Managed by

Integration of Grid Cost Model into ISS/VIOLA Meta-Scheduler environment

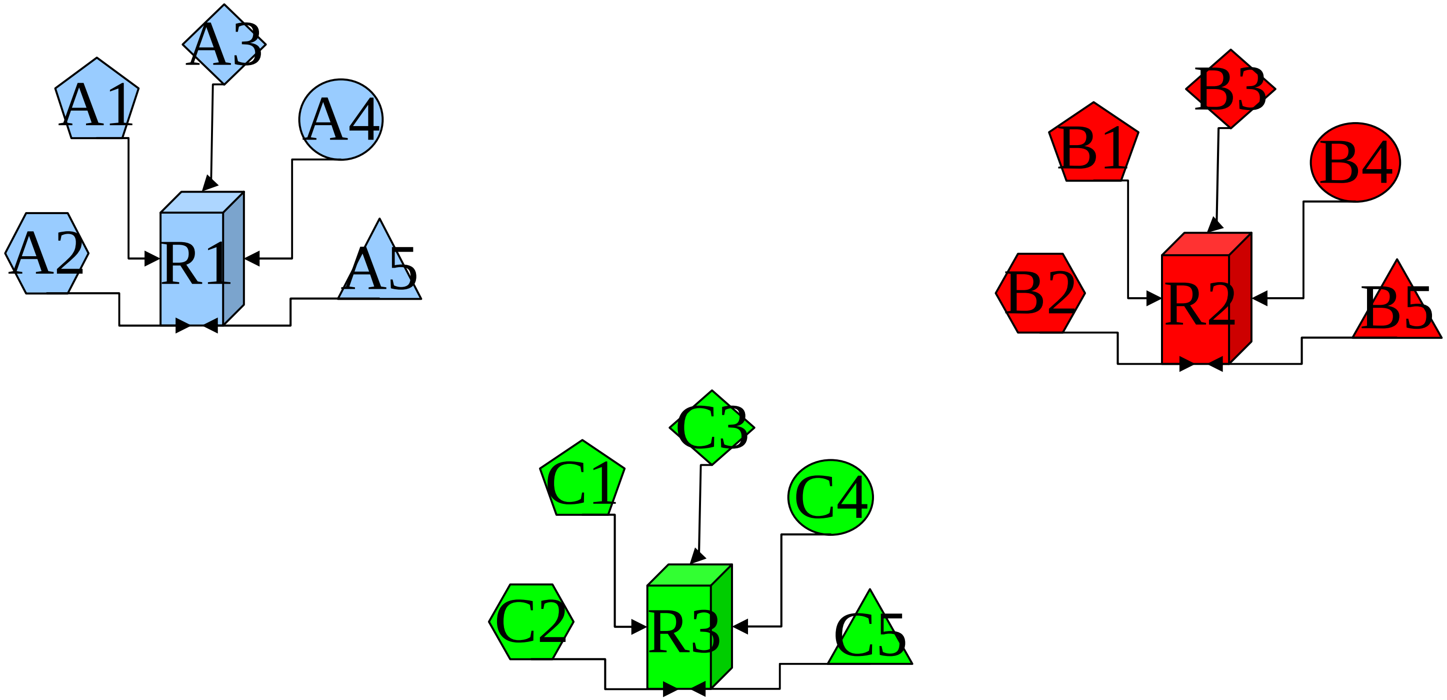
Ralf Gruber*, Vincent Keller*, Michela Thiémard*, EPFL
Oliver Wäldrich*, Wolfgang Ziegler*, SCAI
Philipp Wieder*, Forschungszentrum Jülich
Pierre Manneback*, CETIC

*CoreGRID members

Plan

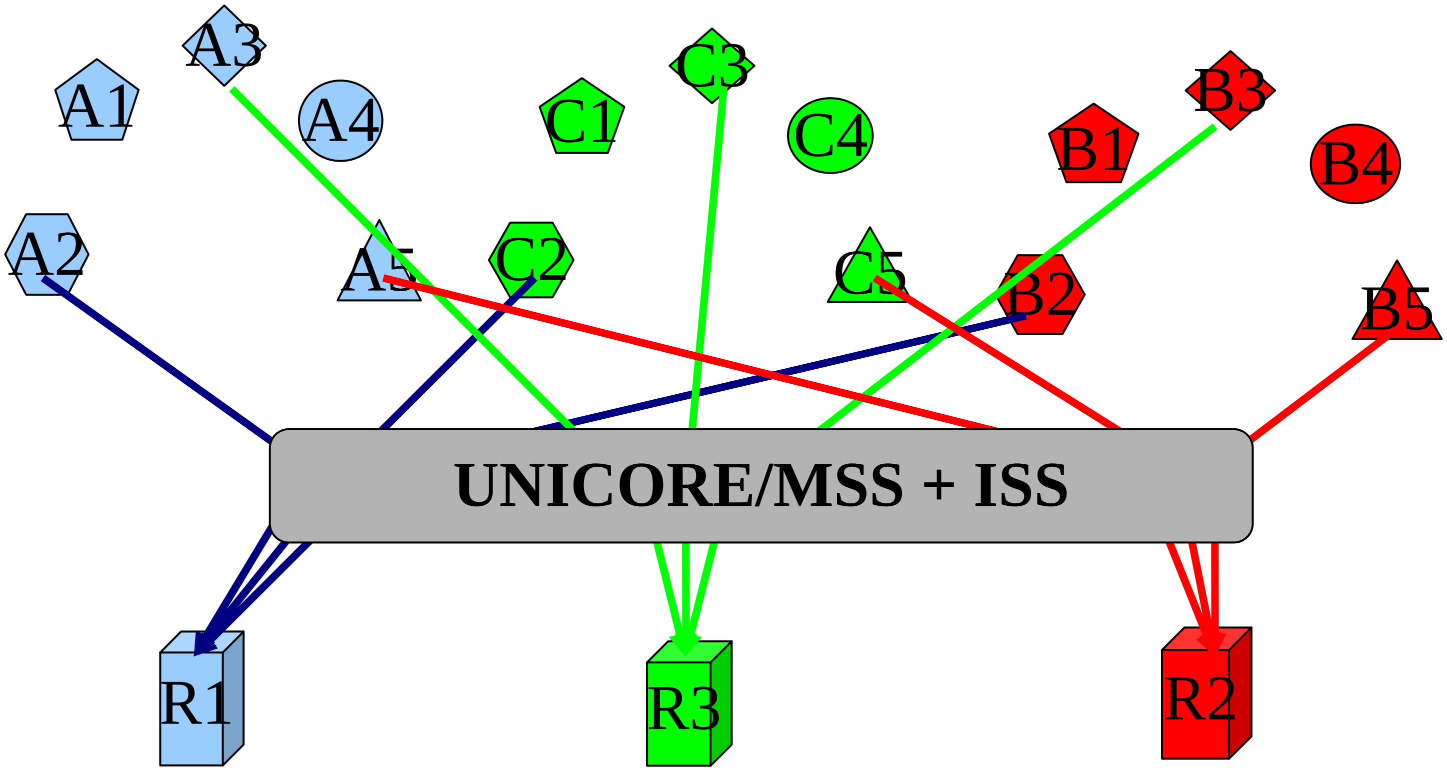
- HPC Situation
- Application structure
- Gamma Model
- ISS Concept
- Cost Model
- UNICORE-VIOLA-ISS Testbed

