Journal of Digital Imaging

Validating DICOM Content in a Remote Storage Model

Pattanasak Mongkolwat, PhD, Pankit Bhalodia, MS, James A. Gehl, BBA, and David S. Channin, MD

Verifying the integrity of DICOM files transmitted between separate archives (eg, storage service providers, network attached storage, or storage area networks) is of critical importance. The software application described in this article retrieves a specified number of DICOM studies from two different DICOM storage applications; the primary picture archiving and communication system (PACS) and an off-site long-term archive. The system includes a query/retrieve (Q/R) module, storage service class provider (SCP), a DICOM comparison module, and a graphical user interface. The system checks the two studies for DICOM 3.0 compliance and then verifies that the DICOM data elements and pixel data are identical. Discrepancies in the two data sets are recorded with the data elements (tag number, value representation, value length, and value field) and pixel data (pixel value and pixel location) in question. The system can be operated automatically, in batch mode, and manually to meet a wide variety of use cases. We ran this program on a 15% statistical sample of 50,000 studies (7500 studies examined). We found 2 pixel data mismatches (resolved on retransmission) and 831 header element mismatches. We subsequently ran the program against a smaller batch of 1000 studies, identifying no pixel data mismatches and 958 header element mismatches. Although we did not find significant issues in our limited study, given other incidents that we have experienced when moving images between systems, we conclude that it is vital to maintain an ongoing, automatic, systematic validation of DICOM transfers so as to be proactive in preventing possibly catastrophic data loss.

KEY WORDS: DICOM, file verification and comparison, PACS

AS THE DEPLOYMENT of Picture Archiving and Communication Systems (PACS) beçomes more prevalent in institutions, the notion of federated systems becomes more important. The concept of federated system implies one or more systems acting together to achieve a common goal or interest. A common model now appearing in clinical PACS sites, an example of a federated system, is the use of off-site, storage service providers (SSP) for long-term archive and disaster recovery purposes. An SSP can be located on-site or off-site. It can be from the same vendor as the PACS or a different vendor. The hardware and software can be purchased outright or the service can be financed on a per study basis. The important fact is that many of these systems are designed or have evolved as independent software systems and therefore must be considered as foreign systems, even if supplied by the same manufacturer.

In Integrating the Healthcare Enterprise (IHE)[1] terms, one can consider that the main PACS is an Image Manager/Image Archive (IM/IA) pair and that the remote service provider acts as a second, perhaps dumber, IM/IA pair. In such a configuration, the main PACS is configured to DICOM (2) store studies to the second IM/IA pair as appropriate. There are numerous other examples wherein a PACS may need to DICOM store images to another DICOM storage service class provider (SCP). DICOM files can be routed to any valid storage SCP on a PACS network. The PACS thus acts as a Service Class User (SCU or client) in the storage transaction while the remote system acts

From the Department of Radiology, Northwestern University, 448 E. Ontario St. Suite 300, Chicago, IL, 60611, USA.

Correspondence to: David S. Channin, MDtel: 312-926-7255; e-mail: dsc@radiology.northwestern.edu

Online publication 12 January 2005
doi: 10.1007/s10278-004-1034-2

as a DICOM SCP (or server). In any event, a DICOM transfer occurs between the two systems. DICOM storage commitment messages may pass between the systems; but both systems may maintain "ownership" of the studies, albeit for different purposes.

Given these scenarios, it becomes evident that a site must be able to verify that the contents of the two or more DICOM storage entities match. It is used to ensure that DICOM studies get to "third-party" storage intact and DICOM data can be retrieved at will. Errors can occur at any point between storage SCU and SCP for many reasons. The two major causes of errors are found in hardware and software.

On the hardware side, networking equipment, network wiring, computer memory, and storage devices may all introduce errors. Even though hardware has significantly improved error detection and correction mechanisms, which enhance the reliability of data, a chance that an error will occur still exists. For example, a computer server that uses parity memory will detect errors detected, but the computer will not be able to correct a failed data bit. Error correction code memory can correct single bit errors, but multiple bit errors, though detected, cannot be corrected. If memory problems go undetected for a long time, they can cause corrupted files and hard disks as well as incorrect computations.

