

Pawel Czarnul, 2014. Comparison of selected algorithms for scheduling workflow applications with dynamically changing service availability. *Journal of Zhejiang University-SCIENCE C (Computers & Electronics)*, **15**(6):401-422.

[doi:[10.1631/jzus.C1300270](https://doi.org/10.1631/jzus.C1300270)]

Comparison of selected algorithms for scheduling workflow applications with dynamically changing service availability

Key words: Scheduling of workflow applications, Workflow management environment, Scheduling algorithms

Corresponding author: Pawel Czarnul
E-mail: pczarnul@eti.pg.gda.pl

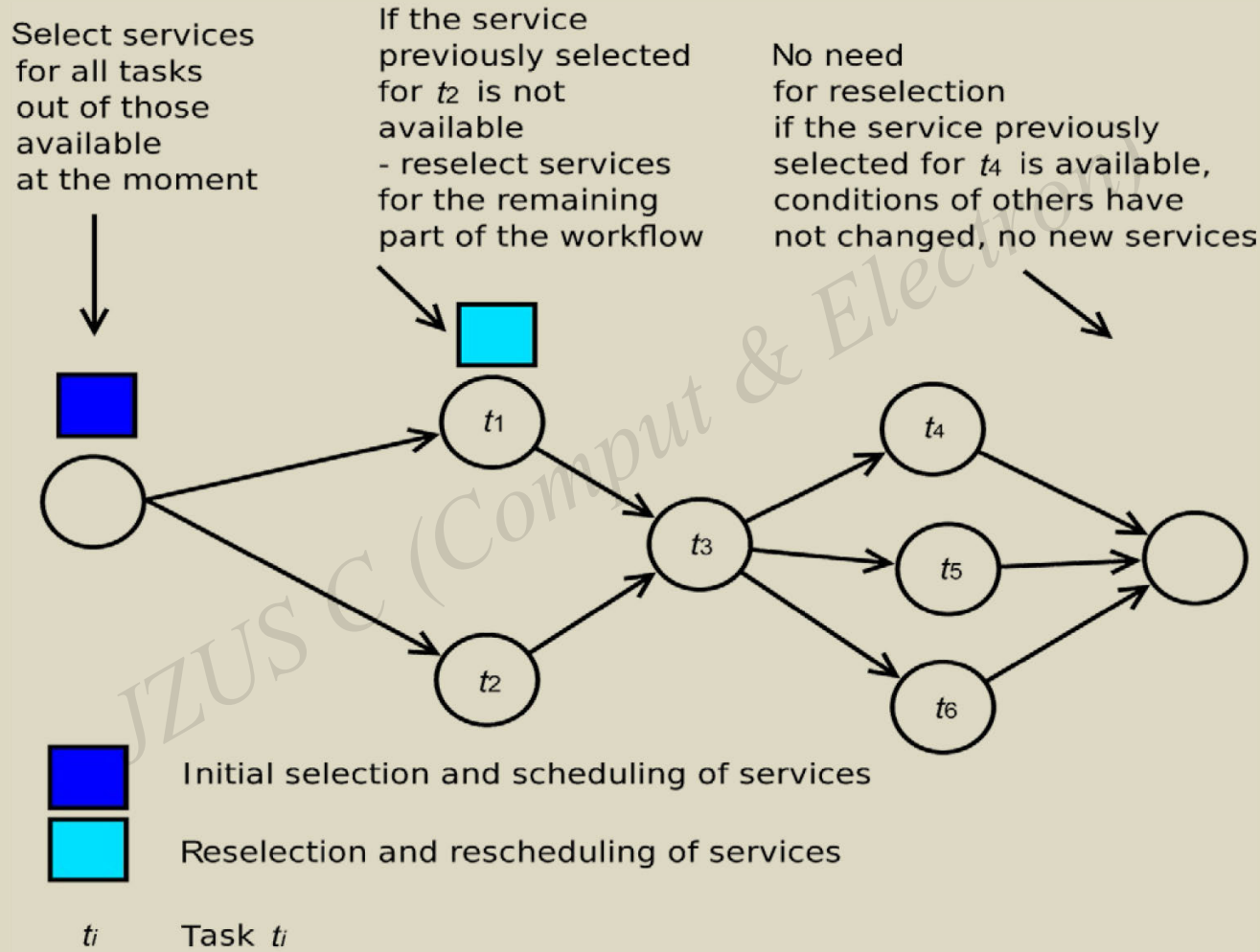
Introduction

- Workflow applications have become widespread in various fields of study and context such as:
 - Scheduling in utility grids
 - Business applications
- There is a need for workflow scheduling algorithms that would cope with dynamic changes in the environment such as variable service availability
- Rescheduling is needed if:
 - Previously selected services have become unavailable
 - New services have appeared
 - Conditions of services have changed

