An Architecture-based Adaptation Framework for Soft Real-Time Applications

BAO xiao-an, ZHANG na, GE qi-wei, GUI ning

Advances in Information Sciences and Service Sciences, Volume 3, Number 7, August 2011

Abstract

Driven by the proliferation of mobile and pervasive computing, real-time systems are increasingly used in dynamic changing environments with variable user needs, hosting real-time applications ranging in number and nature over time. However, existing adaptation frameworks normally are incapable to cope with other competing and changing concerns. This paper proposes an architecture-based framework for managing real-time components’ dependence and lifecycle in an effective and uniform way. A real-time component runtime service is proposed here to maintain the global view, control the whole lifecycle of the components and keep existing real-time components’ promised contracts in the face of run-time changes. This framework is designed to be easily extended with other constraint resolving policies as well as dependence descriptions languages. At end of this paper, the framework is tested from both adaptation and performance aspects by a simulated control application.

Keywords: Real-time Systems, Adaptation Frameworks, Real-time, Components

1. Introduction

Traditional real-time systems such as process control systems typically work in closed and highly predictable environments. Those systems were meant for predefined and fixed services that were not supposed to change. As long as those assumptions and conditions hold, this approach allows hard real-time requirements to be met with very high assurance.

However, some soft real-time systems, such as smart devices, are increasingly operating in changing environments and with diverse user needs. In coping with dynamic and diverse requirements, such devices normally are equipped with open software system that able to host applications ranging in number and nature over time. In such system, the traditional waterflow approach for real-time application development: design, verify, map, and deploy could not work effectively as the system configuration will evolve during the whole system lifecycle.

This open and evolving execution environment requires new approaches in building and manages real-time applications. Thorough off-line system analysis is forbidden by such dynamicity. However, due to scarcity of system resources, these installed applications normally are, to a certain extent, inter-dependent with one another, as one application may competing with other applications for the CPU, memory, networks or other system resources. Failed to deal with such dependence may lead to breaking applications’ real-time contracts which the system designer has supported to satisfy.

Traditional RTOS and real-time schedulers based approach could not cope with that complex dependence, as the scheduler itself has no way to get and could not understand that high level application dependence which can only be interpreted according to current environment. The RTOS is in a sense too “far away” from the application-level concerns to be able to take the right decisions on which components to select and which resources to dispatch.

In order to support the dynamic changing set of real-time applications, from our experience in designing adaptive real-time systems, any effective solution should have the following three important aspects. First, those resource requirements across the currently hosted applications must be obtained and managed during system run-time. Second, those component dependences, including the explicit functional dependencies and the implicit dependencies due to competing for the limited available resources, need to be maintained and managed during run-time. Third, the system’s ability to enhance the system’s adaptability and reactivity of real-time applications require to be fast and efficient.
resources, need to be explicitly modeled into architectural model across application domains. Third, the architectural model should be used to drive the lower-level managements of the corresponding real-time tasks.

Our framework tackles these problems from different perspectives. In this paper, we introduce an architecture-based management framework - the Declarative Real-time Component (DRCom) Framework, spanning through installed application domains. By eliciting each real-time component’ distinguishes attributes from its meta-data and its component management interface, an accurate global architecture model for current real-time system can be built. In order to guarantee certain real-time application performance, this model then is mapped to adjust low-level real-time task’s attributes and states according to different adaptation policies. Benefited from the Service oriented architecture, this framework can be easily extended with other dependence resolving policies to deal with diverse system requirement. The effectiveness of our architecture is demonstrated both from qualitative and a quantitative point of view.

The rest of this paper is structured as follows: in Section 2, a scenario shows the importance for inter real-time application dependence managements. Following section 3 gives a general overview of our architecture. Section 4 describes the component’s declarative real-time meta-data format design. Then in section 5, the component’s dependence resolving process is discussed. Then, our framework is evaluated from two aspects in Section 6. Section 7 provides a comparison with related works. The future work and our conclusions are introduced at the end of this paper.

2. Scenario description

To better illustrate all the complexities in introducing the context knowledge into the application composition process, we introduce an example scenario that will be revisited several times through the course of this paper.

