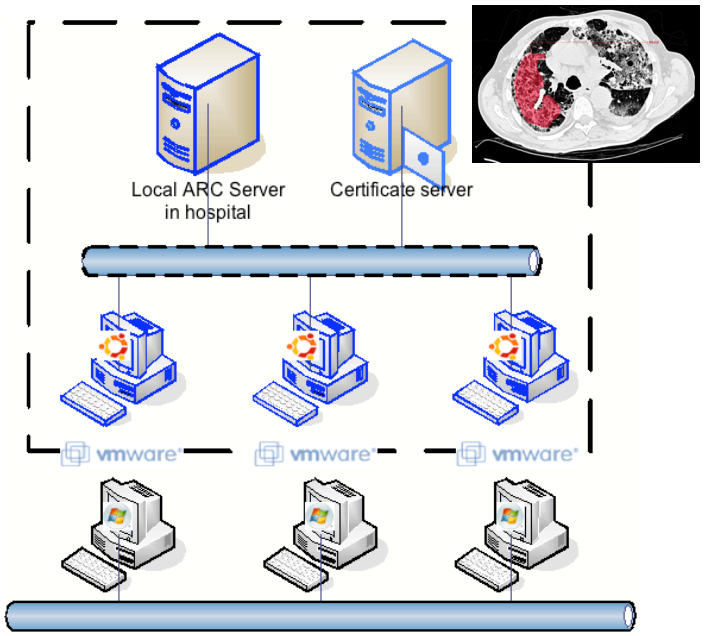
Building a community grid for medical image analysis inside a hospital, a case study



Marko Niinimaki, Xin Zhou,
Adrien Depeursinge (PhD Student),
Antoine Geissbuhler and
Henning Müller,
Medical Informatics Service,
Geneva University Hospitals and
University of Geneva

UNIVERSITÉ DE GENÈVE

HUG
Hôpitaux Universitaires de Genève

KnowARC

Overview

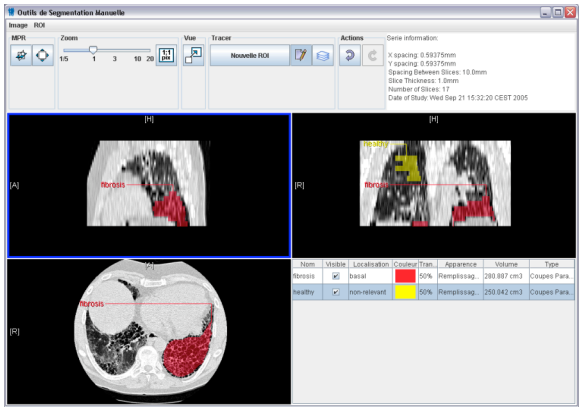
- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