HPC : TODAY



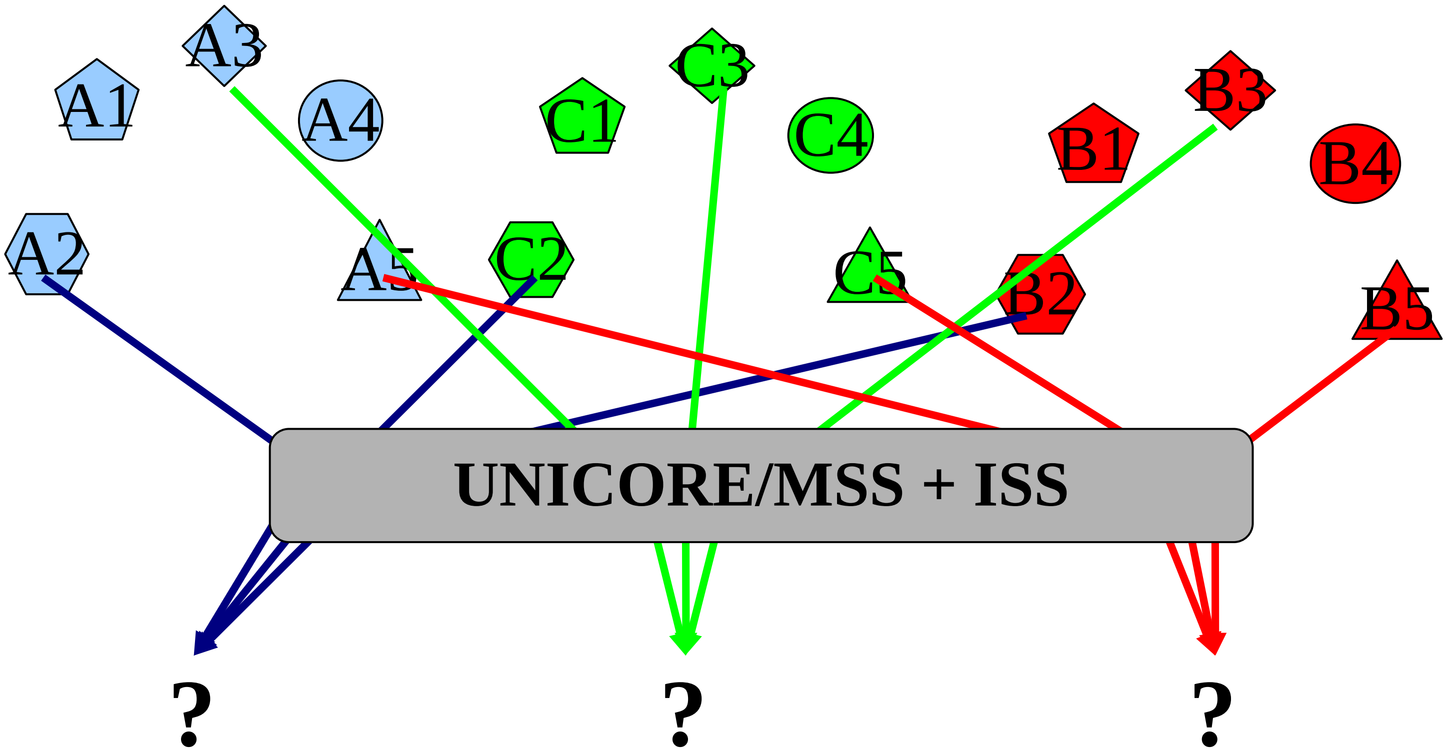
Resources chosen to satisfy needs of all local applications

HPC : TOMORROW



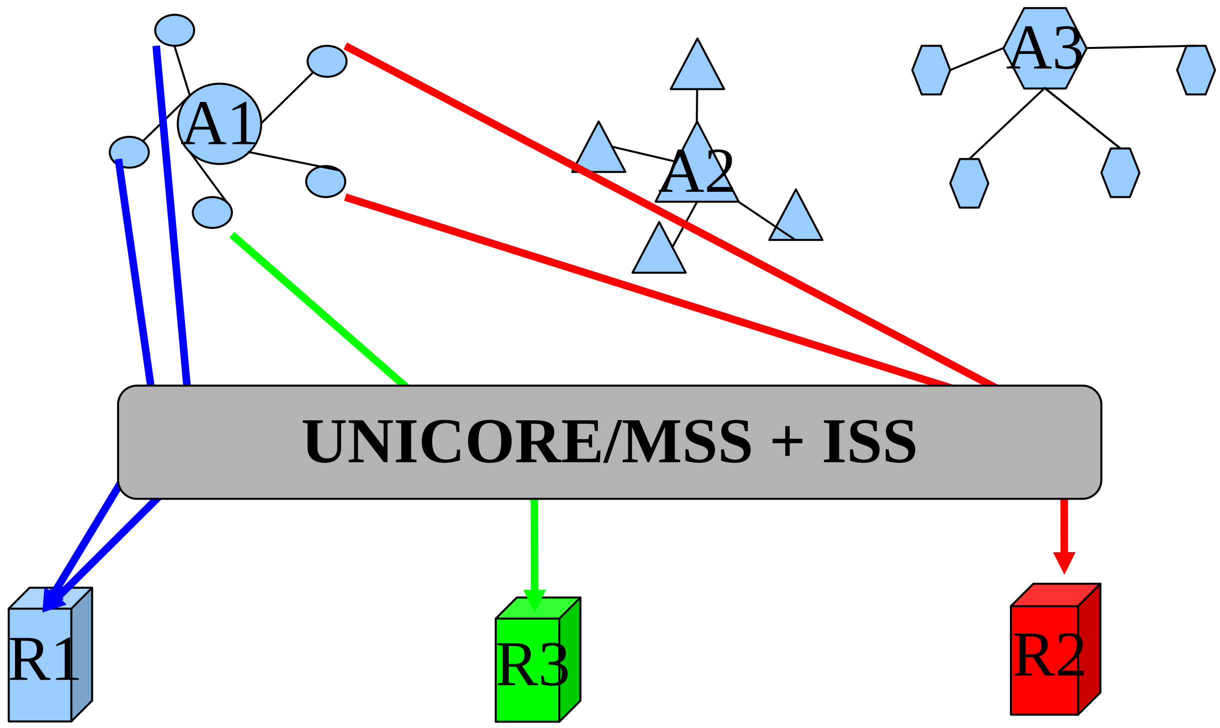
How to **BEST** distribute applications to improve overall performance ?

