

利用 Cloud P2P 构建健康医疗大数据平台^{*}

邢 丹 姚俊明

(济宁医学院医学信息工程学院 日照 276826)

〔摘要〕 分析健康医疗行业集成研究现状和健康医疗大数据平台建设存在的难点, 采用 Cloud P2P 网络构建健康医疗大数据平台, 提出包含资源层、感知/接入层、传输层、服务层和应用层 5 个层次的平台框架。

〔关键词〕 Cloud P2P; 健康医疗; 大数据平台; 数据融合; 数据集成; 系统框架

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.03.004

Construction of Healthcare Big Data Platform Using Cloud P2P XING Dan, YAO Jun-ming, School of Medical Information Engineering, Jining Medical University, Rizhao 276826, China

〔Abstract〕 The paper analyzes the current situation of integration in healthcare industry and the difficulties in the construction of healthcare big data platform, proposes the construction of healthcare big data platform by Cloud P2P network, and the platform framework including the five layers of resource layer, sense/access layer, transfer layer, service layer and application layer.

〔Keywords〕 Cloud P2P; Healthcare; Big data platform; Data fusion; Data integration; System framework

1 引言

医学迅猛发展, 医疗设备更新快, 信息量日益丰富, 单纯依靠传统的医院内部系统和区域间的数据交换与通信模式已远远不能满足现代化医院发展

的步伐。医疗信息的交互与集成研究是现阶段数字化建设的迫切需要, 信息共享是实现价值最大化的途径之一。构建国家医疗卫生信息分级开放应用平台是医院信息化发展的必然趋势^[1]。为了构建跨部门、跨区域的健康医疗大数据平台, 解决传统医学信息系统需要独立的物理计算、存储及网络资源运行的问题, 降低传统云存储服务带来的数据中心风险性, 充分利用和拓展完善现有设施资源, 将移动医疗等规范接入系统中, 本文在分析了健康医疗行业集成研究现状和健康医疗大数据平台建设存在难点的基础上, 提出采用 Cloud P2P 网络构建健康医疗大数据平台, 最终给出健康医疗大数据平台的框架。

2 相关研究概述

〔修回日期〕 2016-12-16

〔作者简介〕 邢丹, 硕士, 讲师, 发表论文 6 篇; 姚俊明, 讲师。

〔基金项目〕 济宁医学院青年基金“基于 P2P 的物联网资源发现机制的研究”; 山东省自然科学基金(项目编号: ZR2011HL002); 济宁医学院校级科研项目(项目编号: JY2013KJ036); 济宁医学院校级教育科学研究课题(项目编号: 16032); 济宁医学院教育信息化试点课题(项目编号: 35)。

2.1 健康医疗行业集成研究现状

2.1.1 系统集成 健康医疗行业系统集成目前多采用联邦式、基于中间件和数据仓库等方法。医疗信息系统的集成方案通常是在医疗机构的子系统之间直接通过数据库互相访问,由于耦合度大,缺乏独立性和通用性,难以推广;基于 HL7 和 DICOM 等通讯标准的系统,具体实现上的差异性和标准的灵活性,无法实现方便的集成^[2];采用工作流、通过引入独立的标准消息中间件与各信息系统分别集成,或采用虚拟数据中心和 Web 服务中间件集成各信息系统分散的数据^[3],这两种集成方式由于缺乏标准支持,也难以实施和推广。

2.1.2 数据集成 目前有模式集成、数据复制方法以及两种方法的综合,研究主要集中在解决数据的语义异构和结构异构问题^[3]。早期数据库的本质是先有模式后有数据,因此数据集成采用全局视图和局部视图中介模式的方法,以自顶向下的方式实现集成^[4]。

