Grid Computing: Research Issues and Challenges

R. K. Ghosh
Dept of CSE, IIT Kanpur

Export Restriction by US

- Computer export from US to India, China, Russia and Middle East based on MTOPs
- Before 2001 – 28,000 MTOPs –> less powerful than a cluster of 10 1.5 GHz/2-way PCs.
- 2001 – 85,000 MTOPs –> less powerful than a cluster of 10 2.2 GHz/4-way PCs.
- 2002 – 195,000 MTOPs –> less powerful than a cluster of 10 3 GHz/8-way PCs.

(source: Xiaodong Zhang, NSF.)

Inadequacy of Client/Server

- 2×10^{18} Bytes/year generated in Internet.
- But only 3×10^{12} Bytes/year available to public (0.00015%).
- Google only searches 1.3×10^{8} Web pages.

(source: Gong, IEEE Internet Computing, 2001.)

Inadequacy of Client/Server

- Asymmetric utilization of services and bandwidth.
  - Clients have mainly passive roles –> computing cycles are unutilized.
  - Servers (popular ones) suffer from traffic congestion.

Characteristics of P2P Model

+ Nodes can leave and join at any time.

- Hetergeneity: service capabilities, storage, network speed, service demand

+ A decentralized system with equal opportunities for all participating nodes.

P2P Model

- Client server.
- Pure P2P system
- Hybrid P2P (directory on top of pure P2P)

Problems in P2P Computing

- Security and Privacy
  - Information leakage, evil codes and viruses, privacy protection (loss of anonymity)
- Weak resource coordination
  - Unbalanced load due to weak/no coordination
  - Lacks communication/schdule monitor –> traffic congestions
  - Rely on self organization.

Ideal P2P Model

- Fast peer service
  - Low diameter region for peer to peer interaction.
  - Dynamically identifying and collecting trusted peers.
  - Adaptive self-organized coordination.
- Allowing peer distrustful peer to exist
  - DoS attack, evil code and viruses, intrusion detection.
  - Exposing identity of peers (communication anonymity)
- Measurable security metrics
  - Benchmarks, stochastic models, quantifying degree of security.

Ideal P2P Model

- Understanding tradeoffs
  - Impact of loss of central control over security.
  - Quantifying security loss, performance loss/gain due to decentralization.
  - Conflict of common and individual objectives.
- Building over existing infrastructures
  - Minimizing new standards and protocols
  - Avoid modifying commonly used and general purposed s/w.
  - Peer oriented processing should be automatic.

Application on P2P

- Document/file sharing: with no or limited central control.
- Instant messaging: immediate voice and file exchange among peers
- Distributed processing: use resources available in other remote peers.

Application Differences: Grid & P2P

- Grid: global problem solving environment for large and critical scientific applications and professional collaborations, where each node is a server.
- P2P: a general and commercial information/computing services, where each peer can be both server and client.

Operation Differences: Grid & P2P

- Grid: direct access to computing, software, and data resources in remote & targeted sites (Servers-based).
- P2P: random accesses to available computing, software, and data resources without a specific target (Clients-based).

Different Participants: Grid & P2P

- Grid: pre-determined and registered clients and servers.
- P2P: clients and servers are not distinguished and registered, which can come and go by their choices.

Different QoS: Grid & P2P

- Grid: guaranteed and reliable services are required for each grid server.
- P2P: only partially reliable, because services from some peers are not guaranteed and trusted.

Security Differences: Grid & P2P

- Grid: authentication, authority, and firewall protection to each grid.
- P2P: privacy, anonymity, authentication, authority, and fire wall protection to each peer is not guaranteed.

