“Execution of a Bioinformatics Application in a Joint IRISGrid/EGEE Testbed”

José Luis Vázquez-Poletti

Eduardo Huedo (huedoce@inta.es)

Rubén S. Montero

Ignacio M. Llorente

Advanced Computing Laboratory
Centre for Astrobiology (INTA-CSIC)

Distributed Systems Architecture and Security group
Complutense University of Madrid

1st LaSCoG Workshop / PPAM 2005 Conference

Objectives

- Demonstrate the feasibility of building *loosely-coupled* Grid environments:
  - based only on Globus services, while
  - obtaining non trivial levels of quality of service through appropriate user-level Grid middleware.
- Resolve the problem of:
  - using several testbeds simultaneously (from an user's viewpoint), and
  - contribute the same resources to more than one project (from an administrator's viewpoint).
- Don’t try to:
  - tailor the core Grid middleware to our needs (since in such case the resulting infrastructure would be application specific), nor
  - homogenize the underlying resources (since in such case the resulting infrastructure would be a highly distributed cluster).

The Grid Philosophy

A grid is a system that...

1) ...coordinates resources that are not subject to a centralized control...

2) ...using standard, open, general-purpose protocols and interfaces...

3) ...to deliver nontrivial qualities of services.

Ian Foster
What is the Grid? A Three Point Checklist (2002)

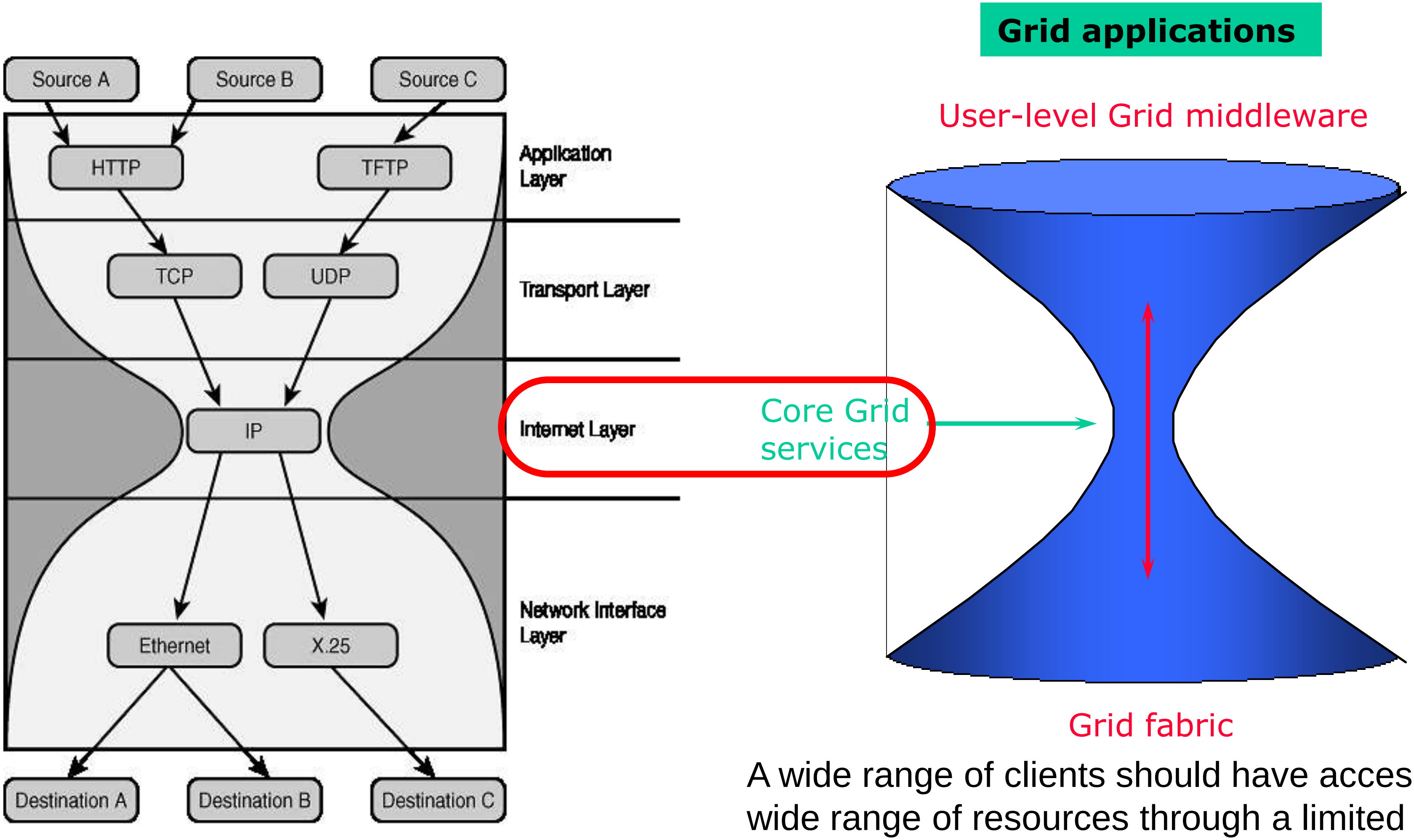
Loosely-Coupled Grids

- In a *loosely-coupled* grid, the different layers of the infrastructure should be separated from each other, being only communicated with a limited and well defined set of interfaces and protocols.
- These layers are:
  - Grid fabric
  - core Grid middleware
  - user-level Grid middleware, and
  - Grid applications.

