"Internet-Wide Global Supercomputing"

Jonathan Giddy
CRC for Enterprise Distributed Systems Technology (DSTC)

David Abramson and Rajkumar Buyya
School of Computer Science and Software Engineering
Monash University, Melbourne, Australia

jon@dstc.edu.au, {davida, rajkumar)@csse.monash.edu.au

AUUG'2000
Canberra

DISTRIBUTED
SYSTEMS
TECHNOLOGY
CENTRE

http://www.csse.monash.edu.au/~rajkumar/ecogrid/

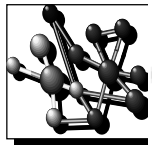
Thanks to:
Jack Dongarra

Agenda

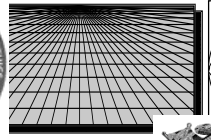
- **Computing Platforms and How Grid is Different ?**
- **Towards Global (Grid) Computing**
- **Grid Resource Management and Scheduling Challenges**
- **Grid Technologies**
  - **Grid Substrates and Interoperability**
  - **Grid Middleware**
  - **Resource Brokers (Nimrod/G) and High Level Tools**
- **Conclusion**

Computing Power (HPC) Drivers

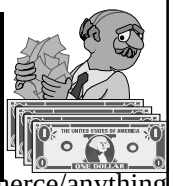
Solving grand challenge applications using computer *modeling*, *simulation* and *analysis*



