文章编号:1000-5404(2005)11-1123-04

论著

# 医院影像归档与存储系统中影像数据长期存储问题的研究

刘仲明[1],王 放[2],郑小林[3] (第三军医大学:[1]科研部,[2]西南医院信息科,重庆 400038;[3]重庆大学生物工程学院,重庆 400044)

**提 要:目的** 解决海量医学影像数据在医院PACS网络中的长期安全存储问题。**方法** 将医院DICOM3.0和非DICOM3.0影像检查设备采集到的不同医学影像转换成影像数据的存储格式、元数据的标识、表示方式等由DICOM3.0标准规定的数字影像,并按照影像数据的使用与存储要求,依据性价比最优的原则,采取在线、近线和离线三级存储的方式,合理配置分级存储设备容量,通过编簇,科学分配影像数据在PACS网络中合理流动。**结果** 本方法构建的医院PACS存储系统,既能满足海量影像数据长期存储与管理,同时具有超强容错能力,实现永久无限制存储,检索查询迅速,存取快速、方便、可靠。**结论** 分级存储是实现PACS海量医学影像数据长期安全存储的有效方法。

**关键词**:PACS;数据存储;医学影像
**中图法分类号**:R197.324;R319 **文献标识码**:A

# Long-term storage of imaging data in hospital picture archiving communication system

LIU Zhong-ming[1], WANG Fang[2], ZHENG Xiao-lin[3] ([1]Division of Scientific Research Affairs, [2]Department of Informatics, Southwest Hospital, Third Military Medical University, Chongqing 400038; [3]College of Bioengineering, Chongqing University, Chongqing 400044, China)

**Abstract**: **Objective** To find a solution to the long-term storage of enormous amount of imaging data in hospital picture archiving communication system (PACS). **Methods** Medical imaging data collected by devices of Digital Imaging and Communications in Medicine-3 (DICOM3) and non-DICOM3 were transformed into the storage format, metadata instance, and expression pattern of digital images standardized by DICOM3. Based on the requirements of data utilization and storage system and the optimal principle of function and value, the reasonable flow of imaging data in PACS network were scientifically allocated by application of on-line, near-line, and off-line storage patterns and by reasonable allocation of the volume and clusters of storage devices. **Results** The PACS storage system established by our method, characterized by super capacity of error tolerance, permanent unrestricted storage, quick in retrieval, and quick, convenient and reliable in storage and reading, could meet the requirements of long-term storage and management of enormous amount of medical imaging data. **Conclusion** Grading storage is the effective method for long-term storage of enormous amount of medical imaging data in PACS.

**Key words**: picture archiving communication system; data storage; medical image

自1986年布鲁塞尔第15届国际放射学术会议首次提出数字化X线成像技术的物理概念及其临床应用以来,揭开了传统X线检查数字化的序幕。随着电子信息技术、传感器技术、光机电一体化及材料工业的发展,面向医院的影像归档与存储系统(picture archiving communication system, PACS)得以迅速发展,备受医院关注。在实际应用中,影像数据的长期保存问题一直是PACS系统必须面对的核心与难点问题。由于医学影像数据量大,要求保存的时间长:1张正位的胸部X线片约28兆字节,通常1个1 500张床位的大型综合性医院,1 d产生的影像数据量约7 G,1年约2.4 T;按相关规定,医学影像保存时间必须大于15年,若按20年计算,则需48 T的存储空间。同时,PACS系统是一个7×24 h不停机系统,对系统的适用性、可靠性和安全性都提出了较高要求。

海量数据的长期存储面临的系列问题:①数字信息技术的变化,将导致历史系统和当前系统的不一致,在编码方式、存储格式、检查设备进步的同时,原有数据在新技术下的存取、显示和转换都是需要考虑与解决的问题;②磁介质和光介质的不稳定性,易造成部分数据的丢失;③人为的破坏、火灾、地震等因素,对数据的长期存储威胁较大;④数字信息的庞大,对备份、管

作者简介:刘仲明(1966-),男,重庆市云阳县人,硕士研究生,助理研究员,主要从事计算机在医学中的应用方面的研究。电话:(023)68752180
通讯作者:郑小林,电话:(023)65103578
收稿日期:2005-04-07;修回日期:2005-04-29

