Real-Time Statechart Semantics

Holger Giese\textsuperscript{1} and Sven Burmester\textsuperscript{2}

Software Engineering Group
University of Paderborn
Warburger Str. 100, D-33098 Paderborn
Germany
[hg,burmi]@uni-paderborn.de

2nd June 2003

Abstract Modelling complex dynamic real-time behavior is a crucial prerequisite when employing object-oriented modelling techniques successfully in the field of complex real-time systems. In this report we therefore review the state of the art of two related lines of research, namely extensions for expressiveness such as Statecharts as proposed in the field of software engineering to support the description of complex behavior and extensions to specify the explicit temporal behavior for state-based behavior such as Timed Automata. We review the Hierarchical Timed Automata model which provides already the most fundamental concepts for structure and time as well as a reasonable semantics for the domain of temporal behavior. To enable the modeling of complex control behavior which incorporates most of the fundamental extensions for expressiveness and time in a manner appropriate for real-time systems, we then propose an extension in form of Extended Hierarchical Timed Automata and Real-Time Statecharts. Different restricted forms reflecting the different requirements of system elements such as physical entities, communication channel abstractions and controllers running on particular platforms are further identified.

Keywords: Real-Time Statecharts, Hierarchical Timed Automata, Timed Automata, real-time systems, distributed systems, behavior specification.
## Contents

1 Introduction 3

2 Requirements 3
   2.1 Real-Time Systems 3
   2.2 Expressiveness 4
   2.3 Temporal Behavior 5
   2.4 Predictability 6
      2.4.1 Feasibility Analysis 6
      2.4.2 Formal Verification 6
   2.5 Realizability 7

3 Timed Automata 8
   3.1 Timed Automata 9
   3.2 Consistency 10
   3.3 Extensions 11
   3.4 Summary 12

4 Hierarchical Timed Automata 12
   4.1 Syntax 12
   4.2 Semantics 14
   4.3 Summary 17

5 Extended Hierarchical Timed Automata 17
   5.1 Extensions 17
   5.2 Syntax 18
   5.3 Semantics 20
   5.4 Summary 24

6 Real-Time Statecharts 24
   6.1 Extension 24
   6.2 Semantics 25
   6.3 Realizability 27
   6.4 Summary 27

7 Related Work 28

8 Conclusion 29
1 Introduction

To model complex reactive real-time systems, state machine specification techniques are one successful approach. However, they have to face a number of problems. On the one hand, a number of extensions for the expressiveness of state machines such as Statecharts [Har87] have been proposed in the field of software engineering to support the description of complex behavior. On the other hand, to specify the required temporal behavior for state-based behavior well-founded formal concepts such as Timed Automata have to be employed. While both research directions have their own merits, a combination of both worlds has to face serious problems. The Statechart extensions for structure and micro- and macro-steps which achieve the required degree of expressiveness assume an infinite fast machine and therefore assumes that the environment will not interfere with its execution. In contrast to this approach assumes the Timed Automata models and related research that temporal properties of the automaton are explicitly modeled to study whether the temporal behavior is appropriate. To obtain an applicable solution for developing real-time control software, it is further required that the specification formalism can be analyzed to proof the required properties (predictability) and used to derive a correct operating software (realizability). We will show that current proposals are not able to cover all four requirements and therefore present a solution that takes expressiveness, temporal behavior, realizability and predictability into account.

First, the basic characteristics of real-time systems and the resulting main requirements for an applicable description technique are identified in Section 2. Then, we review the underlying notion of Timed Automata and its foundations in Section 3. This includes the basic principles of state machines and its extension to the notion of Timed Automata. Extensions with the concepts of Statecharts for expressiveness as proposed with the notion of Hierarchical Timed Automata are then reviewed and defined in Section 4. In Section 5 (ExHTA) the syntax and semantics for the proposed extension in form of the extended hierarchical timed automata model which takes into account all four aspects and addresses the most of the crucial issues identified in Section 2 is described. The report closes with some final conclusion in Section 8.

2 Requirements

To identify the relevant requirements for a temporal state-based formalism that serves expressiveness as well as temporal aspects, we start with reviewing what are the main characteristics and obstacles to modelling of real-time systems in Section 2.1. In Section 2.2 we further review which main structural modelling extensions proposed in the software engineering context are appropriate in the specific domain of real-time systems. Besides reactive behavior and complex state spaces of course the temporal behavior is another prerequisite discussed in Section 2.3. Then the specific question of predictability is addressed in Section 2.4 looking on the analysis opportunities for different timed models. Efforts for the higher-level specification of a system are only reasonable, when the realizability of the concepts as reflected in Section 2.5 is also taken into account.

2.1 Real-Time Systems

Activities within a real-time system are described by a number of independently running tasks that interact with the environment. Tasks can be classified as hard real-time, soft real-time or non real-time constrained relating to a so called utility function. For a hard real-time task the value (utility) of a required result or effect will be zero after its deadline has been passed. For a soft real-time task the value decreases after passing the deadline and for non real-time tasks no deadline effecting the value of the task is given at all.
In general, a real-time system is composed of 3 layers: The hardware, the real-time operating system (RTOS) and the software. The hardware consists of one or multiple controllers and optional resources like memory, sensors or actuators. The RTOS is the interface between hardware and software and provides operations for accessing the resources. In every operating system, all concurrently running tasks try to access the controllers and the resources simultaneously. As every controller can handle only one task, the operating system provides a task manager to regulate this access. RTOS provide a so-called scheduler as task manager, trying to meet the deadlines of the regulated tasks. When accessing shared resources, concurrent access has to be controlled, e.g. in order not to read from and write to a shared resource at the same time.

Therefore, for modeling hard real-time systems it is crucial to have appropriate means to express temporal constraints. While this can be done rather straightforward at the theoretical levels each concrete device can only execute with limited speed and therefore besides specifying what time properties are expected we still have to ensure that the resulting tasks do also meet their deadlines. For hard real-time systems therefore predictability is required. Usually this is only possible when for alternatives simply the worst-case is considered. E.g., usually all tasks are viewed on only by taking their worst-case execution time (WCET) into account.

The reactive nature of most hard real-time systems further make periodic operating tasks the underlying paradigm for real-time systems. Therefore, often multiple independent tasks scheduled either by a real-time operating system (RTOS) or run-time system are employed. The scheduling can be realized using pre-computed static schedules, e.g. controlled by a priority scheduler. Using this scheduler, a priority is assigned to every task determining the order of execution. Note, that even for a single threaded system with explicit ordering of tasks, the resulting worst- and best-case execution give lower or upper bounds on their periodic occurrence which permits to estimate whether the required deadlines will be met.

