EMAIL-P2P GATEWAY TO DISTRIBUTED MEDICAL IMAGING REPOSITORIES

Luís S. Ribeiro, Luís Bastião, Carlos Costa and José Luís Oliveira
University of Aveiro, IEETA/DETI, 3810-190 Aveiro, Portugal

Keywords: PACS integration, Peer-to-peer, DICOM, Medical imaging, Telemedicine, Image communication.

Abstract: For the healthcare professionals the importance of the medical imaging as a diagnostic tool is undeniable. For this reason, industry and research organizations increased significantly their interest in the medical imaging area, trying to deliver solutions for creating, storing, exchanging and displaying medical images. The raise of hardware and software solutions drove the community of vendors to gradually decrease the price of his solutions. As consequence, there was a rise of small imaging centres competing with bigger healthcare institutions. The market offers drives the patients to move across a wide range of healthcare institutions to undergo all the necessary exams. Producing a great amount of medical data dispersed over several institutions. This scenario of isolated islands of images repositories unable of interacting with each other is, in our opinion, propitious to a peer-to-peer (P2P) archive solution. Until now, medical exams (images and studies) have been exchanged through analogue films, media storage devices (CD, DVD, etc), virtual private networks or manual email procedures. This paper describes the Dicoogle P2P system, a distributed PAC system where its users may easily store, search and exchange DICOM files. However, potential peers of the Dicoogle system are usually inside private networks, behind NATs and firewalls, disabling the inter-institutional peer interaction. Therefore, we propose an Email-P2P gateway to Dicoogle that offers a way to exchange DICOM files through these virtual barriers.

1 INTRODUCTION

Nowadays, medical imaging is a valuable and indispensable tool in the healthcare system. To the physicians, they represent a key factor for delivery of high quality decisions. Picture Archiving and Communication System (PACS) concept embraces a set of technologies for the archiving, distribution, visualization and acquisition of medical images over a computer network. Compared with the traditional analogue film, the PACS concept brings significant benefits in the productivity, economy and management of a healthcare institution (Huang, 2004). The PACS concept was boosted by the introduction of the Digital Imaging and Communication in Medicine (DICOM) standard (Mustra et al., 2008), allowing interoperability between PAC systems from different vendors. Moreover, medical imaging modalities and archive solutions are successively becoming less expensive. The result is the proliferation of equipment acquisition by imaging centres, even in small ones (Carlos Costa, 2009). As consequence, the patient's medical data is stored in isolated repositories of the respective healthcare institution. However, the exchange of the medical exams is still a coarse process. Typically, the healthcare institution that performs radiological exams sends the medical data via conventional mail (analogue films, CD, DVD), virtual private networks or manual email procedures. This paper describes a P2P structure built over the Internet Message Access Protocol (IMAP), allowing the search, exchange and store of medical images over a large domain composed by several healthcare institutions. One downside of the P2P architectures is the impossibility of communicating with peers behind virtual barriers (e.g. closed firewalls, Network Address Translation) (Damiani et al., 2002). However, our P2P network is built over IMAP - common firewall configurations do not block the IMAP packages - turning possible communicating with peers behind virtual barriers.

Ribeiro L., Bastião L., Costa C. and Oliveira J. (2010).
EMAIL-P2P GATEWAY TO DISTRIBUTED MEDICAL IMAGING REPOSITORIES.
In *Proceedings of the Third International Conference on Health Informatics*, pages 310-315
DOI: 10.5220/0002744803100315

