算法安全版本 Lauter K、López - Alt A 和 Naehrig M^[19] 提出算法安全版本。值得注意的是其使用的是全同态加密, 即通过这些算法可在任何加密数据上进行任意次数计算。此外有研究使用类同态加密技术保护基因数据隐私安全, 如 Wang S、Zhang Y C 和 Dai W R 等^[20] 提出应用于基因数据的安全精确逻辑回归 (Secure Exact Logistic Regression)。2015 年 Lu W、Yamada Y 和 Sakuma J^[21] 提出利用 HE 进行 GWAS 的隐私保护且将其使用的包装技术与其他研究进行比较。2016 年有研究者使用 Burrows - Wheeler 变换寻找基因组数据集的目标序列。2018 年 Kim M、Song Y 和 Wang S 等^[22] 提出用于真实数据计算优化的新型同态

加密模型, 该模型可用于大规模数据集。翟文斌、王鹏和戴彩艳等^[23] 将同态加密技术应用于中医电子病历隐私保护, 提高中医电子病历信息安全性。

3.2.4 相应问题及未来发展方向 由于同态加密尚处于早期发展阶段, 对于大规模通用基因组数据还无法实现较高效率。例如乘法操作或者 Bootstrap 过程可能需要一定时间, 较难应用于训练机器学习等复杂函数。研究者在使用机器学习方法训练基因组数据时, 数据敏感性较强, 需在加密数据上训练机器学习模型, 因而对同态加密效率要求较高, 基因组数据需要应用不同包装技术从而减少计算时间。

3.3 混淆电路

3.3.1 原理及应用 混淆电路 (Garbled Circuits, GC) 利用密码学原理, 允许在多方之间计算任何函数而彼此隐藏输入^[24]。例如需要在信息保密前提下比较各方年龄, 可以通过 GC 协议得到结果。GC 的安全保证依靠各方通过逻辑电路平等参与, 同时保证输入的保密性。GC 协议主要应用于序列相似性问题, 例如其中一方具有包含基因组序列的数据集而另一方具有敏感的查询序列, 研究者可基于相似性指标查找特定查询序列的相似序列。

3.3.2 相关研究 2008 年 Jha S、Kruger L 和 Shmatikov V^[25] 设计并展示 3 种协议, 可以通过 GC 协议复刻原始编辑距离算法, 但计算时间较长。Wang X S、Huang Y 和 Zhao Y 等^[26] 在 2015 年将该问题定义为“相似病人查询”, 使用公共参考基因组序列对基因组数据集进行预计算, 然后预估查询字符串和数据集之间的编辑距离, 该方法可用于较大数据集原始编辑距离的预估^[26]。公共参考基因组选择可能泄露相关数据分布信息, 同时由于编辑距离预估计算依赖于公共参考基因组, 公共参考基因组选择可能影响结果准确性。近年来较多研究尝试在安全前提下预估编辑距离。例如 Asharov G、Hallevi S 和 Lindell Y 等^[27] 提出一种可用于有效安全地计算“相似病人查询”问题近似值的新方法。

3.3.3 相应问题及未来发展方向 在查询和计算基因组数据时实现隐私保护是应用 GC 的主要目的。虽然目前 GC 技术相对成熟但仍需要在大规模基因组

数据计算中测试。例如对于实现安全编辑距离的预估和计算均考虑较小长度序列,然而在现实中则要求较长基因序列,大规模电路计算对网络的需求尚有待试验和考证。“秘密共享方案”(Secret Sharing Schemes)是另一种安全的多方计算协议,也是安全计算的重要工具,其可在安全协议中用作构建盒,例如用于多方计算的通用协议、拜占庭协议等^[28]。

3.4 置信计算环境

3.4.1 概述 主要是指从计算芯片上隔离出来的一块安全计算区域。在置信计算环境上运行数据可以保证隐私安全(如英特尔 SGX 软件防护扩展),如果底层操作系统或虚拟平台被挟持,SGX 可以防止第 3 方对数据的攻击,因此可以保证在第 3 方进行高性能的安全计算并远程验证第 3 方安全计算环境。但是 SGX 对于内存和攻击有一定要求,仅限于内存 128MB 以及特定算法的旁路攻击^[29-31]。

