in Proc. of Hawaii International Conference on System Sciences,
Software Technology Track, Jan. 1996.

Design of a Real-Time Co-Operating System for Multiprocessor Workstations

Gebran Krikor, Md. Touhidur Raza, David B. Stewart

Dept. of Electrical Engineering
Institute for Advanced Computer Studies
Institute for Systems Research
University of Maryland, College Park, MD 20742
dstewart@eng.umd.edu; http://www.ee.umd.edu/serts

Abstract*: We have designed a Real-Time Co-Operating System (RTCOS) for simultaneously supporting real-time and non-real-time activities on a workstation with two or more processors. The RTCOS is the software equivalent of a co-processor, with a software architecture analogous to the hardware architecture that has been used in many workstations and personal computers. In this paper, we discuss our first software prototype of the RTCOS, which co-exists with Solaris 2.4 on a four-processor Sun SPARCstation 20. We summarize the feasibility of our approach through an experimental characterization of Solaris 2.4. We address the technical issues involved and present the details of our design. The RTCOS is targeted towards real-time applications in the sensor-based control, process control, signal processing, multimedia, and manufacturing domains.*

1. Introduction

We have designed a Real-Time Co-Operating System (RTCOS) for simultaneously supporting real-time and non-real-time activities on a workstation with two or more processors. In this paper, we discuss the design issues, present our solutions, and provide details and preliminary performance results from our first software prototype of the RTCOS. The RTCOS is targeted towards real-time applications in the sensor-based control, process control, signal processing, multimedia, and manufacturing domains.

The RTCOS is the software equivalent of a co-processor, with a software architecture analogous to the hardware architecture that has been used in many workstations and personal computers. These computers have a general purpose processor, such as a SPARC, i486, or M68040. In addition, the computer has one or more co-processors which provide a faster and more efficient means for performing specific tasks. Examples of hardware co-processors include floating point units, graphic accelerators, and digital signal processors.

New multiprocessor workstations, such as the SPARCstation 20 Model 514MP [15] and the Compaq Pentium-based ProLiant 4000 Multiprocessor [1], have several general purpose processors with a distributed shared memory architecture. Solaris 2 is an enhanced version of System V R4 UNIX which has been extended to support these multi-processor platforms. By default, Solaris treats all processors as equal. The goal of the RTCOS is to transform some of the general purpose processors into real-time processing units (RTPUs) which can support both hard and soft real-time applications which require a time resolution of less than one millisecond. Processors not assigned as RTPUs continue to execute time-sharing UNIX applications, without interfering with execution of real-time tasks on the RTPUs.

1.1 Motivation

The primary objective of the RTCOS is to create a real-time operating system environment for desktop workstations and personal computers (both of which are herein referred to as just *workstations*) which will support dynamically reconfigurable and reusable software [12].

Although a workstation is not usually used as the computer of choice for embedded real-time systems, there are many situations for which a workstation is the preferred solution for real-time applications:

Cost: For non-embedded systems, such as many robots, process controllers, and production lines, workstations provide a low cost alternative to expensive real-time hardware.

Performance: Workstations represent the largest market for computing, and as a result, have the most investment into increasing the speed of the computers. As a result, these computers provide the best performance-to-cost ratio.

Education: Workstations are widely available, often in clusters at a university. This provides the opportunity for improved education through hands-on experimentation with real-time systems at the college level, thus better preparing students for the work force.

Software Base: Workstations have a large, evolving software base which uses state-of-the-art technology for such things as editors, debuggers, compilers, graphical user interfaces, software libraries, and CASE tools. They also host distributed file systems, provide security, and offer a variety of networking capabilities. It is highly desirable to leverage as much of this technology as possible for creating equivalent state-of-the-art programming environments for real-time applications, building upon this software base of non-real-time software.

Ubiquitous Computing: There is a technological strive for obtaining a single computer system that can perform all the needs of the users, ranging from non-real-time information system applications to real-time signal processing for multimedia applications such as video conferencing, interactive television, and virtual laboratories. Workstations are the prime candidates for providing the ubiquitous platform, if they can provide the necessary predictability and performance for real-time applications.

There have been several efforts to use off-the-shelf workstation hardware for real-time systems. For example, real-time operating systems (RTOS) such as LynxOS [5] and QNX [8] were designed to use general purpose hardware such as the Intel i486-based computers. However, these systems require that the RTOS take full control of the computer, and completely replace the popular DOS, Windows, or UNIX operating systems. Therefore the workstation can be used for either real-time or non-real-time, but not for both simultaneously.

Other efforts, such as Real-Time Mach [16] and Solaris 2 [15] are UNIX-based operating systems with real-time extensions. There is an attempt to merge real-time tasks with non-real-time tasks in a UNIX environment, by giving the real-time tasks higher priority and performing such procedures as locking pages in memory and reducing the interrupt response time by moving towards a microkernel software architecture.

UNIX-based RTOS with real-time extensions, however, have failed to reach the performance or robustness level of RTOS which have been designed specifically for real-time hardware. Successful RTOS include commercial software such as VxWorks [17], OS-9 [7], and VRTX [6], as well as research projects including Chimera [13] and the Spring Kernel [9]. These systems all require dedicated computers, generally in the form of single-board-computers in an open-architecture backplane. Due to the complexity of these setups, their availability is low and cost is relatively high as compared to the cost of workstations with similar computational power. In some cases, such as with the Spring Kernel, custom hardware is recommended in order to handle the high operating system overhead to provide guaranteed hard real-time performance [9].

