A Hardware Time Manager Implementation for the Xenomai Real-Time Kernel of Embedded Linux

Pierre Olivier, Jalil Boukhobza

A Hardware Time Manager Implementation for the *Xenomai* Real-Time Kernel of Embedded Linux

Pierre Olivier
Université Européenne de Bretagne, France
Université de Brest; CNRS, UMR 3192
Lab-STICC,
20 avenue Le Gorgeu, 29285 Brest, France
pierre.olivier@univ-brest.fr

Jalil Boukhobza
Université Européenne de Bretagne, France
Université de Brest; CNRS, UMR 3192
Lab-STICC,
20 avenue Le Gorgeu, 29285 Brest, France
boukhobza@univ-brest.fr

ABSTRACT

Nowadays, the use of embedded operating systems in different embedded projects is subject to a tremendous growth. Embedded Linux is becoming one of those most popular EOSs due to its modularity, efficiency, reliability, and cost. One way to make it hard real-time is to include a real-time kernel like Xenomai. One of the key characteristics of a Real-Time Operating System (RTOS) is its ability to meet execution time deadlines deterministically. So, the more precise and flexible the time management can be, the better it can handle efficiently the determinism for different embedded applications. RTOS time precision is characterized by a specific periodic interrupt service controlled by a software time manager. The smaller the period of the interrupt, the better the precision of the RTOS, the more it overloads the CPU, and though reduces the overall efficiency of the RTOS. In this paper, we propose to drastically reduce these overheads by migrating the time management service of Xenomai into a configurable hardware component to relieve the CPU. The hardware component is implemented in a Field Programmable Gate Array coupled to the CPU. This work was achieved in a Master degree project where students could apprehend many fields of embedded systems: RTOS programming, hardware design, performance evaluation, etc.

Categories and Subject Descriptors

D.4.8 [**Operating Systems**]: Performances; C.3 [**Special Purpose and Application-based Systems**]: Real-time and Embedded Systems; B.6.1 [**Logic Design**]: Design Style—*logic arrays*; K.3.2 [**Computers and Education**]: Computer and Information Science Education

Keywords

Real-Time Operating Systems, Embedded Linux, Reconfigurable Architectures, FPGA, Embedded Systems

1. INTRODUCTION

Nowadays Real-Time Operating Systems (RTOS) are integrated into many embedded devices. Using an RTOS provides many benefits. First, by abstracting the hardware layer, the OS facilitates the programmers work, and the standardization of the software layer. Using an RTOS also brings an executive software platform, providing some important *services* : 1) *Task scheduling* ; 2) *Time management* : giving the RTOS the notion of time ; 3) *Inter-task communication and synchronization mechanisms* (IPC), etc. Even though RTOSs provide many advantages, the time they spend managing their own structures can be considered as *overhead* that RTOS designers should minimize.

RTOS are generally executed in a constrained environment. First, the RTOS must deal with traditional embedded constraints : limited hardware resources, especially in terms of memory footprint, generally limited computing power, energy consumption constraints, high reliability, and also fast Time To Market. Moreover, RTOS are subject to specific constraints, which are the *predictability* of their behavior, and the *determinism* of the tasks execution time.

RTOSs lie upon a hardware timer periodically generating an interrupt notifying the CPU of a generated clock tick. The corresponding handler is executed and calls the tick management function of the RTOS. Depending on the RTOS this function performs various tasks, but one can identify the following main jobs : 1) maintaining a **system time** variable, counting the number of elapsed clock ticks since the boot of the system ; and 2) notifying different timers that a clock tick has occurred, and performing particular actions for expired timers. Some of those timers are dedicated to manage **periodic tasks**.

A high resolution timer allows the system to be more precise and responsive. Nevertheless, it also means a tick management function executed more frequently, though a higher CPU load. Depending on the load, a weak CPU with a high resolution timer can generate indeterminism in time management and so missed deadlines [7]. A compromise is then to be made between performance and time precision.

In this context, several studies have proposed to migrate some RTOS services from a software toward a hardware implementation. This is performed for two main reasons: 1) A performance improvement, by benefiting from spatial and temporal (pipeline) parallelisms provided by such a hardware implementation. 2) To release the CPU from the corresponding service execution, thus reducing the generated overhead. Such improvements would allow systems design-

ers to use less efficient but less expensive, and power consuming components.

