CREAM and ICE Test Results

Alvise Dorigo – INFN Padova
Massimo Sgaravatto – INFN Padova
EGEE-II JRA-1 All-Hands Meeting
7th-9th November 2006

www.eu-egee.org

Information Society and Media

Updates since last report @ EGEE06 conference in Geneva

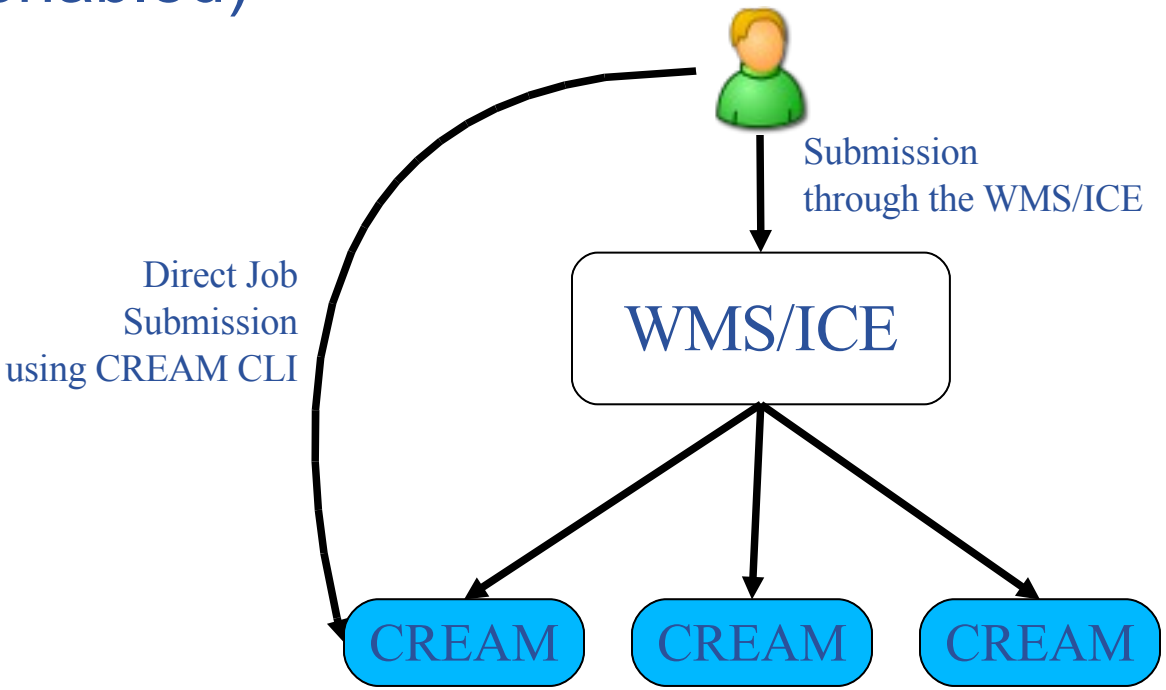
- Several bug fixes and enhancements have been made in CREAM
  - Adoption of **new glexec** (which fixes bug **#20744**) required some code changes in BLAH and CREAM
  - Support for JSDL and BES (see next slide)
- Several bug fixes and enhancements also in ICE
- Installation of UI and WMS 3.1 (ICE enabled) done on the preview testbed (thanks to CNAF's system administrators)
  - Tests reported in this presentation performed on this layout

Basic Execution Service (BES)

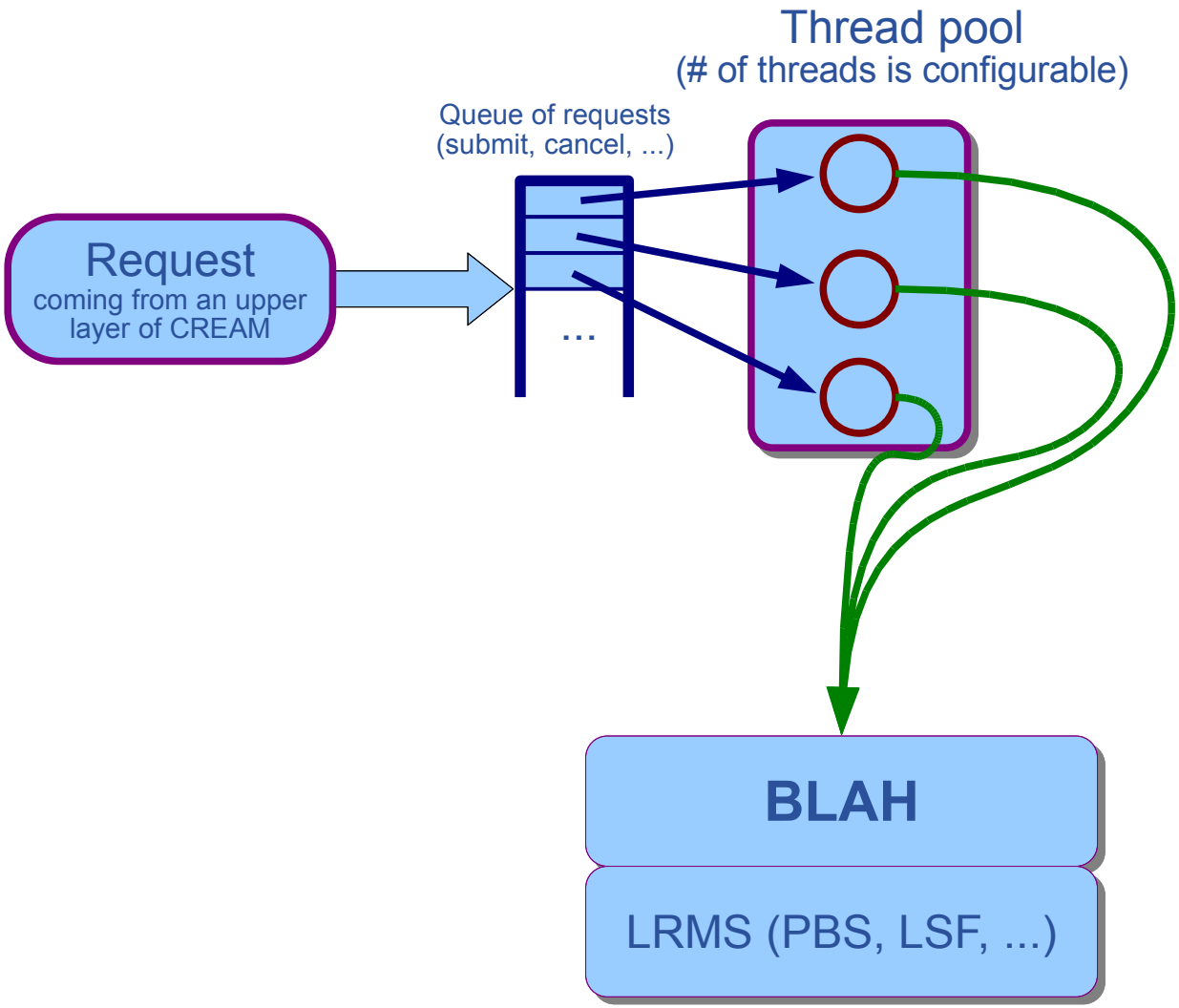
- Basic Execution Service (BES) is a new GGF specification that defines Web Services interfaces for creating, monitoring and controlling computational entities called activities
  - Very primitive now
  - Doesn't cover all CREAM functionality
- **Activities in BES are defined using the Job Submission Description Language (JSDL)**
  - JSDL: GGF specification for describing the requirements of computational jobs for submission to resources in Grid environments.
- **First implementation of BES support done in CREAM**
  - This will be shown at SC'06 (Tampa-FLORIDA) in a interoperability demo with other computational services

