## 7<sup>th</sup> International Workshop on Grid Computing for Complex Problems



## GCCP 2011 Proceedings

Associate action to create national Grid initiative: Making the Grid accessible for electronic science in Slovakia (Sprístupnenie Gridu pre elektronickú vedu na Slovensku)

> October 24 - 26, 2011 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

EGI-InSPIRE - EU FP7 RI project: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe (2010-2014) FP7-261323; Project Research and development of new information technologies for prediction and solution of crisis situations and security of habitants CRISIS ITMS 26240220060; Project Industry research in the area of effective work with large data in user oriented applications RECLER ITMS 26240220029

#### **Program Committee**

Ladislav Hluchý	Institute of Informatics, Slovak Academy of Sciences
Ján Astaloš	Institute of Informatics, Slovak Academy of Sciences
Jozef Černák	Faculty of Science, P. J. Šafárik University, Košice
Miroslav Dobrucký	Institute of Informatics, Slovak Academy of Sciences
Ladislav Halada	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
Peter Kurdel	Institute of Informatics, Slovak Academy of Sciences
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
Jolana Sebestyénová	Institute of Informatics, Slovak Academy of Sciences
Branislav Sitár	FMFI UK Bratislava
Martin Šeleng	Institute of Informatics, Slovak Academy of Sciences
Jarmila Škrinárová	UMB Banská Bystrica
Milan Šujanský	Faculty of Electrical Engineering and Informatics, TU Košice
Viet Tran	Institute of Informatics, Slovak Academy of Sciences
Claudio Vuerli	INAF Trieste Italy
Ivan Zahradník	Institute of Molecular Physiology and Genetics SAS, Bratislava
Peter Závodný	University of Economics in Bratislava

#### **Organizing Committee**

Ladislav Hluchý	Institute of Informatics, Slovak Academy of Sciences
Miroslav Dobrucký	Institute of Informatics, Slovak Academy of Sciences
Peter Kurdel	Institute of Informatics, Slovak Academy of Sciences
Jolana Sebestyénová	Institute of Informatics, Slovak Academy of Sciences
Oľga Schusterová	Institute of Informatics, Slovak Academy of Sciences

Dúbravská cesta 9, 845 07 Bratislava, Slovakia E-mail: {ladislav.hluchy, miroslav.dobrucky, peter.kurdel, sebestyenova, sekr.ui}@savba.sk

#### **Proceeding Editors**

Ladislav Hluchý	Institute of Informatics, Slovak Academy of Sciences
Peter Kurdel	Institute of Informatics, Slovak Academy of Sciences
Jolana Sebestyénová	Institute of Informatics, Slovak Academy of Sciences

ISBN 978-80-970145-5-1

<sup>©</sup> Institute of Informatics SAS and the authors of respective articles, 2011

### Preface

Welcome to the 7<sup>th</sup> International Workshop on Grid Computing for Complex Problems GCCP 2011. The workshop is a three-day combined event for grid users: workshop with invited lectures, plenary discussions, accompanied by tutorial on Cloud computing, which is in the scope of EGI-InSPIRE - EU FP7 RI project: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe (2010-2014) FP7-261323. A special session Crisis situations predictions and solutions is in the scope of project Research and development of new information technologies for prediction and solution of crisis situations and security of habitants, and a special session Querying large data sets using design patterns is in the scope of project Industry research in the area of effective work with large data in user oriented applications.

The topics of the workshop are:

- Bio-applications
- Distributed Computing and Large Scale Applications
- Business benefits and challenges of grid in Financial Markets
- High Performance Distributed Computing and Large Scale Simulations
- Research Infrastructures
- Crisis situations predictions and solutions
- Querying large data sets using design patterns

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 2011 has attracted 41 paper contributions and active participations from Czech Republic, Finland, Italy, Poland, Romania, Ukraine and Slovakia.

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 2011 Bratislava, Slovakia

## **Table of Contents**

## Invited lectures

e-Infrastructure in the Czech Republic Luděk Matyska	12
Activities and Visions in the Italian Grid Infrastructure	22
<b>GPGPU Applications in Theoretical Chemistry</b>	23
An Overview of the Finnish Model for Provisioning of Services for Scientific Computing and its Extending to pan-European e-Infrastructures Per Öster	25
Polish Research Infrastructure and its Relation to e-Infrastructure Norbert Meyer	26
Intelligent Computations for Complex Problem Solving	27

## Section 1 Bio-applications

Modeling of Structure, Folding and Interactions of Biomolecules in the	
Era of GPGPU Computing	36
S. Hayryan, MC. Wu, CK. Hu, Z. Gažová and T. Kožár	
MedDecSup an Intelligent Distributed Medical Decision Support System	45
Adrian Gligor, Barna Iantovics	

## Section 2 Distributed Computing and Large Scale Applications

A Technique to Create Efficient Fortran Programs for Heterogeneous	
Parallel Platforms	54
Anatoliy Doroshenko, Kostiantyn Zhereb, Iurii Tyrchak, and Viet Tran	

<b>On the Performance of Vector Calculations on CPU and GPU</b> Karol Grondžák, Penka Martincová, and Ondrej Šuch	62
Analyzing Performance of Parallel Algorithms for Linear System Equations	70
Peter Hanuliak and Martin Hudik Remarks on GPU usage in Cryptanalysis	78
Matúš Jókay, Milan Plančík, and Pavol Zajac	10
<b>GPU-accelerated Parallel Genetic Algorithms</b>	86

## Section 3 Business benefits and challenges of grid in Financial Markets

Identification of Financial Contagion Outbreaks through Distributed         Systems	94
<b>Endogenous and Exogenous benefits of grid in Financial Markets</b> Calin-Adrian Comes	102
Financial Analysis Computing Architecture in Distributed Environment . Marián Zvada, Marek Domiter, Vladimír Motoška, Lukáš Slebodník, and Tomáš Kasanický	106
Recent Literature and Directions in Non-parametric Methods for Assessing Bank Performance through Efficiency Analysis of Decision Making Units Anca Munteanu	111

## Section 4 High Performance Distributed Computing and Large Scale Simulations

Augmented Reality Application in Parallel Computing System118Branislav Sobota, František Hrozek, Martin Varga, Csaba Szabó, Štefan118Korečko, and Peter Ivančák118

An Agent-based Security Approach for Intrusion Detection Systems Bogdan Crainicu, Barna Iantovics	126
Matrix Multiplying in Virtual Address Space on the 32/64-bit Computer Architecture for Simulation on the HPC	134
Igor Kvasnica, Peter Kvasnica	
Simulation of Uranus and Neptune Formation Using the new Asu-SAV	
Computer Cluster	140
Luboš Neslušan, Marián Jakubík	
Acceleration coefficients impact of the Particle Swarm Optimization algorithm in solving of the scheduling problem using MATLAB	
Distributed Computing Server	148
Ján Zelenka	

## Section 5 Research Infrastructures

Parallel Computing in EGI V. Šipková, M. Dobrucký, and P. Slížik	160
Facility of VOCE Grid Infrastructure for Spatial Data Computing Vladimír Siládi and Ladislav Huraj	168
A Genetic Algorithm for the Multiple Depot Periodic Vehicle Routing Problem Peter Kurdel, Jolana Sebestyénová	176

## Special session Crisis situations predictions and solutions

Computer Simulation of Forest Fires on Complex Terrain Ján Glasa, Peter Weisenpacher, Ladislav Halada	188
<b>Techniques of Parallel Data Mining for Numerical Values Prediction</b> Peter Krammer and Ladislav Hluchý	194
<b>Object-oriented Approach for Cloud Abstraction Layer</b>	202

<b>Optimalizácia výkonu klastra podľa aktuálnych potrieb gridu</b> Miroslav Dobrucký, Ladislav Hluchý	207
HMM Speech Synthesizer in Slovak Sakhia Darjaa, Marian Trnka, Miloš Cerňak, Milan Rusko, Róbert Sabo, Ladislav Hluchý	212
Tool for Creating 3D Slovak Speech Visemes Eva Pajorová	222
Voxel-based Rendering of E-Beam Lithography Processes Robert Andok, Peter Slížik, Pavol Nemec, Anna Benčurová, Viera Jablonská, Adrian Ritomský, and Ladislav Hluchý	228
Computer Fire Simulations by FDS and the Influence of a Domain Decomposition Strategy to Simulation Outcome P. Weisenpacher, L. Halada, J. Glasa	233
Special session Querying large data sets using design patterns	
Using Secure Agent Infrastructure for Management of Network Routers Branislav Šimo, Zoltán Balogh, Emil Gatial, Ladislav Hluchý	242
Using High Performance Data Streaming for Integration and Mining of Environmental Data Ondrej Habala, Ladislav Hluchý, Viet Tran, Peter Krammer, Martin Šeleng	250
Using Parallelization for Simulations of Human Behaviour Š. Dlugolinský, M. Kvassay, L. Hluchý, M. Wrzeszcz, D. Król, J. Kitowski	258
Enron Emails as Graph Data Corpus for Large-scale Graph Querying Experimentation	266
Testovanie Štandardov prístupnosti aplikácií určených pre verejnú	272
Matej Lučenič, Peter Bílik, Michal Ohrablo, Martin Vozár, Eugen Jeckel	212
Testovanie použiteľ nosti webových aplikácií Matej Lučenič, Peter Bílik, Michal Ohrablo, Martin Vozár, Eugen Jeckel	277
Distributed Crawling and Semantic Search in Internet Resources Martin Šeleng, Štefan Dlugolinský, Ladislav Hluchý	283

8

## **Tutorial**

Cloud Computing Guides and Tutorials	292
Viet Tran, Binh Minh Nguyen	

**Invited lectures** 

## e-Infrastructure in the Czech Republic

Luděk Matyska

CERIT-SC, Institute of Computer Science Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic, ludek@ics.muni.cz\*

**Abstract.** With a long history in academic networking and distributed computing support, the Czech Republic is currently in the process of setting up a more complex e-Infrastructure. Through activities of three organizations, it will cover all relevant aspects—communication infrastructure, computing infrastructure covering grids, clouds, and supercomputers, and initial phase of a data infrastructure. The Czech e-Infrastructure is build on an inclusive principle, that gives no exclusivity to the forerunners but expects more parties to be gradually involved in the distributed e-Infrastructure ecosystem.

**Keywords:** e-Infrastructure, Grids, Cloud computing, Storage infrastructure, High speed optical network

#### 1 Introduction

In the past two decades we have been witnessing very fast development of information technology and its widespread adoption by all human activities, with science requirements leading the way. The backbone of these activities is the computer network, in the form of Internet capable of transporting data between hundreds of millions of connected equipment. However, moving data around is simply not sufficient. Nowadays, humanity (and science in particular) is capable to produce ever increasing amounts of data, for which old approaches to data storage, manipulation, and processing are no more adequate. A notion of complex *e-Infrastructure*, a system that connects data depots and data processing (computing) facilities via high speed network, emerged. While in the past computer network (Internet connectivity) was sufficient to provide a proper infrastructure for science, the whole e-Infrastructure is needed now, to support large scale experiments and national and international collaboration.

For the country competitiveness in the research and development activities, the availability, quality, and sustainability of a complex e-Infrastructure is the key factor for adequate position in the international scientific collaboration, access to (or provision of) large scale experimental facilities and the quality of research community itself. The Czech Republic, with its long successful history

<sup>\*</sup> Support of the ESF project CERIT Scientific Cloud, CZ.1.05/3.2.00/08.0144 and the Large infrastructure CESNET is highly appreciated.