1.1 Related Work

Besides the manual exchanges by storage devices (e.g. DVD), accessing from the outside of the institution's firewall is usually possible by the creation of Virtual Private Networks (VPN). This approach may be appropriate for the institution members to access the resources remotely. However, creating a cross-institutional VPN connection is a bureaucratic and a longstanding process. The Medical Image Exchange Platform (MIEP) (Chiu et al., 2007) is a cross-institutional platform intended to replace the ad-hoc and manual exchanges. Through dedicated web services with watermarking mechanism they created a secure system capable of exchanging medical data. However, MIEP does not use the DICOM standard, turning the integration the current healthcare systems more difficult. Blanquer et al (Blanquer et al., 2005) presented a P2P environment that assists image diagnosis apprenticeship and research, enabling radiologists to share DICOM studies and its corresponding diagnosis located on their personal computers. The platform promotes the collaboration within a virtual community of radiologists, sharing anonymous DICOM studies and in this way exchanging knowledge in the radiologic field. Finally, there are many P2P applications for file sharing such as BitTorrent(Qiu and Srikant, 2004), eDonkey(Ghosemajumder, 2002) or Gnutella(Ivkovic, 2001), however they are not suitable for the medical data exchange, for instant, they only allow file name search, in our opinion a content base search is more useful.

2 MATERIALS

The PAC system architecture began mainly on an ad hoc basis, serving small subsets, called modules, of the radiology department. Each module functioned as an independent island, unable to communicate with other modules (Huang, 2004). Later it evolved into a PACS infrastructure solution, integrating the hospital information system and the radiology information system, serving the entire hospital (figure 1). Nowadays, with the emerge of small imaging centres the early PACS design is once again a reality. As a result, a vast repository of images and studies remain stored locally without being shared. In our opinion, the current scenario is propitious to a P2P architecture.

2.1 Peer-to-peer Networks

The P2P networks may follow different topologies that bring different advantages and disadvantages.

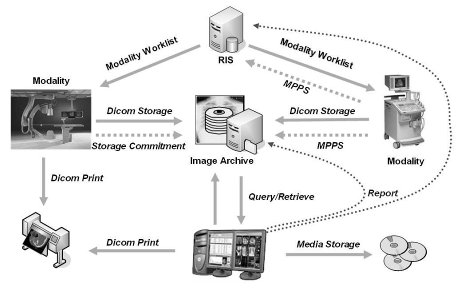

Figure 1: PACS traditional workflow.

The most usual approaches include the classic decentralize structure (e.g. Gnutella v0.4) where every peer is similar and self sufficient, the hybrid approach (e.g. Gnutella v0.6), where some peers of the structure have special responsibilities and, finally, the structure completely centralized (e.g. eDonkey), that relies in at least one central server executing critical tasks for the correct function of the network, keeping the rest of the communications in a P2P basis.

2.1.1 Jxta Framework

Jxta is a network programming and computing platform designed to solve distributed computing problems, especially in the peer-to-peer area. Jxta was designed to fulfil three major concerns (Gradecki and Gradecki, 2002): interoperability, platform independence and ubiquity. These concerns drove Jxta technology into a set of six major protocols: peer discovery protocol, peer resolver protocol, peer information protocol, peer membership protocol and pipe binding protocol. Currently, Jxta is in its 2.5 version and there is a release announced to the end of the year 2009.

2.2 Dicoogle

In 2009 was introduced Dicoogle (Carlos Costa, 2009), a novel approach to manage PAC systems. The main idea of Dicoogle was the replacement or the extension of the traditional relational database in the PAC systems by an indexing and retrieving engine. This approach turned the PACS management more flexible, because it is possible to quickly index new text-based fields without the need of creating new tables or relations like would be necessary in the database supported approach. Furthermore, Dicoogle automatically refreshes the indexes by monitoring the file system events (create, delete or update) targeted to the DICOM files.

3 METHODS

The Dicoogle version 3 aims to an enterprise solution enabling the exchange of DICOM objects (images and studies) between healthcare institutions. Typically, the DICOM objects are stored locally in the PACS of the healthcare institutions that operate over a private network (intranet). Therefore, the most common scenario will be islands of DICOM object repositories. As consequence, the bigger percentage of Dicoogle peers will be delimitated by virtual barriers such as NATs and firewalls (figure 3).

