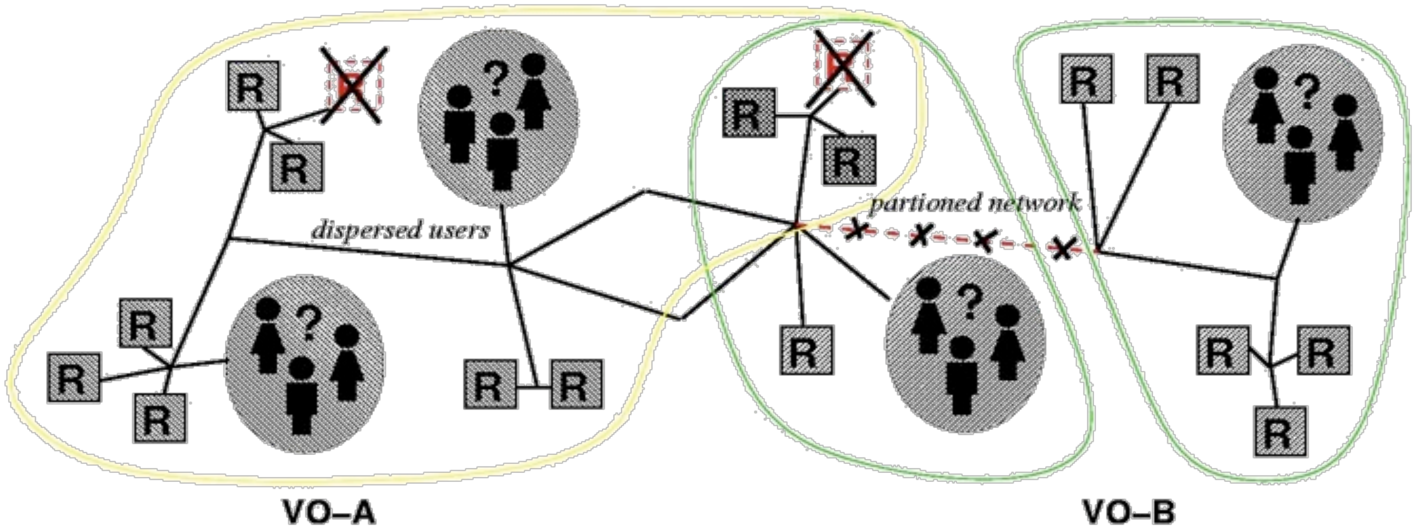
Introduction in Grid Technologies

Dusan Vudragovic
dusan@phy.bg.ac.yu
Scientific Computing Laboratory
Institute of Physics Belgrade, Serbia

Nov. 18, 2008

www.eu-egee.org

Unifying concept: Grid



- **Resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organizations.**

What problems Grid addresses

- Too hard to keep track of authentication data (ID/password) across institutions
- Too hard to monitor system and application status across institutions
- Too many ways to submit jobs
- Too many ways to store & access files/data
- Too many ways to keep track of data
- Too easy to leave “dangling” resources lying around (robustness)

Requirements

- Security
- Monitoring/Discovery
- Computing/Processing Power
- Moving and Managing Data
- Managing Systems
- System Packaging/Distribution
- Secure, reliable, on-demand access to data, software, people, and other resources (ideally all via a Web Browser!)

Ingredients for Grid development

- **Right balance of push and pull factors is needed**
- **Supply side**
  - Technology – inexpensive HPC resources (linux clusters)
  - Technology – network infrastructure
  - Financing – domestic, regional, EU, donations from industry
- **Demand side**
  - Need for novel eScience applications
  - Hunger for number crunching power and storage capacity