TCP/IP and Globus

The IP hourglass model

The Globus hourglass model



Grid applications

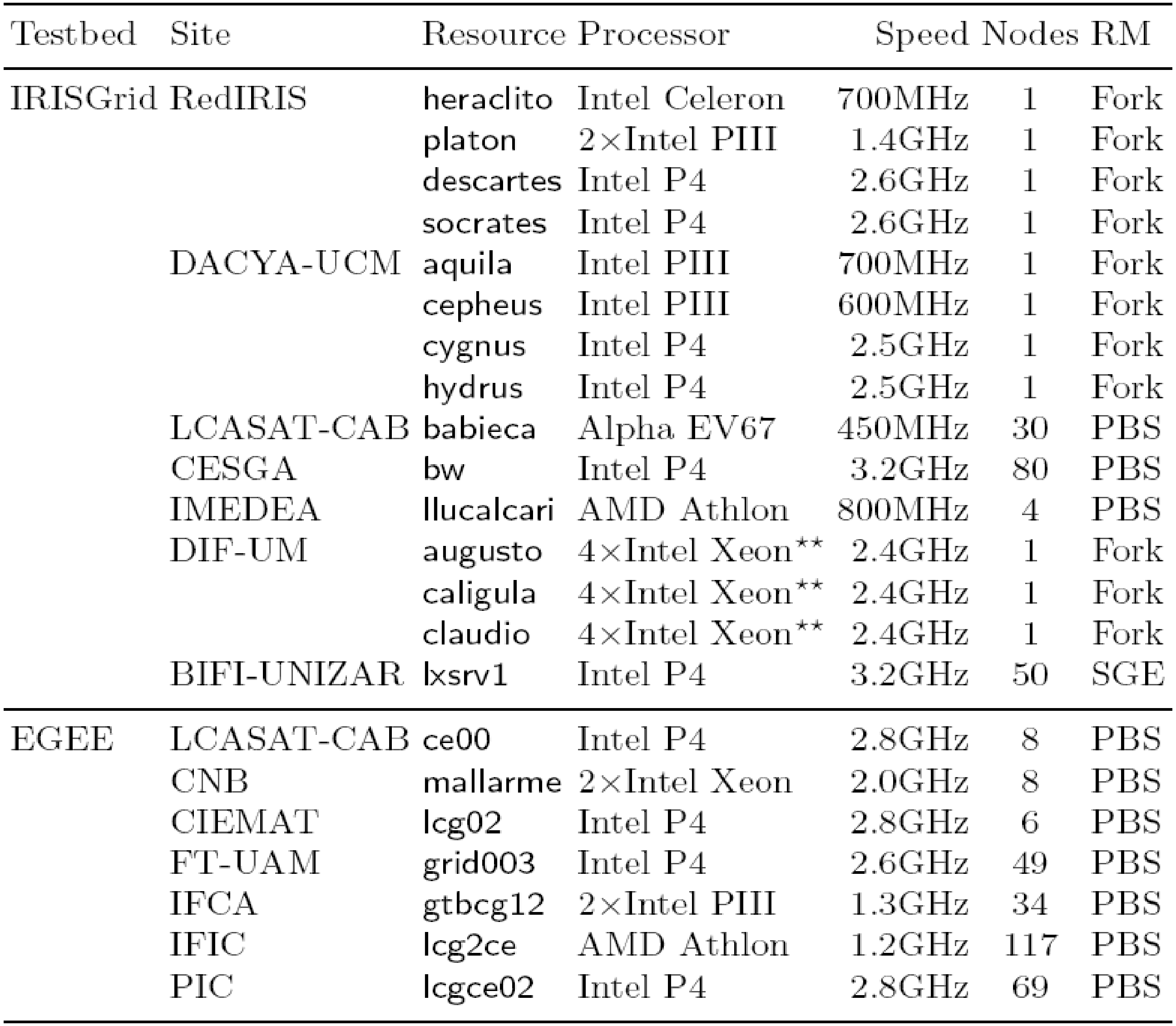
User-level Grid middleware

Core Grid services

Grid fabric

A wide range of clients should have access to a wide range of resources through a limited and standardized set of protocols and interfaces.