Tests performed

- Direct submission to CREAM CE using the command line CREAM UI
- Submission to CREAM CE via gLite WMS (ICE enabled)



Quick overview of CREAM Arch.



Direct submission tests

- **Stress tests:**
  - Submission of an increasing number of jobs from UI @ CNAF (`pre-ui-01.cnaf.infn.it`) to CREAM CE @ Padova (CEId `cream-01.pd.infn.it:8443/cream-pbs-long` with 4 worker nodes)
    - Submission of **100** jobs from 1, 2, 5, 10 parallel threads
    - Submission of **250** jobs from 1, 2, 5, 10 parallel threads
    - Submission of **500** jobs from 1, 2, 5, 10 parallel threads
    - Submission of **1000** jobs from 1, 2, 5, 10 parallel threads
    - Submission of 2**000** jobs from 1, 2, 5, 10 parallel threads
  - CREAM has been configured with **50 threads** (see figure in previous slide)
  - Tests have been made using a pre-delegated proxy
  - Measured values:
    - The number of failed jobs (taking into account the reported failure reasons)
    - The time taken to submit each job to the CREAM CE (i.e. the time needed to get back the CREAM JobID)
    - The time needed to submit the job to the LRMS via BLAH (i.e. the time needed to get the BLAH jobid)

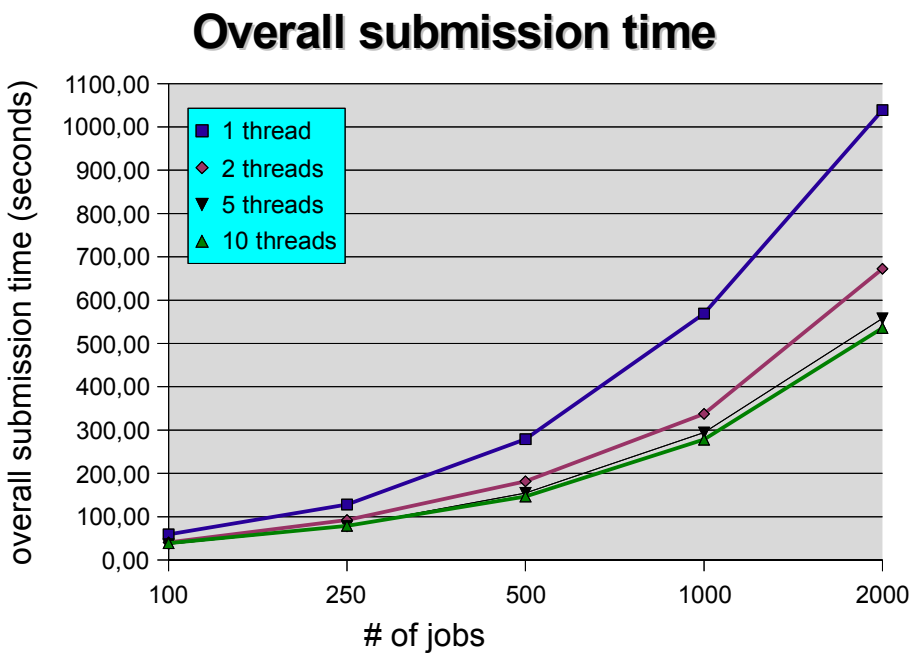
Submission results

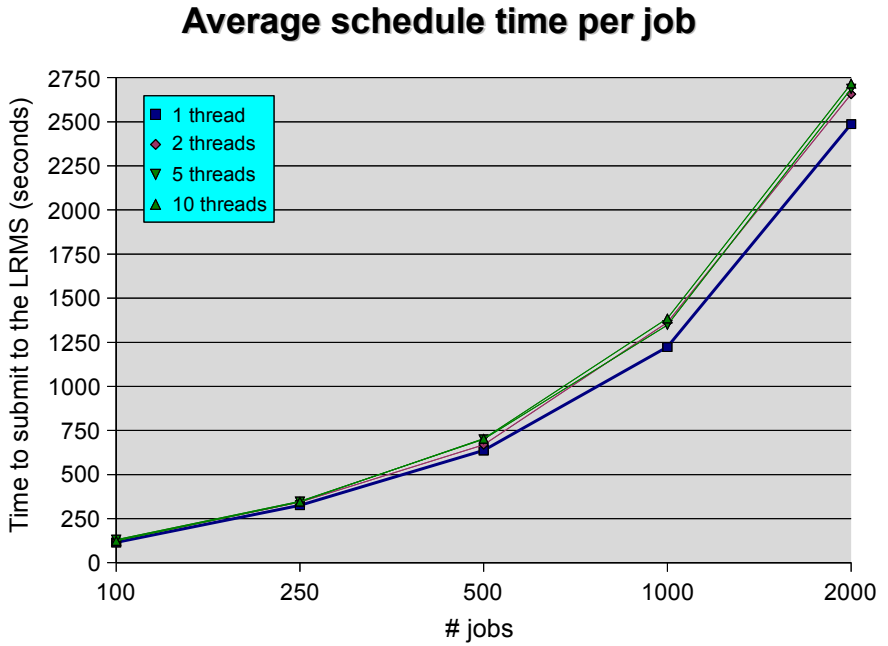
Job JDL:

```
Executable = "test.sh";
StdOutput = "std.out";;
InputSandbox = {"gsiftp://grid005.pd.infn.it/Preview/test.sh"}
OutputSandbox = "out.out";;
OutputSandboxDestURI = {"gsiftp://grid005.pd.infn.itPreview/Output/std.out"};
```

```
#!/bin/sh
echo "I am running on `hostname`"
echo "I am running as `whoami`"
sleep 600
```

- **Results:**





- **… more tests and details in the CREAM web site (`http://grid.pd.infn.it/cream/field.php`) under "Test Results"**

