# Grid Computing Inside Hospitals Using Virtualization Technology: A Secure Solution for Heavy Computing Tasks

Xin Zhou[1], Adrien Depeursinge[1], Marko Niinimaki[1],
Antoine Geissbuhler[1] and Henning Müller[1,2]
[1]Medical Informatics Department, Geneva University Hospitals and University of Geneva, Geneva, Switzerland
[2] Business Information Systems, University of Applied Sciences Sierre, Switzerland

## Abstract:

**This paper describes the evaluation of the implementation of an internal computing Grid inside the Geneva University Hospitals (HUG). The objective is to provide computing power to researchers inside the HUG with a minimum cost and disturbance for the hospital system administrators. The proposed solution makes use of idle resources inside the hospital to fulfil the researchers' computational needs. It is based on virtualization and Grid technologies. The solution was evaluated according to the usability for research purposes and the possible disturbance of the HUG network.**

### Introduction

The ever-increasing production of medical images in digital format requires new tools for the data analysis to use the data up to their full potential [1]. Analysis and automatic processing of large image datasets are computationally expensive [2]. Like most hospitals, the HUG do not have any dedicated research computing infrastructure to support computationally intensive applications at the moment. Using external computing resources may cause legal problems due to the data transfer of patient data. On the other hand, over 6'000 desktop computers are available on the network of the HUG. Even a partial re-use of these resources could help to fulfil the researchers' computational needs. To explore the idle computing resources, virtualization techniques are deployed [3]. The whole infrastructure is built up based on the ARC (Advanced Resource Connector) middleware[1] [4].

### Methods

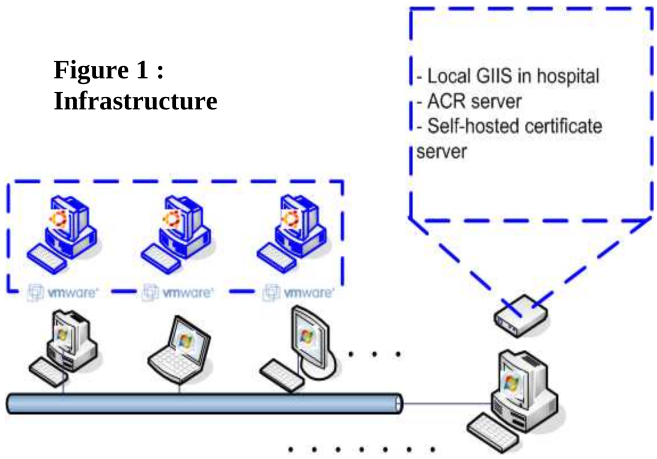
#### Hardware:

To set up our test bed, 20 standard desktop PCs of the HUG were made available for our group. Most of them are old PCs with a 2.8GHz CPU and 768MB RAM. By the end of 2009, all the PC in HUG will be upgraded to at least Dual-core 2.8GHz, 2GB RAM.

#### Software:

On each machine a virtual Linux operating system was installed using VMWare[2]. This virtual machine (VM) serves as a Grid computing node, which separates the computation node from the user's operating system. The back-end middleware in use for each computation node is Condor[3] [5]. No configuration is needed as they are pre-set in the virtual image. All installation components can be installed fully automatically to each computer on the HUG network (standard solution of the HUG based on Microsoft Active Directory). No user interaction is

**Figure 1 : Infrastructure**




[1] http://www.nordugrid.org
[2] http://www.vmware.com
[3] http://www.condor.com

required. Figure 1 provides an overview of the infrastructure. Grid related components such as Grid Information Index System (GIIS), ARC server and certificate server were all installed on an external hard disk.

**Evaluation of the infrastructure:**

The infrastructure is evaluated based on its performance to handle image analysis, the impact on desktop PC users denoting part of resources, and the possible disturbance on the HUG network. For the first evaluation criterion, we compared the time taken for the image analysis to previous experimentation carried out with a local server on the one hand and a remote grid infrastructure on the other hand. The second criterion is to measure the effect of slowing down software in frequent use in the HUG. Network disturbance is evaluated by taking applications, which are already in use as baseline.

