Grid Computing in Projects of IISAS

Ladislav Hluchý
hluchy.ui@savba.sk
www.ui.sav.sk
Tel/Fax: + 421 2 5477 1004

GCCP’05, November 29-December 1, Bratislava
Research and Development Activities

• **Parallel and Distributed Programming:**
  - research into task allocation tools e.g. **mapping** and **dynamic load balancing** that allocate tasks to the processors of distributed parallel computers so that an objective function, modelling the program execution time, is minimized,
  - **performance evaluation** of the network routing with regard to the different communication algorithms and strategies; performance analysis and runtime **monitoring** of parallel programs

• **Large-scale HPCN and Grid applications:**
  - **design and implementation of problem solving environments** (PSE) for modeling and simulation applications (virtual organisations for environmental problems)
  - developing and maintaining of infrastructures for **Grid computing**
  - **design** of parallel application programs and their implementation using MPI

• **Multi-agent systems** (MAS)
  - Architecture, design and development
  - Decision making, communication, negotiation and coordination among agents
  - Ontology, reasoning and their utilization in multi-agent systems
  - Cooperation between multi-agent systems and workflow management/tracking system (WfMS/WfTS)
II SAS projects related to cluster computing

Projects within IST 4FP


SEPP: CIPA-CT93-0251 Software Engineering for Parallel Processing, EU Copernicus (1994-97)

II SAS projects within IST 5FP

• ANFAS: IST-1999-11676 : datA fusioN for Flood Analysis and decision Support (2000-03) When a flood is going to happen, decision makers have to decide what are the most appropriate reactions: evacuation of the population, reinforcement of dykes, etc. The ANFAS project will develop a support decision system for flood prevention and protection, integrating the most advanced techniques in data processing and management. This tool will help decision takers to limit flood damage.


• PELLUCID: IST-2001-34519 A Platform for Organisationally Mobile Public Employees (2002-04) Project will design, develop and validate a flexible software platform for an important kind of knowledge management: to assist organizationally mobile workers at middle and higher levels of public sector organization.
Data Fusion for Flood Analysis and Decision Support

Objectives

• Development of a Geographical Information System database
• Scene modelling from images
• Flood modelling and simulation
• Pilot operations (Vah River, Loire River, Yangzi River)
• System Integration
• Assessment and evaluation
• Dissemination and implementation
Partners

- ERCIM (European Research Consortium for Informatics and Mathematics)
- Matra System & Information (France)
- Bureau de Recherches Geologiques at Minieres (France)
- Reading University (United Kingdom)
- Foundation for Research and Technology - Hellas (Greece)
- Institute of Informatics, Slovak Academy of Sciences (Slovakia)
  - Subcontractor: Water Research Institute (Slovakia)
- Council for the Central Laboratory of the Research Council (United Kingdom)
- Institute of Automation, Chinese Academy of Sciences (China)
- Institute of Remote Sensing Applications, Chinese Academy of Sciences (China)
- Institute of Atmospheric Physics, Chinese Academy of Sciences (China)
- Institut National de Recherche en Informatique et en Automatique (France)
Detailed FESWMS structures

Nonlinear solver
- Newton iteration is used to solve nonlinear equations.
- Prepare next solution
- Linear solver
- Check solution
- OK

Solution schema
- Input files
- Preliminary computations
- Nonlinear solver
- Write solution to the file
- Solution file

Linear solver is the computational kernel of FESWMS and is the most CPU-time consuming part. Therefore, it is the focus of parallelization.

GCCP'05, November 29-December 1, Bratislava
Parallel **iterative** solvers:

- Conjugate Gradient (CG): the most powerful iterative solver which contains only vector and matrix operations → is trivially parallelized.
- Existing libraries with iterative solvers: PINEAPL (developed in ESPRIT IV projects), PETSc, Aztec, …
- Advantages (in comparison with direct solvers):
  - less expensive (in terms of memory and CPU time)
  - higher parallelism, easier to parallelize
- Disadvantages
  - does not guarantee to converge (direct solvers always do)
TIN network at Predmier
The main work of HPCN solution is to implement parallel computational kernel. The existing code in I/O parts of computational module is reused to guarantee compatibility with existing module and save development time. No modification is required for GUI environment. Users will not notice any changes and use the program as normally. GUI (SMS) will run on PC terminals. Parallel computational module (FESWMS) will run on HPCN platform (supercomputers, clusters of workstations) via any standard protocols (FTP, CORBA, HTTP,…). Parallel computational module will run on HPCN platform (supercomputers, clusters of workstations).
Results: flow + water depths
Results: main part affected by highway

Bytca city

Predmier village

GCCP'05, November 29-December 1, Bratislava
Scenario: Water level for current terrain situation (Q-100-year)
Results: water level for highway with 2 bridges (Q-100-year)

Water level is about 30cm higher than for situation without highway.

