

DOI:10.16644/j.cnki.cn33-1094/tp.2018.11.011

健康医疗大数据的安全保障技术研究

黄 婧, 王云光, 皮冰斌

(1. 上海理工大学医疗器械与食品学院, 上海 200093; 2. 上海健康医学院医疗器械学院)

摘 要: 健康医疗大数据的发展是大数据产业驱动的结果,是国家重要的战略部署。为了有效地推动数据开放共享,挖掘医疗行业潜在的数据价值,推动行业的发展进步,就必须加强数据安全保护技术的研究。文章从大数据平台安全的角度入手,重点分析了健康医疗领域相关数据安全保障技术的重难点。研究表明,针对不同的应用场景使用不同的安全保障技术能够有效地加强数据安全保护,提高数据传输的安全性。

关键词: 健康医疗; 大数据; 数据平台; 数据安全

中图分类号: TP301

文献标志码: A

文章编号: 1006-8228(2018)11-45-04

Research on security technology for healthcare big data

Huang Jing, Wang Yunguang, Pi Bingbin

(1. School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. School of Medical Instrument, Shanghai University of Medicine and Health Science)

Abstract: The development of healthcare big data is the result of the big data industry, and it is an important strategic deployment of our country. In order to effectively promote the open sharing of data, tap the potential data value of the medical industry, and promoting the development of this industry, the research on data security protection technology must be strengthened. Starting from the security perspective of big data platform, this paper focuses on analyzing the key and difficult points of data security technology in the field of health care. The researches show that using different security technologies for different application scenarios can effectively enhance the data security protection and improve the security of data transmission.

Key words: healthcare; big data; data platform; data security

0 引言

步入21世纪以来,随着社会网络信息化的高速发展,使得物联网、互联网、云计算和人工智能等新一代信息技术对健康医疗事业的影响日趋显著。美国、英国等发达国家已经将健康医疗大数据的发展作为国家公共事业发展的的重要组成部分投入了大量的人力物力。2016年6月,国务院正式印发了《关于促进和规范健康医疗大数据应用发展的指导意见》,首次将健康医疗大数据定位为“国家重要的基础性战略资源”^[1]。由此可见,国家对健康医疗大数据的重视,而大数据在健康医疗方面的应用与发展也将极大地提升医疗服务质量与效率。由此引来的医疗数据的安全问题也被提上日程。医疗数据保密性很高,数据量庞大,

数据结构复杂且带有一定的专业特殊性,这些都加大了数据安全保护的难度。因此,在进行数据的开放与共享以谋求更大的价值的同时,我们更应该从技术、法律等各方面入手,保障数据的安全。

1 健康医疗大数据研究概述

1.1 健康医疗大数据的概念与特征

健康医疗大数据是涵盖人的全生命周期,既包括个人健康,又涉及医药服务、级别防控、健康保障和食品安全、养生保健等多方面数据的汇聚和聚合^[2]。对改进医疗服务模式以及国家经济社会的发展都具有一定的促进作用。大数据一般都具有数据量大、处理速度快、数据种类多,以及价值密度低的特征。健康

收稿日期:2018-08-16

作者简介:黄婧(1994-),女,江苏泰州人,硕士,主要研究方向:健康医疗大数据安全。

医疗大数据在此基础之上又具有其本身的一些特征,即时效性、不完整性、冗余性以及保密性^[3]。

时效性:信息仅在一定时间内对决策具有影响。患者在就医的不同阶段产生的数据对后续的治疗手段方法等都会带来不一样的影响。

不完整性:因为技术手段、人为因素等导致我们无法全面搜集,记录、处理疾病的全部信息。致使数据出现偏差乃至缺失的情况,造成了数据的不完整性。

冗余性:冗余性是指数据之间的重复,或者同一数据被多次记录的现象。例如,同一患者会因医院不同造成同一检查项的多次检查,造成数据的冗余。

保密性:医疗保密即医务人员在医治患者的过程之中应当保守医疗秘密,不得对外泄露病人的隐私及病情。相关医疗机构不得在未经患者同意的情况下,以任何方式将患者的个人信息透露给外界。