## Results

The following sections provide the results of the evaluation of the infrastructure.

**Usability for researchers:**

Usability was evaluated by using an application that submitted 50 jobs. In each job 1000 images are treated and results are automatically collected. This application was executed on 1 server, remote Grid resources (located mainly in Finland) and the local Grid using the resources inside the HUG. An overall time comparison is shown in **Table 1**. This result exhibits the improvement of the application efficiency using Grid technology [6].

| | Time (min) |
|---|---|
| 1 server with 4 CPU, 4GB RAM | 807 |
| Remote Grid with 37 CPU | 537 |
| Local Grid with 10CPU, 350M x10 RAM | 240 |

**Table 1 :**

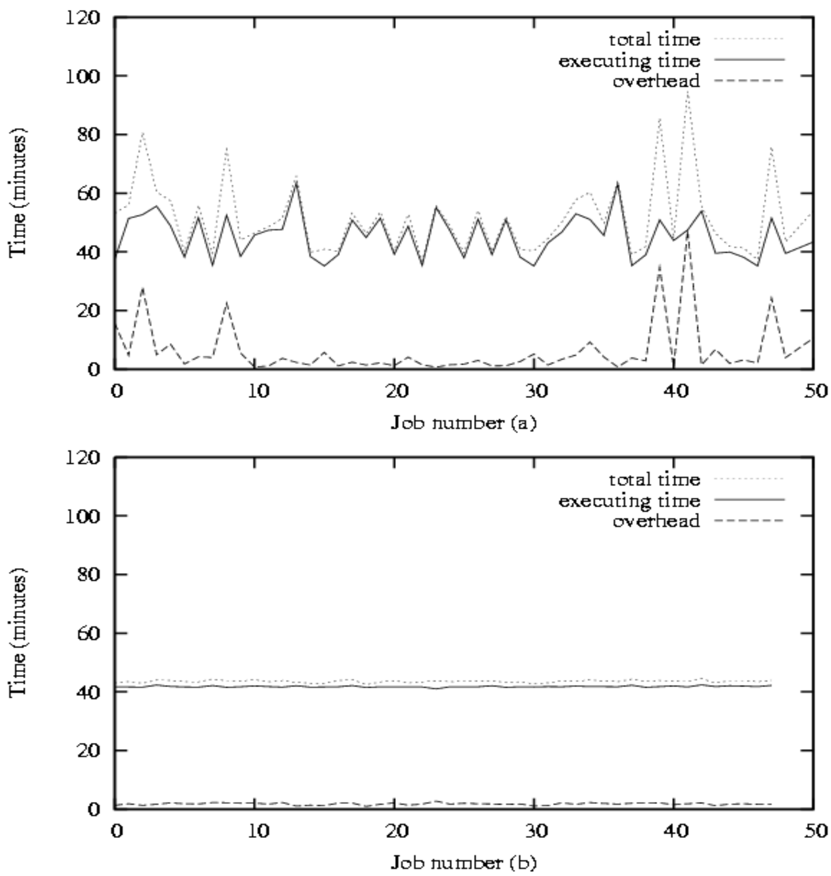



**Figure 2 : Executing time for each job**

Figure 2(a) shows the computing time of each job submitted to the remote Grid resources, the same analysis is performed for the local Grid and the results are shown in Figure 2(b). Compared with the remote Grid resource pools, less CPUs were available in the local Grid resource pool. However, due to the fact that there was less competition for job execution and less communication time, using local resources provided much faster execution.

**Impact on desktop PC users:**
Installing computing nodes on desktop PCs will certainly have certainly an impact on the desktop PC users. The impact is evaluated by frequently used applications in the HUG: Internet Explorer 7, MS office 2003 and MS office 2007. MS office 2003 is currently in use, but will be replaced by MS office 2007 by the end of 2009.

