Sustainable Computing: Informatics and Systems 4 (2014) 241–251

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

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

Reoptimization of the minimum total flow-time scheduling problem[☆]

Guy Baram, Tami Tamir[*]

School of Computer Science, The Interdisciplinary Center, Herzliya, Israel

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form 2 July 2014
Accepted 11 August 2014

Keywords:
Algorithms
Scheduling
Minimum flow-time
Transition cost
Reoptimization

ABSTRACT

We consider *reoptimization* problems arising in dynamic scheduling environments, such as manufacturing systems and virtual machine managers. Due to changes in the environment (out-of-order or new resources, modified jobs' processing requirements, etc.), the schedule needs to be modified. That is, jobs might be migrated from their current machine to a different one. Migrations are associated with a cost – due to relocation overhead and machine set-up times. In some systems, a migration is also associated with job extension. The goal is to find a good modified schedule, with a low transition cost from the initial one. We consider the objective of minimizing the total flow-time.

We present optimal algorithms for the problem of achieving an optimal solution using the minimal possible transition cost. The algorithms and their running times depend on our assumptions on the instance and the allowed modifications. For the modification of machines' addition, we also present an optimal algorithm for achieving the best possible schedule using a given limited budget for the transition.

1. Introduction

This work studies a reoptimization variant of the classical scheduling problem of minimizing the total flow-time on identical machines (denoted in standard scheduling notation by $P||\sum C_j$ [16]). The minimum total flow-time problem can be solved efficiently by the simple greedy Shortest Processing Time algorithm (SPT) [30,9] that assigns the jobs in nondecreasing order by their length. This algorithm, as many other algorithms for combinatorial optimization problems, solves the problem from scratch, for a single arbitrary instance without having any constraints or preferences regarding the required solution – as long as it achieves the optimal objective value. However, many of the real-life scenarios motivating these problems involve systems that change dynamically over time. Thus, throughout the continuous operation of such a system, it is required to compute solutions for new problem instances, derived from previous instances.

Moreover, since the transition from one solution to another consumes energy (used for the physical migration of the job, for warm-up or set-up of the machines, or for activation of the new machines), a natural goal is to have the solution for the new instance *close* to the original one (under certain distance measure).

Solving a *reoptimization* problem involves two challenges:

1 *Computing* an optimal (or close to the optimal) solution for the new instance.
2 Efficiently *converting* the current solution to the new one.

Each of these challenges, even when considered alone, gives rise to many theoretical and practical questions. Obviously, combining the two challenges is an important goal, which shows up in many applications.

Applications: The reoptimization problem of minimizing the total flow-time arises naturally in manufacturing systems, where jobs might be migrated among production lines. Due to unexpected changes in the environment (out-of-order or new machines, timetables of task processing, etc.), the production schedule needs to be modified. Rescheduling tasks involves energy-loss due to relocation overhead and machine set-up times. In fact, our work is relevant to any dynamic scheduling environment, in which migrations of jobs are allowed though associated with an overhead caused due to the need to handle the modification and to absorb the migrating jobs in their new assignment. We describe below several less natural applications in cloud computing and networking.

With the proliferation of cloud computing, more and more applications are deployed in the data centers. Live migration is a common process in which a running virtual machine (VM) or application

☆ A preliminary version of this paper appears in the proceedings of the 1st Mediterranean Conference on Algorithms (MedAlg) December 2012, Ein-Gedi, Israel.

* Corresponding author. Tel.: +972 99602779; fax: +972 9 9568604.
E-mail addresses: guy.baram@gmail.com (G. Baram), tami@idc.ac.il (T. Tamir).

http://dx.doi.org/10.1016/j.suscom.2014.08.011
2210-5379/© 2014 Elsevier Inc. All rights reserved.

moves between different physical machines without disconnecting the client or application [10]. Memory, storage, and network connectivity of the virtual machine are transferred from the original host machine to the destination. Such migrations involve a warm-up phase, and a memory-copy phase. In pre-copy memory migration, the Hypervisor typically copies all the memory pages from source to destination while the VM is still running on the source. Alternatively, in post-copy memory migration the VM is suspended, a minimal subset of the execution state of the VM (CPU state, registers and, optionally, non-pageable memory) is transferred to the target, and the VM is then resumed at the target. Live migration is performed in several VM managers such as Parallels Virtuozzo [25] and Xen [33]. Lot of attention, in both the industry and the academia is given recently to the problem of minimizing the overhead associated with migrations (see e.g., [10,19]). Using our notations, this refers to minimizing the transition costs and the job-extension penalties associated with rescheduling a job. Our work focuses in determining the best possible schedule given these costs. Sequential processing of jobs that might be migrated among several processors is performed also in several implementations of MapReduce (e.g., [5]). These implementations assume that different segments of MapReduce computations can be processed independently on remote computers [12].

Our reoptimization problem arises also in an RPC (Remote Procedure Call) service. In this environment, a cloud of servers can provide service to a limited number of simultaneous users [8]. If the number of requests is high, another virtual server could be temporarily rented, where the cost for using it is per user. The options are to put the RPC in a queue, thus causing latency in the service, or renting more virtual servers, enabling faster service and paying the additional servers' cost. In this application, the transition cost is not due to the migration itself, but due to the activation cost of the additional resources.

Some of our results will be extended to consider modifications that occur after the processing has begun, that is, at time $t > 0$. For this extension (see Section 3) we distinguish between environments in which the currently processed jobs can migrate and be restarted on different machines, and applications in which restarts are not allowed, and a currently processed job must complete its partial processing. The following application describes a system in which restarts are not allowed: In a semiconductor wafers production line, some of the coating methods involve purely physical processes such as high temperature vacuum evaporation (physical vapor deposition – PVD). During the process, a vacuum is created to enable the coating. Once the elements are in a vacuum environment, the process cannot be stopped as if the machine halts, it will be severely damaged [22]. Assume that at time $t > 0$ machines are added. Transferring jobs is costly – to capture the transition overhead and the changes required in programming the machines workplan. Also, the elements that are currently produced, that are already in vacuum state, must complete their production.

1.1. Problem statement and notation

An instance of our problem consists of a set J_0 of n_0 jobs and a set M_0 of m_0 identical machines. Denote by p_j the processing time of job j. A schedule S_0 of the initial instance is given. That is, for every job in J_0, it is specified on which machine it is assigned and on which time interval it is going to be processed. At any time, a machine can process at most one job and a job can be processed by at most one machine.

At time $t \geq 0$, a change in the system occurs. Possible changes include addition or removal of machines and/or jobs, as well as modification of processing time of jobs in J_0. Let J denote the modified set of jobs, and let $n = |J|$. Let M denote the modified set of machines, and let $m = |M|$. Our goal is to suggest a new schedule, S, for the modified instance, with good objective value and small transition cost form S_0. Assignment of a job to a different machine in S_0 and S is denoted *migration* and is associated with a cost. Formally, we are given a *price list* $\theta_{i,i',j}$, such that it costs $\theta_{i,i',j}$ to migrate job j from machine i to machine i'.

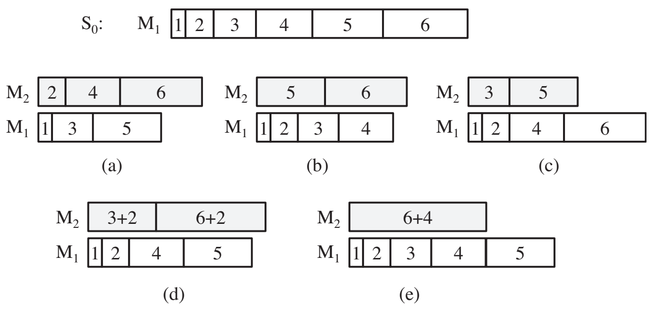

Fig. 1. (Top) An initial assignment, (a) an optimal reassignment achieved with transition cost 3, (b) a possible, and (c) an optimal reassignments achieved with limited budget $B = 2$, (d) an optimal reassignment assuming job-extension penalty $\delta = 2$, and (e) $\delta = 4$.

Moreover, in some systems job migrations are also associated with an extension of the job's processing time. In this extended model, in addition to the transition costs, we are given a *job-extension penalty list* $\delta_{i,i',j} \geq 0$, such that the processing time of job j is extended to $p_j + \delta_{i,i',j}$ when it is migrated from machine i to machine i'. We assume that all the involved parameters (processing times, transition costs and extension penalties) are integers. Some of our results assume unit transition costs, that is, for all j and $i \neq i'$, $\theta_{i,i',j} = 1$.[1]

For a given schedule, let C_j be the completion time of job j, that is, the time when the process of j completes. In this work we consider the problem of minimizing the sum of completion times, denote by $\sum C_j$ and also known as *total flow-time*. In the reoptimization problem, given S_0, J and M, the goal is to find a good schedule for J that is close to the initial schedule S_0.

We consider two problems:

1 Rescheduling to an optimal schedule using the minimal possible transition cost.
2 Given a budget B, find the best possible modified schedule that can be achieved without exceeding the budget B.

