Minimizing the Flow Time without Migration

Baruch Awerbuch *   Yossi Azar †   Stefano Leonardi ‡   Oded Regev §

Abstract

We consider the classical problem of scheduling jobs in a multiprocessor setting in order to minimize the flow time (total time in the system). The performance of the algorithm, both in offline and online settings, can be significantly improved if we allow preemption: i.e., interrupt a job and later continue its execution, perhaps migrating it to a different machine. Preemption is inherent to make a scheduling algorithm efficient. While in case of a single processor, most operating systems can easily handle preemptions, migrating a job to a different machine results in a huge overhead. Thus, it is not commonly used in most multiprocessor operating systems. The natural question is whether migration is an inherent component for an efficient scheduling algorithm, in either online or offline setting.

Leonardi and Raz (STOC'97) showed that the well known algorithm, *shortest remaining processing time* (SRPT), performs within a logarithmic factor of the optimal algorithm. Note that SRPT must use *both* preemption and migration to schedule the jobs. It is not known if better approximation factors can be reached. In fact, in the on-line setting, Leonardi and Raz showed that no algorithm can achieve a better bound.

Without migration, no (offline or online) approximations are known. This paper introduces a new algorithm that does not use migration, works *online*, and is just as effective (in terms of approximation ratio) as the best known *offline* algorithm (SRPT) that uses migration.

1 Introduction

Objectives. One of the most basic performance measures in multiprocessor scheduling problems is the overall time the jobs are spending in the system. This includes the delay of waiting for service as well as the actual service time. This measure captures the overall quality of service of the system. Multiprocessor scheduling problems arise, for example, in the context of server farms accommodating requests for retrieving Web contents over the Internet.

We consider the classical problem of minimizing the flow time in a multiprocessor setting with jobs which are released over time. The performance of the algorithm, both in offline and online settings, can be significantly improved if we allow preemption: i.e., interrupt a job and later continue its execution, perhaps migrating it to a different machine. As shown below, preemption is inherent to make a scheduling algorithm efficient.

While in case of a single processor, most operating systems can easily handle preemptions, migrating a job to a different machine results in a huge overhead. Thus, it is not commonly used in most multiprocessor operating systems. The natural question is whether migration is an inherent component for an efficient scheduling algorithm, in either online or offline setting.

Existing work. Surveys on approximation algorithms for scheduling can be found in [4, 6]. In the non-preemptive case it is impossible to achieve a "reasonable" approximation. Specifically, even for one machine one cannot achieve an approximation factor of $O(n^{\frac{1}{2}-\epsilon})$ unless $NP = P$ where n is the number of jobs [5]. For $m > 1$ it is impossible to achieve $O(n^{\frac{1}{3}-\epsilon})$ approximation factor unless $NP = P$ [7]. Thus, preemptions really seem to be essential.

*Johns Hopkins University, Baltimore, MD 21218, and MIT Lab. for Computer Science. E-mail: baruch@blaze.cs.jhu.edu.
Supported by Air Force Contract TNDGAFOSR-86-0078, ARPA/Army contract DABT63-93-C-0038, ARO contract DAAL03-86-K-0171, NSF contract 9114440-CCR, DARPA contract N00014-J-92-1799, and a special grant from IBM.

†Department of Computer Science, Tel Aviv University, Tel-Aviv, 69978, Israel. E-Mail: azar@math.tau.ac.il.
Research supported in part by the Israel Science Foundation and by the US-Israel Binational Science Foundation (BSF).

‡Dipartimento di Informatica Sistemistica, Università di Roma "La Sapienza", via Salaria 113, 00198-Roma, Italia.
This work was partly supported by EU ESPRIT Long term Research Project ALCOM-IT under contract n. 20244, and by Italian Ministry of Scientific Research Project 40% "Algoritmi, Modelli di Calcolo e Strutture Informative".
E-Mail: leon@dis.uniroma1.it.

§Department of Computer Science, Tel Aviv University. E-Mail: odedr@math.tau.ac.il

Existing work: single processor. Minimizing the flow time on one machine with preemption can be done optimally in polynomial time using the natural algorithm shortest remaining processing time (SRPT) [1].

Existing work: multiple processors with migration. For more than one machine the preemptive problem becomes NP-hard [2]. Only very recently, Leonardi and Raz [7] (STOC'97) showed that SRPT achieves logarithmic approximation for the multiprocessor case, showing a tight bound of $O(\log(\min\{n/m, P\}))$ on $m > 1$ machines with n jobs, where P denotes the ratio between the processing time of the longest and the shortest jobs.

Note that SRPT must use *both* preemption and migration to schedule the jobs. It is not known if better approximation factors can be reached. In fact, in the on-line setting, no algorithm can achieve a better bound [7]. For the easier problem of minimizing the total completion time a constant approximation can be achieved [3].

Our result: multiple processors without migration. Without migration, no (offline or online) approximations are known. We present a new algorithm for minimizing flow time that uses local preemption, but does not migrate jobs between machines. It can be shown to perform as well as the best known *offline* algorithm (SRPT) for the preemptive problem that uses migration. More specifically, our algorithm guarantees, on all input instances, a small performance gap in comparison to the optimal *offline* schedule that allows *both* preemption and migration. Denote by P the ratio between the processing time of the longest and the shortest jobs. Then our algorithm performs by at most $O(\min\{\log P, \log n\})$ factor of the optimal preemptive algorithm that allows migration. The algorithm can be easily implemented in polynomial time in the size of the input instance.

