Scheduling bi-colored chains
Nicolas Vidal

To cite this version:
Nicolas Vidal. Scheduling bi-colored chains. [Internship report] Ecole Normale Supérieure de Lyon - ENS LYON. 2018. hal-01944993

HAL Id: hal-01944993
https://inria.hal.science/hal-01944993
Submitted on 5 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
1 Introduction

Until now, the performance of a supercomputer was mainly measured by its computational power. However, as platforms grow larger and the amount of data involved increases, we encounter new issues. On large scale platform, I/O movement is critical, and recent benchmarks started to measure data movement. Algorithm design is shifting focus from raw computational power to handle the bottleneck due to data management.

To tackle this problem, some approach aim at reducing the amount of data by compressing or preprocessing it. The solution we explore adopts a very different point of view and can be complementary to these approaches. We aim at managing the I/O data in the system.

The current I/O behavior of HPC machines is “best effort”. This strategy does not anticipate congestion, creating slowdown in the performance of applications even though additional enhancements, burst-buffers are added to I/O nodes.

Observations show that some HPC applications periodically alternate between (i) operations (computations, local data-accesses) executed on the compute nodes, and (ii) I/O transfers of data and this behavior can be predicted beforehand. Taking this structural argument, along with HPC-specific applications facts (there are in general very few applications running concurrently on a machine, and the applications run for many iterations with similar behavior) the goal is to design new algorithms for I/O scheduling. The novelty of this class of algorithms for I/O management is that we intend them to be computed statically.

To design a static algorithm, we emphasize on the periodic behavior of most applications. Scheduling the I/O volume of the different applications is repeated over time. This is critical since often the number of application runs is very high.

In the following report, we try do develop a formal background for periodic scheduling approach. First, we define a model, bi-colored chain schedul-
ing, then we go through related results existing in the literature and explore the complexity of this problem variants. Finally, to match the HPC context, we study periodical sub-problems.

2 Model

Our aim is to schedule different applications on an HPC platform with a focus on limiting the congestion. Consider that a given application alternates computation phases and I/O phases. In a HPC platforms, there are specifically dedicated nodes for I/O. Computation and data handling are managed by different nodes. We can therefore model an application as a chain of dependent tasks, alternatively being computation and I/O operations. We call this problem bi-colored chain scheduling.

2.1 Bi-colored chains

For m two-stage jobs $J_1, ..., J_m$ to be processed on machines A and B, verifying the properties below:

- each job consists of a chain of alternating A-operations and B-operations;
- There are no precedence constraints among different jobs;
- Two machines, A and B, of bandwidth $\delta_a$ and $\delta_b$ can run in parallel and can process respectively the A-operations and the B-operations of the job.
- An A-operation (resp. B-operation) cannot start unless the B-operation (resp. A-operation) immediately preceding it on the chain is completed;
- Preemption is not allowed.

We can remark that when all $n_i = 1$, this problem is a 2-Flowshop [18].

The intuition beyond this model is that A-tasks represent computation and B-tasks data management. ($\delta_a$ is the computational power of machine A, and $\delta_b$ is the I/O performance)

In addition, as we focus on the I/O restrictions, we can consider that we can compute an arbitrary number of rigid jobs on A and focusing solely on machine B.

Our aim is mainly to minimize the overall execution time of the applications.
Notations We write chain job \( J_i \) as \( J_i = \prod_{j=1}^{n_i} (a_{i,j}, b_{i,j}) \) where A-type task \( a_{i,j} \) and B-type task \( b_{i,j} \) must be completed before a-type \( a_{i,k} \) and B-type \( b_{i,k} \) for all \( k \geq j \).

\( a_{i,j} \) models the \( j^{th} \) compute phase of application \( i \) and \( b_{i,j} \) its \( j^{th} \) I/O phase.

We write \( n_i \) the length (aka number of A-tasks and B-tasks pairs) of job \( J_i \), \( \alpha_{i,j} \) (resp. \( \beta_{i,j} \)) the starting time of task \( a_{i,k} \) (resp. \( b_{i,j} \)).

2.2 Single processor scheduling with latencies

In the case, where machine A has no limitation (\( \delta_A = \infty \)), the problem is the same as single processor-scheduling with latency (or minimum delay [17]) on a single machine. The problem is the following: Given a set of chains applications, and a single machine P.

- each job consists of a chain of tasks \( b_{i,j} \) to be performed on P.
- each task \( b_{i,j} \) is followed by a latency \( a_{i,j} \), the time between the end of \( b_{i,j} \) and the start of \( w_{i,j+1} \) cannot be lower than \( a_{i,j} \)
- There are no precedence constraints between different jobs
- Preemption is not allowed

\[ J_i = \prod_{j} (b_{i,j}, a_{i,j}) \]

From any instance of the bi-colored chain problem, we can construct an equivalent instance of single processor with latencies.

To do so, we identify B-operations (from bi-colored chains) with tasks and A-operations with latencies and reverse the order of the chains obtained.

This approach of the problem has been discussed in [17] and is easier to study as it focus only on one processor. In the following, we study this form of the problem.

3 Problem presentation

3.1 Objectives

Given a set of jobs with latencies: \( J_i = \prod_{j} (b_{i,j}, a_{i,j}) \)

**Definition 1.** In our context, a schedule \( S \) is a function that assigns a start time to each B-task such that:

- \( S(b_{i,j}) + b_{i,j} + a_{i,j} \leq b_{i,k} \) for all \( i \) and \( j<k \).
- \( \forall i_1, i_2, j_1, j_2, \) either \( \beta_{i_1,j_1} + b_{i_1,j_1} \leq \beta_{i_2,j_2} \) or \( \beta_{i_2,j_2} + b_{i_2,j_2} \leq \beta_{i_1,j_1} \)
In other words, \( S \) respects precedence and latency constraint and two B-tasks cannot be run simultaneously.