Example 1. Assume that six jobs of lengths 1, . . ., 6 are scheduled on a single machine in an optimal SPT order. The total flow time in this schedule is $\sum C_j = 56$. Assume that a second machine is added, and that all migrations have unit transition cost and no job-extension penalties. Fig. 1(a) presents an optimal modified schedule, for which the total flow-time is $\sum C_j = 34$. Migrating jobs appear in grey. The budget required to reach this schedule (or any other schedule with $\sum C_j = 34$) is 3. For a given budget, $B = 2$, it is possible to move, for example, to the modified schedules given in Fig. 1(b) and (c), having total flow-time 36 and 35, respectively. The schedule (c) is optimal for this budget.

Assume further that all migrations are associated with the same job-extension penalty, that is, for all i, i', j, we have $\delta_{i,i',j} = \delta$. An optimal solution for $\delta = 2$ achieving $\sum C_j = 41$ is given in Fig. 1(d). The transition cost to this schedule is 2, however, even with unlimited budget it is impossible to produce a schedule with lower total flow-time. For $\delta = 4$, only one job is migrated in the optimal solution (that achieves $\sum C_j = 45$), given in Fig. 1(e). The optimal algorithm

[1] Note that the constant 1 can be replaced by any other constant.

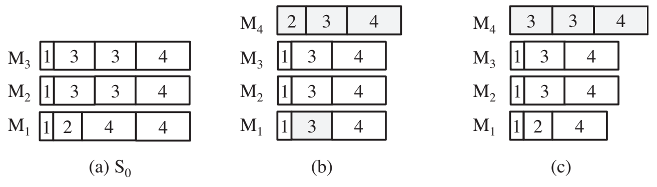

Fig. 2. (a) An initial assignment, (b) an optimal reassignment requires transition cost 4, and (c) an optimal reassignment for $B=3$.

we present in Section 2 solves the problem optimally for arbitrary transition costs and arbitrary job-extension penalties.

Example 2. Fig. 2(a) presents an optimal SPT schedule of 12 jobs on three machines. Assume that a forth machine is added, unit transition costs and no job-extension penalties. Fig. 2(a) presents an optimal modified schedule, for which the total flow-time is $\sum C_j = 55$. The budget required to reach this schedule (or any other schedule with $\sum C_j = 55$) is 4. An important observation is that even though the number of jobs on the added machine is 3, and the initial schedule is optimal, it is not possible to move to an optimal schedule with budget $B=3$. An optimal solution for $B=3$ is presented in Fig. 2(c). It has total flow-time 56.

The above examples demonstrate some of the challenges in solving our reoptimization problems, and the fact that simple natural heuristics do not solve the problem optimally, even if all migrations have unit transition cost. Note first that the natural greedy approach of migrating the long jobs if the budget is low (as in Example 1(b)) is sub-optimal. Another natural approach is prefix-SPT – use the budget to maximize the prefix of the schedule that agrees with an SPT schedule. This approach fails on Example 1 (jobs of lengthes 2 and 4 will be migrated, resulting in $\sum C_j = 37$). Moreover, prefix-SPT as well as suffix-SPT (maximize the suffix of the schedule that agrees with an SPT schedule) are not well-defined since the tie-breaking rule, in case of multiple jobs having the same processing time, is crucial – as demonstrated in Example 2.

1.2. Related work

The 'single-shot' minimum total flow-time on identical machines, $P||\sum C_j$, can be solved in polynomial time by using the shortest processing time (SPT) rule [30,9]. With unrelated machines, that is, when the processing time of a job depends on the machine on which it is processed, the resulting problem (denoted $R||\sum C_j$), is solvable by a reduction to a minimum-weight complete matching problem [7,18].

The work on reoptimization problems started with the analysis of *dynamic graph* problems (see e.g., [13,32] and a survey in [11]). These works focus on developing data structures supporting update and query operations on graphs. Reoptimization algorithms were developed also for some classic problems on graphs, such as shortest-path [24,23] and minimum spanning tree [1].

A different line of research deals with the computation of a good solution for an NP-hard problem, given an optimal solution for a close instance. Among the problems studied in this setting are TSP, [4,6], Steiner Tree on weighted graphs [14] and Knapsack [2]. A survey of other research in this direction is given in [3]. In all of the above works, the goal is to compute an optimal (or approximate) solution for the modified instance. The resulting solution may be significantly different from the original one, since there is no cost associated with the transition among solutions.

The paper [28] suggests the framework we adopt for this work, in which the solution for the modified instance is evaluated also with respect to its difference from the initial solution. This framework is in use also in [27], to analyze algorithms for data placement in storage area network. Considering both the quality of the solution and the transition cost from an initial solution can also be seen as a special case of *multi-objective* optimization problems. In these problems, there are several weight functions associated with the input elements. The goal is to find a solution whose quality is measured with respect to a combination of these weights (see e.g., [26,17]).

Other related work consider several graph algorithms that we apply or adjust in this work. A matching $M \subseteq E$ in a graph $G=(V,E)$ is set of edges such that each node in V appears in at most one edge in M. A Bipartite graph $G=(V,E)$ is a graph in which the vertex set V can be divided into two disjoint subsets V_1 and V_2 such that $E \subseteq V_1 \times V_2$. A complete matching in a bipartite is a matching of size $\min(|V_1|, |V_2|)$. A min-weight complete matching can be found using the Hungarian method [21] or the Push-relabel algorithm [15]. Their runtime is $O(\sqrt{|V|}|E|(\log(|V|^2/|E|))(\log|V|))$ [20]. Another problem whose solution we use as a black box, is the problem of finding a min-cost max-flow in a network. Different approaches have been proposed for solving the min-cost max-flow problem. The minimum mean-cost cycle-canceling algorithm, developed by Goldberg and Tarjan, is a strongly polynomial time algorithm for this problem [15]. Its running time is $O(|V|^2|E|^3 \log|V|)$ [20].

1.3. Our results

In Section 2 we explore the problem of moving to a modified optimal schedule using the minimal required budget. We present an optimal algorithm that returns both an optimal schedule and the minimal budget B required to reach an optimal schedule. The algorithm is suitable for any modification and for arbitrary transition costs and job-extension penalties. The running time of the algorithms is dominated by the time required to find a minimum-weight complete matching in a complete bipartite graph with $O(nm)$ vertices.

The above optimal algorithm assume that the modification takes place at time $t=0$. In Section 3 we describe how and under which conditions it can be extended to handle modifications at time $t>0$. We distinguish between systems in which currently processed jobs can migrate or are obligated to complete their processing in the current machine. We note that our results for change at $t=0$ are not necessarily static, as in many systems, the same workplan is repeated periodically (daily, etc.), thus, every period can be considered as a schedule starting at time $t=0$. A change in the fixed periodic schedule is equivalent to a change at time $t=0$.

Section 4 suggests simple algorithm for restricted instances. In Section 4.2 we present an efficient algorithm for instances with unit migration costs and no extension penalties. The time complexity of this algorithm varies between $O(n)$ (if the initial schedule is optimal) and $O(n \log n)$ (for arbitrary initial schedule). In Section 4.3, we characterize instances for which it is possible to determine the minimal budget in constant time and to find an optimal reschedule in linear-time.

In Section 5 we consider the problem of rescheduling with a limited budget. The goal is to utilize the budget in the best possible way, that is, the modified schedule should have a low total flow-time – the minimal possible among all schedules that can be achieved using the given budget. Our results for this model assume unit migration costs, thus, the budget B gives the maximal number of allowed migrations. The job-extension penalties can be arbitrary. We present an optimal algorithm for the case in which migrations are allowed only to new machines.

We conclude, in Section 6, with a discussion and some directions for future work. We note that our results can be applied also on a sequence of modifications. That is, the environment might change more than once, and the algorithms are performed after each modification.

2. Optimal modified schedule using minimal budget

In this section we consider the problem of moving to a modified optimal schedule with respect to the minimal total flow-time objective, using the minimal possible budget. We present an optimal algorithm for the most general case – with arbitrary modifications, and arbitrary transition costs and job-extension penalties.

Let S_0 be a given initial schedule. We do not assume that S_0 is optimal nor that it has a specific structure or properties. Assume that at time $t=0$, the environment is modified. Possible modifications include addition or removal of machines and/or jobs, and changes in jobs' processing times. The price-list $\theta_{i,i',j}$ specifies for every job j assigned to machine i, how much it costs to migrate j to machine i'. The job-extension penalty list $\delta_{i,i',j}$ specifies for every job j assigned to machine i, by how much its processing time is extended if migrated to machine i'. The goal is to find a new schedule, S, which is optimal with respect to the total flow-time, and has the minimal transition cost from S_0 among all optimal schedules.

We reduce the problem to a minimum weight complete matching problem in a bipartite graph. This approach was suggested by Horn [18], and Bruno, Coffman and Sethi [7] for solving the problem of minimum flow-time on unrelated machines ($R||\sum C_j$). We first review their algorithm.

