# Abstracts (posters)

## 1. DORII (Deployment of Remote Instrumentation Infrastructure)

Norbert Meyer, Marcin Płóciennik (1), Dieter Kranzlmüller, Michael Schiffers (2), Stefano Salon (3), Rainer Keller (4), Anastasios Zafeiropoulos, Ioannis Labotis(5), Paolo Gamba (6), Milan Prica, Roberto Pugliese (7), Davide Adami, Franco Davoli (8), Jesus Marco de Lucas (9), Agustín Monteoliva (10), Ángel David Gutiérrez Barceló (11)

*(1) Poznan Supercomputing and Networking Center, Institute of Bioorganic Chemistry PAS, Poland*
*(2) Ludwig-Maximilians University Munich, Germany*
*(3) Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Italy*
*(4) Universität Stuttgart, Germany*
*(5) Greek Research and Technology Network S.A., Greece*
*(6) European Centre for Training and Research in Earthquake Engineering, Italy*
*(7) Sincrotrone Trieste SCpA, Italy*
*(8) Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Italy*
*(9) Consejo Superior de Investigaciones Cientificas, Spain*
*(10) Ecohydros SL, Spain*
*(11) Universidad de Cantabria, Spain*

The DORII project (RI-213110) is funded by European Union under the Seventh Framework Programme (FP7). It intends to deploy e-Infrastructure for new scientific communities specifically for experimental equipment and instrumentation that are today not or only partially integrated with the European e-Infrastructure. DORII is focusing on the following selected scientific areas:

- earthquake community (with various sensor networks): earthquake early warning system design and simulation, network-centric seismic simulations
- environmental science community: oceanographic and coastal observation and modeling, inland waters and reservoirs monitoring
- experimental science community: on-line data analysis in experimental science coming from increased efficiency of the light sources (synchrotron and free electron lasers) and of the detectors with higher and higher resolution and faster readouts.

The scientific communities targeted by the project are well recognised and organised, even in industry areas represented by SMEs. Working closely with end-users, solutions will be put in place that build upon the success of past and ongoing projects in such areas as remote instrumentation (GRIDCC, RINGrid), virtual laboratories (VLAB), interactivity (int.eu.grid), software frameworks for application developers (g-Eclipse) and advanced networking technologies (GN2) with EGEE-based middleware.

DORII recognises the following strategic goals and objectives:

- To adopt e-Infrastructure functionality across selected areas of science and engineering.
- To deploy and operate persistent, production quality, distributed instrumentation integrated with e-Infrastructure.
- To generalize and deploy a framework environment that can be used for fast prototyping.

The DORII project consists of three phases: The first is the integration and adaptation of the selected products from previous projects that have been successfully carried out by the projects like GRIDCC, Int.EU.Grid, gEclipse or VLAB. It will take advantage of best practices and operate the Remote Instrumentation Infrastructure. The second phase is a deployment of the project applications on the infrastructure and middleware enhancement. The third phase is standardisation and deployment of the results to a wider community outside the project. It is planned to increase the deployment phase on a bigger consortium called MOON (Mediterranean Ocean Observing Network), not being a partner of DORII, which is represented in DORII by OGS. The project will promote standardisation and knowledge transfer via e-IRG and OGF research groups. Project takes active role in a research group in the Open Grid Forum, which exactly focuses on topics presented in the project, i.e. RISGE - Remote Instrumentation Services in a Grid Environment.

Remote control of scientific facilities with usage of virtual laboratories and solution proposed by DORII has the potential of revolutionizing the mode of operation and the degree of exploitation of scientific instruments. Grid technologies will be used to integrate operations with computing farms where complex models and computing coming from instruments could run but also for storing large amount of data. Grid will handle issues related with authorization, resource management, data transfer and storing. Network infrastructure will be used with the available mechanism for QoS handling.

# 2. Support for Cooperative Experiments in VL-e: from Scientific Workflows to Knowledge Sharing

Zhiming Zhao (1), Victor Guevara (1), Adianto Wibisono (1), Adam Belloum (1), Marian Bubak (1,2) and Bob Hertzberger (1)

*(1) Informatics Institute, University of Amsterdam, The Netherlands*
*(2) Institute of Computer Science AGH, University of Science and Technology, Krakow, Poland*

Complex scientific experiments involve distributed scientific data and resources, and are often done by cooperating scientists from different domains. Cooperative experiments involve not only coordination between resources and computing processes, but also the sharing and transfer of knowledge among scientists. Support for cooperative experiments has become, as such, an important requirement for the e-Science middleware. Semantic technologies enhance the storage and query of Grid resources and the high level searching and matching between different resources. Workflow management systems automate experiment processes and integrate different resources. Usage of Web 2 and tools developed by the Computer Supported Cooperative Work (CSCW) [7] society enables cooperative support between scientists. These research lines compliment each other and push forward the support for the e-Science experiments. However, seamlessly integrating such tools in one coherent environment with high usability for scientists from different domain and different background knowledge is still a challenging issue [8].

This paper presents the Dutch Virtual Laboratory for e-Science project which aims at providing generic functionalities that support a wide class of specific e-Science application environments and set up an experimental infrastructure for the evaluation of the ideas. A set of tools are developed: for modeling and managing workflow templates, for browsing resources stored in the VL-e environment, a workflow tool for composing workflows and for managing components, and a framework for coupling VL-e workflows with other legacy workflows. **Templates and components manage tools** [1] include Olingo, a mapping tool between the workflow template description and underlying data presentation, CLAMP, an annotation for VL-e components, and Hammer, a storing and query tools for components. **Virtual resource browser: VBrowser** [2] offers scientists an environment in which they interactively access resources of various types to manipulate data (upload, download, search, annotate, and view), start applications (prepare and execute experiments) and monitor resources (status, control, notification).

**Interactive parameter sweep: FRIPS** [3] aims to support interactive execution of applications that require parameter sweep. It allows scientists to monitor the experiment execution, view intermediate results, and give interactive feedback on a running experiment. **The VL-e workflow system: WS-VLAM system** [4] has a set of client-side applications that allow scientists to design and monitor the execution of the workflows with intuitive interfaces, and provides also server-side applications, including a `workflow engine' that schedules and executes the workflow on the Grid. **Workflow aggregation: VLE-WFBus** [5] provides tools to recognize different workflow descriptions stored in the system and interface to wrap and integrate legacy scientific workflows. Via the tool, a user can execute a workflow via the workflow bus, and integrate it with the other workflows via different connectors.

Compared to the WEeb 2 based cooperative environments, such as myExperiment [6], VL-e tools have clear focus on the runtime issues of the workflow. Currently, there are close discussion going between VL-e and the MyExperiment society, such as proposing WSVLAM workflows as new workflow types which are shared between scientists, and making workflow bus as generic execution interface for different workflows shared over MyExperiment environment.

**Acknowledgements**. This work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl).

References

1. Víctor Guevara-Masís, Konstantinos Krommydas, Adam Belloum, Louis O. Hertzberger, Semantic Workflow Discovery in VL-e, Proceedings of KnowledgeGrid 2006, Workshop, IST 2006 Strategies for Leadership, Helsinki, Finland
2. T. Glatard, K. Boulebiar, P. de Boer, S. Olabarriaga, fMRI analysis on EGEE with the Vbrowser and MOTEUR, 3rd EGEE User Forum, Feb 2008
3. Adianto Wibisono, Zhiming Zhao, Adam Belloum, Marian Bubak: A Framework for Interactive Parameter Sweep Applications. CCGRID 2008: 703
4. A. Wibisono, V. Korkhov, D. Vasunin, V. Guevara-Masis, A. Belloum, C. de Laat, P. Adriaans and L.O. Hertzberger, WS-VLAM: Towards a scalable workflow system on the Grid, Proceeding of the 16th IEEE International Symposium on High Performance Distributed Computing, June 27-29, 2007, Monterey Bay, California, USA
5. Z. Zhao, S. Booms, A. Belloum, C. de Laat and L.O. Hertzberger, VLE-WFBus: A Scientific Workflow Bus for Multi e-Science Domains, 2nd IEEE International Conference on e-Science and Grid Computing, Amsterdam, 2006, Proceedings, p. 11

6. David De Roure and Carole Goble and Robert Stevens , The design and realization of myExperiment Virtual Research Environment for social sharing of workflows, Future Generation Computer Systems, 10 July 2008,
7. Wanda Pratt, Madhu C. Reddy, David W. McDonald, Peter Tarczy-Hornoch, John H. Gennari, Incorporating ideas from computer-supported cooperative work, *Journal of Biomedical Informatics*, *Volume 37, Issue 2*, *April 2004*, *Pages 128-137*
8. Yolanda Gil, Ewa Deelman, Mark Ellisman, Thomas Fahringer, Geoffrey Fox, Dennis Gannon, Carole Goble and Miron Livny,Luc Moreau, Jim Myers, Examining the Challenges of Scientific Workflows, Computer, 12, 2007, 24-32

# 3. Supporting Collaboration by Large Scale Email Analysis

Michal Laclavik, Martin Seleng, Ladislav Hluchy

*Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia*

Mailing lists and email in general is heavily used for collaboration. Collaboration and community mailing lists, contains also lot of valuable information which can be used to support better collaboration and information management. It also contain of peoples' social network and its relations to particular topics, project or problems. We believe information extraction and semantic analysis of email communication can improve information search, collaboration and community or people networking. Personal or organizational email archives are becoming quite large datasets and thus large scale processing is needed.

We have preformed several email archive analysis experiment on Google's MapReduce [1] distributed architecture and its Hadoop [2] implementation. Email analysis extracted social networks as directed and valued graphs, which can be used e.g. email search [3] or recommendation and information management systems such as Xobni [4]

Using semantic annotation and semantic analysis we were also able to transform graph nodes represented by email addresses into different graphs on same data representing people (grouping multiple email addresses), groups, organizations or countries. Information extraction and semantic analysis performed are based on pattern based semantic annotation method Ontea [5].

Results of information extraction and semantic annotation were following:

- Social network of communicating people
- Including level of interaction (how many email's were send and received)
- Social network graph was transformed to different graphs based on email addresses, people or organizations
- Graph can be enriched with other objects mentioned in communications depending on ontology model: e.g. organizations, enterprises, people or geographical locations.

In our previous work [5] we tested semantic annotation and information extraction on Hadoop cluster and we provided overview of Hadoop suitability for such tasks.

Social Network extraction experiments were executed on 5 node MapReduce cluster. We have used Intel(R) Core(TM)2 CPU 2.40GHz with 4GB RAM hardware on all machines. On Hadoop we have analyzed email archives of size 1.8 GB.

| Cores | Data Size | Extracted Metadata | Execution Time |
|---|---|---|---|
| 10 | 1.8 GB | 5625 | 11 min 40 sec |

Email archives processing on Hadoop architecture can scale to Terabytes of data which was proved also by Yahoo!, which use Hadoop in production environment [6] on web data.

Advantage of Map Reduce architecture for social network analysis is also natural way of integrating number of communication among people by Reduce method.

In our work we have shown how social network and related metadata can be extracted from large scale email archives. This metadata can be transformed and integrated with semantic model of application and used for improving information management, search or collaboration.

**Acknowledgements**. This work is supported by projects Commius FP7-213876, SEMCO-WS APVV-0391-06, AIIA APVV-0216-07, VEGA 2/7098/27

References

1. Dean J., Ghemawat S.: MapReduce: Simplified Data Processing on Large Clusters, Google, Inc. OSDI'04, San Francisco, CA (2004)
2. Lucene-hadoop Wiki, HadoopMapReduce, http://wiki.apache.org/lucene-hadoop/HadoopMapReduce (2008)

3. Einat Minkov, Ramnath Balasubramanyan, William W. Cohen: Activity-centric Search in Email; in CEAS 2008 also in Enhanced Messaging Workshop, AAAI 2008, http://www.cs.cmu.edu/~einat/activitySearch.pdf
4. MIT Technology Review: A New Look for Outlook – Xobni makes it easier to find relevant information buried in your inbox; http://www.technologyreview.com/Biztech/19463/?a=f, 2007
5. Michal Laclavik, Martin Seleng, Ladislav Hluchy: Towards Large Scale Semantic Annotation Built on MapReduce Architecture; In Proceedings of ICCS 2008; M. Bubak et al. (Eds.): ICCS 2008, Part III, LNCS 5103, pp. 331-338, 2008
6. Yahoo! Launches World's Largest Hadoop Production Application, Yahoo! Developer Network, http://developer.yahoo.com/blogs/hadoop/2008/02/yahoo-worlds-largest-production-hadoop.html, (2008)

# 4. EUFORIA (EU Fusion fOR Iter Applications)

Francisco Castejon (1), Antonio Gomez-Iglesias (2), Bernard Guillerminet (3), David P. Coster (4), Eric Sonnendruecker (5), Isabel Campos Plasencia (6), Jan Astrom (7), Jan Westerholm (8), Jose Maria Cela (9), Leon Kos (10), Lorna Smith (11), Marcin Plociennik (12), Marcus Hardt (13), Mats Aspnäs (8), Par Strand (14), Rainer Stotzka (13)

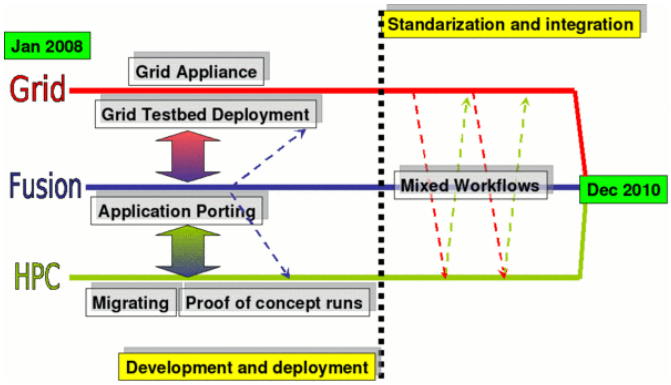
*(1) CIEMAT, Madrid, Spain*
*(2) Extremadura Advanced Research Center (CETA-CIEMAT), Trujillo, Spain*
*(3) Commissariat Energie atomique (CEA), France*
*(4) Max Planck Institute for Plasma Physics (IPP), Germany*
*(5) Universite Louis Pasteur, Strasbourg, France*
*(6) Instituto de Fisica de Cantabria (IFCA), CSIC, Santander, Spain*
*(7) Finnish IT Center for Science (CSC), Finland*
*(8) Abo Akademi University (ABO), Finland*
*(9) Barcelona Supercomputing Center (BSC), Barcelona, Spain*
*(10) University of Ljubljana (LECAD), SI-1000, Ljubljana, Slovenia*
*(11) University of Edinburgh (EPCC), Edinburgh, United Kingdom*
*(12) Poznan Supercomputing and Networking Center (PSNC), Poznan, Poland*
*(13) Forschungszentrum Karlsruhe (FZK), Karlsruhe, Germany*
*(14) Chalmers University of Technology (CHALMERS), Goteborg, Sweden*

EUFORIA is a project funded by European Union under the Seventh Framework Programme (FP7) which will provide a comprehensive framework and infrastructure for core and edge transport and turbulence simulation, linking grid and High Performance Computing (HPC), to the fusion modelling community.

The EUFORIA project will enhance the modelling capabilities for ITER sized plasmas through the adaptation, optimization and integration of a set of critical applications for edge and core transport modelling targeting different computing paradigms as needed (serial and parallel grid computing and HPC). Deployment of both a grid service and High Performance Computing service is essential to the project. A novel aspect is the dynamic coupling and integration of codes and applications running on a set of heterogeneous platforms into a single coupled framework through a workflow engine, a mechanism needed to provide the necessary level integration in the physics applications. This strongly enhances the integrated modelling capabilities of fusion plasmas and will at the same time provide a new computing infrastructure and tools to the fusion community in general.

The EUFORIA Project consists of two different phases that are partly being developed in parallel from the start of the project to become fully integrated in the later stages.