Our approach is distinctly different from any of the approaches listed above, as it does not require taking over the complete workstation (such as QNX, VxWorks, Chimera, and Spring), nor is it obtained by modifying or providing extensions to an existing non-real-time operating system (NRTOS).

Rather, our design is based on the co-existence of two separate operating systems that share the same architectural platform and co-operate in order to provide simultaneous real-time and non-real-time support.

The expected performance and predictability of the RTCOS more closely resembles that obtained by RTOS with dedicated hardware, such as QNX, VxWorks, Chimera, and the Spring Kernel. However, the RTCOS is designed within the constraints imposed by the NRTOS, and thus performance is achieved without the same flexibility as available for the dedicated RTOS. A preliminary performance analysis of the RTCOS convinces us that we can support real-time tasks on a Sun SPARCstation 20 with frequencies over 1000 cycles per second and with an accuracy in the tens of microseconds.

An additional major challenge for supporting real-time on workstations is to provide a software programming environment targeted towards the inexperienced users, rather than requiring highly specialized real-time system engineers to be the only programmers of the system. To address this issue, the RTCOS is being designed to support software assembly through visual programming which is based on the underlying design of dynamically reconfigurable and reusable software components.

In the next section, we present many of the design issues and the constraints imposed by the NRTOS and we introduce some of our solutions. The technical rationale for our design decisions and solutions is in Section 3.

2. Design issues and approach

The objective of our research is to provide the necessary set of RTOS services required to support the predictable execution of reconfigurable real-time tasks on a workstation, without sacrificing any of the workstation's non-real-time functionality.

Depending on the application domain, some real-time tasks may have frequencies greater than 1000 cycles per second, other tasks may require steady communication streams of several megabytes per second, while yet other tasks may interact with advanced graphical interfaces and have soft real-time requirements. Each of these types of real-time tasks has been considered.

In this section, we demonstrate the feasibility of the RTCOS approach, then present many of the technical issues that have been addressed in the design of the RTCOS.

2.1 Feasibility of workstation solution.

In order to determine the feasibility of using the co-operating system approach to obtain improved real-time performance and predictability on a workstation, we performed a detailed experimental characterization of the Solaris 2.4 kernel on a 4-processor SPARCstation 20, Model 514MP.

Our results are summarized in the graphs shown in Figures 1 through 8, and demonstrate the potential of the co-operating system approach. The graphs show the execution time of each cycle of a task τ which computes 100 whetstones per cycle for 500 cycles, under various conditions.

Figure 1 shows the execution time for τ when executing as a non-real-time process on a single processor, in the presence of a typical load in the system. A typical load includes running X windows and a variety of other I/O and CPU-bound jobs. Note the large swing in execution time.

Figure 2 shows the same workload, but with τ scheduled as a real-time Solaris thread. There is a drastic improvement in the execution time of the task. Execution time, however, still varies, which can be seen more clearly in the same but rescaled graph shown in Figure 3. In some cases, the task takes an additional millisecond or more to execute its cycle, due to interrupts and other operating system functions executing at higher priority than the real-time tasks. These interrupts are often side effects of *lower-priority* tasks. For example, a non-real-time process performs a file system *write*, and the processor receives a high-priority interrupt when the *write* is complete, thus interrupting execution of the real-time thread.

For comparison, the ideal execution of task τ is shown in Figure 4. This theoretical best case represents a mathematically generated result if task τ executed in exactly 3.7 msec every cycle, without any processor contention from interrupts, the system clock, or other threads. Figure 4 is *not* an experimental result.

When all four processors are enabled with the Solaris kernel running on each one, real-time execution is not improved, as shown in Figure 5. Solaris load-balances its high-priority interrupt handling and scheduling operations, and the system clock continues to interrupt all processors. This graph shows that simply adding processors to a workstation does not improve real-time performance and predictability.

A significant improvement can be obtained by binding the real-time task τ onto one of the processors, and preventing all other processes from using that processor. The result of this binding is shown in Figure 6. The drawback of doing this in Solaris is that you can then only place one process per additional processor, and thus only run three real-time tasks on a four-processor system. Nevertheless, this graph demonstrates a result that can be applied to the design of our RTCOS, as described later.

The small spikes in Figure 6, which are approximately 80 μsec each, are a result of the system clock still interrupting the processor every 10 msec, and load balancing the Solaris dispatcher and scheduler on every fourth system clock interrupt.

The interrupt overhead shown in the graph in Figure 6 can be eliminated by removing this processor from the pool of processors that can be used for scheduling tasks on the workstation. The result, which is very close to the ideal case shown in Figure 4, is shown in Figure 7. The main difficulty encountered with this case is that it requires a loadable kernel module to implement this feature, and requires super-user permission to install the module. As a result, it is not a viable option for most Solaris users. This approach, however, can be used to install an RTCOS microkernel, and forms part of our design.

To obtain Figure 7, only one Solaris process was bound to the processor, and all other processes prevented from using this processor. As a result, there is never a need to call the local scheduler. If lower-priority processes are allowed to use the processor, then although this processor is not used for running the Solaris dispatcher, the local scheduler can be called, only to realize that there is a real-time process executing. As a result, the real-time process may get interrupted, with the result shown in Figure 8. It is therefore desirable to bind and execute only one real-time Solaris process per processor.

