5th International Workshop on Grid Computing for Complex Problems

GCCP 2009
BOOK OF ABSTRACTS

October 26 – 28, 2009
Bratislava, Slovakia

The workshop is organized by

Institute of Informatics, Slovak Academy of Sciences
Faculty of Electrical Engineering and Informatics, Technical University of Košice
Masaryk University, Brno, Czech Republic

The workshop is supported by

EU FP7 RI project: Enabling Grids for E-sciencE III (2008-2010) FP7-222667

Program Committee

Ladislav Hluchý	IISAS - Institute of Informatics, Slovak Academy of Sciences
Ján Astaloš	IISAS - Institute of Informatics, Slovak Academy of Sciences
Jozef Černák	Faculty of Science, P. J. Šafárik University, Košice
Miroslav Dobrucký	IISAS - Institute of Informatics, Slovak Academy of Sciences
Ladislav Hudec	FIIT Slovak University of Technology in Bratislava
Jacek Kitowski	Cyfronet Cracow Poland
Jan Kmuníček	CESNET Praha, Masaryk University, Brno, Czech Republic
Ján Kollár	Faculty of Electrical Engineering and Informatics, TU Košice
Tibor Kožár	Institute of Experimental Physics SAS, Košice
Robert Lovas	MTA SZTAKI Budapest Hungary
Karol Matiaško	Management Science & Informatics, University of Žilina
Norbert Meyer	PSNC Poznan Poland
Ľudovít Molnár	FIIT Slovak University of Technology in Bratislava
Luboš Neslušan	Astronomical Institute SAS, Tatranská Lomnica
Ján Paralič	Faculty of Electrical Engineering and Informatics, TU Košice
Monique Petitdidier	CNRS - Centre National de la Recherche Scient., Paris, France
Ján Sarnovský	Faculty of Electrical Engineering and Informatics, TU Košice
Branislav Sitár	FMFI UK Bratislava
Jarmila Škrinárová	UMB Banská Bystrica
Milan Šujanský	Faculty of Electrical Engineering and Informatics, TU Košice
Claudio Vuerli	INAF Trieste Italy
Ivan Zahradník	Institute of Molecular Physiology and Genetics SAS, Bratislava
Peter Závodný	University of Economics in Bratislava
Viet Tran	IISAS - Institute of Informatics, Slovak Academy of Sciences

Organizing Committee

Ladislav Hluchý
Miroslav Dobrucký
Peter Kurdel
Jolana Sebestyénová
Oľga Schusterová
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
E-mail: {hluchy.ui, dobrucky.ui, peter.kurdel, sebestyenova, sekr.ui}@savba.sk

Preface

Welcome to the 5th International Workshop on Grid Computing for Complex Problems GCCP 2009. The workshop is a three-day combined event for grid users: workshop with invited lectures, plenary discussions, accompanied by course for users of EGEE Grid sites, which is in the scope of EGEE III project – Enabling Grids for E-science 2008-2010, FP7-222667.

The topics of the workshop are:
- Distributed Computing and Large Scale Applications
- Computational Chemistry & Material Science
- Grid and Service-oriented Computing
- Use of Knowledge and Semantics in Distributed Computing
- Astronomy & Astrophysics and High energy Physics
- Environmental applications and Distributed Computing
- Grid Tutorials.

The next goal of the workshop is an associate action to create national Grid initiative "Sprístupnenie Gridu pre elektronickú vedu na Slovensku" (Making the Grid accessible for electronic science in Slovakia) which will help to improve the e-Science in Slovakia through the creation of virtual organizations for individual science branches. The associate action aims to join Grid specialists with complex application users, to provide a medium for the exchange of ideas between theoreticians and practitioners to address the important issues in computational performance and computational intelligence towards Grid computing.

The workshop on Grid Computing for Complex Problems GCCP 2009 has attracted 28 paper contributions and active participations from Czech Republic, Germany, Italy, Ukraine and Slovakia. This book is a collection of abstracts of papers from International Workshop on Grid Computing for Complex Problems – GCCP 2009. Workshop's papers will be published after the workshop as edited proceeding.

Many people have assisted in the success of this workshop. I would like to thank all the members of the Program and Organizing Committees, the workshop Secretariat for their work and assistance of the workshop. I would like to express my gratitude to all authors for contributing their research papers as well as for their participation in the workshop that made our cooperation more fruitful and successful.

Ladislav Hluchý
October 2009
Bratislava, Slovakia

Table of Contents

Invited lectures

Vendor session

Section 1
Use of Knowledge and Semantics in Distributed Computing

Section 2
Computational Chemistry & Material Science

Section 3
Distributed Computing and Large Scale Applications

Section 4
Grid and Service-oriented Computing

Section 5
Astronomy & Astrophysics and High energy Physics

Section 6
Environmental applications and Distributed Computing

Tutorials

Invited lectures

Grid computing for Astronomy & Astrophysics

Claudio Vuerli

Istituto Nazionale di Astrofisica (INAF), Osservatorio Astronomico di Trieste
e Unità Sistemi Informativi, Via Tiepolo 11, Trieste - ITALY
vuerli@oats.inaf.it

Abstract. In the old days of photographic plates, producing 20 terabytes might take 60 years of observing time, and another ten years of digitization. Current digital sky surveys can produce 20 terabytes in a year. The newest generation of sky surveys will produce 20 terabytes every night for a decade. As data volumes increase dramatically, the importance of computation increases.

.

e-Infrastructures for Science and Industry - Clusters, Grids, and Clouds (the DEISA project)

Wolfgang Gentzsch

EU Project DEISA, http://www.deisa.eu/, DEISApedia:
www.deisa.eu/references/deisapedia, Board of Directors of OGF, http://www.ogf.org/
gentzsch@rzg.mpg.de

Abstract. While Grids allow for direct access to and sharing of distributed resources, for good reasons, clouds are datacenters on the Internet which provide IT services as a utility, on a pay-per-use basis. while Grids stand out because of their flexible, dynamic, feature-rich resources and thus are complex by their very nature, Clouds provide an entirely new business model with its own set of value propositions for (currently mainly) enterprise computing environments, including application scalability, improved economies of scale, reduced costs, resource efficiencies, resource elasticity, faster deployment times, value-based pricing model, disaster recovery and an on-demand infrastructure enabling the truly dynamic data center.

Cloud applications will likely follow similar strategies as grid-enabling ones. Just as challenging, though, are the cultural, mental, legal, and political aspects of clouds. Building trust and reputation among the users and the providers will help in some simple scenarios. But it is still a challenge to imagine users easily entrusting their corporate assets and sensitive data to cloud service providers.

Another question which we will try to answer is how suitable the Cloud services model will be for the capability computing demands of the HPC community, in research and industry. Here, we will look at DEISA, the Distributed European Infrastructure for Supercomputing Applications, to analyze the resource requirements of HPC applications, and check their suitability for the Cloud. We will show how DEISA will have a good chance to be sustainable in the long term, as an e-infrastructure for the computational scientist. And then, we might end up with a DEISA Cloud which will become an external (or public) HPC node within your grid application workflow.

Thus, the aim of the talk will be to elaborate on the main differences between HPC centers, grids and clouds, analyze sustainability with the aid of the DEISA experience, and provide an HPC application check list for Clouds.

H_2O-H_2 interaction: An inter-play of supercomputers and grids

Jozef Noga

Comenius University, Faculty of Natural Sciences, Mlynská dolina, Bratislava, Slovakia
Jozef.Noga@savba.sk

Abstract. The ultimate goal of quantum chemistry is an a priori prediction. Despite indubitable achievements, precise predictions of molecular energies and properties still represent a challenge to method developers. This lecture is an attempt to introduce the world of highly accurate ab initio calculations of small molecules in the perspective of the author's contribution to it and from the perspective of the programmer. A link to astrophysical applications is outlined via an example of the full nine-dimensional potential energy surface of H2O-H2 interaction, which resulted from a combination of about 1000 intensive calculations using supercomputers with almost half a million less intensive calculations spread over within a grid environment.

.

EGI: The Present and the Future of the Pan-European Grid Infrastructure

Ladislav Hluchý

Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
Ladislav.Hluchy@savba.sk

Abstract. Europe has invested heavily in e-science programs over the past years both at the National and the European levels with impressive results. Grid technology is recognized as a fundamental component for e-infrastructures. Many countries have launched National Grid Initiatives (NGI) to establish National grid infrastructures. Driven by the needs and requirements of European research community, the EGI Design Study represents a project for the conceptual setup and operation of a new organizational model of a sustainable pan-European grid infrastructure.

Formal and Heuristic Techniques for Programming Parallel Heterogeneous Platforms

Anatoliy Doroshenko, Mykola Kotyuk, Sergiy Nikolayev, Kostyantyn Zhereb, Olena Yatsenko

Institute of Software Systems of National Academy of Sciences of Ukraine,
Glushkov prosp. 40, 03187 Kyiv, Ukraine
anatoliy.doroshenko@gmail.com, aiyat@i.com.ua,
zhereb@gmail.com

Abstract. The approach based on joint usage of formal algebra-algorithmic and heuristic techniques for specification and development of parallel programs is considered. Algebraic part of the methodology provides the formalized process of parallel program design through high-level algebraic-algorithmic specifications and automated transformations in rewriting style up to program code in a standard programming language. Heuristic part of the system stands for dynamical adjustment of program code for a target platform and its optimization using self-learning code generation and heuristics technologies. An illustrative program example and its evaluation on heterogeneous parallel platform is given to justify the approach advocated.