The first phase is currently ongoing. It handles the development and deployment. It consists of the adaptation and optimization of a selection of codes that cover edge and core physics for grid and HPC environments as appropriate. Inherent within this activity is the deployment of the computational infrastructure. This stage is mainly directed towards code developers and application developers that focus on the detailed implementation and code structures of the physics codes.




The current results are that the first selection of codes has been successfully ported to HPC and to GRID environments. Scalability tests prove the feasibility of the approach taken.

Furthermore, the architecture group has released the architecture-document for the integrated access to HPC and GRID resources. This architecture comprises access via the workflow GUI tool “Kepler” to both Infrastructures (gLite and Unicore) in a seamless way.

Before the second phase (that will run until Dec 2010) will handle the standardisation and integration of our achievements, we will focus on implementation and deployment of the integrated access to HPC and GRID resources.

**Acknowledgements**. This work is supported by EU project EUFORIA INFRA-2007-1.2.2-211804.

# 5. Distributed Computer System for Remote Support of Holistic Rehabilitation of Patients Affected by Stroke

Jacek Kitowski (1), Rafał Wcisło (1), Renata Słota (1), Janusz Otfinowski (2), Maciej Skubis (2), Karolina Probosz (2), Małgorzata Pisula (2), Artur Sobczyk (2), Krzysztof Reguła (2)

*(1) Department of Computer Science, AGH University of Science and Technology, Kraków, Poland*
*(2) Collegium Medicum, Jagiellonian University, Kraków, Poland*

The main goal of the presented project is development of a distributed computer environment serving as a support for a holistic rehabilitation of patients affected by a stroke. One of main parts of this lasting 30 months undertaking is creation and implementation of methods allowing for remote (via Internet) continuation of a rehabilitation process initiated in a medical institution.

The damage of a brain tissue caused by a stroke has organic character and is usually connected with disability of basic brain functions controlling important biological processes. But negative consequences of a stroke are not only limited to the somatic domain – they also affect the intellectual domain of a patient, especially the cognition (memory, concentration, motor coordination, spatial awareness, etc.). This leads to the accumulation of symptoms causing the expansion of therapeutic needs.

The computer seems to be a perfect tool that can be efficiently adopted for improving and exercising of the cognition, reducing motor disabilities [1-4] and, for selected classes of deficiencies, aphasia. Due to ability of creating virtual reality, situations stimulating the cognition can be easily created on a computer display. On the other hand, other peripheral computer devices (joystick, microphone, video camera, VR gloves) can be adopted to improve motor and speech abilities. The computer aided rehabilitation is meant as a supplement to traditional techniques and will enrich the holistic treatment of patients.

In selected by physicians cases the process of rehabilitation can be continued in patient’s home using the internet connection. The designed computer system provides:

- remote setup by leading physician of exercises (selection from a set of hundreds of different exercises, tuning of parameters, assignment of number of repetitions, etc.)
- home rehabilitation in both off-line and on-line modes – in on-line variant the audio and video communication is possible as well as exercise tracking and dynamic parameter tuning
- tracking of patients results (by updating dedicated database located in rehabilitation centre).

For obvious reasons implementation of methods and SOA interactions that will provide security and assure confidentiality is one of major tasks of the presented project.

As a result we expect growth of commonness [5] of the proposed rehabilitation methodology, increase of duration of rehabilitation (very important in this kind of treatment) and reduction of costs of hospitalization.

**Acknowledgements**. This research is financed by Polish Ministry of Education and Science, Project No. N N519 315435.

References

1. Sveistrup H., McComas J., Thornton M., Marshall S., Finestone H., McCormick A., Babulic K., Mayhew A. Experimental studies of virtual reality-delivered compared to conventional exercise programs for rehabilitation. Cyberpsychol Behav. 2003 Jun; 6(3): 245-9.
2. Merians A.S., Jack D., Boian R., Tremaine M., Burdea G.C., Adamovich S.V., Recce M., Poizner H. Virtual reality-augmented rehabilitation for patients following stroke. Phys Ther. 2002 Sep; 82(9): 898-915.
3. Jack D., Boian R., Merians A.S., Tremaine M., Burdea G.C., Adamovich S.V., Recce M., Poizner H. Virtual reality-enhanced stroke rehabilitation. IEEE Trans Neural Syst Rehabil Eng. 2001 Sep; 9(3): 308-18.
4. Broeren J., Bjorkdahl A., Pascher R., Rydmark M. Virtual reality and haptics as an assessment device in the postacute phase after stroke Cyberpsychol Behav. 2002 Jun; 5(3): 207-11.
5. i2010 – A European Information Society for growth and employment. http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm

# 6. Data-aware Composition of Workflows of Web and Grid Services

Ondrej Habala (1), Marek Paralič (2), Viera Rozinajová (3), Peter Bartalos (3)

*(1) Institute of Informatics of the Slovak Academy of Sciences, Bratislava, Slovakia*
*(2) Faculty of Electrical Engineering and Informatics of the Technical University of Košice, Slovakia*
*(3) Faculty of Informatics and Information Technologies of the Slovak University of Technology, Bratislava, Slovakia*

Many previous efforts [1-3] have been dealing with the problem of automatic composition of a workflow of computing tasks. This type of automation is very attractive especially in software engineering applied to scientific research, where complicated simulations and parameter studies often require tens of single steps in order to obtain the solution desired by the scientist. Since the inception of grid computing, workflow composition of grid jobs into complex workflows has also gained prominence with its apparent usefulness, long history of previous works not applied specifically to grid, and robust mathematical theory based mainly on direct acyclic graphs (DAGs). In recent years, advances in semantic web have been applied also in grid computing – creating semantic grid [4] – and specifically in the area of semantically-aided composition of workflows of grid tasks. However, most of the many works on this topic have concerned themselves only with the composition of a workflow of computer processes – represented by grid jobs, calls to web service interfaces, or other tasks – solving the "how", and have omitted the "what" of this problem, in this case "what" being the data, on which these processes operate. This has been left to the user. While the sought-after result is a system, in which the user enters the description of the data he/she requires, and the system composes a workflow able to compute it, most of the existing solutions create only a workflow able to solve a class of problems, and the selection of one unique member of this class via entering the correct data is left to the user.

We have designed, and begun to implement, a system, which tries to deal also with the "what" of automated workflow composition. The proposed system is based on previous work done in the context of the project K-Wf Grid [5], and extends it with tools which are able to determine exactly which data is necessary for which process in the composed workflow in order to get – at the end – the data which the user has described as his/her target. The system is based on semantic description of data and grid services by ontologies. The workflows are modeled as Petri nets, this being a legacy of K-Wf Grid offering very good means to model data (as Petri net tokens). The core of our system is an ontology describing semantics of the services from which the workflows are being composed, as well as of the available data and of the users which use the software, and of course the domain vocabulary, which is easily exchangeable in case of integration of the software with a different application. The main functional part of the software is a workflow enactment engine, able to use the ontology and stored knowledge in construction and execution of workflows of web services. The principal feature of the module is its ability to reuse also already existing data in an automated manner, not requiring the user to enter the data into an already constructed workflow. In such a distributed system user collaboration can be a real problem, and we have designed and described here also a collaboration tool, integrated with the ontology core of the system, which allows the users to exchange data and knowledge, and cooperate in the workflow construction and execution process. The system interacts with the user only to the extent absolutely necessary to acquire data or services which are required for the solution, but currently are not available in the grid.

**Acknowledgements**. This work is supported by the project SEMCO-WS APVV 0391 06, VEGA Nr. 1/3135/06 and INTAS 06-1000024-9154.

References

1. Bubak, M., Gubala, T., Kapalka, M., Malawski, M., Rycerz, K.: Grid Service Registry for Workflow Composition Framework. ICCS 2004, in LNCS vol. 3038, Springer, 2004. ISBN 978-3-540-22116-6, pp. 34-41.
2. VDS – The GriPhyN Virtual Data System. http://www.ci.uchicago.edu/wiki/bin/view/VDS/VDSWeb/WebMain (Accessed Sept. 2008)
3. Krishnan, S., Wagstrom, P., von Laszewski, G.: GSFL: A Workflow Framework for Grid Services. In Preprint ANL/MCS-P980-0802, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, 1L 60439, U.S.A., 2002.
4. Semantic Grid Community Portal. http://www.semanticgrid.org/ (Accessed Sept. 2008)
5. Knowledge-based Workflow System for Grid Applications (K-Wf Grid). EU 6th FP Project, 2004-2007. http://www.kwfgrid.eu (Accessed Sept. 2008).

# 7. WINGS: A Multigrid Wokflow Engine

Carlos de Alfonso, Miguel Caballer, Vicente Hernandez

*Grid and High Performance Computing Research Group, Instituto ITACA – Universidad Politécnica de Valencia*

Grid technology has extended, making computing infrastructures available for the scientists. The infrastructures deployed use different kind of middlewares. Although Globus 4 can be seen as the current de facto standard, many other middlewares are installed (GT2, gLite [1], Fura [2], etc.). So the scientists have to face two problems to create Grid applications: (1) to prepare the algorithms to work in a distributed environment like the Grid, and (2) to know the internal function of the different middlewares used to make the programs work.

In the last years several grid workflow systems have been developed in order to make the migration to the Grid easier. These kind of environments understand an experiment as a sequence of executions, enabling the users to think about the solution to the problem, trying to hide the internal complexity of the grid infrastructures and middlewares.

Recent workflow initiatives have been analyzed [3]: Some use a simple or low-level definition language (DAGMan, WFEE) or describe too much implementation details about the workflow (Kepler). Other are oriented to a specific type of grid deployment (Taverna, K-WfGrid) or enable a set of different kind of deployments, but do not permit extensions to new ones by the user (ASKALON, Triana).

This paper proposes an alternative which provides new features and focuses on high level definition, multigrid and extensibility capabilities. Initially a workflow definition language has been defined. The language, which is named WINGs (Workflow In New-generation GRIDs), is based on four concepts to model a workflow: data sources, operations, activities and executions.

**Data Sources** act so as sources as sinks for data. These Data Sources are used as points for interchange of data among the different executions in the workflow. The data sources enable to apply some filters to the result of the executions, in order to get only the necessary files. It can be filtered using filename wildcards the exit code of an execution, etc.

**Activities** are the abstractions of tasks to be run on the Grid. They describe the functionality of the tasks that will finally be executed by the run time system. They are conceived as an interface to the effective implementation of the task. An activity is defined by: The input and output parameters which define the interface of the activity, and the list of deployments which describe the functionality with a list of different implementations in the different grid systems in which the implementation is available. The number of middlewares to be used as for the deployment of an activity is extensible. It is only needed to develop a java class with the functionality to the selected middleware.

**Executions** are specific instances of an activity: They represent the task which would be actually launched and executed in the grid. The runtime engine is in charge of selecting, from the different available deployments for one activity, the best option in which to execute.

**Operations** are simple executions that will be executed by the workflow runtime, in order to pre-process the information which would be available in the Data Sources, prior to be used by the next tasks or perform some kind of post-process. These operations enable to add simple processing tasks without the need of creating new artificial grid executions such as: split and merge files, arithmetic operations, search operations in files, etc.

The WINGs language enables implicit flow control operations: The data source result filtering and the use of operations to filter the results, enable to create branches on the execution line. Also the system implicitly iterates through the data stored in the data sources, in a similar way to for/foreach instructions. However the language includes two flow control operations:

- If: it enables to use an execution or operation as the condition to select the branch to execute.
- Iterator: it is similar to “if” operator, but it uses an execution or operation as the condition to stop iterating. The iterator enable to establish a number of simultaneous task to launch, enabling (in necessary cases) to sequence all the tasks or execute in parallel.

References

1. EGEE, 2006. gLite Lightweight Middleware for Grid Computing. http://glite.web.cern.ch/glite/
2. GridSystems, 2006. Fura. http://fura.sourceforge.net/
3. Carlos de Alfonso, Miguel Caballer, Vicente Hernández: WINGS: Versatile Workflow for the Grid; Proceedings of ADVCOMP’08, Valencia, September 2008 (to appear).

## 8. Workflow Management with Agent-scheduling Support

*Viet D. Tran*

*Institute of Informatics, Slovak Academy of Sciences, Slovakia*

Over the last years, the development and acceptance of Grid technologies have been forwarded incrementally. Grid technologies connect distributed computational resources of dynamic multi-institutional virtual organization together and provide aggregate computational powers for solving very complex and computation demanding problems. The technologies make the infrastructures for researchers to share resources and knowledge, allows them to collaborate on solving common problems.

Since the infrastructure is becoming more and more powerful each year, the Grid applications also grow in size and complexity. The computation of the applications usually does not consist of a single task but many tasks connected together by data dependences. Workflow management became one of the main focuses of research and developments in Grid computing.

Many applications have workflow in such ways that some of the steps in the workflow are parametric-study tasks (fork-join scheme). The numbers of the tasks in these steps may be very large that managing workflow in the traditional ways (DAG style) are very efficient. Parametric-study tasks are much more elegantly managed by agent-scheduling tools that can provide also fault-tolerance and load balacing.

In this paper, the approach for combination workflow engines with agent-scheduling are introduced. Tasks in workflow will be executed by the worker jobs, that can improve performance and reliability of the workflow execution. That also provide support for complex workflows, where some of its tasks are parametric study.

**Acknowledgment**. This work is supported by VEGA project 2/6103/26.

## 9. Workflow-oriented Performance Monitoring of Grid Applications with the GridMind Monitoring System

Wlodzimierz Funika, Konrad Bula
*Institute of Computer Science AGH, Krakow, Poland*

Contemporary development of grid applications is on the rise due to their capabilities to solve major problems in a more complex range than that allowed by multiprocessor supercomputers or local computer clusters [1]. To optimize the performance of grid applications, tools which provide thevisualization of performance and forecast the behavior of distributed applications are applied. Data obtained from monitoring systems can be used to measure the performance of distributed applications and to control their execution, with analysis and visualization of an application execution, which detect application's weak performance areas and enable to define performance characteristics matched to the application context [2].

The functionality of the "GridMind" monitoring system improves the development of efficient and scalable grid applications by applying Data Mining models collected as the outcome of a designed data flow processing. The GridMind system comprises a tool application (Monitoring System), server application (Monitoring Service), and NET stateful web service (Process Service) implemented on WSRF.NET platform (Web Service Resource Framework). The tool application allows to compose a workflow in a manner similar to SCIRun [3], with an available/extendible set of building blocks like mathematical operations. The workflow is intended to process the monitoring data on the grid application and log them into a relevant database model scheme. The tool application is equipped with mechanisms that analyze the collected data to supprot predicting future values of resources performance. The server application is installed on a machine with a high computing capacity (cycle server) that processes an awaiting workflow and returns the results of performance monitoring to the tool application for further visualization.

The NET stateful service is an access point to the functionality offered by the grid. Its function is to launch processes of a grid application and to provide access to them as grid resources. The allocation of resources ensures to everyone who needs them or everyone demanding the resources to receive exactly what one needs. In addition, it prevents situations of leaving unused resources when commissions wait for realization. The database model scheme in "GridMind" system was implemented by means of Object Relation Mapping (ORM) that takeoffs relational database model on object database model.

The functionality of "GridMind" application aims at improving the task design mechanism for data flow, providing data visualization of monitoring performance of grid application processes, allowing for planning tasks (Scheduler) connected with the workflow, providing statistic reports about grid applications being monitored as well as presenting the inferrence results obtained in form of diagrams of knowledge models such as: clustering, Bayes's rules, association rules, and decision trees usually exploited in Data Mining.

The "GridMind" system, which uses Data Mining and the ability to create tasks for projected data flow in the tool application, allows for detailed valuation of performance, and hence the optimization of monitored distributed application.