One of the RTOS services one can need to optimize and/or configure according to the application needs is the time manager. We propose to enhance this service in the Xenomai real-time Linux framework. We present a configurable hardware architecture for a simple time manager implemented on an FPGA circuit, connected to a CPU executing the Xenomai framework in an embedded Linux.

The *tick-less* mode of some OS allows them not to be interrupted at each timer tick, but only when needed. This is done by carefully loading one-shot timers and responding to various hardware interrupts. The work presented in this paper is done to enhance performance of *non-tick-less* RTOSs, which is the case of most EOS. As Xenomai provides both tick-less and non-tick-less mode, we used the non-tick-less mode.

This project was proposed in a second year Master (Software for Embedded Systems at the Université of Brest - France) course on embedded operating systems. The objective of the project was to cover many domains of embedded systems throughout one project: hardware architecting and programming (FPGA, Hardware Design Languages), real-time systems, embedded operating systems development, device driver and kernel programming and integration, performance evaluation and validation procedures, etc.

In this paper, we first introduce some state-of-the-art studies about hardware implementations of RTOS services. Next, we present the Xenomai real-time framework for embedded Linux. In the third section, we describe the architecture of the hardware time manager, and its integration in the system. We finally give some performance evaluation results before concluding.

2. STATE OF THE ART

A real-time operating system must be able to respond within a deterministic time. The latency generated by the use of the RTOS (*overhead*) must not interfere with the reactivity of the system. Some studies focus on migrating software services to hardware components in order to unload the CPU, thus reducing the RTOS related overhead thereby enhancing the applicative performance.

We can find many hardware implementations of the scheduler in the literature. In [6], the authors present a configurable hardware scheduler, supporting various scheduling policies : *Priority-based*, *Rate Monotonic*, and *Earliest Deadline First*. The *Spring Scheduling Co-Processor* (*SSCoP*) [2] is also a hardware implementation of the scheduler. This system shows an improvement factor of 6.5 as compared to the software version.

Real-time Task Manager (*RTM*) [5] is a component handling scheduling functions, but also time and event management. The authors show that OS (μC/OS-II, NOS) latency and system response time are considerably enhanced with RTM.

Finally, *Fastchart* [8] is a complete hardware real-time kernel. To obtain a fully deterministic EOS, the authors remove features such as CPU pipelines and cache, and DMA. As a software operating system running on such hardware is drastically slower, Fastchart is fully implemented in hardware.

To the best of our knowledge, no study has been realized on the migration of the time manager service of the Xenomai kernel on reconfigurable hardware (FPGA). This allows to have a configurable time manager that can be tuned and dimensioned according to application needs, should the hardware platform contain an FPGA circuit. One of the latest innovation of Intel is precisely the integration of an FPGA and an Atom processor into the same chip (Intel atom E6x5C series).

3. THE XENOMAI REAL-TIME FRAMEWORK FOR LINUX

In this section we briefly present the Xenomai real-time framework for Linux, and the Adaptive Domain Environment for Operating System (Adeos) layer, which allows Xenomai and Linux to run on the same hardware platform.

3.1 Adeos layer

Adeos [9] is a resource virtualization layer, allowing multiple entities called *domains*, that can be seen as complete operating systems, to run simultaneously on the same hardware platform [3].

Adeos domains can compete with each other for receiving system generated *events*. Those events can be incoming external (or virtually generated) interruptions, Linux system calls invocations, or various kernel-code-related events like context switches. Adeos introduces the *event pipeline*, which can be seen as a chain of domains of decreasing priority. The events are consequently propagated throughout the pipeline, distributed firstly to the utmost priority domain, then distributed to lower priority domains.

In the case of a Xenomai RTOS running with a Linux kernel, the Adeos pipeline and the organization of the domains is depicted in Figure 1. In the pipeline, we can see that Xenomai has the highest priority, so it can handle and manage first the events before passing them to the Linux kernel. The events can also be blocked by the interrupt shield, preserving the real-time framework from latencies due to event management by the Linux kernel. Throughout this mechanism, Xenomai framework can provide real-time guarantees.