## Activities and Visions in the Italian Grid Infrastructure

Mirco Mazzucato

Istituto Nazionale di Fisica Nucleare Sezione di Padova Via Marzolo, 8 - 35131 Padova, Italy

The Italian national grid: IGI The relations with the European Grid Initiave – EGI The European Middleware: gLite and EMI Possible future directions for the Distributed Computing Infrastructures in Europe

The Italian Grid Initiave - IGI develops, maintains and operates, on behalf of the constituent organizations and with the aims to satisfy the need of the reference research user communities, the services of the national grid infrastructure which enable, through agreed policies at regional or international level, the remote acquisition, access and sharing of the ICT resources made available by the IGI shareholders or IGI itself.

At the end of 2010 IGI, currently a Joint Research Unit lead by INFN, has received by the Italian Ministry of University and Research- MIUR a "sustainable line of budget" to constitute a new legal organization pursuing on a more solid base the above objectives.

This talk gives an overview of the model and the organizational structure IGI is now developing for providing and consolidating the offer of the grid services required by the research communities. The strategy adopted by IGI to smoothly integrate a Cloud offer without any change in the current center popular best practices, based on the successful batch systems for the machines management, will be presented together with the expectations for an efficient integration with the EGI offer.

IGI is deeply involved in EMI. The status and future of some key components will be discussed together with the IGI vision for the future sustainability of the European Open Grid and Cloud Middleware software and of EGI itself.

### **GPGPU** Applications in Theoretical Chemistry

Stanislav Biskupič

Institute of Physical Chemistry and Chemical Physics, Slovak University of Technology, Bratislava, Slovakia stanislav.biskupic@stuba.sk

General purpose computing on graphical processing units, known as GPGPU (or now often referred to as GPU computing), is the approach of performing computation on the GPU instead of the CPU. GPU computing has been made possible by the increasing programmability and performance of GPUs.

The accuracy and extensibility of computational chemistry methods, particularly those which approximately solve the Schrödinger equation, are ultimately limited by the speed at which computer processors can execute floating point and memory operations. Due to fundamental limitations in processor technology, clock speeds are not singnificantly increasing, and all future increases in computational capability are expected to come from parallelism, which now more than ever can be found within a single processor. Graphics processing units (GPUs) are a type of massively parallel processor in which hundreds of cores can execute many instructions at once, provided they are sufficiently regular. Recently, many theoretical chemistry groups have demonstrated the incredible power of GPUs for scientific applications when sufficient effort is devoted to programming them to exploit their high degree of instruction-level parallelism. The programmability of GPUs has increased dramatically with the NVIDIA CUDA API and associated SDK including CUBLAS and CUFFT (together with other GPU oriented linear algebra libraries e. g. CULA or MAGMA), although these tools require more programming effort to realize the same relative performance as CPUs, especially for irregular algorithms.

To date, many quantum chemistry methods have been implemented on GPUs, including classical molecular dynamics<sup>1.4</sup>, 2-electron atomic integrals<sup>5-6</sup>, DFT<sup>7-10</sup> and SCF<sup>11,12</sup>, perturbation theory<sup>13,14</sup>, and quantum Monte Carlo<sup>15,16</sup>.

<sup>(1)</sup> Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. J. Comput. Chem. 2007, 28, 2618.

<sup>(2)</sup> Anderson, J. A.; Lorenz, C. D.; Travesset, A. J. Comput. Phys. 2008, 227, 5342.

<sup>(3)</sup> Liu, W.; Schmidt, B.; Voss, G.; Møller-Wittig, W. Comput. Phys. Commun. 2008, 179, 634.

<sup>(4)</sup> Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. J. Comput. Chem. 2009, 30, 864.

<sup>(5)</sup> Ufimtsev, I. S.; Martnez, T. J. J. Chem. Theory Comput. 2008, 4, 222.

## An Overview of the Finnish Model for Provisioning of Services for Scientific Computing and its Extending to pan-European e-Infrastructures

Per Öster

CSC - IT Center for Science Ltd, Finland

CSC- IT Center for Science Ltd, a Finnish information technology centre for research, provides modelling, computing and information services for academia, research institutes, the public sector and industry. CSC has also wide activities in data management and maintains Funet, the Finnish university and research network that also connects Finland to NORDUnet and GEANT. CSC is a limited non-profit company. The shares are fully owned by the Finnish state, and governed by the Finnish Ministry of Education and Culture. CSC is the largest national computing centre in Northern Europe with a staff exceeding 200. This talk gives an overview of the Finnish model for provisioning of services for scientific computing and how the national program extends to the pan-European e-infrastructures of EGI, PRACE, EUDAT and other research infrastructures (ESFRI).

## Polish Research Infrastructure and its Relation to e-Infrastructure

Norbert Meyer

Poznań Supercomputing and Networking Center, Poznań, Poland meyer@man.poznan.pl

The e-Infrastructure plays a significant role in European Roadmap for RI (Research Infrastructure) defined by ESFRI (European Strategy Forum on Research Infrastructures) and e-IRG (e-Infrastructure Reflection Group). It is a base for further progress in R&D, co-operation between the science and industry, development of new technologies. We can say that the e-Infrastructure and its services are the major pillar of the European Research Area (ERA). The stakeholders of the infrastructure are technology providers (industry), service providers (NRENs, National Grids, National HPC Centres), founders (usually Ministries of Science and Research in Member Countries, EC) and the scientists (universities, R&D centres).

The Polish research infrastructure consists of several levels: the networking, computing (HPC, clouds and grids), storage and archivisation and platforms of higher level services for e-Science. The goal is to use it in a common way and provide an added value to scientists because of the fact that the national infrastructure is owned by the PIONIER consortium. Several resources are also integrated with European e-Infrastructures, like GEANT, EGI, PRACE, and data infrastructure which is currently built (e.g. EUDAT). But there are also specific services provided by the national academic network PIONIER, e.g. campus computing, remote archivisation, or education scientific HDTV.

There are many new challenges defined for HORIZON 2020, like environment protection, climate simulations, decreasing of energy consumption, regional tourism development, where we have to use these services of RI.

#### **Intelligent Computations for Complex Problem Solving**

**Barna Iantovics** 

Petru Maior University, Faculty of Sciences and Letters, Department of Informatics and Mathematics, Str. Nicolae Iorga, No. 1, 540088, Targu Mures, Romania ibarna@science.upm.ro

Abstract. Quite often, difficulties in the medical problems solving require different kind of computational intelligence (learning capacity for example) of the systems that must solve them. Many results presented in the specialized literature prove, that the intelligence of a computational system can offer advantages in the problems solving versus a system that does not have such intelligence. The agent-based techniques are many times most appropriate for intelligent computation. Cooperative hybrid medical diagnosis systems seem to be well suited for the solving of many difficult medical diagnosis problems. In this paper a novel hybrid medical system, called ComplexMediSys (Complex Medical Hybrid System) is proposed. ComplexMediSys is a highly complex hybrid system composed from clinicians and intelligent agents that can interoperate intelligently in different points of decision. ComplexMediSys is appropriate for many very difficult medical diagnosis problems solving. We will analyze advantages of the cooperative problem solving by the proposed hybrid system. The human and artificial thinking advantages in the diagnostics elaboration are combined by cooperation. The system is able to handle emergent information that rise during the medical problems solving that allows of improving the accuracy of the elaborated diagnostics. Keywords:complex problem, computational intelligence, intelligent agent, multiagent system, hybrid complex system, knowledge-based system, diagnostic accuracy, medical diagnosis system

#### 1 Introduction

In the healthcare there are many complex problems/tasks (in this paper we consider medical diagnosis problems) [8, 11, 12] that are difficult to be solved by clinicians and/or medical computational systems (medical expert systems for instance). As an example of complex problem, we mention the case of a patient that suffers by two illnesses and between the treatments that must be applied to curing of the illnesses there are some dependencies. For example, an effective medicine for the treatment of an illness has a negative effect to the curing of the other illness.

Difficulties in a medical diagnosis problem solving (the establishment of the diagnostic to the illness/illnesses of a patient) appear based on considerations, like: the solving requires a large amount of data (the distributed medical history of the patient for example); the solving requires heterogeneous problem solving knowledge (deSection 1 Bio-applications

# Modeling of structure, folding and interactions of biomolecules in the era of GPGPU computing

S. Hayryan<sup>1</sup>, M.-C. Wu<sup>1</sup>, C.-K. Hu<sup>1</sup>, Z. Gažová<sup>2</sup> and T. Kožár<sup>2</sup>

<sup>1</sup>Insitute of Physics, Academia Sinica, Taipei, Taiwan <sup>2</sup>Department of Biophysics, Institute of Experimental Physics, Slovak Academy of Sciences, Košice, Slovakia tibor@saske.sk

**Abstract.** The recent boom in general-purpose computing on graphics processing units (GPGPU) facilitates simulations with high demands on computer resources. Such simulations are typical for macromolecules and nanoparticles of biological importance. Several proteins, instead of folding into biologically active 3D structures, aggregate together forming large fibril structures called amyloid aggregates. Amyloids are being extensively studied both experimentally and through computer simulations. Since amyloid aggregates are huge molecular complexes composed from hundreds of thousands of atoms, it is clear that their simulations need supercomputing power. GPGPU-based clusters were shown to offer alternative resources for performing molecular dynamics simulations on nanoscale. We were also using one of the newest docking methodology (the AutoDock Vina program) to model the differences in ligand binding to the native insulin and to the unfolded complexes. In addition, virtual lectin arrays were constructed and high-throughput "*In Silico*" screening was performed in order to select the best binders to the particular galectins.

**Keywords:** GPGPU computing, molecular modeling; protein structure and interactions, amyloid aggregation.

#### 1 Introduction

A large variety of computational methods is available to calculate structural and electronic properties of biomolecules and their complexes. High demand for computational resources is common for almost all first-principle quantum mechanical methods in dependence on the level of study. This can be started from the relatively fast semiempirical methods to the *ab initio* level of solving of the Schrodinger equation (either using Hartree/Fock (HF) or Density Functional Theory (DFT)) to the full configuration interaction (CI) protocol with Moller-Plesset (MP) perturbation theory in between. The size of the molecules (the number of atoms and the appropriate selection of the number and type of base functions describing the atomic orbitals) is a key factor influencing the approximation level of the method used in the computational study of molecular properties and behavior. The computation time of the *ab initio* method scales usually with  $n^4$  where n is the number of the atomic base functions. This scal-

## MedDecSup an Intelligent Distributed Medical Decision Support System

Adrian Gligor<sup>1</sup> and Barna Iantovics<sup>2</sup>

<sup>1</sup> Petru Maior University, Electrical Engineering Department, Nicolae Iorga Street, No 1, 540088, Targu Mures, Romania agligor@upm.ro,

<sup>2</sup> Petru Maior University, Mathematics and Informatics Department, Nicolae Iorga Street, No 1, 540088, Targu Mures, Romania barna\_iantovics@yahoo.com

**Abstract.** Recent scientific literature suggests that an important direction for improving health services may be related to the both use of clinical (medical) decision support systems (CDSs) at the patient and general practitioner or specialist medical doctor level. The paper refers to a pilot project that is intended to be implemented at Tg. Mures to assist at various levels of medical decision. The main objective of the proposed system is to provide a solution for reducing time and costs in the medical assistance process and to improve the medical decisions accuracy. The paper focuses on structure, functionality and advantages brought by the proposed solution.

**Keywords:** Healthcare, Medical Decision Support System, Intelligent Agent, Distributed System, Complex Problem, Artificial Intelligence, Complex Medical System, Medical Informatics

#### 1 Introduction

Traditional Medical (Clinical) Decision Support Systems (CDSs) are designed to assist clinicians with decision making tasks. A medical decision support system uses more items of patient data to generate case-specific advice. The main purpose of CDSs is to assist clinicians at the point of care/decission. A clinician would interact with a CDSs to help determine diagnosis, analysis, etc. of patient data. Recently developed CDSs assist the clinician utilizing both the clinician's knowledge and the knowledge detained by them.

