Introduction to GRID Computing and Overview of the Globus Toolkit

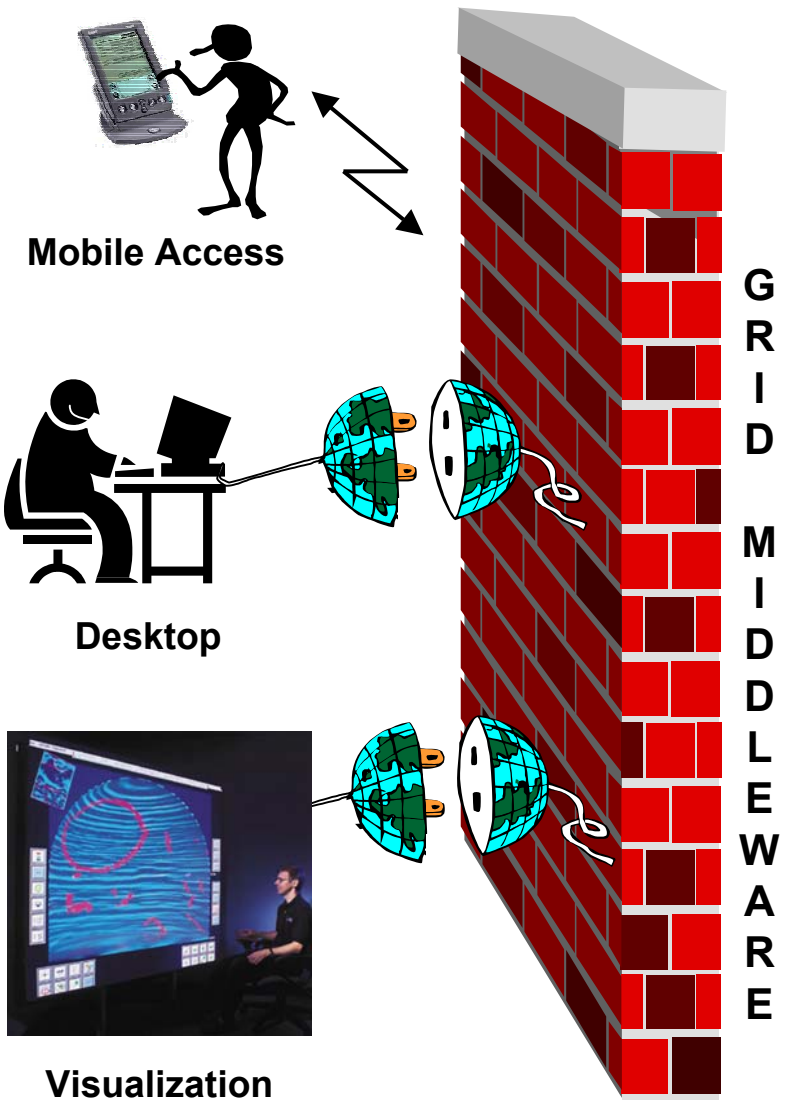
Marcel Kunze
Abteilung Grid Computing und e-Science