These graphs demonstrate some of the capabilities and limitations of using the workstation for real-time execution, and also show what must be done to obtain predictable execution. The design of our RTCOS is based on the conditions that produce the predictability shown in Figure 7. The RTCOS microkernel executes as the only process on each RTPU. It performs its own thread management and uses in-process scheduling, similar to the methods used in creating the kernel for dedicated RTOS such as Chimera, VxWorks and Spring. The technical details of this approach are given later in Section 3.

2.2 Technical Issues

We have addressed many issues relating to the various aspects of design of any operating system. The remainder of this section presents these issues and an overview of our approach. Technical details of our solution for many of these issues are deferred until Section 3.

Memory Management: There are several memory-related issues which have been addressed, including dealing with the cache, virtual memory, and address spaces in a multiprocessor environment. The RTCOS deals with the SPARCstation's

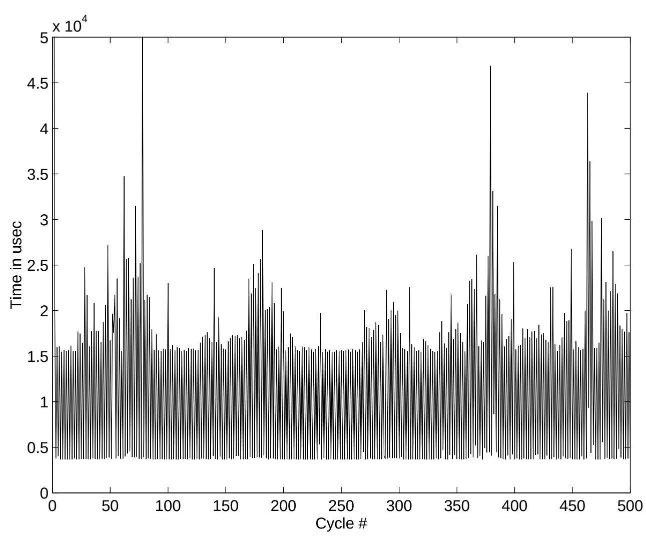

Figure 1: Execution time of each cycle of task τ when scheduled as a Solaris time-sharing thread. Real execution time of task is 3.7 msec.

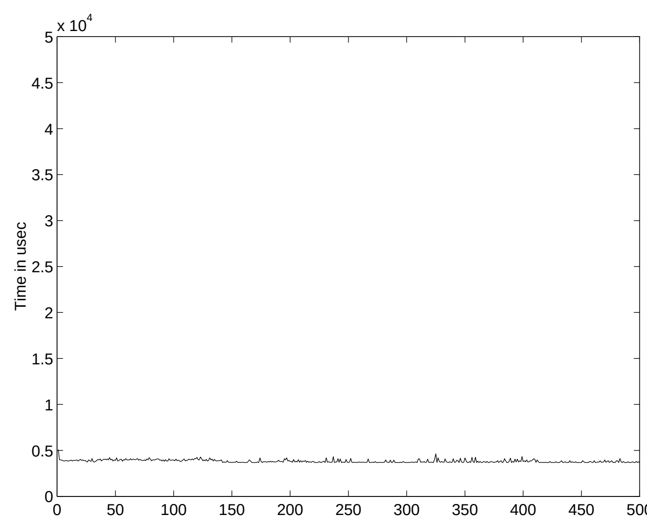

Figure 2: Execution time of each cycle of task τ when scheduled as a Solaris real-time thread on a single processor.

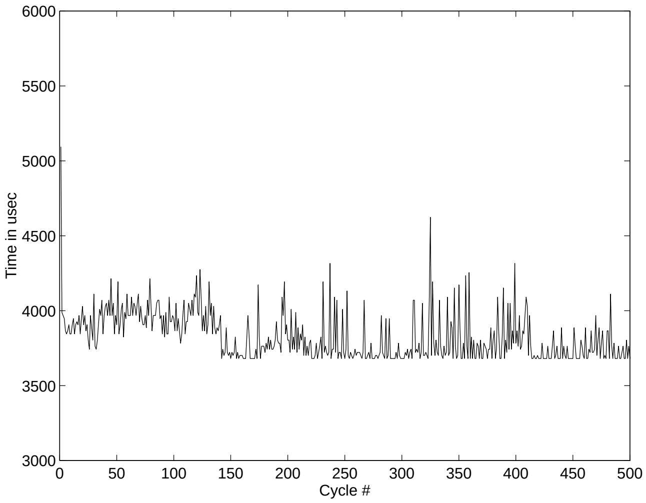

Figure 3: Zoom of graph shown in Figure 2. Note the variation in the execution time when scheduled as a real-time Solaris thread.

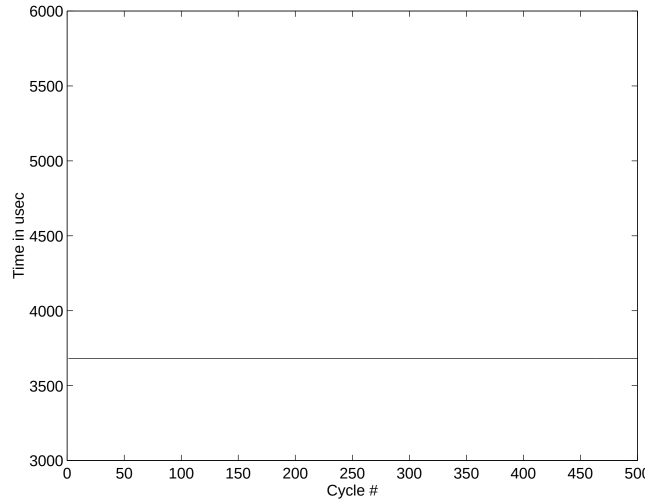

Figure 4: Ideal execution time for task τ. Note this graph was generated based on the real execution time of τ; it was not obtained experimentally.