Objectives

• CrossGrid application development
• Grid application programming environment
• New Grid services and tools
• International testbed organisation
• Dissemination and exploitation
Partners

- CYFRONET Academ.Comp.Centre of Uni of M&M Krakow, PL
- University of Warsaw, Interdiscipl.Centre f.Math.&Comp.Modell., PL
- H.Niewodniczanski Institute of Nuclear Physics, Krakow, PL
- A.Soltan Institute for Nuclear Studies, Warsaw, PL
- Universiteit van Amsterdam, Faculty of Science, Amsterdam, NL
- Institute of Informatics, Slovak Academy of Sciences, Bratislava, SK
- Inst. f.Technische Informatik und Telematik, J.Kepler Uni Linz, A
- Forschungszentrum Karlsruhe GmbH, C.Inf&Comm.Tech.dpt, D
- Uni Stuttgart, Rechenzentrum, D
- Technische Uni Muenchen, L.f.Rechentechnik...,Fak.f.Informatik, D
Partners

- Uni of Cyprus, Dpt.o.Computer Science, Nicosia, CY
- DATAMAT Ingegneria dei Sistemi S.p.A., Roma, I
- Dept. of Computer Science, Trinity College Dublin, IRL
- Consejo Superior de Investigaciones Cientificas, IFCA, Santander, E
- Uni Autonoma de Barcelona, Arq.d.Ordinadors i Sistemes Operatius, E
- Uni de Santiago de Compostela, Inst.d.Informatica, E
- Uni Autonoma de Madrid, Dept. de Fisica Teorica, E
- Aristotle Uni of Thessaloniki, Div.o.Nuclear&Part.Physics, EL
- Lab.de Instrumentacao e Fisica Exp.de Particulas, Comp.C., Lisboa, P
- Algosystems S.A., Applied Research Dept., Piraeus, EL
CrossGrid

1. Interactive biomedical simulation and visualization
2. Flooding crisis team support
3. HEP distributed data analysis
4. Weather forecasting and air pollution modelling
Flood Warning and Forecasting System

Data Collection Network

Data Transmission System
- HF Radios
- Telemetry Telecom System

Meteorological models:
ALADIN/LACE, ALADIN/SLOVAKIA

Rainfall Runoff Models:
NLC, HBV
River Flow Models:
NLN

- Meteorological and hydrological information and forecasts
- System of attendance of clients
- User of information, warning and forecasts
Flood Warning and Forecasting System

Data sources

Meteorological simulation

Hydrological simulation

Hydraulic simulation

Portal

GCCP’05, November 29-December 1, Bratislava
Flood Virtual Organisation data transfer

### Data sources

- Surface automatic meteorological and hydrological stations
- Systems for acquisition and processing of satellite information
- Meteorological radars
- External sources of information
  - Global and regional centers GTS
  - EUMETSAT and NOAA
  - Hydrological services of other countries

### Storage systems

- Databases

### Grid infrastructure

- Meteorological models
- Hydrological models
- Hydraulic models

### High performance computers

### Users

- Flood crisis teams
  - Meteorologists
  - Hydrologists
  - Hydraulic engineers
- River authorities
- Energy
- Insurance companies
- Navigation
- Media
- Public

GCCP'05, November 29-December 1, Bratislava
Results: GridPort

Precipitation forecast for 24.06.2002 00 UTC + 3 View Prev Next

Precipitation from model ALADIN/SLOVAKIA
integration: 24-05-2002 00 UTC
period: +02 to +03 h
Dynamic flood simulation
step 1  time 0
Dynamic flood simulation
step 3    time 0:30
Dynamic flood simulation
step 50   time 12:15
Dynamic flood simulation
step 100        time 24:45
CROSSGRID testbed

Objectives

• Analysis of public sector working environments and tasks
• Definition of generic system architecture
• Development of the interaction layer
• Development of the process layer
• Development of the access layer
• Development of the organisational memory
• Integration of the platform
• Pilot site customisation, operation and evaluation
Partners

