医疗数据隐私保护技术应用研究*

张 茜 王 鹏 闫 慈 管 音 母建康 路正鹏 吴 琛

（新疆医科大学附属肿瘤医院 新疆 830000） （神州数码医疗科技股份有限公司 北京 100000）

孙 刚

（新疆医科大学附属肿瘤医院 新疆 830000）

〔摘要〕 介绍国内外隐私保护现状，分析现阶段数据平台隐私保护问题，阐述基于肿瘤大数据平台的医疗数据隐私保护技术及其实践，包括管理架构、技术架构等各方面，为相关研究提供参考。

〔关键词〕 数据挖掘；数据泄露；隐私保护

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.010.011

Study on the Application of Medical Data Privacy Protection Technology

ZHANG Qian, WANG Peng, YAN Ci, Cancer Hospital of Xinjiang Medical University, Xinjiang 830000, China; GUAN Yin, MU Jiankang, LU Zhengpeng, WU Chen, Digital China Health Technologies Co., LTD., Beijing 100000, China; SUN Gang, Cancer Hospital of Xinjiang Medical University, Xinjiang 830000, China

〔**Abstract**〕 The paper introduces the current situation of privacy protection at home and abroad, analyzes the privacy protection problems of the data platform at the present stage, and expounds the medical data privacy protection technology based on the tumor big data platform and its practice, including the aspects of management architecture, technical architecture and so on, so as to provide references for relevant study.

〔**Keywords**〕 data mining; data leakage; privacy protection

1 引言

1.1 医疗大数据概述

医疗行业关于大数据的研究能够对患者所患疾病隐含规律和相关性提供相关分析，为医生诊断提供极大帮助，如基因测序通过对 DNA 序列的研究分析可以推断出某人是否是肿瘤患者。此外医疗行业对大数据的研究还能对其他行业，如保险、教育、养老、物流等产生推动作用，提供更多就业机会。我国为有效利用此类数据先后发布《促进大数据发展行动纲要》、《“健康中国 2030”规划纲要》、《人口健康信息管理办法（试行）》、《国家健康医疗大数据标准、安全和服务管理办法》等相关文件。医院和相关研究机构管理患者个人信息、治疗信息甚至基因信息，如果运用得当会对医疗行业产

〔修回日期〕 2020-09-13

〔作者简介〕 张茜，副主任医师，发表论文 20 余篇；通讯作者：孙刚，主任医师，博士生导师。

〔基金项目〕 自治区创新环境（人才、基地）建设专项-科技创新基地建设“资源共享平台建设”（项目编号：PT1904）。

生积极作用。

1.2 研究背景

医疗数据采集、存储、应用、分析和共享等过程中在技术和实施方面面临很大挑战，医疗数据量剧增以及云平台应用进一步加大数据隐私安全风险，如果这些信息被泄露会给医院机构以及患者带来巨大影响或严重损失。如何保障医疗行业数据安全问题是重点也是热点。不管国内还是国外，医疗机构成为网络攻击者的重要目标。从 Forgerock 公布的近两年数据来看美国泄露隐私信息达到几十亿条，损失几近 2 万亿美元，其中大部分数据泄露是因为钓鱼网站、勒索病毒及未授权的网页访问[1-2]。温州多家医院医疗信息系统被侵入，导致大量医药信息被泄露；COVID-19 疫情期间国内某医疗公司研发的人工智能系统以及相关训练数据被攻击者盗取，以 4 比特币的价格在网上公开售出；国内某地婴儿和预产孕妇等大量信息被泄露；亚马逊 S3 近 50G 的患者极为敏感数据被泄露。根据 Forgerock 公布的数据，医疗行业在 2020 年第 1 季度隐私数据泄露事件中占一半以上[3]。面对如此严峻的形势，数据隐私安全保护刻不容缓。