Grid Fabric: IRISGrid and EGEE resources



Testbed	Site	Resource	Processor	Speed	Nodes	RM
IRISGrid	RedIRIS	heraclito	Intel Celeron	700MHz	1	Fork
		platon	2×Intel PIII	1.4GHz	1	Fork
		descartes	Intel P4	2.6GHz	1	Fork
		socrates	Intel P4	2.6GHz	1	Fork
	DACYA-UCM	aquila	Intel PIII	700MHz	1	Fork
		cepheus	Intel PIII	600MHz	1	Fork
		cygnus	Intel P4	2.5GHz	1	Fork
		hydrus	Intel P4	2.5GHz	1	Fork
	LCASAT-CAB	babieca	Alpha EV67	450MHz	30	PBS
	CESGA	bw	Intel P4	3.2GHz	80	PBS
	IMEDEA	llucalcari	AMD Athlon	800MHz	4	PBS
	DIF-UM	augusto	4×Intel Xeon**	2.4GHz	1	Fork
		caligula	4×Intel Xeon**	2.4GHz	1	Fork
		claudio	4×Intel Xeon**	2.4GHz	1	Fork
	BIFI-UNIZAR	lxsrv1	Intel P4	3.2GHz	50	SGE
EGEE	LCASAT-CAB	ce00	Intel P4	2.8GHz	8	PBS
	CNB	mallarme	2×Intel Xeon	2.0GHz	8	PBS
	CIEMAT	lcg02	Intel P4	2.8GHz	6	PBS
	FT-UAM	grid003	Intel P4	2.6GHz	49	PBS
	IFCA	gtbcg12	2×Intel PIII	1.3GHz	34	PBS
	IFIC	lcg2ce	AMD Athlon	1.2GHz	117	PBS
	PIC	lcgce02	Intel P4	2.8GHz	69	PBS