3.1 NATs and Firewalls

In order to a machine inside a private network access the internet it requires a public IP address, however the internal machine only holds a private IP address. The NAT process allows the router to translate private IP address into public IP addresses. When an internal computer requests access to the Internet, the requests private IP is placed by the router's public IP address. When the response from a request arrives it is forward to the internal computer. This is only possible when the internal computer starts the process, outside responses with no request are discarded. Besides the NAT there is also the firewall. The firewall job is to restrict traffic coming from the Internet into the private network and many times also restrict the traffic going from the private network to the internet (Gradecki and Gradecki, 2002), by closing all the ports to send data. Only specific ports are commonly open such as the email and the web server ports. These ports are open to send data from the inside to the outside of the intranet, but not the reverse. As result, peers from different private networks cannot communicate "directly". The solution for this problem is the introduction of a relay peer in the middle of the communication between peers in different intranets (figure 4). The relay peer becomes a bridge, enabling the interaction of Dicoogle peers of different private networks. To implement the relay peer we looked which technology is more appropriate based on Dicoogle's relay peer requirements:

1. The relay peer has to be accessible to all peers;
2. The relay peer has to communicate through default open ports of the firewall;
3. The relay has to support secure data transition.

Analyzing the relay requisites and after some tests we concluded that an email server would be suitable for developing the relay peer, because:

1. Email servers are reliable in terms of access;
2. Email may be sent through firewalls without any change in their set rules
3. Email already has encryption protocols for data exchange.

Furthermore, email accounts are free, trustworthy and offer an acceptable storage capacity (e.g. Gmail over 7 Giga bytes of available space). But more important the relay process is really similar to the email paradigm: The peers send addressed messages to relay peer and the relay peer holds the messages until the addressed peer fetch them. Therefore, Dicoogle P2P system implements the relay peer over an email account.

3.2 Components

Dicoogle is a Java open source project developed over some open source libraries. Figure 2 show all the software components used by the Dicoogle framework. Dicoogle uses the Apache Lucene index engine (Hatcher and Gospodnetic, 2004) to index the DICOM files stored in each archive (Dicoogle peer). The P2P functionalities are supported by the Jxta framework (Sun Microsystems, 2004b; Gradecki and Gradecki, 2002). Jxta allows firewall crossing, this is achieved by communicating with a relay peer, outside the intranet, through the port 80 or 8080 (HTTP/S). The specifications of Jxta make it the ideal platform, in theory, to fulfil the Dicoogle's P2P requirements. Although, in the paper, Jxta is a powerful and cohesive idea, the reality is that its implementation is not stable enough for a P2P enterprise solution. In spite of the fact some sources pointed Jxta version 2.5 as unstable (Ferrante, 2008) we analyzed anyway the framework. Our conclusions were concordant with the previous ones, Jxta version 2.5 has still to mature. However, we were able to create a relative stable solution for the private network scenario. Therefore, Dicoogle v3 only uses Jxta for treating the P2P issues within the intranet. We designed a relay peer based in IMAP - using the JavaMail API (Sun Microsystems, 2004a). With the IMAP protocol we created transparent communication channels between the participant intranets enabling inter-institutional peer interaction. Every message sent to the relay peer is encrypted using the Advance Encryption Standard (AES). The messages are encrypted using a private key present in every Dicoogle peer and unreachable to the users. The private key is created using the email password as seed for a key generator algorithm. This way the messages are impossible to read by other entities besides the Dicoogle peers. Finally, to carry the implementation of DICOM services and DICOM standard related issues we used the dcm4che SDK (Zeilinger, 2006).

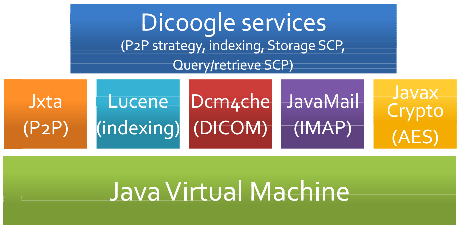

Figure 2: Dicoogle software components.