References
1. M. Matyl, A. Nycz, Grid Computing web page: http://fatcat.ftj.agh.edu.pl/~bater/grid.pdf
2. Core Grid project home page: http://www.coregrid.net
3. SCIRun project page: http://software.sci.utah.edu/scirun.html

# 10. Harmonizing the Management of Virtual Organizations Despite Heterogeneous Grid Middleware – Assessment of Two Different Approaches

Wolfgang Kirchler (1), Michael Schiffers (2,3), Dieter Kranzlmüller (2,3,4)

*(1) Technische Universität München (TUM), Munich/Germany*
*(2) Munich Network Management Team*
*(3) Ludwig-Maximilians-Universität (LMU), Munich/Germany*
*(4) Leibniz Supercomputing Centre (LRZ), Garching/Germany*

## 1. The Problem

Coordinated problem solving and secure resource sharing in dynamic multi-institutional virtual organizations (a.k.a. “the grid problem”) is critically built on the concept of virtual organizations (VO). However, as grid systems continue to grow in scale, exhibit greater dynamics, and become more heterogeneous, managing grid-spanning VOs becomes an increasingly difficult challenge. This is not only caused by different VO-philosophies and different middleware technologies, but also by varying authentication and authorization schemes.

## 2. Two Possible Solutions

In this work we aim at harmonizing VO management despite heterogeneous middleware technologies (Globus Toolkit 4, both in its Web Service (WS) and its pre-WS flavour, the different versions of LCG/gLite, and UNICORE) and emerging federation approaches (Shibboleth). Based on a thorough analysis of the state-of-the-art we have developed two alternative solutions to overcome these difficulties within the German D-Grid Initiative. In the IVOM (Interoperability and Integration of VO Management Technologies) project [1] we proposed an integrated solution following an attribute-based authorization scheme with short-lived credentials. In [2] we investigated a different approach which is based on an additional abstraction layer serving as a proxy between VOs and the grid middleware. Both concepts, the integrated approach and the abstraction solution, have been implemented. We compared them using a grid testbed.

## 3. The Assessment

In both cases, most of the requirements identified in a prior step can be met. However, due to their conceptual differences, both have their individual advantages and disadvantages. While the integrated approach not only requires modifications of the grid middleware but also of policy decision and enforcement points, the abstraction solution creates additional overhead and possibly some fault tolerance issues. Yet, the latter concept is easier to deploy and existing legacy solutions can easily be integrated. The former concept in turn allows for a smoother integration of non-grid authorization schemes (like those known from Shibboleth) and eventually it is more flexible regarding changes in underlying Authentication and Authorization Infrastructures (AAI).

## 4. Conclusions

The presentation and the full conference contribution provide a detailed assessment of both approaches. The result of this assessment indicates that a combined solution may be beneficial.

**Acknowledgements**. This work was funded in parts by the German Federal Ministry of Education and Research as part of the D-Grid Initiative.

References
1. Gietz, P., Grimm, C., Gröper, R., Makedanz, S., Pfeiffenberger, H., Schiffers, M., Ziegler, W.: A Concept for Attribute-Based Authorization on D-Grid Resources, to appear in Future Generation Computer Systems – The International Journal of Grid Computing: Theory, Methods and Application
2. Kirchler, W.: Entwicklung einer einheitlichen Autorisierungs- und Authentifizierungsschnittstelle für heterogene Grids am Beispiel D-Grid, Diploma Thesis (in German) at the Technische Universität München, 2008

# 11. Guarantee and Penalty Clauses for Service Level Agreements

Dominic Battré (1), Georg Birkenheuer (4), Vikas Deora (2), Matthias Hovestadt (1), Omer Rana (2), Oliver Wäldrich (3)

*(1) Technische Universität Berlin, Germany*
*(2) Cardiff University, UK*
*(3) Fraunhofer Institute SCAI, Germany*
*(4) Universität Paderborn, Germany*

On the path of bringing Grid technologies from academia into a commercial environment, Service Level Agreements (SLAs) play a crucial role. They define the service to be delivered, as well as the reward and the penalty for fulfilling or violating the contract. SLAs can be used in various Grid computing scenarios with considerable implications on the type of penalties used and required: computational jobs, service jobs, or even workflow tasks.

WS-Agreement (WSAG) (1), a proposed recommendation by the OGF, is designed to provide a framework for expressing and establishing SLAs. WSAG is domain independent and needs to be supplemented by domain specific extensions, providing extension points for various aspects of SLAs to keep its generality. However, neither structure (in terms of XML structure) nor identifiers (used to reference real world concepts) are prescribed by WSAG.

The Job Submission Description Language (JSDL) (2) became a popular common denominator among several projects using WSAG such as AssessGrid (3), SORMA (4) and VIOLA (5). It allows one to give detailed descriptions about the environment in which a job shall be executed. Guarantees on QoS, however, remained highly domain specific and individual from project to project.

In this paper we suggest a structure to represent guarantees and penalties in SLAs. This structure covers a wide range of scenarios while maintaining generality. Domain specific terms (e.g. to represent the amount of available RAM) remain open for the purpose of generality.

It does not only have to be distinguished between multiple types of SLAs, but also between multiple types of SLA violations. This underlines the focal importance of violation detection for the SLA management, as well as the necessity of a model for expressing guarantees and penalties as well as methods for the fulfillment assessment process, handling guarantees in a very abstract way. We will describe a model, allowing the subscription to KPI (Key performance indicator) data from sensors that monitor the job or service. The SLA needs to specify the frequency at which data are requested, the source and kind of data, and the consequences of failing to deliver data.

These measurements are then aggregated in a second step to scalar values that are passed to a filter. The aggregation might calculate the running average of the ten most recently encountered values or the maximum seen so far. Either after each value or after all values from the input are received the aggregation passes a value to the filter. The filter contains a predicate like “>” to compare the output of the filter against some constant. If the condition is true, the effects are triggered. The effects might modify the reward and penalty and/or terminate the SLA. To enable more complex scenarios, the effects may act as a sensor that creates input for other such chains.

**Acknowledgements**. The authors would like to thank the EC for partially supporting this work within the 6th Framework Programme under “Advanced Risk Assessment and Management for Trustable Grids” (AssessGrid) and “SORMA - Self-Organizing ICT Resource Management”.

References

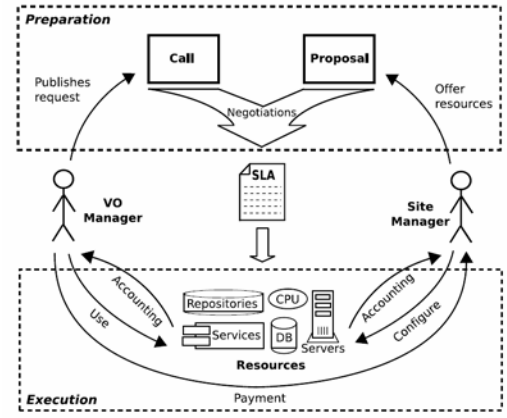
1. Andrieux, A., Czajkowski, K., Dan, A., Keahey, K., Ludwig, H., Kakata, T., Pruyne, J., Rofrano, J., Tuecke, S., Xu, M.: Web Services Agreement Specication (WS-Agreement). Technical report, Open Grid Forum (2007)
2. Anjomshoaa, A., Brisard, F., Drescher, M., Fellows, D., Ly, A., McGough, S., Pulsipher, D., Savva, A.: Job Submission Description Language (JSDL) Specification, Version 1.0. Technical report, Open Grid Forum (2005)
3. AssessGrid – Advanced Risk Assessment and Management for Trustable Grids. http://www.assessgrid.eu (2008)
4. SORMA – Self-Organizing ICT Resource Management. http://www.iw.uni-karlsruhe.de/sormang (2008)
5. VIOLA – Vertically Integrated Optical Testbed for Large Application. http://www.viola-testbed.de (2008)

# 12. SLA Negotiation and Enabling Resources for Users in Grids

Tomasz Szepieniec, Anna Pagacz

*ACC CYFRONET AGH, Krakow, Poland*

The usage of resources managed in decentralized way and delivering non-trivial Quality of Service cannot be fully automatized. The authors believe that Service Level Agreement (SLA) negotiation would remain in grid environments as a human activity related with resource policy. This process could be completed by automatized mechanisms on a different level [1-3]. This paper includes the SLA metrics proposal, model of communication and a tool to standardize and simplify SLA negotiations. Additionally, description is provided for an application of those ideas to a large scale grid environment, namely to EGEE infrastructure [4].




The whole process related to a SLA between a user and a site provider is depicted on the figure. In the preparatory stage, partners, namely user representative (VO manager), and resource provider representative (Site Manager) negotiate a SLA. Later an execution stage follows, in which agreement is implemented in the configuration of resources, then resources are used according to the agreement and a reward for deliver services is provided. What make the problem complicated, is that in practice both actors are simultaneously involved in many such interdependent processes while the pool of resources is limited.

The formal model of this process is described in this work. Additionally, we provide a proposal for metrics that would be helpful in defining SLA. The metric set includes: overall quality of service, computational resources, storage resources, networking resources and additional services. Those metrics are adapted to be measured in EGEE infrastructure or similar.

To support managers dealing with complexity, we need a collaboration tool that would follow the process defined. Therefore we developed a web portal, called Bazaar [5]. The tool provides support for dealing with communication including negotiations, gives a clear view on policies of VOs and resource providers. Furthermore, it enables tracking of contracts and calls execution and gives a possibility to provide feedback about sites and VOs. Bazaar is fully integrated within EGEE environment and becomes part of the main EGEE operational portal [6].

**Acknowledgments**. This work is partially funded by EU Project, EGEE-III INFSO-RI-222667

References

1. Grid Resource Allocation Agreement Protocol WG, web page: https://forge.gridforum.org/sf/projects/graap-wg
2. Pichot: Dynamic SLA-negotiation based on WS-Agreement, Core Grid Technical Report, TR-0082. http://www.coregrid.net/mambo/images/stories/TechnicalReports/tr-0082.pdf
3. G. Di Modica, V. Regalbuto, O. Tomarchio, L.Vita: Enabling re-negotiations of SLA by extending the WS-Agreement specification, IEEE SCC 2007: 248-251
4. EGEE Project, web page: http://www.eu-egee.org/
5. Bazaar, Project Web Page: http://grid.cyfronet.pl/bazaar
6. EGEE CIC Operations Portal, web page: http://cic.in2p3.fr

# 13. Enforcing Rules of Software Licenses in the Chemomentum Grid Infrastructure

Krzysztof Benedyczak (1,2), Piotr Bała (1,2)

*(1) ICM, Warsaw University, Poland*

*(2) Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland*

Applications which are exposed by a grid middleware are quite often subject to commercial or constrained licenses, i.e. are not a free software. Handling of the usage restrictions of such software is typically not addressed by standard grid solutions. However this situation changes as grid middleware matures: number of commercial applications present in a grid environment is increasing. The problem started to be so significant that even a dedicated EU project (SmartLM [1]) was bring into existence.

The Chemomentum project aim is to support the most common requirements of the constrained licenses in a way which is easy to manage. In general it should be transparent for the grid user. As an example of such requirements we can enumerate:

1. applications available only to the selected grid users,
2. applications available with restricted number of concurrent (simultaneous) invocations,
3. applications which can be invoked only at particular grid nodes.

As a solution we propose a system which at lower level can be seen as a generic grid resources authorization service with additional informational and access reservation capabilities. By *grid resources* we understand primarily grid software but the concept can be used also for other resources like databases available on the grid. The service is using XACML [2] as an underlying language for expression of license rules. Usage of standard language allows for exploiting software already available e. g. an XACML processing engine. On the other hand the language is very complicated and unsuitable for the administrators of the grid software. Therefore significant amount of work had to be performed to develop easy to use system interface.

The proposed license system is integrated with the VO infrastructure [3] of the Chemomentum project which is used for obtaining information about grid users. We provide tools for the UNICORE grid middleware to ask for license availability and to enforce its rules upon a grid job execution. It is planned to further integrate the system with broker service.

**Acknowledgments**. This work was funded by EU project Chemomentum (IST-033437).

References

1. SmartLM project, http://www.smartlm.eu (14.09.2008).
2. Tim Moses (ed); eXtensible Access Control Markup Language (XACML) Version 2.0, OASIS Standard, 1 Feb 2005 http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf
3. UVOS system, http://uvos.chemomentum.org/ (14.09.2008). The work was presented as a poster at CGW'07: K. Benedyczak, P. Bała; Chemomentum Virtual Organization Services.

## 14. Semantically Supported Security in Virtual Organizations

Bartosz Kryza (1), Lukasz Dutka (1), Renata Slota (2), Jacek Kitowski (1,2)

*(1) ACC CYFRONET AGH, Krakow, Poland*
*(2) Institute of Computer Science AGH, Krakow, Poland*

One of the main aspects of modern IT infrastructures such as SOA or Grids, especially when applied in commercial settings is the support for advanced and flexible security mechanisms. The security itself is a complex issue which includes among others such aspects as authentication, authorization or trust. While authentication is currently pretty well covered by common standards such as X509 certificates [1] or Shibboleth [2], authorization and trust are issues which raise several open research topics. In general authorization is the process of granting or rejecting access to a resource to a user in a given context. The resource can be either a file, database, service or an other entity that is referenced within the Virtual Organization.. The main difficulty in proper definition of authorization rights lies in how the resources themselves and the rules can be described. Although currently several systems exist, such as PERMIS [3], Akenti [4] and others which allow for definition of quite flexible authorization policies, their main drawback is the fact that there is no standard for definition of resources, which are described by these policies, usually limiting comparison of resources based on simple keyword comparison. This poses a certain problem in heterogeneous environments such as Grid or SOA-based Virtual Organizations, where several organizations need to agree on common rules of resource sharing.

We propose the use of modern technologies from Semantic Web research including ontological description of resources and reasoning in order to make the process of definition and authorizing access rights more flexible. Our solution, FiVO (Framework for Intelligent Virtual Organizations) [5] allows for dynamic and distributed inception of a Virtual Organization based on a collaborative contract negotiation between participating organizations. FiVO uses a distributed knowledge base – Grid Organizational Memory [6] – as an ontological repository. After the contract is agreed upon, an ontological agreement is defined and can be used to create automatically proper access rights and quality of service requirements in order to enforce envisioned operation of the Virtual Organization.

This paper presents early results of applying ontological description and reasoning for the purpose of securing the access to resources in SOA-based environments, based on the FiVO framework. Our approach includes an ontological PDP service (Policy Decision Point) which is used to answer authorization queries within a distributed environment, based on the agreed contract of the Virtual Organization. This component can be used both in Grid based and SOA based environments by means of specially developed set of request interceptors for Globus in case of Grid environments and special Apache plugin for generic SOA environments.

We will present how the semantic description of resource and the access rights themselves, introduce flexibility into the overall security layer of inter-organizational IT infrastructures.

**Acknowledgements**. This work was supported by EU project Gredia IST- 34363 with the related Polish grant SPUB-M.

References

1. V. Welch, I. Foster, C. Kesselman, O. Mulmo, L. Pearlman, S. Tuecke, J. Gawor, S. Meder, F. Siebenlist, X.509 Proxy Certificates for Dynamic Delegation, 3rd Annual PKI R&D Workshop, 2004.
2. Mark Needleman, The Shibboleth Authentication/Authorization System, Serials Review, Volume 30, Issue 3, 2004, Pages 252-253.
3. W. Chadwick and A. Otenko., The PERMIS X.509 role based privilege management infrastructure. Future Generation Comp. Syst., 2, 19, 277-289, 2003, Elsevier
4. M. R. Thompson, A. Essiari, K. Keahey, S. Lang and B. Liu, Fine-Grained Authorization for Job and Resource Management Using Akenti and the Globus Toolkit, CoRR, 2003
5. B. Kryza, L. Dutka, R. Slota, J. Kitowski, Supporting Management of Dynamic Virtual Organizations in the Grid through Contracts, in: M. Bubak, M. Turala, K. Wiatr, Proceedings of Cracow'07 Grid Workshop, Oct 15-17 2007, Cracow, Poland, ACC Cyfronet AGH, 2008, pp.140-147
6. Bartosz Kryza, Renata Slota, Marta Majewska, Jan Pieczykolan, Jacek Kitowski, Grid organizational memory – provision of a high-level Grid abstraction layer supported by ontology alignment, The International Journal of FGCS, Grid Computing: Theory, methods & Applications, vol. 23, issue 3, Mar 2007, Elsevier, 2007, pp. 348-358

# 15. Secure User Management in a Grid Framework

Daniel Harężlak (1), Piotr Nowakowski (1), Marian Bubak (1,2)

*(1) ACC CYFRONET AGH, Krakow, Poland*
*(2) Institute of Computer Science AGH, Krakow, Poland*

The development of modern Grid infrastructures and virtual laboratory framework is heading in the direction of increased involvement of various groups of nonexpert users, typically called *actors*. In lieu of a group of collaborating scientists, with equal expertise in the areas of domain science and computing technologies, modern grid application environments, such as MyGrid Taverna [1] or GridSpace [2] increasingly cater to laymen interested in applications which rely on distributed computing solutions. Thus, a differentiation of user roles becomes a priority in modern virtual laboratory frameworks.