As outlined above, controllers have to cope with the described restrictions, while real world physical entities may behave arbitrary fast. Therefore, a notion for temporal state-based behavior should permit to distinguish both kinds of models and detect whether a model can be executed on a limited device or is a physical abstraction that cannot be realized by a controller. We further name these two different kinds of models

- **limited**, if they can be realized on a man made digital controller or
- **unlimited**, if they describe the behavior of a physical device that cannot be realized on a man made digital controller.

It is to be noted, that certain abstractions might result in a mix of both models. A communication subsystem, for example, consists besides the technical components and software of a transmission media or channel which may result in possible failures or delays that result in an unlimited behavior. For a given hardware, a limited model might still be realizable or not depending on its resource requirements.

### 2.2 Expressiveness

To support the design of complex reactive behavior statecharts [Har87] extended the flat unstructured control state space of traditional finite state machines by supporting specific concepts such as hierarchy, parallel AND-states and deep and shallow history states. Therefore, rather complex state-based behavior can often be described in a compact form. Note that besides modelling comfort also the expressiveness of automata models with or without concepts like AND states vary. Therefore the number of required states to model a specific complex behavior depends on the availability of these higher-level description concepts (cf. [DH94]).
2.3 Temporal Behavior

For Statecharts the temporal behavior can be specified by the \textit{after} and \textit{when} constructs. A transition labeled with \textit{after}(x) fires \(x\) time units after entering the source state, \textit{when}(x) fires at point \(x\) in time. So \textit{after} is used for the declaration of a relative date, \textit{when} for an absolute. In this form, support of an unlimited device is needed to identify the point of times without delay. Besides these explicit constructs which permit to describe timeout behavior as required for the real-time domain, the temporal behavior of Statecharts is usually not specified. Thus for all side-effects and required computations no timing information is present.

Statecharts abstract from concrete timing information and assume an infinite fast device. As outlined in the preceding part about real-time systems, for a controller such an assumption is not realistic. In practice the assumption can be reduced by simply stating that the Statechart is executed fast enough to react on all relevant environment changes. Fast enough is difficult to determine and a model is required that somehow also represents this information. Therefore, an extension with explicit timing information is required.

In the case of the UML state machine [Obj01] the related micro- and macro-step execution order has been adjusted to ensure that the machine does first compute all internal effects of a received event (run-to-completion) before new external events are considered. While this simplifies the implementation of UML state machines on the other hand the delay in reaction to external events can be rather prohibitive for a real-time system. Therefore, when looking for an appropriate real-time modelling notation we have to keep in mind that reasonable worst-case reaction time may be in conflict with some features of the macro- and micro-step semantics and run-to-completion semantics.

2.3 Temporal Behavior

In real world physical entities operations are time-consuming. When they are executed on a limited device, their maximum duration is denoted by the worst case execution time (WCET). \textit{Deadlines} are used to specify that an action has to finish its execution in the given time bound. Typically automata are implemented by a control task, that checks periodically the activation of transitions. So an activated transition is not fired at the point of activation, but rather during the following period of the control task. This results in a delay, caused by the limitation of the device. Specification of the deadline gives – w.r.t. the WCET – an upper bound for the acceptable delay.

When deadlines and WCETs are specified, well-known methods and algorithms can be used to determine if a schedule exists, affecting that all deadlines are met, and to determine this schedule; e.g. using a priority scheduler and assigning the priorities according to the Deadline Monotonic Priority Assignment.

The employed notion of time in a real-time system is usually limited by the maximal clock resolution of the underlying RTOS or hardware. Therefore, the clock values we can compare in a program can always be encoded by a countable set of time points. When instead of absolute clock values only relative time differences are used, a straightforward extension of an automata models towards discrete time exists where different clock values are used to determine time-dependent behavior. However, when using such clock values this usually does not include the assumption that time elapses only in discrete steps. Instead a clock value \(n\) is assumed to represent the interval \([n, n+1)\) rather than a discrete point in time. Therefore, the read clock values and clock value comparisons check whether upper or lower bounds actually hold for a clock value and the underlying interpretation is that the clock therefore is in the time interval determined by the corresponding interval.
2.4 Predictability

The required support for predictability can vary to a great extent. Traditionally, the periodic nature of all real-time tasks is exploited to predict whether all deadlines will be met (feasibility analysis). If more specific system properties should be predicted or the assumption of independent periodic tasks is not valid, however, other techniques for automatic formal verification such as model checking may be employed.

2.4.1 Feasibility Analysis

Most approaches to real-time system design support a feasibility analysis to check whether a given set of periodic tasks can be served in a manner that all deadlines are met. As there are many different kinds of schedulers, for every approach exist different algorithms, for checking the feasibility.

Scheduling can be preemptive or non-preemptive, static or dynamic. One widespread manner is the static, preemptive priority scheduling. Using this scheduler, a priority is assigned to every task. Whenever the scheduler is assigning tasks to processors and has to distinguish between multiple tasks, it chooses the one with the highest priority. If the processor is busy with a low priority task and a high priority task arrives, then the active task is preempted and the processor is assigned to the task with the higher priority. When it is possible to change the priorities during runtime, the scheduling is called dynamic, otherwise static. Thus, the feasibility of a set of tasks strongly depends on the manner of priority assignment.

The Rate Monotonic Approach (RMA) [BR99], that assigns higher priorities to tasks with shorter periods, results in an optimal scheduling. That means, that the set of tasks, having rate monotonic priorities, is schedulable, if any priority assignment exists, that makes the set schedulable. RMA is limited to periodic tasks. A set of tasks, containing aperiodic tasks, the Deadline Monotonic Priority Assignment [BW01] (a task with a short deadline achieves a high priority), is optimal. Both approaches are static, as the priorities are computed in advance of starting the tasks and are not changed during runtime. As these approaches are widespread, there exist a couple of algorithms for accomplishing a feasibility analysis. The theorems from Liu and Layland resp. Lehoccy, Sha and Ding [BR99] confirm the feasibility of a set of tasks scheduled by RMA. In [Jos96] an algorithm is presented, accomplishing an analysis for priority scheduling independent of the manner of priority assignment.

