Clinical Trial Simulation in Grid Environments

Dimosthenis Kyriazis, Andreas Menychtas, Dimitra Dionysiou, Georgios Stamatakos, Theodora Varvarigou

***Abstract*—A constantly increasing number of applications from various scientific sectors are finding their way towards adopting Grid technologies in order to take advantage of their capabilities: the advent of Grid environments made feasible the solution of computational intensive problems in a reliable and cost-effective way. The aim of this paper is to demonstrate how multilevel tumour growth and response to therapeutic treatment models can be used in order to simulate clinical trials, with the long-term intention of better designing clinical studies and understanding their outcome based on basic biological science. For this purpose, a computer simulation model of glioblastoma multiforme response to radiotherapy has been applied to perform the aforementioned simulation in a real Grid environment by also taking into account historical data. The proposed approach yields very good results for the conducted virtual trial since these are in agreement with the outcome of the real clinical study, while the use of Grid technologies demonstrate and highlight their added-value.**

I. Introduction

GRID computing [1] is increasingly considered as an infrastructure able to provide distributed and heterogeneous resources in order to deliver computational power to resource demanding applications in a transparent way [2], [3]. Built on pervasive internet standards, Grids allow organizations to share computing and information resources across department and organizational boundaries in a secure and highly efficient manner. Grids support the sharing, interconnection and use of diverse resources, integrated in the framework of a dynamic computing system.

What we discuss in this paper is a clinical trials simulation in Grid environments. *In silico* ("on the computer") oncology is an emerging interdisciplinary field aiming at mathematically describing and computationally simulating the multi-scale biological mechanisms that constitute the phenomenon of cancer and its response to therapeutic techniques. Within this framework, the "In Silico Oncology Group (ISOG)", National Technical University of Athens, has already developed a four-dimensional simulation model of glioblastoma multiforme (GBM) [4], [5] response to radiotherapy. ISOG has adopted an essentially "top-down" modeling approach and developed a number of hybrid discrete Monte Carlo / cellular automata and continuous differential equation simulation models of tumour growth and response to therapeutic modalities. The aim is a better understanding of biological mechanisms concerning cancer and related therapeutic interventions and, in the long term, a contribution to the design of patient individualized therapies.

For the purposes of the specific work, the ISOG simulation model of glioblastoma multiforme response to radiotherapy has been used in order to perform a virtual clinical trial. The model is based on the clinical, imaging, histopathologic, and molecular data of the patient and numerous fundamental biological mechanisms are incorporated and explicitly described. A prototype system of quantizing cell clusters included within each geometrical cell of a discretizing mesh covering the anatomic area of interest lies at the heart of the proposed simulation approach.

Parameter Sweep Applications (PSAs) are a class of applications that deal with the analysis of a specific simulation for a range of parameter values. Typically, this kind of applications is comprised of a large number of tasks, each of which performs a given simulation over a subset of parameter values. In the context of in silico oncology and radiotherapy simulation, a PSA is comprised of a set of independent tasks, which share some common files (executable code for a specific method of therapy and input files). Due to the fact that tasks are completely independent, the application is viable for efficient large scale execution on a computational grid, thus reducing dramatically the overall time demanded for its completion.

As already described, the simulation environment for performing radiotherapy parameter sweep simulations mainly consists of a real Grid infrastructure provided by the EGEE project [12]. The simulation environment builds on the gLite middleware [13] and provides a web-based Grid portal for enabling interactions with it in a simple and user-friendly way. In order to enable the execution of radiotherapy simulations on the Grid, the legacy code has been suitably migrated to the operating system used on Grid nodes and several scripts have been developed in order to

Manuscript received August 15, 2008.

Dimosthenis Kyriazis is with the Department of Electrical and Computer Engineering, National Technical University of Athens, 9, Heroon Polytechniou Str, 15773 Athens, GREECE (phone: +302107722546; e-mail: dkyr@telecom.ntua.gr).