The GREDIA project is developing an infrastructure for sharing of data and computational resources in the context of banking and media domains. A prerequisite of such application areas is that the system is capable of differentiating actors and delivering functionality according to the role each actor plays in the given application scenario. Moreover, the solution should enable secure sharing of data and secure invocation of computational resources (for instance, when calculating the risk of a banking loan, the solution must ensure that the actual loan scenario is only visible to members of a specific “banking” Virtual Organization, while at the same time retaining the flexibility to be applied in a number of different application contexts). The Appea application development framework enables developers to create grid application scenarios on demand and deliver the functionality listed above, while at the same time retaining a Single Sign-On policy for all operations on grid resources and ensuring that each given application scenario can affect multiple users, according to their roles in a given organization. This functionality is delivered by appropriating existing grid security frameworks, such as Permis [3] and MyProxy [4], as well as the novel Framework for Intelligent Virtual Organizations [5]

This paper presents the structure of the security solution delivered by the Appea platform, explains how the system approaches user authorization and how users can be differentiated and addressed from within the actual application workflows which form the core part of Appea use cases. We also present sample usage scenarios and list the benefits of the Appea solution in comparison to existing business process description tools.

**Acknowledgements**. This work was supported by EU project Gredia IST- 34363 with the related Polish grant SPUB-M.

References

1. T. Oinn et al., Taverna: a tool for the composition and enactment of bioinformatics workflows, Bioinformatics 2004 20(17):3045-3054; doi:10.1093/bioinformatics/bth361
2. P. Nowakowski, D. Harężlak, M. Bubak, A New Approach to Development and Execution of Interactive Applications on the Grid, proc. of CCGrid’08.
3. W. Chadwick and A. Otenko., The PERMIS X.509 role based privilege management infrastructure. Future Generation Comp. Syst., 2, 19, 277-289, 2003, Elsevier
4. The MyProxy project, www.myproxy.ca

5. B. Kryza, L. Dutka, R. Slota, J. Kitowski, Supporting Management of Dynamic Virtual Organizations in the Grid through Contracts, in: M. Bubak, M. Turala, K. Wiatr, Proceedings of Cracow'07 Grid Workshop, Oct 15-17 2007, Cracow, Poland, ACC Cyfronet AGH, 2008, pp.140-147

## 16. A Security Infrastructure for MOCCA Component Environment

Michał Dyrda (1), Maciej Malawski (1), Syed Naqvi (3), Marian Bubak (1,2)

*(1) Institute of Computer Science AGH, Mickiewicza 30, 30-059 Krakow, Poland*
*(2) Academic Computer Center CYFRONET AGH, ul. Nawojki 11, 30-950 Krakow, Poland*
*(3) CETIC, Rue des Freres Wright 29/3, B-6041 Charleroi, Belgium*

The subject of this paper is a detailed analysis and development of security in Grid component systems on the example of MOCCA [1]. MOCCA is a Common Component Architecture compliant framework which supports building and running scientific applications on the Grid. It is based on H2O [2] resource sharing platform which provides a Java container for remote deployment, communication and management of components. Security of such system is a critical issue, because not only data, but also hosts, resources and computations have to be secured from improper access.

In the paper we discuss the security requirements of MOCCA and analyze the existing solutions offered by H2O. Among them such issues as authentication, authorization, single sign-on (SSO) and credential delegation as well as communication security and sandboxing are of interest. Current security mechanisms present in H2O include component sandboxing, pluggable authenticators and SSL-based transport security, but the main drawback is the lack of SSO and proper credential delegation. These features are important for distributed applications in the case of components which need to run on shared resources and initiate actions on user's behalf. One of the optional, but important requirements was the compatibility with the existing solutions widely accepted in large-scale Grid infrastructures such as EGEE.

Based on the detailed analysis of H2O security model and on the review of available security technologies (such as Shibboleth, GSI, MyProxy, OpenID) we designed and implemented a new solution which is the GSI Authenticator for H2O. It is based on Grid Security Infrastructure (GSI) [3] developed for Globus Toolkit and being used by e.g. EGEE infrastructure. The proposed solution makes use of X.509 proxy certificates to support SSO and delegation. The authenticator was integrated with H2O and MOCCA and its usability was successfully demonstrated.

The developed GSI Authenticator was subject to threat analysis regarding potential sources of weakness and possible attacks. The performance of the authenticator was also studied in detail in various configurations. The obtained results suggest that although the GSI-based solution introduces a considerable performance overhead, the benefits of stronger security and facilitated usage (SSO, delegation) can outperform other solutions, particularly in the case of having access to existing Public Key Infrastructure as is the case of European Grid infrastructures. For more detailed description of the GSI Authenticator please refer to [4].

**Acknowledgements**. This work was supported by EU project GREDIA with the related Polish grant SPUB-M.

References

1. Maciej Malawski, Dawid Kurzyniec, Vaidy Sunderam: MOCCA - towards a distributed CCA framework for metacomputing. In Proceedings of the 10th International Workshop on High-Level Parallel Programming Models and Supportive Environments (HIPS2005) in conjunction with International Parallel and Distributed Processing Symposium (IPDPS 2005). IEEE Computer Society, 2005.
2. Dawid Kurzyniec, Tomasz Wrzosek, Dominik Drzewiecki, and Vaidy Sunderam: Towards Self-Organizing Distributed Computing Frameworks: The H2O Approach. Parallel Processing Lett., 13(2):273-290, 2003.Maciej Malawski, Dawid Kurzyniec, and Vaidy Sunderam.
3. Von Welch, Frank Siebenlist, Ian Foster, John Bresnahan, Karl Czajkowski, Jarek Gawor, Carl Kesselman, Sam Meder, Laura Pearlman, Steven Tuecke: Security for Grid Services, *HPDC*, p. 48-57, 12th IEEE International Symposium on High Performance Distributed Computing (HPDC-12 '03), 2003
4. Michał Dyrda. Security in Component Grid Systems. Master's thesis, AGH University of Science and Technology, Krakow, Poland, 2008.

## 17. Threat Model for MOCCA Component Environment

Jan Meizner (1), Maciej Malawski (1), Syed Naqvi (3), Marian Bubak (1,2)

*(1) Institute of Computer Science AGH, Mickiewicza 30, 30-059 Krakow, Poland*
*(2) Academic Computer Center CYFRONET-AGH, ul. Nawojki 11, 30-950 Krakow, Poland*
*(3) CETIC, Rue des Freres Wright 29/3, B-6041 Charleroi, Belgium*

Nowadays, no widely available network services are able to exist without providing proper security mechanisms. Our goal is to asses potential threats against MOCCA [1] components running inside H2O [2] containers and to propose new authentication and authorization mechanisms compatible with those used in the ViroLab project. H2O is a Java-based software that allows users to deploy and run pluglets in containers called kernels located on various nodes of a distributed system. It provides all necessary transport protocols (e.g. RPC, SOAP) as well as basic security mechanisms (like SSL encryption and access control mechanism based on users and groups). MOCCA is the CCA framework enabling its users to build complex distributed applications from interconnected components.

Since it is crucial to protect security software from any vulnerabilities that might be used to compromise it, we decided to perform advanced analysis of this aspect by creating the Threat Model [3] [4] for the MOCCA/H2O. In this model, we show security requirements of the system, protected assets, and a selected use case of the sample testing MOCCA component. We analyze possible entry points to the system, trust levels and create STRIDE classification of the threats by specifying their types such as spoofing, tempering, repudiation, information disclosure, denial of service and elevation of privilege. We also show relations between the entry points and the threats that are affected by them as well as possible attack scenarios and mitigations for the described threats.

Based on the knowledge gained during creating this model, we were able to safely plan further development of MOCCA security features. The goal was to combine MOCCA and Shibboleth [5] security infrastructure used in ViroLab. Shibboleth is Web-based federated SSO authentication and authorization framework. It allows many users working at various institutions to be authenticated at their organizations (called here Home Organization) by component called Identity Provider, and to be authorized to use, on the basis of attributes assigned to them, services supplied by other member of the federation using component called Service Provider. Shibboleth is based on a very well established security standard - Security Assertion Markup Language (SAML) used to exchange both authentication information and attributes used for authorization decisions.

The features of Shibboleth made it suitable for ViroLab project, but there was a need to enhance Shibboleth software to include support for interaction not just between a human and a machine like in standard Shibboleth scenario, but also between just machines e.g. between H2O kernels. For that purpose we created specific software enabling users to authenticate against Identity Provider protection mechanism without need to use any web browser. Our current research is focused on a solution for delegating Identity Provider's authorization decisions and attributes assertions. It is based on GridShib [6], which allows integrating Shibboleth and GSI [7] - X.509 certificates based security mechanism used by Globus Toolkit. It is able to generate GSI proxy certificates with SAML assertions embedded in the certificate as non-critical extensions. Combining Shibboleth and GSI will supply a homogeneous method of secure access to MOCCA/H2O software for the large group of users.

**Acknowledgements**. This work was supported by EU project ViroLab IST-027446 with the related Polish grant SPUB-M.

References

1. Maciej Malawski, Dawid Kurzyniec, Vaidy Sunderam: MOCCA - towards a distributed CCA framework for metacomputing. In Proceedings of the 10th International Workshop on High-Level Parallel Programming Models and Supportive Environments (HIPS2005) in conjunction with International Parallel and Distributed Processing Symposium (IPDPS 2005). IEEE Computer Society, 2005.
2. Dawid Kurzyniec, Tomasz Wrzosek, Dominik Drzewiecki, and Vaidy Sunderam: Towards Self-Organizing Distributed Computing Frameworks: The H2O Approach. Parallel Processing Lett., 13(2):273-290, 2003.
3. Amit D. Lakhani, Erica Yang, Brian Matthews, Ian Johnson, Syed Naqvi, Gheorghe C. Silaghi. Threat Analysis and Attacks on XtreemOS: a Grid–enabled Operating System. Towards Next Generation Grids, Proceedings of the CoreGRID Symposium 2007, p. 53-62, Springer, 2007
4. Syed Naqvi, Michel Riguidel. Threat Model for Grid Security Services, Advances in Grid Computing –- EGC 2005, European Grid Conference, Amsterdam, The Netherlands, February 14-16, 2005, Revised Selected Papers, LNCS 3470, p. 1048-1055, Springer, 2005.
5. http://shibboleth.internet2.edu/
6. http://gridshib.globus.org/
7. http://www.globus.org/security/overview.html

# 18. Towards Analytic Workload Models for Improving Grid Scheduling

Paul Heinzlreiter, Jens Volkert
*GUP, Joh. Kepler University Linz, Austria*

Nowadays grid computing has evolved into a valuable tool for day-to-day research offering computational and data services to e-science applications from various different domains. Grid resource brokers are commonly used to assign jobs to resources. However the quality of the assignments is heavily dependent on the amount of information available to the resource broker. Runtime estimations for specific job / resource combinations are specifically valuable, since they enable to estimate when a resource will be free again. Another important application is given by scheduling of distributed pipelines for example within visualization scenarios to avoid bottlenecks and maximize the overall throughput [4].

There are several factors which are influencing the runtime such as the runtime complexity of the algorithm, the size of the input data and the performance of the executing resource. The best runtime estimation would of course be provided by measuring the time consumed by a test run. However since the runtime is mainly influenced by input data size which may vary considerably, an analytic model which can be used to calculate a runtime estimation for a specific input data size and resource without actually executing the job is valuable tool for supporting grid scheduling decisions.

We have chosen to model the workload of an algorithms execution rather than runtime, since for scheduling decisions a workload model can be used to quickly compare the fitness of different resources for a specific job.

Our workload models are comparable to other methods such as the big O notation for the runtime complexity of an algorithm [3] or the analytic workload models for some visualization algorithms described by Bowman [1]. However our workload models represent the workload of algorithms more precisely for a specific input data set compared to the big O notation which only expresses the asymptotic behavior of an algorithm and are more detailed and better validated then the ones given by Bowman et al.

Since a distributed visualization pipeline represents a good application example, three common visualization algorithms have been modeled as a first step: A vector glypher, an isosurface extraction algorithm, and a streamline generator. The workload of these algorithms is linearly dependent on the input data size according to runtime measurements on several different resources, therefore the models are expressing the workload as a linear combination of input data characteristics. While the workload of the vector glypher is only dependent on the input data size, the runtime of the isosurface extractor is also dependent on the size of the output data. The characteristics relevant for the streamline generation are even more complex, since this model mainly depends on the number of integration steps required to calculate a streamline.

The validation of the workload models has been done by comparing the actual CPU time consumed during test runs with the quotient of the workload and the computational capacity of the executing resource, which is determined using the HINT benchmark [2]. The median deviations of the workload predictions delivered by our modules from the workloads measured on a set of grid resources with diverse hardware characteristics range from six percent for the streamliner model to eight percent for the isosurface extractor.

Future work will investigate workload models for well-known parallel algorithms using the message-passing paradigm as for example matrix multiplication, sorting, or n-body problems.

**Acknowledgements**. This work was partly supported by the Austrian Grid Project funded by the Austrian Federal Ministry of Science and Research under contract GZ BMWF-10.220/0002-II/10/2007. The results discussed in this work were partially achieved using grid resources provided by Austrian Grid and the EU funded projects Interactive European Grid and EGEE-II.

References

1. Bowman, J. Shalf, K. Ma, W. Bethel: Performance Modeling for 3D Visualization in a Heterogenous Computing Environment; Tech. Rep. LBNL 56977, Visualization Group, Lawrence Berkley National Laboratory, 2004
2. J.L Gustafson, Q.O. Snell: HINT: A New Way to Measure Computer Performance; Proceedings of the 28th Hawaii International Conference on System Sciences, pp. 392-401, Jan., 2005
3. D.E. Knuth: Big Omicron and big Omega and big Theta; ACM SIGACT News, vol. 8, no. 2, pp. 18-23, Apr.-Jun., 1976
4. M. Zhu, Q. Wu, N.S.V. Rao, S. Iyengar: Optimal Pipeline Decomposition and Adaptive Network Mapping to Support Distributed Remote Visualization; Journal of Parallel and Distributed Computing, vol. 67, no. 8, pp. 947-956, Aug., 2007

# 19. Distributed Dynamic Load Balancing for Iterative-Stencil Applications

Gérard Dethier (1), Pierre-Arnoul de Marneffe (1), Pierre Marchot (2)

*(1) EECS Department, University of Liège, Belgium*
*(2) Chemical Engineering Department, University of Liège, Belgium*

In the context of jobs executed on heterogeneous clusters or Grids, load balancing is essential. Indeed, a slow machine must receive less work than a faster one or the overall job termination will be delayed. This is particularly true for Iterative-Stencil Applications' Jobs where tasks are run simultaneously and are interdependent. The problem of assigning coexisting tasks to machines is called mapping.