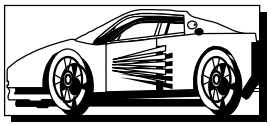
Life Sciences



Aerospace



E-commerce/anything



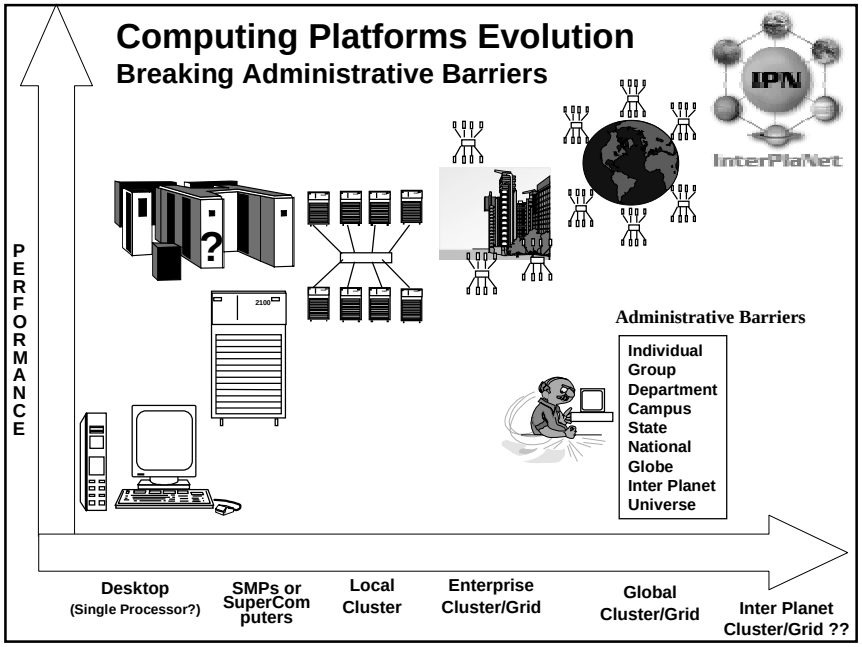
CAD/CAM

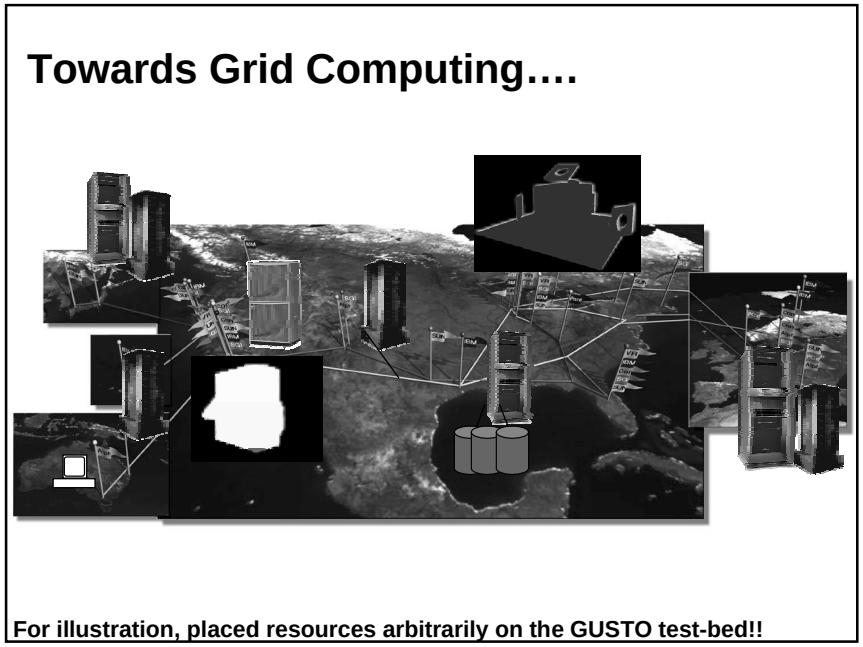
Digital Biology

Military Applications

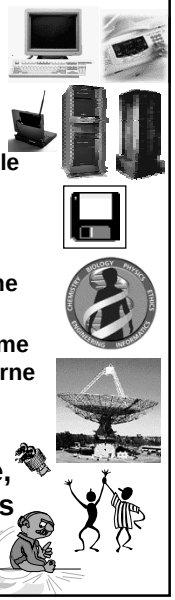
How to Run App. Faster ?

- **There are 3 ways to improve performance:**
  - **1. Work Harder**
  - **2. Work Smarter**
  - **3. Get Help**
- **Computer Analogy**
  - **1. Use faster hardware: e.g. reduce the time per instruction (clock cycle).**
  - **2. Optimized algorithms and techniques**
  - **3. Multiple computers to solve problem: That is, increase no. of instructions executed per clock cycle.**


Computing Platforms Evolution
Breaking Administrative Barriers
IPN
InterPlaNet
PERFORMANCE
Administrative Barriers
Individual
Group
Department
Campus
State
National
Globe
Inter Planet
Universe
Desktop
(Single Processor?)
SMPs or
SuperCom
puters
Local
Cluster
Enterprise
Cluster/Grid
Global
Cluster/Grid
Inter Planet
Cluster/Grid ??

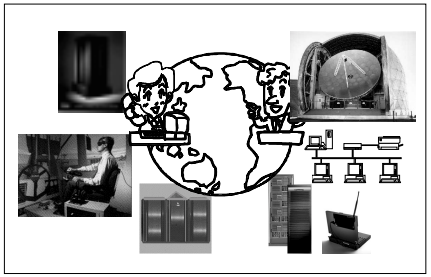

Towards Grid Computing….
For illustration, placed resources arbitrarily on the GUSTO test-bed!!

What is Grid ?

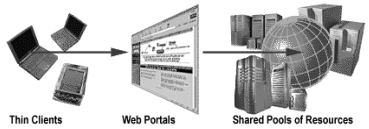


- **An infrastructure that couples**
  - **Computers (PCs, workstations, clusters, traditional supercomputers, and even laptops, notebooks, mobile computers, PDA, and so on) …**
  - **Software ? (e.g., ASPs renting expensive special purpose applications on demand)**
  - **Databases (e.g., transparent access to human genome database)**
  - **Special Instruments (e.g., radio telescope--SETI@Home Searching for Life in galaxy, Austrophysics@Swinburne for pulsars)**
  - **People (may be even animals who knows ?)**
- **across the local/wide-area networks (enterprise, organisations, or Internet) and presents them as an unified integrated (single) resource.**

Conceptual view of the Grid



Leading to Portal (Super)Computing



http://www.sun.com/hpc/

Grid Application-Drivers

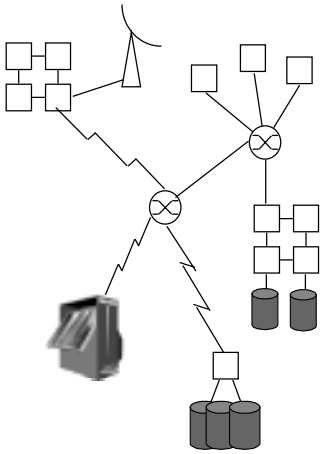
- Old and New applications getting enabled due to coupling of computers, databases, instruments, people, etc:
  - (distributed) Supercomputing
  - Collaborative engineering
  - high-throughput computing
    - large scale simulation & parameter studies
  - Remote software access / Renting Software
  - Data-intensive computing
  - On-demand computing

The Grid Vision: To offer

"Dependable, consistent, pervasive access to

[high-end] resources"