HPC : TOMORROW EVENING



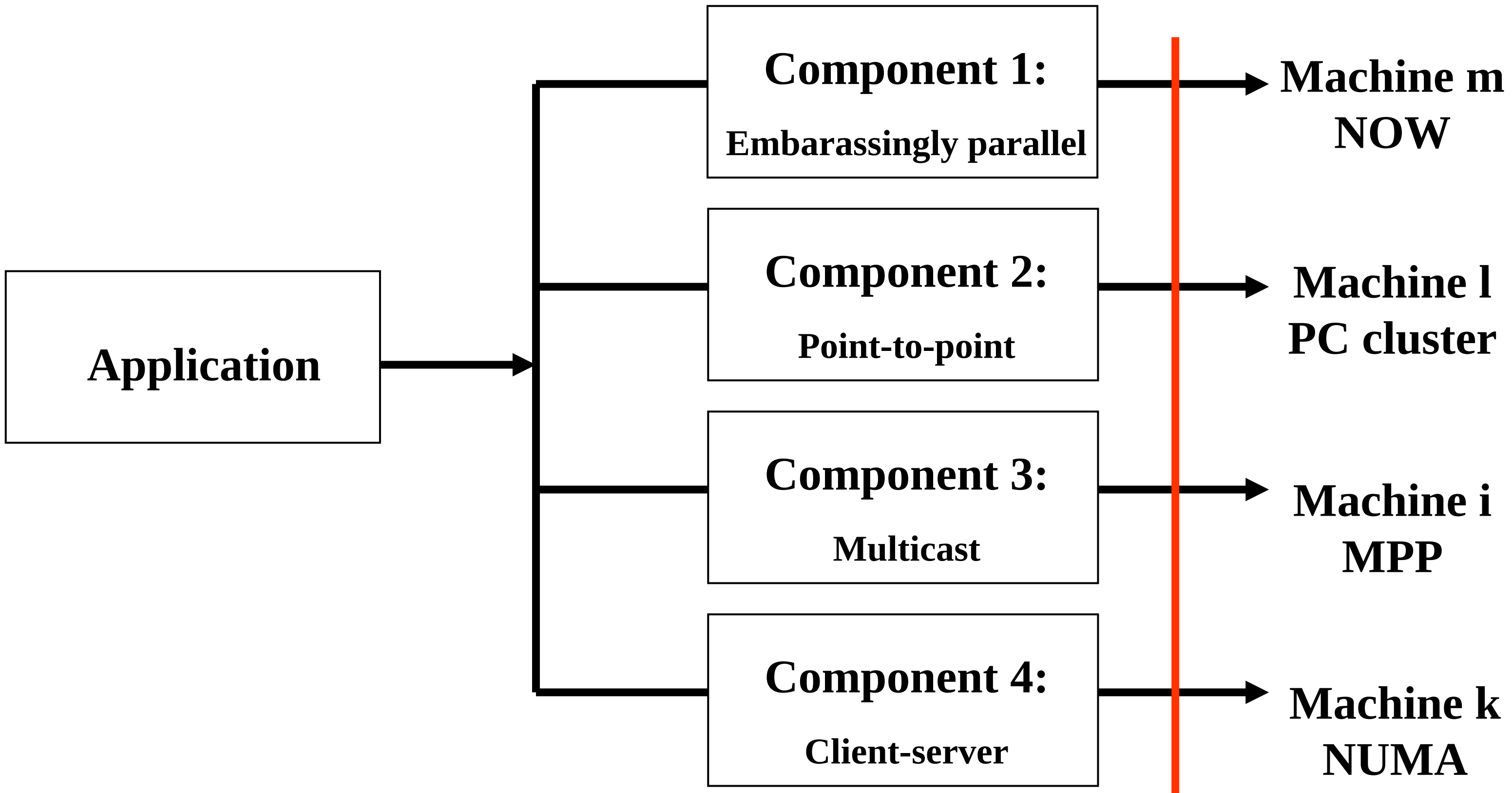
What resources needed to satisfy applications components
WITHIN LOWEST COSTS ?

HPC : AFTER TOMORROW



What to do to BEST distribute application COMPONENTS to improve overall performance ?

Application structure



New feature to VIOLA/Meta-Scheduler environment:
Submit to well-suited machines for application components

Γ model

Characterizes application components and parallel machines

$$\Gamma = \frac{E}{1-E}$$ = CPU time over communication time

Γ parameters of **application components** for one machine:
CPU time
Communication time
Size of messages