Our algorithm is also on-line. We note that in the proof of the $\Omega(\log P)$, $\Omega(\log n/m)$ lower bounds of [7] for online algorithms, the optimal algorithm does not use migration. Hence, the (randomized) lower bound holds also for a non-migrative algorithm. This implies that our algorithm is optimal with respect to the parameter P, i.e., no on-line algorithm can achieve a better bound both when the offline algorithm is or is not allowed to migrate jobs (the first claim is obviously stronger), while there is still a small gap between the $O(\log n)$ upper bound and the $\Omega(\log n/m)$ lower bound. A matching $O(\log n)$ lower bound for the algorithm is omitted in this extended abstract.

Unlike SRPT our algorithm may continue to run a job on a machine even when a shorter job is waiting to be processed. This seems essential in the non-migrative setting since being too eager to run a shorter job may result in an unbalanced commitment to machines. A non-migrative algorithm has to trade off between the commitment of a job to a machine and the decrease in the flow time yielded by running a shorter job. Our algorithm runs the job with the shortest remaining processing time among all the jobs that were already assigned to that machine, and a new job is assigned to a machine if its processing time is considerably shorter than the job that is currently running.

The model: We are given a set J of n jobs and a set of m identical machines. Each job j is assigned a pair (r_j, p_j) where r_j is the release time of the job and p_j is its processing time. In the preemptive model a job that is running can be preempted and continued later on any machine. Our model allows preemption but does not allow migration, i.e., a job that is running can be preempted but must later continue its execution on the same machine on which its execution began. The scheduling algorithm decides which of the jobs should be executed at each time. Clearly a machine can process at most one job in any given time and a job cannot be processed before its release time. For a given schedule define c_j to be the completion time of job j in that schedule. The flow time of job j for this schedule is $F_j = c_j - r_j$. The total flow time is $\sum_{j \in J} F_j$. The goal of the scheduling algorithm is to minimize the total flow time for each given instance of the problem. In the off-line version of the problem all the jobs are known in advance. In the on-line version of the problem each job is introduced at its release time and the algorithm bases its decision only upon the jobs that were already released.

2 The algorithm

A job is called alive at time t for a given schedule if it has already been released but has not been completed yet. Our algorithm classifies the jobs that are alive into classes according to their remaining processing times. A job j whose remaining processing time is in $[2^k, 2^{k+1})$ is in class k for $-\infty < k < \infty$. Notice that a given job changes its class during its execution. The algorithm holds a pool of jobs that are alive and have not been processed at all. In addition, the algorithm holds a stack of jobs for each of the machines. The stack of machine i holds jobs that are alive and have already been processed by machine i. The algorithm works as follows:

- Each machine processes the job at the top of its stack.
- When a new job arrives the algorithm looks for a machine that is idle or currently processing a job of a higher class than the new job. In case it finds one, the new job in pushed into that machine's stack and its processing begins. Otherwise, the job is inserted into the pool.
- When a job is completed on some machine it is popped from its stack. The algorithm compares the class of the job at the top of the stack with the minimum class of a job in the pool. If the minimum is in the pool then a job that achieves the minimum is pushed into the stack (and removed from the pool).

Clearly, when a job is assigned to a machine it will be processed only on that machine and thus the algorithm does not use migration. In fact, the algorithm bases its decisions only on the jobs that were released up to the current time and hence is an on-line algorithm. Note that it may seem that the algorithm has to keep track of all the infinite number of classes through which a job evolves. However, the algorithm recalculates the classes of jobs only at arrival or completion of a job.

3 Analysis

We denote by A our scheduling algorithm and by OPT the optimal off-line algorithm that minimizes the flow time for any given instance. For our analysis we can even assume that OPT may migrate jobs between machines. Whenever we talk about time t we mean the moment after the events of time t happened. For a given scheduling algorithm S we define $V^S(t)$ to be the volume of a schedule at a certain time t. This volume is the sum of all the remaining processing times of jobs that are alive. In addition, we define $\delta^S(t)$ to be the number of jobs that are alive. $\Delta V(t)$ is defined to be the volume difference between our algorithm and the optimal algorithm, i.e., $V^A(t) - V^{OPT}(t)$. We also define by $\Delta\delta(t) = \delta^A(t) - \delta^{OPT}(t)$ the alive jobs difference at time t between A and OPT. For a generic function f (V, ΔV, δ or $\Delta\delta$) we use $f_{\geq h, \leq k}(t)$ to denote the value of f at time t when restricted to jobs of classes between h and k. Similarly, the notation $f_{=k}(t)$ will represent the value of function f at time t when restricted to jobs of class precisely k.

Let $\gamma^S(t)$ be the number of non-idle machines at time t. Notice that because our algorithm does not migrate jobs, there are situations in which $\gamma^A(t) < m$ and $\delta^A(t) \geq m$. We denote by $\mathcal{T}$ the set of times in which $\gamma^A(t) = m$, that is, the set of times in which none of the machines is idle. Denote by P_{min} the processing time of the shortest job and by P_{max} the processing time of the longest job. Note that $P = P_{max}/P_{min}$. Denote by $k_{min} = \lfloor \log P_{min} \rfloor$ and $k_{max} = \lfloor \log P_{max} \rfloor$ the classes of the shortest and longest jobs upon their arrival.