Now, the Set-top devices are equipped with open system can be harnessed to provide services in addition to its core application for temporal or long-term base. As a Set-top device, its basic function should be TV application. In brief, this application will received Stream Video data from remote server, decode this data and output to the TV port. As an open system, Set-top can also install certain applications to enhance its usability. For example, it can install a new application which transcode a recorded High Definition TV program to IPod/IPhone format for late display on those mobile devices. Figure 1 shows the simplified component graph for those two applications.

As a typical multi-task system, if a user starts those two applications, a Set-top may try to execute the two applications simultaneously whenever the Set-top device has enough resources. If not, it may eventually lead to possibly transient timing problems of TV decoding task including missing frame, data overflows etc. These kinds of time breaches can results in the poor video quality and bad user experience. The typical schedulers in the RTOS itself has no way to determine which actions to be taken as it only support very simple attribute for scheduling, e.g. Priority or Earliest Deadline First(EDF).

If these two applications are constructed by reflective real-time components, an external framework can elicit different real-time tasks characteristics from different application domains and automatically determine which actions should be taken according to current context. In our scenario, one possible strategy can choose to suspend or stop a computational intensive component and reserve enough system resources for TV application because a user normally prefers to give highest priority to those application that matter his/her experience most. When the user turns off the TV, the architecture should able to sense the changes and resume/start the transcoding component.
3. Architectural Framework

The above case study shows that, without the architectural-based adaptation supports, it is very hard to effectively deal with conflicts across application domains. In order to deal with these challenges, a framework based on declarative real-time component is proposed to build the adaptive real-time application. Firstly, the system design goal is analyzed.

3.1. System design Goal

To deal with highly dynamic environments while providing explicit real-time support, we describe an integrated architecture that provides support for the following.

- Real-time components' dynamicity support
  The dynamicity of real-time component is supported by taking control of all requests for creation, configuration, reconfiguration and un-initialization – that is, the whole of the components’ lifecycle – from underlying system. This allows having a complete picture of all the lifecycle requests taking place in the system as a whole, as well as a snapshot of current system configuration.

- Real-time components’ dependence support
  Another problem we have to tackle is the intrinsic dependencies that each component has on other components – both in hardware and software. We distinguish two types of dependences, functional (structural) dependencies regarding the business logic and non-functional dependences regarding the QoS requirements. Our framework resolves and manages both above two types of dependencies.

- Customizable global adaptation support
  Given a description of the current system configuration and of the functional and non-functional dependencies between components, rather than using a general adaptation scheme, here we make use these descriptions to support the definition of custom adaptation strategies. Depending upon the dynamic availability of resources and adaptation policy (for instance, performance first, safety first or mission duration first …) different algorithms for constraint resolution may be desired. The following section describes the system architecture design of our solution.

3.2. System Requirements Analysis

In our model, the distinguishing real-time aspect of DRCom is declared in a meta-data file. Intensive studies have been carried out on how non-functional properties of a component implementation are specified, ranging from simple name-value pair based SPDF (Simple prerequisite Description Format) [14] to modern QoS specification languages such as CQML+ [6] etc. According to specific domains, different ways of specifying real-time components’ prerequisites and dependence may be used.
The same is true for the adaptation logic. Due to the real-time applications’ domain-specific and application specific features, rather than try to propose a silver bullet general adaptation logic, which is rather unlikely, this architecture need to be flexible and easy to be tailored to different contexts.

Our framework is designed in a ways that different kinds of description languages and adaptation logics could be used. Three key modules were designed to fulfill different aspects requirements – the Description, Reasoning and Action modules. Dependencies and configurations are represented in the description modules and reasoned upon in the reasoning module while the corresponding actions are taken in the Action module. Contrary to [16] and most of the state-of-art studies, in which the execution runtime statically integrates these three modules, our framework implemented them as service providers by employing the Service-oriented model. They are loosely coupled by their service interfaces. Future more sophisticated QoS descriptions as well as those adaptation algorithms to be easily plugged into our framework as long as they implemented the same service interface.

3.3. Architectural Framework

Figure 2 presents a schematic view of the major elements of our architecture. This framework is composed of three main parts corresponding to the modules described in the previous section plus other ancillary services, such as resource management for monitoring system resource changes.