+

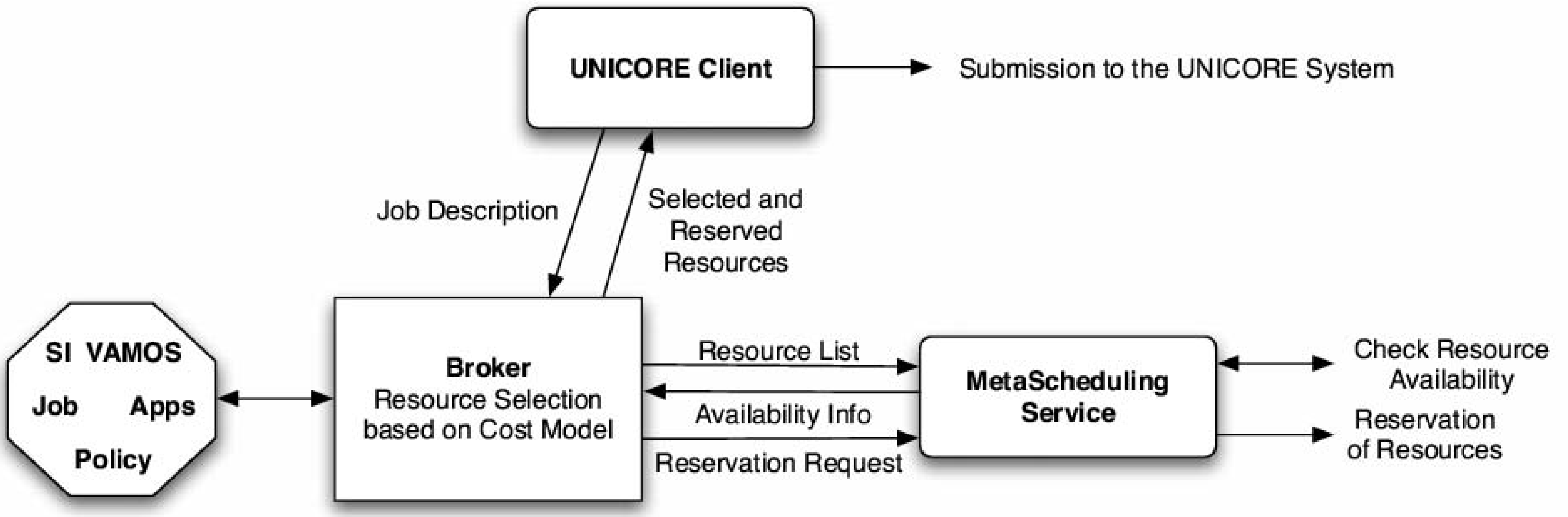
Parameters on one **machine**:
Memory bandwidth
Processor performance
Network performance

+

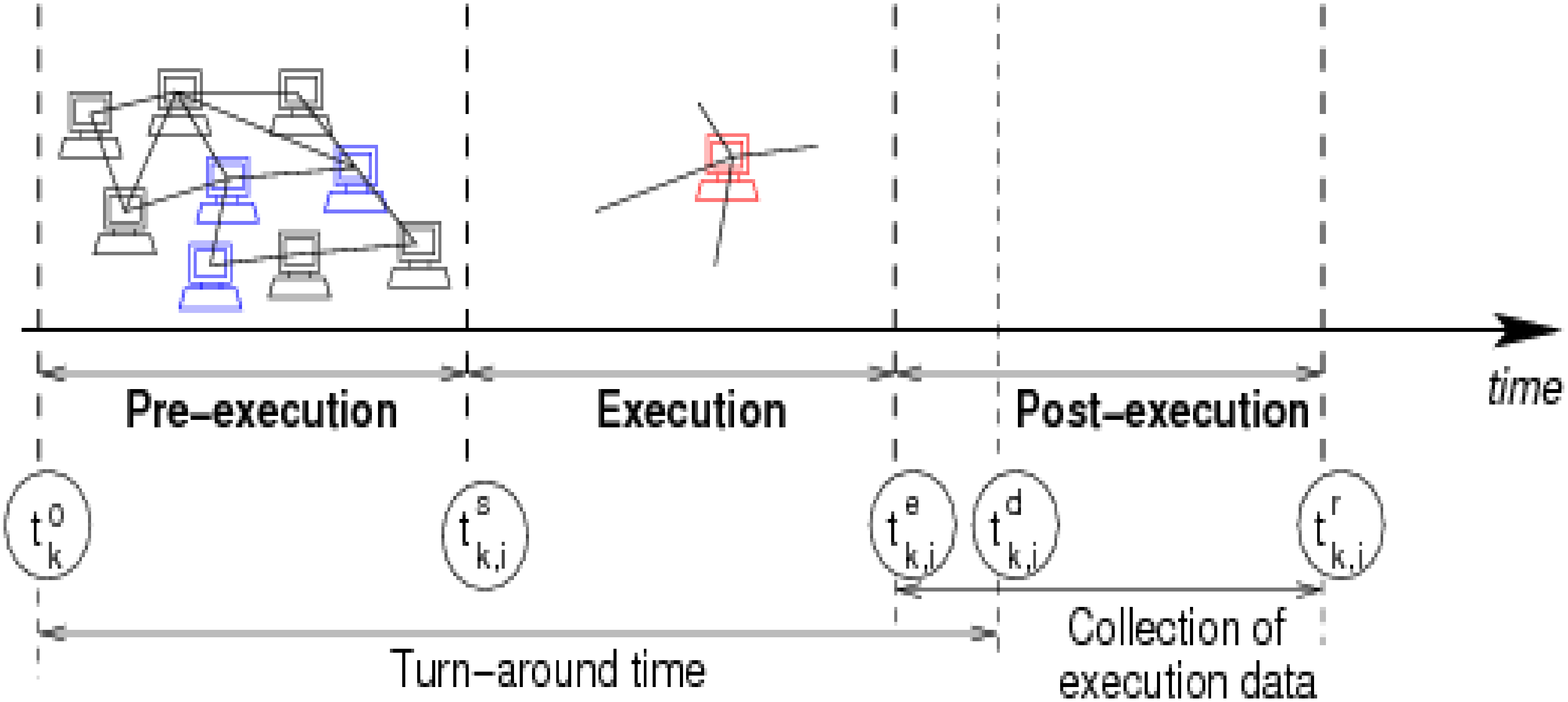
Parameters on other machine:
Memory bandwidth
Processor performance
Network performance

Γ parameters of application component on other machine

Framework


UNICORE Client
Submission to the UNICORE System
Job Description
Selected and Reserved Resources
SI VAMOS
Job
Apps
Policy
Broker
Resource Selection based on Cost Model
Resource List
Availability Info
Reservation Request
MetaScheduling Service
Check Resource Availability
Reservation of Resources

ISS concept



time
Pre–execution
Execution
Post–execution
t^o_k
$t^s_{k,i}$
$t^e_{k,i}$
$t^d_{k,i}$
$t^r_{k,i}$
Turn–around time
Collection of execution data

Pre-execution

Resource discovery

Prologue

Decision

Queing

Prologue

Find eligible machines: Yes on

Machine up?
Access rights?
Program exists?
Enough memory?

