Trace analysis report AuverGrid

This is the trace analysis report (generated by reportgen.py) for the AuverGrid system. The trace data was taken from the filename auvergrid_jobs.gwf, which contains job data obtained from Local resource manager. Below is a summary of the contents of the trace data:

- Date first entry: Sun Jan 01 00:00:24 2006
- CPU time consumed by jobs: 277y 226d 7h 8m 7s
- Number of sites in the system: 5
- Number of CPUs in the trace: 475
- Number of jobs in the trace: 404176
- Number of users in the trace: 405
- Number of groups in the trace: 9

System-wide characteristics

System utilization

We define the overall system utilization as the ratio between the total CPU time consumed by users, and the total CPU time available to the users. We compute the total CPU time consumed by users as the sum of CPU time consumed by each job in the system; for failed jobs, only those that have effectively spent resource time are considered. We compute the total CPU time available as the number of CPUs multiplied by the duration of a fixed time interval, c.q. 10 minutes.

Below we show the statistical properties of both the overall system utilization and the overall system for non-zero values, that is, excluding all intervals that have system utilization equal to zero. This excludes values that may account for downtime of the system.

Figure 1 shows System utilization over time.

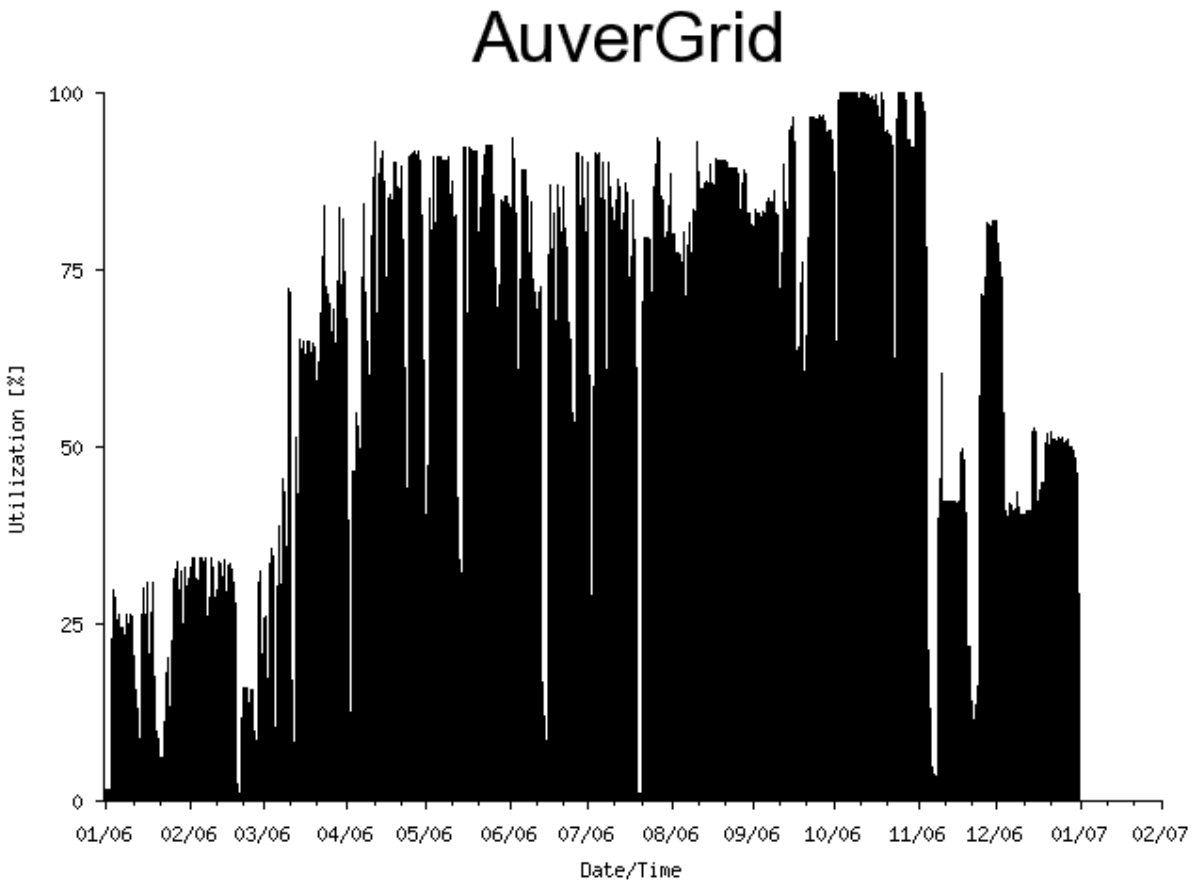

Figure 1: System utilization over time

Overall system utilization

- Minimum: 0.0 percent
- Maximum: 100.0 percent
- Average: 58.481 percent

Overall system utilization for non-zero values

- Minimum: 0.139 percent
- Maximum: 100.0 percent
- Average: 58.595 percent

Job arrival rate

We define the job arrival rate as the number of jobs that are submitted to the system in a fixed time interval. We compute the arrival rate for every hour by counting the all jobs that are recorded in the trace during that hour. This includes failed jobs and jobs that are cancelled before execution. Below we list the time periods in which the highest number of jobs were submitted to the system. We also summarize statistical properties for all job arrival rate values, and the statistical properties for arrival rate higher than zero. This excludes time periods that may account to downtime of the system.

Figure 2 shows Overall job arrival rate during hourly intervals.

