Report 11 GridPP: James Cunha Werner Manchester, 1/04/2007

Jan-Mar/2007: EasyGrid product development

1. EasyGrid and grid scheduling optimisation using Genetic algorithms.

There are thousands of papers applying genetic algorithms to scheduling optimization of resource brokers, grid resources and load balance, job shop scheduling, manufacturing, etc – paramount in a overload computational environment with heterogeneous resources such ours. It is a class of NP hard problems with huge solution space, impossible to be solved searching exhaustively all the possibilities. Modern heuristics have been successfully applied with constraints due unfeasible solutions.
I developed an always-feasible representation to cope with genetic operators disruptive effect in problem's constraints, and compared Monte Carlo technique and genetic algorithms to find the best schedule of 161 benchmarks (fig. 1). EasyGrid submitted 322 jobs, and recovered 312 without problem and 10 grid error listings for further analysis. Each MC job created thousands of millions instances to obtain the same solution obtained with GA evaluating 50 million instances for each benchmark (4 days running in 322 computers at Manchester tear2, instead years in my desktop computer!).
There were 10 jobs that expended more than 2 days (2000 decisions), and PBS killed them. There is a bug in LCG resource broker that resubmit 3 times the job again when it fail by processing time, making things worse (PBS will kill them 3 times after 2 days of processing). This error was reported 3 years ago, and still not solved.
Despite GA is considered a good search tool by scientific community, my implementation is not able to find the global optimal solution for all benchmarks, with a clear correlation with complexity. There are analysis discussing the crossover disruptive effect, and early convergence to local maximum (possible to explain due the low density of better states in Markov chain process and its stochastic characteristics). Eventually my policy keeping best individuals from different generations increases early convergence.
Another important information is model decision-making complexity and chromosome size effect (2^(10*number of decisions)) and fitness function smoothness.

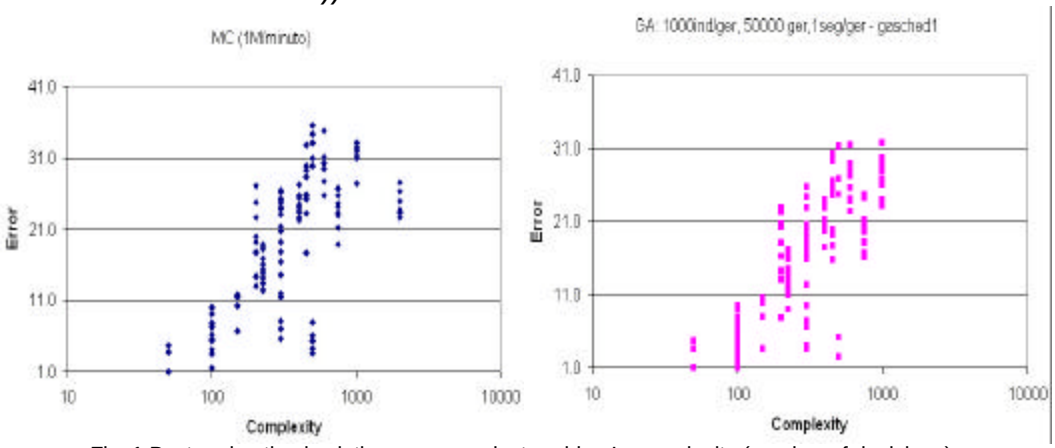

Fig. 1 Best and optimal solutions error against problem's complexity (number of decisions).

2. Grid results visualisation.

The amount of grid results (when you are working with EasyGrid) is astonishing and complex to analyse when working with genetic programming. Instead to have ntuples, which users just have to add in one consolidate plot, genetic programming provides several knowledge-based models.
Usual tools (such as Mathematica) could help in consolidating the results from several models using their functional programming capabilities. However there are problems, such as their smooth multi parametric plots add information where they are not available, making analysis difficult.
To overcome these difficulties, I am developing a framework in Visual C++, integrating adaptive algorithms, genetic algorithm, and genetic programming with a more flexible visualisation tool that will be able to merge functional models in a larger functional set for truth analysis. I am able to import solutions from grid, and use them as start point for further analysis, or even improve the combined solution obtained previously (known as island model).