1.2 健康医疗大数据的来源与隐私保护价值意义

1.2.1 数据来源

健康医疗大数据从数据结构上可以分为结构化、半结构化和非结构化三类。各式各样的数据来源于不同的地方,按照数据产生的来源,可将健康医疗大数据分为临床诊疗、医院管理、医学研究、公共卫生和个人健康五类。如表1所示。

表1 健康医疗大数据分类

类别	数据来源
临床诊疗	个性化诊疗、疾病早期诊断、不良事件预警
医院管理	绩效评价、医疗质量管理、流程优化
医学研究	疾病相关分析、疾病精准分析、生物标记筛查
公共卫生	人口健康平台、大众健康、传染病控制
个人健康	个性化医疗、远程健康监控、智能穿戴设备数据分析

1.2.2 健康医疗大数据安全保护价值意义

在我国,多方需求共同推动健康医疗大数据的发展,首先社会需求加快了大数据的应用。据统计,我国人口老龄化进程显著加快。预计到2035年60岁以上人口将增至4.18亿,约占人口比例的29%^[4]。

我国慢性病人群众庞大,已经被确诊大的患者高达2.6亿人且每年仍以8.9%的速度在递增^[5]。根据北京市药监局西城分局对辖区内五个街道的过期药品回收状况的调查显示,91.8%的家庭有过期药品,70.1%的家庭存储过期药品超过半年。

我国医疗领域需求庞大,医疗资源分配不均衡;医疗信息不对称、不透明、不开放、不共享,也导致了

信息的冗余,患者不能参与到医疗过程之中,医生也不能根据以往有用信息迅速的做出精准的判断,往往会造成治疗的延误导致错失治疗的最佳时期。

大数据在健康医疗领域的有效应用将大大减少上述问题对我们的困扰。而保障数据的安全有效,杜绝信息的泄露是大数据在健康医疗领域的应用的最基本问题,也是人们最关注的一个问题。因此,保障好数据的安全,是当下我们迫切需要解决的基本问题之一。

2 健康医疗大数据平台安全关键技术

医疗大数据平台为大数据在医疗信息领域的应用和发展提供了有利的支持保障。平台的安全体系建设则更加有利于健康医疗大数据的发展。涉及的关键技术有身份认证技术、数据隔离技术、访问控制技术以及审计技术。如图1所示。

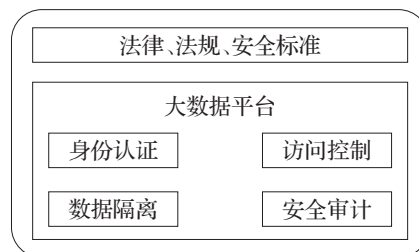


图1 大数据安全保护

2.1 医疗信息系统的身份认证技术

目前身份认证技术主要包括:以口令为基础的认证技术,智能卡认证技术,密码认证技术、多因子认证技术。

口令认证技术是最简单方便快捷的认证技术,其优势在于成本低,速度快,但该方式的安全性较低。智能卡认证技术中智能卡具有硬件加密功能,安全可靠,但是该方法增加了成本开销,需要在每个终端都安装读卡设备,且对于一些信息系统而言该方法不适用。密码认证中较为经典的认证系统有两个,一个是Kerberos认证系统另一个是PKI II CA系统。Kerberos认证系统支持分布环境下的认证服务和双向认证服务,能够为网络中的实体提供一个集中的、统一的认证管理机制。该系统解决了密钥管理的问题,也解决了执行效率的问题。PKI II CA系统的认证鉴别机制安全性较好,适合网上的安全认证,但是该系统也存在不足之处。例如无法验证用户提供信息的真实性,用户私有密钥保存的安全问题等。多因子安

全认证技术相比于传统的认证技术在安全凭证方面添加了多种因素,进一步的加强了安全认证的可行性,但是该方法太过复杂不便于操作。因此,在身份认证方面,需要一种可行方法,在保证安全性能的同时,提升执行效率与可行性,这方面仍有待研究。