In general, different tasks may access the same resources (like memory). Obviously, the access to these shared resources, has to be controlled as there will be problems, when a task, writing to a memory address, is interrupted by an higher prioritized thread, which writes to the same address. To avoid this problem a task, accessing a shared resource, blocks all (higher prioritized) tasks, that try to access this resource. Unfortunately, this solution results in an effect called Priority Inversion. This occurs when due to blocking the highest prioritized task finishes its execution at last (see [Jos96, BR99] for a complete example). Because of this effect, protocols have been investigated to minimize the blocking times. Widespread protocols are the Priority Inheritance Protocol and the Priority Ceiling Emulation [BR99].

2.4.2 Formal Verification

A behavioral modelling technique with well defined semantics permits to analyze the required temporal model properties using, e.g., temporal logic [Pnu77]. Given a specification in a temporal logic $\phi$ and a model $M$ there exist a number of techniques to prove that the model fulfils the
specification \((M \models \phi)\). Following [Lam77] such specifications can be further separated into safety and liveness conditions. While safety conditions ensure that something bad never happens, do liveness conditions ensure that something good will happen. Note, that thus liveness includes progress (something will happen).

The basic notion of temporal logic does only cover causal ordering, but does not permit to express properties with explicit real-time constraints. Therefore, a number of extensions with different time models have been proposed.

A first natural approach is to employ and extend available techniques for un-timed systems such as model checking. If a countable set of time points are sufficient a discrete notion of time can be used where, for example, the natural numbers denote all possible time points. When instead of absolute values only relative time differences are used, a straightforward extension of an automata model towards discrete time exists and available techniques for its analysis can be employed (cf. [EMSS90]).

A Tool for such a discrete timed model is, for example the model checker RAVEN [RK99, RK00] which employs clocked computation tree logic (CCTL) as specification language. For a timed step with a delay of more than one minimal time unit in such a discrete interpretation of time the question arises what is the correct interpretation of the intermediate states. Do these intermediate states correspond to “copies” of the source state, “copies” of the target state, or is in fact nothing known about these intermediate states. Therefore, in [LS01] a temporal logic and related model checking techniques are proposed that assume the latter while in RAVEN the first case is assumed.

However, the assumption that time elapses only in discrete steps is rather strong. A more serious problem in practice is to choose a sufficient small minimal discrete time value. On the one hand the choice has to ensure that the above assumption of a correct time abstraction holds. On the other hand the applicability of analysis techniques crucially depends on a chosen minimal discrete time value large enough. Therefore a systematic way to determine such a minimal discrete time value is required, before these approaches can be safely employed to automatically predict and analyze the required system properties.

As outlined in [HNSY92], when a digital clock value \(n\) is assumed to represent the interval \([n, n+1)\) rather than a discrete point in time, for restricted cases of models or temporal logic formulas results proven for the digitalization will also hold for the more realistic continuous time model. When also stronger specifications in the digital model are used and the digitization granularity is refined we can further approximate step-wise the required continuous time property.

To further automate this task, an approach that automatically determines a finite abstraction of a given real-time system is more appropriate than a manual approximation. For a continuous model of time, timed automata [AD90] as automata theoretic approach provides such a solution. It supports continuous clocks and employs the property that the clocks are only tested for a finite number of rational constants for only rational constants to construct of a finite abstraction in form of a region-graph (cf. [AD90]). With COSPAN [AK96], HyTECH [HHWT95a], KRONOS [DOTY96], and UPPAAL [LPPY97, PL00] a number of tools for model checking Timed Automata exist. Some of them further support hybrid and parameterized verification.

### 2.5 Realizability

A real-time system consists of hardware (controller, resources, actuators, sensors etc.) and its software. Concerning the integration between software and hardware there are 2 common platforms: (1) a real-time operating system (RTOS) is used to decouple the real-time application from the hardware and (2) the special I/O architecture of a micro-controller is directly employed to connect
the real-time application with its environment. The advantages of the RTOS approach are that it provides operations with uniform access to the underlying hardware and can control the scheduling of multiple tasks which includes their access to the (shared) resources. Alternatively, when a micro-controller is employed, the whole logic of the application has to be covered by a single task. In this case the scheduler, respectively the RTOS, is not needed. One of the advantages of the multi-task architecture is, that the tasks can control each other and handle errors. In a single thread application, this is not possible.

Real-time systems are often also *embedded systems* and thus besides the described heterogeneity w.r.t. scheduling also the available resources and execution restrictions can vary. When modelling a controller device these different forms (programmable logic controller (PLC), micro-controller (MC), and computer) restrict the delay and procedure the software can observe and control its environment. E.g., for PLC we usually have a bounded cycle time, while computer equipped with a RTOS permit a great variety of task periods which are served by the scheduler.

For the considered case of high-level behavior models, a suitable mapping to an implementation is required. Whether a mapping is possible or not has thus to be detected also at the high-level model. Otherwise, a delayed detection can result in rather wasted design efforts. Therefore, feasibility analysis at the level of the high-level behavior model is required.

### 3 Timed Automata

To discuss the foundations for extended state machine models for modelling temporal behavior, we introduce the required terminology and formal treatment by first reviewing the basic definition and terms for a finite automaton.

The finite automaton is the most fundamental concept for state-based behavior. Their main elements are a finite set of control states (locations) denoted by $S$, a contained set of start locations $S_0 \subseteq S$ and a set of transitions $T$, where each $t \in T$ further describes which changes between the locations are permitted depending either on internal choice or the environment. In a hardware oriented view the environment is modelled by input $i \in I$ and output $o \in O$ signals. The resulting output $o$ can be further described either as function of $S$ (Moore automata) or $S$ and $I$ (Mealy automata). In the considered software context instead a Mealy automata sending and receiving possible events $a \in A$ rather than permanently provided input and output signals is usually employed.

**Definition 1** A finite automaton $M$ is a 4-tuple $(S, S_0, A, T)$ with
- $S$ a finite non empty set of locations,
- $S_0 \subseteq S$ the subset of start locations,
- $A$ a finite set of events including the internal event $\tau \in A$.
- $T \subseteq S \times A \times S$ a finite set of transitions $t = (s, a, s') \in T$ with
  - $s \in S$ the source location,
  - $a \in A$ the related action, and
  - $s' \in S$ target location.

By convention for a transition $t = (s, a, s') \in T$ we also write $s \xrightarrow{a} s'$. Using $S$ as the set of possible configurations the resulting transition system is as follows.

**Definition 2** For a finite automaton $M = (S, S_0, A, T)$ the set of all possible configurations is simply $S$. The initial possible configurations are the elements of $S_0$. For a finite automate with
3.1 Timed Automata

In the simply case of a finite automaton the distinction between the syntactical notion of the location set and the possible configurations is rather superficial. However, in the following we will study extended concepts for automata where both notions are not synonymous any more.

