The DataTAG Project

http://www.datatag.org/

Presentation at University of Twente, The Netherlands
17 September 2002

J.P. Martin-Flatin and Olivier H. Martin
CERN, Switzerland

Project Partners

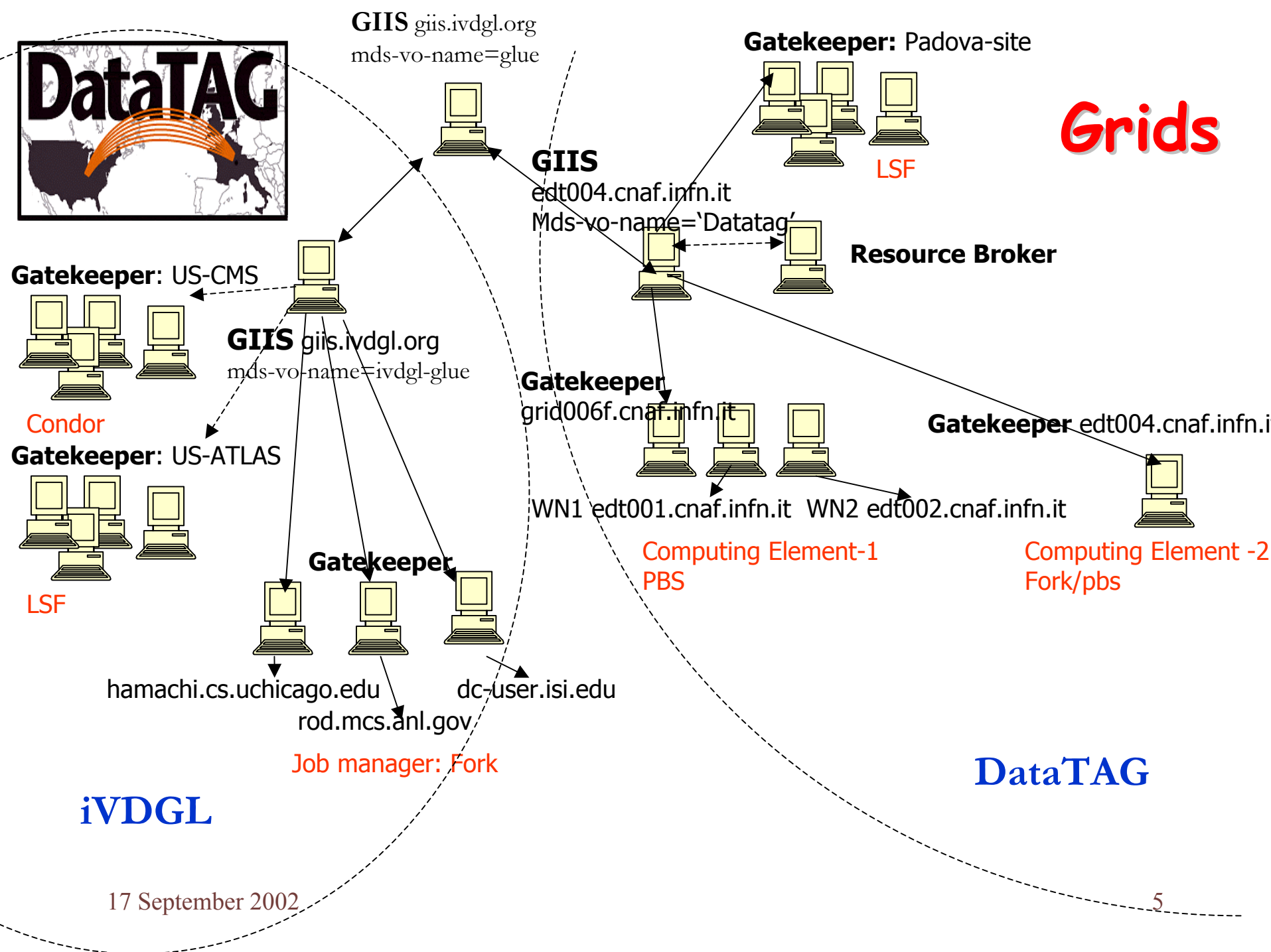
- EU-funded partners: CERN (CH), INFN (IT), INRIA (FR), PPARC (UK) and University of Amsterdam (NL)
- U.S.-funded partners: Caltech, UIC, UMich, Northwestern University, StarLight
- Associate partners: SLAC, ANL, FNAL, Canarie, etc.
- Project coordinator: CERN
  - contact: datatag-office@cern.ch