We start by observing the simple fact that the flow time is the integral over time of the number of jobs that are alive (for example, see [7]):

Fact 3.1 *For any scheduler S,*

$$F^S = \int_t \delta^S(t)dt.$$

First we note that the algorithm preserves the following property of the stacks:

Lemma 3.2 *In each stack the jobs are ordered in a strictly increasing class order and there is at most one job whose class is at most k_{min}.*

Proof: At time $t = 0$ the lemma is true since all the stacks are empty. The lemma is proved by induction on time. The classes of jobs in the stacks change in one of three cases. The first is when the class of the currently processed job decreases. Since the currently processed job is the job with the lowest class, the lemma remains true. The second case is when a new job arrives. In case it enters the pool there is no change in any stack. Otherwise it is pushed into a stack whose top is of a higher class which preserves the first part of the lemma. Since the class of the new job is at least k_{min} the second part of the lemma remains true. The third case is when a job is completed on some machine. If no job is pushed into the stack of that machine the lemma remains easily true. If a new job is pushed to the stack then the lemma remains true in much the same way as in the case of the arrival of a new job. ■

Corollary 3.3 *There are at most $2 + \log P$ jobs in each stack.*

Proof: The number of classes of jobs in each stack is at most $k_{max} - k_{min} + 1 \leq 2 + \log P$. ■

We look at the state of the schedule at a certain time t. First let's look at $t \notin \mathcal{T}$:

Lemma 3.4 *For $t \notin \mathcal{T}$, $\delta^A(t) \leq \gamma^A(t)(2 + \log P)$.*

Proof: By definition of $\mathcal{T}$, at time t at least one machine is idle. That implies that the pool is empty. Moreover, all the stacks of the idle machines are obviously empty. So, all the jobs that are alive are in the stacks of the non-idle machines. The number of non-idle machines is $\gamma^A(t)$ and the number of jobs in each stack is at most $2 + \log P$ according to corollary 3.3 ■

Now, assume that $t \in \mathcal{T}$ and let $t_0 < t$ be the earliest time for which $[t_0, t) \subset \mathcal{T}$. *We denote the last time in which a job of class more than k was processed by t_k.* In case such jobs were not processed at all in the time interval $[t_0, t)$ we set $t_k = t_0$.

Lemma 3.5 *For $t \in \mathcal{T}$, $\Delta V_{\leq k}(t) \leq \Delta V_{\leq k}(t_k)$.*

Proof: Notice that in the time interval $[t_k, t)$, algorithm A is constantly processing on all the machines jobs whose class is at most k. The off-line algorithm may process jobs of higher classes. Moreover, that can cause jobs of class more than k to actually lower their classes to k and below therefore adding even more to $V^{OPT}_{\leq k}(t)$. Finally, the release of jobs of class $\leq k$ in the interval $[t_k, t)$ is not affecting $\Delta V_{\leq k}(t)$. Therefore, the difference in volume between the two algorithms cannot increase. ■

Lemma 3.6 *For $t \in \mathcal{T}$, $\Delta V_{\leq k}(t_k) \leq m2^{k+2}$.*

Proof: First we claim that at any moment $t_k - \epsilon$, for any $\epsilon > 0$ small enough, the pool does not contain jobs whose class is at most k. In case $t_k = t_0$, at any moment just before t_k there is at least one idle machine which means the pool is empty. Otherwise, $t_k > t_0$ and by definition we know that a job of class more than k is processed just before t_k. Therefore, the pool does not contain any job whose class is at most k.

At time t_k jobs of class at most k might arrive and fill the pool. However, those jobs increase both $V^{OPT}_{\leq k}(t_k)$ and $V^{A}_{\leq k}(t_k)$ by the same amount, so jobs that arrive exactly at t_k do not change $\Delta V_{\leq k}(t_k)$ and can be ignored.

Since the jobs in the pool at time t_k can be ignored, we are left with the jobs in the stacks. Using lemma 3.2, $\Delta V_{\leq k}(t_k) \leq m(2^{k+1} + 2^k + 2^{k-1} + \ldots) \leq m2^{k+2}$. ■

Lemma 3.7 *For $t \in \mathcal{T}$, $\Delta V_{\leq k}(t) \leq m2^{k+2}$.*

Proof: Combining Lemma 3.5 and 3.6, we obtain $\Delta V_{\leq k}(t) \leq \Delta V_{\leq k}(t_k) \leq m2^{k+2}$ ■

The claim of the following Lemma states a property that will be used in the proof of both the $O(\log P)$ and the $O(\log n)$ approximation result.

Lemma 3.8 *For $t \in \mathcal{T}$, for $k_{min} \leq k_1 \leq k_2 \leq k_{max}$, $\delta^{A}_{\geq k_1, \leq k_2}(t) \leq 2m(k_2 - k_1 + 2) + 2\delta^{OPT}_{\leq k_2}(t)$.*

