The Fruitful Application of Volunteer Computing to Regions with Low Scientific Funds: Extremadura as an Example

Miguel Cárdenas on behalf of Collaboration

Ibergrid 2008
Porto / 12 - 14 May

Summary

- Why Volunteer Computing and why in Extremadura
- Radio Network Design
- Neural Network Simulation
- Virtualization
- Conclusion

Why Volunteer Computing

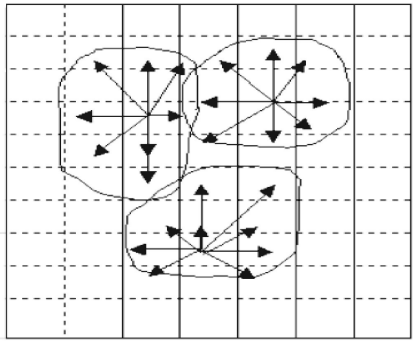
- Volunteer Computing emerges as a natural way to extract an extra life to computing.
- To enrich the research in a region with low level of public funds.
- This work explains the works in VC for convincing politicians to adopt it.
  - To proof (to politicians) that VC is not intrusive nor destructive (not a lot).
  - To proof (to local researchers) that VC is useful for research.

Why Volunteer Computing

- Volunteer Computing emerges as a natural way to extract an extra life to computing, specially taking into account the 70 k PCs in secondary schools at Extremadura.
  - Identical operating system (GNU/Linex, a variant of GNU/Linux).
  - Network connectivity.
  - There are technicians in every centre of secondary school.
  - Central software installation facility.

RND

- Radio Network Design is an NP-hard combinatorial problem that consists in determining a set of locations for placing radio antennas (BTS) in a geographical area in order to offer high radio coverage using the smallest number of antennas:
  - its resolution by means of evolutionary algorithms is very appropriate,
  - carrying out thousands of experiments



$$f(x) = \frac{CoverRate(x)^a}{NumberTransmittersUsed(x)}$$

RND

- In these experiments:
  - Many different evolutionary algorithms (Population-Based Incremental Learning, Differential Evolution,...) are possible and it is necessary to find an evolutionary algorithm that obtains good solutions to this problem.
  - Generally, all these evolutionary algorithms have a high number of configuration parameters, and it is important to obtain the best configuration of these parameters.
  - In order to obtain general results, it is necessary to test the RND problem with numerous variants (different antennas type, different coverage radius, etc.).
  - Takes in being executed from minutes to hours.

RND

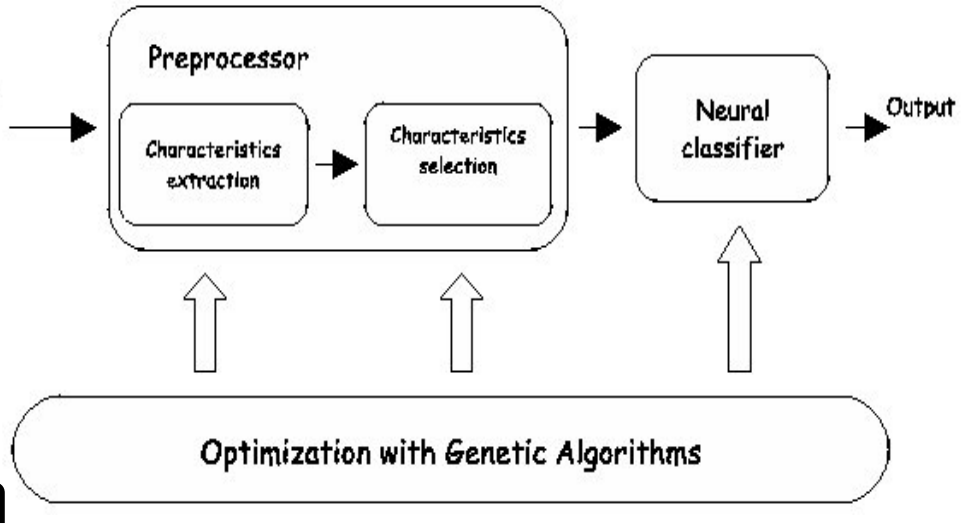
- Solving this problem:
  - 100 computers of University were attached to RND project,
  - during the month of August all the BOINC clients were available 24 hours per day, but this is not true for the rest of months,
  - 25280 different experiments planned: different antenna types, and different coverage radius (20 and 30),
  - 32454 hours (3.71 years) necessary in order to perform it,
  - all experiments concluded in 6 months.
  - During August of 2006, the 100 computers used were running 24 hours per day, a 5% of hardware breaks (1 processor and 4 power supply units burned) were produced.

NNSIMU

- In this section we describe the problem of optimising neural classifers and its application to diagnosis of microcalcification clusters (MCCs) in mammograms.
- The problem deals with optimising the input space to neural classifiers using Genetic Algorithms. It is important to remark that we do not train the Neural Network using Genetic Algorithms.