As long as finite behavior is considered additionally often a set of accepting locations $S_f \subseteq S$ is defined. The valid finite traces $s_0 \xrightarrow{a} s_1 \cdots s_n$ with $s_0, s_1, \ldots, s_n \in S^*$ are given by all finite sequence of transitions starting with any start location $s_0 \in S_0$ which end at an accepting location $s_n \in S_f$ (acceptance condition). For infinite traces $s_0 \xrightarrow{a} s_1 \cdots$ with $s_0, s_1, \ldots \in S^\omega$, however, Büchi automata [Büc62] have to be used where instead for any possible infinite trace all states $s_a$ of the accepting location set $S_a$ have to be visited infinite often ($|\{i|s_i = s_a\}| = \infty$) to fulfill the acceptance condition.

This way to describe acceptance of infinite behavior is rather elegant, however, not really intuitive when modelling a system. This mismatch is related to the following underlying fact: For the acceptance of finite or non finite regular languages in form of an automata angelic non-determinism [HU79] is employed, which essentially assumes that choices are always made with global knowledge. When modelling system behavior in a constructive fashion by composing multiple independently operating subsystems, however, erratic non-determinism [Bro86] is natural, where each subsystem is assumed to behave arbitrary. Therefore, a different technique to model the liveness and progress properties of the system is required when modelling systems rather than characterizing the acceptance of a language.

3.1 Timed Automata

Automata can be extended to also describe the temporal behavior using a continuous notion of time. The term timed automata also includes a number of automata theoretic approaches with real-value clocks. The main characteristic is that for the models often a finite abstraction can be constructed, which results in a number of related model checking approaches. One basic notion of Timed Automata also named timed graphs [ACD90] uses real-value clocks to further restrict the occurrence of transitions. This notion, however, does not exclude that the automata may idle forever in any reached configuration. The original proposal [AD90] avoids this problem extending the concepts of Büchi automata [Büc62] and thus accepts only such infinite runs which also infinitely often visits states marked as accepting. While this notion can be used to specify progress conditions in [HNSY92] a notion for Timed Automata also named timed safety automata is proposed that avoids to require accepting states and instead introduces clock constraints for control states to model progress guarantees explicitly. We will further use this last notion as basic notion of timed automata and refer to [AD94, Pet99] for more detailed overviews.

The basic extension w.r.t. finite automata is that besides the control locations $S$ a set of real-value clocks $C$ is also part of the transition system configuration in form of a clock binding $\nu \in C_\nu = [C \to R_+]$. $\nu^0$ is the special binding which assigns to each clock variable 0. To control the temporal behavior clock constrains $C_c = [C \to R] \to \{true, false\}$ in form of time guards for transitions and time invariants for locations are employed. To reset the clock values as required specific clock updates $C_u = [C_\nu \to C_\nu]$ are used.

\[\text{For a more detailed discussion on the different forms of non-determinism see [WM97].}\]
Definition 3 A Timed Automaton $M$ is a tuple $(S, S_0, T, \text{Inv}, A, C)$ where

- $S$ a finite non empty set of locations,
- $S_0 \subseteq L$ the subset of start locations,
- $\text{Inv} : S \to C_c$ assigns each state a clock constraint,
- $A$ a finite set of events including the internal event $\tau \in A$,
- $C$ a finite set of clocks and,
- $T \subseteq S \times A \times C_c \times C_u \times S$ a finite set of transitions $t = (l, a, g, r, l') \in T$ with
  - $l \in S$ the source location,
  - $a \in A$ the related action,
  - $g \in C_c$ a clock constraint named time guard,
  - $r \in C_u$ a clock update, and
  - $l' \in S$ target location.

In the resulting transition system two forms of transitions are present. The transitions associated with each transition of $T$ occur in a timeless step. Additionally, so called time steps which remain within the same control state (location) are possible where any arbitrary time delta $d$ elapses.

Definition 4 For a Timed Automaton $M = (S, S_0, T, A, C)$ the state space is build by all pairs $(l, \nu) \in S \times C_{\nu}$. The initial possible states are all such pairs $(l, \nu^0)$ with $l \in S_0$. For a Timed Automaton in configuration $(l, \nu)$ we further define enabling by $\text{Enabled}((l'', \nu), (l, a, g, r, l')) \iff l'' = l \land g(\nu) \land \text{Inv}(l')(r(\nu))$ and the resulting configuration transition $\Gamma_{(l, a, g, r, l')}(l, r(\nu)) := (l', r(\nu))$. Thus, for a transition $t \in T$ the following action rule results:

$$\frac{\text{Enabled}((l, \nu), t)}{(l, \nu) \xrightarrow{t} \Gamma_t(l, \nu)} \text{ action.}$$

Time can further elapse in so called delay steps and for the corresponding update function $\Gamma_d((l, \nu)) := (l, \nu + d)$, where $\nu + d$ stands for the current clock assignment plus the delay for all the clocks, we have

$$\frac{\text{Inv}(l)(\nu + d)}{(l, \nu) \xrightarrow{\Delta} \Gamma_d(l, \nu)} \text{ delay.}$$

Typically the updates and constraints for the continuous clock variables are further syntactically restricted to ensure that a finite abstraction [AD90] exists. For the transition updates $C_u$ hold that clocks are usually only reset to 0 or integer numbers. For the time constraints in form of time invariants for locations and time guards for transitions usually the expressions are limited for $x, y$ clocks and constants $n$ to the conjunctive combination of the form $x \square n$ or $x - y \square n$ with $\square \in \{ <, \leq, =, \geq, > \}$.

For this restricted notion of time constraints the reachability problem for Timed Automata is decidable when the coefficients $(n)$ in the guards are rational numbers [AD90]. The reachability problem becomes as soon undecidable as the coefficients are chosen, for example, from the set $\{1, \sqrt{2}\}$ [Pur99]. When additive clock constraints are used, already for no more than 4 clocks the emptiness problem is in general undecidable [BD00].

3.2 Consistency

The described Timed Automata model can further result in subtle semantic problems. For example, for the employed notion of a Timed Automaton a configuration without any further possible timeless step named deadlock may be reached.
3.3 Extensions

Definition 5 A configuration \((l, \nu)\) is a deadlock iff \(\forall t \in T \ hold \ \neg \text{Enabled}((l, \nu), t)\).

A configuration without further external visible timeless steps still describes a reasonable behavior. If, however, the related time invariant of the control state (location) expires and further time elapses an inconsistent temporal behavior has been specified.