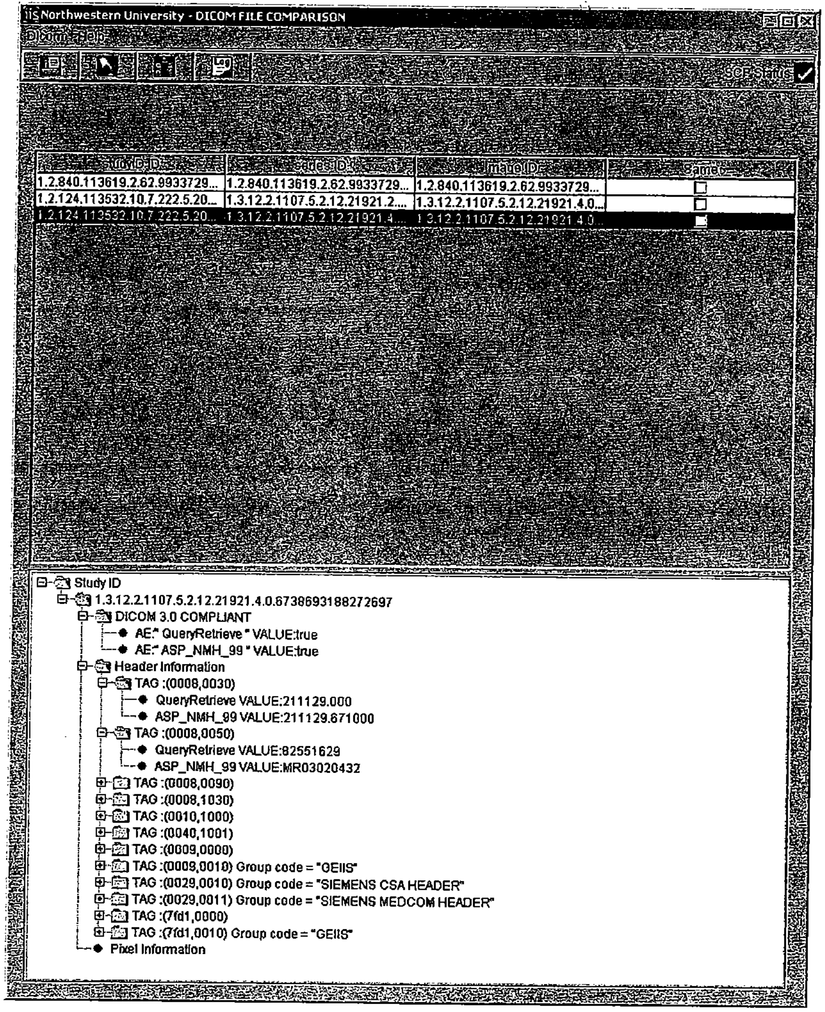
Errors that often occur on the software application side stem from human pitfalls. For a standard like DICOM, different groups of software developers can and do implement DICOM related software differently. They may implement proprietary algorithms and information for their internal use, and these may sometimes conflict with the standard or the interpretation of the standard by another software product. Additionally, software bugs persist despite the best "good manufacturing" processes. Simple file size and date comparisons are not sufficiently robust for validating DICOM delivery and storage in a PACS environment. The DICOM header information may be modified to insert or remove information and DICOM headers may be reconstructed from database information on-the-fly. Inconsistency in DICOM files can be introduced during this file reconstruction, justifying the need for complete DICOM file validation. These inconsistencies can range from relatively benign discrepancies in non-critical header fields to discrepancies in mandatory fields such as modality, patient's name, or the study UID. In either case, operators should be notified of these inconsistencies so they can be appropriately resolved.

MATERIALS AND METHODS

As part of our deployment of an off-site SSP model, we developed a DICOM file comparison and validation tool. Our system is comprised of a DICOM Q/R module, DICOM store SCP, directory archives, DICOM comparison module, and graphical user interface for real-time interaction between the comparison systems and users. We implemented this software using the JBuilder IDE (Borland, Scotts Valley, CA), Java J2SE 1.3.x (Sun Microsystems, Inc., Mountain View, CA), and the MergeCOM-3'[4], Java DICOM toolkit (Merge-Efilm, Milwaukee, WI).

