Centro Comunicaciones CSIC RedIRIS.
Spanish NREN

Distributed Systems Architecture and Security Group
Dpto. de Arquitectura de Computadores y Automática
Universidad Complutense de Madrid

Advanced Computing Laboratory
Centro de Astrobiología
Associated to *NASA Astrobiology Institute*
CSIC – INTA

A Grid Scheduling Algorithm Considering Dynamic Interconnecting Network

Antonio Fuentes,   Eduardo Huedo,   Ruben S. Moreno,   Ignacio Martín Llorente

Antonio Fuentes

antonio.fuentes@rediris.es

OutLine

Motivation

1. Need use dynamic bandwidth. Why? The network is not only for Grid Computing.

2. Need Selection resources strategy considering interconnecting Network.

   1. Evaluation the transfer costs

Is my job in the best resource?

Why does it spend a lot of time?

*Is the sheduled considering the **bandwith**?*
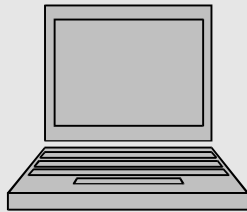
*Is the scheduled considering the **latency**?*

*Is the scheduled considering **the best way**?*

Applications and FrameWork

GridWay FrameWork

Provides an **easier** and more **efficient** execution (*submit & forget*) on heterogeneous and dynamic Grids
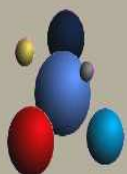

Grid *Way*

Grid

- Easily **adaptable** (modular design)

- Easily **scalable** (decentralized architecture)

- Easily **deployable** (user, standard services)
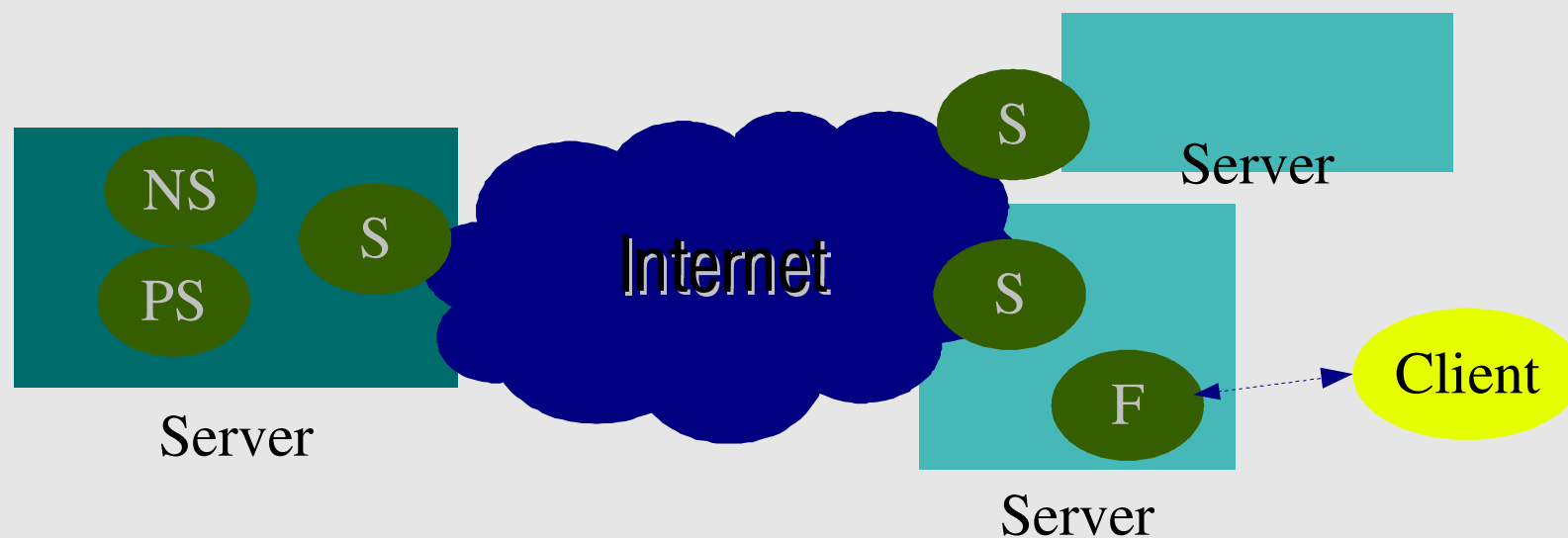
- Easily **applicable** (wide range of applications)

Aplications and Framework

Iperf

Measure Bandwidth between hosts. Point to Point.

