Article Compilation

Nuno Reis

December 8, 2023

1 Introduction

For the state of the art, I have chosen to focus on 10 different articles, all related to the topic of Energy Aware Scheduling. To begin, I will provide a concise summary of each article's objective and the methodology employed in their respective studies.

1.1 Energy aware mixed tasks scheduling in real-time systems

The objective of this paper is to address the problem of scheduling mixed tasks, which include hard real-time periodic tasks with shared resources and soft aperiodic tasks. They want to reduce energy consumption and decrease the response time of aperiodic tasks.

They use two scheduling algorithms SMTS (Static Mixed Task Scheduling) and DMTS (Dynamic Mixed Task Scheduling). The SMTS algorithm schedules periodic tasks with optimal speed and aperiodic tasks with maximum speed. On the other hand, the DMTS algorithm reclaims dynamic slack time (assigns priorities to processes based on their slack time) generated from the early completion of periodic tasks and uses a CBS (Constant Bandwidth Server) to reduce energy consumption. The algorithms combine Dynamic Voltage Scaling (DVS) and Dynamic Power Management (DPM) techniques. The document concludes that the DMTS algorithm reduces average energy consumption $\times$ response time by 7.18% and 53.66% compared with SMTS and BL algorithms, respectively.

1.2 Energy-Aware Scheduling for High-Performance Computing Systems: A Survey

The document aims to explore various aspects of energy-efficient task scheduling on heterogeneous computing systems with a focus on performance and energy optimization.

I need to read more on this later.

1.3 Performance and energy aware scheduling simulator for HPC: evaluating different resource selection methods

The objective of the study is to evaluate and compare different resource selection methods for high-performance computing (HPC) using the PEAS (Power and Energy Aware Scheduling) simulator. The focus is on optimizing response time and power consumption in HPC centers.

The study employs the PEAS simulator to implement and test various resource selection algorithms, including the non-dominated sorting genetic algorithm-II (NSGA-II), first-fit, and multiobjective heterogeneous earliest finish time (MOHEFT). Simulations are conducted on different HPC clusters, and performance metrics such as average waiting time, average response time, average execution time, and power consumption are analyzed.

The results indicate that NSGA-II outperformed first-fit and MOHEFT in terms of response time and power consumption in the majority of cases (nine out of 12 simulations). The study concludes that NSGA-II is a recommended algorithm for modern scheduling systems in HPC centers, as it demonstrates superior behavior in saving money, energy, and time.

1.4 Energy-aware Scheduling of Multi-version Tasks on Heterogeneous Real-time Systems

The focus is on embracing heterogeneity, considering both different CPU types and GPU-style accelerators, and incorporating a fine-grained energy model and in-application frequency switching.

The methodology involves the development of an energy-aware scheduling heuristic called eFLS (energy-aware Forward List Scheduling). The approach combines heterogeneous CPU and GPU, multi-version dependent tasks, and a detailed energy model.

The eFLS scheduling approach outperforms a standard FLS scheduler and a state-of-the-art ARSH-FATI scheduler by an average of 45.6% in terms of energy consumption.

1.5 Energy Efficient Scheduling Based on Marginal Cost and Task Grouping in Data Centers

The focus is on minimizing the total energy consumption of servers, cooling systems, and state transitions in data centers. The proposed strategy, referred to as EATS (Energy-Aware Task Scheduling), aims to optimize the energy efficiency of servers and cooling systems by considering the energy costs associated with state transitions.

The document concludes that the proposed strategy, EATS, is effective in optimizing data center energy consumption and improving resource utilization. The experiments demonstrate that EATS outperforms other algorithms (Tech-1 and Tech-2) in terms of reducing total energy consumption under different scenarios, task arrival intensities, numbers of cooling modules, and numbers of servers.

It may not be that useful to my thesis. Maybe the marginal cost method will prove useful.

1.6 Energy consumption prediction using machine learning; a review

Some of the references listed on this document may be useful for me to develop a prediction algorithm. I need to read this better

1.7 An Empirical Approach of Performance and Energy-Aware Scheduling [PEAS] Mechanism in the HPC-Cloud Model

The objective of the document is to propose a PEAS (Power-Efficient and Adaptive Scheduling) mechanism for optimal task scheduling and resource utilization in HPC (High-Performance Computing) clusters. The focus is on optimizing makespan, energy consumption, and average power for HPC-aware scientific workflows, specifically cybershake and montage.

The conclusion summarizes the findings of the research. It highlights the marginal improvements observed by the PEAS mechanism in terms of energy utilization, average power, and makespan for both cybershake and montage workflows. The document emphasizes the importance of optimal task scheduling and resource utilization in HPC clusters and suggests that while there are marginal improvements.

1.8 Simulation-Based Performance Evaluation of an Energy-Aware Heuristic for the Scheduling of HPC Applications in Large-Scale Distributed Systems

The objective of the document is to conduct a simulation-based performance evaluation of an energy-aware heuristic designed for scheduling High-Performance Computing (HPC) applications in large-scale distributed systems.

The document uses simulation as the primary methodology to evaluate the performance of the proposed energy-aware heuristic. It specifically focuses on scheduling HPC applications in the context of large-scale distributed systems. The document compares the performance of ENRG-MaxMin with a baseline policy, MaxMin, in terms of metrics like Job Guarantee Ratio and Total Energy Consumption. In the first scheduling step of the ENRG-MaxMin policy, a parameter called the Selection Margin is used. The document explores the impact of different SM values on the heuristic's performance.

The simulation results indicate that the ENRG-MaxMin heuristic outperforms the MaxMin baseline. Specifically, when the selection margin is set to 25%, ENRG-MaxMin achieves the highest guarantee ratio and the lowest energy consumption.

1.9 An energy-aware scheduling algorithm for budget-constrained scientific workflows based on multi-objective reinforcement learning

The objective of the paper is to address the challenge of energy consumption in cloud computing by proposing a multi-objective reinforcement learning algorithm, called EnMORL, for scheduling scientific workflows. The focus is on simultaneously minimizing makespan and energy consumption while adhering to budget constraints.

The authors propose the EnMORL algorithm, which uses a Chebyshev scalarization function for weight selection. The methodology involves addressing the weight selection problem and ensuring the feasibility of solutions within a budget constraint. The algorithm is compared with two popular multi-objective meta-heuristics, NSGA-II and MOPSO, across different workflows. The evaluation includes makespan, energy consumption, and the hypervolume indicator.

The experimental results demonstrate that EnMORL outperforms NSGA-II and MOPSO in terms of makespan, energy consumption, and the hypervolume indicator across various workflows.

While the focus of the paper is on scientific workflows and budget constraints, the principles and methodologies proposed, the use of multi-objective reinforcement learning, may provide useful on helping on the scheduling of irregular tasks.

1.10 Integrating Dynamic Pricing of Electricity into Energy Aware Scheduling for HPC Systems

The primary objective of the document is to propose a novel job power-aware scheduling design with the aim of reducing the electricity bill of HPC systems. The focus is on considering the individual power profiles of HPC jobs and varying electricity prices throughout the day to optimize scheduling and achieve cost savings.

My thesis focuses on irregular tasks, so this probably wont be that useful.