



Inserm

Institut national
de la santé et de la recherche médicale

My after-Dutch life

Tristan Glatard
in daily collaboration with Sorina Camarasu-Pop

August 26th 2009
Science Park, Amsterdam

Creatis
LRMN

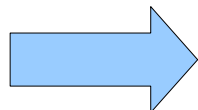
from vlemmed to biomed

VL-e med

- ~ 11 CEs/SEs
- Weekly “ct-grid” meetings
- Job success ~ 80 %
- Queue time up to 15 min
- Monitored VO activity

Biomed

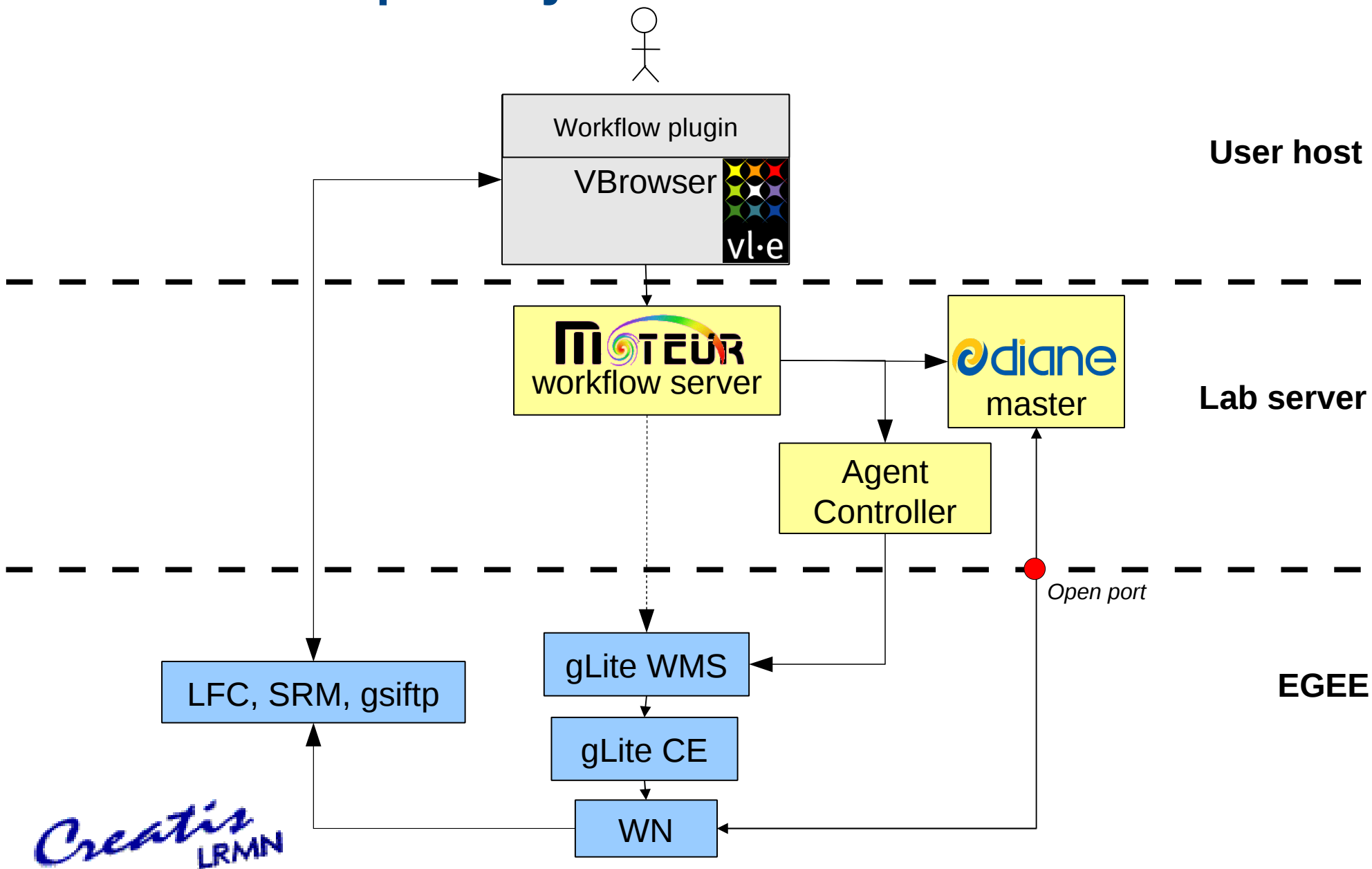
- ~ 120 CEs/SEs
- No formal contact with sites
- Job success ~ 40 %
- Queue time up to hours
- Unknown VO activity



Outline

- **Developments**
 - Pilot jobs
 - QoS with site monitoring and pre-selection
 - Interoperability (D-Grid and Nordugrid)
 - Monitoring and activity report
- Applications
- Modeling (prospective)

DIANE pilot jobs in VL-e med soft.



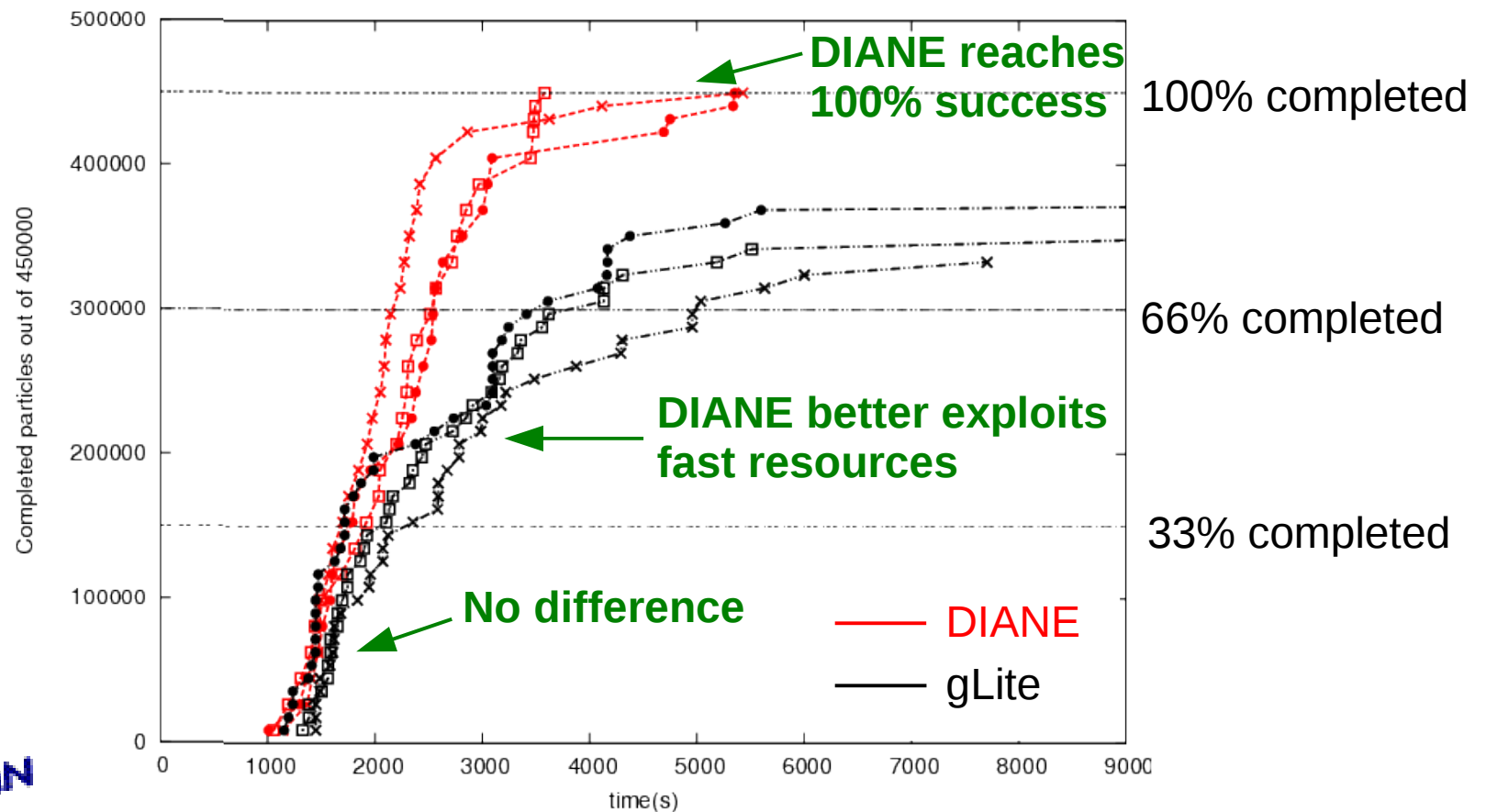
MOTEUR/DIANE interface

- Starts DIANE masters
 - One per user, on open port (one port per user)
- Task submission/monitoring CLI
 - `diane-job-{submit,status,cancel}`
- “Live” stdout/err transfers
- Control of agent submission
- Monitoring interface

Actions	User	MASTERS				AGENTS				TASKS					STATS			
		Date launched	Workspace	Port	Alive	Submitted	Running	Registered	Removed	Submitted	Scheduled	Running	Done	Failed	Worktime	Running	Download	Upload
Kill!	Tristan Glatard/	Jun 12 10:41	/var/www/diane/runs/0293	23004	true	1	0	0	0	0	59	0	0	0	0h:0m:0s	%	%	%
Kill!	Carlos Gines Fuster/	Jun 11 23:16	/var/www/diane/runs/0292	23005	true	782	7	373	366	0	0	1	1950	181	127h:31m:4s	16.8%	78.6%	4.5%
Kill!	Sorina Camarasu/	Jun 10 16:44	/var/www/diane/runs/0275	23001	true	38	0	21	21	0	0	0	5	9	3h:0m:36s	95.6%	3.3%	.9%

DIANE vs gLite-only (GATE appli)

- $\#\{\text{gLite jobs}\} = \#\{\text{DIANE agents}\} = \#\{\text{tasks}\}$
 - No job resubmission
- Ex. for 75 agents/jobs/tasks:



Agent controller

- Algorithm currently in use

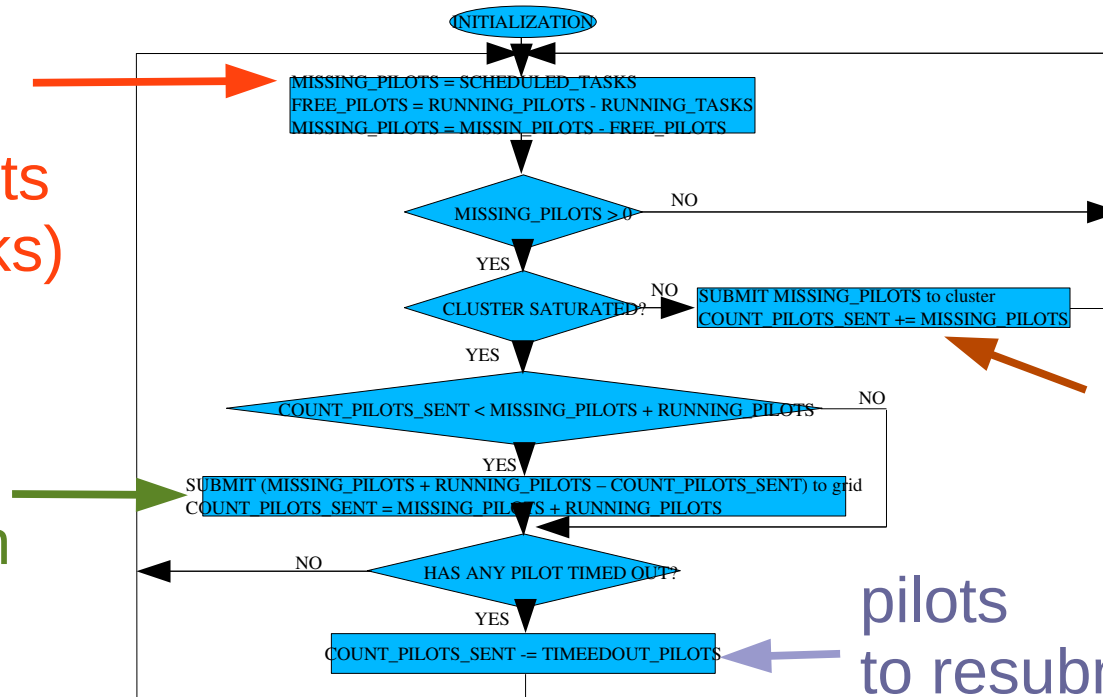
```
submit init pilots
while master is alive do
  sleep s seconds
  n = number of scheduled tasks in master
  submit sub=min(maxSub,n) pilots
  s = defaultSleep+sub*factor
end while
```