Network Weather Service

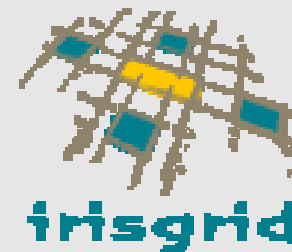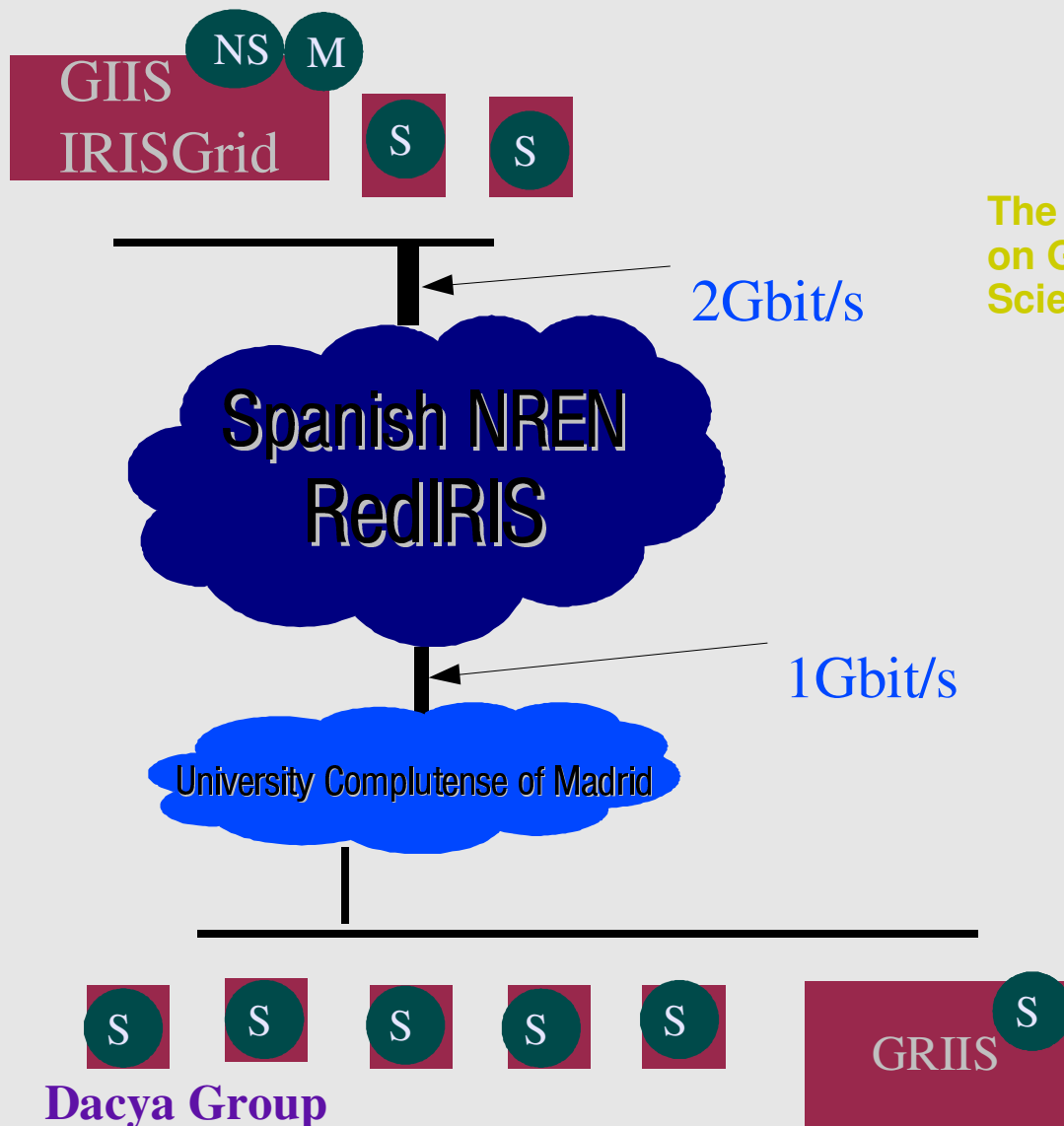1. Distributed Resource Performance Forecasting Service for Metacomputing.
2. Provide accurate forecasts of dynamically changing performance characteristics.

Environment of the experiment. IRISGrid

RedIRIS Local Network

GIIS IRISGrid NS M S S

2Gbit/s

Spanish NREN RedIRIS

The Spanish Thematic Network on Grids in the framework of e-Science Initiatives

1Gbit/s

University Complutense of Madrid

S S S S S GRIIS S

Dacya Group

NS NWS NameServer

M NWS Memory

S NWS Sensor

Host.Globus 2.4
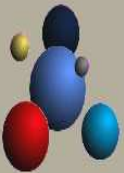
Resources Selection Algorithm

Rank calculated with an **estimated submission time**:

$$-\text{Rank} = T_{sub}(h_n, t_n) = T_{exe}(h_n, t_n) + T_{xfr}(h_n, t_n)$$

Dispatch Manager

Globus Monitoring and Discovery Service (MDS)

Static Information (S.O., architecture)

- User-provided Requirements
- Authorization test

Resource Discovery

Filtered LDAP search

GRIS

Dynamic Information (CPU load, …)

- Rank expression
- User provided executable
- Characterize discovered hosts. Get a ordered list of resources by rank

Rank Calculation

GRIIS

GRIIS

LDAP queries

GRIIS

Network Measurements

Experimental TestBed. RedIRIS-UCM

TestBed Description

Host	Model	Speed	OS	Memory	Domain
aquila	Pentim III	700Mhz	Linux 2.4	128M	dacya.ucm.es
cygnus	Pentium IV	2.5Ghz	Linux 2.4	512M	dacya.ucm.es
cepheus	Pentium III	600Mhz	Linux 2.4	256M	dacya.ucm.es
hydrus	Pentium IV	2.5Ghz	Linux 2.4	512M	dacya.ucm.es
aristoteles	Pentium III	1.4Ghz	Linux 2.4	1G	rediris.es
platon	Pentium III	1.4Ghz	Linux 2.4	1G	rediris.es
heraclito	Celeron	700Mhz	Linux 2.4	256M	rediris.es

Experiment.