irisgrid

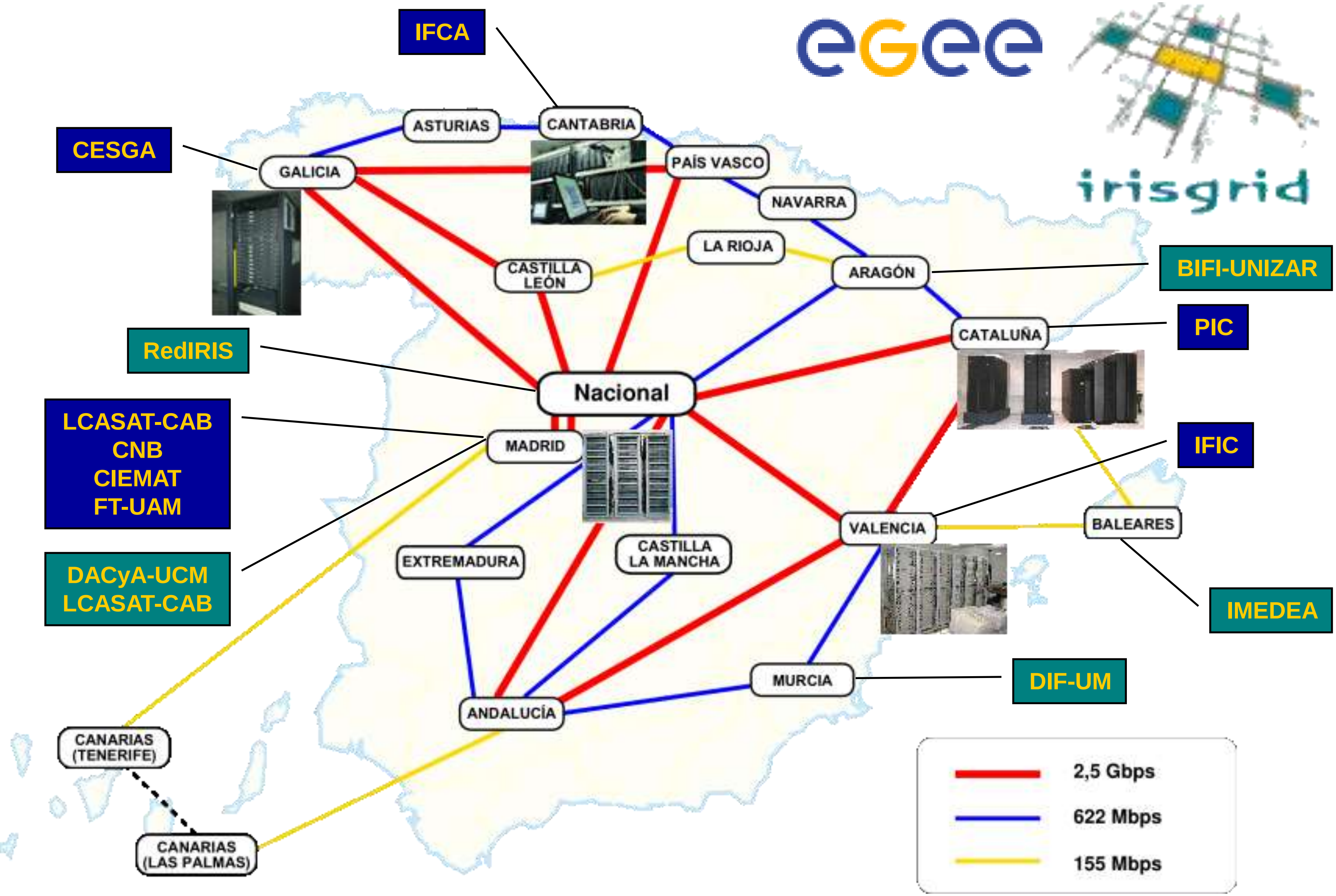
7 sites and 195 CPUs

egee

7 sites and 333 CPUs

Total: 13 sites and 528 CPUs. Limitation of 4 running jobs per resource (64 CPUs)

Grid Fabric: IRISGrid and EGEE resources



IFCA

CESGA

RedIRIS

LCASAT-CAB
CNB
CIEMAT
FT-UAM

DACyA-UCM
LCASAT-CAB

BIFI-UNIZAR

PIC

IFIC

IMEDEA

DIF-UM

🟥 2,5 Gbps

🟦 622 Mbps

🟨 155 Mbps