2 国内外隐私保护现状

2.1 国外

如何在保护个人数据隐私的前提下实现数据共享以及最大限度地挖掘数据内在价值是当前研究的关键。国内外关于数据隐私保护主要是在法律法规和技术保护两个方面。国外，美国《健康保险流通与责任法案》(Health Insurance Portability and Accountability Act，HIPAA)[4]对个人信息隐私以及安全性规范制定国家级标准，在行政、物理、技术方面保障信息系统中个人数据隐私性、一致性、可用性；欧盟颁布《通用数据保护条例》(General Data Protection Regulation，GDPR)[5]，只要是处理欧盟用户相关数据就要受到 GDPR 的制约，是现阶段保护欧盟用户数据最严格的法案；日本颁布的《个人信息保护法》以及加拿大颁布的《隐私法》和《个人信息保护及电子文档法案》(Personal Information and Protection and Electronic Documents Act，PIPE-DA)[6]等都体现出发达国家对国民隐私保护的重视。

2.2 国内

我国目前尚无独立完整的数据隐私保护法律体系，但是在其他诸多法律法规中，如《网络安全法》、《传染病防治法》、《艾滋病防治条例》等，均有个人数据收集、传输、储存、共享等过程中涉及的隐私保护方面相关规定[7-8]。2020 年 7 月 2 日我国发布《中华人民共和国数据安全法（草案)》，该草案与欧盟公布的 GDPR 有相似之处。这些法案交织成保护个人隐私信息的强大后盾。大数据时代由于数据边界模糊不清，在保护数据网络安全和隐私安全方面加大了难度。目前数据隐私保护技术方面，有数据访问权限技术如属性加密、密钥加密机制等；匿名保护技术如 K-匿名、L-多样性匿名[9]等；数据脱敏技术如替代、数据变换、删除等；此外还有差分隐私等技术。以上隐私保护技术均有各自适用性与局限性，面对日益增加的数据量，区块链等新型数据保护技术是未来发展方向[10]。

3 基于肿瘤大数据平台的数据隐私保护技术实践

3.1 概述

医疗数据采集、传输、储存和共享等技术的发展推动医疗大数据崛起，数据处理需要平台支撑[11-12]。应用支撑平台标准规范体系包括系统标准、技术标准、应用集成、业务标准、管理标准规范等。标准规范体系是规范系统功能开发、部署、集成、应用和管理的重要依据。为加强项目数据管理，规范数据使用，防止数据丢失和泄密，保障数据安全，应从行政管理规范、物理保护规范和技术规范 3 个方面建立更加健全、完善的数据保密机制，保障项目数据安全。

3.2 医疗大数据管理架构

为保障数据隐私安全，行政管理方面需要对接

触数据的人员做权限划分，如设置监察员，负责安全策略落实，不断完善相关安全制度，对所有涉及数据信息的人员进行安全意识培训；备案员，对数据操作流程进行备案记录；审判员，从项目角度评估数据使用申请的合规性；管理员，具有数据处理账户的最高权限。所有接触数据的工作人员必须接受数据隐私安全相关培训，培养数据隐私安全保护意识。必须签署相关保密协议，对于被授权访问数据的操作人员，需要按照最小授权原则，即完成任务的最少够用信息的操作权限。数据安全管理、具体操作、审批人员要进行分离设置，对各员工操作权限定期进行评估审查。进行物理保护时，对于临时存储介质要进行加密和备案，使用后彻底格式化，对数据进行异地容灾备份。

3.3 医疗大数据技术架构

医疗数据处理包括数据采集、应用、分析展示。大数据平台由数据仓库、挖掘分析、数据清洗、采集储存、信息安全等系统组成，其业务架构[13]，见图 1。数据采集、传输、存储、共享、应用等环节容易出现疏漏，给隐私安全带来极大挑战[14-16]。为保证数据库安全，在平台数据中心上层部署网闸、流量控制器和防火墙，实现流量控制、熔断机制和单向数据流向；医院上传端同时部署虚拟专用网（Virtual Private Network，VPN），防火墙提供对外访问的出入口并严格控制访问授权，定期更新和查杀病毒库[17]。网络安全技术是保障医疗大数据平台数据安全的重要手段，使用更好的安全加密系统、先进算法以及存储手段能进一步保护平台系统中患者数据隐私安全，如同态加密可以更好地保护患者数据在云端的安全，此类技术应该是近阶段的最优选择。