Some developed medical decisions support systems have been proved useful in medical cases were was necessary decision support of the medical personnel. Decision support systems could improve benefits in healthcare increasing the accuracy of solutions and minimizing the human effort [4][5][6][7][10][11]. Recent implementations are limited in intelligent support of very difficult decisions. The next

Section 2 Distributed Computing and Large Scale Applications

## A Technique to Create Efficient Fortran Programs for Heterogeneous Parallel Platforms

Anatoliy Doroshenko<sup>1</sup>, Kostiantyn Zhereb<sup>1</sup>, Iurii Tyrchak<sup>1</sup>, Viet Tran<sup>2</sup>

<sup>1</sup>Institute of Software Systems of National Academy of Sciences of Ukraine, Glushkov prosp. 40, 03187 Kyiv, Ukraine

doroshenkoanatoliy2@gmail.com, zhereb@gmail.com, tyrchak@dep27.isofts.kiev.ua <sup>2</sup>Institute of Informatics, Slovak Academy of Sciences viet.ui@savba.sk

**Abstract.** In the paper we describe our approach of transforming existing sequential Fortran programs into their parallel equivalents. We use rewriting rules technique to automate transformation process. Sequential source code is transformed into parallel code for one of target platforms: shared-memory parallel systems (such as multicore processors), distributed memory systems (e.g. clusters) and GPU-based systems. Parallelizing and optimizing transformations are formally described as rewriting rules which facilitates their reuse. Using high-level algebraic models allows describing program transformations in more concise manner. Performance measurements demonstrate high efficiency of obtained parallel programs.

**Keywords:** rewriting rules technique, algebraic program models, heterogeneous platforms, multicore processors, Fortran, OpenMP.

#### **1** Introduction

Despite being one of the first programming languages, Fortran is still widely used, in particular for solving scientific and engineering computation-intensive problems. Its popularity is due to its relative simplicity and lack of complex facilities (e.g. pointers), closeness to mathematical description of problem and efficiency of generated binary code. Another reason for continued use of Fortran is that in more than 50 years of its existence a vast repository of programs, libraries and routines for solving different scientific problems has been created. Algorithms implemented in such programs are still valuable, however there is a need to adapt this legacy code to new computational platforms, especially parallel systems such as clusters, Grids, cloud computing systems. Furthermore, due to size and complexity of existing code, manual adaptation is not a practical option: there is a need of automated tools to facilitate conversion of legacy code to modern parallel platforms [2], [3].

In this paper we describe our approach of transforming existing sequential Fortran programs into their parallel equivalents. We use rewriting rules technique to automate transformation process, similar to our previous work [6]. Sequential source code is transformed into parallel code for one of target platforms: shared-memory parallel systems (such as multicore processors), distributed memory systems (e.g. clusters)

# On the performance of vector calculations on CPU and GPU

Karol Grondžák<sup>1</sup>, Penka Martincová<sup>1</sup> and Ondrej Šuch<sup>2</sup>

 Faculty of Management Science and Informatics, University of Žilina, Univerzitna 1, 010 26 Zilina
 Joint Institute of University of Matej Bel and Mathematical Institute of Slovak Academy of Sciences, Dumbierska 1, 974 01 Banska Bystrica

Karol.Grondzak@fri.uniza.sk
Penka.Martincova@fri.uniza.sk
ondrej.such@gmail.com
http://www.fri.uniza.sk

**Abstract.** Modern processing devices manufacturers (either Central Processing Units (CPU) or Graphical Processing Units (GPU)) are oriented towards producing multiple core devices. It is a natural trend for improving overall performance of the devices.

In this paper we will briefly summarize the possibilities to employ modern processors (either CPU or GPU) for scientific calculations. On the problem of simple vector calculation we demonstrate step–by–step procedure to design efficient GPU algorithm which outperforms the CPU algorithm.

Key words: vector calculation, CUDA, parallel computing

#### 1 Introduction

Modern processing devices manufacturers (either Central Processing Units (CPU) or Graphical Processing Units (GPU)) are oriented towards producing multiple core devices. It is a natural trend for improving overall performance of the devices. Another trend is to unify the access to all computing devices available in a computer, e.g. CPUs, GPUs and others.

To utilize the modern processors architecture, the developers must understand it and embrace it. Despite some similarities, the CPU and GPU architecture differ. In the next paragraphs we compare them and note significant properties of both CPU and GPU units.

Modern CPU designers improve performance by creating multicore designs. From the point of view of the parallel processing multicore setup can be described as shared memory system. All cores access the same main memory. Such architecture allows relatively simple data sharing and synchronization. To utilize this architecture, the consortium of producers has proposed a scalable, multi-platform, parallel programming Application Programming Interface (API) - OpenMP ([1]).

### Analyzing Performance of Parallel Algorithms for Linear System Equations

## <sup>1</sup>Hanuliak Peter, <sup>2</sup>Hudik Martin

<sup>1</sup>Polytechnic institute, Dubnica nad Vahom, Slovakia, phanuliak@gmail.com <sup>2</sup>University of Zilina, Faculty of Control and Informatics, Zilina, Slovakia, hudikm@gmail.com

**Abstract.** In many cases the most effective way how to increase the performance of parallel algorithms (parallel applications) is to use the parallel principles. Therefore the paper describes the developing steps of parallel algorithms and then it summarised the basic concepts for parallel complexity of linear system equations.

Current trends in high performance computing (HPC) and grid computing (Grid) are to use networks of workstations (NOW) as a cheaper alternative to traditionally used massively parallel multiprocessors or supercomputers. For effective use of parallel machines it is crucial to know it limitations (bottlenecks). This paper describes basic communication bottlenecks as they are in many parallel systems the limiting factor of the maximum rate of parallelization. The paper focused on parallel algorithms for linear system equations, specifically exact methods.

In development process the decomposition technique plays the key role for the effective parallelization of sequential algorithm. Choosing the right decomposition method is dependent on various factors such as run-time environment, parallel machine and algorithm. Usually development process result in testing where the results shows how efficient and scalable the algorithm is. The fastest sequential algorithm for a given problem is the best sequential algorithm. But determining the best parallel algorithm is considerably more complicated. A lot of variables are needed to be taking in mind. Isoefficiency analysis helps us determine the best algorithm/architecture combination for a particular problem without explicit analysing all possible combinations under all possible conditions.

#### Keywords

parallel computer, parallel algorithm, performance modelling, system of linear equations, complexity

#### 1 Introduction

The development of computers has shown that increasing computing performance by increasing the performance of one processor is not effective and so parallel organizations of processors, cores or independent computers and a use of various forms of parallel processes [4, 9, 11]

## Remarks on GPU usage in cryptanalysis

Matúš Jókay, Milan Plančík, Pavol Zajac \*

UIM FEI STU, Ilkovičova 3, Bratislava, Slovakia

**Abstract.** In this paper we compare the effectiveness of GPU computations in the cryptanalysis. In the first part, we examine the speedup of the brute force attack on AES. In the second part, we compare the brute-force attack with the algebraic cryptanalysis approach. Although the algorithms used in algebraic cryptanalysis are relatively unsuitable for GPU usage, we show that specific optimization techniques can make GPU platform competitive even in this case.

#### 1 Introduction

Cryptanalysis is, simply said, a process of extracting secret information from the cryptosystem by the unauthorized user. In a basic scenario (known-plaintext attack), the attacker knows a ciphertext, and a corresponding plaintext, respectively, and wants to compute an unknown key. The simplest form of cryptanalysis is the extraction of the key by guess and verify approach (the brute-force attack). It is not possible to prevent this kind of attack. In the present era of strong cryptography it is often the only real possible attack available for most of the attackers. The attack can be especially devastating, if the encryption key is derived from a weak password. For the attacker, it is crucial to employ as much computing power as possible for his given budget. Practical attack possibilities are demonstrated in [3]. We note that the guess and verify approach is very suitable for (SIMD-type) parallelization, and it is also well suited for the computing on GPUs (graphic cards).

A special consideration in the present is focused on the so called algebraic cryptanalysis (ACA) [9]. As an alternative to brute-force approach, we can try to use advanced algebraic techniques to recover the secret key. The cryptanalytic problem is transformed into a problem of solving a system of non-linear Boolean equations (which is NP-hard). Although the complexity of the attack is not known, we know it is upper bounded by the complexity of the brute-force attack. If we guess the key bits (or some of them), and substitute the guessed bits it in the system of equations, the problem becomes easy to solve (in extreme case, it can be solved in linear time by the spreading of constants). We can trade some complexity between guessing, and algebraic part. It is also important to use parallel computing to be able to verify more guesses. It is then crucial to implement algebraic part as fast as possible. However, due to different algorithms used

<sup>\*</sup> This material is based upon work supported under the grant VEGA 1/0244/09, and by the project "GPU Výpočty" funded by Slovak University of Technology ("Projekt na podporu mladých výskumníkov", 2010).

## **GPU-accelerated parallel genetic algorithms**

Andrej Kozák, Ivan Hujsi, and Jozef Rodina

Faculty of Electrotechnics and Informatics, Slovak University of Technology, Bratislava, Slovakia {andrej.kozak,ivan.hujsi,jozef.rodina}@stuba.sk

**Abstract.** Computational and memory demanding algorithms and applications often require solutions utilizing all available resources in order to complete in acceptable time. Parallel computing offers an approach to solve such problems by processing independent tasks at the same time. Since GPU exceeds CPU with number of cores and so enables massive parallelization, GPGPU (General-Purpose Computing on Graphics Processing Units) is becoming more popular in the field of parallel processing.

This paper deals with parallel genetic algorithms (PGA) and the possibility to optimize them for processing on GPU. The need to process a great amount of data similarly makes GA suitable for GPGPU to reduce the most time consuming part of fitness calculation. The focus is on the crucial aspects limiting the performance of the algorithm – data transfers and resource utilization. Based on an implementation of synchronous PGA, the asynchronous approach is introduced as a possible way, how to fix the weak point discovered in the synchronous solution.

Keywords: GPU, GPGPU, genetic algorithms, CUDA, accelerating.

#### 1 Introduction

The power of GPU platform pushed GPU far past the borders of graphic processing, therefore many applications and algorithms can now benefit from the speedup on the GPU. Achievements and significant advances in parallel computing and GPGPU force developers to port their applications on the GPU platform. However, there is no measurement tool to calculate the performance of any algorithm on the GPU and the suitability to solve the particular problem can only be predicted.

#### 2 **GPGPU** overview

SIMD architecture of the GPU designates the applications to organize bulks of data to be processed on the GPU at the same time whilst the CPU manages this process. The crucial aspects that should be taken into consideration when designing or optimizing the application for GPU are data transfers and program

Section 3 Business benefits and challenges of grid in Financial Markets

# Identification of financial contagion outbreaks through distributed systems

Bradut-Vasile Bolos, Adrian Gligor,

<sup>1</sup> Petru Maior University, Finance and Accounting Department, Strada Nicolae Iorga nr 1, 540088, Targu Mures, Romania bradut.bolos@ea.upm.ro,

<sup>2</sup> Petru Maior University, Electrical Engineering Department, Strada Nicolae Iorga nr 1, 540088, Targu Mures, Romania agligor@upm.ro

**Abstract.** As financial contagion spreads throughout the business network, so does the information regarding economic hard facts, but at a slower speed. The paper explores means to unravel and disarm rumors and false information and present decision makers with options to tackle the contagion. The delay between fundamentals generation and market attitude can be a solution to disarm contagion, and a mix between advanced accounting and distributed systems are the means.

Keywords: Financial contagion, economic epidemiology, financial crisis, distributed systems

#### 1 Introduction

The economic and financial crises are often seen as rare and abnormal conditions of the financial system. However, history shows that they are neither rare nor abnormal. The blueprint for such crisis was set a long time ago by the so called "tulip mania" in Netherlands, in 1637, and occurrences of documented cases are present ever since, in different countries and different marketable assets.