3.4.2 相关研究 2012 年 Canim M、Kantarcioglu M 和 Malin B^[32]首次将置信计算环境应用到基因组数据隐私保护,构建通过加密硬件集成来存储和处理敏感生物学数据的框架,从而保护基因组数据隐私。随着英特尔 SGX 软件防护扩展的到来,基因组数据安全保护技术受到广泛关注。2016 年 Chen F、Dow M 和 Ding S 等^[33]使用 SGX 设计 PREMIX 框架并发明一种安全的基因亲缘关系分析方法。同年 Chen F、Wang S 和 Jiang X Q 等^[34]针对分布在不同大陆的罕见疾病遗传数据,利用 SGX 设计分析的系统,提出一种可实现高效、安全的基因数据存储和计算的框架^[35]。

3.4.3 相应问题及未来发展方向 置信计算环境例如英特尔 SGX 或 AMD 内存加密仍属新技术。尽管其可以解决在数据共享和存储中的隐私安全问题,但仍需要时间检验。由于置信计算环境可以成为同态加密的有效代替,使用第 3 方可信组件或依赖于单个密钥的安全风险仍需要进一步测试。

3.5 联邦学习

3.5.1 概述 联邦学习(Federated Learning)使用独立分布的本地数据进行模型训练,即数据并非收集在单个中央服务器上而是由不同方提供。

3.5.2 相关研究 2013 年 Wang S、Jiang X Q 和 Wu Y 等^[36]发表全球第 1 篇关于医疗在线联邦学习的论文,开发可用于分布式隐私保护在线学习的 EXPLORER 模型。2016 年谷歌研究院提出联邦学习概念^[37-39],旨在分布于不同方的数据集上实现建模且同时保护数据隐私^[40]。Du W 和 Zhan Z^[41]以及 Vaidya J 和 Clifton C^[42]提出用于垂直分割数据的安全多方决策树算法。Vaidya J 和 Clifton C^[43]提出安全关联规则挖掘算法可用于垂直分割数据、K-means 算法^[44]以及朴素贝叶斯分类器^[45]。Yu H、Jiang X 和 Vaidya J 等^[46]提出用于垂直分割数据和水平分割数据^[47]的安全支持向量机算法。Du W、Han Y S 和 Chen S^[48]提出安全协议可用于多方线性回归和分类。2018 年 Lee J、Sun J 和 Wang F 等^[49]将联邦学习应用于医疗健康领域,提出一种算法可以帮助在机构中有效搜索相似患者,从而在联合数据分析过程中保护隐私。同年 Brisimi T S、Chen R 和 Mela T 等^[50]使用分布式算法解决二元监督分类问题,以预测因心脏事件而产生的住院治疗数,同时建立一个新框架,可以通过多方协作来预测模型而无需交换原始数据。2020 年冯涛、焦滢和方君丽等^[51]提出基于联邦学习的可用于医疗健康的安全模型。

3.5.3 相应问题及未来发展方向 联邦学习能够用于基因组数据隐私保护,可以利用大规模医疗数据构建无偏决策模型,反映个人生理状况,对罕见疾病具有敏感性同时保护个体隐私。但实施联邦学习仍需要严格的技术考量,以确保算法在不影响安全性和患者隐私前提下进行最佳处理。同时联邦学习尚无法解决部分医学数据固有问题。成功的模型训练依赖于很多因素,如数据质量、偏差和标准化等因素。无论对于联邦学习或者非联邦学习来说,这些问题都有待解决^[52]。

3.6 差分隐私

3.6.1 概述 差分隐私(Differential Privacy, DP)是一种隐私保护方式,主要用于在发布有关数据库统计信息时实现隐私保护。涉及差分隐私和基因组数据的大部分研究工作均围绕 GWAS。