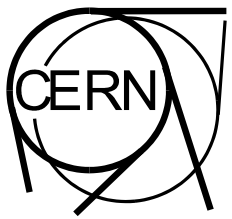
Budget of EU Side

- EUR 3.98M
- Funded manpower: 15 FTE/year
  - 21 FTE recruited
- Start date: January 1, 2002
- Duration: 2 years

Three Objectives

- Build a testbed to experiment with massive file transfers across the Atlantic
- High-performance protocols for gigabit networks underlying data-intensive Grids
- Interoperability between several major Grid projects in Europe and USA


GIIS giis.ivdgl.org
mds-vo-name=glue
Gatekeeper: Padova-site
Grids
LSF
GIIS
edt004.cnaf.infn.it
Mds-vo-name='Datatag'
Resource Broker
Gatekeeper: US-CMS
GIIS giis.ivdgl.org
mds-vo-name=ivdgl-glue
Gatekeeper
grid006f.cnaf.infn.it
Gatekeeper edt004.cnaf.infn.i
Condor
Gatekeeper: US-ATLAS
WN1 edt001.cnaf.infn.it
WN2 edt002.cnaf.infn.it
Computing Element-1
PBS
Computing Element -2
Fork/pbs
Gatekeeper
LSF
hamachi.cs.uchicago.edu
dc-user.isi.edu
rod.mcs.anl.gov
Job manager: Fork
iVDGL
DataTAG



Testbed

Objectives

- Provisioning of 2.5 Gbit/s transatlantic circuit between CERN (Geneva) and StarLight (Chicago)
- Dedicated to research (no production traffic)
- Multi-vendor testbed with layer-2 and layer-3 capabilities:
  - Cisco
  - Alcatel
  - Juniper
- Testbed open to other Grid projects
- Collaboration with GEANT