- Dependable: Can provide performance and functionality guarantees
- Consistent: Uniform interfaces to a wide variety of resources
- Pervasive: Ability to "plug in" from anywhere



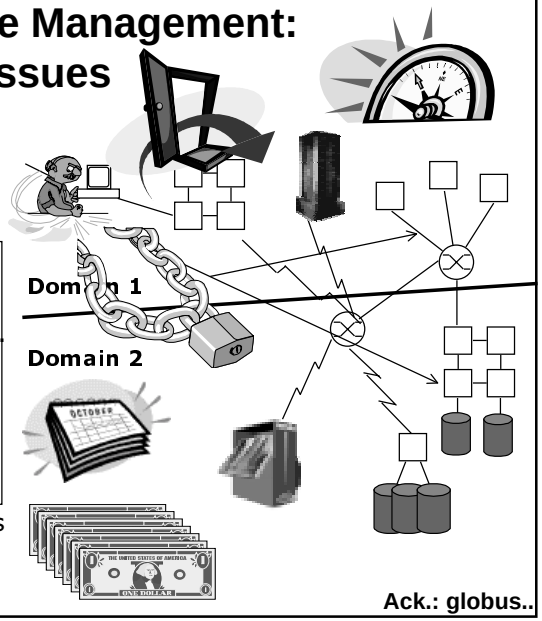
Source: www.globus.org

Sources of Complexity in Grid Resource Management

- **No single adminstrative control**
- **No single policy**
  - **each resource owners have their own policies or scheduling mechanisms.**
  - **Users must honor them (particularly external users of the grid).**
- **Heterogenity of resources (static and dynamic)**
- **Unreliable - resource may come or disappear (die)**
- **No uniform cost model (it cannot be)**
  - **varies from one user to another and time to time.**
- **No Single access mechanism**

Grid Resource Management: Challenging Issues

- Authentication (once)
- Specify simulation (code, resources, etc.)
- Discover resources
- Negotiate authorization, acceptable use, Cost, etc.
- Acquire resources
- Schedule Jobs
- Initiate computation
- Steer computation
- Access remote data-sets
- Collaborate on results
- Account for usage

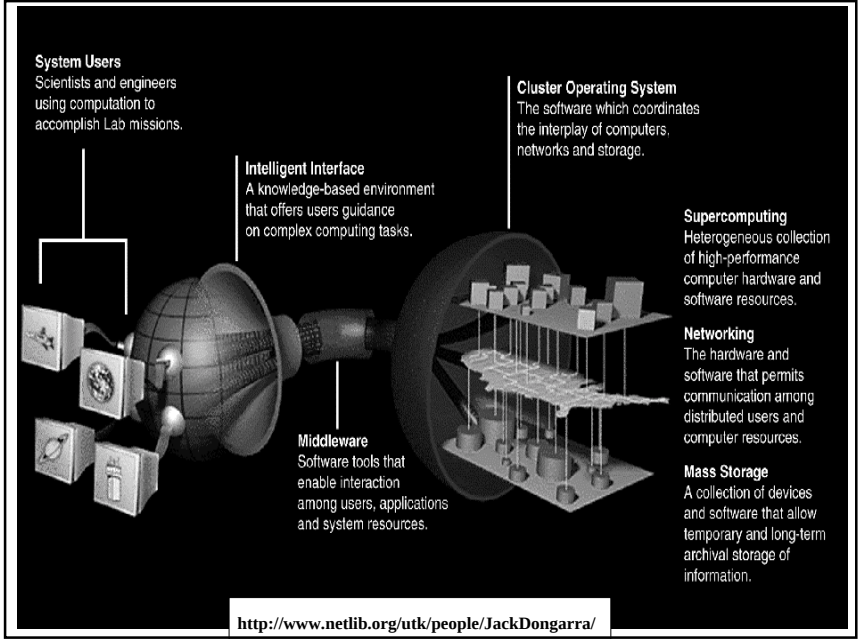


Ack.: globus..

