医疗数据隐私保护研究综述

陈磊[①]

摘　要　随着医疗数据的迅速电子化，数据的存储变得越来越方便，与此同时，医疗数据的隐私性、安全性问题也随之而来，特别是当这些数据需要发布在网上供二次使用的时候。研究了医疗数据隐私保护的意义、研究进展以及面临的威胁，最后针对中国医疗信息化现状对医疗数据隐私保护提出了几点建议。

关键词　医疗数据　隐私保护　安全保护

Doi:10.3969/j.issn.1673-7571.2013.11.031

Overview of Medical Data Privacy Protection / CHEN Lei //China Digital Medicine.-2013 8(11): 95 to 98

Abstract　Along with the rapid digitalization of medical data (e.g. Electronic Health Records), there is an increasing concern on maintaining data privacy and security while garnering the benefits, especially when the data are required to be published for secondary use. This paper study the signification, research progress and threats of medical data privacy protection and finally some suggestions about the medical data privacy protection were put forward for the actuality of the chinese medical informationization.

Keywords　medical data, privacy protection, safety protection

Corresponding author　The 413th Hospital of PLA, Zhoushan 316000, Zhejiang Province, P.R.C.

1　引言

近年来，随着互联网以及电子病历的大量普及，医疗数据的安全以及隐私性问题变得越来越重要，因为这些数据对病人来说是极其敏感的。同时，伴随着大量的医疗数据的电子化，人们越来越将关注集中在把这些医疗数据公开发布用于更合理的使用，例如医学研究、公共健康、政府管理以及其他一些相关的卫生保健服务[1]。当前，对病人医疗数据隐私性保护的研究主要集中在几个方面：防止被授权的用户滥用、被没有授权的用户访问以及通过公开发布的病人医疗数据进行再识别等等。针对以上威胁主要采取的防护措施有：访问控制、加密技术、文件完整性检查、防火墙技术、泛化[2]或者匿名化处理[3]等。为了解决公共网络中的医疗数据隐私泄密问题，还需要对隐私攻击和威胁作全面的了解，并对其做出讨论分析。

2　医疗数据隐私保护的重要意义

2.1 体现了对患者人格和尊严的尊重　一个典型的电子病历是由一系列的标志属性（如姓名、病历号等）以及敏感属性组成。因此，在诊疗过程中，围绕患者疾病情况和诊疗行为会形成大量的关于患者个人隐私的信息，这些隐私信息由于其特殊性涉及到患者的人格和尊严，一旦泄露，会对患者的声誉及生活带来极大的影响，甚至可能引起严重的伦理道德问题。

2.2 医疗数据隐私保护具有特殊的社会价值　医疗领域中存储着大量、丰富的数据，通过进行信息整合、共享和深度分析，对促进医学研究、政府决策均具有重要意义。如何在利用好医疗信息的同时防止隐私的泄漏有着重要的意义。例如目前各地在新型农村合作医疗中普遍采取统一标准的筹资模式，但这种模式忽略了人群在收入、家庭方面的差异性，通过对电子病历中参合家庭人员组成、收入和健康情况的综合分析和评价，可以设计差异化的、更为合理的筹资标准。这样电子病历

①解放军第413医院信息科，316000，浙江省舟山市定海区文化路98号

中的信息就得到了更大限度的利用，但是在进行数据分析时往往会对参合家庭的隐私信息造成威胁。

2.3 医疗数据隐私保护具有特殊的经济价值 医疗数据中同样蕴含着巨大的商业价值。目前，医疗卫生机构的信息化建设大多采用外包给第三方企业的形式进行，以达到降低建设成本，提高信息化建设质量和效率的目的。但是采用外包的方式进行信息系统的开发和维护有可能增加隐私数据泄露的风险。