- On-going extension

[Internship Alejandro Tovar de Duenas
March – June 09]

computes
number of
missing pilots
(sched. tasks)

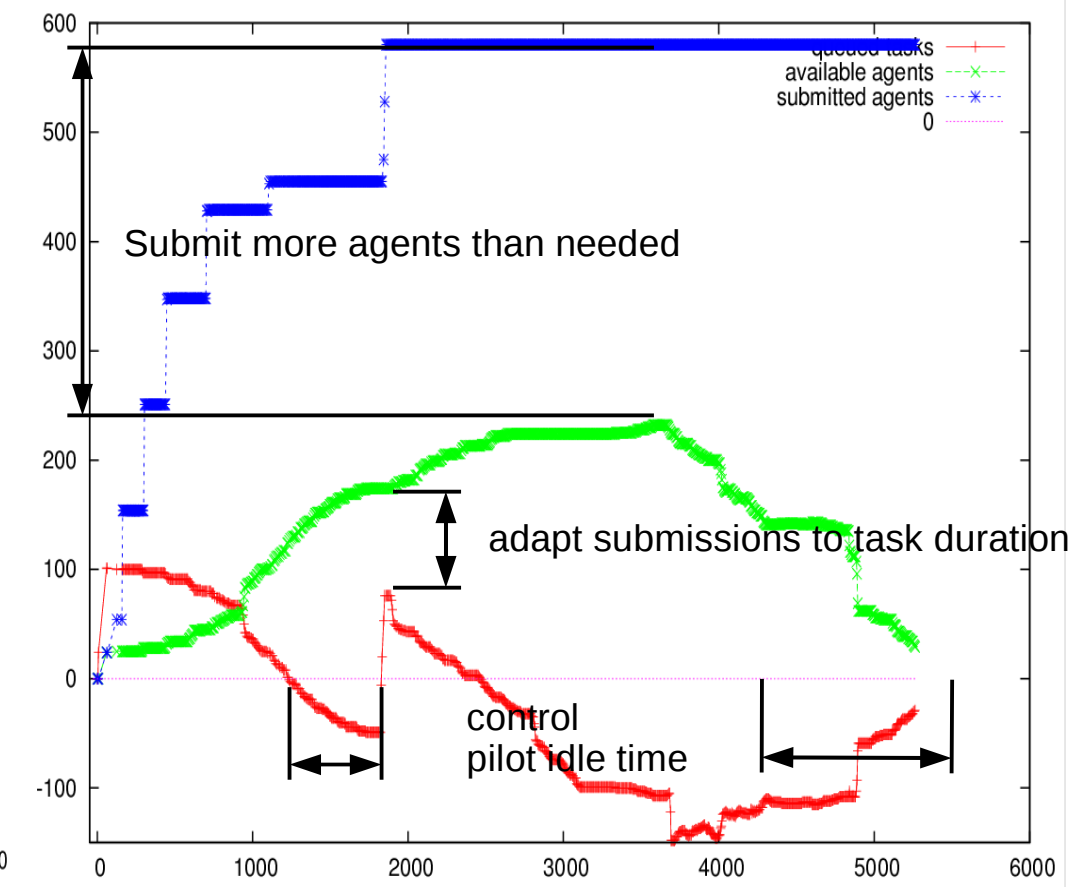
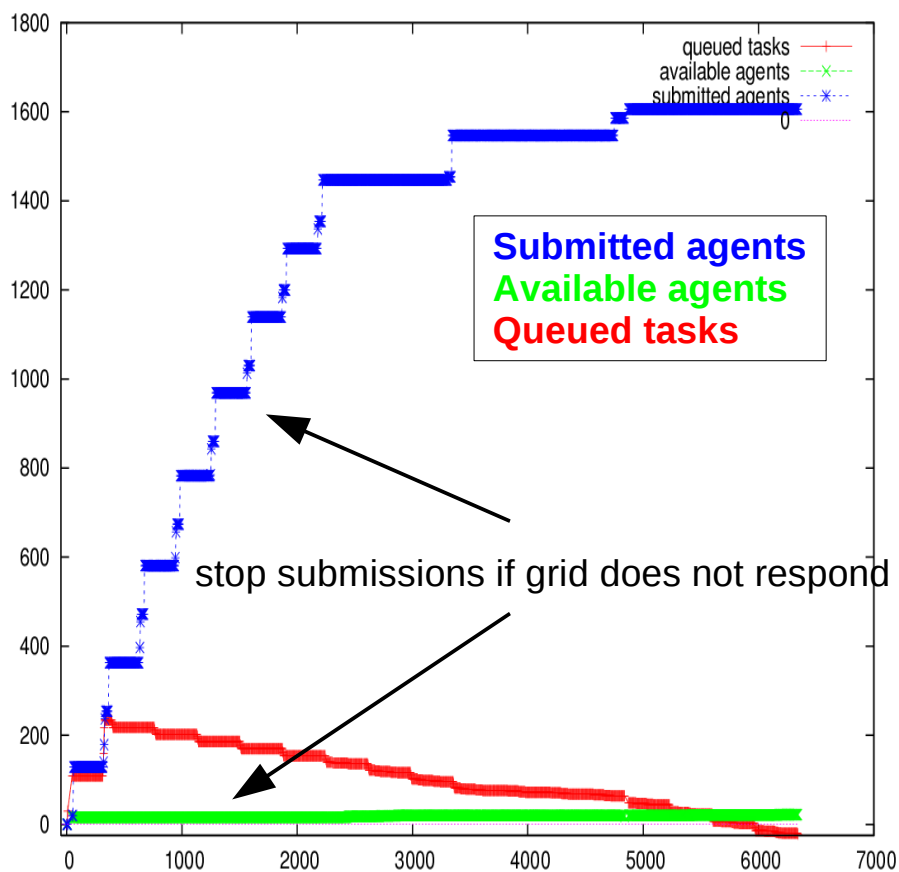
submit pilots
to the grid
(if not enough
in queue)



submit pilots
to cluster
(if not saturated)

pilots
to resubmit (timed-out)

Agent controller: extensions

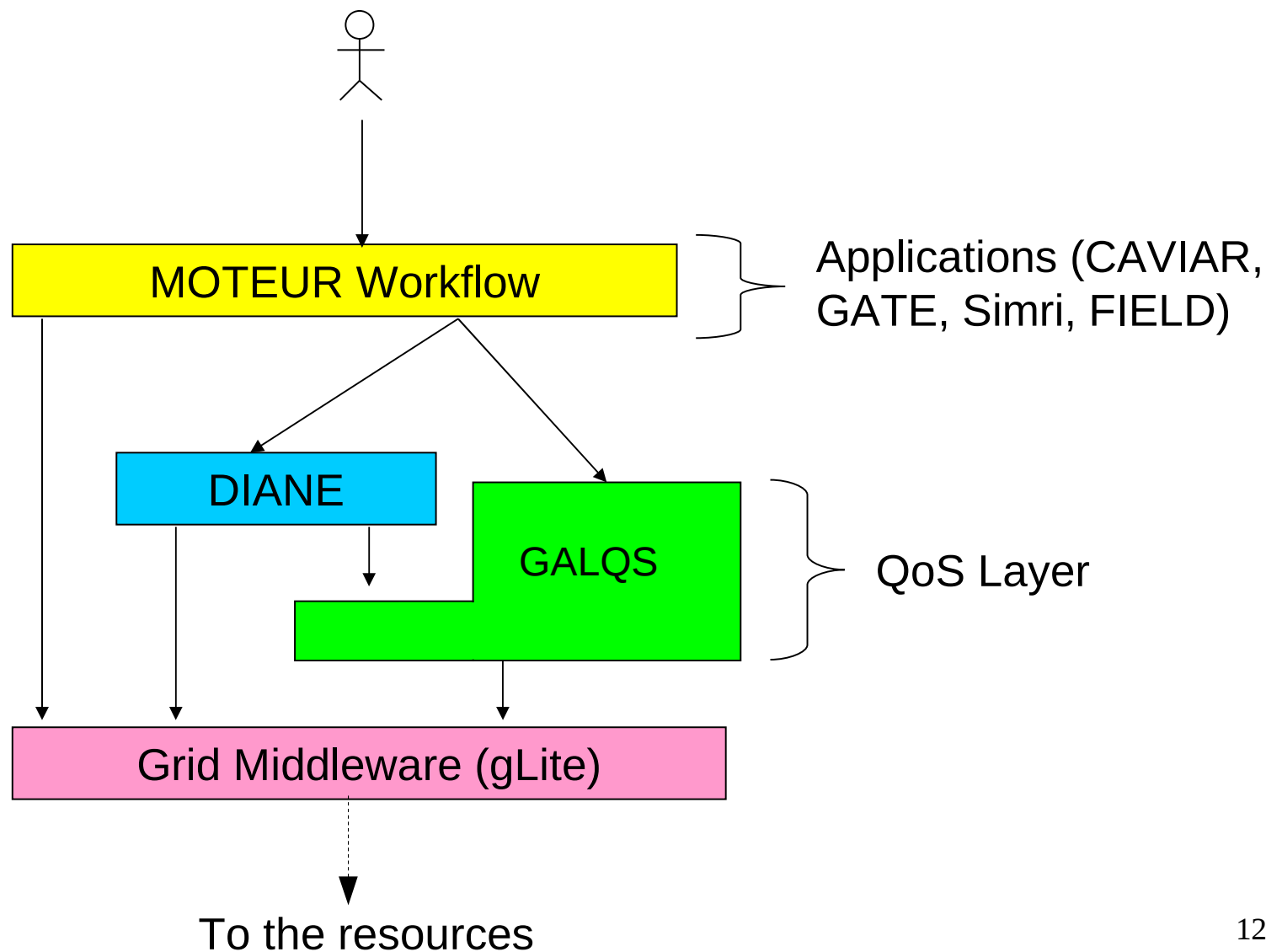


Current limits of pilot jobs

- Task requirements
 - Tasks are not gLite jobs (i.e. no GLUE requirements)
 - Master do not know about pilot (GLUE) requirements
- Introduces (yet) another layer in execution
 - Control of agent submission
 - Handle ports openings
- Interoperability with other middleware
 - e.g. with ARC (data transfer by CE VS late task-to-resource binding)
- Security (?)
 - Pilot GSI authentication to master (ok with DIANE)
 - Risk of compromised master (~compromised UI ?)