Direct submission tests: failures

# of threads	100 jobs	250 jobs	500 jobs	1000 jobs	2000 jobs
1	0/100	0/250	0/500	7/1000 (0,7%)	22/2000 (1,1%)
2	0/100	0/250	1/500 (0,2%)	4/1000 (0,4%)	26/2000 (1,3%)
5	3/100	1/250 (0,4%)	4/500 (0,8%)	4/1000 (0,4%)	24/2000 (1,2%)
10	1/100 (1%)	3/250 (1,2%)	1/500 (0,1%)	8/1000 (0,8%)	19/2000 (0,9%)

- 96 jobs failed because of gridftp problems when transferring input sandbox
- 29 jobs failed because of problems when using glexec

Some conclusions on the direct job submission tests

- **Submission time to CREAM looks good**
- **Submission time to LRMS (schedule time) is not too good**
  - Necessary to perform some profiling to better understand where most of the time is spent
  - Some tests increasing/decreasing the number of CREAM threads (and see how performance is impacted) are needed as well
- **Overall efficiency is good but can be improved**
  - gridftp problem when transferring input sandbox is not completely understood, even if we suspect on BLAH bug #20357

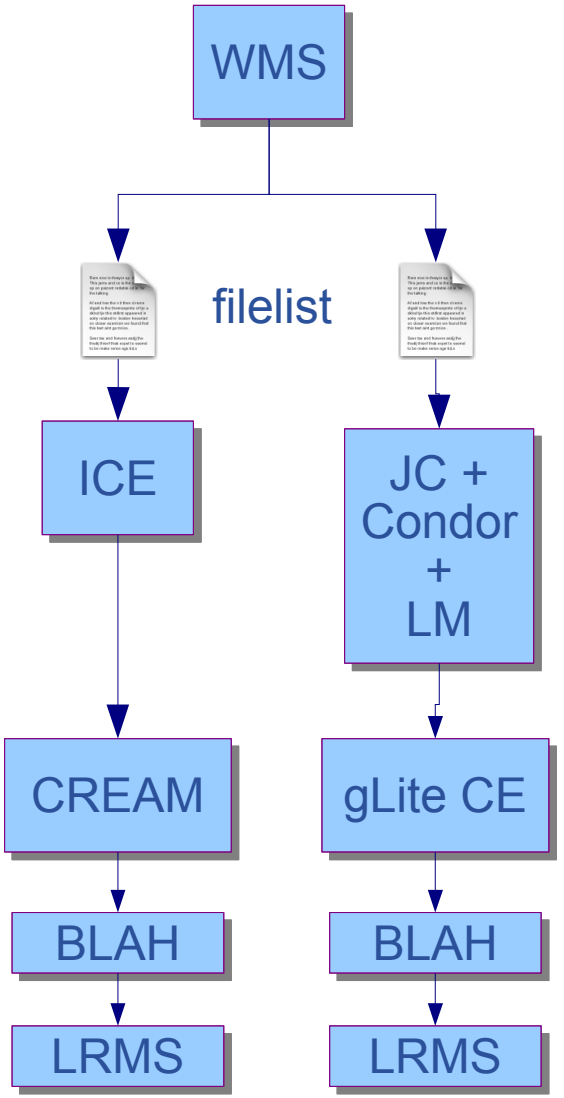
Submission through WMS (ICE vs JC+Condor+LM)

- **Testbed configuration**
  - WMS, BDII, UI @ INFN-CNAF
  - A single CREAM CE @ INFN-PADOVA configured with **50 threads** (**`cream-01.pd.infn.it:8443/blah-pbs-long`**)
  - A single gLite 3.0 CE @ INFN-PADOVA (**`cert-04.pd.infn.it:8443/blah-pbs-long`**)

- **What has been measured**
  - Efficiency (reporting the number of failed jobs along with the failure reasons)
  - For both JC+Condor+LM and ICE: for each job, the time needed for the submission to the LRMS and the corresponding JC+Condor+LM/ICE throughput
  - Only for ICE: the time needed for the submission to the CREAM CE and the corresponding ICE throughput
    - Not straightforward to distinguish submission to CE vs submission to LRMS in the JC+Condor+LM scenario

Submission through WMS (ICE vs JC+Condor+LM)

- **How the tests have been performed**
  - ICE/JC is turned OFF
  - Submission of 1000 jobs to the WMS in order to fill the ICE/JC filelist
  - ICE/JC is turned ON, so it can start to satisfy the submission requests
  - deep and shallow resubmission were disabled
- **How the measurements have been performed**
  - **Tstart** = LB timestamp of **first ICE/JC dequeued event** (i.e. request removed from the filelist)
  - **Tstop** = LB timestamp of the **last "Transferred OK to CE"** event (when measuring throuput to submit to CE for ICE scenario) or timestamp of submission event in the BLAH accounting log file* (when measuring throughput to submit to LRMS for both ICE and JC+Condor+LM scenarios)
  - **Throughput = # jobs / (Tstop - Tstart)**



Job JDL:

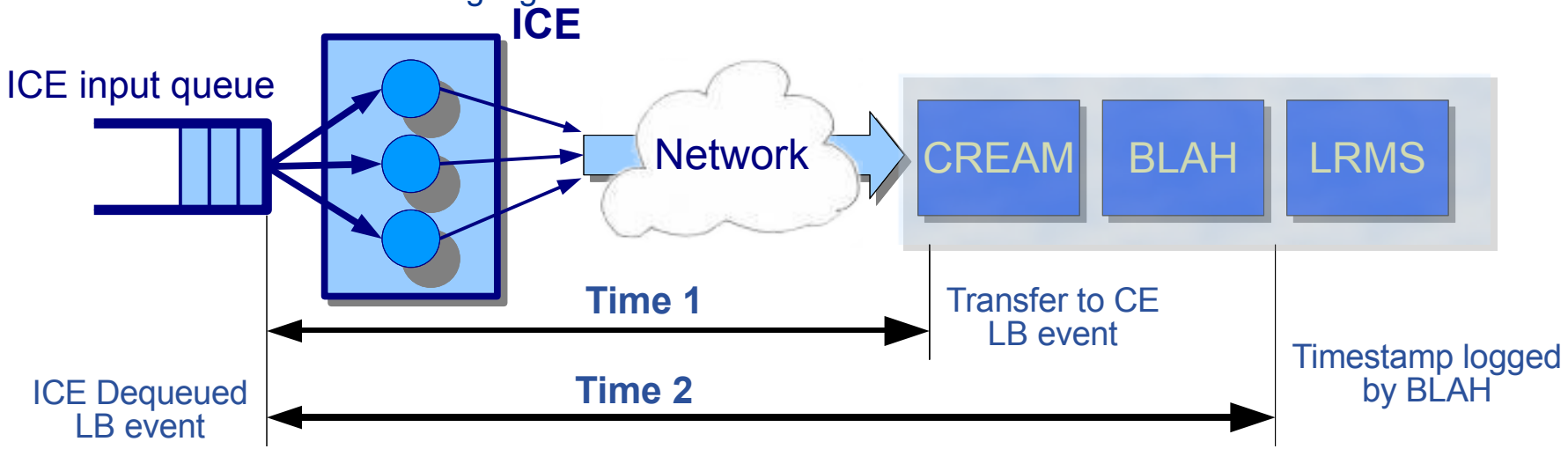
```
Executable = "test.sh";
StdOutput = "std.out";;
InputSandbox = {"gsiftp://grid005.pd.infn.it/Preview/test.sh"}
OutputSandbox = "out.out";;
OutputSandboxDestURI = {"gsiftp://grid005.pd.infn.itPreview/Output/std.out"};
RetryCount = 0;
ShallowRetryCount = 0;
```

```
#!/bin/sh
echo "I am running on `hostname`"
echo "I am running as `whoami`"
sleep 600
```

