Complex real-life data sets in Grid simulations

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

{xklusac,hanka}@fi.muni.cz

Cracow Grid Workshop 2009, October 12 - 14, 2009

Introduction

- Both production or experimental scheduling algorithms have to be heavily tested
- Usually, through a simulation using synthetic or real-life workloads
- Reviewers often don't like "home-made" synthetic workloads
  - Workload fits to the proposed solution
- Popular real-life based workloads
  - Parallel Workloads Archive (PWA)
  - Grid Workloads Archive (GWA)

PWA and GWA workloads (1)

- Provide variety of different workloads
- Workload typically contains
  - Job_id
  - Submission time
  - Execution start time
  - Completion time
  - # of requested CPUs
  - Queue_id
  - … and more (differs according to the used workload)

PWA and GWA workloads (2)

- But for proper simulation, information about machines are needed too
- Often tricky to get for both PWA and GWA workloads
- Example: original SHARCNET workload file

```
# Generated by get-clean-logs.py ($Revision: 0.1$) on 2007-05-31 13:15:35.343000
# ------------------------------------------------------
#
# System name:    SharcNet
# System info:      ?
# Sites:  10
# Processors:       6828
# CPU Info:   ?
# Memory:     ?
# Disk space:?
# Network:    ?
# Log source:?
#
#------------------------------------------------------
# Format documentation: Grid Workload Format (http://gwa.ewi.tudelft.nl/)
```

PWA and GWA workloads (3)

- Some machine parameters may be recovered from different sources
- Example: SHARCNET website contains quite reasonable machine description
- Sadly, this is not the case for many workloads
- **Question**: If someone uses a GWA/PWA workload, what machine parameters does he or she use? Random? Identical? Intuitive? Benchmark-based? Is it discussed in his or hers paper?
- To reproduce someone's result, jobs' description is not enough

Other problems

- Changing state of the system, additional constraints
  - Maintenance (failures/upgrades)
  - Dedicated machines, background load
  - Additional jobs' requirements (required machine properties)
- Without such knowledge, resulting simulation may be very far from the original execution
- Example: NorduGrid workload in the GWA

> ...the grid/non-grid job ratio is around 20 percent, the majority of running jobs being local jobs. Locally submitted jobs do not appear in the NorduGrid traces present in the GWA...