Proof: $\delta^{A}_{\geq k_1, \leq k_2}(t)$ can be expressed as:

$$
\begin{aligned}
& \sum_{i=k_1}^{k_2} \delta^{A}_{=i}(t) \\
\leq \quad & \sum_{i=k_1}^{k_2} \left(\frac{\Delta V_{=i}(t) + V^{OPT}_{=i}(t)}{2^i}\right) \\
\leq \quad & \sum_{i=k_1}^{k_2} \frac{\Delta V_{\leq i}(t) - \Delta V_{\leq i-1}(t)}{2^i} + 2\delta^{OPT}_{\geq k_1, \leq k_2}(t) \\
\leq \quad & \frac{\Delta V_{\leq k_2}(t)}{2^{k_2}} + \sum_{i=k_1}^{k_2-1} \frac{\Delta V_{\leq i}(t)}{2^{i+1}} \\
& - \frac{\Delta V_{\leq k_1 - 1}(t)}{2^{k_1}} + 2\delta^{OPT}_{\geq k_1, \leq k_2}(t) \\
\leq \quad & 4m + \sum_{i=k_1}^{k_2-1} 2m + \delta^{OPT}_{\leq k_1 - 1}(t) + 2\delta^{OPT}_{\geq k_1, \leq k_2}(t) \\
\leq \quad & 2m(k_2 - k_1 + 2) + 2\delta^{OPT}_{\leq k_2}(t).
\end{aligned}
$$

The first inequality follows since 2^i is the minimum processing time of a job of class i. The second inequality follows since the processing time of a job of class i is less than 2^{i+1}. The fourth inequality is derived by applying Lemma 3.7, observing that $\Delta V_{\leq k_1 - 1}(t) \geq -V^{OPT}_{\leq k_1 - 1}(t)$ and that 2^{k_1} is the maximum processing time of a job of class at most $k_1 - 1$. The claim of the lemma then follows. ■

The following corollary of Lemma 3.8 is used in the proof of the $O(\log P)$ approximation ratio of Theorem 3.10

Corollary 3.9 *For $t \in \mathcal{T}$, $\delta^A(t) \leq 2m(4 + \log P) + 2\delta^{OPT}(t)$.*

Proof: We express

$$
\begin{aligned}
\delta^A(t) &= \delta^{A}_{\leq k_{max}, \geq k_{min}}(t) + \delta^{A}_{< k_{min}}(t) \\
&\leq 2m(k_{max} - k_{min} + 2) + 2\delta^{OPT}(t) + m \\
&\leq 2m(4 + \log P) + 2\delta^{OPT}(t)
\end{aligned}
$$

The second inequality follows from the claim of Lemma 3.8 when $k_2 = k_{max}$ and $k_1 = k_{min}$, and from the claim of Lemma 3.2 stating that the stack of each machine contains at most one job of class less than k_{min}. The third inequality is obtained since $k_{max} - k_{min} + 5/2 \leq \log P + 4$. ■

Theorem 3.10 $F^A \leq 2(5 + \log P) \cdot F^{OPT}$*, that is, algorithm A has a $2(5 + \log P)$ approximation factor even compared to the optimal off-line algorithm that is allowed to migrate jobs.*

Proof:

$$
\begin{aligned}
F^A &= \int_t \delta^A(t)dt \\
&= \int_{t \notin \mathcal{T}} \delta^A(t)dt + \int_{t \in \mathcal{T}} \delta^A(t)dt \\
&\leq \int_{t \notin \mathcal{T}} \gamma^A(t)(2 + \log P)dt \\
&\quad + \int_{t \in \mathcal{T}} (2m(4 + \log P) + 2\delta^{OPT}(t))dt \\
&\leq (2 + \log P) \int_{t \notin \mathcal{T}} \gamma^A(t)dt \\
&\quad + 2(4 + \log P) \int_{t \in \mathcal{T}} m dt + 2 \int_{t \in \mathcal{T}} \delta^{OPT}(t)dt \\
&\leq (8 + 2\log P) \int_t \gamma^A(t)dt + 2 \int_t \delta^{OPT}(t)dt \\
&\leq 2(5 + \log P) \cdot F^{OPT}
\end{aligned}
$$

The first equality is from the definition of F^A. The second is obtained by looking at the time in which none of the machines is idle and the time in which at least one machine is idle separately. The third inequality uses Lemma 3.4 and Corollary 3.9. The fifth inequality is true since $\gamma^A(t) = m$ when $t \in \mathcal{T}$. Finally, $\int_t \gamma^A(t)dt$ is the total time spent processing jobs by the machines which is exactly $\sum_{j \in J} p_j$. That sum is upper bounded by the flow time of OPT since each job's flow time must be at least its processing time. ■

We now turn to prove the $O(\log n)$ approximation ratio of the algorithm. A different argument is required to prove this second bound. The main idea behind the proof of the $O(\log P)$ approximation ratio was to bound for any time $t \in \mathcal{T}$, the alive jobs difference between A and OPT by $O(m \log P)$. A similar approach does not allow to prove the $O(\log n)$ approximation ratio: It is possible to construct instances where the the alive jobs difference is $\Omega(n)$.

Leonardi and Raz [7] proved the $O(\log n/m)$ approximation ratio for SRPT when migration is allowed arguing that the worst case ratio between the SRPT flow time and the optimal flow time can be raised only if a "big" alive jobs difference is kept for a "long" time period. This observation yields also for our non-migrative algorithm. This is formally stated in Lemma 3.11 for any time $t \in \mathcal{T}$ when no machine is idle and in Lemma 3.12 for any time $t \notin \mathcal{T}$. These Lemmas prove that the minimum remaining processing time of a set of unfinished jobs is exponentially decreasing with the size of the alive jobs difference between A and OPT. Thus, either new jobs are released at a rate exponential in the size of the alive jobs difference, or the ongoing processed jobs are consumed and the alive jobs difference is reduced.

