保障医疗大数据安全及其实践*

王红迁 汪 鹏 王 飞 罗 浩

（第三军医大学西南医院 重庆 400038）

〔摘要〕 基于目前医疗大数据平台中数据安全存在的问题，首先从安全形势入手分析我国医疗大数据安全保障的重难点和常见保护措施，结合西南医院医疗大数据平台实践，提出目前医疗大数据平台数据安全保障策略，包括管理安全、脱密安全、存储安全、网络安全等方面。

〔关键词〕 医疗大数据；大数据安全；隐私保护；安全保障体系

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2017.12.010

Guaranteeing Security of Medical Big Data as Well as Its Practice *WANG Hong－qian, WANG Peng, WANG Fei, LUO Hao, South West Hospital Third Military Medical University, Chongqing 400038, China*

〔**Abstract**〕 Based on current data security problems existing in medical big data platform, the paper starts from security situation first, analyzing the key points, difficulties and common protection measures of security guarantee of domestic medical big data. Combining with the medical big data platform practice implemented by Southwest Hospital, it puts forward security strategy for data security of current medical big data platform, including management security, decryption security, storage security, network security, etc.

〔**Keywords**〕 Medical big data; Big data security; Privacy protection; Security assurance system

1 引言

目前，全球发生骇人听闻的健康信息数据泄露事件越来越多，后果越来越严重。如 2016 年白桦林全国联盟共接到来自 30 个省区市的 275 例艾滋病感染者因个人信息发生泄露而导致的诈骗，2016 年英国国家医疗服务体系（National Health Service, NHS）就因数据安全问题停止使用 Care Data 健康医疗大数据平台[1]。类似的安全事件，给医院信息化工作者敲响了警钟，总体来看目前医疗大数据的安全形势不容乐观，发生数据泄露的主要原因如下[2]：（1）网络攻击频发[3]。美国非营利性组织身份盗用资源中心的统计数据显示，医疗健康信息系统成为黑客首要攻击目标，占漏洞总数的 43.8%。（2）安全意识不健全。我国相关的隐私保护等法律法规系统不完善，居民和相关机构对数据安全意识淡薄，各方面的基础建设也相应不足。（3）新技术、新挑战。随着云存储、云计算、云服务、虚拟化等技术的广泛应用，过去的安全防护方案已成为制约数据安全的瓶颈。

最近，国家颁布的“健康中国 2030”规划纲要和《国务院办公厅关于促进和规范健康医疗大数据应用发展指导意见》都明确提出推进健康医疗大数据应用，重点提出加强健康医疗大数据安全保障和

〔修回日期〕 2017－06－13

〔作者简介〕 王红迁，硕士，工程师，发表论文 2 篇；通讯作者：汪鹏。

〔基金项目〕 重庆市社会民生项目（项目编号：cstc2015shmszx120 025）；智慧医疗重大领域项目（项目编号：SWH 2016ZDCX4102）。

患者的隐私保护。发展健康医疗大数据应首先保障数据安全，保护患者隐私，唯有如此才真正实现医疗数据互联互通、共享和规范应用。

2 常见医疗大数据平台体系

目前医疗大数据架构路线最典型的共有 4 种[4]：第 1 种是采用 MPP 架构的新型数据库集群；第 2 种是基于 Hadoop 的技术扩展和封装；第 3 种是大数据一体机，这是一种专为大数据的分析处理而设计的软、硬件结合的产品；第 4 种是采用常规的数据仓库技术。医疗行业既有结构化数据又有非结构化数据，如何设计架构才能保障数据集中平台的最优化和高性价比值得探讨。任何一种技术实现，必然是为更好地体现产品或服务的价值，医疗机构都选择基于 Hadoop 技术扩展和封装的医疗大数据平台，一个特别重要的原因是 Hadoop 开源，有着高可靠性、高扩展性、高效性、高容错性的口碑，能够降低成本的同时确保技术的可靠与发展的延续。