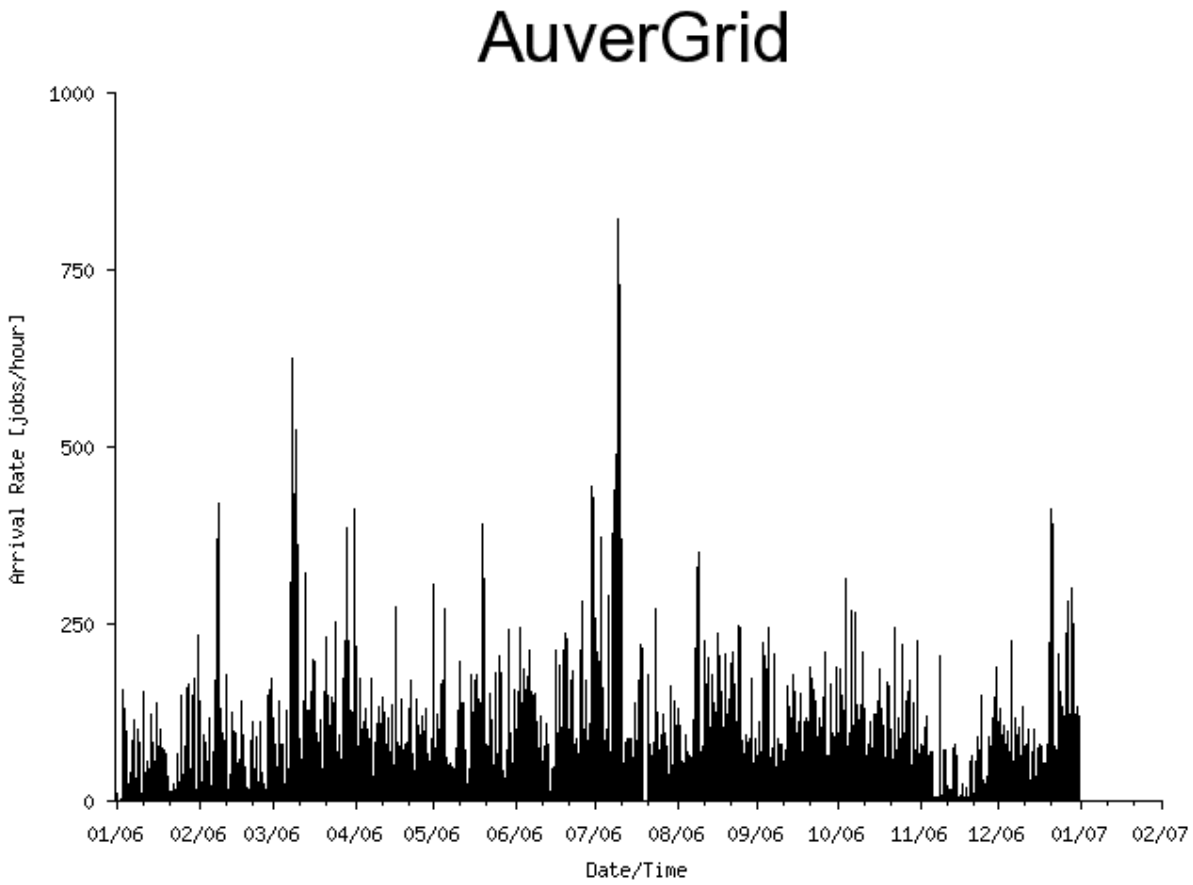

Figure 2: Overall job arrival rate during hourly intervals

Busiest time periods in terms of number of job submissions

- Busiest day: 2006-07-10
- Busiest week: 2006-27
- Busiest month: 2006-07

Overall job arrival metrics

- Minimum: 0.00 jobs/hour
- Maximum: 823.00 jobs/hour
- Average: 46.13 jobs/hour

Overall job arrival metrics for non-zero values

- Minimum: 2.00 jobs/hour
- Maximum: 823.00 jobs/hour
- Average: 48.31 jobs/hour

Job characteristics

We compute three important characteristics of jobs in the trace: number of CPUs used, the runtime of the job and the amount of memory used. Below we summarize the statistical properties for single jobs in the trace. We do not include jobs that were cancelled before execution, because those jobs did not consume resources from the system.

Figure 3 shows CDFs of the most important job characteristics.