Andreas Menychtas is with the Department of Electrical and Computer Engineering, National Technical University of Athens (e-mail: a_menychtas@telecom.ntua.gr).

Dimitra Dionysiou is with the Department of Electrical and Computer Engineering, National Technical University of Athens (e-mail: dimdio@central.ntua.gr).

Georgios Stamatakos is with the Department of Electrical and Computer Engineering, National Technical University of Athens (e-mail: gestam@central.ntua.gr).

Theodora Varvarigou is with the Department of Electrical and Computer Engineering, National Technical University of Athens (e-mail: dora@telecom.ntua.gr).

automatically conduct parameter sweep radiotherapy simulations. The application – portal is enhanced with added functionality in order to simplify the job submission process and automate the interaction with the Grid services. In that way more users are able to access the computational resources while the administrators manage the application and monitor the operational status.

The remainder of the paper is as follows: Section 2 presents related work in the field of *in silico* oncology, while Section 3 concisely presents the radiotherapy simulation model that has been originally implemented. The virtual clinical trial is included thereafter in Section 4. The way Grid technologies can be applied to the specific field of science and the aforementioned simulations are discussed in Section 5. In the last section (Section 6) we present the results obtained for specific parameter sweep simulations. Finally, Section 7 concludes with a discussion on future research and potentials for the current study.

II. Related Work

In general, enabling applications execution on Grid environment has been a research topic since the distributed nature of a Grid-based infrastructure makes feasible the solution of computational intensive problems in a reliable and cost-effective way. To this direction, literatures [14], [15] and [16] present the work performed for various application domains (biocomputational, learning and medical), while in [17] the authors describe an approach for deploying, testing and analyzing multiphysics codes in Grid-based computing environments.

In many cases, computational Grids serve the needs of PSAs, which typically consist of a large set of independent tasks that run the same code over different input parameter values. High performance parametric modeling has been identified as a killer application for the Grid [22]. Grid-Enabled PSAs have been recently developed in a number of scientific and engineering fields, like Bioinformatics [23], Particle Physics [24], Computational Fluid Dynamics (CFD) [21], etc. The main challenge of porting PSAs to a Grid environment is to provide efficient execution and scheduling mechanisms that have the ability to adapt to the dynamic and heterogeneous nature of Grids, as also discussed in [25] and [26].

Taking into consideration that the end-users are not familiar with Grid infrastructures, the final success of any attempt to apply Grid technologies to clinical trial simulations will primarily depend on its real adopters; the end users whose main demand refers to usability and user friendliness. Therefore, we consider the interface of major importance for the application and the web portal approach as a user interface has been adopted as the best option to fulfill the aforementioned users' requirements. In the domain of Grid technologies, research on the web portals that offer Grid services is a very challenging topic. For the EGEE Grid infrastructure there are currently two portal implementations under development, the GENIUS Grid Portal [27] and the P-GRADE Portal [28]. Furthermore, BEinGRID Project [18], [19] dedicates specific effort to web portals as a means to access Grid infrastructures. These portals are available to the research community and include many advantages such as workflow construction and execution.

III. Simulation Model

As also described in detail in [4], [5], [6] the simulation model is basically based on a procedure, during which the geometrical mesh covering the tumour region is scanned every T units of time. The elementary cubic volume of the mesh is called "Geometrical Cell (GC)". In each time step, the updated state of a given GC is determined on the basis of a number of algorithms describing the behavior of the cells constituting the tumour. More specifically, each GC of the mesh belonging to the tumor contains biological cells, which are distributed in a number of "classes" (compartments), each one characterized by the phase in which its cells are found (within or out of the cell cycle: G1, S, G2, M, G0, Necrosis, Apoptosis). A cycling tumour cell passes through the subsequent phases G1, S, G2, and M. Cell death due to necrosis or apoptosis is incorporated into the model. Cell killing by irradiation is described by the Linear Quadratic or LQ Model:

$$S(D) = \exp[-(\alpha D + \beta D^2)] \qquad (1)$$

where S(D) is the surviving fraction after a (uniform) dose D (Gy) of radiation to a population of cells. The parameters α (Gy-1) and β (Gy-2) are called the radiosensitivity parameters of the LQ model.