理都是一种巨大压力。

## 1 材料与方法

### 1.1 PACS系统的组成与结构

一个典型的PACS包含3个主要部分:图像采集、图像存储与管理、图像显示与处理,见图1[1]。它以高速计算机为基础,旨在全面解决医学图像的获取、显示、存储、传输和管理,用高速网络链接各种影像设备和相关科室,利用大容量磁、光存储技术,以数字化的方式存储、管理、传送和显示医学影像及其相关信息,具有影像质量高,存储、传输无失真、传送迅速以及图像资料可共享等特点,是近年来国内外新兴的医学影像信息技术,是一项综合了计算机、通信、医学影像学等多项技术的集成,是实现医院影像信息管理(RIS)现代化的重要组成部分[2],也是医院无胶片诊断需求的不断实现和具体化。在医院医疗、教学和科研上有着广泛的应用前景,对于降低医疗成本、提高远程医疗普及率具有重要的意义。

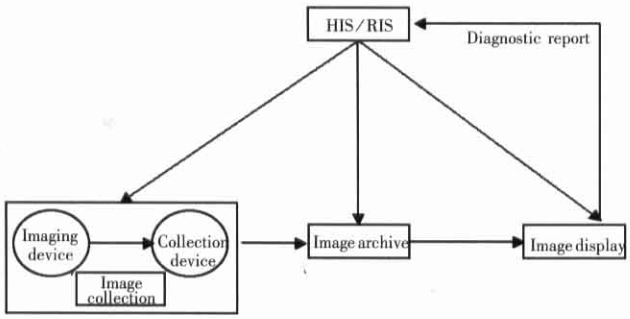



**图1 PACS的组成部分**

**Fig 1 Components of PACS**

### 1.2 影像数据来源

从信息源数据形态来看,可分为:①原有X线片、CT胶片等介质形态扫描转换为数字影像;②直接由CR、DR、CT、DSA等影像检查设备产生;③由胃镜、肠镜、腹腔镜、宫腔镜、喉镜、纤支镜等内窥镜及显微镜、B超等检查设备产生的视频模拟信号影像转换为数字影像;④其他数据格式的数字影像转换而来。

### 1.3 影像数据存储格式

影像数据的存储格式、元数据的标识、表示方式等由DICOM3.0(digital imaging and communications in medicine, DICOM)[3,4]标准规定。在影像文件中,保存了影像的产生日期(image date)、时间(image time)、设备类型(modality)、设备厂家(manufacturer)、病人姓名(patient's name)、病人ID号(patient ID)、出生日期(date of birth)、性别(sex)、检查惟一标识(study instance UID)、序列标识(series instance UID)等,利用这些信息建立影像管理数据库对影像数据进行管理。影像数据一旦形成就不会再改变,对影像的标注、解释等可通过另外保存数据实现。

### 1.4 影像数据的使用与存储系统要求

1.4.1 影像数据的使用要求 一方面,在影像数据的使用频度和存取速度要求上,与影像数据产生的时间密切相关:①正在做检查的门诊病人影像数据,要求立即传送到诊断工作站和相关的门诊工作站;②在院病人的影像,医生会频繁调用、查看,要求同一个病人的影像能够在多个场所同时调用并能快速存取调用影像数据;③出院后病人的影像,可能会回访调用;④出院后10年以上的病人影像偶有调用。另一方面,医生对影像有如下4方面的需求:①在门诊,自己诊治的病人检查影像,要求诊治时能立即观看;②在病房,病人本次住院的历次检查影像要求诊治时能立即观看;③在中心读片室,当天的检查影像要求写诊断报告时能立即观看;④在科研、教学工作中,按要求检索出全院所有相关检查影像资料。

1.4.2 影像数据存储系统要求 ①高效。通过在网络上分布存储当前病人的影像数据,满足病人诊治对影像数据的要求;通过控制磁带机实现磁头快速定位和数据高速读写;实现数据迁移和回迁。②开放。系统中迁移策略、任务调度和数据迁移模块与硬件设备相对独立。在硬件的支持下,系统可以挂接各种主流磁带库和光盘库。③扩展性强。影像文件的共享节点、分发服务器、存储设备可以动态扩展。④稳定性高。归档服务器由分中心服务器、中心服务器组成,中心服务器采用双机热备方式。⑤高性价比。既能满足海量影像数据的长期安全存储,同时,相比进口国外的存储设备和存储管理软件具有优越的性价比。⑥数据保护功能。通过CRC长校验保证数据写入安全;通过在线扩充,降低近线磁带的使用次数;具有介质重建功能。⑦高度灵活性。多重推送和预取机制保证系统规模可以根据用户的需要灵活组织。扩充系统规模时,无需改动现有软硬件结构。且不会影响系统正常工作。