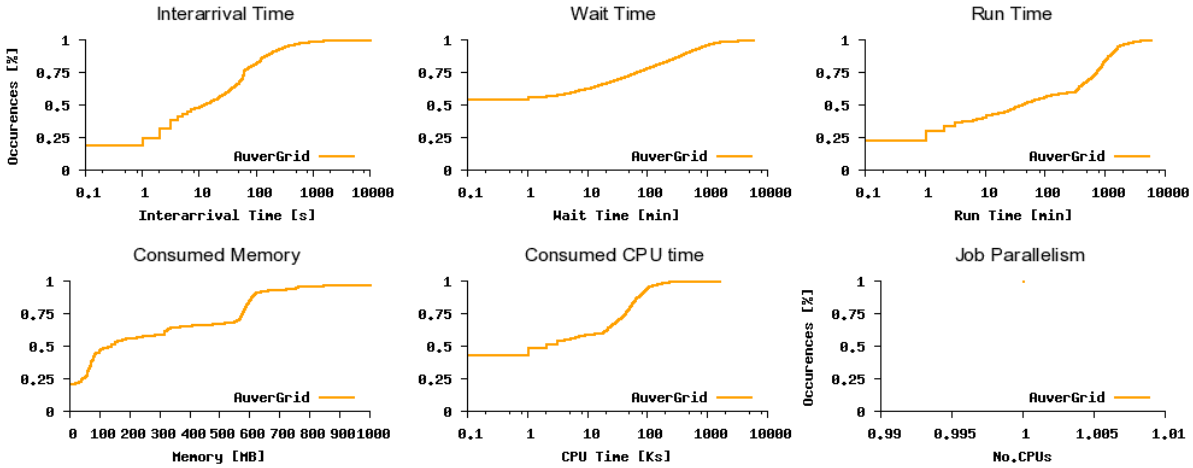

Figure 3: CDFs of the most important job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 1 processors
- Average: 1.000 processors
- Standard deviation: 0.000
- Coefficient of variation: 0.000

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 1575814.00 seconds
- Average: 25186.27 seconds
- Standard deviation: 40780.303
- Coefficient of variation: 1.619

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 3667.65 MB
- Average: 295.58 MB
- Standard deviation: 342.991
- Coefficient of variation: 1.160

Sequential vs. Parallel jobs

Below we summarize the resource usage of all sequential and all parallel jobs, that is all jobs that use more than one processor. First we calculate the number of sequential jobs and the number of parallel jobs that are submitted to the system. Furthermore, we compute the consumed CPU time by multiplying the runtime of a job by the number of processors allocated to the job. Again, this is divided into parallel and sequential jobs. For the number of jobs and the consumed CPU time, the percentage of all jobs is displayed.

Number of jobs

- Sequential: 347611 jobs (86.00 percent)
- Parallel: 0 jobs (0.00 percent)

Consumed CPU Time

- Sequential: 8755024087 seconds (100.00 percent)
- Parallel: 0 seconds (0.00 percent)

User and group characteristics

User characteristics

Figure 4 shows The number of submitted jobs and the consumed CPU time by user.

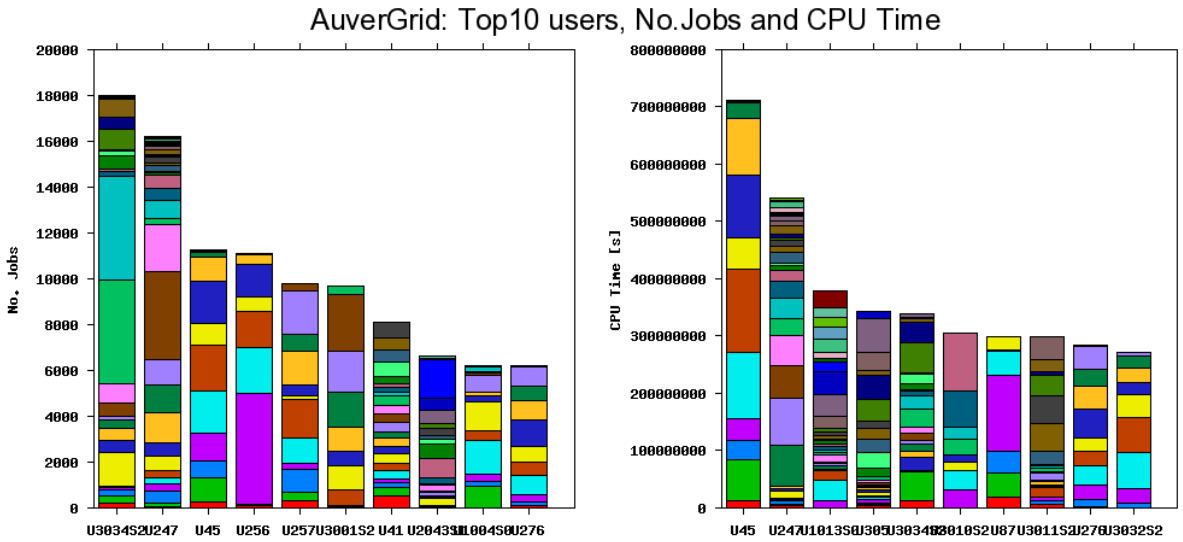

Figure 4: The number of submitted jobs (left) and consumed CPU time (right) by user. Only the top 10 users are displayed. The horizontal axis depicts the user's rank. The vertical axis shows the cumulated values, and the breakdown per week. Users have the same labels in the left and right sub-graphs

Top 10 users by number of job submitted to the system

Table 1 shows Top 10 users by number of jobs submitted to the system.

Table 1

Rank	UserID	Number of jobs	Percentage
1	U3034S2	18021	4.46%
2	U247	16218	4.01%
3	U45	11259	2.79%
4	U256	11083	2.74%
5	U257	9781	2.42%
6	U3001S2	9663	2.39%
7	U41	8082	2.00%
8	U2043S1	6619	1.64%
9	U1004S0	6220	1.54%
10	U276	6200	1.53%

Table 1			
Rank	UserID	Number of jobs	Percentage
11	Other	301030	74.48%
12	Total	404176	100.00%

System utilization

- Minimum: 0.0 percent
- Maximum: 88.744 percent
- Average: 18.186 percent

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 399.00 jobs/hour
- Average: 11.86 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 1 processors
- Average: 1.000 processors
- Standard deviation: 0.000
- Coefficient of variation: 0.000

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 504299.00 seconds
- Average: 26204.47 seconds
- Standard deviation: 38996.798
- Coefficient of variation: 1.488

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 3667.65 MB
- Average: 316.87 MB
- Standard deviation: 426.304
- Coefficient of variation: 1.345

Top 10 users by consumed CPU time

Table 2 shows Top 10 users by consumed CPU time (in seconds).