* This timestamp logged by BLAH to the accounting log for DGAS is not very accurate, but the overall results shouldn't be affected too much

ICE throughput test: submission of 1000 jobs to the WMS

- Performed tests using different configuration of ICE (5, 10, 15, 20, 25, 30 threads)
- Each ICE thread performs the following operations:
  1. parsing of classad request
  2. setting up the user credentials (gsoap-plugin)
  3. connection to endpoint (trustmanager authentication) and user proxy delegation (`getProxyReq()`, `GRSTx509MakeProxyCert()`, `putProxy()`)
  4. connection and dial with CREAM service
  5. Event logging to LB (`jobRegister`, `wms_dequeued`, `cream_transfer_start`, `cream_transfer_ok`, etc...)
- **Time_1**: time between the first "`dequeued by ICE`" LB event and the last "`Transferred to CE`" LB event (see previous slide)
- **Time_2**: time between the first "`dequeued by ICE`" LB event and the last timestamp logged by BLAH in the accounting log file



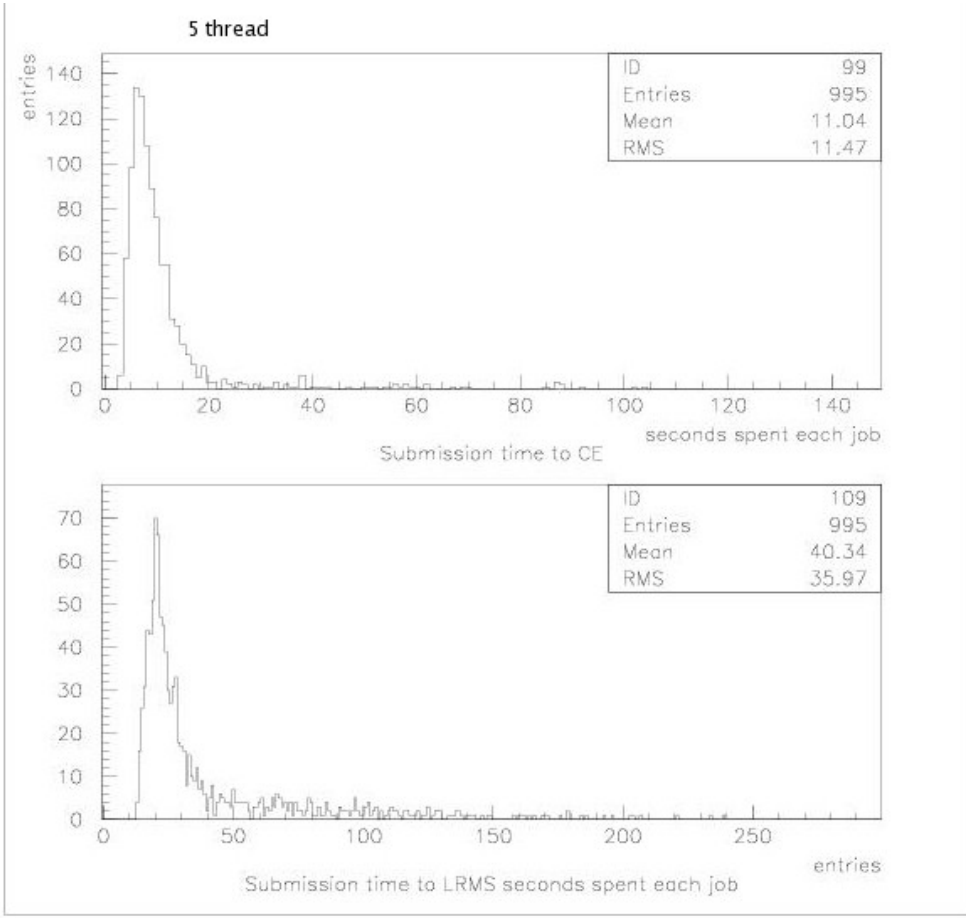
Test results

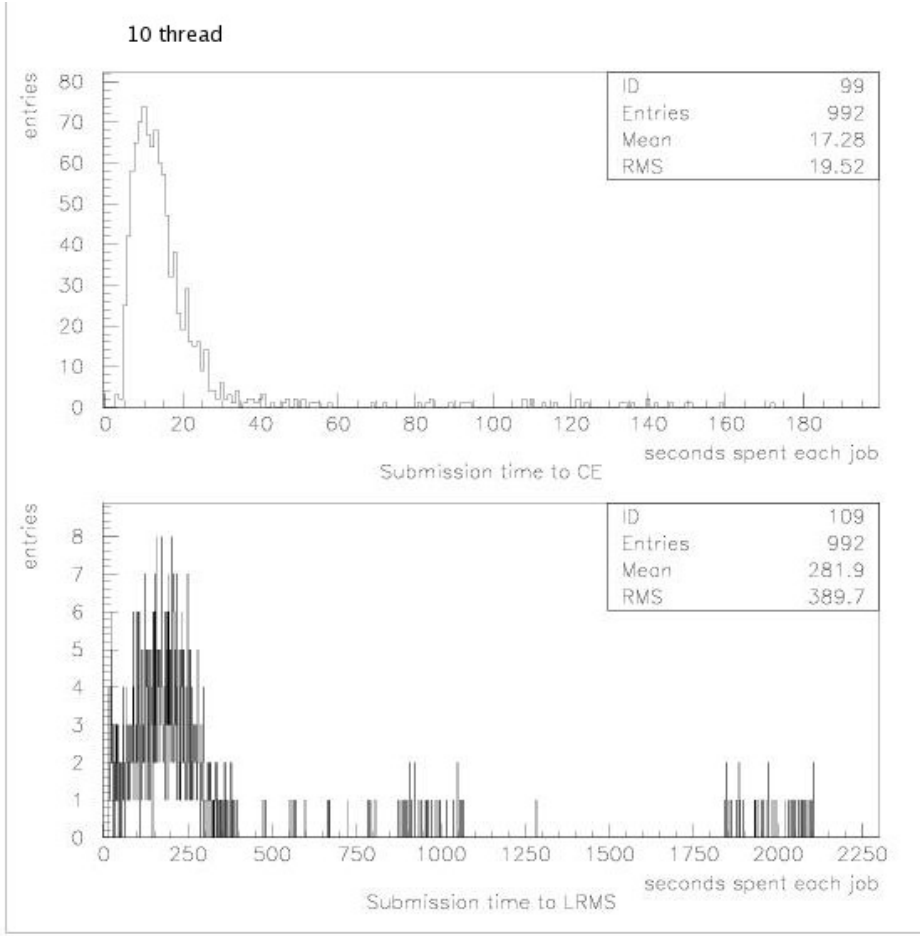
# Threads	Submission rate to CE (job/min)	Submission rate to LRMS (job/min)	% success (considering all jobs)	% success (considering only jobs managed by ICE)	# jobs failed for gridftp problems when transferring ISB	# jobs failed for problems when using glexec	# jobs failed for blah submission failure	# jobs never transferred to ICE*
5	16,2	16,2	99,1	99,6	4	0	0	5
10	21,6	19,8	98,3	99,1	9	0	0	6
15	22,8	19,8	99,2	99,3	6	1	0	1
20	24,6	19,8	98,4	99,0	5	2	3	6
25	25,8	19,8	99,2	99,5	3	0	2	3
30	29,4	19,8	98,9	99,7	6	1	2	12

