健康医疗大数据的安全机制研究

肖忠良[1], 李默轩[2], 李　晶[3]

(1.湖南娄底职业技术学院, 湖南　娄底　417000; 2.北京陆融通达科技有限责任公司, 北京　100020;
3.娄底市第一人民医院, 湖南　娄底　417000)

摘　要: 在国家政策的引导下和人民健康诉求的促进下, 以健康医疗大数据为基础的产业链将迎来巨大的发展空间和机遇, 但如影随形的安全问题, 同样也是健康医疗大数据发展过程中的一个重要制约因素。文章分析了医疗大数据的特性, 提出了分级分类安全管理模型, 并从数据的存储、访问控制及数据的管理方面探讨了健康医疗大数据面临的风险及应对策略。

关键词: 健康医疗大数据; 分级分类; 存储安全; 访问控制; 数据管理

随着健康医疗大数据的汇聚, 以及数据挖掘、数据分析、人工智能等技术的不断革新, 利用大数据进行分析、预测、科研的场景会越来越多[1]。大数据将为医疗相关行业的诊疗和决策提供重要的辅助依据, 决策的方式也会从之前的“经验即决策”, 到现在的“数据辅助决策”, 至将来的“数据即决策”[2]。尽管医疗大数据可以产生许多有用的信息和价值, 但其作为医疗领域产生的数据具有数据量大、敏感性高等特点, 要实现医疗大数据的融合共享, 首先要警惕数据安全, 因此, 保证健康医疗大数据的安全是医疗行业开展大数据技术的重要前提[3]。本文提出了分级分类安全管理模型, 并从存储、访问和管理3方面探讨健康医疗大数据存在的安全隐患及对应的策略。

1　健康医疗大数据的特性

不同于一般行业的数据, 医疗数据具有其特殊的敏感性和重要性。医疗数据的来源和范围也非常广泛和多样, 涵盖医院诊疗、医疗保险、医学实验、科研数据等[4]。这些数据不仅关系到数据主体的隐私、行业发展, 甚至关系到国家安全。比如, 2016年艾滋病感染者个人信息遭泄露的事件, 让诈骗集团有机可乘, 并引起了世界卫生组织驻华代表处和联合国艾滋病联合规划署驻华代表处的关注。

随着信息化的普及和医疗数据的逐步集中, 企业、研究机构及公众对数据访问的需求将变得迫切, 如果不提升安全防护水平, 大规模数据泄露的风险将会增加[5]。警惕数据安全, 保护患者隐私, 才能真正实现数据融合共享、开放应用。

2　健康医疗大数据面临的挑战及应对策略

医疗行业是数据密集型行业, IDC Digital预测截至2020年, 医疗数据量将达到40万亿GB。由于健康和医疗数据的高度敏感性, 对其进行集中存储和管理后, 一方面会引起恶意人员的高度关注, 另一方面, 一旦发生数据泄露其影响面非常广, 对于健康医疗大数据的安全和个人相关的隐私保护, 必须予以高度重视。基于数据的存储、访问和管理方面, 提出了3层的分级分类安全管理模型, 如图1所示。

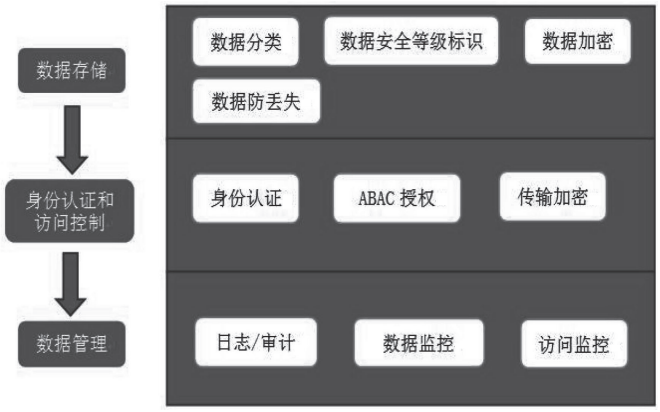

图1　分级分类安全管理模型

2.1 数据存储

数据存储是否安全高效, 关乎隐私性、医疗相关业务的连续性、医疗大数据的应用价值, 系统一旦出现故障, 首先考验的是数据的存储和恢复能力。为避免数据丢失问题, 需对数据进行定期备份, 并定期进行数据恢复验证测试, 确保备份数据的可恢复性。

在网络架构方面, 依据医疗大数据的特点, 提出分级分类存储解决方案, 根据数据的时效性、访问频率、容量、性能等指标, 将数据进行分级管理, 采取不同的存储方式分别存储在不同性能的存储设备上, 以获得更好的性价比。根据数据的隐私性为数据设置不同的安全标识, 进行安全分类, 为下一步访问控制策略提供控制依据。此外, 通过相关加密算法和密钥对数据进行加密存储, 可以从数据源层面保护敏感信息不被泄露。

2.2 数据访问

由于医疗健康大数据的特殊性, 将多个数据池中的数据进行组合时, 隐私风险也将成倍增加, 这是由于人们很难从单条数据中推断出用户的身份, 但是当对多条数据进行组合分析时, 推断出用户身份特征的概率将大大增加, 进一