Reducing $R||\sum C_j$ to a min-weight complete-matching problem: An instance of n jobs and m machines is given by an $n \times m$ matrix, where for all $1 \le i \le n$ and $1 \le j \le m$, $p'_{i,j}$ denotes the processing time of job j if processed by machine i. The algorithm constructs a bipartite graph $G'=(V, E)$, where $V=J \cup U$. The set J represents the set of n jobs (a single node per job). The set U consists of mn nodes, $q_{i,k}$, for $i=1, \ldots, m$ and $k=1, \ldots, n$, where node $q_{i,k}$ represents the kth from last position on machine i. The edge-set E includes an edge $(v_j, q_{i,k})$ for every node in J and every node in U (a complete bipartite graph). The edge-weight function $w\prime : E \to \mathbb{N}$ is defined to be $w\prime(v_j, q_{i,k}) = kp\prime_{i,j}$. As shown in [18,7], a min-weight complete matching in G' corresponds to a schedule achieving minimum total flow-time. The proof is based on the following property:

Property 2.1. *If job j is the kth from the last job to run on machine i, it contributes exactly k times $p'_{i,j}$ to the sum of completion times.*

We now return to our reoptimization problem. The following algorithm solves the problem optimally. The reoptimization problem is reduced to a min-weight matching problem in a bipartite graph G, that has the same set of vertices and edges as the bipartite graph G' described above. The graphs G and G' differ in their edge-weights. Note that edge weights in G (determined in Step 2) consist of two components: a dominant component (with large factor Z) corresponding to the contribution of a job assigned in a specific position to the total flow-time, and a minor component corresponding to the associated transition cost. Both components are combined to form a single weight. Fig. 3 illustrates the bipartite and the edges corresponding to a single job.

Algorithm 1. – An optimal algorithms for rescheduling using minimal budget

1. Let $\theta_{i,i',j}$ be a price list, i.e., it costs $\theta_{i,i',j}$ to migrate job j from machine i to machine i'. In particular, for all i, j, $\theta_{i,i,j}=0$.

   Let $\delta_{i,i',j}$ be a job-extension penalty list, i.e., the processing time of job j is extended to $p_j+\delta_{i,i',j}$ when it is migrated from machine i to machine i'. In particular, for all i, j, $\delta_{i,i,j}=0$.

   Let $\Delta = \max_{i,i',j} \theta_{i,i',j}$, and let Z be an integer larger than $n\Delta$.

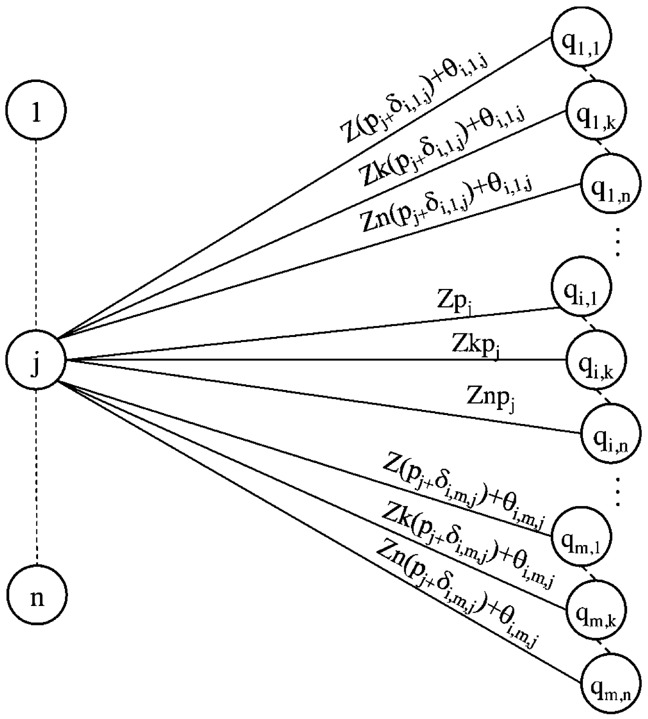
1
j
n
$q_{1,1}$
$q_{1,k}$
$q_{1,n}$
$q_{i,1}$
$q_{i,k}$
$q_{i,n}$
$q_{m,1}$
$q_{m,k}$
$q_{m,n}$
$Z(p_j+\delta_{i,1,j})+\theta_{i,1,j}$
$Zk(p_j+\delta_{i,1,j})+\theta_{i,1,j}$
$Zn(p_j+\delta_{i,1,j})+\theta_{i,1,j}$
Zp_j
Zkp_j
Znp_j
$Z(p_j+\delta_{i,m,j})+\theta_{i,m,j}$
$Zk(p_j+\delta_{i,m,j})+\theta_{i,m,j}$
$Zn(p_j+\delta_{i,m,j})+\theta_{i,m,j}$

Fig. 3. The bipartite graph for Algorithm 1. The figure corresponds to a configuration in which job j is assigned to machine i in S_0, thus the weights of edges $(v_j, q_{i,k})$ do not include a penalty component.

2. Construct a complete bipartite graph $G=(V, E)$, where $V = \{v_j | \forall 1 \le j \le n\} \cup \{q_{i,k} | \forall 1 \le i \le m, 1 \le k \le n\}$. Set the edge weights $w : E \to \mathbb{N}$ as follows: For every job j assigned to i in S_0,
   - Let $w(v_j, q_{i,k}) = Zkp_j$.
   - For every $i' \neq i$, let $w(v_j, q_{i',k}) = Zk(p_j + \delta_{i,i',j}) + \theta_{i,i',j}$.
3. Find a min-weight complete matching in G. Let H denote the set of edges in this matching.
4. Return the schedule corresponding to H. That is, for every $(v_j, q_{i',k}) \in H$, assign j in the kth from last position on machine i'. The minimum transition cost is $\sum_{(v_j, q_{i',k}) \in H} \theta_{i,i',j}$, where i is the machine on which j is assigned in S_0.

Theorem 2.2. *Algorithm 1 returns an optimal schedule using the minimal possible transition cost from S_0.*

Proof. The proof consists of two claims, the first concerning the optimality with respect to the total flow-time, and the second concerning the optimality with respect to the transition cost.

Consider the modified instance and assume that transitions are not associated with costs, but only with job-extensions. Due to the job-extensions, the resulting problem is an $R||\sum C_j$ problem on unrelated machines, where for any machine i' and job j it holds that $p'_{i',j} = p_j + \delta_{i,i',j}$. Let G' denote the bipartite graph built for solving the corresponding min-weight matching problem [18,7]. The graph G' has the same vertex-set and the same edge-set as the graph G built in Algorithm 1. The graphs G and G' differ in the edge weights. As justified by Property 2.1, the weight $w\prime(v_j, q_{i',k})$ of an edge $(v_j, q_{i',k})$ in G', corresponding to a job j assigned to machine i in S_0 is $k(p_j + \delta_{i,i',j})$.

We say that a schedule induced by a set of edges H is feasible if H is a complete matching and for every machine i, either no vertex $q_{i,k}$ belongs to H, or there exists an index n_i such that all vertices $q_{i,k}$ for $1 \le k \le n_i$ belong to H. Informally, this implies that all the jobs are assigned and the jobs are assigned to consequent slots on the machines.

Claim 2.3. *The set of edges H found in Step 3, induces a feasible schedule S with minimum total flow-time.*

Proof. We show that the set of edges H is a minimum weight matching also in G' – and thus, as shown in [18,7], it corresponds to a feasible schedule with minimum total flow-time. First, since G and G' differ only in edge weights, H is a legal matching in G'. Assume by way of contradiction that H is not minimal with respect to the weights $w\prime$ in G', and that H^* is a complete matching in G' with a lower weight. Since all the weights are integers,

$$1+\sum_{e\in H^*} w\prime(e) \le \sum_{e\in H} w\prime(e). \quad (1)$$

By definition of w and $w\prime$, the weight of H^* in G is

$$\sum_{e\in H^*} w(e) = Z\sum_{e\in H^*} w\prime(e) + \sum_{e=(v_j,q_{i,k})\in H^*} \theta_{i,i\prime,j}.$$

By definition of Z, it holds that $\sum_{e=(v_j,q_{i,k})\in H^*}\theta_{i,i\prime,j} \le n\Delta < Z$. Therefore,

$$\sum_{e\in H^*} w(e) < Z(1+\sum_{e\in H^*} w\prime(e)). \quad (2)$$

Since H is a min-weight matching with respect to the weights w, it holds that

$$\sum_{e\in H} w(e) \le \sum_{e\in H^*} w(e). \quad (3)$$

Combining Eq. (1) (multiplied by Z),the definitions of w and $w\prime$, (3) and (2), we get the following contradiction:

$$Z(1+\sum_{e\in H^*} w\prime(e)) \le Z\sum_{e\in H} w\prime(e) \le \sum_{e\in H} w(e) \le \sum_{e\in H^*} w(e) < Z(1+\sum_{e\in H^*} w\prime(e)).$$

□

We conclude that the schedule S returned by the algorithm is a feasible schedule minimizing the total flow-time, and turn to show it also minimizes the transition cost from S_0.