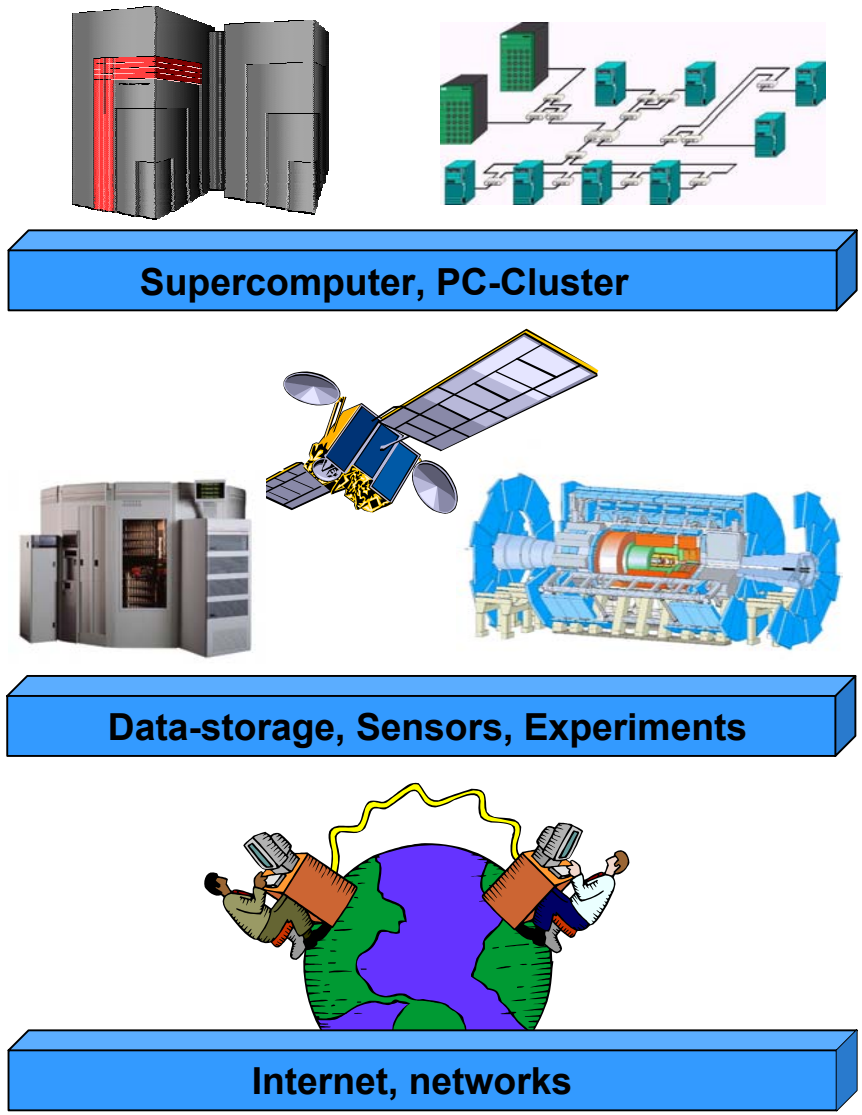
GES

Overview

- What is GRID computing ?
- Why GRIDs ?
- GRID projects world wide
- The Globus Toolkit
- The Open Grid Services Architecture (OGSA)

