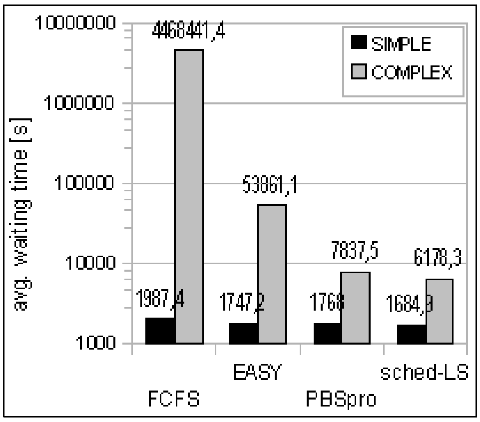
Complex real-life data sets in Grid simulations

Dalibor Klusáček, Hana Rudová
Faculty of Informatics, Masaryk University, Brno, Czech Republic

Widely used workload sources such as Parallel Workloads Archive (PWA) [4] or Grid Workloads Archive (GWA) [5] often provide data sets that are still unrealistic. Typically, very limited information is available about the grid/cluster parameters such as architecture, speed, RAM size or resource specific policies. Moreover, no information concerning background load, resource failures, or specific user's requests are available. In heterogeneous environments, users often specify a small subset of machines that are suitable to perform the jobs. This subset is usually defined either by resource owner's policy (user is allowed to use such cluster) or by user's requirements (user specifies some property offered by some clusters only) or by both factors. When one tries to create a good scheduling algorithm and compare it with current approaches such as PBSpro [3], all such information and constraints are crucial, since they make the algorithm design much more complex.

So far, we have been able to collect complex real-life data set from the Czech national Grid infrastructure MetaCentrum [6] that covers many previously mentioned issues, representing 5 months of execution, involving 103 620 jobs completed on 14 heterogeneous clusters having 806 CPUs. This data set includes exact machine parameters involving both hardware setup (speed, architecture, RAM size) and supported properties (e.g., hardware/institution/queue based restrictions). For each job, a queue where the job was originally submitted is known, as well as the maximal runtime after which the job would be killed. Also, the numbers, types, time limits and priorities of all queues are known. Moreover, machine failures and restarts are known together with the list of temporary dedicated, thus unavailable, machines. In contrast to the GWA or PWA sources, our data set allows to perform far more realistic simulations involving all precedent features.

We have studied behavior of several objective functions in our solution that cover typical requirements such as average job slowdown, response time, waiting time and the algorithm's runtime. We use schedule-based algorithms involving Local Search (sched-LS) which we have been developing for couple of years [1], as well as widely used queue-based solutions such as FCFS, EASY Backfilling [2] or PBSpro algorithms [3] representing multiple-queues priority based backfilling. In our experiments, we have focused on two scenarios. SIMPLE scenario does not simulate dedicated resources or failures. Moreover, all jobs can be executed on any cluster (if enough CPUs are available), thus SIMPLE represents the typical amount of information available in the GWA or PWA data sets. On the other hand, COMPLEX scenario uses every additional information available in our data set such as queue priorities, machine failures, as well as additional machine and job properties that define the job-to-machine suitability. As observed during the experiments, the difference between SIMPLE and COMPLEX setup is dramatic as shown in the figure depicting the average waiting time (in log. scale). In case of SIMPLE, the differences between algorithms are quite small while the COMPLEX data set introduces huge differences among algorithms, causing that previously acceptable FCFS or EASY backfilling now degrade dramatically. Similar behavior was visible for all above mentioned objective functions.



It is clear that complex and "rich" data set influences the algorithms' performance and causes significant differences in the values of objective functions. We suggest that—beside the PWA and the GWA—complex data sets should be also used to evaluate existing and newly proposed algorithms under harder conditions. Therefore, at the time of the conference our data set will be publicly available for further open research.

Acknowledgements. We appreciate the support of the Ministry of Education, Youth and Sports of the Czech Republic under the research intent No. 0021622419 and the support of the Grant Agency of the Czech Republic with grant No. 201/07/0205.

1. Klusáček, D.; Rudová, H.; Baraglia, R.; Pasquali, M.; Capannini, G. *Comparison of Multi-Criteria Scheduling Techniques.* In Grid Computing Achievements and Prospects, Springer, 2008.
2. Skovira, J.; Chan, W.; Zhou, H.; and Lifka, D. A. *The EASY - LoadLeveler API Project*. In IPPS '96: Proceedings of the JSSPP Workshop. Springer. 1996.
3. Jones, J. P. 2005. *PBS Professional 7, administrator guide*. Altair. 2005.
4. Parallel Workloads Archive: http://www.cs.huji.ac.il/labs/parallel/workload/
5. Grid Workloads Archive: http://gwa.ewi.tudelft.nl
6. MetaCentrum: http://meta.cesnet.cz