Definition 6 A configuration \((l, \nu)\) is a time stopping deadlock iff it is a deadlock and for any \(d > 0\) hold \(\text{Inv}(l)(\nu + d) = \text{false}\).

Another problem of the presented approach to model continuous time behavior is caused by the employed interleaving of time steps and discrete action steps. While for arbitrary delta \(d\) and \(d'\) time steps still describe a reasonable behavior, a sequence \((l_0, \nu_0) \xrightarrow{d_0} (l_1, \nu_1) \xrightarrow{d_1} (l_2, \nu_2) \xrightarrow{d_2} (l_3, \nu_3) \xrightarrow{d_3} (l_4, \nu_4) \ldots\) with alternating time and action step is not reasonable when infinitely many discrete action steps occur in a finite amount of time \((\exists d = \sum_{i=1}^{\infty} d_i)\).

Definition 7 An infinite sequence of configurations and steps \((l_0, \nu_0) \xrightarrow{x_0} (l_1, \nu_1) \xrightarrow{x_1} (l_2, \nu_2) \xrightarrow{x_2} (l_3, \nu_3) \xrightarrow{x_3} (l_4, \nu_4) \ldots\) with \(x_i \in (T \cup \mathbb{R}^+)\) is considered to be zeno iff the number \(|\{i | x_i \in T\}|\) is infinite while for the elapsed time a finite upper bound \(d\) exists such that \(\sum_{x_i \in \mathbb{R}^+} d_i \leq d\).

The problem of zeno behavior will usually only occur for unlimited systems, because there exists a lower bound \(t_i\) for all the times required to execute a transition in a specific situation. Therefore, as long as besides the zeno behavior a reasonable non-zeno behavior exists it is usually assumed to simply exclude the traces with zeno behavior. When, however, unlimited systems and their physics is considered, the result of an infinite but convergent chain of transitions and time steps may indeed determine the correct behavior of the system.

3.3 Extensions

Besides this basic notion of Timed Automata several extensions have been proposed to simplify the specification of temporal behavior.

For a configuration \((l, \nu)\) any enabled transition \(t\) may occur or not as long as the time invariant of the current location is valid. The only case when some transition will be executed definitely is when the invariant does not permit any further delay. Therefore, the concept of urgent transitions [HHWT95b, DY95, BLL∗95] has been proposed where time can only elapse when no urgent transition is enabled.

Formally for \(T_u \subseteq T\) the subset of all urgent transitions which are often further restricted to have empty guards \((g = \text{true})\) we can define whether an urgent transition is enabled \(\text{UrgentEnabled}((l, \nu)) := \exists t \in T_u \\text{Enabled}((l, \nu), t)\) and simply adjust the delay rule accordingly.

To simplify the modelling of complex state spaces besides explicit control states also integer variables [LPY97] have been invented. The resulting reachable system configurations simply have to include a binding for the variables, too.

Note, that in the presented basic notion of Timed Automata the labels \(a \in A\) do not effect the execution of a transition. For a system of multiple parallel executed Timed Automata synchronous
communication using channels in a Calculus of Communicating Systems (CCS) [Mil89] style has been proposed in [LPY97]. This permits to coordinate two parallel Timed Automata in a single atomic step. Two transitions with compatible channel events $e!$ and $e?$ are combined into a single transition of the combined system.

Another way to let multiple parallel operating Timed Automata interact are shared variables [LPY97]. In contrast to the synchronous communication do shared variables facilitate an asynchronous exchange of information between parallel Timed Automata.

3.4 Summary

A system of Timed Automata with synchronous communication and shared variables provides already the required fundamental concepts to describe the complex real-time behavior of large systems consisting of a number of components. For reasonable small systems of Timed Automata model checking techniques further provide the required analysis and help to predict whether required properties hold. Therefore, the requirements temporal behavior and predictability are covered. However, the presented basic formalisms are still limited with respect to expressiveness and realizability. In the next section we therefore consider an approach that further extend the only flat structures supported by Timed Automata towards complex state spaces as supported by Statecharts.

4 Hierarchical Timed Automata

As identified in Section 2.2 the fundamental extensions to cope with complex behavior are hierarchy, parallelism and history as provided by statecharts [Har87] and its UML derivation UML state machines [Obj01]. The notion of a Hierarchical Timed Automata model [DMY02] provides the most fundamental structural concepts as well as a reasonable continuous time semantics. This notion supports model checking by a mapping to plain Timed Automata as employed in the UPPAAL tool [LPY95]. As part of the AIT-WOODDES research project a corresponding UML state machine profile is further planed. In the following we define Hierarchical Timed Automata as introduced in [DMY02] and summarize their exact limitations w.r.t. the in Section 2 identified requirements.

4.1 Syntax

Definition 8 A Hierarchical Timed Automaton is a tuple $(S, S_0, \delta, \sigma, V, C, \text{Inv}, \text{Ch}, T)$ where

- $S$ is a finite set of locations. root $\in S$ is the root.
- $S_0 \subseteq S$ is a set of initial locations.
- $\delta : S \rightarrow 2^S$ is a function, mapping a location $l$ to all possible substates of $l$ and generating a tree structure with root as the root. Applying $\delta$ on sets, delivers the intuitively expected results.
- $\sigma : S \rightarrow \{\text{AND}, \text{XOR}, \text{BASIC}, \text{ENTRY}, \text{EXIT}, \text{HISTORY}\}$ associates a type to every location.
- $V, C, \text{Ch}$ are sets of variables, clocks, and channels and are used for the specification of Guard, Reset, Sync, and Invariant as described in the following.
4.1 Syntax

- \( \text{Inv} : S \rightarrow \text{Invariant assigns an invariant to every location.} \)

- \( T \subseteq S \times (\text{Guard} \times \text{Sync} \times \text{Reset} \times \{\text{true, false}\}) \times S \) is the set of transitions. A transition is composed of a source \( (l) \) and a target location \( (l') \), a guard \( g \), an assignment \( r \) (including clock resets), and an urgency flag \( u \). The notation \( l \xrightarrow{g,r,u} l' \) is used for a transition \( t = (l, s, g, r, u, l') \). We omit \( s, g, r, u \) when they are necessarily absent (or \text{false}, in the case of \( u \)).

Guards, synchronizations, resets, and assignments use Guard, Reset and Sync, as described in the following:

- \( V \) is the finite set of integer variables. \( V(l) \subseteq V \) is the set of integer variables local to a superstate \( l \).