3.6.2 相关研究 2013 年 Jiang X、Ji Z 和 Wang S

等^[53]提出一种新机制以实现满足 ϵ -差分隐私的数据发布,同时通过成分分析提高数据可用性。同年 Johnson A 和 Shmatikov V^[54]提出差分隐私方法以回答 GWAS 统计查询,包括 DNA 中特定位置与一种特定疾病相关;DNA 中两个位置之间的关系;与特定疾病相关的位置数。2014 年 Fei Y、Rybar M 和 Uhler C 等^[55]对该问题进行扩展,运用逻辑回归将差分隐私应用于检测多个位置之间的关联,同时将这一差分隐私过程应用于 GWAS 数据研究。2015 年 Tramèr F、Huang Z 和 Hubaux J P 等^[56]提出一种新的差分隐私机制,面对不同不利因素均可以实现隐私保护。2016 年 Simmons S 和 Berger B^[57]通过将其建模为优化问题,完善上述两个在 DNA 中找到重要位置的问题。2016 年 Simmons S、Sahinalp C 和 Berger B^[58]提出针对 EigenStrat 和基于线性混合模型、存在人口分层的 GWAS 差分技术。另外还有很多对于 GWAS 的使用差分隐私手段进行数据分发或共享的研究^[59-61]。2017 年 Aziz M、Ghasemi R 和 Waliullah M 等^[62]针对基因信标^[11-12]的隐私攻击提

出一种差分隐私解决方案,通过仔细检查原始攻击参数,针对不同情况进行分析并提出一种简单的差分隐私“随机响应”算法。2018 年任倬辉^[63]设计一个基于差分隐私的分析系统,可在保证数据安全的前提下进行医疗数据分析。2020 年陈磊磊、陈兰香和穆怡^[64]提出一种基于差分隐私的新型双向隐私保护方法,为医疗诊断隐私安全提供保障。

3.6.3 相应问题及未来发展方向 差分隐私在人类基因领域隐私保护中得到广泛应用。然而由于误差或噪声校准问题以及对基因组研究的高精度要求,该方法需要在不同研究问题中进行分析。差分隐私技术为数据输出提供隐私保证。例如对差分私有数据执行计算则可以证明输出数据遵循相同差分隐私保证。此定义有助于避免在某些不需要安全计算情况下的安全计算技术(如 HE 或 SGX)开销。此外在机器模型隐私保护领域的广泛研究有利于差分隐私在基因组数据安全领域的应用。

3.7 隐私保护技术总结(表 1)

表 1 隐私保护技术总结

隐私保护技术	基本原理	适用性	特色	不足
基因信标	查询某个数据集中是否存在特定对象,为数据共享提供一个安全、易于实现的标准化接口	被广泛应用于基因数据共享中	为基因组数据共享提供便利	虽然基因数据信标仅提供有限信息量,仍存在隐私泄露风险
同态加密	经过同态加密处理的数据可以直接进行计算而无需解密,且其结果等同于直接使用明文数据计算	适用于基因数据存储和公共云服务中的隐私保护	加密计算,通信成本低	数据量过大时效率偏低
混淆电路	密码学原理,通过 GC 协议计算	主要应用于序列相似性问题	在相似序列查询时实现隐私保护	还未在大规模基因组数据计算中测试
置信计算环境	从计算芯片上隔离出来的一块安全计算区域	英特尔的 SGX	可以在第 3 方进行高性能的安全计算	技术较新,仍需时间检验
联邦学习	使用来自于多方的独立分布的本地数据训练模型	可适用于大规模医疗数据	无需将数据收集在单个中央服务器上,保护多方隐私	无法解决医学数据的固有问题
差分隐私	在发布有关数据库的统计信息时实现统计数据库中记录的隐私保护	主要应用于 GWAS	可实现隐私保护和信息利用的平衡	由于误差或噪声校准问题以及对基因组研究的高精度要求,该方法需要在不同的研究问题中进行分析

4 结语

基因与生命体紧密相连,其储存个体全部信息。随着更低成本、更精确序列方法和更复杂攻

击识别时代的到来,对基因组数据安全性和隐私保护的需求增强,需要对基因信息数据安全性和隐私保护问题进一步探索,以保证基因组数据存储、共享、分析过程隐私安全,促进基因领域整体发展。