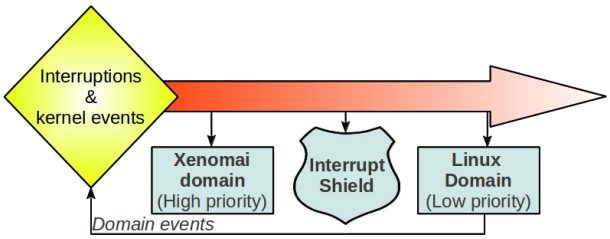

Figure 1: Adeos Event pipeline when running the Xenomai with Linux kernel (adapted from [3]).

3.2 Xenomai

Xenomai [4] provides a kernel-based Application Programming Interface (API) for real-time applications. A user-land API is also available, at the cost of longer latencies. Xenomai introduces the concept of *skins*. Skins are source codes emulating proprietary APIs used for porting real-time applications from various RTOSs such as *VxWorks*, *pSOS*, etc. to Xenomai. When designing a Xenomai application from scratch, the *native* Xenomai skin can be used.

All Xenomai skins rely on the *nucleus*, the core of the RTOS, implementing all algorithms for real-time functional-

ities. Xenomai provides all standard services one can expect to find in a RTOS[1] : task management, multiple scheduling algorithms, IPCs, etc.

3.3 Time management in Xenomai

We focused in this project on periodic real-time tasks that make extensive use of timers. A commonly used skeleton for those tasks in Xenomai as follows :

```
void myRealTimeTask(void *arg) {
  SetPeriodic(myself, period);
  while(1) {
    /* Do something */
    wait_period();
  }
}
```

The `SetPeriodic()` function creates and starts a timer related to the calling task, with a period (in clock ticks) equals to the specified parameter. The global tick management function notifies the timer each time a clock tick occurs. When reaching wake up time, the timer executes a handler placing the task in the ready state.

4. ARCHITECTURE AND INTEGRATION OF THE HARDWARE TIME MANAGER

In this section we present the architecture of the proposed hardware time manager component, and the way it is integrated into the Xenomai framework.

From the RTOS point of view, the hardware time manager should provide the following basic time management operations :

- **GetTime** and **SetTime**: read/modify the value of the system time ;
- **TaskDelay**: load a counter of a delayed task for a given amount of clock ticks ;
- **GetTasksToWake**: obtain the identifiers of tasks that need to be awaken (if any) ;
- **ClearTask**: acknowledge from the CPU indicating the awakening of a task.

4.1 Hardware architecture and integration

The designed hardware time manager is composed of two main modules, which represent the two main functions we want to support : 1) maintaining the global system time and 2) allowing tasks to suspend their execution for a specific amount of time. Thus, the two main modules composing the hardware time manager are the system time module and the waiting tasks array as seen in Figure 2.

The *system time module* is a simple counter which value is initialized to zero when the system starts (i.e. when the bitstream is loaded on the FPGA). This counter is then incremented each FPGA clock cycle.

The *array of waiting task module* is an array of counters. The number of counters it contains is equal to the number of tasks that can potentially be put in a waiting state (there is space for optimization for this module). When a task needs

[1]For more information about Xenomai, one can browse the Xenomai website : `http://www.xenomai.org`

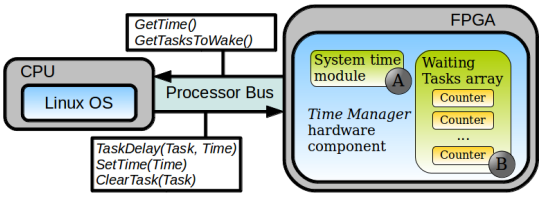

Figure 2: Block view of the hardware time manager and its integration in a Linux-based system.

to be put in a waiting state, the corresponding counter is loaded with the appropriate number of ticks. This counter is then decremented at each clock cycle.

All counters in the component are 64 bit width which corresponds to a very high frequency/precision counter. Counter size can be reduced according to the needed precision.

The architecture was realized in an incremental approach. Students were first provided with some simple hardware component to test, then a very simple counter, and finally the whole module was introduced.