An Iterative-Stencil Application (ISA) can be represented by an undirected graph (the ISA graph) where vertices maintain a state and edges indicate bidirectional data flows. Each vertex sends informations about its current state to its neighboring vertices and waits informations from them to update its state. This process is generally repeated several times. We consider the class of ISAs where the vertices all have the same weight (they represent the same amount of work). The edges of the ISA graph are weighted by the amount of data transferred during the execution.

An heterogeneous cluster can also be represented by an undirected graph (the machine graph). Each vertex is weighted with the power of the corresponding machine. The edges are not weighted, which means an homogeneous bandwidth is assumed for all network links.

If each machine runs exactly one task, the load balancing problem is solved with a good mapping of the ISA graph on the machine graph (partitioning the ISA graph and assigning each partition to a machine). A good mapping has the following features: (1) the amount of vertices of the ISA graph associated to a machine is proportional to its power and (2) the inter-machine data flow is minimized. This optimization problem was proven to be NP-Complete.

When the access to a cluster is managed by a middleware, the subset of the cluster's machines that is available is not known at submission time. Also, the machine graph representing this subset can change over time, either in the number of vertices (new machines become available, others become unavailable), or in the weight of the vertices (a machine can become slower because of the background load). Rebalancing is needed each time the machine graph changes.

Some heterogeneous mapping methods currently available (Quick-Quality Map [1], MiniMax [2], Fastmap [3]) address even more general problems but are designed for static mapping. Furthermore, only Fastmap features a distributed mapping scheme and, therefore, is potentially scalable. However, a prerequisite to Fastmap is the existing hierarchical organization of schedulers.

We propose a method resulting from the combination of existing algorithms (namely the distributed spanning tree construction algorithm [4] and the Tree-Walking Algorithm (TWA) [5]) to achieve fast load balancing and, more importantly, rebalancing for ISAs in the context of dynamic heterogeneous clusters. A spanning tree is first constructed using a distributed algorithm. Then, the ISA graph is initially homogeneously distributed across the machines of the spanning tree (and therefore, of the machine graph). The TWA, initially intended for parallel scheduling, is then used to balance the load. Finally, each machine refines its partition using a local optimization method by exchanging nodes with its neighbors. The only initial information that must be available to each machine is its power and its neighborhood (required to build the spanning tree) defined by the machine graph.

Our approach achieves very good load balancing and acceptable partitions quality but, more importantly, runs very fast. Another interesting feature is the iterative nature of a remapping. If the machine graph has only changed a little, the partitions resulting from the remapping are close to the previous partitions, which leads to a small number of ISA graph vertices migrations.

References

1. Panu Phinjaroenphan, Savitri Bevinakoppa, Panlop Zeephongsekul, “A Heuristic Algorithm for Mapping Parallel Applications on Computational Grids”, in Proc. EGC 2005, LNCS 3470, pp. 226-236, Springer, 2005.
2. Shailendra Kumar, Sajal K. Das, Rupak Biswas, “Graph Partitioning for Parallel Applications in Heterogeneous Grid Environments”, IPDPS, p. 0066, International Parallel and Distributed Processing Symposium - Symposium Volume, 2002.
3. Amit Jain, Soumya Sanyal, Sajal K. Das, Rupak Biswas, "FastMap: A Distributed Scheme for Mapping Large Scale Applications onto Computational Grids", *CLADE*, p. 118, 2004.
4. Radia Perlman, “An algorithm for distributed computation of a spanningtree in an extended LAN”, ACM SIGCOMM Computer Communication Review, Vol. 15, Issue 4, pp. 44-53, 1985.
5. Wei Shu, Min-You Wu, “Runtime Incremental Parallel Scheduling (RIPS) on Distributed Memory Computers”, IEEE Transactions on Parallel and Distributed Systems, Vol. 07, no. 6, pp. 637-649, June, 1996.

# 20. Biz2Grid: An Implementation of Market-based Grid Scheduling

Jochen Stößer, Thomas Meinl
*Institute of Information Systems and Management, Universität Karlsruhe (TH), Germany*

In recent years, many papers and research projects have proposed and investigated the usage of economic principles (in particular auctions and negotiations) in allocating Grid resources (e.g. [1, 2, 3]). The underlying idea is that markets perform well in allocating scarce resources among self-interested market participants. For instance, in Grid settings, prices (i) cause the users to make efficient use of the resources and (ii) provide incentives to resource owners to contribute their scarce resources to the Grid. However, only few of these projects actually *implemented* economic logics into Grid middleware, like e.g. the Globus Toolkit (GT).

One aim of the Biz2Grid project (www.biz2grid.de), funded by the German D-Grid Initiative, is to bridge this apparent gap between theoretical research and practical implementation by integrating an economic framework into GT. Two essential components in this framework are

- *Market-based scheduling logics*: Existing, purely technical scheduling algorithms are enriched to take into account the participants' economic preferences (such as the valuation of users and the costs of resource providers) to generate efficient allocations in the economic sense.
- *Intelligent tools*: Participants (users and provides) cannot be expected to continuously monitor and interact with the market. Consequently, configurable bidding agents are needed that automatically interact with the market on behalf of the participant, thus shielding parts of the system's complexity.

This paper presents a conceptual architecture for integrating these components into GT, cf. Figure 1. The scheduling component is an extension of the GridWay metascheduler, which is already integrated into GT. GridWay's technical scheduling policies are replaced with a policy that takes into account both the technical attributes of jobs and resources as well as the participants' economic preferences. Technical attributes are taken from JSDL files and GT's GRAM component, respectively, while economic attributes are sourced from an SQL database. Comparable to the eBay approach, the bidding agents run on the market server and are configured via an AJAX-based Web Interface (cf. Figure 2). The bidding agents populate the database with the economic attributes on behalf of the participants and based on their policy-based configuration.

This architecture has two main benefits: (i) The economic extensions to GT and GridWay are purely complementary. The system can thus be easily configured to operate as usual, i.e. without any economic logic or component. (ii) Because of the chosen implementation of the bidding agents, the end-user's application does not need to be changed. However, due to the use of Web Service interfaces, the bidding agent could also be run on the client-side without having to change the system as such.

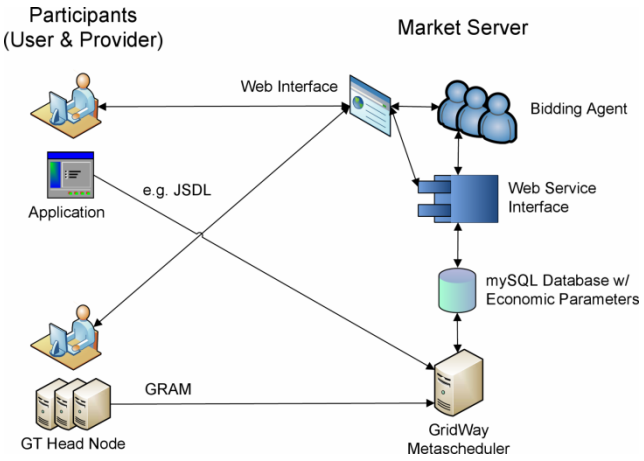



**Figure 1. Conceptual Architecture**

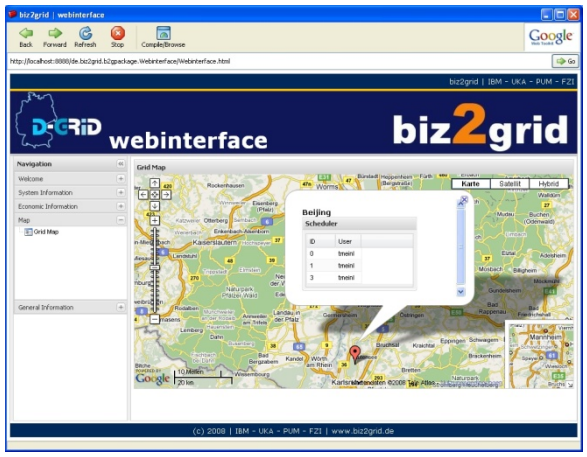



**Figure 2. Web Interface**

**Acknowledgements**. This work was supported by the German D-Grid Initiative under grant "Biz2Grid".

References

1. Rajkumar Buyya, David Abramson, Jonathan Giddy, Heinz Stockinger: Economic Models for Resource Management and Scheduling in Grid Computing; Concurrency and Computation: Practice and Experience, vol. 14, no. 13-15, pp. 1507-1542, 2002.
2. Dirk Neumann, Jochen Stößer, Christof Weinhardt, Jens Nimis: A Framework for Commercial Grids – Economic and Technical Challenges; Journal of Grid Computing, vol. 6, no. 3, pp. 325-347, 2008.
3. Jochen Stößer, Arun Anandasivam, Nikolay Borissov, Dirk Neumann: Economic Virtualization of ICT Infrastructures; Proceedings of the Cracow Grid Workshop '06, p. 392, Oct., 2006.

## 21. << *cancelled* >>

## 22. Magrathea—Scheduling Virtual Grids with Preemption

Jiri Denemark (1,2), Mirek Ruda (1,3)

*(1) Cesnet, z. s. p. o., Prague, Czech Republic*
*(2) Faculty of Informatics, Masaryk University, Brno, Czech Republic*
*(3) Institute of Computer Science, Masaryk University, Brno, Czech Republic*

Virtual clusters, virtual grid or cloud computing have recently become well-known names for utilizing a single technology in grids—virtualization. Virtualization provides users with more flexible grid environment, where each of them can use her own environment, often optimized and tailored for her applications. This makes grids very flexible for end users. The price which is paid for higher flexibility is a slightly higher overhead and more scheduling complexity. However, using preemption it may enable resource-efficient coexistence of long-running jobs with services which only occasionally need to be awakened while consuming large portion of system resources.

In this paper we describe the Magrathea system we have developed for enabling a batch scheduling system to schedule jobs into virtual machines and how various types of preemption techniques with respect to Xen virtual machine monitor may be used for running services or to allow scheduling of high priority jobs. Techniques, such as suspending (freezing) a virtual machine and reducing memory and CPU power usable to a virtual machine are described together with their advantages and disadvantages. Preemption overhead, speed of preemption and resumption under different conditions including large memory and CPU intensive computations are presented together with solutions we have developed for the Magrathea system to reduce the overall overhead. Our work is also compared to related work of others, such as Nimbus (or Virtual Workspaces) or OpenNebula.

In the future work, we will concentrate on integrating the Magrathea system with a virtual cluster system for enabling seamless coexistence between virtual clusters and normal jobs and allowing a single scheduler to manage both entities at the same time to achieve better resource efficiency.

## 23. Billing Resources in Scientific Grid Networks

Arun Anandasivam (1), Dirk Neumann (2), Christof Weinhardt (1)

*(1) Institute of Information Systems and Management, University of Karlsruhe, Germany*
*(2) Department of Information Systems, University of Freiburg, Germany*

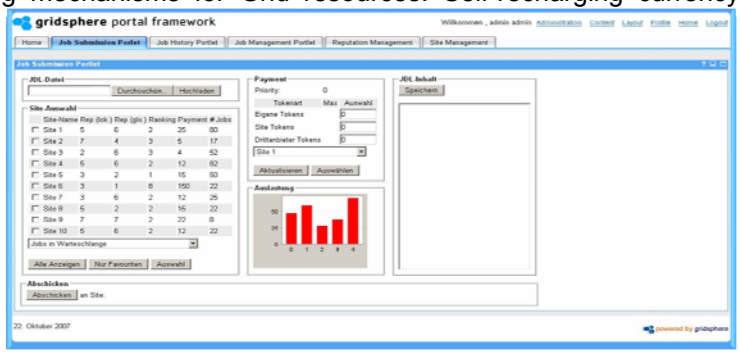
Since the Large Hadron Collider has recently gained attention in the public, stable scientific Grid networks are required to analyze the mass of upcoming data. Computation jobs are sent to the Grid to receive insights of the collected data. Due to the mass of data, computation and storage resources can get scarce. An efficient usage of the Grid infrastructure can be optimized by allocation mechanisms of jobs and resources. Economic models seem to be a promising approach for an efficient allocation of resources. However, the application of economic models for scientific environments faces some restrictions [1]. Unlike the commercial environment science has does avoid real money or hoarding of money. A fair contribution sharing of resources is desirable. Nevertheless, institutes in a scientific Grid network sometimes behave selfish, because resource sharing induces fewer resources for themselves.

The aim of the project *Billing the Grid*[1] is to provide a billing infrastructure for a fair exchange of resources in scientific Grid networks. The billing infrastructure is enhanced by incentive mechanisms to enforce an obedient and fair of the participants in the network. Three mechanisms were implemented and compared:

- *Grid Credits*: this is a simple substitute for real money. Every institute gets an initial amount of money, which they can use to buy resources. An entire economic system is required, which controls inflation, tries to avoid hoarding and subsidizing institutes without resources appropriately. From the real world it is known, how complex these systems get.
- *Resource-based exchange*: Participants can use resources based on their resource provision to the Grid network. The more they contribute, the more they can consume. This tit-for-tat billing is quite simple, but it lacks of avoid hoarding. Institutes without resource can hardly participate.
- *Self-recharging currency*: Every institute has a certain amount of money. As soon as the money is spent, the budget is recharged over time [3]. This approach avoids hoarding of money, but it does not set incentives to provide resources, since no compensation is paid. However several modifications of the system are possible, which were analyzed in the project.

[1] http://www.billing-the-grid.org/

The Grid Middleware in the particle physics domain is gLite[2]. We implemented all three mechanisms and proved the practical usage of different billing mechanisms for Grid resources. Self-recharging currency seems to be the most promising approach. However, since incentives for providing resources are missing we enhanced this mechanism by introducing a reputation system to evaluate resource provision [2]. Reputation of a provider can be evaluated by users or by independent institutions. We propose a metric based on the availability of the resources, because many providers tend to bulkhead servers of the network to support their own demand. We present how this reputation is incorporated with the self-recharging currency and how it enhances the social welfare of the entire network.




The billing system is running independently from the underlying middleware. The implementation of a middleware adapter enables the access the gLite system. The billing system is installed on a central server. Users can see their current account status and send jobs via a web-interface based on Gridsphere.

References

1. A. Anandasivam, D. Neumann (2008): Reputation-based pricing for Grid Computing in eScience. 16th European Conference on Information Systems (ECIS).
2. C. Dellarocas (2003). The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. Management Science 49 (10) 1407–1424.
3. Irwin, D.; Chase, J.; Grit, L.; Yumerefendi, A. (2005). Self-recharging virtual currency. In Proceeding of the 2005 ACM SIGCOMM.
4. Jurca, R. and Faltings, B. (2005). Reputation-based pricing of P2P services. P2PECON '05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of P2P systems.

## 24. Blending Routing Metrics for Optimalization the Best Path Selection in Multiple Protocol Environment

Marian Knezek (1,2)
*(1) Department of Computer Systems and Networks, Babyland, Bratislava, Slovakia*
*(2) Department of Informatics and Software Engineering, Babyland, Bratislava, Slovakia*

This paper represents a new solution for selecting the best path in particular the shortest path selection of several protocols at layer 3 OSI model. The first rule of multiple path selection is based only on protocol's reliability, which may cause that router decides for the less favorable route. Our proposed solution consists in the establishment of blended metrics by the application of comparative analyses of routing protocols. Comparison of metrics originated by various protocols for particular destination is created independently of administrative distance. Therefore the routers operate in a more objective way.

Routing intermediate data among grids influences total performance in virtual organization. The approach presented could be adapted to routing of computing elements as encapsulated packets in virtual organization operating arbitrary jobs in grid computing.

In more extensive networks, coexistence of several protocols is needed for routing at network layer. Protocols at network layer process of selecting appropriate paths to packets destinations [2, 3]. The best path selection influence overall network efficiency. Although the best path selection methods are relatively sophisticated within particular protocols, there are no warranty that router really chooses the best path which includes results among optimal routes particular protocols. On the present, the most common technique for the shortest path selection among several routing protocols is based on assigning administrative distance (AD). AD designates routing protocol reliability [2].

