IT Licentiate theses
2000-004
MRTC Report 00/21

A Formal Approach to Analysis of Software Architectures for Real-Time Systems

ANDERS WALL

UPPSALA UNIVERSITY
Department of Information Technology

MRTC
MÄLARDALEN REAL-TIME RESEARCH CENTRE

UPPSALA UNIVERSITY

A Formal Approach to Analysis of Software Architectures for Real-Time Systems

BY
ANDERS WALL

September 2000

DEPARTMENT OF COMPUTER SYSTEMS
INFORMATION TECHNOLOGY
UPPSALA UNIVERSITY
UPPSALA
SWEDEN

Dissertation for the degree of Licentiate of Philosophy in Computer Systems
at Uppsala University 2000

A Formal Approach to Analysis of Software Architectures for Real-Time Systems

Anders Wall
anders.wall@mdh.se

Department of Computer Systems
Information Technology
Uppsala University
Box 337
SE-751 05 Uppsala
Sweden

http://www.it.uu.se/
http://www.idt.mdh.se/

ISSN 0346-8887, ISSN 1404-3041
Printed by University Printers, Uppsala University, Sweden

Licentiate theses from the Department of Information Technology

2000-001 Katarina Boman: *Low-Angle Estimation: Models, Methods and Bounds*

2000-002 Susanne Remle: *Modeling and Parameter Estimation of the Diffusion Equation*

2000-003 Fredrik Larsson: *Efficient Implmentation of Model-Checkers for Networks of Timed Automata*

2000-004 Anders Wall: *A Formal Approach to Analysis of Software Architectures for Real-Time Systems*

UPPSALA
UNIVERSITY

A Dissertation submitted for the Degree of Licentiate of Philosophy in Computer Systems at Uppsala University, September 2000

ABSTRACT:

Wall A. 2000: A Formal Approach to Analysis of Software Architectures for Real-Time Systems. IT Licentiate thesis 2000-004 , ISSN 0346-8887. MRTC Technical report 00/21, ISSN 1404-3041.

A software architecture is a high-level design description of a software system. In terms of the architecture, early design decisions can be analyzed to improve the quality of a real time software system, which depends very much on how it is structured rather than how it is implemented. Architectural analysis techniques vary in their degree of formality. The least formal is based on reviews and scenarios, whereas the most formal analysis methods are based on mathematics. In this thesis, we propose to use a formal approach to software architectural analysis. We use networks of timed automata to model the architecture of real time systems and transform architectural analysis problems to reachability problems that can be checked by the existing tools for timed automata. The typical properties that can be handled using this approach are schedulability and safety properties.

As the first technical contribution, we extend the classic model of timed automata with a notion of real time tasks. This yields a general model for real-time systems in which tasks may be periodic and non-periodic. We show that the schedulability problem for the extended model can be transformed to a reachability problem for standard timed automata, and thus it can be checked by existing model-checking tools, e.g. UPPAAL for timed automata. As the second contribution, we present a method to check general high level temporal requirements e.g. timing constraints on data flowing in multi-rate transactions, which can not be handled by traditional approach to schedulability analysis. We have developed an algorithm that given a data dependency model and a schedule for a transaction constructs a timed automaton describing the behavior of the transaction. Thus, by using existing verification tools we can verify that a given architecture is schedulable and more importantly, it is correctly constructed with respect to the high level temporal constraints.

Anders Wall, Department of Computer Engineering, Mälardalen University, Box 883, S-721 23 Västerås Sweden, and Department of Computer Systems, Information Technology, Uppsala University, Box 325, S-751 05 Uppsala, Sweden

ISSN 0346-8887, ISSN 1404-3041

Printed by University Printers, Uppsala, Sweden

Distributed by *Department of Computer Engineering, Mälardalen University, Box 883, S-721 23 Västerås Sweden, and Department of Computer Systems, Information Technology, Uppsala University, Box 325, S-751 05 Uppsala, Sweden*

Acknowledgments

I would like to thank my supervisors Dr. Wang Yi at the Department of Computer Systems at Uppsala University and Dr. Christer Norström at the Department of Computer Engineering at Mälardalen University for their guidance and constructive feedback on my work. Your support is highly appreciated.

I also want to thank my colleagues here at the Department of Computer Engineering at Mälardalen University, especially the people at the systems design lab. So far, it has been a great pleasure to work together with you.

Finally, I would like to thank Kristian Sandström and Henrik Thane for taking on the burden of reviewing this thesis.

This work has been supported by ARTES.

Västerås, May 2000

Anders Wall

1 Thesis contents

A. Wall A., Software Architectures for Real-Time Systems, Technical Report MRTC 00/20 ISSN 1404-3041, Mälardalen Real-Time Research Centre, Mälardalen University, May 2000

B. Norström C., Wall A., and Yi W., Time Automata as Task Models for Event-Driven Systems, In proceedings of the 6th International Conference on Real-Time Computing Systems and Applications

C. Wall A., Sandström K., Mäki-Turja J., Norström C., and Yi W., Verifying Temporal Constraints on Data in Multi-Rate Transactions using Timed Automata, Submitted to Real-Time Systems Symposium 2000.

2 Thesis introduction

A real-time system is a software system where the correctness depends not only on correct functionality, but also the time when the computation is performed, i.e. temporal correctness. The functionality in a real-time system is usually divided into several concurrent processes called *tasks*. Each task in a real-time system has its own temporal requirement regarding, for instance, when to start its execution and when the execution should be finished. The collected set of temporal attributes that can be specified for a task constitutes a task model. The task model provides the information necessary for scheduling the task set.

Depending on the criticality of the temporal correctness, real-time systems are divided into *hard*- and *soft* real-time systems. If the system is considered hard, all temporal constraints must be satisfied. Typically, such systems reside in nuclear power plants, aircraft, vehicles, etc. On the contrary, tasks in soft real-time systems may occasionally violate their temporal constraints. Many soft real-time systems exist in our surrounding, e.g. home appliances, toys, and telecommunication.

Traditionally, when designing real-time systems, models are constructed out of the temporal constraints on tasks for the purpose of verifying the temporal correctness. The temporal correctness of real-time systems is of great importance but there exist other important properties, e.g. *safety*, *dependability*, *reusability*, *maintainability*, usually referred to as *x-abilities*. In order to verify, or predict, how well a designed system complies with such quality requirements, models of more than the temporal constraints are needed.

One approach to attacking this problem is to employ the software architecture and software analysis techniques available. Software architectures are high-level descriptions of software systems. On its highest level of abstraction, software architectures describe the components in a system and their interconnection, i.e. the structure of the system. Many of the properties mentioned above can be estimated by analyzing the architecture of the system solely. There exist several techniques for architectural analysis, all with varying level of formality reaching from experience-based reasoning to mathematical methods. Basically, the modeling language determines the level of formality of the analysis.

Languages for modeling of the architecture of a system are called an *architectural description language* (ADL). On its simplest form, an ADL may only describe the components in a system and the connection between them. Formal models are usually considered to model the behavior of a software system on a lower level than its architectural structure. However, by applying formal modeling and verification techniques in the architectural design, some of the quality properties can be verified using mathematics.

As time is of great significance for real-time systems, it is important that the language used for modeling such systems has a notion of time. In this thesis, the formal modeling language *timed automata* is used. By making models of real-time systems in timed automata, the temporal correctness, as well as safety properties, can be formally verified.

We describe how formal modeling languages, such as timed automata, can be used in software architectural design and analysis for real-time systems. More specifically, we

describe how to verify the temporal requirements that are derived from the high-level requirements by *model-checking* architectural models constructed in timed automata. For this purpose we have extended the classical definition of timed automata with a notion of tasks. Moreover, high-level temporal requirements for a system do not necessarily have their correspondence in the temporal attributes available in the task model. For instance, there could be a requirement on the end-to-end deadline for a task transaction, while the task model only allow specification of period times and deadlines for the individual tasks. In order to verify the temporal correctness, schedulability analysis is performed. However, the schedulability analysis is based on the mapping of the high-level temporal requirements to the temporal attributes provided by the used task model. Consequently, a system can be schedulable, i.e. all deadlines for each task is met, even though the temporal behavior is incorrect, as the high-level requirements might not be fulfilled. In the method proposed in this thesis, we aim at verifying that, not only is a schedulable architecture constructed, but also that the correct architecture is constructed with respect to the high-level requirements. The high-level temporal requirements we consider are on data flowing through task transitions. Such requirements typically specify minimum or maximum age on sampled data used in computations.

3 Results

Paper A: Software Architectures for Real-Time Systems

In this report, the state of the art in the field of software architectures, with a focus on software architectures for real-time systems, is described. Software architectures is a part of software engineering concerned with high-level design and analysis. The objective of architectural analysis is to verify quality requirements on software. Typically examples of such quality requirements are maintainability, reliability and reusability. For real-time systems, the temporal correctness is a crucial quality property. The report discusses architectural description languages, architectural view and architectural analysis. Moreover, quality properties valid for real-time systems and methods for analyzing such properties are described.

Paper B: Timed Automata as Task Models for Event-Driven Systems

In this paper, the classical model of timed automata is extended with a notion of real-time tasks. The main idea is to associate each discrete transition in a timed automaton with a task. Intuitively, a discrete transition in the extended timed automaton denotes an event releasing a task and the guards on the transition specify the possible arrival times of the event. This yields a general model for real-time systems in which tasks may be periodic and non-periodic. Moreover, the paper shows that the schedulability problem for the extended model can be transformed to a reachability problem for standard timed automata and thus it is decidable. This allows us to apply model-checking tools for timed automata to verify the schedulability of a event-driven system. In addition, based on the same models of a system, properties such as safety and, to some extent functionality, can be verified.

Paper C: Verifying Temporal Constraints on Data in Multi-Rate Transactions using Timed Automata

This paper describes how to verify temporal constraint on data flowing through a set of collaborating tasks that runs at different frequencies. Such a set of tasks is called a multi-rate transaction. Typically temporal constraints on data flowing through such a transaction is maximum time from input to output or a maximum time difference between inputs. Such constraints are of great importance to guarantee the correct functioning of the designed system. But normally they cannot be checked using the traditional approach to schedulability analysis. The paper describes how to use timed automata and reachability analysis to verify such temporal constraints on data in transactions. By making a timed automaton model of the data dependencies in a transaction, we enable automatic verification of timing constraints such as end-to-end latency. The model can handle different computational models and any non-preemptive execution order of the tasks in the transaction.

4 Future work

As future work we will implement tools that, based on existing model-checkers, support the methods proposed in this thesis. Furthermore, as timed automata is used, we will consider to implement a tool that automatically generates code from such models. Thus, the gap between formal models and their implementations, which is a source of possible divergences, is reduced.

Today in industry, products are developed based on existing platforms and applications. The platforms must be constructed in such a way that they easily can adopt new functionality, i.e. they should be flexible. We will investigate how the temporal dimension, present in the domain of real-time systems, affects the analysis of quality properties such as flexibility, and how flexible platforms should be constructed.

5 Papers produced

A. Wall , Software Architectures for Real-Time Systems, Technical Report MRTC 00/20 ISSN 1404-3041, Mälardalen Real-Time Research Centre, Mälardalen University, May 2000

C. Norström, A. Wall , and W. Yi, Time Automata as Task Models for Event-Driven Systems, In proceedings of the 6th International Conference on Real-Time Computing Systems and Applications

A. Wall, K. Sandström, J. Mäki-Turja, C. Norström, and W. Yi, Verifying Temporal Constraints on Data in Multi-Rate Transactions using Timed Automata, Submitted to Real-Time Systems Symposium 2000.

H. Thane, and A.Wall, Formal and Probabilistic Arguments for Reuse and Reverification of Components in Safety-Critical Real-Time Systems, Technical Report, Mälardalen Real-Time Research Centre, Mälardalen University, March 2000

Software Architectures for Real-Time Systems

By

Anders Wall

Technical Report MRTC 00/20 ISSN 1404-3041, Mälardalen Real-Time Research Centre, Mälardalen University May 2000

Software Architectures for Real-time Systems

Anders Wall
Department of Computer Engineering
Mälardalen Real-Time research Center
Mälardalen University
Sweden
anders.wall@mdh.se

ABSTRACT

The solution to the complex nature of developing software is software engineering. Software engineering provides techniques for structured design, formal- and informal analysis, and software metrics. The part of software engineering concerned with high-level design and analysis is called software architectures. The objective of architectural analysis is to verify quality requirements on software. It can be applied on any level in the design but it focuses on the structure of the software. While the architecture provides a high-level abstraction of the software, divergences between the designed system and the requirements can be detected early in the design phase. However, the structure of the software alone does not always provide enough information in order to analyze all requirements put upon a software system. Additional information about the software construction is provided by different architectural views. The number of views, and their contents varies depending on the system domain and the required quality properties to analyze.

In this report, the state of the art in the field of software architectures is described. The survey is focused on software architectures for real-time systems but many of the described techniques can be applied to general software systems.

Contents

1 Introduction

The number of projects in industry developing software is constantly increasing. Software is not only replacing old and well-established technologies, but also increasing in size and complexity. To manage the complexity, engineering methods for constructing software needed, i.e. *software engineering*. Software engineering has been established as a broad discipline that covers topics ranging from requirements capture, design, implementation, and software metrics, to maintenance, verification and validation. An established engineering practice is taken for granted in many engineering disciplines but not in the software community. In order to be considered an engineering practice, we must be able to construct models that can be analyzed and verified. Moreover, design methods are needed including established techniques that have been proven successful as well as tools supporting the methods. The part of software engineering that focuses on high-level design and analysis is called *software architectures*.

Edsger Dijkstra pointed out in a paper from 1968 the importance of partitioning and structuring software, in contrast to just focusing on programming to produce the correct functionality [dijk68]. This is what software architecture, and software architectural analysis is about as it deals with how to structure a software system and how to evaluate that structure with respect to different quality properties. The interest in the software architecture field has increased lately due to the increased functionality provided by software systems, the increased size and complexity, and the increased cost of developing and maintaining software products. Today, industry is aware of the benefits of being able to analyze and verify software constructions in an early phase of the development process. If a software development project diverges from the functional requirements or the quality requirements, and if those divergences are not detected early, the cost of revising the design in the end of the project will be significant due to redesign. Almost 80 percent of the cost for developing a software product are spent after the initial design and implementation phases [Clem96b]. These 80 percent spent on maintenance, which includes error detection, correction and evolutionary development.

Not only does a structured description of a software system constitute a basis for architectural analysis, it can also improve the productivity of new members in a project. The architecture provides a simple and holistic view of the whole system. This is very important since complex system usually engage a lot of people, all with unique competencies, at different stages of the development process. Since designing real-time systems usually require multi-disciplinary knowledge, it is very important to have an architectural description that can be understood by software engineers as well as control and mechanical engineers. Furthermore, many software projects employ a lot of consultants. Consultants may have little knowledge of a company's product line and need a quick briefing in order to get productive and cost efficient.

The complexity of software systems also causes problems when maintaining and correcting errors in a software product. It is seldom possible to, in advance, be aware of all the side effects that particular a correction may give rise to. If an architectural description is at hand, it could give some guidance on what modules are most likely to be affected by the correction. This is highly related to evolutionary development. If the architecture of the software construction is violated, it ceases to exist in its former

shape. The construction still has an architecture, but as long as the architecture is not explicitly, and correctly described, it is of no use. Consequently, the architectural description may, and should, evolve as the construction that it describes evolves.

1.1 Towards a definition

There are almost as many definitions of software architecture in the literature as there are software architects and designers. We mention a few examples:

The software architecture of a program or computing system is the structure or structures of the system, which compromise software components, the external visible properties of those components, and the relationships among them [BCK98].

In [Paul94] the following definition is given:

Software architecture not only reflects how the functional requirements are met, but addresses:

1. *non-functional requirements*
2. *design rationale*
3. *architecture style*

Yet another definition is provided in [Clem96a]:

A view of a system that includes the system's major components, the behavior of those components as visible to the rest of the system, and the ways in which the components interact and coordinate to achieve the system's mission.

One property that seems to be common among almost every proposed definition is that the software architecture describes a system by a composition of its software components and their interrelationships. In addition, software architectures should provide a high level description, i.e. a more abstract level than the level that algorithms and data structures provide. However, defining a software architecture only as a syntactical representations of components and their interconnections in the software systems is not sufficient. To be useful, additional information must be present in the description, in particular the semantics of components and connections. Different domains of software systems have different semantics of their software architectural description. A domain defines the class of application to wich a product belongs, e.g. desktop applications and industrial control applications. As a consequence, there will be variations in the definitions of software architectures depending on the domain. Furthermore, the definition also depends on the aim of the architectural description, e.g. support for architectural analysis, representation or description of the designed system. It is probably impossible to unify software designers in one single definition as it depends on the aim of the architecture and the domain in which it is used. What we can state is that software architecture is a description of the software structure and methods to evaluate and compare design solutions.

1.2 Open research areas

As software architecture is an immature research area, a lot of open questions still exist. Most of the ongoing research in the field of software architectures is focused on description languages and analysis of architectures for non-real-time systems. The

analysis methods are still informal in their nature. As the analysis methods are informal they provide rough metrics and estimations. We believe that formality can be added to architectural models. Thus, the models can provide means for formal verification of some of the quality properties that are listed in this survey.

Most of the material on software architectural analysis found in the literature ignores the temporal aspects. By adding the temporal dimension on software, completely new problem arises. As an example, components developed for real-time systems, i.e. system for which correctness depend on both the functionality and the temporal correctness, can not be reused in new environments unless at least the temporal constraints are still fulfilled. Quality properties such as flexibility, i.e. the ability of a software system to adopt new, or remove old functionality, are also important. As real-time systems are restricted to resources such as processors, communication busses, etc., a lot of additionally parameters must be taken into account in such an analysis.

The tool support for architectural design and analysis is poor. Tools that support the complete process of developing an architecture are needed. Today, architectural tools for real-time systems almost exclusively focus on schedulability analysis. As indicated in this report, there are a lot of other important properties of real-time systems software. However, implementing such tools is non-trivial. One approach is to use existing tools for automatic verification. This can be done if the problem of analyzing a specific quality property can be transformed into a property that can be verified using that tool. Examples of existing tools for formal verification are UPPAAL and KRONOS [LPY97][DaYo95].

1.3 Outline

Chapter 2 discusses architectural description languages and desired properties of such. In Chapter 3, the architectural view necessary for an architectural analysis of real-time software architectures is discussed. Architectural analysis is dealt with in Chapter 4. Finally, Chapter 5 concludes the report. Terminology used in the paper is explained as it is used. Appendix A provides, however, a complete list of the vocabulary together with a short explanation.

