Discrete Applied Mathematics 196 (2015) 20–27

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Scheduling under dynamic speed-scaling for minimizing weighted completion time and energy consumption

Christoph Dürr [a,b], Łukasz Jeż [c,d], Oscar C. Vásquez [b,e,*]

[a] CNRS, France
[b] Sorbonnes Université, LIP6, Université Pierre et Marie Curie, Paris, France
[c] Blavatnik School of Computer Science, Tel Aviv University, Israel
[d] Institute of Computer Science, University of Wrocław, Poland
[e] Industrial Engineering Department, University of Santiago of Chile, Chile

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form 1 August 2014
Accepted 4 August 2014
Available online 22 August 2014

Keywords:
Scheduling
Energy management
Quality of service
Optimization
Mechanism design

ABSTRACT

Since a few years there is an increasing interest in minimizing the energy consumption of computing systems. However in a shared computing system, users want to optimize their experienced quality of service, at the price of a high energy consumption. In this work, we address the problem of optimizing and designing mechanisms for a linear combination of weighted completion time and energy consumption on a single machine with dynamic speed-scaling. We show that minimizing linear combination reduces to a unit speed scheduling problem under a polynomial penalty function. In the mechanism design setting, we define a cost share mechanism and study its properties, showing that it is *truthful* and the overcharging of total cost share is bounded by a constant.

1. Introduction

Humanity has entered a period when natural resources become rare. This situation triggered consciousness in responsible consumption, and many countries, companies and individuals aim to minimize their energy consumption. Minimizing energy consumption is a relatively new topic in decision theory, giving rise to new problems and research areas.

An area of increasing interest is the energy consumption minimization of computing systems with dynamic speed-scaling, allowed in modern microprocessors by technologies such as Intel SpeedStep, AMD PowerNow!, or IBM EnergyScale. The theoretical energy consumption model has been introduced in [17] and triggered the development of offline and online algorithms; see [1] for an overview.

In these systems, minimizing energy consumption of the machines and minimizing waiting times of the users are opposed goals [12]. Small waiting times improve the quality of service experienced by the users, which generally comes at the price of high energy consumption.

The online and offline optimization problem for minimizing flow time, while respecting a maximum energy consumption, has been studied for a single machine in [2,5,15,7] and for parallel machines in [3]. For the variant where an aggregation of energy and flow time is considered, polynomial time approximation algorithms have been presented in [6,4,13].

In this paper, we study this problem from a decentralized and realistic perspective. Here in order to optimize the objective function, a service operator needs specific information from the users together with the characteristics of the submitted jobs. This situation creates the need for a truthful mechanism.

* Corresponding author at: Industrial Engineering Department, University of Santiago of Chile, Chile. Tel.: +56 22 27184102; fax: +56 22 27799723.
E-mail addresses: christoph.durr@lip6.fr (C. Dürr), lje@cs.uni.wroc.pl (Ł. Jeż), oscar.vasquez@usach.cl, oscar.vasquez@lip6.cl (O.C. Vásquez).

http://dx.doi.org/10.1016/j.dam.2014.08.001

Specifically, we consider a simplified computing system with a single shared machine using dynamic speed-scaling, meaning that it can run at a variable continuous speed to influence the completion times of the jobs. Users submit jobs to this system, each job has some workload, representing a number of instructions to execute and a delay penalty factor. All jobs are available from time 0 on. During the submission of a job, only the workload is publicly known, while the players might announce false delay penalties in order to influence the game towards their interest.

The machine is controlled by an *operator* who aims to minimize the sum of the total weighted completion time and of the energy consumption cost. To this end he decides both on the speed function of the machine and the order in which jobs are to be scheduled. The energy cost consumed by the schedule needs to be charged to the users. The individual goal of each user is to minimize the sum of waiting time and the energy cost share. Therefore it is the charging scheme chosen by the operator that influences the players' behavior.