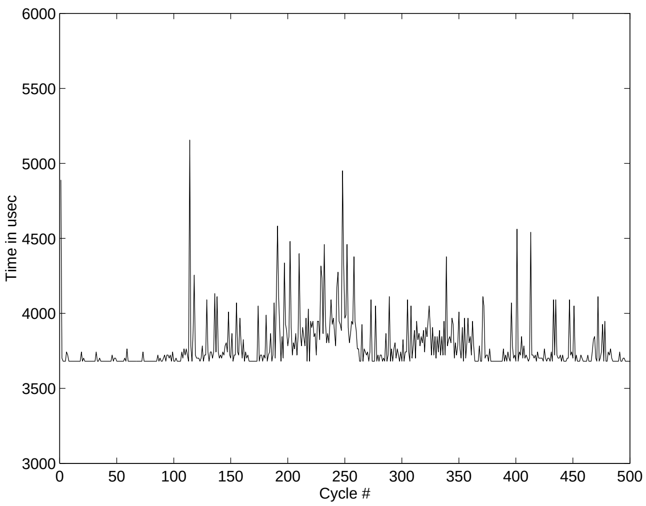

Figure 5: Execution time of each cycle of task τ when scheduled as a Solaris real-time thread in a multi-processor environment.

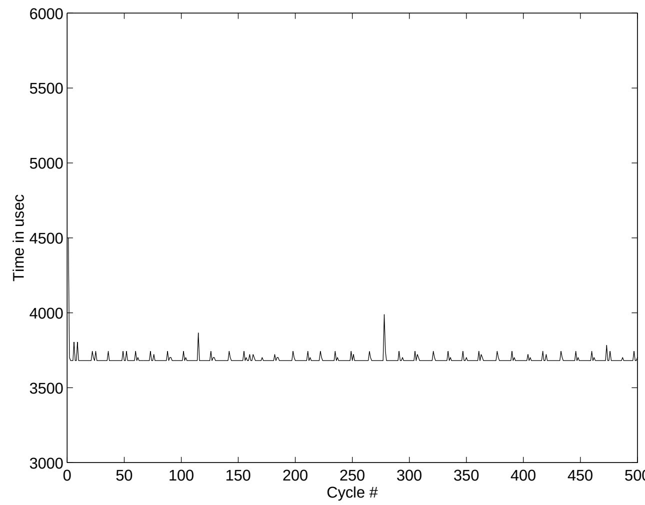

Figure 6: Execution time of each cycle of task τ when executed as only process on a separate processor from the non-real-time threads.

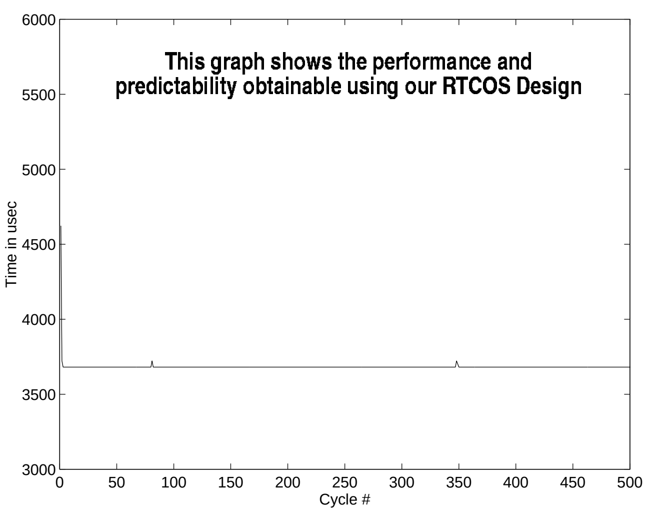

Figure 7: Execution of task τ when executed as only process on a separate processor, and interrupt threads disabled.

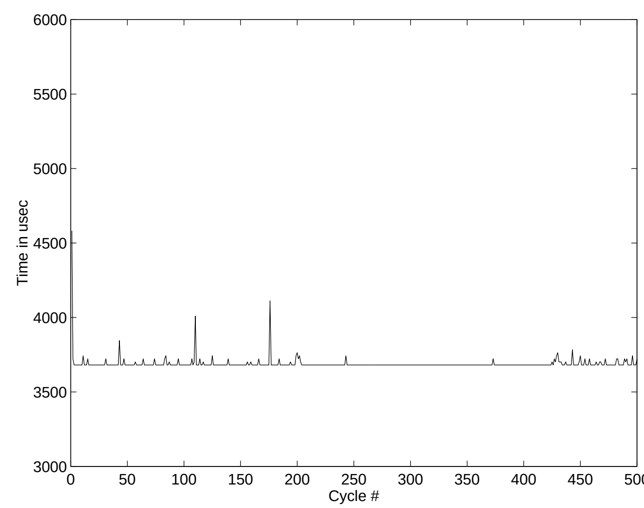

Figure 8: Effect of disabling interrupt threads from processor, but allowing other lower-priority real-time processes to also use the processor.