3 医疗大数据安全新威胁

3.1 大数据时代带来的新威胁

在新时代用户面临在毫不知情的情况下个人信息被收集和利用，存在信息使用知情权的风险。同时用户也面临在未经同意的情况下被推送信息和服务，存在个人被“骚扰”的风险。最严重的是用户个人数据信息被随意共享和交易，面临信息被随意支配的风险。在此背景下用户面临严重个人信息数据泄露的危机[5]。

3.2 新模式带来的新威胁

IT 变得越来越复杂，IT 基础架构和应用模式也随着云计算时代的到来而变革。数据的云端集中存储增加了数据泄露风险。大数据集群的分布式架构导致网络层安全策略、交互机制都面临新的挑战。同时大数据不同于传统数据库，数据量、架构、存储模式和查询模式，需重新实现或全面设计多数据安全工具的架构[4]。基于 Hadoop 的大数据架构本身的安全机制并不成熟，很多机构采用与第 3 方合作构建大数据平台，该模式将数据的所有权转移给服务提供商，用户失去了对数据资源的直接控制。

4 医疗大数据安全保护的重难点

4.1 概述

在健康医疗大数据背景下，健康医疗大数据将呈现日益活跃的“流动”趋势，在“流动”中发挥价值。作为医院信息化工作者，保障医疗数据安全的意识应贯穿数据的全生命周期，从数据产生、采集、传送、存储、管理、分析、发布、交易、使用、销毁等各个环节都要重视[6]。

4.2 访问控制

首先确定谁能访问，其次是对访问者身份进行认证，最后也是最重要的一点是对访问者的权限控制和访问后的行为追溯，这是数据控制的重要目的。

4.3 全方位保护

数据全方位保护也要做 3 个方面的工作，分别是对信息完整性的保护、防泄露的机密性保护、安全有序的可控“流动”。

4.4 个人隐私脱密

医疗数据中个人隐私难以脱密，主要原因在于医学数据类别众多，各种数据中都隐含敏感信息且没有明确的规则。

4.5 安全共享

主要原因在于医疗大数据融合多个机构数据后，如何在保障数据安全的前提下使用数据是个难题。

5 医疗大数据安全保护措施

5.1 概述

现今，大数据技术在医疗卫生行业的应用越来

越广泛，将患者数据存储在云端也是一种趋势，如何充分利用大数据技术又保障医疗数据的安全，做到安全和发展相辅相成，“防”与“用”两手都要硬，已成为亟待解决的问题。

5.2 法律层面

医院信息化工作者和使用该平台的医务工作者应有良好的法律意识，对患者的信息保密，在员工入职时签订保密协议，明确员工对患者信息保密的法律义务与责任。所以对于医务从业者而言保持良好的法律意识和做出相应的行动，对于患者的数据和信息安全具有重要意义。

5.3 管理层面

医院应建立相应的内控机制，使医疗大数据平台中患者数据在医院的内部体系处于可控范围。在这个内控系统内对涉及患者数据和信息的流程建立相应的控制措施，明确每个环节数据保护的责任人，在管理层面给予相应的重视。

5.4 技术层面

要建立数据的安全等级、访问控制机制、追踪审查机制以及大数据集群内部安全防范机制。医院应有大数据技术人员通过技术手段限制对医院各种数据的非法访问，限制医院大数据平台中数据的导出，同时严格控制技术人员对医院大数据的访问权限，对其进行严格实时监控和审计，确保数据不被泄露。

5.5 合作层面

在与第 3 方机构合作构建医疗大数据平台中，医院信息化工作人员应特别注意对医院数据进行保护，做到医院大数据平台内部数据无法出网，防止在合作过程中导致的医院数据的不慎泄露。

6 西南医院医疗大数据安全实践

6.1 概述

面对日益严峻的个人医疗信息被侵害的风险，建立全方位、立体化的数据管理框架，是解决医疗大数据平台数据安全的重要手段。西南医院整个医疗大数据平台采取私有云和公有云混合的模式，采取医院内部自建大数据集群，这样的既可保障医疗数据的高安全性同时又满足数据的开放使用的需求。