- Let \( C \) be the finite set of clock variables. \( C(l) \subseteq C \) contains the clocks local to a superstate \( l \). If \( l \) is a history state, \( C(l) \) contains only clocks, that are declared as forgetful. The other clocks of \( l \) (local clocks) belong to \( C(\text{root}) \).

- Let \( \text{Ch} \) be the finite set of synchronization channels. \( \text{Ch}(l) \subseteq \text{Ch} \) contains the channels that are local to a superstate \( l \), i.e., it is not possible to synchronize along a channel \( c \in \text{Ch}(l) \) between one transition inside \( l \) and one outside \( l \).

- \( \text{Ch} \) leads to the finite set of channel synchronizations(\( \text{Sync} \)). It holds: \( c \in \text{Ch} \Rightarrow c?, c! \in \text{Sync} \). If \( s \in \text{Sync} \), then \( \overline{s} \) denotes the corresponding complementary (e.g. \( \overline{c!} = c? \) and \( \overline{c?} = c! \)).

- Data constraints are boolean expressions of the form \( A \square A \) with \( A \) is an arithmetic combination of elements from \( V \) and \( \square \in \{<, >, =, \leq, \geq\} \). Clock constraints are boolean expressions of the form \( x \square n \) or \( x - y \square n \), with \( x, y \in C \), \( n \in N \) and \( \square \in \{<, >, =, \leq, \geq\} \). A clock constraint is called downward closed, if \( \square \in \{<, =, \leq\} \). A guard is a finite conjunction over data and clock constraints. An invariant is a finite conjunction over downward closed clock constraints. The set of guards and invariants, joint with \( \{\text{true, false}\} \) are called Guard respectively Invariant.

- Clock resets are of the form \( x := 0 \), with \( x \in C \). Data assignments are of the form \( v := A \), with \( v \in V \) and \( A \) is an arithmetic combination of elements from \( V \). Reset is the set of clock resets and data assignments.

The function \( \text{TYPE} : S \rightarrow \{\text{true, false}\} \) is used as an abbreviation for the expression \( \sigma(l) = \text{AND} \). As the type \( \text{HISTORY} \) is an important one, the function \( \text{HENTRY} \) is defined similarly as \( \text{HENTRY}(l) = \text{ENTRY}(l) \lor \text{HISTORY}(l), l \in S \).

For support, we define the following functions:

- the parent function \( \delta^{-1}(l) = l' \) with \( \delta(l') = l \) and \( l \neq \text{root} \).

- \( \delta^+(l) \) returns the set of all nested locations of \( l \).

- \( \delta^{-*}(l) \) denotes the set of ancestors of the location \( l \), including \( l \).

- \( \delta^x(l) = \delta^+(l) \setminus \{l\} \)

- \( \overline{\delta}(l) = \{ n \in \delta(l) \mid \text{BASIC}(n) \lor \text{XOR}(n) \lor \text{AND}(n) \} \) to denote the children-locations, being of Type \( \text{BASIC}, \text{XOR}, \text{or AND} \).

- \( V^+(l) = \bigcup_{n \in \delta^{-*}(l)} V(l) \) to denote all variables in the scope of the location \( l \). \( C^+(l) \) and \( \text{Ch}^+(l) \) are defined analogously.
In [DMY02] a given hierarchical Timed Automaton is only well-formed when the following conditions are fulfilled.

**Definition 9** A Hierarchical Timed Automaton is well-formed iff

1. The Hierarchical Timed Automaton is rooted at root: $S = \delta^*(\text{root})$,
   - Basic nodes are empty: $\text{BASIC}(l) \leftrightarrow \delta(l) = \emptyset$,
   - substates of AND superstates are not basic: $\text{AND}(l) \land n \in \delta^*(l) \Rightarrow \neg \text{BASIC}(n)$, and
   - invariants of ”pseudo-locations” are trivial: $\text{ENTRY}(l) \lor \text{EXIT}(l) \Rightarrow \text{Inv}(l) = \text{true}$.

2. $S_0$ belongs to a situation of consistent and proper control, i.e., $\text{root} \in S_0$ and for every $l \in S_0$ the following holds:
   - $l$ is a proper location: $\text{BASIC}(l) \lor \text{XOR}(l) \lor \text{AND}(l)$,
   - $S_0$ does contain all parents: $l = \text{root} \lor \delta^{-1}(l) \in S_0$,
   - only one alternative of an XOR-state is present: $\text{XOR}(l) \Rightarrow |\delta(l) \cap S_0| = 1$, and
   - all substates of an AND-state are proper states: $\text{AND}(l) \Rightarrow \delta(l) \cap S_0 = \delta(l)$.

3. There are no conflicts in assignments in synchronizing transitions: $l_1 \xrightarrow{r,c,i,x,s,u} l_1', l_2 \xrightarrow{r',c',i',x',s',u'} l_2' \in T \Rightarrow \text{vars}(r) \cap \text{vars}(r') = \emptyset$, where $\text{vars}(r)$ is the set of integer variables occurring in $r$. A similar constraint is used to handle transitions originating in the entry of an AND superstate. For $l \xrightarrow{r,c,i,x,s,u} l' \in T, g, r$ are defined over $V^+(\delta^{-1}(l)) \cup C^+(\delta^{-1}(l))$ and $s$ is defined over $Ch^+(\delta^{-1}(l))$ to ensure a proper static scoping.

4. Let $e \in S, \text{ENTRY}(e)$.
   - If XOR$^{-1}(l))$, then $T$ contains exactly one transition $e \xrightarrow{l'} l'$.
   - If $\text{AND}^{-1}(l))$, then $T$ contains exactly one transition $e \xrightarrow{e, l} l'$ for every proper substate $l \in \delta^{-1}(l))$, and $e, l \in \delta(l_i)$.
   - In case of $\text{HISTORY}(e)$, the default history locations are specified by the outgoing transitions. If $\text{AND}(s) \lor \text{HISTORY}(s)$, then every substate $l$ of $s$ has to provide either a history entry or a default entry.

5. Transitions, connecting States, that are neither entries nor exits, at different hierarchical levels, are disallowed. Note that transitions are not allowed to lead directly from entries to exits. A Transition $l \xrightarrow{r,c,i,x,s,u} l'$ is called ”pseudo-transition” when $\text{ENTRY}(l)$ or $\text{EXIT}(l')$ holds. If $\text{ENTRY}(l)$ holds, only the form $l \xrightarrow{e, l} l'$ is valid. In the case of $\text{EXIT}(l')$, only the form $l \xrightarrow{e, l} l'$ is allowed. If $\text{EXIT}(l) \land \text{EXIT}(l')$ holds, the form is restricted to $l \xrightarrow{e, l} l'$.