2.2 医疗信息系统数据隔离技术

虚拟化技术的负面作用之一是削弱了数据间的物理隔离,致使数据间的边界很模糊,每个用户都有成为发起攻击节点的潜在条件,对数据的安全构成了极大的威胁^[6]。因此,开发数据隔离机制来保证用户之间的数据不可见是解决问题的关键。在隔离技术中较为主流的有以下几种。

(1) 分离表架构:该方法中每个用户都拥有属于个人的数据库表,系统共享时只会共享相同的数据。

(2) 共享表架构:通过字段来确定数据之间的关系,系统共享时,共享相同的数据实例和数据库表。此架构在降低硬件成本的同时,极大地利用了数据实例的存储能力,缺点是复杂程度增加了,产生了高昂的容灾备份成本。

(3) 分离数据库架构:这种架构能够高效实现数据隔离和容灾备份,但是硬件成本也相对较高。

2.3 数据访问技术

对大规模的医疗数据资源进行管理时,为降低安全风险,可根据用户的需求和数据的保密程度赋予用户和数据不同的等级权限。针对普通医疗数据的访问控制,可以通过属性加密和角色控制两种方法。而针对对用户访问需求不明确的情况,出现了一种新型的风险自适应访问模型。

2.3.1 基于角色挖掘的访问控制方法

角色挖掘与传统的角色设计的根本不同之处在于角色挖掘是“自下而上”的从已有的用户-权限分配关系中来自动化地实现角色定义和管理工作,以减小对管理员地依赖^[7]。在保证系统已有用户-权限分配关系准确的情况下,目前已有的研究方法有利用聚类进行角色挖掘的方法,用子集枚举的角色挖掘算法等。上述方法都能够在一定程度上降低对管理员的依赖。

2.3.2 基于属性加密的访问控制

基于属性加密的访问控制是一种利用密文机制实现客体访问控制的方法,主要分为两种:基于密钥策略的属性加密(KP-ABE)和基于密文策略的属性加

密(CP-ABE)^[8]。KP-ABE 主要用来访问静态数据,CP-ABE 因为可以灵活的控制用户访问数据,所以被广泛地应用于云计算地访问控制。为解决传统方案中,密文与密钥长度都与属性个数线性相关从而使得计算开销增加的问题。Sreenivasa 和 Ratna 提出了一种多权限分散的 CP-ABE 机制,利用最小授权集加密数据,因此密文大小与访问结构中的最小属性集呈线性关系,且在解密期间双线性配对操作数是不变的^[9]。Chen 等提出了一种用于云计算的具有定长密文的多权限 CP-ABE 访问控制方案,密文的长度和解密过程中的配对操作数都是不变的,与访问结构中设计的属性数也无关,在相对较强的安全模型中保持了高效率^[10]。

2.3.3 基于风险自适应的访问控制

研究者注意到仅仅基于风险的访问控制的判定是不合理的,在医疗信息系统中,紧急情况发生,风险较大的访问请求被简单的拒绝可能会延误治疗的时期,对病人、医院造成不可挽回的损失。一种弹性的风险判断方式被人们所研究采纳,即风险带的概念。有研究者采用了一种风险自适应访问控制实施办法,在严格拒绝和弹性拒绝之间有着一个细分的风险容忍区域,可以根据访问行为的风险系数在其中的位置来调整权限,从而提高了访问控制判定的灵活性。也有研究者从算法模型的角度进行相关的研究,文献[11]以诚实医生访问行为的熵作为系统可承受风险的基准值,对所有医生的访问行为使用 EM 算法进行进一步的分析,对不同医生的访问行为的概率分布进行了区分,利用风险量化,监测和控制对于医疗记录的过度访问以及特殊情况下的访问请求。研究证明,该方法确实能够有效提高风险评估的准确性。

2.4 大数据审计

大数据处理平台也采用安全审计技术来对安全事件进行跟踪,以及时发现安全违规事件,便于进行安全事件追责^[12]。安全审计首先搜集原始的系统状态信息,然后将原始状态信息和已有的安全记录(包括已经发生的安全问题及其他类似系统发生的安全问题)进行汇总整理,以此为基础通过数理统计导出相应的结论,在结论分析基础上,制定安全等级,采取相应的安全应对措施,预防可能会发生的安全问题^[13]。目前大数据平台主要通过审计日记记录平台中所有数据操作。Hadoop 生态的几个常用组件都可以配置具有审计功能。