QoS with site monitoring and pre-selection

Grid Application-Level QoS Service



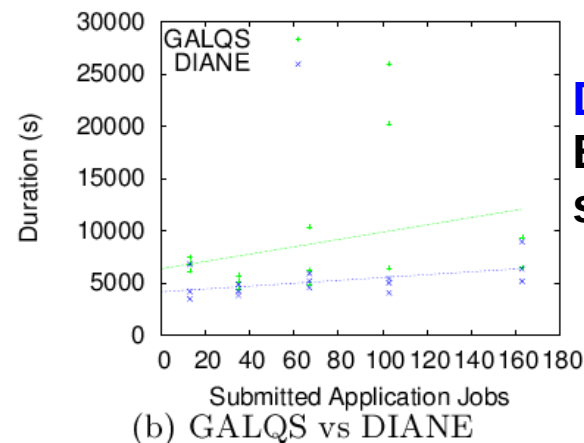
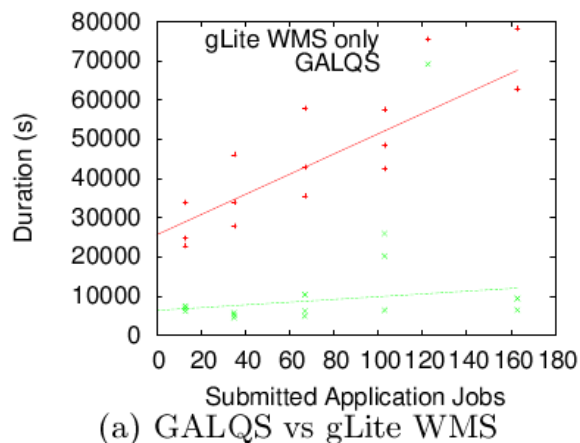
GALQS

- **Monitoring**
 - One probe per application and per CE
 - Probes submitted once per day (at 1am)
 - Probes check application execution time/correctness
 - Maintain database of error ratios and latencies per CE
- **Site pre-selection**
 - Select only CEs with 100% reliability
 - Select x% fastest CEs
 - Performed before submission to WMS -> JDL requirements adaptation

GALQS evaluation

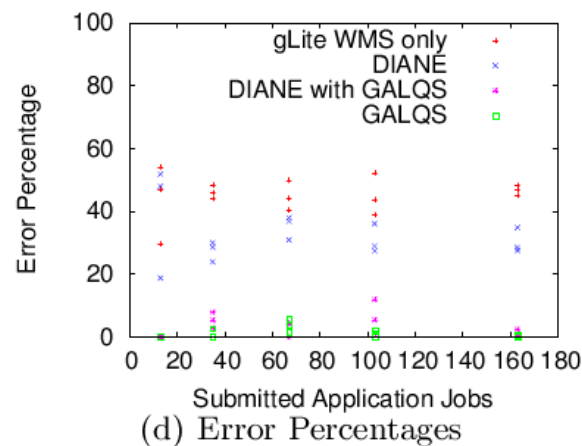
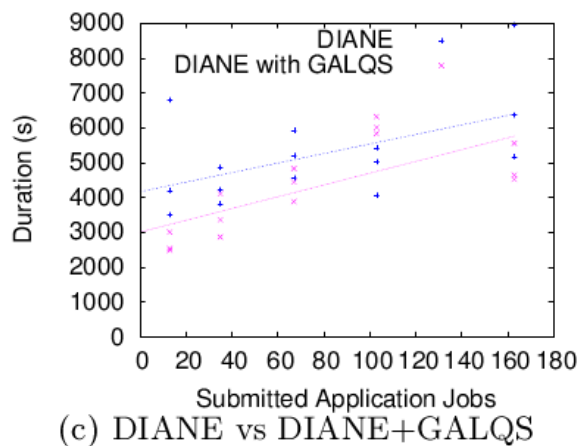
- Application: CAVIAR (~12-min jobs)
- Evaluation metric: time to reach 100% success

WMS vs GALQS
Throughput x 5



DIANE vs GALQS
Equivalent in some cases

DIANE vs DIANE+GALQS



Without GALQS:
~40% error

With GALQS:
~5%-10% error

QoS: next steps

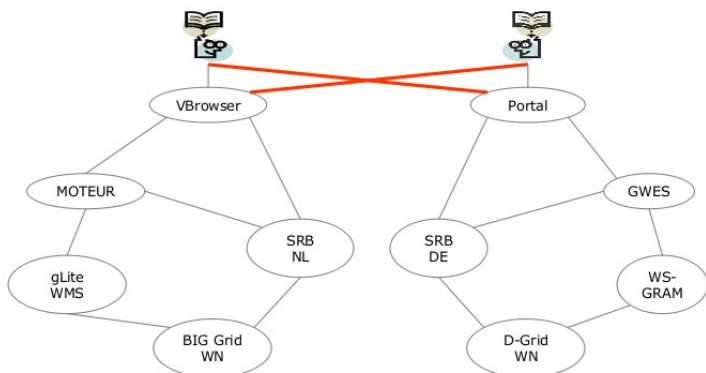
- Application feedback
- Automatic site selection
- Data probes
- => Merge with Java Job Submission
 - Add WMS support
 - Enable application-specific DBs
 - Interface with MOTEUR and Ganga

<http://cc.in2p3.fr/docenligne/269>

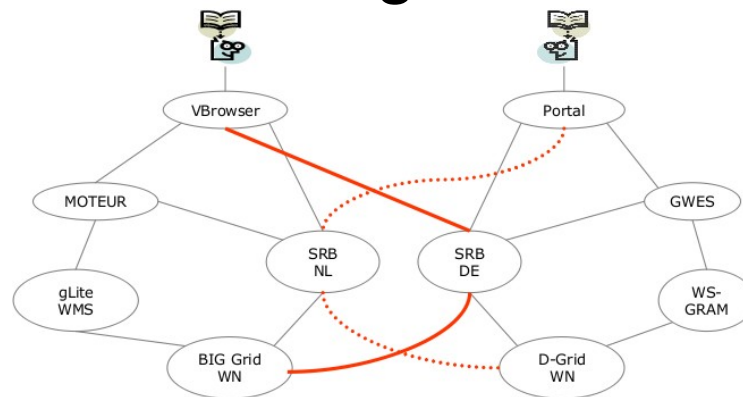
Interoperability

Interoperability with D-Grid

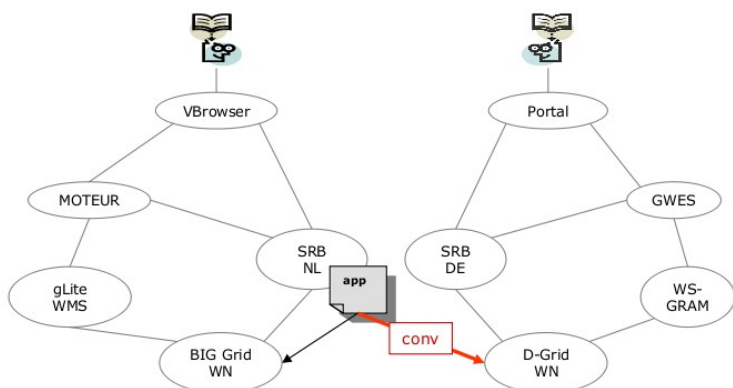
- Crossing users



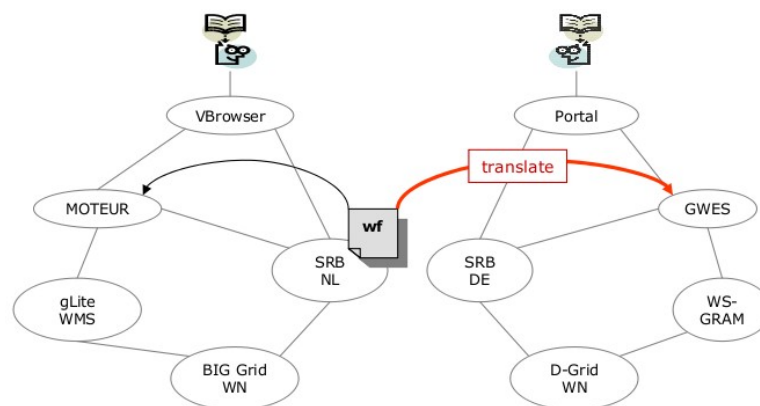
- Crossing files



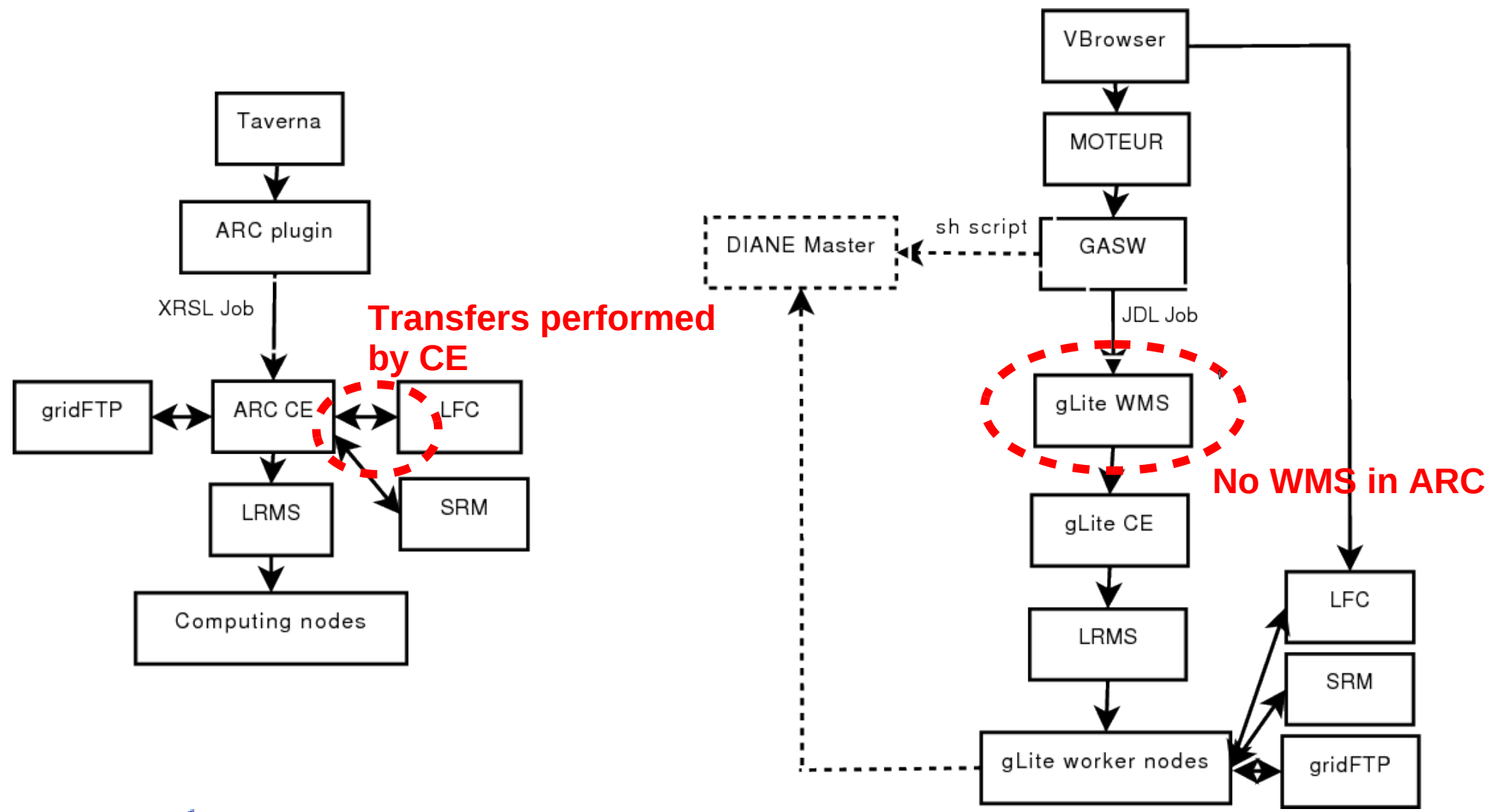
- Crossing software



- Crossing workflows



gLite VS ARC



Interface with ARC comp. resources

- Jobs

- Pilot jobs hardly usable on ARC resources
- MOTEUR submits XRSL jobs instead of JDL
- ARC client installed in 2 minutes on PoC UI
- Problems with proxies having 2 VOMS extensions

- Data

- Use ARC's support for LFC
- Only worked for non-DPM EGEE SEs
- In ARC, SRM output directory path has to be specified (only SE name with gLite)

Data transfers

- From EGEE LFC to user's Desktop (via srm/gsiftp)

- ARC client (native)
- UVA's VBrower (Java)
- GSAF ? / JavaGAT ? (not tested)

- KB/s, without VPN

	data on EGEE		data on ARC (RLS)	
	download	upload	download	upload
VBrower	4523	1022	X	X
ARC-client	4451	997	4301	911