Keywords: algebras of algorithms, parallel computation, code generation, heuristics, heterogeneous platforms.

Implementation and Testing of Multiple Walkers Approach Based Free Energy Calculations in the Grid Environment

Zora Střelcová [1], Petr Kulhánek [1], Jan Kmuníček [2], Jaroslav Koča [1] and Luděk Matyska [2]

1)National Centre for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic
2)CESNET, z. s. p. o., Zikova 4, CZ-16000 Praha 6, Czech Republic

Keywords: VOCE, Charon Extension Layer, free energy, Multiple Walkers Approach, Adaptive Biasing Force method, ABF, supramolecular chemistry, pseudorotaxanes, cucurbit[n]urils

Abstract

Our long standing effort is focused on the application and implementation of novel methods for the free energy calculations. The free energy is an important thermodynamical quality used in biochemical and chemical disciplines. It forms an essential connection between theoretical models and experimentally observed data. For example, it can be used in the prediction of binding affinities of various compounds to biological molecules. Another application is the validation of proposed reaction mechanisms. Unfortunately, its calculation using computer simulations faces several problems. The most serious problem is that very long simulations are required to obtain converged and reliable results. To circumvent this so called sampling problem, a lot of strategies were developed and suggested in the past. One of them is Multiple Walkers Approach[1] (MWA) connected with Adaptive Biasing Force (ABF) method[2,3,4,5].

The Adaptive Biasing Force method calculates the free energy along a prescribed reaction coordinate (order parameter). Thus it is especially suitable in the study of reaction mechanisms. The method works as follows. The raw estimate of free energy along the reaction coordinate ξ is calculated and applied back to the system. As a result, the system moves freely along the reaction coordinate and the free energy estimate converge to the final free energy. To reach reasonable convergence, very long molecular dynamics simulations are required, especially if the multidimensional reaction coordinates are used.

The Multiple Walkers Approach (MWA) is a simple method accelerating the free energy calculations. It is based on the exchange of the reconstructed free energy potential among nearly independent molecular dynamics simulations (walkers) biased by ABF method. As the walkers are independent the resulting free energy is reconstructed almost N times

fasters than without MWA (where *N* is the number of walkers). Moreover, the free energy exchange can be irregular and with longer period than an integration time step used in molecular dynamics simulations. Such small requirements open up a straightforward way for the utilization of MWA method in the grid environment. The MWA method was implemented as a weakly bound client/server, where communication between ABF clients (walkers) and MWA server is over simple TCP/IP connection. In current implementation, the MWA server is run outside of the grid as it is usually long time job. Moreover, it collects valuable data and a single failure would be devastating for an entire calculation.

Since recent grid setups support MPI jobs, the individual walkers, which are molecular dynamics simulations, were run in parallel to further speed up the whole calculation. According to our results, this multilevel parallelization scheme significantly decreases the time needed for obtaining converged free energy results.

The implemented methods were applied in the study of the cucurbit[n]uril molecules (CBn) and its complexes. The CBn are wheel like molecules, which can be threaded over molecules with string shape (guests). If the guest molecule is long enough, the CBn can shuttle over it. Moreover, this molecular shuttle can be controlled by various physico-chemical conditions. The shuttle process was quantified by the calculation of the free energy profile as a function of position of CBn along a guest.

In conclusion, an advanced way how to perform large-scale free energy calculations within worldwide EGEE/EGI environment is presented. Our extension to the approach already implemented and fully functional in a local cluster environment allows us to adapt it for routine utilization within worldwide Grid subsequently boosting the research into new quality and quantity level through obtaining deeper insight into interactions taking place in supramolecular and/or biomolecular structures.

References

1. Raiteri, P. et al. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *JOURNAL OF PHYSICAL CHEMISTRY B* **110**, 3533-3539(2006).
2. Darve, E., Rodriguez-Gomez, D. & Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **128**, 144120-13(2008).
3. Rodriguez-Gomez, D., Darve, E. & Pohorille, A. Assessing the efficiency of free energy calculation methods. *JOURNAL OF CHEMICAL PHYSICS* **120**, 3563–3578(2004).
4. Darve, E., Wilson, M. & Pohorille, A. Calculating free energies using a scaled-force molecular dynamics algorithm. *MOLECULAR SIMULATION* **28**, 113-144(2002).
5. Darve, E. & Pohorille, A. Calculating free energies using average force. *THE JOURNAL OF CHEMICAL PHYSICS* **115**, 9169–9183(2001).

The dynamics of outer trans-Neptunian objects from its simulation for 2 Gyr.

Marián Jakubík[1], Luboš Neslušan[1], and Giuseppe Leto[2]

[1] Astronomical Institute, Slovak Academy of Sciences, 05960 Tatranská Lomnica, Slovakia,
ne@ta3.sk; mjakubik@ta3.sk

[2] Catania Astrophysical Observatory, Via Santa Sofia 78, I-95123 Catania, Italy,
gle@oact.inaf.it

Abstract. An intensive study of trans-Neptunian region started quite recently, in the middle of 19 nineties. It is a very young discipline of astronomical research. To contribute to our understanding of the origin and evolution of small bodies situated beyond the orbit of Neptune, we perform the simulation of the dynamical evolution of 14 799 test particles, representing the once-existing proto-planetary disc, for 2 Gyr. The particles surviving beyond Neptune, but inside the inner border of the comet Oort cloud, can be identified to the trans-Neptunian bodies and their dynamical properties can be analysed.

To perform the simulation, we use the GRID computational system, Virtual Organisation for Central Europe (VOCE), TriGrid, and Consorcio COMETA. In more detail, we use 240 CPUs in the first stage and 200 CPUs in the second stage of our calculations.

The formation and acquired structure of outer and inner parts of the Oort cloud, as yielded from our simulation, was presented in our previous papers. In this contribution, we use our output data to present the predicted structure of the "near" TN populations: classical Edgeworth-Kuiper belt, its resonant component, as well as scattered disc. The model of non-migrating planets does not seem to be sufficient to explain the observed structure of the TN populations. In the region of heliocentric distances 34−35.5 AU, 36−40 AU, and beyond 42.3 AU, more than 90% of TPs remain on their almost initial, dynamically very cold orbits ($e < 0.05$; $i < 0.05$ rad). We refer to this sub-population as "very cold Edgeworth-Kuiper belt" (VCEKB). The existence of the VCEKB in our model is the largest discrepancy between this model and observed reality. Our model qualitatively provides all observed mean-motion resonances with Neptune. However, individual abundances of TPs in the resonances and distribution of orbital characteristics differ from their observed counterparts. The difference does not seem to be only a consequence of the observational selection.

With respect to the observational selection effects biasing the observed structure of the scattered disc, a relatively good agreement appears between the structure of the disc in our model and observed disc. The model well describes the characteristic behaviour of the distribution of disc's eccentricity. As well, it provides the distribution of inclination within the observed range of values.

The number density of TPs in the VCEKB is about an order of magnitude higher than the number density of other TN populations. This fact in combination of non-existence of real VCEKB evokes an idea of disappearance of VCEKB via collisional grinding, in the first stage of its evolution. Actually, if we were able to explain the VCEKB disappearance, then our model would be in a much better agreement with the observed TN structure. The classical Edgeworth-Kuiper belt would have a relatively sharp outer edge at about 50 AU, as observed. Unfortunately, our attempts to set the free parameters in a collisional-grinding model has not been successful. The VCEKB either does not completely disappear or the structure of other TN populations is changed improperly. This negative result thus supports the concept of the formation of TN populations within the scenario of migrating giant planets.
Some animations of TP positions at the end of the simulation (2 Gyr) as well as the evolution of astrodymically interesting distributions and dependencies of the orbital elements are also presented.

Vendor session

Pohľad Hewlett-Packard na HPC - High-performance computing

Juraj Laifr

Hewlett-Packard Slovakia, s.r.o., Galvaniho 7, 820 02 Bratislava 22, Slovakia
Juraj.Laifr@hp.com

Abstract. Ako spoločnosť HP vníma svet HPC, ako pomáha svojim zákazníkom v tejto oblasti s ponukou komplexných služieb UCP - Unified Cluster Portfolio. Silnou stránkou tohto riešenia je rýchla integrácia nových technológií spolu s dramatickým zvýšením výpočtového výkonu, stabilitou a robustnosťou. Základná idea tohto riešenia spočíva v prepojovaní klastrov a zjednodušeného zdieľania distribuovaných zdrojov. Novinky v oblasti škálovateľného úložiska dát ako ExDS Exadata Storage a SFS - Scalable File Share.

HPC architektúra od spoločnosti Sun Microsystems

Dalibor Kubaček

Sun Microsystems Slovakia s.r.o., Bratislava, Slovakia
Dalibor.Kubacek@Sun.COM

IBM technológie pre GRID riešenia
IBM Technologies for GRID solutions

Marian Kovačik

IBM Slovakia, s.r.o., Bratislava, Slovakia
marian.kovacik@sk.ibm.com

Section 1
Use of Knowledge and Semantics in Distributed Computing

Semantic web services based crisis information system exploiting automated workflow composition

Peter Bartaloš[1], Martin Gažák[2]

[1]Slovak University of Technology in Bratislava, Faculty of Informatics an Information Technologies, Institute of Informatics and Software Engineering
Ilkovičova 3, 842 16, Bratislava, Slovak Republic
[2]MicroStep-MIS spol. s r.o., Čavojského 1, 84108 Bratislava, Slovak Republic

bartalos@fiit.stuba.sk, martin.gazak@microstep-mis.com

Abstract. The process of a pollution assessment and prediction of the consequences in case of radiological emergence usually results in complex data- and work- flows among databases, models and simulation tools (geographical databases, meteorological and dispersion models, etc.). The architecture and inherent heterogeneity of the system, computational complexity and interfaces to other systems and services make it well suited for the decomposition of a system into the set of web and grid services. In this paper we deal with the possibility of exploiting (semantic) web service technologies in a context of a crisis information system.