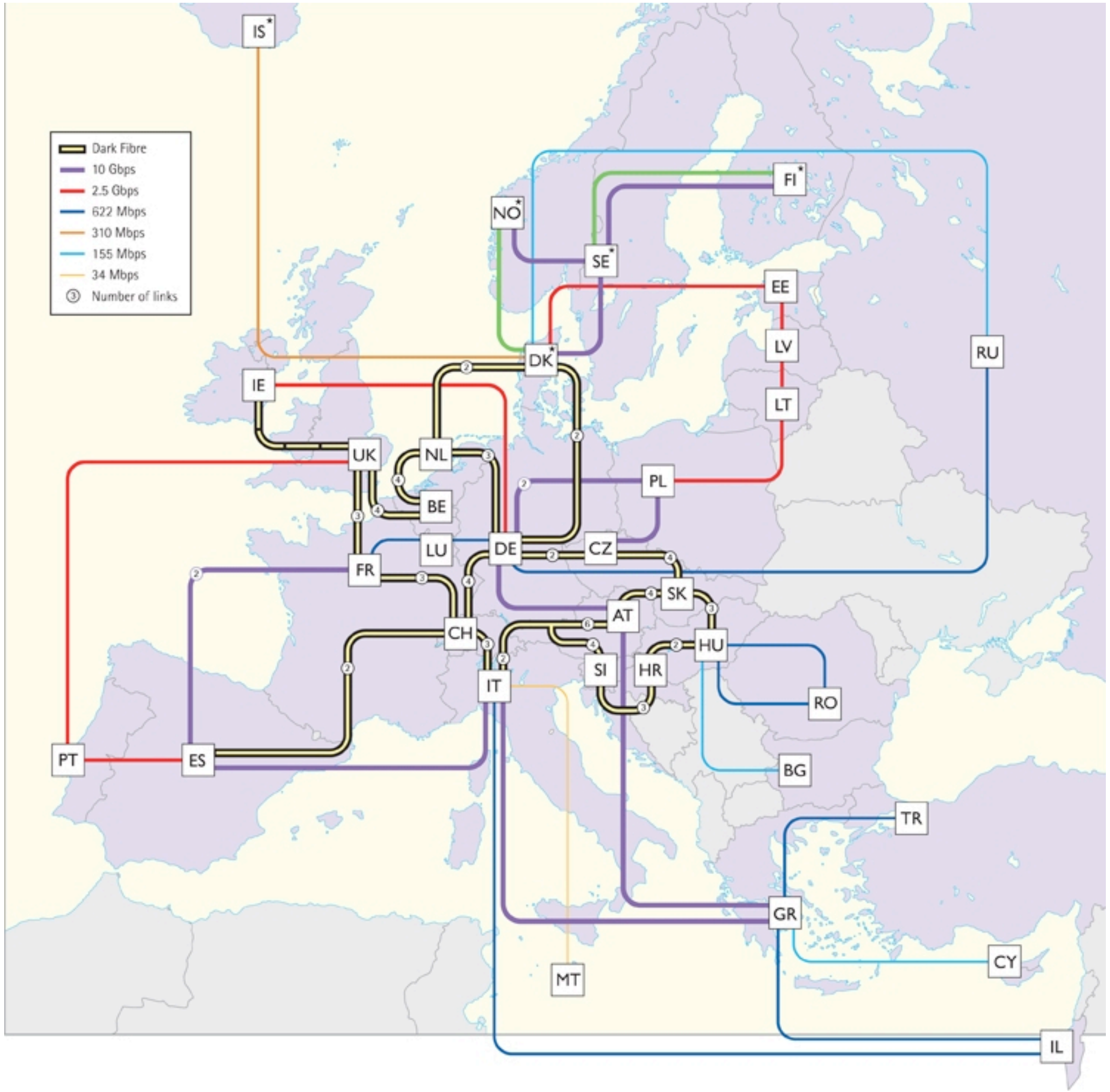
Supply side - cluster

- **The cheapest supercomputers – massively parallel PC clusters**
- **This is possible due to:**
  - Increase in PC processor speed (> Gflop/s)
  - Increase in networking performance (1 Gbs)
  - Availability of stable OS (e.g. Linux)
  - Availability of standard parallel libraries (e.g. MPI)
- **Advantages:**
  - Widespread choice of components/vendors, low price (by factor ~5-10)
  - Long warranty periods, easy servicing
  - Simple upgrade path
- **Disadvantages:**
  - Good knowledge of parallel programming is required
  - Hardware needs to be adjusted to the specific application (network topology)
  - More complex administration
- **Tradeoff: brain power ← → purchasing power**
- **The next step is GRID:**
  - Distributed computing, computing on demand
  - Should “do for computing the same as the Internet did for information” (UK PM, 2002)