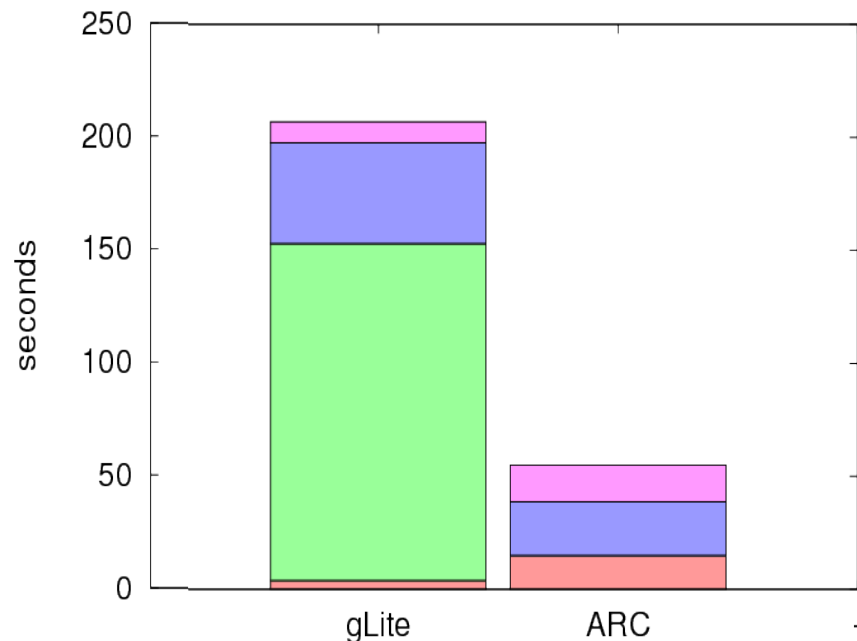
- KB/s, with VPN (inside Univ. Hospitals of Geneva)

	data on EGEE		data on ARC (RLS)	
	download	upload	download	upload
VBrower	339	116	X	X
ARC-client	361	110	345	112

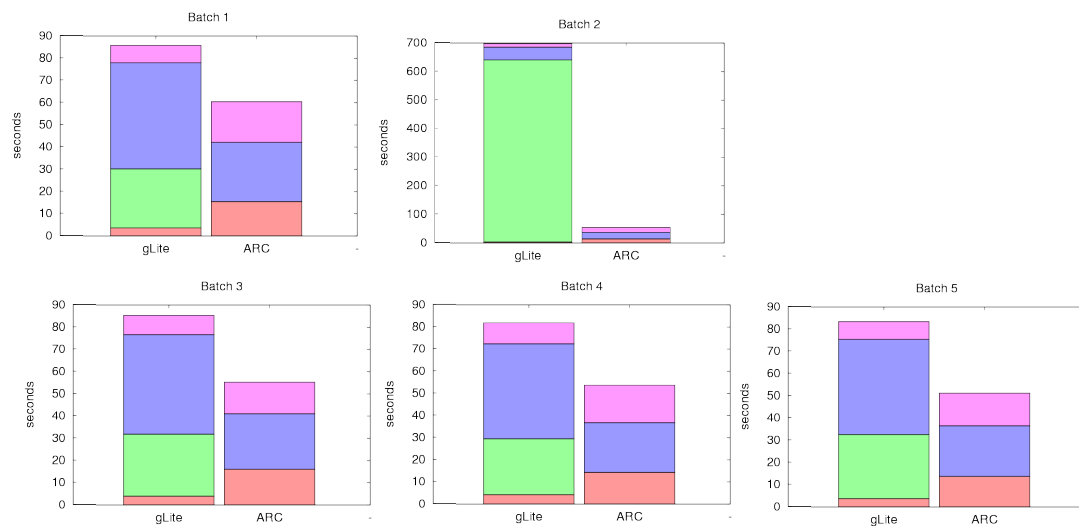
gLite VS ARC: job submission

- GATE application (50 jobs, ~5min each, 3 CEs)

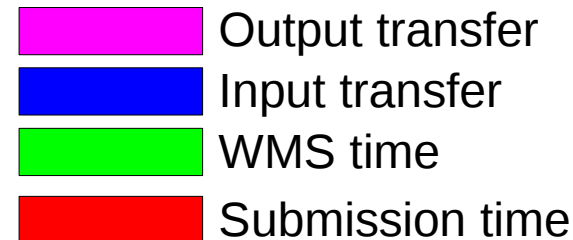
Average



All 5 batches



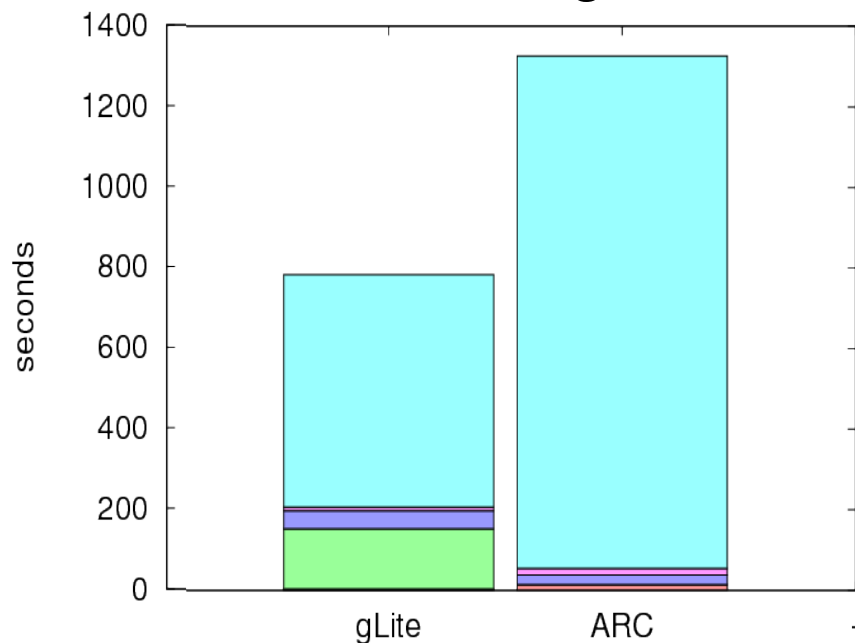
- gLite subm. + WMS > ARC subm.
- ARC data trsf ~ gLite's



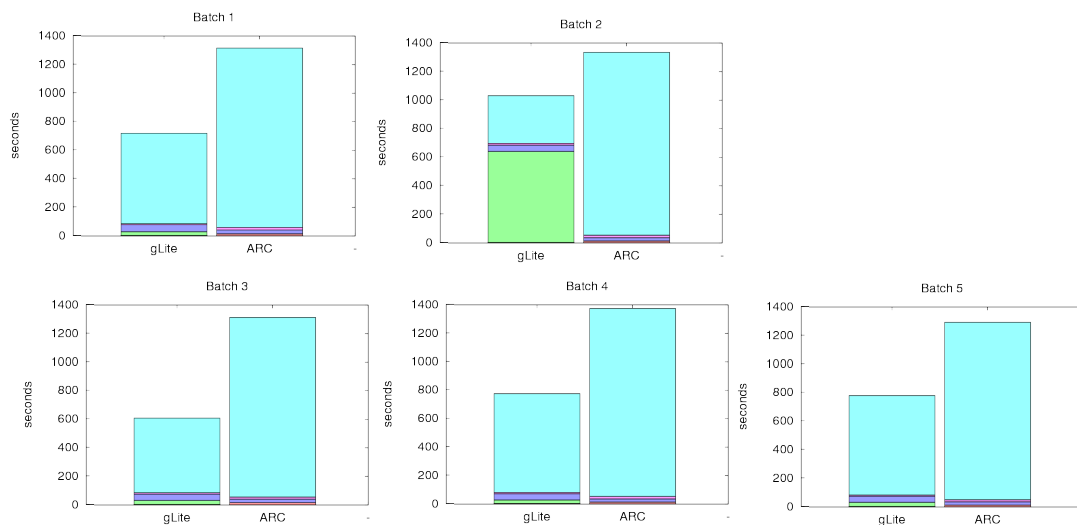
gLite VS ARC: job submission

- Middleware overhead = {reported running time} – {actual time on WN}

Average

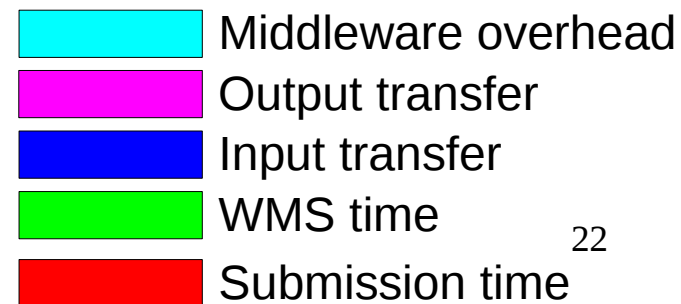


All 5 batches



- Middleware overhead dramatically high (side note: this time is saved by pilot jobs)

- (needs further investigation)



Monitoring and activity report

Workflow dashboard

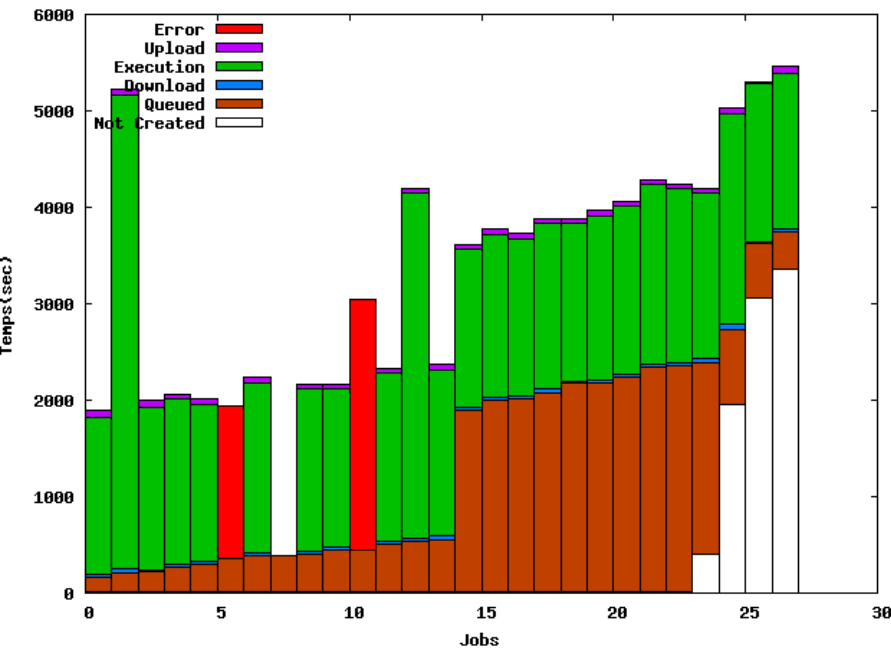
- Current user activity

Report from Jul 1th to the last day of Jul

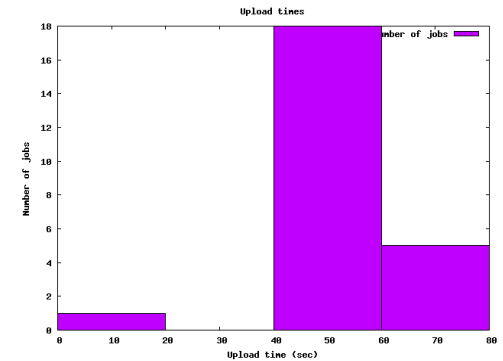
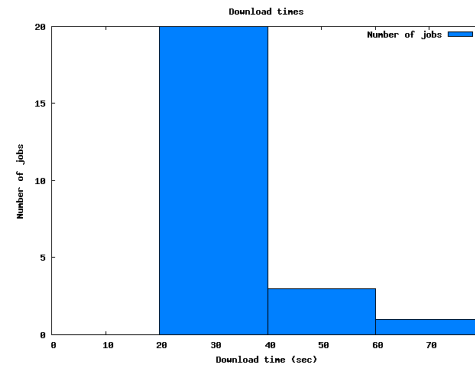
Actions	Date	User	Workflow ID	Total grid jobs	Done application jobs	Failed application jobs	Global success ratio	Application success ratio
Kill	Jul 7 08:54	Tristan Glatard/	workflow-XZ73UQ	29	21	0	72.00%	100.00%
Cleanup	Jul 6 18:44	Tristan Glatard/	workflow-DQ2ZGT	59	50	0	84.00%	100.00%
Cleanup	Jul 6 17:01	Tristan Glatard/	workflow-UUhzJF	52	1	0	1.00%	100.00%

- Detailed statistics (DIANE only)