Core Grid Middleware: Globus

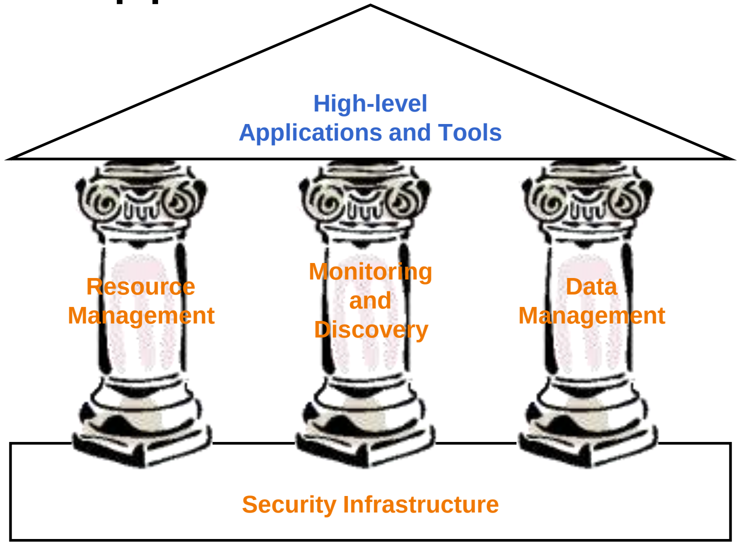
Globus allows **secure remote operation** over **multiple administration domains** with different **resource management systems** and **access policies**.

Globus is…

- a set of services, commands, libraries and APIs
- a *software* infrastructure, or ***middleware***.

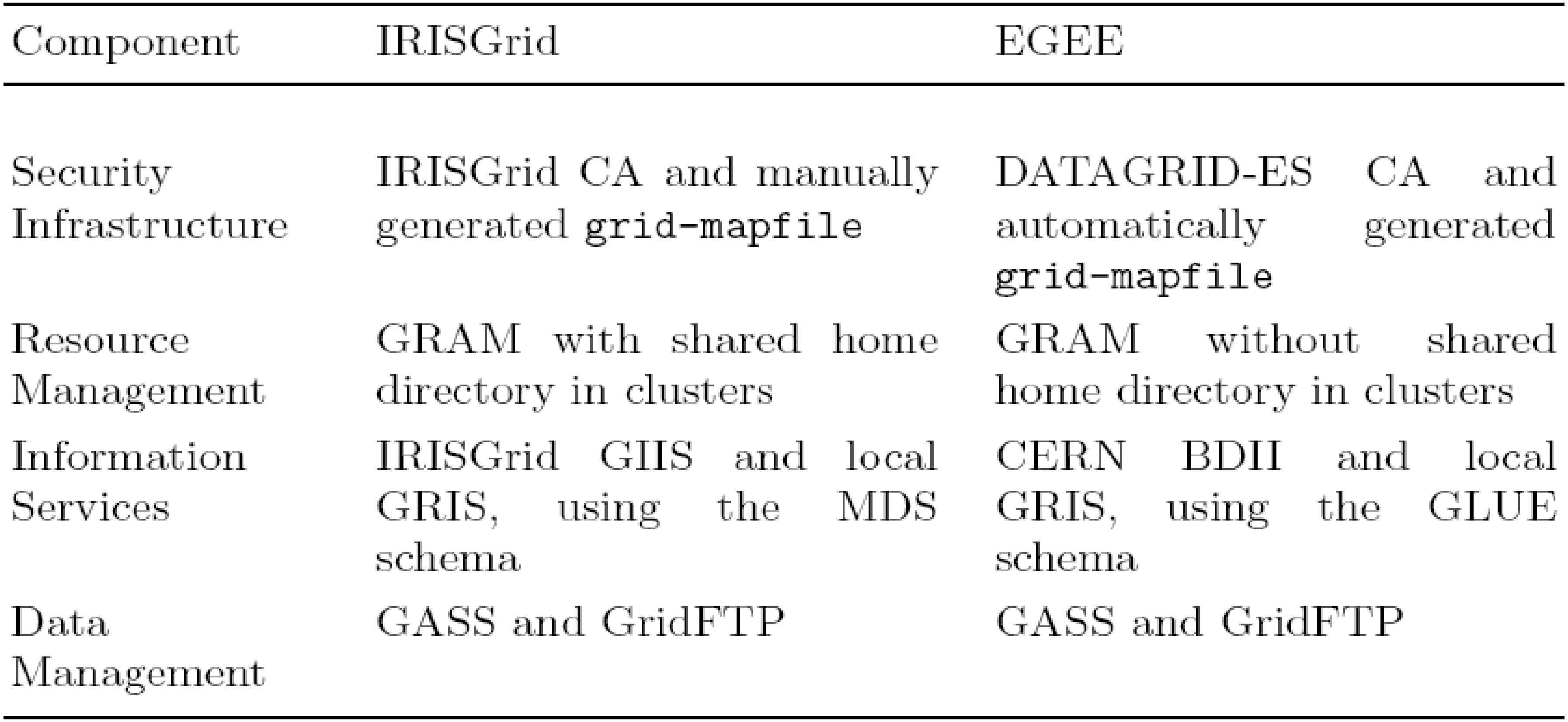
Globus is **NOT**...