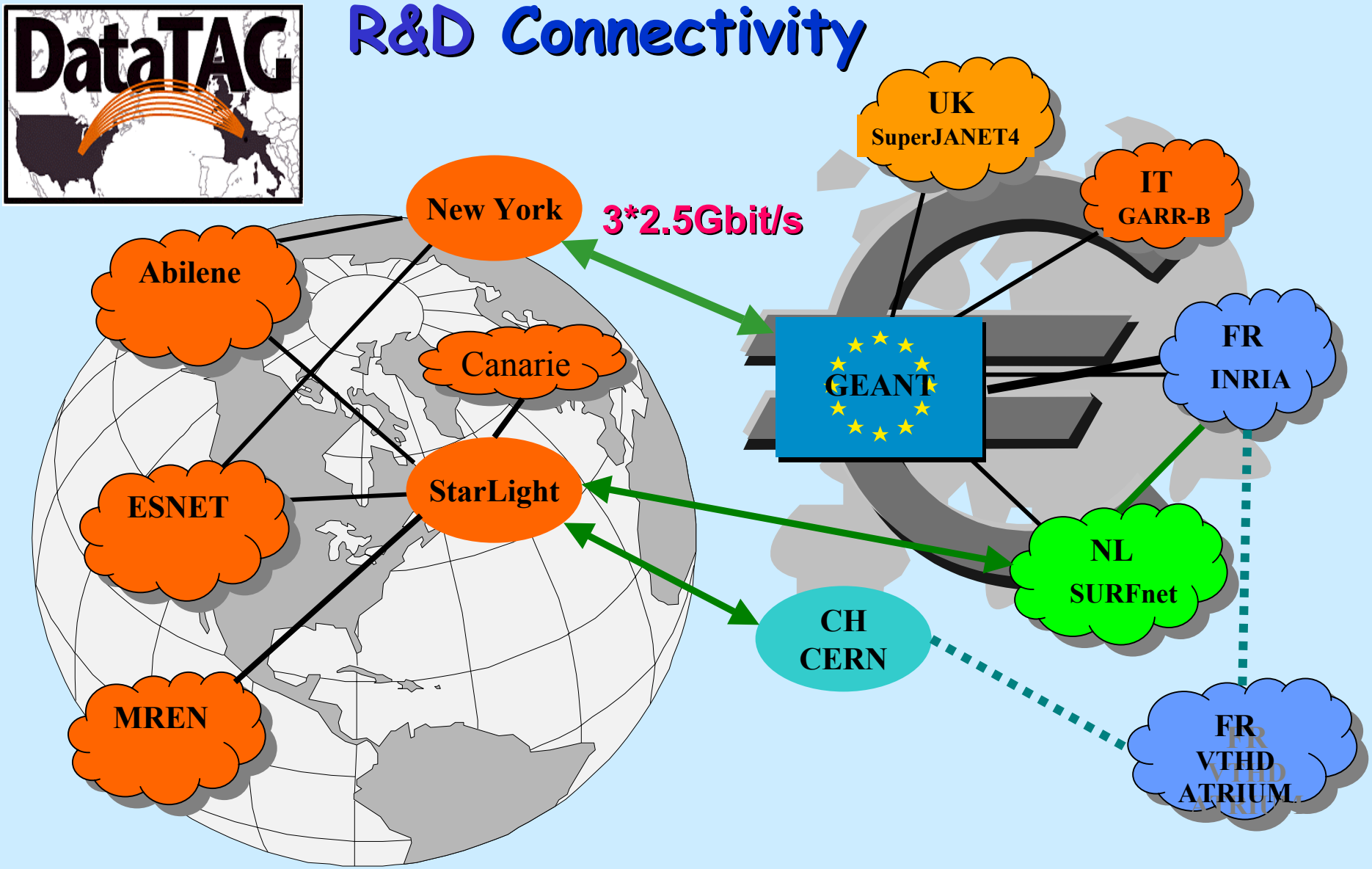
2.5 Gbit/s Transatlantic Circuit

- Operational since 20 August 2002 (T-Systems)
- Delayed by KPNQwest bankruptcy
- Routing plan developed for access across GEANT
- Circuit initially connected to Cisco 76xx routers (layer 3)
- High-end PC servers at CERN and StarLight:
  - SysKonnect GbE
  - can saturate the circuit with TCP traffic
- Layer-2 equipment deployment under way
- Full testbed deployment scheduled for 31 October 2002

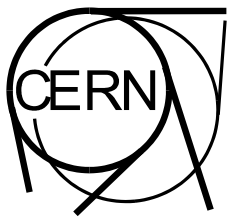
Why Yet Another 2.5 Gbit/s Transatlantic Circuit?

- Most existing or planned 2.5 Gbit/s transatlantic circuits are for production
  - not suitable for advanced networking experiments
- Need operational flexibility:
  - deploy new equipment (routers, GMPLS-capable multiplexers),
  - activate new functionality (QoS, MPLS, distributed VLAN)
- The only known exception to date is the Surfnet circuit between Amsterdam and Chicago (StarLight)