The Grid Computing Metaphor

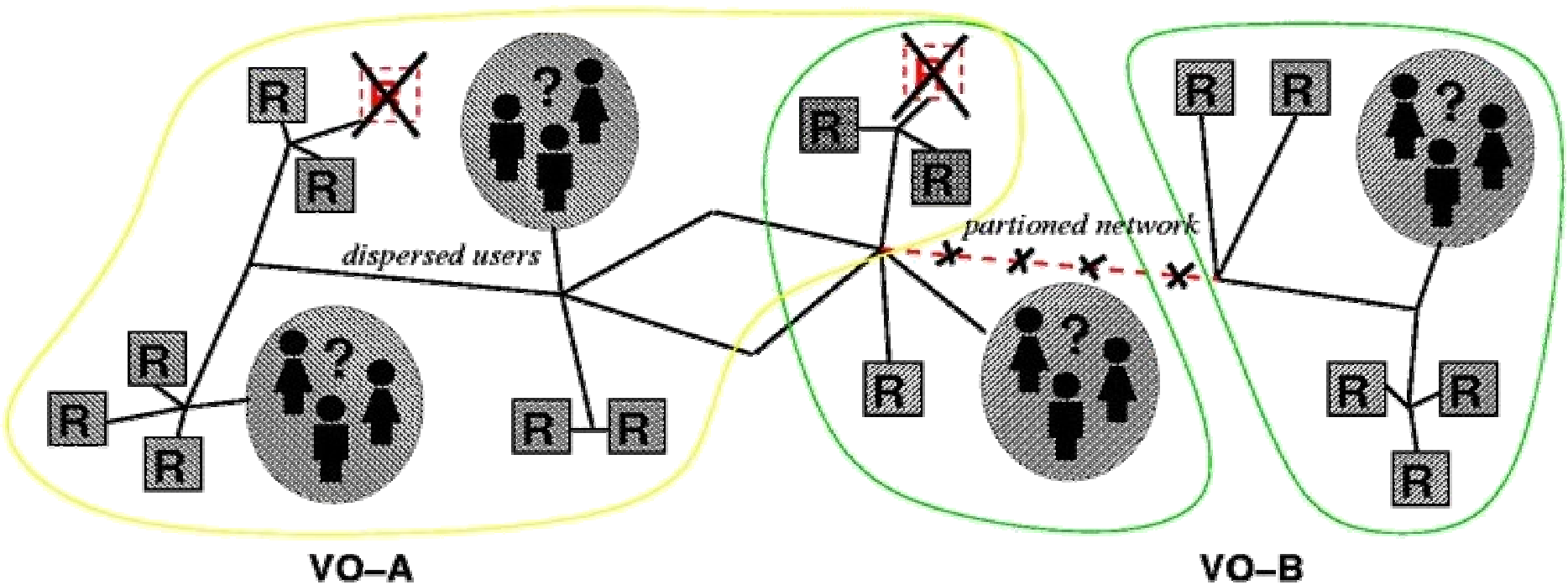




Hoffmann, Reinefeld, Putzer

What is Grid Computing?

Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations



Why Grids? (1) *eScience*

- A biochemist exploits 10,000 computers to screen 100,000 compounds in an hour
- 1,000 physicists worldwide pool resources for peta-op analyses of petabytes of data
- Civil engineers collaborate to design, execute, & analyze shake table experiments
- Climate scientists visualize, annotate, & analyze terabyte simulation datasets
- An emergency response team couples real time data, weather model, population data

Why Grids? (2) *eBusiness*

- Engineers at a multinational company collaborate on the design of a new product
- A multidisciplinary analysis in aerospace couples code and data in four companies
- An insurance company mines data from partner hospitals for fraud detection
- An application service provider offloads excess load to a compute cycle provider
- An enterprise configures internal & external resources to support eBusiness workload

Grids: Why Now?

- Moore's law ⇒ highly functional end-systems
- Ubiquitous Internet ⇒ universal connectivity
- Network exponentials produce dramatic changes in geometry and geography
  - 9-month doubling: double Moore's law!
  - 1986-2001: x340,000; 2001-2010: x4000?
- New modes of working and problem solving emphasize teamwork, computation
- New business models and technologies facilitate outsourcing

Elements of the Problem

- Resource sharing
  - Computers, storage, sensors, networks, …
  - Heterogeneity of device, mechanism, policy
  - Sharing conditional: negotiation, payment, …
- Coordinated problem solving
  - Integration of distributed resources
  - Compound quality of service requirements
- Dynamic, multi-institutional virtual orgs
  - Dynamic overlays on classic org structures
  - Map to underlying control mechanisms

GRID computing at work

- user's identity has to be certified by (mutually recognized) national Certification Authorities (accessing resources belonging to different domains requires identity to be certified).
- secure access to resources is required (security framework to allow resources access only to certified, identified users (X.509 Public Key Infrastructure)).
- resources (node machines) have to be certified by CAs
- temporary delegation from users to processes to be executed "in user's name" (proxy certificates).
- Common agreed policies for accessing resource and handling user's rights across different domains in within the same Virtual Organization a user belongs to.