1MByte cache by treating it as a local memory, and limiting the size of real-time tasks which collectively execute on a processor to less than 1MByte. Virtual memory is circumvented by locking all cached and shared memory pages into real memory. Multiprocessor pointer addressing in the RTCOS is simplified by using a uniform 32-bit address space for all RTPUs. Details are given in Section 3.1.

Communication: Solaris 2.4 only has non-real-time interprocess communication (IPC) mechanisms. Although POSIX-compliant interfaces for real-time mechanisms are defined in the Solaris documentation, there currently is no implementation. To obtain predictable IPC, the Chimera IPC package was ported to our RTCOS. It includes mechanisms such as spin-locks, remote semaphores, prioritized typed message passing, and global state variable tables. Express mail is also implemented as the underlying IPC mechanism between the RTCOS microkernels executing on each RTPU. The RTCOS IPC mechanisms are further discussed in Section 3.2.

Resource Sharing: The file system, display, network, and other workstation hardware must all be shared by both the NRTOS and RTCOS. However, it is not acceptable to have real-time tasks block and wait for non-real-time tasks. We address this issue by adopting the global/local operating system separation as designed into the Chimera RTOS [13], and using express mail and remote procedure calls to implement system calls and an extended file system. We also improve upon the Chimera design by implementing the servers as multiple real-time Solaris threads on the NRTOS processor, as described in Section 3.3.

Scheduling: Real-time scheduling is always an issue for obtaining predictable execution. Our original design used a loadable Solaris scheduling module; however, for the reasons described in Section 2.1, we could not obtain the desired predictability, nor obtain a resolution better than the Solaris 10 msec system clock. In our revised design, we opted for in-process user-level scheduling with a policy/mechanism separation, which provides us with many advantages, as discussed in detail Section 3.5. A maximum-urgency-first scheduler [10] is used as the default RTCOS scheduler, as it provides the maximum flexibility by supporting both static, dynamic, andmixed priority scheduling, as well as support for guaranteed scheduling of both hard and soft real-time tasks. The RTCOS scheduler also provides deadline failure detection and handling. Continuous monitoring of real-time tasks is provided through an automatic task profiling mechanism with better than 10 µsec resolution [11], as described in Section 3.6.

Process Management: Solaris provides process management through lightweight processes (LWP). Each process allocates a set of LWPs used to schedule the threads it creates. The Solaris operating system only manages the LWPs and not the threads within a process. A LWP, upon being scheduled, chooses the highest priority thread from the set of runnable threads and executes that thread. Solaris only guarantees that enough threads will be active so that the process can continue to make progress, which is why many Solaris applications are designed with a one-to-one mapping between LWP and threads. These *bound threads* guarantee that a LWP will be available to execute the thread as soon as it becomes runnable. As highlighted in Section 3.5, more than one LWP on a processor reduces the predictability of the real-time tasks. Therefore, the RTCOS microkernel defines only a single LWP per processor, and manages its own threads internally, independent of Solaris. Solaris is instructed never to disturb the RTCOS LWP running on the processor except when the RTCOS is directly affected by an event, such as timer interrupt signalling a missed deadline.

Clocks and Timers: The RTCOS must be designed within the constraints of not only the hardware timers and clocks available, but also within the constraint of how the NRTOS has them programmed. In Section 3.5, we present a solution that maximizes timer resolution while minimizing overhead resulting from timer interrupts.

Global Error Handling: The RTCOS uses the Chimera *global error handling* facility for fault detection and handling. Solaris uses the standard UNIX mechanisms for error handling. However, as discussed in [11], it is preferable to support global error handling, such that exception handling code can be separated from the main code. Global error handling also removes much of the redundant error checking typical in most software systems, and is required to support the design of component-based software. The Chimera RTOS already supports global error handling, which ensures that return codes from system calls are always checked and errors handled appropriately. This improves application development and debugging times by ensuring proper error handling and reducing the amount of code needed to handle errors. If the default error handling is not sufficient, the application developer can specify the handler for a specific error or class of errors on a per-module or per-scope basis. This results in error handling code being separate from the main thread of execution, and allows error handling to be modularized and reused.

Portability: One of the major issues in developing the RTCOS is to create a programming environment that has minimal hardware dependence, and does not require custom versions of the NRTOS, which for our software prototype is Solaris 2.4. To maximize portability, we constrained our design space by not allowing any modifications to the NRTOS kernel or operating system modules. First, this ensures that all applications which run on the NRTOS can continue to run, even when the RTCOS is installed. Second, it ensures that the RTCOS can run with any version of the NRTOS without recompilation of the NRTOS. Third, the RTCOS never directly access any workstation hardware. Any hardware dependent accesses are implemented using a policy/mechanism separation, allowing the minimum amount of code to be modified if porting to a different hardware platform.

Security: Security is generally not an issue in RTOSs because real-time systems are generally single user applications. In Solaris, creating a real-time process requires superuser privileges, which limits the availability of real-time processing to a very few users. The requirement for superuser permissions is used to prevent one user from completely taking over all available processing time on the machine. The RTCOS package contains only one process which is run with superuser permissions. This process promotes itself to the highest priority Solaris real-time task in the system, allocates itself an infinite time quantum and then executes all user tasks with normal user priority. The RTCOS also prevents real-time tasks from being created on all of the processors in the system, which ensures that the system is always in an operational state.