3 研究进展

目前，国内外针对医疗数据的隐私保护研究主要从法律和技术两个方面展开。

法律层面上，欧美等发达国家已建立了相对完善的政策法规体系以加强隐私保护。早在1974年，美国正式制定了《隐私权法》，被视为美国隐私保护的基本法。1996年美国国会颁布《健康保险携带和责任法案》（Health Insurance Portability andAccountability Act，HIPAA），针对医疗信息化中的交易规则、医疗服务机构的识别、从业人员的识别、医疗信息安全、医疗隐私、健康计划识别、患者识别等问题制定了详细的法律规定，以保护医疗数据安全和患者隐私权[4]。2000年，美国卫生和福利部(HHS)依据该法授权制定《个人可识别健康信息的隐私标准》，标志着美国已为保护患者医疗隐私构建起一个完整且具有可操作性的法律体系[5]。韩国在其第2个卫生信息系统10年计划（2001—2010）中，重点通过标准化和完善司法制度等基础工作加强隐私保护工作[6]。欧盟正在着手建立覆盖全欧盟范围的数字医疗体系，并对数据交换过程中安全和隐私保障问题给予了高度关注。2000年左右我国开始推广使用电子病历，但目前尚未对其规范性、有效性和法律作用等做出统一的规定。直至2010年《电子病历基本规范(试行)》、《电子病历基本架构与数据标准(试行)》、《卫生系统电子认证服务管理办法(试行)》、《病历书写基本规范》等重要的政策规范才陆续出台。但对于电子病历在医疗纠纷中的法律地位、存档管理、使用人员身份标识、使用权限分级管理等关键问题，却未提出具体的、可操作性的解决方案。

技术层面上，基于访问控制的技术。通过限制用户对各类信息资源权限管理，来防止越权使用资源，使各类数据在合法范围内使用。目前大多数研究集中在基于角色的访问控制。通过用户口令等实现登录控制，进而杜绝访问控制列表、配置文件等进行资源访问授权。Blanquer提出了利用本体进行自动授权，提高隐私保护水平和效率[7]。刘逸敏等分析了基于视图的访问控制和使用数据标签的访问控制在医疗隐私数据保护中存在的问题，并探讨解决方法[8]。基于角色的访问控制应用范围广，但灵活性较差，代价较高。针对上述这些问题，也有研究开始探讨利用规则引擎技术等来完成隐私保护。基于数据加密的技术。数据分析、处理过程中隐藏敏感数据的方法，在分布式应用环境中有着广泛的应用。在分布式环境下实现隐私保护要解决的首要问题是通讯的安全性，Maglogiannis等人提出在中间件基础上的远程加密框架，保证点对点通信中的加密[9]。同时在分布式应用中，如分布式关联规则挖掘和分布式聚类等，数据加密也起着重要的作用。关联规则挖掘的加密技术主要研究如何在统计频繁项集的过程中保护隐私信息[10]。分布式聚类的重点则在于如何计算加密后数据间的距离[11]。基于匿名化的技术，通过对数据的隐藏和泛化等操作来保护隐私。经典的匿名化技术是一种针对链接攻击的K匿名模型[12]。Mohammed等针对香港红十字会血液传输过程中设计的隐私问题进行研究，找出了影响传统匿名方法应用的主要因素。在此基础上，提出了一种基于匿名算法的LKC隐私模型来解决香港红十字会血液传输过程中的隐私保护问题。现实数据分析表明，该方法可以有效保留隐私数据中的相关信息供数据分析使用，且用于大规模的匿名数据集[13]。Alhaqbani等使用假名来替代患者真实身份，让患者能够控制自己的隐私信息，并用实例证明该架构在数据；真实性和患者隐私之间取得了较好的平衡[14]。高爱强等讨论了基于数据可用性的隐私保护匿名方法，主要讨论数据分析任务。如果对属性顺序敏感环境下的数据处理方法，基于多维数据匿名化概念，讨论了一种方法来进行基于数据可用性的数据发布共享和匿名性处理方法[15]。

4 面临的威胁

4.1 基于属性的再识别攻击 基于属性的再识别攻击[16]最初是基于对关系数据库的隐私保护。当涉及到个人隐私的数据库记录被公布时，可用于鉴别身份的字段往往被匿名化。然而，攻击者仍可以通过其他方式，如使用公开字段的组合而获取记录的身份信息。

4.2 基于网络拓扑结构的再识别攻击 基于公共网络拓扑结构的再识别（structural re-identification）[17]攻

击。公共网络中，被发布用于科研的医疗数据往往偏重于网络结构，而只包含非常少的节点属性字段，因而很难进行属性再识别攻击。然而，由于公共网络数据的性质，在发布的网络拓扑结构中也包含大量有用的信息，可以被用来识别用户身份。康奈尔大学的研究人员在文献[17]中第一次明确提出了此类问题，即尽管节点属性被删除，但是攻击者依然可以通过与节点有关的网络拓扑特性推断节点的身份。提出了主动和被动两种攻击方式。在主动攻击中，攻击者在数据发布之前故意注册一些伪账号，并且控制他们之间的朋友关系。这就相当于在公共网络G中植入了一个已知拓扑结构的子图H。同时，用这些新账号联系攻击目标，试图成为他们的朋友。当公共网络结构G发布以后，首先通过拓扑结构信息在其中找到被植入的子图H，然后利用已知的目标账号和H内节点的朋友关系，在H的邻域中识别目标账号对应的匿名节点。在被动攻击中，攻击者首先与k—1个已知用户串谋。对任意目标，利用此目标与串谋的k个用户之间的连接关系，即有机会在匿名化以后的数据里将此用户对应的节点识别出来。