Table 2			
Rank	UserID	CPU seconds	Percentage
1	U45	711866619	8.13%
2	U247	541985743	6.19%
3	U1013S0	378342891	4.32%
4	U305	342114526	3.91%
5	U3034S2	339168487	3.87%
6	U3010S2	304774893	3.48%

Table 2			
Rank	UserID	CPU seconds	Percentage
7	U87	298862747	3.41%
8	U3011S2	298685551	3.41%
9	U276	283279419	3.24%
10	U3032S2	272064853	3.11%
11	Other	4983878358	56.93%
12	Total	8755024087	100.00%

System utilization

- Minimum: 0.0 percent
- Maximum: 88.361 percent
- Average: 25.366 percent

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 399.00 jobs/hour
- Average: 9.51 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 1 processors
- Average: 1.000 processors
- Standard deviation: 0.000
- Coefficient of variation: 0.000

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 1535085.00 seconds
- Average: 45594.80 seconds
- Standard deviation: 49758.277
- Coefficient of variation: 1.091

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 2123.31 MB
- Average: 291.12 MB
- Standard deviation: 267.298
- Coefficient of variation: 0.918

Group characteristics

Figure 5 shows The number of submitted jobs and consumed CPU time by group.

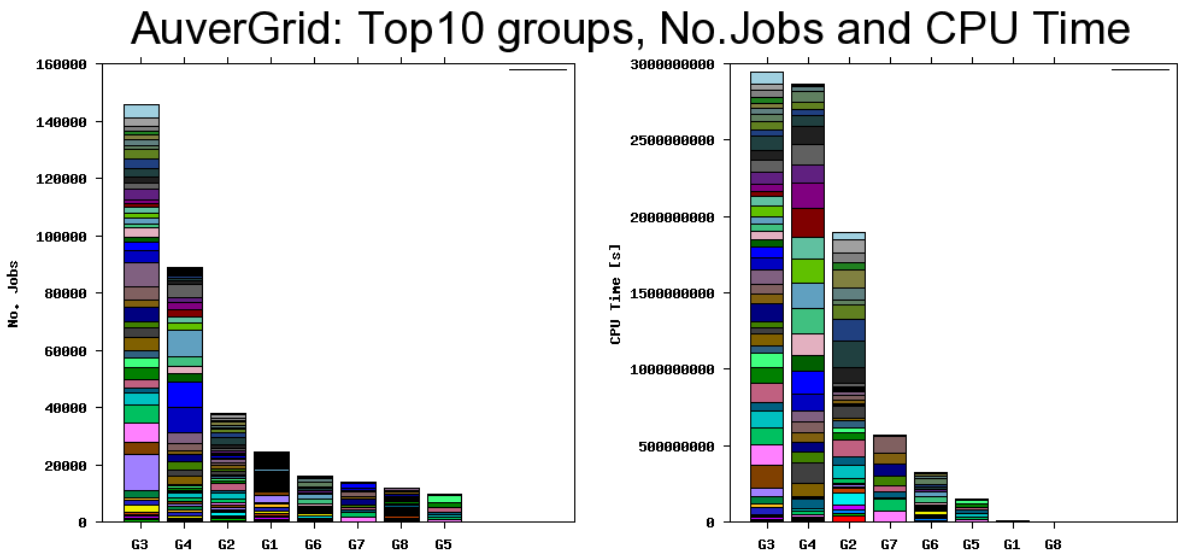

Figure 5: The number of submitted jobs (left) and consumed CPU time (right) by group. Only the top 10 groups are displayed. The horizontal axis depicts the groups rank. The vertical axis shows the cumulated values, and the breakdown per week. Groups have the same labels in the left and right sub-graphs

Table 3 shows Top 10 groups by number of jobs submitted to the system.

Table 3

Rank	GroupID	Number of jobs	Percentage
1	G3	145508	36.00%
2	G4	88681	21.94%
3	G2	37792	9.35%
4	G1	24311	6.01%
5	G6	15924	3.94%
6	G7	13790	3.41%
7	G8	11903	2.95%
8	G5	9702	2.40%
9	Other	56565	14.00%
10	Total	404176	100.00%

Table 4 shows Top 10 Groups by consumed CPU time (in seconds).

Table 4

Rank	GroupID	CPU seconds	Percentage
1	G3	2942799943	33.61%
2	G4	2866445504	32.74%
3	G2	1894964091	21.64%
4	G7	566862443	6.47%

Table 4			
Rank	**GroupID**	**CPU seconds**	**Percentage**
5	G6	322790712	3.69%
6	G5	151943132	1.74%
7	G1	6955681	0.08%
8	G8	2262581	0.03%
9	Other	0	0.00%
10	Total	8755024087	100.00%

Performance analysis

Waiting and running jobs

Figure 6 shows The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility.