2.2 云、P2P 及其在健康医疗领域研究现状

将分布式存储和虚拟化技术结合的云存储为大数据存储提供了有效的解决方案,针对海量数据存储问题,开发了将对等(P2P)技术与云存储结合的系统^[5]。由于构建于结构化 P2P 拓扑上的可扩展云能够分担存储负载,因此可避免中心服务器造成的性能瓶颈问题^[6-8]。云医疗服务模式突破了医院有限的物理空间,可实现慢性病的日常监测^[9]。随着移动互联网、社交网络的发展,用户可以通过终端设备便捷地获取云端服务^[10-12]。为解决传统的集成方式导致集成平台难以维护的问题,也有研究采用分布式计算解决收集、存储和分析上的难题^[10]。

3 基于 Cloud P2P 网络构建健康医疗大数据平台

3.1 健康医疗领域大数据平台问题分析

3.1.1 异构性 原有健康医疗领域的软件采用不

同公司、不同厂商的产品,基于不同的软硬件平台,在底层技术和业务流程上存在广泛的异构性,各部门内及区域之间的基础设施组建方式不同,系统之间的异构性导致形成信息孤岛,使得不同的用户群体之间实现数据共享非常困难。

3.1.2 数据类型多样、标准各异 数据来源广泛、内容多样,既涵盖结构化数据,又具有图形、图像、视频等非结构化数据,具有海量性、异构性等特点。采用的标准各异,还需将可穿戴设备、智能健康电子产品、健康医疗移动应用等产生的数据资源规范接入平台。

3.2 基于云的健康医疗大数据平台

3.2.1 云计算技术 不同的用户需要独立的客户端设备、操作系统、应用程序和配置文件访问不同信息系统,在健康医疗大数据平台中采用云技术帮助降低基础设施建设和运维成本,能够保证提供更好的系统性能、安全性和用户体验。同时,健康医疗行业大数据的增长使传统的关系型数据库在扩展性、容错性等方面遇到了瓶颈,而云计算技术依靠其高扩展性、高可用性、容错性等特点,成为大规模数据管理的有效方案。

3.2.2 将 P2P 技术作为补充加入云存储结构中 医疗机构在内部利用云计算建立私有云,通过公共云来获得需要的信息技术服务,这样可以充分利用云计算资金投入少、风险低、即买即用等优势^[13]。但是也存在不足之处,基础资源由第 3 方运营商保管,医疗机构无法控制平台服务的稳定性和可靠性;同时,由服务器集群所组成的云存储增加了数据中心的危险性,一方面系统搭建费用和管理成本都集中在运营商,另一方面断电、自然灾害及人为攻击等因素增加了数据中心的危险性。这种云存储服务在遇到灾难性事件后很可能造成数据丢失,进而导致严重的后果。为此,一些研究使用 P2P 技术作为一种补充加入到云存储结构中^[14-15]。P2P 网络强调去中心化,弱化了服务器的功能,将闲散的用户存储空间以 P2P 的形式组织起来,构成节点存储层,加入到由服务商提供的云存储服务器集群中,形成一个 Cloud P2P 云存储结构,见图 1。

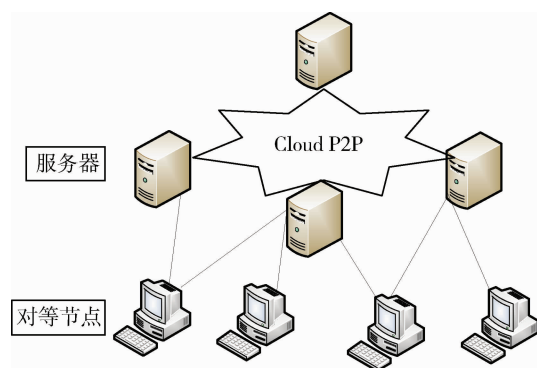


图 1 Cloud P2P 网络结构

3.3 基于 Cloud P2P 的健康医疗大数据平台

3.3.1 Cloud P2P 模式 既是一种计算模式，又是一种存储方案。计算模式是将网络核心的云数据中心和网络边缘的终端节点上的各类资源有机聚合成更大规模的资源池，以供上层应用调用。云端数据存储系统中蕴含的节点及可用资源范围更广、数量更大，更具可伸缩性，更易满足用户日益增加的存储需求，能消除性能瓶颈，实现负载均衡和多副本冗余备份，降低数据中心能耗和成本以及避免集中数据存储带来的安全风险。Cloud P2P 云存储结构包括节点存储层和云中心两部分。节点存储层主要由运行在用户节点上的客户端程序和用户的共享存储空间组成，当云中心备份服务器发生故障时，节点存储层所提供的副本备份仍可以满足用户使用。云中心由服务提供商的服务器集群组成。采用 Cloud P2P 技术，这类网络成本低、容灾性好、可扩展性强，且具有较强的负载均衡能力^[16]。