If the number of tumour cells contained within a given GC drops below a given threshold, then a procedure which attempts to "unload" the remaining biological cells in the neighboring GCs takes place. Cells are preferentially placed within the neighboring GCs with the maximum available free space. If the given GC becomes empty, it is assumed to disappear from the tumour. An appropriate shift of a chain of GCs, intended to fill the "vacuum", leads to tumor shrinkage. This can happen after the killing of a number of cells by irradiation. On the other hand, if the number of alive and dead cells within a given GC exceeds a given threshold, then a similar procedure attempting to unload the excess cells in the surrounding GCs takes place. In case that the unloading procedure fails to sufficiently reduce the number of cells, then a new GC emerges. Its position relative to the "mother" GC is determined using a random number generator. An appropriate shifting of a chain of adjacent GCs leads to a differential expansion of the tumor.

IV. The Virtual Clinical Trial

The "*Radiation Therapy Oncology Group (RTOG) 83-02*" clinical study [7] has been selected to serve as a realistic paradigm of clinical trials in which modeling could play a useful role; a series of simulations corresponding to its various arms have been performed. This was a randomized Phase I/II study of escalating doses for Hyperfractionated radiotherapy (HF, 1.2Gy twice daily to doses of 64.8, 72, 76.8, or 81.6Gy) and Accelerated Hyperfractionated radiotherapy (AHF, 1.6Gy twice daily to doses of 48 or 54.4Gy) with carmustine (BCNU) for adults with supratentorial GBM or anaplastic astrocytoma. The study has revealed that GBM patients who received the higher HF doses had survival superior to the patients in the AHF arms or lower HF doses.

In order to create the *in silico* counterpart of the real study cohort, a thorough literature review concerning critical parameters influencing GBM growth and response to radiotherapy has been performed. The involved parameters where the cell cycle duration (TC) and the α and β radiosensitivity parameters of the linear quadratic (LQ) model. For the cell cycle duration a parameter range has been identified (20h-120h) [8], [9], [10] whereas for the LQ model pair (α,β) a number of experimentally determined values have been considered reflecting specific molecular profiles [11]. The parameter values considered seem to cover the whole range of values of TC and (α, β) that have been reported in literature for GBM and therefore are assumed to reflect GBM interpatient variability. For practical reasons the continuous range of TC values has been divided into a number of equal intervals of 10h. All combinations of TC and (α,β) pair values have been considered. All other parameters of the simulation model (e.g. cell density, cell loss factor, oxygen enhancement ratio etc.) have been selected so as to reflect a typical value for GBM tumors as dictated in the relevant literature. More specifically, the study comprised a total of 462 radiotherapy simulations based on the 6 radiotherapeutic schemes of the RTOG 83-02 clinical study. 11 values of the cell cycle duration and 7 pair values of the radiosensitivity α and β parameters of the LQ model have been considered.

V. Applying Grid Technologies

Exploitation of Grid technologies is imperative for in silico oncology for the following reasons:

-- exponential increase of required computational resources when considering a more dense discretization of the space-time (4D) Grid of the biological problem

-- heterogeneity of required data (imaging, histopathologic, genetic) with different preprocessing requirements

-- large number of involved patients.

The vast resources provided by a Grid infrastructure may lead to a better understanding of the biological and clinical behavior of cancer and especially solid tumours. Furthermore, computer simulation may be employed in order to optimize treatment of cancer, by conducting a number of simulations for different therapeutic schemes based on the individual data of a patient. A restraining factor is that simulations need to be conducted in clinically accepted computational time. As the number of possible therapeutic schemes and consequently the number of simulations increases, the time required for evaluating and comparing the effects of the different schemes may become forbiddingly high. Exploiting Grid computing is a very attractive solution, as the resources provided in a Grid infrastructure may be efficiently used to reduce overall required execution time in a cost-effective and efficient manner.