Quality of Semantic Grid Services

Peter Bednár and Tomáš Kasanický

Introduction

One of the approaches how to implement automatic discovery, compositions and execution of services in the GRID environment is based on the semantic technologies adopted for the description of the services. There are many emerging proposals for semantic web services (OWL-S, WSMO, SAWSDL) which define how to describe functional properties of the services, i.e. relations between the inputs and outputs and how to describe behavior of the services during the execution (i.e. how to describe interaction between the service requester and the service or between the services orchestrated into the composite service). These functional properties define the capability of the service, which has to fulfill the specified goal in order to provide outputs required by the user for the specified inputs.

It is possible that many services fulfill functional properties so it is necessary to select one particular candidate for the invocation. Selection can be automatic or semi-automatic and it is based on the non-functional properties of services, which cover all additional aspects important for the service execution, including contracting, information about the service provider, and QoS quantitative measures.

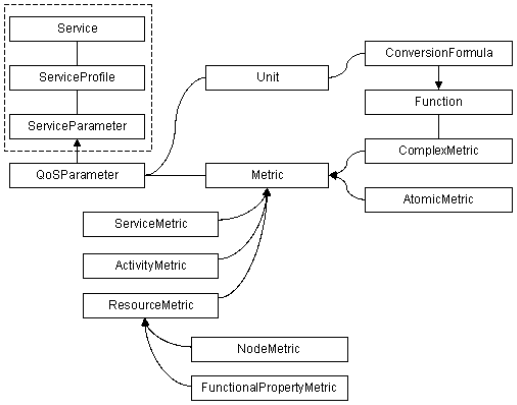
In this paper we described approach for the selection of candidates according to non-functional properties organized in the QoS ontology. The paper is divided to the following chapters: first chapter describes the main concepts of the proposed QoS ontology, the second chapter describes monitoring environment for monitoring of the quantitative characteristics and the last chapter describe the module for the semi-automatic selection of the services.

QoS ontology

There is a necessary to choose the services witch is able to return the result witch we required. The many criteria can be used in the the selection process. Process of the automatic services selection can be divided in to the two parts:

- Comparison of the functional properties of the services
- Comparison of the quality indicators

We need to conceptualize all the services for the automatic comparison. The picture number one describe the ontology, our conceptual model

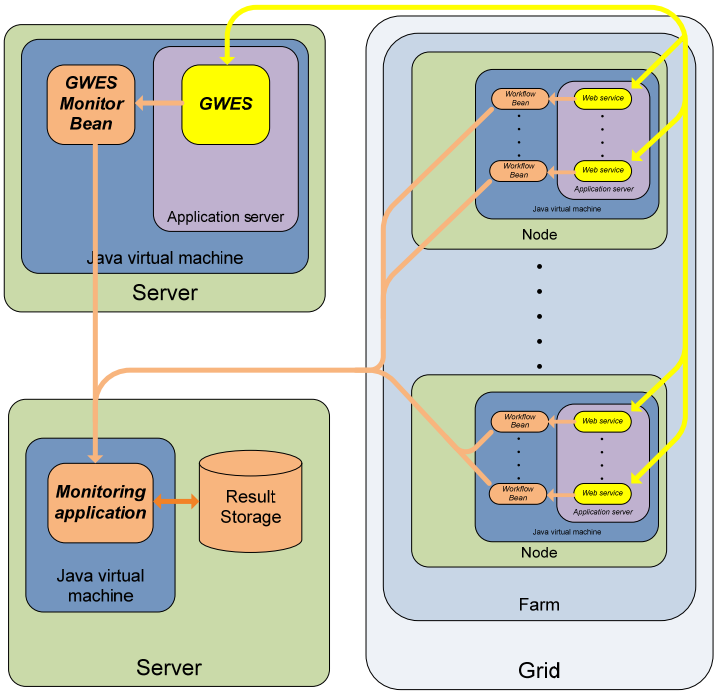


Architecture for monitoring of the GRID execution environment

Approach described in this paper is based on the indirect diagnostic of the services quality. The Grid Workflow Execution Service is used as execution environment. If java language is used for development of services and as we know GWES is programmed in java, than we can used JMX (Java Management Extension) for monitoring of the JVM (Java Virtual Machine). There are two types of the monitoring used in presented work:

- GWES monitoring, activity monitoring module
- Client server application for the detail resources monitoring and result storing

Detailed functionality is shown on picture number two.



Selection of service candidates

The selection of the candidate for the invocation can be automatic or semiautomatic with the additional interaction with the user. The automatic selection is performed in two steps:

- In the first step the service candidates are ordered according to multiple quantitative criteria inferred from the QoS metrics.
- In the second step, candidates are additionally filtered according to the semantic query specified for the declarative non-functional properties such properties about the provider, security, contracting, pricing etc.

The first step is based on the sorting of the services according to the quantitative QoS metrics aggregated per service. Sorting is weighted, i.e. user can specify his/her preferences (“weight”) for each criterion included in the sorting (if weight for the metric is 0, then it is excluded from the sorting and it is not taken into account during the candidate selection).

The semiautomatic approach is based on the two-dimension visualization of the quantitative QoS metrics. The goal is to visualize the QoS metrics in the way, where the user can simply identify services with good performance or identify potential bottlenecks. For both axes, it is possible to specify weights for the linear combination of QoS metrics (aggregated for the service or “raw” for each invocation) or specify if the inverted value is included in the visualization (i.e. 1/value). In this way, user can visualize base two-dimensional graphs for two metrics or use linear multidimensional scaling in order to visualize multiple metrics (for example the linear combination of the functional property metrics, i.e. service inputs on the x-axis and the execution time on the y-axis).

Semantic Composition of Web and Grid Services

Peter Bartaloš[1], Martin Gažák[2], Marek Paralič[3], Ondrej Habala[4]

[1]Slovak University of Technology in Bratislava, Faculty of Informatics an Information Technologies, Institute of Informatics and Software Engineering
Ilkovičova 3, 842 16, Bratislava, Slovak Republic
[2]MicroStep-MIS spol. s r.o., Čavojského 1, 84108 Bratislava, Slovak Republic
[4]Institute of Informatics SAS, Bratislava, Slovak Republic

bartalos@fiit.stuba.sk, martin.gazak@microstep-mis.com

Prediction of Meteorologically Significant Events Using Data Mining

Martin Gažák, Juraj Bartók, Ján Paralič, Peter Bednár,
Ladislav Hluchý, Ondrej Habala

MicroStep-MIS spol. s r.o., Čavojského 1, 84108 Bratislava, Slovak Republic
Institute of Informatics SAS, Bratislava, Slovak Republic

martin.gazak@microstep-mis.com

Abstract. The project Data Mining Meteo starts in September 2009 for two years. In its course we will design, deploy, and assess advanced data mining and data integration methods in the context of several environmental scenarios. These scenarios have been selected by an industrial partner, who is active in development and deployment of monitoring and risk management systems. Initially, three scenarios are envisaged: short-term fog prediction, precise rainfall detection using radar imagery, and low cloud cover potentially dangerous to aviation. In this paper we present the details of these scenarios, and an outline of the data integration and data mining methods we plan to use.

Application of Data Integration and Mining to Environmental Scenarios

Ondrej Habala, Martin Šeleng, Viet Tran, Ladislav Hluchý

Institute of Informatics SAS, Bratislava, Slovak Republic

ondrej.habala@savba.sk

Abstract. Application of Data Integration and Mining to Environmental Scenarios Abstract: As part of the ongoing EU FP7 project ADMIRE we have developed a set of scenarios from the environmental domain, which use advanced data integration and data mining techniques. We have devised a generic framework for data mining and integration of spatio-temporal data, and in its context have developed and deployed a set of OGSA-DAI services for extraction, conversion, cleaning, integration, and mining of data from several independent providers. The scenarios which use these data are from the hydrological, meteorological, and water management domains.

Tools for Advanced Data Mining

Branislav Simo, Michal Laclavik, Ivan Janciak, Ladislav Hluchy

Institute of Informatics SAS, Bratislava, Slovak Republic

Branislav.simo@savba.sk

Abstract. This paper describes the tools supporting advanced data mining and integration that are being developed in the context of the EU ADMIRE project. It briefly describes the architecture and infrastructure used and then continues with the description of the tools that allow users to exploit the infrastructure. The tools - Process Designer, Data preparation tool and Semantic Knowledge Sharing Assitant - are integrated in the ADMIRE Workbench, which is a Java fat client based on the Eclipse application framework.