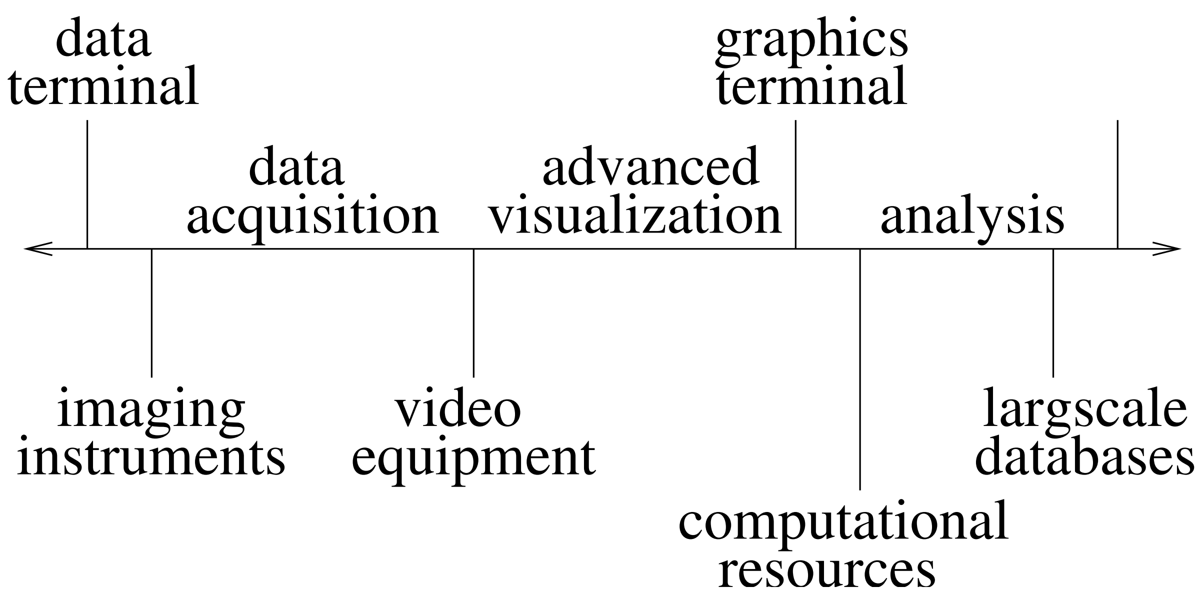
Different Controls: Grid & P2P

- Grid: centralized control plays important role in resource monitoring/allocations and job scheduling.
- P2P: limited or no central controls, mainly rely on self-organization.

Grid Computing

- Term was coined around 1995 to denote (a proposed) distributed computing infrastructure for science & engineering.
- Extended to commercial computing applications.
- Dynamically links resources together for execution of large scale, resource intensive distributed applications.
- Integrates networking, communication, computation and information into a virtual platform for computation and data management.

Grid Computing



Grid Computing

- Similar to a utility grid.
- Seeks to and is capable of adding an infinite number of computing devices.
- Capabilities can be added within the operational environment.
- Collaboration at global level −> huge talent pool.
- Takes distributed computing to next evolutionary level.
- Creates an illusion of a simple but large self managing virtual computer.

Grid Computing

- Complicated global computing environment that leverages many open standards and technologies in a wide variety of implementation schemes.
  - UDDI, XML, SOAP, HTTP, WSDL, WSFL
  - Globus, Linux, Java

Grid Computing

- Ubiquitous platform – so far as usability scenarios and virtual organizations indicate.
- Virtual organizations
  - Financial forecasting models (e.g. deciding on new factory location)
  - Feasibility studies (e.g. multi-disciplinary simulation of aircarft)
  - Crisis management (e.g. mitigation of chemical spills)
  - Data grid (e.g. high energy physics - 178,368 peta bytes of data)
  - Internet games (e.g. virtual world - adding to population)
  - Impact of drug on performance of brain (low level chemical simulation across different databases)

Grid Candidates

- A cluster system on local area network –> just a resource.
  - a centralized control over the hosts that it manages.
- Web Service is generic solution for interoperability over distributed environment (Internet).