In order for *in silico* oncology to be efficiently transferred to a Grid infrastructure, certain aspects need to be addressed regarding its adaptation to the Grid programming model:

-- mechanisms for automatic simulation submission and monitoring

-- data management and result aggregation

-- development of efficient Grid workflows, taking into consideration the characteristics of the different simulation models

-- provide some basic QoS to the user (e.g. provide an optimal response time for various simulation models – always taking into consideration the characteristics of the model)

As the number of parameter value combinations needed to simulate a clinical trial is rather high, execution of the aforementioned code in Grids constitutes as an attractive solution for keeping the simulation time within practical limits. Therefore, a toolkit enabling the execution of the simulations on Grid infrastructures has been designed and developed.

The Grid-enabled environment has been developed for use on the EGEE Grid infrastructure [13] and thus builds on the gLite middleware [14]. In order to enable the execution of radiotherapy simulations on the Grid, the legacy code has been suitably migrated to the operating system used on Grid nodes and several scripts have been developed in order to automatically conduct parameter sweep radiotherapy simulations (parameter sweep applications – PSAs: a class of applications that deal with the analysis of a specific simulation for a range of parameter values). In the context of in silico oncology and radiotherapy simulation, a PSA is comprised of a set of independent tasks, which share some common files (executable code for a specific method of therapy and input files). Due to the fact that tasks are completely independent, the application is viable for efficient large scale execution on a computational Grid, thus reducing dramatically the overall time demanded for its completion.

The simulation environment provides a web-based Grid portal for enabling interactions in a simple and user-friendly way. The application-specific portal was based on a multi-tier architectural approach, defining different layers for the

operations and functions of the application framework, and thus simplifying the installation and the maintenance processes. The simulation application consists of four layers: the presentation layer (which includes all the functionality for the interaction with the end users and is presented to the users through web pages), the portal services layer (which includes all the functionality of the application and establishes the connection between the presentation and the gLite and database layers), the gLite layer (which includes all the functionality for the communication with the Grid services and resources) and the database layer (which includes a database keeping all the data regarding the application and user management).

The architecture of the simulation environment, along with implementation details are presented in the following figure (Figure 1).

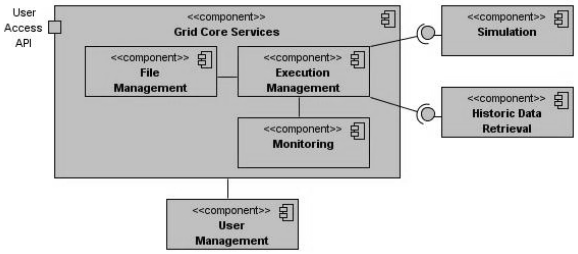

Figure 1. Architectural Component Model.

The basic services of the Grid middleware used refer to *File Management* (for transferring the input files including the simulation data to the Grid resources and obtaining the results), *Execution Management* (for submitting a "job" for execution to the infrastructure) and *Monitoring* (for monitoring the "jobs" status). We have also implemented and deployed three (3) services that reside on the Grid Service Provider but are not considered to be part of the Grid middleware. These services refer to *User Management* (for managing different classes of users: end-users, administrators, application workflow providers, etc), Simulation (which is the actual service that implements the simulation model and performs the simulations) and *Historic Data Retrieval & Analyzer*. The latter is used to compare the data obtained by the simulation service with real data in order to evaluate the simulation method for patients with the same "medical information / record". Based on this comparison, an extra value is produced, which is also presented to the end-user, as a value that shows the fidelity of the simulation result.

