Dynamic Scheduling of Real-Time Tasks on Multicore Architectures
Thomas Megel, Damien Chabrol, Vincent David, Christian Fraboul

To cite this version:

HAL Id: cea-00451284
https://hal-cea.archives-ouvertes.fr/cea-00451284
Submitted on 29 Jan 2010

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.
Dynamic Scheduling of Real-Time Tasks on Multicore Architectures

Thomas Megel, Vincent David, Damien Chabrol
CEA LIST, Embedded Real Time Systems Laboratory
Point Courrier 94, Gif-sur-Yvette, F-91191 France
Email: Firstname.Lastname@cea.fr

Christian Fraboul
IRIT - ENSEEIHT, Toulouse University
2, rue Camichel 31000 Toulouse - France
Email: Christian.Fraboul@enseeiht.fr

Abstract—We present a new dynamic scheduling on multicore architectures. This is an improvement of the Optimal Finish Time (OFT) scheduler introduced by Lemerre[7] reducing preemptions. Our result is compared with other schedulers and we show that our algorithm can handle with more general scheduling problems.

I. INTRODUCTION

We are interested in embedded multiprocessor architecture for real-time systems, i.e. dealing with safety-critical systems, for industrial purposes such as avionics, automotive or nuclear industry or domotics. We present a dynamic optimal scheduling for real-time tasks which means that if the task set is feasible, our scheduling meets all deadlines. It applies to tasks with same start time and deadline, periodic and time-triggered tasks and can handle with non-predictable tasks thanks to its dynamic (online) behaviour and optimal finish time characteristic. After some definitions in section 2, we present our optimal algorithm in section 3 and conclude in section 4 on ongoing works.

II. DEFINITIONS

We consider a system task \( \Gamma \) composed of several tasks. A task \( T \) releases several finite or infinite consecutive jobs. Each job \( J \) is characterized by \( J.r \) its release time, \( J.d \) its relative deadline and \( J.e \) its worst-case execution time. Task parallelism allows jobs to be executed in parallel on different cores/processors. We use in a first step tasks with same start time and deadline (i.e. same \( J.r \) and \( J.d \) \( \forall J \)).

Reconfiguration This is an operation changing the task system currently executed. Its includes global reconfiguration (migration) for distributed systems, when there is a memory transfer from one memory to another, to execute a job in a different node (core, processor). We call local reconfiguration in shared memory systems, when job is just running in a different processor/core.

III. OPTIMAL SMP REAL-TIME SCHEDULING

Given a set \( Q \) of \( N \) independant jobs with same release time and deadline on \( M \) identical processors sharing the same memory. We propose a real-time scheduler based on the Lemerre’s[7] algorithm with constant time complexity: intra-job parallelism is forbidden here, preemption and local reconfiguration are allowed but their costs are not considered.

A. Algorithm description

\( Q \) is a deque containing jobs ordered by increasing durations. \( T_T \) is the interval duration and we suppose that \( N>M \). The first part is a setup phase to set the \( M \) smallest jobs of \( Q \) on the \( M \) processors and remove them from \( Q \). The jobs currently in execution remain active until its ending or until the biggest jobs of \( Q \) become urgent (i.e. when its laxity are null: see for example job \( J_8 \) at \( t=1, J_7 \) at \( t=3 \) in Figure 1). We exclusively reserved processors to them by stopping the biggest jobs currently executing (they return then in the first places of \( Q \)). This is the main difference between Lemerre’s algorithm which preferred to stop the smallest jobs currently executing and migrate them on others processors. We do less operations and reduce preemptions by a factor 2. Our algorithm builds the schedule incrementally, determining the next preemption instant. The next instant is calculated by the minimum between remaining time of the smallest job in execution and laxity of the biggest job (the first value is 1 in example of Figure 1). If there are no urgent jobs, when a job is ending, we replace it by the next first waiting job from \( Q \) (see job \( J_4 \) replace \( J_1 \) at \( t=1.5 \) in Figure 1).

B. Scheduling example

Figure 1: an example of an execution of our scheduler for 8 jobs (\( J_1, J_2, J_3, J_4, J_5, J_6, J_7, J_8 \)) sorted by increasing
durations, respectively (1.5,2,4,4,5,7,9) (the sorting is made off-line) on 4 identical multiprocessors sharing the same memory (there are 2 preemptions and 2 local reconfigurations). Job duration is given by its weight multiplying by the interval duration (\( T_\gamma = 10 \)).