We need to introduce more notation. Recall that $\mathcal{T}$ is defined to be the set of times in which $\gamma^A(t) = m$. We denote by $T = \int_{t \in \mathcal{T}} dt$ the size of set $\mathcal{T}$. For any $t \in \mathcal{T}$, define by $\delta^{A,P}(t)$ the number of jobs in the pool of algorithm A, i.e., not assigned to a machine, at time t, and by $\Delta\delta^P(t) = \delta^{A,P}(t) - 2\delta^{OPT}(t)$ the difference between the number of jobs in the pool of algorithm A and twice the number of jobs not finished by the optimal algorithm. For any machine l, time t, define by $\delta^{A,l}(t)$ the number of jobs assigned to machine l at time t in the schedule of algorithm A. Moreover, define by $\mathcal{T}^l = \{t | \delta^{A,l}(t) > 0\}$, the set of times when machine l is assigned with at least one job, and by $T^l = \int_{t \in \mathcal{T}^l} dt$ the size of set $\mathcal{T}^l$.

Lemma 3.11 *For any time $t \in \mathcal{T}$, if $\Delta\delta^P(t) \geq 2mi$, for $i \geq 3$, then the pool of algorithm A contains at least $2m$ jobs of remaining processing time at most $\frac{V^A(t)}{m2^{i-3}}$.*

Proof: Let $\overline{k}$ be the maximum integer such that $\delta^{A,P}_{\geq \overline{k}}(t) \geq 2m$. Then,

$$2m \leq \delta^{A,P}_{\geq \overline{k}}(t) \leq \delta^{A}_{\geq \overline{k}}(t) \leq \frac{V^A(t)}{2^{\overline{k}}}$$

thus yielding $2^{\overline{k}} \leq \frac{V^A(t)}{2m}$. In particular, the last inequality follows since $2^{\overline{k}}$ is the minimum processing time of a job of class $\overline{k}$.

By the definition of $\overline{k}$, we have:

$$\begin{aligned} \Delta\delta^P_{\leq \overline{k}}(t) &= \delta^{A,P}_{\leq \overline{k}}(t) - 2\delta^{OPT}_{\leq \overline{k}}(t) \\ &= \delta^{A,P}(t) - \delta^{A,P}_{> \overline{k}}(t) - 2(\delta^{OPT}(t) - \delta^{OPT}_{> \overline{k}}(t)) \\ &= \Delta\delta^P(t) - \delta^{A,P}_{> \overline{k}}(t) + 2\delta^{OPT}_{> \overline{k}}(t) \\ &\geq 2m(i-1). \end{aligned}$$

where the last inequality follows since $\delta^{A,P}_{> \overline{k}}(t) < 2m$.

From Lemma 3.8, for any integer $k' \leq \overline{k}$, we get:

$$\begin{aligned} \Delta\delta^P_{\leq \overline{k}}(t) &= \delta^{A,P}_{\leq \overline{k}, \geq k'}(t) + \delta^{A,P}_{< k'}(t) - 2\delta^{OPT}_{\leq \overline{k}}(t) \\ &\leq \delta^{A}_{\leq \overline{k}, \geq k'}(t) + \delta^{A,P}_{< k'}(t) - 2\delta^{OPT}_{\leq \overline{k}}(t) \\ &\leq \delta^{A,P}_{< k'}(t) + 2m(\overline{k} - k' + 2), \end{aligned}$$

thus yielding $\delta^{A,P}_{< k'}(t) \geq 2m(i + k' - \overline{k} - 3)$. Consider now the maximum integer k' such that $\delta^{A,P}_{< k'}(t) < 2m$. For such k' we get $2m > 2m\ (i + k' - \overline{k} - 3)$ and thus $k' \leq \overline{k} - i + 3$. It follows that there exist at least 2m jobs of class at most $k' \leq \overline{k} - i + 3$ in the pool of the algorithm. The remaining processing time of those $2m$ jobs is bounded by $2^{\overline{k} - i + 4} \leq \frac{V^A(t)}{m2^{i-3}}$, thus proving the claim. ■

Lemma 3.12 *For any machine l, time $t \in \mathcal{T}^l$, if $\delta^{A,l}(t) \geq i$ for $i \geq 1$ then there exists a job with remaining processing time at most $\frac{T^l}{2^{i-2}}$ assigned to machine l at time t.*

Proof: For any time $t \in \mathcal{T}^l$, there is at most one job assigned to machine l for every specific class. Assume machine l is assigned with a job of highest class k, obviously satisfying $T^l \geq 2^k$. If $\delta^{A,l}(t) \geq i$ then there is a job of class at most $k - i + 1$ assigned to machine l, with processing time at most $2^{k-i+2} \leq \frac{T^l}{2^{i-2}}$. ■

We partition the set of time instants $\mathcal{T}$ when no machine is idle into a collection of disjoint intervals $I_k = [t_k, r_k)$, $k = 1, \ldots, s$, and associate an integer $i_k \geq 1$ to each interval, such that for any time $t \in I_k$, $2m(i_k - 1) < \Delta\delta^P(t) < 2m(i_k + 1)$ for $i_k > 1$ and $\Delta\delta^P(t) < 2m(i_k + 1)$ for $i_k = 1$.