Claim 2.4. *Among all schedules achieving minimum total flow-time, the schedule S induced by H has the minimal transition cost from S_0.*

Proof. Let H^* be any perfect matching in G, corresponding to a schedule, S^*, achieving minimum total flow-time. We show that the transition cost from S_0 to S^* is not lower than the transition cost to S. We know that H is a min-weight complete matching in G, therefore,

$$\sum_{e\in H} w(e) \le \sum_{e\in H^*} w(e). \quad (4)$$

Also, since both matchings induce schedules achieving minimum total flow-time and the weights $w\prime$ in G' reflect the total flow-time without the transition costs,

$$\sum_{e\in H} w\prime(e) = \sum_{e\in H^*} w\prime(e). \quad (5)$$

The definition of w implies that for every matching $\hat{H}$, it holds that

$$\sum_{e\in \hat{H}} w(e) = Z\sum_{e\in \hat{H}} w\prime(e) + \sum_{e=(v_j,q_{i,k})\in \hat{H}} \theta_{i,i\prime,j}, \quad (6)$$

where the second term is exactly the transition cost from the initial schedule to the schedule induced by $\hat{H}$. Therefore, by applying Eq. (6) on both H and H^*, and using Eqs. (5) and (4), we get:

$$\sum_{e=(v_j,q_{i,k})\in H^*} \theta_{i,i\prime,j} - \sum_{e=(v_j,q_{i,k})\in H} \theta_{i,i\prime,j} = \sum_{e\in H^*} w(e) - \sum_{e\in H} w(e) \ge 0.$$

We conclude that the transition cost to S^* is not lower than the transition cost to S. □

Theorem 2.2 follows from combining Claims 2.3 and 2.4. □

Remark: The above algorithm can be extended to an environment of unrelated machines by setting the edge-weights in the bipartite G to be $w(v_j, q_{i\prime,k}) = Zk(p\prime_{i\prime,j} + \delta_{i,i\prime,j}) + \theta_{i,i\prime,j}$, where $p'_{i'j}$ is the given processing time of job j on machine i'.

3. When the modification occurs at time $t > 0$

In this section we extend the algorithm to consider systems that are modified after the processing has begun, that is, at time $t > 0$. Denote by J_t the set of jobs processed at time t, and let, for every machine i, $\gamma_i \ge 0$ denote the time required to complete the job from J_t processed at time t on machine i. As detailed in the introduction, in some systems, the processing of a job $j \in J_t$ must complete on its current machine. In other systems, currently processed jobs can be migrated to another machine and *restart* their processing, that is, after the migration, the job must be processed for $p_j + \delta_{i,i',j}$ time units, independent of its processing before the migration. We present different algorithms for the two settings.

3.1. Restarts are not allowed

When restarts are not allowed and machines are removed, then the problem is not well-defined for the jobs that are processed at time t – since these jobs are obligated to complete their processing on the to-be-removed machines. Thus, we assume that the modifications are machines' addition and/or changes in the set or processing times of jobs. The goal is to determine the schedule of jobs whose processing did not begin before time t. An optimal algorithm for this case is based on the observation that for the modified schedule machine i is available starting at time γ_i. Algorithm 1 can be generalized by setting the weights $w_{noR} : E \to \mathbb{R}$ in the bipartite graph (determined in Step 2) as follows: For every job j assigned to i in S_0,

- Let $w_{noR}(v_j, q_{i,k}) = Z(kp_j + \gamma_i)$.
- For every $i' \ne i$, let $w_{noR}(v_j, q_{i\prime,k}) = Z(k(p_j + \delta_{i,i\prime,j}) + \gamma_{i\prime}) + \theta_{i,i\prime,j}$.

The difference from the case in which the modification occurs at time $t = 0$ is the fact that machine i is available only from time γ_i. Thus, if job j is the kth from last job to run on machine i, it contributes exactly γ_i plus k times its processing time to the sum of completion times. As in Algorithm 1, the weights consist of a dominant component (with large factor Z) ensuring that schedule achieves minimum total flow-time, and a minor component ensuring the minimal possible transition cost.

The availability time of machine i is added to the dominant component, as it affect the flow-time of the jobs assigned to it. The proof of the following theorem follows directly the proof of Theorem 2.2.

Theorem 3.1. *Algorithm 1 with weights w_{noR} returns an optimal schedule using the minimal possible transition cost from S_0, when restarts are not allowed.*

3.2. Restarts are allowed

When restarts are allowed, a job $j \in J_t$ might complete its processing on its current machine, but can also migrate to a different machine. If migrated, the corresponding transition cost is applied and the job must *restart*. We assume that preemptions are not allowed.[2] Another possibility for a job $j \in J_t$ is to remain on its current machine, but delay its processing – letting jobs migrating from other machines precedes it. In this case, the job must restart, but no transition cost is applied, as no migration is performed.

Recall that for every machine i, $\gamma_i \geq 0$ denotes the time required to complete the job from J_t processed at time t on machine i. Our algorithm assumes that the initial schedule, S_0, is optimal and that the modification includes machines' addition. We also assume no job-extension penalties.

Algorithm 1 can be generalized by setting the weights $w_R : E \rightarrow \mathbb{R}$ in the bipartite graph (determined in Step 2) in the following way:

- For every job $j \in J_t$ that is currently processed on machine i let $w_R(v_j, q_{i,k}) = Zk\gamma_i$.
- For every job $j \notin J_t$ that is assigned to machine i, let $w_R(v_j, q_{i,k}) = Zkp_j$.
- For every $i' \neq i$, let $w_R(v_j, q_{i',k}) = Zkp_j + \theta_{i,i',j}$.

Note that the processing time of the currently processed job j on i is assumed to be γ_j even though j might not be assigned to be the first job on machine i. The next lemma justifies these assumptions.

Lemma 3.2. *For every machine i, let $j \in J_t$ be the job processed by M_i at time t, then there is an optimal schedule in which j is not processed second or later on machine i.*

Proof. Assume that there exists an optimal reschedule S^* in which j is scheduled second (or later) on the machine M_1 on which it was assigned in the initial configuration. Recall that γ_1 is the remaining processing time of j on M_1. Since in any optimal schedule, the jobs on every machine are sorted from shorter to longer, there must be at least one job, j', not longer than γ_1, before j. If $p_{j'} = \gamma_1$ then by swapping j and j', the value of the total flow-time is unchanged and j is assigned first as required. Thus, we assume that $p_{j'} < \gamma_1$. Since we assume that the initial schedule is SPT, and j was processed at time t by M_1, the job j' also belongs to J_t. Given that two jobs from J_t are on M_1 in S^*, by the pigeonhole-principle, there must be a machine, M_2, on which no job from J_t is assigned in S^*. Let j^* be the first job on M_2 in S^* (see Fig. 4(a)). Since $j^* \notin J_t$ it must be that $p_{j^*} \geq \gamma_1 > p_{j'}$. Denote by k_1 and k_2 the number of jobs on M_1 and M_2 in S^*, respectively.

If $k_1 > k_2$, move j' to be first on machine M_2 (see Fig. 4(b)). The contribution of j' to the total flow-time before the migration is $k_1 p_{j'}$. The contribution of j' to the total flow-time after the migration is $(k_2 + 1)p_{j'}$. For any $k_2 < k_1$, this migration does not increase the total flow-time. Moreover, it might save the transition cost of j' (if its original machine is M_2), thus, the resulting schedule is either better than S^*, contradicting its optimality, or has the same total flow-time and transition cost as S^*, and it satisfies the requirement that j is first on M_1.

If $k_1 \leq k_2$, move j and j^* to be the first and second jobs on M_1, and move j' to be the first job on M_2 (see Fig. 4(c)). The contribution of these three jobs to the total flow-time before the change is $k_1 p_{j'} + (k_1 - 1)\gamma_1 + k_2 p_{j^*}$. The contribution of these three jobs to the total flow-time after the change is $k_1\gamma_1 + (k_1 - 1)p_{j^*} + k_2 p_{j'}$. The total flow-time reduced by $(p_{j^*} - p_{j'})(k_2 - k_1) + p_{j^*} - \gamma_1$, which is positive for any $p_{j^*} \geq \gamma_1 > p_{j'}$ and $k_2 \geq k_1$. Thus, the resulting schedule has lower total flow-time, contradicting the fact that S^* is optimal. □

We conclude that Algorithm 1 with the weights w_R solves optimally the reoptimaization problem with modifications at time $t > 0$ and restarts allowed. Note that the output of the algorithm is an SPT schedule, therefore, the algorithm can also handle a sequence of modifications.

4. Efficient algorithms for identical migration costs

In this section we consider systems with identical migration costs and no job-extension penalties, that is, for all i, i', j, it holds that $\theta_{i,i'j} = 1$ and $\delta_{i,i'j} = 0$. We present efficient algorithms for finding an optimal modified schedule using the minimal possible budget. The algorithms are based on some properties of the SPT algorithm for $P||\sum C_j$. For completeness, we first review this algorithm and highlight some properties of SPT schedules.

