GridMOSI VO – a Grid enabled Research Infrastructure in Modelling, Simulation and Optimisation

GABRIEL NEAGU
Department of Research and Development
National Institute for Research and Development in Informatics - ICI
8-10 Averescu Ave., 011455 Bucharest
ROMANIA

Abstract: The GridMOSI project funded by the National Research of Excellence Programme aimed at providing Grid enabled solutions for high performance modelling, simulation and optimisation problems. The adopted project approach is in line with current orientations at the national and European levels in building up advanced research infrastructures based on the eInfrastructure concept.
The paper presents the components of the GridMOSI pilot architecture, stressing on the project Grid infrastructure (coordination of administration and operation activities - DrC. Alexandru STANCIU – ICI Bucharest) and the gridified applications in the following fields: computer-aided advanced system modelling and control - scientific coordinator Dr. V. SIMA, ICI Bucharest; unconstrained optimisation - Dr. N. ANDREI, ICI Bucharest; complex numerical modelling and MDO in industrial applications - Dr. C. NAE, National Institute for Aerospace Research Bucharest; evolutionary multiobjective optimisation - Prof.dr. Dana PETCU, West University of Timisoara; applications scheduling using genetic algorithms - Prof.dr. V. CRISTEA, University "Politehnica" of Bucharest; cryptographic and cryptanalytic algorithms - Prof.dr. Rodica POTOLEA, Technical University of Cluj-Napoca. The GridMOSI project results configure an open framework that should be further extended in both Grid capacity and application diversity respects.

Key-Words: research infrastructure, Grid resources, virtual organisation, computer-aided system modelling and control, unconstrained optimization, complex numerical modelling, evolutionary multiobjective optimization, genetic algorithms, cryptology.

1 Introduction

Modelling, simulation and optimization (MOSI) procedures and tools are widely used for design and operation of advanced man-made systems, as well as for investigation or improvement of all kind of complex technological, economical or biological processes.
Grid computing is considered one of the most promising approaches to solve in a convenient way large scale computing intensive problems. The Grid concept referrers mainly the coordinated and transparently distributed manner of problem solving capacity at the level of virtual, dynamic, multi-institutional organisations.
According to a basic definition given in [1], virtual organisations (VOs) "enable disparate groups of organizations and/or individuals to share resources in a controlled fashion, so that members may collaborate to achieve a shared goal". Shared resources are physical (computers and storage provided through the Grid infrastructure) and logical ones (data, application software supporting VO community objectives). The physical resources may be shared between several VOs, while logical resources are specific to a given VO.
At the European level, the Gird and associated technologies are considered an essential pillar to build the eInfrastructure kind support for implementing the European Research Area concept [2]. At the national level, the strategic document entitled "*Pilot project for the implementation of the national Grid infrastructure*", states about the Grid infrastructure as an essential component of the knowledge-based society [3].
The objective of the GridMOSI project was to create a VO based on Grid technology for high performance modelling, simulation and optimization aiming to provide a convenient access of various users and organizations to powerful computing and computational resources and associated software tools [4]. This community includes both application developers using Grid infrastructure resources and application end-users accessing available MOSI applications to solve their concrete problems.
The project consortium includes research and academic organisations with a sound expertise at the

national level in project areas of competence (Grid infrastructure administration and operation, Grid computing, mathematical modelling, simulation and optimisation): National Institute for Research and Development in Informatics – ICI Bucharest (Department of Research and Development), University "Politehnica" of Bucharest - UPB (National Centre for Information Technology), National Institute for Aerospace Research – INCAS Bucharest (Subsonic Wind Tunnel Laboratory), Technical University of Cluj-Napoca – UTCN (Department of Computer Science) and West University of Timisoara – UVT (Department of Computer Science).

The paper presents the project experience as the first attempt at the national level to implement a Grid enabled research infrastructure in case of a largely referred research topics: modelling, simulation and optimisation. Section 2 introduces the project approach in achieving the project objective, considering project resources, available competences at the consortium level, current trends and results in scientific international cooperation in this field. Section 3 presents the project results in implementing a powerful and operational Grid infrastructure, including 5 Grid resource centres and the virtual organization. Section 4 is devoted to the most significant project output: the set of gridified MOSI applications, putting the stress on problem description, grid related performance and benefits, potential scientific and social impact. Some concluding remarks are formulated in the final section.