Each maximal interval of times $[t_b, t_e)$ contained in $\mathcal{T}$ is dealt with separately. Assume we already dealt with all times in $\mathcal{T}$ which are smaller than t_b, and we have created $k-1$ intervals. We then define $t_k = t_b$ with $i_k = 1$. Given t_k and i_k, define $r_k = \{min\{t, t_e\} | t > t_k, \Delta\delta^P(t) \geq 2m(i_k+1)$ or $(i_k > 1, \Delta\delta^P(t) \leq 2m(i_k-1))\}$, that is, r_k is the first time $\Delta\delta^P(t)$ reaches the value $2m(i_k+1)$, or the value $2m(i_k-1)$ if $i_k > 1$. In case $r_k < t_e$, we set $t_{k+1} = r_k$ and $i_{k+1} = \lfloor \frac{\Delta\delta^P(t_{k+1})}{2m} \rfloor$ if $\Delta\delta^P(t_{k+1}) \geq 2m(i_k+1)$, $i_{k+1} = \lceil \frac{\Delta\delta^P(t_{k+1})}{2m} \rceil$ if $\Delta\delta^P(t_{k+1}) \leq 2m(i_k-1)$. It is straightforward from this definition that indeed for any time $t \in I_k$, $2m(i_k-1) < \Delta\delta^P(t) < 2m(i_k+1)$ for $i_k > 1$ and $\Delta\delta^P(t) < 2m(i_k+1) = 4m$ for $i_k = 1$.

Denote by $x_k = r_k - t_k$ the size of interval I_k, and define $\mathcal{T}_i = \{\cup I_k | i_k = i\}$, $i \geq 1$, as the union of the intervals I_k with $i_k = i$. We indicate by $T_i = \int_{t\in\mathcal{T}_i} dt$ the size of set $\mathcal{T}_i$. We also denote by $D = \max\{T, \max_{t\in\mathcal{T}}\{V^A(t)/m\}\}$. The following lemma relates the number of jobs, n, and the values of T_i.

Lemma 3.13 *The following lower bound holds for the number of jobs:*

$$n \geq \frac{m}{8D} \sum_{i\geq 4} T_i \, 2^{i-4}.$$

Proof: Consider the generic interval I_k, with the corresponding i_k. The interval starts when $\Delta\delta^P(t_k)$ reaches $2mi_k$, from above or from below. This interval ends when $\Delta\delta^P(r_k)$ reaches $2m(i_k+1)$ or $2m(i_k-1)$. In the first case we have the evidence of $n_k = m$ jobs finished by OPT (recall that $\Delta\delta^P(t) = \delta^P(t) - 2\delta^{OPT}(t)$). In the second case we have the evidence of $n_k = 2m$ jobs that either leave the pool to be assigned to a machine by algorithm A or arrive to both A and OPT. In both cases we charge $n_k \geq m$ jobs to interval I_k. We can then conclude with a first lower bound $n_k \geq m$ on the number of jobs charged to any interval $I_k \in \mathcal{T}_i, i_k \geq 4$.

Next, we give a second lower bound, based on Lemma 3.11, stating that during an interval $I_k = [t_k, r_k)$ there exist in the pool $2m$ jobs with remaining processing time at most $\frac{D}{2^{i_k-4}}$, since $\Delta\delta^P(t) > 2m(i_k-1)$ for any $t \in [t_k, r_k)$. This implies that all the m machines are processing jobs with remaining processing time at most $\frac{D}{2^{i_k-4}}$. We look at any subinterval of I_k of length $\frac{D}{2^{i_k-4}}$. For each machine, during this subinterval, either a job is finished by the algorithm or a job is preempted by a job of lower class. Therefore, we can charge at least m jobs that are either released or finished with any subinterval of size $\frac{D}{2^{i_k-4}}$ of I_k. A second lower bound on the number of jobs charged to any interval is then given by $n_k \geq m\lfloor\frac{x_k 2^{i_k-4}}{D}\rfloor$.

Observe now that each job is charged at most 4 times, when it is released, when it is assigned to a machine by A, when it is finished by A and when it is finished by OPT. Then,

$$\frac{m}{2D}\sum_{i\geq 4} T_i \, 2^{i-4} \leq \sum_{k|i_k\geq 4} m \max\{1, \lfloor\frac{x_k 2^{i_k-4}}{D}\rfloor\} \leq 4n,$$

where the first inequality is obtained by summing over I_k's instead of over T_i's and the simple fact that $\alpha \leq \max\{1, \lfloor 2\alpha \rfloor\}$ and the second inequality follows from the lower bounds we have shown on the charged jobs. The lemma easily follows. ∎

We next bound the number of jobs that are assigned to a machine l during the time instants of $\mathcal{T}^l$. We partition the set of time instants $\mathcal{T}^l$ into a set of disjoint intervals $I_k^l = [t_k^l, r_k^l)$, $k = 1, \ldots, s^l$, and associate an integer i_k^l to each interval, such that for any time $t \in I_k^l$, $\delta^{A,l}(t) = i_k^l$. Consider a maximal interval of times $[t_b^l, t_e^l)$ contained in $\mathcal{T}^l$. Assume $t_k^l = t_b^l$. Given t_k^l and i_k^l, define $r_k^l = \{min\{t, t_e^l\} | t > t_k^l, \delta^{A,l}(t) \neq i_k^l\}$. In case $r_k^l < t_e^l$, we set $t_{k+1}^l = r_k^l$. Denote by $x_k^l = r_k^l - t_k^l$ the size of interval I_k^l. Define by $\mathcal{T}_i^l = \{\cup I_k^l | i_k^l = i\}$, $i \geq 1$, the union of the intervals I_k^l when the number of jobs assigned to machine l is exactly i, and by $T_i^l = \int_{t\in\mathcal{T}_i^l} dt$ the size of set $\mathcal{T}_i^l$.

