Marek Tudruj, Ph D, D Sc,
Polish-Japanese Institute of Information
Technology
Institute of Computer Science
of the Polish Academy of Sciences

Review of a PhD thesis of Mr Krzysztof Rządca:

Resource Management Models and Algorithms for Multi-Organizational Grids

1. The Context and the Scope of the Thesis

Many scientific, engineering and commercial computational problems have permanently been hindered by insufficiency of processing power. Since the long-lasting development of microprocessors performance by a brute–force increase of the processor clock frequency has been recently limited by the high heat dissipation in processor chips , the computation parallelization seems to start playing a crucial role in achieving high performance of current computing systems. Highly parallel systems and clusters built from commodity computer components, have been widely and successfully used since at least 1990s.

Although clusters are modular and easy to extend, from a certain point of view, a number of limitations appear which also restrains the single cluster sizes. Firstly, the budget of a funding institution is usually limited. With the increase of cluster size, the job service time and operating system latencies become not negligible. Consequently, in order to build a system of a very large scale, many clusters and supercomputers of different types can be interconnected in wide area agglomerations of generalized computing resources called Grids. In such largely distributed and heterogeneous systems, consisting of many computer clusters owned and controlled by different parties, a crucial aspect for computing efficiency is job and task scheduling as well as management of resources.

Grids are more and more popular in a number of scientific, engineering and commercial projects. Many existing systems have already achieved production-level quality and a large scale. The organizational decentralization of the Grid determines many new problems in the domain of scheduling and resource management. In such distributed

systems, the increased latency and limited bandwidth of communication should also be taken into account both by the users and the resource management policy. Diverse, and often contradictory goals of resource owners, administrators and users should be respected.

The subject matter of the thesis is concerned with parallel processing, operating systems and operational research. The thesis concentrates on scheduling and resource management in Grids, which are composed of resources belonging to different administrative entities (called organizations). The author considers diverse types of control the Grid scheduler might have over the resources (centralized control when organizations completely grant their resources to the system; limited control, when organizations can quit the system; and very weak global control, when organizations can arbitrary change the schedule on the local resources). The analysis is applied to three different types of systems (dedicated processors, identical processors with divisible loads and multiple parallel processors). Depending on the level of control, different mathematical modeling is used (equitable multi-objective optimization, constrained equitable multi-objective optimization and game theory). Using both mathematical modeling and simulation, the thesis analyses the cost of increased decentralization in terms of the loss of system's efficiency, compared to classic centralized systems. To conclude, the problems discussed in the thesis (analysis of program scheduling algorithms by the game theory and the axiomatic theory of fairness) belong to a novel, interesting field of research which is important both for the theory and practice.

2. Evaluation of the goals of the Thesis

The goal of the thesis was to demonstrate that the complete organizational decentralization in Grids can lead to an unacceptable loss of performance of the system. The hypothesis has been formulated that if Grid participants agree to follow certain rules and participate in algorithms that balance the system load, it is possible to distribute the load so that the system is used efficiently as a whole and, at the same time, the individual participants are treated fairly and the system respects their individual goals. The author also claimed that it was possible to construct such algorithms without the need of using external forms of compensation (money, tokens, ...) for accessing resources, commonly proposed in the literature. The aim of the thesis was to propose a number of such

algorithms for various types of systems considered. The goals of the thesis have been formulated in a clear way and the results shown in the thesis enable to state that these goals have been fulfilled.

2.1 The Contents of the Thesis

The doctoral dissertation by Mr Rządca is composed of 6 chapters and a bibliography consisting of over 70 entries. The overall contents of the dissertation can be subdivided into 3 parts. In the first part (Chapter 1), the author presents the context of the work and formulates the goals and the scope of the thesis. In the second part (Chapter 2), after formally defining the terminology and introducing the notations, the framework model of the system is proposed. This model, applied to various concrete systems is analyzed in the third part of the thesis (Chapters 3-5).

Chapter 1 contains an introduction to the various techniques used later in the thesis. After discussing the motivation, the scope and the goal of this work, the author introduces tools used in the theoretical analysis of the proposed models: multi-criteria optimization, the axiomatic theory of fairness, game theory and approximation algorithms. Then, the overview of related works is presented. The author divides the broad literature into four subject sections: Grid systems, which describes the most important approaches to build such systems and discusses two typical Grids (EGEE and Grid'5000); scheduling theory; economic approaches to Grid resource management, in which the author presents various problems commonly not taken into account by those approaches; and, finally, game theory, in which game-theoretic methods oriented to scheduling are analyzed. The chapter is completed by the summary of contributions to Grid modeling and program scheduling done in the sequel of the thesis.

Chapter 2 formally introduces the model of the Grid used in the remaining parts of the thesis. The model is based on the notion of an organization: an independent entity that grants a resource to the Grid, but expects in return that the jobs produced by its local users would be treated fairly. The author introduces also definitions of the concepts used to describe various types of parallel systems and the measures of performance that an organization can use to assess the performance of executed jobs. The efficiency of the solutions produced by the Grid scheduler depends on the control it has over local scheduling in the resources. The author considers three possible control levels,

corresponding to the increasing decentralization of the system: fully centralized control with a complete control of local scheduling; limited control, where organizations cannot change the Grid schedule but can quit the system; and almost not centralized control, where organizations can change the schedule on the local resource in an arbitrary way if profitable. For each type of the system, appropriate mathematical modeling is proposed (equitable multi-objective optimization, constrained equitable multi-objective optimization, and game theory). Finally, the author discusses how the proposed model can be applied to the real-world Grid systems considered in the previous chapter.