Reconfigurable and reusable software: the primary objective of the RTCOS is to provide a computing environment to support dynamically reconfigurable and predictable real-time applications. The first aspect of reusing software is to allow the RTCOS to use the same software libraries that are available to Solaris real-time threads. Reconfigurable software at the operating system level is obtained by supporting loadable modules and dynamically bound policy/mechanism separation for operating system communication and scheduling functions. Reconfigurable software at the user level is obtained by supporting the port-based object abstraction defined in the Chimera Methodology [12]. Chimera provides a set of library functions for supporting reconfigurable software, such as the *cfig()* utility, which allows for rapid development of code for reading configuration files. The set of libraries available to the RTCOS will be a union of the Solaris MT-Safe libraries and the Chimera libraries.

Graphical User interfaces: Solaris provides process priority inheritance (PPI) to prevent priority inversion from occurring when a high-priority process and a low-priority process are competing for the same resource. The PPI system works by monitoring the system for cases when a low-priority process is blocking a high priority process by holding a critical resource and automatically boosting the priority of the low priority process to that of the blocked process, hopefully freeing the critical resource sooner than it would normally be freed. This allows, for example, an X-Windows based RTCOS real-time process to perform updates ahead of other non-real-time processes. The use of the PPI system in the RTCOS allows for practical real-time displays and other graphical interfaces. The RTCOS provides access to the standard X-Windows interfaces, allowing applications developers full access to the Solaris platform, without giving up real-time advantages. These real-time windowing facilities are used by Onika to provide a user-friendly interface to developing and executing real-time applications.

Programming environment: One of the goals of the RTCOS is to minimize replication of the programming environment available for software development on the NRTOS. In that respect, the RTCOS uses all the same tools, including compilers, debuggers, window systems, and graphical tools as the NRTOS. The RTCOS is designed with a message passing interface to the user interface, so that command-line, program, and graphical user interfaces can be interchanged. The support for reconfigurable software and loadable modules also allows the use of software assembly-based interfaces for rapid visual programming of real-time applications [2].

Compilers/Linkers/Debuggers: Since Solaris uses a standard dynamically linked executable format generated by all Solaris 2.x compilers, the application developer can use their compiler of choice, whether it is the standard Sun C/C++ compiler, the GNU C/C++ compiler, or another third party compiler, to create software components. External commercial or freeware libraries can be used to develop applications and simply included in the link stage. The use of dynamic linking results in smaller executable files and lower memory usage for running an executable.

2.3 Section Summary

In this section, we described the many issues that were addressed in the design of the RTCOS, and gave an overview of the solutions employed in the design and implementation of the RTCOS. In the next section, the technical details of our solutions are presented.

3. Technical Discussion

In this section we discuss the technical details of our design, and the advantages and repercussions of our approach, for various aspects of the RTCOS.

3.1 Memory Management

Caches: Making effective use of the 1 MByte level 2 cache associated with each processor requires preloading each real-time task into the cache and keeping the process size to under 1 Mbyte. Each text and data page in a task running under the RTCOS is accessed before execution, ensuring that the page is in the cache. Care is taken to prevent pages from being removed from the cache. Proper use of the cache is highly dependent upon the underlying architecture and the RTCOS uses a method which provides high performance using the Sparcstation 20 architecture.

Virtual memory: The RTCOS does not use virtual memory. The *plock(2)* system call is used to disable virtual memory and page swapping for the RTCOS and all associated real-time tasks. Effective cache use limits the RTCOS to using 1 megabyte of memory, so disabling virtual memory for the RTCOS does not unduly affect the memory requirements of the rest of the system. All RTCOS memory pages are locked into physical memory and cannot be swapped to disk during the lifetime of the RTCOS kernel. All other tasks in the system can continue to use virtual memory as usual.

This increases the predictability of the system as well as the overall speed.

Memory Map: All RTCOS processes have both a private memory area used to for internal data storage and a shared memory area, common to all RTCOS processes. The shared memory area is at the same base address in all of the RTCOS processes and is reserved at RTCOS start-up by the “mmap” system call. This allows pointers to data in the shared memory areas to be passed between RTCOS processes without translation, which improves the speed and flexibility of the system.

3.2 IPC Mechanisms

The RTCOS has implemented the Chimera IPC suite, including remote semaphores, dynamically allocatable shared memory, prioritized message queues, and the global state variable table [13]. This decision is based on the fact that, although Solaris has real-time IPCs defined in the manual, in practice, the functions currently return *ENOSYS* (i.e. "not implemented in this system" error). Therefore the Solaris IPCs still cannot be used. We therefore ported the Chimera IPCs to the RTCOS. To make the RTCOS IPCs more compatible with the POSIX.4 specification for real-time operations, a translation layer has been added to provide compatibility with the IPCs defined in POSIX.4. The Chimera IPCs are a superset of those specified by the POSIX.4 API, which simplified the process of implementing the translation layer. The RTCOS IPC mechanisms are implemented on top of express mail.

The IPC implementation uses a policy/mechanism separation that is based on the Chimera reconfigurable device drivers [11]. This separation allows for each of the various IPC mechanisms to be replaced while the system is running. For example, the replacement of the semaphores require shifting the system to a *safe* state where there is no contention between RTCOS tasks for the resources (no blocked *P* operations on semaphores). As long as the system is in this safe state, the IPC mechanism can be replaced or upgraded. This implementation serves as the basis for evolutionary design, where a system that cannot be taken off-line can be upgraded and dynamically modified while on-line.