2 Architecture description languages

Communication among software engineers is crucial. Without means for communication, important information into- and from the design phase might accidentally get lost, resulting in misinterpretations. Moreover, a system designer must be able to communicate with customers, other project members and management in an unambiguous way. An unambiguous architectural description is also a necessary condition for performing architectural analysis. A parable is the building trade, where building architects transform the customer requirements into a design. This design must be described in a way the building constructor understands in order to do mechanical strength calculus and for building workers to use as a blueprint. When developing software, a software engineer formalizes the customer requirements. Based on the requirements, a high-level design is described in a language that is commonly understood by customers and designers. The common language is a necessity in order to communicate and discuss design solutions. As output from the high-level design phase, one or several candidate architectural solutions are produced.

To verify that the quality requirements of the system are met by the architectural solutions, the architecture has to be analyzed. Hence, the description language used in the high-level design must support the analysis methods. Once a software architecture is constructed that fulfils the requirements, the architectural description is used as a ”blueprint” when implementing the system. In addition, an architectural description makes maintenance easier since it facilitates the understanding how parts of software systems cooperate. Thus, the parts of a software system, i.e. components and sub-systems, affected by a correction are detected in advance.

2.1 Desired properties of an architecture description language

Languages for architectural description are called Architecture Description Languages (ADL). There is an abundance of ADL:s, each of them with its own specific syntax, semantics, expressiveness and purposes[EHLS94][LKAV93][Vest94]. An ideal ADL should however, provide six classes of properties: *composition*, *abstraction*, *reusability*, *configuration*, *heterogeneity* and *analysis* [SHGA96]. By *composition* is meant that a software system should be described as a composition of components and connections. Furthermore, components and connections must also be described in a way that clearly and explicitly describes the exact role of each element, i.e. modeled on an appropriate level of *abstraction.*

As components are *reused* in different applications that are described using different description languages, the architectural description must be able to adopt to reuse. That is**,** it should be possible to reuse components, connectors describing the interconnection between components and architectural patterns in different architectural descriptions. Related to reusability is *heterogeneity*. Heterogeneity is the possibility of combining different heterogeneous architectural descriptions.

Configuration means that the architectural structure among components in the system should be separated from the structure in the components. The language should also support dynamic reconfiguration. As will be discussed in Chapter 3, the structural

view describes all components and connections, whereas the module view unveils the structure of each component.

Finally, as high-level design analysis is one of the primer justifications for using software architectural techniques, the architectural description must support different kinds of *analyses*.

Considering the desired properties of an architectural description above, how can a software architecture be described? One possibility is a plain textural description in a natural language. However, natural languages tend to be ambiguous, making them really hard to interpret in a consistent manner. By using a formal language an unambiguous description is obtained. With formal languages it is possible to use mathematics when modeling and verifying the architecture. The disadvantage of using formal languages as architectural descriptions are that most of them requires a lot of experience and mathematical skill. Consequently, such a description may be sufficient and useful at some stage in the design process but not for communication with partners in a project without a computer science background. By relaxing the formality, a semi-formal, graphical representation may be obtained. Even inexperienced people can get a feeling for how a system is constructed by interpreting a graphical representation. The semi-formal description also permits analyses and quality predictions to be made as described later in this report. The graphical approach has been adopted by many of the available ADL, where the software design is

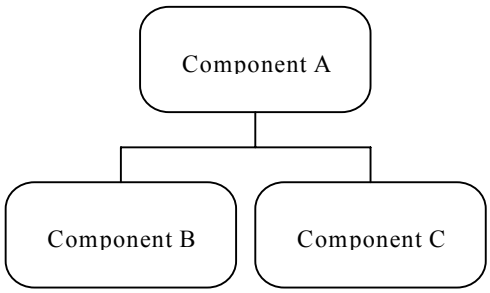


Figure 1. A graphical software architecture description.

constructed using components and their interconnections in a 4th generation language manner as illustrated in Figure 1.

2.2 Semantics of an ADL

The architectural description in Figure 1 provides only the information that there are tree components in the system, which are connected to each other. The connections could indicate a class hierarchy or a network communication link over a distributed hardware architecture. As stressed by Clements and Northrop [Clem96b], it must be known exactly what the components are, what the connections mean and what the position of the components imply, i.e. a well-defined semantics. If the semantics is not clear the architectural description is quite useless.

One single architectural description language can not fit the desired level of abstraction for every different software domain and application. There is for example

a big difference between designing a real-time system with hard- and soft temporal requirements compared to designing an administrative application with database management and transactions. Consequently, we need a unique description language for every application domain.

Even though there must be differences in the architectural description depending on the application domain, there might exist a least common denominator. Such a least common denominator could, for instance, consist of components and connections. But the significance of a connection or a component could be domain specific.

If the ADL has an unambiguous semantics, design tools for architectural analyses can be developed [ERGUSA97] [LPY97]. However, analysis of quality properties usually requires more information than just the architectural structure. This additional information is provided by the architectural views and is discussed in Chapter 3.

2.3 Examples of existing architectural description languages

There exist several architectural description languages for real-time systems. Typically they differ in their expressiveness and formality. As an example of a formal modeling language that can be used for describing architectures for real-time systems we use *timed automata* [ALDI92]. Architectures are described in timed automata as a network of finite state machines, where a process or a component is one state machine. Synchronization channels connect processes in timed automata to each other. A synchronization channel defines the name of the signal used for synchronization. Thus, architectural interconnections are described using synchronization. Below is a more rigorous description of timed automata.

A timed automaton is a finite state machine extended with real-valued clocks that increases uniformly. Moreover, transitions in a timed automaton are decorated with guards and actions. Guards are clock constraints that enables or disables a transition, i.e. if the guard is true then the transition can be taken. In Figure 2, the transition from *S1* to *S2* can be taken if the clock *x* has a value greater than 10 time units.

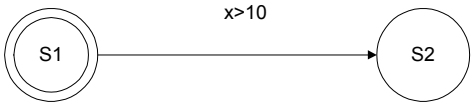


Figure 2. A simple timed automaton

Actions enable synchronization between different timed automata in a rendezvous manner, i.e. processes halts until both participating processes can synchronize. This indicates that a complete system is modeled by a set of timed automata, such a set is called a network and consists of the parallel composition of the included processes. Consider Figure 3 where a small network is displayed consisting of two processes, *A* and *B*.

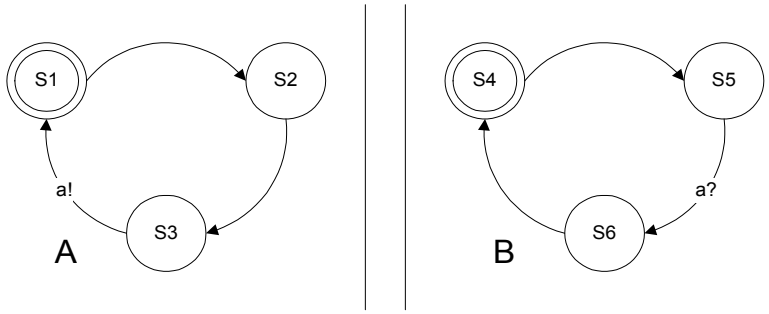


Figure 3. A network of timed automata processes

Whenever process *B* is in state *S5*, it will wait for another process to send a signal on channel *a*. The question mark after the channel name indicates that *B* is waiting for the signal. As long as no signal is being sent on channel *a*, *B* is stuck in state *S5*. As soon as process *A* reaches state *S3*, the processes can synchronize and the processes progress to *S1* and *S4* respectively. During this transition, no clocks are progressing, i.e. it is a discrete transition. Models of timed automata can be constructed and automatically verified using existing model-checking tools, e.g. UPPAAL and KRONOS [LPY97][DaYo95].

Another example of an ADL for real-time systems is *MetaH* [Vest94]. This is a language that models a system on level of abstraction higher than timed automata. MetaH provides means for specifying real-time processes, referred to as tasks, that can be either periodic or aperiodic, communication among tasks, modes and composites of processes and modes that are called macros. Furthermore, the hardware allocation of processes and characteristics of the hardware such as channels that are used for communication among processors can be specified. As the temporal properties of tasks and modes are provided in the models, MetaH support different kinds of real-time analyses such as schedulability analysis. There exist a graphical tool that supports the modeling in MetaH and analysis of real-time software architectures described in MetaH. The schedulability analysis in this tool is based on rate-monotonic [LILA73].

3 Architectural views

Architectural views constitute an important part of a software architectural description as they expose architectural information apart from only the structure. In Figure 4, architectural description languages for different software families (domains), are viewed as an inheritance graph. The top node includes description primitives shared by all domains (compare with a virtual base class in the object orientation community). Two common description primitives could, for example, be syntactical symbols representing components and the connections between components. This means that components and links can describe the structure of any sub-domain of software applications. However, the component primitives and the connection primitives have no semantics in the top node. Semantics and new syntactical symbols will be added while moving down in the inheritance hierarchy. For instance, the semantics of a component in a real-time system is probably a task, and the links are the communication among tasks or precedence relations. In an administrative software application on the other hand, components are most certainly databases or user interfaces, and connections denote database transactions.

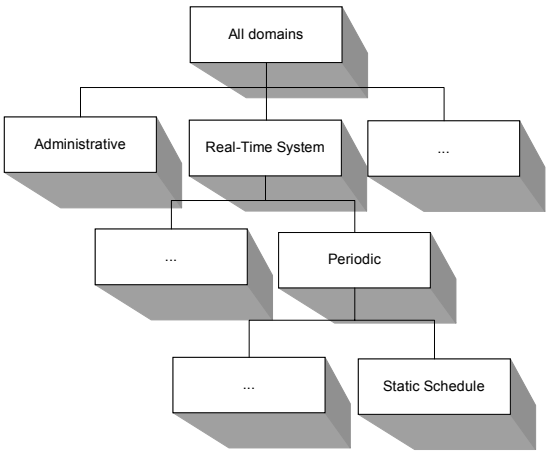


Figure 4. Architectural description and view hierarchy.

The nodes in Figure 4 are intentionally displayed in a 3-dimensional manner. Each side of a cubical node is a metaphor for a unique view of the architecture. There might be arbitrary many different views, depending on the needs for verification and analysis in a software development project.

In this Chapter, views important for the real-time systems domain are discussed. Note that not all views must be modeled in a project developing real-time systems. Only the views sufficient for the analyses required must be present in the architectural description. The names of the views and their contents are not standardized but we propose the following:

- Structural view
- Module view
- Logical view
- Hardware view

- Temporal view
- Communication view
- Synchronization view

Structural view

The structural view describes the overall architectural design and style, providing the highest level of abstraction. This is the natural starting point for an architect designing a software system. The structural view consists of software modules and their interconnections, i.e. the interfaces between them. The syntactical representation of modules and connections is optional but should be uniform within the development project for the sake of communication among engineers.

As design on this level is rather rapid, it is possible to design several competing architectures for evaluation and comparison. Once a software architecture satisfying the quality requirements is selected, it is settled. Depending on the required analyses, more views might have to be modeled in order to make a correct decision. For instance one or more of the views proposed in this chapter could be considered.

In Figure 5, the structure of a system consisting of four components is displayed. The arrows between the components represent function calls through the component interfaces.

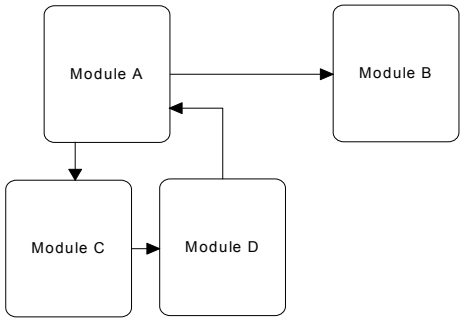


Figure 5. The structural view of the software architecture.

In the design methodology called *Module Approach to Software Construction, Operation and Test* (MASCOT), the structural view is modeled with a diagram called the *decomposed component level view* [Masc87]. This view provides a decomposition of a sub-system into its main constituents, i.e. its tasks.

The object-oriented methodology for real-time systems called *Hard Real-Time Hierarchical Object-Oriented Design* (HRT-HOOD), also has a structural view that is provided by the so-called parent-objects [BuWe94]. A parent-object is a component on its highest-level that may be further decomposed.

The corresponding abstraction for networks of timed automata is the processes. A system modeled in a network consists of a set of automata (processes). Each of these processes could be seen as a component. The interconnections are modeled using synchronization actions. Interconnections visualize the data flow. Information of the control flow is given by the logical view, which is discussed in Chapter 3.3.

Module view

The module view exposes all the functions, methods or sub-modules in all the components modeled in the structural view. A software component is a software module, which is further, decomposed into functions and sub-modules in order to unveil the division of functionality. This view should also describe the interactions between the functions. It is, for example, desirable that the interaction between functions in different components is held to a minimum. Some communication between components is necessary, but the communication must be performed trough well-defined interfaces that conceal the underlying functionality.

Hierarchical methods such as MASCOT and HRT-HOOD both provide means for component decomposition. In MASCOT the module view becomes the structural view as each component is refined, while in HRT-HOOD, the module view is described by child-objects derived from each parent-object.

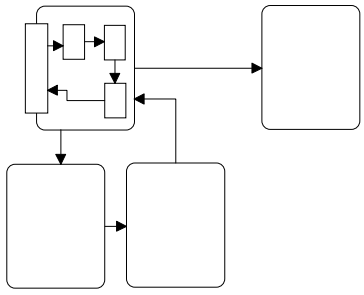


Figure 6. The Module view of components.

Logical view

In this view the functions from the module view is described in more logical details. It serves as a model of the actual implementation, which can be used as a low-level description or constitute the basis for formal verification. Some possible descriptions are state machines or algebra like CCS [Miln87]. These are all different ways of describing the functionality of software formally. State machines can be of different types depending on the application. For example, timed automata can be used for real-time systems as it provides a notion of time as well as concurrency [ALDI92]. If time is of no concern, an ordinary state machine can be used. CCS is a process algebra with which it is possible to model concurrent systems. Such algebra is useful when modeling communication and synchronization, which is essential when designing real-time systems.

In Figure 7, The logical view for the sub-components is modeled using time automata. The upper sub-module synchronizes with the lower sub-module by sending signal *a*.

From the software architecture perspective, the logical view may be on a far too detailed level since software architectures are descriptions of software systems on a higher level than algorithms. However, this view will eventually be implemented, if not in logic so in the chosen programming language which in itself is a formal description of the specification.

The logical view is of no interest when settling the architectural style. It provides a basis for formal verification and in the end the program source code.

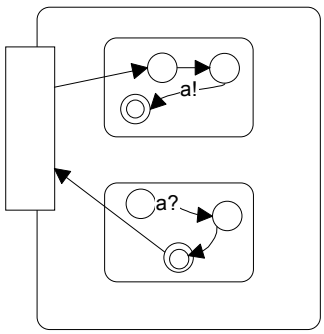


Figure 7. The logical view.

Hardware view

If the system is distributed, i.e. a set of interconnected and geographically separated CPUs, or a multi-processor system, i.e. a set of interconnected and geographically collected CPUs, there might be requirements of pre-allocated functionality among the nodes in the system. Such an allocation will affect the final architecture and the performance of the application. Yet another reason for having a hardware view description in the software architecture is the issue of portability. If software should be easy to move between different types of platforms, the dependencies to the hardware and the operating systems must be encapsulated from the rest of the software system. One can discuss whether this is a software architectural view or not, but as long as hardware has an impact on the software architecture, we consider it a view.

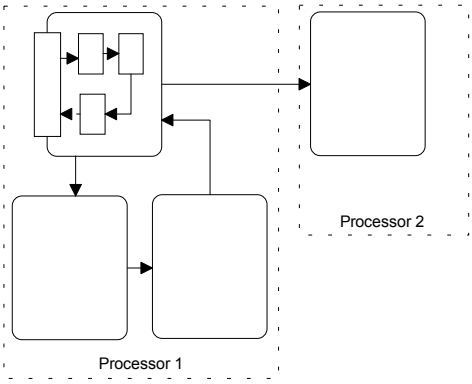


Figure 8. The processor allocation in the hardware view.

In the Yourdon Structured Method (YSM), the allocation of functions to hardware processors is called the *processor environment model* [Cool91]. Besides the function allocation, this view reveals the data that will be communicated among the processors.

Temporal view

The views discussed so far are common among different software families and consequently reside in the topmost node in the architectural hierarchy shown in Figure 5. The temporal view is, however, domain specific. As the correctness of a real-time

system not only depends on correct function, but also correct timing, the temporal constraints must be present in the architecture. By correct timing we mean not too early and not too late. In order to verify whether or not tasks in a real-time application will be schedulable, i.e. all temporal constraints are fulfilled such as all deadlines are met, we need a view of the temporal requirements.

The temporal view contains data such as release time i.e. the earliest start time of a task, the deadline i.e. the latest completion time of a task, the period time (the frequency) of a task, etc. We say that a *task model* determines the exact content of the temporal view. The exact appearance of a task model varies depending on the *execution strategy*. The execution strategy defines the rules that determine what task to execute.

As an example of a variation in the temporal view, consider a periodic task that samples a sensor in a process. As the sampling should be performed with some specific frequency in order to obtain a correct view of the process, a period specifying the interval between two consecutive executions of the sampling must be specified. In contrast, if the application is purely event trigged, i.e. tasks have arbitrary release times, there is no need for specifying period times. Instead, the minimum inter-arrival times must be specified for the tasks.

HRT-HOOD has a temporal view that is divided into two parts, one that describes the execution strategies for a class and one that provides the temporal attributes. The execution strategy can be either *cyclic* or *sporadic*. Depending on the execution strategy, classes can be assigned, e.g. period times, minimal inter-arrival times, and deadlines.

In timed automata, clocks and guards on clocks describe the temporal view.

Communication view

For telecommunication systems, and for real-time systems in general, it is desirable to model communication among tasks and processes. Communication is typically performed using messages and signals that are sent back and forth in the system, either locally on one processor or among nodes in a distributed system. For this purpose the communication view can be used. In Figure 9, the communication is visualized with Message Sequence Charts (MSC). The vertical line in each process depicts time which increase downwards. The horizontal lines between the processes depict the messages or signals.