参考文献

- Margulies M, Egholm M, Altman W E, et al. Genome Sequencing in Microfabricated High - density Picolitre Reactors [J]. *Nature*, 2005, 437 (7057): 376 - 380.
- Becich M, Santana - Santos L, Gullapalli R, et al. Next Generation Sequencing in Clinical Medicine: challenges and lessons for pathology and biomedical informatics [J]. *J Pathol Inform*, 2012 (3): 40.
- Shapiro E, Biezuner T, Linnarsson S. Single - cell Sequencing - based Technologies Will Revolutionize Whole - organism Science [J]. *Nature Reviews Genetics*, 2013 (14): 618 - 630.
- 王博, 刘芳, 张二春, 等. 国家基因库: 共有、共为、共享 [J]. *遗传*, 2019, 41 (8): 761 - 772.
- Aziz M M A, Sadat M N, Alhadidi D, et al. Privacy - preserving Techniques of Genomic Data - a survey [J]. *Brief Bioinform*, 2019, 20 (3): 887 - 895.
- Akgün M, Bayrak A O, Özer B, et al. Privacy Preserving Processing of Genomic Data: a survey [J]. *Journal of Biomedical Informatics*, 2015, 56 (C): 103 - 111.
- Malin B, Sweeney L. Determining the Identifiability of DNA Database Entries [C]. New York: Proc AMIA Symp, 2000: 537 - 541.
- Gottlieb S. US Employer Agrees to Stop Genetic Testing [J]. *BMJ*, 2001 (322): 449.
- Homer N, Szelling S, Redman M, et al. Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High - density SNP Genotyping Microarrays [J]. *PLOS Genetics*, 2008, 4 (8): e1000167.
- Wang R, Li Y F, Wang X F, et al. Learning Your Identity and Disease from Research Papers; Information Leaks in Genome Wide Association Study [C]. Chicago: Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS09), 2009: 534 - 544.
- Shringarpure S, Bustamante C. Privacy Risks from Genomic Data - sharing Beacons [J]. *American Journal of Human Genetics*, 2015, 97 (5): 631 - 646.
- Raisaro J L, Choi G, Pradervand S, et al. Protecting Privacy and Security of Genomic Data in i2b2 with Homomorphic Encryption and Differential Privacy [J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2018, 15 (5): 1413 - 1426.
- Thenen N V, Ayday E, Cicek A E. Re - identification of Individuals in Genomic Data - sharing Beacons via Allele Inference [J]. *Bioinformatics*, 2019, 35 (3): 365 - 371.
- Paillier P. Public - key Cryptosystems Based on Composite Degree Residuosity Classes [C]. Czech: Advances in Cryptology - EUROCRYPT99, 1999: 223 - 238.
- Kantarcioglu M, Jiang W, Liu Y, et al. A Cryptographic Approach to Securely Share and Query Genomic Sequences [J]. *IEEE Transactions on Information Technology in Biomedicine*, 2008, 12 (5): 606 - 617.
- Aziz M, Hasan M Z, Mohammed N, et al. Secure and Efficient Multiparty Computation on Genomic Data [C]. Montreal: International Database Engineering & Applications Symposium. ACM, 2016.
- Ghasemi R, Alaziz M M, Mohammed N, et al. Private and Efficient Query Processing on Outsourced Genomic Databases [J]. *IEEE J Biomed Health Inform*, 2017, 21 (5): 1466 - 1472.
- Ayday E, Raisaro J L, McLaren P J, et al. Privacy - preserving Computation of Disease Risk by Using Genomic, Clinical, and Environmental Data [EB/OL]. [2013 - 08 - 12]. <https://dl.acm.org/doi/10.5555/2696523.2696524>.
- Lauter K, López - Alt A, Naehrig M. Private Computation on Encrypted Genomic Data [C]. Brazil: Third International Conference on Cryptology and Information Security, 2014: 3 - 27.
- Wang S, Zhang Y C, Dai W R, et al. HEALER: homomorphic computation of ExAct Logistic rEgRession for secure rare disease variants analysis in GWAS [J]. *Bioinformatics*, 2016, 32 (2): 211 - 218.
- Lu W, Yamada Y, Sakuma J. Efficient Secure Outsourcing of Genome - wide Association Studies [C]. San Jose: 2015 IEEE Security and Privacy Workshops, 2015: 3 - 6.
- Kim M, Song Y, Wang S, et al. Secure Logistic Regression Based on Homomorphic Encryption: design and evaluation [J]. *JMIR Medical Informatics*, 2018, 6 (2): e19.
- 翟文斌, 王鹏, 戴彩艳, 等. 基于同态加密的中医电子病历安全存储方法 [J]. *产业科技创新*, 2020, 2 (36): 37 - 39.
- Yao A C. Protocols for Secure Computations [C]. Washington: Proceedings of 23rd IEEE Symposium on Foundations of Computer Science, 1982: 160 - 164.
- Jha S, Kruger L, Shmatikov V. Towards Practical Privacy for Genomic Computation [C]. Oakland: 2008 IEEE Symposium on Security and Privacy (SP 2008), 2008: 216 - 230.
- Wang X S, Huang Y, Zhao Y, et al. Efficient Genome - wide Privacy - preserving Similar Patient Query based on Private Edit Distance [C]. Denver: ACM Sigsac Conference on Computer & Communications Security, 2015: 492 - 503.
- Asharov G, Halevi S, Lindell Y, et al. Privacy - preserving Search of Similar Patients in Genomic Data [J]. *Proceedings on Privacy Enhancing Technologies*, 2018 (4): 104 - 124.