- a scheduler, a resource broker or an application
- an end-user tool.



Core Grid Middleware: Globus

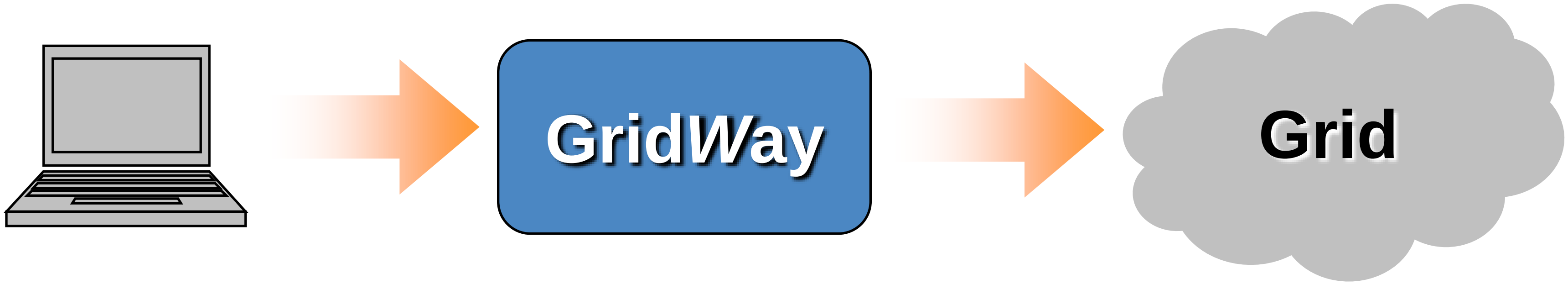
Globus Toolkit (GT2.X y GT3.X), with the following core pre-WS Grid services:



Component	IRISGrid	EGEE
Security Infrastructure	IRISGrid CA and manually generated `grid-mapfile`	DATAGRID-ES CA and automatically generated `grid-mapfile`
Resource Management	GRAM with shared home directory in clusters	GRAM without shared home directory in clusters
Information Services	IRISGrid GIIS and local GRIS, using the MDS schema	CERN BDII and local GRIS, using the GLUE schema
Data Management	GASS and GridFTP	GASS and GridFTP

User-Level Grid Middleware: GridWay

Easier and efficient execution in dynamic and heterogeneous grids in a ***submit & forget*** fashion.



Functionality:

- **Adaptive scheduling**
- **Adaptive execution**
- **High throughput apps.**

Design Guidelines:

- **Adaptable/extensible** (modular design)
- **Scalable** (decentralized architecture)
- **Deployable** (user, standard services)
- **Applicable** (wide application range)

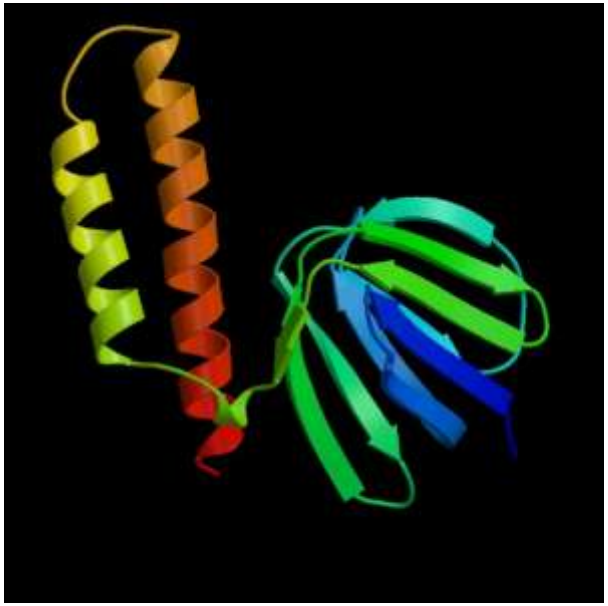
Grid Application: Computational Proteomics

Protein structure prediction and thermodynamic studies from their aminoacid sequences by means of *threading* methods.