R&D Connectivity



Major R&D 2.5 Gbit/s circuits between Europe & USA



Network Research

DataTAG Activities

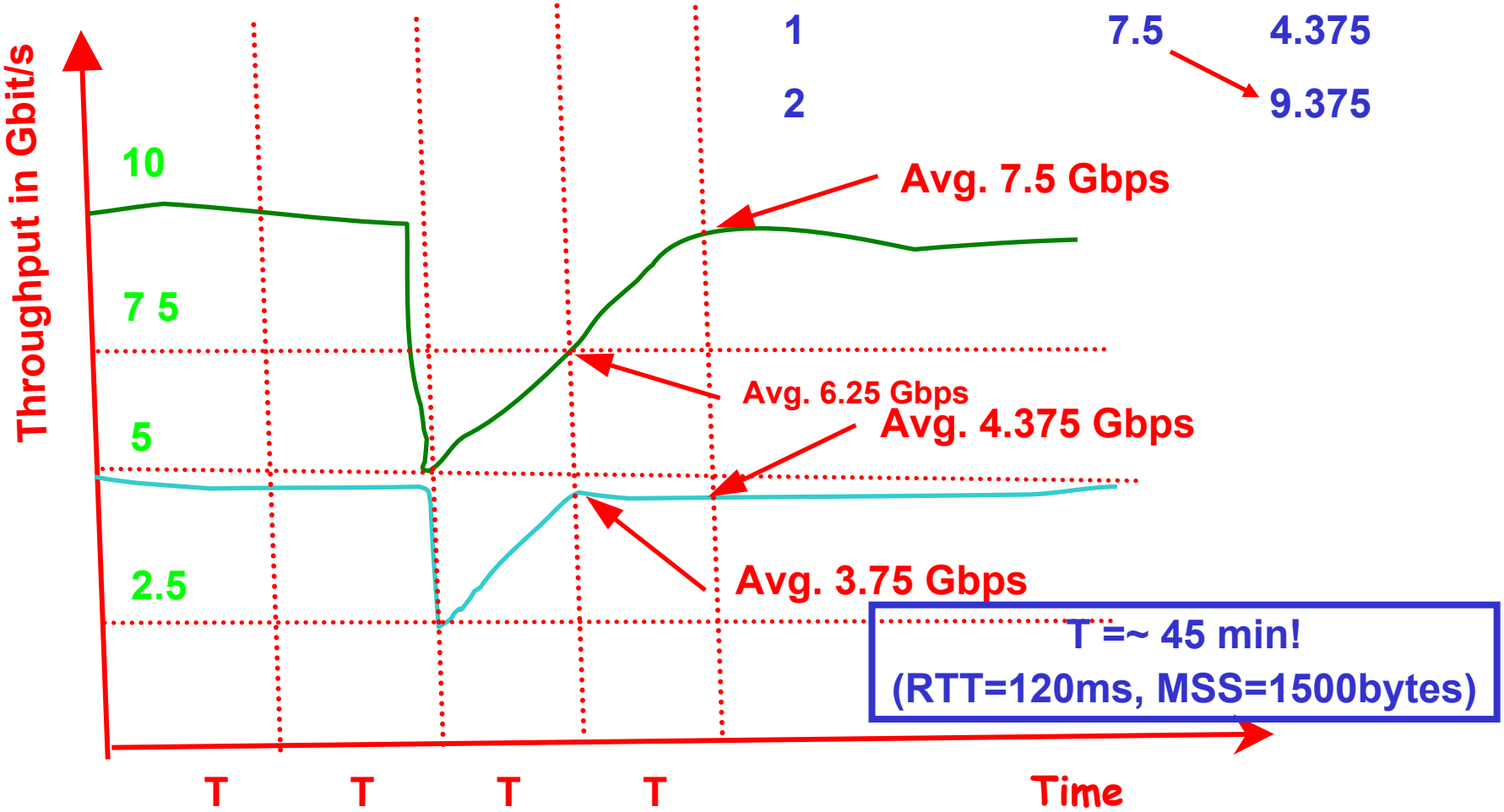
- Enhance TCP performance
  - modify Linux kernel
- Monitoring
- QoS
  - LBE (Scavenger)
- Bandwidth reservation
  - AAA-based bandwidth on demand
  - lightpath managed as a Grid resource