- 28 Beimel A. Secret - sharing Schemes: a survey [C]. Qingdao: International Conference on Coding and Cryptology, 2011: 11 - 46.
- 29 Ming - Wei Shih, Sangho Lee, Taesoo Kim, et al. T - sgx: eradicating controlled - channel attacks against enclave programs [C]. San Diego: Proceedings of the 2017 Annual Network and Distributed System Security Symposium (NDSS), 2017.
- 30 Wang W, Chen G, Pan X, et al. Leaky Cauldron on the Dark Land: understanding memory side - channel hazards in SGX [C]. Dallas: Conf Comput Commun Secur, ACM, 2017: 2421 - 2434.
- 31 窦佐超, 汪诚弘, 邓杰仁, 等. 生物医疗大数据隐私与安全保护的应对策略与技术 [J]. 中华医学图书情报杂志, 2019, 28 (11): 11 - 15.
- 32 Canim M, Kantarcioglu M, Malin B. Secure Management of Biomedical Data with Cryptographic Hardware [J]. IEEE Trans Inf Technol Biomed, 2012, 16 (1): 166 - 175.
- 33 Chen F, Dow M, Ding S, et al. PREMIX: privacy - preserving estimation of individual admixture [J]. AMIA Annu Symp Proc, 2017 (10): 1747 - 1755.
- 34 Chen F, Wang S, Jiang X Q, et al. PRINCESS: privacy - protecting rare disease international network collaboration via encryption through software guard extensions [J]. Bioinformatics. 2017, 33 (6): 871 - 878.
- 35 Chen F, Wang C, Dai W, et al. PRESAGE: privacy - preserving genetic testing via software guard extension [J]. BMC Med Genomics, 2017 (10): 48.
- 36 Wang S, Jiang X Q, Wu Y, et al. Expectation Propagation Logistic Regression (EXPLORER): distributed privacy - preserving online model learning [J]. Journal of Biomedical Informatics, 2013, 46 (3): 480 - 496.
- 37 Konečný J, Brendan M H, Ramage D, et al. Federated Optimization: distributed machine learning for on - device intelligence [EB/OL]. [2019 - 07 - 23]. <https://www.cnblogs.com/lucifer1997/p/11229862.html>.
- 38 Konečný J, Brendan M H, Yu F X, et al. Federated Learning: strategies for improving communication efficiency [EB/OL]. [2018 - 08 - 13]. <https://dblp.uni-trier.de/rec/journals/corr/KonecnyMYRSB16.html>.
- 39 Memahan B, Moore E, Ramage D, et al. Communication - efficient Learning of Deep Networks from Decentralized Data [J]. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017 (54): 1273 - 1282.
- 40 Yang Q, Liu Y, Chen T, et al. Federated Machine Learning: concept and applications [J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10 (2): 1 - 19.
- 41 Du W, Zhan Z. Building Decision Tree Classifier on Private Data [C]. Darlington: Proceedings of the IEEE International Conference on Privacy, Security and Data Mining, 2002: 1 - 8.
- 42 Vaidya J, Clifton C. Privacy - preserving Decision Trees over Vertically Partitioned Data [C]. Storrs: 19th Annual IFIP WG 11. 3 Working Conference on Data and Applications Security, 2005: 139 - 152.
- 43 Vaidya J, Clifton C. Privacy Preserving Association Rule Mining in Vertically Partitioned Data [C]. Edmonton: Proceedings of the 8th ACM SIGKDD (Special Interest Group on Knowledge Discovery and Data Mining) International Conference on Knowledge Discovery and Data Mining, 2002: 639 - 644.
- 44 Vaidya J, Clifton C. Privacy - preserving K - means Clustering over Vertically Partitioned Data [C]. Washington: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003: 206 - 215.
- 45 Vaidya J, Clifton C. Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data [C]. Florida: Proceedings of the Fourth SIAM International Conference on Data Mining, 2004: 522 - 526.
- 46 Yu H, Jiang X, Vaidya J, et al. Privacy - preserving SVM Using Nonlinear Kernels on Horizontally Partitioned Data [C]. Dijon: Proceedings of the 2006 ACM Symposium on Applied Computing, 2006: 603 - 610.
- 47 Yu H, Vaidya J, Jiang X. Privacy - preserving SVM Classification on Vertically Partitioned Data [C]. Singapore: 10th Pacific - Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006), 2006: 647 - 656.
- 48 Du W, Han Y S, Chen S. Privacy - preserving Multivariate Statistical Analysis: linear regression and classification [C]. Orlando: Proceedings of the Fourth SIAM International Conference on Data Mining, 2004: 222 - 233.
- 49 Lee J, Sun J, Wang F, et al. Privacy - preserving Patient Similarity Learning in a Federated Environment: development and analysis [J]. JMIR Med Inform, 2018 (6): e20.
- 50 Brisimi T S, Chen R, Mela T, et al. Federated Learning of Predictive Models from Federated Electronic Health Records [J]. International Journal of Medical Informatics, 2018, 112 (4): 59 - 67.
- 51 冯涛, 焦滢, 方君丽, 等. 基于联盟区块链的医疗健康数据安全模型 [J]. 计算机科学, 2020, 47 (4): 305 - 311.
- 52 Rieke N, Hancox J, Li W, et al. The Future of Digital Health with Federated Learning [J]. NPJ Digit Med, 2020 (3): 119.