Supply side - network

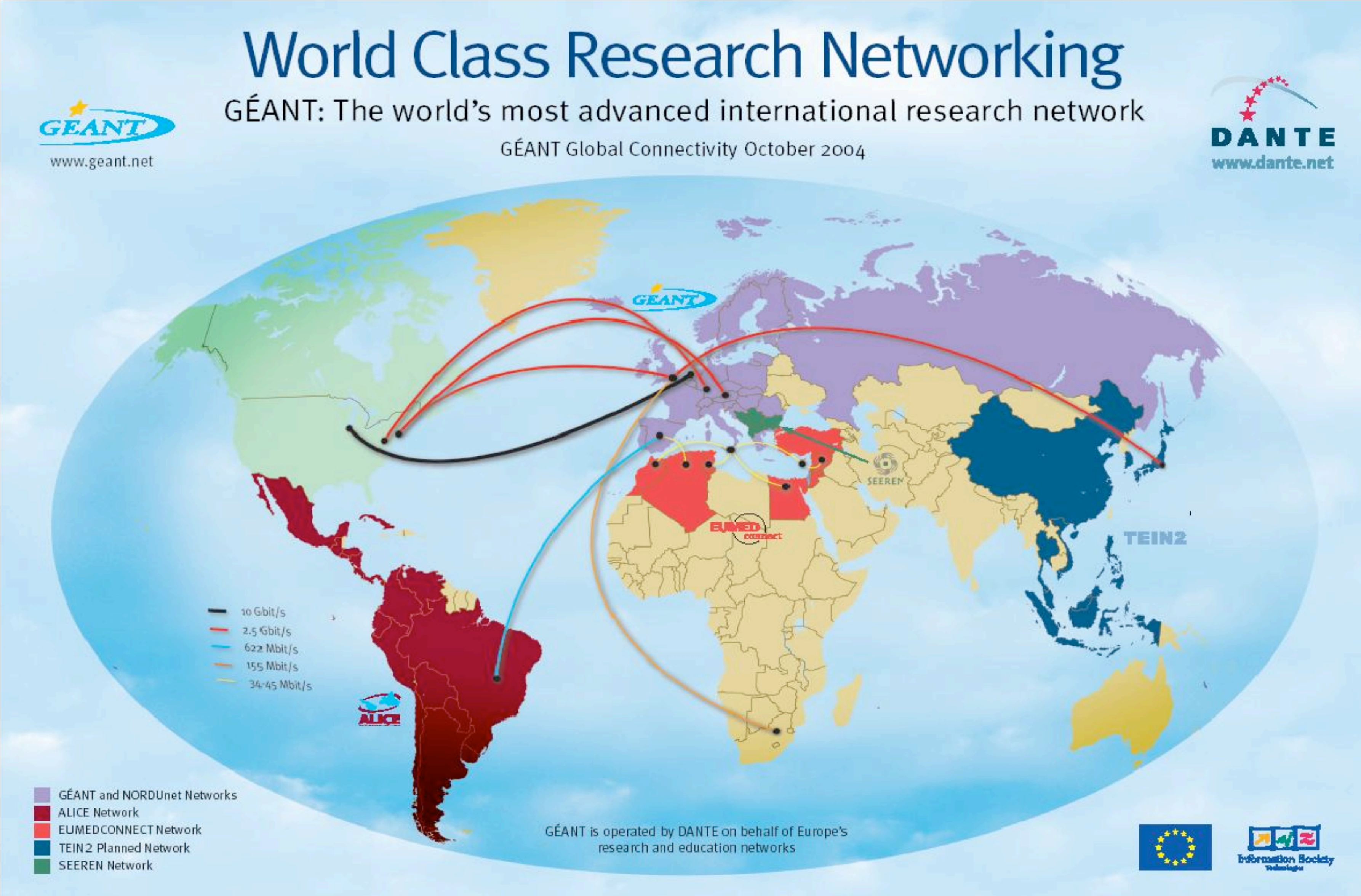
- **Needed at all scales:**
  - World-wide
  - Pan-European (GEANT2)
  - Regional (SEEREN2, …)
  - National (NREN)
  - Campus-wide (WAN)
  - Building-wide (LAN)
- **Remember – it is end user to end user connection that matters**