2 Project Approach

The major aim of the project was to provide an integrated view on Grid technology capabilities to support the advanced research infrastructure facilities: shared resources, transparent but secure access for all interested users, distributed architecture, support for virtual collaboration. The project activity was structured according to the following specific objectives:

• consolidate the Grid resource centres at the level of partner organizations, operate and maintain the project Grid infrastructure;

• set-up the virtual organization for the MOSI domain as the first contribution of this kind to the national research infrastructure, demonstrate the feasibility and benefits of this approach;

• migrate to the Grid environment a relevant set of the available MOSI applications selected from the project partners' offer.

Therefore the following components were included in the GridMOSI pilot architecture: Grid resource centres, virtual organisation, gridified applications, and project portal [4].

The configuration of the Grid resource centres was decided with the main idea to provide a balanced infrastructure, taking advantage of existing resources at the project consortium members level and of their geographical distribution.

The virtual organization as architectural component of the Grid infrastructure is an innovative contribution of the project, being the first attempt of this kind at the national level. The main reason for adoption of this architectural solution was the VO support to set-up a user community and to provide to its registered members a flexible and secure distributed environment to use their data and applications.

As regards the gridified applications, the decision of their migration to the Grid environment was mainly justified by the necessity to access more computing and storage capacity for solving large scale problems, to better fit in with the distributed nature of some target problems, to improve the quality of services provided by the Grid infrastructure itself to the application developers and users, to increase the applications visibility and accessibility for potential user community [5].

The portal has been playing the role of the main dissemination tool about project evolution and results. Interested users are provided with information about the VO membership procedure and MOSI applications, as well as links to web based interfaces for applications access.

Also, during the project life the portal provided useful support for collaborative work at the level of geographically distributed project team.

3 GridMOSI Infrastrucure

The project infrastructure was set up with the active contribution of all partners and includes the following sites, according to their national code: RO-01-ICI, RO-03-UPB, RO-05-INCAS, RO-08-UTCN and RO-09-UVT.

All sites had previously incubated their administration, operation and interoperability capabilities in the regional Grid infrastructure for South-East Europe, administrated by the EC funded SEE-GRID project.

At the project start an important budget share was allocated to extend their configurations and improve their networking connectivity. Overall, the GridMOSI sites currently account for 91 CPUs with an equivalent of 130 kSI2k computing power, and 9 TB memory.

During the project life, the objective was to improve the performance indicators of this infrastructure in order to demonstrate the feasibility of the future national Grid infrastructure. Thus, common administration, operation and usage rules were adopted, such as *Service Level Agreement* between Grid sites and Grid Infrastructure enforcing site availability and reliability measures, or *Acceptable Use Policy* stating about the conditions of the expected and acceptable usage of VO or Grid resources. The idea is that every VO running on the infrastructure must have its own acceptable use policy, and must enforce its acceptance by each VO member. A related document is the *VO Security Policy* which defines the responsibilities of the VO Manager and VO members as regard to the secure usage of the VO resources.

Also, most of the core infrastructure and VO services were implemented on project sites in order to improve the operational autonomy of the infrastructure.

The GridMOSI VO was registered into the VO repository of the Europe wide Grid infrastructure for eScience with the aim of promoting the hosted MOSI applications as support for further development of scientific cooperation in this domain.

The Grid infrastructure administration and operation was coordinated by DrC. Alexandru Stanciu – ICI.

4 MOSI Applications

The following MOSI domains have been addressed in the project: computer-aided advanced system modelling and control (scientific coordinator Dr. V. Sima - ICI), unconstrained optimisation (Dr. N. Andrei - ICI), complex numerical modelling and MDO in industrial applications (Dr. C. Nae – INCAS), evolutionary multiobjective optimisation (Prof.dr. Dana Petcu - UVT), applications scheduling using genetic algorithms (Prof.dr. V. Cristea - UPB), cryptographic and cryptanalytic algorithms (Prof.dr. Rodica Potolea - UTCN).

4.1 Computer-aided advanced system modelling and control

The general control systems-related topics, including implementation issues, are dealt with in [6]. Often, these models are too large to be used effectively and economically. Given large order, state-space models, the **GridModRed** application for **model order reduction** enables to find reduced order models which preserve the essential dynamical properties of the original models [5]. GridModRed includes algorithms working on linear time-invariant dynamical system models. Given a state space system of order n, $G := (A,B,C,D)$, with the *transfer function matrix* (TFM) $G(\lambda) = C(\lambda I-A)^{-1}B + D$, where $A, B, C,$ and D are $n\times n$, $n\times m$, $p\times n$, and $p\times m$, respectively, and λ is either the complex variable s of the Laplace transform, for continuous-time systems, or the complex variable z of the Z-transform, for discrete-time systems, GridModRed enables to find an approximate model $G_r := (A_r,B_r,C_r,D_r)$ of order r ($r < n$), with TFM $G_r = C_r(\lambda I-A_r)^{-1}B_r + D_r$. Balance and truncate methods, singular perturbation approximation, and Hankel norm approximation methods are implemented.