Decision

Collect data on their availabilities

Evaluate cost function

Submit the job

Cost function

time costs → $\beta K_w(C_k, R_i, p_k)$

HW+SW costs → $\sum_{k=1}^{n} \Im_{C_k}(R_i, p_k)$

$$\min(z) = \beta K_w(C_k, R_i, p_k) + \sum_{k=1}^{n} \Im_{C_k}(R_i, p_k)$$

such that $\forall 1 \le k \le n$

$$\sum_{k=1}^{n}(K_e(C_k, R_i, p_k) + K_l(C_k, R_i, p_k)$$

$$+ K_{eco}(C_k, R_i, p_k) + K_d(C_k, R_i, p_k)) \le KMAX$$

$$\max_{i,k}(t_{k,i}^{d}) - \min_{k}(t_k^{0}) \le TMAX$$

$$(R_i, p_k) \in \Re(C_k)$$

Cost function

$$\Im_{C_k}(R_i, p_k) = \alpha_k(K_e(C_k, R_i, p_k) + K_l(C_k, R_i, p_k))$$
$$+ \gamma_k(K_{eco}(C_k, R_i, p_k))$$
$$+ \delta_k(K_d(C_k, R_i, p_k))$$
$$\alpha_k, \beta, \gamma_k, \delta_k \geq 0$$
$$\alpha_k + \beta + \gamma_k + \delta_k > 0$$

Free parameters α_k, β, γ_k, δ_k

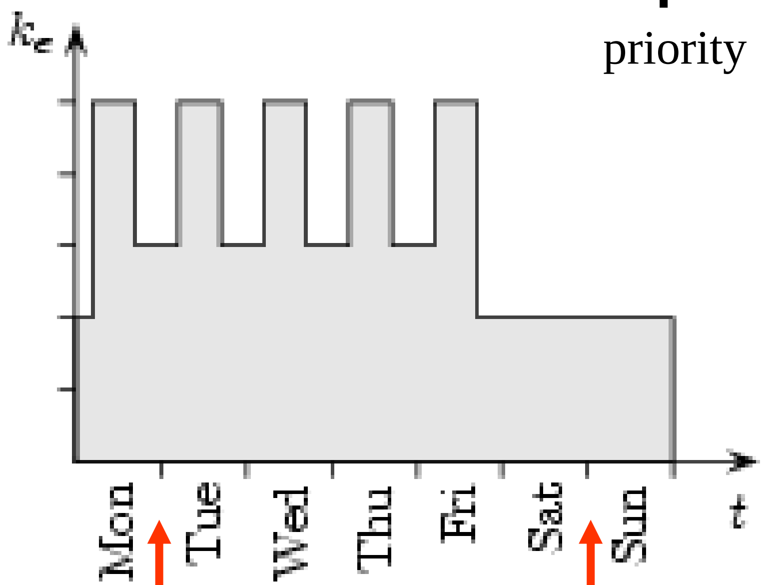
Minimize turn-around time: $\beta= 1$, KMAX=∞, α_k γ_k, $\delta_k = 0\ \forall k$

Minimize hardware costs: $\beta= 0$, TMAX=∞, α_k γ_k, $\delta_k \geq 0$

CPU costs K_e

$$K_e(C_k, R_i, p_k) = \int_{t^s_{i,k}}^{t^e_{i,k}} k_e(C_k, R_i, p_k, \varphi, t)\,dt$$