Lemma 3.14 *The following lower bound holds for the number of jobs assigned to machine l:*

$$n^l \geq \frac{1}{4T^l}\sum_{i\geq 1} T_i^l \, 2^{i-2}.$$

Proof: We will proceed charging at least one job with every interval I_k^l. Every job will be charged at most twice, when it is assigned to a machine by A and when it is finished by A.

Consider the generic interval I_k^l, with the corresponding i_k^l. The interval starts when $\delta^{A,l}(t_k)$ reaches i_k^l, from above or from below. The interval ends when $\delta^{A,l}(r_k)$ reaches $i_k^l + 1$ or $i_k^l - 1$. In the first case we have the evidence of one job that is assigned to a machine by A and in the second case of one job that is finished by A. In both cases we charge one job to interval I_k^l.

Next we give a second lower bound, based on Lemma 3.12. Lemma 3.12 states that during an interval $I_k^l = [t_k^l, r_k^l)$, machine l is constantly assigned with a job of remaining processing time at most $\frac{T^l}{2^{i_k^l-2}}$. We look at any subinterval of I_k^l of length $\frac{T^l}{2^{i_k^l-2}}$. During this subinterval, either a job is finished by the algorithm or a job is preempted by a job of a lower class. In any case, a job that is assigned or finished during any subinterval of size $\frac{T^l}{2^{i_k^l-2}}$ is charged. A second lower bound on the number of jobs charged to any interval is then given by $n_k \geq \lfloor\frac{x_k^l 2^{i_k^l-2}}{T^l}\rfloor$.

Observe now that each job is considered at most twice, when it is assigned to machine l and when it is finished by A. Then, from the following inequalities:

$$\frac{1}{2T^l}\sum_{i\geq 1} T_i^l\ 2^{i-2} \leq \sum_{k|i_k^l\geq 1} \max\{1, \lfloor\frac{x_k^l 2^{i_k^l-2}}{T^l}\rfloor\} \leq 2n^l,$$

the claim follows. ■

Theorem 3.15 *$F^A = O(\log n)F^{OPT}$, that is algorithm A has an $O(\log n)$ approximation ratio even compared with the optimal off-line algorithm that is allowed to migrate jobs.*

Proof:

$$\begin{aligned}
F^A &= \int_t \delta^A(t)dt \\
&= \sum_{l=1}^{m}\int_t \delta^{A,l}(t)dt + \int_{t\in\mathcal{T}} \delta^{A,P}(t)dt \\
&= \sum_{l=1}^{m}\int_{t\in\mathcal{T}^l} \delta^{A,l}(t)dt \\
&\quad + \int_{t\in\mathcal{T}} (2\delta^{OPT}(t) + \Delta\delta^P(t))dt \\
&\leq \sum_{l=1}^{m}\int_{t\in\mathcal{T}^l} \delta^{A,l}(t)dt + 2F^{OPT} \\
&\quad + \sum_{i\geq 1}\int_{t\in\mathcal{T}_i} 2m(i+1)dt \\
&\leq \sum_{l=1}^{m}\sum_{i\geq 1} i\,T_i^l + 2F^{OPT} + 2m\sum_{i\geq 1}(i+1)\,T_i \\
&= \sum_{l=1}^{m}\sum_{i\geq 1} i\,T_i^l + 2F^{OPT} + 2m\sum_{i\geq 1}(i-3)\,T_i \\
&\quad +2m\sum_{i\geq 1} 4\,T_i \\
&\leq \sum_{l=1}^{m}\sum_{i\geq 1} i\,T_i^l + 2F^{OPT} + 2m\sum_{i\geq 4}(i-3)\,T_i \\
&\quad +8F^{OPT} \\
&\leq \sum_{l=1}^{m}\sum_{i\geq 1} i\,T_i^l + 10F^{OPT} + 2m\sum_{i\geq 1} i\,T_{i+3}
\end{aligned}$$

The second equality is obtained by separately considering for any machine l the contribution to the flow time due to the jobs assigned to machine l, and the contribution to the flow due to jobs in the pool. The fourth inequality is obtained by partitioning $\mathcal{T}$ into the $\mathcal{T}_i$'s, $i \geq 1$, such that at any time $t \in \mathcal{T}_i$, $\Delta\delta^P(t) < 2m(i+1)$. The seventh inequality is obtained by observing that $m\sum_{i\geq 1} T_i \leq \sum_j p_j \leq F^{OPT}$, since all machines are busy processing jobs at any time $t \in \mathcal{T}$.

To complete the proof of the $O(\log n)$ approximation ratio, we will show that:

1. $F(n) = 2m\sum_{i\geq 1} i\,T_{i+3} = O(mD\log\frac{n}{m})$. We give an upper bound to $F(n)$ under the constraint $\sum_{i\geq 1} T_{i+3} \leq D$, and the constraint on n given by Lemma 3.13.

   $$\begin{aligned}
   \max_{\{T_1,\ldots\}} F(n) &= 2m\textstyle\sum_{i\geq 1} i\,T_{i+3} \\
   n &\geq \tfrac{m}{8D}\textstyle\sum_{i\geq 1} T_{i+3}\ 2^{i-1}; \\
   D &\geq \textstyle\sum_{i\geq 1} T_{i+3}.
   \end{aligned}$$