Section 2
Computational Chemistry & Material Science

Material tension stress-strain curve determination via inverse analysis using finite element method in computational Grids

Ladislav Écsi[1] , Pavel Élesztős[1], Viera Šipková[2] , Miroslav Dobrucký[2], Ján Astaloš[2]

[1] Institute of Applied Mechanics and Mechatronics, Faculty of Mechanical Engineering,
Slovak University of Technology in Bratislava, Námestie slobody 17
812 31 Bratislava, Slovakia
ladislav.ecsi@stuba.sk

[2] Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9
845 07 Bratislava, Slovakia
viera.sipkova@savba.sk

Abstract. The material tensile stress-strain curve represents one of the most important material characteristics for elastic-plastic finite element analysis (FEA). While the curve can be determined quite accurately in small strain small deformation conditions using standard tension tests, in large strain large deformation conditions, due to non uni-axial stress state at near notch region of a specimen, the accurate stress-strain curve determination is almost impossible. In this paper the authors present an alternative method of the aforementioned material characteristic measurement using an inverse analysis in a Grid computing environment. The idea is based on a repeated finite element analysis (FEA) execution until the best possible results are reached. In the numerical simulation the finite element method (FEM) using large strain/ large deformation formulation and the updated Lagrange method were used. The material yield function f , the co-rotational Cauchy stress tensor $\hat{\boldsymbol{\sigma}}$, the objective fourth order elastic material tensor $\mathbb{C}$, the second order yield surface normal tensor $\mathbf{N}$ and the corresponding evolution equations are expressed in terms of the accumulated plastic strain rate $\dot{\varepsilon}^p$, the accumulated plastic strain ε^p and a rotating tensor $\mathbf{R}$ as follows

$$f(\hat{\boldsymbol{\sigma}},\sigma_y)=\sqrt{3J_2}-\sigma_y\le 0\ ,\ J_2=J_2\left(\hat{\boldsymbol{\Sigma}}\right)=\frac{1}{2}\hat{\boldsymbol{\Sigma}}:\hat{\boldsymbol{\Sigma}}\ , \tag{1}$$

$$\hat{\boldsymbol{\sigma}}^{n+1}=\Delta t\left(\mathbf{C}^{\nabla J}:\hat{\mathbf{d}}^{n+\frac{1}{2}}\right)+\hat{\boldsymbol{\sigma}}^{n}, \quad \text{in elastic region} \tag{2}$$

$$\hat{\boldsymbol{\sigma}}^{n+1}=\Delta t\left[\mathbb{C}^{\nabla J}-\frac{\left(\mathbb{C}^{\nabla J}:\mathbf{N}\right)\otimes\left(\mathbb{C}^{\nabla J}:\mathbf{N}\right)}{\mathbf{N}:\mathbb{C}^{\nabla J}:\mathbf{N}+H}\right]:\hat{\mathbf{d}}^{n+\frac{1}{2}}+\hat{\boldsymbol{\sigma}}^{n} \quad \text{in plastic region} \tag{3}$$

$$\mathbb{C}^{\nabla J}_{abcd} = \frac{1}{{}^{n+\frac{1}{2}}J} R^{n+\frac{1}{2}}_{am} \cdot R^{n+\frac{1}{2}}_{bn} \cdot R^{n+\frac{1}{2}}_{co} \cdot R^{n+\frac{1}{2}}_{dp} \cdot F^{n+\frac{1}{2}}_{mi} \cdot F^{n+\frac{1}{2}}_{nj} \cdot F^{n+\frac{1}{2}}_{ok} \cdot F^{n+\frac{1}{2}}_{pl} \cdot \mathbb{C}_{ijkl}\,, \tag{4}$$

$$H = H\left(\varepsilon^p\right) = \frac{\partial \sigma_y}{\partial \varepsilon^p}\,,\ \mathbf{N} = \sqrt{\frac{3}{2}}\frac{\hat{\boldsymbol{\Sigma}}}{\|\hat{\boldsymbol{\Sigma}}\|}\,,\ \ \hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\sigma}} - \frac{1}{3} tr(\hat{\boldsymbol{\sigma}})\mathbf{1}\,, \tag{5}$$

$$\hat{\boldsymbol{\sigma}}^{n+1} = \mathbf{R}^{n+1T}\boldsymbol{\sigma}^{n+1}\mathbf{R}^{n+1}, \hat{\boldsymbol{\sigma}}^{n} = \mathbf{R}^{nT}\boldsymbol{\sigma}^{n}\mathbf{R}^{n}\,,\ \dot{\varepsilon}^{p} = \sqrt{\frac{2}{3}\hat{\mathbf{d}}^{p} : \hat{\mathbf{d}}^{p}} = \dot{\lambda}\,,\ \varepsilon^{p} = \int_0^{t+\Delta t} \dot{\varepsilon}^{p} dt\,, \tag{6}$$

$$\mathbf{R}^{n+\frac{1}{2}} = \exp\left[\frac{\Delta t}{2}\mathbf{W}^{n+\frac{1}{2}}\right]\mathbf{R}^{n}\,,\ \mathbf{R}^{n+1} = \exp\left[\Delta t \mathbf{W}^{n+\frac{1}{2}}\right]\mathbf{R}^{n}\ . \tag{7}$$

In the stress update calculation the Jaumann objective rate in the form of the Green-Naghdi objective rate was used. In the presented paper some aspects of the mathematical model implementation are discussed and a few preliminary calculations on selected test examples are presented using straight and notched specimens. The calculation results will be briefly discussed and compared with available experimental results.

Acknowledgement

Funding using the VEGA grant No. 1/4103/07 resources is greatly appreciated.

References

1. Belytschko, T., Liu, W. K., Moran, B.: Nonlinear finite elements for continua and structures, John Wiley & Sons LTD, Chichester, 2000.
2. Crisfield, M. A.: Non-linear finite element analysis of solids and structures, Essentials, Vol 1, John Wiley & Sons LTD, Chichester, 2000.
3. Crisfield, M. A.: Non-linear finite element analysis of solids and structures, Advanced Topics, Vol 2, John Wiley & Sons LTD, Chichester, 2001.
4. Haupt, P.: Continuum mechanics and Theory of Materials, Springr, Berlin, 2000.
5. Mardsen, J. E., Hughes, T. J. R., Mathematical foundations of elasticity, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1983.
6. Oden, J. T.: Finite elements of nonlinear continua, McGraw-Hill, NY, 1972.
7. Simo, J. C., Hughes, T. J. R.: Computational inelasticity, Springer, NY, 1998.

The distributed Parameter Model of the Electrostatic-Actuated Gas-Damped MEMS-based Devices for the Simulation on the Parallel Computer Systems

Ivan Plander[1], Michal Stepanovsky[1]
[1]Alexander Dubcek University of Trencin, Trencin, Slovakia;

Corresponding Author: I. Plander, A. Dubcek University of Trencin, Dep. of Informatics
Studentska 2, 911 50 Trencin, Slovakia; plander@tnuni.sk

Extended abstract

In the paper the distributed parameter model of the electrostatic-actuated gas-damped micro-electro-mechanical-systems (MEMS)-based devices is presented. The model is suitable for the simulation only on the parallel computer systems because of their complexity. The modeling and simulation is an important aspect of design of various devices, especially when the fabrication of such products is expensive. Moreover, the simulation results provide detailed information about behavior of the proposed device. Thus, the optimization process can be realized before fabrication. The model is represented by the system of partial differential equations

$$\frac{\partial^2 \mathbf{\Pi}_1}{\partial t^2} + \frac{\partial \mathbf{\Pi}_2}{\partial t} + \mathbf{\Pi}_{3,1} + \mathbf{\Pi}_{3,2} = \mathbf{Q}, \tag{1}$$

where

$$\mathbf{\Pi}_1 = \begin{pmatrix} \rho_M u_x \\ \rho_M u_y \\ \rho_M u_z \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{\Pi}_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \rho_G \\ \rho_G v_x \\ \rho_G v_y \\ \rho_G v_z \\ \rho_G E \\ 0 \end{pmatrix}, \quad \mathbf{\Pi}_{3,1} = \frac{\partial}{\partial x}\begin{pmatrix} -\sigma_{xx} \\ -\sigma_{xy} \\ -\sigma_{xz} \\ \rho_G v_x \\ \rho_G v_x^2 + p \\ \rho_G v_x v_y \\ \rho_G v_x v_z \\ \rho_G H v_x \\ \varepsilon \dfrac{\partial V_e}{\partial x} \end{pmatrix} + \frac{\partial}{\partial y}\begin{pmatrix} -\sigma_{yx} \\ -\sigma_{yy} \\ -\sigma_{yz} \\ \rho_G v_y \\ \rho_G v_y v_x \\ \rho_G v_y^2 + p \\ \rho_G v_y v_z \\ \rho_G H v_y \\ \varepsilon \dfrac{\partial V_e}{\partial y} \end{pmatrix} + \frac{\partial}{\partial z}\begin{pmatrix} -\sigma_{zx} \\ -\sigma_{zy} \\ -\sigma_{zz} \\ \rho_G v_z \\ \rho_G v_z v_x \\ \rho_G v_z v_y \\ \rho_G v_z^2 + p \\ \rho_G H v_z \\ \varepsilon \dfrac{\partial V_e}{\partial z} \end{pmatrix}, \tag{2}$$

$$\mathbf{\Pi}_{3,2} = \frac{\partial}{\partial x}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\tau_{xx} \\ -\tau_{yx} \\ -\tau_{zx} \\ -\sum\limits_{k=x,y,z} \tau_{xk} v_k - k\dfrac{\partial T}{\partial x} \\ 0 \end{pmatrix} + \frac{\partial}{\partial y}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\tau_{xy} \\ -\tau_{yy} \\ -\tau_{zy} \\ -\sum\limits_{k=x,y,z} \tau_{yk} v_k - k\dfrac{\partial T}{\partial y} \\ 0 \end{pmatrix} + \frac{\partial}{\partial z}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\tau_{xz} \\ -\tau_{yz} \\ -\tau_{zz} \\ -\sum\limits_{k=x,y,z} \tau_{zk} v_k - k\dfrac{\partial T}{\partial z} \\ 0 \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} f_{M,x} \\ f_{M,y} \\ f_{M,z} \\ 0 \\ \rho_G f_{G,x} \\ \rho_G f_{G,y} \\ \rho_G f_{G,z} \\ \rho_G \left(f_{G,x} v_x + f_{G,y} v_y + f_{G,z} v_z \right) \\ q_e \end{pmatrix},$$

where ρ_M is the mass density of the material, u_i is the displacement , σ_{ij} is the Cauchy stress tensor component with $i, j = \{x, y, z\}$, where ρ_M, u_i and σ_{ij} are defined on the solid mechanics domain Ω_M; ρ_G, E, T, k, p, v_i, resp. τ_{ij} where $i, j = \{x, y, z\}$ defined on the gas dynamics domain Ω_G are the gas density, internal energy, gas temperature, thermal conductivity, gas pressure, gas velocity, resp. stress tensor component of the gas; ε, V_e, resp. q_e defined on the electrostatics domain Ω_e are the permitivity, electric potential, resp. charge density; H is the entalpy defined on the Ω_G:

$$H = E + \frac{p}{\rho_G}. \tag{3}$$

In the above, regarding the Hook's law for linear isotropic material:

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} \lambda_L + 2\mu_L & \lambda_L & \lambda_L & 0 & 0 & 0 \\ \lambda_L & \lambda_L + 2\mu_L & \lambda_L & 0 & 0 & 0 \\ \lambda_L & \lambda_L & \lambda_L + 2\mu_L & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_L & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_L & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_L \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ 2\varepsilon_{yz} \\ 2\varepsilon_{xz} \\ 2\varepsilon_{xy} \end{pmatrix}, \tag{4}$$

where ε_{ij} with $i, j = \{x, y, z\}$ is the strain; λ_L and μ_L are Lame's parameters:

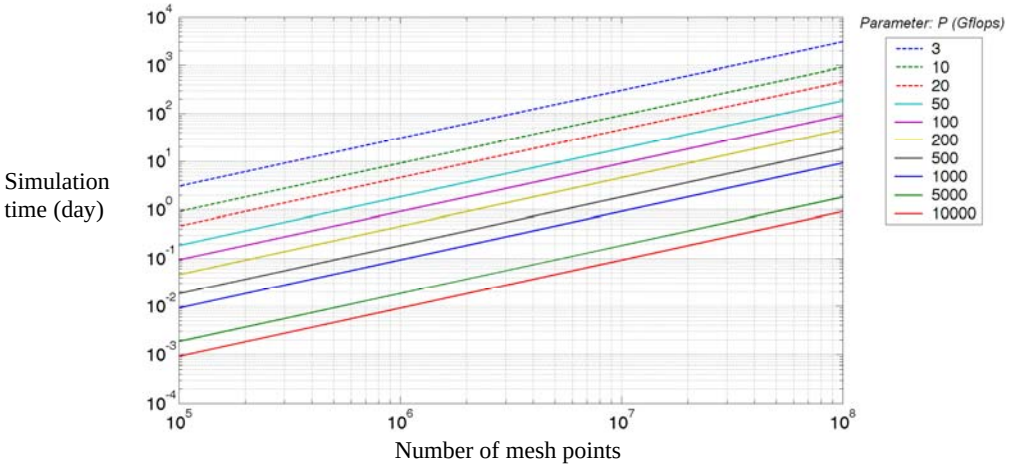
$$\lambda_L = \frac{\nu E}{(1+\nu)(1-2\nu)} \quad \text{and} \quad \mu_L = \frac{E}{2(1+\nu)}, \tag{5}$$

where E is Young modulus of elasticity and ν is Poisson constant.

The deviatoric stresses τ_{ij} are defined as follows

$$\tau_{ij} = \mu\left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i} - \frac{2}{3}\delta_{ij}\nabla^T \mathbf{v}\right). \tag{6}$$

The simulation time of the presented model is highly depending on the number of the mesh points of the discretized model. For different problems the different number of mesh point are desired. The dependence of the simulation time on the number of mesh points is shown in the next figure. The parameter P denotes the system performance in GFlops.



References:

[1] KALTENBACHER, Manfred. *Numerical Simulation of Mechatronic Sensors and Actuators*. Berlin (Germany): Springer-Verlag, 2004. 311 s.

[2] ZIENKIEWICZ, O.C. – TAYLOR, R.L.: *The Finite Element Method: Fluid Dynamics*. Oxford: Butterworth-Heinemann, 2000. 334 p.,Vol.3, 5th edition.

[3] YOUNIS, M.: *Modeling and Simulation of Microelectromechanical Systems in Multi-Physics Fields*. PhD thesis. Faculty of Virginia Polytechnic Institute and State University. 2004.

[4] FELIPPA, Carlos A.: *Introduction to Finite Element Methods*. Colorado (USA): Department of Aerospace Engineering Sciences and Center for Aerospace Structures, University of Colorado, 2004.

Section 3
Distributed Computing and Large Scale Applications

Grid–enabled maximum clique algorithm

[1]Karol Grondžák and [2]Peter Kortiš

[1]Faculty of Management Science and Informatics, University of Žilina
Slovak Republic
[2]Faculty of Electrical Engineering, University of Žilina
Slovak Republic

Karol.Grondzak@fri.uniza.sk,Peter.Kortis@fel.uniza.sk

Abstract. There are many scientific problems where the evaluation of maximal or maximum cliques is needed. In the coding theory maximum clique problem is solved to find largest binary code able to correct prescribed number of errors. In computational biology problem of maximum or maximal cliques is applied for solving the problems of genome mapping or 3-D protein structure alignment. Some tasks in computer vision and pattern recognition areas include the problem of finding either maximum or maximal cliques. The maximal or maximum clique determination is proved to be NP-hard problem. This problem has been studied for a long time, but because of its nature, it is still being actively studied and it attracts researchers all around the world. In this paper, we will present a proposal for a grid–enabled algorithm for maximum clique search. The algorithm is tested on a problem of PIN code generation.

Key words: PIN code generation, security, maximum clique search, grid algorithm

Parallelization Techniques for the Matrix Test Precomputation

Extended Abstract

Matúš Jókay and Pavol Zajac *

KAIVT FEI STU, Ilkovičova 3, Bratislava, Slovakia

A new test for randomness using the index μ and period λ of Boolean matrices was proposed in [2]. For a practical implementation of the test, a significant computing effort is required to precompute the test statistics. If the parameter of the test is n, we need to precompute indexes and periods for every matrix in the whole set of $n \times n$ Boolean matrices. The raw complexity of the precomputation effort is thus 2^{n^2} operations, where every operation is a computation of the index and period of a given Boolean matrix. The computation of the index and period can be transformed to a cycle finding problem with complexity $O(\mu + \lambda)$ matrix multiplications [1]. A partial speedup is only possible if large random access memory is available. Using 2^{n^2} memory blocks, we can store statistics for whole sequences, reducing the time complexity to 2^{n^2} matrix multiplications. Statistics in the paper [2] were precomputed for values of n up to 6. Using the internal parallelization techniques and GRID computing we would like to extend these results to cases $n = 7$, and $n = 8$, respectively.

A Boolean matrix is a matrix containing only elements from the set $\{0, 1\}$. We define a binary operation $\cdot$ on Boolean matrices as a matrix multiplication with the operation OR used instead of addition and operation AND used instead of multiplication. A set of all $n \times n$ Boolean matrices along with operation $\odot$ is a semigroup. This semigroup was studied already in [3] in the context of binary relations, with the operation $\odot$ representing the composition of relations.

Let S be a finite semigroup. Let us construct a sequence $x, x^2, \ldots$ For any $x \in S$ we can find $1 \leq a < b$ such that $x^a = x^b$. The smallest positive numbers $\mu = \mu(x), \lambda = \lambda(x)$ such that $x^\mu = x^{\mu+\lambda}$ are called the index, and the period of x, respectively. According to the Euler-Fermat Theorem for Finite Semigroups [3], there exist universal exponents $\Lambda = \mathrm{lcm}\{\lambda(x)|x \in S\}$, $M = \max\{\mu(x)|x \in S\}$, and for every $x \in S : x^M = x^{M+\Lambda}$. For a semigroup of $n \times n$ Boolean matrices the universal exponents are given in [4].

Given a Boolean matric A we want to compute its index and period. A basic algorithm is a direct enumeration. We compute and store $A, A^2, ...$, where $A^{i+1} = A^i \odot A$. After computing A^{i+1} we compare it with every A^j for $j = 1, 2, ...i$. If we find such j that $A^j = A^{i+1}$, than the index is $\mu(A) = j$, and period is $\lambda(A) = i + 1 - j$. The straightforward method requires $\lambda + \mu - 1$ Boolean matrix multiplications and the same number of matrices stored in memory. The number of comparisons is $1/2(\lambda + \mu - 1)(\lambda + \mu - 2) + \mu$, i.e. $O\left((\lambda + \mu)^2\right)$.

* This research was partially supported by project NIL-I-004. Distributed computing power was provided by the GRID lab at FIIT STU under grant VEGA 1/0649/09.

Alternative algorithms to compute μ, λ are provided by a multitude of cycle finding algorithms, e.g. by Brent's cycle finding algorithm [1]. This algorithm requires a small memory storage (for 2 matrices only), and both $O(\lambda + \mu)$ multiplications and comparisons. However, the speedup only applies if μ and λ are sufficiently large. In our case most of the matrices have only small $\mu+\lambda$, thus the more advanced cycle finding algorithms are actually slowing the computation.