1. A mechanism (Custom language parsing Service, working as Description module) for resource dependence representation lets developers specify component dependencies in their own custom language
2. A constraint resolving service (working as reasoning module) for QoS constraints resolving policy lets users or administrators specify the domain-specific system adaptation policy.
3. The Declarative Real-time Component Run-time (DRCR) execution environment (action module) works at the core of the framework. It manages the DRCom instances and performs the actions reasoned by resolving service. It also takes the responsibilities to monitoring the change events sent by Resource Management Service and OSGi system. A DRCom instance registry is designed to keep the installed real-time component instances’ information and it is maintained by DRCR.

![Figure 2. Architectural Framework](image)

Based on the decisions made by the constraint resolving service, DRCR takes full control of the component lifecycle. In this framework, a complete component’s lifecycle model and a correspondent real-time component management interface is designed. The adaptation process will be triggered by component state changes and the notification from external event sources such as Resource Manager or a simple periodic time. In the following sections, we provide a more in-depth description of key elements and processes of this architecture.
4. Component description

Real Time Component description is defined in meta-data files attached with component implementation. This document specifies the distinguished characteristics of the specific real-time component. Here we use the hybrid real-time component model introduced in our previous work [7], while an important feature is added to support more complex description languages.

4.1. Meta-data description design

In our implementation of DRCom, a real-time component configuration represents a component instance’s functional contracts as well as non-functional requirements. As discussed in section 3, the framework should support different kinds of component prerequisite languages. However, from an implementation point of view, a purely separated approach means little or no pre-knowledge of the component’s meta-data format. This will introduce considerable implementation complexity and bring unnecessary performance overhead. Compared to the complex non-functional requirements, component’s functional part is comparably well structured and stable. In order to strike a balance between performance and complexity, we use a two phase mode in describing and processing the component’s functional and non-functional dependences.

1. The functional part is defined by XML schema and parsed by DRCR.
2. For the component’s resource prerequisite description (non-functional part), the job of parsing this information is delegated to an external custom parsing service. The custom resolving service will process the parsed data and send results to the DRCR.

Figure 3 shows a fragment of meta-data file which contains the functional and non-functional contract of a smart camera component that can return regions of interests (subsets from a frame image data) on demand.

4.2 Functional Requirement Description

The component element has the following attributes:
• name: The name of a component must be globally unique because it is used as a component reference.
• enabled: Controls whether the component is enabled when the bundle is started. The default value is true. If enabled is set to false, the component is disabled until system enables it.

The implementation element is required and defines the name of the component implementation class. The component instances will be created by DRCR by referring to this attribute. The import and outport elements define the communication methods from which the inter-components data are shared. The component may have 0 or more inports or outports.

• outport: outport specifies what the required output for the real-time component is. It consists of a name attribute, which is used as the communication reference. In addition to name, the input port should have:
  • interface - the communication interface. It can be shared memory, real-time FIFO pipe line or Mailbox.

```xml
<? xml version="1.0" encoding="UTF-8"?>
<drt:component name="camera" desc="this is a smart camera controller" enabled="true">
  <implementation bincode="ua.pats.demo. smartcamera.RTComponent"/>
  <outport name="images" interface="RTAI.SHM" type="Byte" size="400"/>
  <inport name="windposition" interface="RTAI.SHM" type="Integer" size ="1" optional="false"/>
  <property name="windowsize" type="Integer" value="40" />
```
An Architecture-based Adaptation Framework for Soft Real-Time Applications

BAO Xiao-an, ZHANG Na, GE Qi-wei, GUI Ning

Advances in Information Sciences and Service Sciences, Volume 3, Number 7, August 2011

Figure 3. Component sample configuration

- **type** - the transported data type, which can be integer or byte.
- **size** - the cardinality of this type.

Compared to **outport**, **inport** element has one additional attribute: The **optional** attribute shows whether this port dependence is mandatory needed for the component’s execution.

A component description can also have a number of properties. The property element can be used to configure the real-time component instance via the management interfaces. It can be used to query the attribute values via component management interface described in our previous work [5].

4.3. Non-functional Prerequisite Description

The non-functional prerequisite description is described in the **service-specific** element. This element has only one attribute – **language**, which designates which language is employed in describing the resource prerequisites. The DRCR itself will not try to parse the value of this element. It will just parse the **language** attribute and store the value of **service-specific** element as a normal string in the component instance registry and delegate it to the custom parser service to parse when necessary. To demonstrate the validation of our framework, we designed the Simple Task Description Format (STDF) for describing DRCom real-time task’s characteristics.