There is a wide spread debate regarding causes of crisis, ranging from human nature related issues, to systemic issues, social order issues, and complex intercorrelations. We do not intend to approach reasons in this study as they change from case to case. Instead, our research focuses are crisis patterns and usage of modern technology and science for detecting, researching and countering the negative effects.

The hypothesis in our research is that any asset can be subject of over-valuation, any risk can be under-evaluated, and recent events tend to gather more credibility than more distant past events, leading to repetitive heard behavior based upon them.

Any "treatment" of crisis is based on valid information and clear focus on sensitive paths of contagion. In modern times, information overwhelms decision-makers; as a consequence focusing attention becomes more and more critical. As technology

## Endogenous and Exogenous benefits of grid in Financial Markets

Călin-Adrian Comes

Petru Maior University, Nicolae Iorga, 1, Tîrgu-Mureş, 540088, MUREŞ, Romania calin.comes@ea.upm.ro http://www.upm.ro

Abstract. Value at Risk - VaR [4] represent a measure of portolio to different risks in financial environment and can be used for portfolio optimization process. Endogenous and Exogenous Grid for Financial Markets - EEGFM use almost Kirchhoff law for one node  $\mathcal{N}_j$ , almost Kirchhoff law for one net  $\mathcal{N}$  in idea to simulate one stock market  $\mathcal{SM}$  starting from real time series data, respectiv almost Kirchhoff law for one net of m stock markets  $\mathcal{SM} = \oplus \mathcal{SM}_m$ .

**Keywords:** Endogenous, Exogenous, Portfolio, Grid, Financial Markets, EEGFM

#### 1 Introduction

Business Risk can be defined as the volatility of unexpected outcomes, witch represent the value of the assets [12]. Exogenous Business Risk - ExBR could be from demographical, social, political, financial nature as:

- fixed exchage rate system broke in 1971;
- oil price shocks from 1973;
- black monday, october 19, 1987, U.S. stocks collaps 23 %;
- asian turmoil 1997, failure of a big hegde fund;
- october 11, 2011, terrorist attack in WTC.

Endogenous Business Risk - EnBR came from inside organization:

- capacity planning;
- resource(financial, assets, human, data) optimization;
- corporate governance.

In idea to avoid the exogenous and endogenous business risks we propose a simulation of Financial Markets in grid environment.

Grid computing has led researchers and developers to apply the technology on different range of domains such as bioinformatics [8], [13], [14], [3] quantum chemistry [2], [10] high energy physics [5], [15], neurosciences [16], [7], language processing [17], earth sciences [1], and weather forecasting [6].

## Financial Analysis Computing Architecture in Distributed Environment

Marián Zvada<sup>1,2</sup>, Marek Domiter<sup>3,4</sup>, Vladimír Motoška<sup>3</sup>, Lukáš Slebodník<sup>1</sup>, and Tomáš Kasanický<sup>1,5</sup>

<sup>1</sup> FURT Solutions, s.r.o., Košice, Slovak Republic
 <sup>2</sup> Karlsruhe Institute of Technology, Karlsruhe, Germany
 <sup>3</sup> SORS Security, s.r.o., Košice, Slovak Republic
 <sup>4</sup> Faculty of Electrical Engineering and Informatics, TU Košice, Slovak Republic
 <sup>5</sup> Institute of Informatics, SAS, Bratislava, Slovak Republic
 {zvada,domiter,motoska,slebodnik,kasanicky}@sors.com

**Abstract.** Nowadays, any computing infrastructure designed for specific problem solving faces big expansion of a modern computing approach which becomes an issue from different aspects like scalability, reliability, data consistency, storage and transparency to users/clients. In this paper we introduce architecture of a flexible heterogeneous computer system that was created for support of the analysis in the financial area. We will describe in detail technologies which were used for the creation of the presented specific computing environment, we will also provide detailed description of chosen hardware and software solution.

**Keywords:** grid computing, cloud computing, financial market behaviors, financial analysis, condor, dynamic resource allocation, RedHat, MRG

#### 1 Introduction

Instead of building own huge computing centers with computational resources, most of statistical analysis of any kind can use grids and clouds in various areas of the commercial business, industry or research. One of the biggest consumers of such environment is LHC Computing Grid [1]. We would like to concentrate on well-known computing techniques introduced in last decade [2] and used by many scientists, where designers/developers didn't think deeply about enabling such a computational power for multiple purposes and people are still reluctant inherit from this idea due sophisticated technology back-end. Many applications are surely hardware and architecture dependent, but most of them can be driven by open source projects replacing heavy commercial products often limiting exploration of new ideas in the research community by its complex design and license policy. Thus, we provide architectural approach to model environment with high elasticity for solving complex problems, where we primary choose as application domain statistical and computational finance.

## Recent literature and directions in non-parametric methods for assessing bank performance through efficiency analysis of decision making units

Anca Munteanu<sup>1,2</sup>

<sup>1</sup> Petru Maior University, N. Iorga Street, No 1 Mures, Romania anca.munteanu@ea.upm.ro
<sup>2</sup> Bucharest Academy of Economic Studies, Mihail Moxa Street, No5-7 Bucharest, Romania

**Abstract** This paper presents a review of the recent neural network literature concerned with the problem of efficiency assessment of banking activity. Artificial neuronal networks algorithms (NN) and Data envelopment analysis (DEA) represent two of the most prominent methods that can be used to represent the relationship between input and output variables of the production process. Our aim is to highlight some important aspects regarding multi-criteria performance assessment in the context of efficient resource utilization and also to point out some future research directions by establishing best practice guidelines.

#### Introduction

A great deal of attention is paid to the performance of banks due to the fact that banks are seen as special given their major role in providing credit to enterprises. The role of banking institutions in the process of reallocation of financial resources is even more important if other elements of the financial sector are underdeveloped. Thus, in this situation banks contribute in a larger scale to the optimal allocation of financial resources in the real sector.

In this context, the evaluation of organizational performance represents a key element considering the importance of generated information regarding competitive advantages and future development of an enterprise.

Non-parametric methods are defined by the fact that no assumptions are made concerning the functional form that links the inputs and outputs used to describe an operating process. This paper discusses recent approaches in assessing bank performance through efficiency analysis. Artificial neuronal networks algorithms (NN) and Data envelopment analysis (DEA) represent two of the most prominent methods that can be used to represent the relationship between input and output variables of the production process.

The conclusions focus on strengths and opportunities offered by NN algorithm in terms of practical research dilemma in the field of bank performance assessment through qualities such as robustness and flexibility.

Section 4 High Performance Distributed Computing and Large Scale Simulations

# Augmented reality application in parallel computing system

Branislav Sobota<sup>1</sup>, František Hrozek<sup>1</sup>, Martin Varga<sup>1</sup>, Csaba Szabó<sup>1</sup>, Štefan Korečko<sup>1</sup>, Peter Ivančák<sup>1</sup>

<sup>1</sup> Technical University of Košice, Faculty of Electrical Engineering and Informatics, Department of Computers and Informatics 042 00 Košice, Letná 9, Slovakia {branislav.sobota, frantisek.hrozek, martin.varga, csaba.szabo, stefan.korecko, peter.ivancak}@tuke.sk

**Abstract.** This paper describes some parts of semi-markerless augmented reality (AR) application based on recognition of selected patterns of square or rectangular shape. The paper also presents possible directions where parallelization can be implemented (fine-grained or coarse grained) because some stages of this AR application needs a lot of computational power. This problem can be solved with parallel computing environment based on cluster or GPGPU technologies. Such as distributed computer system called D-GTS which aims on GPGPU was developed at Department of Computers and Informatics at Technical University of Košice.

Keywords. Augmented reality, semi-markerless system, parallel computing.

#### 1 Introduction

There are several technologies applicable when creating VR applications. One of these technologies is mixed reality. Mixed reality (MR) is an area of computer science research that aims at combination of real world with computer generated data (virtual reality). Such computer generated graphical objects are blended in to real environment in real time. Real world data are provided through sensors in real time back to mixed reality system [1]. Mixed reality can according to [2] originate from at least one of these technologies: augmented reality (AR) and augmented virtuality (AV).

AR environment consists of real world components as well as virtual (synthetic) components. For example person that uses AR system is equipped with human interface (semitransparent glasses, head mounted display, data display helmet or combination of monitor and camera). Via this interface person sees real world with superimposed computer generated objects on top of real ones.

Augmented virtuality is technologically similar to augmented reality. However its principle is opposite to AR. Most of a world (scene) that is displayed is virtual and real objects are inserted in the world. If observer in inserted into scene he or she is

### An agent-based security approach for Intrusion Detection Systems

Bogdan Crainicu<sup>1</sup>, Barna Iantovics<sup>1</sup>

Petru Maior University, Department of Mathematics and Informatics, Nicole Iorga street, No. 1, 540088, Tg. Mures, Romania {cboqdan@upm.ro, barna iantovics@yahoo.com}

Abstract. An Intrusion Detection System (IDS) monitors network traffic and user activity in order to detect malicious traffic and abnormal activity that violates security policies of an organization. This paper presents a proposed intelligent agent-based security approach for deploying IDSs in a distributed environment. A decentralized, agent-based IDS allocates tasks to agents for collecting, analyzing and delivering together data needed for recognizing threats and for taking effective actions. The advantages of this approach are scalability, coping with network latency and load, no single point of failure, representing a very suitable IDS model for large and heterogeneous network environments. Keywords: intelligent agent, multiagent system, intrusion detection system, distributed system

#### 1. Introduction

An intrusion detection system is usually a software application that monitors network and/or system activities for malicious activities and produces reports to a station [6]. IDSs typically record information related to observed events, notify security administrators of important observed events, and produce reports [6]. Intrusion detection systems are primarily focused on identifying possible incidents, logging information about them, and reporting unwanted attempts. Some organizations use IDSes for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies [6].

This paper proposes an intelligent agent-based security approach for deploying IDSs in a distributed environment. A decentralized, agent-based IDS allocates tasks to agents for collecting, analyzing and delivering together data needed for recognizing threats and for taking effective actions.

The upcoming part of the paper is organized as follows: Section 2 describes the intelligent agents as a solution for many problems solving, Section 3 presents some Intrusion Detection Systems, In Section 4 our proposal is described, Section 5 outline the advantages of the solution proposed by us.

## Matrix Multiplying in Virtual Address Space on the 32/64-bit Computer Architecture for Simulation on the HPC

Igor KVASNICA<sup>1</sup>, Peter KVASNICA<sup>2</sup>

 <sup>1</sup> Regional Department for Environment Issues of Trenčín, Hviezdoslavová č. 3, 911 33 Trenčín, Slovak Republic, kvasnica.igor@tn.kupz.sk
 <sup>2</sup> Alexander Dubček University in Trenčín, Center of information technologies and Faculty of Special Technology, Department of Informatics, Študentská č. 2, 911 50 Trenčín, Slovak Republic, peter.kvasnica@tnuni.sk

**Abstract.** The paper deals with the method of multiply matrix for simulation in high performance computing or large simulation in address space. A matrix form is often employed for definition tasks in science and engineering. Availability this architecture is certainly an area, which is essential to parallel computing. The result of simulation of mathematical models of the economic, control, climate changes and etc. can be created using parallel computer architecture. Users priorities associated with the a simulation matrix multiply requiring them essential to maximize hardware utilizing return on investment.

**Keywords:** multiplying matrix, operating system, 32-bit and 64-bit technologies, address space, virtual address space, mathematical model.

#### 1 Introduction

An  $m \ge n$  matrix is a rectangular array of m n numbers arranged in m rows and n columns and is written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = a_{ij}$$

An *m* x *n* matrix may therefore be considered as made up of *m* row vectors  $a_i$  (i = 1, 2, ...m) or of *n* column vectors  $a_j$  (j = 1, 2, ...n) [1].