Table 2 shows the starting time for each application. Two configurations were tested. One is based on a computer with an old CPU and little RAM, another is a relatively new machine. No significant impact was detected even when the VM was running a Grid job. Furthermore, it is up to the user to decide whether to donate part of the resources or not, by starting or stopping the virtual Linux system. A user can take back the resources at any time they like.

| | VM not running | VM running but idle | VM running Grid jobs |
|---|---|---|---|
| Old machine ( CPU 2.8GHz, 768MB RAM ) | | | |
| Internet Explorer 7 | 2.7 | 3.2 | 5.8 |
| MS office 2003 | 2.8 | 3.3 | 8.1 |
| New machines ( Dual-core 2.8GHz, 2GB RAM ), All machines will be upgraded to this configuration by the end of 2009. | | | |
| Internet Explorer 7 | 2.7 | 2.7 | 2.8 |
| MS office 2003 | 2.4 | 2.5 | 3.0 |
| MS office 2007 | 4.2 | 4.5 | 6.1 |

**Table 2 : Start time comparaison**

**Possible disturbance on the network:**
Grid usage might generate frequent data transfers. It is thus important to measure the possible disturbance on the network.

| | PC idle, VM not running | PC idle, VM started | ProVision and IE 7 in use | Grid running with 25M input file |
|---|---|---|---|---|
| Measurement during 10 minutes | | | | |
| bytes w/o job files | 29,000 | 59,800 | 4,500,000 | 61,000 |
| job files | | | | 26,009,600 |
| bytes/s | 48 | 100 | 7,500 | 43,451 |
| bits/s | 387 | 797 | 60,000 | 347,608 |
| Measurement during 1 hour | | | | |
| bytes/s | 76 | 105 | 7,300~15,000 | 27,290 |
| bits/s | 608 | 840 | 61,600 | 218,320 |

**Table 3 : Network charge**

In Table 3, the network charge of 4 use cases is compared. ProVision is the standard image viewer and IE 7 is the standard browser. These two applications were put in use to simulate an ordinary work situation. Results show that with the Grid setup installed and no job running, the network charge is negligible. When job execution requires data transfer, the network charge is only slightly higher than the ordinary usage of the Internet. Finally, the network charge can be limited by the configuration of the computing node.

## Conclusions

The proposed solution shows to be robust and allows faster computation time than an expensive server used for comparison. The data transfers do not disturb the hospital network and can also be limited if necessary. On recent desktop computers with dual core processors and 2GB of RAM) the delay is almost unrecognizable by the desktop user.

To help developers to quickly adapt their applications to Grid execution, a workflow management software (Taverna[4]) with Graphical User Interface (GUI) was also used. It allows users to design the execution sequence and workflow for application graphically. An ARC plug-in[5] takes care of job submission, job status checking, as well as result collection. Figure 3 shows an example of using the ARC Grid via this tool.