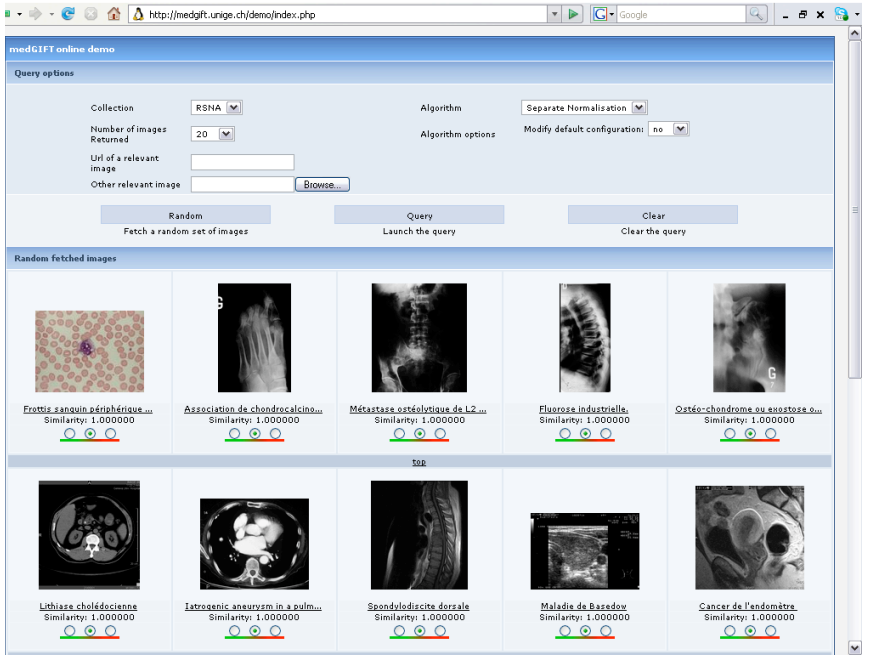
Overview

- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

Motivation: Medical image analysis

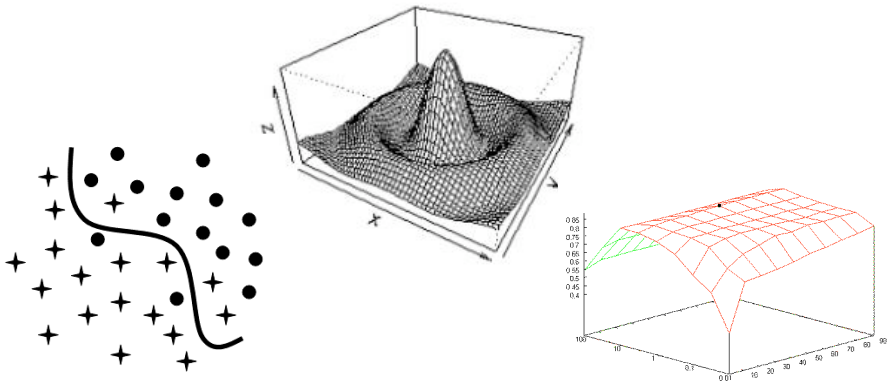
- Computerized medical image analysis provides tools for radiologists inside hospitals
  - Content-based image retrieval (CBIR)
    - Find similar images/cases
  - Computer-aided diagnosis
    - Automatic detection of abnormal patterns
  - ...