Notations. We write \( \beta_{i,j} = S(b_{i,j}) \) the starting time of task \( b_{i,k} \) and \( \alpha_{i,j} = S(b_{i,j}) + b_{i,j} \) its end of execution (start of latency period).

A scheduling problem consist to find an schedule \( S \) that optimizes a given objective.

Makespan We first define the makespan of a schedule as the total execution time.

**Definition 2 (Makespan).** \( MS(S) = \max_i (\alpha_{i,n_i} + a_{i,n_i}) \)

**Definition 3 (Makespan problem).** Given a set of job with latencies, \( J_i = \prod_j (b_{i,j}, a_{i,j}) \) Is there an algorithm that provides an schedule \( S \) such that \( MS(S) \) is minimal ?

This objective is the "easiest" one, and the first classically used in scheduling problems.

Whereas the first objective focus on the global performance of the platform, in real life, different actors can run applications at the same time. We might also consider fairness as an additional objective. Dilation is a user-oriented objective which focus on relative progress rates of applications.

Dilation Given a schedule \( S \), the dilation is the largest slowdown imposed to each application.

**Definition 4 (Dilation).** \( \text{dil}(S) = \max_{i=1...m} \frac{\sum_{k} \alpha_{i,k} + b_{i,k}}{\alpha_{i,n_i} + a_{i,n_i}} \)

**Definition 5 (Dilation problem).** Problem: Given a set of job with latencies, \( J_i = \prod_j (b_{i,j}, a_{i,j}) \) Is there an algorithm that provides an schedule \( S \) such that \( \text{dil}(S) \) is minimal ?

4 Complexity and Algorithms

We have previously seen that schedule to reduce I/O congestion can be model by a 2-stage bi-colored chain scheduling problem. This problem is known to be NP-hard. In this part, we present some restrictions of this problem and ways to solve them.
4.1 Chains of length 1

To begin with, we consider the sub problem where all chains contain only one two-stage element. In this particular case, bi-colored chains scheduling is a classic two stages flow-shop problem.

To lighten the notation, $b_{i,j}$ can become $b_j$ and we relabel the $j$ to match the scheduling order (ie $i<j$ if and only if $b_i$ is executed before $b_j$).

As there are no precedence constraints, all tasks $b$ can be run whenever convenient on machine $B$. Hence the following result:

**Lemma 4.1** ([17],[18]). For problem $J_i = (b_i, a_i)$, $C_{max} = \sum_{j<i} b_j + a_i$ and an optimal schedule is reach by greedily executing the largest $a_i$ first (Johnson order).

4.2 Chains of length at least 2

In this part, we discuss the difficulty of finding an optimal schedule with chains of equal length.

In order to do so define the problem duplicated subset sum as follow.

**Definition 6** (Duplicated subset sum). Given an integer set $A = \{x_1, x_2, ..., x_t\}$ and a integer $T$. Are there disjoints subset $S_1$ and $S_2$ such that:

- $S_1, S_2 \subseteq A$
- $S_1 \cap S_2 = \emptyset$
- $\sum_{x_i \in S_1} x_i + 2 \sum_{x_j \in S_2} x_j = T$

**Theorem 4.2.** Duplicated subset sum is NP-complete.

**Proof.** Reduction from 3-sat. The classical reduction from 3-sat to subset sum as defined in [14] can be adapted for duplicated subset-sum. See Appendix A.

**Theorem 4.3.** 2-chains is NP-complete.

**Proof.** Let $A = x_1, x_2, ..., x_t$, $T$ be an instance of duplicated subset sum. Consider that $2\Sigma x_i > T$.

Build an instance of 2-chains:

- $J_i = (x_i, 1)^2$ for $i = 1...t$
- $J_{t+1} = J_{t+2} = (T + 1, T + 1)^2$

The idea is to build an optimal solution of the 2-chains problem by scheduling the jobs $J_{t+1}$ and $J_{t+2}$. Then fill the lateness at the end of the schedule (of length $T+1$) with as many tasks as possible.
These two proofs can be generalized for any integer $k$.

**Corollary 4.3.1.** $k$-subset sum is NP complete.

**Theorem 4.4.** $k$-chains is NP-complete.

### 4.3 All tasks of all chains are identical:

In this section, we consider the case where all tasks are identical but spread into different chains. Minimizing the makespan of this sub-problem can be done in polynomial time.

Consider a set of $m$ Jobs with the form $J_i = (b, a)^{n_i}$ sorted by non-increasing $n_i$. We want to find a schedule $S$ that minimizes the makespan $MS(S)$. Given a sequence, we write $b^t_i$, the $t^{th}$ occurrence of task $b_i$

**Algorithm 1** Hierarchical Round-robin

1. procedure HRR($J_i = (b, a)^{n_i}$)
2. Sort jobs by non-increasing $n_i$
3. Initialize $n_0$ blocks containing $b_0$
4. while $\exists J_i$ of length $n_0$ do
5. Schedule one of its task within each block
6. $i \leftarrow 0$
7. while $\exists J_j$ remaining, $n_j < n_0$ its length do
8. for $t = 0, t < n_j - 1$ do
9. Insert $b^t_j$ at the end of block $i + t \mod n_0$
10. $i \leftarrow i + n_j \mod n_0$

**Theorem 4.5.** Hierarchical Round-robin is optimal for $J_i = (b, a)^{n_i}$

The proof follows the steps below:

1. We define of a cost function $C$ such that $\forall A$, $MS_A \geq C(A)$
2. $\forall A$ algorithm, we show $C_{min}(A_{opt}) \geq C(HRR)$
3. We conclude by $MS_{HRR} = C(HRR)$

In the following, we consider jobs to be sorted by non-increasing $n_i$.

**Lemma 4.6.** There is an optimal schedule which starts with $J_0$.

*Proof.* Consider an optimal schedule, $b_i$ its first task. If $i \neq 0$, then consider the first $b^0_0$ such that there is no $b_i$ between $b^0_0$ and $b^{t+1}_i$ Such a task exists because $n_i \leq n_0$. Write $b^k_i$ the last occurrence of $b_i$ before $b^0_0$. As all tasks are identical, we can permute $b^k_i$ and $b^k_0$ for $k<i$ without increasing the makespan.