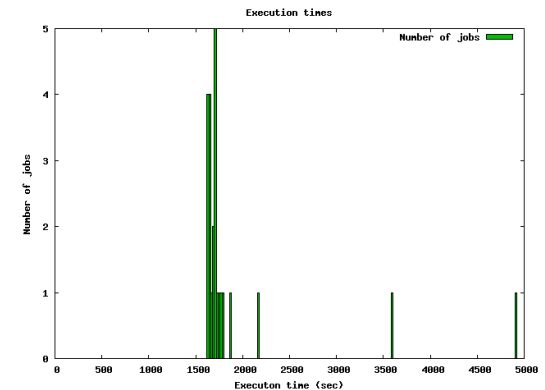
Job flow



Data transfers histograms

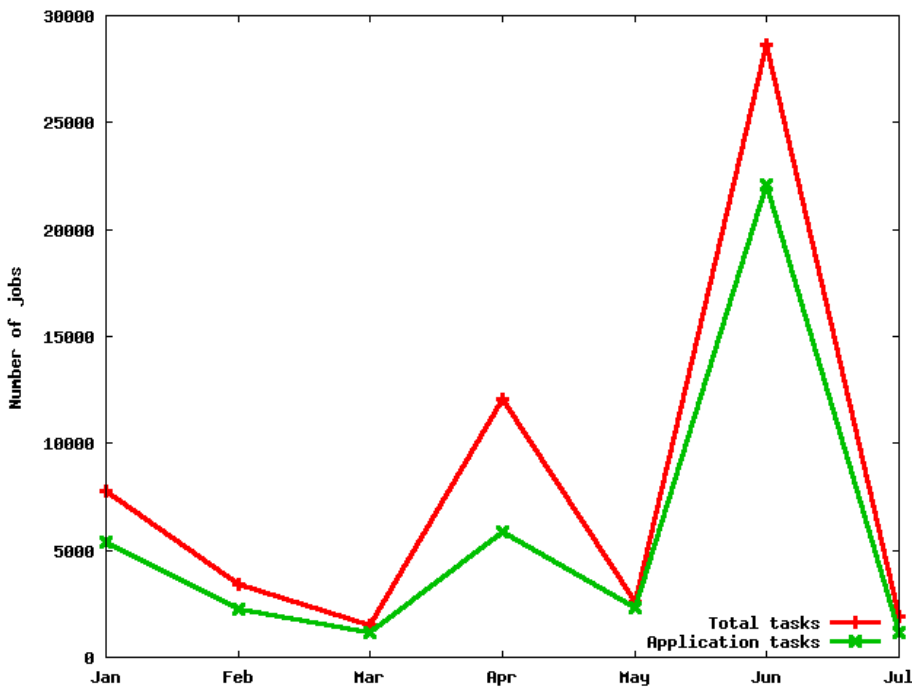


Running times histogram

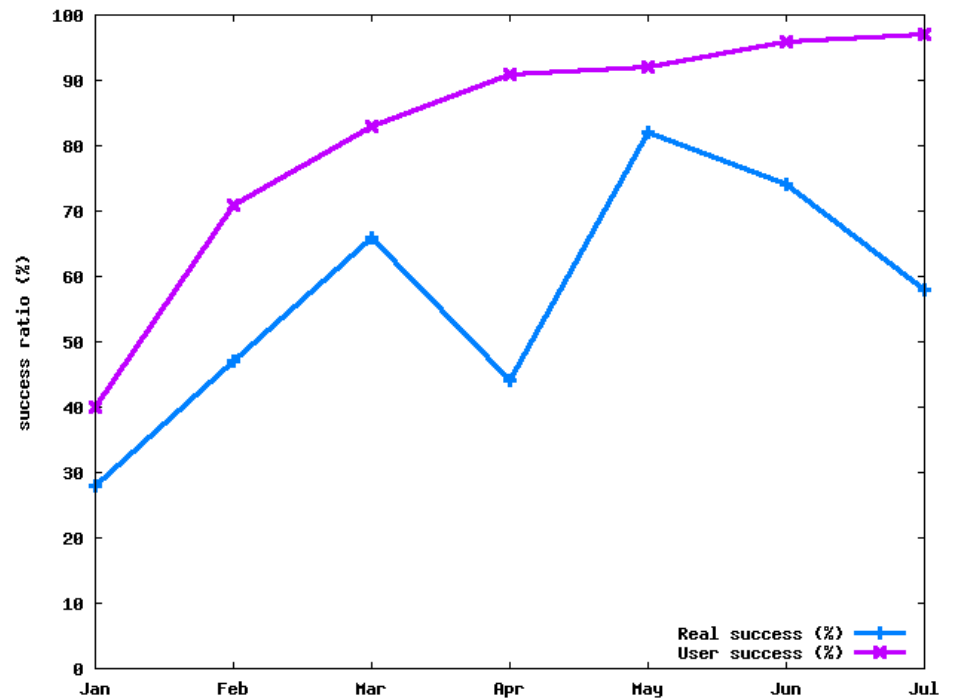


Activity report (1/1/09 - 7/7/09)

- 3 “real” users (+ 4 “fakes”)
- Job numbers



Success rate



- Average 7 workflows per day

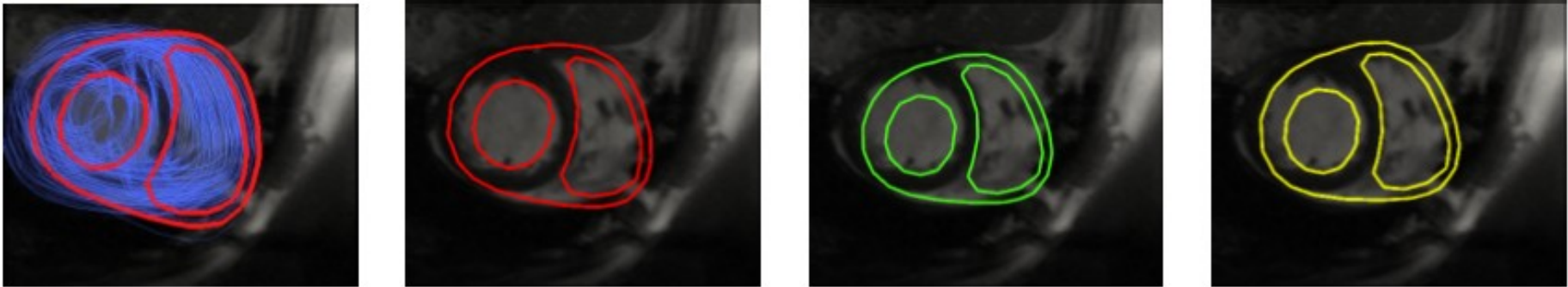
Outline

- Developments
- Applications
 - Cardiac segmentation
 - FIELD US simulation
 - GATE hadrontherapy simulation
- Modeling (prospective)

Cardiac segmentation

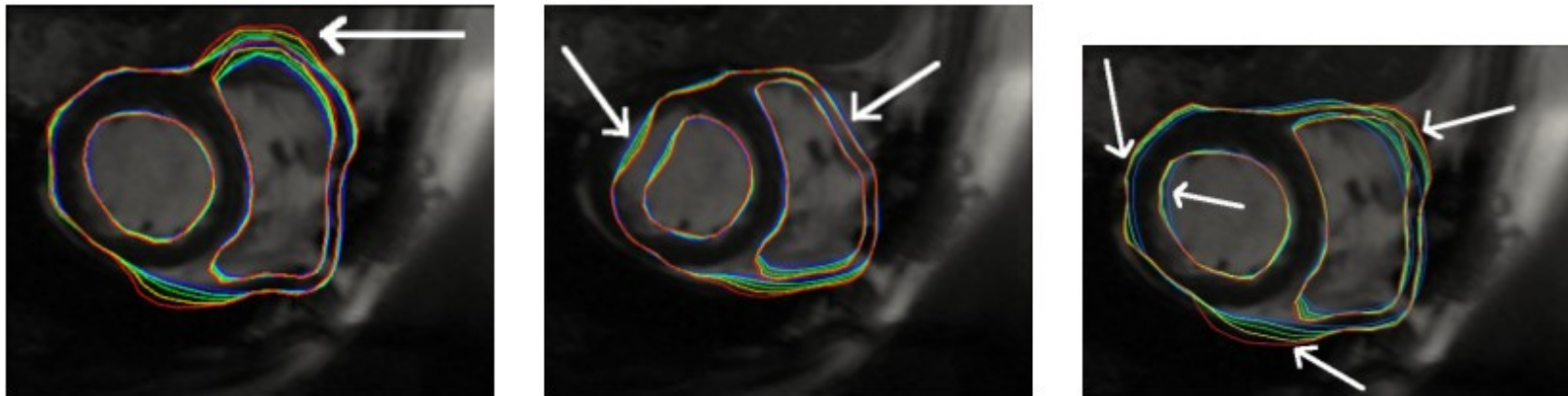
- Variations of initialization parameters

~ 3 min each



- Variations of the deformation force factor

~ 6.5 min each

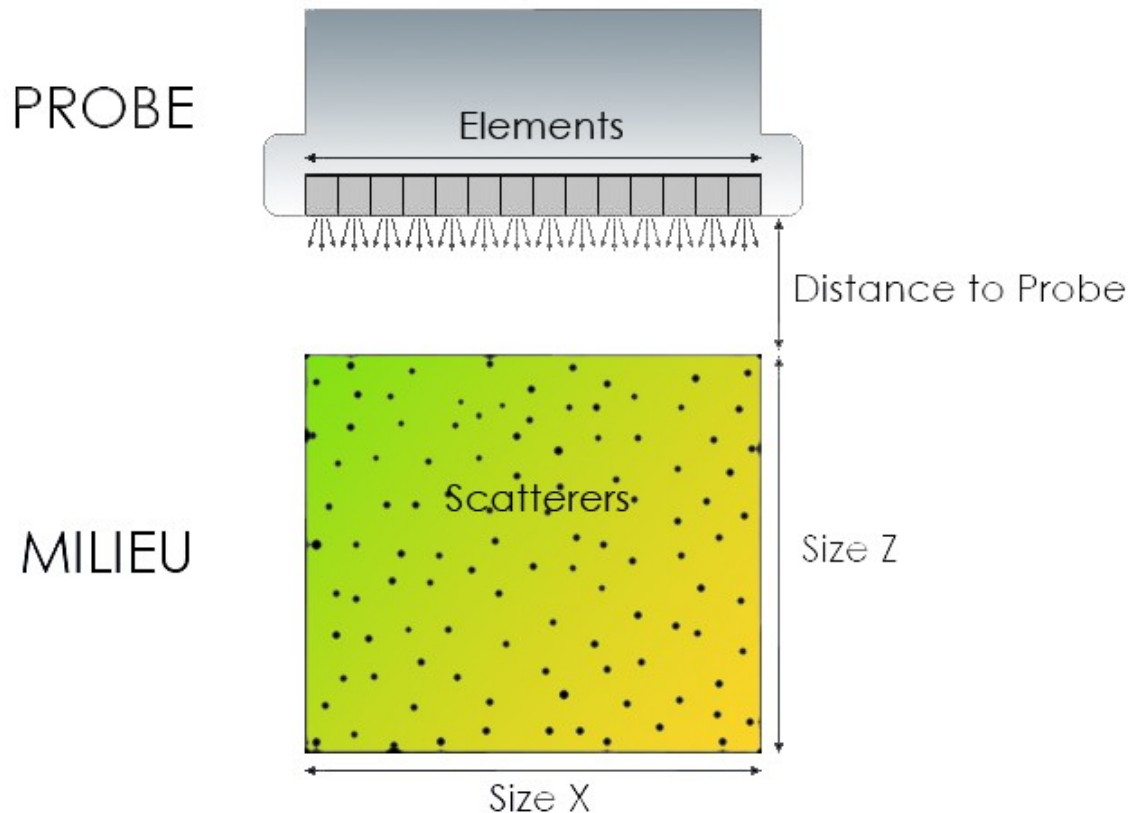


- Challenges

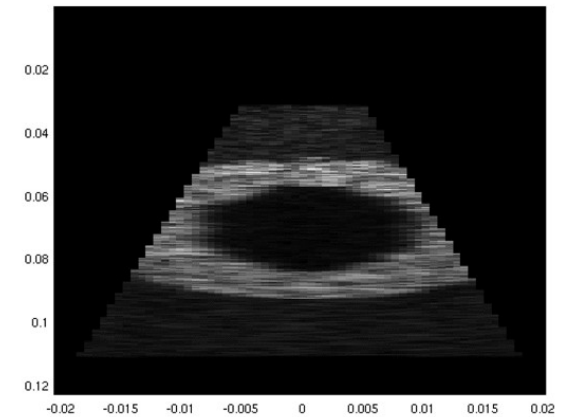
- Quantitative segmentation quality criterion ?
- Workflow language enhancements (Moteur2)