Shorter route can be routed by using less reliable routing protocol. Solutions based on AD behave router could decide for less favorable route in this best path selection from routing table what can affect less overall network efficiency.

We present the solution to finding the best route among more than one routing protocol problem. Our proposed solution supports communication among routing protocols in case of sharing metrics results. It provides that protocol metrics results are comparable and routing device can decided more accurately. An implemented solution makes decisions automatically.

[2] http://www.glite.org

One of benefits is 100% compatibility with existing network topologies. We can implement Blended Metrics Comparative Algorithm (BMCA) only on network nodes, where we want to reach better results. There is no necessity to change all router in network topology, because only influence individual router behavior.

The second benefit is versatility. Our solution can be implemented in each TCP/IP network, such as LAN and Internet, however our feature must be implemented in router's operating system. Local administrators make determination on which segment are blended metrics implemented. They can also determine range of use.

Currently, we are working on the evaluation of proposed solution on more extensive topologies. For the present, our experimental test acknowledges our theoretical findings that our method finds best routes more objective.

References


1. Orallo E., Carbo J.: Fast and efficient routing algorithms for delay-bounded and dependable channels, Comput. Commun. 2007
2. Lammle T.: CCNA Study Guide, 5th Edition, Sybex, Alameda, California, USA, 2005, ISBN: 9780782143911
3. Timm C., Wade E.: CCNP: BSCI Study Guide, 2nd Edition, Sybex, Alameda CA, USA, 2003, ISBN: 978-0782142938
4. Moy J.: OSPF Version 2, in: IETF RFC 2328, 1998
5. Callon P.: Use of OSI IS-IS for routing in TCP/IP and dual environments, in: IETF RFC 1195, 1990
6. EIGRP, IP Routing, IGRP. Available on (19.2.2008): http://www.rhyshaden.com/eigrp.htm


## 25. High Performance Data Access Aspects in National Data Storage


Renata Słota (1) Darin Nikolow (1), Marcin Jarząb (1), Jacek Kitowski (1,2)

*(1) Institute of Computer Science, AGH, Kraków, Poland*
*(2) Academic Computer Center CYFRONET AGH, Kraków, Poland*


Today, in the age of information, where computers are everywhere, the computer stored data are one of the most important things for the modern enterprises and other institutions. In the case of complete or even partial data loss the enterprise is usually quickly out of business. That is why some data backup method should be considered. Traditionally data are backed-up on removable media and kept in safe place. Such solutions can be very expensive, dependant on the level of automation and data availability deployed. Many institutions are looking for a cost efficient way to safely store their valuable data. Outsourcing is a popular method of cost reducing.

In Poland there is a constantly growing demand for highly available, reliable, fast and secure data storage services. A national project named KMD (National Data Storage) [1] has been started in order to meet these demands. The project concerns implementation of a geographically distributed storage system intended to provide high quality storage service using disk arrays and HSM systems located in the main computer centers in Poland and using fast network as an internal back bone. The system relies on the network facilities available within another national project – Pionier [2]. Pionier is intended to provide high bandwidth network to the scientific community and high performances computer centers in Poland.

In the KMD project besides implementing basic backup and archiving functionality a task for high performance data retrieval has been added. High performance data access in a distributed storage system can be achieved by using data replication techniques. For that purpose data are migrated or replicated automatically in order to have low access times. Proper monitoring is a must for such systems. Many parameters need to be monitored in order to be able to estimate the performance of particular storage subsystems. As part of our previous work a model for storage system monitoring with special attention taken to HSM systems has been developed [3].

This paper presents our experience of implementing replication techniques within the KMD project in order to achieve high performance distributed data access. The proposed solution uses JIMS as a monitoring framework. A JIMS plug-in for monitoring various storage systems is being developed. Due to the method of storing the monitoring values into database it is possible to implement different replication strategies according to the user requirements.


**Acknowledgements**. This research is supported by the MNiSW grant nr R02 055 03 and AGH grant nr 11.11.120.777.


References


1. *National Data Storage project*, Polish MNiSW grant nr R02 055 03, https://kmd.pcss.pl
2. *Pionier - Polish Optical Internet*, http://www.pionier.gov.pl

3. D. Nikolow, R. Slota, and J. Kitowski, *Grid Services for HSM Systems Monitoring*, in: R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski (Eds.), Proceedings of 7-th International Conference, PPAM 2007, Gdansk, Poland, September 2007, LNCS 4967, Springer 2008, pp.321-330.

# 26. Flexible and Scalable Grid Based Network Storage Protocol for Exabyte Scale

Karol Romanowski (1), Adam Nowaczyk (1), Lukasz Dutka (2) and Jacek Kitowski (1,2)

*(1) Institute of Computer Science AGH-UST, Krakow, Poland*
*(2) ACC CYFRONET AGH, Krakow, Poland*

The amount of storage grows day by day all over the world and simple disk matrices are unable to handle this demand. With this rapid growth it became obvious that there is a need for specialized storage data centers, which will administrate many geographically spread storage devices in order to provide one huge storage space. With this new approach new obstacles arise, just to mention scalability and efficiency as most important issues. Architects of such systems try to solve this problems using latest hardware solutions and faster network connections. Although this causes the increase in the available bandwidth, the increase in the throughput is not that significant. This is because the hardware is not the only important aspect. Software plays a great role in the whole process. Data transfer protocols are becoming bottlenecks for many applications. Especially in scientific computing and data intensive grid applications the problem of not efficient data transfer protocol is very common. Within the reliable transfer protocols being used in the Internet TCP is the undisputed leader. It is used extensively by many of the Internet's most popular application protocols and resulting applications, including the World Wide Web, E-mail, File Transfer Protocol, Secure Shell, and some streaming media applications. However its window based congestion control mechanism contains some drawbacks that prevent its use in high bandwidth-delay product (BDP) environments. Many scientists have been working on improving TCP or designing other alternative solutions to the problem.

We also took a closer look on the data transfer issues and we propose new flexible and scalable protocol able to work in the environment of high speed wide area networks, such as grid based networks. There are several possible approaches to design such protocol. We could write new implementation of TCP protocol with a better congestion control algorithm, however switching the current implementation of TCP protocol would not be easy. Updating the kernel of existing system could cause a lot of problems, as well as changing the configuration of existing routers wouldn't be a pleasant job to do. Because of that we chose another solution, to write the new protocol on top of the well known UDP protocol. Our job is to implement the connectivity and reliability, but keeping in mind the fact, that the protocol is designed to work for high speed wide area networks. Unfortunately the ease of using UDP as underlying protocol has its price. Since we are coding in the user space and we cannot modify kernel code, some extra memory coping occurs and we had to focus on the CPU usage and optimization. Nevertheless the gain of adjusting the protocol to the environment results in better throughput. Main feature of our protocol is use of negative acknowledgments, opposed to the positive acknowledgments in TCP. In our protocol receiver knows at the beginning how many bytes it will receive, so it can easily realize, which packets are missing and then ask sender to resend this packet.

The protocol is called IDP (IEBS Data Protocol) and is a part of the IEBS (Intelligent ExaByte Storage) system. IEBS system is the software for big scale, geographically spread storage data centers. It is able to create the network of specially managed disk (or disk matrices) in order to obtain one huge storage space. IDP protocol is used to transfer data between user and the system.

**Acknowledgments**. This work is being developed as the topic of the master thesis and is a part of the IEBS (Intelligent ExaByte Storage) system with the cooperation of the ACC CYFRONET AGH.

References

1. Yunhong Gu and Robert L. Grossman: Optimizing UDP-based Protocol Implementations, PFLDnet 2005
2. Ryan X. Wu, Andrew A. Chien, Matti A. Hiltunen, Richard D. Schlichting, Subhabrata Sen: A High Performance Configurable Transport Protocol for Grid Computing
3. D. Katabi, M. Hardley, and C. Rohrs: Internet Congestion Control for Future High Bandwidth-Delay Product Environments, ACM SIGCOMM 2002
4. RFC793: Transmission Control Protocol
5. RFC768: User Datagram Protocol
6. Barbara Miłoś, Tomek Miłoś, Karol Romanowski, Adam Nowaczyk: Intelligent ExaByte Storage (https://iebs.icsr.agh.edu.pl/)

## 27. Scalable Metadata Model for Large-Scale Storage Systems

Barbara Milos/Palacz (1), Tomasz Milos (1), Lukasz Dutka (2) and Jacek Kitowski (1,2)

*(1) Institute of Computer Science AGH-UST, Krakow, Poland*
*(2) ACC CYFRONET AGH, Krakow, Poland*

The need for large-scale, reliable and simply manageable storage had been growing exponentially in recent years. In order to cope with this problem, multiple storage systems have emerged, providing not only storage but also accessibility and fault-tolerance. Systems of such characteristics require great amounts of metadata, describing the user data that is being retained in the storage. Efficient organization and maintenance of the metadata is critical in terms of performance, as well as scalability. This paper introduces a novel approach to creating an extensible metadata model, which is used in a system called IEBS – Intelligent ExaByte Storage ([1], [2]).

IEBS is a highly scalable, geographically distributed system, supplying useable storage of Exabyte order of capacity [1]. The more storage the system provides, the more metadata it needs to describe it. In fact, the amount of metadata in IEBS is so vast, that the number of records of one kind is too large to be efficiently maintained in a single database table. Therefore, it is necessary to decompose the data with the use of a distributed table – a single database table is replaced with multiple partial tables, which store only a portion of the metadata. This concept is known in database terminology as *partitioning* however it has been tailored for IEBS system's specific needs.

Since the system is distributed geographically, it consists of multiple managing nodes, which share the responsibility of metadata management. To ensure scalability the metadata model has to support sharing and transferring metadata between those nodes. This can be achieved by exchanging entire partial tables, however to minimize the communication between managers it is important to have data from these tables as separable as possible.

Due to the distribution of partial tables another issue needs to be faced: how to retrieve a particular record, without knowing its location within the system? Resolution of this problem requires two things: introducing hierarchical primary keys and adding "routing" tables to the databases. The hierarchical primary keys are related to metadata separability – the shorter the common prefix of two records' primary keys, the more separated they are. Using such keys leads us to partial tables containing all records with common prefix of their primary keys. At this point routing tables are used to track the location of partial tables. A routing table, in its basic form, consists of prefixes of records' primary keys and locations of partial tables that hold them. These locations may be names of concrete partial tables (local), or addresses of nodes that manage the data (remote). Prefixes in routing tables may be of variable length what allows for aggregation – tables with common prefix kept on the same remote node may have only one entry in routing table. The entry with longest matching prefix points to the location of partial table containing requested record. This concept is similar to routing in IP networks.

The IEBS metadata model is highly scalable and allows for the system's future growth. Many parts of this solution can be adopted by other large-scale systems that require great amounts of metadata, which must be well organized and easily maintained in order for the system to function properly and efficiently.

References

1. Lukasz Dutka, Barbara Palacz, Jacek Kitowski: IEBS – Intelligent ExaByte Storage based on Grid Approach; Workshop Proceedings of the CGW '06, Krakow, October 2006.
2. Barbara Palacz, Tomasz Milos, Lukasz Dutka, Jacek Kitowski: IEBS Ticketing Protocol as Answer to Synchronization Issue; Proceedings of PPAM'07, Gdansk, September 2007, LNCS.

## 28. Comparing two Lustre Implementation Scenarios – Based on Storage Servers and Enterprise SAN Disk Arrays

Łukasz Flis, Patryk Lasoń, Marek Magryś, Marek Pogoda, Grzegorz Sułkowski, Maciej Twardy

*ACC CYFRONET AGH, Krakow, Poland*

Lustre, a distributed object-oriented file system, is one of the most popular storage solutions for HPTC clusters, as it offers high scalability, throughput and allows a large number of clients to perform parallel I/O operations without performance loss. Lustre gives the possibility to use almost any kind of storage hardware, starting from base level storage servers with SATA disk drives, up to enterprise SAN disk arrays, as back-end block devices.

This paper presents results obtained during Lustre deployment in Academic Computer Centre Cyfronet AGH. Authors have focused on two specific Lustre hardware configurations, based on SUN x4540 servers and HP EVA 8000 disk array. Performance tests that were carried out were focused on typical storage access patterns of grid jobs running on Cyfronet's computational resources. This document contains detailed

descriptions of hardware configurations that were used, Lustre parameters, test results and discussion on choosing an optimal setup that would suit best for grid clusters.

References


1. M. Pogoda, G. Sułkowski, M. Twardy, *Preparing Storage Infrastructure to Meet the Requirements of the Grid - Environment*, Proceedings of the CGW06, Kraków 2006
2. *Lustre: A scalable, high-performance file system.Whitepaper,* Cluster File System, Inc. http://www.lustre.org/docs/lustre.pdf.
3. Gary Grider. *Scalable i/o, file systems, and storage netwoks:* R&D at Los Alamos, May 2005.


## 29. Infrastructure Monitoring System for an NGI


Małgorzata Krakowian, Marcin Radecki

*ACC CYFRONET AGH, Krakow, Poland*


Current organization of operations in EGEE project is going to change in the near future, the three-level structure will be broken into two-level one consisting only of a central EGI and a bunch of autonomous NGIs, without the regional level. This transition entails need for careful adapting not only operational procedures but also tools. One of the most important tool essential in keeping Grid infrastructure stable is a monitoring system. The results of monitoring are used on all levels of operational support and to assess overall quality of the service.

With the movement to EGI/NGI model which likely happen in 2010 the monitoring system can no longer be run centrally, but the responsibility for maintain it will be handed over to NGIs. However, there is still a need to collect monitoring results in one place to facilitate overall assessment of the infrastructure as well as it is easier for development of the tools requiring such data like monthly availability/reliability report generator. It also gives the possibility to use a regional VO for monitoring instead of one artificial VO spanning over all Grid sites. Going further this path we could envision using supported VOs for monitoring rather than those dedicated only for services testing. With introducing NGI-level monitoring system there is also a room for NGI-specific requirements.

From the NGI support structures point of view it is important to ensure precise diagnostic of problems i.e. to raise an alarm against the entity which is causing problem e.g. to spot network or hardware problem first rather than a problem with the service running on it. To do that, checking and visualization of trends is needed. Here is vital to mention about implementation of testing hierarchical model where dependences between tests are set – when one test fail those dependent on it will not be even executed. For example in the first place access to hardware would be tested, and after that higher level services. NGI is also interested in total computing power provided to Grid which can be done within monitoring system as well.

In this paper we present a short overview of monitoring system used in Grid right now, analyze requirements coming from NGI on infrastructure monitoring system, then we follow up with design of a system according to the requirements and suitable in EGI/NGI model. We also elaborate changes in operational procedures that are necessary with appearance of new model.

## 30. Improvements and Measurements of their Influence on Grid Information Service Performance in EGEE

Wojciech Ziajka, Marcin Radecki
*ACC CYFRONET AGH, Krakow, Poland*

Information system in Grid contains data about all grid resources available in the infrastructure. The nature of this service is similar to Domain Name System which is also being asked frequently and by many clients, thus must be distributed but synchronized with other instances. In EGEE grid infrastructure the data about resources contains tens of megabytes and a piece of it may be required at any time by any of tens of thousands users' jobs running concurrently. Reliability of the information system is crucial for proper operation of the grid infrastructure and, in consequence, for providing scientists with robust environment for performing their work.

The information system in EGEE is called a BDII [1] (from Berkeley Database Information Index). It is based on LDAP protocol which make use of Berkeley DB transactional backend. There are several instances of the service around the infrastructure handling more than 250 production grid sites. It is not feasible to deploy as many as will service instances due to unacceptable load on site' information systems thus the most suitable way to tackle the scalability problem was to investigate improvements to the performance of the service itself.