□
We introduce the notion of block:

**Definition 7.** Given a schedule $S$, the block $B^S_t$ is the subsequence $S$ starting from $b^0_t$ (included) until $b^0_{t+1}$. $B^S_{n_0}$ the subsequence starting from $b^0_0$ until the end of the schedule. Its length $|B^S_t|$ is its number of tasks. Define its cost:

$$C(B^S_{n_0}) = |B^S_{n_0}| b + a$$

and

$$C(B^S_t) = \max(b + a, |B^S_t| b)$$

if $t \neq n_0$.

**Proposition 1.** $B_t \mapsto C(B_t)$ is an increasing function of $|B_t|$

**Definition 8.** Given a schedule $S$, define $C(S) = \sum_i C(B^S_i)$

As we work given a fixed schedule, we write abusively $B_t$ for $B^S_t$.

**Proposition 2.** For any $S$, $M_S \geq C(S)$.

**Proof.** Write $I_S$ the total idle time of the processor during the full execution of $S$, $I_{S,0}$ the idle time that occurs immediately before a task of $J_0$.

$M_S = \sum_i b_i^I + I_S$

with $I_S \leq I_{S,0}^I$.

By definition of blocks $I_{S,0}^I = \sum_t \left(C(B^S_t) - |B^S_t| b\right)$ Hence the expected inequality.

**Definition 9.** For $k \in \{1, 2, \ldots, n_0\}$ define $q_k = \left\lfloor \frac{N - n_0 - k}{n_0 - 1} \right\rfloor$, $r_k = (N - n_0 - k) \mod (n_0 - 1)$. Let $S_{\text{min}}$ be a schedule such that its first $r_k$ blocks have size $q_k + 1$, the $n_0 - 1 - r_k$ following have size $q_k$, and the last has size $k$.

$C_{\text{min}}(k) = C(S_{\text{min}})$

**Proposition 3.** $\forall S$, $C(S) \geq C_{\text{min}}(|B^S_n|)$
Proof. \( S \) an arbitrary schedule. To lighten a little the notation \( q = q_{0}B_{n_{0}}^{S}, r = r_{0}B_{n_{0}}^{S} \), By recurrence on the smallest size \( p \) of a \( S \) interior block. \( p = \min_{t < n_{0}} |B_{t}^{S}| \).

Initialization: If \( p=q \). If there is \( n_{0} - 1 - r_{0}B_{n_{0}}^{S} \) blocks of size \( q \), then \( C(S) = C_{\text{min}}(|B_{n_{0}}^{S}|) \) by identity. Assume that if there is at most \( k \) blocks of size \( q \), then \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \). If there is \( k+1 \) blocks of size \( q \), then there exists \( B_{i} \) and \( B_{j} \) two blocks such that \( |B_{i}| = q, |B_{j}| = p_{\text{max}} \) is maximal. Create blocks \( \tilde{B}_{i} \) and \( \tilde{B}_{j} \) by moving task \( b, b \in B_{j}, b \not\in B_{i} \), from \( B_{j} \) to \( B_{i} \).

- if \( (a + b) \geq p_{\text{max}}b \) then \( C(B_{j}) + C(B_{i}) - C(\tilde{B}_{j}) + C(\tilde{B}_{i}) = 0 \)
- if \( p_{\text{max}}b \geq (a + b) \geq (p_{\text{max}} - 1)b \) then \( C(B_{j}) + C(B_{i}) - C(\tilde{B}_{j}) + C(\tilde{B}_{i}) = p_{\text{max}}b - (a + b) \)
- if \( p_{\text{max}} - 1 \geq (a + b) \geq (q+1)b \) then \( C(B_{j}) + C(B_{i}) - C(\tilde{B}_{j}) + C(\tilde{B}_{i}) = b \)
- \( (q + 1)b \geq (a + b) \geq qb \) then \( C(B_{j}) + C(B_{i}) - C(\tilde{B}_{j}) + C(\tilde{B}_{i}) = qb - a \)
- if \( qb \geq (a + b) \) then \( C(B_{j}) + C(B_{i}) - C(\tilde{B}_{j}) + C(\tilde{B}_{i}) = 0 \)

In all cases give \( (|\tilde{B}_{i}| + |\tilde{B}_{j}|) \leq (|B_{i}| + |B_{j}|) \)

Hence, if there is \( k+1 \) blocks of size \( q \), then \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \). By recurrence, if the smallest block in \( S \) has size \( q \), then \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \).

Hypothesis: For an arbitrary \( p_{0} \), if the smallest block in \( S \) has size at least \( p_{0} \), then \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \).

Consider that the smallest block in \( S \) has size \( p = p_{0} \).

Then there exists \( B_{i} \) and \( B_{j} \) two blocks such that \( |B_{i}| = p, |B_{j}| \) is maximal. Create blocks \( \tilde{B}_{i} \) and \( \tilde{B}_{j} \) by moving task \( b, b \in B_{j}, b \not\in B_{i} \), from \( B_{j} \) to \( B_{i} \).

With similar cases as above: \( |\tilde{B}_{i}| + |\tilde{B}_{j}| \leq (|B_{i}| + |B_{j}|) \)

Hence if the smallest block in \( S \) has size at \( p = p_{0} - 1 \), then \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \).

By recurrence, \( C(S) \geq C_{\text{min}}(|B_{n_{0}}^{S}|) \) for all \( p \).

Proposition 4. \( l = |\{J_{i}, \text{ length of } J_{i} \text{ is } n_{0}\}|, S_{\text{opt}} \text{ an optimal schedule then } |B_{n_{0}}^{S_{\text{opt}}}| \geq l. \)

Proof. Using the same argument as in lemma 4.6

Proposition 5. \( MS_{HRr} = C_{\text{min}}(l) \)

Proof. Blocks obtained by HRr have the same shape as blocks of \( C_{\text{min}}(l) \).