The problem of computing the indexes and periods for the whole set of matrices can be distributed easily by partitioning the set of matrices to disjunct set and computing statistics of each set on a separate computer/processor. There is no communication during the computation of statistics of the subset. Finally we need a short post-processing phase where we add the results from each computer together. The optimal speedup is obtained if every computer can compute its assigned task in the same total time.

On each processor we can also exploit an internal parallelization using a so-called bit-slicing technique. To implement the operation $\odot$ we only need bit operations AND, OR working on individual bits in the matrix. An implementation of these operations on a typical processor works with the whole vector of 32-bits or 64-bits at once. Using SSE2 (Streaming SIMD Extension 2) registers and operations, we can even work with 128-bit vectors within one tact of the processor. A bit-sliced implementation stores b-matrices in $n \times n$ b-bit words, where bits of the first matrix are stored in LSB bit of every word, and so on. The implementation of b-parallel matrix multiplications is then straightforward. We have also developed an efficient algorithm for a comparison of the matrices and a computation of the λ, μ statistics.

We have implemented two bit-sliced versions of the basic algorithm in the C programming language. One is working with classical registers and operations only. The second one is using SSE2, that is now available on most of the processors. We have compared various implementation and computed the speedup of the internal parallelization for two distinct scenarios: a packing of random matrices, and a packing of matrices in the specific order given by the enumeration algorithm. The first case is typical for an actual application of the Matrix set for a given tested sequence. The second case is used in the actual computation of the statistics of the whole set of Boolean matrices. Our experiments show that the second case provides a higher speedup, with highly regular statistical results. This might indicate an interesting topic for a further theoretical research.

References

1. Brent, R.: An improved Monte Carlo factorization algorithm. BIT 20, 176–184 (1980)
2. Grošek, O., Vojvoda, M., Krchnavý, R.: A new matrix test for randomness. Computing 85, 21–36 (2009)
3. Schwarz, Š.: On the semigroup of binary relations on a finite set. Czech Math J 20(95), 632–679 (1970)
4. Schwarz, Š.: The Euler-Fermat theorem for the semigroup of circulant boolean matrices. Czech Math J 30(105), 135–141 (1980)

TSP Models for cluster computing on the base of genetics algorithms

Jarmila Škrinárová and Filip Zelinka

Metej Bel University Banská Bystrica, Faculty of Natural Science, Department of Informatics
skrinar@fpv.umb.sk, zelinkaf@gmail.com

Abstract. The Traveling Salesman problem has received great attention from the operational research and computer science communities because it is very easy to describe but very hard to solve. The problem can be formulated saying that the traveling salesman must visit every city in his territory exactly once and then return to the starting point. Given the cost of travel between all cities, he should plan his itinerary for a minimum total cost of the entire tour. The most direct solution would be to try all permutations and find out the cheapest tour using brute force search. The running time for this approach lies within a polynomial factor of O(n!), the factorial of the number of cities, so this solution becomes impractical even for only 20 cities. One of the earliest applications of dynamic programming is an algorithm that solves the problem in $O(n^2 2^n)$ [1]. This task is extremely suitable for high performance computing.

Various heuristics and approximation algorithms, which quickly yield good solutions, have been developed. For example, algorithm of constructive heuristics, the nearest neighbour algorithm, so-called greedy algorithm. This algorithm chooses the nearest unvisited city for his next move and quickly yields an effectively short route. Results for *N* cities are obviously approximately 25% worse than the shortest tour. Some algorithms are based on Lin–Kernighan heuristics. The pair wise exchange, so-called *2-opt* technique iteratively removes two edges and replaces these with two different edges that reconnect the fragments created by edge removal into a new and shorter tour. Known and improved algorithms on this base are using *k-opt* and *V-opt* heuristics. Randomized improvements are obviously reached by optimized Markov chain, random path change or some general heuristics such as genetic algorithms.

In this paper a TSP problem solution on the base of genetic algorithm and its implementation on the computer cluster as a small part of computer grid is presented. There are various implementations introduced and compared. Their goal is to find out appropriate methods of decomposition and adequate practices for attaching good solution for high performance computation of TSP task.

This paper consists of some parts. First part is concerned on genetic algorithms in general and design of TSP solution on the base of genetic algorithms. Genetic algorithms consist of following steps: Generate random population of *n* suitable solutions for the problem (chromosomes, individualities). Evaluate the fitness *f(x)* of each chromosome *x* in the population. Create a new population by replacing the most individualities of population by new individualities reached, for example by mutation, or other techniques. Replace and use new generated population for the further run of algorithm. Test the final condition and if satisfied, stop, and return the best solution in

current population. In opposite case jump back to the evaluation and repeat the whole process.

In our task, the population was created by generating permutations of validated TSP tours on the base of Knuth's shuffle. The identifier of city is a number. Tour starts and finishes in city marked 0. The tour is represented as an array of numbers (visited cities). This method produces permutations from sorted array. Convenience of every individual (tour) is evaluated by the fitness function. In TSP it is the length of tour. If the length of tour is shorter, then is the better individual. One third of these best individuals will be in new population. Second third is newly generated population and last third of the children population is created by mutation of the best individuals.. The old population is replaced by new one. If the final condition is satisfied, algorithm is stopped and returns the best solution (or certain number of the best solutions) in current population. These partial results are needed for following processing [2].

Second part of this paper introduces computer cluster, as a part of computer grid, used for implementation [3]. In third part various models of solution are described. Designed models are assumed to be suitable for acceleration of calculation and reaching better results. It means, for example, to find out a tour which is the closest to the best known tour. First model of implementation is designed as a distribution on the particular nodes and chooses the best tour from every used node. Another models are based on improvements of partial results by Lin–Kernighan and others heuristics. In next parts performance of these models are verified and compared. There are also some methods for high performance computing for TSP composed and generalized.

Reference

[1] BELLMAN, R., 2006 *Combinatorial Processes and Dynamic Programming, Combinatorial Analysis*, Proceedings of Symposia in Applied Mathematics 10,, American Mathematical Society, pp. 217–249

[2] GOLDBERG, D., E., 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st edition Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, pp. 372 ISBN:0201157675

[3] ŠKRINÁROVÁ, J., MELICHERČÍK, M., 2008: *Measuring concurrency of parallel algorithms.* In: Proceeding of the 2008 1st International Conference on Information Technology Gdansk : IEEE computer society, 2008, p. 289-292, IEEE Katalog Number: CFP0825E-PRT. ISBN 978-1-4244-2244-9. Library of Congress 2008901868.

Section 4
Grid and Service-oriented Computing

User-friendly Access to Grid using g-Eclipse

Peter Kurdel, Jolana Sebestyénová

Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
peter.kurdel@savba.sk, sebestyenova@savba.sk

Abstract. The central idea of g-Eclipse is to simplify access to the grid to build an integrated workbench framework to access the power of existing grid infrastructures. The framework was built on top of the reliable eco-system of the Eclipse community to enable a sustainable development. The framework provides tools to customize grid users' applications, to manage grid resources and to support the development cycle of new grid applications.

Brief Introduction to Cloud Computing

Minh B. NGUYEN
Institute of Informatics, Slovak Academy of Sciences
Dubravska cesta 9, 845 07 Bratislava, Slovakia
Email: minh.ui@savba.sk

From the beginning of the 21st century, the term "Cloud Computing" had started to be mentioned. In 2007, Google, IBM and a number of research institutes saw increased activities of Cloud computing, they have researched and developed the technologies. Cloud computing have been becoming a hot topic by mid - 2008 and numerous cloud computing events had been scheduled. This paper tries to take one overview of Cloud computing. This paper will provide a short introduction to cloud computing and also try to compare the similarities and contrast between Cloud Computing with Grid Computing through characteristics of both.

What is the Cloud Computing?

"Cloud computing" is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet, allow users to access technology without knowledge of, expertise with, or control over the technology infrastructure that supports them [1].

This concept may be described simply that the computational power as software, services … are located in the virtual server on internet instead of location in the PC (on the ground) for everybody or businesses connect and use whenever they need. With the available services of Cloud, businesses have not to purchase and maintain a lot of computers and software therefore they will save expenditure and only concentrate for productions [7]. Some examples of cloud computing are the following: Amazon.com, Amazon EC2 [2], [3], Salesforce.com, Google's Apps, PayPal, Data Centres (IBM, Microsoft, and others) [4], Window Azure and operation system on Web was developed by Microsoft and many more.

Who uses Cloud Computing?

- Enterprise Cloud Computing: This is the primary business use of Cloud Computing. Medium to large companies who want to more efficiently run their operations in Cloud.
- Government Cloud Computing Applications: Governments are starting to look a little more closely about leveraging Cloud to run their government applications.
- Healthcare Applications: Large healthcare organizations are starting to look at leveraging Cloud Computational power. Obviously with all the sensitive medical info at stake, there are security implications and issues that need to be worked out.
- Social Media Cloud Computing: When we use Cloud Computing to collaborate with others.
- Personal Cloud Computing: Everybody. These applications are powered by Cloud that enables us to perform our tasks.

Why choose Cloud Computing?

However, users will be depended on technology which was provided by servers or manufacturers that would make reduce flexibility and creativity in the activities of them. Also they will feel constrain when they only have right to perform within scope of services or providers allowed. But why you willingly relinquish to control of resources and allow them to virtually exist in the Cloud? The following are main reasons:

- Users (customers) can access to applications and their data store at anywhere has connected to internet and any time.
- If the customers are company, Cloud will help them to save office space for servers.
- Cloud Computing helps to reduce hardware and software costs.