6.2 管理安全

牢固树立法律意识，医院定期组织相关人员学习相关的保密法规，增强全体医务人员的法律意识。在构建医疗大数据平台的时候，把保障数据安全放在首要位置，成立专门安全团队设计整个安全框架，制定规范的安全操作流程。

6.3 脱密安全

医疗数据本身就有强烈的敏感性，为避免敏感数据因泄露等原因造成严重的后果，医院医疗大数据集群医疗大数据集群中采取数据脱密和加密存储的保障机制，目前采用 HIPAA Section 164.51（b）（2）法案的脱密机制，具体的脱密策略可以根据具体需求和具体科室灵活的调整。具体的脱密字段，见表 1。

表 1 脱密字段

身份标识	政策	身份标识	政策
姓名	***	身份证号	***
地理位置	保留至城市，邮编保留前 4 位	许可证书	***
出生日期	保留到年	车辆号码	***
电话号码	***	设备标识符和序列号	***
传真号码	***	网址	***
邮件地址	***	IP 地址	***
社会保障号码	***	生物识别标识符	***
病历号码	MD5 加密	面部生理图像	***
健康计划受益人号码	***	其他识别信息	***

注：*** 表示信息隐藏。

6.4 存储安全

医疗大数据平台采取私有云架构，保障外部人员无法获取医院医疗大数据集群数据。大数据存储采用分布式云存储技术，为保证大数据安全存储，采取对称秘钥和非对称秘钥相结合的加密技术，同时保证秘钥和加密数据存储在不同的节点，并且采取多重备份数据保证容灾机制。具体设计中采取批量导入的方式解决数据抽取的问题，即原有的信息系统的数据通过 ETL 每天定时拷贝到大数据集群[7]。再用实时获取的方式优化现有的数据抽取方法，即在各业务系统数据库中建立拥有访问数据库权限的用户，通过数据同步工具和 ETL 工具，将历史数据和增量数据同步到医院内部云集群中。整个大数据集群部署在双机架 3 备份的原则，同时服务器上的数据文件采用“谁使用谁有权限”的原则加以控制。系统使用文件层加密，保证恶意用户尝试访问数据节点直接获得文件也无法看到里面的数据，同时使用密钥管理服务分发密钥和证书，为每个组应用程序和用户设置不同的密钥，这样进一步限制攻击者获得加密密钥的难度[4]。具体的文件加密模型，见图 1。

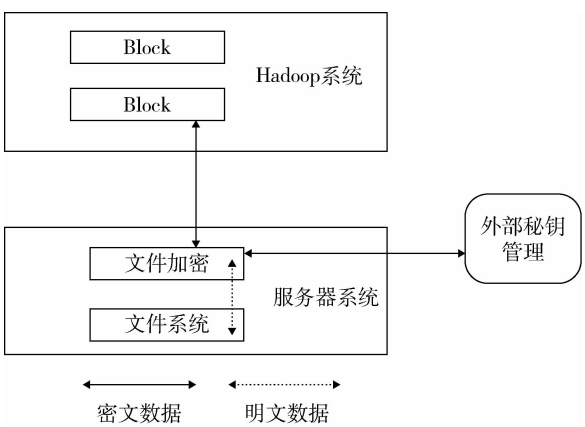

图 1 文件加密模型

6.5 网络安全

西南医院医疗大数据集群内部采取基于 Kerberos 的认证机制[4]，保证只有合法身份才能进入集群内部。同时采取数据传输中加密、安全认证隔离等保护措施。大数据集群外部采用流量控制、网闸、防火墙、VPN 等措施保证数据不外泄。具体设计中，Hadoop 平台内部采用 Kerberos 认证，该方法可以很好地融合到 Hadoop 的基础设施环境，它提供 Kerberos 可有效验证服务间通信，阻断集群中的恶意节点和应用程序。保护 Web 控制台的访问，使得管理通道难以被攻击[8]。同时大数据平台网络架构中为保证医院内部数据库安全和大数据平台中数据安全，采用如图 2 所示的网络架构。在医院内部军卫网之间部署网闸和防病毒网关，单向控制只允许军网数据定时备份到大数据集群；大数据集群上层部署网闸、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；医院同时部署 VPN 提供对外访问的出入口，并且严格控制访问授权。