By construction, all tasks but those from \( J_{0} \) verify their precedence constraints as soon as they appear in the schedule. Hence \( MS_{HRr} = \sum l C(B_{t}^{HRr}) \)
proof of the theorem. We have seen that in proposition 4 that \( |B_{n_0}^{\text{opt}}| \geq l \). Thus \( C(S_{\text{opt}}) \geq C_{\text{min}}(l) \) by proposition 3. This bound is attained by Hierarchical Round-robin hence the optimality.

4.4 Approximations

We have seen that finding a schedule of jobs \( J_i = (b_i, a_i)^{n_i} \) with an optimal makespan is NP-complete. Therefore, we try to find algorithms that provide an acceptable approximation.

**Definition 10.** A \( \lambda \)-approximation algorithm is an algorithm whose execution time is polynomial in the instance size and that returns an approximate solution guaranteed to be, in the worst case, at a factor \( \lambda \) away from the optimal solution.

**Proposition 6.** Any list-heuristic provides a 2-approximation for \( J_i = (b_i, a_i)^{n_i} \).

**Proof.** Given a schedule obtained by a list-based heuristic, by definition, if the B-processor is idle, it means that no B-task is available. This total idle time is \( t_{\text{idle}} = MS_{\text{list}} - \sum_{i} n_i b_i \). All jobs are currently over or performing a A-task, hence the following bound with \( J_i \) the last executing job. This gives \( t_{\text{idle}} \leq \sum_j a_{ij} \leq \sum_j (a_{ij} + b_{ij}) \leq \text{opt} \).

**Lemma 4.7.** This bound is tight.

**Proof.** The straightforward heuristic would be to prioritize jobs with the most work remaining on processor B. This heuristic provides a approximation factor arbitrary close to 2 on the instance: \( J_1 = (\epsilon, \epsilon)^k \rightarrow (\epsilon, 1-\epsilon) \), \( J_2 = (1,0) \) and \( \epsilon \) small enough.

5 Periodic Algorithms

In this section, we study periodic restriction of the problem. Indeed, in HPC, applications tend to have periodic or pseudo-periodic in the I/O behavior. HPC applications alternate between computation and I/O transfer, this pattern being repeated over-time. In addition, fault-tolerance techniques (such as periodic checkpointing) also add to this periodic behavior.

Late publications such that [1] and [8] propose I/O scheduling strategy with periodic approach.

Some of our objectives were to provide theoretical backgrounds to these algorithm and in a second time, to discuss their pertinence.
5.1 n-chains: $J_i = (b_i, a_i)^n$

We have proved that this problem is NP-hard (theorem 4.5).

As all chains length are equals and we want to ensure fairness between the different jobs. Let’s consider a schedule built by ordering one task from each job in a periodic sequence and repeating this period n times. We call this kind of policy 'periodic schedule' (Periodic), First, we discuss the way of ordering tasks within a period and then discuss the performance of such scheduling algorithms.

Notations.

- In the following, I call "idle time" of a schedule $S$, the time $t_i(S) = MS_S - \sum b_i$

- In Periodic, all jobs have only one task in each period. We can define the order $\prec$: $i \prec j$ if and only if $b_i$ appears before $b_j$ in the period.

The overall idle time of the periodic schedule is:

$$t_i(\text{Periodic}) = (n - 1) \max_i \left(a_i - \sum_{j \neq i} b_j\right) + \max_k \left(a_k - \sum_{k < j} b_j\right)$$

The order within a period does not change the overall idle time, therefore we can sort tasks by non-increasing $A$-task length in a period with gives:

$$t_i(\text{Periodic}) \leq (n - 1) \max_i \left(a_i - \sum_{j \neq i} b_j\right) + \max_k (a_k)$$

Given an optimal schedule, the idle time is:

$$t_i(\text{OPT}) \max_i \left(na_i - n \sum_{j \neq i} b_j + \sum_{j \prec i} b_j\right) \geq n \max_k \left(a_k - \sum_{j \neq k} b_j\right)$$

where $i$ is the last task running on $A$ and $i_1$ is its first iteration.

Therefore, using straightforward bounds, the difference between these periodic and opt is at most:

$$n \max_i \left(a_i - \sum_{j \neq i} b_j\right) \leq n \max_i (a_i)$$

The optimal makespan is at least $n \max_i (a_i + b_i)$ Hence the following proposition.

Definition 11. Define Periodic as the schedule obtain by running in each period one task of each job following an order between jobs (for example Johnson order see Lemma 4.1).

Proposition 7. Periodic is a $\rho$-approximation for $J_i = (b_i, a_i)^n$. With $\frac{11}{10} \leq \rho \leq 2$

Remark. The development above ensure that periodic is at worst a 2-approximation. However it may not be intuitive that Periodic does not provide an optimal schedule. The following example provides the other bound:

- 3 chains of length 2
• $J_i = (b_i = 2, a_i = 2)$ for $i=1,2$

• $J_3 = (b_3 = 1, a_3 = 0)$ for $i=3$

The best periodic algorithms give a makespan of 11. (1-2-3-1-2-3)

An optimal ordering has makespan 10. (1-2-1-2-3-3)

Lemma 5.1. Periodic is asymptotically optimal for $J_i = (b_i, a_i)^n$

Proof. Write $T$ a period, $nT \leq MS_{Periodic} \leq nT + a_{max}$

By construction: $T = \sum b_j + \max (a_i - \sum_{j \neq i} b_j)$

$T = \max (a_i + b_i, \Sigma b_j)$

$n * \max (a_i + b_i, \Sigma b_j) \leq MS_{opt} \leq MS_{Periodic} \leq n * \max (a_i + b_i, \Sigma b_j) + \max (a_i)$

Which gives: $\frac{MS_{Periodic}}{MS_{opt}} \leq 1 + \frac{1}{n}$

5.2 Periodic repetition: $J_i = [(b_i, a_i)^{k_i}]^n$

This problem is equivalent to the general case. However, as discussed before, the notation emphasizes the method of resolution. We aim to build a period containing $k_i$ tasks of each job $i$ using list scheduling and to repeat it $n$ times.