3.3.2 Cloud P2P 健康医疗网络架构 为保证健康医疗大数据平台性能的稳定性，由面向运营商的 Cloud P2P 网络架构根据节点的性能，将传统云端服务器作为超级节点，将原有健康医疗基础设施中的高性能机器作为备份节点，目的是当故障发生时由备份节点代替超级节点，普通节点由用户计算机组成。基于 Cloud P2P 健康医疗网络架构，见图 2。参考移动互联网的分层^[17]，将医疗云 Cloud P2P 业务平台分为 Cloud P2P 核心服务层和面向 Cloud P2P 应用的特殊服务层两层。Cloud P2P 网络中的资源通常来自于服务器集群和对等节点，核心服务层提供基础功能，如具有服务与资源目录、管理模块、调度模块以及安全保障机制和虚拟化机制。其中管

理模块提供管理和服务，主要负责管理可用的计算、存储、信息资源和服务，如节点启动引导和基于 DHT 协议的资源查找功能，配合安全机制对节点的身份和角色进行认、授权和管理；调度模块负责接收用户提交的服务请求，根据用户请求动态地部署、配置和回收资源。为满足 Cloud P2P 运营商提供电信级服务的需求，为运营商提供网络监控、可靠性管理等功能，Cloud P2P 核心服务层引入了监控模块，具有两个基本功能：（1）监视网络状况。由运营商控制的 Peer 节点或监控服务器监测网络状况，如系统资源的使用情况，处理异常，实现节点配置、负载均衡和资源监控。（2）网络控制功能。当网络抖动过于剧烈时，向网络中增加一定数量的稳定节点或改变 Peer 节点的备份策略。动态适配模块根据底层网络的状况，动态地改变 P2P 协议的参数或 P2P 算法适应网络的变化，如在网络抖动剧烈时，用非结构化网络拓扑替换 DHT 拓扑获得更好的性能。可靠性模块为提供电信级高可靠性业务，意味着 Peer 节点可以在任何时候获得需要的内容，而 Cloud P2P 网络在抖动性高的情况下很难满足上述要求，这使得系统需要提供相应的备份机制，这些备份机制充分考虑节点的在线时间、节点失效的概率等因素。目前研究提出采用网络层析机制实现对全局流量和网络拓扑结构变化的分析^[18]，针对节点的频繁加入和离开问题，也是防范攻击的重要方法，主要有拓扑推测、链路损失推测和链路时延推测 3 类。

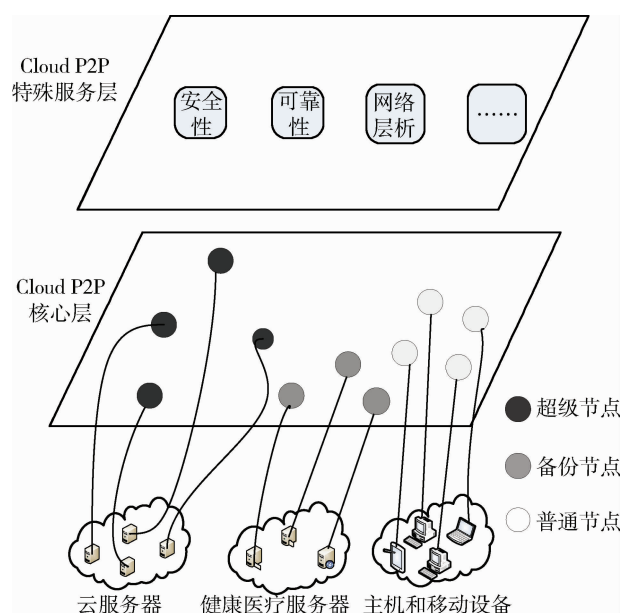


图 2 健康医疗 Cloud P2P 网络体系架构

面向 Cloud P2P 应用的特殊服务层提供数据管理、信誉度管理、激励机制、数据版权管理等功能,为在该平台上开发新的 Cloud P2P 应用提供相应的开发接口。