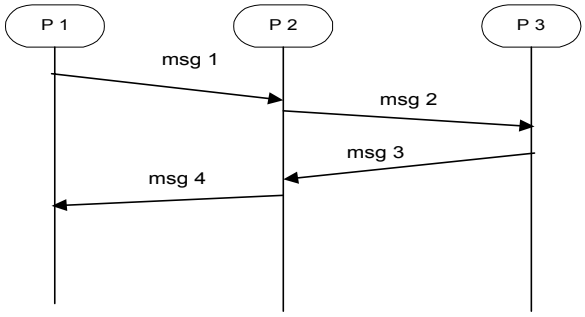


Figure 9. Message Sequence Chart

The MSC can be translated into ordinary finite state automata which makes it possible to formally verify them using, for instance, temporal logic [LaLe94].

Synchronization view

As real-time systems often are multi-tasking systems having several tasks running concurrently, it is necessary to synchronize access to shared resources in order to avoid inconstancy. Tasks that uses a shared resource must mutually exclude each other, i.e. only one task can use the resource at the time. There exist several techniques for handling mutual exclusion in real-time systems, e.g. semaphores, signals or separation of task in time. In addition, to guarantee precedence relations, i.e. requirements of the execution order among tasks in a system, synchronization is necessary.

What synchronization technique to choose depends on the provided infrastructure, i.e. the real-time operating system (RTOS), and the available task models. For instance, if the system is pre-run-time scheduled, i.e. a pre-runtime generated table defines the execution order of the tasks, time-vise separation of tasks can be used. On the other hand, if the system is event-triggered, and semaphores are the only means for synchronization provided by the infrastructure, the semaphore approach must be used.

The information unveiled in the synchronization view is implicitly present in other views discussed in this section. For instance, if synchronization is resolved by separation in time, this is visible in the temporal view, or if signals are used, this is visible in the communication view.

In MASCOT, communication and synchronization is modeled using *paths* along which entities communicate. A path can indicate a dependency to commonly used data, or a dependency to another entity that results in a sending/receiving of messages.

Communication and synchronization between processes can be modeled in timed automata by using synchronization actions.

3.1 Discussion

All the different views should not be designed in the beginning of a development project. Instead an iterative process is often preferred. For some applications, some of the views can be excluded. For instance, if there is no distribution and no requirements regarding portability, the hardware view may be excluded.

There exist relations among different architectural views. The relation between the structural view and the module view is obvious as the module view provides a decomposition of the architecture specified in the structural view. The logical view defines the ”low-level-design”, specified in some formal language suitable for formal verification of, for instance, the communication and synchronization among the modules in the software system. The schedulability of a distributed real-time system depends on how tasks are allocated, i.e. how the tasks are distributed. The allocation affects the utilization of each processor and the time spent on communication between tasks allocated on different processors.

4Architectural analysis

The main incitement for using software architecture notation when designing a software system is the ability to analyze and verify the design in an early stage of the development process. By comparing different candidate architectures, confidence in early design decisions is achieved. Such a comparison is done by listing pros and cons for each architectural solution according to the quality requirements put on the system. Furthermore, architectural analysis enables the possibility to get software metrics based on the high-level design, e.g. the level of coupling and cohesion within and between the different modules that constitute the software system [Fenton96].

In this report, the software system quality properties are divided into two different classes, functional and nonfunctional. Functional quality properties are those concerned with the runtime behavior of the software, e.g. performance or reliability, whereas nonfunctional quality properties are concerned with the quality of the software itself, e.g. maintainability or reusability. Most of these software quality properties are qualitative rather than quantitative, thus being practicable only for comparison between different architectures.

4.1 Methods for architectural analysis

An architectural analysis process is divided into two stages, *questioning* and *measuring*. The questioning phase generates questions that are answered by the measuring phase. Len Bass et. al. [BCK98], have categorized the questioning stage in architectural review and evaluation into three different classes namely *Scenario-based*, *checklist-based* and *questionnaire-based*.

Scenarios are a set of cases where the software architect asks a lot of ”what if” questions that reflect the requirements. It is however not a trivial task to construct the right questions and to know when to stop generating scenarios. This requires a lot of experience and knowledge, which can be achieved by being involved in many design projects. A *scenario* is always system specific, i.e. tailor-made for a particular application in a domain, whereas questions that are valid for all architectures in a particular domain resides in a checklist. The items in the checklist can either generate scenarios or be verified in the measuring stage directly. As an example, consider the domain of safety-critical real-time systems. The checklist contains the following items:

1. Is the system schedulable?
2. Is there error recovery code in the system to clean up after error detection?

The first item is verified directly by performing a mathematical schedulability analysis. The second item is too general and therefore it must be formalized into a set of scenarios before it can be answered. As scenarios are system specific, they can stress different types of errors in specific modules residing in the system. One possible scenario is: “*What happen when division by zero occurs in the control task*”. The scenarios can than be verified by, for instance, simulation or scenario execution, both described later in this chapter.

The questionnaire-based questioning typically stresses general logistical software architecture questions. These questions have usually very little to do with the quality of the software itself, but rather focusing on issues such as documentation, and how the architecture was generated. Although the logistical questions do not examine the quality of the software product itself, it has impact on the quality since good quality requires a mature development process. Examples of such questions are: "Is a standard architectural description language used?", or "Is the intended work distribution supported by the architecture?".

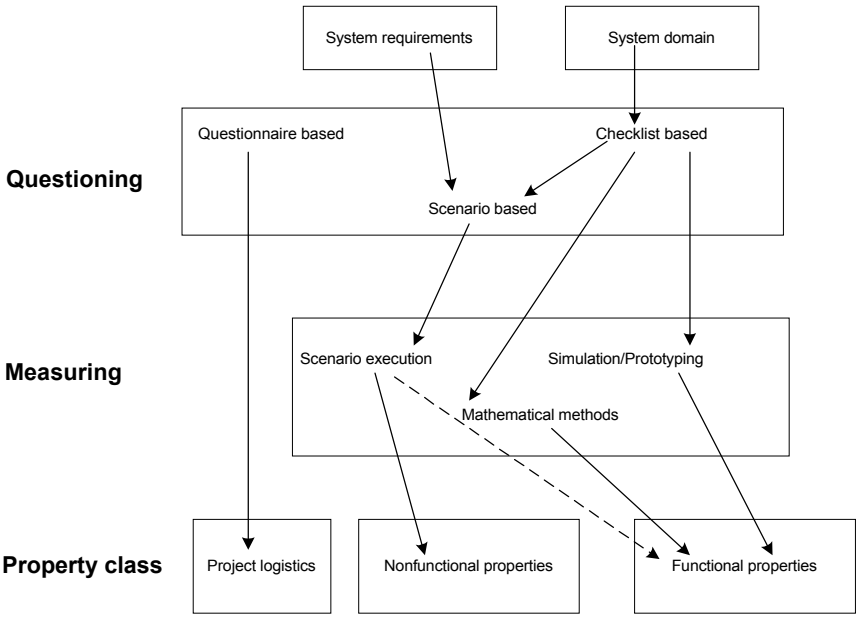
There are a couple of measuring techniques available for architectural analysis namely *scenario execution, simulation* and *prototyping, mathematical methods* and *experience based knowledge reasoning*. The idea with *scenario execution* is to "execute" the question stated by a scenario on the architecture. By executing a scenario is meant that the effects on the architecture imposed by a scenario is investigated. This method is particular suited for analysis of non-functional quality properties.

Simulation requires a prototype implementation of the architecture. Such a prototype should be as small as possible, containing only the information needed for the analysis to be performed. Simulation is a method targeting on analysis of functional quality properties.

Experienced-based reasoning can be used for any of the two classes of quality properties. Actually, experienced-based reasoning is usually how the software architecture evaluation is done in industry today, although in a relatively unorganized manner. As an organization and its development process mature, more of the formal evaluation techniques will be adopted.

Mathematical methods can be used provided that a mathematical model of the architecture exists. Such a model is provided by, e.g. timed automata. More examples of mathematical measuring techniques are the schedulability test for real-time systems and statistical reliability modeling. These methods give a clear yes- or no answer, or a quantitative value that is comparable among all different types of software applications.

Figure 10 provides a schematic picture of how the different evaluation techniques relate.



Although measuring techniques might give quantitative values, these values must be treated carefully. The quantitative values should be used as relative values when comparing competing software architectures. Moreover, if scenarios or experienced reasoning was used to obtain the values, the exact same set of scenarios and reasoning must be used when evaluating the competing or refined architecture. Otherwise, the measures are not comparable. Consequently, it is impossible to compare measured quality of a software architecture across the application domain i.e. within the same class of products but in different environments or applications.

4.2 Functional analysis

There exist functional quality properties in abundance, among which the properties of particular interest when designing safety-critical real-time system is listed in Table 1.

Figure 10. Schematic picture of the relations between the evaluation techniques

	time
Safety	The property of the system that it will not endanger human life or the environment
Security	The ability of a software system to resist malicious intended actions
Availability	The probability of a system functioning correctly at any given time
Temporal constraints	Real-time attributes such as deadlines, jitter, response time, worst case execution times (wcet), etc.

Table 1. Functional quality properties

Performance

Certain functional properties of a software system are tricky or even impossible to predict using the architectural description level only, e.g. performance. Performance estimations must have the algorithmic solutions as input. As discussed in the introduction, software architecture is a description of the system on a higher level of abstraction than algorithmic solutions and data structures. However, by using prototyping and simulation techniques, performance in terms of ,for instance, event throughput or queuing length for events in a system, can be estimated [GRBO]. Since such a performance measure is not absolute, it can only be used when comparing two different architectural solutions, not when estimating, for instance, the worst execution time for handling an event in the system.

Reliability

There are mathematical methods based on probability theories such as Markov models for assessing reliability [Tram95]. However, these theories are developed for hardware where failures often are caused by physical wear such as corrosion, overheating, etc. Such failures are probabilistic in nature whereas software failures are mistakes (errors), made in the specification, the design or in the implementation. These types of failures are certainly not probabilistic according to some distribution over time. Furthermore, software can never be worn out. Attempts have been made to apply the methods from the hardware community to software. In software, the statistics are the numbers of errors in the program or the likelihood of a failure in a point of time based upon the error distribution in the past [Fenton96]. To get such failure estimations, there must be an implementation of the application or at least a prototype. Anyhow, a description of the application on a lower level than the architecture is needed. With some heuristics from similar applications developed earlier experienced engineers can estimate the expected number of errors in the components. Such estimations are very complex, giving rough metrics. An alternative to directly measure the reliability of the architecture is to measure the testability. The testability is a function of the effort required in order to assure the required level of reliability or availability.

There are three different approaches to handle faults in order to achieve a reliable system [Lapr92]:

- Fault avoidance
- Fault removal
- Fault tolerance

Fault avoidance is about designing error free systems. This implies the use of structured design methodologies such as formal methods or semi-formal methods. Formal methods are based on mathematical models of the software system and the requirement specification. These models form the basis when proving correctness of the model with respect to the system specification. There exists a wide area of formal methods and formal modeling languages, each supporting different system domains. Semi-formal methods are, as the name suggests, less formal, i.e. they do not support techniques to exhaustively prove correctness of the models. Instead, they offer a structured way of reasoning, both when designing models of the system and when analyzing the models. The methods are usually based on some "formal" notation, e.g. *Unified Modeling Language* (UML)[BRJ98], ADLs, etc., representing the system model. Examples of such methods are *object-oriented analysis and design* (OOA/OOD), and software architecture techniques in general.

No matter how accurate the models are analyzed, there may still be errors in the implementation. These errors usually originate from the specification and from the mismatch when mapping the models to the source code. In order to improve reliability in the program, *fault removal* techniques can be applied. Fault removal is basically the task of finding the errors by testing and removal of them by error correction. Under the assumption that no new errors are introduced, the reliability will grow as errors are corrected. This assumption is, unfortunately, seldom true, implying that the whole system has to be re-tested after each increment. The results from testing and re-testing can be used for statistically forecasting of the failure rate (and consequently the reliability), of a software system. Such a method is the *reliability growth model*, first proposed for software by Jelinsky et. al. [JEMO72]. There exist an abundance of

different approaches to model reliability growth; they are all based on data collected during testing, but differ in the way the statistical model is made.

Some faults are impossible to avoid regardless of how accurate the design and the tests are performed. If it is particularly important that a certain module in the system does not fail, fault-tolerance can be introduced. *Fault-tolerance* is a technique which can be interpreted in two different ways: it could be the ability of a software system to tolerate faults from its environment, e.g. the operator, hardware errors, etc., or it could mean that the system should be tolerant against design faults in the software itself. The two different fault-tolerance approaches are, naturally, solved using different techniques. For instance, to be fault-tolerant against hardware errors such as electromagnetic distortion, redundant hardware can be used, each with equivalent software running on them. This solution will however not tolerate software faults. Different approaches to be tolerant against software faults are *recovery blocks* and *N-version programming* [Storey96][CA78].

Recovery blocks are based on acceptant tests of the calculated values. If the processed value is not accepted the program tracks back to a recovery point where it is safe to continue the execution after having restored the system's state.

N-version programming is achieved by developing *N* different versions of the software; each developed by different and isolated design teams. All *N* different versions run in parallel at runtime and their respective results are voted upon. This technique has, however, been proven not so successful since all different versions of the software start out from the same specification, and since most design errors originate from the specification, they will contain common errors.

Even if the source code is absolutely correct, the compiler may still produce erroneous binaries. Faults introduced by the compiler can be tolerated by using the N-version approach. Each version has exactly the same code, but they are all compiled using different compilers.

It is important to note that the different techniques discussed above can be applied at any stage in the development process. For instance fault removal can be used when verifying the designed architecture against the system specification. Fault-tolerance is also a matter of architectural design. The techniques for fault-tolerance discussed above are all achieved using different architectural solutions.

Safety

Safety seems, at a first glance, very similar to reliability. There is however a clear distinction as safety is only concerned with failures that endangers human life and the environment, i.e. hazards, whereas reliability deals with all failures regardless of their consequences. However, before any safety analysis of the architecture can be performed, the hazards must be identified. This is done in a *hazard analysis* that is a reasoning based method for finding all hazards in the system that is going to be designed [Leve95].

There exist several techniques for assessing safety properties in software designs. Most of them are scenario based and work either backward or forward. If the method works backwards, the analysis starts with the hazard as a scenario, trying to trace

down the responsible component. On the contrary, if the method works forward, the effects of an error in a component is investigated.

Some of the most well known forward methods are Failure Mode and Effects Analysis (FMEA) and Hazard and Operability studies (HAZOP). Both methods analyze the consequences of failures in the components. One commonly used backward technique is called Fault Tree Analysis (FTA)[Storey96]. FTA starts with a hazard, trying to determine its origin among the components. This kind of analyses give an understanding of where in the architecture fault-tolerance techniques should be introduced, or if already introduced, verifying whether the intended fault-tolerance is achieved or not.

Depending on the results from the safety analysis, changes in the design may have to be performed. Different design approaches to avoid catastrophic failures can be applied based on the severity of an accident caused by the hazard. The different approaches are [Leve95]:

- Hazard elimination
- Hazard reduction
- Hazard control
- Damage minimization

The severity is a quantified value that makes it possible to compare and rank hazards. Typically, the severity is given in terms of the cost or, lost lives, for the stakeholder if the accident occurs.

Substitution, *decoupling*, and *simplifications* achieve *hazard elimination*. By substitute a dangerous design possibility by a functionally equivalent, but not dangerous solution, the hazard itself is eliminated. For instance, if the system involves a very toxic chemical liquid, substituting the liquid with a non-toxic one eliminates the hazard. Moreover, by decoupling safety-critical parts of the software from non-critical software, the risk for an error in the non-critical part to propagate into the safety-critical parts is eliminated. There exist some known architectural solutions based on decoupling, e.g. safety kernels, firewalls, hierarchical architectures [Storey96].

Hazard reduction reduces the likelihood of the occurrence of a hazard. It might not be feasible or even possible, to eliminate the hazards. Then the designer has to design the system in such a way that the hazard is not very likely to occur. An example of hazard reduction is to erect a fence around an industrial robot, preventing humans to come close enough in order to get hurt.

Hazard control is applied in order to reduce the likelihood of an accident if a hazard arises. This can be achieved using *fail-safe design*, i.e. the system should be designed to detect the hazard and then transfer it into a safe state if such exists. There are, however systems where no safe state exists. A typically example of such a system is airplanes. These systems must keep operating even if something goes wrong. This is achieved using fault-tolerance such as *redundancy*. It is essential that an airplane keeps flying even if one engine breaks down by using the second engine. The performance will of course be reduced, but the airplane can still be maneuvered to its safe state on the ground.

Yet, if an accident still occurs, the consequences and losses must be reduced. This is achieved with damage minimization that strives to minimize the exposure of the accident to the environment or human beings.

Availability

Reliability and availability are strongly correlated. According to the definitions given in Table 1, reliability is the probability of a software system functioning correctly over a given *period* of time and availability is the probability of a software system functioning correctly at *any* given time. More generally, reliability is equivalent to Mean-Time-Between-Failure (MTBF) and the availability is a percentage figure given by the formula below:

$$Availability = 1 - \frac{MTTR}{MTBF}$$

MTTR is an abbreviation for *Mean-Time-To-Repair*, i.e. time spent on service. The relation is shown graphically in Figure 11 below. If any point of time is picked randomly along the y-axis, there is a probability of having correct functionality, i.e. the availability of the software system.

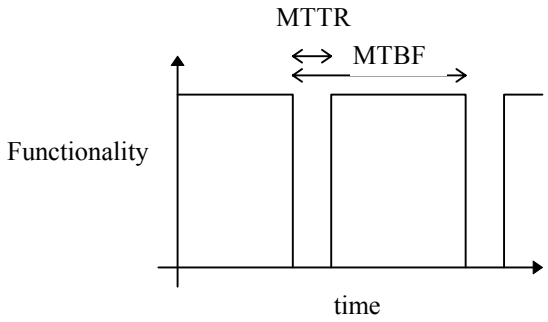


Figure 11. Availability and reliability

Security

Security is concerned with protecting a software system from malicious intended actions, e.g. intrusion by unauthorized users or locking out unintended accesses to safety-critical parts of the system. This can be achieved by different architectural solutions: safety/security kernels, firewalls, etc. which all are different ways of restricting the access to the system or sub-systems. As security can be achieved using different architectural solutions, it can be assumed that security is assessable by architectural analysis. A scenario-based method can be used. Typically, such a scenario could reason about what happens if an operator or a sub-module tries to access a protected region of the system. Another possible way of analyzing software architectures from the security point of view, is simulation, provided that the logical view of the software architecture contains sufficient information regarding rules for authorization and identification.