Northwestern Memorial Hospital operates a large GE Centricity PACS (GE Healthcare Information Technologies, Mount Prospect, IL). The system currently contains over 1.8 million studies representing over 80 million images. The system had used on-site magneto-optical disk jukeboxes for long-term archive. Approximately 16 months ago, we began using an off-site storage service (GE Enterprise Archive [EA] SSP, GE Healthcare Information Technologies, Mount Prospect, IL) for long-term storage and disaster recovery. The data in the magneto-optical jukeboxes is being migrated to the off-site storage server as well.

We retrieve a list of DICOM Study UIDs (universal identifier) to be validated from the PACS database. Our software reads this list of DICOM Study UIDs. For each study UID selected, it then queries both the PACS and the off-site EA archive for a pre-defined percentage of the images in the study. The user is allowed to define a percentage of the number of files to be compared ranging from 10% to 100% for related study and series and image UIDs. When the percentage number is less than 100, images within a study are randomly selected for comparison. The system then retrieves the images



Study ID	Series ID	Image ID	Same
1.2.840.113619.2.62.9933729...	1.2.840.113619.2.62.9933729...	1.2.840.113619.2.62.9933729...	□
1.2.124.113532.10.7.222.5.20...	1.3.12.2.1107.5.2.12.21921.2....	1.3.12.2.1107.5.2.12.21921.4.0...	□
1.2.124.113532.10.7.222.5.20...	1.3.12.2.1107.5.2.12.21921.4....	1.3.12.2.1107.5.2.12.21921.4.0...	■

```
⊟ Study ID
   ⊟ 1.3.12.2.1107.5.2.12.21921.4.0.6738693188272697
      ⊟ DICOM 3.0 COMPLIANT
         ├─● AE:" QueryRetrieve " VALUE:true
         └─● AE:" ASP_NMH_99 " VALUE:true
      ⊟ Header Information
         ⊟ TAG :(0008,0030)
            ├─● QueryRetrieve VALUE:211129.000
            └─● ASP_NMH_99 VALUE:211129.671000
         ⊟ TAG :(0008,0050)
            ├─● QueryRetrieve VALUE:82551629
            └─● ASP_NMH_99 VALUE:MR03020432
         ⊞ TAG :(0008,0090)
         ⊞ TAG :(0008,1030)
         ⊞ TAG :(0010,1000)
         ⊞ TAG :(0040,1001)
         ⊞ TAG :(0009,0000)
         ⊞ TAG :(0009,0010) Group code = "GEIIS"
         ⊞ TAG :(0029,0010) Group code = "SIEMENS CSA HEADER"
         ⊞ TAG :(0029,0011) Group code = "SIEMENS MEDCOM HEADER"
         ⊞ TAG :(7fd1,0000)
         ⊞ TAG :(7fd1,0010) Group code = "GEIIS"
      └─● Pixel Information
```

Fig 1. The user interface of the DICOM validation software.

using a DICOM C-MOVE operation to a temporary directory. The system can retrieve these studies from a configurable list of any number of distinct DICOM Storage SCPs.

For each image, the software first verifies for DICOM 3.0 compliance. It then compares DICOM data elements and pixel data, using byte-by-byte checking for pixel data. Header data comparisons are done by creating two collections of data elements from each DICOM file. The two collections are then compared, starting with data elements with the same tag number. Data elements found in one file but not in the other will be recorded. Discrepancies in the two data sets will be displayed and stored in a log file with data elements (tag number, value representation, value length, and value field) and pixel data (pixel value and pixel location). Figure 1 depicts an example of a comparison. An image was selected that contains some differences in tag values and no pixel data differences.

The application also permits users to submit a DICOM Q/R at patient root, study root, or patient/study levels for interactive retrieval and checking. It allows users to create and save a collection of DICOM storage SCP peers. The system can be configured to run at a certain time interval to avoid adding workload to a PACS during busy hours. It also has a display mechanism for reviewing previously stored log files.