This method provides at least a 2-approximation.

Using the same argument as in lemma 6

6 Model Evaluation

Experimental measurement The HPC-oriented team of Inria Bordeaux are working with ATOS. In the scope of this collaboration, we received a collection of data gathered by monitoring the execution of different applications on ATOS platform.

The application are rather short to suit a periodical behavior with execution time between 2h40 and 5 minutes. Given the complete log of I/O operation, we started to gather relevant timestamps (those who show significant amounts of operations) into phases. Then, we output the amount of such phases, their mean volume standard deviation, and their mean period and standard deviation. Additionally, we also compute the percentage of time taken by the first I/O phase and the one elapsed before the last in order to emphasized start up and shutdown behavior. For the applications that have more than 8 I/O phases, we also used the python function stats.normaltest to measure the distance with a normal distribution. It appears that these applications have a nearly periodical comportment with regular I/O volumes. see Appendix for extended view on the data.
7 Collaboration with the CLARISSE project

CLARISSE is a middleware designed to enhance data-staging coordination and control in the HPC software storage I/O stack [9]. It first uses a prediction tool to anticipate I/O behavior, then when congestion is anticipated, it allocates inactive nodes of the platform to applications to accelerate their computation and desynchronize their I/O.

CLARISSE currently uses first come - first served policy to choose which applications to speed-up. Using the prediction tool to produce an instance of our problem, the heuristics proposed in section 5.1 (prioritize jobs that have the largest computational requirements) would provide a schedule (ie an order of jobs). CLARISSE can then use this order in place of the current policy.

An other approach could be that our algorithm provides time frames where each I/O can be executed. Any I/O operations submitted outside this time frame would be delayed.

Our periodic schedule theoretically provides better performance than FIFO. Experiments with CLARISSE would help quantify this amelioration, validate our model and open new perspectives for algorithm design.

In practice, to be discussed

INPUT: The expected behavior of jobs: computation and I/O durations

- Our algorithm computes a schedule and provides two possible kind of output

OUTPUT: 1. an order between jobs to use in place of FIFO

2. for each I/O phase, a "release date" which is the moment when it can start.

3. for each I/O phase, a timeframe in which it should be triggered. I/O requests that arrive outside these dates are delayed.

8 Related Work

Surprisingly enough, periodic problems is not so abundant among the scheduling literature despite their interest.

Most of them are very application-oriented so the model is not easily exportable. We are not an exception, as our model is designed to represent I/O issues of HPC applications.

Periodic scheduling are also studied in the scope of cyclic scheduling ie considering that the tasks being scheduled have periodical release date ([4],[16],[5]) which does not cover the latency issue.
Figure 2: Scheduling three applications

In this figure, we show an example of three applications competing for I/O bandwidth. On the top part, we see the applications doing computations without any constraint. However at the end of their computations, all applications need to transfer some volume of I/O. The bandwidth is therefore used, in turn, by each application. The order of these turn is what we intend to provide with our algorithms.

The following table is a non-exhaustive survey of problem that approach ours:

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
<th>Reference</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_i = (a_i, b_i)$</td>
<td>Polynomial</td>
<td>[17] [18]</td>
<td></td>
</tr>
<tr>
<td>$J_i = (a_i, b_i)$</td>
<td>Polynomial</td>
<td>Round-Robin</td>
<td></td>
</tr>
<tr>
<td>$J_i = (a_i, b_i)$</td>
<td>NP-complete</td>
<td>[17]</td>
<td></td>
</tr>
<tr>
<td>$J_i = (a_i, b_i)$</td>
<td>NP-complete</td>
<td>Theorem 4.4</td>
<td>Periodic [(1+1/n)-approx]</td>
</tr>
<tr>
<td>$J_i = (a_i, b_i)$</td>
<td>NP-complete</td>
<td>Open</td>
<td>2-approximation</td>
</tr>
<tr>
<td>$J_i = \Pi_{j=0}^{n_i} (a_{i, j}, b_{i, j})$</td>
<td>NP-complete</td>
<td>list-scheduling</td>
<td>2-approx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
<th>Reference</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P</td>
<td>2FL</td>
<td>C_{max}$</td>
<td>Polynomial</td>
</tr>
<tr>
<td>$P_m</td>
<td>2FL</td>
<td>C_{max}$</td>
<td>Pseudo-polynomial</td>
</tr>
<tr>
<td>Periodic $C_{max}/S_{max}$</td>
<td>NP-complete</td>
<td>[2]</td>
<td>3-partition</td>
</tr>
<tr>
<td>$F_2[\text{chains}]C_{max}$</td>
<td>NP-complete</td>
<td>[11]</td>
<td></td>
</tr>
<tr>
<td>$F_2[\text{chains}, \text{pmnt}]C_{max}$</td>
<td>NP-complete</td>
<td>Lenstra</td>
<td>cf. [3]</td>
</tr>
<tr>
<td>$P_m[p_{ij}=1, \text{chains}]\Sigma w_i C_i$</td>
<td>NP-complete</td>
<td>[13]</td>
<td></td>
</tr>
<tr>
<td>$P[p_{ij}=1, \text{prec}]C_{max}$</td>
<td>NP-hard</td>
<td>[12, 15]</td>
<td></td>
</tr>
<tr>
<td>$O_2[\text{chains}]S_{max}$</td>
<td>strongly NP-hard</td>
<td>[13]</td>
<td></td>
</tr>
<tr>
<td>$O_2[\text{pmnt}]\Sigma w_i C_i$</td>
<td>strongly NP-hard</td>
<td>Lenstra</td>
<td></td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=1 \text{ chains}</td>
<td>C_{max}$</td>
<td>P</td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=2, \ldots, 1 \text{ chains}</td>
<td>C_{max}$</td>
<td>NP-complete</td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=2, 2 \ldots, 1 \text{ chains}</td>
<td>C_{max}$</td>
<td>NP-complete</td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=3 \ldots, 1 \text{ chains}</td>
<td>C_{max}$</td>
<td>strongly NP-complete</td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=3, 2 \ldots, 1 \text{ chains}</td>
<td>C_{max}$</td>
<td>strongly NP-complete</td>
</tr>
<tr>
<td>$1</td>
<td>\text{min delays, } k=3, 2 \ldots, 1 \text{ chains}</td>
<td>C_{max}$</td>
<td>NP-complete</td>
</tr>
</tbody>
</table>
References