### 1.5 影像数据长期保存方法

大存储容量的获得是影像数据长期保存的基础。磁盘阵列、磁带库、光盘库都能提供几十TB的存储空间。基于光接口的磁盘阵列可达2 GB/S的带宽,I/O次数可达10万次;磁带库中,LTO格式的磁带驱动器,其数据带宽为10~20 MB/S,磁带介质使用寿命可达30年,1 GB空间约4元;从性价比和可靠性等方面考虑,分级存储是实现PACS影像数据长期保存的有效方法。

1.5.1 分级存储管理策略 当前病人的影像数据,它侧重的是共享方式下的使用,要求高的传输速度和便捷的显示切换;而出院病人的影像数据,它侧重于保存和再利用,以大数据量为特征。

影像特征及属性的管理必须依托数据库进行,方便影像的检索和查询;通过特征(属性)与存储管理系统的有效关联,完成对影像数据的分级存储与管理。影像文件除具有DICOM3.0存储格式定义的属性外,另外增加关于存储/归档属性,如影像文件收到日期/时间、在线存放路径、近线对象标识、近线归档日期/时间、离线对象标识、离线归档日期/时间等。存储管理系统负责将影像文件记录在近线(离线)介质上,并负责对介质进行加载管理。

病人一旦在门诊挂号或办理入院,该病人近15年的历次检查影像从近线存储设备中迁移到在线存储设备中,并推送到相应的PACS工作站;影像文件一旦产生,在在线存储设备中保存1份;同时,通过调度服务将影像文件分发到PACS网络的其它工作站。在每日0:00时开始将前1 d产生的影像数据在近线存储设备上保存1份,保存期限为15年;在每月第10个工作

日将上 1 个月的影像数据在离线存储介质上保存 1 份，保存期为无限期。

对在线存储的影像进行清除时，首先判断该病人是否在院，然后判断影像生成时间是否超过保存期；对于既不在院，又超过保存期的病人影像进行批量删除。对于近线存储的影像，超过保存期即删除。

1.5.2　分级存储设备

1.5.2.1　定义　　从存取速度、存取方式、是否需要人工干预等指标上，可将存储设备划分为 3 级：在线存储设备、近线存储设备和离线存储设备。

在线存储设备：存储介质以磁盘阵列和磁盘为代表，具有读取速度快，吞吐量大等优点，这类存储介质一直处于在线状态，随时可以存取。

近线存储设备：存储介质以磁带库和光盘塔为代表，它有其共同特点：第一，在任一时刻，设备中只有少数几个存储介质处于在线状态，而大多数存储介质是离线的，当需要的数据在这些离线的存储介质中时，需要把这些存储介质加载到驱动器里，与在线存储相比这往往需要较长的时间。第二，对数据的定位以及读取时间也比磁盘要长，磁带库更是如此。第三，设备中的离线存储介质，通过控制机械手自动加载到驱动器中。

离线存储设备：当磁带或光盘离开磁带库或光盘塔后，对磁带或光盘都只能进行手工管理，当数据存储在这些介质上，需要手工将所需磁带或光盘加入磁带库或光盘中，数据才能读、写。

1.5.2.2　连接方式(存储结构)　　存储设备从连接方式可分为：以服务器为中心的直连存储(direct attached storage，DAS)、以数据为中心的附网存储(network attachment storage，NAS)、以网络为中心的存储区域网(storage area network，SAN)。

DAS 被定义为直接连接在各种服务器或客户端扩展接口下的数据存储设备。它完全以服务器为中心，寄生在相应服务器或客户端上，其本身是硬件的堆叠，不带有任何存储操作系统。NAS 的概念是来自文件服务器(file Server)[5]，就是替客户端(client)提供文件管理的服务器(client/server 的架构)。NAS 设备为网络用户提供独立的存储空间，对存储对象的管理可以到文件级，依用户可以设置对文件或目录不同的存取权限。

SAN 是一个高速的光纤网络[6]，以铜轴线缆、单模或多模光纤作为传输介质，传输距离最大可达 10 公里，传输速度可达 2 Gb/S；SAN 存储网络以 SAN 交换机(集线器)为核心，可以连接带光纤连接接口的磁带库、磁盘阵列、服务器等组成一个 SAN，非光纤连接的存储设备(如基于 SCSI 的存储设备等)可以通过网桥、路由器等互联设备连接到 SAN 中。目前占主导优势的 DAS 架构将逐渐为性能更加优越的 SAN 架构所代替。而 DAS 架构的设备通过接口转换和增加交换机等网络设备可以方便地升级为 SAN 架构。