VI. Simulation Results

Indicative results drawn from the whole series of simulations are presented in the following figures (Figure 2, Figure 3). In Figure 2 the clear advantage of HF schedules and in particular high-dose HF schedules is evident for a hypothetical case of a moderately radiosensitive GBM tumour with TC=40h and $(\alpha,\beta) = (0.31Gy^{-1}, 0.04Gy^{-2})$. In Figure 3 the comparative tumor cell kill effect of the HF 72Gy radiotherapy scheme for hypothetical clinical cases differing in their cellular radiosensitivity is depicted. The increased cell survival in the more radioresistant tumours is evident.

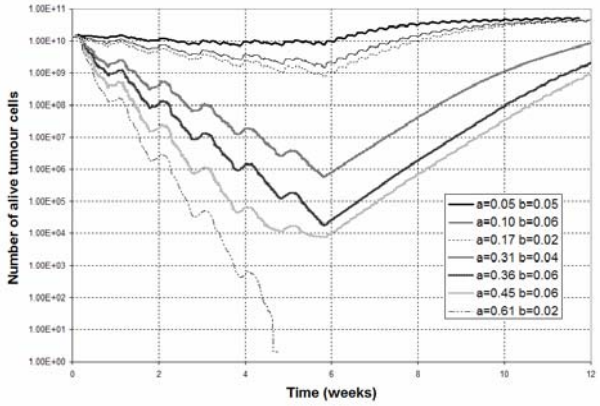

Figure 3. The number of alive tumour cells as a function of time from the start of the radiotherapy treatment for hypothetical cases of a GBM tumours with the same cell cycle duration, TC=40h, but differing in their cellular radiosensitivity and treated according to the HF 72Gy schedule. α is in Gy^{-1}, β is in Gy^{-2}.

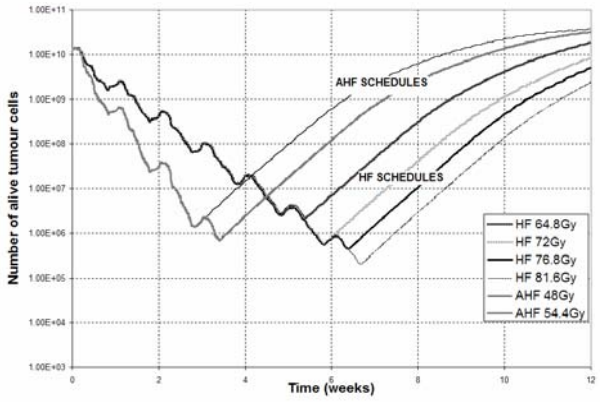

Figure 2. The number of alive tumour cells as a function of time from the start of the radiotherapy treatment for a hypothetical case of a GBM tumour with TC=40h and (α,β) = (0.31Gy-1, 0.04Gy-2). HF: HyperFractionation, AHF: Accelerated Hyperfractionation. For each scheme the total dose is given.

The simulation results are in agreement with the results of the RTOG 83-02 clinical study, as they reveal that for GBM radiotherapy the use of high-dose hyperfractionation schemes is advantageous in terms of tumour cell kill compared to low-dose hyperfractionation schemes or accelerated hyperfractionation schemes. This holds true for the whole range of parameter values that have been tested and is therefore interpreted as being valid for GBM tumours in general, under the assumption mentioned before that the values of the parameters used cover the whole spectrum of GBM tumours.

In the table that follows (Table I), we present execution results in terms of the time needed for the simulations. The

results indicate that by performing the parameter sweep simulations on the Grid, a considerable speedup may be achieved. This is very important in the context of *in silico* oncology, since the computational requirements of the simulations become overwhelmingly large as the required detail of simulation grows and because of the large number of potentially involved patients. Grid computing is a very appealing solution in the context of *in silico* oncology, since the vast resources provided by a Grid may be efficiently used for providing timely and accurate results.

