Grid applications
Computing GRID: the issue ....

- Supercomputer, cluster, ...
  - How to extract the 99,999999% of the computing power of my limited powered expensive environment

- GRID environment
  - How to extract the very power I need from the theoretically infinite powered cheap environment

- Consequence
  - Speedup/efficiency curves are not any more relevant information..
Grid vs. Cluster computing from application view

- **Cluster**
  - Have applications, build a cluster for those applications
  - High efficiency but expensive

- **Grid infrastructure**
  - Have existing platforms, find applications that can efficiently run on those platforms
  - Cheap but not well tailored to every application
Types of Grid applications

- Type 1:
  - Traditional HPC applications running within a site (VO)
    - Using traditional models (MPI, PVM,…)
    - Ready-to-run, no need to modify/re-compile
    - Role of the Grid middleware
      - Resource discovery
      - Deploy and run the application remotely, securely on the discovered resource
Types of Grid applications

- Type 2:
  - New HPC applications running across multiple sites (VOs)
    - Require new programming models/tools
    - Multiple level parallelism
    - Embracing parallelism
    - Example: bio-informatics, parameter sweeping
    - Huge speedup can be achieved
  - Very few applications
  - Role of the Grid middleware
    - Resource discovery
    - Resource allocation and co-allocation
    - Application supporting services
    - Dynamic deployments and executions of application components
Issues

- Missing high-level services
  - QoS of resources

- Heterogeneity

- Code portability
  - Binary/Byte code or source code?

- Resource connectivity
  - Firewall/NAT/ Virtual IP

- Fault tolerance
  - Resource volatility

- Data protection
  - Protect sensitive data from stealing
Grid in the world
**United Kingdom**

- **E-Science, 215M€ over 5 years**
  - e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet.

1 - National Network of Grid Centers
2 - Development of Generic Grid Middleware
3 - Support for e-Science Projects.

- **e-Science support centre**
- **Grid Network Team**
- **Grid Engineering Task Force**
Virtual Laboratory for e-science (VL-e), 55 M€ over 5 years
- 21 partners in 19 institutions

The mission of the VL-E project is:
- To boost e-Science by the creation of an e-Science environment and doing research on methodologies.

The strategy will be:
- To carry out concerted research along the complete e-Science technology chain, ranging from applications to networking, focused on new methodologies and reusable components.

The essential components of the total e-Science technology chain are:
- e-Science development areas,
- a Virtual Laboratory development area,
- a Large Scale Distributed computing development area, consisting of high performance networking and grid parts.
ACI-GRID, ~8 M€ 2001-2004
  - Based on « call for proposals »
  - Use RENATER network

GRID 5000
  - Building a nation wide experimental platform for Grid researches (like a particle accelerator for the computer scientists)
  - 10/11 geographically distributed sites, every site hosts a cluster (from 256 CPUs to 1K CPUs)
  - All sites are connected by RENATER (French Academ. Network)
  - RENATER hosts probes to trace network condition load
  - Design and develop a system/middleware environment for safely test and repeat experiments
  - Only experimental platform (no production)
Europe - CERN

- DATAGRID 10M€, ended beginning 2004
  - 21 partners
  - Feasibility project, final test bed
    - 1000 computers, 15 Terabytes on 25 sites
  - Followed by…

- EGEE, 4 years, 40 M€ for the first two years
  - 70 partners in 27 countries
  - To provide the necessary storage and computing infrastructure to LHC (and others..)
...and at HCMUT....
**VN-Grid:** toward a national-scale computing Grid

- **Main focus:** *infrastructure*
  - High-level services
    - Resource discovery and reservation
    - Scheduling
    - VO and policy management
  - OGSA and WSRF compliance
  - Programming support
    - MPI
    - POP-C++

- We do not develop from scratch!
  - Using GT for providing base services
A VN-Grid scenario

Site

Resource Discovery

Learning

Scheduling

Data warehouse

Monitoring

Resource Allocation/Reservation
Our first prototype-to-be-built

- Keep in mind the heterogeneity the dynamics
  - “Virtual site” concept (VSite)
  - Combine the flexibility of P2P technologies (partial view assumption) with the efficiency of centralized management on each VSite
  - Flexible to involve more resources
  - Flexible security management
    - Multiple level authentication and authorization (VO, user, …)

- Programming supports
  - Parallel object model (POP-C++)
  - MPI

- Applications
  - Oil exploitation (geo-physic data computation of oil fields)
  - Supraconductor study
  - Aviation
  - Chip Design