4.3 基于背景知识的攻击 在实际应用中，更为常见的情况是攻击者在发布前未做任何处理，只有在网络发布后进行攻击。对发布后的网络进行攻击，由于之前没有做任何标识，所以攻击者只能依靠从其他途径获得的外部信息作为参考来对匿名的社会网络中目标节点进行识别。攻击者可获得的关于目标对象的一切信息都称为背景知识。因此，将这类攻击称为基于背景知识的攻击。

4.4 信息聚集攻击 文献[18]提出了信息聚集攻击（information aggregation attack）：在许多情况下，攻击者可以在公共网络中收集目标用户有意无意泄露的各种零碎的个人信息，尽管一小片信息看上去并无价值，但是如果攻击者将这类信息关联起来（即信息聚集），往往会导致严重后果。此外，攻击者可以在不同的社会网络中发现个人信息中明显的标志，以此为桥梁，将同一个目标用户在不同社交网站的账号联系起来。在这种情况下，同一用户的私人信息就可以从多个来源收集起来，从而实现跨网站的信息聚集攻击，严重伤害用户的隐私。

4.5 推理攻击 推理攻击（inference attack）[19-20]是研究一种通过社会关系间接泄露私人信息的方法。在很多公共网络中，用户可以选择公开或者隐藏自己的私人信息，然而，即使用户自己的信息被隐藏，有些属性仍然可以通过公开的朋友信息而被泄露。文献[20]运用贝叶斯网络来实现这种推理，他们研究影响推理准确性的因素，并建议有选择地隐藏社会关系(朋友)，那样有利于保护自己的隐私。文献[19]也研究这种混合了公开和隐藏档案的社会网络，发现不仅仅是朋友关系，用户加入的在线群或团体信息也可能被用来推断敏感的隐藏属性，而这些群的成员名单是公开的。

5 讨论与建议

近年来，各级卫生行政管理机构、医疗机构十分重视医疗信息化的建设，分分加大投入、积极探索，努力实现医疗信息化建设大跨越、大发展，目前已取得了一定的成效。其中，隐私保护是医疗信息化建设可持续发展的重要保障，必须高度重视，建立相应的组织机构，加大资金投入，明确隐私保护基本原则，建立配套的法律法规体系以及良好的组织协调机制，使之与医疗信息化建设同步、协调发展，才能最大程度发挥医疗信息系统的综合效益。

参考文献

[1] Safran C,Bloomrosen M, Hammand WE, et al,Detmer DE: Toward a national framework for the secondary use of health data:an American Medical Informatics Association white paper[J].J Am Med Inform Assoc, 2007(14):1-9.

[2] Emam K,Dankar F,Issa R,et al.A globally optimal k-anonymity method for the de-identification of health data[J].Journal of the American Medical Informatics Association: JAMIA 2009,16(5):670-682.

[3] Li N,Li T,Venkatasubramanian S: t-Closeness:privacy beyond k-anonymity and l-diversity[C].Proceedings of the 23rd International Conference on Data Engineering 2007,106-115.

[4] Bradley KJ,Melinda Cline,Carl SGuynes. HIPPA,privacy and organizational change:a challenge for management[J].Computers and Society,2007,37(1):12-17.

[5] 邢小云.美国医疗信息隐私保护立法介绍与启示[J].护理学杂志,2007(5):72-74.

[6] 中华人民共和国卫生部.基于电子病历的医院信息平台建设技术解决方案[S].2010,11.

[7] I.Blanquer V.Hem.D.Segrelles.Enhancing Privacy and Authorization Control Scalability in the Grid through On-tologies[J].IEEE Transactions on Information Technology in Biomedicine,2009,13(1): l6-24.

[8] 刘逸敏,王志勇,乔晋,等.细粒度访问控制技术在医疗数据库中的应用与展望[J].中国数字医学,2008,3(11):45-49.