In Chapter 3, the framework model is applied for job scheduling in a system with dedicated uni-processors. Each job in the system must be computed on a specific processor, not necessarily the one owned by the producer of the job and the off-line model is assumed in respect of the release time of jobs. The author proves that the decision version of the resulting unconstrained multi-criteria optimization problem is NP-complete, and that the number of Pareto-efficient strategies can be exponential with the number of jobs. Then, three algorithms are proposed for the problem: exhaustive search, an exact dynamic programming and a greedy heuristics based on taboo-search. Next, the model is analyzed in terms of game-theory: the author proves that the worst-case loss of efficiency can be proportional to the number of jobs in the system. Finally, the thesis shows how the algorithms can be applied to constrained optimization problems. The algorithms are tested by extensive simulation experiments.

In Chapter 4, the framework multi-organizational Grid model is applied to systems processing so-called divisible load (jobs that can be divided into a large number of fragments, processed independently and in parallel). The considered job time model is on-line in respect to the system life. The chapter starts with the analysis of the expected delay a foreign job may cause on a local job. The author derives a cost function, which expresses the expected delay as a function of the size of the local and the foreign loads. Using this function, it has been shown that the perfect load balancing is equitably-optimal, and thus efficient from the system's point of view. Yet, as the receiving organization has a non-zero cost, load balancing will not be performed in the fully-decentralized scenario. Then, the limited control scenario is considered. When resources are similarly loaded, it is sufficient to force organizations to collaborate during a longer

time period in order to achieve full efficiency of the system. Then, the author considers situations when some of the resources are overloaded. It is demonstrated that if organizations decide whether to cooperate observing their known load, cooperation will never appear. Finally, the author presents a load balancing algorithm that is appropriate for systems with clusters, which show significantly different loads. It relies upon a centralized, fully-informed control of the foreign load queues and the strategy of prioritary controlling the least-loaded resources. All the algorithms are validated by simulation.

In Chapter 5, the proposed multi-organizational Grid model has been applied to the classic, NP-hard problem of scheduling parallel jobs on multiprocessor resources. This chapter proposes new scheduling algorithms based on list scheduling approach and discusses their worst-case performance. Firstly, the author proposes to sort the jobs by the decreasing number of required processors, which guarantees a satisfactory use of resources. Then, the author shows that in the case of a fully centralized system, the worst-case performance of the list scheduling algorithm is increased in comparison with a single multiprocessor. In the fully decentralized scenario, the author shows that the system can achieve the performance of at most twice the optimal. Finally, a scheduling algorithm is presented, which has a fixed, worst-case guarantee on the global makespan of the system, while at the same time does not increase the makespans of individual organizations. The proposed algorithms are validated experimentally.

The thesis is concluded by the obtained results summary.

2.2 Main Contributions of the Thesis

The main contributions of the thesis consist of:

- a model of the Grid that allows to take into account different levels of the organizational decentralization;

- an analysis of the fairness of job scheduling algorithms using the axiomatic theory of equity, this approach allows to balance equity and efficiency;

- demonstration that the job scheduling in fully-decentralized systems with dedicated processors is analogous to the Prisoner's Dilemma, thus, such systems will work inefficiently;

- a load balancing algorithm for divisible load processing that delivers equitable, efficient load distribution even when loads of processors differ significantly;

- demonstration that the performance of the widely used, greedy list-scheduling algorithm degrades on multiple multiprocessors;

- a scheduling algorithm for parallel, rigid jobs on many multiprocessors that has a fixed, worst-case global performance and, at the same time, does not degrade the makespan of any organization;

- the experimental evaluation of the efficiency of the proposed algorithms.

2.3 Remarks

After having read the text of the dissertation no major flaws nor errors have been detected. The results have been obtained using the correct methods based on the multi-criteria optimization and game theory approaches. The following questions in the dissertation remain unclear:

- The theoretical study elaborated in the thesis relating to game theory is based mainly on one-shot games. What kind of results one can expect if models of repetitive games would be applied ?

- It has been stated that the developed game-theoretical model is "similar" (Chapter 1) or "analogous" (Chapter 3) to Prisoner's Dilemma (PD) game. To what degree these two models are similar or analogous in the light of the pay-off function?

- How the notion and values of the Price of Anarchy (Chapter 3) can be interpreted in terms of a Grid organization ?

- The author focused on the theoretical part of the problem and evaluation by simulation, it would be interesting to see the implementation of the proposed algorithms in real-world systems.

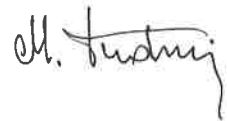## 2.4 Evaluation of the Style of the Dissertation

Generally, the dissertation is written using a clear, scientific language. The arguments are straightforward to follow. The text is supported by a number of helpful figures and tables.

3. Conclusions

The remarks enumerated above do not decrease the overall positive evaluation of the dissertation under review. The thesis contains original interesting theoretical results, supported by extensive experimental evaluation and I consider that the presented results are of very high quality. Parts of the presented results have been published in three papers by well known publishing houses like ACM, IEEE and Springer, and in one paper accepted for publication in European Journal of Operational Research. The results were also presented at international scientific conferences.

Therefore, I can state that the presented to me thesis of Mr Krzysztof Rządca fulfills with an excess all requirements concerning the PhD dissertations, defined by the Polish law, and I recommend admitting Mr. Rządca and his thesis to the public defense.

Due to the very high quality of the theoretical and practical results of the thesis, high significance of the obtained results for the theory of program scheduling in Grid systems and the high rank publication of the results of the thesis (publications in journals from the Philadelphia list) I recommend considering this PhD thesis for distinction.