GÉANT2 Pan-European IP R&E network

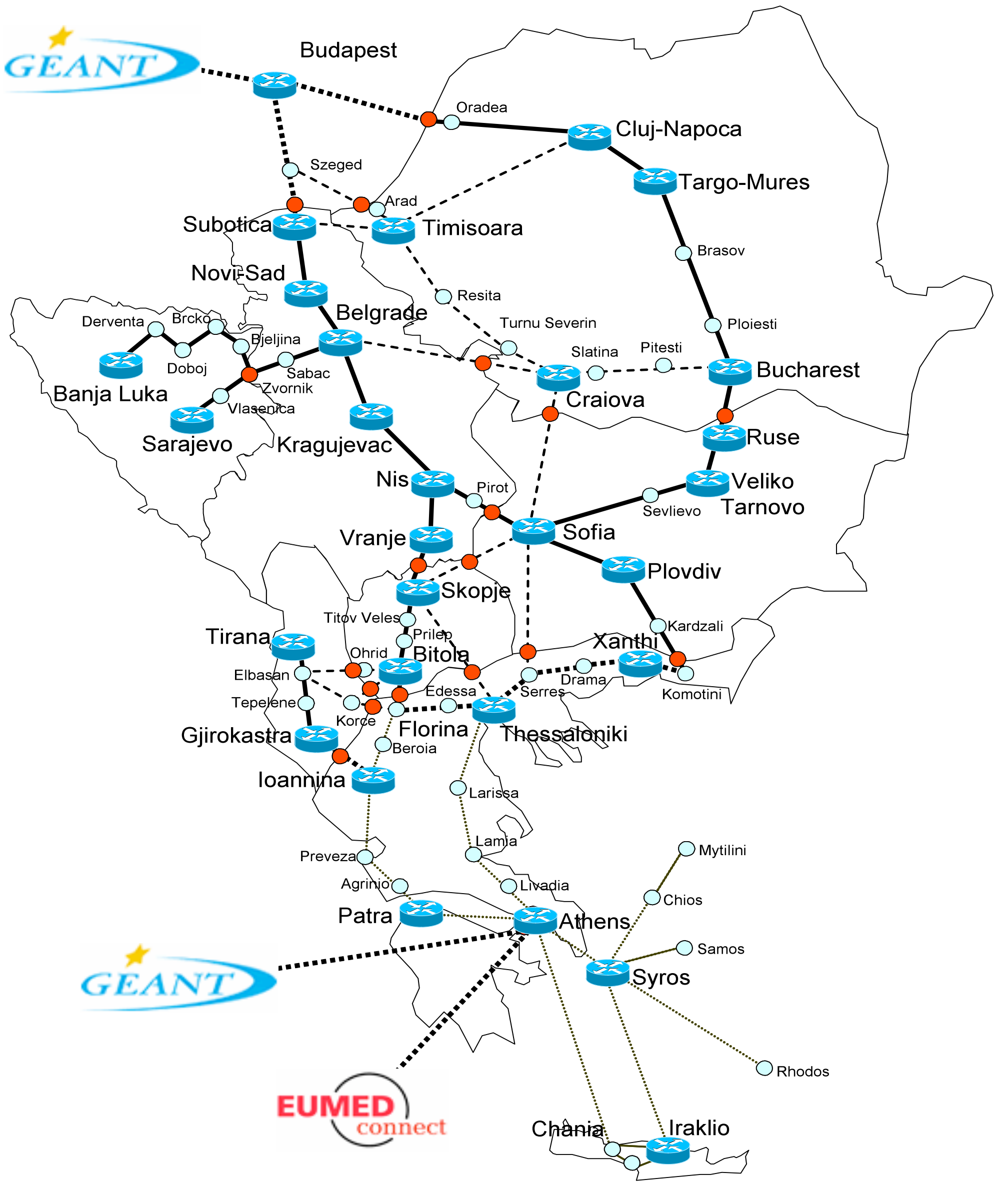


GÉANT2 Global Connectivity

Enabling Grids for E-sciencE



Future development of regional network



Supply side - financing

- **National funding (Ministries responsible for research)**
  - Lobby gvnmt. to commit to Lisbon targets
  - Level of financing should be following an increasing trend (as a % of GDP)
  - Seek financing for clusters and network costs
- **Bilateral projects and donations**
- **Regional initiatives**
  - Networking (HIPERB)
  - Action Plan for R&D in SEE
- **EU funding**
  - FP6 – IST priority, eInfrastructures & GRIDs
  - FP7
  - CARDS
- **Other international sources (NATO, …)**
- **Donations from industry (HP, SUN, …)**

Demand side - eScience

- **Usage of computers in science:**
  - Trivial:
    text editing, elementary visualization, elementary quadrature, special functions, ...
  - Nontrivial:
    differential eq., large linear systems, searching combinatorial spaces, symbolic algebraic manipulations, statistical data analysis, visualization, ...
  - Advanced:
    stochastic simulations, risk assessment in complex systems, dynamics of the systems with many degrees of freedom, PDE solving, calculation of partition functions/functional integrals, ...
- **Why is the use of computation in science growing?**
  - Computational resources are more and more powerful and available (Moore's law)
  - Standard approaches are having problems
    Experiments are more costly, theory more difficult
  - Emergence of new fields/consumers – finance, economy, biology, sociology
- **Emergence of new problems with unprecedented storage and/or processor requirements**

Demand side - consumer

- **Those who study:**
  - Complex discrete time phenomena
  - Nontrivial combinatorial spaces
  - Classical many-body systems
  - Stress/strain analysis, crack propagation
  - Schrodinger eq; diffusion eq.
  - Navier-Stokes eq. and its derivates
  - functional integrals
  - Decision making processes w. incomplete information
  - …
- **Who can deliver? Those with:**
  - Adequate training in mathematics/informatics
  - Stamina needed for complex problems solving
- **Answer: rocket scientists (natural sciences and engineering)**