基金项目: 娄底职业技术学院院级课题; 项目名称: 个人信息泄露防范技术探究; 项目编号: 2012zk005。

作者简介: 肖忠良(1974—), 男, 湖南涟源人, 高级工程师, 硕士; 研究方向: 信息安全与项目管理。

步可能根据获得的信息对患者进行预测和预判，危害无可估量。

医疗大数据汇集后，需要相对开放的共享给内部不同团队或外部机构使用，才能发挥大数据的价值。在访问过程中存在两种威胁：一是在信息使用传递过程中发生的泄露，可能包括科学研究的过程，区域性平台数据交互等；二是基于健康医疗信息的敏感性，对访问者的访问权限控制和对医疗信息的隐私保护。

在传输过程中的加密依赖于网络安全协议。收集到的海量数据供个人、企业或有关机构访问时，首先通过对称加密的方式加密传输的数据，然后使用非对称加密的方式传递对称加密所使用的密钥，这样既能保证数据传输的效率，也能保证数据的安全。

引入Kerberos网络认证协议进行身份认证，可有效保证用户身份的可靠性以及数据源的不可否认性，用户通过身份认证后可获得访问大数据平台的资格。然后以分级分类存储的数据为基础，通过一个多元组对ABAC（Attribute Based Access Control）访问控制机制的属性进行描述，包含用户的实体属性，如年龄、姓名等；数据安全属性，如病历文档、B超图片、CT影像等数据的安全标识；操作权限属性，如对数据的读、写、删除等；环境属性，如用户访问的时间，网络位置等，通过定义完备的属性—权限之间的对应关系，制定细粒度的访问限制规则，可控制到被访问对象的字段级别，通过Kerberos身份认证和ABAC访问控制来管理不同用户对不同资源的访问许可。

2.3 数据管理

要保证医疗大数据的安全，必须做好数据的管理工作，一是根据数据的敏感性、关联风险和业务要求等对数据进行分类分级管理，如姓名、证件号、联系方式等信息应进行严格的管控和保护，而对于诊疗过程数据、病历信息等健康医疗数据，则可以在做好访问控制的前提下供授权者访问。二是从大数据特性层面对数据进行标记（例如数据源、数据类型、访问频率、访问角色、处理方式等维度），了解数据流的流向、使用方式、使用对象等，这些有助于数据发现的管理，并为数据访问控制策略提供依据。此外，掌握敏感数据在大数据平台中分布情况，并监控其使用情况，适时地调整访问策略，是能否做到全面保护数据安全的关键。

虽然通过数据保护、身份认证、授权及访问控制等各种方式可以一定程度保证健康医疗大数据平台的安全。但大数据平台仍然有可能会受到非法访问和特权用户的访问，因此，我们需要根据预先定义的规则对大数据平台的一切活动进行审计和监控并生成告警信息，对其中的可疑活动进行记录，分析和生成各种安全报告。如用户登录和身份验证事件、授权错误、敏感数据操作等异常事件。只有全面收集在大数据平台中的一切活动，才有机会捕捉可能会发生的安全事故及进行事后分析时有机会进行回溯分析，追踪事故根源。

3 结语

综上所述，随着医疗数据的汇聚和健康产业的发展，医疗大数据可以产生许多有用的信息和价值，但其高度的隐私性和敏感性，使得数据的安全问题日益突出，一旦准备和配套不足，很有可能引发全局性安全风险，影响健康医疗大数据整体产业布局和发展。如何更好地保护敏感信息及病人隐私，成为实现数据融合共享、开放应用的一大难题。本文分析了健康医疗大数据应用中可能存在的风险，建立基于数据分级分类的安全管理模型，从数据存储、访问控制和数据管理方面提出了相关的安全策略。

[参考文献]

[1]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.
[2]王艺，任淑霞.医疗大数据可视化研究综述[J].计算机科学与探索，2017（5）：681-699.
[3]许培海，黄匡时.我国健康医疗大数据的现状、问题及对策[J].中国数字医学，2017（5）：24-26.
[4]马诗诗，于广军，崔文彬.区域卫生信息化环境下健康医疗大数据共享应用思考与建议[J].中国数字医学，2018（4）：11-13，25.
[5]代涛.健康医疗大数据发展应用的思考[J].医学信息学杂志，2016（2）：2-8.

Research on security mechanism of healthcare big data

Xiao Zhong Liang[1], Li Moxuan[2], Li Jing[3]
（1.Loudi Vocational and Technical College, Loudi 417000, China; 2.Beijing Lurong Tongda Technology Co., Ltd., Beijing 100020, China; 3.The First People's Hospital of Loudi City, Loudi 417000, China）

Abstract: Under the guidance of national policies and the promotion of people's health appeals, the industry based on healthcare big data will usher in unprecedented opportunities. However, the security is an important constraint in the development of big data. This paper has a research on features of healthcare big data, and builds up a multi-level classified security model, and discusses the risks for healthcare big data in terms of data storage, access control, and data management, and puts forward the strategies.

Key words: healthcare big data; multi-level classified; data storage; access control; data management