Real-time requirements

When designing real-time systems it is important to ensure the temporal correctness of tasks in the application. The timing must be just perfect, neither too fast nor too slow. The information necessary for the verification of temporal constraints is provided by the temporal view of the architecture. A typical example of such an analysis are schedulability test, i.e. analyzing whether the task set is schedulable or not given the resources and temporal constraints given as release times, deadlines, worst case execution times (wcet), jitter, etc. The resources taken into account when analyzing the schedulability of a system are typically CPUs, communication busses, actuators, etc.

There exist a lot of mathematical methods for verifying the temporal behavior of a real-time system, all having different assumptions on the scheduling strategy and the task model [LILA73][ABDTW95]. A task model defines the temporal requirements put upon a task, i.e. priorities, period times, etc. The task model and the scheduling strategy is strongly coupled since the task model provide the input to the schedulability analysis.

In Figure 12, a classification of different scheduling strategies is illustrated.

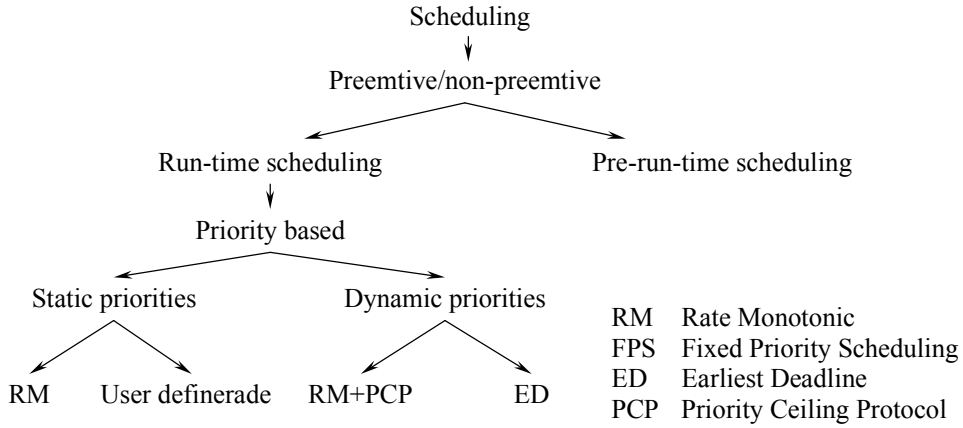


Figure 12. Classification of scheduling strategies.

4.3 Nonfunctional analysis

The number of nonfunctional quality properties is, as the functional quality properties in the previous chapter, very large. In Table 2, a subset of all such quality properties is listed, all being important in a mature and modern design process for real-time systems.

Cost	The cost for performing any action such as development, evolution and verification
Testability	How easy it is to prove correctness of the system by testing
Reusability	The extent to which the architecture can be reused
Portability	How easy it is to move the software system to a different hardware- and/or software platform

Maintainability	The aptitude of a system to undergo repair and evolution
Modifiability	How sensible the architecture is to changes in one or several components

Table 2. Nonfunctional quality properties

A very simple but yet powerful method for analysis of nonfunctional quality properties is execution of scenarios. Several of the direct and indirect quality properties listed in Table 2 can be examined and analyzed by using scenarios. By direct we mean an attribute that focus on the software only such as the reusability of a module or subsystem or the portability i.e. how easy or hard it is to move the system to another operating system or hardware platform. An indirect property is one that depends on a direct one. A typical example is the cost. The cost is always related to the action, for instance the cost associated with testing, development, maintenance, etc.

Cost

As discussed above, cost is an indirect quality property, always depending on other quality properties of the system. Typically, after a system has been released and been running for a while, new functionality is required from the customer or new features and improvements are desired within the organization. Then the cost is probably dependent on the reusability, maintainability and testability of the software. Cost estimations are probably one of the hardest tasks for every development project. The cost estimation for the design of a completely new system is extremely hard to achieve. Usually such estimations are based only upon historical experiences with similar systems. If no such experience is available, the estimation gets even more imprecise. The software architecture description could help illuminate the cost of developing a system or adding new functionality to an existing system. Partly by being a structured description of the application, helping the designer to get a full perspective of the application scope, but also by providing techniques for analyzing the effects of adding new features to an existing software system.

Testability

Testing is essential in order to prove functional correctness of a software system. It is also used for obtaining some confidence in functional quality properties such as reliability, performance, etc. A lot of time and consequently, money is spent in the testing phase of software development. To reduce the amount of time needed for testing of the software, the architecture must be designed so that it is easy to test, i.e. having high testability. The testability is dependent on three individual properties: *observability, controllability*, and for concurrent systems and systems dependent on time, *reproducibility* [Bind94]. Testability is consequently an indirect quality property as well.

In order for a test case to be useful, the result of it must be observed. If the components in the architecture are seen as “black boxes”, i.e. the structural view, only

the interfaces are observable. The bigger interface, the more visibility. Apparently, bigger interfaces give higher *observability,* thus higher testability.

When performing a test, a particular input is given to the system or a sub-system. This input is the only way in which the test engineer can control the path taken in the program. If the path taken only depends on the input itself, maximum controllability is achieved. This is of course not the case in general. There are often data dependencies between different modules such as global variables etc. If those data dependencies, which are not controllable by the test input data, affect the control flow, the controllability is decreased, giving lower testability.

Finally, when testing concurrent system or real-time systems in general, the order in which different processes in the system are executed will influence the observed result from a test. For instance, in a system controlling the water level in a tank, there is one process sampling the actual water level and one process calculating how to adjust the water level based on the measured value and some set value. If the control process executes twice without any intermediate execution of the sampling process, the result of the control decision will be different in the second invocation than if the water level was re-sampled in between. To get high testability, the order in which processes execute must be controllable or deterministic, i.e. high reproducibility [ThHa99].

Reusability

Reusing a software component to its full extent, without any modifications, is extremely difficult if not the domain in which the reuse is intended is the exact domain of the component origin. When a component or architecture is reused in the same application domain we call it a domain-dependent reuse. When containers are reused, i.e. lists, arrays, sets, etc., they can be reused across different application domains. An example of such reuse is the Standard Template Library (STL) for the object-oriented language C++. Reuse, which is possible across the application domains, is consequently called domain-independent reuse.

When analyzing the level of reusability of a component or a part of the architecture, one must consider not only the original application domain, but also how isolated and independent it is from rest of the system. The less dependencies, the more reusable, and vice versa.

The focus on reuse, in industry, has been intensified due to the potential cuts of cost. The time spent on implementation decreases when reusing components. Furthermore, components can be bought from third-party developers. Such components are called Commercial-Off-The-Shelf components (COTS).

Portability

To be able to analyze software architectures with respect to portability, the platform on which the system is going to run on has to be modeled as well. This to unveil the dependencies between the software components in the system and the *platform*. As platform we consider the hardware, e.g. processors, A/D converters, as well as the software providing the infrastructure e.g. operating systems. If the amount of direct dependencies, i.e. the number of components having a direct connection to the

platform, is low, then the architecture as whole is quite insensible to a change of platform. Thus, having a high degree of portability.

Maintainability

Kazman et. al. [KAC96], have proposed a methodology for visualizing the amount of changes required in the modules or in the architecture when adding or changing functionality in the system. The amount of changes in the software architecture enforced by adding new functionality or error corrections, are referred to as *maintainability*. By using scenarios developed from the requirements of the new function, the existing architecture is analyzed.

The concept, direct scenarios, were introduced meaning scenarios that are directly supported by the existing architecture i.e. no major architectural changes are required. In opposite, an indirect scenario exposes the need for architectural changes, which is more difficult and costly to achieve. Remember that there is a difference between a direct or indirect scenario and the direct and indirect quality properties introduced earlier in this chapter. After having mapped the scenarios on the architectural structure and determined if the scenario is direct or indirect, scenario interaction should be revealed. Two or more indirect scenarios are said to interact if they affect the same module.

To make the potential architectural violations and changes in the system visible, graphical representation of modules were scaled in the ADL according to the amount of indirect scenario interactions.

5 Architectural design

Architectural analysis can, and should, be used as guidance when designing a software system. A software system can be implemented in several ways, all having different architectural solutions. By using architectural analysis, the architecture that fulfills the requirements best can be chosen. The workflow for designing architectures for a system is shown in Figure 13.

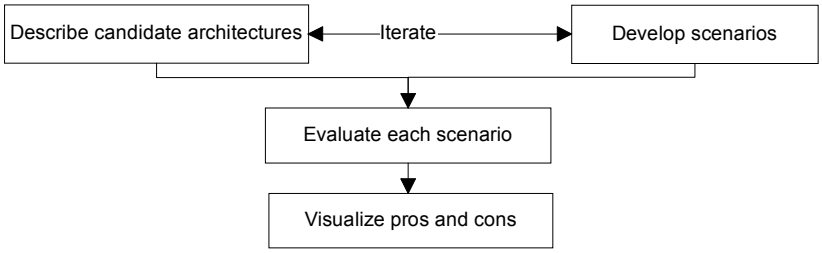


Figure 13. Architecture development and analysis process.

The first phase when developing a software system is to develop candidate architectures and a set of scenarios that reflects the requirements on the system. The number of scenarios to develop is related to the generation of ordinary test cases. Eventually, a state is reached where the added value of a new scenario is less then the effort required to develop the scenario itself. When this point in time is reached or when the development budget is violated, the scenario generation should stop.

Now we have the candidate architecture and a set of scenarios. By executing the scenarios on the architecture a table with the desired quality attributes can be constructed. In the table, all requirements are marked with plus signs and minus signs representing how well the architecture fulfills the requirements. If the result from the analysis is satisfactory, the next phase is to do low-level design and implementation. However, if the analysis results are not satisfactory, an alternative architecture must be developed on which exactly the same scenarios are executed. Consequently, the evaluation must be done all over again. The work of finding a sufficient architecture is highly iterative, meaning that the architecture can evolve by small steps until a reasonable solution is found. Consequently, changes suggested by the analysis may result in a complete redesign using a completely different architectural style or minor modifications in subsystems only.

The table produced in the analysis phase containing all the analyzed quality properties constitutes the input to a *tradeoff analysis*. In a tradeoff analysis the set of competing architectures is compared or the result from a refined architectural solution is compared with the result from the analysis of the preceding generation of the architecture. The objective of the tradeoff analysis is to choose the architectural alternative that best complies with the ranking among the quality properties.

A method for tradeoff analysis called *Architecture Tradeoff Analysis Method* (ATA) has been developed at the Software engineering institute (SEI) at Canegie Mellon university [Kazm98]. It is an iterative development method that is similar to the process shown in Figure 13.

A method called *Software Architecture Analysis Method* (SAAM) is also developed at SEI. The purpose of SAAM is to analyze software quality attributes by examining competing architectures [KBAW94]. To do so, they partitioning the functionality in the architecture i.e. identifies were in the different architectures the functionality of the system is allocated. The functional partitioning is system domain specific. Some domains already have a well-defined functional partitioning; a typical example of such a domain is compilers. Compilers are built with a front-end, a parser, a code generator etc. However, nothing is assumed about how functions are organized and structured, i.e. the architecture of the compiler. This partitioning gives a common description and common modules, each with the same functionality but organized in different ways. The communal description is an absolute condition for the comparison, which aims to unveil how well a certain quality attribute, is adopted by the architecture. Again, the analysis is based on scenarios, constructing input for a tradeoff analysis.

5.1 An example

As an example of how an architecture is constructed, analyzed and transformed in order to better comply with the requirements consider a real-time system that controls the water level in a tank. The system samples a water level sensor, takes a decision whether to let water out, or pour water into the tank. The system actuates a pump or a valve if the level has to be adjusted. As it is a real-time system, the temporal constraints on the system must be fulfilled, i.e. there is a functional quality requirement on timing. Moreover, the system should easily be modified to run on different platforms (real-time operating system and hardware), i.e. portability.

First, the structural view of the architecture is developed, identifying the components in the system and their interconnections. In this case, the interconnections represents transportation of data among the task using services provided by the RTOS. While portability is crucial, the operating system and the hardware view is modeled as well. The first candidate architecture is shown in Figure 14.

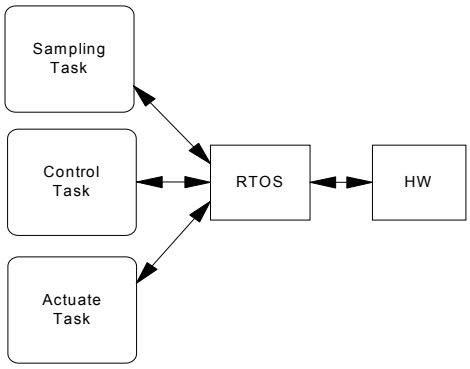


Figure 14. The first candidate architecture for a water tank controller

Next the compliance between the architecture and the required quality properties must be analyzed. Verifying the temporal behavior requires the temporal view of the architecture. For this particular application, the period time, the estimated worst execution time (wcet), and the deadlines for the three tasks is shown in Table 3.

Task	Period time (T)	wcet (C)	Deadline (D)
Sampling task	1 ms	50 μs	60 μs
Control task	2 ms	200 μs	1 ms
Actuate task	2 ms	50 μs	1 ms

Table 3. The temporal view

The temporal behavior is verified using *exact analysis* where the worst case response time for all tasks is calculated. If the response times are less than the specified deadlines for all tasks, the system is schedulable [JOPA86]. Exact analysis requires priorities to be assigned to the tasks. In this particular example, priorities are assigned according to the *rate monotonic* algorithm where the task with the shortest period gets highest priority [LILA73]. Rate monotonic gives the sampling task high priority, the control task medium priority and the actuating task low priority. The exact analysis formula is recursive and calculates the worst case response time with respect to interference of the execution of tasks with higher priorities. The recursion stops when two subsequent calculations result in the same response time, i.e. a fix-point is reached. The formula is shown below:

$$R_i^{n+1} = C_i + \sum_{\forall j \in hp(i)} \left\lceil \frac{R_i^n}{T_j} \right\rceil C_j$$

$\forall j \in hp(i)$ Denotes all tasks j with higher priority than task i.

The response times for the sampling task is 50 μs as no other task interferes with it since it has the highest priority. The response time for the control task is 250 μs. Finally, the actuate task has a response time of 300 μs. If the calculated response times are compared to the specified deadlines, it could easily be verified that the system is schedulable as the response times for all tasks are less than corresponding deadlines.

To assess portability, scenarios can be used. For the matter of simplicity, only one scenario is used in this example, namely: ”*Move the system to another platform*”. The idea is to execute this scenario on the proposed software architecture to estimate the number of component being subjects to changes. As portability is the issue, the number of affected components should be held to a minimum. In the architecture suggested in Figure 14, all the components interact with the real-time operating system. Consequently, there are a lot of platform specific system calls embedded in each and every component, giving poor portability since every component has to be changed as a result of a changed platform. To increase the portability, architectural transformations have to be performed, i.e. the software architecture has to be refined. One possible transformation is to introduce a *proxy-component* between the task components and the real-time operating system. This transformation is shown in Figure 15.

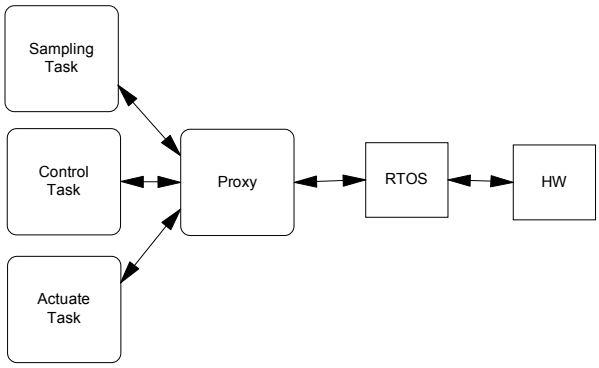


Figure 15. The architecture after the transformation

The proxy provides the tasks with all necessary services in order for them to perform their intended tasks, while hiding the actual system calls. To verify the new architecture according to the requirements, the scenario has to be re-executed. Now the proxy component is the only one affected by a changed platform, i.e. a maximal portability is achieved. However, the portability is achieved at the expense of an increased overhead for system calls. Therefore, the worst case execution times for the individual task components must be re-estimated and the exact analysis must be done all over again to verify the temporal behavior of the system. The phenomena that quality properties might affect each other in a negative manner, is referred to as *tradeoff*.

6 Conclusions

Software architecture is part of what generally is referred to as software engineering. Software engineering also includes a lot of other techniques like software metrics, formal methods, test methodologies, etc. Thus, software engineering is an umbrella for all techniques and methods needed to establish a ”science of engineering” practice in the software community. Software architectures are an important part of software engineering since it deals with high-level modeling and evaluation. The software architecture community is still very young, but the recent interests from the industry have launched a lot of research activities in academia. Especially relevant are the software architecture analysis methods as the analysis provides the information for early design decisions.

To make architectural analysis possible, the architecture must be described in a language with well-defined semantics. A language that describes software architectures is called Architectural Description Language (ADL). There exists a lot of different ADL:s, but few of them have received any particular attention since it is very difficult to design a language with syntax and semantics powerful enough to cover all possible application domains and that can be interpreted by all stakeholders in a project. As a consequence, software developers use their own description languages. An important property of an ADL is the architectural views, providing detailed information needed for the analysis. The number of views and the contents of each view will vary between different application domains and the required analyses. Finally, a description language with a well-defined semantics is also a necessary condition for developing tools that support architectural development and evaluation.

This report has described existing techniques for describing and evaluating software designs based on information mainly provided by the high level description, i.e. the software architecture. The ability to evaluate early design decisions is very important since early design decisions are crucial for the final result, both regarding correct functionality and cost. The earlier design mistakes are detected, the less time has to be spent on redesign. The properties analyzed using software architectures are called quality properties. In this survey, the quality properties are divided into two separate classes, functional and nonfunctional. Functional quality properties are concerned with the run-time behavior of the software system, for instance performance and reliability. In contrast, nonfunctional quality properties are concerned with the quality of the software itself. Examples of nonfunctional properties are reusability, maintainability, and testability.

Tool support for architectural development and evaluation is poor. It is possible to formalize knowledge in frameworks, guiding the designer in both architectural transformations and in the tradeoff analysis. There exist tools for some of the analyses, for instance tools for verifying the temporal behavior in a real-time system [ERGUSA97], but these tools are still islands in the ocean called software engineering. We need to discover, or build new islands and connect them to each other in order to get complete suits of tools, supporting the complete software development- and maintenance process. In mature engineering disciplines, such tool support is taken for granted. Software engineering tools will probably appear as the software

community gets more mature, it is still very young, at least when compared to other traditional engineering disciplines.