- When companies use applications of Cloud, they have not to employ IT managers therefore they will save expenditure.
- If inside of Cloud is a "Grid" in which resources connected together as Gird Computing, users will utilise full ability of processing for solving very large task.

Comparing between Cloud Computing and Grid Computing

Cloud Computing is new technology that has features of Grid Computing and also has individualizations, however the visions of Cloud and Grid are the same – to reduce the cost of computing, increase reliability [5][6][9]. For better to understand about Cloud Computing, in this paper we will compare between Cloud and Grid Computing from several characteristics as economic model, architecture, resources and virtualization e.g.

Conclusions

Cloud Computing is a next big step of field computing and storage. Through this paper, we have an overview about Cloud Computing and technologies its. Although there are some concerns about security [8], however providers are researching on various solutions for response to customer's concerns and then we will have benefits for user, science and businesses.

References

[1] "What is Cloud Computing", Wikipedia.com. http://en.wikipedia.org/wiki/Cloud_computing, 2009.

[2] Amazon Elastic Compute Cloud (Amazon EC2), http://aws.amazon.com/ec2, 2009.

[3] Amazon Simple Storage Service (Amazon S3), http://aws.amazon.com/s3, 2009.

[4] IBM Cloud Computing http://www.ibm.com/cloud/, 2009.

[5] Open Cloud Consortium. http://www.opencloudconsortium.org, 2009.

[6] Borja Sotomayor, Lisa Childers Morgan Kaufmann, Globus Toolkit 4: Programming Java Services; First Edition, December 2005.

[7] "Twenty Experts Define Cloud Computing", SYS-CON Media Inc, http://cloudcomputing.sys-con.com/read/612375_p.htm, 2008.

[8] J. Brodkin, Gartner: "Seven cloud-computing security risks", http://www.networkworld.com/news/2008/070208-cloud.html, 2008.

[9] B. Allcock, J. Bester, J. Bresnahan, A. L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnal, S. Tuecke."Data Management and Transfer in High PerformanceComputational Grid Environments", Parallel ComputingJournal, Vol. 28 (5), May 2002.

Distributed GPGPU – new trends

Ján PERHÁČ, Branislav SOBOTA, Csaba SZABÓ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

jan.perhac@tuke.sk, branislav.sobota@tuke.sk, csaba.szabo@tuke.sk

Abstract — More and more effort has been made to increase the computational power of latest generations of graphical hardware. Nowadays, this hardware represents a massive parallel computational architecture and with its latest improvements in graphic processor unit programmability the graphical hardware can be utilized not only in classic image or scene manipulation or rendering but also in advanced physics simulation and in other computational or scientific areas as an high-performance mathematical coprocessor. This fact represents one of latest possibility how to increase the overall power of computer systems not only in present times and also in the future. This paper focuses on graphics hardware utilization possibilities especially in distributed computer systems.

Keywords —CPU, GPU, GPGPU, distributed computer systems, CUDA

1. Introduction

Desktop computers, notebooks, netbooks, server workstations, game consoles, mobile devices. They all are equipped not only with central processor unit (GPU), but also with a high performance graphics processor unit (GPU) chip, which comes mostly with large dedicated memory. The CPU is designed to run the operating system of any device and applications programs of any kind, written in many computer (high level) languages like C, C++, C#, JAVA, FORTRAN, PYTHON and so on because of its multipurpose architecture design. Compared to CPU the GPU has more specialized design, which better provides realization of graphical tasks like rendering of 3D scenes, image rasterization and other transformations. For many GPU generations the functionality or programmability of GPU processors was very limited. In nowadays, the GPU processors are able to compute ten millions of vertices and to rasterize hundreds of millions or more fragments per second. The computational time for GPU is significantly shorter the computational time of the same problem on a CPU processor. But GPU processors are not able to perform any kind of program for general purpose tasks, like the CPU

processor can. For many years the GPUs were used only for acceleration of processing of some parts of graphical calculation.

The classical architecture of GPUs was changed dramatically towards vertex and fragment shader architecture (SA) which was enhanced into unified shader architecture (USA). SA and USA allows now to perform also non-graphical operations on graphical hardware with an ease. The emerge of SA and USA also led to a new computer branch called General-Purpose computing on Graphics Processing Units (GPGPU), which utilize the computational power of specialized high-performance computer hardware based on SA or USA GPU processors. This new branch, about three years of age, gave impulse for creation of new programming languages, suitable especially for GPU processors. Some of these GPU based languages are: High Level Shader Language (HLSL), OpenGL Shading Language (GLSL), C for graphics (Cg), Compute Unified Device Architecture (CUDA), Close to Metal (CTM) and many others. Graphical processor is designed especially for processing of graphical operations and so its functionality is limited for operations execution and programmability. Considering the construction character of graphical hardware, the GPU is effective only in solving of problems, which can be executed in parallel. This means, that the GPU processor is able to concurrently process large amount of independent vertices and fragments. Basically, the GPU processor is a stream processor, which is able to execute one kernel in parallel on top of many data concurrently. The stream represents a dataset which requires identical computations. Streams are creating data parallelism. A kernel is a function, which is applied on every single element in the stream. Each vertices and fragments represents elements of stream, on which vertex and fragment of unifies shaders are executing given kernel.

The development of graphical processors is rising with high tempo and is motivated by three main aspects, which are depicted on fig. 1. Firstly it is the semiconductor industry, which doubles the number of transistors on chip every 18 months. The result of transistor count doubling is higher computational power, which is almost two times higher. This phenomenon is known as Moore law and is providing faster and cheaper hardware. Second aspect of GPU development is notable increase of computations count, which simulates ambient world. Human brain processes pictures of 3D world, which are perceived by eyes with high precision and sharpness. Probably we’ll never reach the point, when the computer graphics will be as real as is the reality. The reality is simply too real. Last strong puller of GPU development is our desire for strong visual experiences. This puller is a force, which connects the source (simulation and hardware) in a way, which provides the visual reality more real than whenever before.

The multi-dimensional model of grid computing

Viera Šipková

Ústav informatiky, Slovenská akadémia vied
845 07 Bratislava, Dúbravská cesta 9

Extended Abstract

Processors recently have added explicit parallelism in the form of multiple cores, and the number of cores is predicted to be increasing exponentially over time. Along this line processors will continue to scale exponentially in performance, but processors themselves are no longer significantly scaling their clock rate. Hence, software applications that are not parallelised at all possible levels will not benefit. Programs are required to exploit the multiple heterogeneous cores on a chip, multiple processors within a cluster, and the additional processors and clusters available across a network. This situation implies a need for new algorithms and program structures able to perform many operations at once, to take advantage of the new hardware. Terms *concurrency, locality, scalability, modularity*, and *portability* become fundamental requirements for an algorithm and application design. For the software development the selection of an appropriate programming model that matches the hybrid architecture of the available computing units, is crucial, it should expose maximum amounts of parallelism in an abstract way.

In this paper we have focused on the software engineering techniques of developing a distributed-parallel application intended for running on hardware resources like multi-core machines, clusters, and grids. The paper outlines how to apply and combine together different strategies, originated from HPC (High Performance Computing), HTC (High Throughput Computing), grid technologies and high-level tools, in order to design and compose large parallel programs that satisfy user requirements for correctness and performance.

Section 5
Astronomy & Astrophysics and High energy Physics

Visualization of Grid based astronomical simulation Flight towards the Sun

Eva Pajorová[1], Ľuboš Neslušan[2]

[1] Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 84507 Bratislava, Slovakia
utrrepaj@savba.sk
[2] Astronomical Institute Slovak Academy of Sciences,
05960 Tatranská Lomnica,
Slovakia
ne@ta3.sk

Abstract. The Design of the on-line visualization tool (VT) for Grid-based Parametric Studies as a sequence application in Grid environment has been developed. The design is tested on the applications those have the character as a Parametric studies. Examples are the Astronomical simulations. One of them is the simulation of the Oort-cloud formation. The dynamical evolution of the test particles was followed via numerical integration, in the GRID, for the period of 1 and 2 Gyr. The main problem is Visual control in time when the Grid application is still running. VT is actual tested on visualization "Flight towards the Sun "- the visualization of stellar surround and Oort-cloud in time 1 Gyr represent as a Flight of the virtual space ship from the long distance towards the Sun.

Keywords: grid computing, parameter studies, visualization tool

Execution of grid workflows by agent-scheduling tools

Viet D. Tran

Institute of Informatcis, Slovak Academy of Sciences
Viet.ui@savba.sk

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.

On the other hand, applications with parametric study have large amount of independent tasks without data dependence among them. These applications are usually executed in master-worker tools (also called agent-scheduling tools), where worker jobs will download tasks from master and execute them until all tasks are executed. Such tools are very efficient because they can reduce the overhead of Grid job submission, and also provide load balancing and fault tolerance.

As workflow may have large number of small tasks, the idea is to use agent-scheduling tools for execution the tasks from workflows. Although tasks in workflow may have data dependence, the tasks that are ready for execution do not have any dependence among them (if a task has data dependence on another unfinished tasks, it can not be ready). Therefore, they can be executed by agent-scheduling tools.

This approach can work efficiently also for small workflows (smaller numbers of long-time tasks). If jobs are submitted to heavy-loaded grid infrastrutures, they may have to wait for long times (hours, days) in queues before executions. It means that in classical workflow execution, once a task finishes, its successive tasks will be submitted to grid, and wait long times before starting. In agent-based scheme, an agent can execute more tasks, so the sucessive tasks can be executed immediately by the same agent, which can reduce total execution time of workflows.