4 健康医疗大数据平台框架

4.1 框架图

健康医疗大数据平台自底向上包含 5 个层次,即资源层、感知/接入层、传输层、服务层和应用层,见图 3。健康医疗行业相关的标准、规范和安全机制贯穿所有层次,建立统一的疾病诊断编码、临床医学术语、检查检验规范、药品应用编码、信息数据接口和传输协议等相关标准,促进健康医疗大数据产品、服务流程标准化。

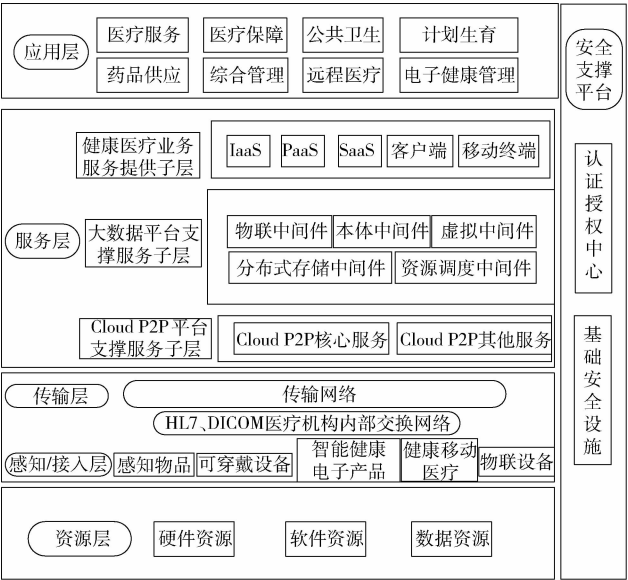


图 3 健康医疗大数据平台框架

4.2 资源层

为大数据平台的运行基础,提供包括存储服务器、计算服务器等高性能硬件,结构化和非结构化的数据、平台和模型/服务等软件资源^[19]。

4.3 感知/接入层

通过各类感知设备获取可穿戴设备、智能健康电子产品、健康医疗移动应用等产生的大数据,通过无线传感网络、4G-LTE 等途径收集资源、规范

接入平台。

4.4 传输层

为能使原有系统“无缝”接入大数据平台,医疗机构内部系统的集成依然采用传统的 HL7 和 DICOM 标准和协议,外网通过传输层传输。

4.5 服务层

Cloud P2P 平台支撑子层包括核心服务和其他服务。核心服务如拓扑结构算法、查找算法、索引技术、节点复制策略等;大数据平台支撑服务子层涵盖资源调度中间件、分布式存储中间件(如 Hadoop 等)、物联中间件、本体中间件、虚拟中间件。资源调度中间件提供负载均衡、容错调度等支撑服务,对底层资源进行全面整合和集中管理,为各类行业应用提供资源支持。本体中间件通过采用本体建模规范行业语义,构建健康领域模型,使用推理验证完成本体推理;健康医疗业务服务提供子层使用传统的客户端、移动终端等提供 IaaS、PaaS、SaaS 服务。

4.6 应用层

通过实现医疗服务、医疗保障、公共卫生、计划生育、药品供应、综合管理等应用,完成信息系统数据采集、集成共享和业务协同,提供远程会诊、远程影像、远程病理、远程心电诊断服务;采用居民电子健康档案应用,推动预防、治疗、康复和健康管理覆盖全生命周期的一体化电子健康服务。