TCP Performance Issues

- TCP's current congestion control (AIMD) algorithms are not suited to gigabit networks
  - long time to recover from packet loss
- Line errors are interpreted as congestion
- Delayed ACKs + large window size + large RTT = problem

Single vs. Multiple Streams: Effect of a Single Packet Loss

Streams/Throughput	10	5
1	7.5	4.375
2		9.375



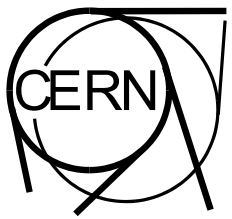
Responsiveness

Capacity	RTT	# inc	Responsiveness
9.6 kbit/s	40 ms	1	0.6 ms
10 Mbit/s	20 ms	8	150 ms
622 Mbit/s	120 ms	~2,900	~6 min
2.5 Gbit/s	120 ms	~11,600	~23 min
10 Gbit/s	120 ms	~46,200	~1h 30min

inc size = MSS = 1,460

Research Directions

- New fairness principle
- Change multiplicative decrease:
  - do not divide by two
- Change additive increase
  - binary search
  - local and global stability
- Caltech technical report CALT-68-2398
- Estimation of the available capacity and bandwidth*delay product:
  - on the fly
  - cached

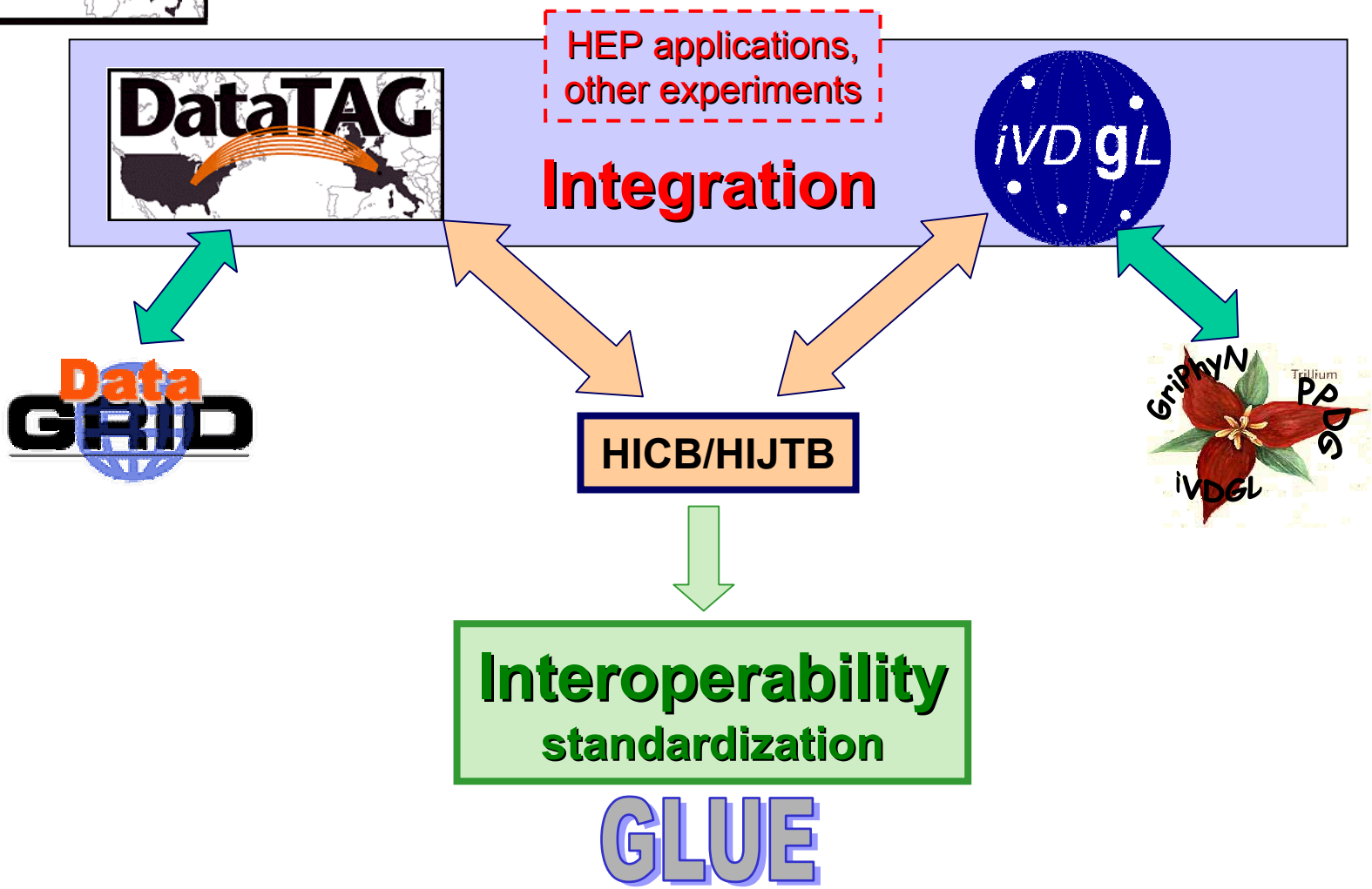


Grid Interoperability

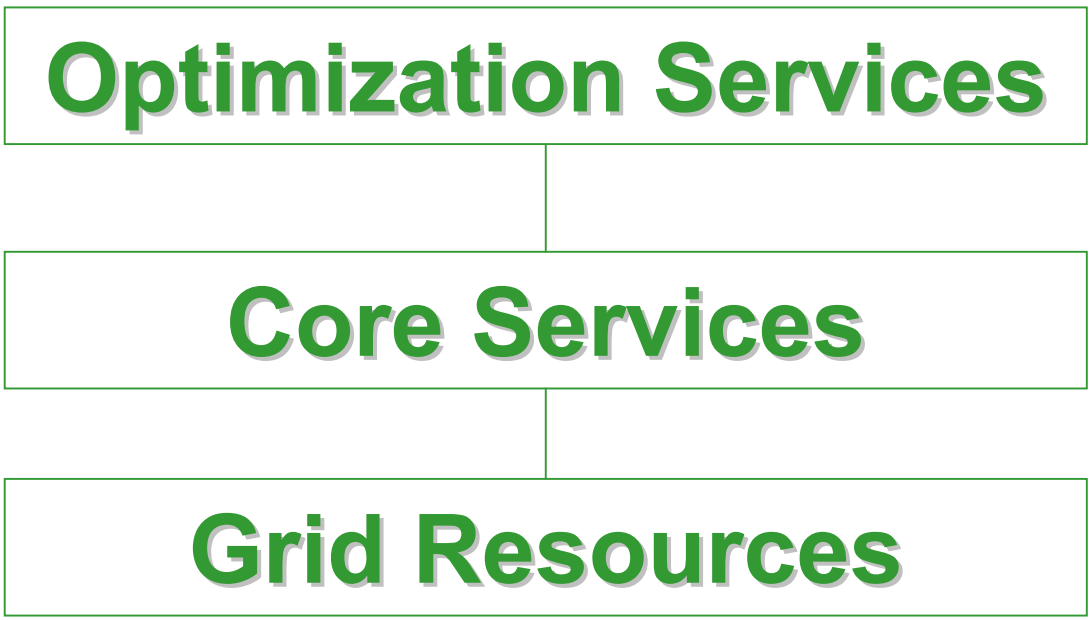
Objectives

- Interoperability between European and US Grids
  - Middleware integration and coexistence
  - GLUE = Grid Lab Uniform Environment
    - integration & standardization
    - testbed and demo
- Enable a set of applications to run on the transatlantic testbed:
  - CERN LHC experiments: ATLAS, CMS, Alice
  - other experiments: CDF, D0, BaBar, Virgo, Ligo, etc.