### 4.2 Semantics

The semantics of the Hierarchical Timed Automata model is further defined in [DMY02] using a configuration of the form $(\rho, \mu, \nu, \theta)$ which contains the active locations, the integer variable values, the clock values, and the history.

- $\rho : S \rightarrow 2^S$ delivers a set containing all active substates of a superstate $s$. An inactive state $s$ results in $\rho(s) = \emptyset$. We define another function for support: $\text{Active}(l) = l \in \rho^x(\text{root})$, where $\rho^x(l) = \rho \setminus \{l\}$. Notice that $\text{Active}(l) \Leftrightarrow l \in \rho(\delta^{-1}(l))$. 

• \( \mu : S \rightarrow (\mathbb{Z})^* \) gives rise to the values of the variables of a given superstate \( l \). \( \mu \) returns a finite tuple of integer numbers. It is required that \( |\mu(l)| = |V(l)| \) and that \( \mu \) is consistent with respect to the value of shared variables (i.e., always maps to the same value). \( \mu(l)(a) \) defines the value of variable \( a \in V(l) \). If \( \neg\text{BASIC}(l) \) the tuple is extended by local variables initialised by a default value (0). We use the abbreviation \( 0^V(l) = (0, 0, ..., 0) \) with arity \( |V(l)| \). We demand \( \neg\text{Active}(l) \Rightarrow \mu(l) = \lambda \) (the empty tuple).

• \( \nu : S \rightarrow (\mathbb{R}^+)^* \) gives \(-\)similar to \( \mu \) rise to the values of the clocks of state \( l \). It is required that \( |\nu(l)| = |C(l)| \) and \( \neg\text{Active}(l) \Rightarrow \nu(l) = \lambda \).

• \( \theta = (\theta_{\text{state}}, \theta_{\text{var}}) \) denotes the substate(s) and local variables of a superstate that will become valid by entering this superstate via a history entry. Thus, \( \theta_{\text{state}} \) is used to restore \( \rho(l) \), while \( \theta_{\text{var}}(l) \) returns a tuple of integer values to restore \( V(l) \). Clocks are not influenced by the history.

The existence of a history entry is handled with the predicate \( \text{HasHistory}(l) = \exists n \in \delta(l)[\text{HISTORY}(n)] \). \( \text{HEntry}(l) \) is the unique history entry of a superstate \( l \) if \( \text{HasHistory}(l) \) holds, the default entry of \( l \) otherwise. If \( \text{BASIC}(l) \) then \( \text{HEntry}(l) = l \) holds. \( \text{HEntry}(l) \) is undefined for all other cases. Initially, \( \forall l \in \{S|\text{HasHistory}(l)\} \Rightarrow \theta_{\text{state}}(l) = \text{HEntry}(l) \wedge \theta_{\text{var}}(l) = 0^V(l) \) holds.

The function \( \text{Targets}_\theta \) to describes the locations reached by following a fork is defined as
\[
\text{Targets}_\theta(l) = L \cup \bigcup_{l' \in L} \{n|n \in \theta_{\text{state}}(l) \wedge \text{HISTORY}(l') \} \cup \{n|(l, \tau, r, n) \in T \wedge \text{ENTRY}(l)\}.
\]

\( \text{Targets}_\theta(l) \) is used as abbreviation for \( \text{Targets}_\theta([l]) \). \( \text{Targets}_\theta^* \) is the reflexive transitive closure of \( \text{Targets}_\theta \).

Firing a transition \( t : l \xrightarrow{a,b,c,d} l' \) can be split up into 3 steps: 1. executing a join to exit \( l \), 2. taking the proper transition \( t \) itself, and 3. executing a fork at \( l \). In case of \( \text{BASIC}(l) \) (respectively \( \text{BASIC}(l') \)), part 1. (respectively 3.) is trivial. The 3 steps define a run-to-completion step. A run-to-completion step is formally represented by a transformation function \( \Gamma_t \), where \( t \) is a transition. The 3 parts of the step are described as follows:

1. join : \( (\rho, \mu, \nu, \theta) \) is transformed to \( (\rho^1, \mu^1, \nu^1, \theta^1) \) as follows: \( \rho \) is updated to \( \rho^1 := \rho \forall n \in \rho^*(l)[n \rightarrow \emptyset] \), \( \mu \) is updated to \( \mu^1 := \mu \forall n \in \rho^*(l)[n \rightarrow \lambda] \), \( \nu \) is updated to \( \nu^1 := \nu \forall n \in \rho^*(l)[n \rightarrow \lambda] \). \( \Gamma_t \) is recorded. Let \( H \) be the set of superstates \( h \in \rho^*(\delta^{-1}(l)) \), where \( \text{HasHistory}(h) \) holds. Then \( \theta^1_{\text{state}} := \theta_{\text{state}}[\forall h \in H|h \rightarrow \text{HEntry}(\rho(h))] \) and \( \theta^1_{\text{var}} := \theta_{\text{var}}[\forall h \in H|h \rightarrow \mu(h)] \). If \( \neg\text{EXIT}(l) \) or \( H = \emptyset \), then \( \theta^1 := \theta \).

2. proper transition part : \( (\rho^1, \mu^1, \nu^1, \theta^1) \) is transformed to \( (\rho^2, \mu^2, \nu^2, \theta^2) := (\rho^1[l'], r(\mu^1), r(\nu^1), \theta^1) \). \( r(\mu^1) \) denotes the updated values of the integers after the assignments and \( r(\nu^1) \) the updated clocks after the resets.

3. fork : \( (\rho^2, \mu^2, \nu^2, \theta^2) \) is transformed to \( (\rho^3, \mu^3, \nu^3, \theta^3) \) by moving the control to all proper locations reached by the fork, i.e. those in \( \text{Targets}_\theta(l') \). Note that \( \rho^3(n) = \emptyset \) for all \( n \in \delta^*(l') \). Thus we can compute \( \rho^3 \) from \( \rho^2 \) by modifying it iteratively for all \( n \in \text{Targets}_\theta(l') \) for \( \text{ENTRY}(n) \) such that \( \rho^3(\delta^{-2}(n)) := \rho^3(\delta^{-2}(n)) \cup \{\delta^{-1}(n)\} \) or else \( \rho^3(\delta^{-1}(n)) := \{n\} \). \( \mu^3 \) is derived from \( \mu^2 \) by first initializing all local variables of the superstates \( s \) in \( \text{Targets}_\theta(l') \), i.e. \( \mu^3(|V(s)|) := 0^V(s) \). If \( \text{HasHistory}(s), \theta_{\text{var}}(s) \) is used instead of \( 0^V(s) \). Then all variable assignments and clock-resets along the pseudo-transitions belonging to this fork are executed to update \( \rho^3 \) and \( \nu^3 \). The history does not change, \( \theta^3 \) identical to \( \theta^1 \).
As mentioned above, part 1. and 3. correspond to the identity transformation, if \( \text{BASIC}(l) \) and \( \text{BASIC}(l') \) hold. The transformation function \( \Gamma_t \) of a transition \( t \) returns the result of the 3 steps: \( \Gamma_t(\rho, \mu, \nu, \theta) = (\rho^3, \mu^3, \nu^3, \theta^3) \). \( \rho^t = \rho^3 \) and \( \nu^t = \nu^3 \) is used if the context is definite.