This paper continues the work we started on the BDII service and presents propositions of improvements in information system and results of the implementation. The ideas for improving the service are not only technical, but they also deal with Grid organizational aspects (Virtual Organization-level BDII), service architecture (master-slave, hierarchic approach) which also seems to better conform to an EGI/NGI model. This document also presents framework for performing the service stress tests. This framework use grid infrastructure for making load on the information system and measuring the results. Using this framework we tried simulate real work environment of Top Level BDII as framework use queries of type asked by real tools.

References

1. J. Astalos, L. Flis, M. Radecki, W. Ziajka: Performance Improvements to BDII – Grid Information Service in EGEE; CGW'07 Workshop Proceedings, Krakow, 2007.
2. BDII service wiki page https://twiki.cern.ch/twiki//bin/view/EGEE/BDII

## 31. JXTA-based Collaboration in the SemMon Performance Monitoring System

Wlodzimierz Funika, Arkadiusz Knapik, Michał Lozinski, Pawel Chrzaszcz
*Institute of Computer Science AGH, Krakow, Poland*

When monitoring an application execution for different purposes, a need emerges for a kind of cooperation between users, especially when one of them is evaluating performance, another is debugging, the third is performing checkpoining. Even when all of them are performing the same activity on the same application, they may need co-ordinate their actions, so not to distrurb each other, but rather help. One of the most common ways the users of the monitoring system collaborate is remote conference. One user can start a monitoring session and other geographically distributed users can join it, watch and influence its progress, participate in discussions and decision-making on performance issues, leave the conference.

We are going to show an idea how to extend the SemMon monitoring system [1] by the capability of conference-like collaboration. Semmon tool is an all-in-one monitoring system for distributed applications. This system uses many structures like metrics and ontology description for performance evaluation and relevant actions triggered in the underlying physical monitoring system. The semantic knowledge used in SemMon can greatly contribute to a multi–layer and multi–source monitoring process giving the end-user an easy and efficient tool for providing assistance on monitoring and semi–automatic performance data analysis.

The development of collaboration means between end-users within an already existing application is a complex task. Our task was not only to add some new features to SemMon, but also to make some research for new capabilities and constrints, which may occur when using the JXTA technology [2] and other specific solutions. We focused on creating a multi-user remote control panel for remote conferences. The idea is quite simple. One end-user creates a conference with a given topic. Another end-user may join this conference through browsing a conference list by its topics. Any changes in the SemMon system on the conference's owner side is saved and sent to the joined listeners.

Adding the multi-user control panel to the Semmon system requires modifications in the architecture of the software at least in the GUI layer. All events from application are captured through an additional listener which saves this event to the history list of events. At the moment new end-users sign up as interested in

participation, their GUI is set to the state in which the GUI of the conference owner is in. As the conference progresses, each event occurring at the sender's side is broadcast to all listeners with updating their GUI accordingly.

The use of the JXTA technology opens new opportunities but also forces some specific methodology. The most important gain when using the JXTA technology is the transparency of the conference – users can be in different networks and behind firewalls. The most important limitation due to the JXTA technology is the need of rendezvous server for communication between different networks. For the need of broadcast we use html protocol over unicast transmission. This decision is aimed to support further safe-connection development.

The use of the command pattern solution for multi-user remote panel also creates new opportunities but also forces some specific approach. As for now we can not watch the progress of the conference after it had finished and therefore we are not able to analyse the steps that had lead us to the conclusions, and possibly finding any possible mistakes in the process. With the existing command pattern solution we can, however, easily imagine a system in which all this is possible. By having the list of all commands called out inside one conference we could browse through it and use it as a tutorial. We could even create a whole library of online tutorials. It would need some further modifications inside the existing solution. We can easily manage the people who are participating in the conference, and have control over approving new listeners, and for instance removing ones who appear to cause more troubles than help in dealing with the problem at hand. Each user can disconnect from the conference at any moment, and to proceed the monitoring of process on his/her own sparing from the state in which he/she was when leaving the conference. At the same time nothing stands on the way to start a new conference after quiting the last one – since each event is being stored, even the ones that had been sent to us from the conference's host, once we leave the conference, the state of the GUI is just as if we have made all of the proceeding steps ourselves. The main disadvantage of the command pattern solution is the need to cover all new events from new modules of the system by the listener. Any changes in the GUI layer have to include an update to the commands pattern listener. This can be laborious, but still it is much more simpler than building synchronisation between the logic layers of SemMon.

Our further efforts will be aimed at extending the functionality of the conference-like collaboration in SemMon. First of all it is the optimisation of the events that have to be sent to users joining the conference – we store the whole record of the conference, but if the user is not interested in it we could optimise the list of events which occurred, and send out only the ones that have an actual impact on the current state of the GUI (for instance if we had only opened and closed the window without making any meaningful actions, we can easily omit the events that were linked to it, since they don't bring any important data). This way we could save the amount of data that needs to be transferred over the network. Next, in order to be able to rewind the conference we would need to log the data from the metrics that had been run during the conference. Finally, we aim to enable a chat between the participants of the conference, sending the messages to specific users, and creating a general channel that allow the owner of the conference to explain the task at hand, and the participants to give their suggestions about how to solve the problem.

References


1. Piotr Godowski, Piotr Pegiel, Adaptive monitoring of distributed Java applications, M.Sc. thesis, 2007, Krakow, Poland
2. JXTA community home page: https://JXTA.dev.java.net/


# 32. Integration of the SemMon Semantic Monitoring Tool into the ProActive Platform


Wlodzimierz Funika, Mateusz Kupisz, Paweł Koperek
*Institute of Computer Science AGH, Krakow, Poland*


The importance of distributed computing using parallelism techniques continues to grow. Because of the demand for a fast development of such applications, which in many cases are very complex, the use of production-ready frameworks is essential. One of such frameworks is ProActive [1]. It enables one to develop parallel and concurrent applications and includes support for asynchronous communication, load balancing and computation tasks migration. Easy deployment on various infrastructures is provided as well by a specific layer of framework. ProActive exploits the concept of Active Objects – small, basic units of activity (each has its own thread and execution queue) which can be easily managed and distributed across the environment. Using Java both as a programming language and runtime environment, ProActive is a cross-platform solution: it can be run in any environment in which JVM works.

One of characteristics of distributed applications is the demand for continuous monitoring. Java runtime provides mature and production-ready technology, enabling observation and management of virtual machine and applications that run in it – Java Management Extensions (JMX). A graphical monitoring system bundled with ProActive (IC2D – Interactive Control and Debugging of Distribution) uses JMX to control and monitor

applications. It also allows one to obtain extended benchmarking and profiling information concerning computing, provided in the ProActive TimIt service. Such a solution, albeit very useful in monitoring complex systems, floods the user with information and hardens the process of system state analysis. Any higher level processing of such information has to be performed manually by the researcher. Knowledge about effective and authoritative methods of examination isn't shared between the system users. No collaboration tools are provided. Therefore a new solution for monitoring distributed systems is required.

The SemMon project [2] is aimed to create a tool providing semantic analysis of data coming from monitoring of a distributed system. Knowledge about effective tool usage (relevance and accuracy of metrics used during system examination) may be shared within the research team with the use of a built-in scoring system. All resources and metrics that are used are elements of a specific ontology, which provides a semantic description of gathered data. Based on such an approach to data processing, SemMon has the ability to interpret data and can trigger indispensable actions, e.g. provide some suitable notifications (alarms) to users. For communication with the resources under monitoring the JMX technology is used, which allows the tool to be used in conjunction with any external software based on Java environment, including ProActive library. To provide a better integration with ProActive, a new web service, which registers library's specific resources, has been developed. One central repository for these resources, called *registry*, is provided as well. These two components combined enable the connection with active instances of ProActive runtime. For the connection itself ClientConnector, an extended JMX connector – part of ProActive framework, is used. Once dynamically loaded on a connection event, a specialized monitoring MBean (Managed Bean) from JIMS [3] library is used to gather data from the monitored system. Extended information can be gathered for each Active Object with the use of the benchmarking services supplied within the framework. The base ontology used in SemMon has been extended to provide means of representation of ProActive specific resources for analysis and storage purposes.

Semantic-based monitoring facilities provide a controllable solution to the problem of the overwhelming amount of information gathered from systems under monitoring. SemMon's modern design, in conjunction with support for the widely used ProActive framework, is aimed to help users and administrators to cope with management and improvement of wide and complex distributed systems, e.g. grids. With such an approach common problems like network bottlenecks or overloaded machines can be instantly diagnosed. The main goal of the project is to provide an intelligent, extendable solution which will enable to discover and control performance problems in complex distributed environments.

References


1. ProActive project's web page: http://proactive.inria.fr
2. Piotr Godowski, Piotr Pegiel: Adaptive monitoring of distributed Java applications, AGH-UST, 2007
3. Krzysztof Zielinski, Marcin Jarzab, Damian Wieczorek and Kazimierz Balos, JIMS Extensions for Resource Monitoring and Management of Solaris 10, in Vassil N. Alexandrov and G. Dick van Albada and Peter M.A. Sloot and Jack Dongarra, editors, Computational Science – ICCS 2006, 6th International Conference, Reading, UK, May 28-31, 2006, Proceedings, Part IV, volume 3994 of Lecture Notes in Computer Science, pages 1039-1046, Springer, 2006.


## 33. Self-healing Oriented Technologies in Autonomous Monitoring Systems


Wlodzimierz Funika, Piotr Pegiel

*Institute of Computer Science AGH, Krakow, Poland*


Nowadays systems are becoming very complex. They are, in fact, very frequently built with many components which are working on different machines, in a “distributed environment”. It is impossible to monitor such systems manually, there are too many different indicators to check (resources states, network traffic, operated system,…). This is the reason why distributed monitoring systems were developed. They help user in managing the system – usually user is able to see all interesting data in the monitoring system presentation layer. The next stage was introduced when autonomous monitoring systems were developed. Such systems do not need user interaction to make a good decision what should be monitored in the current situation. Decision is based on the knowledge gathered from the previous monitoring results. The Monitoring system could also guide user what should be checked in the next step. To fulfill such requirements monitoring systems became ‘intelligent’. From this point it was a straight way for enabling self-healing – the decision made by the monitoring system could make monitored system to behave more stable, reliable and predictable. In this paper we are going to present a study of monitoring systems and techniques used in monitoring systems for self-healing.

Two main aspects of self-healing monitoring systems can be distinguished when the self-healing systems are concerned. The first aspect is related to the physical layer of the system (like computers, resources, network), while the second aspect is related to the logical layer (applications, operating system).

Monitoring the physical layer is usually more intuitive. We can imagine situation when the operating system can make a decision to automatically offline a faulty resource. This functionality could be even implemented on the system level – like it is in the Solaris 10 [1].

In the second approach a self-healing functionality can be injected into the monitored system. In this situation technologies like Aspect Oriented Programming can be used [2,3]. Using the AOP techniques we can cross-cut the business logic of the application to inspect its state. When an incorrect state is detected a monitoring system can perform a recovery action. There are also efforts to extend the existing monitoring systems to enable self-healing – e.g. Nagios system can be used to automate recovery after service failures [4].

We can also consider self-healing in two different contexts. The first is described above – self-healing of the monitored system. The second is also quite interesting – similarly the monitoring system can be self-healing – we can imagine that the system is composed of many autonomous agents which can make a decision to disable unstable agent.

In the paper we aim also to briefly describe some other approaches to self-healing systems. The first one is the self healing in the context of digital libraries [5]. The second one is not directly related to the computers – it is a self healing infrastructure for transferring energy [6], this can inspire some ideas related to computer science.

References


1. Predictive Self-Healing in the Solaris™ 10 Operating System – A Technical Introduction September 2004
2. Towards Self-adaptable monitoring framework for self-healing, CoreGRID TR-0150, July 3, 2008
3. R. Griffith and G. Kaiser. Adding self-healing capabilities to the common language runtime. Technical report, Columbia University, 2005.
4. Using Nagios to monitor faults in a self-healing environment, Mikko A.T. Pervilä, 2007.
5. Towards a self healing information system for digital libraries, AMBATI Vamshi, REDDY Raj, Aug 2005
6. Toward self-healing energy infrastructure systems, IEEE Computer Applications in Power, ISSN 0895-0156/01/$10.00©2001 IEEE


## 34. Acceptance of Desktop Grid Computing amongst SME's and the General Public


Ad Emmen, Leslie Versweyveld

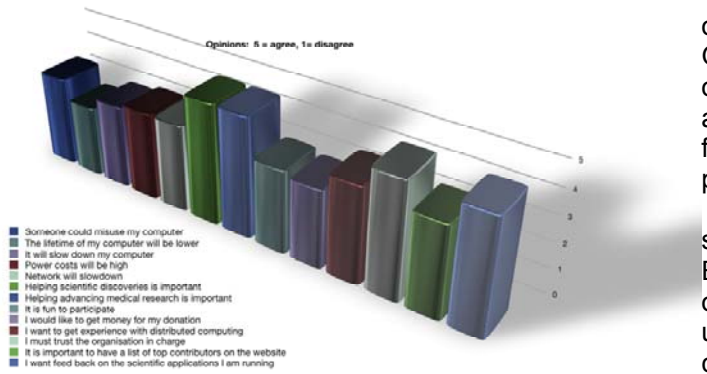
*Stichting AlmereGrid, Almere, The Netherlands*


Desktop Grids are widely deployed as volunteer computing platforms. The idea is that people donate unused computing cycles to science. Over the world several millions of computers are active in a volunteer computing Grid. Examples are AlmereGrid [1], SZTAKI Desktop Grid [2,3], and World Community Grid [4]. Until now, most of the persons donating computing time, belong to the early adopters: they are computer savvy. The early majority still seems a bit reluctant to donate computing time. To verify this we conducted an in-depth survey amongst SME's and private persons all around Europe, as part of the EDGeS project [5]. Here we present results of this survey.

The survey was conducted using standard questionnaire-form techniques. We designed a form that was pre-tested by a number of persons first to see whether the questions were clear. Then we asked a number of persons, both private persons and representatives of SME'S (Small and Medium Sized Enterprizes) to fillout the form. The form was presented to them by a partner of EDGeS that could also provide explanation of the questions. We opted for this method to get a reliable sample – which one cannot get from widely used web surveys – and because the interviewers could provide explanations of the concepts in Desktop Grid computing. The sample size was selected so the results provide better than 20% accuracy. In total we got 51 filled out questionnaires.

The main results are: There is interest in Desktop Grid computing in Europe: only 18% answered they are not willing to use it . Amongst companies and organisations there is also considerable interest in looking into setting up an own Grid: 58%. However, that people are willing to change their current practice and say that they want to participate in Grid efforts does not mean that they are actually going to do that. Communication is a means to try to convince them to do so.

People want to donate computing time to science (>3 on a scale of 5) but not to defense applications (1,9 on a scale of 5). Detailed results will be presented in a table/diagram.




| | Someone could misuse my computer | The lifetime of my computer will be lower | It will slow down my computer | Power costs will be high | Network will slowdown | Helping scientific discoveries is important | Helping advancing medical research is important | It is fun to participate | I would like to get money for my donation | I want to get experience with distributed computing | I must trust the organisation in charge | It is important to have a list of top contributors on the website | I want feed back on the scientific applications I am running |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | 3,6 | 2,8 | 3,3 | 3,5 | 3,5 | 4,8 | 4,8 | 3,3 | 3,1 | 3,8 | 4,6 | 3,6 | 4,5 |
| Standard deviation | 1,5 | 1,6 | 1,5 | 1,5 | 1,6 | 0,5 | 0,5 | 1,3 | 1,4 | 1,4 | 0,9 | 1,5 | 0,9 |

We also wanted to get a feeling of the opinions over and feelings about Desktop Grid computing. Hence we asked a number of questions to which the respondents could agree/disagree. See the diagram on the left for part of the answers. More will be presented in the full paper.

The results of the survey will be used to shape the communication strategy of EDGeS. It can also be used by other Grid operators to adapt their Grid service to the users. Future work we are looking at is a comparison with other surveys such as the BOINC survey, and repeating the survey after one or two years to compare with the current results.