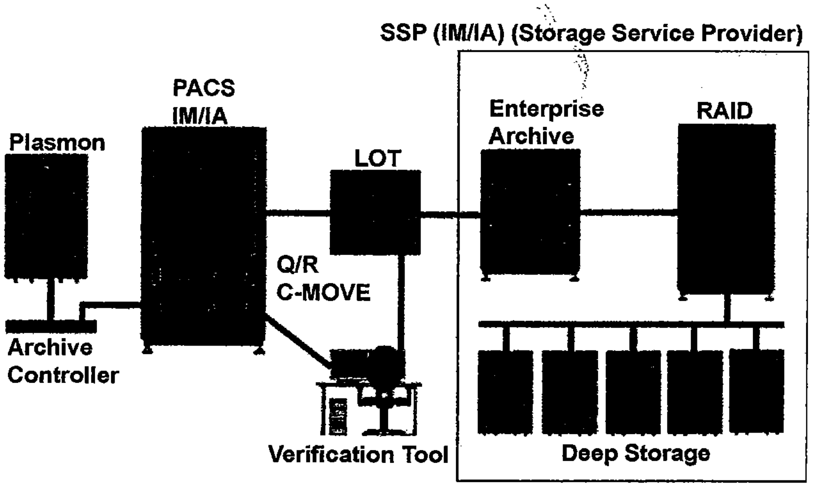


Fig 2. The systems architecture showing the relationship between the PACS, the offsite storage and the validation tool.

RESULTS

The software described was put into use for a particular period of time during the conversion to SSP operations at Northwestern Memorial Hospital (NMH). The system was used to Q/R studies from the NMH PACS and the GE SSP. This configuration is depicted in Fig. 2. It ran in parallel with the vendor's DICOM file comparison software. We selected over 50,000 study UIDs based on studies that were located in the short-term archive and already stored in the long-term archive during the months of March through June 2003. We randomly selected 15% of the study UIDs to be validated. We found two pixel data mismatch cases and 831-header data mismatched and/or missing. Both pixel data mismatches were corrected on re-transmission to our server. We randomly selected additional 1000 study UIDs in March 2004 for comparisons. All images in each study were compared. We found 985-header data mismatches and no pixel data mismatches.

DISCUSSION

In our environment, the GE PACS product acts as the primary IM/IA system. It communicates via DICOM transactions to a distinct GE product, the "Enterprise Archive," which acts as a "dumb" long-term archive. Because these are two distinct product lines, developed by two separate groups within GE, for practical purposes they can be considered the equivalent of two separate systems. That is, given that they do not share hardware, database, or even information models, they must be validated as if they were from separate vendors.

By "dumb" we mean that the long-term, off-site archive does not reconcile patient or study information in the DICOM files. Rather, the main PACS does this management and the off-site storage acts as glorified "coat check" system, returning precisely what it was provided. Similarly, when querying the main PACS, for example, for a married patient's studies, all the patient's studies can be retrieved, including those performed under the patient's maiden name, by virtue of this data management. Yet the DICOM headers will still reflect the original DICOM information as acquired at the modalities. This has significant implications when considering a migration from one PACS to another, yet explains clearly why we should expect very few header element mismatches.

All of the header discrepancies encountered were expected, given known changes to DICOM header data by the two systems in our configuration. Specifically, the SCP must identify itself in one of the header elements and this will be, by definition, different between the two systems. This can easily be filtered by the software to reduce false-positive alarms.

CONCLUSIONS

When transferring images between a clinical PACS and any peer system, a quality-control mechanism must be in place to ensure the accuracy and integrity of the handoff of information between the two systems. As more and more PACS and PACS environments become more and more federated, tools of this nature will become more prevalent. Although we did not find significant issues in our limited study, given other incidents that we have experienced when moving images between systems, we conclude that it is vital to maintain an ongoing, automatic, systematic validation of DICOM transfers so as to be proactive in preventing possibly catastrophic data loss.

REFERENCES

1. Integrating the Healthcare Enterprise. Radiological Society of North America/Healthcare Information and Management Systems Society, Oak Brook, IL, 2003

2. Digital Imaging and Communications in Medicine. The American College of Radiology and National Electrical Manufactures Association, Rosslyn, VA, 2003

3. MergeCOM-3 Advanced Integrator's Tool Kit. Java Class Library Reference Manual Volume 1 and 2: Core Library Classes, Merge-Efilm, Milwaukee, WI, 2002

4. MergeCOM-3 Advanced Integrator's Tool Kit. Java User's Manual, Merge-Efilm, Milwaukee, WI, 2002