Contribution

- A detailed side-by-side performance-cost comparison of the following algorithms is made:
 - Integer Linear Programming (ILP), heuristic (ILPHEU)
 - Divide-and-conquer (DaC)
 - Genetic algorithm (GA)
 - GAIN
- Comparison is performed in a dynamic environment with changeable service availability and potentially other QoS parameters
- The optimization goal is to minimize the workflow execution time with an upper bound on the cost of selected services

Dynamic workflow scheduling



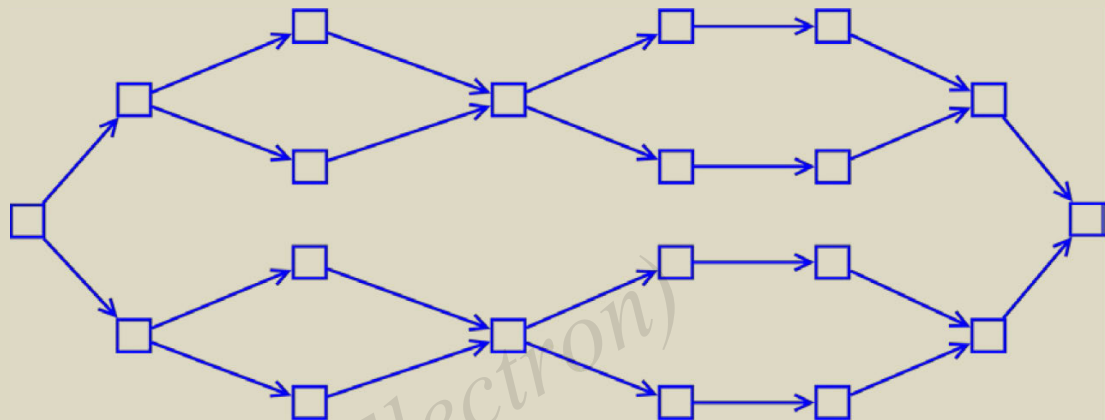
Workflow application scheduling with changing service availability and parameters

Testbed workflow applications and environment

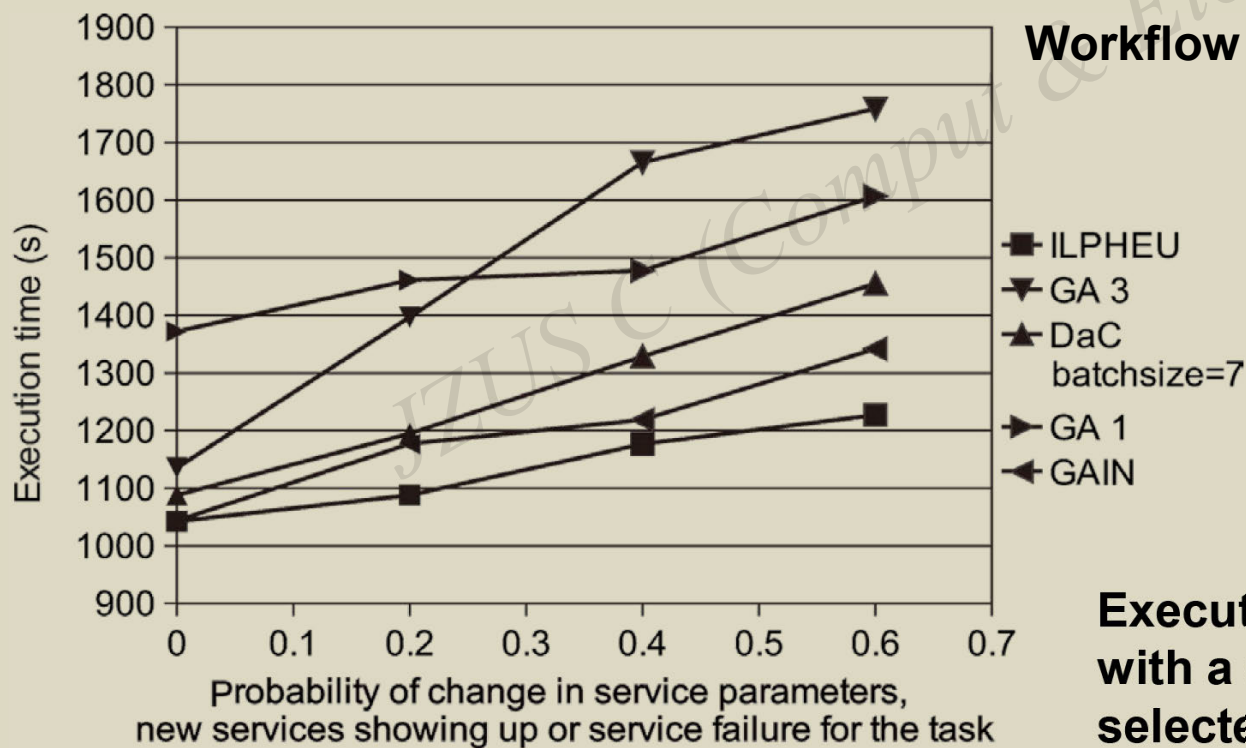
- Tests are performed for several balanced and imbalanced workflow applications:
 - Number of tasks (nodes) ranging from 6 up to 102
 - Number of services ranging from 30 to 418
 - Probability of change in service parameters or service unavailability before task is to be executed from 0 to 0.6
 - Execution times of workflow scheduling algorithm considered in the workflow execution time
- Comparison is performed in real environment
BeesyCluster – middleware for integration of services from various HPC and off-the-shelf servers and systems