A K-subset sum

Adapted from [14]

To show that \( k\)-SUBSET-SUM is in NP, for an instance \(< A, B >\) of the problem, we let the subsets \( S_1 \) and \( S_2 \) be the certificate. Checking whether \( B = \sum_{a_i \in S_1} a_i + 2 \sum_{a_j \in S_2} a_j \) can be accomplished by a verification algorithm in polynomial time.

We now show that 3-SAT can be reduced to SUBSET-SUM. Given a 3-SAT formula \( \phi \) over variables \( x_1, x_2, \ldots, x_n \) with clauses \( C_1, C_2, \ldots, C_k \), each containing exactly three distinct literals, the reduction algorithm constructs an instance \(< A, B >\) of the duplicated subset-sum problem such that \( \phi \) is satisfiable if and only if there are subset \( S_1 \) and \( S_2 \) of \( A \) such that \( B = \sum_{a_i \in S_1} a_i + 2 \sum_{a_j \in S_2} a_j \).

Without loss of generality, we make two simplifying assumptions about the formula \( \phi \). First, no clause contains both a variable and its negation, for such a clause is automatically satisfied by any assignment of values to the variables. Second, each variable appears in at least one clause, for otherwise it does not matter what value is assigned to the variable.

The reduction creates two numbers (and their duplicate if need be) in set \( A \) for each variable \( x_i \) and two numbers in \( A \) for each clause \( C_j \). We shall create numbers in base 10, where each number contains \( n+k \) digits and each digit corresponds to either one variable or one clause. Base 10 has the property we need of preventing carries from lower digits to higher digit.

We construct set \( A \) and target \( B \) as follows. We label each digit position by the either a variable or a clause. The least significant \( k \) digits are labeled by the clauses, and the most significant \( n \) digits are labeled by variables.

- The target \( t \) has a 1 each digit labeled by a variable and a 3 in each digit labeled by a clause.
- For each variable \( x_i \), there are two integers, \( v_i \) and \( v'_i \) in \( A \). Each has \( A \) in the digit labeled by \( x_i \) and O’s in the other variable digits. If literal \( x_i \) appears in clause \( C_j \), then the digit labeled by \( C_j \) in \( v_i \) contains a 1. If literal \( \neg x_i \) appears in clause \( C_j \), then the digit labeled by \( C_j \) in \( v'_i \) contains a 1. All other digits labeled by clauses in \( v_i \) and \( v'_i \) are 0.

So far all \( v_i \) and \( v'_i \) values in set \( A \) are unique. Why? For \( l \neq i \), no \( v_l \) or \( v'_l \) values can equal \( v_i \) and \( v'_i \) in the most significant \( n \) digits. Furthermore, by our simplifying assumptions above, no \( v_i \) and \( v'_i \) can be equal in all \( k \) least significant digits. If \( v_i \) and \( v'_i \) were equal, then \( x_i \) and \( \neg x_i \) would have to appear in exactly the same set of clauses. But we assume that no clause contains both \( x_i \) and \( \neg x_i \) and that either \( x_i \) or \( \neg x_i \) appears in some clause, and so there must some clause \( C_j \) for which \( v_i \) and \( v'_i \) differ.
For each clause $C_j$, there are two integers, $s_j$ and $s'_j$ in $S$. Each has 0’s in all digits other than the one labeled by $C_j$. $s_j$ and $s'_j$ have a 1 in the $C_j$ digit. These integers are "slack variables", which we use to get each clause-labeled digit position to add to the target value of 3.

\[ \begin{array}{cccccccc}
\text{v}_1 & \rightarrow & x_1 & 0 & 0 & 0 & 1 & 1 \\
\text{v}_2 & \rightarrow & 0 & 1 & 0 & 0 & 0 & 1 \\
\text{v}_3 & \rightarrow & 0 & 0 & 1 & 0 & 0 & 1 \\
\text{v}_4 & \rightarrow & 0 & 0 & 0 & 1 & 1 & 0 \\
\text{v}'_1 & \rightarrow & 0 & 0 & 1 & 1 & 0 & 0 \\
\text{v}'_2 & \rightarrow & 0 & 0 & 1 & 1 & 0 & 0 \\
\text{v}'_3 & \rightarrow & 0 & 0 & 0 & 0 & 0 & 1 \\
\text{v}'_4 & \rightarrow & 0 & 0 & 0 & 0 & 0 & 1 \\
\text{s}_1 & \rightarrow & 0 & 0 & 0 & 1 & 0 & 0 \\
\text{s}_2 & \rightarrow & 0 & 0 & 0 & 1 & 0 & 0 \\
\text{s}_3 & \rightarrow & 0 & 0 & 0 & 0 & 1 & 0 \\
\text{s}_4 & \rightarrow & 0 & 0 & 0 & 0 & 1 & 0 \\
t & \rightarrow & 1 & 1 & 1 & 3 & 3 & 3 \\
\end{array} \]