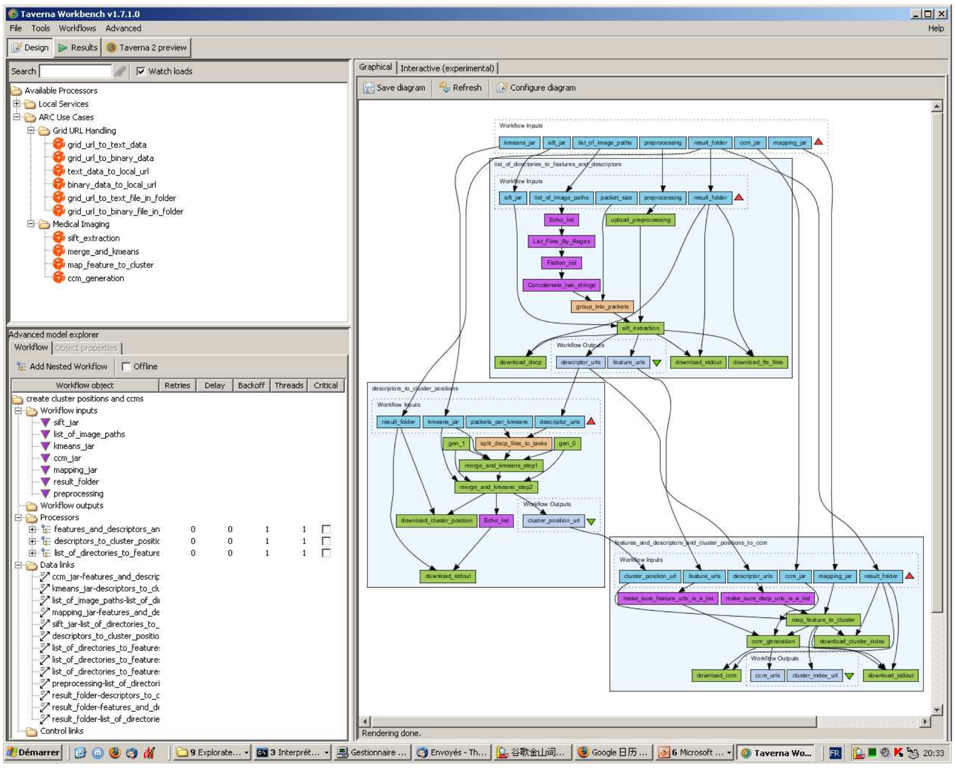



**Figure 3 : an application in Taverna**

An internal computation Grid was created to support researchers with computing power. Three medical imaging applications, general content-based image retrieval, lung image retrieval, and fracture retrieval were “griddified” and tested. Further extension of the Grid system seems possible and particularly it can be foreseen to make the infrastructure available to a larger number of researchers, for new applications.

## Acknowledgements

This work was partially funded by the European Union 6th Framework program in the context of the

[4] http://taverna.sourceforge.net/
[5] http://taverna.nordugrid.org/

KnowARC[6] research project (contract number 032691). We would also like to thank the DIG for their support throughout the project and the quick responses we often received.

## References

[1] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content–based image retrieval systems in medicine – clinical benefits and future directions. Int. Journal of Medical Informatics, 73:1–23, 2004.

[2] J. Montagnat, V. Breton, and I. E. Magnin. Partitioning medical image databases for content-based queries on a grid. International Journal of Supercomputer Applications, 44(2):154-160, 2005.

[3] M. Niinimäki, X. Zhou, A. Depeursinge, A. Geissbuhler, and H. Müller. Building a community grid for medical image analysis inside a hospital, a case study. In *MICCAI-Grid'08*, pages 3–12, 2008.

[4] M. Ellert, M. Grønager, A. Konstantinov, B. Kónya, J. Lindemann, I. Livenson, J. Langgaard Nielsen, M. Niinimäki, O. Smirnova, and A. Wäänänen. Advanced resource connector middleware for lightweight computational Grids. FGCS, 23(2):219–240, 2007.

[5] M. Litzkov, M. Livny, and M. Mutka. Condor — a hunter of idle workstations. In Proceedings of the 8th international conference on distributed computing, pages 104–111, 1988.

[6] X. Zhou, M. Pitkanen, A. Depeursinge, and H. M¨uller. A medical image retrieval application using grid technologies to speed up feature extraction. In ICT4Health, Manila, Philippines, February 2008.

[7] Hajo Krabbenhöft, Steffen Möller, and Daniel Bayer. Integrating arc grid middleware with taverna workflows. Bioinformatics Applications Note, 2008.

## Correspondence

Xin Zhou
Geneva University Hospitals and University of Geneva, Medical Informatics Department
24, rue Gabrielle-Perret-Gentil, CH-1211 Geneva 14, Switzerland
xin.zhou@sim.hcuge.ch, http://www.sim.hcuge.ch/medgift/
Tel ++ 41 22 372 6279, fax ++ 41 22 372 8879

[6] http://www.knowarc.eu/