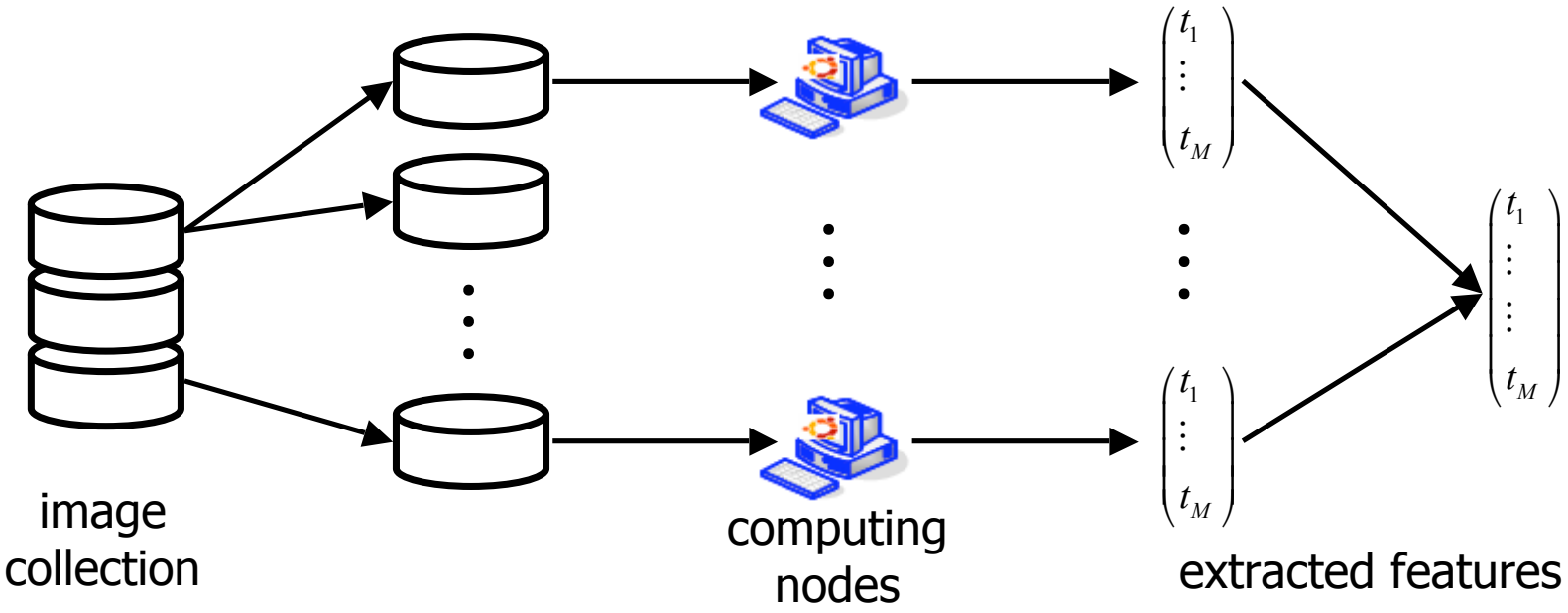
Motivation: Medical image analysis

- Medical image analysis can be computationally intensive:
  - Modern hospitals produce enormous amount of medical images:
    - ~70'000 images/day in Geneva radiology in 2007
  - 2D to 4D high-resolution images
  - Using latest image processing and machine learning algorithms
    - 2D to 4D convolutions
    - High-dimensional feature spaces
    - …



Motivation: Medical image analysis

- Grid computing for medical image analysis:
  - Most of image processing and machine learning algorithms are parallelizable
  - Offline feature extraction for CBIR over large image collections is easy to parallelize:



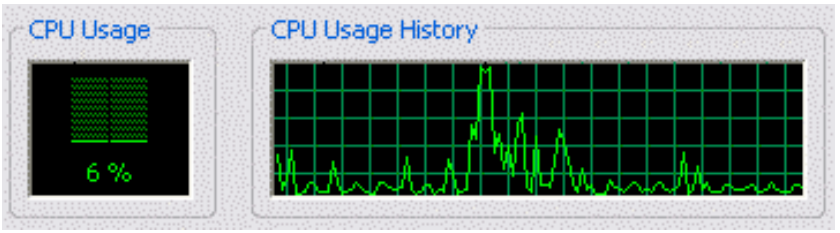
Motivation: Medical image analysis

- Previously: feature extraction for the GNU image finding tool (GIFT) CBIR system was parallelized using external resources provided by the KnowARC project

- Goal: use local resources within hospitals
- Benefits:
  - Policy conformance for handling confidential data
  - Connect resources at different sites within the organisation
  - Interoperability with external resources: one unique interface

Motivation: Medical image analysis

- Local resources at the University Hospitals of Geneva (HUG):
  - No central computing infrastructure for research
  - ~6000 desktop computers
    - Mainly used to access patient records from central databases



⇨ **High potential computing resources**

Overview

- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

Challenges inside hospitals

- Low priority for research compared to clinical routine:
  - Do not exhaust hospitals network resources as people's lives can depend on them
  - Respect data security and confidentiality
  - Use solutions requiring little maintenance for the user
- Main challenges are categorized as:
  - Infrastructures (I)
  - Grid software (G)
  - Political problems (P)

Challenges inside hospitals

- Infrastructures (I)
  - I1: Ports are blocked by a rigid firewall
    - Requirement: do not rely on external resources
  - I2: Research applications and Grid middlewares are mostly running on Linux whereas hospitals desktops have Windows XP
    - Req.: do not assume a Linux environment

Challenges inside hospitals

- Grid software (G)
  - G1: Grid software can be hard to install and maintain
    - Req.: use external expertise and plan time for learning
  - G2: Certificates are difficult to obtain
    - Req.: use own certificates authority if needed
- Political problems (P)
  - P1: Grids are relatively unknown by hospitals ITs: rejection of technology
    - Req.:
      - Involve hospitals IT
      - Create a test bed
      - Create solutions that requires little IT personnel's involvement