priority



$t^s_{k,i}$ $t^e_{k,i}$

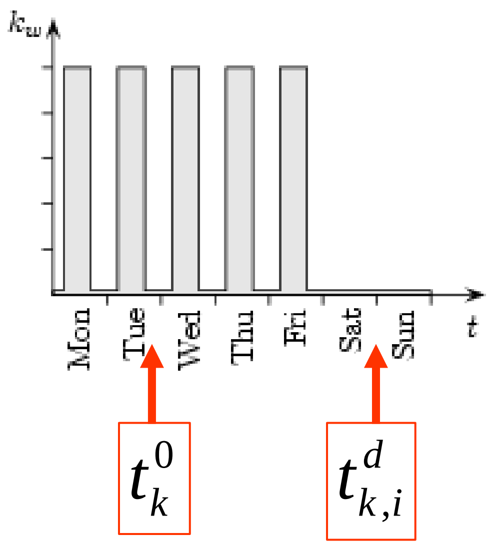
License fees K_l

$$K_l(C_k, R_i, p_k) = \int_{t_{i,k}^s}^{t_{i,k}^e} k_l(C_k, R_i, p_k, t)\,dt$$

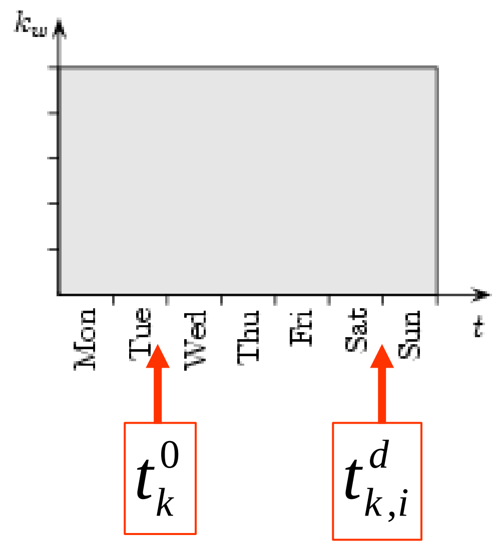
Costs of turn-around time K_w

$$K_w(C_k, R_i, p_k) = \int_{\min_k t_k^0}^{\max_k t_{k,i}^d} k_w(t)\,dt$$

salary



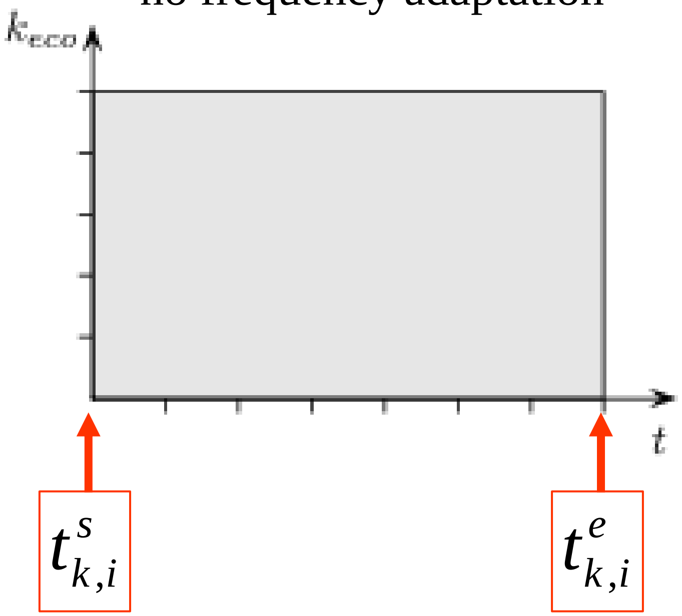
time-to-market



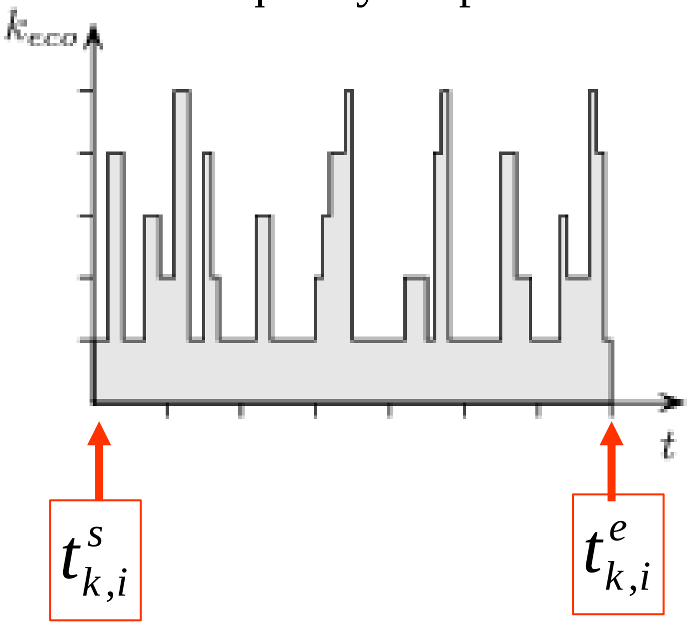
Energy costs

$$K_{eco}(C_k, R_i, p_k) = \int_{t_{k,i}^{s}}^{t_{k,i}^{e}} k_{eco}(C_k, R_i, p_k, t)\,dt$$