Goals of Application Development Support

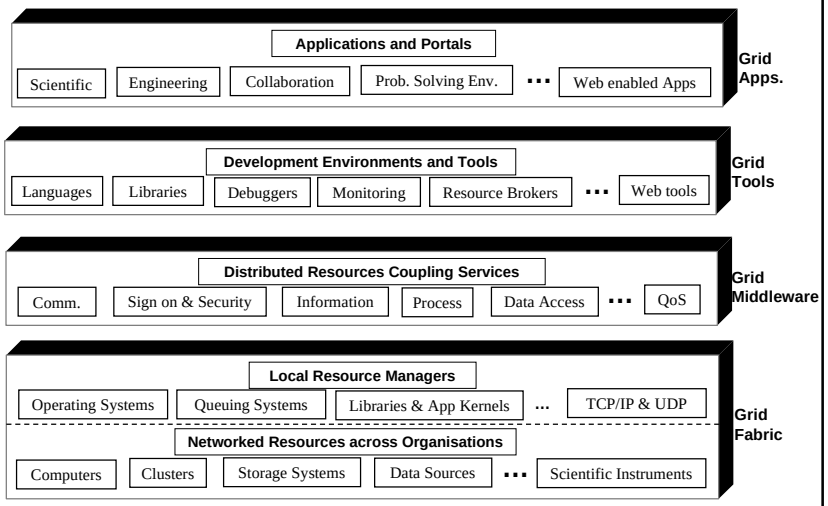
- **Applications should**
  - be easy to develop
  - be portable
  - achieve high performance
    - » as close to what is possible by hand
- **Application Developer**
  - should be able to concentrate on problem analysis and decomposition
- **System**
  - should handle details of mapping abstract decomposition onto computing configuration
- **Developer and System**
  - should work together to debug and tune the program

http://www.netlib.org/utk/people/JackDongarra/



http://www.netlib.org/utk/people/JackDongarra/

Grid Components

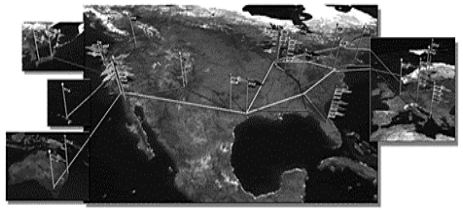


Many GRID Projects and Initiatives

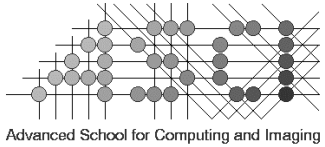
- PUBLIC FORUMS
  - Computing Portals
  - Grid Forum
  - European Grid Forum
  - IEEE TFCC!
  - GRID'2000 and more.
- Australia
  - Nimrod/G
  - EcoGrid and GRACE
  - DISCWorld
- Europe
  - UNICORE
  - MOL
  - METODIS
  - Globe
  - Poznan Metacomputing
  - CERN Data Grid
  - MetaMPI
  - DAS
  - JaWS
  - and many more...
- Public Grid Initiatives
  - Distributed.net
  - SETI@Home
  - Compute Power Grid
- USA
  - Globus
  - Legion
  - JAVELIN
  - AppLes
  - NASA IPG
  - Condor
  - Harness
  - NetSolve
  - NCSA Workbench
  - WebFlow
  - EveryWhere
  - and many more...
- Japan
  - Ninf
  - Bricks
  - and many more...

http://www.gridcomputing.com/

Many GRID Testbeds...



GUSTO



Distributed ASCI Supercomputer

NASA IPG

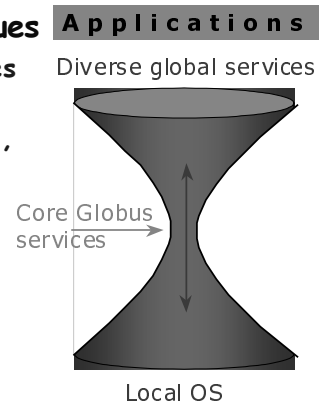
Globus Project: Argonne National Lab & Information Sciences Inst. (Foster & Kesselman)

- Enable high-performance applications that use resources from a "computational grid"
  - Computers, databases, instruments, people
- Through
  - Basic research in grid-related technologies
  - Development of Globus toolkit: Core services for grid-enabled tools & applications
  - Construction of large-scale grid testbed: GUSTO
  - Extensive application experiments

http://www.globus.org/

Globus Approach

- **Focus on architecture issues**
  - **Propose set of core services as basic infrastructure**
  - **Use to construct high-level, domain-specific solutions**
- **Design principles**
  - **Keep participation cost low**
  - **Enable local control**
  - **Support for adaptation**