- 9 李贺, 余璐, 许一明, 等. 解释水平理论视角下的社交网络隐私悖论研究 [J]. 情报学报, 2018, 37 (1): 1-13.
- 10 Stern T, Salb D. Examining Online Social Network Use and Its Effect on the Use of Privacy Settings and Profile Disclosure [J]. Bulletin of Science, Technology & Society, 2015, 35 (1): 25-34.
- 11 王瑜超. 在线医疗社区用户健康隐私信息披露意愿的影响因素研究 [J]. 信息资源管理学报, 2018, 8 (1): 93-103.
- 12 Krasnova H, Spiekermann R, Koroleva R, et al. Online Social Networks: why we disclose [J]. Journal of Information Technology, 2010, 25 (2): 109-125.
- 13 Chen R. Living a Private Life in Public Social Networks: an exploration of member self-disclosure [J]. Decision Support Systems, 2013, 55 (3): 661-668.
- 14 Lutz C, Strathoff P. Privacy Concerns and Online Behavior - Not so Paradoxical after All? Viewing the Privacy Paradox Through Different Theoretical Lenses [EB/OL]. [2019-12-18]. https://www.researchgate.net/publication/259470061_Privacy_Concerns_and_Online_Behavior_-_Not_so_Paradoxical_After_All_Viewing_the_Privacy_Paradox_through_Different_Theoretical_Lenses.
- 15 Trope Y, Liberman N. Temporal Construal and Time-dependent Changes in Preference [J]. Journal of Personality and Social Psychology, 2000, 79 (6): 876-889.
- 16 Wang T, Dong T D, Chen C C. Intention to Disclose Personal Information via Mobile Applications: a privacy calculus perspective [J]. International Journal of Information Management, 2016, 36 (4): 531-542.
- 17 姜又琦. 在线医疗网站用户个人健康信息披露意愿影响因素研究 [D]. 武汉: 武汉大学, 2017.
- 18 Metcalfe J, Mischel W. A Hot/Cool - system Analysis of Delay of Gratification: dynamics of willpower [J]. Psychological Review, 1999, 106 (1): 3-19.
- 19 朱侯, 王可, 严芷君, 等. 基于隐私计算理论的 SNS 用户隐私悖论现象研究 [J]. 情报杂志, 2017, 36 (2): 134-139.
- 20 Dinev T, Xu H, Smith J H, et al. Information Privacy and Correlates: an empirical attempt to bridge and distinguish privacy-related concepts [J]. European Journal of Information Systems, 2013, 22 (3): 295-316.
- 21 Kehr F. Feeling and Thinking: on the role of intuitive processes in shaping decisions about privacy [D]. St. Gallen: University of St. Gallen, 2015.
- 22 Dienlin T, Trepte S. Is the Privacy Paradox a Relic of the Past? An In-depth Analysis of Privacy Attitudes and Privacy Behaviors [J]. European Journal of Social Psychology, 2015, 45 (3): 285-297.