FIELD US simulation

- Principle



- Example on 2D beating heart

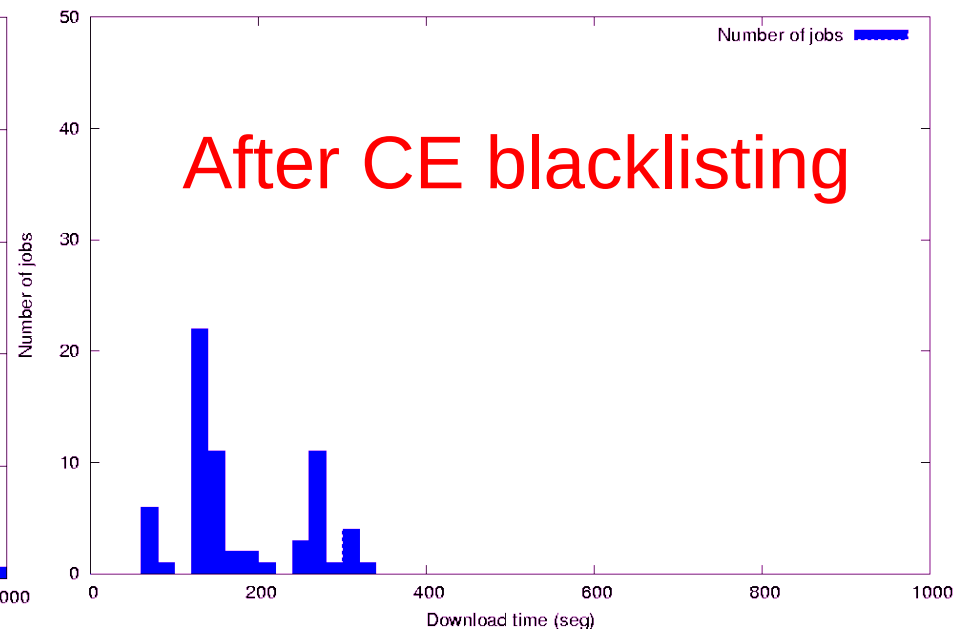
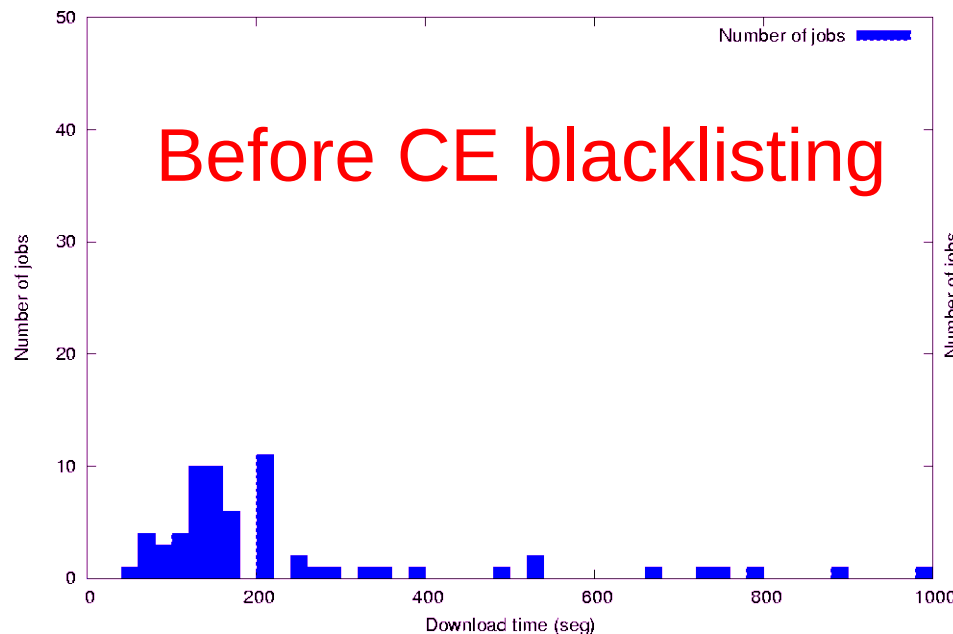


- 1920 lines (30 milieus x 64 lines)
- $> 16h \Rightarrow < 3h$
- 12% error (first try) \Rightarrow 2% (fine-tuning)

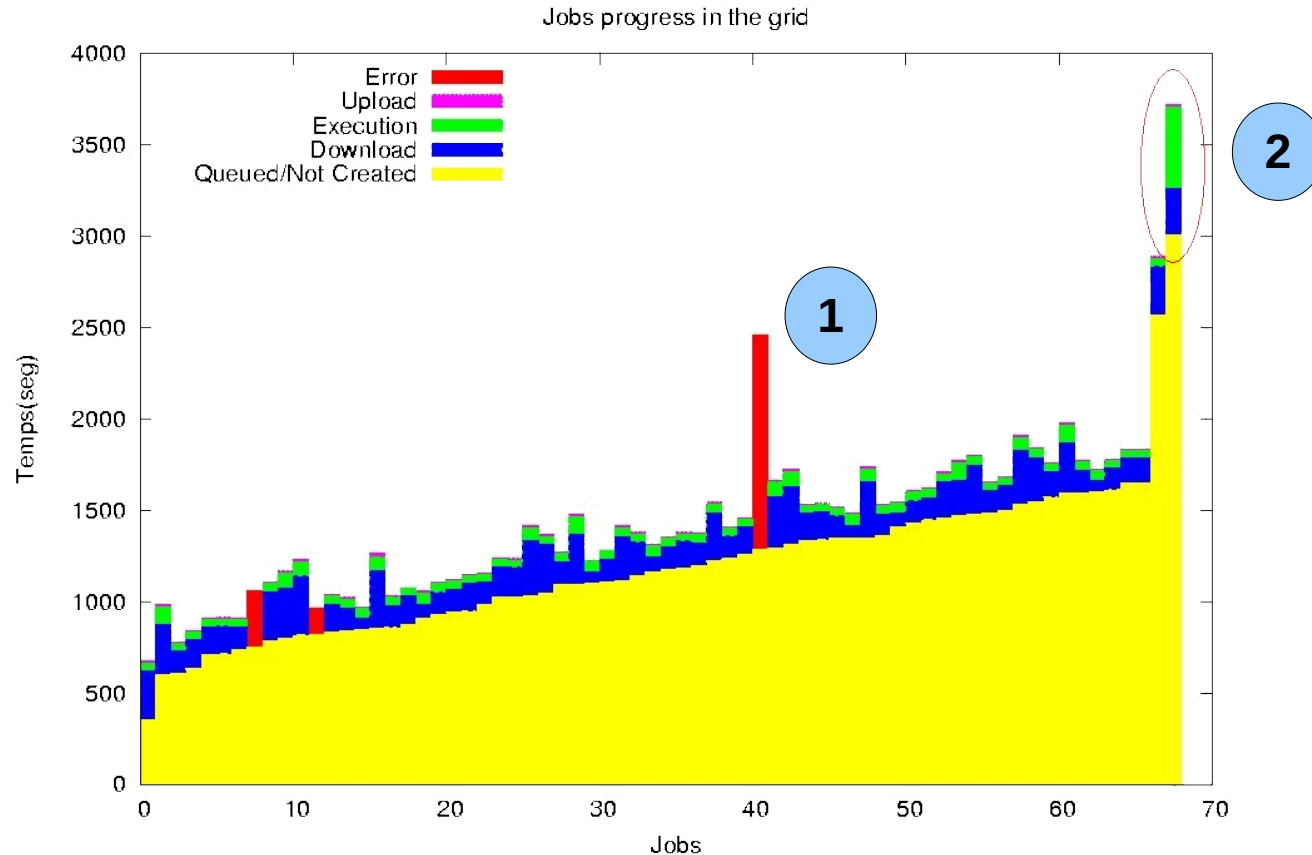
- Parallelism on lines
- Parallelism on milieus

FIELD execution on EGEE biomed

- Matlab code
 - Compiled (mcc, €500) and run with (free) Matlab Compiler Runtime (MCR)
 - On-the-fly MCR download and install
- Data transfer histograms (download)

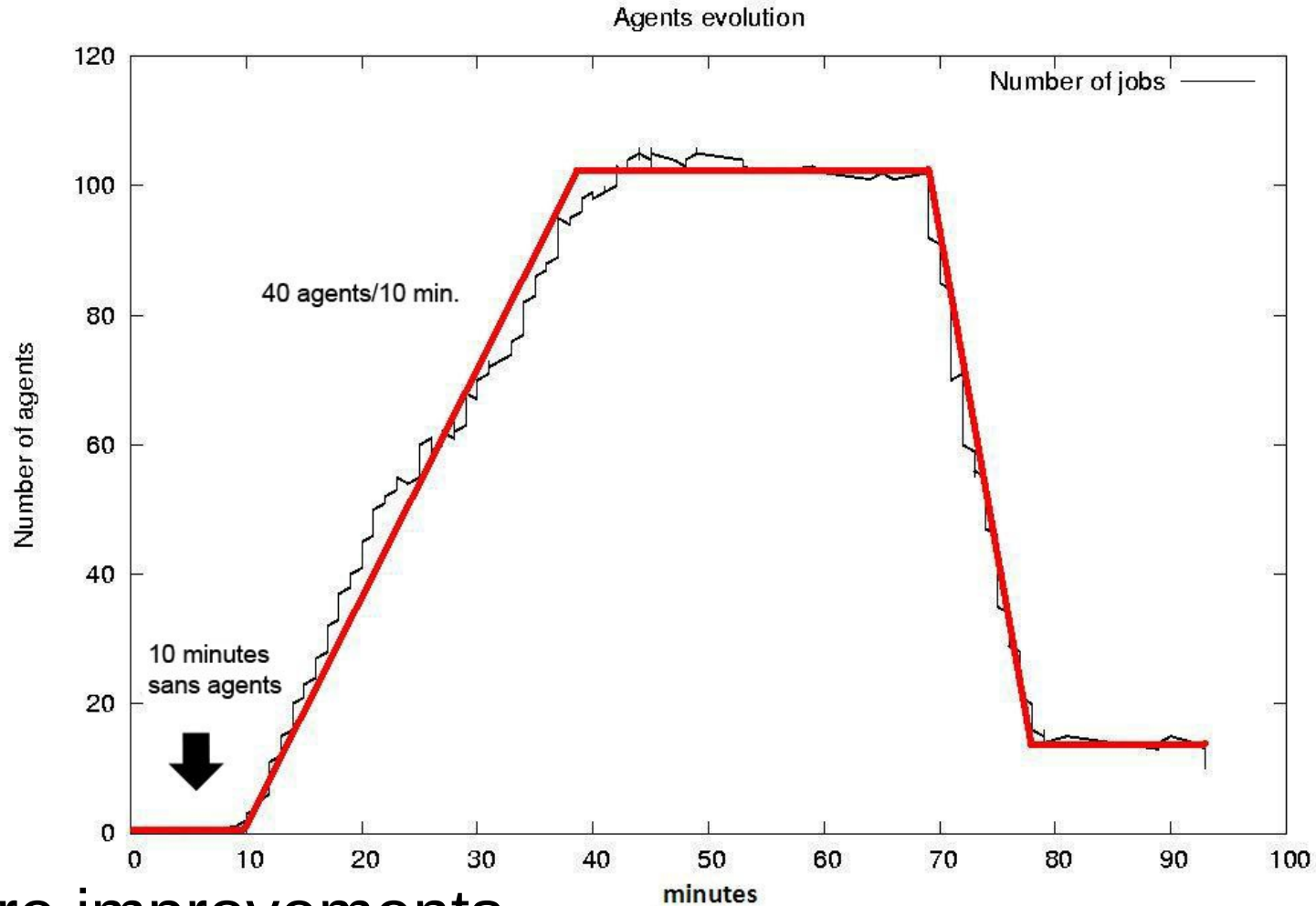


FIELD: other performance losses



- **1** Lately detected errors (result upload)
 - Upload test before execution (done)
- **2** Merge cost
 - On-the-fly merge ? (to be studied)