7 References

[ABDTW95] N. C. Audsley, A. Burns, R. I. Davis, K. Tindell, and A. J. Wellings, Fixed Priority Pre-emptive Sceduling: An Historical Perspective, Real-Time Systems 8(2-3):173-198, 1995

[ALDI92] R. Alur, and D. L. Dill, A theory of timed automata, 1992

[BCK98] L. Bass, P. Clements, and R. Kazman, Software Architecture in Practice, Addison Wesley 1998

[Beng97] PO. Bengtsson, and J. Bosch, Scenario-based Software Architecture Reengineering, University of Karlskrona/Ronneby 1998

[Bind94] R. V. Binder, Design for Testability in Object-Oriented Systems, Communications of the ACM, Volume 37, No 9, pp. 87-101, 1994

[BRJ98] G. Booch, J. Rumbaugh, and I. Jacobson, The Unified Modeling Language User Guide, Addison Wesley ISBN 0-201-57168-4, 1998

[BuWe94] A. Burns, and A. Wellings, HRT-HOOD, a Structured Design Method for Hard Real-Time Systems, 1994

[CA78] L. Chen, and A. Avizienis, N-version programming: a Fault Tolerant Approach to Reliability of Software Operation, In proceedings of 8th Annual International Conference on Fault Tolerant Computing, pp. 3-9, 1978

[Clem96b] P. C. Clements, and L. M. Northrop, Software Architecture: An Executive Overview, Technical report CMU/SEI-96-TR-003 1996

[Clem96a] P. C. Clements, Coming Attractions in Software Architecture, Technical report CMU/SEI-96-TR-003 1996

[DaYo95] C. Daws, and S. Yovine, Two examples of verification of multirate timed automata with KRONOS, In proceedings of 16th IEEE Real-Time Systems Symposium, PP 66-77, 1995

[dijk68] E. W. Dijkstra, The Structure of ”THE”-Multiprogramming System, ACM on Operating System Principles 1967

[EHLS94] S. Edwards, W Heym, T. Long, M. Sitarman, and B. Weide, Specifying Components in RESOLVE, Software Engineering Notes, vol. 19, no. 4, 194

[ERGUSA97] K. Sandström, C. Eriksson, and M. Gustafsson, RealTimeTalk - a Design Framework for Real Time Systems - a Case Study, SNART 1997

[Fenton96] N.E. Fenton, and S. Lawrence Pfleeger, Software Metrics, International Thomson Computer Press 1996

[Garl93] D. Garlan, and M. Shaw, An Introduction to Software Architecture, Advances in Software Engineeri Vol 1 World Scientific Publishing Company 1993

[GHJV94] E. Gamma, R. Helm, R. Johanson, and J. Vlissides, Design Patterns - Elements of Reusable Object-Oriented Software, Addison-Wesley 1994

[GRBO] H. Grahn, and J. Bosch, A Simulation Approach to Predict and Evaluate the Performance of Software Architectures, University of Karlskrona/Ronneby 1998

[JEMO72] Z. Jelinsky, and B.P. Moranda, Software Reliability Research, Statistical Computer Performance Evaluation, pp 465-484, New York , SA, Academic Press, 1972

[JOPA86] M. Joseph, and P. Pandya, Finding Response Times in a Real-Time System, The Computer Journal, Volume 29, No. 5, pp. 390-395, 1986

[KAC96] R. Kazman, G. Abowd, L. Bass, and P. Clements, Scenario-Based Analysis of Software Architecture, IEEE Software 1996

[KBAW94] R. Kazman, L. Bass, G. Abowd, and M. Webb, SAAM: A Method for Analyzing the Properties of Software Engineering, Int. Conf. On Software Engineering IEEE Computer Science Press pp. 81-90, 1994

[Kazm98] R. Kazman, M. Klein, M. Barbacci, T. Longstaff, H. Lipson, and J. Carriere, The Architecture Tradeoff Analysis Method, Submitted to the 1998 International Conference on Software Engineering

[LaLe94] P. B. Ladkin, and S. Leue, What Do Message Sequence Charts Mean?, IFIP-Transactions-C:-Communication-Systems.n C-22, pp 301-316, 1994

[Lapr92] J.C. Laprie, Dependability: Basic Concepts and Associated Terminology, Dependable Computing and Fault-Tolerant Systems, vol. 5, Springer Verlag, 1992

[Leve95] N.G. Leveson, Safeware, System Safety and Computers, Addison Wesley 1995

[LILA73] C. L. Liu, and J. W. Layland, Scheduling Algorithms for Multiprogramming in a Hard Real-Time Environment, Journal of ACM, Volume 20, Nr. 1, pp. 46-61, 1973

[LKAV93] D. Luckham, J. Kenney, L. Augustin, J. Vera, D. Bryan, and W. Mann, Specification and Analysis of System Architecture Using Rapide, Stanford University technical report, 1993

[LPY97] K. G. Larsen, P. Pettersson, and W. Yi, Uppaal in a Nutshell, In Springer International Journal of Software Tools for Technology Transfer 1(1+2), 1997

[Masc87] The official handbook og MASCOT, Version 3.1, Issue 1, 1987

[Miln87] R. Milner, Communication and Concurrency, Prentice Hall 1989

[Paul94] F. Paulisch, Software Architecture and Reuse – An Inherent Conflict?, Proceedings of 3 rd International Conference on Software Reuse, pp 214, 1994

[SHGA96] M. Shaw, and D. Garlan, Software Architecture - Perspective on an emerging Disipline, Prentice Hall 1996

[Storey96] Neil Storey, Safety-Critical Computer Systems, Addison-Wesley 1996

[ThHa99] H. Thane, and H. Hansson, Towards Systematic Testing of Distributed Real-Time Systems, In proceedings of the 20th IEEE Real-Time Systems Symposium, 1999

[Tram95] C. Trammell, Quantifying the reliability of software: statistical testing based on a usage model, In 'Experience and Practice', Proceedings., Second International IEEE Software Engineering Standards Symposium,. , pp. 208 –218, 1995

[Vest94] S. Vestal, Mode Changes in a Real-Time Architecture Description Language, Proceedings of 2nd International Workshop on Configurable Distributed Systems, 1994, Page(s):136-146

Appendix A - Terminology

ADL - Architectural Description Language, Language for describing software architectures

Architectural style - Standard types of architectures identified with names and patterns

Architectural view - Provide the architecture description with information needed when analyzing it. The components and their interconnections are shown in the structural view.

Architectural transformation – Changing the architecture in order to obtain required functionality and quality

Availability - The probability of a system functioning correctly at any given time

Checklist based questions – Domain specific questions used when evaluating a software architecture

Cost - The cost for performing any action such as development , evolution and verification

COTS - Commercial Off The Shelf components

Design patterns - Named object oriented solutions in the object oriented community

Design space - A N-dimensional space where every axis represents a design parameter, scaled with the different design options possible for that particular parameter

Direct scenario - A scenario that is directly supported by the architecture

Fault-tolerance - The ability of software to detect and tolerate errors in the design and/or from its environment

Framework - An architectural pattern for a particular domain, widely used in the object oriented community.

Functional quality property – Quality properties concerned with the run-time behavior of the software system

Indirect scenario – A scenario that requires an architectural transformation to be supported by the architecture

Maintainability - The aptitude of a system to undergo repair and evolution

Modifiability - How sensible the architecture is to changes in one or several components

MTBF – Mean-Time-Between-Failure

MTTR – Mean-Time-To-Repair

Nonfunctional quality property - Quality properties concerned with the software itself

Performance - How fast or slow the system performs its functions measured in time or the systems capacity measured in event-throughput

Portability - How easy it is to move the software system to a different hardware- and/or software platform

Reference style - Architectural styles widely used in particular application domains, e.g. the pipe-and-filter Architecture used in compilers.

Reliability - The probability of a system functioning correctly over a given period of time

Reusability - The extent to which the architecture can be reused

Safety - The property of the system that it will not endanger human life or the environment

Scenario based questions – Application specific questions used when evaluating a software architecture

Scenario execution - Method for analyzing an architecture by asking "what if" questions

Security - The ability of a software system to resist malicious intended actions

Temporal constraints - Real-time attributes such as deadlines, jitter, response time, worst case execution times (wcet), etc

Testability - How easy it is to prove correctness of the system by testing

Tradeoff - A relation between two or more quality attributes where an increased level of on property results in a decrease of another property.

Questionnaire based evaluation – Questions used when evaluating project logistic properties of software architectures

B

Time Automata as Task Models for Event-Driven Systems

By

Christer Norström, Anders Wall, and Wang Yi,
In proceedings of the 6th International Conference on Real-Time Computing Systems and Applications, 1999

Timed Automata as Task Models for Event-Driven Systems

Christer Norström[1] and Anders Wall[1]
[1]Mälardalen University
Department of Computer Engineering
P.O. Box 883, S-721 23 Västerås, Sweden
{awl,cen} @mdh.se

Wang Yi[1,2]
[2]Uppsala University
Department of Computer Systems
P.O. Box 325, S-751 05 Uppsala, Sweden
yi@docs.uu.se

Abstract

In this paper, we extend the classic model of timed automata with a notion of real time tasks. The main idea is to associate each discrete transition in a timed automaton with a task (an executable program). Intuitively, a discrete transition in an extended timed automaton denotes an event releasing a task and the guard on the transition specifies all the possible arriving times of the event (instead of the so–called minimal inter-arrival time). This yields a general model for hard real-time systems in which tasks may be periodic and non-periodic.

We show that the schedulability problem for the extended model can be transformed to a reachability problem for standard timed automata and thus it is decidable. This allows us to apply model-checking tools for timed automata to schedulability analysis for event-driven systems. In addition, based on the same model of a system, we may use the tools to verify other properties (e.g. safety and functionality) of the system. This unifies schedulability analysis and formal verification in one framework. We present an example where the model–checker UPPAAL *is applied to check the schedulability and safety properties of a control program for a turning lathe.*

1. Introduction

The traditional approach to the development of hard real-time system is often based on scheduling theory. There are various methods [5, 12, 7] e.g. rate monotonic scheduling, which have been very successful for the analysis of time-driven systems as tasks are *periodic.* To deal with *non-periodic* tasks in event–driven systems, the standard method is to consider non-periodic tasks as periodic using the *minimal* inter-arrival times as *task periods.* Clearly, the analysis result based on such a task model would be pessimistic in

many cases, e.g. a task set which is schedulable may be considered as non-schedulable as the inter-arrival times of the tasks may vary over time, that are not necessary minimal.

In recent years, in the area of formal methods, there have been several advances in formal modeling and analysis of real time systems based the theory of timed automata due to the pioneering work of Alur and Dill [2]. Notably, a number of verification tools have been developed (e.g. KRONOS and UPPAAL [6, 4]) in the framework of timed automata, that have been successfully applied in industrial case studies (e.g. [3, 13, 11]). Timed automata have proved expressive enough for many real-life examples, in particular, for event-driven systems. The advantage with timed automata is that one may specify very relaxed timing constraints on events (i.e. discrete transitions) than the traditional approach in which events are often considered to be periodic. However, it is not clear how the model of timed automata can be used for schedulability analysis. In this paper, we present an extended version of timed automata with real-time tasks to provide a model for event-driven systems. We show that the extended model can be used for both schedulability analysis and verification of other properties, e.g. safety and liveness properties of timed systems. This unifies schedulability analysis and formal verification in one framework.

The main idea is to associate each discrete transition in a timed automaton with a task (or several tasks in the general case). A task is assumed to be an executable program with two given parameters: its worst execution time and deadline. Intuitively, a discrete transition in an extended timed automaton denotes an event releasing a task and the guard (clock constraints) on the transition specifies all the possible arrival times of the associated task. Whenever a task is released, it will be put in the scheduling queue for execution. We assume that the tasks will be executed according to a given scheduling strategy e.g. earliest deadline first. Then a delay transition of the timed automaton corresponds to the execution of the task with earliest deadline and idling for the other waiting tasks.

Thus, the sequences of discrete transitions of an extended timed automaton will correspond to the sequences of *arrivals* of non-periodic tasks. We say that such a sequence of tasks is schedulable if all the tasks can be executed within their deadlines. Naturally an automaton is *schedulable* if all the task sequences are schedulable. We shall show that under the assumption that the tasks are non-preemptive, the schedulability problem can be transformed to a reachability problem for ordinary timed automata and thus it is decidable. This allows us to apply model-checking tools for timed automata to schedulability analysis for event-driven systems. We present an example where the model–checker UPPAAL is applied to check the schedulability and safety properties of a control program in control applications.

The rest of this paper is organized as follows: Section 2 presents the syntax and semantics of the extended timed automata with tasks. Section 3 shows how to transform the scedulability analysis problem for extended model to a reachability problem for ordinary timed automata, and thus schedulability analysis may be performed by the existing verification tools for timed automata. Section 4 provides an example to illustrate our approach. Section 5 concludes the paper with summarized results and future work.

2. Timed Automata with Real-Time Tasks

The theory of timed automata was first introduced in [2] and has since then established as a standard model for real time systems. We first give an brief review to fix the terminology and notation and then present an extended version of the model with tasks.

2.1. Timed Automata

A timed automaton is a standard finite-state automaton extended with a finite collection of real-valued clocks. The transitions of a timed automaton are labelled with a *guard* (a condition on clocks), an *action*, and a *clock reset* (a subset of clocks to be reset). Intuitively, a timed automaton starts execution with all clocks set to zero. Clocks increase uniformly with time while the automaton is within a node. A transition can be taken if the clocks fulfill the guard. By taking the transition, all clocks in the clock reset will be set to zero, while the remaining keep their values. Thus transitions occur instantaneously. Semantically, a state of an automaton is a pair of a control node and a *clock assignment*, i.e. the current setting of the clocks. Transitions in the semantic interpretation are either labelled with an action (if it is an instantaneous switch from the current node to another) or a positive real number i.e. a time delay (if the automaton stays within a node letting time pass).

For the formal definition, we assume a finite set of alphabets Act for actions and a finite set of real-valued variables C for clocks. We use a, b etc to range over Act and X_1, X_2 etc. to range over C. We use $\mathcal{B}(C)$ ranged over by g and later by ϕ etc, denote the set of conjunctive formulas of atomic constraints in the form: $X_i \sim m$ or $X_i - X_j \sim n$ where $X_i, X_j \in C$ are clocks, $\sim \in \{\leq, <, \geq, >\}$, and m, n are natural numbers. The elements of $\mathcal{B}(C)$ are called *clock constraints*.

Definition 1. *A timed automaton over actions Act and clocks C is a tuple $\langle N, l_0, E\rangle$ where*

- *N is a finite set of nodes,*
- *$l_0 \in N$ is the initial node, and*
- *$E \subseteq N \times \mathcal{B}(C) \times Act \times 2^C \times N$ is the set of edges.*

When $\langle l, g, a, r, l'\rangle \in E$, we write $l \xrightarrow{g,a,r} l'$. □

Formally, we represent the values of clocks as functions (called clock assignments) from C to the non–negative reals $\mathbb{R}_{\geq 0}$. We denote by $\mathcal{V}$ the set of clock assignments for C. A semantical *state* of an automaton is now a pair (l, u), where l is a node of the automaton and u is a clock assignment and the semantics of the automaton is given by a transition system with the following two types of transitions (corresponding to delay–transitions and action–transitions):

- $(l, u) \xrightarrow{d} (l, u + d)$
- $(l, u) \xrightarrow{a} (l', u')$ if $l \xrightarrow{g,a,r} l'$, $u \in g$ and $u' = [r \mapsto 0]u$

where for $d \in \mathbb{R}_{\geq 0}$, $u + d$ denotes the clock assignment which maps each clock X in C to the value $u(X) + d$, and for $r \subseteq C$, $[r \mapsto 0]u$ denotes the assignment for C which maps each clock in r to the value 0 and agrees with u over $C \backslash r$. By $u \in g$ we denote that the clock assignment u satisfies the constraint g.

2.2. Extended Timed Automata with Tasks

We shall view a timed automaton as an abstract model of a running process. The model

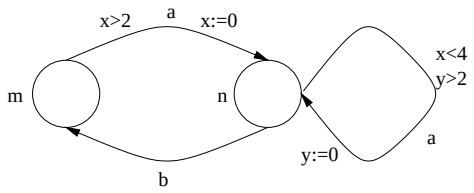


Figure 1. An Example Timed Automaton with Tasks.

describes the possible events (alphabets accepted by the automaton) that may occur during the execution of the process and the occurrence of the events must follow the timing constraints (given by the clock constraints). But the model gives no information on how these events should be handled. In many cases, for example in a control system, when an external event occurs, some computation must be performed to handle the event. A more concrete example is an interrupt handling system. Whenever an interrupt signal occurs, the associated interrupt handling program will be executed.

Now, assume that each action symbol in a timed automaton is associated with a program called *task*. Let P ranged over by p etc, denote the set of tasks. We further assume that the *worst case execution time* and *hard deadline* of the tasks in P are known. We shall use clock constraints to specify the arrival times of the tasks. Thus, each task p in P is characterized as a pair (c, d) of natural numbers with $c \leq d$ where c is the execution time of p and d is the relative deadline for p.

The deadline d is a relative deadline meaning that when task p is released, it should finish within d time units.

Definition 2. *An extended timed automaton with tasks (TAT), over actions Act, clocks C and tasks P is a tuple $\langle N, l_0, E, T\rangle$ where*

- *$\langle N, l_0, E, T\rangle$ is a standard timed automaton,*
- *$T : Act \hookrightarrow P$ is a partial function assigning tasks to actions.*

□

Semantically, an extended automaton may perform two types of transitions just as an ordinary timed automaton. In addition, an action transition will release a new instance of the task associated with the action. Assume that there is a queue holding all the task instances generated by action transitions and ready to run. The queue corresponds to the ready queue in an operating systems. A semantic state of an extended automaton is a triple consisting of a *node* (the current control node), a *clock assignment* (the current setting of the clocks) and a *task queue* (the current status of the ready queue).

Consider the automaton of Figure 1. Let $p1$ and $p2$ be tasks handling the interrupt signals a and b respectively. Assume that the initial state is $(m, [x = 0, y = 0], [])$ where the clocks are 0 and the task queue is empty. Then the automaton may demonstrate the following sequence of transitions:

$$\begin{aligned}(m, [x = 0, x = 0], []) &\xrightarrow{3} (m, [x = 3, y = 3], [])\\ &\xrightarrow{a} (n, [x = 0, y = 3], [p1])\\ &\xrightarrow{a} (n, [x = 0, y = 0], [p1, p1])\\ &\xrightarrow{3} (n, [x = 3, y = 3], [p1, p1])\\ &\xrightarrow{a} (n, [x = 3, y = 0], [p1, p1, p1])\\ &\xrightarrow{1} (n, [x = 4, y = 1], [p1, p1, p1])\\ &\xrightarrow{b} (m, [x = 4, y = 1], [p1, p1, p1, p2])\\ &\ldots\end{aligned}$$

Note that several instances of the same task may be released. However, the number of copies may be bounded by the clock constraints. For example, in state $(n, [x = 4, y = 1], [p1, p1, p1])$, no more instance of $p1$ will be released because the clock values will not satisfy the constraint $x < 4$ and $y > 2$, but

an instance of $p2$ may be released by the b-transition (which has no timing constraint).