[9] Maglogiannis I,Kazatzopoulos L, Delakouridis K ,et a1.Enabling Location Privacy and Medical Data Encryption in Patient Tele-monitoring Systems[J].IEEE Transactions on Information Technology in Biomedicine,2009,

13(6): 946-954.
[10] Clifton C,Kantarcioglou M,Lin X,et a1. Tools for Privacy Preserving Distributed Data Mining[J].ACM SIGK-DD Explorations, 2002,4 (2):28-34.
[11] Jagannathan G,Pillaipakkamnatt K, Wright RN.A New Privacy-preserving Distributed k-Clustering Algorithm [C].Proceedings of the 2006 SIAM International Conference on Data Mining,2006:492-496.
[12]Yan Zhu,Lin Peng.Study on K—anonymity Models of Sharing Medical Information[C].2007 International Conference on Service Systems and Service Management: 1-8.
[13] Mohammed N,Benjamin CM. Fung,Patrick CK.Hungy.Anonymizing healthcare data:a case study on the blood transfusion service[C].Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2009:1285-1293.
[14] Bandar Alhaqbani,Colin Fidge.Privacy-Preserving Electronic Health Record Linkage Using Pseudonym Identifiers[C].Proceedings of the 10th IEEE International Conference on e-Health Networking. Applications and Service,2008:108-117.
[15] 高爱强,刁麓弘.医疗数据发布中属性顺序敏感的隐私保护方法[J].软件学报,2009, 20(zk):314-320.
[16] GROSS R,ACQUISTI A.Information revelation and privacy in online social network[C].Proc of ACM Workshop on Privacy in the Electronic Society.2005:71-80.
[17] BACKSTROM L,DWORK C, KLEINBERG J.Wherefore art thour3579x: anonymized social networks,hidden patterns, and structural steganography[C].Proc of the 16th International Conference on World Wide Web.2007:181-190.
[18] LUO Be,LEE D.On protecting private information in social networks: 8proposal[C]. Proc of IEEE International Conference on Data Engineering.2009:1603-1606.
[19] ZHELEVA E,GETOOR L.To join or not to join:the illusion of privacy in social networks with mixed public and private user profiles [C]. Proc of the 18th International Conference On World Wide Web.2009:531-540.
[20] HE Jian-ming,CHU WW.Protecting private information in online social networks [C].Proc of Intelligence and Security Informatics.Berlin: Springer, 2008:249-273.

【收稿日期：2013-04-22】
【修回日期：2013-05-19】
（责任编辑：刘华）

业界观察

实现“IT新型态” 惠普工作站开启全新Z时代

本刊记者刘华报道 2013年10月10日， 2013年惠普商务IT新型态之全球工作站解决方案大会在北京隆重举行。会上，惠普响应市场变化和用户需求，推出了具备强劲性能、前沿创新、专业可靠、移动互联的全线13款Z家族工作站新品，从而引领整个工作站行业迈入一个专注顶级、专业应用、追求极致的“Z时代”，进而帮助专业计算领域的用户实现“IT新型态”，让无限创想，一站实现。

惠普全球高级副总裁、打印与信息产品集团（PPS）中国区总裁仪晓辉、惠普打印与信息产品集团（PPS）商业解决方案业务部全球产品营销总监Josh Peterson、惠普打印与信息产品集团（PPS）亚太及日本地区工作站业务部总经理Stephen KHOO、惠普打印与信息产品集团（PPS）中国区增值产品业务部总经理徐行、梦工厂动画公司技术交流与战略联盟主管Kate Swanborg 女士，以及英特尔等众多合作伙伴一同出席了此次盛会。

惠普全球高级副总裁，打印与信息产品集团中国区（PPS）总裁仪晓辉表示：“随着移动互联网、云计算、大数据等IT发展新趋势的出现，不仅企业的IT应用和管理模式发生了变化，终端用户的工作、生活和消费方式也在随之改变。信息技术的使用变得更加简便灵活、速度更快、成本更低，人们的工作和生活也越来越移动互联，呈现出一种前所未有的‘IT新型态’。”

惠普打印与信息产品集团中国区（PPS）增值产品业务部总经理徐行表示：“在专业计算领域，为了实现‘IT新型态’，许多用户已经不满足于在桌面使用工作站这种传统方式，而对资源分配、信息安全、协作、移动等方面提出了更高的要求。具备‘创新、高性能、可靠性和移动性’的惠普工作站Z家族就将满足用户的全面需求。可以说，惠普Z家族的全新亮相将引领整个工作站行业迈入一个专注顶级、专业应用、追求极致的‘Z时代’。”

在此次大会上，多达13款的Z家族新品亮相，其中包括更轻薄便携的ZBook移动工作站，支持英特尔Ivy Bridge 和Thunderbolt的Z系列台式工作站，以及采用第二代IPS面板的专业图形显示器。与此同时，现场还展示了搭载不同工作站产品的多种行业解决方案。

OBSERVATION