Figure 3 shows an STDF meta-data for real-time task characteristics in the **service-specific** element. The Machine-type specifies that this component’s real-time execution parts is a periodic task, running on CPU 1, priority 1 with period 10000ns (at rate 100Hz) and execution time for each round of execution is 1000ns and deadline for overrun is 2000ns.

5. Constraint Resolving Process

One of the key processes in the declarative real-time component framework is the constraint resolving process. It takes responsibility to identify the relationship between the real-time components as well as resources and to reason about possible reactions according to the current system context. Complying with the two phrase

When system configuration changes, e.g. because a new real-time component arrives, or the resource management service prompts that there is a noticeable resource change, DRCR will be activated. Firstly it will check existing DRCom’s constraints and dependences. After the functional part was successfully checked and satisfied, the DRCR will determine whether the component’s non-functional requirements could be guaranteed without jeopardizing the validity of other components’ real-time contracts. If this is indeed the case, this new coming component is now ready to be activated.

Complying with two phase DRCom configuration, two phase constraint resolving was used. One is the Functional dependence resolving phase, which checks the functional dependence for any real-time component when system change occurs. In current prototype, the DRCR integrate the Function dependence resolving modules to achieve better performance and reduce implementation complexity. The other phase is the customized constraints resolving process. As the framework uses the OSGi’s service-oriented model, system administrators can change the system resolving behavior by registering their own customized resolving service to fit specific system configurations.

5. 1. Functional Dependence Resolving

The function dependence resolving consists of the following two main steps:

A. **Circular Check**: Real-time components in applications can be modeled as a component graph,
which normally form a directed acyclic graph. In a task route, a task at a component has a certain set of outputs, which is a set of inputs for the task at next component along the task route. It’s possible for a set of component descriptions to create a circular dependency. In this case at least one component configuration should not be satisfied. Circular reference is detected by DRCR when it attempts to satisfy component configuration. When this happens, the DRCR runtime logs an error message and does not activate any component configuration which may create circular dependencies.

B. Port Compatibility Check:
In order to satisfy the functional dependency, all enabled and activated component’s input ports should have corresponding output ports. This means their port names, interface types, data types and data sizes should match exactly. In Figure 3, the smart camera port needs an input port called “windowsposition”. In order to satisfy the port compatibility check, one activated component instance with name = "ua.pats.trajectory.windows" and outport defined as follows

```xml
<outport name="windposition" interface="RTAI.SHM" type="Integer" size="1"/>
```

must be already deployed and enabled in the registry.

5.2. Custom Resolving process

In practical environments, according to dynamic availability of resources and system’s specific requirements, different algorithms for the resource allocation resolution may be desired. Thus the Custom Resolving service is designed aligning with service oriented mode to give customizable support. By using different customized resolving service providers, system behavior can be easily altered without the need to change the application main functional logic. This eased the need to recompile and re-load the whole application. One basic resolving policy can be – all the basic resource prerequisites must be satisfied before the component configuration is activated. However, with context-specific or domain-dependence knowledge, more flexible and complex resolving processes can be employed. One example can be: to stop low priority tasks or those tasks with optional port requirements to release their resource for incoming high priority real-time components. Certain rule engines, for instance Jess [17], instead of hard coded logics can also be integrated into the resolving service.

In our system, several custom resolving service providers may co-exist in one system. DRCR will choose one candidate service provider which matches current system context. In current implementation, the system context is a global property managed by administrator. Our ongoing work is to employ the context reasoning scheme to get more accurate and flexible context information so as to make use of a more appropriate custom resolving service.

5.3. Adaptation Scenario.

Figure 4 shows a sequence diagram for adaptation. In this scenario, component B needs component A’s output to satisfy its functional constraints. As an instance (ComAInstance) of component A configuration already exists, the DRCR will successfully solve B’s functional constraints. Then, the internal resolving service and the external customized service will be consulted for real-time resolving process. When both services return positive results, the DRCR will create and activate the component B’s instance (ComBInstance) according to its configuration. If ComAInstance is stopped, the DRCR gets notified about this event and consults its internal and external custom service again to check for newly (un)satisfied component instances. When it gets the results, the DRCR will find the ComBInstance is now unsatisfied and should be disabled.