4.1. Review and properties of SPT algorithm

SPT algorithm [30,9] produces a schedule achieving minimum total flow-time of n jobs on m identical machines. We describe a specific application of this algorithm below. Initially, dummy jobs of length 0 are added such that the total number of jobs is a multiple of m. Specifically, if n is not a multiple of m, then $m - (n \bmod m)$ jobs of length zero are added to the instance. The dummy jobs can be scheduled on arbitrary machines and (when rescheduled) their migration cost is 0. Given that n is a multiple of m, SPT algorithm proceeds as follows: First, sort the jobs in non-decreasing order of processing time (break ties arbitrarily). Next, partition the jobs into n/m *rounds* of m jobs each. The rth round consists of the jobs indexed $(r - 1)m + 1, \ldots, rm$ in the sorted list. Schedule on each machine one job from the first round, followed by one job from the second round, etc.

Throughout this work we use the following known properties of SPT schedules:

The *round* property: the internal assignment of jobs from a particular round to the machines does not affect the total flow time, that is, any schedule in which the m jobs of round r are assigned on the r-th slots on the m machines is optimal.

The *load* property: in any optimal schedule the number of non-dummy jobs on any machine is either $\lfloor n/m \rfloor$ or $\lceil n/m \rceil$.

4.2. Efficient optimal algorithm

The first algorithm we present can be applied for addition or removal of machines and/or jobs, as well as changes in jobs' processing times. Let L be the set of job-lengths in the modified instance. Recall that dummy jobs of length 0 are added such that the total number of jobs is a multiple of m. The set L includes at most n distinct values. By the round-property of SPT schedules, an optimal schedule can be characterized by the numbers $n_{\ell,r}$, for all $\ell \in L$ and $1 \leq r \leq (n/m)$, where $n_{\ell,r}$ is the number of jobs of length ℓ in round r, in any optimal schedule. Moreover, the problem of finding an optimal schedule using minimum transition cost reduces to the problem of finding a schedule obeying the optimal $n_{\ell,r}$ values with a minimal number of migrations from the initial schedule. The following is an overview of our optimal algorithm:

Algorithm 2. – An efficient optimal algorithm for rescheduling with identical migration costs.

1. For every length $\ell \in L$ and round $1 \leq r \leq (n/m)$, calculate $n_{\ell,r}$, the number of jobs of length ℓ in round r, in any optimal modified schedule.

[2] Enabling preemptions affects all the jobs of the instance, thus causing the problem to be intractable [29].

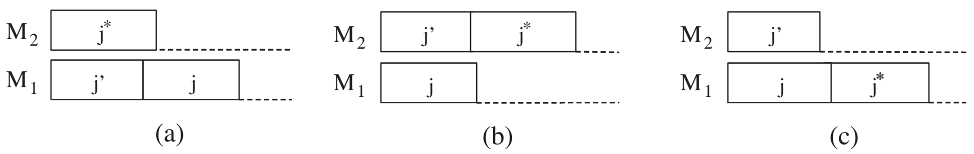

Fig. 4. (a) The assumed schedule, (b) A better schedule if $k_1 > k_2$, and (c) a better schedule if $k_1 \leq k_2$.

2. Partition L into two sets of job lengths: Let $L_1 \subseteq L$ be the set of lengths such that $\ell \in L_1$ if and only if $n_{\ell,r} > 0$ for a single round r. Let $L_2 = L \setminus L_1$ be the set of lengths such that $\ell \in L_2$ if and only if $n_{\ell,r} > 0$ for more than a single round.
3. For every round $1 \leq r \leq (n/m)$, schedule a maximal number of non-migrating jobs in round r. First, assign jobs having lengths in L_1, then in L_2. When assigning jobs from L_2, give higher priority to short jobs.
4. Schedule migrating jobs.

The idea is to assign first a maximal number of non-migrating jobs, and then assign the migrating jobs. When assigning the non-migrating jobs, we first assign the more restricted jobs – having lengths in L_1, and must be assigned in a specific round, and then the more flexible jobs whose lengths are in L_2 (and can be assigned in more than one specific round).

Denote by S the schedule built by the algorithm. Steps (3–4) are implemented as follows: Denote by $S_{i,r}$ the slot in the rth round on machine i. Initially, for all $1 \leq i \leq m$, $1 \leq r \leq (n/m)$ it holds that $S_{i,r}$ is available ($=\emptyset$). During Steps (3–4) some slots are assigned to non-migrating jobs. Whenever a job j of length ℓ is assigned to the r-th slot on machine i, the corresponding variable $S_{i,r}$ is set to j, and the corresponding counter of $n_{\ell,r}$ is reduced by one. Specifically, Steps (3–4) are implemented as follows:

Step 3: Step 3 consists of n/m iterations. In iteration r, the algorithm assigns non-migrating jobs into slots of round r. Consider a slot $S_{i,r}$. Let $ForFree(i, r)$ denote the set of jobs that can be assigned to $S_{i,r}$ with no migration. Formally, $j \in ForFree(i, r)$ if and only if (i) $n_{p_j,r} > 0$, (ii) j is assigned to M_i in S_0, and (iii) j was not assigned to M_i in earlier rounds.

In Step 3, if possible, the algorithm assigns to $S_{i,r}$ a job from *ForFree*(i, r) giving priority to lengths in L_1, and then to shorter lengths in L_2. Formally,

```
For $r = 1$ to $n/m$
  For $i = 1$ to $m$
    Calculate $ForFree(i, r)$.
    If $ForFree(i, r) \neq \emptyset$
      If there exists $j \in ForFree(i, r)$ such that $p_j \in L_1$. Set $S_{i,r} = j$, $n_{p_j,r} = n_{p_j,r} - 1$.
      Else, let $j$ be the shortest job in $ForFree(i, r)$ such that $p_j \in L_2$.
        Set $S_{i,r} = j$, $n_{p_j,r} = n_{p_j,r} - 1$.
```

Step 4: Step 4 consists of n/m iterations. In iteration r, the algorithm assigns, with migrations, jobs to slots $S_{i,r}$ for which $ForFree(i, r) = \emptyset$. Formally,

```
While there exist $\ell$, $r$ such that $n_{\ell,r} > 0$,
  Assign any unassigned job $j$ of length $\ell$ to any machine $i$ s.t. $S_{i,r} = \emptyset$.
    Set $S_{i,r} = j$, $n_{\ell,r} = n_{\ell,r} - 1$.
```

The number of migrations is the number of non-dummy jobs assigned in Step 4. This number is the minimal budget required to reach an optimal schedule. We prove the optimality of the algorithm by combining two lemmas.

Lemma 4.1. *The algorithm produces an optimal schedule with respect to the total flow-time.*

Proof. The schedule S satisfies the $n_{\ell,r}$ values calculated by SPT algorithm, therefore it must be optimal. Since these values were calculated according to the amounts of jobs in the modified instance, all jobs are assigned, that is, in Step 4, while there exist ℓ, r such that $n_{\ell,r} > 0$, it is guaranteed that there is an available empty slot for a job of length ℓ in round r. □

Lemma 4.2. *Every schedule minimizing the total flow-time requires at least the same number of migrations as the number of migrations applied by the algorithm.*

Proof. We prove the following greedy choice property: *for every round r there exists an optimal solution minimizing the total number of migrations, in which the non-migrating jobs assigned to round r are identical to those selected by the algorithm.* The following simple observation will be used to analyze the assignment of jobs having lengths in L_2.

Observation 4.3. *For every round r, there are at most two lengths $\ell_1, \ell_2 \in L_2$ such that $n_{\ell_1,r} > 0$ and $n_{\ell_2,r} > 0$.*

Proof. By definition, jobs of lengths in L_2 span across more than one round in any optimal schedule. Another known property of SPT schedules is that all job lengths in round r are not shorter than job lengths in round $r - 1$ and not longer than job lengths in round $r + 1$. It is not possible to have three different lengths, all spanning over round r and an additional round, since in order to preserve the above SPT property, jobs of the middle length, must all be assigned to round r. □

We prove the greedy choice property for round r: Assume that an optimal schedule agrees with the algorithm in rounds earlier than r, and consider the assignment to round r. For every machine i, if $ForFree(i, r) = \emptyset$ then this is valid also for the optimal assignment, and a migration from another machine to $S_{i,r}$ is inevitable. If $ForFree(i, r)$ includes at least one job then we use exchange argument to show that any selection of job to $S_{i,r}$ that is different from the algorithm's choice can be changed to the algorithm's choice without hurting the total number of non-migrating jobs. Let $j \in ForFree(i, r)$ be the job assigned by the algorithm to $S_{i,r}$. Let $j' \neq j$ be the job assigned in the optimal schedule to $S_{i,r}$. If $j' \notin ForFree(i, r)$, then by switching j and j', we can only reduce the number of non-migrating jobs. If $j' \in ForFree(i, r)$, we distinguish between two cases:

1 $p_j \in L_1$. In this case, j must be assigned to round r, and assigning it to $S_{i,r}$ is the only way to assign it for free. By switching the assignment of j' and j in the optimal assignment, we avoid the migration of j, and cause a migration to j', thus, the total number of migrations does not increase.