The product of an  $m \cdot p$  matrix A with an  $p \cdot n$  matrix B is the matrix C, whose elements are given by

## Simulation of Uranus and Neptune formation using the new Asu-SAV computer cluster

Luboš Neslušan, Marián Jakubík

Astronomical Institute, Slovak Academy of Science, 05960 Tatranská Lomnica, Slovakia ne@astro.sk,mjakubik@astro.sk

**Abstract.** We describe astrophysical background and computational demands of our study of the Uranus and Neptune formation, which still remains a puzzle in the cosmogony of the Solar System. The simulations of the last stage of this formation were performed using both small, 32 CPUs (with hypertrading), cluster and larger, 128 CPUs, cluster recently installed at the Astronomical Institute of the Slovak Academy of Sciences (Asu-SAV computer cluster). Our experience with the usage of these facilities for a large series of sequential jobs as well as some experiments with parallel tasks is discussed.

**Keywords:** planetary systems – planets and satellites: formation – planets and satellites: individual: Uranus, Neptune – protoplanetary disks cluster computing – grid computing – parallel computing

#### 1 Introduction

The accretion of Uranus and Neptune is a long-standing problem in planetary science. In the past, Safronov (1969) was the first to point out that the accretion of these two planets from a planetesimal disk at their current locations would have taken implausibly long timescales. This problem was confirmed by Levison et al. (2001) using modern numerical simulations.

There is now a consolidated view that the giant planets were closer to each other in the past (probably all within 12 AU from the Sun) and that they moved to their current orbits after their formation (e.g. Tsiganis et al., 2005). Thus, it is no longer necessary to construct a model capable of explaining the formation of Uranus and Neptune at their current, remote locations.

Forming Uranus and Neptune within 12-15 AU from the Sun is in principle easier than forming them at 20-30 AU because the density of solid material was probably higher and the dynamical timescale (i.e. the orbital period) was shorter. However, forming 10-15 Earth mass (M<sub> $\oplus$ </sub>) cores from a planetesimal disk turns out to be difficult at *any* location.

In this paper, we present the results of our attempts to form the Uranus and Neptune in early Solar System history, using a N-body simulations with taking into account the migration of giant planets (Jupiter and Saturn) and also the influence of gas disk. These simulations was performed on our computational

## Acceleration coefficients impact of the Particle Swarm Optimization algorithm in solving of the scheduling problem using MATLAB Distributed Computing Server

#### Ján Zelenka

Institute of Informatics / Slovak Academy of Sciences, Dúbravská cesta 9, 845 07 Bratislava, Slovak Republic { jan.zelenka@savba.sk }

**Abstract.** Currently, optimization processes requirements focusing on several parameters are emphasized. Algorithms allowing to find an optimal (near-optimal) solution, are in most cases moving in the large area of possible solutions. This is why stochastic search techniques such as evolutionary algorithms, an artificial immune system or particle swarm optimization algorithm have found application in scheduling field. Their application to the scheduling process reduces computation time, but at the cost of finding the optimal solution. Their running requires strong computational support and hunger solution programs to run on multi-core workstation, clusters, grid and clouds. In this article serial and parallel computing of the scheduling problem by using MATLAB Distributed Computing Server is compared. The second part of this article refers the PSO algorithm acceleration coefficients influencing on the solution. The simulation with different coefficients is compared.

Keywords: particle swarm optimization algorithm, MATLAB distributed computing server.

#### 1 Introduction

Stochastic search techniques such as evolutionary algorithms, an artificial immune system AIS or particle swarm optimization algorithm PSO have found application in various fields (microstrip filter dimension designing [17], scheduling field [1], [2], [4], [6] etc. ). The problem of scheduling is concerned with searching for optimal (or near-optimal) schedules subject to a number of constraints. A variety of approaches have been developed to solve the problem of scheduling. The principles of several dynamic scheduling techniques, their application and comparisons, namely dispatching rules, heuristics, meta-heuristics, artificial intelligence techniques and multi-agent systems are described in many publications [8]. A multi machine jobshop scheduling problem is to assign each operation to a machine and to find a sequence of jobs (operations) on machines that the maximal production time is minimized [5]. Scheduling is defined as the allocation of resources to jobs over time. It is a decision-making with the goal of optimizing one or more objectives [8]. The objectives can be the minimization of the completion time of jobs (makespan), mean

## Section 5 Research Infrastructures

## Parallel Computing in EGI

V. Šipková, M. Dobrucký, and P. Slížik

Ústav informatiky, Slovenská akadémia vied 845 07 Bratislava, Dúbravská cesta 9 http://www.ui.sav.sk/ {Viera.Sipkova, Miroslav.Dobrucky, Peter.Slizik}@savba.sk

**Abstract.** EGI.eu is a foundation established in February 2010 to create and maintain a *pan-European Grid Infrastructure* (EGI) so as to guarantee the long-term availability of a generic e-infrastructure for all European research communities and their international collaborators. Its work builds on previous EU-funded grid projects: LHC, DataGrid, EGEE, a.o. EGI does not develop the software deployed in the grid infrastructure, all upgrades and new programs are produced with external technology providers. Concerning the compute area, the major highlights of the first release of the EGI middleware - EMI 1, are improvements in the CREAM service, extensions in the JDL language, and the support for the user defined fine-grained mapping of processes to physical resources. This makes possible for grid applications to employ various parallel programming models. This work presents the overview of the current job management facilities provided by the EGI middleware components.

#### 1 Introduction

**EGI.eu** is a foundation established in February 2010 to create and maintain a *pan-European Grid Infrastructure* (EGI) [1], in collaboration with National Grid Initiatives (NGIs) and European International Research Organizations (EIROs), so as to guarantee the long-term development, availability and sustainability of grid services and e-infrastructure for all European research communities and their international partners. Its work builds on previous EU-funded projects which raised this goal from the initial concept of a scalable, federated, distributed computing system.

The distributed computing grid was originally conceived in 1999 to analyze the experimental data produced by the particle accelerator Large Hadron Collider (LHC) at CERN (European Organization for Nuclear Research). The research and development of grid technologies started in January 2001 within the European Data Grid [2] project which proved the successful application of the grid in research fields of the high energy physics, earth observation and bioinformatics. Upon its completion in 2004, a new project, called Enabling Grid for E-science (EGEE) [3], took over the grid's further development. EGEE allowed researchers the access to computing resources on demand, from anywhere in the world and at any time of the day. By April 2010 when the last project phase

## Facility of VOCE Grid Infrastructure for Spatial Data Computing

Vladimír Siládi $^1$  and Ladislav Huraj $^2$ 

<sup>1</sup>Matej Bel University, Faculty of Natural Sciences Tajovského 40, 974 01 Banská Bystrica, Slovak Republic vladimir.siladi@umb.sk,
<sup>2</sup>University of SS. Cyril and Methodius in Trnava, Faculty of Natural Sciences Nám. J. Herdu 2, 917 01 Trnava, Slovak Republic ladislav.huraj@ucm.sk

Abstract. Nowadays, state of computer technology offers significant opportunity to achieve high computing performance. This possibility are used to solve numerous problems which need this high computing performance. The geoprocessing and geographical information systems (GIS) is such an area of computing science research which are able to use these possibilities. An interpolation is one of fundamental task in GIS. The interpolation algorithms can be performed on grid infrastructure. This infrastructure represents powerful tool of high-performance computing (HPC). However, the facility of computers and clusters aggregated in grid differ. In paper, we present a comparison of offered clusters, which are aggregated in VOCE grid infrastructure. The inverse distance weight interpolation is used as a criterion for the comparison. The outputs of interpolation are raster files for GIS Grass visualization.

**Keywords:** Grid computing, Inverse distance weighting, Interpolation method, Geographical information systems

#### 1 Introduction

Geographical information systems manipulate with huge amounts of spacial data, for example data in climatology [7]. Data exactly measured in terrain represent the smaller part of the visualized data due to unreality to place gauging stations in all ground points, which are projected into pixels of the raster map. Other data are usually obtained by interpolation methods. These data represents the major part of the data set. Spatial interpolation is a fundamental task in the geoprocessing performed by GIS. A common characteristic of the geospatial interpolation methods is time-consumption. Apart from variety of the methods, the current state of the art in computer technology, especially massive parallel hardware offers several possibilities to speed up the processing, from multicore processors through graphics processing units (GPUs) to supercomputers and computer clusters. These partial segments of high performance computing could be aggregated in a high-performance computer grid infrastructure. Grid

## A Genetic Algorithm for the Multiple Depot Periodic Vehicle Routing Problem

Peter Kurdel, Jolana Sebestyénová

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

**Abstract.** Deriving the optimal cash deployment strategy for a network of ATMs involves the analysis of logistics costs, inventory policies as well as of the routing of replenishment vehicles. The optimal cash deployment strategy focuses on the reduction of cash-related expenses provided that ATMs do not run out of cash. Parallel computing technologies offer engineers the means to accelerate solutions of their computational problems by using multiple hardware resources. The ability to solve very large problems by scaling computer programs to run on multi-core workstations, clusters, grids, and clouds can help engineers gain significant research and competitive advantages.

#### 1 Introduction

Parallel computing technologies offer engineers the means to accelerate solutions of their computational problems by using multiple hardware resources. The ability to solve very large problems by scaling computer programs to run on multi-core work-stations, clusters, grids, and clouds can help engineers gain significant research and competitive advantages.

Deriving the optimal cash deployment strategy for a network of ATMs involves the analysis of logistics costs, inventory policies as well as of the routing of replenishment vehicles. Thus the problem is twofold requiring first a conceptual framework to derive the optimal cash deployment strategy for a network of ATMs and second an assessment of potential benefits of sophisticated cash management software. Given the state of the ATM industry, the optimization objective is clearly to minimize costs.

Consequently, the optimal cash deployment and replenishing strategy focuses on the reduction of cash-related expenses provided that ATMs do not run out of cash. The study [5] develops a conceptual framework to derive the optimal cash deployment strategy for a network of ATMs and assesses potential benefits of sophisticated cash management software: logistics costs, inventory theory, routing (Travellingsalesman problem, Vehicle routing problem).

One of the most common problems related to logistics and inventory theory is the design of routes for vehicles and goods. The operations research literature distinguishes between designing optimal routes that visit a set of given places (node-covering) and routes that traverse all the edges (edge-covering).

The Vehicle Routing Problem (VRP) is a well known problem in operational research where customers of known demands are supplied by one or several depots. The

## Special session CRISIS Crisis situations predictions and solutions

## Computer Modeling of Forest Fires on Complex Terrain

Jan Glasa, Peter Weisenpacher, Ladislav Halada

Institute of Informatics, Slovak Academy of Science, Dubravska cesta 9, 84507 Bratislava, Slovakia {Jan.Glasa, Peter.Weisenpacher, Ladislav.Halada}@savba.sk

**Abstract.** In this paper, the use of two advanced fire simulators for forest fire simulation in complex terrain is discussed. Results of post-fire reconstruction of a forest fire in complex, hardly accessible conservation areas in Slovak Paradise National Park by the system FARSITE and air flow simulation by the system FDS indicate a good potential of both fire simulators for forest fire simulation in specific Slovak conditions.

Keywords: Forest fire simulation, complex terrain, wind, FARSITE, FDS.

#### 1 Introduction

Forest fires and fires in wildland-urban interface (WUI) often cause great damages of nature and property and threaten people's lives. Society also suffers from bad devastation of natural landscape scenery and damages of environment and nature ecosystems in regions afflicted by fire.

Advantages in computers and information technologies stimulate the research and development of computer systems capable to simulate fire spread and spatial and temporal fire behaviour to support crisis management decisions. There are many difficulties which complicate forest fire modeling by computers, such as complicated geometry of forest and large amount of hardly accessible fuel information, complexity of description of chemical and physical dynamics of combustion, capturing meteorological conditions and their dependence on fire-induced air flows, computational complexity of numerical solution of differential equations systems, etc.