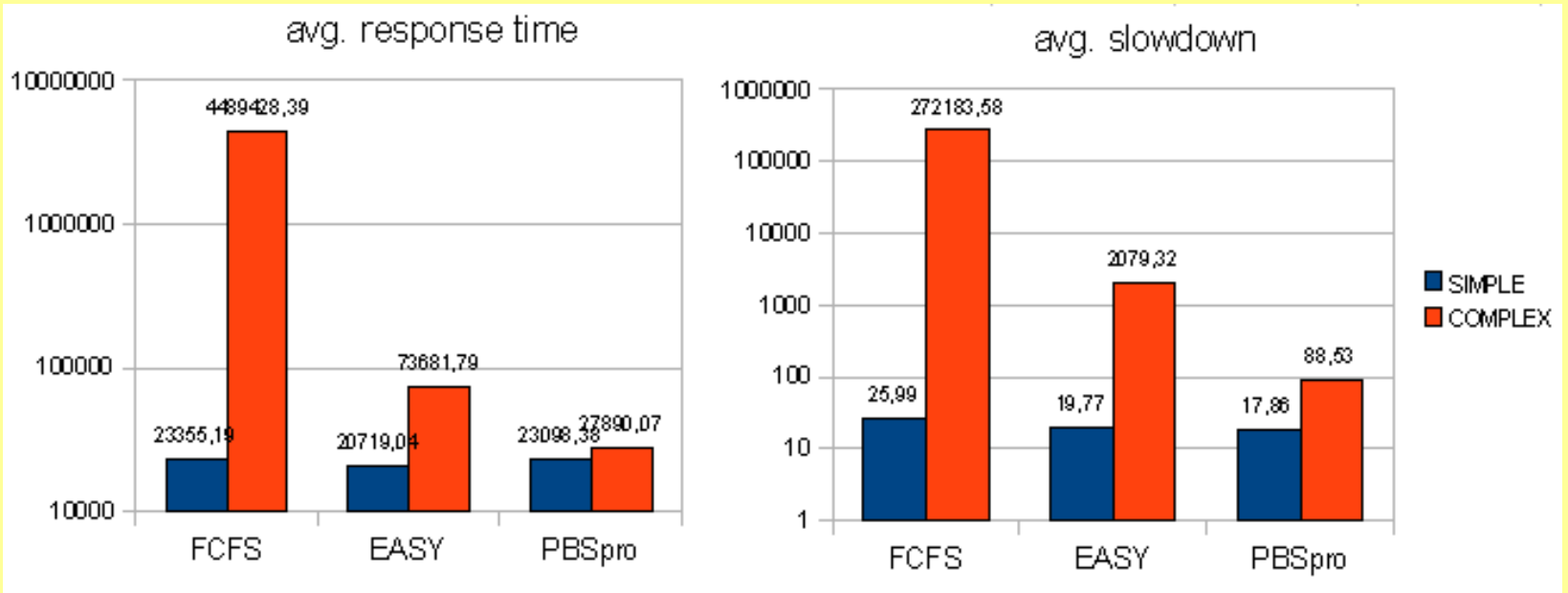
Complex workload from MetaCentrum

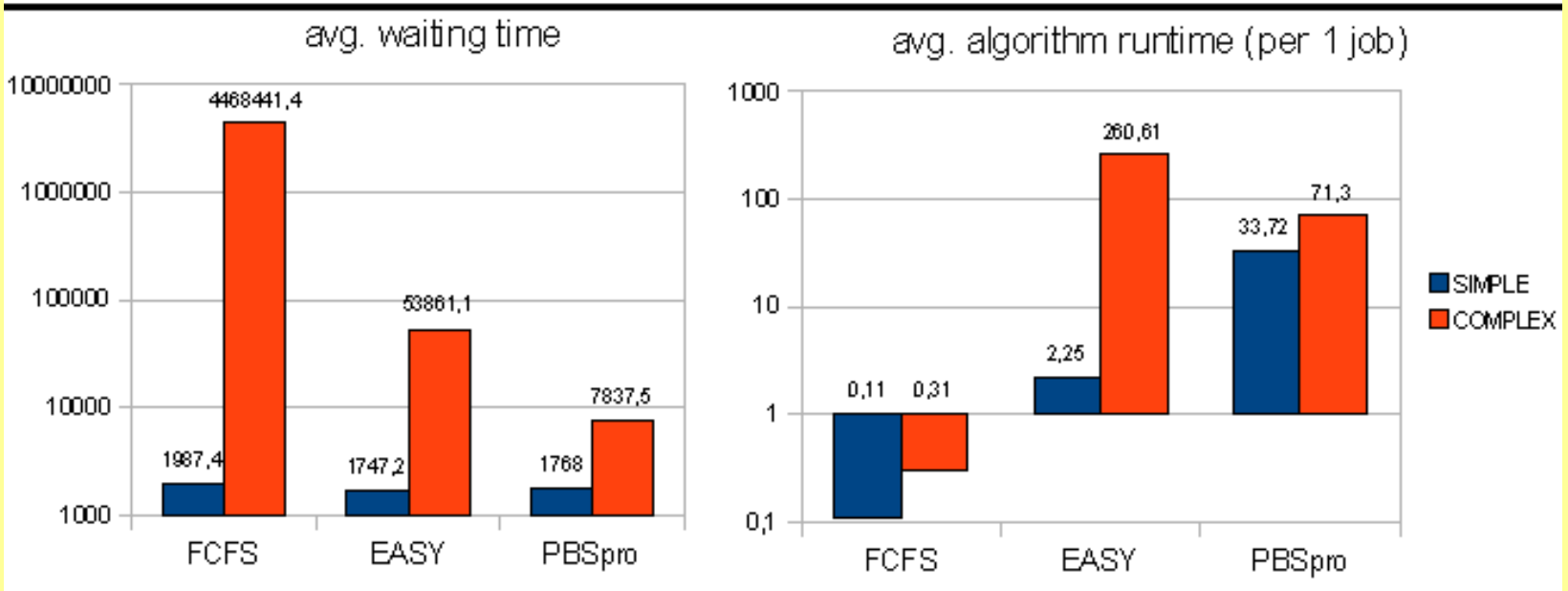
- MetaCentrum is the Czech national Grid infrastructure
- We were able to collect complex data set
  - 5 months of execution during January – May 2009
  - 103 656 jobs, 14 clusters (806 CPUs)
  - Job having different requirements concerning target machine
  - Machines having different parameters (job-to-machine suitability)
  - Queues + priorities + time limits
  - Descriptions of machines in maintenance (failures & upgrades)
  - Descriptions of dedicated and reserved machines
  - No ignored background load
  - Clusters' benchmark results included (SPEC CPU2006)

Evaluation through Alea Scheduling Simulator

- 3 simulated algorithms (FCFS,EASY Backfilling and PBSpro like algorithm). Two experimental setups:
  - **SIMPLE** (GWA/PWA like simulation)
    - No additional job requirements, only #CPUs is used
    - No maintenance, failures, dedicated machines
  - **COMPLEX**
    - Job requirements supported
    - Maintenance and dedicated machines simulated
- **Question**: Do the additional information and constraints influence solution and algorithms' behavior?
- **Answer**: Yes, dramatically...

Evaluation – graphs





Conclusion & Future Work

- Complex and "rich" workload influenced algorithms' performance significantly
- Beside PWA/GWA, also complex workloads should be used to evaluate algorithms under harder conditions
- Based on the knowledge of the complex workload, we plan to add additional synthetically generated parameters to the PWA/GWA workloads and evaluate it experimentally w.r.t. the original PWA/GWA traces
- Our workload is freely available for further open research:

  http://www.fi.muni.cz/~xklusac/workload