In a companion paper [9], we study a similar game, where the players announce a strict deadline for their job, while keeping the delay penalty factor private. This way the players control the quality of service guaranteed by the operator, leaving to the operator the goal of optimizing the consumed energy under this constraint.

2. The model

Consider a non-cooperative game with n users and a computing system with a single shared machine using dynamic speed-scaling. Each player has a single job i with a positive workload w_i and a positive delay penalty p_i. When the player submits his job, he announces the workload and some delay penalty $\hat{p}_i$. The announced value might differ from the real value, in order to influence the game toward his advantage, while the workload has to be the true value. The latter assumption makes sense, since it is a quantity observable by the operator, who could punish players in case they lied on the workload.

The machine is controlled by an operator, who upon reception of the jobs has to produce a schedule. This schedule generates some energy consumption, and the controller needs to charge this value to the players, according to some *charging scheme*. This charging scheme is known in advance to all users.

A schedule is defined by an execution order π and an execution speed. Following [17] it is assumed that every job i is scheduled at a constant speed, and for the purpose of simplifying notation we rather specify the execution length ℓ_i of every job i, rather than its speed s_i, which is w_i/ℓ_i. Two costs are associated with a schedule: the energy cost

$$E(\boldsymbol{\ell}, \boldsymbol{w}) := \sum_i \ell_i s_i^\alpha = \sum_i w_i^\alpha \ell_i^{1-\alpha}$$

defined for some fixed physical constant $2 \le \alpha \le 3$, and the weighted flow time, representing the quality of service delivered to the users,

$$F(\pi, \boldsymbol{\ell}, \boldsymbol{p}) := \sum_i p_i C_i,$$

where C_i is the completion time of job i, defined as $\sum_{j:\pi(j)\le\pi(i)} \ell_j$, with $\pi(j)$ being the rank of job j in the schedule.

Ideally the operator would like to minimize the total cost $E(\boldsymbol{\ell}, \boldsymbol{w}) + F(\pi, \boldsymbol{\ell}, \boldsymbol{p})$, which we call the *social cost*. When adding the two costs, we consider the conversion of energy and completion time into monetary values and assume for simplification that the conversion factors are hidden in the penalty factors. When optimizing the social cost, some balance between the two components has to be found, because the energy cost is small when the execution lengths are large, while for the weighted flow time it is exactly opposite. Several problems arise in this situation.

First, the game operator knows only the announced penalty factors $\hat{\boldsymbol{p}}$, so the game has to be *truthful*. This means that every player optimizes his cost by announcing the true value $\hat{p}_i = p_i$.

Second, no polynomial algorithm for finding a schedule $\pi, \boldsymbol{\ell}$ that minimizes $E(\boldsymbol{\ell}, \boldsymbol{w}) + F(\pi, \boldsymbol{\ell}, \boldsymbol{p})$ for arbitrary value α is known. In fact, it is also not known whether this problem is NP-hard. However, a PTAS is known [13]. Dominance properties for this problem have been established in [10]. We note that the solution boils down to finding the right order of scheduling, since once π is fixed, the optimum durations (speeds) for processing each job can be easily determined, cf. [13] or Section 3.

The operator charges to every player i some value b_i, depending on the submitted jobs and on the constructed schedule. This charge has two roles. On one side it is supposed to cover the energy cost of the schedule, and on the other side it influences the players behavior, as every player i wants to minimize the sum of the weighted completion time of his job plus his cost share $p_i C_i + b_i$.

2.1. Desirable properties

In summary, we want to design a cost sharing mechanism that is

- *truthful*, meaning that every player minimizes his penalty by announcing his true value, i.e. $\hat{\boldsymbol{p}} = \boldsymbol{p}$. This implies that the strategy profile $\boldsymbol{p}$ is a pure Nash equilibrium.
- *β-budget-balanced* for some constant β, meaning that the sum of cost shares is at least the energy cost and at most β times this value.
- *efficient*, meaning that the social cost of the Nash equilibrium is close to the social cost optimum.