1.5.2.3　性价比　　在线存储：存取的速度最快，但所提供的存储空间有限，每 GB 的存储价格最高；近线存储：存取的速度较快，所提供的存储空间较大，每 GB 的存储价格介于两者之间；离线存储：存取的速度慢，自动化程度不高，需要人工干预，但所提供的存储空间无限，每 GB 的存储价格最低。因此，要获得优良的性价比，就必须合理配置在线、近线、离线设备，即在线存储设备用于存储正在使用的影像数据，近线存储设备用于存储经常使用的影像数据，离线存储设备用于存储归档的影像数据。

1.5.3　影像数据在分级存储设备上的存储

1.5.3.1　影像数据在 PACS 中的流动　　在 PACS 中，定义分发服务器和它所管辖的影像工作站组成一个簇，簇内工作站上的影像数据都相同。其数据流向示意图见图 2。从使用角度看，同一临床科室的工作站编入一个簇，医生可以在任一台机器上完成查看影像和诊断报告；同一门诊属性的诊室编入一个簇，医生可以互换诊室和操作台，方便了上岗医生的安排和医生的合理调配。作为特例，各分中心服务器、中心服务器和各部门的诊断工站组成一簇，由分中心服务器担当分发服务器。

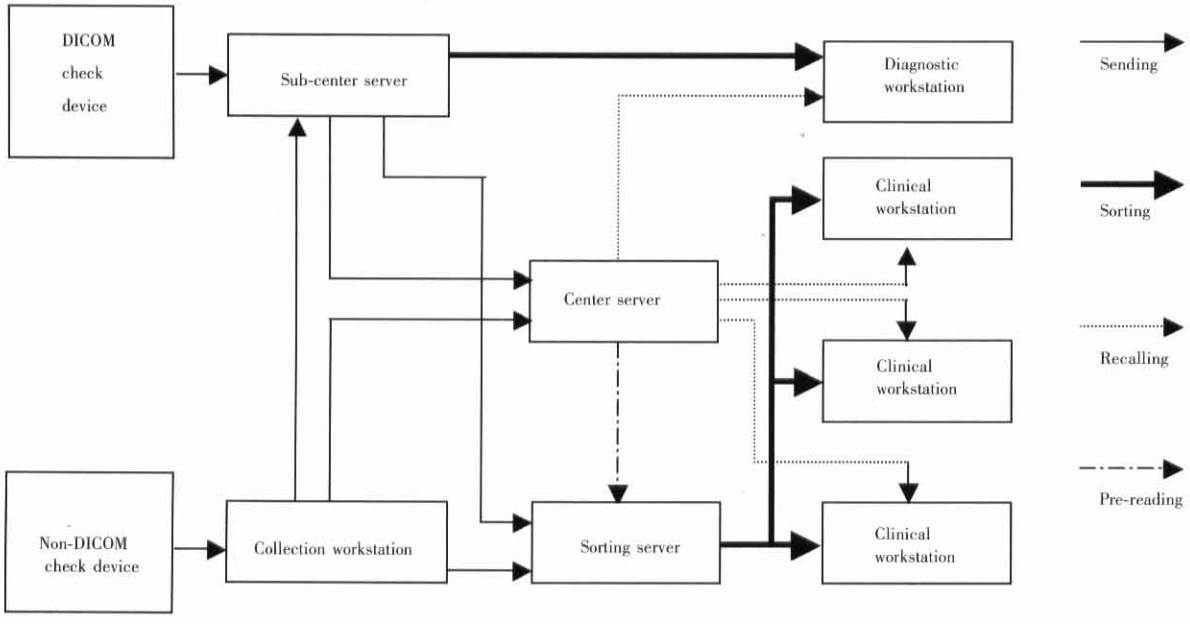



**图 2　PACS 网络中数据流向示意图**

**Fig 2　Schematic diagram of dataflow in PACS network**

分发由分发服务器完成,它将收到的影像发送给该服务器所管辖的其它影像工作站。

推送由分中心服务器完成,它将收到的影像发送到由该病人检查申请的开单科室决定的分发服务器上。一个检查影像只需推送到一个分发服务器上,具有过滤作用,减少了数据在主干网上的流量。

预取由中心服务器完成,病人入科时触发中心服务器的预取过程。中心服务器在归档数据库中按 ID 号查出历次检查记录,将相应检查影像从磁带库中检索出来,并发送给病人所在科室的分发服务器。

调用由各影像工作站提出请求,中心服务器完成在影像库中查得影像并传回给提出请求的影像工作站。中心服务器首先在磁盘阵列查找相应检查记录的 UID 影像文件;如果不存在,则在归档数据库中查找相应的影像归档记录,并从磁带库中检索出影像文件放入磁盘阵列;然后发送给提出请求的影像工作站。