3.3 Architecture

Dicoogle's P2P distributed system is formed by multiple Dicoogle peers, located in different intranets, and a relay peer. Inside each participant healthcare institution (intranet) it is at least one Dicoogle peer. One may classify Dicoogle's P2P topology as hybrid, however we consider that Dicoogle's topology as two different levels. At the intranet level the Dicoogle peers have a totally decentralized topology. They automatically discover each other and create the communication channels between them. While at the internet level the Dicoogle peers interact with a centralize approach. The Dicoogle peers announce their existence on the relay peer and all the communication between peers in different intranets passes through the relay peer. The main reason to use a P2P centralized approach is to fulfil one system requirement: Dicoogle must pass through virtual barriers. Figure 3 shows the overall architecture of the Dicoogle P2P distributed system.

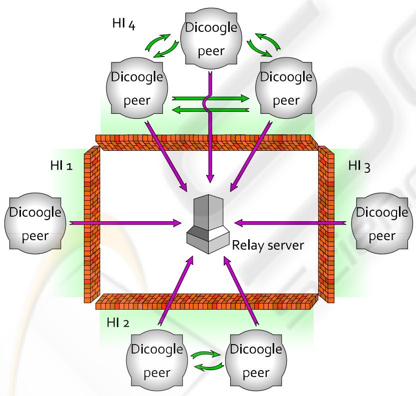

Figure 3: Dicoogle's common usage scenario. Several peers disperse over several healthcare institutions (HI). Peers within the HI communicate directly. Peers in different HIs communicate through the relay peer.

3.3.1 Searching DICOM files

Each Dicoogle peer is responsible for managing a PACS archive. To do so, it indexes every DICOM file of the repository. The indexes are the foundations of Dicoogle's search mechanism. As soon as the Dicoogle peer is initiated over a PACS archive it starts by indexing the detected DICOM files. After this phase, the Dicoogle peer monitors the file system looking the system calls create, delete and update targeting the DICOM files. When a relevant system call is detected Dicoogle automatically refreshes the peer indexes.

Dicoogle search mechanism has three modes: local search, private network search and entire network search. Each mode represents a search domain, according to its needs the peer user may define which search mode it prefers. Also the peer administrator may specify which search mode will be executed when the peer is contacted by a non-Dicoogle station - using the DICOM query service. The search modes function over the peer, or peers, indexes executing the queries entered by the user, as follows:

1. **Local Search:** The query is executed in the local peer and the results are displayed on the peer to the user.
2. **Private Network Search:** The query is sent to the peers in the same private network (using the Jxta's multicast mechanism). When the other Dicoogle peers receive the search query, they execute it locally over their indexes then send the result to the calling peer. Finally the calling peer merges the received results and displays them to the user.
3. **External Network Search:** The query is sent to every Dicoogle peer outside the private network. The communication channel is built through the relay peer. The calling Dicoogle peer sends the search query to the relay peer, addressed to all Dicoogle peers not present in its private network. This is accomplished because every private network has a unique identifier (cluster ID), therefore if a search query arrives to the relay peer the Dicoogle peers only fetch the query if the message has a different cluster ID. This way, only the peers outside private network of the calling peer will receive the message. After this, the external Dicoogle peers fetch the message, execute the query locally and send a message with the result to the relay peer - the message is addressed to the calling peer. Finally, the calling peer fetches the results, merges the multiple results and displays them to the user.

Furthermore, it is also possible to combine the different search modes, for instance to search the entire P2P network the Dicoogle peer triggers the three search modes at the same time.