Figure 3: Example: The reduction of 3-SAT to SUBSET-SUM. The formula is $\phi = C_1 \land C_2 \land C_3 \land C_4$, where $C_1 = (x_1 \lor \neg x_2 \lor \neg x_3)$, $C_2 = (\neg x_1 \lor \neg x_2 \lor \neg x_3)$, $C_3 = (\neg x_1 \lor \neg x_2 \lor \neg x_3)$ and $C_4 = (x_1 \lor x_2 \lor x_3)$. A satisfying assignment of $\phi$ is $\langle x_1 = 0, x_2 = 0, x_3 = 1 \rangle$. The set $S$ produced by the reduction consists of the base-10 numbers shown; reading from top to bottom, $S = 1001001, 1000110, 100001, 101110, 10011, 11100, 1000, 1000, 100, 100, 10, 10, 1, 1$. The target $t$ is 1113333. The subset $S' \subset S$ contains $v'_1, v'_2, v'_3$, corresponding to the satisfying assignment. It also contains slack variables $s_1, s'_1, s'_2, s'_3, s'_4$ and $s'_4$ to achieve the target value of 3 in the digits labeled by $C_1$ through $C_4$.

Note the the greatest sum of digits in any one digit position is 8, which occurs in the digits labeled by clauses (six 1’s from the $v_i$ and $v'_i$ and their duplicates, plus two 1 for $s_j$ and $s'_j$). Interpreting these numbers in base 10, therefore, no carries can occur from lower digits to higher digits. The reduction can be performed in polynomial time. The set $A$ contains $2n+2k$ values, each of which has $n+k$ digits, and the time to produce each digit is polynomial in $n+k$. The target $t$ has $n+k$ digits, and the reduction produces each in constant time.

We now show that the 3-SAT formula $\phi$ is satisfiable if and only if there are subsets $S_1$ and $S_2$ of $A$ whose sum is $B$. First, suppose that $\phi$ has a satisfying assignment. For $i = 1, 2, ..., n$, if $x_i = 1$ in this assignment, then include $v_i$, in $S_1$. Otherwise, include $v'_i$. In other words, we include in $S_1$ exactly the $v_i$ and $v'_i$ values that correspond to literals with the value 1 in the satisfying assignment. Having included either $v_i$ or $v'_i$, but not both, for all $i$, and having put 0 in the digits labeled by variables in all $s_j$ and $s'_j$, we see that for each variable-labeled digit, the sum of the values of $S_1$ must be 1, which matches those digits of the target $t$. Because each clause is satisfied, there is some literal in the clause with the value 1. Therefore, each
digit labeled by a clause has at least one 1 contributed to its sum by a \(v_i\) or \(v'_i\) value in \(S_1\). In fact, 1, 2, or 3 literals may be 1 in each clause, and so each clause-labeled digit has a sum of 1, 2, or 3 from the \(v_i\) and \(v'_i\) values in \(S_1\). We achieve the target of 3 in each digit labeled by clause \(C_j\) by including in \(S_1\) the appropriate nonempty subset of slack variables \(s_j, s'_j\). Since we have matched the target in all digits of the sum, and no carries can occur, the values of \(S_1\) sum to \(B\).

Now, suppose that there are subsets \(S_1\) and \(S_2\) of \(A\) such that \(B = \sum_{a_i \in S_1} a_i + 2 \sum_{a_j \in S_2} a_j\). The subset \(S_1\) must include exactly one of \(v_i\) and \(v'_i\) for each \(i = 1, 2, \ldots, n\), for otherwise the digits labeled by variables would not sum to 1. For the same reasons \(S_2\) must be empty. If \(v_i \in S_1\), we set \(x_i = 1\). Otherwise, \(v'_i \in S_1\), and we set \(x_i = 0\). We claim that every clause \(C_j\), for \(j = 1, 2, \ldots, k\), is satisfied by this assignment. To prove this claim, note that to achieve a sum of 3 in the digit labeled by \(C_j\), the subset \(S_1\) must include at least one \(v_i\) or \(v'_i\) value that has a 1 in the digit labeled by \(C_j\), since the contributions of the slack variables \(s_j\) and \(s'_j\) together sum to at most 2. If \(S_1\) includes a \(v_i\) that has a 1 in that position, then the literal \(x_i\) appears in \(C_j\). Since we have set \(x_i = 1\) when \(v_i \in S_1\), clause \(C_j\) is satisfied. If \(S_1\) includes a \(v'_i\) that has a 1 in that position, then the literal \(\neg x_i\) appears in \(C_j\). Since we have set \(x_i = 0\) when \(v'_i \in S_1\), clause \(C_j\) is again satisfied. Thus, all clauses of \(\Phi\) are satisfied, which completes the proof.