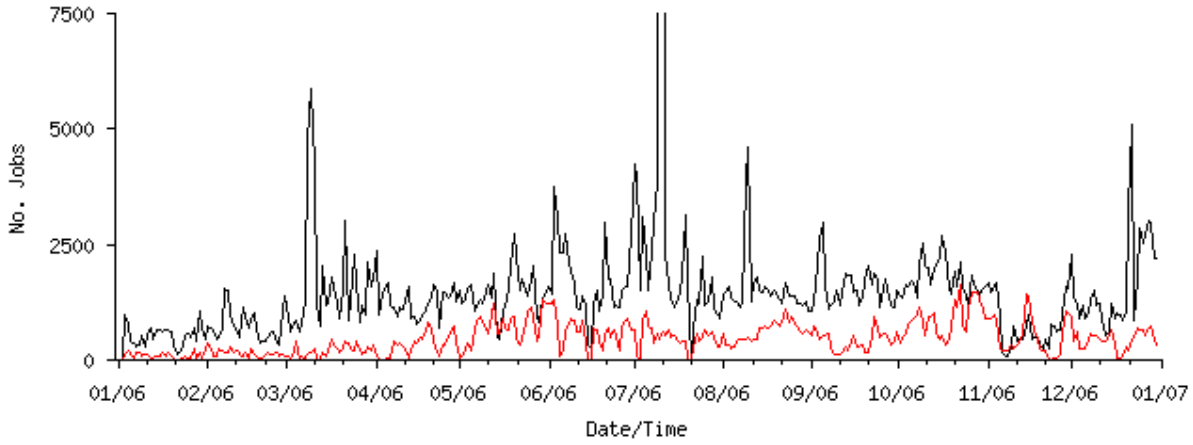

Figure 6: The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility

We compute the number of running and waiting jobs by considering a fixed time interval. In each time interval, we count in the trace the amount of jobs that have been submitted but not yet started, that is, waiting. We also count the number of jobs that have been submitted, and have started executing in the time interval, but did not finish executing, and thus are running. Below we show the values for an interval value of 3600 seconds, summarized in amounts per day. Also the summary for values higher than zero are displayed, which excludes the possible effect of downtime of the system.

Number of waiting jobs per day

- Minimum: 0 jobs
- Maximum: 1612 jobs
- Average: 451.41 jobs

Number of waiting jobs per day (non-zero values)

- Minimum: 1 jobs
- Maximum: 1612 jobs
- Average: 453.89 jobs

Number of running jobs per day

- Minimum: 3 jobs
- Maximum: 13325 jobs
- Average: 1431.89 jobs

Number of running jobs per day (non-zero values)

- Minimum: 3 jobs
- Maximum: 13325 jobs
- Average: 1431.89 jobs

Throughput

We compute the job throughput by considering a fixed time interval. In each time interval, we count in the trace the amount of jobs that have been submitted, started and finished executing. Below we show the values for an interval value of 3600 seconds, summarized in amounts per day. Also the summary for values higher than zero are displayed, which excludes the possible effect of downtime of the system.

Figure 7 shows Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility.

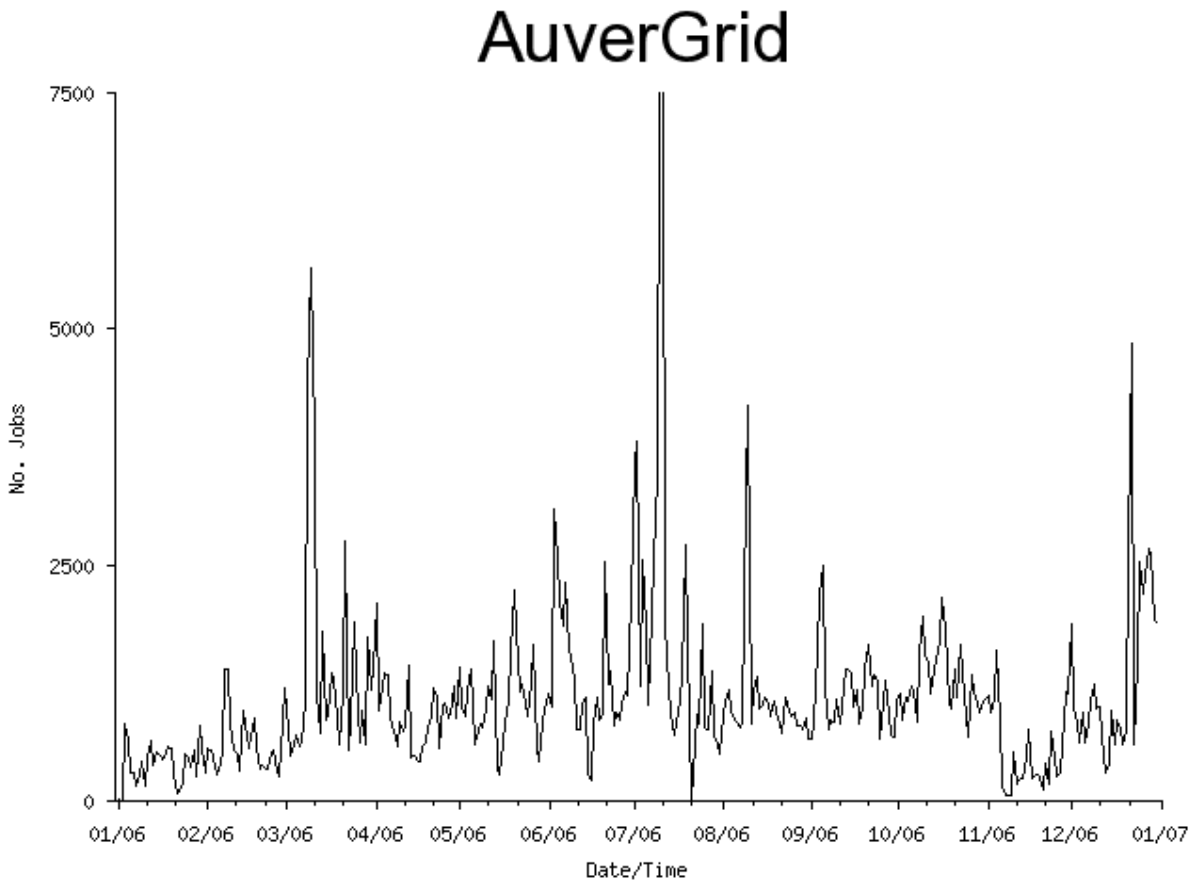

Figure 7: Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility

Throughput per day

- Minimum: 0 jobs
- Maximum: 12594 jobs
- Average: 1102.30 jobs

Throughput per day (non-zero values)