TABLE I
EXECUTION TIMES OF RADIOTHERAPY PARAMETER SWEEP SIMULATIONS

Scheme #	Mean Job Execution Time	Overall Schema Execution Time
1	~32 minutes	~58 minutes
2	~31 minutes	~59 minutes
3	~34 minutes	~59 minutes
4	~38 minutes	~72 minutes
5	~36 minutes	~66 minutes
6	~21 minutes	~47 minutes

In comparison with other sequential computing schemes, the main difference and added value of using Grid infrastructures for clinical trial simulations is that these simulations can be conducted in parallel in different Grid nodes. For example the time needed for a simulation is ~30 minutes and for 100 simulations in a sequential computing scheme, the time would be 3000 minutes. However, in Grid environments with more than 100 nodes, these simulations could be conducted in parallel and therefore the overall time would be ~30 minutes.

VII. CONCLUSION

In silico oncology is a multidisciplinary field that aims to model the multi-scale biological mechanisms that constitute the phenomenon of cancer and evaluate its response to therapeutic techniques by computer simulations. Due to the exponential increase in the complexity of the simulation as the density of discretization of the 4D Grid of the biological model increases and the heterogeneity of required data and their preprocessing needs, as well as the large number of potentially involved patients, the large scale execution capabilities and vast computational capacities offered by a Grid may prove exceptionally beneficial.

In the work presented in this paper, a Grid-enabled toolkit for in silico oncology Simulations has been developed. The toolkit builds on the gLite middleware and enables the execution of radiotherapy simulations on the Grid infrastructure deployed by the EGEE project. Several mechanisms for automatically creating parameter sweep simulations have been implemented. The toolkit provides a web-based portal that acts as a user-friendly way for the non-Grid expert doctor or researcher to access the resources of the Grid.

Several parameter sweep simulations have been conducted using the toolkit. Exploitation of Grid resources has made possible the simulation and comparison of different therapeutic schemas that would be extremely time consuming in case of execution on a conventional computer. Obtained results, which are in accordance to the clinical studies, have provided the required data for a thorough comparison of the simulated therapeutic schemas. Execution times prove that a considerable speedup may be achieved by applying Grid technologies. These can also provide solutions in case that comparative results for therapeutic schemas are needed in real time (as described with the use of Historical Data).

It is expected that during the next few years, such simulation models and mechanisms supported by new generation algorithms are likely to be used. The computer infrastructure needed for these simulations will be significant and to this direction, Grid infrastructures can serve as a means to support this endeavor.

REFERENCES

[1] I. Foster and C. Kesselman, "The Grid: Blueprint for a Future Computing Infrastructure", Morgan Kaufmann Publishers, USA, 1999.

[2] I. Foster, C. Kesselman, S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations" International Journal Supercomputer Applications, Vol. 15, No. 3, 2001.

[3] W. Leinberger, V. Kumar, "Information Power Grid: The new frontier in parallel computing?" IEEE Concur., Vol. 7, No. 4, pp. 75-84, October-December 1999.

[4] D . Dionysiou , G . Stamatakos , N . Uzunoglu , K . Nikita, "A computer simulation of in vivo tumour growth and response to radiotherapy: New algorithms and parametric results, Computers in Biology and Medicine", 2006.

[5] G. Stamatakos, D. Dionysiou, E. Zacharaki, N. Mouravliansky, K. Nikita, N. Uzunoglu, "In silico radiation oncology: combining novel simulation algorithms with current visualization techniques". IEEE Proceedings: Special Issue on Bioinformatics: Advances and Challenges, 2002.

[6] D. Dionysiou, G. Stamatakos, N. Uzunoglu, K. Nikita, A. Marioli, "A Four Dimensional In Vivo Model of Tumour Response to Radiotherapy: Parametric Validation Considering Radiosensitivity", Genetic Profile and Fractionation. J. theor. Biol. 230, 2004/

[7] Werner-Wasik, M., Scott, C. B., Nelson, D. F., Gaspar, L. E., Murray, K. J., Fishbach, J. A., Nelson, J. S., Weinstein, A. S. and Curran, W. J.: Final Report of a Phase I/II Trial of Hyperfractionated and Accelerated Hyperfractionated Radiation Therapy with Carmustine for Adults with Supratentorial Malignant Gliomas: RTOG 83-02, 1996.