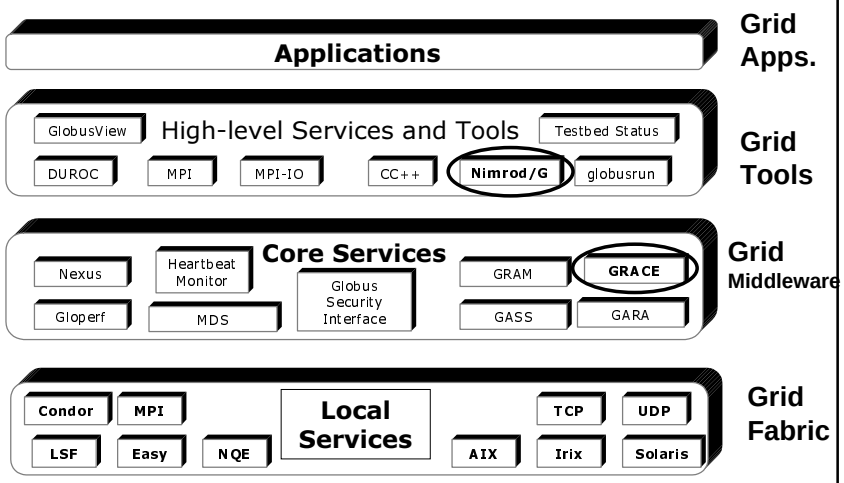


Globus Toolkit: Core Services

- **Scheduling (Globus Resource Alloc. Manager)**
  - Low-level scheduler API
- **Information (Metacomputing Directory Service)**
  - Uniform access to structure/state information
- **Communications (Nexus)**
  - Multimethod communication + QoS management
- **Security (Globus Security Infrastructure)**
  - Single sign-on, key management
- **Health and status (Heartbeat monitor)**
- **Remote file access (Global Access to Secondary Storage)**
- **Reservation of Resources in Advance (GARA)**

Grid Components and our Focus:

Resource Brokers and Computational Economy

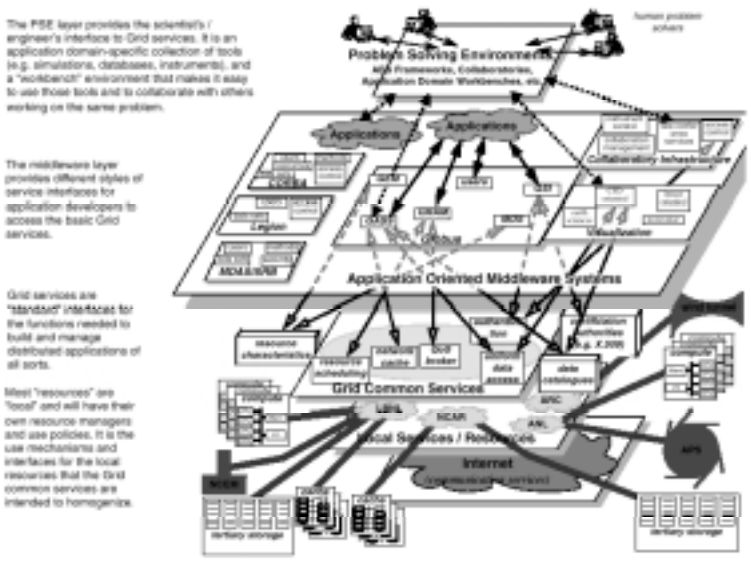


Source: Globus

NASA's Information Power Grid

- Developing and evolving the technologies needed into a computational and data grid, providing the infrastructure for widely distributed systems.
- Specific IPG requirements come from analyzing NASA Aerospace Engineering Systems and Earth Sciences, Data Assimilation Office applications.

- Grids will provide:
  - Application capabilities
  - Distributed resource access
- Grids must provide:
  - Scalability
  - Usability



1/99

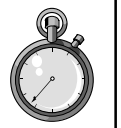
12

http://nas.nasa.gov/~wej/home/IPG

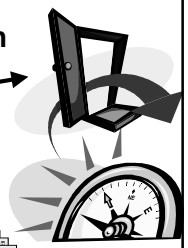
Nimrod/G Resource Broker

Nimrod/G Approach to Resource Management and Scheduling

What is Nimrod/G ?