The Grid World: Current Status

- Dozens of major Grid projects in scientific & technical computing/research & education
  - Deployment, application, technology
- Considerable consensus on key concepts and technologies
  - Open source Globus Toolkit™ a de facto standard for major protocols & services
  - Far from complete or perfect, but out there, evolving rapidly, and large tool/user base
- Global Grid Forum a significant force
- Industrial interest emerging rapidly

GRID projects world wide

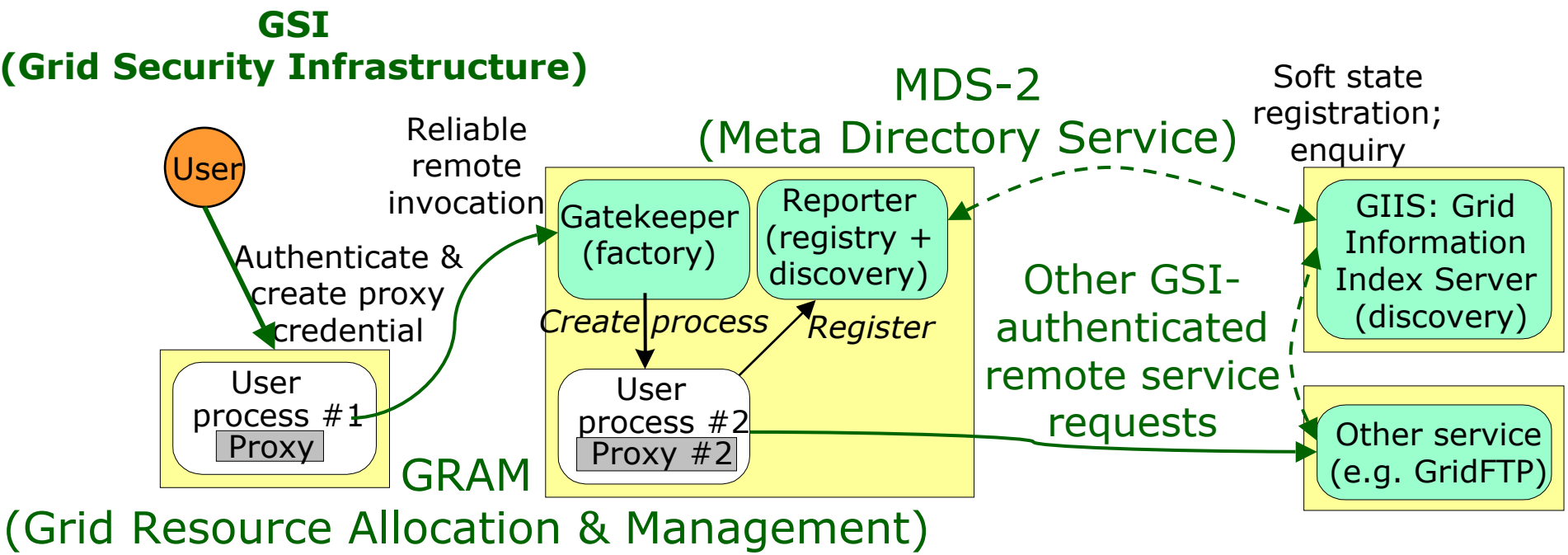
- EU
  - EDG (EU-IST) – R&D EU GRID project [www.edg.org]
  - CrossGRID – QoS – Real Time apps. [www.crossgrid.org]
  - DataTAG
  - GLUE (EU-USA) [www.datatag.org]
  - LCG – The LHC Computing GRID – Deployment [cern.ch/lcg]
  - The new 16,2 B Euro EU VI Framework Prog. GEANT based GRID projects
- USA
  - GriPhyN
  - iVDGL-VDTv1
  - PPDG (NSF, DoE)

  [www.griphyn.org] [www.idvgl.org] [www.ppdg.org]
- Asia
  - ApGrid
  - Pragma (USA-Asia)

  [www.apgrid.org]

The Globus Toolkit in One Slide

- Grid protocols (GSI, GRAM, …) enable resource sharing within virtual orgs; toolkit provides reference implementation
  (= Globus Toolkit services)



- Protocols (and APIs) enable other tools and services for membership, discovery, data mgmt, workflow, …