- SADIEL, Spain
- Softeco, Italy
- Cyfronet, Poland
- CCLRC, UK
- II SAS, Slovakia
- Comune di Genova, Traffic & Mobility Directorate, Italy
- Mancomunidad de Municipios del Bajo Guadalquivir, Spain
Pellucid Architecture

ACCESS LAYER
- Dynamic Process Plan Database
- Static Process Plans Database
- Process Plan Data Warehouse
- Analysis Agent
- Backup

PROCESS LAYER
- Monitoring Agent
- Task Assistant Agent
- Workflow Enabled Agent
- Personal Agent
- Action Monitoring
- Plan Retrieval
- Relevant Plan Modification
- Recommendations
- Delegation of tasks and subtasks
- Workflow Process
- Results Extraction
- Task Decomposition

INTERACTION LAYER
- Environment Sensing
- Environment Monitoring
- Reporting
- Delegation
- Environment
- Human Agent (User)
- Real World Processes Functioning

GCCP’05, November 29-December 1, Bratislava
Results: Pellucid Ontology
II SAS projects within IST 6FP

  Project started on April 1, 2004 (70 partners from Europe and USA). The vision of the EGEE Integrated Infrastructure Initiative (III) is to create and deploy Grid technologies to enable the widespread uptake of e-Science applications throughout the European Research Area. To achieve this vision, EGEE will focus on four key objectives: * integrating Grid technological developments from across Europe; * establishing a Europe-wide Grid infrastructure for science and industry with a focus on heterogeneity and interoperability; * enabling the creation of e-Science applications from across the scientific and industrial spectrum; * ensuring the timely delivery of the projects programme of work, guided by the needs of academic and industrial partners.

• **EGEE II proposal is in the hearing stage (April 2006 – March 2008)**
EGEE Project Structure

32 Million Euros EU funding over 2 years starting 1st April 2004

24% Joint Research
JRA1: Middleware Engineering and Integration
JRA2: Quality Assurance
JRA3: Security
JRA4: Network Services Development

28% Networking
NA1: Management
NA2: Dissemination and Outreach
NA3: User Training and Education
NA4: Application Identification and Support
NA5: Policy and International Cooperation

48% Services
SA1: Grid Operations, Support and Management
SA2: Network Resource Provision

Emphasis in EGEE is on operating a production grid and supporting the end-users.

GCCP’05, November 29-December 1, Bratislava
Grid Applications

- Medical/Healthcare (*imaging, diagnosis and treatment*)

- Bioinformatics (*study of the human genome and proteome to understand genetic diseases*)

- Nanotechnology (*design of new materials from the molecular scale*)

- Engineering (*design optimization, simulation, failure analysis and remote Instrument access and control*)

- Natural Resources and the Environment (*weather forecasting, earth observation, modeling and prediction of complex systems*)
CERN: Data intensive science in a large international facility

• The Large Hadron Collider (LHC)
  – The most powerful instrument ever built to investigate elementary particles physics

• Data Challenge:
  – 10 Petabytes/year of data !!!
  – 20 million CDs each year!

• Simulation, reconstruction, analysis:
  – LHC data handling requires computing power equivalent to ~100,000 of today's fastest PC processors!
Mediterranean Grid of Multi-Risk Data and Models

Main objectives:
1. Define a multi-risk assessment platform based on distributed spatial digital data and data processing models
2. Work on the standardization of the structure of data sets that can be useful for risk models testing and multi-risk assessment
3. Create a validation framework for models developed in the context of previous R&D projects
4. Interface to the content of distributed networks of disaster data and
5. Develop web based risk assessment applications, using distributed data

Consortium:
1. ALGOSYSTEMS SA, Kalithea, Greece
2. Associação para o Desenvolvimento da Aerodinâmica Industrial, Coimbra, Portugal
3. EIPFEI/CEREN, Gardanne, France
4. TECNOMA SA, San Sebastián de los Reyes, Spain
5. Institute of Informatics SAS, Bratislava, Slovakia
6. University of Newcastle upon Tyne, United Kingdom
Applications – Flood modeling

• Consists of several simulation models (meteorological, hydrological and hydraulics) and appropriate post-processing tools:
  – MM5 meteorological model forecasts precipitation
  – HSPF hydrological model computes the discharge of the river
  – DaveF hydraulics model computes the possible flood and flooded area
  – All the models generate binary output data, which are then used by post-processing tools to generate pictures visualizing the situation
Applications – Landslides