1000 jobs using gLite UI

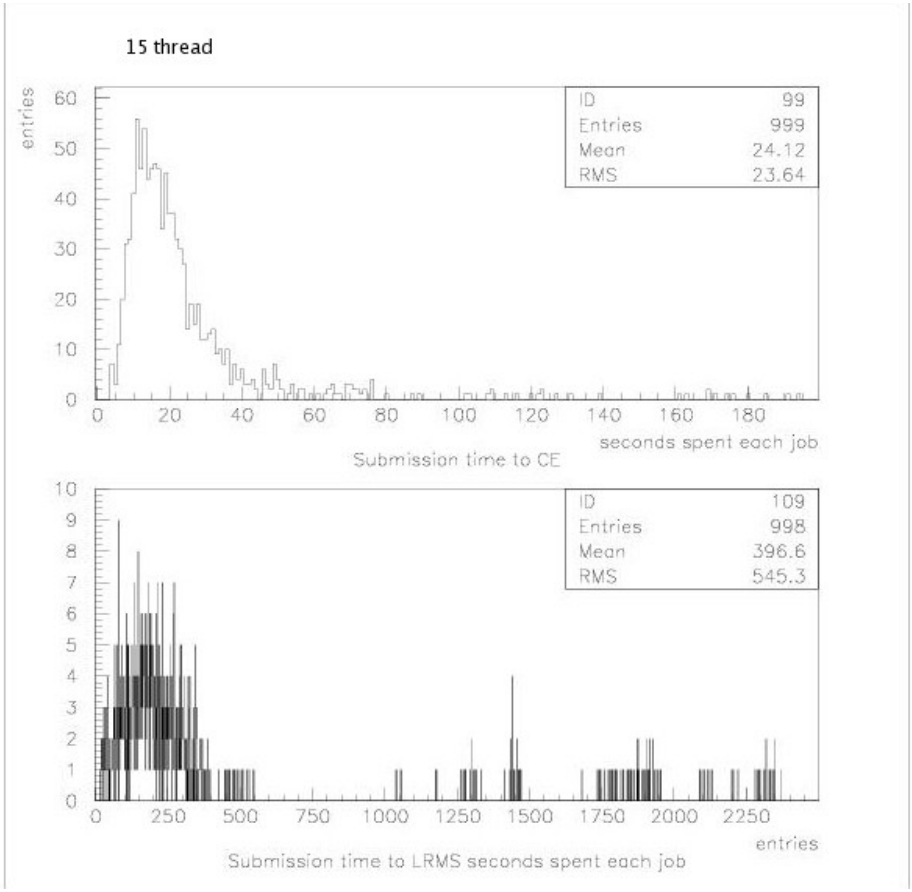
(*) These jobs remain in “Waiting” state, due to a problem in WMproxy-LBproxy interaction; relevant developers have already been informed

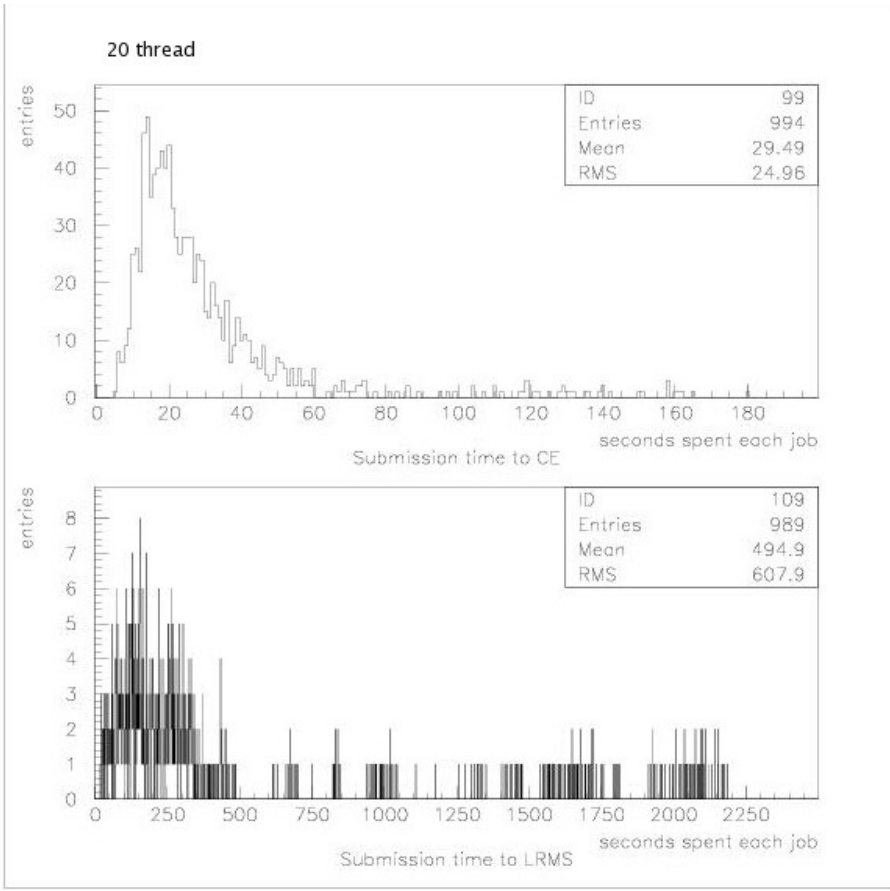
Submission times to CE and to LRMS (ICE scenario)