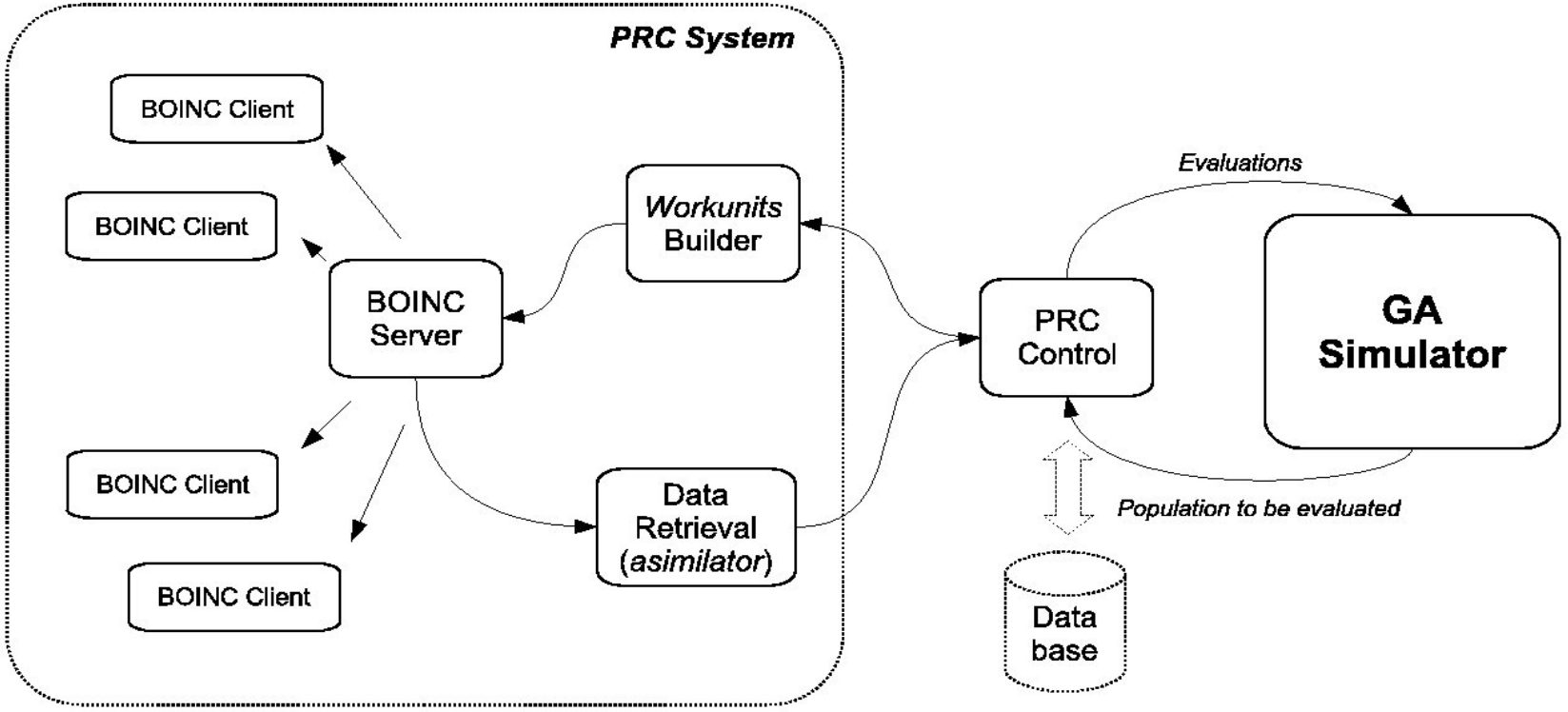
NNSIMU

- In our case of MCCs diagnosis, the input to the system consists in one or more regions (patches) around the pixel that we want to diagnose (classify), while the output is the probability for that region to be considered as malignant. The sizes of the patches are parameters to be optimised.
- The characteristics extraction is carried out by using Independent Component Analysis (ICA). However, before applying ICA, Principal Component Analysis (PCA) is used with the objective of reducing the initial dimension of the problem.



NNSIMU

- Pure BOINC architecture.