The **GridIdent** application for **finding mathematical models via system identification** includes algorithms operating on existing or generated (by dedicated experiments) input-output data of dynamical systems or processes. The application generates models of linear time-invariant (LTI) discrete-time multivariabile dynamical systems, described in state space. Subspace techniques, MOESP (*Multivariable Output Error state SPace*), N4SID (*Numerical algorithm for Subspace State Space System IDentification*), and their combination are used.

The **GridWident** application determines multivariable discrete-time state-space models of nonlinear Wiener-type dynamical systems using input-output data records. Specialized Levenberg-Marquardt algorithms are implemented for estimating the parameters of the nonlinear part and then of the whole system.

4.2 Unconstrained optimisation

SCALCG and CGALLP belongs to the high performance unconstrained optimization library BIBFR, developed in the GridMOSI environment and providing the fundamental support for **large-scale unconstrained optimization**, which incorporate major algorithmic developments and state-of-the-art solutions for a board range of unconstrained problems [7, 8].

SCALCG implements a **scalable conjugate gradient algorithm** in order to solve the unconstrained optimization issues.

CGALLP is a **nonlinear unconstrained optimization package** implementing a number of 23 conjugate gradient algorithms. The scalar product is implemented in a parallel manner.

Both BIBFR components proved effectiveness in solving real industrial issues. For CGALLP the following use cases are included in the BIBFR library: Elastic-Plastic Torsion Problem, Pressure Distribution in a Journal Bearing Problem, Optimal Design with Composite Materials Problem, Inhomogeneous Superconductors. 1-dimensional Ginzburg-Landau Problem, Steady State Combustion. Solid Fuel Ignition, Lennard-Jones Cluster Problem, Minimal surface area.
The user is provided with a number of easy to access parameters allowing customized optimization conditions. He/she has to type a subroutine which calculates function value and its gradient, as well as the initial point.

4.3 Complex numerical modeling and MDO

The **OpTG - Optimizer based on Grid Technology** application is dedicated for large optimization problems where the computation of the cost function needs evaluation performed by complex applications requesting high performance computing resources [5].
Cost function definition and initial set-up are generally related to the user experience and several templates might help in order to proper correlate the optimization problem with the cost function definition and algorithm choice.
Evaluation of the cost function is based on local integration of computed data from field codes, where the basic methodology is related to domain discretization and primitive variables evaluation using partial differential equation systems.
The implementation on Grid is highly justified. At international level this is an interesting and cost-effective alternative to sensitive high performance computing applications implemented only on very expensive supercomputers [9]. At national level, in aerospace community, this is a major step beyond the state of the art tools and methods, mainly due to the complexity of the problems that may be considered, geometries, simulations and computational requirements. It gives the possibility to move from experimental design towards real industrial application in aerospace industry [10]. The current implementation contains the following algorithms: a gradient based method as standard for optimization process, a genetic algorithm as an option for complex optimization, templates for several optimization problems based on cost functions expressed as integrals of flowfield variables (e.g. pressure) on surfaces that are computed using field simulation based on geometrical discretization of the domain (e.g. CFD analysis), scripts for specific computations as on-demand services on the Grid for simulation codes.

MFCC – Modern Field Code Cluster is a cluster of applications using state of the art numerical models in flow physics and advanced techniques in domain decomposition and parallelization. MFCC aim is to enable that any flow physics and related problem may be solved using existing implemented resources. This includes all three phases in such activities: pre-processing, solver and post-processing. Existing codes in various experimental development phases were considered for integration based on specific adaptation and gridification performed in GridMOSI project.
The basic orientation is towards CFD (Computational Fluid Dynamics) simulations based in advanced discrimination of the domain. Therefore, as pre-processing phase, dedicated tools and instruments are included for geometry definition, mesh generation, parcelation and domain decomposition.
Solver is based on Navier-Stokes equation solver, so that a wide choice of turbulence models can be accommodated.
Post-processing tools are intended for detailed analysis using state of the art dedicated software (e.g. TecPlot) through a set of templates and scripts.
The basic need for a MFCC is coming from the increasing demand for extended simulation capabilities using a common core of tools and methods. Grid implementation is a natural solution that enables both unified access and optimization of existing resources involved in complex simulations.