5 结语

基于 Cloud P2P 的健康医疗大数据平台旨在解决传统医疗系统集成问题,充分利用和完善现有医疗信息化基础设施,整合多源、异构平台。该技术的使用避免了完全由运营商维护云平台的安全风险,构建了健康医疗大数据平台框架的功能结构。该平台具有高度的灵活性、可扩展性,具备按需扩展的存储资源与计算资源的支撑、低成本投入及维护的保障。全面建成互通共享的医疗,为健康医疗信息分级开放应用平台的数据互联融合、开放共

享, 消除信息孤岛起到借鉴的作用。

参考文献

- 1 国务院办公厅. 国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见 [Z]. 2016-06-24.
- 2 万歆, 姚晴虹. 基于异构系统的数据集成平台的搭建和应用 [J]. 医疗卫生装备, 2016, 37 (2): 61-62.
- 3 苏琪, 刘西林, 王军. 基于 Web Service 的数据集成研究及应用 [J]. 计算机技术与发展, 2014, 24 (8): 51-54.
- 4 孟小峰, 杜治娟. 大数据融合研究: 问题与挑战 [J]. 计算机研究与发展, 2016, 53 (2): 231-246.
- 5 Xu H, Shi Y, Liu Y, et al. Integration of Cloud Computing and P2P: a future storage infrastructure [C]. 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012: 1489-1492.
- 6 徐小龙, 龚培培, 章韵, 等. 一种基于移动 Agent 的云端 (Cloud-P2P) 数据复合销毁机制 [J]. 计算机科学, 2015, 42 (10): 138-179.
- 7 Chen G, Hu T, Jiang D, et al. BestPeer++: a Peer-to-Peer based large-scale dataprocessing platform [C]. Washington DC: Proceedings of the IEEE 28th International Conference on Data Engineering (ICDE), 2012: 582-593.
- 8 Zhao P, Huang T, Liu C, et al. Research of P2P Architecture Based on Cloud Computing [C]. 2010 International Conference on Intelligent Computing and Integrated Systems (ICISS), 2010.
- 9 廖湘庆. 基于云平台的公立医院服务模式创新研究 [D]. 武汉: 华中科技大学, 2015.
- 10 姚琴. 面向医疗大数据处理的医疗云关键技术研究 [D]. 杭州: 浙江大学, 2015.
- 11 徐小龙, 李硕, 顾厉森, 等. 基于 Cloud-P2P 网络平台的泛知识云模型 [J]. 计算机应用研究, 2014, 31 (2): 408-409.
- 12 王伟阳. 基于云框架的移动医疗信息化服务平台的研究与实现 [D]. 青岛: 青岛科技大学, 2014.
- 13 秦鹏. Peer-to-Peer Cloud 关键技术研究 [D]. 武汉: 华中科技大学, 2014.
- 14 Rahman SMM, Masud MM, Hossain MA. Privacy Preserving Secure Data Exchange in Mobile P2P Cloud Healthcare Environment [J]. Peer-To-Peer Networking and Applications, 2016, 9 (5): 894-909.
- 15 Rajkumar R, Iyengar NCSN. Dynamic Integration of Mobile JXTA with Cloud Computing for Emergency Rural Public Health Care [J]. Osong Public Health and Research Perspectives, 2013, 4 (5): 255-264.
- 16 金顺福, 王晨飞, 陈玲玲, 等. Cloud-P2P 云存储结构的模型建立与性能分析 [J]. 通信学报, 2015, 36 (3): 3-4.
- 17 李伟, 徐正全, 杨铸. 应用于移动互联网的 Peer-to-Peer 关键技术 [J]. 软件学报, 2009, 20 (8): 2202-2203.
- 18 林文敏. 云环境下大数据服务及其关键技术研究 [D]. 南京: 南京大学, 2015.
- 19 宫夏屹, 李伯虎, 柴旭东, 等. 大数据平台技术综述 [J]. 系统仿真学报, 2014, 26 (3): 492-493.

关于《医学信息学杂志》启用 “科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力, 端正学风, 维护作者权益, 《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”, 对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库, 可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为, 本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度, 自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部