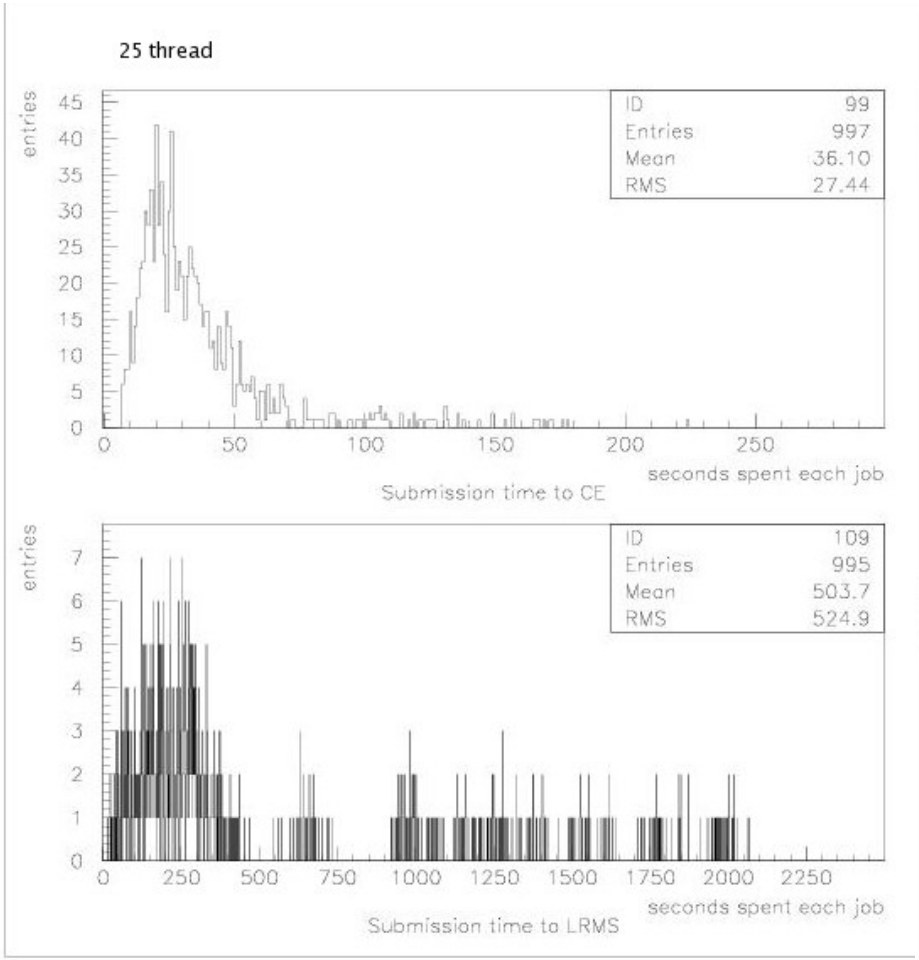


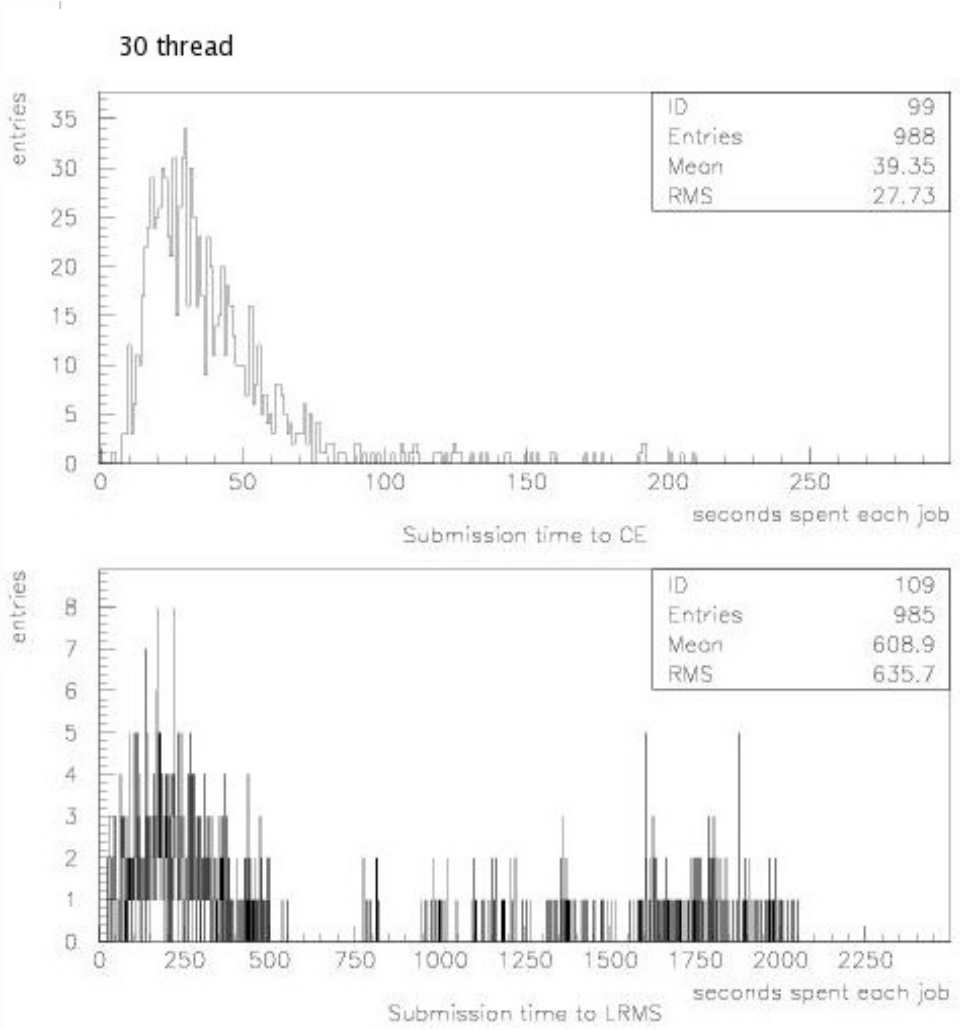
Submission times to CE and to LRMS (ICE scenario)





Submission times to CE and to LRMS (ICE scenario)