(上接第9页)

- 53 Jiang X, Ji Z, Wang S, et al. Differential - private Data Publishing through Component Analysis [J]. Trans Data Priv, 2013, 6 (1): 19-34.
- 54 Johnson A, Shmatikov V. Privacy - preserving Data Exploration in Genome - wide Association Studies [EB/OL]. [2016-03-31]. <https://www.doc88.com/p-2292365878124.html?r=1>.
- 55 Fei Y, Rybar M, Uhler C, et al. Differentially - private Logistic Regression for Detecting Multiple - SNP Association in GWAS Databases [C]. Ibiza: International Conference on Privacy in Statistical Databases, 2014: 170-184.
- 56 Tramèr F, Huang Z, Hubaux J P, et al. Differential Privacy with Bounded Priors: reconciling utility and privacy in genome - wide association studies [C]. New York: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015: 1286.
- 57 Simmons S, Berger B. Realizing Privacy Preserving Genome - wide Association Studies [J]. Bioinformatics, 2016, 32 (9): 1293-1300.
- 58 Simmons S, Sahinalp C, Berger B. Enabling Privacy - preserving GWASs in Heterogeneous Human Populations [J]. Cell Syst, 2016, 3 (1): 54-61.
- 59 Uhler C, Slavkovic A B, Fienberg S E. Privacy - preserving Data Sharing for Genome - wide Association Studies [J]. Journal of Privacy & Confidentiality, 2012, 5 (1): 137.
- 60 Wang S, Mohammed N, Chen R. Differentially Private Genome Data Dissemination through Top - down Specialization [J]. BMC Med Inform Decis Mak, 2014 (S2): 14.
- 61 Froelicher D, Egger P, Sousa J S, et al. UnLynx: a decentralized system for privacy - conscious data sharing [J]. Proceedings on Privacy Enhancing Technologies, 2017 (4): 152-170.
- 62 Aziz M, Ghasemi R, Waliullah M, et al. Aftermath of Bus-tamante Attack on Genomic Beacon Service [J]. BMC Medical Genomics, 2017, 10 (S2): 43.
- 63 任倬辉. 基于差分隐私保护的医疗数据分析系统的设计与实现 [D]. 北京: 北京邮电大学, 2018.
- 64 陈磊磊, 陈兰香, 穆怡. 医疗辅助诊断系统中新型的双向隐私保护方法 [J]. 密码学报, 2020, 8 (1): 167-182.