4.4 Evolutionary multiobjective optimisation

DEMO/G - Distributed Evolutionary Multiobjective Optimization on Grid addresses the problem of efficiently solving multi-objective optimization problems by using parallel and distributed implementations of evolutionary algorithms [5]. Let $f_1,f_2,\ldots,f_r$ be r objective functions each one depending on n decision variables. The aim is to find all decision vectors $x=(x_1,\ldots,x_n)$ which are nondominated with respect to all r criteria; x is considered nondominated if there does not exist any vector y such that y is better than x with respect to all criteria.
The current implementation of DEMO/G allows the execution on different environments: single nodes, clusters and grid [11]. Each colony contains a set of communicating populations, can be characterized

by its specific evolutionary algorithm and control parameters, and can be executed on a different site in Grid. DEMO/G has a modular structure having specific modules corresponding to: the evolutionary algorithms, the communication strategies and the concrete problem descriptions. Currently there are implemented two different multiobjective evolutionary algorithms (MOEAs), five variants of dividing the search space and four communication strategies.

The implementation on Grid of multi-objective evolutionary algorithms allows to: conduct experimental design of MOEAs in order to efficiently explore the stochastic character of these algorithms; reduce the high computational cost of sequential implementations by dividing a large population of candidate solutions in smaller subpopulations; deal with problems having a distributed character (e.g. mining of distributed data).

4.5 Applications scheduling using genetic algorithms

DIOGENES - DIstributed near-Optimal GENEtic algorithm for grid applications Scheduling aims to achieve a distributed, fault-tolerant, scalable and efficient method for optimizing task assignment in Grid environments [12, 13].

For real-time task scheduling, the monitoring information on resource description is used. The general assumptions are heterogeneous and dynamic environment, decentralized scheduling. A near-optimal schedule is computed by the Scheduler based on the Scheduling requests and the Monitoring data provided by the Grid Monitoring Service (MonALISA). The schedule is then sent as a Request for task execution to the Execution Service. The user receives feedback related to the solution determined by the scheduler, as well as to the status of the executed jobs in the form of the Schedule and task information.

The scheduler interacts with Grid Monitoring, Execution and Discovery Services. The Grid Monitoring Service has the specific purpose to obtain real-time information in a heterogeneous and dynamic environment such as a Grid. The Execution Service can send execution requests to an already installed batch queuing system on the computing node to which a particular group of tasks was assigned. Lookup processes are triggered by the Discovery Service and determine the possibility of achieving a decentralized schedule by increasing the number of hosts involved in the genetic scheduling.

The main innovative contributions of DIOGENES include: a decentralized scheduling method for tasks scheduling in distributed systems, based on genetic algorithms; an agent oriented, fault tolerant scheduling platform; a scalable scheduling system for dynamic environments; a new algorithm for DAG Scheduling (*ICPDP-Improved Critical Path using Descendant Prediction*); multiple user support offering a simple user interface and a special agent (broker) for user management.

New application features or functional benefits obtained from the grid include: multi-criteria optimization of Grid scheduling, task classes with complex dependencies, new scheduling algorithms for real-time scenarios, backup and recovery from service fails (re-scheduling), optimized file transfer solving the co-scheduling problem.

4.6 Computer-aided advanced system modelling and control

The following main classes of cryptographic and cryptanalytic algorithms were studied and adapted for grid execution in the **CryptoGrid** library: block ciphers, stream ciphers, public key algorithms, hash functions, random number generators, randomness testing algorithms and factoring algorithms [14, 15]. Each class contains several algorithms that are well known and very popular in the research and industrial community: block ciphers (Mars, RC6, Rijndael, Serpent and Twofish), stream ciphers (RC4), public key algorithms (RSA - key generation, encryption, decryption), hash functions (SHA-1, SHA 224, SHA 256, SHA 384 and SHA 512), random number generators (LCG, MODEXP, BBSG, MSG, XORG, QCG1 and CCG), randomness testing algorithms (Block Frequency, Bit Stream, 3D Spheres, DNA Test, Squeeze, Runs and Birthday), factoring algorithms (QS and GNFS). A new taxonomy for the Grid, derived from Flynn's taxonomy, was proposed and eight basic execution modes (two serial modes and six parallel modes) derived from the taxonomy were introduced, as most significant innovative contribution of this GridMOSI application.

The selected algorithms were executed in scenarios based on several execution modes, which were identified and clearly defined based on the proposed taxonomy. Other algorithms are being considered for further implementation. Performance measurements were made, and performance increase was studied.

The practical results of this application consist in: implementing the cryptographic and cryptanalytic algorithms mentioned above and identifying ways to optimize them by parallel execution on the Grid; finding of the optimal read buffer size.