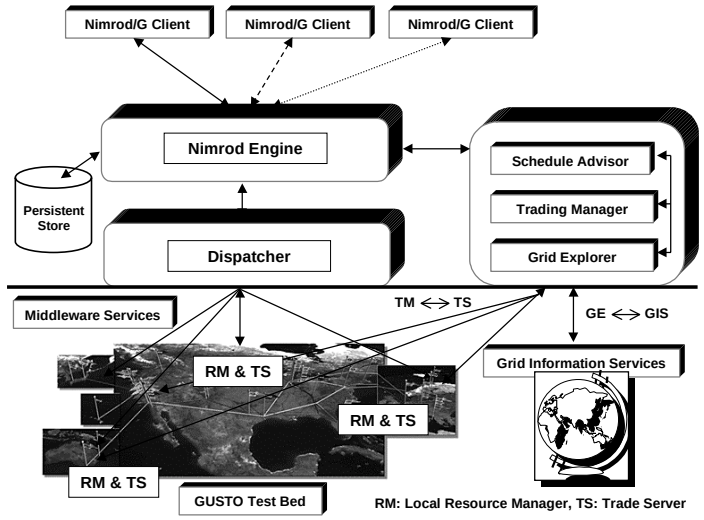


- A global scheduler for managing and steering task farming (parametric simulation) applications on computational grid based on deadline and computational economy.
- Key Features
  - A single window to manage and control whole experiment
  - Resource Discovery
  - Trade for Resources
  - Scheduling
  - Steering & data management
- Leverages Globus Services
- Other parallel application models can be supported easily (MPI & DUROC co-allocation)

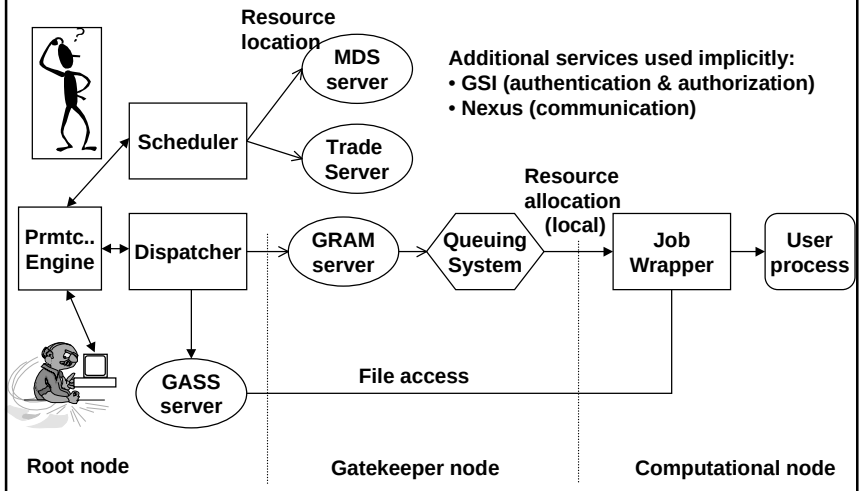


A Nimrod/G User Console

Nimrod/G Architecture

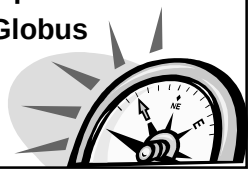


Nimrod/G Interactions



Resource Location/Discovery

- Need to locate suitable machines for an experiment
  - Speed
  - Number of processors
  - Cost
  - Availability
  - User account
- Available resources will vary across experiment
- Supported through Directory Server (Globus MDS)



Computational Economy

- Resource selection on based real money and market based
- A large number of sellers and buyers (resources may be dedicated/shared)
- Negotiation: tenders/bids and select those offers meet the requirement
- Trading and Advance Resource Reservation
- Schedule computations on those resources that meet all requirements

Cost Model

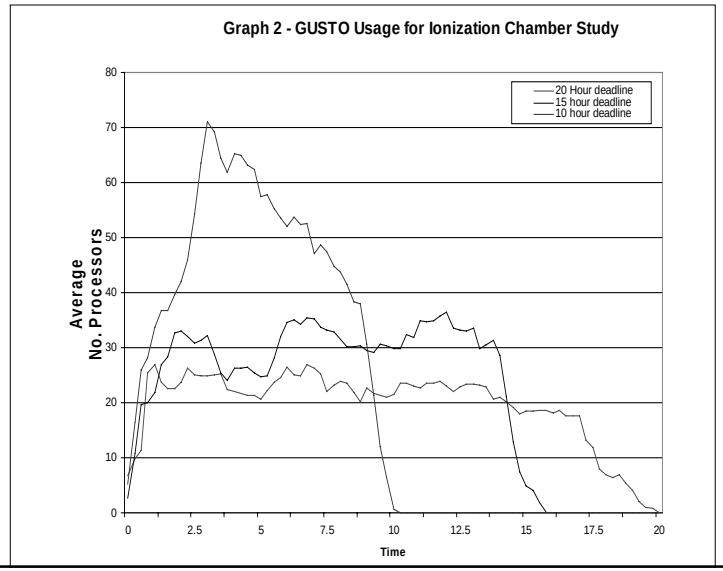
- non-uniform costing
  - time to time
  - one user to another
  - usage duration
- encourages use of local resources first
- user can access remote resources, but pays a penalty in higher cost.