**Remark.** The very same proof can be adapted for \(k\)-subset-sum.
B Traces

Figure 4: Sample of data provided by ATOS
<table>
<thead>
<tr>
<th>id</th>
<th>total_duration</th>
<th>total_volume</th>
<th>phases</th>
<th>starting_IO</th>
<th>terminating_IO</th>
<th>mean (period)</th>
<th>sd (period)</th>
<th>pvalue</th>
<th>mean (vol)</th>
<th>sd (vol)</th>
<th>pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>10246</td>
<td>335000</td>
<td>1472542576</td>
<td>2</td>
<td>95.52%</td>
<td>100.00%</td>
<td>1.50E+04</td>
<td>2.50E+03</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4799</td>
<td>280000</td>
<td>1472541830</td>
<td>3</td>
<td>10.71%</td>
<td>100.00%</td>
<td>1.25E+04</td>
<td>4.91E+08</td>
<td>5.11E+04</td>
<td>5.43E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54171</td>
<td>694000</td>
<td>108571914971</td>
<td>26</td>
<td>0.50%</td>
<td>99.93%</td>
<td>2.72E+05</td>
<td>4.98E+04</td>
<td>4.18E+09</td>
<td>9.66E-07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10256</td>
<td>270000</td>
<td>1472542576</td>
<td>3</td>
<td>11.11%</td>
<td>98.15%</td>
<td>1.25E+04</td>
<td>2.50E+03</td>
<td>4.91E-08</td>
<td>4.33E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10235</td>
<td>325000</td>
<td>1472542576</td>
<td>3</td>
<td>13.85%</td>
<td>100.00%</td>
<td>1.25E+04</td>
<td>2.50E+03</td>
<td>4.91E-08</td>
<td>4.33E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51682</td>
<td>161500</td>
<td>4086112341</td>
<td>19</td>
<td>2.79%</td>
<td>99.07%</td>
<td>8.22E+04</td>
<td>7.19E+04</td>
<td>2.15E+09</td>
<td>5.43E+09</td>
<td>9.33E-09</td>
<td></td>
</tr>
<tr>
<td>10248</td>
<td>330000</td>
<td>1472544180</td>
<td>3</td>
<td>15.15%</td>
<td>100.00%</td>
<td>1.56E+04</td>
<td>5.00E+03</td>
<td>4.91E-08</td>
<td>4.33E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10260</td>
<td>335000</td>
<td>1472542576</td>
<td>2</td>
<td>94.03%</td>
<td>98.51%</td>
<td>1.56E+04</td>
<td>4.30E+04</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47579</td>
<td>57770000</td>
<td>22627161837</td>
<td>15</td>
<td>0.05%</td>
<td>99.57%</td>
<td>4.60E+04</td>
<td>3.54E+04</td>
<td>1.97E+09</td>
<td>1.95E+09</td>
<td>1.57E-04</td>
<td></td>
</tr>
<tr>
<td>47818</td>
<td>214500</td>
<td>103103082</td>
<td>65</td>
<td>0.23%</td>
<td>99.97%</td>
<td>3.30E+04</td>
<td>4.64E+03</td>
<td>1.59E+09</td>
<td>8.66E+09</td>
<td>2.17E-24</td>
<td></td>
</tr>
<tr>
<td>10258</td>
<td>250000</td>
<td>1472542576</td>
<td>2</td>
<td>94.00%</td>
<td>100.00%</td>
<td>1.56E+04</td>
<td>3.08E+06</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47458</td>
<td>955000</td>
<td>1790292653</td>
<td>8</td>
<td>0.16%</td>
<td>99.84%</td>
<td>1.27E+06</td>
<td>2.16E+09</td>
<td>2.93E+09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48565</td>
<td>214500</td>
<td>124731940062</td>
<td>1</td>
<td>100.00%</td>
<td>0.00%</td>
<td>1.25E+11</td>
<td>1.25E+11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56181</td>
<td>312500</td>
<td>873233292</td>
<td>8</td>
<td>0.96%</td>
<td>100.00%</td>
<td>4.56E+04</td>
<td>3.73E+04</td>
<td>1.09E+09</td>
<td>1.77E+09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10253</td>
<td>335000</td>
<td>1472542576</td>
<td>2</td>
<td>94.03%</td>
<td>98.51%</td>
<td>1.56E+04</td>
<td>1.54E+04</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54176</td>
<td>462000</td>
<td>295325749</td>
<td>1</td>
<td>100.00%</td>
<td>0.00%</td>
<td>1.65E+05</td>
<td>1.75E+04</td>
<td>3.68E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6012</td>
<td>836500</td>
<td>8061124443</td>
<td>6</td>
<td>0.24%</td>
<td>99.76%</td>
<td>1.65E+06</td>
<td>1.77E+04</td>
<td>2.95E+08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10261</td>
<td>335000</td>
<td>1472542576</td>
<td>2</td>
<td>95.52%</td>
<td>100.00%</td>
<td>1.56E+04</td>
<td>3.94E-04</td>
<td>1.34E+10</td>
<td>6.91E+09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10259</td>
<td>255000</td>
<td>1472541830</td>
<td>2</td>
<td>94.12%</td>
<td>100.00%</td>
<td>1.56E+04</td>
<td>3.94E-04</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47752</td>
<td>129500</td>
<td>7269460956</td>
<td>6</td>
<td>1.54%</td>
<td>99.23%</td>
<td>2.45E+05</td>
<td>1.92E+05</td>
<td>7.36E+08</td>
<td>7.11E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10255</td>
<td>270000</td>
<td>1472541830</td>
<td>2</td>
<td>92.59%</td>
<td>100.00%</td>
<td>2.00E+04</td>
<td>1.64E+05</td>
<td>8.74E+08</td>
<td>1.11E+10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47512</td>
<td>720000</td>
<td>6120350852</td>
<td>7</td>
<td>0.69%</td>
<td>98.61%</td>
<td>1.08E+06</td>
<td>8.19E+09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10247</td>
<td>495000</td>
<td>8192000000</td>
<td>1</td>
<td>100.00%</td>
<td>0.00%</td>
<td>8.19E+09</td>
<td>8.19E+09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10251</td>
<td>400000</td>
<td>8192000000</td>
<td>1</td>
<td>100.00%</td>
<td>0.00%</td>
<td>8.19E+09</td>
<td>8.19E+09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10233</td>
<td>325000</td>
<td>902803474</td>
<td>1</td>
<td>100.00%</td>
<td>0.00%</td>
<td>9.82E+08</td>
<td>9.82E+08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10252</td>
<td>340000</td>
<td>1472542576</td>
<td>3</td>
<td>16.18%</td>
<td>98.53%</td>
<td>1.25E+04</td>
<td>2.50E+03</td>
<td>4.91E-08</td>
<td>4.33E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10250</td>
<td>380000</td>
<td>1472542576</td>
<td>3</td>
<td>13.16%</td>
<td>98.68%</td>
<td>1.25E+04</td>
<td>2.50E+03</td>
<td>4.91E-08</td>
<td>4.33E+08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55335</td>
<td>168500</td>
<td>8733177695</td>
<td>7</td>
<td>2.97%</td>
<td>100.00%</td>
<td>2.73E+05</td>
<td>2.50E+03</td>
<td>1.25E-09</td>
<td>2.89E-09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Statistical analysis of ATOS data.
duration in ms, operations in occurrence number