[8] Hegedues, B., Czirok, A., Fazekas, I., Babel, T., Madarasz, E., Viscsek, T.: "Locomotion and proliferation of glioblastoma cells in vitro: statistical evaluation of videomicroscopic observations", J. Neurosurgery, 2000.

[9] T.Hoshino "Cell kinetics of glial tumors", Rev. Neurol. (Paris) 148, 396-401, 1992.

[10] L. E. Dillehay, "A model of cell killing by low-dose-rate radiation including repair of sublethal damage, G2 block, and cell division", Radiat. Res. 124, 1990.

[11] D.A.Haas-Kogan, G.Yount, M.Haas, D.Levi, S.S.Kogan, L.Hu, C.Vidair, S.F.Deen, W.C.Dewey, M.A.Israel, Int. J. Radiat. Oncol. Biol. Phys. 36, 95-103, 1996.

[12] The EGEE Project, http://www.eu-egee.org/

[13] The gLite middleware, http://glite.web.cern.ch/glite

[14] Katsaloulis, E. Floros, A. Provata, Y. Cotronis, T. Theoharis, "Gridification of the SHMap Biocomputational Algorithm",

International Special Topic Conference on 'Information Technology in Biomedicine' by IEEE and EMB, Ioannina, 2006.
[15] Ardaiz, L. Diaz de Cerio, A. Gallardo, R. Messeguer, K. Sanjeevan, "ULabGrid Framework for Computationally Intensive Remote and Collaborative Learning Laboratories", IEEE International Symposium on Cluster Computing and the Grid, 2004.
[16] Tristan Glatard, Johan Montagnat, Xavier Pennec, "Grid-enabled workflows for data intensive medical applications", Computer Based Medical Systems (CBMS'05), Special Track on Grids for Biomedicine and Bioinformatics, Dublin, Ireland, June 23-24, 2005.
[17] Toan Nguyen, Lizhe Wang, Vittorio Selmin, "Virtual Private Environments for Multiphysics Code Validation on Computing Grids", International Conference East West High-Speed Flow Field (EWHSFF'2005), Beijing, 2005.
[18] BEinGRID Project, www.beingrid.eu
[19] Gridipedia, http://www.gridipedia.eu/102.html
[20] Computational Fluid Dynamics Automatic Optimization using OpenFOAM in Grid Environments
[21] Gregory Katsaros, Francisco Campos, Domosthenis Kyriazis, Theodora Varvarigou, "Computational Fluid Dynamics Automatic Optimization using OpenFOAM in Grid Environments", OpenFOAM Conference, 2008.
[22] Abramson, D., Giddy, J., and Kotler, L., High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid?, IPDPS'2000, Mexico, IEEE CS Press, USA, 2000.
[23] Giovanni Aloisio, Massimo Cafaro, Sandro Fiore, Maria Mirto: ProGenGrid: A Workflow Service Infrastructure for Composing and Executing Bioinformatics Grid Services. CBMS 2005.
[24] B. Beeson, S. Melnikoff, S. Venugopal, and D. G. Barnes, A Portal for Grid-enabled Physics, Proceedings of the 2005 Australasian workshop on Grid computing and e-research, 2005.
[25] H. Casanova, G. Obertelli, F. Berman, and R. Wolski. The AppLeS Parameter Sweep Template: User-level middleware for the grid. In Proceedings of the Super Computing Conference, pages 75–76, 2000.
[26] E. Huedo, R. S. Montero, and I. M. Llorente. Experiences on adaptive grid scheduling of parameter sweep applications. In 12th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP'04), 2004.
[27] GENIUS Grid Portal: https://genius.ct.infn.it
[28] P-GRADE Portal: http://portal.p-grade.hu