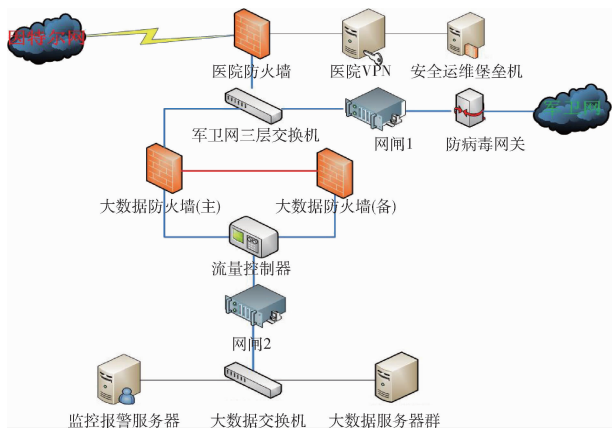

图 2 西南医院大数据平台网络结构

6.6 应用安全

通过堡垒机服务器保证用户访问权限、多级别访问控制、审计和追溯机制，所有服务器登录都必须通过堡垒机（SSH/RDP），从而构造强大的数据访问体系、账号时效性体系、完整的审查机制。具体堡垒机的访问控制界面，见图 3，通过该堡垒机可以实现完整的运维功能、报表统计、资源管理、系统配置等管理功能。通过堡垒机去访问集群服务器方便快捷，与操作人员访问本地服务器操作是一样，减少使用成本。实现所有进入内网人员的“拍照”，做到完备的审查机制。

6.7 监控安全

医疗大数据平台可从集群状态实时安全监测、安全事件管理、API 攻击防御等角度保护集群免受

非法的攻击，且采用 Ganglia + Nagios 分布式监控，实现定制化监控、告警，更好、更快维护集群可靠性，并且部署短信和邮件告警服务，以便于运维人员第一时间介入解决集群服务的问题。大数据集群管理运维界面，见图 4。

图 3 堡垒机访问控制界面

图 4 大数据集群管理运维界面

7 结语

目前，医疗大数据的实现高度依赖低成本的集群，相对而言安全机制相对滞后。要规范医疗大数据安全体系就应体现积极防御的思想。要结合互联网技术发展的方向和趋势，以技术为手段、以组织为保证、以管理为灵魂，在数据的建、管、用全生命周期中加强数据的安全保障手段，加强预案管理和应急处置工作，形成系统化精细化的医疗大数据安全保障体系。针对医院的具体情况，在以后的医疗大数据建设过程中也会继续加大对数据的保护措施，进一步规范数据使用：（1）制定医院内部和合作医院之间的数据共享机制。（2）制定更加细致的数据分级分类。（3）制定更细粒度的数据访问权限。（4）构建基于 Hadoop 的可信数据分析环境。（5）构建基于风险的数据访问控制。

参考文献

1 吕欣，韩晓露. 健全大数据安全保障体系研究［J］. 信息安全研究，2015，1（3）：211－216.

2 魏凯敏，翁健，任奎. 大数据安全保护技术综述［J］. 网络与信息安全学报，2016，2（4）：1－11.

3 陈文捷，蔡立志. 大数据安全及其评估［J］. 计算机应用与软件，2016，33（4）：34－38.

4 杨湘华，王芳，张静. 架构在 Hadop 上的安全医疗大数据生态系统［J］. 湖北中医杂志，2016，38（6）：74－75.

5 程学旗，靳小龙，杨婧，等. 大数据技术进展与发展趋势［J］. 科技导报，2016，34（14）：49－59.

6 曹珍富，董晓蕾，周俊，等. 大数据安全与隐私保护研究进展［J］. 计算机研究与发展，2016，53（10）：2137－2151.

7 汪鹏，王飞，王毅琳，等. 医疗大数据临床应用的探索与实践［J］. 中国数字医学，2016，11（9）：8－10.

8 K Ren，C Wang，Q Wang. Security Challenges for the Public Cloud［J］. IEEE Internet Computing，2012，16（1）：69－73.