**Property** Reconfiguration and preemption only happen when a local laxity of the biggest job is null, but at each time we use exclusively a processor for it, so there is at most \( M-1 \) preemptions and local reconfigurations. If there are more urgent jobs, it means that the task system was unfeasible.

<table>
<thead>
<tr>
<th>Algo.</th>
<th>Preemptions</th>
<th>Online complexity</th>
<th>Alloc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>McN[8]</td>
<td>( \leq M-1 )</td>
<td>( O(N) )</td>
<td>static</td>
</tr>
<tr>
<td>Level[6]</td>
<td>( \leq N^2(M-1) )</td>
<td>( O(MN^2) )</td>
<td>dynamic</td>
</tr>
<tr>
<td>SCH[5]</td>
<td>( \leq M-1 )</td>
<td>( O(N+M\log M) )</td>
<td>dynamic</td>
</tr>
<tr>
<td>Lem.[7]</td>
<td>( \leq 2M-2 )</td>
<td>( O(1) )</td>
<td>dynamic</td>
</tr>
<tr>
<td>IZL</td>
<td>( \leq M-1 )</td>
<td>( O(1) )</td>
<td>dynamic</td>
</tr>
</tbody>
</table>

Table 1: Optimal scheduling comparative (tasks with same release time and deadline on identical processors; alloc(ation): static, dynamic), our scheduler is IZL (Incremental Zero Laxity).

We propose an improvement of Lemerre’s[7] algorithm which allows to be as competitive as MacNaughton’s one (bin-packing approach[8]) for static scheduling and better (in complexity) or similar (in number of preemptions) than Gonzalez’s[5] and Horvath’s[6] one for dynamic scheduling (see Table 1). We can use the same approach as in [7] to prove our algorithm.

**C. Applications to more complex models**

We use now the implicit-deadline periodic model: a job \( J_i \) arriving periodically is characterized by a period \( p_i \), a deadline equivalent to the period and an execution time \( \tau_i \); \( J_i = (\tau_i, p_i) \). We define the hyperperiod as the least common multiple of all jobs periods from the periodic task system \( \Gamma \), \( H_\Gamma = \text{lcm}(p_i, \forall J_i \in \Gamma) \). The job utilization is \( u_i = \tau_i / p_i \).

Given a set of \( N \) periodic independent jobs on \( M \) identical processors sharing the same memory. We decompose the hyperperiod into intervals to schedule job sets. Each interval starts at the end of the last interval and finishes when the first job deadline is met, each interval begins (respectively ends) at job boundaries. To construct a list of jobs, it simplifies the problem by setting the weight of each released job to its utilization (done off-line, in \( O(N) \) time). These weights are constant on each interval, only the duration of these intervals vary: we apply our algorithm on each interval which jobs have same release time and deadline.

Pfair class of algorithms[1][2][9] -dealing with periodic task model- has generally \( M \) preemptions per time quanta \( (M^2\text{lcm}(p_i)/\min(p_i) \text{ with BF}[9]) \) and at best linear complexity in run-time (with \( PD^2[1] \)). We propose an interesting alternative to the Pfair scheduling class (our online complexity: \( O(M) \) setup time then \( O(1) \) per system call, preemptions: \( M-1 \) max. per interval).

These techniques are not exclusively used for periodic task model and can be extended to time-triggered tasks (such OASIS task model[3]) even if weights are different between two intervals. It is possible because the next boundary job is known (one cannot schedule on-line in multiprocessors architecture without a priori knowledge of the next jobs characteristics[4]).

**IV. Conclusion and Future Work**

Our goal of our research is to propose real-time energy-efficient scheduling for embedded many-core architecture with more general recurring task model which can be more appropriate for industrial purpose and to take into account the new trends, for example using processor/core groups. The algorithm presented here aims to minimize preemptions in shared memory systems. We are also working on dynamic scheduling with hierarchical memory and allowing both local and global reconfiguration. We would like to define how to evaluate costs for both of them. Another way is to work on thread parallelism: it is a part of a job \( J_i \) executed simultaneously on several coresprocessors sharing the same memory. It allows intra-job parallelism: it is commonly forbidden in real-time scheduling but the trend may change with parallel computing (OpenMP, MPI). This is a new research field, and it would be interesting to handle with. Moreover, we would take into account fault-tolerance in critical systems, how to define an execution and architecture model allowing error detections including recovering techniques.

**References**