In the above example, we have only shown that the task queue is growing due to action transitions. Now we discuss the effect of delay transitions on task queue. We shall see that the queue will be shrinking due to delay transitions. Let $p1 = p2 = (2, 8)$ i.e. the computation time of both $p1$ and $p2$ is 2 and the deadline is 8. We assume that there is a processor running the task instances according to a certain scheduling strategy. A delay transition with t time units is to execute the tasks in the queue with t time units. After the transition, a task will be removed from the queue (shrinking) if its computation time becomes 0 and the deadlines of all tasks in the queue will be decreased by t (since time has progressed by t). Now we have a precise description on the state changes for the above transition sequence:

$$\begin{aligned}(m, [x = 0, x = 0], []) &\xrightarrow{3} (m, [x = 3, y = 3], []) \\ &\xrightarrow{a} (n, [x = 0, y = 3], [(2, 8)]) \\ &\xrightarrow{a} (n, [x = 0, y = 0], [(2, 8), (2, 8)]) \\ &\xrightarrow{3} (n, [x = 3, y = 3], [(1, 5)]) \\ &\xrightarrow{a} (n, [x = 3, y = 0], [(1, 5), (2, 8)]) \\ &\xrightarrow{1} (n, [x = 4, y = 1], [(2, 7)]) \\ &\xrightarrow{b} (n, [x = 4, y = 1], [(2, 7), (2, 8)]) \\ &\dots\end{aligned}$$

More precisely we have the following assumptions on the underlining execution model:

1. A ready queue holding the task instances released and waiting for execution. A task instance will be removed from the queue when its computation time becomes 0.

2. An on-line scheduler Sch sorting the queue according to a given scheduling strategy. It will report $\perp$ if the queue becomes non-schedulable when a new task instance is added.

3. A single processor executing the tasks according to the ordering of the queue. It will always execute the task in the first position. The tasks are executed non-preemtive.

Further we use $\mathsf{Run}(q, t)$ to denote the resulted task queue after t time units of execution. The meaning of $\mathsf{Run}(q, t)$ should be obvious. For example, let $q = [(2, 7), (2, 8)]$ and $t = 3$ then $\mathsf{Run}(q, t) = [(1, 5)]$ in which the first task is finished and the second has been executed for 1 time unit. Now we are ready to present the transitional rules for extended timed automata.

Definition 3. *The semantics of an extended automaton is a transition system defined by the following transition rules (corresponding to release of new task and execution of existing tasks):*

- $(l, u, q) \xrightarrow{a} (l', u', \mathsf{Sch}(q'))$ *if* $l \xrightarrow{g,a,r} l'$, $u \in g$, $u' = u[r \mapsto 0]$, *and* $q' = q :: T(a)$
- $(l, u, q) \xrightarrow{t} (l, u + t, \mathsf{Run}(q, t))$

We shall write $(l, u, q) \longrightarrow (l', u', q')$ *if* $(l, u, q) \xrightarrow{a} (l', u', q')$ *for an action* a *or* $(l, u, q) \xrightarrow{d} (l', u', q')$ *for a delay* d. □

Finally, to handle concurrency and synchronization, parallel composition of extended timed automata may be introduced in the same way as for ordinary timed automata (e.g. see [10]) using the notion of synchronization function [8]. For example, consider the parallel composition $A||B$ of A and B over the same set of actions Act. The set of nodes of $A||B$ is simply the product of A's and B's nodes, the set of clocks is the (disjoint) union of A's and B's clocks, the edges are based

on synchronizable A's and B's edges with enabling conditions conjuncted and reset-sets unioned. Note that due to the notion of synchronization function [8], the action set of the parallel composition will be Act and thus the task assignment function for $A||B$ is the same as for A and B.

3. Schedulability Analysis as Reachability Analysis

Traditionally, the temporal attributes for a real-time computer systems are derived from their environment, e.g. period times, etc. These attributes are used for constructing a model of the system in terms of its temporal behavior. Such a temporal model is often called a task model, which is used to verify whether the system is schedulable or not, but other properties such as functional and safety properties can not be verified based on such a model. In our approach, we may construct a model for the whole system including the environment and tasks in the control system. The parallel composition of these models give us the possibility of not only verifying temporal constraints, but also its other aspects such as synchronization between tasks and simple computations within tasks etc.

Normally, a system is said to be schedulable if all tasks can always be executed within their deadlines, i.e. no deadlines are violated. The objective of the schedulability analysis is to verify that there are no violation of deadlines in all situations where the system may evolve to. Now we formalize the notion of schedulability for extended timed automata.

Definition 4. *An extended timed automaton A is non-schedulable if it may reach a non-schedulable state, that is: $(l_0, u_0, q_0) \longrightarrow^* (l, u, \perp)$ where (l_0, u_0, q_0) is the initial state of A, and $\longrightarrow^*$ is the transitive closure of $\longrightarrow$. We say that A is schedulable if and only if all its reachable states are schedulable.* □

Thus, the schedulability of extended automata can be checked by reachability analysis, to prove that $(l, u, \perp)$ is not reachable in the automaton. However, it is not obvious that the reachability problem for extended automata is decidable. In fact, the decidability of this problem is closely related to the preemptiveness of the tasks P. The following is one of our main results in this paper.

Theorem 1. *The problem of checking schedulability for extended timed automata over non-preemptive tasks P is decidable.*

Proof idea: It is based on the fact that the problem of schedulability checking for extended timed automata can be transformed to the reachability problem for standard timed automata, which is known to be decidable [1]. See the following subsection for details on the transformation. □

3.1. Transformation from TAT to ordinary timed automata

The idea is to construct a timed automaton simulating a ready queue and a scheduler that code all possible scenarios of the system described by a TAT, including the tasks in the queue and schedules. For example, consider the temporal attributes of the two tasks p_a and task p_b, where p_a had a worst-case-execution time (wcet), of 4 time units (tu), and a deadline (d), of 7 tu. The second task p_b has a wcet of 3 tu and a deadline of 5 tu.

Intuitively for a system to be schedulable, the ready queue can contain only a finite number of task instances. More precisely, there can only be MNT_i instances of task i, where MNT_i is given by:

$$MNT_i = \left\lfloor \frac{d_i}{c_i} \right\rfloor$$

where d_i denotes the deadline for task i and c_i denotes the computation time.

By calculating the maximum length of the ready queue, we know that to be schedulable, the queue in our example can only contain one instance of p_a and one instance of p_b. If at any time point, there are more than one instances of a particular task in the ready queue waiting for execution, we know for sure that the system is non-schedulable and the error state should be reached. This ensures a finite number of states in our model of the scheduler and the ready queue. Now, we use the above example to present the algorithm for constructing the scheduler and queue automaton, which can be generalized easily to the general case.

1. Create three different nodes, one node in which the ready queue is empty, one for which there exists task instances in the ready queue and, finally an error node.

2. Create transitions from the empty node to the running node, one for every action associated with a task. Furthermore, tasks can arrive while in the run node, consequently we need one transition from run back to run for every possible task instance as well. In order to keep track of every new task instance, a unique semaphore for every instance is introduced (denoted as taska and taskb in Figure 2). We also need a unique deadline clock for every instance in order to know which task to execute and to detect deadline violations.

3. According to EDF, the task having least time left until its deadline should be executed. For all possible task instances, create a transition from run to run which compares its relative deadline to all the other ready tasks. In our example p_a should be executing if $7 - d_a < 5 - d_b$, and p_b if $5 - d_b < 7 - d_a$ where d_a and d_b are the deadline clocks. In order to keep track of execution time of the running task, a clock is reseted on every release of a task. In our example, this clock is denoted as c. Furthermore, as we consider the non-preemptive case, no task can start to execute while another task already is executing. Thus we need a semaphore to know whether the processor is idle or not (denoted *r* in Figure 2).

4. Introduce one transition from run to run for every possible instance which terminates the task whenever c becomes equal to its specified execution time and its deadline clock is less or equal to its specified deadline. Termination is modeled by resetting the instance semaphore.

5. If ready queue gets empty, i.e. no tasks instances are present in the queue a transition to the empty node should be taken.

6. For each possible task instance we introduce a transition from run to error if:

   - An action A occurs, making the number of instances of A exceeding MNT_a
   - The executing task has overrun its deadline
   - A task pending for execution in the ready queue has exceeded its deadline

Figure 2 shows the result from transforming our example system shown. This is an ordinary timed automata for which decidability has been proven in [1].

For the general case, the scheduler and queue automata is illustrated in Figure 3 where q denotes a queue, r is the executing task, c measures how long time the executing task

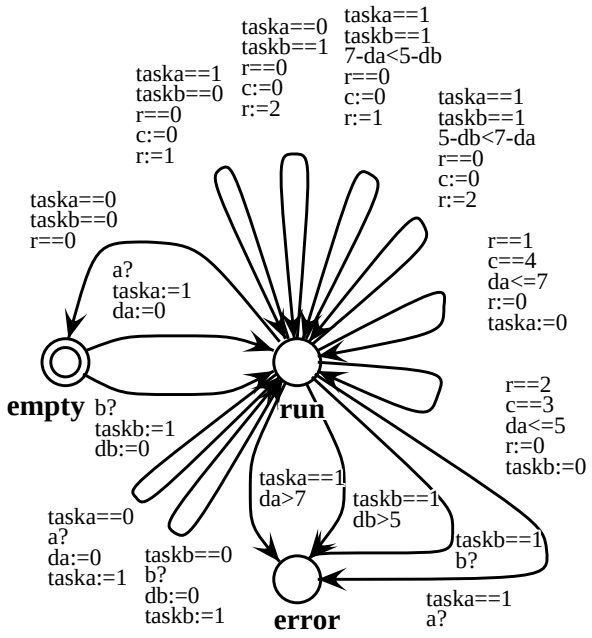


Figure 2. A model of the ready queue and the scheduler using ordinary timed automata

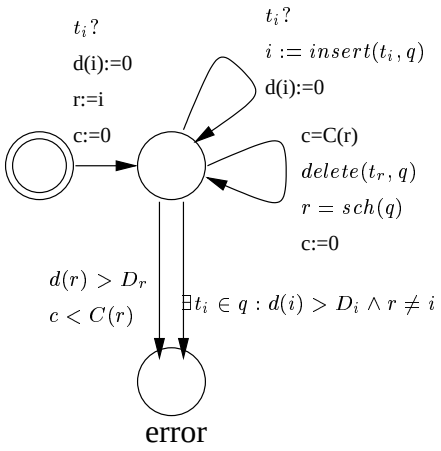


Figure 3. A general model of the scheduler using ordinary timed automata

has been running and d(i) is a vector keeping track of the time elapsed since the tasks entered the ready queue. C(i) is a vector holding the worst case execution time of all tasks. Both are vectors are finite as been discussed above. Moreover, the function sch() returns the instance among all tasks residing in the queue having least time left until its deadline. Task *i* is returned by sch() if the predicate $\bigwedge_{\forall m \in q: m \neq i} d(i) - d(m) \leq D_i - D_m$ is true, where D_i denotes the relative deadline specified for task *i*.

4. A Case Study with UPPAAL

UPPAAL is a model-checker for timed automata [9]. As shown in the previous section, the scheduler and ready queue can be modeled as an ordinary timed automaton. In this section, we present an example showing how to use UPPAAL for schedulability checking.

Our example system is a event-driven application controlling the speed of the shaft in a turning lathe. The objectives of the formal verification is to verify that the system is schedulable and the safety requirement that the engine is not turned on by the control task while the emergency stop is active. An event reports the current speed of the shaft and a control task is checking that the speed is within the speed limits (in our example speed=3). If the speed is to high (over 3), the engine is turned off and if the speed is to low (below 3), the engine is turned on. There is also an emergency stop function which is implemented in software. The setup is shown in Figure 4.

As shown in Figure 4, the parts belonging to the systems environment are the shaft having an optical sensor generating an event on every complete revolution, the emergency stop button having two states: up or down and the engine, being either on or off. Consequently, we have to model all these parts as a network of TATs. Moreover, we have two software tasks, the control task and the emergency stop

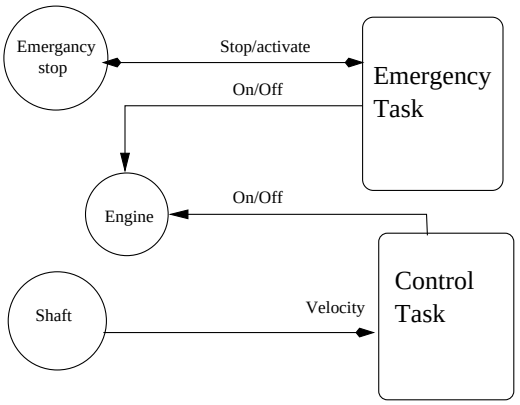


Figure 4. The setup for our example system

handler. These parts also have to be modeled in TATs belonging to the network constituting the complete system.

4.1. Modeling the system

We start by modeling the environment, i.e. the shaft, the emergency stop button and the engine. This can for instance be done as shown in Figure 5, 6.

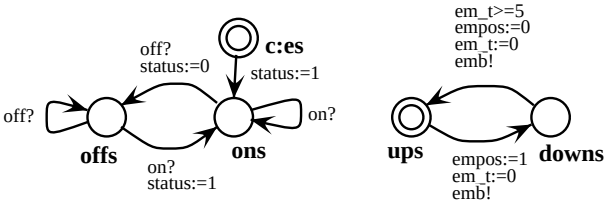


Figure 6. A model of the engine and the emergency stop button

If the engine is on, the shaft makes a complete revolution in between 4-8 time units, and an event is generated every time the optical sensor detects a complete revolution.

Next to model is the emergency stop handler and the control task. The control task has a calculated wcet of 2 tu and a hard deadline of 3 tu (Figure 8).

As for the control task, a deadline and a wcet must be specified for the emergency stop handler. According to our imagined requirement specification, it must respond within 2 tu, i.e. it has a deadline at 2 tu. The wcet estimation result in a wcet of 1 tu (see Figure 7). Furthermore, two subsequent activations/deactivation of the emergency stop can not be less than 5 tu in between. This gives us a minimum inter-arrival time for the emergency stop handler of 5 tu.

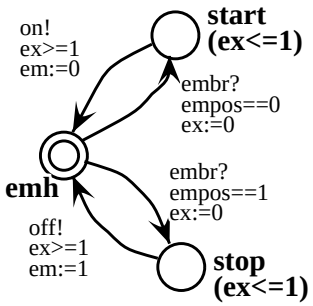


Figure 7. A model of the emergency handler in timed automata

The model of the scheduler is omitted in the paper. However, this process will be generated automatically by UPPAAL according to the algorithm given in Section 3.1 and will be invisible for the designer.

4.2. Verifying schedulability and safety

We use model checking and reachability analysis on our network of TAT for this purpose. UPPAAL uses a timed CTL language for specifying properties to verify. To verify that the system is schedulable, we must show that the

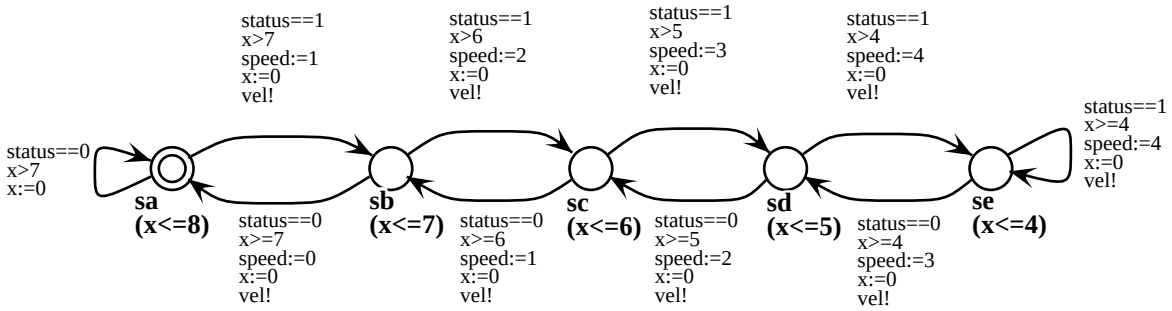


Figure 5. A model of the shaft in timed automata

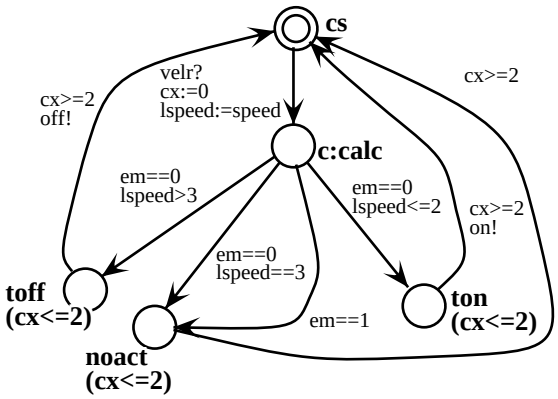


Figure 8. A model of the control task in timed automata

error state is never reachable. We will use the *always* predicate in our example as *always* not α is equivalent to *never*. This property is specified as shown in the formula below, *scheduler.error* means the state error in the process named scheduler:

$$\forall\Box not\ scheduler.error$$

For the safety property we need to verify that the system never reach a state where the control task is in position to turn the engine on while the emergency stop has been activated. For our model, such an expression looks like the formula given below:

$$\forall\Box not(control\_task.ton\ and\ em = 1)$$

First we will verify the schedulability property. As a result UPPAAL tells us that the property is not satisfied by giving a counter example. Consequently, the system is not schedulable. In order to obtain a schedulable system, the temporal constraints on the tasks have to be modified. The counter example given by UPPAAL, shows that the emergency handler task misses its deadline if this event happens just after the control task has been invoked. By changing the deadlines for the control task and the emergency stop handler to 4 tu, the system becomes schedulable. This is verified by the same property, but with an updated scheduler model. The model of the scheduler must be updated since now there can exist two instances of the control task and four instances of the emergency handler simultaneously in the ready queue.