3.3.2 Transfering DICOM Files

Being able of searching the resources of a P2P network is important, but useless unless it is possible to obtain the contents returned by the search. Therefore, the Dicoogle system allows the exchange of DICOM files between the participant peers. Typically, after a search the peer user triggers a file transfer. At this stage the Dicoogle peer has all the information required regarding the file location. Two different scenarios may occur, according to the file location:

- If the wanted DICOM file is in the same private network of the calling peer the file transfer is performed through Jxta's Bidipipes. A Bidipipe is bidirectional communication channel created between two peers. The calling peer starts by sending a file request message to the peer that holds the DICOM file. Then the called peer sends the file blocks until the file transfer succeeds.
- If the file is located outside the private network the file transfer is performed through the relay peer. The transfer protocol is similar to the intranet scenario although with the relay peer abstraction (figure 4). The calling peer sends a file request message to the relay peer, addressed to the peer that holds the file. The called peer that periodically checks the email account detects that has a message addressed to it. As consequence, the called peer fetches the message. Besides specifying the file to transfer, the file request message contains a session name. If the file transfer is accepted the called peer creates a new folder in the email account with the received session name. The DICOM file is transferred to the session folder. In the other side, the calling peer monitors the email account waiting for the session folder creation. If the session folder is created it means the file transfer was accepted and the file will be sent, otherwise the file transfer was denied and the file exchange fails by timeout. Furthermore, the DICOM file cannot be sent directly to the session folder. Because the relay peer is built over an email account the DICOM file is usually bigger than the maximum size allowed for the message attachments. Therefore, the DICOM file is divided into small blocks and sent, block by block, to the session folder. Finally, the calling peer fetches the file blocks, reassembles the file blocks into the original DICOM file and saves it in its repository.

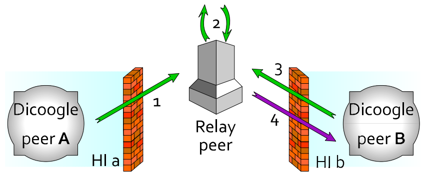

Figure 4: Inter-institutional communication. Dicoogle peer A wants to send a message to the peer B, the two peers are in different private networks and between them there are virtual barriers. Communication protocol through the relay peer: (1) - Peer A sends the message(s) to the relay peer addressing the peer B. (2) - The relay peer holds the message in its file system. (3) - Peer B is periodically checking the relay peer for messages addressed to it. (4) - Peer B detects the message sent by the peer A, downloads message and removes it from the relay.

4 RESULTS

We created an efficient P2P platform oriented to medical imaging services, by reusing the available resources in the healthcare institutions. The Dicoogle peers are installed over the existent PACS archive and, besides functioning as a typical PACS archive management system, it also extends the contents of the local PACS archive, by interacting with the other Dicoogle peers. This way medical data exchange process is facilitated and as consequence increasing the productivity of healthcare procedure. Dicoogle also provides to its users a way to cross the network's virtual barriers expanding the P2P potential users normally excluded from this type of networks.

From a simple email account a new virtual community of data sharing is created. In other words, one email account represents a new Dicoogle universe, to belong to one of these universes the Dicoogle peer has to know the respective email account. Otherwise, the Dicoogle peer may only communicate at the intranet level.

The Dicoogle system expedites the *modus operandi* of the healthcare institutions, for instance, a healthcare institution composed by several small imaging centres, dispersed over a big geographical area. By adoption the Dicoogle system the dispersed PACS archives could easily connect as a global PACS archive, serving the entire healthcare institution as a whole. Dicoogle system offers a comfortable user interface to manage the local Dicoogle peer and to explore the contents of the entire Dicoogle distributed system. Figures 5 and 6 show some views of Dicoogle's interface.