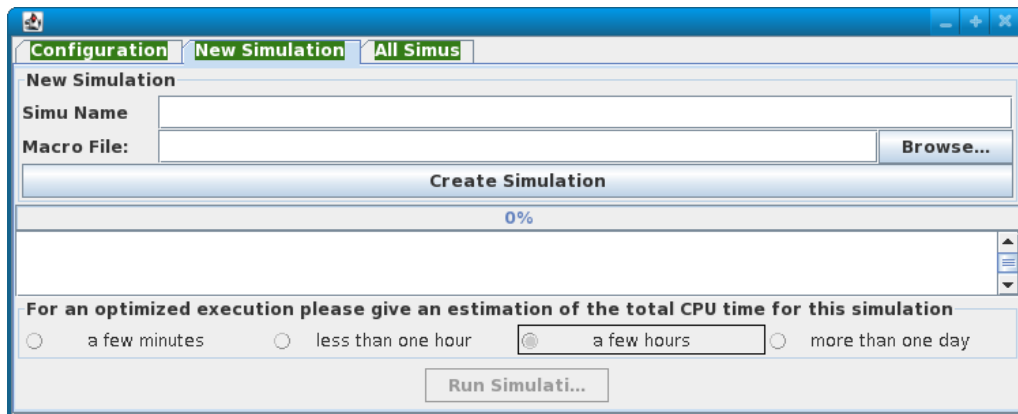
FIELD: DIANE agent registration



- Future improvements
 - More (close) resources
 - Improve agent controller

GATE VBrowser plugin

- Features [S. Camarasu-Pop]
 - Parses/checks GATE input files
 - Uploads inputs on the grid
 - GATE release selection
 - Workflow submission
 - Monitoring
 - Simulation history

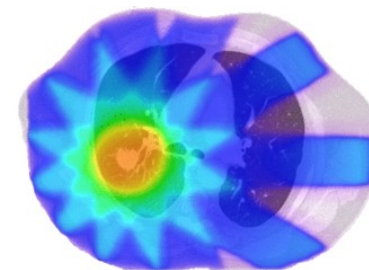


Submission interface
and history mgt



Monitoring interface

GATE load-balancing

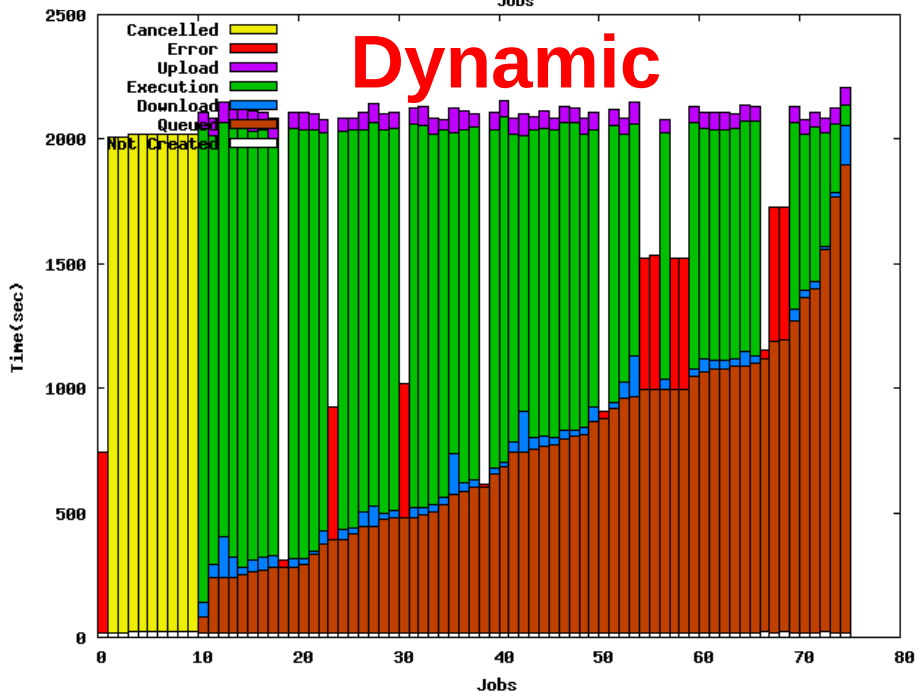
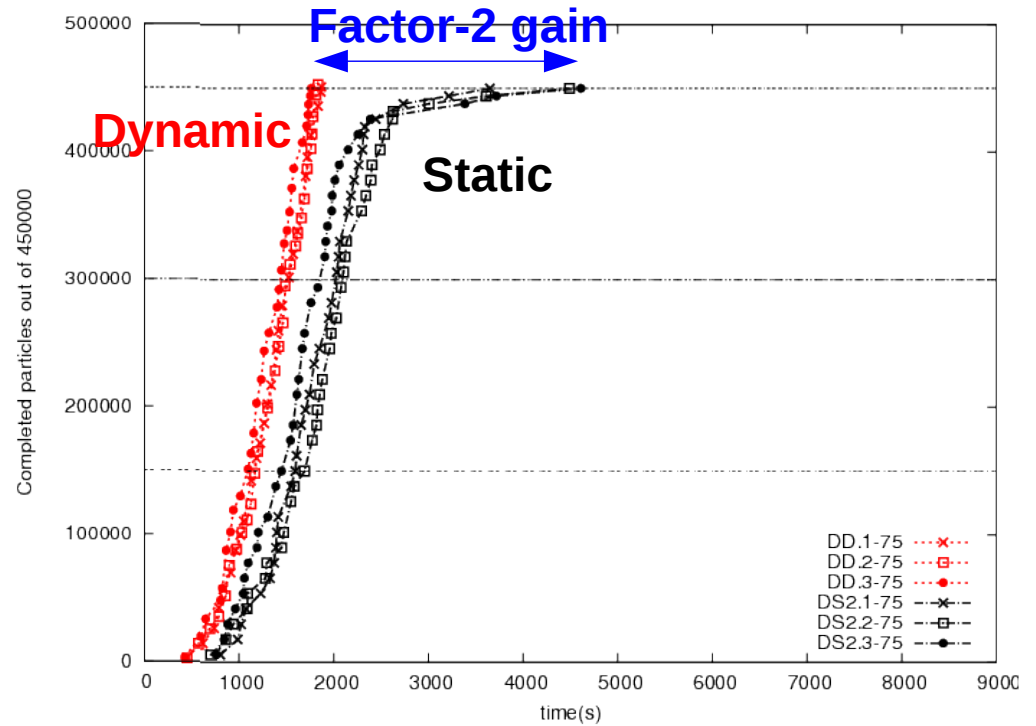
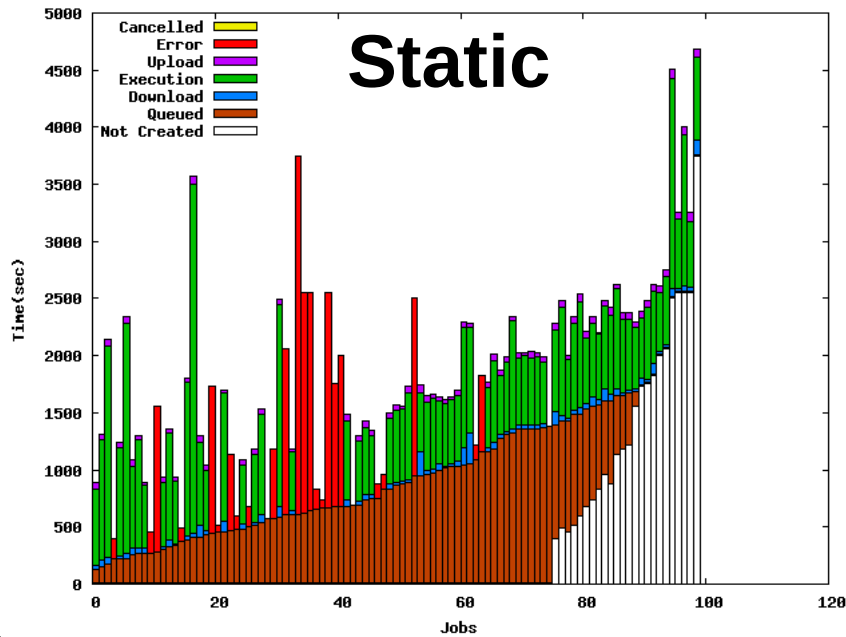


GATE hadrontherapy simulation

- GATE Monte-Carlo simulations
 - Simulation of particle trajectories
 - A few million of (independent) particles
- Dynamic particle distribution among pilots
 - Pilots
 - Receive the whole simulation
 - Compute until 'stop' signal received from master
 - Finally upload results
 - Master
 - Periodically sums up computed particles
 - Sends 'stop' signals when total is reached
 - Resubmit tasks in case of upload errors

[S. Camarasu-Pop]

GATE: dynamic load-balancing



- Reduced error impact
- Better resource exploitation
- Elapsed time divided by 2
- Extend it to other applications ?

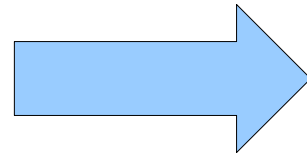
Outline

- Developments
- Applications
- Modeling (prospective)

A model for pilot-job applications

- Problem

- n pilots submitted at $t = 0$
- grid latency L is random, $P(L < t) = F_L(t)$ is measured



how many
available pilots at time t ?
 N

- Modeling

$$P(N = k) = \binom{n}{k} F_L(t)^k (1 - F_L(t))^{(n-k)}$$

Choose k among n For those k , $L < t$
For the others, $L > t$

- N follows a binomial distribution, thus:

$$E_N(t) = nF_L(t)$$

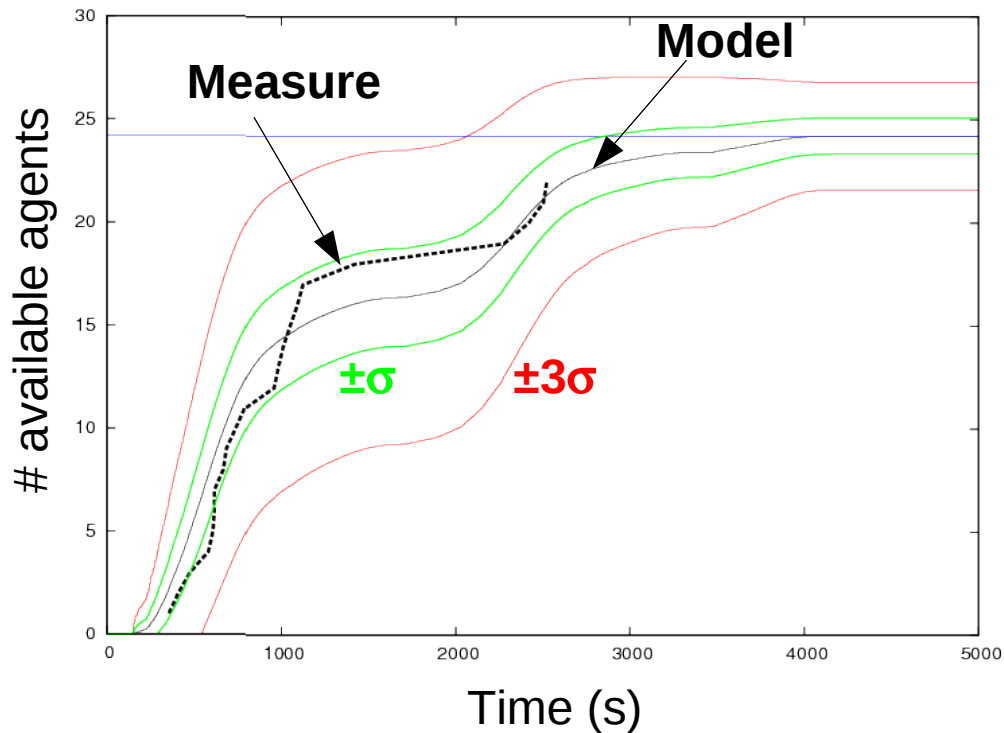
$$\sigma_N(t)^2 = n(1 - F_L(t))F_L(t)$$

- Job error rate also considered

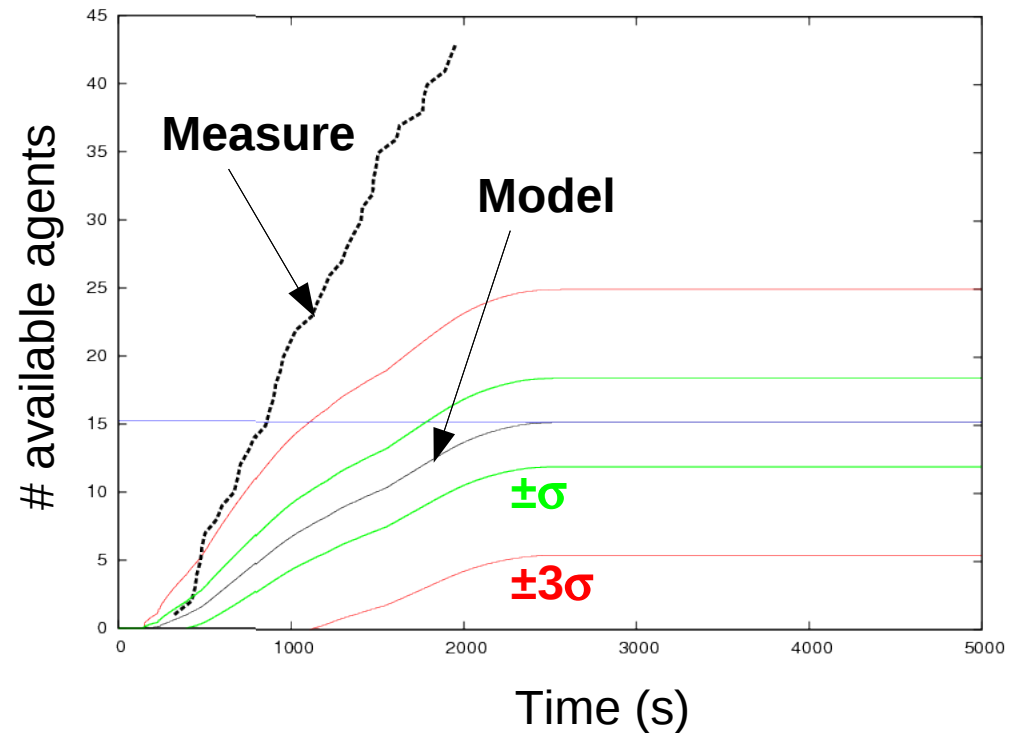
Model evaluation

- Model built from probe jobs

Error rate < 15%



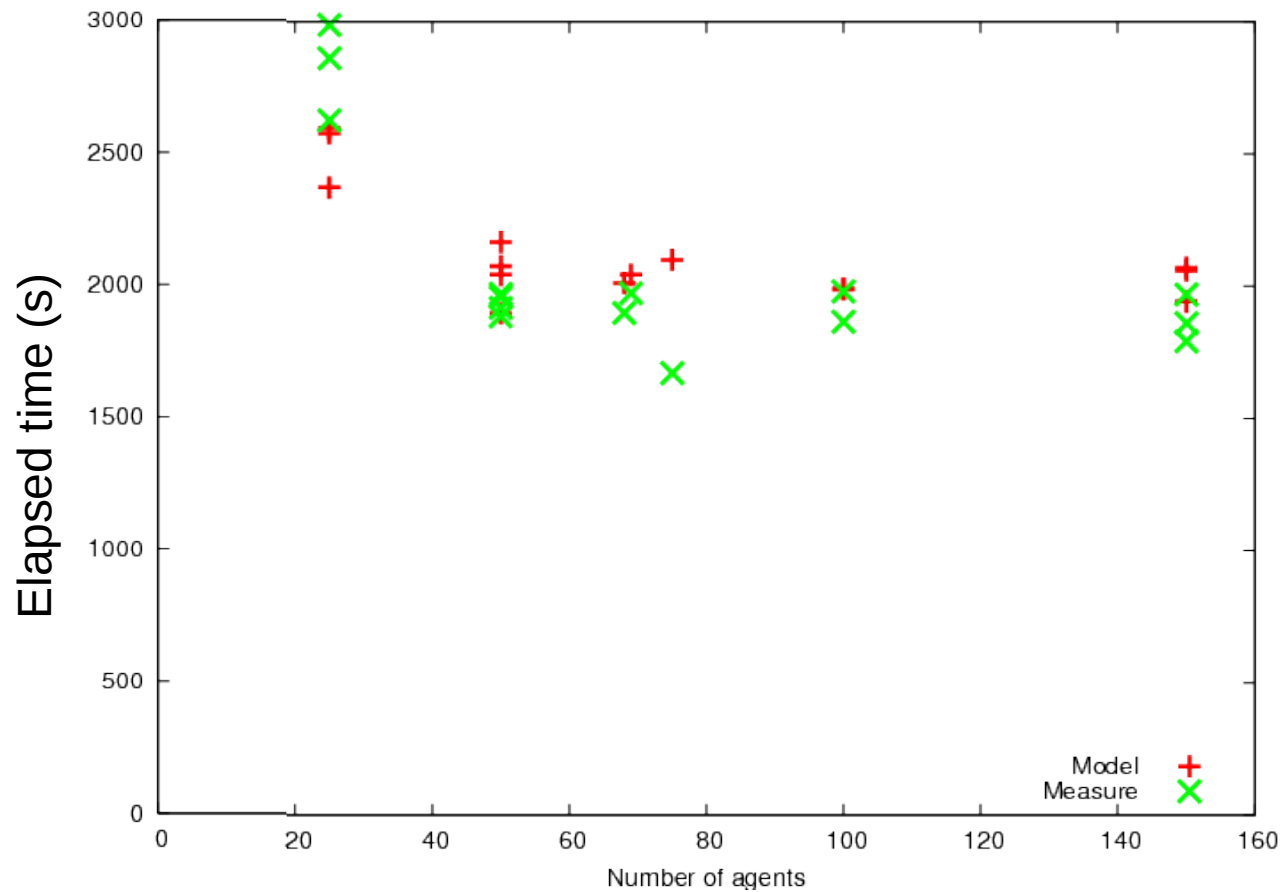
Error rate > 15%



- Use it to steer pilot submission ?

Application to elapsed time estimation

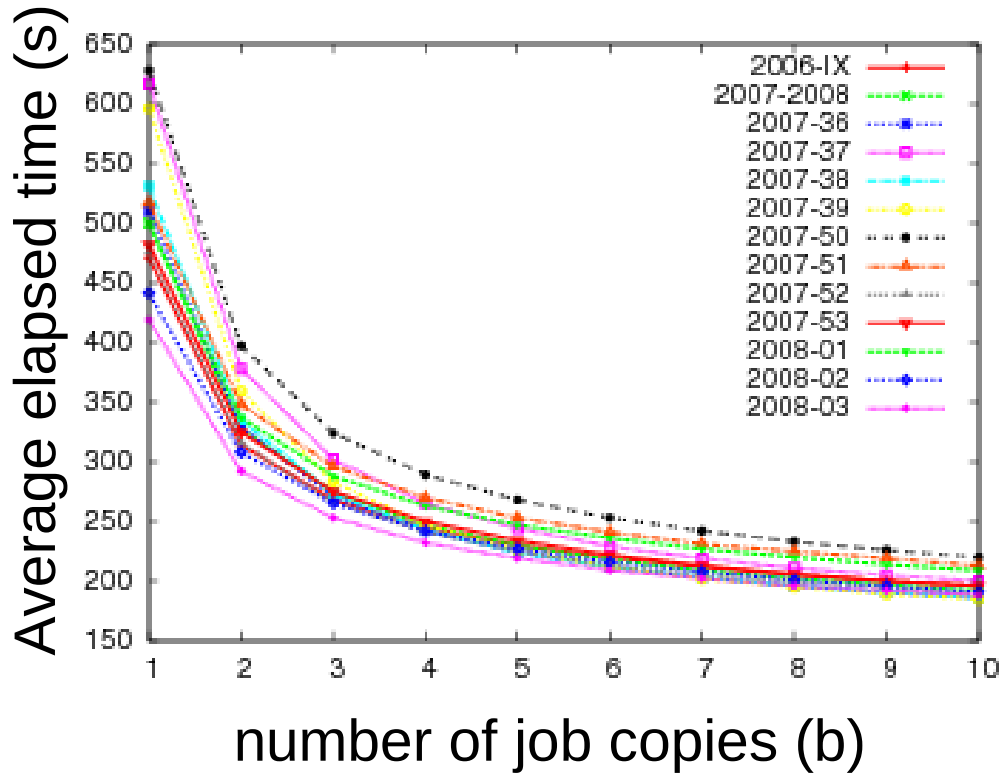
- Tested on a 6h45min GATE simulation



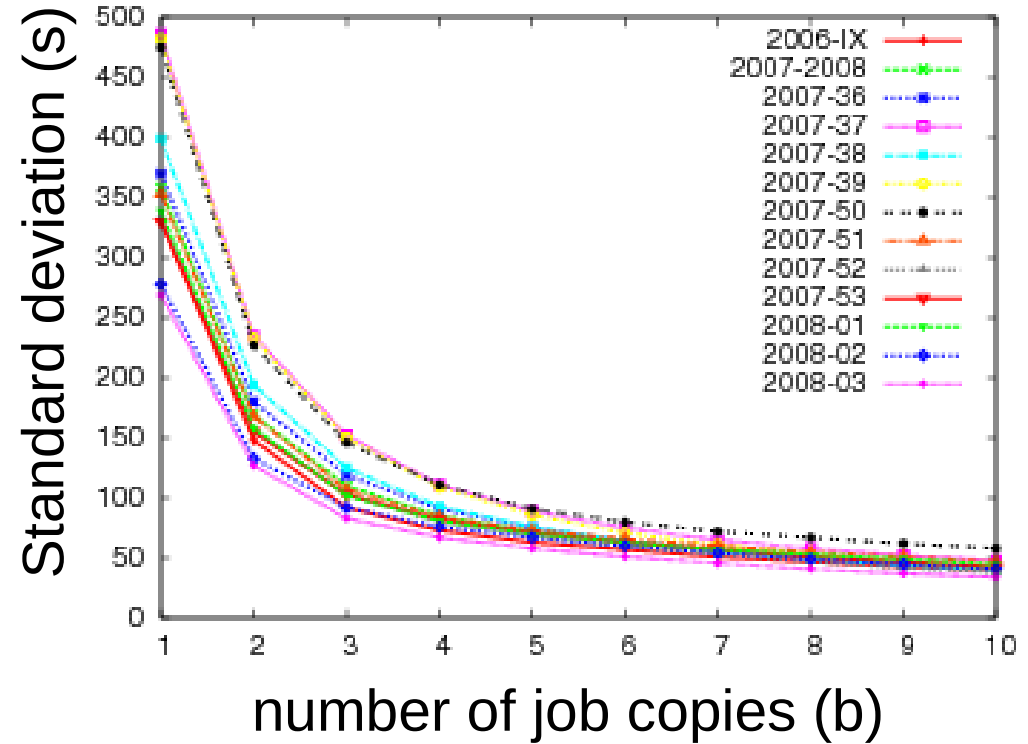
Model for redundant job submission

- Same method (probabilistic)

Mean



Stdev



- Not validated in real conditions yet

Credits



Inserm

Institut national
de la santé et de la recherche médicale



Creatis

Sorina Camarasu-Pop
Mark Chin Chuen Teck ; *QoS extensions*
Patrick Clarysse ; *Cardiac segmentation*
Christopher Casta ; *Cardiac segmentation*
Denis Friboulet ; *US simulation*
Carlos Gines Fuster ; *FIELD grid porting, monitoring tools*
Hervé Liebgott ; *US simulation*
David Sarrut ; *GATE simulation*
Wen-Jun Tan ; *GALQS*
Alejandro Tovar de Duenas ; *Agent controller*

VI-e medical software

Silvia D. Olabbarriaga ; *AMC Amsterdam*
Piter T. de Boer ; *Universiteit Van Amsterdam*
Spiros Koulouzis ; *Universiteit Van Amsterdam*

Pilot jobs (DIANE)

Jakub T. Moscicki ; *CERN*

ARC interoperability

Henning Müller ; *University Hospitals Geneva, CH*
Oxana Smirnova ; *Lund University, Sweden*
Xin Zhou ; *University Hospitals Geneva, CH*

D-Grid interoperability

Silvia D. Olabbarriaga ; *AMC Amsterdam*
Andreas Hoheisel ; *Fraunhofer Berlin*
Dagmar Krefting ; *Charité hospitals Berlin*

Modeling

Diane Lingrand ; *University of Nice Sophia-Antipolis*
Johan Montagnat ; *CNRS, University of Nice Sophia-Antipolis*

Grid support

<https://gus.fzk.de>