2 $p_j \in L_2$. Since the algorithm gives priority to jobs whose lengths are in L_1, it must be that all job lengths in $ForFree(i, r)$ are in L_2 and in particular, $p_{j'} \in L_2$. By Observation 4.3, p_j, p_j' are the only lengths of jobs in $ForFree(i, r)$. Among lengths in L_2, the algorithm gives priority to shorter jobs, therefore, $p_j < p_{j'}$. Moreover, r is the last round in which jobs of length p_j will be assigned, as otherwise, the SPT order is not preserved (given that jobs of length $p_{j'}$ are assigned on both r and $r + 1$). Therefore, assigning j to $S_{i,r}$ is the only way to assign it for free. By switching the assignment of j' and j in the optimal assignment, we avoid the migration of j, and cause a migration to j', thus, the total number of migrations does not increase.

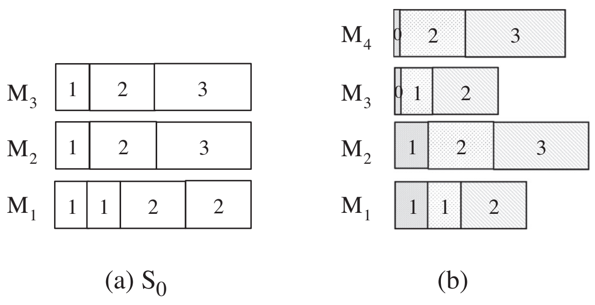

Fig. 5. (a) An initial assignment, (b) an optimal reassignment (with dummy jobs) achieved with transition cost 2. Different fill-patterns denote different rounds.

We conclude that any optimal assignment can be modified such that it agrees with the algorithm's choice, without hurting the number of migrations. Thus, the algorithm produces an optimal assignment. □

Example: Consider the schedule S_0 on three machines, given in Fig. 5(a). Assume that a forth machine is added. We first add two dummy jobs of length 0 to have $n=12$, which is a multiple of four. In every optimal schedule on four machines we have $n_{0,1}=2$, $n_{1,1}=2$, $n_{1,2}=2$, $n_{2,2}=2$, $n_{2,3}=2$, $n_{3,3}=2$. Thus, $L_1=\{3\}$ and $L_2=\{1, 2\}$. The optimal reassignment produced by the algorithm is given in Fig. 5(b). The algorithm assigns first the non-migrating jobs: In the first round, two jobs of length 1 on M_1 and M_2; in the second round, two jobs of length 1 on M_1 and M_3, and one job of length 2 on M_2. In the third round two jobs of length 2 are assigned on M_1 and M_3, and one job of length 3 on M_2. Finally, the two remaining jobs (of lengthes 2 and 3) are assigned on M_4 – these are the only migrating jobs. The resulting schedule is optimal and the minimum transition cost is 2.

Time complexity analysis: Algorithm 2 consists of four steps. In order to calculate the $n_{\ell,r}$ values in Step 1, the jobs should be sorted by processing times. Such a sorting takes $O(n \log n)$ time. The partition of job lengths into L_1, L_2 in Step 2 is clearly linear. Step 3 iterates on the rounds and in each round assigns jobs using the already sorted list. The *ForFree* structure can be implemented using a list of pointers. Since *ForFree* jobs are assigned in a non-decreasing order and since, by Observation 4.3, at most two different lengthes from L_2 are considered in each round, Step 3 takes $O(m)$ for each round and a total of $O(m(n/m))=O(n)$. In Step 4, the algorithm assigns the remaining jobs in time $O(n)$.

We conclude that the time complexity of the algorithm is $O(n \log n)$.

We note that the sorting in the first step can be performed faster under various assumptions: If the initial schedule is optimal, that is, in SPT order, and the modification does not include jobs' length modification, then the algorithm only needs to sort the jobs of each round in S_0 separately, and concatenate the resulting lists. As there are m_0 jobs in each round we get an $O(n \log m_0)$ time algorithm. Moreover, if in the initial SPT schedule the jobs are assigned sequentially on the machines, or if m_0 is a constant, then Step 1, and the whole algorithm, take $O(n)$ time.

4.3. Linear-time algorithms for some special cases

Algorithm 2 uses the minimal budget required to achieve an optimal schedule assuming unit migration cost. For several cases of m_0, m, and when the original schedule is optimal (SPT) we can calculate an optimal solution in linear time and determine the required budget in constant time. We consider the two cases of adding or removing machines, where in each case we explore several options and show the analysis of the minimal budget bound. Recall that M_0 denotes the initial set of machines and

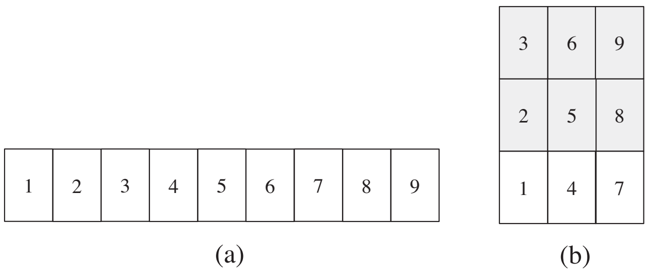

Fig. 6. (a) The initial schedule (b) An optimal modified schedule when $m'=2m_0$. Every block represents m_0 jobs.

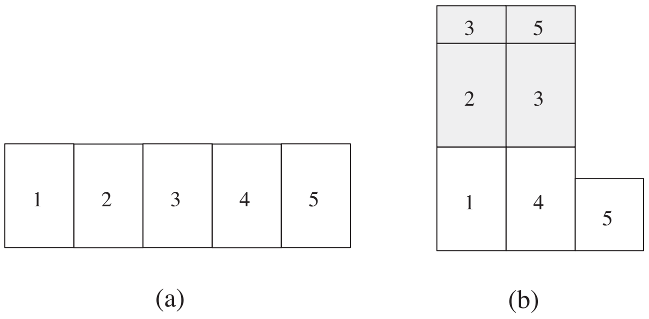

Fig. 7. (a) The initial schedule (b) An optimal modified schedule when $m_0 < m'$

$m_0 = |M_0|$, M denotes the modified set of machines, and $m = |M|$. Let M' denote the set of added / removed machines, and $m' = |M'|$. Thus, in machines' addition, $m'=m-m_0$ and in machines' removal, $m'=m_0-m$. Denote by S_0, S the original and the modified schedule, respectively. Let R_{0r}, R_r denote the r^{th} round in the initial and in the modified schedule, respectively.

4.3.1. Adding machines

The case $m_0 \le m'$: we show that the optimal schedule for this case can be achieved using only migrations to the new machines (there are no internal migrations). Specifically, we show that there exists an optimal schedule in which no job scheduled on M_0 migrates to a different machine in M_0. Thus, the minimal budget is the minimum number of jobs on the new machines in an optimal schedule, which is $m' \lfloor n/m_0 + m' \rfloor$. As S_0 is optimal (SPT), there exists an optimal schedule S where for all r, each of the jobs of R_{0r} is scheduled in S in round that is not higher than any round on which a job from R_{0r+1} is assigned.

If m' is a multiple of m_0, that is, for some integer x, $m'=xm_0$, there exists an optimal schedule where every $x+1$ rounds of S_0 unite to one round of S. Therefore, in every round R_r it is possible to assign m jobs on their original machines (see Fig. 6 for $x=2$). Specifically, assigning jobs from the first round of the $x+1$ rounds on their original machines in S. Clearly, such an assignment is optimal and as all machines $m \in M_0$ are assigned with non migrating jobs, there is no optimal schedule with less migrations.

If $m_0 \le m'$, and m' is not a multiple of m_0, it might be that jobs from one round in S_0 will end up in different rounds in S. Still, we show that there exists an optimal schedule S achieved using no internal migrations. Thus, a budget of $m' \cdot \lfloor n/m_0 + m' \rfloor$ is sufficient. Since $m_0 < m'$, the number of jobs in every round in S is more than $2m_0$, therefore, every round of S includes at least one whole round of S_0. This implies that in every round of S it is possible to assign m jobs on their original machines. Clearly, such assignment is optimal and as all machines $m \in M_0$ are assigned with non migrating jobs, there is no optimal schedule with less migrations. Such an assignment is demonstrated in Fig. 7. The second round includes jobs from R_{03},

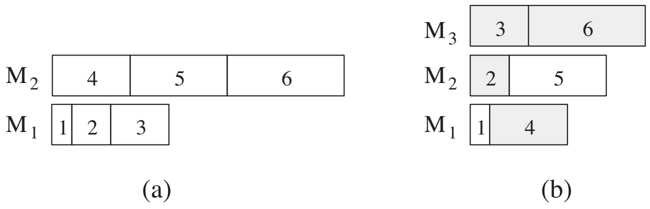

Fig. 8. (a) An initial non optimal assignment, (b) M_3 is added, an optimal reassignment requires 4 migrations.