4.2 Communication between CPU and IP

Communication between the CPU (executing the operating system) and the hardware time manager is performed through a register-based interface.

On our evaluation board, the FPGA chip is directly connected to the CPU through a dedicated bus. We defined some control registers (i.e. FPGA registers) in the hardware part that were mapped to the device driver's process address space. A kernel driver was written in C, as well as a user-space one. The memory mapping is performed using the `mmap()` system call in user-land, and `ioremap()` in kernel space. One function is provided for each basic operations supported by the hardware component.

When the counter corresponding to a waiting task reaches zero, the related task needs to be awakened at the OS level. The component then triggers an (hardware) output signal at a high level and waits for an acknowledgment from the CPU meaning that the task is actually awake. This specific signal is plugged on the interrupt output of the CPU.

Here again, students interfaced the hardware component with the RTOS part in an incremental approach. They first insured a communication in user-land before exploring the kernel code and integrating the driver.

4.2.1 Integration into the Xenomai Nucleus

All the code modifications and the integration of the hardware time manager calls were performed at the *Nucleus* level. Doing so, we ensured the compatibility of the performed modifications with 1) all Xenomai skins and 2) kernel and user-land based Xenomai real-time applications. We included our driver into the Nucleus code, and modified the `wait_period()` primitive, which now calls the hardware time manager. This allows to launch a timer with an initial value corresponding to the task period, and then suspends the calling task.

When the delay is elapsed, the hardware component produces an interrupt signal. Thus, we implemented a handler executed each time the interrupt is received. This function retrieves the identifier of the task(s) needing to be woke up. Then, it places it (them) in the ready state after sending an acknowledgment to the component indicating that the task

is actually awaken.

5. PERFORMANCE EVALUATION

In this section, after introducing the evaluation platform, we give some results describing the benefits gained from replacing the software time manager service by the hardware version. We implemented the hardware time manager on a development board and measured the time manager latencies. From these results we computed the corresponding CPU load for a given set of tasks. We also present the cost of the hardware component, in terms of FPGA resources. In the following section we refer to the software default Xenomai time manager as the *software mode*, as opposed to the *hardware mode*.

5.1 Evaluation platform

Our evaluations were achieved on the *Armadeus Systems APF27* development board [1]. It is equipped with a *Freescale i.MX27* microprocessor clocked at a frequency of 400 Mhz. This processor is coupled to a *Xilinx Spartan 3A* FPGA chip on which we synthesized a small version of the hardware time manager. This component is able to manage up to 12 real-time tasks. The hardware time manager was clocked at a frequency of 102 MHz, giving one tick every 9.8 ns. For the software part we used the native Xenomai skin, in a **non-tickless mode**, with a base period of 10 ms. We used the user-land mode for the defined Xenomai real-time tasks.

5.2 Performance evaluation results

In this part, we investigate the performance of the proposed design throughout the reactivity, saved CPU load, and the FPGA resource cost.

5.2.1 Reactivity and execution time

Time measurement methodology.

In the next sections, we present results based on various time measurements. Those measurements were made, using the `GetTime` operation implemented by the hardware time manager as follows :

```
/* 1. Measure the calibration time */
c1 = GetTime(); c2 = GetTime();
calibration_value = c2 - c1;
/* 2. Measure the execution time */
t1 = GetTime();
call_to_measured_operations();
t2 = GetTime();
result = (t2 - t1) - calibration_value;
```

By subtracting the calibration value from the measured time, we took off the overhead due to the `GetTime` function itself (measured by two consecutive GetTime operations). Each time measure were performed 10 times and we considered the average value. The measures were always performed just after the system boot, thus insuring the same initial conditions.The hardware time manager bitstream were loaded at boot time (U-Boot) before the kernel starts up.

System reactivity.

We investigate the system reactivity by quantifying the imprecision between the moment a task needs to be woke up (i.e. the occurrence of the tick corresponding to the end of its delay period) and the moment this task is really woke up. To do so, we measured the execution times of the tick management function in software mode, and compared them to the execution time of the task wake up interrupt handler in hardware mode. Results for different number of tasks are shown in Figure 3.