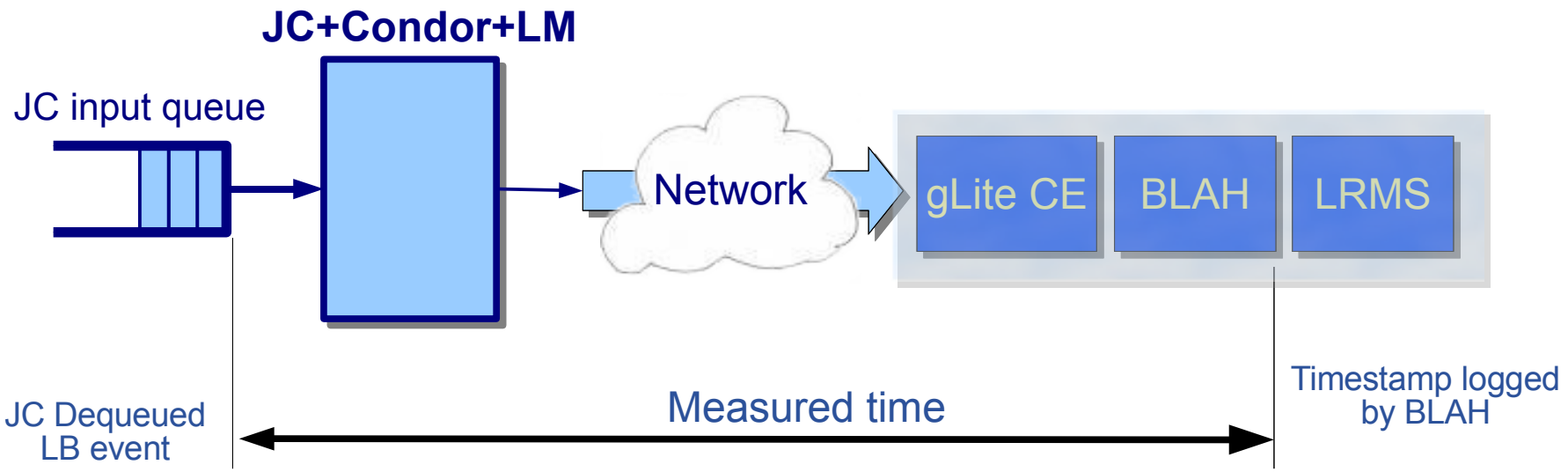


Some conclusions on the submission via WMS-ICE tests

- **Submission of a single job takes some time**
  - ICE performs a new proxy delegation for each new job
  - Delegation consists of a `getProxyRequest` (I/O bound), signing with user proxy, a `putProxy` (I/O bound)
  - ICE calls a `JobRegister` SOAP function (I/O bound)
  - Each relevant operation is logged to LB (I/O bound)
  - Each operation involves authentication, encrypt-decrypt and XML SOAP-based dialog with several services. This takes a while...
- **Some optimization can be performed:**
  - E.g. `putProxy` and `JobRegister` in the same call
  - Code profiling already on-going to identify problems and optimize performance
- **Considering multiple threads which perform in parallel several of these operations increase the overall throughput ...**
- **... but sooner or later CREAM saturates**
- **Please note that in our test we considered a single CREAM CE, but a single ICE istance can of course interact with multiple CEs**
  - A larger number of ICE threads can help when dealing with a larger number of different CREAM CEs
- **Efficiency is not too bad, but can be improved**
  - Basically same failures reasons and same considerations done for the direct job submission tests

Submission through WMS (JC+Condor+LM scenario)

- Submitted 1000 jobs to the WMS with JC switched off
- Restarted JC when the JC input queue was full
- Measured Time: time between the first "dequeued by JC" LB event and the last timestamp logged by BLAH in the accounting log file
- Calculated throughput
- Performed 3 tests



Submission through JC+Condor+LM

- **In the tests considering JC+Condor+LM scenario we observed that at any given time only about 100 jobs were in CE's schedd (and PBS).**
  - Therefore the resulting throughput is pretty low (see next slides)
- **Tried to play (thanks to FrancescoP) with Condor config files (to see if there are some limits defined in Condor configuration)**
- **Contacted Condor developers: they reported it is a bug in Condor**
  - `"I have tracked the problem down to a bug in the gridmanager. It ends up ignoring GRIDMANAGER_MAX_SUBMITTED_JOBS_PER_RESOURCE for condor-c jobs. Unfortunately, there is no workaround. We'll have it fixed for the next release."` reported by Jamie Frey (Condor Team)

JC+Condor+LM scenario test results

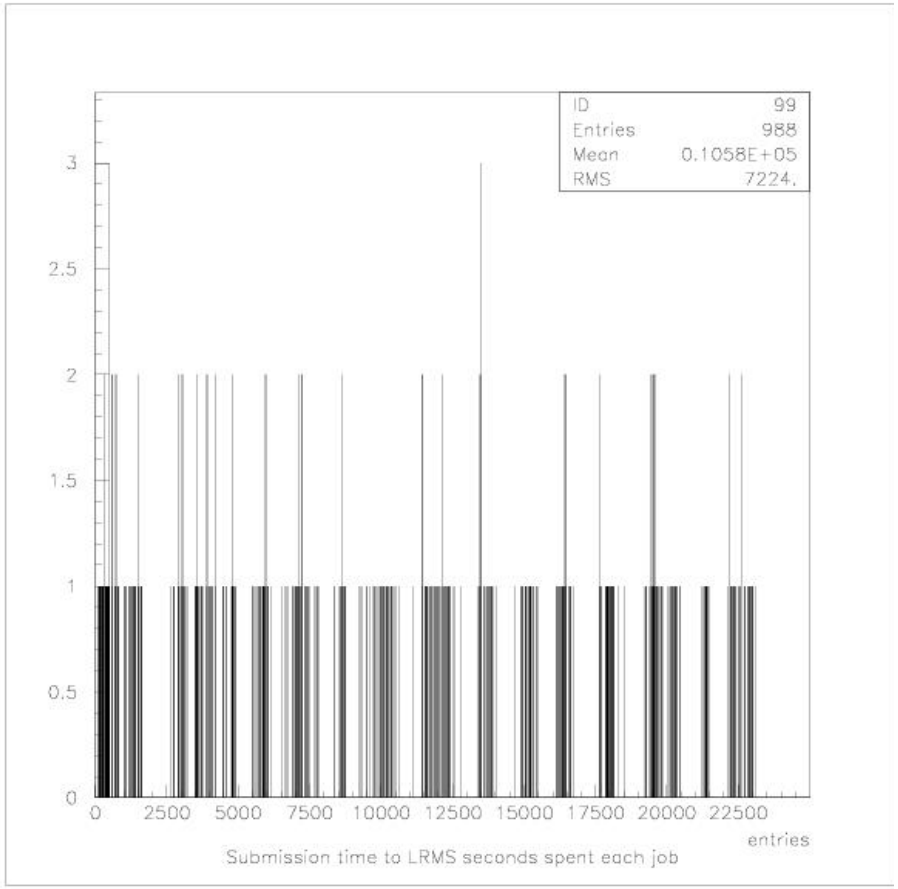
Try No.	Submission rate to LRMS (job/min)	% success (considering all jobs)	% success (considering only jobs managed by JC+Condor+LM)
1	2,4	94,3[1]	94,6
2	2,4	93,9[2]	93,9
3	2,4	96,0[3]	96