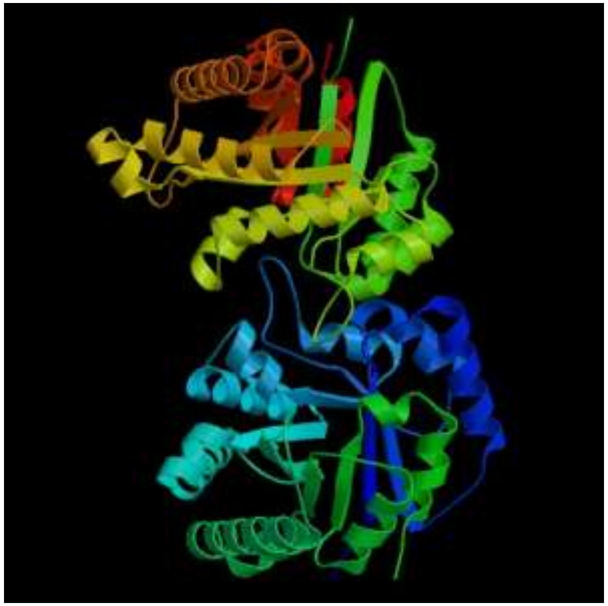
Application to families of **orthologous roteins** ⇒ **High Throughput**

–MTYHLDVVSAEQQMFSGLVEKIQVTGSEGELGIYPGHAPLLTAIKP
GMIRIVKQHGHEEFIYLSGGILEVQPGNVTVLADTAIRGQDLDEARA
MEAKRKAEEHISSSHGDVDYAQASAELAKAIAQLRVIELTKK

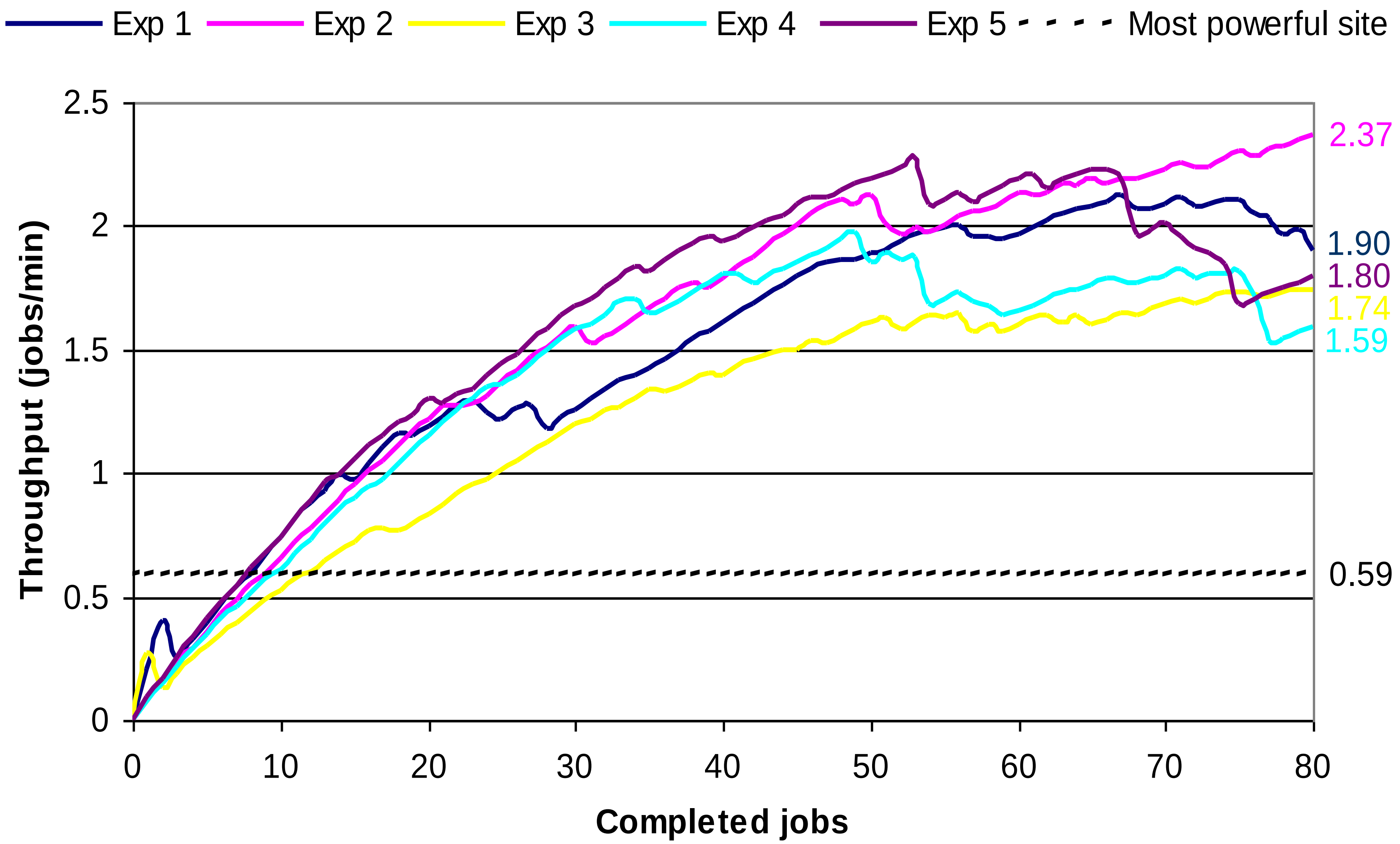
ATP Sintase (ε chain)