<https://beesycluster.eti.pg.gda.pl/ek/Main>

Sample results (I)

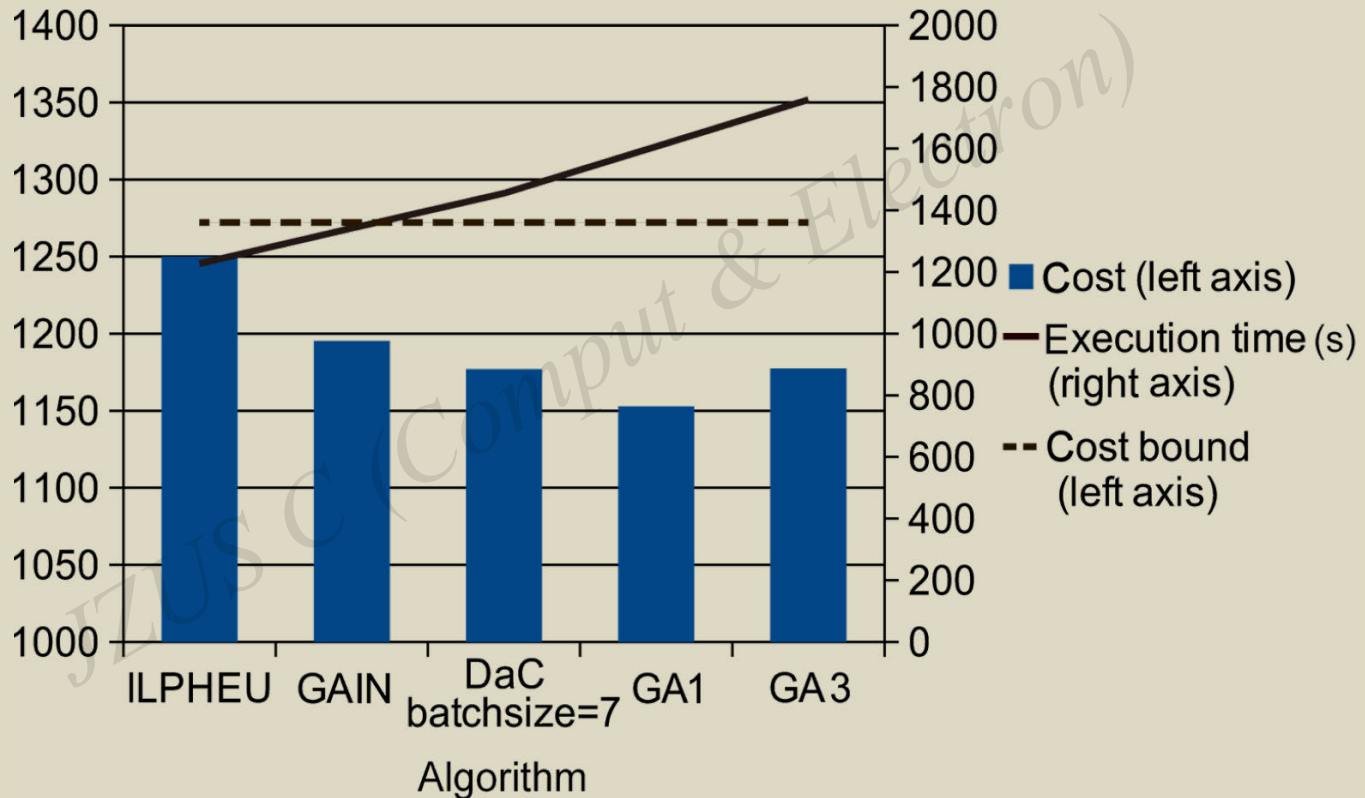


Workflow with 20 nodes, 200 services



Execution time of the workflow with a constraint on the cost of selected services

Sample results (II)



Execution times and costs for algorithms with 20 nodes, 200 services

Conclusions

- For small problems, the algorithms from best to worst are: ILP, DaC, GAIN, GA.
- For mid-sized problems up to 40 workflow nodes, the algorithms from best to worst are: ILPHEU, GAIN, DaC, GA.
- For large workflows with over 100 nodes and several hundred services, the algorithms from best to worst are: ILPHEU, DaC, GAIN, GA.