For implementation, we are choosing K-WF Grid workflow management system [1], which has workflow representation based on Petri net and dynamic workflow composition. The system is strongly service-oriented, and it has been recently adapted to gLite for EGEE infrastructure [2]. For agent-scheduling tools, we choose DIANE [3] which is written in Python scripting languague and is very flexible. Advanced users can modify the source code and add new features if required.

The implementation details are straightforward. Once a task in a worklow is ready for execution, instead of submitting it directly to gLite, the K-WF Grid engine will store the task description in a file and send it to DIANE. The DIANE task creation

script is modified so it will check new tasks created by the workflow engine and add them to execution.

References

1. Knowledge-based workflow system for Grid applications. http://www.kwfgrid.eu/
2. gLite middleware. http://glite.org
3. DIANE: Distributed Analysis Environment. http://cern.ch/diane.

Using DIANE for parametric applications

Viet D. Tran

Institute of Informatcis, Slovak Academy of Sciences
Viet.ui@savba.sk

In this paper, we present our experiences with using DIANE for astrophysics applications. The applications have the characteristics of parametric study: many small independent tasks with the same code and different data. Using DIANE can improve the reliability and response time of the applications in comparison with standard gLite parametric jobs. This will help developers with parametric applications to port their applications to DIANE quickly

In the astrophysics EGEE cluster, we have four different applications dealing with simulations of different aspects of solar systems. The applications have a common characteristics: each simulation has many small independent tasks with the same code and different input data. Although gLite provides support for parametric jobs, experiments show a high rate of failure when too many jobs are submitted. Furthermore, submitting small tasks as Grid jobs is inefficient because the overhead of management. The more suitable approach for such applications is to use agent-scheduling tools, which work in master-worker style and can provide fault-tolerance and load balancing. We choose DIANE for our applications because it is in RESPECT program of EGEE and has a long development history.

DIANE is a lightweight job execution control framework for parallel scientific applications. DIANE improves the reliability and efficiency of job execution by providing automatic load balancing, fine-grained scheduling and failure recovery. model is based on master-worker architecture. This approach is also known as agent-based computing or pilot jobs in which a set of worker agents controls the resources. The resource allocation is independent from the application execution control and therefore may be easily adapted to various use cases. DIANE uses the Ganga interface to allocate resources by sending worker agent jobs, hence the system supports a large of computing backends: LSF, PBS, SGE, Condor, LCG/EGEE Grid.

DIANE (and its execution backend Ganga) is written in Python scripting language and the users need to write their applications as scripts in the programming language, too. However, recent version of DIANE 2.0-beta17 has greatly improved the usability by introducing ExecutableApplication module, that allows users keeps their applications in standard forms (scripts, binaries) and just to set the parameters to the modules.

Installing and using DIANE is straightforward. The developers provide simple installation script that will download DIANE (and also Ganga) and install it to user-specified directory on a machine with gLite User interface installed. The installation process can be done without root privileges (except for opening specific port in

firewall configuration for omniORB if firewall is installed). A simple tutorial is also provided for porting simple parametric applications to DIANE.

While simple use, DIANE can also provide nearly unlimited possibilities for configuration for advanced users. Master process can run on a separate machine without gLite user interface, i.e. it can run on any Unix-based platforms. Whole tool is written in scripting language so the application developers can change the scheduling scheme, adding priorities or making other modification according to their own needs.

Our experiments with DIANE for astrophysics applications show good performance of DIANE. The installation and configuration process of DIANE (without installing glite UI) can be done in few minutes and first version of parametric applications can be created in half hour with good performance.

Section 6
Environmental applications and Distributed Computing

New approach in Licensing for Simulation Climate Changes in GRID

Igor Kvasnica[1], Peter Kvasnica[2]

[1]Regional Department of the Living Environment Trenčín, Hviezdoslavová č. 3, 911 33 Trenčín, Slovak Republic
[2]Alexander Dubček University in Trenčín, Faculty of Mechatronics, Department of Informatics, Študentská č. 2, 911 50 Trenčín, Slovak Republic

kvasnica.igor@tn.kupz, kvasnica@tnuni.sk

Abstract. The paper deals with the method of licensing today in grid computing. Open Grid Service Architecture framework has been proposed how service oriented approach in emission air and climate change induce their additional utilization. Security this architecture is certainly an area, which is essential to grid computing. The result of a change to the emission factor that reflects the existing situation in modeling methods and technologies. Users priorities asso-ciated with the applications requiring them is essential to maximize software li-cense return on investment. Commercial grids today are mostly deployed at the enterprise level (i.e. within the intranet).

A Grid Computing Approach to a Depth of Snow Cover Modelling

Vladimír Siládi, Jozef Siláči, Norbert Polčák, and Ladislav Huraj

Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 974 01 Banská Bystrica, Slovak Republic
{siladi,silaci,polcak}@fpv.umb.sk
huraj@ucm.sk
http://www.fpv.umb.sk

Abstract. The global warming has an effect on changes of a snow cover over wintertime. This effect is observed in Slovak ski resorts, too. A prediction of these trends is important to build new and keep the existing ski resorts.
We are able to analyze the snow cover depth in detail. This analyze is based on the many continuos observations and measurements at the specific climatological stations of Slovak Hydrometeorogical Institute. However, these climatological stations do not render accurately all ski places, which could be examined. The aim of this work is the depth of the snow cover computing in the desired point based on the geographical characteristics of a specific geographical point in a modelled area. This computation is time-consuming.
In this paper, we present the use of the grid infrastructure to accelerate this computation. We designed a parallel application, which allows estimating the state of the snow cover depth the geomorphological entity Zvolenská kotlina. The output of our aplication are the matrices, which contain the computed depth of the snow cover values for the chosen period. The outputs are visualized by the GIS Grass.

Key words: Grid computing – Geographical data – Hydrogeographi – Model of snow cover depth – GIS

1 Introduction

The depth of the snow cover is very variable meteorological element in the landscape. It is many factors depended , mainly to the snow precipitation, altitude, air temperature, profile of relief, solar power, cloudiness, air temperature inversion, etc. The measurement of the snow cover is taken by meteorological, climatological and precipitation stations. The total depth of snow is stored, i.e. the dept of snow and the depth of new snow cover. We are able to analyze the depth of the snow cover in detail. This analyze is based on the many continuos observations and measurements at the specific climatological stations. We can geographical strictly characterize all these gauging places by the altitude, latitude and longitude, as well as by the detailed characteristic of the relief shape.

The computing of the solar power in the assigned geographical point is possible from a digital relief (terrain) model. This computing is based on the terrain slope and terrain orientation. We are able to describe an arbitrary geographical point in landscape by the geographical properties mentioned above (altitude, latitude, longitude, characteristic of relief shape, solar power). However, we are not able to identify the depth of snow cover without performing direct observations and measurements.

The aim of this work is the depth of the snow cover computing in the arbitrary point based on the geographical characteristics of a specific geographical point in a modelled area. The result is derived from the available data, which ware obtained from meteorogical stations, climatological stations and rain gauge stations from a defined landscape area.

As an example of the application of our designed method we have chosen the geomorphological entity Zvolenská kotlina, which is exactly defined by its borders [1]. We use the digital terrain model of this entity, which has been done. The 17 meteorogical stations of Slovak Hydrometeorogical Institute are situated in this defined area. The Table 1 contains the snow cover data for the month of January from these stations. We decided for the period years 1990–2009.

The input data are stored in the large matrices. The output values are depended on the time-consuming computing process. We try to speed up this process by using the parallelization. The parallel implementation was tested on small computer cluster with the 8 working nodes. After that we sent the tasks to run on the computation grid infrastructure. The output of our aplication are the matrices, which contain the computed depth of the snow cover values for the chosen period. Each element of the each matrix carries the value of the particular geographical point. This approach makes it possible to display desired data. The data are visualized through interactive maps. We use for the visualization GIS Grass. The GIS Grass used to filter, analyze and display a variant information, too. They are information such as: the average snow cover depth map on January during the period 1990–2009, the average snow cover depth map on January during chosen partial period, the average snow cover depth map on January during the chosen day, map of the area with snow cover depth over the 1 cm (5 cm, 10 cm, 20 cm, 50 cm), etc.

References

1. MAZÚR, E., LUKNIŠ, M.: *Geomorfologické jednotky. Map 1:500 000.* In: Atlas SSR. Bratislava: GÚ SAV a SÚGK, 1980.

Tutorials and Courses

Training course (including tutorial) for Grid users and application developers

Miroslav Dobrucký, Viera Šipková, Viet D. Tran

Institute of Informatics, Slovak Academy of Sciences, Bratislava

Participants of the hands-on tutorial are advised to bring their own laptops for this session with having an SSH client installed (e.g. PUTTY on Windows, SSH terminal on Linux).

Programme schedule:

- Current state of Grid projects: EGEE and EGI (European Grid Initiative) (20 min / Dobrucký)
- Grid security and getting access to the Grid (30 min / Dobrucký)
- Overview of Grid middlewares and high-level Grid tools (20 min / Tran)
- Development of Grid applications (20 min / Tran)
- Demonstration of DIANE (Lightweight Job Execution Framework) (30 min / Tran)
- gLite middleware (30 min / Šipková)

Hands-on tutorial
Practicals with the gLite middleware (60 min)
(Proxy generation and basic operations of job and data management using the gLite and LCG CLI on the Grid UI)