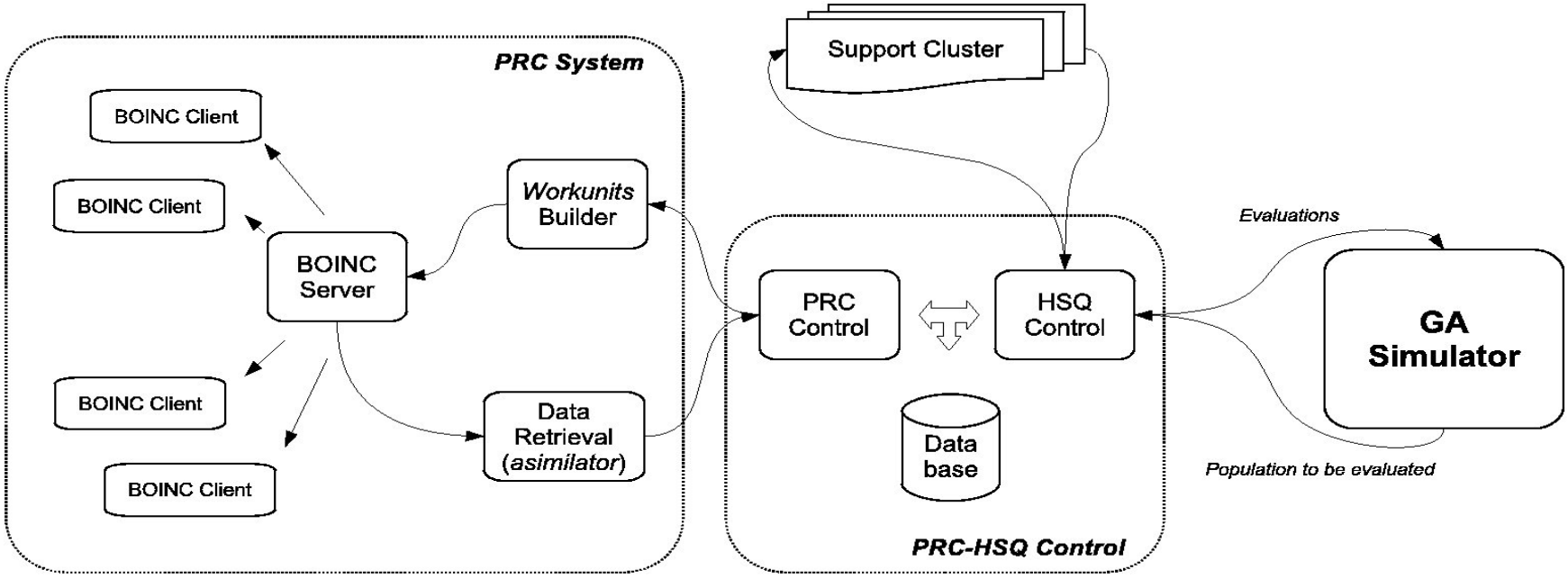


Modification of the GA simulation library for executing the evaluation function corresponding to each individual for the active generation

GA sends the whole population to the PRC control module and waits until this module returns individuals evaluation.

NNSIMU

- BOINC architecture with support cluster.



Schema similar to the previous one, but it includes a modified PRC control module (PRC-HSQ control) and a centralized and defined characteristic. As previously stated, the function of this module is to decide if the work must be assigned to the support cluster instead of submitting it to the BOINC based PRC system.

NNSIMU

- At this moment, our work is centred on porting and testing the Neural Network (NN) simulation application under BOINC.
- The application should train a NN with a specific input configuration and give back the result over the validation set. This value will be used as fitness

NNSIMU

- The NN simulation, based on the Stuttgart Neural Network Simulator library, has been ported to BOINC. The most important steps to carry out this port were:
  - Input and output files: changing the library functions for using BOINC API.
  - Validation: NN training is itself an iterative process for finding a local minimum, and the training usually does not provide exactly the same values. So a specific validator developed.
  - Submission and assimilation: submission and assimilation subsystems developed, suitable for our application.

Virtualization

- Virtualization techniques can solve important issues when deploying a BOINC production infrastructure.
  - Urgent finishing of a production stage. Consider a scenario in which a production stage must be completely finished in order to generate the new set of jobs for a next stage based on the previous one. The volatile nature of BOINC resources makes impossible to assure a definitive deadline for a set of declared jobs. Virtualized clients can endorse the finishing of the production stage in the time scheduled.
  - Quick deployment of an infrastructure. Virtualization techniques allow the setting up of a whole production infrastructure in a few minutes.
  - Dynamic configuration of infrastructures. On grid environments, in situations on which most of the Working Nodes are not used BOINC clients can be quickly deployed.

Virtualization

- Virtualization provides the ability for quick deployment of a large number of BOINC clients and for taking advantage of unused resources with BOINC computation.
- For this virtualization, XEN was used.
- This set of virtual clients have been used in projects such as: ZIVIS, RND, NNSIMU and others, in the early stage of them and when became public.

Conclusions

- In this paper we have presented the works done by several groups from University of Extremadura and CETA-CIEMAT using volunteer computing.
- As our results show, volunteer computing is a very interesting option for grid computing, enabling to obtain important scientific results in regions with low financial resources, like Extremadura.
- These works have been used in order to convince politicians to adopt Volunteer Computing as a way to develop R+D+i (Research + Development + Innovation), beside the scientific value of the activities carried out.

Conclusions

Muito Obrigado

Gracias

Thanks