1.5.3.2 存储方法 医院影像检查设备在对病人进行辅助检查时产生影像数据,影像数据具有逐日生成的特点,对影像数据的使用、管理也与时间密切相关。因此,同一天的影像数据文件存储在同一个目录下较好管理。即在阵列上,以 image 为根目录存储影像和标识文件;以年月日为二级目录,在该目录下存储该日生成的影像文件和标识文件管理影像文件的数据库存于 db 目录下,其路径为:Program Files/Medvision/db。在磁带上,影像文件用 LTO 格式的磁盘保存,管理磁带上影像文件的存放信息的数据库也存于 db 目录下。同时,为了保证数据的安全,将 db 的备份存于指定的磁带上。

1.5.3.3 影像数据回迁 在线存储设备是确保 PACS 系统运行的关键设备,在线数据的缺失对医疗工作的影响最大,快速恢复在线数据是 PACS 系统必须要考虑的问题。

一旦在线存储设备恢复正常,从分中心服务器中恢复近期影像数据,其他影像数据从近线存储中恢复;近线数据的缺失,可以从在线存储设备中恢复部分数据,其他影像数据从离线存储中恢复。

医学影像特征及属性数据库、存储管理数据库的保护和重建已是比较成熟的技术了[3,4,7],就不再赘述。

## 2 结果

通过合理配置各级存储设备容量构建的医院 PACS 在线、近线和离线三级存储系统,是"从临床中来到临床中去"的产品,它不仅综合了临床医生的实践经验和对疾病检查治疗的要求,而且满足海量影像数据长期存储与管理,同时具有超强容错能力,解决了影像数据的恢复问题,确保影像数据在在线、近线存储上恢复,不存在不能恢复的影像数据。实现永久无限制存储,检索查询迅速,存取快速、方便、可靠,性价比优越。以此方法建立的 PACS 存储系统已经在西南医院稳定运行近 2 年,收到了良好的经济和社会效益。

## 3 讨论

医学影像是医学临床诊断中应用最普遍、最重要的诊断依据之一,也是 PACS 系统在临床医学领域中得以迅速发展的原因。我国 PACS 技术的研制和发展进入实质性阶段是近几年才开始的,但真正能做到图像信息共享和异地实时在线诊断的医院还为数不多,尤其是海量数据安全存储问题,一直是困扰 PACS 快速发展的制约瓶颈,现在有一些医院购买了一些国外的存储设备和存储管理软件,但这些软件片面追求大而全,造成功能上的浪费,而且价格昂贵,限制了其广泛应用。本研究采取在线、近线和离线三级分级存储方式,由于三级存储设备容量配置科学合理,该存储系统既能满足海量影像数据的长期安全存储,又比进口国外的存储设备和存储管理软件价格低得多,相比之下,具有优越的性价比。总之,医院存储架构要纳入医院信息系统建设体系进行总体规划。根据存储界摩尔定律,单位数据存储价格降低速度很快,性能容量增长速度则更翻倍,因此存储系统建设也要在总体规划前提下分步购买,购买原则是满足需要,适度超前,切不可盲目攀比造成投资浪费。所以结构的开放、可扩展性就尤为重要。

## 参考文献:

[1] 胡阳秋, 高小榕, 高上凯. DICOM 医学图像采集的方法与实践[J]. 北京生物医学工程, 2001, 20(1): 1-5.

[2] 何友全, 方 磊. 医学影像计算机传输与处理[J]. 西安交通大学学报, 2003, 38(3): 267-270.

[3] 李桂祥, 刘运定, 李刚荣. 基于 ORACLE 数据库的医院信息系统正常运行的保障方法[J]. 医疗设备信息, 2004, 19(2): 10-11, 56.

[4] 杨铁林. Oracle 数据库中 Standby Database 的搭建和使用[J]. 计算机与现代化, 2003, (4): 4-6.

[5] 冯 丹, 袁圣权, 张江陵. 磁盘阵列附网存储技术的研究[J]. 计算机工程, 2002, 28(10): 172-173.

[6] 蔡皖东. 基于 SAN 的高可用性网络存储解决方案[J]. 小型微型计算机系统, 2001, 22(3): 284-287.

[7] 于鸿飞, 黄 昊, 柯新华, 等. 医院信息系统备份与恢复方案的实现方法[J]. 医疗设备信息, 2003, 18(4): 20-21.

(编辑 汪勤俭)