3 未来展望

未来,大数据在医疗健康领域的应用将会越来越广泛。数据的开放共享是使数据价值利用最大化的根本途径。在数据安全保护方面,依然是研究的重点,亟待解决的问题仍有很多。相关的法律法规和政策的制定都应该以保护数据安全,推动大数据在医疗健康领域的健康发展为主。其次要思考解决如何实现数据安全与数据共享的均衡问题;数据共享与数据隐私保护的均衡问题等。按需制定访问控制策略、保障数据有效加密的同时又不影响执行效率。


4 结束语

大数据作为国家重要的战略性基础资源,在健康医疗领域的应用会推动该产业翻天覆地的变化。数据安全问题也得到了学术界和产业界的高度重视。本文对健康医疗领域的 data 安全问题进行了研究,从大数据平台的角度入手对目前身份认证技术、数据隔离技术、数据访问控制技术和数据审计几个方面所使用的较为先进的安全保障技术进行了研究,总结了每个方法的优劣性,分析得出大数据安全保障技术下一步要在数据安全与数据共享、数据共享与隐私保护等问题上做进一步的探讨。

参考文献(References):

- [1] 卢朝霞,姚勇,尹新等.健康医疗大数据理论与实践[M].电子工业出版社,2017.
- [2] 卢朝霞,姚勇,尹新等.健康医疗大数据理论与实践[M].电子工业出版社,2017.
- [3] 戴明凤,孟群.医疗健康大数据挖掘与分析[J].中国卫生信息管理,2017.14(2):126-130
- [4] 中华人民共和国国家统计局.中华人民共和国2015年国民

经济和社会发展统计公报.中国统计,2015.

- [5] 国家卫生和计划生育委员会统计信息中心.2013中国卫生服务调查研究.中国协和医科大学出版社,2015.
- [6] 王丹,赵文兵,丁浩明.大数据安全保障关键技术分析综述[J].北京工业大学学报,2017.43(3):335-349
- [7] 李昊,张敏,冯登国等.大数据访问控制研究[J].计算机学报,2017.1:72-91
- [8] 陈兴蜀,杨露,罗永刚等.大数据安全保护技术[J].四川大学学报(工程科学版),2017.5:1-12
- [9] Rao Y S,Dutta R.Decentralized ciphertext-policy attribute-based encryption scheme with fast decryption[C]//Proceedings of the 14th IFIP TC 6/TC 11 International Conference on International Conference on Communication and Multimedia Security.Magdeburg: Springer-Verlag,2013:66-81.
- [10] Chen Yanli,Song Lingling,Yang Geng.Attribute-based access control for multi-authority systems with constant size ciphertext in cloud computing[J].China Communications,2016.13(2):146-162
- [11] Hui Zhen, Li Hao, Zhang Min, Feng Deng-Guo. Risk-adaptive access control model for big data in healthcare.Journal on Communications,2015.36(12): 190-199(in Chinese).
- [12] BAUMG RTNER L, STRACK C, HOBACH B. Complex event processing for reactive security monitoring in virtualized computer systems[C]//Proceedings of the 9th ACM International Conference on Distributed EventBased Systems. Oslo: ACM,2015:2233
- [13] SOOKHAK M,GANI A,TALEBIAN H,et al. Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues[J]. Computing Surveys, 2015.47(4):134 

(上接第44页)

on the Web[EB/OL]. <https://www.w3.org/2016/04/blockchain-workshop/>,2016-06-29

- [10] 刘海英.“大数据+区块链”共享经济发展研究[J].技术经济与管理研究,2018.1:91-95

- [11] 胡凯,白晓敏,高灵超等.智能合约的形式化验证方法[J].信息安全研究,2016.2(12):1080-1089
- [12] 郁莲,那恩艳.区块链技术[J].中国计算机学会通讯,2017.13(5):10-15 