CPU intensive artificial WorkLoad.
Initially, the job submision from cygnus

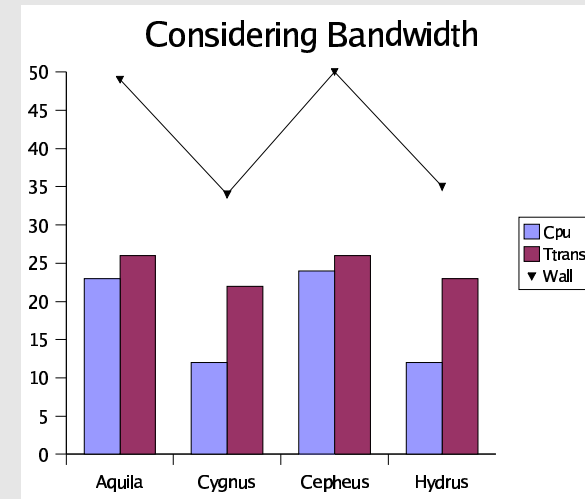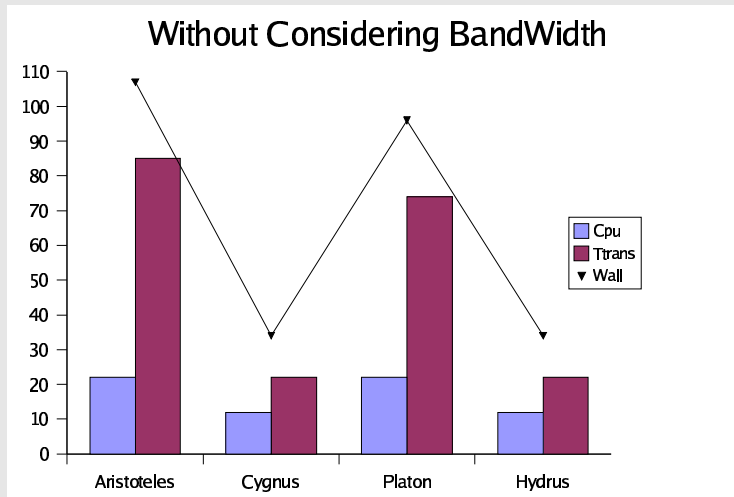Experiment with intensive artificial workload with diferents:

$$T_{xfr}(h_n,t_n) \ / \ T_{exe}(h_n,t_n)$$

$T_{xfr}(h_n,t_n)$ = Transfer Time $T_{exe}(h_n,t_n)$= Execution Time

Example. 4 Jobs Running. $T_{xfr}(h_n,t_n) / T_{exe}(h_n,t_n)=0.01$

Without Considering BandWidth

Considering Bandwidth

JID	AID	TID	DM	SM	GSM	STIME	ETIME	CPUTIME	XFRTIME	EXIT	TEMPLATE		HOST
0	0	0	zomb	done	--	13:55:13	13:55:47	00:12	00:22	0	job_template		cygnus.dacya.ucm.es
1	0	1	zomb	done	--	13:55:13	13:55:48	00:12	00:23	0	job_template		hydrus.dacya.ucm.es
2	0	2	zomb	done	--	13:55:13	13:56:02	00:23	00:26	0	job_template		aquila.dacya.ucm.es
3	0	3	zomb	done	--	13:55:13	13:56:03	00:24	00:26	0	job_template	**90seg**	cepheus.dacya.ucm.es

JID	AID	TID	DM	SM	GSM	STIME	ETIME	CPUTIME	XFRTIME	EXIT	TEMPLATE		HOST
0	0	0	zomb	done	--	13:56:49	13:57:23	00:12	00:22	0	job_template		cygnus.dacya.ucm.es
1	0	1	zomb	done	--	13:56:49	13:57:23	00:12	00:22	0	job_template		hydrus.dacya.ucm.es
2	0	2	zomb	done	--	13:56:49	13:58:36	00:22	01:25	0	job_template	**187seg**	aristoteles.rediris.es
3	0	3	zomb	done	--	13:56:49	13:58:25	00:22	01:14	0	job_template		platon.rediris.es

Conclusion and Future Work

Conclusion

Relevance of **resource proximity** in the resource selection process to reduce the cost of **file staging**.

Future Work

1. Extension experiment at all IRISGrid Resources.

2. Integration of Migrations Jobs.

3. Integration of Dynamic Reservation of bandwidth. Traffic Prioritiy.

Thank you for your attention