Several advanced simulators for forest fires and fires in WUI have been developed. Mostly they are based on empirical or semi-empirical fire spread models suitable for simulation of more extensive forest fires (on dozens or hundreds of hectares) which are developed generally for currently available computers. Semi-empirical models describe fire spread by functions obtained by approximation of experimental data gathered by laboratory and field experiments. They do not include complete physical mechanism of fire and do not involve non-linear physical dynamics inherently connected with fire (e. g. abrupt eruptions). Such models do not take into account fireatmosphere interaction. Therefore, they are not suitable to capture specific fire behaviour in strongly heterogeneous environment, but are efficient for extensive

## Techniques of Parallel Data Mining for Numerical Values Prediction

Peter Krammer<sup>1</sup>, Ladislav Hluchý<sup>1</sup>

<sup>1</sup> Institute of Informatics of the Slovak Academy of Sciences Dubravska cesta 9 84507 Bratislava, Slovakia

peter.krammer@savba.sk, ladislav.hluchy@savba.sk

**Abstract.** Presented paper in data mining area is focused on the prediction of numerical attribute issues in the hydrology. Models based on feed-forward neural networks consisting of perceptrons are used for training and predictions. The paper presents several strategies of parallelization, suitable for training of neural networks, applicable for implementation in Grid and cluster infrastructures. This paper describes the results of the experiments; it also compares the quality of a centralized-trained global model (trained on all available data), with the quality of models created by aggregation of submodels trained over divided data. Finally, it provides further opportunities for parallelization on computing nodes level, with description and comparison of achieved results.

**Keywords:** Data Mining, Neural Networks, Perceptrons, Parallel Training, Numerical Prediction, Hydrology

#### **1** Introduction

Modeling and prediction of numerical variables is encountered in engineering practice very often. They occur in different sectors, such as ecology, hydrology, meteorology, economics, physics, chemistry, etc. In general, both the input attributes and the target attribute are numeric values represented by real numbers. Tasks of this type can be described by many different models - regression trees, functional regression, K-nearest neighbours, isotonic regression, Gaussian process and many others.

One of the best models is the neural network model, but its training process requires excessive time. Overall, the time aspect of data mining is often a significant factor. For these reasons, parallelization looks like the best choice in many cases, but parallel training of neural networks still remains the key issue. Several studies define

## **Object-oriented Approach for Cloud Abstraction Layer**

Binh Minh Nguyen<sup>1</sup>, Viet Tran<sup>1</sup>, Ladislav Hluchy<sup>1</sup>

<sup>1</sup> Institute of Informatics, Slovak Academy of Sciences Dubravska cesta 9, 845 07, Bratislava, Slovakia {minh.ui, viet.tran, Ladislav.Hluchy}@savba.sk

**Abstract.** In this paper, we will present an object-oriented approach for abstraction of cloud resources. The abstraction will allow users to manipulate virtual machines as objects and simplify the process of porting applications to cloud computing.

Keywords: cloud computing, interoperability, abstraction, object-oriented programming

#### **1** Introduction

In recent years, the term "Cloud Computing" has been mentioned many times in scientific papers and research works. The Cloud services also have been extensively used in every field of daily life. However, there are no definitions of Cloud Computing seems to be widely accepted yet, although this question has already been tried to address by several authors [1], [2] and [3]. Conceptually, Cloud Computing is the conjunction of key features like high availability, flexibility, elasticity, virtualization resources, pay-by-use, reduce total cost for both users and providers. With Cloud technologies, the illusion of unlimited resource that is really interesting for short-term testing and development, as well as for long-term flexible infrastructures. More and more institutions, companies have started to build private or public Cloud for the purpose of use or commerce.

Nowadays, there are several major enterprises like Amazon, Microsoft, Google, ElasticHosts, etc. provide services base on Cloud technologies. Each vendor turns towards one or more diverse type of Cloud. There is also open-source Cloud middleware such as Eucalyptus [4] and OpenNebula [5] as well as proprietary software from VMWare, Citrix, IBM, and so on. Unfortunately, the software are often incompatible with each other that may increase cost of porting applications to Cloud environment, restrict the ability to collaborate between users with other. Moreover, it is difficult for customers to evaluate work performance while using the service of a single provider. There are efforts to standardize cloud middleware, mostly notable by Open Grid Forum with OCCI [6] (Open Cloud Computing Interface).

In this paper, we present an object-oriented approach for abstraction layer of cloud computing. The abstraction layer could simplify the creation and use of virtual machines in cloud, and also make interoperability between providers from the view of users. The abstraction also enables opportunities for creating optimization

## Optimalizácia výkonu klastra podľa aktuálnych potrieb gridu

Miroslav Dobrucký<sup>1</sup> and Ladislav Hluchý<sup>1</sup>

<sup>1</sup> Ústav informatiky SAV, Dúbravská cesta 9, 84507 Bratislava, Slovensko {Miroslav.Dobrucky,Ladislav.Hluchy}@savba.sk

Abstrakt. Výpočtové klastre sú často zaťažované nepravidelne a nárazovo, preto je žiaduce dynamicky prispôsobiť výkon klastra, čiže počet pracujúcich výpočtových uzlov, podľa aktuálnych požiadaviek používateľov. Autor navrhol a implementoval jednoduchý systém pozostávajúci z dvoch procesov volaných pravidelne v určenom čase. Proces zapínania využíva možnosť zobudenia po sieti (wake on LAN), ktorá sa dá povoliť v BIOSe a je spúšťaný v prípade, ak nejaké úlohy čakajú v nejakej fronte dávkového spracovania úloh (PBS). Zapína potrebný počet uzlov ohľadom na veľkosť paralelnej úlohy (MPI, OpenMP a pod.). Proces vypínania vyhľadá voľné uzly, na ktorých nebežia úlohy z dávkového spracovania ani ďalšie procesy (aktualizovanie softvéru administrátorom a pod.) a tie vypne, pričom ponechá z nich ešte pracovať aspoň toľko, aby sa mohli rýchlo vykonať zadané rezervácie (v plánovači MAUI). V závere príspevku autor analyzuje cenu zaplatenú za ušetrenie peňazí za energie, a to veľkosť réžie tohto systému, ktorá predlžuje dobu vykonania niektorých úloh a tiež vplyv tohto systému na gridový midlvér - na možné zníženie atraktivity takto riadeného klastra pre gridové úlohy zadávané cez WMS (workload management system) a na reportovanie poskytnutej výpočtovej kapacity, ktorá sa nerovná potenciálnej kapacite (commitment, čiže ktorá bola prisľúbená).

**Kľúčévé slová:** výpočtový klaster, dávkové spracovanie úloh PBS/torque, zobudenie po sieti wake-on-LAN, dynamické riadenie spotreby a výkonu.

#### 1 Motivácia

V súčasnosti sa čoraz viac uplatňuje virtuálne počítanie "v oblakoch" (cloud), kde si potrebnú výpočtovú silu sám používateľ zakúpi podľa svojich aktuálnych potrieb a finančných možností. Poskytovatelia výpočtových prostriedkov (cloud providers) sa ich istotne snažia efektívne spravovať a zapínať ich podľa trendu okamžitého dopytu. Podobne by sa mala správať aj akademická organizácia pri prevádzkovaní svojho klasického výpočtového klastra, ktorý môže ale nemusí byť zapojený v gride, kvôli finančnej úspore.

Výpočtové klastre sú často zaťažované nepravidelne a nárazovo, čo súvisí s pracovnou dobou, sviatkami, dovolenkovými obdobiami a podobne. Preto je vhodné

### HMM speech synthesizer in Slovak

Sakhia Darjaa, Marian Trnka, Miloš Cerňak, Milan Rusko, Róbert Sabo, Ladislav Hluchý Institute of Informatics of the Slovak Academy of Sciences, Dubravska cesta 9, 845 07 Bratislava, Slovak Republic {utrrsach, trnka, milos.cernak, milan.rusko, robert.sabo, ladislav.hluchy}@savba.sk http://ui.sav.sk

Abstract. The Department of Speech Analysis and Synthesis of the Institute of Informatics of the Slovak Academy of Sciences has already built many types of Slovak speech synthesizers – based on phoneme/subphoneme concatenation, LPC, RELP, Diphone synthesis and Unit Selection approaches. In the effort to introduce emotional load in the synthesized speech the authors decided to create a synthesizer of new generation, based on Hidden Markov modeling (HMM) of spectral features, fundamental frequency contours and phoneme lengths. Using the same speech database, as was used in the Unit Selection synthesizer Kempelen 2.0, a new HMM synthesizer for Slovak language Kempelen 3.0 was developed. The paper presents its emotionally neutral version, evaluates it and compares with the Unit Selection and Diphone synthesizers. Emotional speech modeling using HMM synthesis should offer better results, as the model is capable of representing nuances of timbre changes and prosodic details. Nevertheless the high naturalness of Unit Selection synthesized speech makes this type of synthesizer a strong competitor. The answer to the question which of the approaches is more suitable for the given task will be probably known only after the emotional speech synthesis database is finished and used for building of emotional synthesizers, which is the aim of the following phase of the project CRISIS. Iterative training and testing the HMM models during the synthesizer optimization is highly computationally demanding and memory consuming process, which could only be accomplished by parallel computing techniques. These calculations were realized on the computer cluster of the Institute of Informatics. The paper proposes data parallelization for fast automatic detection of bad speech transcription using the forced alignment technique, and for rapid testing of synthetic voices in terms of speech intelligibility comparison.

#### 1 "Expressive speech synthesis" activity in the CRISIS project

The European project Crisis is aimed at solving critical situations when the inhabitants are endangered. The goal of the "Expressive speech synthesis" activity is to perform basic research, applied research and development of a system which would be capable of generating information system messages and dialogue system replicas in natural speech with considerable content of paralinguistic and extra-linguistic information representing properties such as warning tone, urgency, but also soothing and reassuring speech tone. The application result would be represented by a prototype of new speech synthesizer using large speech databases and modelling using hidden Markov models.

## **Tool for Creating 3D Slovak Speech Visemes**

Eva Pajorová

Institute of Informatic, Slovak Academy of Sciences, Dúbravská 9 84507 Bratislava Slovakia Eva Pajorová, utrrepaj@savba.sk

**Abstract.** A viseme is a representational unit used to classify speech sounds in the visual domain. A "viseme" describes the particular facial and oral positions and movements that occur alongside the voicing of phonemes. Design tool for creating Slovak speech visemes is composed from 5 modules. Modul for creating phonemes, modul for creating 3D slovak speech visemes, modul for facial expression and modul for synchronization between phonemes and visemes and last one modut to generate slovak speech triphones . Lot of system are based on anglish visemes, but till now **no exist slovak speech visemes**. Slovak speech have lot of difference speech-sounds and for that reason it is needed to develop accomplished slovak speech visemes.

#### **1** Introduction

For generating of new sentences, we use a triphonebased approach [6]. Triphones are short pieces of motion sequence that span three phonemes, so each viseme is stored with its context and therefore captures all of the coarticulation effect caused by the direct neighbors. Our similarity measure is easily extended from visemes to triphones, and we can thus find the best overlapping triphone sequences in our database that match any new sentences that needs to be synthesized. Our work is based on dense 3D surface scans, which makes it more versatile than image-based techniques [6]. Related Work

Facial animation is facing three different challenges:

- producing corect and realistic face shapes in every single frame of the animation
- creating a dynamically realistic face motion over time
- creating corect and realistic lip-speech animation

Lot of models [4,5,6,7] may be based on marker point positions, 3D scans or images. This approach facing the problem of defining how the parameters of the model vary over time. For speech synthesis, this involves the problem of coarticulation. Consecutive new approach[8] define dominance functions of phonemes that control the interaction between subsequent phonemes as applied to muscle-based systems [1]. Same systems are based on Hidden Markov Model [9] to learn the dynamics of speech from audio, and transfer this information to a face model. Another approach [10]uses regularization techniques to compute smooth curves for the model parameters over time. In this model, coarticulation is due to the smoothness of the

## Voxel-Based Rendering of E-Beam Lithography Processes

Robert Andok, Peter Slížik, Pavol Nemec, Anna Benčurová, Viera Jablonská, Adrian Ritomský, and Ladislav Hluchý

> Ústav informatiky SAV, Dúbravská cesta 9, 845 07 Bratislava, Slovak Republic

**Abstract.** Electron-Beam Lithography (EBL) is a technology of creating patterns in a thin film of material (resist) by exposing the resist with the electron beam and subsequently chemically removing the exposed (or non-exposed) parts. In order to display the changes in the structure of resist, voxel-based visualization techniques can be used. This paper describes the problem area of ebeam lithography and its visualization. Some tentative insights into the visualization techniques and data formats are presented.