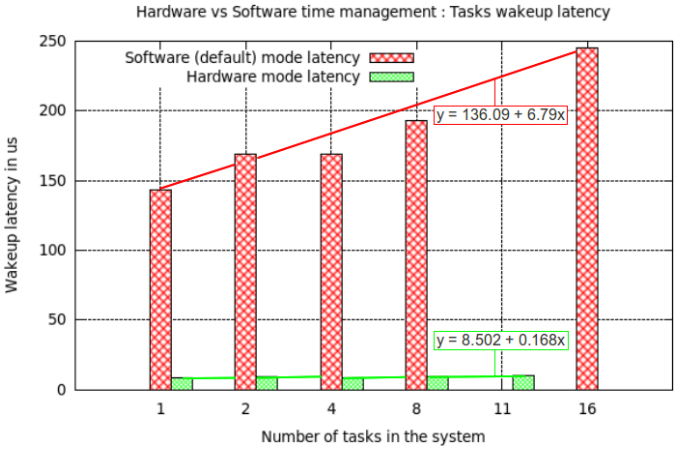

Figure 3: System reactivity, wake-up latency per task.

We can see that in software mode, the task wake-up latency is highly related to the number of tasks in the system. Indeed, at each clock cycle the master timer handler must inform all the delayed tasks timers that a clock cycle occurred. In the hardware mode, this latency is much less dependent on the number of tasks (see the estimated slope equation), thus making it more stable. The inferred wake up latencies are highlighted in Figure 3.

Offloading the CPU.

Based on previously measured values (see Figure 3), we can estimate the CPU overhead for delayed tasks management in both software and hardware modes for a duration of *one second*. We assumed a master timer base period of 10 ms in the software mode. In order to simplify the figure, we also assumed that all tasks have the same period (no impact on the results).

In software mode, we computed the CPU time dedicated to the delayed tasks management using the estimation equation in Figure 3 and multiplying the result by the number of clock ticks in one seconds : 100 (1 sec divided by 10 ms) in our case. In hardware mode, this time value corresponds to the number of task awakenings in a 1 second period multiplied by the time measured of the task wake up interrupt handler (in Figure 3). In Figure 4, we present the improvement factor (speedup) given by using the hardware mode. This speedup is obtained by dividing the time for task delay management in software mode by the time taken in hardware mode.

We can observe that the speedup is very high when the number of task wake-ups is low (i.e. the task period is long), because in hardware mode we spend CPU time only when needed (interrupt driven) in order to wake up tasks. Conversely, in software (polling) mode, we have to iterate over the delayed tasks timer list at each timer tick. One must keep in mind that for the hardware timer the precision is 3

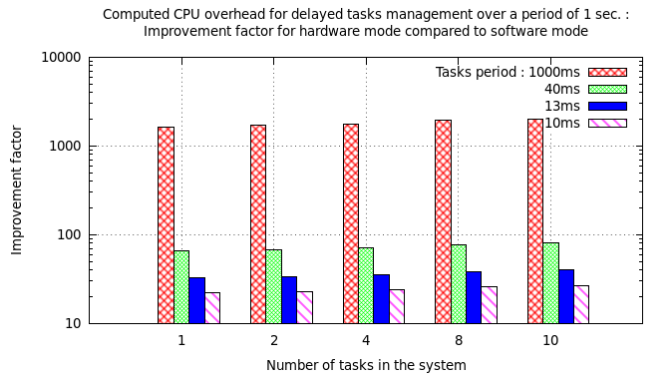

Figure 4: Improvement factor for the CPU overhead due to delayed tasks management. Improvement factor = (CPU overhead in software) / (CPU overhead in hardware)

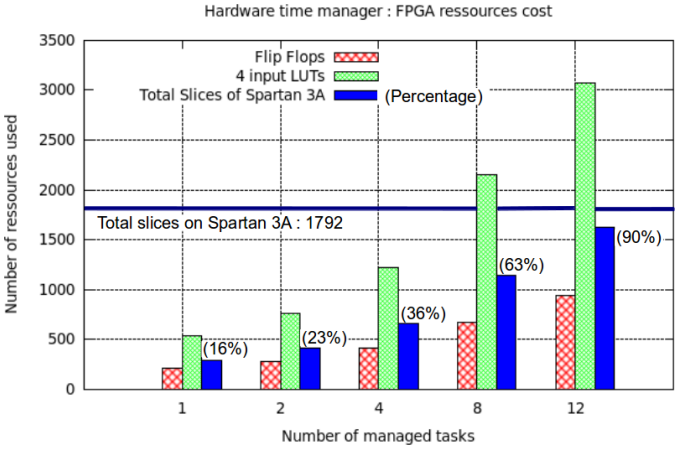

Figure 5: FPGA resources cost for various versions of the hardware timer according to the number of managed tasks

orders of magnitude better than the software version for all the performed measures.