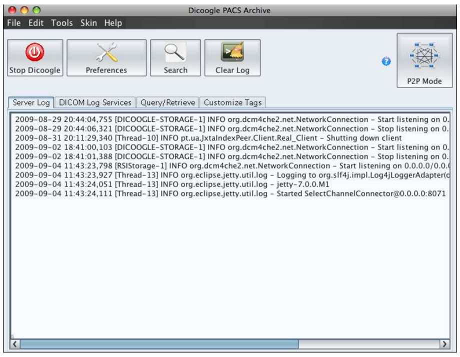

Figure 5: Dicoogle's control console

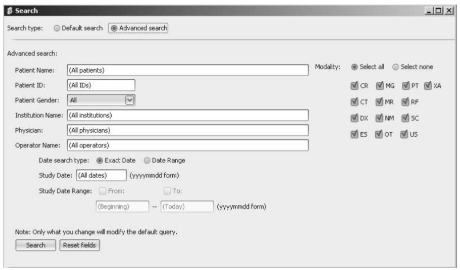

Figure 6: Dicoogle's Advance search view

5 CONCLUSIONS

In this paper we present the Dicoogle distributed PACS and its Email-P2P Gateway interface capable of searching and sharing DICOM files over a wide scenario. Dicoogle is a turn-key solution that may be easily installed in a personal computer or in a central server. And its integration with the existent medical imaging repository of a healthcare institution is facilitated. Because, each Dicoogle peers is also a service class provider of the DICOM services storage, query and retrieve. This way the surrounding stations of the healthcare institution may interact with the system without any reconfiguration.

The system uses a P2P decentralize topology within the private network and a centralize topology for remote communications, maximizing this way the benefits of each topology, turning Dicoogle in a scalable system. Furthermore, the relay peer is built over an email account and not a common web server. Therefore, the cost of owning a web server to act as a relay peer is eliminated and, because the email ports are usually open, it also avoids the need to change the firewall's default configuration. Dicoogle offers to his users an even more flexible P2P system than others known P2P applications (e.g. eMule) because it does not leave behind the peer inside the private network incapable of communicating directly with the P2P network due to the virtual barriers. As consequence, Dicoogle extends the range of its potential users increasing the resources available in the network.

REFERENCES

Blanquer, I., Hernandez, V., and Mas, F. (2005). A peer-to-peer environment to share medical images and diagnoses providing context-based searching. In *Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP'05)*, pages 42–48.

Carlos Costa, Augusto Silva, J. L. O. (2009). Indexing and retrieving dicom data in disperse and unstructured archives. volume 4, pages 71–77.

Chiu, D. K., Hung, P. C. K., Cheng, V. S. Y., and Kafeza, E. (2007). Protecting the exchange of medical images in healthcare process integration with web services. pages 131–131.

Damiani, E., di Vimercati, D. C., Paraboschi, S., Samarati, P., and Violante, F. (2002). A reputation-based approach for choosing reliable resources in peer-to-peer networks. In *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, pages 207–216. ACM.

Ferrante, M. (2008). *The JXTA way to Grid: a dead end?* PhD thesis, University of Genova.

Ghosemajumder, S. (2002). Advanced peer-based technology business models. *MBA Master's Thesis, Massachusetts Institute of Technology, Sloan School of Management, Cambridge, MA.*

Gradecki, J. and Gradecki, J. (2002). *Mastering JXTA: building Java peer-to-peer applications*. Wiley.

Hatcher, E. and Gospodnetic, O. (2004). Lucene in action. *Action series. Manning Publications Co., Greenwich, CT*.

Huang, H. (2004). *PACS and imaging informatics: basic principles and applications*. Wiley-IEEE.

Ivkovic, I. (2001). Improving gnutella protocol: Protocol analysis and research proposals. *Retrieved May*, 12.

Mustra, M., Delac, K., and Grgic, M. (2008). Overview of the DICOM standard. In *ELMAR, 2008. 50th International Symposium*, volume 1.

Qiu, D. and Srikant, R. (2004). Modeling and performance analysis of BitTorrent-like peer-to-peer networks. In *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 367–378. ACM New York, NY, USA.

Sun Microsystems, I. (2004a). Java Mail API. *URL (07.07.2009): http://java.sun.com/products/javamail*.

Sun Microsystems, I. (2004b). Jxta. *URL (07.07.2009): https://jxta.dev.java.net/*.

Zeilinger, G. (2006). dcm4che, a DICOM implementation in Java. *URL (07.07.2009): http://www.dcm4che.org*.