Next to verify is our safety property, i.e. the control task should not be able to turn the engine on as long as the emergency stop is activated. In this case UPPAAL reports that the property is satisfied and consequently, the safety requirement is fulfilled.

It is of course possible to verify other functional properties. For instance, we can verify

that the shaft eventually will rotate with the set value. In our model, the set value is the speed of 3, i.e. *the speed is eventually equal to 3*. The corresponding formula given in UPPAAL logic is:

$$\exists \Diamond speed = 3$$

5. Conclusions

An important step in the development of embedded real-time systems is "schedulability analysis" that is to check whether all tasks in a system can be executed within the given deadlines in all possible scenarios. The traditional approach to schedulability analysis is often based on scheduling theory and a task model, which has been very successful for periodic tasks, but less successful for event-driven tasks.

In this paper, we have developed an extended version of timed automata with real-time tasks to provide a model for event-driven systems, which can be used for modeling, schedulability analysis, formal verification, and code generation. The main idea is to associate each discrete transition in a timed automaton with a task (an executable program e.g. written in C) with its worst case execution time. Intuitively, a discrete transition in an extended timed automaton denotes an event releasing a task and the guard on the transition specifies all the possible arriving times of the event (instead of the so–called minimal inter-arrival time). This yields a general model for hard real-time systems in which tasks are non-periodic. In this model, an automaton is used to model control structure of a systems and associated tasks are used to perform computation. Thus, code generation for such a model is reduced to transform the automaton into a runnable program with procedure-call. However, a critical problem is to guarantee that all the tasks associated with the automaton can be executed within their deadlines. This is the so-called schedulability checking problem. As the main result of this paper, we have shown that the schedulability checking problem for the extended timed automata with real time tasks can be transformed to a reachability problem for standard timed automata and thus it is decidable. This result allows us to apply model-checking tools for timed automata to schedulability analysis for event-driven systems. In addition, based on the same model of a system, we may use the tools to verify other properties (e.g. safety and functionality) of the system. This unifies schedulability analysis and formal verification in one framework.

As future work, we plan to extend the UPPAAL model checker for schedulability analysis. Future work also include code generation which is to translate extended timed automata with tasks into executable programs.

References

[1] R. Alur. Model-checking in dense real-time. *Information and computing*, 1993.

[2] R. Alur and D. Dill. Automata for modelling real-time systems. In *Proceedings of ICALP'90*, volume 443 of *Lecture Notes in Computer Science*. Springer, 1990.

[3] Bengtsson, Griffioen, Kristoffersen, Larsen, L. an d Pettersson, and Yi. Verification of an audio protocol with bus collision using uppaal. In *Proceedings of CAV'96*, volume 1102, 1996.

[4] J. Bengtsson, K. G. Larsen, F. Larsson, P. Pettersson, and W. Yi. UPPAAL in 1995. In *Proc. of the 2*nd *Workshop on Tools and Algorithms for the Construction and Analysis of Systems*, number 1055 in Lecture Notes in Computer Science, pages 431–434. Springer–Verlag, Mar. 1996.

[5] G. C. Buttazzo. *Hard Real-Time Computing Systems*. Kluwer Academic Publishers, 1997.

[6] C. Daws and S. Yovine. Two examples of verification of multirate timed automata with KRONOS. In *Proc. of the 16*th *IEEE Real-Time Systems Symposium*, pages 66–75, Dec. 1995.

[7] M. L. Dertouzos. Control robotics: The procedural control of physical processes. *Information Processing*, 1974.

[8] H. Hüttel and K. G. Larsen. The use of static constructs in a modal process logic. In *Logic at Botik'89*, number 363, pages 163–180. Springer–Verlag, 1989.

[9] K. G. Larsen, P. Pattersson, and Y. Wang. UPPAAL in a nutshell. *Springer International Journal of Software Tools for Technology Transfer*, 1, 1997.

[10] K. G. Larsen, P. Petterson, and Y. Wang. Compositional and symbolic model-checking of real-time systems. In *Proceedings of the 16th Real-Time Systems Symposium*, pages 76–87. IEEE Computer Society Press, 1995.

[11] M. Lindahl, P. Pettersson, and W. Yi. Formal design and analysis of a gear controller. *Lecture Notes in Computer Science*, 1384:281–297, 1998.

[12] C. Liu and J. Layland. Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment. *Journal of the Association for Computing Machinery*, 2, 1973.

[13] H. Lönn, P. Pettersson, and W. Yi. Formal Verification of a TDMA Protocol Start-Up Mechanism. In *Proceedings of 1997 IEEE Pacific Rim International Symposium on Fault-Tolerant Systems*, pages 235–242, 1997.

C

Verifying Temporal Constraints on Data in Multi-Rate Transactions using Timed Automata

By

Anders Wall, Kristian Sandström, Jukka Mäki-Turja, Christer Norström, and Wang Yi

Submitted to Real-Time Computing Systems and Applications 2000.

Verifying Temporal Constraints on Data in Multi-Rate Transactions using Timed Automata

Anders Wall[1], Kristian Sandström[1], Jukka Mäki-Turja[1], and Christer Norström[1]
Mälardalen University[1]
Mälardalen Real-Time research Center (MRTC)
P.O Box 883, S-721 23 Västerås, Sweden
{awl,ksm,jma,cen}@mdh.se

Wang Yi[1,2]
Uppsala University[2]
Department of Computer Systems
P.O Box 325, S-751 05 Uppsala, Sweden
yi@docs.uu.se

Abstract

Transactions involving multiple tasks, possibly with different period times, are common constructs used in the design of real-time systems. Data flowing through a transaction is usually subject to temporal constraints, such as maximum time from input to output or a maximum time difference between inputs. Such constraints are of great importance to guarantee the correct functioning of the designed system. But normally they cannot be checked using the traditional approach to schedulability analysis. In this paper we describe how to use timed automata and reachability analysis to verify such temporal constraints on data in transactions. By making a timed automaton model of the data dependencies in a transaction, we enable automatic verification of timing constraints such as end-to-end latency. The model can handle different computational models and any non-preemptive execution order of the tasks in the transaction. Our experiences from industrial case studies indicate that in a substantial number of applications, the transactions are of sizes that can be handled using this approach.

1 Introduction

Designing safety-critical real-time systems involves assessment of functionality, temporal requirements and dependability. The temporal requirements on such systems may come in many forms, examples are end-to-end deadlines, jitter constraints, and latency constraints. Such constraints can for example be found in multi-rate control systems, where sampling, control, and actuation may execute with different frequency. In such a system there are requirements on the delay from sampling to actuation in the feedback loop. More precisely, for a particular actuation one wants to know which sample the calculation is based on. The feedback loop delay is then defined as the time difference between actuation and sampling. To be able to fulfill such constraints, the designer has to have some means to express and preferably also verify them. However, computational models like fixed priority scheduling [1,2], and pre-run-time scheduling [3], used in the industry, cannot directly express such constraints. For instance, jitter constraints are handled by manual transformations into release times and deadlines of individual tasks. Translating these constraints to the attributes of the computational model is non-trivial and the schedulability analysis does not verify that the translation itself is correct.

In this paper, we will show how to verify that this mapping is correct. We do this by presenting an algorithm that transforms the data flow and available timing information of an application, or part of an application, into timed automata. In addition, we construct an automaton modeling the execution strategy that defines the execution orders of the involved tasks. By composing these two automatons and by using model checking, we can verify timing constraints such as latency. The benefit of this is two folded. First, the result from the scheduler can be checked, and second the high-level requirements from the specification can be verified. We also believe that this is the starting point for integrating real-time scheduling and timed automata to enable efficient design and verification techniques of both time-triggered and event-triggered systems in one framework.

Several researchers including Mok, Gerber, and Kim have provided specific computational models that directly allow specification of latency constraints [4, 5, 6]. Our approach makes few assumptions about the computational model, and can therefore be applied to different computational models. Furthermore, it also gives the possibility to model the functional behavior of tasks and to efficiently integrate handling of event-triggered tasks by defining an environment model, as reported in [7].

The paper is structured as follows: In section 2 we define transactions, data dependency, and execution strategies. Section 3 describes the construction of a timed automaton that models data dependencies and how to verify temporal constraints on data. Finally, in Section 4 we present our conclusions.

2 Transactions, Data Dependency Model and Execution strategies

A transaction is a set of tasks, collaborating in order to provide some desired function. For instance, consider a control transaction consisting of three tasks, *sample*, *control*, and *actuate*. The task *sample* reads an input value from the process, performs some filtering, and thereafter sends the value to the *control* task. The control task consumes the sample value, reads a reference value, and calculates a new control signal. Finally, the actuator task consumes the new control signal and imposes it on the controlled process. In our model, each task in a transaction has an input – calculate – output behavior. That is, when the task starts its execution it first consumes all its input data, performs the computations, and before completion, it outputs the results.

The execution of tasks is considered to be non-preemptive. Apart from being non-preemptive, tasks may execute according to any strategy, e.g., time driven or event driven. Furthermore, transactions can be of multi-rate nature, i.e., tasks in the transaction may execute with different rates. In Figure 1, a transaction consisting of four tasks is displayed, where 1, 2, and 3 are input to the transaction and the arrows describe the data flow through the transaction.

We will represent a transaction as a data dependency graph. A data dependency graph is a set of nodes, $n_0,\ldots,n_p$, that represents the inputs and the tasks in the transaction, and a set of edges that represents the data flow in the transaction. The initial nodes of the dependency graph model the inputs to a transaction. If a task τ_q consumes several

different data from task τ_p, only one edge between those nodes is needed. Moreover, if a task reads several inputs, there is need for only one initial node representing those inputs. This is not a restriction but a consequence of the input – calculate – output behavior of tasks, in which a task reads all its inputs at the beginning of its execution.

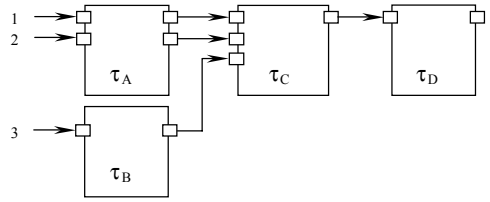


Figure 1. An example transaction.

Figure 2 illustrates the data dependency graph of the transaction in Figure 1, where task τ_C depends on data produced by both τ_A and τ_B. Moreover, τ_D depends on τ_C. Note that τ_A consumes data from input 1 and 2, which is represented in the dependency graph as a single initial node n_0.

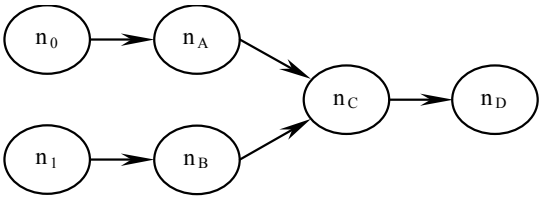


Figure 2. The data dependency graph for the example transaction.

Definition 1. A data dependency graph for a transaction is a Directed Acyclic Graph (DAG) defined by a tuple $\langle N, E, N_0, n_{end} \rangle$, where

- N is a finite set of nodes representing tasks in the transaction.
- $E \subseteq N \times N$ denotes the edges between nodes.
- $N_0 \subseteq N$ is a finite set of initial nodes denoting the inputs.
- $n_{end} \in N$ represents the last task in the transaction.

□

Note, that for each initial node $n_0 \in N_0$ there exists only a single edge e_0 to another node. If several tasks in a transaction read the same input, two or more initial nodes in the data dependency graph can be used to model that input. Each node $n_p \in N$ has an execution time specified as an interval $C(n_p)=[bcet, wcet]$, where *bcet* is the best-case execution time and *wcet* is the worst-case execution time for task τ_p.

Furthermore, as tasks in the transaction may execute in any arbitrary order, the dependencies do not imply a precedence relation between tasks. The execution order depends upon the execution strategy, e.g., event triggered tasks with fixed priorities or time triggered pre-run-time scheduled tasks. Formally, an execution order is defined as follows:

Definition 2. An execution order σ is a sequence of pairs $\langle t, s \rangle$ where $t \in$ denotes the start time and $s \in N^+$ is a sequence of one or more tasks, thus $\sigma \in (\mathbb{N} \times N^+)^*$.

□

The start time for the first task in the sequence s is equal to t, whereas the remaining tasks in s start as soon as the preceding tasks complete. Thus, the start time of a task, that is not the first task in s, is determined by the start time of the preceding task τ_p and the execution time interval ranging from *bcet* to *wcet*. An example of an execution order involving four tasks is $\{\langle 0, \tau_A \cdot \tau_B \cdot \tau_C \rangle, \langle 12, \tau_A \cdot \tau_D \rangle\}$.

From the data dependency graph the dependencies for each task in the transaction can be derived. The data dependencies for task τ_p are represented as the set $L(n_p)$ of independent paths from the set of initial nodes to node n_p. Formally a data dependency is defined as follows:

Definition 3. A data dependency relation is

- $L(n_0) = \varnothing$ where $n_0 \in N_0$
- $L(n_q) = \bigcup_{(p,q) \in E} \{\mu \cdot p \mid \mu \in L(n_p)\}$

□

Note that μ denotes a path from $n_0 \in N_0$ to node n_p. For instance, the dependency set $L(n_C)$ for node n_C in the data dependency graph depicted in Figure 2 is given as $L(n_C) = \{n_0 \cdot n_A, n_0 \cdot n_B\}$.

We will denote the set of all the data dependencies for the tasks in a transaction as L, which is a union of all data dependency sets.

3 Verifying temporal constraints using timed automaton

Timed automata has been recognized as a basic semantic model for specifying and verifying timing constraints for real-time systems. Here we give a brief introduction to the model of timed automata. For details, we refer to [9].

A timed automaton is a standard finite-state automaton extended with a finite set of real-valued clocks. On each transition there are constraints (guards) on clocks, synchronization action, and clocks to be reset. Whenever the guard is satisfied of the current values of the clocks, the transition can be taken, i.e., the synchronization action is performed and the clocks to be reset are set to 0. A state of a timed automaton can be considered as a tuple containing the current node of the finite automaton, and the current values of the clocks. Informally, the semantics of a timed automaton is given by two transition rules. First of all, it can stay in the current node letting time pass (delay), i.e. the clocks are updated and the current node remains unchanged. Secondly, it can take the transition instantaneous resulting in a state with a new node. In recent years, there have been a number of software tools developed e.g. KRONOS and UPPAAL [8,10] for automated analysis of logical properties of timed automata.

In this paper, we are aiming at using the existing tools to verify timing constraints on transactions by transforming the data dependency model to timed automata. We will

refer to a timed automaton that describes data dependencies as a *data dependency automaton*.

The temporal constraints that can be verified using the approach proposed in this paper are:

- End-to-End timing constraint, i.e., minimum and maximum time from readings of inputs until the end of the transaction.
- Variation in End-to-End timing, i.e. output jitter.
- Input synchronization, i.e., minimum and maximum time difference between input readings used by the transaction in order to produce a result.

3.1 From Data Dependency Graphs to Timed Automata: an Example

In this subsection, we use an example to show the main idea and intuition of the translation algorithm. The transaction for the example, illustrated in Figure 3, consists of the input k and the three tasks τ_A, τ_B, and τ_C. The task executes according to a non-preemptive time-triggered strategy. We want to verify that the data that τ_C uses to produce the result for the transaction does not origin from an input reading that is older than 10 time units.

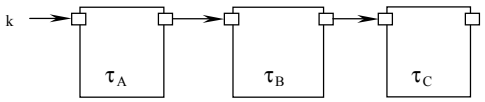


Figure 3. The example transaction with its three tasks.

The data dependency graph representing the transaction in Figure 3 will consist of four nodes and is displayed in Figure 4. The node n_{end} represents the last task τ_C.

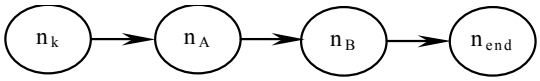


Figure 4. The data dependency graph.

According to definition 3, all tasks in the transaction depend on input data k. The data dependency relation sets for the nodes are $L(n_k) = \{\ \}$, $L(n_A) = \{n_k\}$, $L(n_B) = \{n_k \cdot n_A\}$, $L(n_{end}) = \{n_k \cdot n_A \cdot n_B\}$.

In the data dependency automaton we measure the age of data when an input data instance has been processed by the last task in the transaction (τ_{end}). Therefore we use time stamps to measure the time elapsed since a particular data entered the transaction. We denote a time stamp for inputs represented by the initial node n_k in the dependency graph as X^k. As the transaction might be of a multi-rate nature, more than one instance of an input could exist simultaneously in the transaction, all with different age. Consequently, for all initial nodes in the dependency graph, there must be one or more associated time stamp instances. The actual number of time stamp instances for an initial node is correlated to the number of paths from that initial node to all other reachable nodes in the dependency graph, excluding n_{end}. For instance, the number of time stamp instances needed for X^k in this example is two, since there is one path from n_k to n_A and one additional path from n_k to n_B (see $L(n_A)$ and $L(n_B)$).

We use clocks, which can only be reset, to implement time stamp instances in timed automata. Therefore, when a task that reads an input executes, the corresponding time stamp instance is reset. When consumers of data produced by that task execute, the time stamp instance is distributed. Since several time stamp instances may be needed for an input, nodes in the data dependency graph may use any of these instances. To ensure that the time stamps are consistent, a state in the data dependency automaton is used to keep track of the time stamp instances currently used by each node, i.e., one state in the data dependency automaton models the assignments of time stamp instances for all nodes. This gives for all nodes the age of all data that the tasks, represented by the nodes, have read at this point. Table 1 presents the assignment of time stamp instances for state *S0* to *S3* in the data dependency automaton displayed in Figure 5.

node	n_A	n_B
S0	X_1^k	X_1^k
S1	X_2^k	X_1^k
S2	X_2^k	X_2^k
S3	X_1^k	X_2^k

Table 1. Time stamp instance assignments for state *S0* to *S3* in the dependency automaton.

Since several nodes simultaneously can use the same time stamp instance, a time stamp instance cannot be reset without considering possible multiple uses. Assume that a task τ_p executes, and as a consequence a time stamp for an input should be reset. If the time stamp instance used by node n_p is used by at least one other node, then node n_p will have to use a time stamp instance that is not assigned to any node. Consequently, a transition has to be made to a state in which node n_p is assigned the new time stamp instance. As an example of such an instance replacement consider the transition from *S0* to *S1* in Figure 5 where the time stamp instance for node n_A changes from X_1^k to X_2^k. If, on the other hand, no other node in the data dependency graph uses X_i^k, it can be reused. The transition *S1* to *S1* in Figure 5 is an example of reusing a time stamp instance. In this case X_2^k is reused in order to reflect the most recent reading of input *k*.