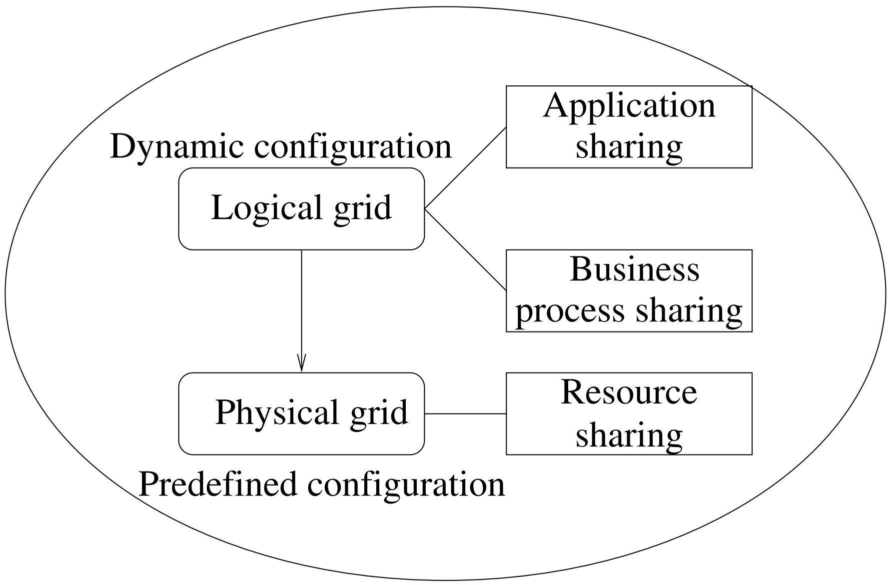
Web Service & Grid Computing

- Grid –> extension of WS to solve computing problems in scientific and business domain.
- OGSA (open grid service architecture) is a distributed interaction and computing architecture
  - Leverages WS to define WSDL interfaces for Grid service.
  - Assures interoperability on heterogeneous systems based arond Grid services.