Globus Toolkit: Evaluation (+)

- Good technical solutions for key problems, e.g.
  - Authentication and authorization
  - Resource discovery and monitoring
  - Reliable remote service invocation
  - High-performance remote data access
- This + good engineering is enabling progress
  - Good quality reference implementation, multi-language support, interfaces to many systems, large user base, industrial support
  - Growing community code base built on tools

Globus Toolkit: Evaluation (-)

- Protocol deficiencies, e.g.
  - Heterogeneous basis: HTTP, LDAP, FTP
  - No standard means of invocation, notification, error propagation, authorization, termination, …
- Significant missing functionality, e.g.
  - Databases, sensors, instruments, workflow, …
  - Virtualization of end systems (hosting envs.)
- Little work on total system properties, e.g.
  - Dependability, end-to-end QoS, …
  - Reasoning about system properties

"Web Services"

- Increasingly popular standards-based framework for accessing network applications
  - W3C standardization; Microsoft, IBM, Sun, others
- WSDL: Web Services Description Language
  - Interface Definition Language for Web services
- SOAP: Simple Object Access Protocol
  - XML-based RPC protocol; common WSDL target
- WS-Inspection
  - Conventions for locating service descriptions
- UDDI: Universal Desc., Discovery, & Integration
  - Directory for Web services

Transient Service Instances

- "Web services" address discovery & invocation of persistent services
  - Interface to persistent state of entire enterprise
- In Grids, must also support transient service instances, created/destroyed dynamically
  - Interfaces to the states of distributed activities
  - E.g. workflow, video conf., dist. data analysis
- Significant implications for how services are managed, named, discovered, and used
  - In fact, much of our work is concerned with the management of service instances

Open Grid Services Architecture

- Service orientation to virtualize resources
- From Web services:
  - Standard interface definition mechanisms: multiple protocol bindings, multiple implementations, local/remote transparency
- Building on Globus Toolkit:
  - Grid service: semantics for service interactions
  - Management of transient instances (& state)
  - Factory, Registry, Discovery, other services
  - Reliable and secure transport
- Multiple hosting targets: J2EE, .NET, "C", …

The Grid Service

- A (potentially transient) Web service with specified interfaces & behaviors, *including*
  - Creation (Factory)
  - Global naming (GSH) & references (GSR)
  - Lifetime management
  - Registration & Discovery
  - Authorization
  - Notification
  - Concurrency
  - Manageability

Globus Toolkit Refactoring

- Grid Security Infrastructure (GSI)
  - Used in Grid service network protocol bindings
- Meta Directory Service 2 (MDS-2)
  - Native part of each Grid service:
    - Discovery, Registry, RegistryManagement, Notification
- Grid Resource Allocation & Mngt (GRAM)
  - Gatekeeper -> Factory for job mgr instances
- GridFTP
  - Refactor control channel protocol
- Other services refactored to used Grid services

Summary: Evolution of Grid Technologies

- Initial exploration (1996-1999; Globus 1.0)
  - Extensive appln experiments; core protocols
- Data Grids (1999-2002; Globus 2.0+)
  - Large-scale data management and analysis
- Open Grid Services Architecture (2002-??, Globus 3.0)
  - Integration w/ Web services, hosting environments, resource virtualization
  - Databases, higher-level services
- Radically scalable systems (2003-??)
  - Sensors, wireless, ubiquitous computing

Summary

- The Grid problem: Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations
- Grid architecture: Protocol, service definition for interoperability & resource sharing
- Globus Toolkit a source of protocol and API definitions—and reference implementations
  - And many projects applying Grid concepts (& Globus technologies) to important problems
- Open Grid Services Architecture represents next step in evolution

For More Information

- The Globus Project™
  - www.globus.org
- Grid architecture
  - www.globus.org/research/papers/anatomy.pdf
- Open Grid Services Architecture
  - www.globus.org/research/papers/ogsa.pdf
  - www.globus.org/research/papers/gsspec.pdf