- Minimum: 10 jobs
- Maximum: 12594 jobs
- Average: 1111.43 jobs

Completed jobs

Figure 8 shows the number of completed jobs during hourly intervals.

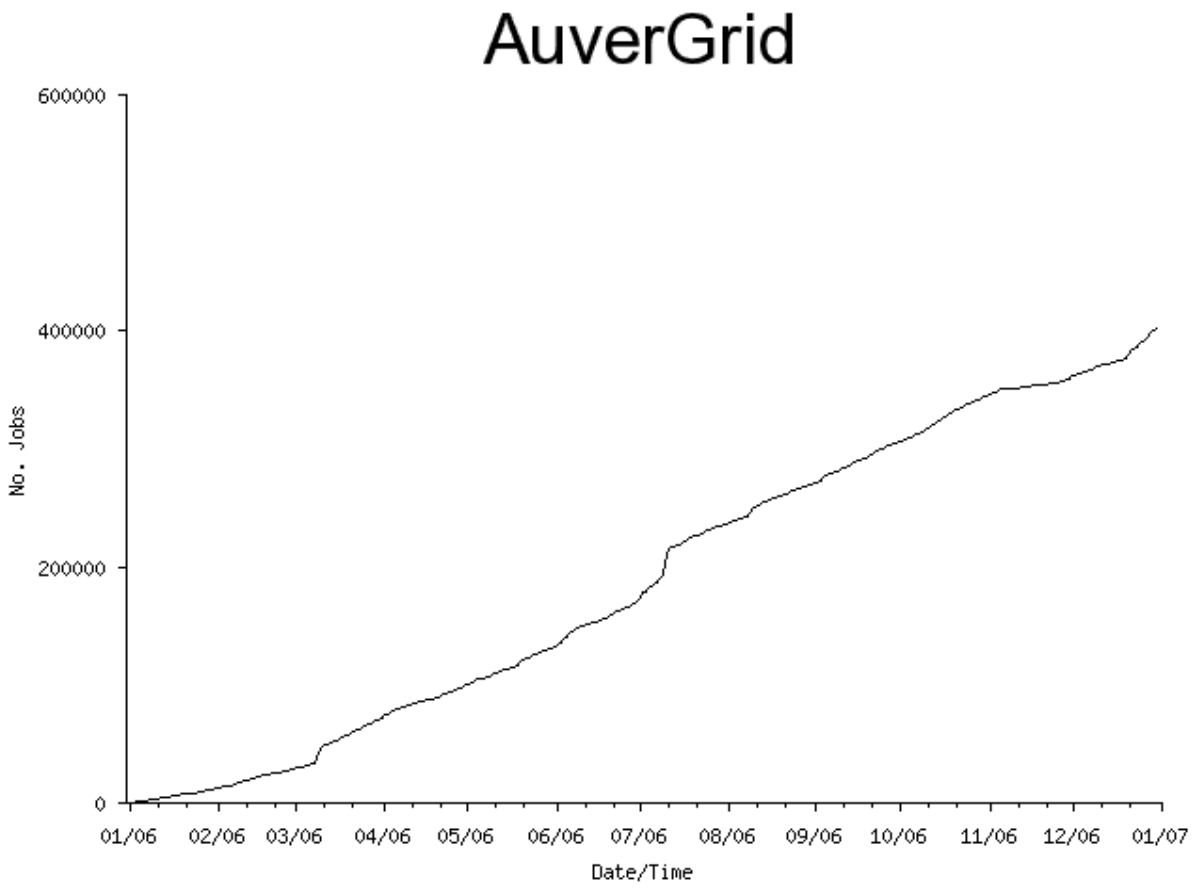

Figure 8: The number of completed jobs during hourly intervals

Workload model

This section contains the workload model for the analyzed trace. The workload model consists of several parameters: job size, job runtime, requested runtime and interrarivals of jobs. These parameters are modeled by fitting well-known distributions to the data obtained from the trace. In all cases, first a logarithmic transformation was performed on the dataset to diminish the effect of outliers and speed up the modelling process. The fitting was performed using the maximum likelihood estimation method, which tries to maximize the log-likelihood function of each distribution given a dataset.

Job runtime

Figure 10 shows Cumulative distribution function for the logarithm of the job runtimes, with fitted distributions.