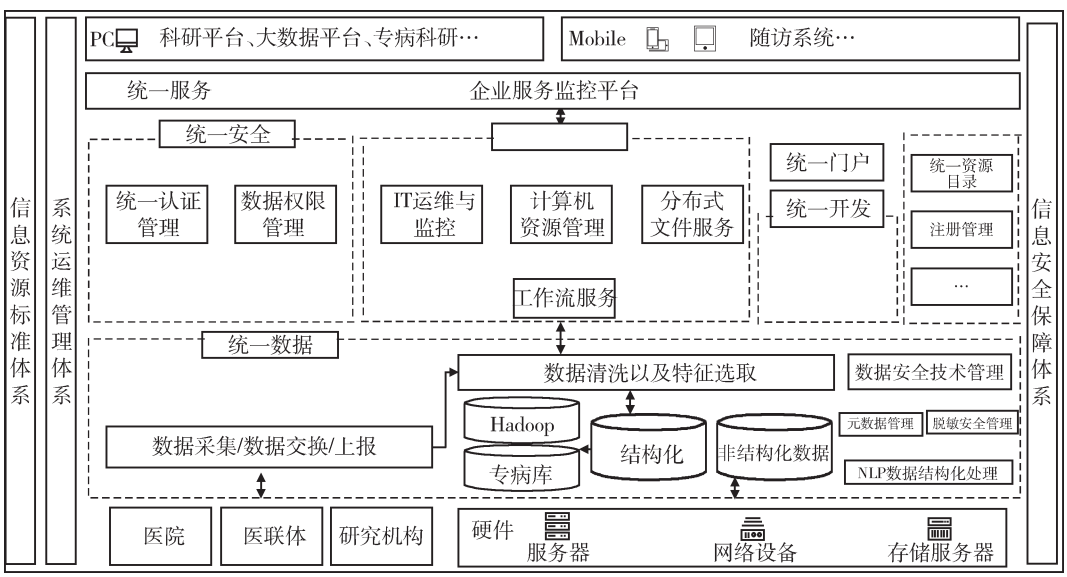

图 1 大数据平台业务架构

3.4 数据去标识化

去标识化是指通过技术手段对个人信息处理后达到无法识别相关信息主体的一种数据处理方式。通过去标识化去除数据中敏感信息，在一定程度上防止数据再识别。常用去标识化方法有屏蔽、抑制、假名化、泛化、加密、数据合成等技术，相应去标识化模型有 L-多样性、K-匿名、差分隐私等[18]。去标识化原则应遵循：合规合法，满足我国相关法律法规对隐私信息安全保护的规定；安全优先，在保护数据安全前提下处理数据；技术和管理手段相结合；技术不断改善，定期对数据进行再识别风险评估。去标识化过程主要有目标确定、目标识别、处理目标和数据导出 4 个重要环节。目标确

定主要是要明确去标识化字段、风险识别安全阈值以及去标识化实施方案；目标识别主要确定数据中的直接和间接标识符，可以通过数据规律建立自动化程序，自动识别相应数据标识符；处理目标主要是对数据进行预处理、选择相对应模型、数据去标识化、再识别风险计算等；数据导出是指去标识化后的数据在使用前需要管理层校验和批准[19]。为应对数据去标识化，国际标准化组织（International Organization for Standardization，ISO）与国际电工委员会（International Electrotechnical Commission，IEC）联合发布相关术语、技术和使用规范。国内的《信息安全技术——个人信息去标识化指南》以及《信息安全技术——个人信息安全规范》，在个人信息去标识化研究论述的基础上提出符合我国大数据发展的去标识化指南。