• Using SHETRAN model
  – Physically based, spatially distributed, integrated surface/subsurface modelling system for water flow and sediment transport in river basins
  – A component is available for modelling shallow landslide erosion and sediment yield
  – can be applied to a single complete basin or to parts of a basin or to groups of contiguous basins up to an area of about 5000 km² using a grid resolution of 500 to 2000 m
Applications – Soil erosion

- Provides answer to the highest soil erosion risks after the passage of a forest fire in definite climate conditions
Applications – Fire Danger and Propagation

• Calculates the danger of occurrence of a forest fire in a known geographical target area and simulates the propagation of such a fire, which is deemed to have started within this area.
• Two main modules:
  – Danger Calculation Module yields the possibility of there being a fire.
  – Simulation Module simulates what will happen if a fire actually occurs.
Applications – Forest fire

• Simulation of fire spread over complex topography
• Semi-empirical model for fire rate of spread, which takes as input local terrain slope, parameters describing fuel properties as well as the wind speed and direction
• Two different models are implemented for the simulation of the wind field
• Outputs: the time evolution of the fire shape, fire rate of spread, fire intensity and other related parameters
• Secondary outputs: the Fire Weather Indexes and the 3D wind field calculation
K-WfGrid

www.kwfgrid.net

- Fraunhofer FIRST (Berlin, Germany)
- UIBK (Innsbruck, Austria)
- IISAS (Bratislava, Slovak Republic)
- CYFRONET (Cracow, Poland)
- LogicDIS S.A. (Athens, Greece)
- Softeco Sismat SpA (Genoa, Italy)

 GCCP’05, November 29-December 1,
 Bratislava
K-Wf Grid Objectives

• Integrating services into coherent application scenarios

• Enabling automatic construction and reuse of workflows with knowledge gathered during operation

• Involving monitoring and knowledge acquisition services in order to provide added value for end users

Technologies: WSRF & service-oriented Grid architecture, workflows, ontologies, dynamic instrumentation
K-Wf Grid Model

- Execute workflow
- Monitor environment
- Analyze information
- Capture knowledge
- Reuse knowledge
- Construct workflow

GCCP’05, November 29-December 1, Bratislava
Architecture and Flow of Actions

- **User**
  - User interaction through the Portal

- **Web Portal**
  - Grid Workflow User Interface
  - User Assistant Agent
  - Guidances for the user

- **Knowledge**
  - Grid Organizational Memory
    - Ontological store of knowledge
  - Knowledge Assimilation Agent
    - Analysed and extracted knowledge
    - Information about workflow execution
    - Information about performance of particular resources
    - Information about resources and environment

- **Workflow Orchestration and Execution**
  - Automatic Application Builder
  - Workflow Composition Tool
  - Scheduler
  - Performance Analysis
  - Information on available resources and their description

- **Grid Resources**
  - Execution of chosen Grid services

- **Low Level Grid Middleware (WS-RF)**
  - Grid Performance Monitoring and Instrumentation Service
  - Information about resources and environment

Flow of Actions:
- User interaction through the Portal
- Workflow composition and execution visualization
- User’s decisions in crucial points of execution
- Information on available resources and their description
- Analysed and extracted knowledge
- Information about workflow execution
- Information about performance of particular resources
- Information about resources and environment
- Execution of chosen Grid services
- User’s decisions in crucial points of execution
Pilot Applications
CTM application
LogicDIS – Enterprise resource planning

- ERP counting thousands of installations in Greece
- Microsoft technology (SQL Server, COM business logic)
- Three Web Services (MS .NET) Available:
  - Calculate New Order
  - Insert New Order
  - Insert New Customer
- Six additional WS wrappers (J2EE) implemented:
  - User request input through a HTML form
  - Visualization of processing results
Flood Forecasting Application

- Flood forecasting based on a series of simulations
- Several simulation models for
  - Meteorology
  - Hydrology
  - Hydraulics
- 2D/3D visualization
- Implementation using GT4 WSRF
- Being extended
  - More models
  - More services – towards risk management
Planed next projects within 5\textsuperscript{th} call of 6FP

\begin{itemize}
\item EGEE II
\item EU-Interactive Grid
\item GRISK
\item ScodeGrid
\item DEGREE
\item GENEVA
\item GRISYS
\end{itemize}
Ďakujem za pozornost’

Vaše otázky?