Overview

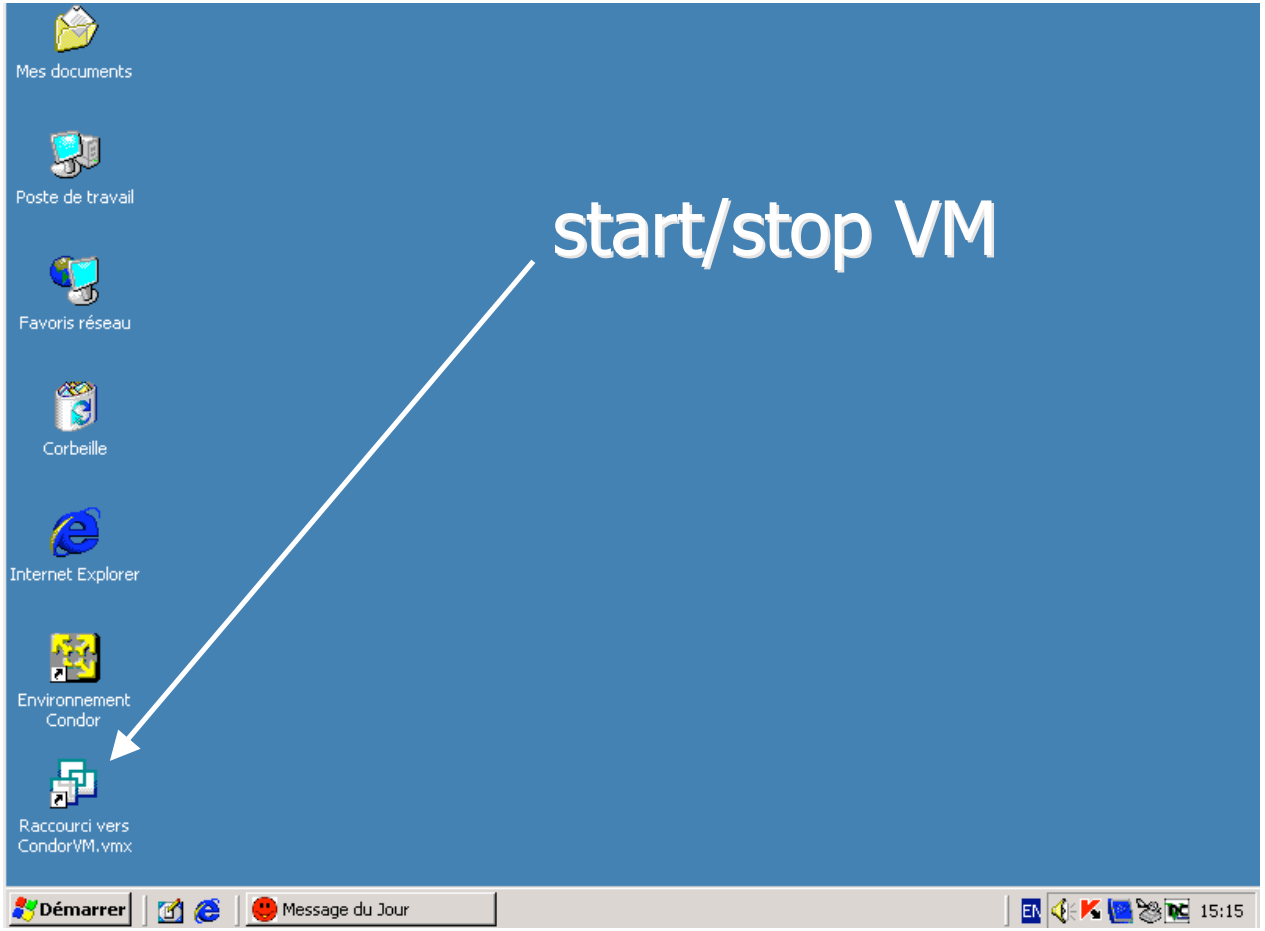
- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

