s

PROPOSITION DE SUJET DE THESE
Campagne 2020/2021

A remplir en français ou en anglais en fonction de la langue qui sera utilisée pour la thèse

Date: 11 Décembre 2019

ECOLE DOCTORALE: INFOMATHS

TITLE OF RESEARCH SUBJECT: **Energy-aware strategies for periodic scientific workflows under reliability constraints on heterogeneous platforms**

Research team: Equipe ROMA, laboratoire LIP
http://www.ens-lyon.fr/LIP/ROMA/

Supervisor: Yves Robert, Professeur ENS Lyon
Email : yves.robert@ens-lyon.fr

Co-supervisor: Jing Liu, Professeur ECNU Shanghai
Email : jliu@sei.ecnu.edu.cn

Co-supervisor: Frédéric Vivien, Directeur de Recherche INRIA
Email : frederic.vivien@inria.fr

Lab Language: English

Abstract:

This PhD thesis focuses on energy-aware mapping and scheduling strategies for real-time workflows of scientific applications that execute on high-performance computing platforms subject to transient errors. Consecutive copies of the same workflow, expressed as a task directed acyclic graph with precedence constraints, are periodically input to the platform. The optimization problem is to minimize the expected energy consumption while enforcing the period and matching reliability constraints.

Research topics: Algorithms, mapping, scheduling, resilience, transient errors, reliability threshold, energy consumption.

Advisor: The thesis will be conducted within the ROMA project-team at ENS Lyon, under the co-supervision of Professor Yves Robert and Dr. Frédéric Vivien in Lyon, and Professor Jing Liu in Shanghai.

Thesis description: This PhD thesis focuses on energy-aware mapping and scheduling strategies for real-time workflows of scientific applications that execute on high-performance computing platforms. Consecutive copies of the same workflow, expressed as a task directed acyclic graph (DAG) with precedence constraints, are periodically input to the platform, which is subject to transient failures.

The DAG is composed of N tasks. Instance number i is released at time i.P and must complete by time (i+1)P, where P is the global period of the problem. The target parallel platform is composed of M heterogeneous processors with different (and possibly unrelated) characteristics, including speed profile, energy cost, and failure rate.

The objective is to minimize the expected energy consumption while enforcing a given reliability threshold and meeting all periodic deadlines. Each task is subject to transient failures, with some probability that is typically evaluated as a function of the task worst-case execution time on each different processor, and on the failure rate of that processor. A reliability threshold is enforced for the workflow, and must be converted into a threshold for each task. The first problem is to design efficient algorithms to perform this conversion.

Once we have a reliability threshold for each task, we have to decide: (i) how many replicas to use; (ii) on which processor to map each replica; and (iii) when to schedule each replica on its assigned processor. Different mappings achieve different levels of reliability and consume different amounts of energy. Scheduling matters because once a task replica is successful, the other replicas of that task are cancelled, which calls for minimizing the amount of temporal overlap between any replica pair.

On each processor, the actual duration of each task is a random variable whose support is upper bounded by its worst-case execution time on that processor.

The energy consumed has a static part that only depends upon which are the enrolled processors, and a dynamic part that depends upon the actual execution times and successes or failures of each replica. The objective is to minimize the expectation of the total consumed energy.

This complicated optimization problem can be addressed via the design of several algorithms for the different phases: computing the reliability thresholds, computing the number of replicas and the assigned processors for each of these replicas, scheduling the replicas on each processor.

As a first step, the thesis will consider simple dependence graphs such as chains or fork-joins.

General methodology: The work will be conducted along three main lines:

- Synthesis of relevant literature;

- Design and analysis of mapping and scheduling algorithms;

- Evaluation through simulations and using benchmark workflows.

Pre-requisite:

- Basic background in computer science, probability theory;

- Interest in algorithm design and complexity.

Bibliography:

(a) General references:

- T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, Introduction to Algorithms. The MIT Press, 2001.

- M. Mitzenmacher, E. Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, 2005.

- J. Leung editor, Handbook of Scheduling: Algorithms, Models and Performance Analysis, CRC Press 2004.

- H. Casanova, A. Legrand, and Y. Robert, Parallel Algorithms, CRC Press 2008.

(b) Reference on real-time workflows:

- L. Han, L.-C. Canon, J. Liu, Y. Robert, and F. Vivien, “Improved energy-aware strategies for periodic real-time tasks under reliability constraints,” in 40th IEEE Real-Time Systems Symposium (RTSS), 2020.

- G. Xie, G. Zeng, R. Li, and K. Li, Scheduling Parallel Applications on Heterogeneous Distributed Systems. Springer Singapore, 2019.