Grid Candidates

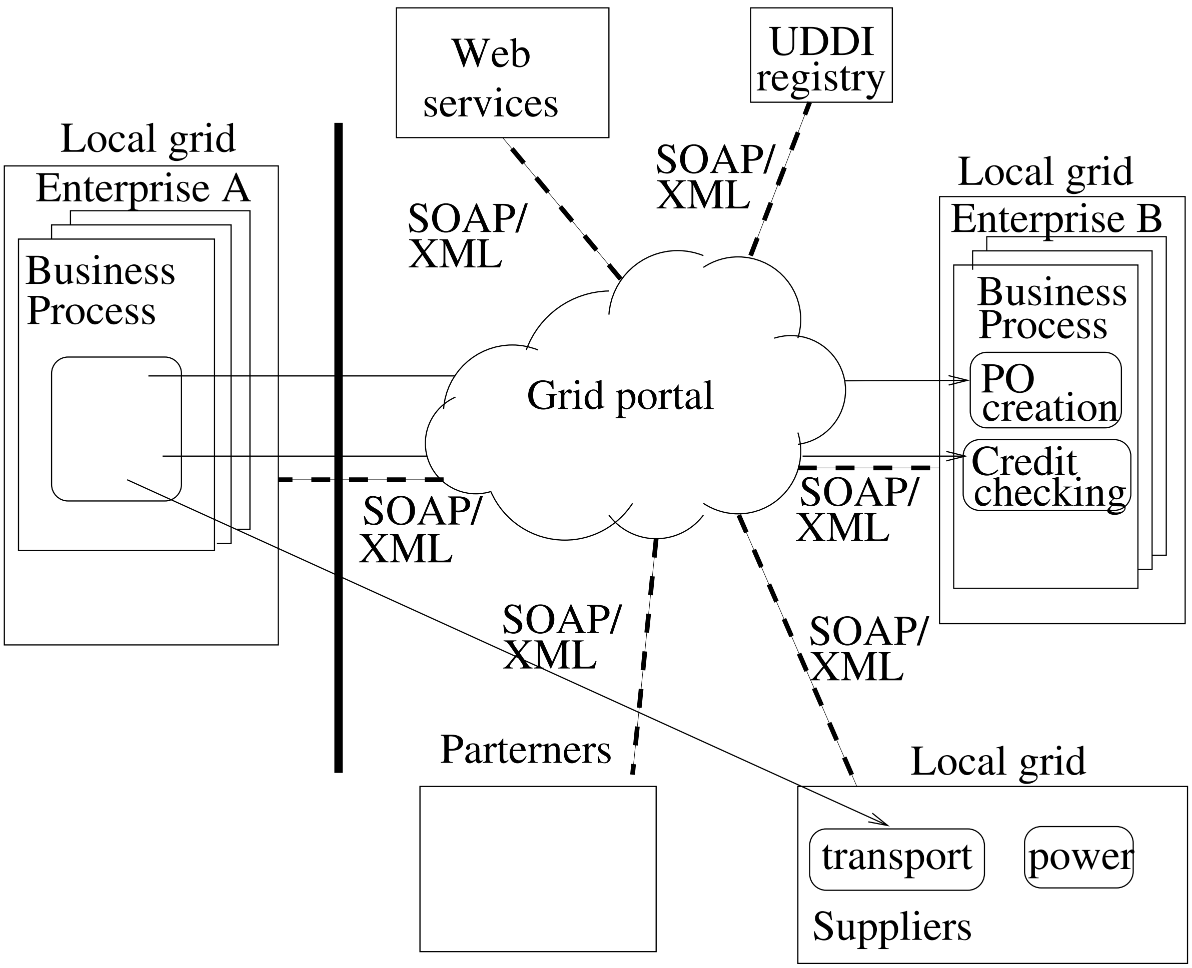
- Multi-site schedulers can reasonably be called (first-generation) Grid
- Distributed computing systems provided (Condor, Entropia, and United Devices) which harness idle desktops
- Peer-to-peer systems (such as Gnutella) which support file sharing among participating peers;
- A federated deployment of the Storage Resource Broker, which supports distributed access to data resources.
- The protocols used in these systems are too specialized though each integrates distributed resources in the absence of centralized control, and delivers interesting qualities of service in narrow domains.

Grid Solution Sphere



- Logical grid refers to s/w and appl. sharing as well as business process sharing. Configured dynamically.
- Physical grid refers to computer power shared over distributed n/w for a specific task –> predefined.

BPO: Logical Grid



Remote Execution

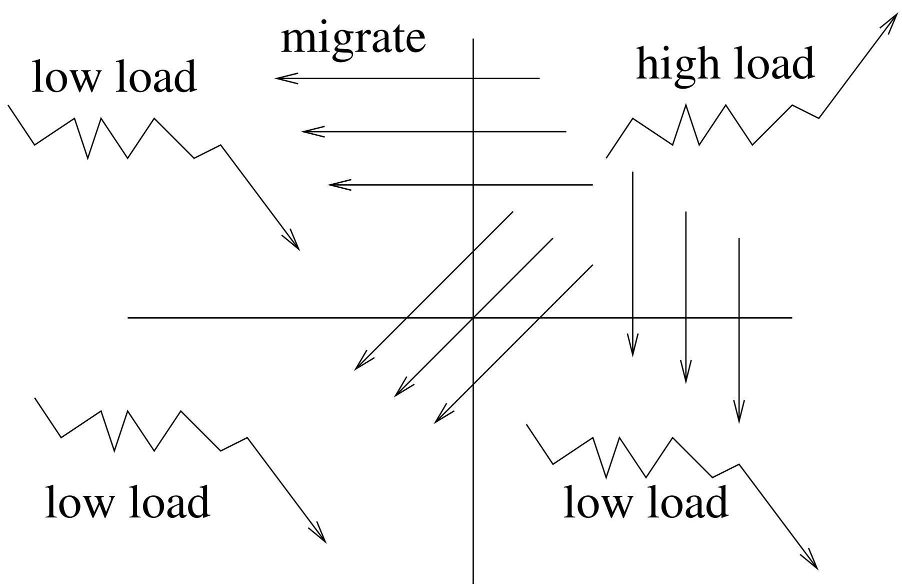
- Easiest use of grid computing is to run existing application on a different m/c.
- Pre-req for this:
  - Application must be executable remotely with undue overhead.
  - Remote m/c could meet all special h/w, s/w or other resource requirements.
- Using remote m/c for word processing (interactive jobs) does not make sense. But for batch jobs it is ok.
- There are under-utilized computing resources (desktops are busy only 5% of time). Grid computing can make use of these.
- Grid can make data highly available. Most computer has lot of storage, so data can be replicated.