Solutions guidelines

- A test bed was created within the HUG
- Infrastructures (I)
  - A local Grid information system was employed
  - Installation package was created containing a virtual machine (VM) with:
    - Debian Linux and image processing software
    - A local resource management system (LRMS)
  - Installation package was distributed using existing software distribution of the hospitals
  - The user can start/stop the VM when desired

Solutions guidelines

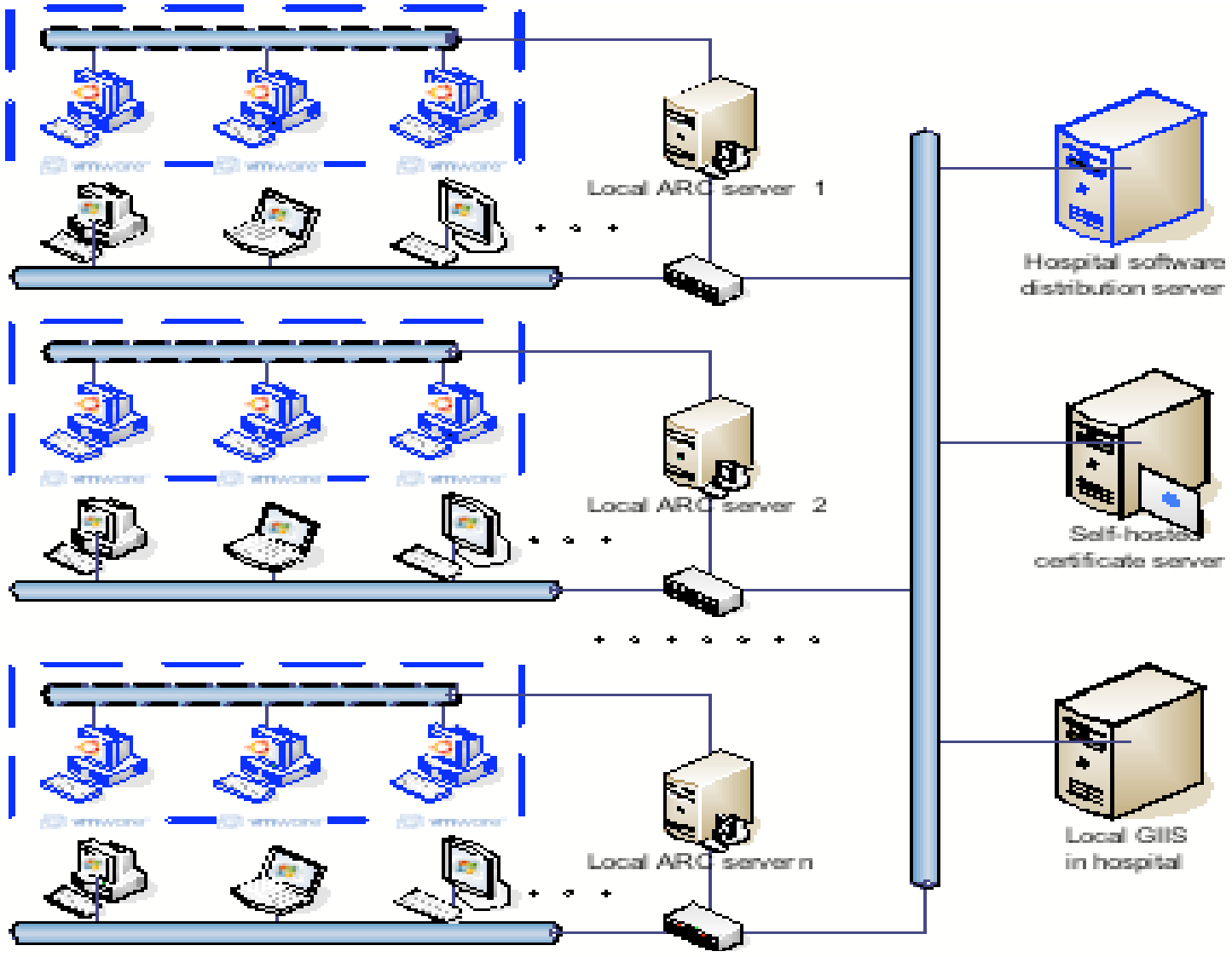
- Screenshot of users's desktop:



Solutions guidelines

- Grid software (G)
  - The Advanced Resource Connector (ARC) Grid middleware and Condor LRMS were used
  - Own certificate authority (CA)
- Political problems (P)
  - A test bed of 20 PCs was created before deploying a larger scale Grid
  - The PCs gain their IP addresses from main hospital area's address pool
    - VM acquire a MAC address at installation, and the DHCP has allocated IP addresses for those

Solution guidelines: setup

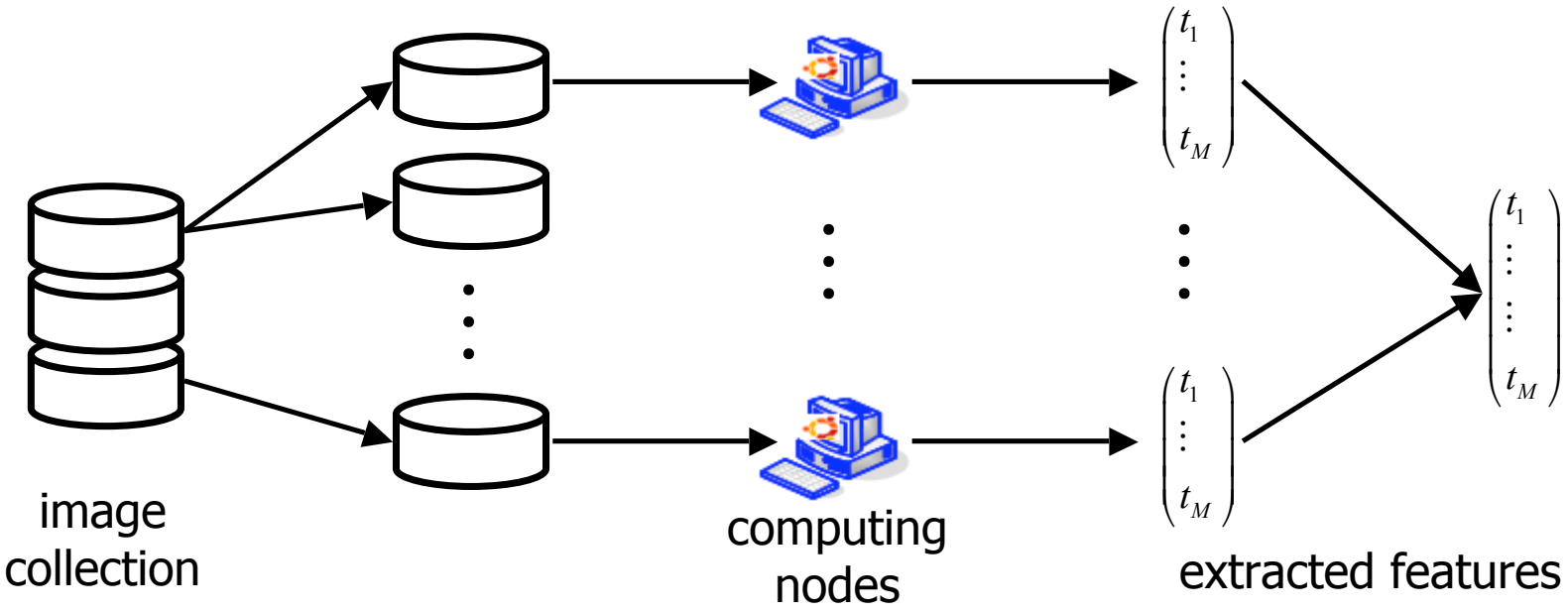


Overview

- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

An application

- Offline feature extraction for the GIFT system was parallelized
  - N packages containing images and their processing instructions are sent to computing nodes
  - results are recovered from the nodes