   We rewrite the problem using variables $Y_i = \sum_{j\geq i} T_{j+3}$, $i \geq 1$:

   $$\begin{aligned}
   \max_{\{Y_1,\ldots\}} F(n) &= 2m\textstyle\sum_{i\geq 1} Y_i \\
   n &\geq \tfrac{m}{8D}\textstyle\sum_{i\geq 1}(Y_i - Y_{i+1})\ 2^{i-1}; \\
   D &\geq Y_1 \geq Y_2 \geq \ldots.
   \end{aligned}$$

   A relaxation of the second constraint can be rewritten as
   $$n \geq \frac{1}{2}\frac{m}{8D}(\sum_{i\geq 1} Y_i\ 2^{i-1}),$$
   and then the function is upper bounded by assigning $D = Y_1 = \ldots = Y_k$, and $0 = Y_{k+1} = Y_{k+2} = \ldots$ where k is the minimum integer such that the constraint is tight or violated, namely it is the minimum integer such that $\frac{mD}{16D}(\sum_{i=1}^{k} 2^{i-1}) \geq n$.
   We then obtain for k the value $k \leq 5 + \log\frac{n}{m}$, and then $F(n) = O(mD\log\frac{n}{m})$.

2. For any machine l, $F^l(n^l) = \sum_{i\geq 1} iT_i^l = O(T^l\log n^l)$. We give an upper bound to $F^l(n^l)$ under the constraint $\sum_{i\geq 1} T_i^l \leq T^l$, and the constraint on n^l given by Lemma 3.14.

   $$\begin{aligned}
   \max_{\{T_1,\ldots\}} F^l(n^l) &= \textstyle\sum_{i\geq 1} iT_i^l \\
   n^l &\geq \tfrac{1}{4T^l}\textstyle\sum_{i\geq 1} T_i^l\ 2^{i-2}. \\
   T^l &\geq \textstyle\sum_{i\geq 1} T_i^l
   \end{aligned}$$

   We rewrite the problem using variables $Y_i = \sum_{j\geq i} T_j$, $i \geq 1$:

   $$\begin{aligned}
   \max_{\{Y_1,\ldots\}} F^l(n^l) &= \textstyle\sum_{i\geq 1} Y_i \\
   n^l &\geq \tfrac{1}{4T^l}\textstyle\sum_{i\geq 1}(Y_i - Y_{i+1})\ 2^{i-2}; \\
   T^l &\geq Y_1 \geq Y_2 \geq \ldots.
   \end{aligned}$$

   A relaxation of the second constraint can be rewritten as
   $$n^l \geq \frac{1}{4}\frac{1}{4T^l}(\sum_{i\geq 1} Y_i\ 2^{i-1}),$$

and then the function is upper bounded by assigning $T^l = Y_1 = \ldots = Y_k$, and $0 = Y_{k+1} = Y_{k+2} = \ldots$ where k is the minimum integer such that the constraint is tight or violated, namely it is the minimum integer such that $\frac{T^l}{16T^l}(\sum_{i=1}^{k} 2^{i-1}) \geq n^l$.

We then obtain for k the value $k \leq 5 + \log n^l$, and then $F^l(n^l) = O(T^l \log n^l)$.

We finally express the total flow time of algorithm A as:

$$\begin{aligned} F^A &\leq \sum_{l=1}^{m} \sum_{i \geq 1} i T_i^l + 10F^{OPT} + 2m \sum_{i \geq 1} i\, T_{i+3} \\ &\leq \sum_{l=1}^{m} F^l(n^l) + 10F^{OPT} + F(n) \\ &= \sum_{l=1}^{m} O(T^l \log n^l) + 10F^{OPT} + O(mD \log \frac{n}{m}) \\ &= O(\log n) \sum_{l=1}^{m} T^l + 10F^{OPT} + O(\log \frac{n}{m}) F^{OPT} \\ &= O(\log n) F^{OPT}. \end{aligned}$$

The fourth equality follows since $mD \leq \max\{mT, \max_{t \in \mathcal{T}}\{V^A(t)\}\} \leq \sum_{j \in J} p_j \leq F^{OPT}$. The fifth equality follows since $\sum_{l=1}^{m} T^l = \sum_j p_j \leq F^{OPT}$. ■

References

[1] K.R. Baker. *Introduction to Sequencing and Scheduling*. Wiley, 1974.

[2] J. Du, J. Y. T. Leung, and G. H. Young. Minimizing mean flow time with release time constraint. *Theoretical Computer Science*, 75(3):347–355, 1990.

[3] L. Hall, D. Shmoys, and J. Wein. Scheduling to minimize average completion time: Off-line and on-line algorithms. In *Proc. of 7th ACM-SIAM Symposium on Discrete Algorithms*, pages 142–151, 1996.

[4] L.A. Hall. Approximation algorithms for scheduling. In D.S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, pages 1–45. PWS publishing company, 1997.

[5] H. Kellerer, T. Tautenhahn and G.J. Woeginger. Approximability and nonapproximability results for minimizing total flow time on a single machine. *Proc. of the 28th Annual ACM Symposium on the Theory of Computing*, pp. 418-426, 1996.

[6] E.L. Lawler, J.K Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys. Sequencing and scheduling: algorithms and complexity. In *Handbooks in operations research and management science*, volume 4, pages 445–522. North Holland, 1993.

[7] S. Leonardi and D. Raz. Approximating total flow time on parallel machines. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 110–119, El Paso, Texas, 1997.