The generic IPC framework also allows the use of other IPC mechanisms not currently defined or implemented. This mechanism can be used to add user-defined functionality to the RTCOS kernel. Note that user-level dynamically reconfigurable modules which are not accessed through the RTCOS kernel do not use this mechanism. Instead, they use the support for reconfigurable software, based on the Chimera methodology.

3.3 Resource Sharing

The Solaris kernel is reentrant and thus can have multiple processes (and processors) executing kernel code simultaneously. This allows multiple applications to be inside the filesystem and networking code (and other areas of the kernel) while avoiding race conditions. However, this can affect the predictability of real-time tasks.

To avoid potential problems with using the Solaris method of permitting direct filesystem access, the RTCOS uses the same global/local operating system separation used by Chimera. Remote procedure calls based on express mail are used to provide networking, file system, and user I/O support for real-time applications. Express mail is a unique message passing mechanism for distributed shared memory systems, which allow the non-blocking communication between real-time and non-real-time operating systems. The mechanism passes all service requests for network or filesystem access to the NRTOS. The RTCOS concentrates exclusively on providing real-time performance and predictability, while the NRTOS handles those aspects which can compromise the predictability of real-time tasks, such as security related issues and hardware interrupts.

For example, if the RTCOS allowed direct filesystem access, a low priority real-time task could begin a disk access and then be preempted by a higher priority task. Then, when the disk is ready and issues an interrupt, the higher priority real-time process will be preempted to handle the interrupt, even though the interrupt was caused by a lower priority task. The remote procedure calls to the NRTOS prevent this type of priority inversion.

3.4 Interrupt handling

The RTCOS controls the interrupt generation and handling on each of its processors, so that only interrupts which directly affect the real-time application are handled by RTPUs.

In the Solaris system, interrupts are handled by very high-priority non-preemptible threads which are members of a special scheduling class. These *interrupt* threads are higher priority than Solaris real-time tasks, and if a processor has interrupt processing enabled, these interrupt threads can run on a processor and preempt even a real-time task. This was the primary reason for the unpredictable execution times for the task that was shown in Figures 3 through 6, and Figure 8. The RTCOS uses the Solaris kernel functions *cpu_enable_intr* and *cpu_disable_intr* to control interrupt processing on the RTCOS processors. Access to these functions is provided to the RTCOS through a loadable kernel module. These functions prevent Solaris dispatcher from stealing CPU cycles from an RTCOS real-time thread. The Solaris dispatcher services are not needed by the RTCOS because the RTCOS is the only Solaris process on each RTPU. The processing of interrupts from other sources, such as the hard disk or mouse, are also not allowed. Controlled use of these and other interrupt-related functions allows the RTCOS to provide low latency, highly predictable services.

3.5 Scheduling and Timing

The original design of the RTCOS used the Solaris dispatcher in combination with a MUF scheduler loadable kernel module to control all task switching between RTCOS threads. The loadable scheduler implemented a new scheduling class which was given a higher priority than Solaris real-time threads. In other words, threads of the RTCOS class could preempt even threads in the Solaris real-time class. This was done to ensure that when an RTCOS process was "bound" to a processor, it would never be preempted by another process in any scheduling class. The binding of the RTCOS to a particular processor was done using the Solaris *processor_bind(2)* system call.

This design suffered from several limitations, namely accuracy and predictability. Solaris 2.4 provides only a 10ms clock for scheduling, which limits the accuracy and increases the latency for tasks running under the RTCOS. In other words, a task asking to run at T = 100 msec will actually run at $T = 100$ msec ± 10 msec. Tasks running at frequencies greater than 100Hz will be hurt by tasks which miss their deadlines because the scheduler will not find out about the missed deadline for (worst case) 10 msec. This low timing resolution and accuracy is not acceptable for most control applications. Even if the RTCOS defined a custom Solaris scheduling class, task execution is not predictable; it would result in real-time execution that resembles the output shown in Figure 4.

Our current design of the RTCOS scheduler uses user-level context switching and an alpha-version of the *timer14* device driver [3] developed and supplied to us by Sun Microsystems to provide high-resolution, low latency task switching in the RTCOS. The RTCOS is still bound to a specific processor, as it was in the previous design. It executes as the highest priority real-time task in the system and has been given an infinite time quantum. This is done using the *priocntl(2)* system call. To Solaris, the RTCOS appears to be a single process with a single thread of execution, with one RTCOS process running on each RTPU, but never on processor 0 which also executes the NRTOS. Task switching is done using the Solaris *getcontext(2)*, *setcontext(2)*, and *makecontext(2)* system calls.

User-level context switching allows for very low overhead context switches and fast response times when a formerly-blocked high-priority task is moved onto the ready-queue. Very high accuracy (timings accurate to within 10 microseconds) data collection for automatic task profiling is also made possible with this scheduling mechanism. Automatic task profiling is described more below.

The source code to this device driver, which is to be distributed with Solaris 2.6, has been modified and provides the RTCOS with a loadable kernel module to generate timer interrupts with microsecond resolution. The RTCOS scheduler has been designed to program the *timer14* driver only when necessary, and thus the RTCOS is interrupted only when process rescheduling is needed. By not having a periodic clock tick, we avoid the periodic noise shown that was shown in Figure 5.

This design allows the RTCOS to provide support for tasks running at 1000Hz, which was one of the design criterion for the system. All hardware dependent portions of the system are encapsulated in one small set of files and can easily be replaced when moving the RTCOS to other platforms. The use of this design improves the speed and portability of the system.