Collaboration

- Simplifies collaboration.
- Users can be organized dynamically into number of VOs each having different policy requirements. But they share resources collectively.
- Sharing can be in data (files, databases) – by replication, striping etc.
- Equipment, s/w, services, licenses all can be shared.
- But sharing calls for strong security rules.

Balanced Resource Utilization

- Better balancing of resource utilization.

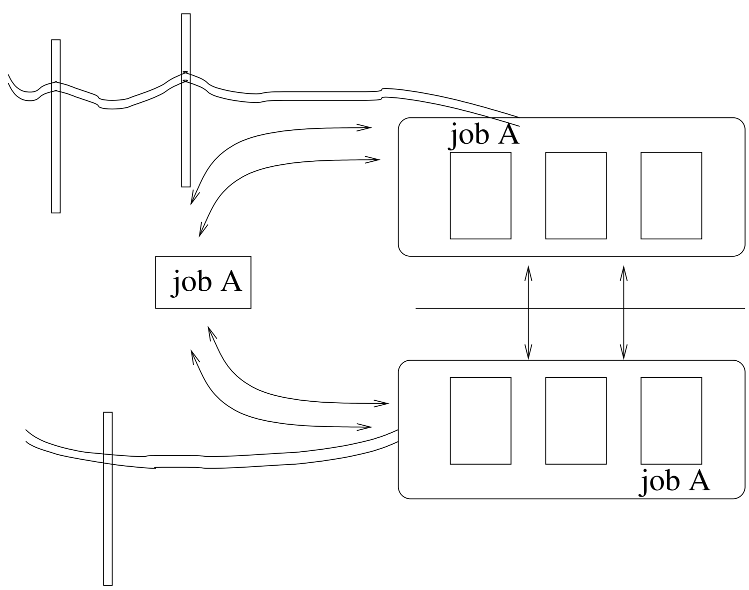


Parallel Computing

- Parallel CPU capacity is another attractive feature.
- All applications can not be transformed to run in parallel.
- The number of independent running parts into which an application can be split is the major difficulty.
  - All application can not be transformed to run in parallel on a grid and achieve scalability.
  - There are no tools for transforming an arbitrary application to exploit parallel capabilities of a grid.

Reliability & Fault Tolerance

- Redundancy in conventional system must be built explicitly and expensive (both h/w and s/w).
- Inherent redundancy in grid configuration allows fault tolerance and reliability without any extra costs.