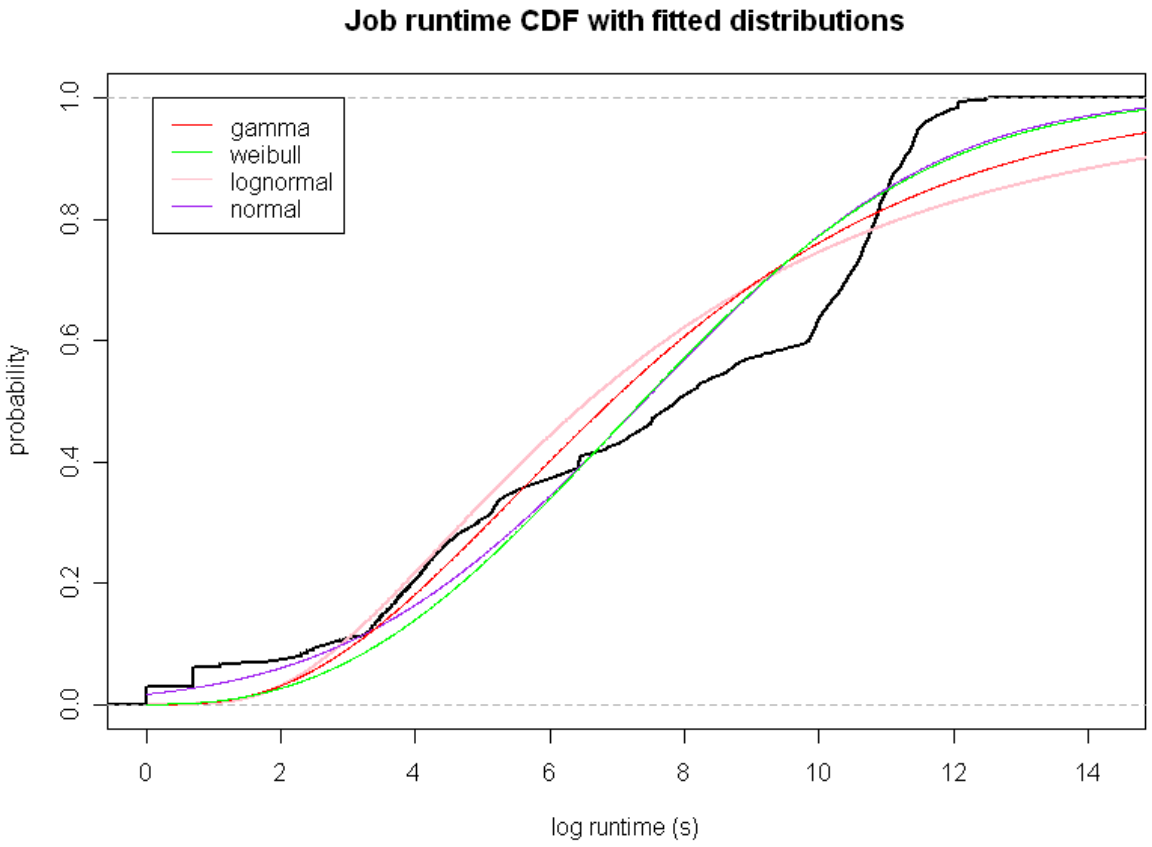

Figure 10: Cumulative distribution function for the logarithm of the job runtimes, with fitted distributions

Job requested runtime

Figure 11 shows Cumulative distribution function for the logarithm of the job requested runtimes, with fitted distributions.