5 Conclusion

The project represents the first attempt at the national level to make available for a potentially large user community the benefits of using Grid infrastructure facilities and services in the field of modelling, simulation and optimization. Grid resource centres and VO services are illustrative for the successful implementation of the eInfrastructure concept, based on the network and Grid technologies.

GridMOSI applications are relevant for the scientific expertise of the project partners in this field, for the diversity of application migration approaches, and for the usability potential of this kind of solutions.

The GridMOSI project results configure an open framework that should be further extended in both Grid capacity and application diversity respects, with the contribution of either current partners or other organizations interested to take advantage of the accumulated expertise in setting up and using this advanced research infrastructure solution.

References:

[1] I. Foster, C. Kesselman, S. Tuecke. The Anatomy of the Grid Enabling Scalable Virtual Organizations, *International J. Supercomputer Applications*, 15(3), 2001.

[2] *** Changing the way research is done. The user perspective on eInfrastructure, *EC, DG INFSO,* Bruxelles, December 2003, p. 3.

[3] *** Pilot project for the implementation of the national Grid infrastructure – Strategic Plan, *ANCS Task Force for Grid Infrastructure*, Bucharest, February 2006, p. 10.

[4] G. Neagu, N. Andrei, V. Sima, A. Stanciu, N. Tapus, V. Cristea, C. Nae, R. Potolea, D. Petcu, GridMOSI Virtual organization – component of the National Research Infrastructure, *Proceedings, AMCSIT Annual Conference on Research of Excellence,* Brasov, 22-24.10.2006, Printech 2006, Paper L4-23 (in Romanian).

[5] G. Neagu, N. Andrei, V. Sima, V. Cristea, C. Nae, Rodica Potolea, Dana Petcu, A. Stanciu, "Grid Enabled Applications for Modelling, Simulation and Optimization", *Proceedings, Conference CEEX 2007 "Excellence Research as a way to E.R.A", Nicolae Vasiliu, Lanyi Szabolcs (Eds.),* Brasov, Oct. 25-26, 2007, 95-1 - 95-6. Editura Tehnica. ISSN 1843-5904.

[6] S. Van Huffel, V. Sima, A. Varga, S. Hammarling, F. Delebecque, Development of high performance numerical software for control, *IEEE Control Systems Magazine*, Vol. 24, No. 1, Feb. 2004, pp. 60-76.

[7] N. Andrei, A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Applied Mathematics Letters*, Vol. 20, 2007, pp. 645-650.

[8] N. Andrei, Performance of conjugate gradient algorithms on some MINPACK-2 unconstrained optimization applications, *Studies in Informatics and Control*, Vol. 15, No. 2, 2006, pp.145-168.

[9] W. Rehm, C. Nae and B. Wang, Modern Computational Fluid Dynamics of Complex Flows with Supercomputing for Safety Research, *Int. Conference on Supercomputing in Nuclear Applications-SNA'2003*, Paris, 22-24.09.2003.

[10] C. Nae, V.M. Pricop, E. Munteanu, Buffeting Alleviation using Active Flow Control, *2nd European Conference for Aero-Space Sciences–EUCASS*, Brussels , 1-6.07.2007.

[11] S. Panica, Daniela Zaharie, Dana Petcu, Evolutionary Multiobjective Optimization in Grid Environments, *Proceedings of IASTED Int. Conference on Parallel and Distributed Computing and Networks*, Innsbruck, 12-14.02.2008.

[12] F. Pop, V. Cristea, Intelligent strategies for DAG scheduling optimization in Grid environments, *Proceedings of the 16th International Conference on Control Systems and Computer Science*, Bucharest, May 2007, pp. 98-103, ISBN 978-973-718-743-7.

[13] F. Pop, C. Dobre, G. Godza, V. Cristea, A Simulation Model for Grid Scheduling Analysis and Optimization, *Proceedings of Int. Symposium on Parallel Computing in Electrical Engineering (PARELEC'06)*, September 13-17, 2006, Bialystok, Poland, Published by IEEE Computer Society, pp. 133-138, ISBN: 0-7695-2554-7.

[14] Rodica Potolea, A. Suciu, A. Mascasan, Benchmarking the GridMOSI Library, *Proceedings of the Int. Conference on eChallenges*, The Hague, Oct. 24-26, 2007.

[15] Rodica Potolea, A. Suciu, Finding the Optimal Read Buffer Size for Grid Applications, *Proceedings of 9th Int. SYNASC*, *Workshop on Grid Computing Applications Development*, Timisoara, 26-29.09.2007, pp. 51-54.