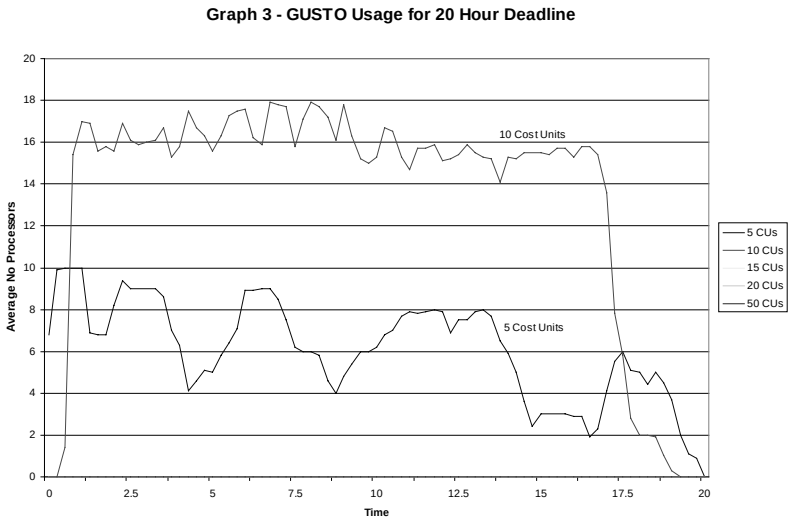
	Machine 1			Machine 5	
User 1	1			3	
User 5	2			1	

Nimrod/G Scheduling Algorithm

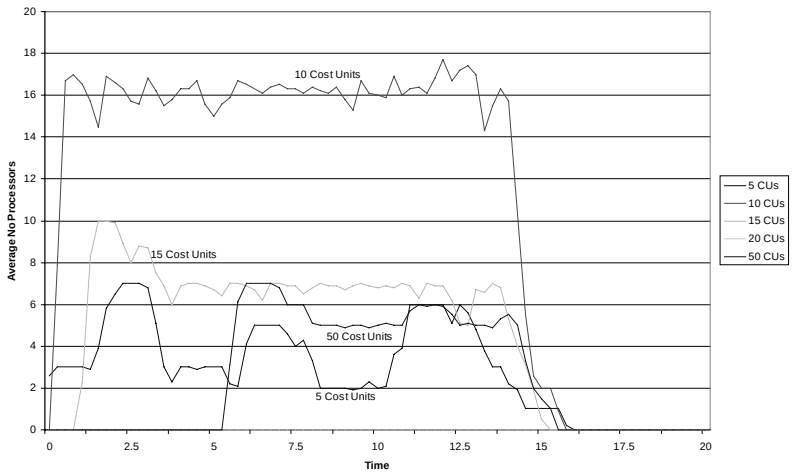
Find a set of machines (MDS search)
Distribute jobs from root to machines
Establish job consumption rate for each machine
For each machine
  Can we meet deadline?
  If not, then return some jobs to root
  If yes, distribute more jobs to resource
If cannot meet deadline with current resource
  Find additional resources

Resource Usage (for various deadlines)