6. Implementation & simulation results

In this section, we will use a scenario-based simulation to evaluate our adaptation approach across different application domains. Our framework has been validated both from a qualitative and a quantitative point of view. Firstly, we will discuss the system implementation.
6.1. System implementation

We use the OSGi framework as our implementation platform. OSGi technology serves as the platform for universal middleware ranging from embedded devices to server environments. The open standard simplifies the introduction of new capabilities and services. Equinox, a popular, free, open source OSGi Platform developed by the Eclipse organization, is used as our basic development platform. In the component implementation, we use our DRCom model proposed in our previous work [5,7]. DRCom was originally designed for the construction of dynamically configurable & reflective real-time systems.

The DRCom model has a basic management interface enabling a uniform and coherent way to control a component’s lifecycle and attributes. Each component is associated with a meta-data file which enlists its abstractions – for instance, interface for communication, attributes as well as reference constraints. This approach provides a meta-data level reflection and introspection which is very effective in component run-time composition. Each component is tagged with task related information for configuration and validation.

6.2. Simulation configuration

As shown from Figure 1, all six modules’ execution parts are implemented as periodic tasks. Decoding component of TV application is implemented with a real-time task with period 33.37 ms with priority 2, execution time about 8 ms and deadline is 12 ms. The execution part of Transcoder module is implemented as with a real-time task with period 50ms with priority 1, execution time about 10 ms and deadline is 30ms. The schedule policy used by underline RTAI system is FIFO. In order to show the interference between these two components, others are implemented with little CPU time. The transmission methods are asynchronized (shared memory) so we can focus on the two coding module’s performance.

As the video decoding execution time may vary according the contents of video streams, in order to more clearly demonstrate the mutual-influence among applications, we substituted the decoding function into an calculate function which use comparably constant CPU time for each round of calculation. So the execution influence of inter-influence can be more clearly indentified from the results. In Figure 1, the color of component shows state of each component after adaptation process.
6.3. Custom resolving policy

As described in section 2, when TV application and transcoding application are executed simultaneously, system may not have enough resources to guarantee their performance. Here, the custom resolving services will be employed to deal with the competing requirements. The strategy employed in our simulation is to protect the TV application as it normally matter user’s experience most.

The custom resolving service will track the decoding component’s overrun attribute via getproperty() defined in the component management interface. If the overrun increase too fast (by recording overrun values and time of retrieving the values), it will try to disable the component which consume most CPU time (executiontime/period), the algorithm 1 describe the adaptation process.

Algorithm 1: Guarantee TV application

Requires: Installed components’ CPU usage information

1. Ensure: Allocate enough CPU time to the TV application
2. If Decoder. getProperty(overrun) increase > Max_threshold
3. for all cmp in SatisfiedComponents do
4. if component not from TV application
5. get CPUusage from cmp attributes
6. record the cmp with highest CPU usage.
7. end if
8. end for
9. put cmp to disabled component list(dcl)
10. return dcl

After this custom resolving process, the Transcoder component will be disabled. Then, due to functional dependence constraint violation, the File Writer in the Transcoding application will also be disabled in functional resolving process. The different color in Figure 1 shows the states of installed components after these adaptation processes.

6.4. Simulation results

We perform 6000 observations for the execution time of Decoding Component in TV application. Table 1 shows the execution time in case of with and without adaptation framework support. In order to eliminate the transitory effects of cold start and adaptation, we collect measurements after the system has started and renders a steady execution. The execution time is express in micro-second(µs).

<table>
<thead>
<tr>
<th>Context/Execution Results</th>
<th>With Adaptation</th>
<th>No Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Exe. Time (µs)</td>
<td>8127.297</td>
<td>10647.804</td>
</tr>
<tr>
<td>AVEDEV (µs)</td>
<td>4.15</td>
<td>3410.9</td>
</tr>
<tr>
<td>MIN (µs)</td>
<td>8121.8</td>
<td>8125.3</td>
</tr>
<tr>
<td>MAX (µs)</td>
<td>8175.1</td>
<td>18296.7</td>
</tr>
<tr>
<td>Overrun (times)</td>
<td>0</td>
<td>1880</td>
</tr>
</tbody>
</table>