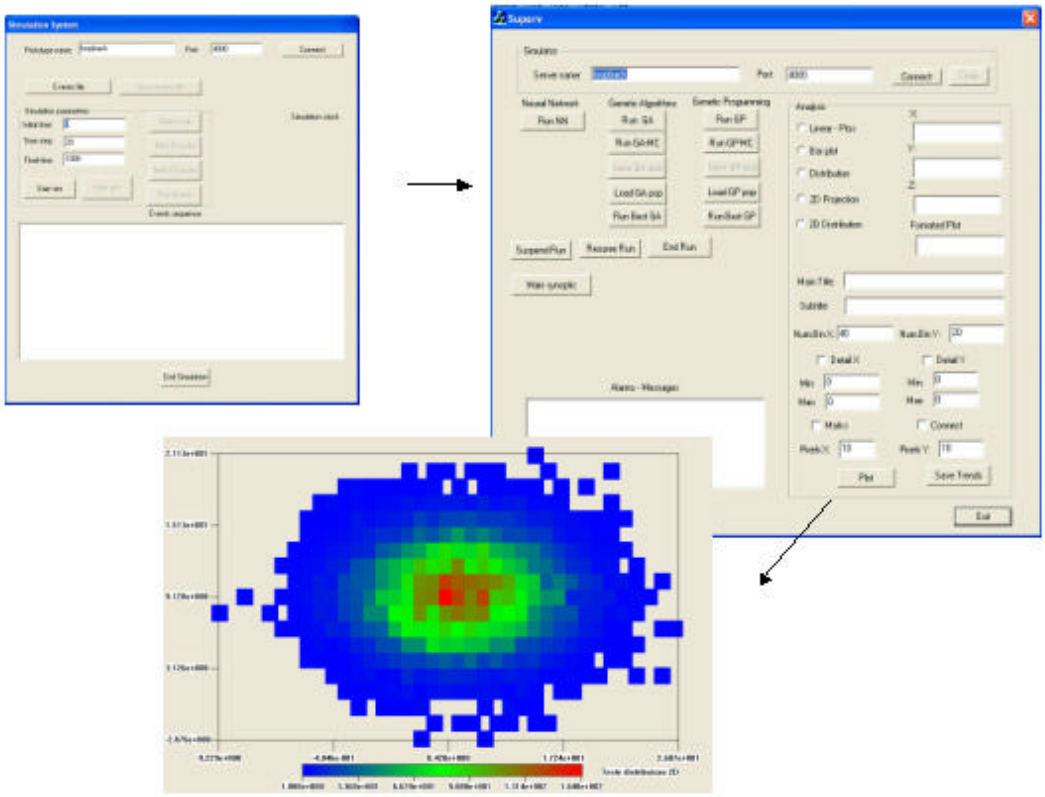

Fig 2 Simulation and visualization software for grid results analysis.

3. EGEE User Forum

The plenary talk 'EasyGrid: a job submission system for distributed analysis using grid" was accepted, and my slides are available at http://indico.cern.ch/materialDisplay.py?contribId=9&sessionId=20&materialId=slides&confId=7247

The talk shows all successfully results in all benchmarks and users data, not only in data gridification but also in task gridification with genetic algorithms and genetic programming. Neutral pion discriminate function applied to discriminate real neutral pions from background will support my proposal to study Higgs to gamma gamma using the same approach. This project will be amazing, because this is a clean decay without neutrinos, and Higgs mass and characteristics will be determined without any doubt.

4. MPI training at Manchester Computing.

I have attended the following free courses in MPI:

-MPI advanced
-MPI One sided communication
-MPI IO

I am developing a MPI version of my genetic algorithm and genetic programming codes, to replace the PVM version (obsolete due time consuming openssh access with X509 certificates).

5. Other activities

-Gridpp18 at Glasgow.
-poster in Gridpp stand at IoP2007.

6. Future activities

-Develop one chaotic optimisation problem using genetic problem to cover all optimisation taxonomy. I have developed several others applications in previous postdoctoral projects. Portfolio optimisation using quadratic economic function and low volatility could be a very interesting problem to model stock selection decision-making process using genetic programming.

-Develop a complete framework for genetic and adaptive algorithms in grid, with all libraries for data maintenance and ontology model, Backus-Naur parser to model grammatical structures in genetic programming, and graphical interface for parameterisation and visualization.

-Write a better project about Evolutionary discriminate functions to discriminate Higgs from background in Atlas experiment. I would appreciate if BaBar software and data is available **for real** at Manchester Tier 2, to make better plots of hadronic tau decays with neutral pions and have better ground for arguments next month when I will go fight for funding.

-Laughing at Roger's pathetic alibabar and slashgrid waste of taxpayer's money. I hope Roger will finally realise he does not have competence to develop my work, and will extend my contract and allow me to do the job he offered me to convince me leave Imperial College 3 years ago. Otherwise, find a new job where I could use my skills to solve LCG pathetic performance and data bottleneck, conventional software gridification, and support real users.