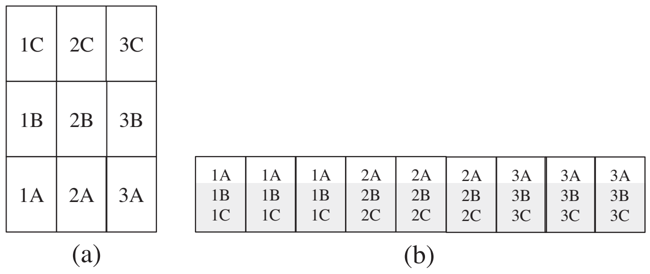

Fig. 9. (a) The initial schedule (b) An optimal modified schedule when m'=2m.

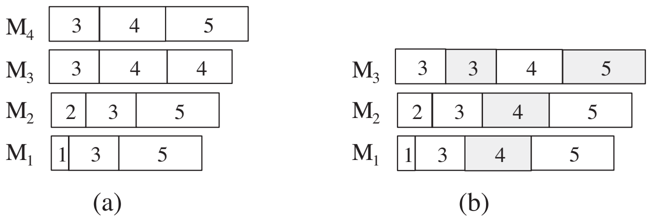

Fig. 10. (a) An initial optimal assignment, (b) M_4 is removed, an optimal reassignment requires 4 migrations.

R_{04}, R_{05} where R_{04} is fully included in R_2. It is possible to assign all the jobs in R_{04} on their original machines.

The case $m_0 > m'$**:** In this case, as demonstrated in Fig. 2 in the introduction, it might be inevitable to have internal migrations within M_0. We cannot bound the minimal required budget by $m' \cdot \lfloor n/m_0 + m' \rfloor$, and Algorithm 2 should be applied.

Finally, we note that as demonstrated in Fig. 8, if the initial schedule S_0 is not optimal or if the modification includes changes of job lengths, the bound is not valid even in the simplest case m'=1.

4.3.2. *Removing machines*

When m' **is a multiple of** m**:** Assume that for some integer x, $m = m'/x$. We show that the minimal budget required for achieving an optimal schedule in this case is the number of jobs on the removed machines. Specifically, we show that there exists an optimal schedule in which no job from $M_0 \setminus M'$ migrates.

As $m' = xm$, there exists an optimal schedule where every round of S_0 forms $x+1$ subsequent rounds of S. For all r, the m jobs of R_{0r} that are assigned on machines that are not removed might spread on $x+1$ different rounds. Still, according to the *round*-property of SPT schedules, the assignment of these jobs is flexible, and each of these jobs can remain on its machine (see Fig. 9 for $x=2$, the jobs forming the set 1A might spread along three rounds). Clearly, such an assignment is optimal and, as all machines $m \in M_0$ are assigned with non migrating jobs, there is no optimal schedule with less migrations.

When $(m' \bmod m) \neq 0$**:** In this case internal migrations within M might be inevitable. For $m > m'$, consider the example in Fig. 10. The initial schedule on $m_0 = 4$ machines is optimal. Assume that M_4 is removed. Any optional modified schedule must satisfy $n_{34} = 3$, thus one internal migration is inevitable – a job of length 4 must leave M_3.

For $m \le m'$, consider the example in Fig. 11(a) and (b). The initial schedule on $m_0 = 5$ machines is optimal. Assume that 3 machines, M_3, M_4, M_5 are removed. Any optional modified schedule must satisfy $n_{32} = 1$ and $n_{22} = 1$, thus one internal migration is inevitable – one job must leave M_1.

Thus, when $(m' \bmod m) \neq 0$ for both $m < m'$ and $m > m'$, internal migrations might be inevitable and a budget of $(m_0 - m') \cdot \lfloor n/m_0 \rfloor$ might not be sufficient.

Finally, we note that as demonstrated in Fig. 11(c) and (d), if the initial schedule S_0 is not optimal or if the modification includes changes of job lengths, the bound is not valid even in the simplest case m'=1.

5. Rescheduling with a limited budget – unit migration costs

In this section we consider the rescheduling problem assuming a limited budget. Naturally, the goal is to utilize the budget in the best possible way, that is, the modified schedule should have a low total flow-time – the minimal possible among all schedules that can be achieved using the given budget. We assume unit migration costs, that is, $\theta_{i,i',j} = 1$, independent of the job j and the involved machines. Thus, the budget B gives the maximal number of allowed migrations. Clearly, the problem is of interest only if $B < n$, as otherwise an optimal schedule can be found by ignoring the migration costs. We present an optimal algorithm for the case when the system's modification consists of machines addition and the only allowed migrations are to the new machines. This scenario arises in practice when the system is upgraded with new machines that are ready to receive tasks, while the old machines are not capable to accept new tasks.

The job-extension penalties can be arbitrary. That is, for all i, i', j, $\delta_{i,i',j} \ge 0$ gives the extension of job j if migrated from machine i to machine i'.

The algorithm is based on a reduction to a min-cost max-flow problem. An illustration of the flow network is given in Fig. 12. Each edge is labeled by its capacity and the cost of one flow unit.

An overview of the flow network: The set of nodes $v_{i,k}$ for $1 \le i \le m_0$, $1 \le k \le n$ correspond to positions on the initial machines, where $v_{i,k}$ denotes the kth-from-last position on machine i. Similarly, the set of nodes $q_{i',k}$ for $1 \le i' \le m'$, $1 \le k \le B$ correspond to positions on the added machines. All the q-nodes are connected to node d. The capacity of the edge (d, t) is the budget B. This limited capacity guarantees that the total number of slots occupied on the new machines will not exceed B. The set of nodes $1 \le j \le n$ correspond to the jobs. Every job j that is assigned to machine i in S_0 is connected to the nodes corresponding to positions on machine i and to all the q-nodes. The capacities of all edges except for (d, t) are 1. The cost of an edge connecting job j to a node corresponding to a kth from last position on the machine i on which it is assigned is kp_j, while the cost of an edge connecting job j to a node corresponding to a kth from last position on a new machine i' is $k(p_j + \delta_{i,i',j})$. All other edges have cost 0.

Theorem 5.1. *A minimum-cost maximum-flow (of value n) in G corresponds to an optimal schedule without exceeding the budget B.*

Proof. First, note that every valid schedule corresponds to a maximum-flow in G. On the other hand, not every maximum-flow in G corresponds to a schedule, since a job might be assigned to the kth from last position in some machine, while less than k jobs are assigned to that machine. However, such a maximum-flow is clearly not of minimal cost – a better matching can be obtained by shifting the $k' < k$ jobs assigned to that machine into the k' last slots. Therefore, a schedule of minimum total flow-time corresponds to

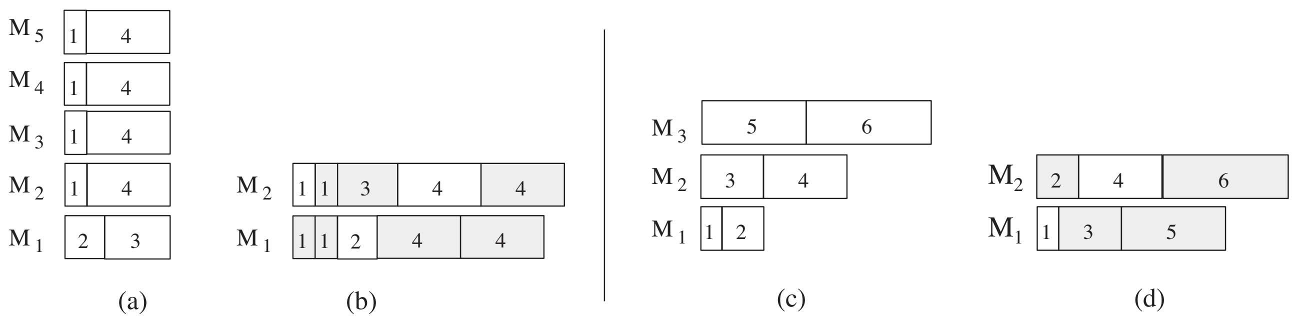

Fig. 11. (a) An initial optimal assignment, (b) M_3, M_4, M_5 are removed, an optimal reassignment requires 7 migrations. (C) An initial non-optimal assignment, (d) M_3 is removed, an optimal reassignment requires 4

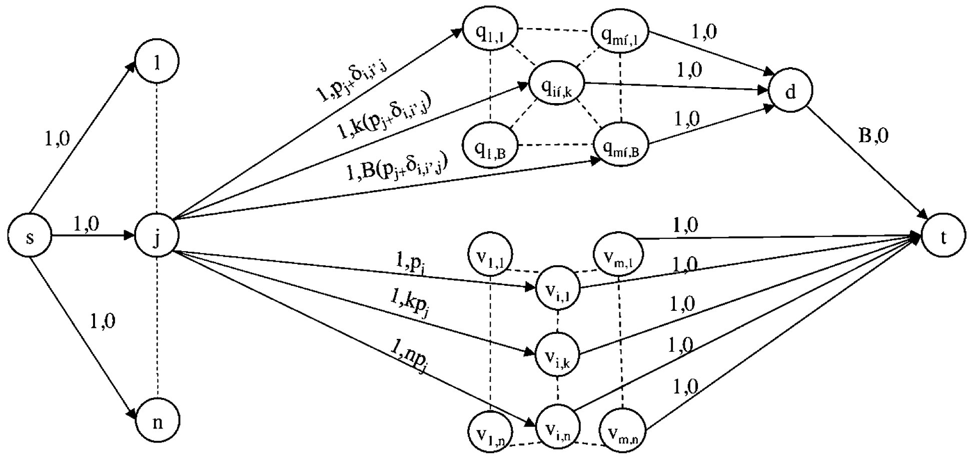

Fig. 12. The flow network built for the rescheduling with limited budget problem.

a minimum-cost maximum-flow in G. As the capacity of (d, t) is B, while all other edges' capacity is 1, at most B q-nodes have incoming flow. These nodes correspond to migrating jobs. Thus, a minimum-cost maximum-flow in G corresponds to an optimal schedule without exceeding the budget B. □

This algorithm can be extended for the case in which the systems' modification occurs at time $t > 0$ – similar to the extensions described in Section 3. If restarts are allowed, then our extension assumes that every currently processed job is the shortest job on its machine (which is true if the initial schedule is optimal, or if the schedule is a result of our algorithm – even on a sequence of modifications). If restarts are not allowed then our extension is valid for any initial schedule.