3.5 患者信息安全存储以及第 3 方安全技术

对于患者纸质或电子病历信息，相关管理人员需提高法律保护意识，最大限度地避免患者数据泄露、损坏和丢失。在使用病历过程中，对使用的人员进行身份审查以及备案。对于保存电子病历的电子系统，保障其不被勒索病毒、黑客入侵，定期对平台系统进行检测。共建和合作关系会牵涉数据共享问题，需要受到相关伦理委员会审查，一定要符合相应法律规范，不能因为商业利益将患者数据提供给其他商业研究机构。医院在储存患者病历、档案以及检查报告图像等信息时需要上传云端，建议在上传云端之前将数据文件切割加密。相关服务器会根据收到的数据特点和用户特征生成密钥，对切割后的数据文件用高级加密标准（Advanced Encryption Standard，AES）算法进行加密储存。使用 AES 算法时单向不可恢复加密，在没有通过用户特征获取密钥之前无法访问数据。目前我国相关医疗产品在分布式拒绝服务（Distributed Denial of Service，DDos）和 CC 攻击（Challenge Collapsar）方面有很大劣势，因此加入第 3 方安全厂商漏洞报告组织机构是一个特别好的选择[20]。另外随着技术发展，医院和患者之间若是存在第 3 方存储及监督管理机构，对医疗行业海量的数据信息进行分级存储管理，有助于建成实时存取的跨区域医疗数据系统应用平台，降低违规操作的可能性，保障信息系统服务器及后台服务安全性。

4 隐私保护问题分析

4.1 医疗数据使用基本原则

如果不能处理好数据隐私安全和伦理问题就无法消除数据孤岛、实现数据共享。为更好地实现数据去标识化，可加大在数据标准化方面的研究，通过对医疗数据的标准化建设实现医疗行业数据的科学管理以及效率提升，保护医疗数据信息安全。

4.2 技术方面问题

医疗数据标准化建设可以首先从我国大数据隐私安全标准化体系入手，制订更加规范的标准来约束数据共享和应用，从而保障数据安全使用；其次规范行业内实践，《信息安全技术——个人信息去标识化指南》已发布使用，要根据行业前沿研究成果及时转化为数据安全使用规范，与时俱进地升级去标识化规范，以引导企业、研究机构开展相关工作；再次引进国际上最新的去标识化技术研究成果，将成熟优质的去标识化模型和机制做成标准化流程来提升信息保护力度；最后企业应积极使用国家制定的标准，结合自身特点建立健全去标识化管理及实施规范等。在个人隐私保护方面，目前的技术难以满足数据隐私保护需要，应建立法律法规、技术应用、经济发展紧密结合的共同体。现阶段我国没有专门的隐私保护法案，去标识化技术运算效率低下，要加大算法优化研究力度，完善相关隐私安全保护条例，规范数据采集、传输、储存、共享使用等行为，尽快构建管理、技术、监督等全方位信息保护体系[21-23]。

5 结语

大数据飞速发展的时代机会和挑战同时存在，大数据产生巨大市场红利的同时也给信息隐私保护工作带来困难。目前去标识化、同态加密、安全多

方计算、置信计算、联邦学习[24-25]等技术手段是隐私保护研究重点。以这些关键技术为突破点，结合相关法律法规是完善数据安全体系、发展数据产业关键所在。

参考文献

1 汤啸天. 个人健康医疗信息和隐私权保护［J］. 同济大学学报：社会科学版, 2006, 17（3）：117－123.

2 颜延, 秦兴彬. 医疗健康大数据研究综述［J］. 科研信息化技术与应用, 2014, 5（6）：3－16.

3 McGraw D. Building Public Trust in Uses of Health Insurance Portability and Accountability Act De－identified Data［J］. Journal of the American Medical Informatics Association, 2013, 20（1）：29－34.