**Acknowledgements**. This work was supported by EU project EDGeS, that is supported by a Grant from the European Commission's FP7 IST Capacities programme under grant agreement RI-211727. EDGeS partners contributed to executing this survey.

References

1. AlmereGrid, http://AlmereGrid.nl (visited September 2008)
2. SZTAKI Desktop Grid: Building a scalable, secure platform for Desktop Grid Computing, Attila Csaba Marosi, Gábor Gombás, Zoltán Balaton, Péter Kacsuk,Tamás Kiss, CoreGRID Technical Report, Number TR-0100, August 28, 2007.
3. SZTAKI Desktop Grid – a Hierarchical Desktop Grid System, P. Kacsuk, A. Marosi, J. Kovács, Z. Balaton, G. Gombás, G. Vida, and Á. Kornafeld, CGW'06 Proceedings, Editors: M. Bubak, M. Turala, K. Wiatr, p42, 2006.
4. World Community Grid, http://www.worldcommunitygrid.org/ [visited September 2008]
5. EDGeS: bridging Desktop and Service Grids, Miguel Cardenas-Montes, Ad Emmen, Attila Csaba Marosi, Filipe Araujo, Gabor Gombas, Gabor Terstyanszky, Gilles Fedak, Ian Kelley, Ian Taylor, Oleg Lodygensky, Peter Kacsuk, Robert Lovas, Tamas Kiss, Zoltan Balaton, and Zoltan Farkas, IberGrid 2008.

## 35. Belorussian National Grid-Initiative

Sergey Ablameyko, Uladimir Anishchanka, Anatoliy Krishtofik, Oleg Tchij
*UIIP NAS of Belarus, Minsk, Republic of Belarus*

National grid-initiative is set of interconnected actions on learning and deployment technologies of grid in science, education, social sphere and manufacture. Main purpose of grid-initiative is to create computing resources in global grid environment and supporting wide range user's application. Initiator of grid-initiative is United Institute of Informatics Problems (UIIP) NAS of Belarus.

Main goals:

- learning and assimilation of foreign country experience in creating and using grid-infrastructure;
- integrating existing resources with international grid projects;
- creating national grid-infrastructure for science research;
- development national telecommunication infrastructure;
- new technologies staff training;
- using grid technologies in various areas;

Reaching these goals open up possibilities for resource-intensive applications such as engineering modeling, device inventing, nuclear physics, pharmacological research and molecular dynamics, meteorology, global climate fluctuation prediction, space and engineering research. Implementation of national grid-initiative is accomplished in following directions:

- fulfillment of "SKIF-GRID" program;
- participation in European Commission programs;
- performance of tasks within bounds of international programs;
- grid-technologies user involvement;
- establishment of resource centers and integrating them with grid-infrastructure;

"SKIF-GRID" program stipulates creating of persistent and reliable site of national grid net. This site consists of two segments. First segment is intended for gLite middleware for integrating resources with international grid-infrastructures. Second one is intended for UniCore middleware for national grid net implementation. "SKIF-GRID" program also stipulate integration various grid-segments with pan-European grid nets via BalticGrid. Integration with elements of grid-infrastructure of Russian Federation is planned. National research network BASNET has become a full participant of TERENA. This will speed up integrating Belarusian research and education networks with pan-European research high-bandwidth telecommunication infrastructure, and will let BASNET become an associate member of GEANT3 project. Development of infrastructure of application design is started in following area: high-energy physics, mechanical engineering, bioinformatics, and material science. All these actions permit development and deployment of transnational application as international scientific collaboration.

Basic computing resources of national grid-infrastructure are resources of supercomputing center of UIIP NAS of Belarus. Resources of main technical universities (BNTU, BSUIR, JIPNR, GRSU) and some others organization will be used additionally.

Integration works with Ukrainian academic grid within international programs are executed.

Main activity within 7th framework programme BalticGrid-II is aimed at integration resources of UIIP NAS of Belarus with pan-European grid-infrastructure, creating certificate authorities, user involvement and personnel education.

Implementation of these steps allows create and efficiently use national grid infrastructure.

## 36. Design of Grid Operations Database for NGI/EGI Model

Lukasz Flis, Marcin Radecki
*ACC Cyfronet AGH, Kraków, Poland*

Largest production Grid infrastructure in Europe is developing towards more decentralized and autonomous. Transition that is to be made with EGI/NGI model will cover not only financial aspect but will also affect structural and operational organization of the Grid.

In the new model Grid will be composed of National Grid Initiatives (NGIs) being autonomous systems coordinated at a country level and one central body called EGI. It is likely that we will have a flat hierarchy with all NGIs at equal level rather than an intermediate, regional level which is present now in EGEE project. The EGI body itself is intended as a "glue" between NGIs ensuring proper interoperability and coherence for the benefit of user communities.

One of the important aspects of compatibility interface between NGIs is access to the infrastructure related information. EGEE project have developed a solution called GOCDB which is a central database used to collect and share the information required for management and operations. NGI as antonomous systems will develop or adapt their own infrastructure information management systems for that purpose. Although top-down (central) approach (like GOCDB) was sufficient for EGEE environment it is not suitable for EGI which requires rather bottom-up solution in order to avoid data redundancy and assure proper data flow direction.

Presented work discusses requirements for infrastructure information management system as seen from NGI point of view. Analysis and comparison of EGI and EGEE project characterisics are presented and new concept of information management system architecture (including NGI and EGI levels) is proposed. Document presents guidelines and concepts of two systems NGI infrastructure management database and EGI (inter-NGI) information management database with their common interface and data workflows.

References

1. The European Grid Initiative – www.eu-egi.org

## 37. Evaluation of Grid eLearning: eLGrid User Evaluation Experiment Results

Kathryn Cassidy (1), John Walsh (1), Brian Coghlan (1), Declan Dagger (2)
*(1) Computer Architecture and Grid Research Group, Trinity College Dublin, Ireland*
*(2) Knowledge and Data Engineering Group, Trinity College Dublin, Ireland*

The Computer Architecture and Grid Research Group at Trinity College Dublin has developed an adaptive Grid eLearning system, the eLGrid [1]. This system offers personalised courses to learners and integrates with a virtualised Grid training infrastructure (t-Infrastructure).

The eLGrid uses the APeLS Adaptive eLearning Service [2] to provide personalised eLearning to Grid users. This allows us to take advantage of recent developments in personalised and adaptive eLearning and enables courses to be developed and presented using sound pedagogical principles.

While other Grid eLearning systems do exist, only very limited evaluation appears to have been done to determine their effectiveness. Many face-to-face Grid courses rely on reactionnaires [2], which can only measure user satisfaction, for evaluation. Many Grid eLearning based systems include no form of user evaluation.

We argue that in order to demonstrate the effectiveness of the eLearning system a range of factors must be considered ranging from user satisfaction to measuring the actual learning achieved. We present the results of two experiments to evaluate the eLGrid eLearning system. These results show that on the whole, the system is an effective mechanism for teaching new users about the Grid.

References


1. Kathryn Cassidy, Jason McCandless, Stephen Childs, John Walsh, Brian Coghlan, Declan Dagger. Combining a virtual Grid testbed and Grid eLearning courseware. In Proc. Cracow Grid Workshop 2006 (CGW06), Academic Computer Centre CYFRONET AGH, Cracow, Poland, October 2006.
2. Owen Conlan. The Multi-Model, Metadata driven approach to Personalised eLearning Services; PhD Thesis, Trinity College Dublin, 2004.
3. Leslie Rae: Assessing the Value of your Training. Gower, Hants, England, 2002.


## 38. Setting up SKIF-UNICORE Experimental Grid Section


Sergey Ablameyko, Uladzimir Anishchanka, Anatoliy Krishtofik, Oleg Tchij
*UIIP NAS of Belarus, Minsk, Republic of Belarus*


Belarusian experimental grid section SKIF-UNICORE is being created on the basis of computing and data storage resources of Republican Supercomputer Multi-Access Center, additional resources of the United Institute of Informatics Problems of the National Academy of Sciences of Belarus (UIIP NASB) and resources of all interested national parties. The main directivity of the national experimental grid section is grid application development and deployment for scientific, social and industrial spheres.

UNICORE middleware was chosen over other systems (gLite, Globus, MPICH-G2, X-com и Condor-G2) because of its unique suitability for Belarus specific needs. UNICORE's best features are:

- open source nature,
- cross-platformity,
- continuous development,
- easy installation, development and operating,
- MPI support,
- integrability with national security standards,
- minor functionality in comparison with gLite however allows easier modification for specific needs.

National experimental grid section will provide:

- computing resources (SKIF cluster systems, servers and workstations, grid applications);
- file memory resources (available data storage systems of all kinds);
- network resources (telecommunication infrastructure provided by Belarusian Research and Education Network BASNET and by other interested parties within Belarus);
- grid applications.

Development of UNICORE grid infrastructure is carried out in two ways synchronously:

- development and establishment of national grid infrastructure;
- development and establishment of Belarusian-Russian grid segments.

Functional structure of national experimental grid section answers the purpose of its development work and includes:

- the UIIP NASB's grid segment based on computing and data storage of Republican Supercomputer Multi-Access Center;
- regional distributed grid segments SKIF-UNICORE;
- engineering grid segments of industrial organizations.

It is also planned to draw in computing resources of institutes of higher education; colleges or universities.

National experimental grid section will address specific needs of new scientific communities such as bioinformatics, nano-science, material science, microelectronics, mechanical engineering, machine tool building industry, medicine and genetics.

Main line of the establishment of SKIF-UNICORE experimental grid section is defined by the UIIP NASB that runs following tasks:

- general and operative management and coordination;
- development of new grid technologies and applications;

- establishing and running grid central services;
- establishing and running grid security services;
- establishing and supporting central grid software repository;
- establishing and supporting national grid-related website and running grid information dissemination activities;
- training grid users and professionals by organizing seminars, conferences, season schools, and workshops;
- involvement and encouragement of users to take part in grid projects.

## 39. Running MapReduce Type Jobs in Grid Infrastructure

Marek Ciglan, Martin Šeleng, Marian Babík, Michal Laclavík, Ladislav Hluchý
*Institute of Informatics, Slovak Academy of Sciences*

This paper presents our approach to enable execution of MapReduce [1] type jobs, using Hadoop framework [2], on commodity clusters in grid infrastructure. Nowadays production grid infrastructures, such as EGEE [3], are excellent for high-throughput computing, executing large number of sequential, independent compute tasks. Other approaches to distributed and parallel computing are not widely supported. For example, gLite, the grid middleware powering EGEE infrastructure, supports only one implementation of MPI [4] – MPICH ; moreover the MPICH support on individual sites in grid is sparse (e.g. only 31% of EGEE supporting VOCE VO provide environment for executing MPICH jobs). In this paper we describe our work on enabling running of MapReduce jobs in the gLite powered grid infrastructure.

MapReduce is a programming model suitable for distributed computation for processing of large data sets. MapReduce programs include definition of map and reduce functions, where map function process input set of key-value pairs and outputs intermediate result containing transformed key-value pairs. The reduce function merges the intermediate results of map functions. MapReduce programs can be automatically parallelized, executing map functions on separate worker nodes.

Hadoop is an open-source implementation of MapReduce programming model, written in JAVA programming language. It is designed to run in conjunction with distributed file system (HDFS)

and takes advantage of data location in HDFS to schedule computation of map tasks on resources containing or close to required data.

There already exist an effort to integrate Hadoop framework with commodity clusters manage by batch systems. The implementation is called Hadoop-On-Demand (HOD) [5] and it allows to submit Hadoop jobs to compute cluster managed by Torque batch system. Even though Torque is batch system of preference for many sites in nowadays grids, direct integration of HOD with gLite job submission mechanism is not straitforward.

The goal of our work is to enable submission of Hadoop jobs via standard gLite submission mechanism using JDL for job description and gLite client toolkit for running the job in the grid. When JDL file is processed at the resource broker, the sites supporting Hadoop jobs must be identifiable. When the job is scheduled, it is required to allocate multiple nodes at the execution site. When the job execution starts (with multiple nodes allocated) it is necessary to configure environment for Hadoop on all nodes involved in the job execution and initialize Hadoop prior to the start of Hadoop processing itself. In addition, data preparation (files insertion into Hadoop HDFS) must take place. In the paper, we describe each step in detail and provide examples. We also provide brief overview of Hadoop architecture to explain implementation decisions of our approach.

A great help for our implementation was the work done in Int.Eu.Grid (I2G) [6] project on MPI jobs support. MPI-Start [4] mechanism developed in I2G allows usage of different MPI implementations, support for multiple schedulers and provide the possibility to run user-defined scripts (hooks) before and after the execution of the binary. The latter allows us to set up the environment for Hadoop execution on the allocated nodes in cluster.

We conclude the paper by discussing other possible approach to the integration of Hadoop and grid middleware and we highlight the directions for the future work.

References

1. Dean J., Ghemawat S.: MapReduce: Simplified Data Processing on Large Clusters, Communications of the ACM, Volume 51, Issue 1 (January 2008), Pages 107-113, ISSN:0001-0782
2. Lucene-hadoop Wiki, HadoopMapReduce, http://wiki.apache.org/lucene-hadoop/HadoopMapReduce (2008)
3. EGEE project, http://public.eu-egee.org/
4. Dichev K., Keller R., Stork S., Fernandez E.: MPI on the Grid, JIn Computing and informatics. ISSN 0232-0274, 2008, vol. 27, no. 2.
5. Hadoop On Demand: http://hadoop.apache.org/core/docs/r0.17.2/hod.html

6. Marco, J. et al.: The interactive European Grid: Project objectives and achievements. In Computing and informatics. ISSN 0232-0274, 2008, vol. 27, no. 2.

# 40. Establishing SKIF-gLite Grid Infrastructure

Sergey Ablameyko, Uladzimir Anishchanka, Anatoliy Krishtofik, Oleg Tchij
*UIIP NAS of Belarus, Minsk, Republic of Belarus*

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (UIIP NASB) has based its international grid cooperation on development and deployment of grid segments running gLite middleware. In order to provide an effective start the UIIP NASB has started establishing of experimental grid section composed of four separate grid segments: BY-UIIP, BY-JIPNR, BY-BSUIR, BY-BNTU (named after hosting institutions).

On the basis of the experimental grid section the UIIP NASB has planned to deploy within next two years several grid applications in the field of bioinformatics, high energy physics, material science, engineering sciences, etc.

In order to ensure that users of grid applications have a secure and convenient access to grid infrastructure the UIIP NASB will create three virtual organizations and will provide them with all proper support from Belarusian Grid Certification Authority that has been already created.

The UIIP NASB is going to make available for international use these kinds of resources:

- computing resources (SKIF cluster systems, servers and workstations, grid applications);
- file memory resources (available data storage systems of all kinds);
- network resources (telecommunication infrastructure provided by Belarusian Research and Education Network BASNET);
- grid applications.

The UIIP NASB also hosts several grid central services and provide following functionality:

- establishment and registration of grid sites, getting appropriate certificates and connecting SKIF computing resources and data storage systems into international grid networks;
- establishing and running grid certificate authority;
- further distribution of grid infrastructure onto resources of other interested organizations and communities;
- issue of certificates for users and resources of Belarusian gLite grid infrastructure;
- general and operative management and coordination of Belarusian grid participants;
- establishment and support of grid software repository;
- establishment and support of grid.by website and running various grid information dissemination activities;
- participation in international grid projects;
- training grid users and professionals by organizing seminars, conferences, season schools, and workshops;
- involvement and encouragement of users to take part in grid projects.

The establishment of experimental grid section is the first step of grid technology in the Republic of Belarus. It is expected that this first step will facilitate mastering of new technologies and will accelerate propagation of the technologies all around Belarus. It is already obvious that creation of technological basis for development of new and science intensive competitive production can easily be intensified by power of grid computing.