#### **1** Introduction

There are many applications in microtechnology where 3D structures are required. Examples include MEMS, NEMS, optoelectronic devices, photonic band gap (PBG) crystals, diffractive optical elements, blazed gratings, etc. It is known that the performance characteristics of such structures are highly sensitive to their dimensional fidelity. Therefore, it is essential to have a fabrication process by which such 3D structures can be realized with high dimensional accuracy.

One of the tools of choice for writing micro- and nanostructures on a wide variety of materials is the e-beam lithography. This is due to the fact that modern EBL machines are capable of writing nanometer-sized structures on areas up to mm square. Today's requirements in microelectronic production for critical dimension in the range of 45 nm are connected with the improvement and optimization of the electron beam lithography. The resist thickness is 40–150 nm in order to achieve a realistic aspect ratio. There the resist-profile-relief simulation using computer models plays an increasingly important role. Only correct selection of the exposure and development conditions can ensure the necessary higher resolution and the desired resist profile and dimensions. Therefore the details of the computer simulation and visualization should be reconsidered.

## Computer fire simulations by FDS and the influence of a domain decomposition strategy to simulation outcome

P. Weisenpacher<sup>1</sup>, L. Halada<sup>1</sup>, J. Glasa<sup>1</sup>

<sup>1</sup> Institute of Informatics, Slovak Academy of Sciences, Dubravska cesta 9, 84507 Bratislava, Slovakia, {Peter.Weisenpacher, Ladislav.Halada, Jan.Glasa}@savba.sk

**Abstract.** Computational Fluid Dynamics (CFD) has become an important tool in many branches of scientific research as well as in practical applications. One of these application is a simulation of fire and smoke transfer in human structures. Correct parallel implementation of CFD techniques requires solution of nontrivial numerical problems. This paper describe the use of a parallel version of FDS (Fire Dynamics Simulator) for smoke transfer simulation in a short road tunnel. The impact of various computational domain decomposition and different numbers of MPI processes on simulation precision is tested and analysed.

Keywords: fire dynamics, FDS, numerical simulation, domain decomposition

#### 1 Introduction

Computer simulations of complex phenomena plays a significant role in many research areas filling in the gap between experimental and theoretical approach. During the past few years, rapid advance in computing hardware and technologies have allowed simulations of the most challenging and complex scientific problems. One of such discipline is applying the Computer Fluid Dynamics (CFD) theory in simulations of the fire dynamics and its related phemonena. One of the most successful and widespread program system in this field is Fire Dynamics Simulator (FDS) [1,2] developed at National Institute for Standard and Technology (NIST), USA. In this paper, we describe the FDS simulation of smoke transfer in a short road tunnel.

There are several articles in the literature dealing with similar computer simulations of fire in a tunnel [3,4,5,6]. Most of them are focused on the course of combustion and its parameters (temperature, pressure, smokiness, etc.). Our aim in this work is to show some characteristic features of simulation outputs, which are the results of different manners of FDS parallelization. This parallelization requires a decomposition of computational domain into computational meshes, which affect the simulation outputs. Although the parallel calculation of the FDS is already in use, its verification with the respect of the impact of this decomposition is not yet satisfactory completed. In [6], we analyzed some features of this problem, mainly, the dissipation of behaviour pattern on computational mesh boundaries, temperature increase lag for finely decomposed computational domains with many mesh boundaries and

## Special session RECLER Querying large data sets using design patterns

# Using secure agent infrastructure for management of network routers\*

Branislav Simo<sup>1</sup>, Zoltan Balogh<sup>1</sup>, Emil Gatial<sup>1</sup>, Ladislav Hluchy<sup>1</sup>

<sup>1</sup> Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia {branislav.simo, zoltan.balogh, emil.gatial, ladislav.hluchy} @savba.sk

**Abstract.** This paper presents the usage of secure agent infrastructure and individual agents for monitoring and management of specialized network routers called multi-bearer routers. The paper introduces the secure agent infrastructure, its architecture, security concept and possible use case. Then it describes the integration of the agent platform and agents into the multi-bearer routers oftware stack and how are the agents used for monitoring the routers' state and management of certain functionality.

#### 1 Introduction

One of the challenges of the communication infrastructures for crisis management is to add new smart functions to existing services which would make the communication more effective and helpful for users. The aim is to provide smart functions via distributed IT systems which should provide a secure distributed paradigm to achieve confidentiality and access to resources. Such infrastructure should further provide a smart negotiating system for parameterization and independent handling of access requests to achieve rapid reaction. By fulfilling the above stated goals a pervasive and trusted *communication infrastructure* satisfying the requirements of crisis management authorities and ready for immediate application could be introduced.

This article presents the requirements, design and architecture of a distributed agent platform as a core part of this communication infrastructure and describes how this infrastructure is used to monitor and manage distributed network of specialized router devices called multi-bearer routers.

#### 2 Existing Agent-based Platforms For Crisis Management

Mobile agents provide a distributed computing-based paradigm for code mobility that has already demonstrated high effectiveness and efficiency in IP-based highly

<sup>\*</sup> This work is supported by projects SeCriCom FP7-218123, project ITMS: 26240220029, project ITMS: 26240220060, VEGA No. 2/0211/09.

## Using High Performance Data Streaming for Integration and Mining of Environmental Data

Ondrej Habala\*, Ladislav Hluchý, Viet Tran, Peter Krammer, and Martin $\check{\mathrm{Seleng}}$ 

Institute of Informatics of the Slovak Academy of Sciences, Dúbravská cesta 9, 84507 Bratislava, Slovakia

**Abstract.** We have performed several experiments with applying data mining techniques to a set of carefully chosen meteorological and hydrological scenarios. These experiments are part of the FP7 project AD-MIRE, and additionally to serve as an experimental platform for meteorologists and hydrologists, we have used them as a testing platform for a suite of advanced data integration and data mining (DMI) tools, developed within this project. The idea of the project ADMIRE is to develop an advanced DMI platform accessible even to users who are not familiar with data mining techniques. To this end, we have designed a novel DMI architecture, supported by a set of software tools, managed by DMI process descriptions written in a specialized high-level DMI language called DISPEL, and controlled via several different user interfaces, each performing different set of tasks and targeting different user group.

Keywords: Data mining, data integration, environmental data management

#### 1 Introduction

Environmental risk management research is an established part of the Earth sciences domain, already known for using powerful computational resources to model physical phenomena in the atmosphere, oceans and rivers [5]. In this chapter we explore how the data-intensive processes mentioned above can be applied to benefit the experts who produce daily weather predictions, as well as rarely needed, but crucial and often time-critical risk assessments for emerging environmentally significant events. We illustrate the possibilities on a simple scenario from the hydro-meteorological domain, and then describe how this scenario extends to provide meteorologists and hydrologists with new data and insights currently not routinely available. These examples illustrate the complexity of working with real data from multiple sources and lead to a series of "lessons learned" at the end of the chapter.

<sup>\*</sup> This work is supported by projects DMM VMSP-P-0048-09, Projekt ITMS: 26240220029, SMART II ITMS: 26240120029, VEGA No. 2/0211/09.

## Using Parallelization for Simulations of Human Behaviour

## Š. Dlugolinský<sup>1</sup>, M. Kvassay<sup>1</sup>, L. Hluchý<sup>1</sup>, M. Wrzeszcz <sup>2</sup>, D. Król<sup>2</sup>, J. Kitowski<sup>2</sup>

<sup>1</sup>Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia {stefan.dlugolinsky, marcel.kvassay, hluchy.ui}@savba.sk

<sup>2</sup>AGH University of Science and Technology, ACC Cyfronet AGH, Krakow, Poland {wrzeszcz, dkrol, kito}@agh.edu.pl

*Abstract* This article shows how simulations of human behaviour can benefit from the use of high-performance parallel infrastructures. We start with a brief overview of human behaviour modelling and Data Farming experiments in the context of the EDA project A-0938-RT-GC EUSAS. We then proceed to analyse structural adaptations required for exploiting parallel infrastructures, and finally conclude with an outline of various physical realisations permitted by our generic and flexible parallelization approach.

#### 1. Introduction

Human Behaviour Modelling (HBM) is an important area of computational science with implications not only for social sciences, but also for economics, epidemiology and other fields. We encountered HBM in the EDA project A-0938-RT-GC EUSAS, which is financed by 20 nations under the Joint Investment Program Force Protection of the EDA and focuses on asymmetric security threats in urban terrain and aims at combining mission analysis with virtual training of security forces in a highly realistic 3-D cyber environment [1]. This goal is to be achieved by a detailed modelling and simulations of the behaviour of individuals and crowds on the basis of latest findings deriving from psychology, relying on the PECS reference model.

The acronym "PECS" stands for Physical conditions, Emotional state, Cognitive capabilities and Social status [2]. According to [3], "PECS is a multi-purpose reference model for the simulation of human behaviour in a social environment," with emphasis on the "emergent behaviour... typical of the formation of groups and societies." The context for this model is provided by the so-called "agent world," which comprises three kinds of entities: the environment, the connector and the agents. Put briefly, the agents communicate through the connector and interact in

## Enron Emails as Graph Data Corpus for Large-scale Graph Querying Experimentation

Michal Laclavík<sup>1</sup>, Martin Šeleng<sup>1</sup>, Marek Ciglan<sup>1</sup>, Ladislav Hluchý<sup>1</sup>

<sup>1</sup>Institute of Informatics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 07 Bratislava, Slovakia {laclavik.ui, martin.seleng, marek.ciglan, hluchy.ui}@savba.sk

**Abstract.** In this paper we describe Enron email corpus in graph/network data format. Nodes of the graph are emails connected with named entities (NE) extracted from text like people, email addresses, telephone numbers. Edges are links between NE representing concurrence in same email part, paragraph, sentence or composite NE. Enron Graph corpus contains a few millions of nodes and it is quite large corpus for experimenting with various graph querying techniques like graph traversing or spread of activation on graph. The idea is to make this data available for future experiments.

Keywords: graph corpus, email, experimentation, graph querying.

#### **1** Introduction

Graph data and graph databases are becoming quite popular nowadays. Emerging semantic web and its LinkedData<sup>1</sup> is in its bases another type of graph representation build on triples – typed edges – instead of nodes and edges model. Social networks present in social network sites or in telecommunication operator databases are also important source of graph data which need to be processed and queried. The need of graph data processing and querying is growing, but so far scalable solutions, public graph corpuses or common benchmarking for graph querying is not so much present.

In this paper we would like to describe another source of graph data created from email communication, namely from well known Enron email corpus<sup>2</sup> [1]. Email communication analysis allows the extraction of social networks with links to people, organizations, locations, topics or time. Social Networks included in email archives are becoming increasingly valuable assets in organizations, enterprises and communities, though to date they have been little explored. We believe that email communication and its links to other organizational resources can be valuable source of information and knowledge for knowledge management or business intelligence.