6. Conclusions and future work

We studied reoptimization problems arising in production planning, in which the goal is to combine the objective of finding a schedule with low total flow-time, with the goal of efficiently converting a given initial schedule to the modified one. We presented the first positive results in this framework. We presented algorithms for finding an optimal schedule achieved using the minimal possible transition cost, and algorithms for optimal utilization of a limited number of migrations.

Several interesting important problems remain open:

1 Our algorithm for the problem of achieving the best possible reschedule using limited budget (Section 5) assumes uniform transition costs and no internal migrations. For arbitrary instances the complexity status of the problem should be studied. Hardness results or efficient algorithms should be developed.
2 Identify the range of budget B for which it is guaranteed that an optimal reschedule can be achieved using no internal migrations. It is easy to see that this range is included in $m' < B \leq m' \cdot (n/m_0 + m')$.
3 Another open research direction is to consider different objective functions. In particular, minimizing the makespan of the schedule, given by the last completion time of some job. Since the problem is NP-hard, the reoptimization problem is clearly also NP-hard. The goal is to develop an algorithm for the reoptimization problem whose approximation-ratio is similar to the best approximation-ratio known for the original problem. For the minimum total-flow problem, we were able to reduce the reoptimization problem on identical parallel machines to the same problem on unrelated machines ($R||\sum C_j$). It seems that a similar reduction can be applied also for the minimum makespan problem. The best approximation ratio for the resulting problem ($R||C_{max}$) is $2 - 1/m$, and it is based on solving an LP problem [31]. We believe that a simpler greedy algorithm tailored for the reoptimization problem can have a similar performance. Note that the order of the jobs assigned to a specific machine is not important. Thus, some of the challenges involved in scheduling remainders of currently processed jobs as first on their machines are not relevant in this problem.

References

[1] G. Amato, G. Cattaneo, G.F. Italiano, Experimental analysis of dynamic minimum spanning tree algorithms, in: Proceedings of the 8th SODA, 1997.

[2] C. Archetti, L. Bertazzi, M.G. Speranza, Reoptimizing the 0–1 knapsack problem, Discrete Appl. Math. 158 (17) (2010) 1879–1887.
[3] G. Ausiello, V. Bonifaci, B. Escoffier, Complexity and approximation in reoptimization, in: B. Cooper, A. Sorbi (Eds.), Computability in Context: Computation and Logic in the Real World, Imperial College Press/World Scientific, 2011.
[4] G. Ausiello, B. Escoffier, J. Monnot, V.T. Paschos, Reoptimization of minimum and maximum traveling salesmans tours, J. Discrete Algorithms 7 (4) (2009) 453–463.
[5] J. Berlinskaa, M. Drozdowskib, Scheduling divisible MapReduce computations, J. Parallel Distrib. Comput. 71 (3) (2011) 450–459.
[6] H.J. Böckenhauer, L. Forlizzi, J. Hromkovič, J. Kneis, J. Kupke, G. Proietti, P. Widmayer, On the approximability of TSP on local modifications of optimally solved instances, Algorithmic Oper. Res. 2 (2) (2007) 83–93.
[7] J.L. Bruno, E.G. Coffman, R. Sethi, Scheduling independent tasks to reduce mean finishing time, Commun. ACM 17 (1974) 382–387.
[8] A.D. Birrell, B.J. Nelson, Implementing remote procedure calls, ACM Trans. Comput. Syst. 2 (1984) 39–59.
[9] R.W. Conway, W.L. Maxwell, L.W. Miller, Theory of Scheduling, Addison-Wesley (1967).
[10] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt, A. Warfield, Live migration of virtual machines, in: The 2nd Symposium on Networked Systems Design and Implementation (NSDI), 2005.
[11] C. Demetrescu, I. Finocchi, G.F. Italiano, Dynamic graph algorithms, in: J. Yellen, J.L. Gross (Eds.), Handbook of Graph Theory, CRC Press Series, in Discrete Math and Its Applications, 2003.
[12] J. Dean, S. Ghemawat, MapReduce: simplified data processing on large clusters, in: Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI), 2004, pp. 137–150.
[13] D. Eppstein, Z. Galil, G.F. Italiano, Dynamic graph algorithms, in: M.J. Atallah (Ed.), Algorithms and Theoretical Computing Handbook, CRC Press, 1999.
[14] B. Escoffier, M. Milanič, V.T. Paschos, Simple and Fast Reoptimizations for the Steiner Tree Problem. DIMACS Technical Report 2007-01, 2007.
[15] A.V. Goldberg, R.E. Tarjan, Finding minimum-cost circulations by canceling negative cycles, J. Assoc. Comput. Mach. 36 (4) (1989) 873–886.
[16] R.L. Graham, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, Optimization and approximation in deterministic sequencing and scheduling: a survey, Ann. Discrete Math. 5 (1979) 287–326.
[17] F. Grandoni, R. Zenklusen, Optimization with more than one budget, in: Proceedings of the ESA, 2010.
[18] W. Horn, Minimizing average flow-time with parallel machines, Oper. Res. 21 (1973) 846–847.
[19] S. Hacking, B. Hudzia, Improving the live migration process of large enterprise applications, in: The 3rd International Workshop on Virtualization Technologies in Distributed Computing (VTDC), 2009.
[20] J.R. Kenney, Solving unweighted and weighted bipartite matching problems in theory and practice (Ph.D. thesis), Stanford University, 1995.
[21] H.W. Kuhn, The Hungarian method for the assignment problem, Naval Res. Logist. Q. 2 (1955) 83–97.
[22] D. Mattox, Handbook of Physical Vapor Deposition (PVD) Processing, 2nd ed., Elsevier, 2010.
[23] E. Nardelli, G. Proietti, P. Widmayer, Swapping a failing edge of a single source shortest paths tree is good and fast, Algorithmica 35 (2003) 56–74.
[24] S. Pallottino, M.G. Scutella, A new algorithm for reoptimizing shortest paths when the arc costs change, Oper. Res. Lett. 31 (2003) 149–160.
[25] Parallels Virtuozzo, http://www.parallels.com/products/pvc
[26] R. Ravi, M.X. Goemans, The constrained minimum spanning tree problem, in: 5th Workshop on Algorithm Theory, 1996, pp. 66–75.
[27] H. Shachnai, G. Tamir, T. Tamir, Minimal cost reconfiguration of data placement in storage area network, Theor. Comput. Sci. 460 (2012) 42–53.
[28] H. Shachnai, G. Tamir, T. Tamir, A theory and algorithms for combinatorial reoptimization, in: Proceedings of the 10th LATIN, 2012.
[29] R.A. Sitters, Two NP-hardness results for preemptive minsum scheduling of unrelated parallel machines, in: Proceedings of the 8th IPCO, 2001, pp. 396–405.
[30] W.E. Smith, Various optimizers for single-stage production, Naval Res. Logist. Q. 3 (1956) 59–66.
[31] E.V. Shchepin, N. Vakhania, An optimal rounding gives a better approximation for scheduling unrelated machines, Oper. Res. Lett. 33 (2005) 127–133.
[32] M. Thorup, D.R. Karger, Dynamic graph algorithms with applications, in: Proceedings of the 7th SWAT, 2000.
[33] Xen Project, http://www.xenproject.org/

Guy Baram completed his MSc in the school of Computer Science at the Interdisciplinary Center (IDC) Israel, in 2013.

Tami Tamir joined the school of Computer Science at the Interdisciplinary Center (IDC) Israel, in 2004, and is serving as the school's dean since 2012. She received her PhD from the CS department at the Technion in 2001. Her research interests include design and analysis of algorithms, resource allocation problems, and algorithmic graph theory. Prior to her PhD studies, she was a member of the performance enhancement group of Intel in Haifa. After graduation, she spent two years as a lecturer and postdoctoral fellow at the University of Washington in Seattle.