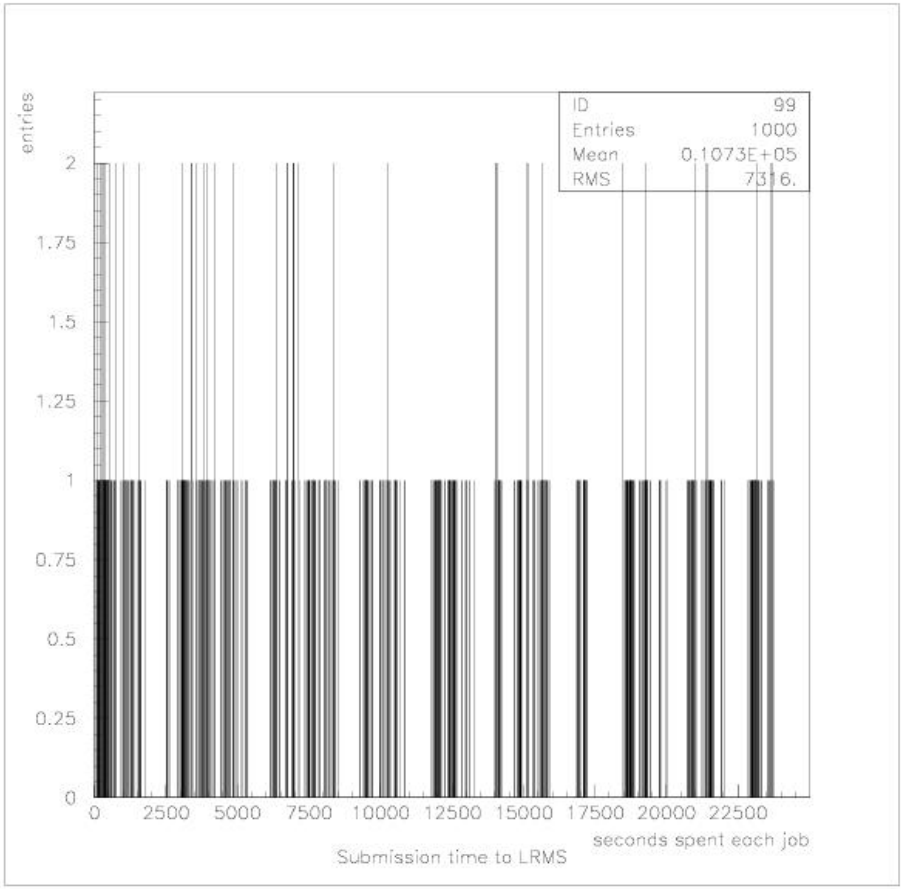
- **([1]) 50 jobs aborted,**
  - 2 Submission to Condor failed
  - 13 File not available. Cannot read JobWrapper output, both from Condor and from Maradona
  - 5 Job got an error while in the CondorG queue
  - 28 Standard output does not contain useful data. Cannot read JobWrapper output, both from Condor, and from Maradona
  - 2 Removal retries exceeded

  7 jobs in waiting

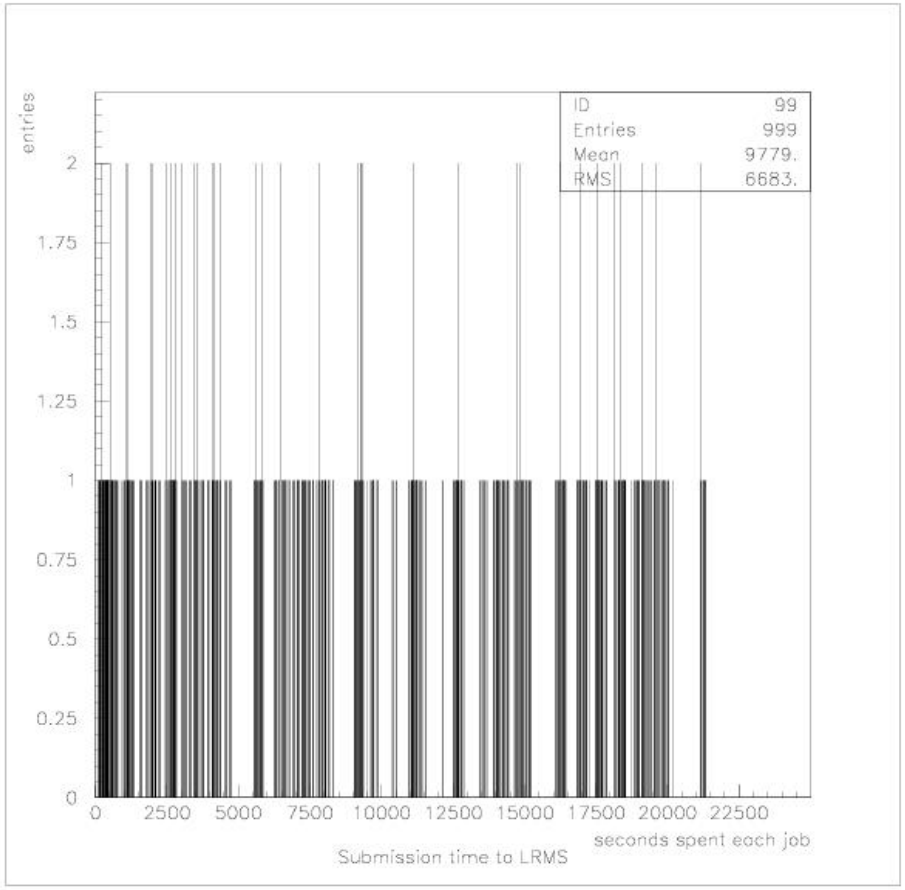
- **([2]) 61 jobs aborted**
  - 21 Jobs got an error while in the CondorG queue
  - 40 Standard output does not contain useful data. Cannot read JobWrapper output, both from Condor and from Maradona
- **([3]) 40 jobs aborted**
  - 1 Submission to Condor failed
  - 25 Standard output does not contain useful data. Cannot read JobWrapper output, both from Condor and from Maradona
  - 14 Jobs got an error while in the CondorG queue

Submission times to LRMS (JC+Condor+LM scenario)





Submission times to LRMS (JC+Condor+LM scenario)



Next steps

- **Continue testing and debugging of ICE and CREAM**
  - perform testing of ICE considering more than a single CREAM CE
    - Another CREAM CE is being installed in Prague for the Preview testbed
- **Continue code profiling to better understand where the bottlenecks are**
- **2 known critical bugs affecting CREAM**
  - **#18244(just fixed)**: there was a bug in VomsServicePDP of gJAF
    - Because of this bug we needed to specify all user DNs in the grid-mapfile
    - developers committed the fix yesterday: to be tested (hopefully today)
  - **#20357**: race condition in BLAH
    - This causes problems in concurrent submissions (in particular when done by different persons)
- **When these 2 bugs get fixed, we should be ready to open the Preview testbed to users interested to test ICE & CREAM**
- **More information: CREAM web site: http://grid.pd.infn.it/cream**