PREFACE

Miroslaw Thor – Marek Tudruj

The goal of this special issue of *The Journal of Supercomputing* is to present current advances in parallel and distributed computing. We have invited the authors of best papers presented at WORLDCOMP'09 Conference on Parallel and Distributed Processing and Applications to submit extended, improved versions of their papers. It was a two-level selection. First, 14 papers were preliminary chosen from over hundred papers presented in 15 conference sessions on parallel aspects of computer architecture, distributing processing and applications. Next, after the invited authors have submitted their new, extended papers, only 7 papers with the highest reviewing rating were finally accepted for this special issue of the journal.

The papers cover a large range of important topics in parallel and distributing computing and applications like scalable, distributed and multi/many core applications and systems, grid computing, different architecture issues, communication strategies, routing and protocols.

Let us characterize shortly all included papers. In the paper "*A moving threads processor architecture MTPA*", a new kind of approach for multi-core processor architectures is proposed. In contrast to traditional approach where each thread stays in the core where it has been created and data is moved from the main memory via caches to each core and thread, in the moving threads approach, each core can access only a certain portion of the main memory via its local memory, and thus extremely lightweight threads are moved between the cores. Therefore, all kinds of cache coherence problems and the need for read reply messages are eliminated. The overall structure, operation, instruction set and thread management mechanism are described and the proposed architecture is evaluated with different functional unit settings with simulations as well as early silicon area and power consumption are estimated.

For molecular dynamics simulation, the paper entitled "*Exploiting Hierarchical Parallelisms for Molecular Dynamics Simulation on Multi-core Clusters*" describes the parallelization techniques applied at three different levels: inter-node via message passing, intra-node via multithreading, and data-level parallelization on each core by the use of SIMD instructions. After shortly explaining the application and the techniques, speedup results on Bluegene machines are presented for inter-node, and on Intel Xeon machines for other techniques. The paper also presents a model on how to estimate the effect of SIMD usage given different memory access latencies, and shows the usefulness by comparing with measurement results. The paper presents a full range of parallelization techniques and issues for a given application to be run on current highly hierarchical parallel systems, and thus is interesting to read.

The paper "*Statistical Measures for Quantifying Task and Machine Heterogeneities*" presents a new way in dealing with machine and task heterogeneity in resource allocation problems.
Finding the optimal mapping of tasks to machines in a heterogeneous computing (HC) system has been shown to be, in general, an NP-complete problem. Therefore, heuristics have been used to find near-optimal mappings. The performance of allocation heuristics can be affected significantly by such factors as task and machine heterogeneities. A number of statistical measures to quantify the heterogeneity of HC systems are proposed. The authors identify different statistical measures used to quantify the heterogeneity of HC systems, and show the correlation between the performance of the

heuristics and these measures through simple mapping examples and synthetic data analysis. In addition, it is illustrated how regression trees can be used to predict the most appropriate heuristics for an HC system based on its heterogeneity. In the paper, the impact that the heterogeneity measures may have on the performance of five different heuristics was demonstrated through simple examples.

In the paper "*Multi-CMP System with Data Communication on the Fly*" a system architecture based on multiple Chip Multiprocessor (CMP) modules interconnected by a global network with a special new feature of the communication on the fly inside the CMP modules is proposed and described. Communication on the fly can be an important data exchange mechanism for execution of computational algorithms in which strong data sharing appears among parallel fragments of programs. It enables strong reduction of data traffic on busses which lead from processor cores to shared L2 data caches and main memory modules. This type of communication should be embedded in special CMP modules meant for execution of numerical fragments of parallel programs.
Generally this paper presents a new architectural model for group communication in multi-core systems. Its basic idea is to allow several cores, organized as a cluster, to prefetch data in the same time. As it is presented in the paper, this mechanism applies between L2 and L1 caches. Considering the huge interest in the effective use of resources in multi-core systems, this paper is of actuality. A lot of research work is carried out especially regarding power management and parallel execution.

The next two papers concern Grid Computing. In the paper "*Power Efficient Scheduling Heuristics for Energy Conservation in Computational Grids*" the authors propose a power-aware scheduling scheme, which reduces power consumption by changing the states of the workstation to hibernating or offline. Different solutions for reducing the cost factor in the operations of a data center, at different levels: component, workstation, cluster and grid -- are presented. The power aware scheduling algorithm is an extension of the PRISM scheduling algorithm proposed by authors for the MARS (Management Architecture for Resource Services) framework. After discussing the approach taken in the new algorithm development, the authors describe three phases of the algorithm: clustering, including workstations with overall higher utilization and workstations with high utilization for the current job; schedule and design of alternate schedules; schedule selection from all the candidates. The description highlights similarities with the PRISM algorithm and motivates different decisions taken in the algorithm design. In next part of the paper, the authors discuss the evaluation results obtained for power conservation and response time. The scheduling algorithm performance is presented and more specifically the time taken to decide on the schedule of a job. The authors observe that the implementation at minimum cost has a very high response time, while the minimum response time heuristics uses more power.

In the paper "*A Multilevel Scheduler for Batch Jobs on Large-scale Grids*" a priority-based two-level schedule is presented with the following three major objectives: load balancing between clusters, hardware and software (licenses) usage optimization and respecting job deadlines. The first level (meta) of the scheduler computes the priorities of jobs at their submission and the second cluster level (local) readjust these priorities at every arrival/completion of a job. Different heuristics are proposed at each level (meta and local). For validation, the proposed scheduler has been evaluated using an event-based simulator. The obtained results demonstrate the efficiency of the proposed approaches.

However, multiprocessor nodes and multi-core architectures, as well as programmable NICs (Network Interface Cards) provide new opportunities to take advantage of the available multi-gigabit per second link bandwidths, and achieving adequate communication performance levels efficient parallel processing of network tasks and interfaces should be considered. Therefore, in the paper "*Improving*

IPS by network processors" the authors describe and evaluate an intrusion prevention system (IPS) based on a multithreaded network interface. It makes it possible to take advantage of the parallelism implemented in network processors to improve not only the latency, but also the bandwidth of legitimate traffic that shares the same communication path with the corrupted traffic. The authors take the advantage of the IXP28xx features. With a multithreaded network interface they have developed a IPS which can be placed in two different positions (the Micro Engines or the host) so a better and fair comparison can be established among the IPS processing done on a host general purpose CPU or on a network processor. The authors indicate that placing the IPS close to the network, by using specialized network processors, gives many times lower latency and higher bandwidth available to the legitimate traffic. An important result is that the authors' prototype has shown a huge benefit to the legitimate traffic of moving the IPS from a CPU to Micro-Engines, and therefore in a hybrid system, using the power of a network processor for certain tasks and the flexibility of a CPU for others, can become a high performance combination.

Miroslaw Thor, Quest Editor
Telemark University College, Bø i Telemark, Norway

Marek Tudruj, Co-Quest Editor
[1] Institute of Computer Science of the Polish Academy of Science, Warsaw, Poland
[2] Polish-Japanese Institute of Information Technology, Warsaw, Poland