4 Bishop S K, Winckler S C. Implementing HIPAA Privacy Regulations in Pharmacy Practice［J］. Journal of the American Pharmaceutical Association, 2002, 42（6）：836－846.

5 Zhang R, Liu L. Security Models and Requirements for Healthcare Application Clouds［C］. Florida：IEEE, International Conference on Cloud Computing, 2010：268－275.

6 AlShwaier A A, Emam A Z, Arabia－Riyadh S. Data Privacy on E－Health Care System［J］. International Journal of Engineering, Business and Enterprise Applications, 2013, 3（2）：89－99.

7 Chan K S, Fowles J B, Weiner J P. Electronic Health Records and the Reliability and Validity of Quality Measures：a review of the literature［J］. Medical Care Research and Review, 2010, 67（5）：503－527.

8 Cavallaro S, Paratore S, De Snoo F, et al. Genomic Analysis：toward a new approach in breast cancer management［J］. Critical Reviews in Oncology/Hematology, 2012, 81（3）：207－223.

9 李玲娟, 郑少飞. 基于数据处理的数据挖掘隐私保护技术分析［J］. 计算机技术与发展, 2011, 21（3）：94－97.

10 徐国海. 面向中文医疗文本的命名实体识别研究［D］. 上海：华东师范大学, 2019.

11 王兰成, 李超. 论档案信息共享中的隐私保护及新技术［J］. 档案学研究, 2017, 24（4）：41－45.

12 王伟. 我国信息网络安全立法问题研究［D］. 西安：西安理工大学, 2006.

13 黄小燕. 欧盟网络个人信息法律保护研究［D］. 广州：暨南大学, 2018.

14 冯登国, 张敏, 李昊. 大数据安全与隐私保护［J］. 计算机学报, 2014, 37(1)：246－258.

15 朱晓勃. 我国医院信息化建设现状与发展对策研究［J］. 现代仪器, 2015, 21（1）：76－79.

16 沈昌祥. 加快推进信息安全等级保护工作［J］. 信息网络安全, 2008, 8（5）：4－5.

17 王红梅, 宗慧娟, 王爱民. 计算机网络信息安全及防护策略研究［J］. 价值工程, 2015, 34（1）：209－210.

18 Tomes J P. The Health Insurance Portability and Accountability Act of 1996：understanding the anti－kickback laws［J］. Journal of Health Care Finance, 1998, 25（2）：55－62.

19 Lin C, Wang P, Song H, et al. A Differential Privacy Protection Scheme for Sensitive Big Data in Body Sensor Networks［J］. Annals of Telecommunications, 2016, 71（9－10）：465－475.

20 Sweeney L. K－anonymity：a model for protecting privacy［J］. International Journal of Uncertainty, Fuzziness and Knowledge－Based Systems, 2002, 10（5）：557－570.

21 Link M P, Hagerty K, Kantarjian H M. Chemotherapy Drug Shortages in the United States：genesis and potential solutions［J］. Journal of Clinical Oncology, 2012, 30（7）：692－694.

22 Knoppers B M. International Ethics Harmonization and the Global Alliance for Genomics and Health［J］. Genome Medicine, 2014, 6（2）：13.

23 Curtmola R, Garay J, Kamara S, et al. Searchable Symmetric Encryption：improved definitions and efficient constructions［J］. Journal of Computer Security, 2011, 19（5）：895－934.

24 Ahrendt S A, Decker P A, Doffek K, et al. Microsatellite Instability at Selected Tetranucleotide Repeats is Associated with p53 Mutations in Non－small Cell Lung Cancer［J］. Cancer Research, 2000, 60（9）：2488－2491.

25 Xhafa F, Feng J, Zhang Y, et al. Privacy－aware Attribute－based PHR Sharing with User Accountability in Cloud Computing［J］. Journal of Supercomputing, 2015, 71（5）：1607－1619.