An application

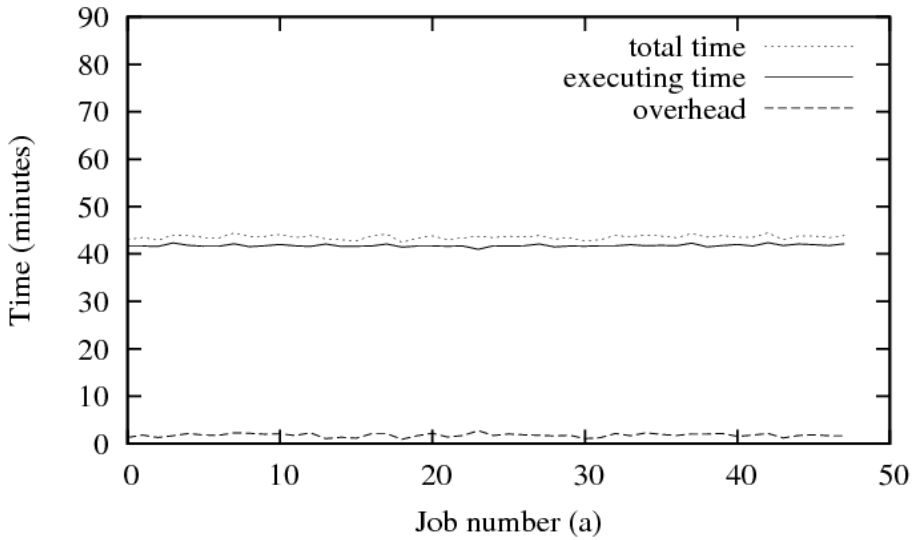
- Comparison of execution times for 50 packages of 1000 images:

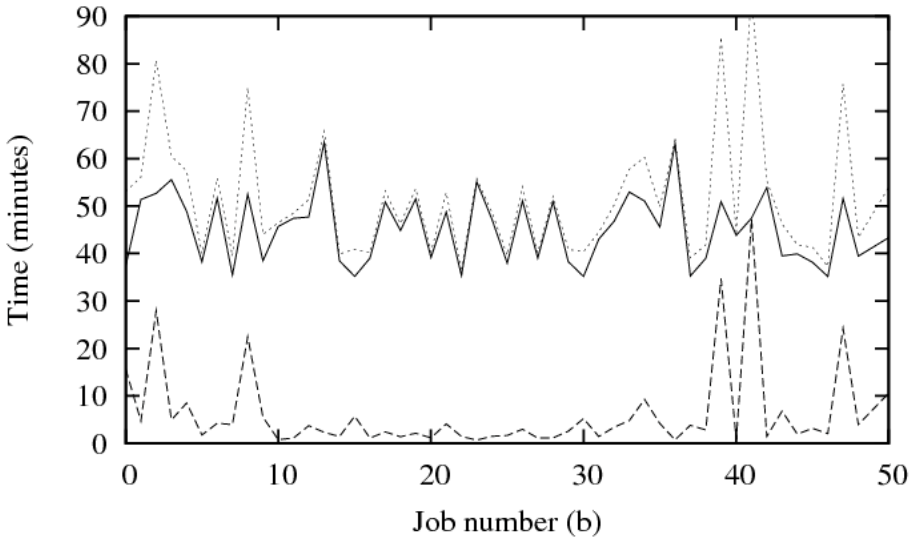
Local server	709 min
37 CPUs remote clusters (KnowARC resources)	537 min
Local cluster, 10 CPUs	240 min

- ✓ Execution time was reduced
- ✓ Though the local cluster had less CPU's, the results were better – less transfer and queue time (no waiting time with local resources)

An application

- Local (a) versus remote (b) execution:





- ✓ Overhead was reduced from 10 min per package to 1 min
- ✓ Higher constancy of the execution times

Overview

- Motivation: Grid computing for medical image analysis
- Challenges inside hospitals
- Solution guidelines
- An application
- Conclusions

Conclusions

- A desktop Grid-like infrastructure was deployed inside the hospitals
- Infrastructures, Grid software and political problems were identified
- Little manual intervention was required because of virtualization and automatic software distribution
- Efficiency was improved compared to using external resources