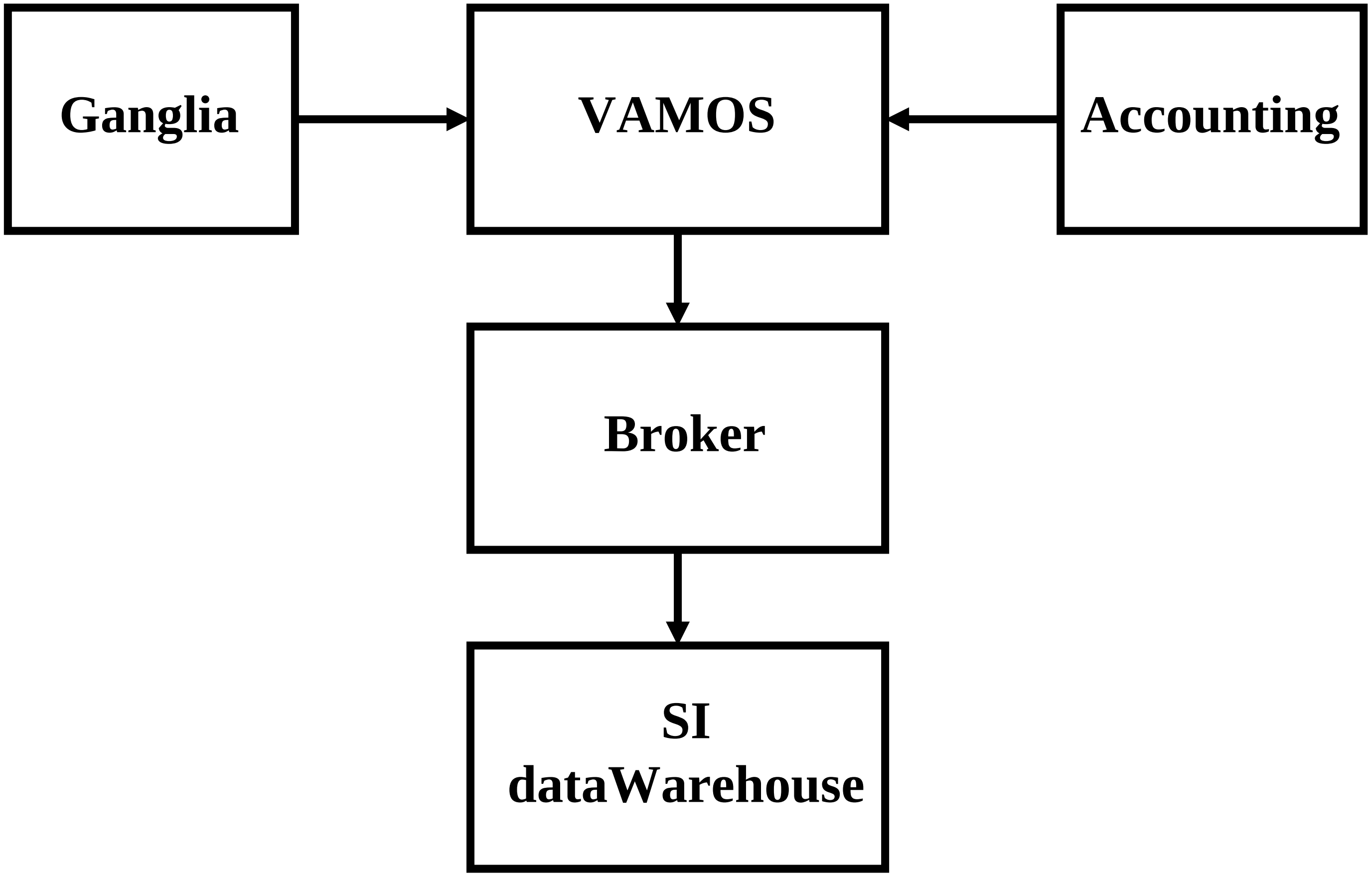
no frequency adaptation



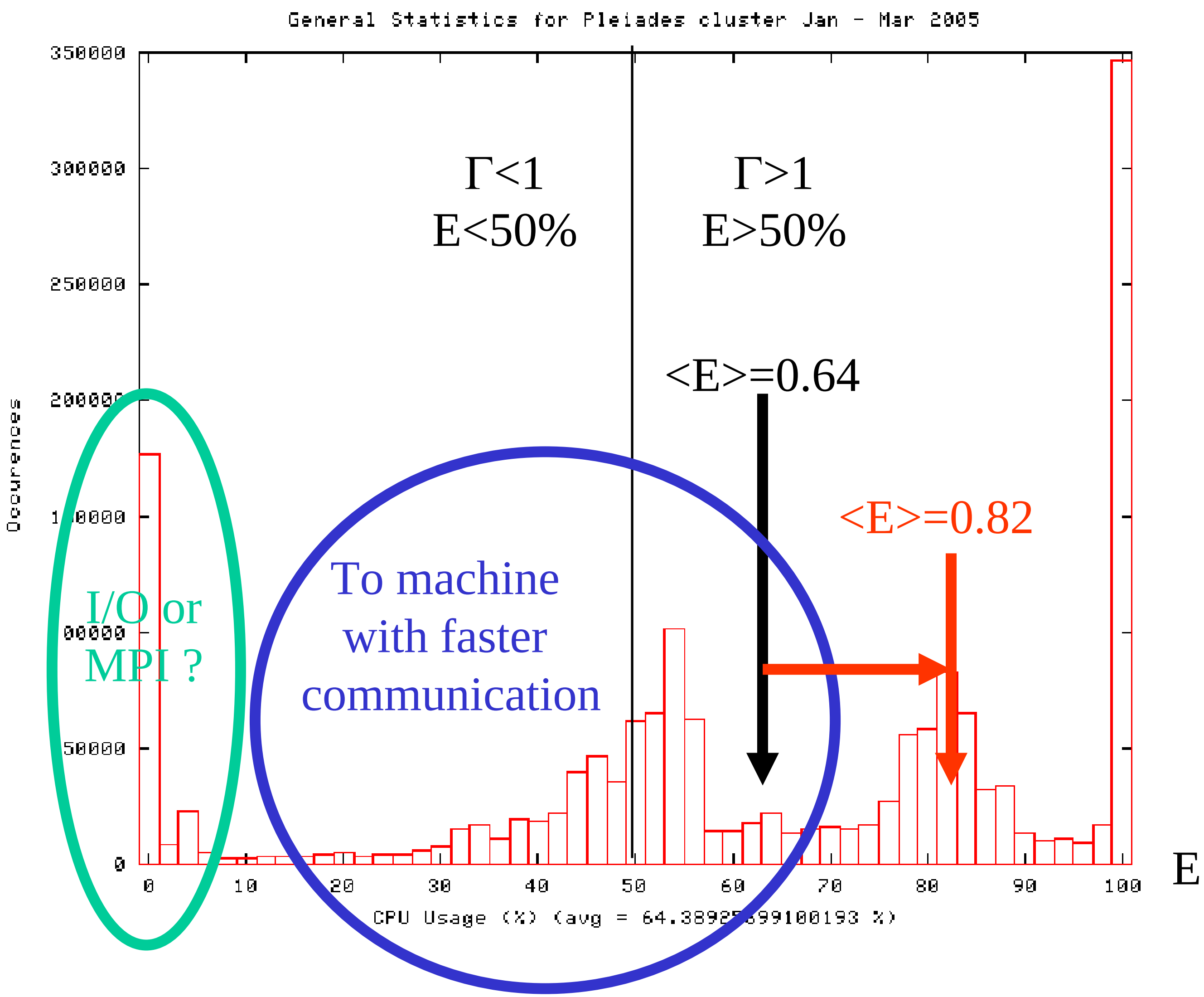
with frequency adaptation



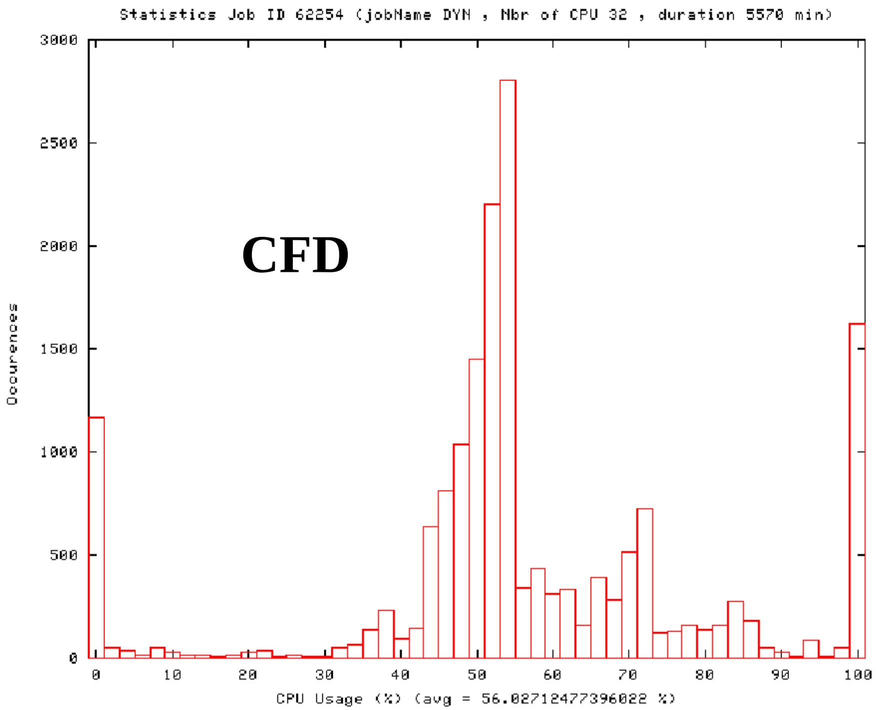
Epilogue

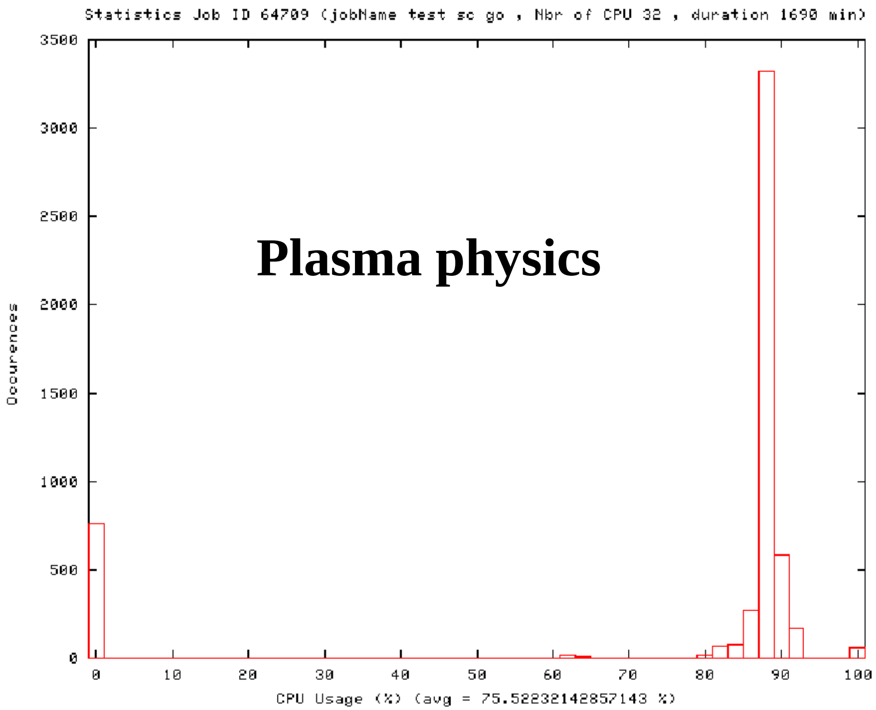


Pleiades: Ganglia data



VAMOS: single job profiles





First VIOLA/UNICORE/Meta-Scheduler/ISS testbeds

UNICORE
VIOLA
MSS
ISS

LIN-EPFL:
Pleiades1
Pleiades2
Pleiades3

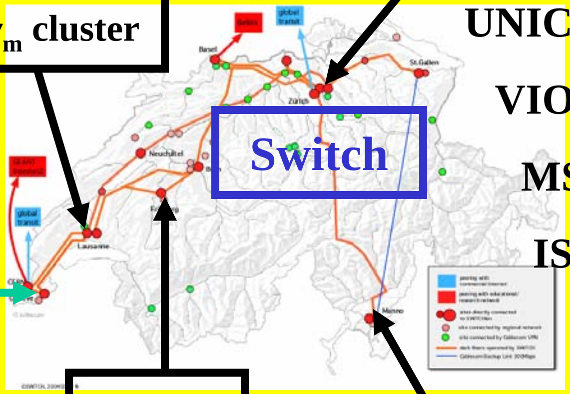
DIT-EPFL:
Mizar
Condor NOW
BlueBrain

12.2006: EPFL HPCN installations

EPFL:
SMP/NUMA
High γ_m cluster

ETHZ:
SMP/NUMA
High γ_m cluster

UNICORE
VIOLA
MSS
ISS



Switch

CERN:
egee Grid

EIF:
NoW

CSCS:
SMP/vector
Low γ_m cluster

12.2007: Swiss HPCN Grid initiative

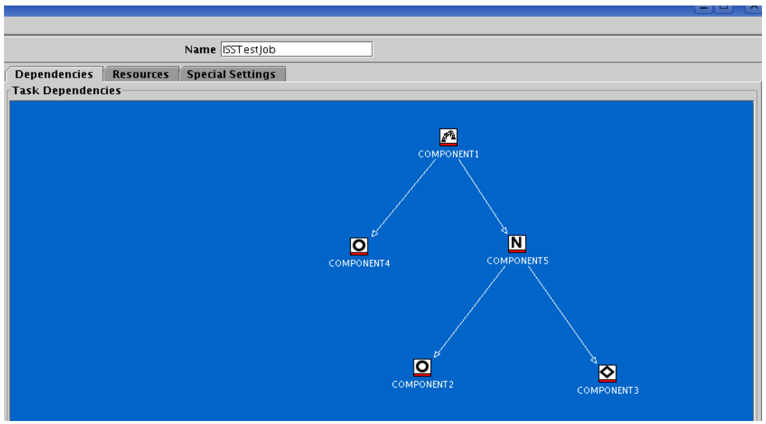
Outlook: Simulator

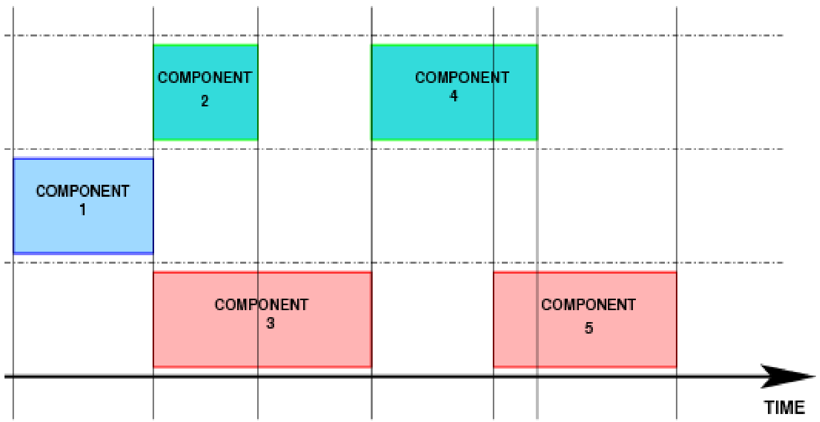
Understanding the behavior of the simulated grid depending of the value of the free parameters.

- Usage of old executions data on 2 departemental clusters
  - each job was monitored. Data stored in a mysqlDB
  - Mapping of ganglia and localscheduler (torque) info (VAMOS service)
- simulation of the real situation with UNICORE middeware and the VIOLA MSS simulated on top of it
- Using the real job traces.
- training of the system (broker service) .
- Metric used to show the improvement of the grid is the utilization of the machines.

Outlook: Component dependency

- Each component of a workflow is executed on the well suited machine





- **Hard problem :** need new ideas