Relationships



Interoperability Framework

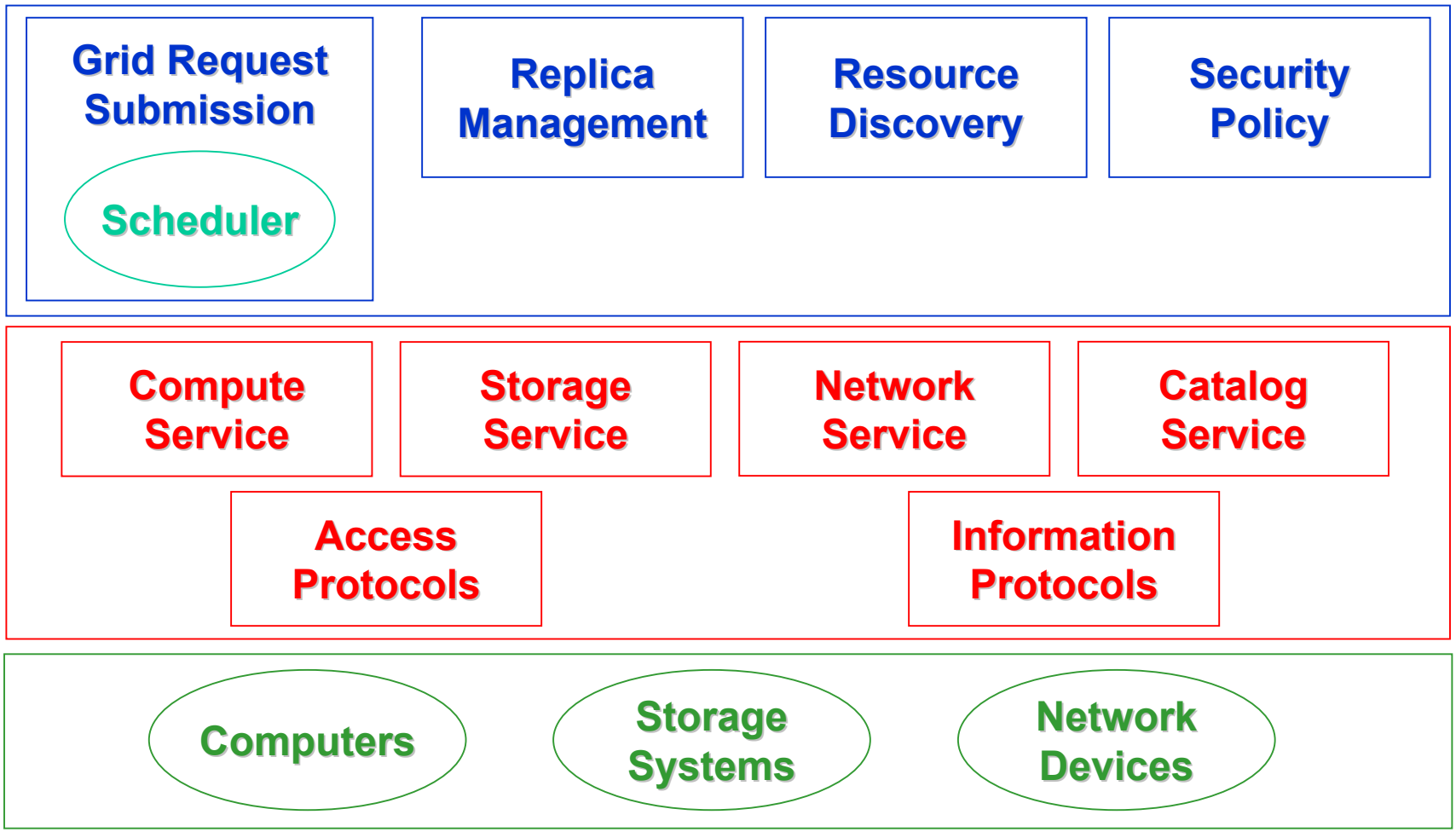


must coexist

must be common to all Grids

CE, SE, NE

Grid Software Architecture

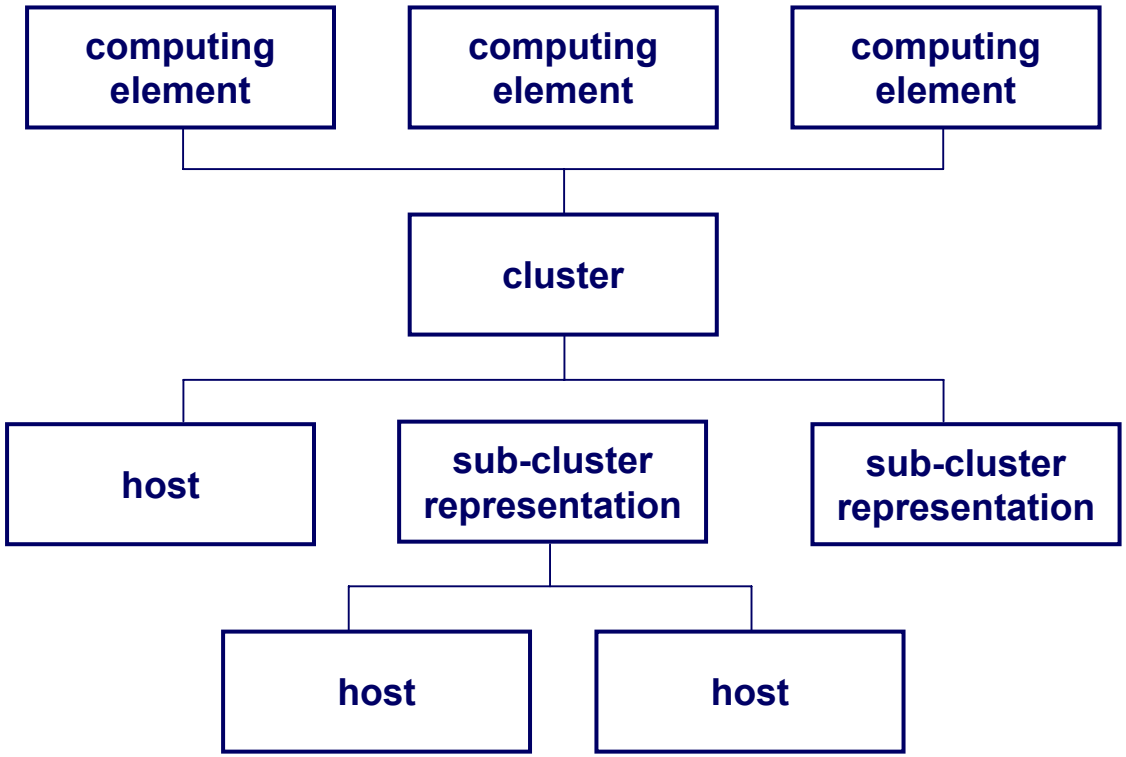


Status of GLUE Activities

- Resource discovery and GLUE schema
  - computing element
  - storage element
  - network element
- Authentication across organizations
- Minimal authorization
- Unified service discovery
- Common software deployment procedures

Resource Discovery and GLUE Schema

Computing Resources Structure Description



Entry point into queuing system

Container groups subcluster or nodes

Homogeneous collection of nodes

Physical computing nodes

Future GLUE Activities

- Data movement:
  - GridFTP
  - replica location service
- Advanced authorization:
  - cross-organization, community-based authorization

Demos

- iGrid 2002
  - US16 with University of Michigan
  - US14 with Caltech and ANL
  - CA03 with Canarie
- IST 2002
- SC 2002

Summary

- Gigabit testbed for data-intensive Grids:
  - Layer 3 in place
  - Layer 2 being provisioned
  - Modified version of TCP to improve performance
- Grid interoperability:
  - GLUE schema for resource discovery
  - Working on common authorization solutions
  - Evaluation of software deployment tools
  - First interoperability tests on heterogeneous transatlantic testbeds