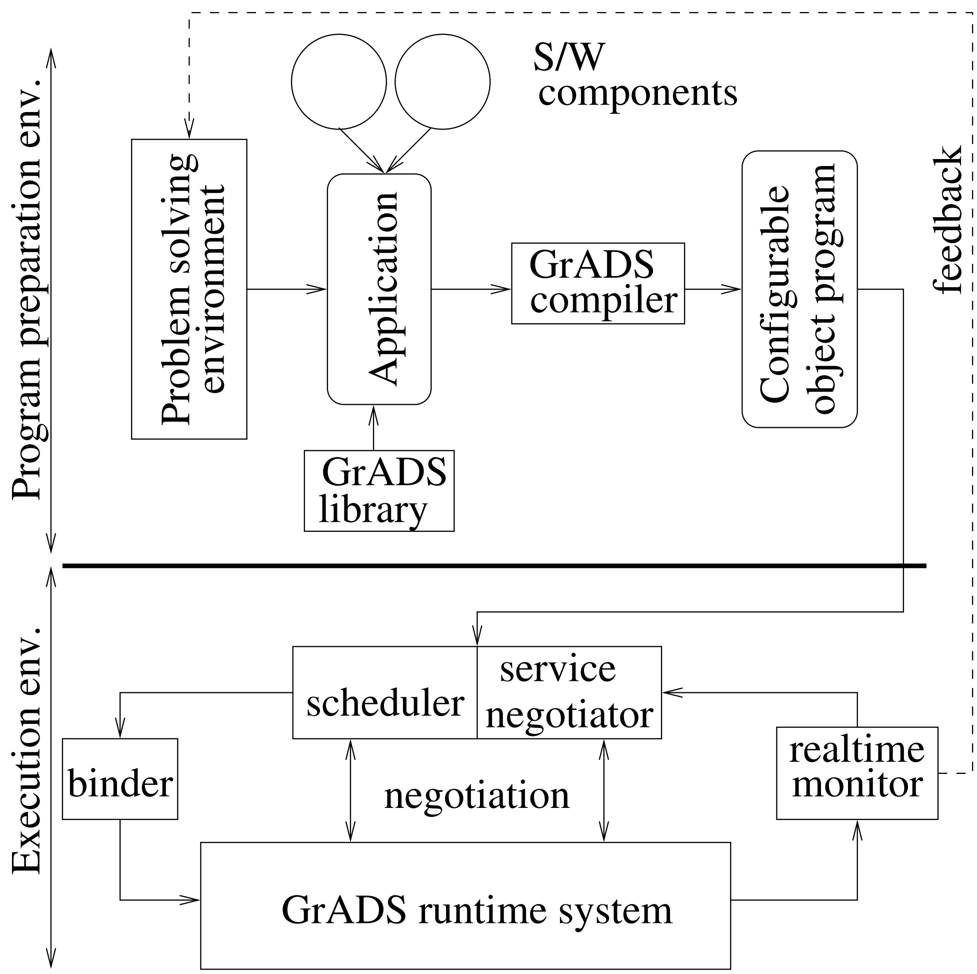
Grid Management

- Controlled expenditures for computing resources over a larger organization.
- Priorities among different projects can be better managed
- Aggregated utilization enhances ability to anticipate future upgrades and eased maintenance (reroute jobs from maintenance sites).
- Autonomic computing tools (recovery from various grid outages, failures) can be deployed.

Application Development

- Grid Application Development Software (GrADS) provides tools and execution environment.
- Ongoing since 1999. Participating universities: UC San Diego, U Tennessee Knoxville, UIUC, Univ of Houston.
- Development framework has two distinct parts
  - GrADS program preparation system (GrADS PPS).
  - GrADS execution environment (GrADS EE).

Development Framework



Challenges

- Comprehensive administration
- Resource provisioning
- Adapative application intergration
- Flexible data sharing and access.
- Activity monitoring.
- Policy-based grid management mechanisms.

References

[1] F. Berman, et al. New Grid Scheduling and Reschduling Methods in the GrADS Project, International Journal of Parallel Programming (to appear), 2005.

[2] X. Zhang. Research Issues for Building and Integrating Peer-Based and Grid Systems. In http://www.nesc.ac.uk/talks/china_meet/zhang_beijing_talk.pdf

[3] Ian Foster, et. al. The Anatomy of a Grid. International Journal of Supercomputer Applications, 2001.

[4] Jen-Yao Chung, Liang-Jie Zhang. Business Grid: Grid Computing Infrastructure for e-Business Solutions, OMG's 2nd Workshop On Web Services Modeling, Architectures, Infrastructures And Standards.

[5] Li Gong. A Software Architecture for Open Service Gateways. IEEE Internet Computing 5(1): (2001).