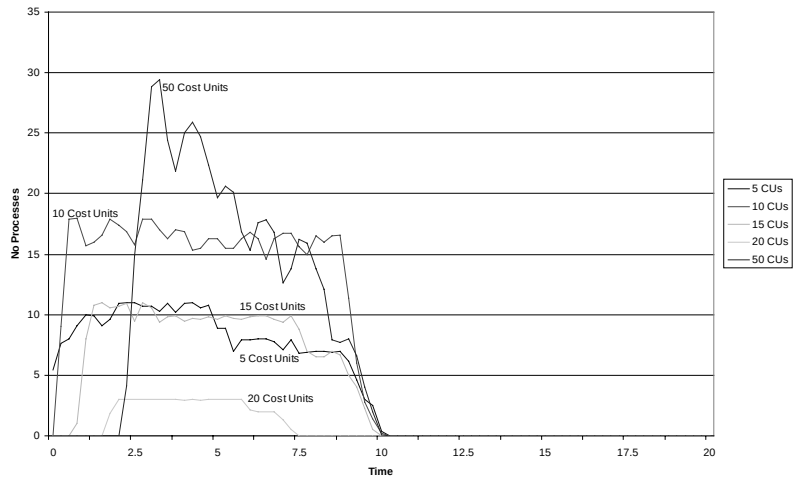




Graph 4 - GUSTO Usage for 15 Hour Deadline



Graph 5 - GUSTO Usage for 10 Hour Deadline



Scheduling Methods for Global Resource Allocation

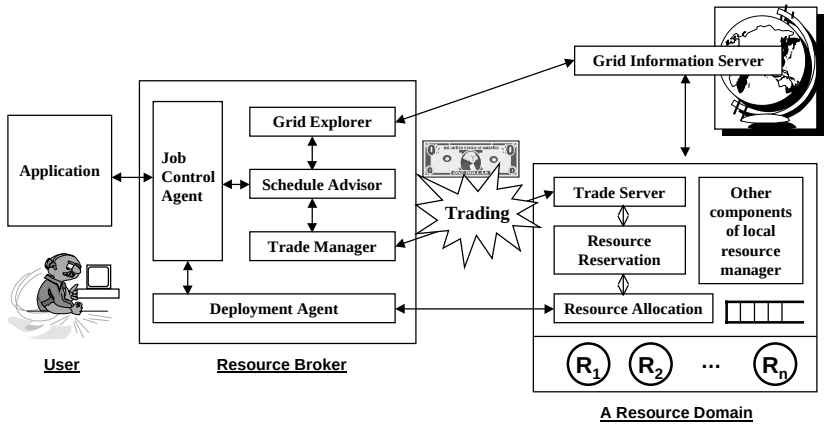
1. Equal Assignment, but no Load Balancing
2. Equal Assignment and Load Balanced
3. (2) + deadline (no worry about cost)
4. (2) + deadline (minimize the cost of computation) ✓
5. (4) + budget --> deadline + cost (up front agreement)
   - I am willing to pay $$$, can you complete by deadline.
6. (5) + Advance Resource Reservation
6. (6) + Grid Economy - dynamic pricing - use Trading/Tender/Bid process
7. (6) + Grid Economy - dynamic pricing - use auction technique
8. Genetic/Fuzzy Logic, etc algorithms

GRACE

Grid Architecture for Computational Economy

- GRACE aims help Nimrod/G overcome the current limitations.
- GRACE middleware offer generic interfaces (APIs) that other developers of grid tools can use along with Globus services.

Grid Resource Management Architecture (Economic/Market-based)



Why Computational Economy in Resource Management ?

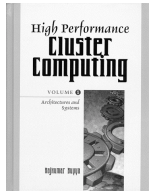
"Observe Grid characteristics and current resource management policies"

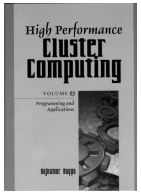
- **Grid resources are not owned by user or single organisation.**
- **They have their own administrative policy**
- **Mismatch in resource demand and supply**
  - **overall resource demand may exceed supply.**
- **Traditional System-centric (performance matrix approaches does not suit in grid environment.**
  - **System-Centric --> User Centric**
- **Like in real life, economic-based approach is one of the best ways to regulate selection and scheduling on the grid as it captures user-intent.**
- **Markets are an effective institution in coordinating the activities of several entities.**

Conclusions

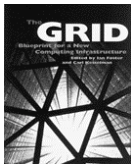
- The Emergence of Internet as a Powerful connectivity media is bridging the gap between a number of technologies leading to what is known as "Everything on IP".
- Cluster-based systems have become a platform of choice for mainstream computing.
- Economic based approach to resource management is the way to go in the grid environment.
- The user can say "I am willing to pay $…, can you complete my job by this time…"
- Both sequential and parallel applications run seamless on desktops, SMPs, Clusters, and the Grid without any change.
- Grid: A Next Generation Internet ?

Further Information





- Cluster Computing Infoware:
  - `http://www.buyya.com/cluster/`
- Grid Computing Infoware:
  - `http://www.gridcomputing.com`
- Millennium Compute Power Grid/Market Project
  - `http://www.ComputePower.com`
  - You are invited to join CPG project!
- Books:
  - High Performance Cluster Computing, V1, V2, R.Buyya (Ed), Prentice Hall, 1999.
  - The GRID, I. Foster and C. Kesselman (Eds), Morgan-Kaufmann, 1999.
- IEEE Task Force on Cluster Computing
  - http://www.ieeetfcc.org
- GRID Forums
  - http://www.gridforum.org | http://www.egrid.org
- GRID'2000 Meeting
  - `http://www.gridcomputing.org`



GRID'2000