In addition to the states that are needed to represent the use of time stamp instances, there must exist states that represent that the transaction is completed, i.e., corresponding to the last task, τ_{end}. We will refer to such a state as an *end-state*. Upon a transition from a state *S* to an end state a time stamp is reset, thereby making it possible to verify the end-to-end age constraint on the data. Note that this transition does not affect the assignment of time stamp instances, thus there is always a transition back to state *S* in the automaton.

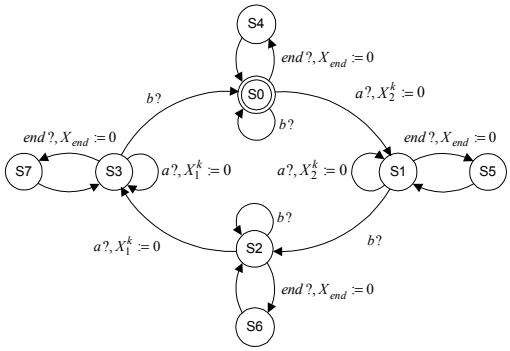


Figure 5. The data dependency automaton for the example.

In order to verify temporal constraints for a transaction, an automaton describing the execution strategy is needed. The execution strategy of tasks can be time triggered or event triggered. For time triggered systems the translation of the execution order to a timed automaton is straightforward. If the system is event triggered, the system environment that generates the events must be modeled as well [7]. For the purpose of illustration, a simple execution scenario for the tasks in the example transaction is depicted in Figure 6. All tasks have a *bcet* equal to 1 time unit and a *wcet* equal to 2 time units. The start time for the two instances of τ_A and the second instance of τ_B is fixed, whereas the start times for the rest of the task instances are relative to the preceding task. The complete execution sequence is repeated as soon as the second instance of τ_C has completed.

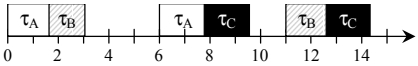


Figure 6. A possible execution scenario in the three-task transaction.

The execution scenario depicted in Figure 6 results in the automaton illustrated in Figure 7.

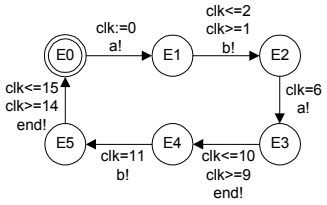


Figure 7. The execution order automaton.

The execution order automaton in Figure 7 give rise to the following state transitions in the data dependency automaton in Figure 5:

$$S0 \xrightarrow{a?, X_2 := 0} S1 \xrightarrow{b?} S2 \xrightarrow{a?, X_2 := 0} S3 \xrightarrow{end?, X_c := 0} S7 \longrightarrow S3 \xrightarrow{b?} S0 \xrightarrow{end?, X_c := 0} S4 \longrightarrow S0$$

The transaction is completed when the data dependency automaton reaches the end-states S7 and S4. Consequently, the age constraint is satisfied if, when in those states, a related reading of input *k* occurred no longer than ten time units earlier. By related reading we mean a reading of data that has propagated through the entire transaction. Thus, in order to verify the age constraint the following invariant must hold.

$$\forall \Box \left((S4 \rightarrow (X_{end} \geq X_1^k - 10)) \wedge (S5 \rightarrow (X_{end} \geq X_1^k - 10)) \wedge (S6 \rightarrow (X_{end} \geq X_2^k - 10)) \wedge (S7 \rightarrow (X_{end} \geq X_2^k - 10)) \right)$$

That is, in all possible execution scenario, restricted by the execution automaton, the time difference between an instance of the time stamp corresponding to the reading of input k (x_1^k or x_2^k), and the time stamp corresponding to the completion of τ_C, is never greater than 10 time units.

1.2 From Data Dependency Graphs to Timed Automata: the Translation Algorithm

In this section we present how to construct a timed automaton that represents the data dependencies for a given transaction. The translation considers the data dependencies in the transaction, represented by the L set. Transactions are assumed to have a N-to-1 topology. That is, there can be multiple tasks that may read inputs, but only one task that produces outputs. A transaction with multiple tasks producing outputs can be represented as several different transactions, each with a single output-task. Furthermore, the execution of tasks is considered to be non-preemptive.

As discussed in Section 3.1, several instances of a time stamp may be needed in order to keep track of the age of data flowing through a multi-rate transaction. We denote a particular instance i of time stamp X^k as x_i^k . The number of time stamp instances num_k needed to measure the age of input data represented by an initial node n_k is finite and equal to the number of nodes depending on n_k, i.e., the size of $\{n_k \cdot \sigma | n_k \cdot \sigma \in L\}$where $n_k \in N_0$. As the tasks in the transaction can execute in arbitrary order, they can in particular execute in a manner that gives each task a unique age of the input data it depends on. If the number of time stamps needed for an initial node n_k is num_k, the instances will be enumerated from 1 to num_k.

The age of an input k, that a task depends on, is represented as a pair consisting of the time stamp for k, and the path from the initial node in the data dependency graph that represents k to the node that represents the task. This path uniquely identifies a dependency and distinguishes different dependencies of the same data. For each task τ_p in a transaction, the set $A(n_p)$ contains all such pairs for the data that τ_p depends on.

Definition 4. The set $A(n_p)$ contains pairs $\langle X^k, s \rangle$ where X^k is the time stamp of input, represented by the initial node n_k, and s is a path from n_k to the node n_p, representing task τ_p.

$$A(n_p) = \bigcup_{n_k \cdot \mu \in L(n_p)} \left\langle X^k, n_k \cdot \mu \right\rangle$$

□

Note that for each pair $\langle X^k, s \rangle$ in $A(n_p)$, the path s is static whereas the particular instance i of time stamp X^k may vary between states in the data dependency automaton. We will use A to denote the set containing the sets $A(n_p)$ for all tasks in the transaction excluding the last task τ_{end}.

A state in the data dependency automaton represents a unique time stamp instance assignment for the set A. Moreover, the number of states in the data dependency automaton is finite. Thus, the problem of verifying temporal constraints using reachability analysis is decidable.

Proposition 1. The number of states in the data dependency automaton is finite and given by:

$2 * \prod_{n_p \in N} \prod_{\langle X^k, s \rangle \in A(n_p)} num_k$ where $n_p \neq n_{end}$.

PROOF: Since every time stamp X^k in $A(n_p)$ can be one of num_k instances, there are $\prod_{\langle X^k, s \rangle \in A(n_p)} num_k$ ways of constructing $A(n_p)$.

The total number of possible states for the data dependency graph excluding n_{end} is then given as all possible ways of combining the time stamp instances for all nodes. Consequently, the total number of combinations is given by:

$$\prod_{n_p \in N} \prod_{\langle X^k, s \rangle \in A(n_p)} num_k$$

Since from every state there must be a transition to a unique end state, the total number of states in the data dependency automaton is $2 * \prod_{n_p \in N} \prod_{\langle X^k, s \rangle \in A(n_p)} num_k$

□

We will now present the rules for constructing the time automaton representing the data dependency graph for a given transaction that complies with the assumptions given earlier in this section. The automaton is constructed starting from an initial state *S* and the rules *G1* to *G5* decides what action to take and how the states changes when a task τ_p executes. Two basic rules R_1 and R_2 constitute the basis for *G1* to *G4*, whereas *G5* corresponds to completion of the transaction. If node n_p is an immediate successor to the initial node n_k and if node n_p uses the time stamp instance X_j^k to represent the age of data read by task τ_p, then R_1 is satisfied if X_j^k is not used by any other node. Moreover, R_1 is also satisfied if node n_p does not depend upon an initial node. The second rule R_2 is satisfied if every immediate predecessor to node n_p in the data dependency graph uses the same time stamp instance as n_p itself. R_2 is also satisfied if n_p has no dependencies to other nodes.

$$R_1: \langle X_j^k, n_k \rangle \in A(n_p) \rightarrow \langle X_j^k, n_k \cdot \mu \rangle \notin A(n_q) \vee n_q = n_p$$

$$R_2: \langle X_j^k, n_k \cdot \mu \cdot n_q \rangle \in A(n_p) \rightarrow \langle X_j^k, n_k \cdot \mu \rangle \in A(n_q)$$

In Figure 8, the transitions corresponding to rules *G1* to *G5* are displayed. For each node in the data dependency graph, one out of four possible transitions, *G1* to *G4*, should be present in every state of the resulting timed automaton. The rules *G2* to *G4* result in a change of the time stamp instance assignment in the set *A* and therefore a transition must be made to another state in the timed automaton that represents the assignment of time stamp instances. *G1* on the other hand, does not change time stamp instance assignments, and consequently, a transition to a new state in the timed automaton is superfluous. Finally, *G5* corresponds to completion of the transaction by reaching an *end-state*. The rules should be applied in all states for all nodes in the dependency graph.

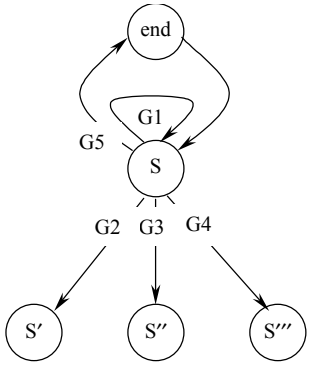


Figure 8. The rules for constructing the data dependency automaton.

The rules are described by a condition consisting of a composition of the basic rules R1 and R2, an action describing the transition taken in the timed automata and which time stamp instance, if any, that should be reset. Furthermore, the initial state S is formed as:

$S = A$ where $\left\langle X_j^k, s \right\rangle \in A(n_p) \Rightarrow j = 1$

That is, initially tasks that depend on the same input data uses the same time stamp instance for that data.

Transition G1. The time stamp instance for the input (if any) that the task reads is not used by any other task, and there is no updated time stamp for the data that the task consumes (if any). The time stamp for the input is updated.

Condition: $R_1 \wedge R_2$

Action:

$S \xrightarrow{p?, X_j^k := 0} S'' \; iff \left\langle X_j^k, n_k \right\rangle \in A(n_p)$

$S \xrightarrow{p?} S'' \; iff \left\langle X_j^k, n_k \right\rangle \notin A(n_p)$

Transition G2. The time stamp instance for the input that the task reads is used by at least one other task. If the task consumes data from other tasks, there is no updated time stamp, i.e., the tasks use the same time stamp. As a consequence, the state S changes to S'.

Condition: $\neg R_1 \wedge R_2$

Action: $S \xrightarrow{p?, X_j^k := 0} S' \; where \left\langle X_j^k, n_k \cdot \mu \right\rangle \notin A(n_q) \; \text{for all} \; n_q \in N$

Transition G3. The time stamp instance for the input (if any) that the task reads is not used by any other task, but there are one or more updated time stamps for the data that the task consumes. The state S changes to S'' as one or more time stamp instances has to be changed considering the data dependency.

Condition: $R_1 \wedge \neg R_2$

Action:

$S \xrightarrow{p?, X_j^k := 0} S'' \; iff \left\langle X_j^k, n_k \right\rangle \in A(n_p)$

$S \xrightarrow{p?} S'' \; iff \left\langle X_j^k, n_k \right\rangle \notin A(n_p)$

Transition G4. The time stamp instance for the input that the task reads is used by at least one other task, and there are one or more updated time stamps for the data that the task consumes. The new state S''' reflects the fact that we need both a unique time stamp instance for the input and that one or more time stamp instances have to be changed considering the data dependency.
Condition: $\neg R_1 \wedge \neg R_2$
Action: $S \xrightarrow{p?, X_j^k := 0} S''' \; where \left\langle X_j^k, n_k \cdot \mu \right\rangle \notin A(n_q) \; \text{for all} \; n_q \in N$

Transition G5. The last task in the transaction executes and completes. The time stamp X_{end} is reset on the completion of τ_{end}.
Condition: τ_{end} completes.
Action: $S \xrightarrow{end?, X_{end} := 0} end \longrightarrow S \;\; where \; end = n_{end}$

1.3 From Data Dependency Graphs to Timed Automata: The example revisited

As an example on how to apply the rules *G1-G5* in order to construct a data dependency automaton, reconsider the transaction of the example in Section 3.1. The data dependency graph for that transaction is equal to the graph in Figure 9.

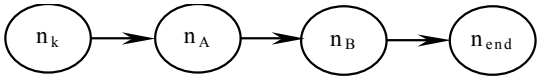


Figure 9. The data dependency graph.

The L sets for the nodes are $L(n_k) = \{\ \}$, $L(n_A) = \{n_k\}$, $L(n_B) = \{n_k \cdot n_A\}$, $L(n_{end}) = \{n_k \cdot n_A \cdot n_B\}$. Using Definition 4 gives the A sets $A(n_A) = \{\langle X^k, n_k \rangle\}$ and $A(n_B) = \{\langle X^k, n_k \cdot n_A \rangle\}$.

For the initial node S the sets $A(n_p)$ is assigned time stamp instances according to $S = A$ where $\left\langle X_j^k, s \right\rangle \in A(n_p) \Rightarrow j = 1$, which gives $A(n_A) = \{\langle X_1^k, n_k \rangle\}$ and $A(n_B) = \{\langle X_1^k, n_k \cdot n_A \rangle\}$.

Now we will explore the rules for constructing a data dependency automaton by deducing the transitions from the initial state S.

Transition G1 ($R_1 \wedge R_2$)
Starting with node n_A, rule R_1 is not satisfied since $\langle X_1^k, n_k \cdot n_A \rangle \in A(n_B)$ and thereby no transition is taken. For node n_B, both rule R_1 and R_2 is satisfied and therefore the transition $S \xrightarrow{B?} S$ is taken. Since $\langle X_1^k, n_k \rangle$ is not in $A(n_B)$ no time stamp instance is reset.

Transition G2 ($\neg R_1 \wedge R_2$)
For node n_A, $\neg R_1$ is satisfied as well as rule R_2, and thereby the transition $S \xrightarrow{A?, X_2^k := 0} S'$ is taken, where $\langle X_2^k, n_k \cdot \mu \rangle \notin A(n_q)$ for all $n_q \in N$. For state S' $A(n_A) = \{\langle X_2^k, n_k \rangle\}$ whereas $A(n_B) = \{\langle X_1^k, n_k \cdot n_A \rangle\}$ remains unchanged. For node n_B, $\neg R_1$ is not satisfied and therefore no transition is taken.

Since neither *G3* ($R_1 \wedge \neg R_2$) nor *G4* ($\neg R_1 \wedge \neg R_2$) is satisfied for any of the nodes, there are no transitions for these cases.

Transition G5

Finally, according to *G5*, the transition $S \xrightarrow{end?, X_{end}:=0} end \longrightarrow S$ is added to the automaton.

The part of the data dependency automaton constructed so far is displayed in Figure 10. Repeating the procedure above for state *S'* will eventually complete the automaton, resulting in the automaton of Figure 5.

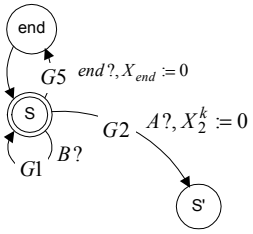


Figure 10. The partial data dependency automaton.

4 Conclusions

A temporal constraint can for instance be the time from input to output or the time difference between several inputs to a transaction. Such constraints of a transaction are not always possible to express in the task models at hand. Thus, the designer has to map such a constraint manually onto the temporal attributes in the existing task model, e.g. period times, deadlines, offsets, etc. The schedulability analysis only verifies whether the mapped system description can be realized or not. It does not verify that the requirement mapping itself is correct. In this paper we have described how to use timed automata to verify temporal constraints on data in transactions. By constructing a timed automaton model of the data dependencies in a transaction, we enable verification of, for instance, end-to-end constraints using model-checking.

As the model is general, it can handle arbitrary computational models and execution orders of the task in the transaction. The main contribution of the paper is the rules for automatically generating such a model in timed automata. Although the size of the constructed automaton grows, in the worst case, exponentially with the number of tasks in a transaction, we believe that the method is suitable and applicable to real-world applications. The method can be applied to one transaction in isolation, i.e., modeling and verification can be performed on one transaction at a time. Consequently, only the transactions of interest in the complete system have to be verified. Our experiences from industrial case studies indicate that, in a substantial number of applications, the majority of the transactions are of size feasible for this method.

As future work we will extend the method to also include preemptive execution strategies and we will implement a tool that, from a given system description, generates data dependency automata and timed automata modeling the tasks' execution. The model of the tasks' execution will be derived from an existing scheduler, and for model-checking we will use the existing model-checkers, e.g. UPPAAL [10]. Furthermore, we plan to investigate the possibility of making timed automata models of the functional behavior of tasks at some appropriate level of

abstraction. Such models enable verification of, not only temporal correctness, but also functional properties such as safety and functional correctness.

5 References

[1] Liu C. L. and Layland J. W. Scheduling Algorithms for Multiprogramming in a Hard Real-Time Environment. Journal of ACM 20(1), 1973.

[2] Audsley N. C., Burns A., Davis R. I., Tindell K., and Wellings A. J. Fixed Priority Pre-emptive Scheduling: An Historical Perspective. Real-Time Systems 8(2-3): 173-198, 1995.

[3] Xu Jand Parnas D. L. Scheduling Processes with Release Times, Deadlines, Precedence and Exclusion Relations. IEEE Transaction on Software Engineering, Vol. 16 No. 3, March 1990.

[4] Mok A. K., Tsou D., and De Rooij R. C. M., The MSP.RTL Real-Time Scheduler Synthesis Tool, In proceedings of 17th IEEE Real-Time Systems Symposium, pp. 118-128, 1996

[5] Gerber R., Hong S., and Saksena M. Guaranteeing Real-Time Requirements with Resourse-Based Calibration of periodic Processes. IEEE Transactions on Software Engineering, 21(7), 1995.

[6] Kim N. A Scheduling Technique for Real-Time Systems with End-to-End Timing Constraints, In proceedings of RTCSA, 1996.

[7] Norström C., Wall A., and Yi W., Timed Automata as Task Models for Event-Driven Systems, In proceeding of RTCSA, 1999.

[8] Daws C. and Yovine S., Two examples of verification of multirate timed automata with KRONOS, In proceedings of 16th IEEE Real-Time Systems Symposium, PP 66-77, 1995

[9] Alur R. and Dill D. A theory of timed automata, Theoretical Computer Science vol. 126 pp. 183-235, 1994

[10] Larsen K. G., Pettersson P. and Yi W., UPPAAL in a Nutshell, In Springer International Journal of Software Tools for Technology Transfer 1(1+2), 1997.