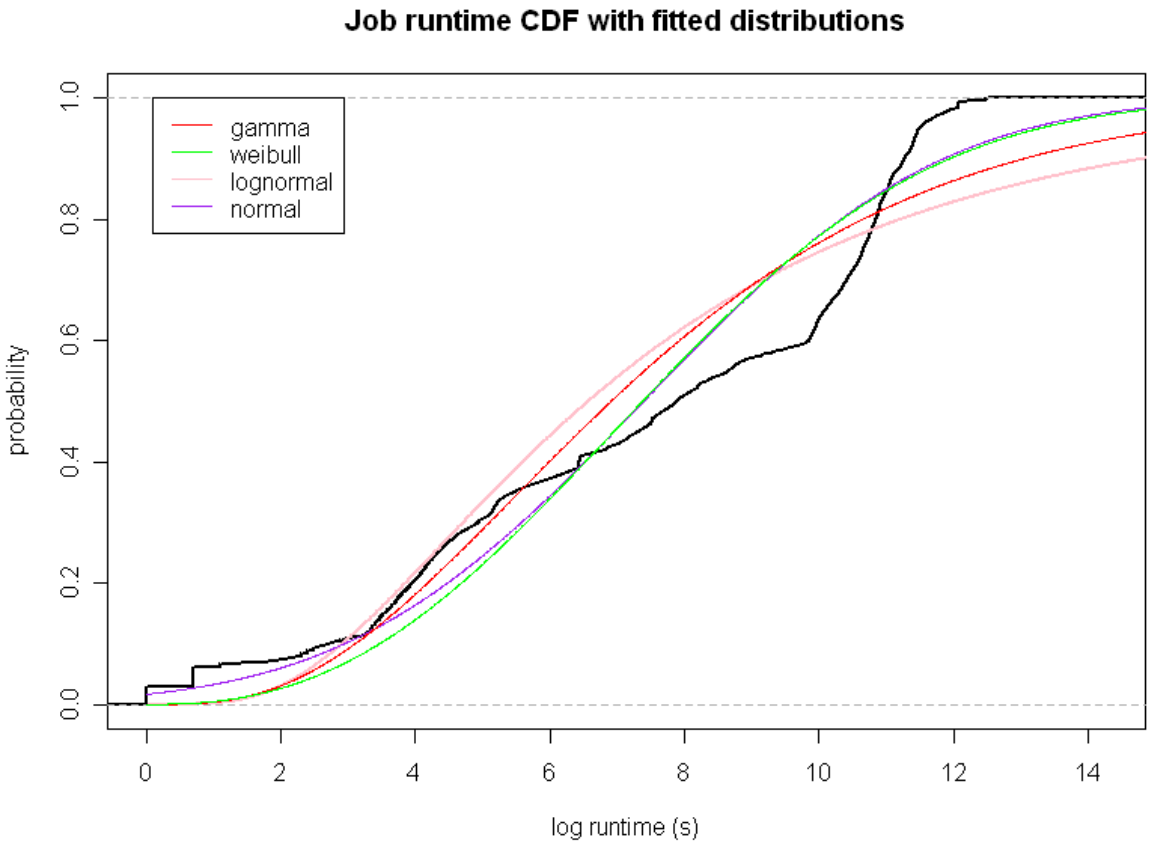

Figure 11: Cumulative distribution function for the logarithm of the job requested runtimes, with fitted distributions

Parameters of fitted distributions

Lognormal distribution
- meanlog: 2.87663735104998
- sdlog: 0.0786064065578723

Exponential distribution
- rate: 0.0561740604932808

Normal distribution
- sd: 1.12533696177834
- mean: 17.8018108575152

Weibull distribution
- shape: 45.2512565165272
- scale: 18.0698251996298

Hyperexponential distribution
- p: 1.11603020791087
- rate2: 0.0461080021132448
- rate1: 0.132681757003409

Gamma distribution
- shape: 188.121298078541
- rate: 10.5675341197646

Goodness-of-fit (Kolmogorov-Smirnov test)

Table 7 shows for each distribution the results for the Kolmogorov-Smirnov test, which gives a measure for the distance of the distribution to the original dataset (lower distance => better fit).

Table 7	
Distribution	**Distance**
Lognormal	0.471175710123947
Exponential	0.570398538763021
Normal	0.470558968708816
Weibull	0.459531505712732
Hyperexponential	0.499039770441242
Gamma	0.471470468074848

Job interarrival

Figure 12 shows Cumulative distribution function for the logarithm of the job interarrival, with fitted distributions.

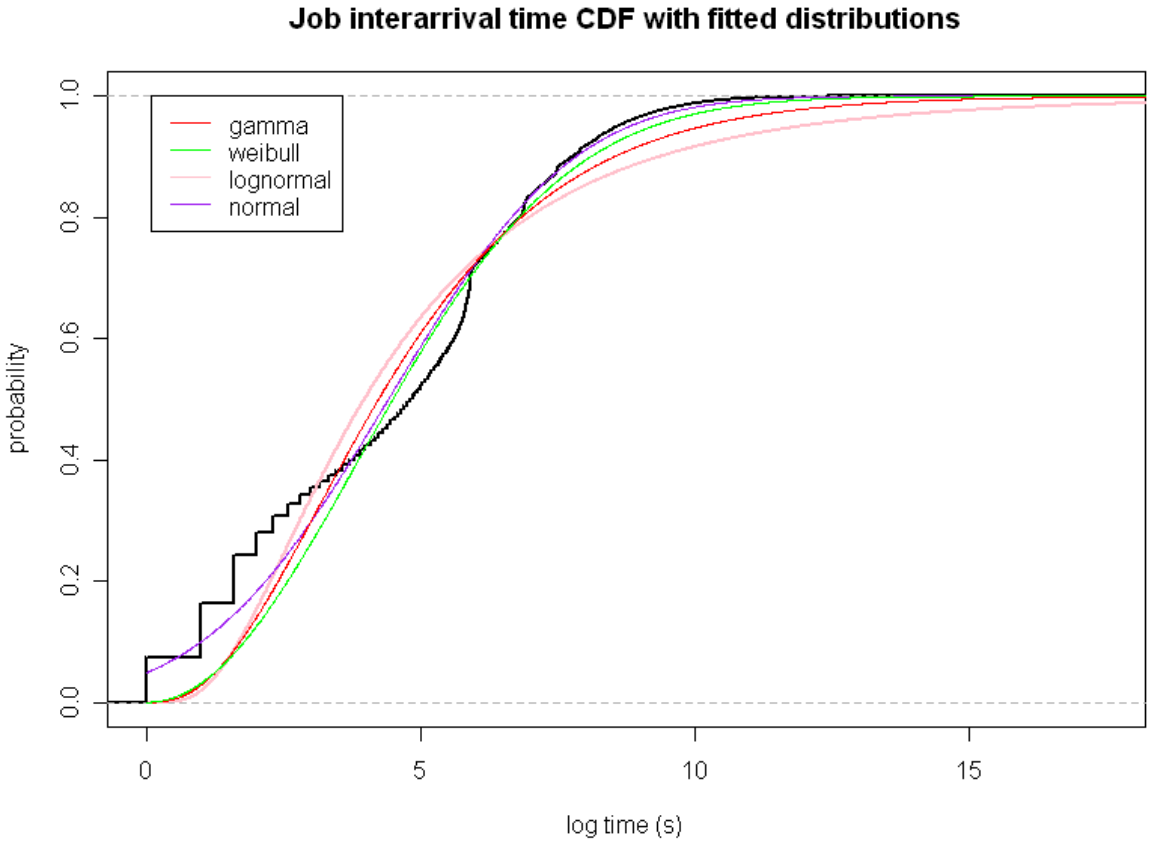

Figure 12: Cumulative distribution function for the logarithm of the job interarrival, with fitted distributions