<sup>&</sup>lt;sup>1</sup> http://linkeddata.org/

<sup>&</sup>lt;sup>2</sup> http://www.cs.cmu.edu/~enron/

# Testovanie Štandardov prístupnosti aplikácií určených pre verejnú správu

Matej Lučenič<sup>1</sup>, Peter Bílik<sup>1</sup>, Michal Ohrablo<sup>1</sup>, Martin Vozár<sup>1</sup>, Eugen Jeckel<sup>1</sup>

<sup>1</sup> ANASOFT APR, spol. s r.o., Mlynská dolina 41 81102 Bratislava, Slovenská republika {recler}@anasoft.sk http://www.anasoft.sk

**Abstract.** Štandardy pre prístupnosť aplikácií verejnej správy sú definované v dokumente Výnos Ministerstva financií Slovenskej republiky z 8. septembra 2008 č. MF/013261/2008-132 o štandardoch pre informačné systémy verejnej správy. Každým rokom k tomuto výnosu vychádzajú metodické pokyny, ktoré upresňujú a podrobnejšie popisujú štandardy definované vo výnose.

V rámci projektu "Priemyselný výskum v oblasti efektívnej práce s rozsiahlymi dátami v používateľsky orientovaných aplikáciách" sme sa sústredili na štandardy týkajúce sa ergonómie a použiteľnosti webových stránok. Štandardy je možné rozdeliť na základe rôznych kritérií na viacero skupín. Niektoré je možné otestovať vizuálne z webovej stránky, iné je potrebné overiť kontrolou zdrojového kódu.

Skutočnosť, či aplikácie spĺňajú štandardy sme overovali dvoma základnými spôsobmi. Pomocou vhodne zvolených automatických nástrojov, druhý spôsob kontroly realizovali priamo technicky zdatní odborníci. V niektorých prípadoch boli oba spôsoby a výsledky navzájom porovnané.

Na testovanie boli vybrané webové projekty určené pre široké použitie, napríklad portál <u>www.portal.gov.sk</u>. V rámci každého portálu boli určené konkrétne stránky, ktoré boli podrobené testovaniu. Typovo boli vybrané stránky, ktoré sa vyskytujú vo väčšine internetových aplikácií (napríklad prihlásenie/registrácia).

Kritériom výberu odborníkov, ktorí realizovali druhý spôsob testovania bola znalosť HTML kódu, skúsenosti s tvorbou webových stránok, schopnosť čítať a analyzovať zdrojový kód webových aplikácií. Štandardy boli rozdelené do oblastí, zoradené do prehľadných formulárov, do ktorých boli po pretestovaní zapisované výsledky. Na základe získaných informácií budú formulované odporúčania pre tvorbu webových aplikácií určených nielen pre verejnú správu.

## Úvod

Testovanie štandardov webových aplikácií pre verejnú správu bolo realizované v rámci projektu Priemyselný výskum v oblasti efektívnej práce s rozsiahlymi dátami v používateľsky orientovaných aplikáciách (ITMS 26240220029), aktivita 2.1. Výskum v oblasti skvalitnenia práce v rôznymi formátmi údajov. Hlavným zámerom

## Testovanie použiteľ nosti webových aplikácií

Matej Lučenič<sup>1</sup>, Peter Bílik<sup>1</sup>, Michal Ohrablo<sup>1</sup>, Martin Vozár<sup>1</sup>, Eugen Jeckel<sup>1</sup>

> <sup>1</sup> ANASOFT APR, spol. s r.o., Mlynská dolina 41 81102 Bratislava, Slovenská republika {recler<u>}@anasoft.sk</u> http://www.anasoft.sk

Abstract. Testovanie použiteľnosti webových stránok by malo byť súčasťou procesu tvorby webových aplikácii určených pre frekventované použitie. V rámci jednej s aktivít projektu "Priemyselný výskum v oblasti efektívnej práce s rozsiahlymi dátami v používateľsky orientovaných aplikáciách sme sa zaoberali testovaním vybraných webových riešení určených pre použitie v oblasti verejnej správy. Špecifiká tejto oblasti sme zohľadnili pri návrhu metodiky využitej pri testovaní vybraných aplikácií.

Navrhnutá metodika mala nasledovné vlastnosti:

- Testovanie používateľmi z rôznych vekových skupín, sociálnych skupín s rôznou úrovňou skúseností s prácou s počítačom
- Neanonymné testovanie za prítomnosti koordinátora testovania
- Samostatná práca s aplikáciou pri počítači (plnenie jednoduchých úloh v aplikácii)
- Asistovaná práca s aplikáciou pri počítači (plnenie jednoduchých úloh v aplikácii)
- Vyplnenie dotazníka s otázkami k praktickej časti testovania

Boli vytvorené testovacie postupy, formuláre s otázkami, určené oblasti testovania, spoločné znaky testovaných aplikácii, ktoré sa skúmali (napríklad registrácia používateľa), definované skupiny používateľov rozdelené podľa veku, sociálneho zaradenia, podľa skúseností s používaním internetu a počítačov. Výsledkom testovania sú vyplnené dotazníky, ktoré budú následne analyzované a vyhodnotené. Na základe vyhodnotenia bude možné formulovať odporúčania pre podobné typy testovania, aj pre vývoj webových aplikácií určených nielen pre verejnú správu.

## Úvod

Projektová aktivita 2.1. Výskum v oblasti skvalitnenia práce v rôznymi formátmi údajov je zameraná na výskum možností zlepšenia prístupnosti softvérových produktov pre používateľov z pohľadu používateľskej efektivity a jednoduchosti práce

# Distributed crawling and semantic search in Internet resources

Martin Seleng, Stefan Dlugolinsky, Ladislav Hluchy

Institute of Informatics Slovak Academy of Sciences Dubravska cesta 9 845 07 Bratislava 45 Slovak republic {Martin.Seleng , Stefan.Dlugolinsky, Ladislav.Hluchy}@savba.sk

**Abstract.** In this paper, we describe our work in progress in the scope of information extraction and information retrieval utilizing distributed computing. Large Internet resources cannot be crawled by one computer, we need to use a cluster of computers. In our approach we are using MapReduce paradigm to fulfill the demand of crawling large Internet resources. Other thing we want to address in our paper is the semantic search. If we want to do a semantic search we also need to address the information extraction/retrieval and semantic analysis.

**Keywords:** distributed web crawling, information extraction, information retrieval, semantic search

#### 1 Introduction

Building a robust web-scale search service cover plenty of problems. All these problems result from the fact that the Web content is very diverse. There are many heterogeneous sources of information in different languages and there are also many different formats of information representation. According to W3C surveys, there are about 65.5% of top 1 million websites using XHTML markup language, while the rest is using HTML. The trend of XHTML usage is slightly growing, but the semantic standards like RDF, GRDDL, RDFa, SPARQL, OWL, RIF or SKOS are used sporadically in XHTML documents. Therefore general solutions cannot be applied. We need to do information extraction and semantic analysis of the crawled data to allow "intelligent searching" over this data. Another aspect of a robust web-scale search service is the scalability, which can be fulfilled by distributed architecture. [2]

## **Tutorial**

### **Cloud Computing Guides and Tutorials**

Viet Tran<sup>1</sup>, Binh Minh Nguyen<sup>1</sup>

<sup>1</sup>Institute of Informatics, Slovak Academy of Sciences Dubravska cesta 9, 845 07, Bratislava, Slovakia {minh.ui, viet.tran}@savba.sk

**Abstract.** This tutorial will guide participant through various sections which provide an understanding about cloud computing from definitions to classification as well as its actors and roles. Furthermore, it also will give a short description about cloud techniques, together with analyses. And at the last, let's take the view of a few existing cloud systems and problem of developing and deploying appliances/services in cloud environment.

#### **1** Introduction to Cloud Computing

The term of cloud computing may not be strange to scientific communities as well industry nowadays, as it grows very fast in the last five years with the support of infrastructures over network. Users' computer and companies have gradually changed their habit ways of using computational resources to develop and maintain their own data. A simple example, at present, data can be stored and managed in a personal computer or maybe in a centralized third party provider, which already have all the resources what users want and on their demands when they need. The development of cloud computing is sooner or later a trend when John McCarthy<sup>2</sup> opined in 1960 that "computation may someday be organized as a public utility". The term "cloud" has begun to come into commercial use in the early 1990s and up to today. Currently, there are a lot of cloud computing definitions. In the most general way, cloud Computing can be defined as follows:

"A standardized IT capability (services, software, or infrastructure) delivered via the Internet in a pay-per-use and self-service way". [1]

Conceptually, user acquire computing platform or IT infrastructures from computing clouds and then run their applications inside. Therefore, users thus can on-demand subscribe to their computing infrastructures with requirements of hardware configuration, software installation and data access demands.

The cloud computing distinguishes itself from other computing paradigms, like Grid computing, Global computing, Internet Computing in the following aspects: user-centric interfaces, on demand service, QoS guaranteed offer, autonomous system organization, scalability and flexible services. On the economic front, there are

<sup>&</sup>lt;sup>2</sup> John McCarthy (computer scientist) who received the Turing Awards in 1971 for his major contributions to the field of Artificial Intelligence (AI)

### **Author Index**

Andok Robert 228 Balogh Zoltán 242 Benčurová Anna 228 Bílik Peter 272, 277 Biskupič Stanislav 23 **Bolos** Bradut-Vasile 94 Cerňak Miloš 212 Ciglan Marek 266 Comes Calin-Adrian 102 Crainicu Bogdan 126 Darjaa Sakhia 212 Dlugolinský Štefan 258, 283 Dobrucký Miroslav 160. 207 Domiter Marek 106 Doroshenko Anatoliy 54 Gatial Emil 242 Gažová Z. 36 Glasa Ján 188, 233 Gligor Adrian 45. 94 Grondžák Karol 62 Habala Ondrej 250 Halada Ladislav 188. 233 Hanuliak Peter 70 Hayryan S. 36 Hluchý Ladislav 194, 202, 207, 212, 228, 242, 250, 258, 266, 283 **Hrozek** František **118** Hu C.-K. 36 Hudik Martin 70 Hujsi Ivan 86 Huraj Ladislav 168 Iantovics Barna 27, 45, 126

Ivančák Peter 118 Jablonská Viera 228 Jakubík Marián 140 Jeckel Eugen 272, 277 Jókay Matúš 78 Kasanický Tomáš 106 Kitowski J. 258 Korečko Štefan 118 Kozák Andrej 86 Kožár Tibor 36 Krammer Peter 194, 250 Król D. 258 Kurdel Peter 176 Kvasnica Igor 134 **Kvasnica** Peter 134 Kvassav Marcel 258 Laclavík Michal 266 Lučenič Matej 272, 277 Martincová Penka 62 Matyska Luděk 12 Mazzucato Mirco 22 Meyer Norbert 26 Motoška Vladimír 106 Munteanu Anca 111 Nemec Pavol 228 Neslušan Luboš 140 Nguyen Binh Minh 202, 292 Ohrablo Michal 272, 277 Öster Per 25 Pajorová Eva 222 Plančík Milan 78 Ritomský Adrian 228

Rodina Jozef 86 Rusko Milan 212 Sabo Róbert 212 Sebestyénová Jolana 176 Siládi Vladimír 168 Slebodník Lukáš 106 Slížik Peter 160, 228 Sobota Branislav 118 Szabó Csaba 118 Šeleng Martin 250, 266, 283 Šimo Branislav 242 Šipková Viera 160 Šuch Ondrej 62 Tran Viet 54, 202, 250, 292 Trnka Marian 212 Tyrchak Iurii 54 Varga Martin 118 Vozár Martin 272, 277 Weisenpacher Peter 188, 233 Wrzeszcz M. 258 Wu M.-C. 36 Zajac Pavol 78 Zelenka Ján 148 Zhereb Kostiantyn 54 Zvada Marián 106 **Sponsors** 



DATALAN, a. s.



## **European Union**

European Regional Development Fund





Podporujeme výskumné aktivity na Slovensku / Projekt je spolufinancovaný zo zdrojov EÚ.