The RTCOS scheduler is implemented using the policy/mechanism separation which is the basis for Chimera reconfigurable device drivers. The implementation of the scheduler uses Solaris dynamic loading to allow for safe, dynamic replacement of the RTCOS scheduler while the system is running.

3.6 Automatic Task Profiling

The RTCOS has been designed to provide continuous monitoring of real-time tasks, based on the automatic task profiling (ATP) designed for the Chimera RTOS. The RTCOS version of ATP provides 5 μsec measurement accuracy, as opposed to the 1 millisecond accuracy in the Chimera implementation.

The RTCOS ATP is implemented as part of the scheduler, since the scheduler always knows which task is executing at any given time and is responsible for switching task contexts. Data points are also taken upon entry to and exit from interrupt handlers, thus providing for accurate readings even in the presence of interrupts.

ATP is an extremely useful RTOS feature because one of the assumptions made by most real-time scheduling algorithms is that the execution time of a task is known. In the general case, manual methods of profiling a task are required in order to determine how long executing a cycle of the task takes, which can be a long and tedious task. If changes are made to the code, then this profiling must be performed again. In the past, task sampling has been used to obtain task profiles. However, the results are not necessarily accurate due to the coarse grain of the sampling and using a finer grain results in too much system overhead for the profiling. In our RTCOS, the profiling provides statistical feedback to the user, the executing task, and to the on-line schedulers, so that, if necessary, they may adapt to account for the actual execution times of the system, and not the estimated worst-case execution times.

Implementation of the ATP in the RTCOS is based on the Solaris *gettimeofday(3C)* system function, which performs a system trap and returns (on a Sparcstation 20) the current time with better than 10 microsecond resolution.

4. Summary

This paper describes our design of a Real-Time Co-Operating System which transforms a general purpose processor into a real-time co-processor. It is used to simultaneously support real-time and non-real-time activities on a workstation with two or more processors. The RTCOS is the software equivalent of a co-processor, with a software architecture analogous to the hardware architecture that has been used in many workstations and personal computers. In this paper, we discuss our first software prototype of the RTCOS, which co-exists with Solaris 2.4 on a four-processor Sun SPARCstation 20. We summarize the feasibility of our approach through an experimental characterization of Solaris 2.4. We address the technical issues involved and present the details of our design.

The RTCOS is a step in the direction of ubiquitous computing, which is the technological strive for obtaining a single computer system that can perform all the needs of the users, ranging from non-real-time information system applications to real-time signal processing for multimedia applications such as video conferencing, interactive television, and virtual laboratories.

5. Acknowledgments

The research described in this paper is supported by Sun Microsystems, Inc.; the Institute for Advanced Computer Studies (UMIACS), the Institute for Systems Research (ISR), the Graduate Research Board, and the Electrical Engineering Department at University of Maryland.

6. References

[1] Compaq Computer Corp., *http://www.compaq.com/*.

[2] M. W. Gertz, D. B. Stewart, and P. K. Khosla, "A Human-Machine Interface for Distributed Virtual Laboratories," *IEEE Robotics and Automation Magazine*, Vol. 1, No. 4, Dec. 1994.

[3] M. Hamilton, "Extending the realtime capabilities of solaris." Sun Microsystems internal working document, April 1995.

[4] C. L. Liu, and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard real time environment," *J. of the Association for Computing Machinery*, v.20, n.1, pp. 44-61, Jan. 1973.

[5] Lynx Real-Time Systems, Inc., *http://www.lynx.com/*.

[6] Microtech Research: *http://www.mri.com/*.

[7] Microware Systems Corp. *http://www.microware.com/*.

[8] QNX Software Systems Ltd., *http://www.qnx.com/*.

[9] J. A. Stankovic and K. Ramamritham, "The design of the Spring kernel," in *Proc. of Real-Time Systems Symposium*, pp. 146-157, December 1987; *http://www-ccs.cs.umass.edu/*.

[10] D. B. Stewart and P. K. Khosla, "Real-time scheduling of sensor-based control systems," in *Real-Time Programming*, ed. by W. Halang and K. Ramamritham, (Tarrytown, New York: Pergamon Press Inc.), 1992.

[11] D. B. Stewart, *Real-Time Software Design and Analysis of Reconfigurable Multi-Sensor Based System*s, Ph.D. Dissertation, ECE Dept., Carnegie Mellon University, Pittsburgh, PA 15213, April 1994.

[12] D. B. Stewart, P. K. Khosla, "The Chimera Methodology: Design of Dynamically reconfigurable real-time software using port-based objects", in To appear in *Int'l Journal of Software Engineering and Knowledge Engineering*, May 1996.

[13] D. B. Stewart, D. E. Schmitz, and P. K. Khosla, "The Chimera II real-time operating system for advanced sensor-based control applications," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1282-1295, November/December 1992.

[14] H. Tokuda and C. Mercer, "ARTS: A distributed real-time kernel," *ACM Operating Systems Review*, vol. 23, No. 4, July 1989.

[15] Sun Microsystems, Inc., *http://www.sun.com/*.

[16] H. Tokuda, T. Nakajima, and P. Rao, "Real-time Mach: Towards a predictable real-time system," in *Proc. of the USENIX Mach Workshop*, Oct. 1990.

[17] Wind River Systems, Inc., *http://www.wrs.com/*.