- Co-allocation is now made manually.
- With ISS in the future : automatically.

Publications

Pierre Manneback, Guy Bergère, Nahid Emad, Ralf Gruber, Vincent Keller, Pierre Kuonen, Sébastien Noël, and Serge Petiton (2005), "Towards a scheduling policy for hybrid methods on computational Grids", CoreGRID Meeting, Pisa (28-30 November, 2005), to appear in Lecture Notes in Computer Sciences (Springer)

Ralf Gruber, Vincent Keller, Pierre Kuonen, Marie-Christine Sawley, Basile Schaeli, Ali Tolou, Marc Torruella, and Trach-Minh Tran (2005), "Intelligent GRID Scheduling System", PPAM 2005, Poznan, Poland, Lecture Notes in Computer Sciences (Springer) 3911, p. 751-757

Vincent Keller, Kevin Christiano, Ralf Gruber, Pierre Kuonen, Sergio Maffioletti, Nello Nellari, Marie-Christine Sawley, Trach-Minh Tran, Philipp Wieder, and Wolfgang Ziegler, "Integration of ISS into the VIOLA Meta-Scheduling Environment", CoreGRID Meeting, Pisa (28-30 November, 2005), to appear in Lecture Notes in Computer Sciences (Springer)

Ralf Gruber, Pieter Volgers, Alessandro De Vita, Massimiliano Stengel, and Trach-Minh Tran, "Parameterisation to tailor commodity clusters to applications", Future Generation Computer Systems 19 (2003) 111-120

Ralf Gruber, Vincent Keller, Michela Thiémard, Oliver Wäldrich, Philipp Wieder, Wolfgang Ziegler, and Pierre Manneback, "Integration of Grid Cost Model into ISS/VIOLA Meta-Scheduler environment", (2006) to appear.

Information Society Technologies

CoreGRID

THANK YOU