Triose Phosphate Isomerase

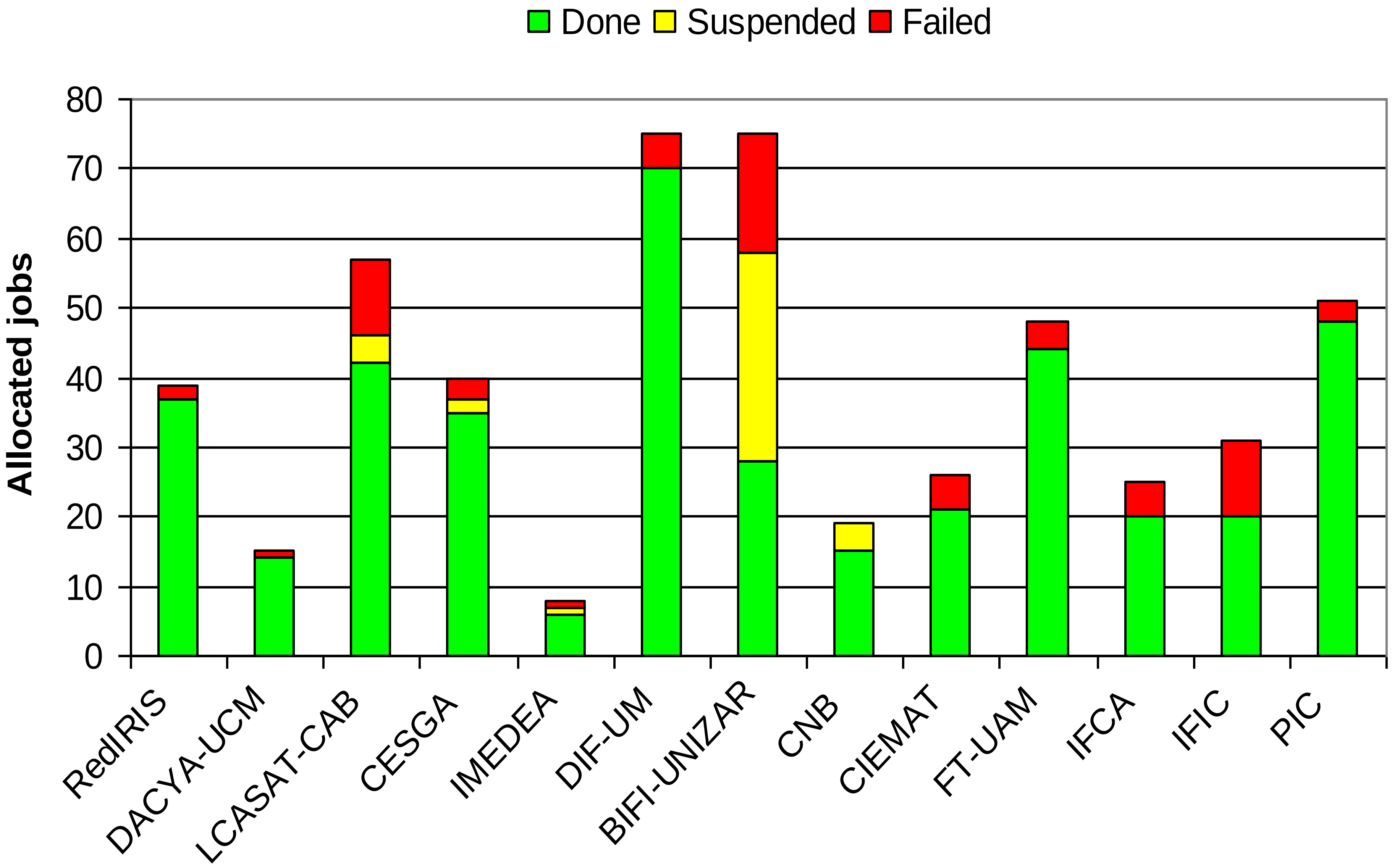


Results: Dynamic Throughput



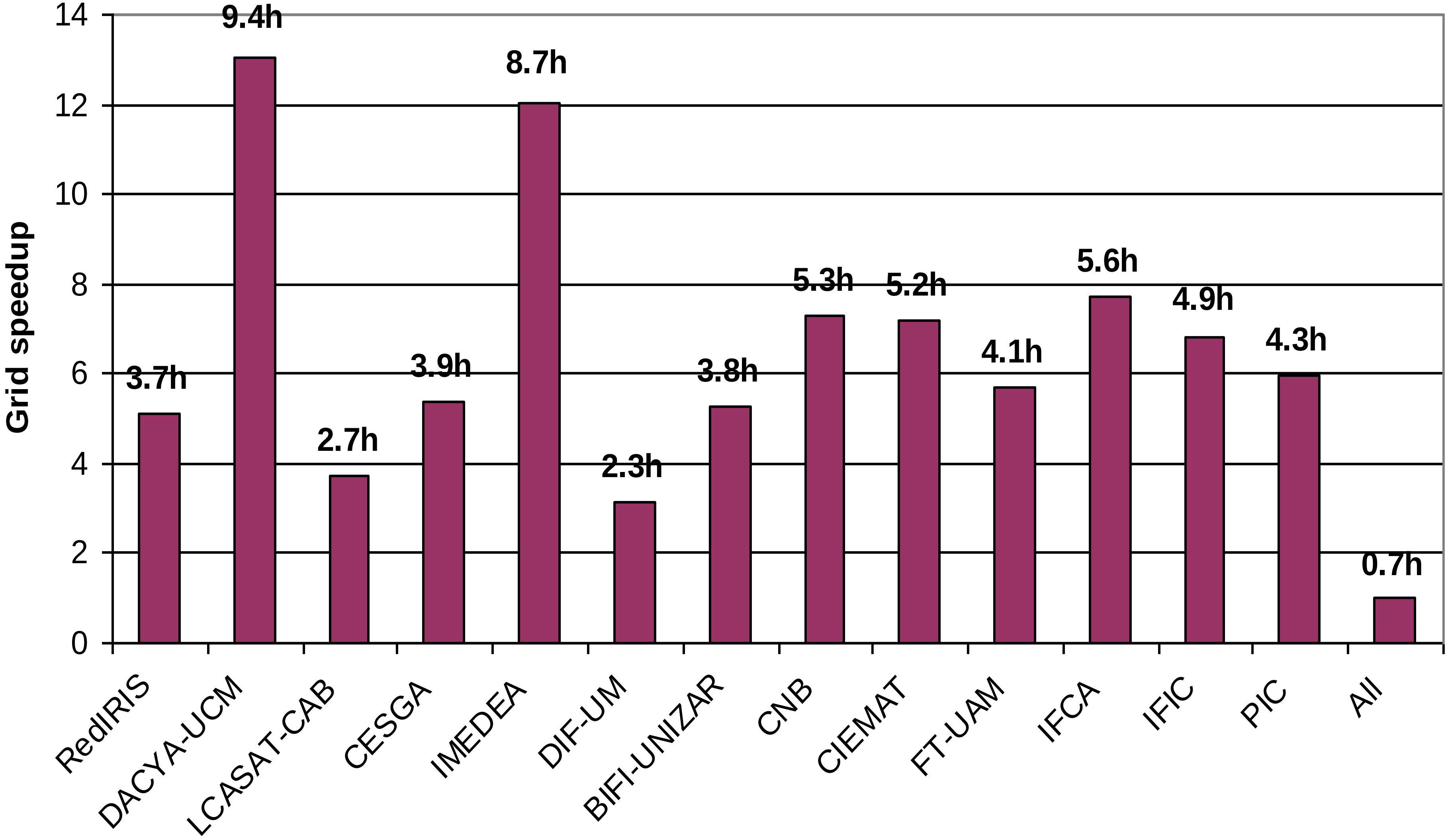
Result: Schedule

Aggregated schedule performed during the five experiments



Results: Grid Speedup

$S_{Grid}=T_{site}/T_{Grid}$



Conclusions

- Grid*W*ay, as user-level Grid middleware, can work with Globus, as a standard core Grid middleware, over any Grid fabric in a *loosely-coupled* way.
- The Grid*W*ay approach (the Grid way), based on a modular and decentralized architecture, is appropriate for the Grid.
- Advantages of *loosely-coupled* grids:
  - They allow a straightforward resource sharing, since resources are accessed and exploited through de facto standard protocols and interfaces, similar to the early stages of the Internet.
  - They allow an easier, scalable and compatible deployment.
  - They reduce the firewall configuration to a minimum, which is also welcome by the security administrators.