In the formal case, the execution time of decoder is largely around 8ms, the biggest execution time is about 8.18 ms when the custom resolving service exist (it will disable transcoding module). In contrast, if no such service available, the system will try to run these two applications simultaneously. The jitter of decoder task’s execution time is very big. As we can see from Table 1, the average deviation is 3410.9µs which is significantly worse than 4.15µs when our framework resolving mechanism exists. This may eventually lead to possibly transient timing problems of TV decoding task including missing frame, data overflows etc.
7. Related work

Our platform follows the approach of an architecture-based management framework for real-time system. That is: the real-time guarantee is not directly implemented within components' application code but provided by the container runtime environment according to additional component descriptors.

The OMG’s CORBA Component Model (CCM) [11] forms the basis for many functional concepts of our component model, but it does not address special problems related to non-functional requirements—for instance, component dynamicity and system downtime-free evolution.

CIAO [10] builds a QoS-enabled CORBA Component Model (CCM) implementation on top of TAO [12]. The project adhered to existing OMG specifications such as RT/CORBA and CCM. In contrast, our focus is on the challenges of simultaneously supporting real-time control and non real-time component adaptation management while keeping the implementation of our component model as lean as possible. The considerable overhead of implementing or extending a fully compliant CCM infrastructure would have been counterproductive to our system goals.

In the real-time component designs, Stewart et al. [13] designed a reconfigurable port-based object framework whose real-time communication scheme is similar to ours. Their pure real-time design philosophy makes this system hard to develop.

In the real-time Java specification, RTSJ [2] introduces the concepts of timeliness, schedule ability, and real-time synchronization to Java-based applications. It aims to enhance the Java platform with capabilities required by real-time applications. However, dynamicity of modules is not addressed, which would prevent an implementation of our concepts on top of this platform.

In order to deal with component dynamicity, Cervantes and Hall [15] propose a service-oriented component based framework for constructing adaptive component-based applications. The key part of the framework is the Service Binder which automatically controls the relationship between components. Our approach mimics theirs in dealing with component’s dynamicity. However, the static dependence description and resolving policy limits its application in normal OSGi applications and could not be applied to the real-time domain.

Aleš, Loiret et al. propose a component framework [16] for Java-based Real-time systems based on RTSJ. It provides continuum between the design and implementation process. By automatically generating an execution infrastructure, it mitigates complexities of RTSJ development and transparently manages the real-time concerns. Their approach tries to solve the non-intuitive rules and restrictions introduced by RTSJ. However, their work lacks an automatic component dynamicity control scheme and their component dependence support is rather static and confined to the RTSJ domain specific knowledge.

Hartig and Zschaler designed and implemented enforceable component-based real-time contracts [17]—the extension of component-based software engineering technology that comprehensively supports adaptive real-time systems from specification all the way to the running system. Their work is the most similar to our framework. Their system runs large and complex parts in a classic non-realtime environment and only small, predictable parts in a real-time environment. However, although they propose the concept of adaptation manager for parameter adjustment and profile change, there is no formal design for how to deal with the dynamicity of component’s availability which is crucial for downtime-free systems.

8. Conclusion and future work

This paper describes the experience in building a framework that supports run-time adaptation in response to the dynamic evolution and continuous deployments of modern complex real-time systems. In this framework, the real-time contract is specified in the component’s meta-data. The component instance is managed by the system for the dependence resolving and real-time contract. Global view of current system configuration is managed by the DRCR service. Such system reasons about the changes in the system configuration and performs certain actions while still guaranteeing the designated real-time components’ real-time contracts. It sheds the burden for each real-time component to monitor the system status and maintains the reference to the other dynamically available real-time components.
Although our experience was done based on the OSGi middleware, we believe our findings to be general to be used in other architecture-based management systems.

Although we showed an integrated approach to providing real-time properties to component-based applications, areas remain open for further research. One ongoing work is to design languages to describe those component-specific adaptation interfaces. By using meta-data to describe those component-specific adaptation interfaces as well as its capabilities, we can achieve more flexible and more powerful adaptation capability.

9. Acknowledgement

This work is partially supported by the Zhejiang Provincial Natural Science foundation of China under Grant No, Y1100289, Y1100726, and Science Technology Department of Zhejiang Province Grant No. 2011C33G2060108.

10. References