5.2.2 *FPGA hardware resource cost*

The FPGA resource cost is given in terms of FPGA 4 input LUTs, and flip-flop slices used by the hardware time manager on the Spartan3A chip. To obtain these values, we lied upon the outputs of the Xilinx ISE design suite. We synthesized various versions of the component, each one able to manage a given number of tasks and we studied the variation of resource utilization. Even though the given implementation, students worked on, is very naive and can be substantially optimized, we found it interesting to show the results.

Results about FPGA resource costs are presented in Figure 5. A maximum of 12 tasks managed on the Spartan3 may seem low, indeed we presented the timer with the higher granularity (64 bit counters). The cost in terms of hardware resources for the component can be optimized by reducing the granularity of the timer or optimizing the architecture.

6. CONCLUSION AND FUTURE WORKS

In this study we presented a simple architecture model for a flexible/configurable hardware time manager component, designed to replace the software version of the Xenomai kernel of embedded Linux. As our performance evaluation shows, this implementation drastically enhances the reactivity of the system by more than a factor 10, and allows to have a flexible and extremely precise timer for real-time tasks. Furthermore, we also considerably reduced the CPU overhead due to the delayed tasks management.This is not without a quantified FPGA hardware cost, but the flexibility and precision offered by such a component can be worth it. This work can be extended by testing the reconfigurability feature of some FPGAs (not the Spartan3A) to dynamically reconfigure the timer in terms of precision and number of managed tasks according to the application evolution. We plan to reduce the cost of the hardware component in terms of FPGA resources, in order to be able to manage a more important number of tasks

This project was given to Master degree students to introduce them to embedded systems research. It was well received by students and considered as challenging as it allowed them to apprehend many domains of embedded systems such as: 1) embedded Linux tools handling, 2) Xenomai real time kernel installation and use, 3) device driver development and integration, 4) kernel code exploration and programming, 5) hardware programming and interfacing, and 6) performance evaluation and validation.

All the hardware VHDL and component driver sources together with a Xenomai patch will be available on-line.

7. REFERENCES

[1] Armadeus Systems. Apf27 board datasheet, 2012. `http://www.armadeus.com/_downloads/apf27/documentation/datasheet_apf27.pdf`.

[2] W. Burleson, J. Ko, D. Niehaus, K. Ramamritham, J. A. Stankovic, G. Wallace, and C. Weems. The spring scheduling co-processor: A scheduling accelerator. In *IEEE Transactions on VLSI*, 1993.

[3] P. Gerum. *Life with adeos*. 2005.

[4] J. Kiszka. The real-time driver model and first applications. In *7th Real-Time Linux Workshop, Lille, France*, 2005.

[5] P. Kohout, B. Ganesh, and B. Jacob. Hardware support for real-time operating systems. In *Proceedings of the 1st IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis*, CODES+ISSS '03, pages 45–51, New York, NY, USA, 2003. ACM.

[6] P. Kuacharoen, M. Shalan, and V. M. A configurable hardware scheduler for real-time systems. In *in Proceedings of the International Conference on Engineering of Reconfigurable Systems and Algorithms*, pages 96–101. CSREA Press, 2003.

[7] J. Labrosse. *MicroC/OS-II: the real-time kernel*. Newnes, 2002.

[8] L. Lindh. Fastchart-a fast time deterministic cpu and hardware based real-time-kernel. In *Real Time Systems, 1991. Proceedings., Euromicro '91 Workshop on*, pages 36 –40, jun 1991.

[9] K. Yaghmour. Adaptive domain environment for operating systems. *Opersys inc*, 2001.