A state can only be exited, if all its parallel substates can synchronize on this exit. The function \( \text{PreExitSets}(l) \) returns a family of sets. Every element \( X \) of \( \text{PreExitSets}(l) \) is a set of exits. By firing the transitions of all exits in \( X \), all substates synchronize.

\[
\text{PreExitSets}(l) = \begin{cases} 
\bigcup_{n_1, \ldots, n_k \in \mathbb{N}} \text{PreExitSets}(n_i), \text{ where } k = \left\lceil \delta^{-1}(l) \right\rceil, \{n_1, \ldots, n_k\} \subseteq \delta^{-1}(l_i), \\
\forall i. \text{EXIT}(n_k) \land n_i \rightarrow l \in T \\
\delta^{-1}(\{n_1, \ldots, n_k\}) = \delta(l) \\
\{\{l\} \cup \bigcup_{m \in S} \text{PreExitSets}(m)\}, \text{if EXIT}(l) \land \text{XOR}(\delta^{-1}(l)) \\
\{\}, \text{if BASIC}(l) 
\end{cases}
\]

The operator \( \oplus \) is the below described product over families of sets which is like intuitionelly expected extended for a given arbitrary finite number of arguments.

\[
\oplus : \left(2^{2^S}\right) \times \left(2^{2^S}\right) \rightarrow \left(2^{2^S}\right) \quad \oplus \{A_1, \ldots, A_n\}, \{B_1, \ldots, B_n\} = \{A_1 \cup B_1, \ldots, A_n \cup B_n\}.
\]

To denote the active leaves of the tree, described by \( \rho \), \( \text{Leaves}(\rho, l) = \{n \in \rho^x(l) | \rho(n) = 0\} \) is used. \( \text{Leaves}(\rho, l) \) is used to define predicates for the transition system rules for \( \text{JoinEnabled} \).

The predicate, expressing that all the substates of a state \( l \) can synchronize on a join, is:

\[
\text{JoinEnabled}(\rho, \mu, \nu, l) = \text{BASIC}(l) \lor \exists X \in \text{PreExitSets}(l) \cdot \forall n \in \text{Leaves}(\rho, l) \cdot \exists n' \in X. n \xrightarrow{\rho^x} n' \land g(\mu, \nu)
\]

The invariants of a location are denoted by \( \text{Inv}_\nu : S \rightarrow \{\text{true, false}\} \), which evaluates the invariant of a location dependant on a clock evaluation \( \nu \). The predicate \( \text{Inv}(\rho, \nu) = \wedge_{n \in \rho^x(\text{root})} \text{Inv}_\nu(n) \) is used to express, that for the state denoted by \( \rho \) and for the clock valuation \( \nu \) all invariants hold. The predicate \( \text{Enabled} : T \rightarrow \{\text{true, false}\} \) describes whether \( t \) is enabled:

\[
\text{Enabled}(t : l \xrightarrow{g \cdot \Delta} l', \rho, \mu, \nu) = \neg \text{EXIT}(l') \lor g(\mu, \nu) \land \text{JoinEnabled}(\rho, \mu, \nu, l) \land \text{Inv}(\rho^t, \nu^t).
\]

As urgency has precedence over delay, it has to be well-known wether an urgent transition is enabled. This information is delivered by the predicate \( \text{UrgentEnabled} \) over a configuration.

\[
\text{UrgentEnabled}(\rho, \mu, \nu) = (\exists t : l \xrightarrow{g \cdot \Delta} l' \land u \land \text{Enabled}(t, \rho, \mu, \nu)) \lor (\exists t_1 : l_1 \xrightarrow{g \cdot \Delta^1} l_1' \land l_1 \notin \delta^\nu(l_1) \land \text{Enabled}(t_1, \rho, \mu, \nu) \land \text{Enabled}(t_2, \rho, \mu, \nu) \land (u_1 \land u_2)).
\]

The resulting rules for the transition system steps \( \text{action} \) for a transition \( t \in T \) (action), a time step (delay), and synchronization (sync) are:

\[
\text{action} \quad \text{delay} \quad \text{sync}
\]

\[
\frac{\text{Enabled}(t : l \xrightarrow{g \cdot d} l', \rho, \mu, \nu)}{\Gamma_t(\rho, \mu, \nu, \theta)} \quad \frac{\text{Inv}(l)(\rho, \mu, \nu)}{\Gamma_t(\rho, \mu, \nu, \theta)} \quad \frac{\text{Enabled}(l_1 : l_1 \xrightarrow{g \cdot \Delta^1} l_1', \rho, \mu, \nu)}{\Gamma_t(\rho, \mu, \nu, \theta)} \quad \frac{\text{Enabled}(l_2 : l_2 \xrightarrow{g \cdot \Delta^2} l_2', \rho, \mu, \nu)}{\Gamma_t(\rho, \mu, \nu, \theta)}
\]
4.3 Summary

While the Hierarchical Timed Automaton covers to a certain degree the three requirements temporal behavior, predictability and expressiveness a number of limitations remain: At first the approach is rather limited w.r.t. the data model. Only integer variables and their assignment are supported. Therefore, it is not possible to include complex data and operations even at the semantics level. The supported synchronous event model is also a rather strong limitation. Often asynchronous events rather than synchronous ones are used to decouple the detection and reaction on a specific external event. The formalism also does not support priorities for transitions. The only way to ensure that a certain precedence between two transitions activated at the same time holds will be the use of the urgency flag. However, usually if the first transition is not enabled we would expect the second one to fire immediately.

5 Extended Hierarchical Timed Automata

The most proposed modelling techniques focus either on structural expressiveness or temporal properties. The in the last section defined notion of Hierarchical Timed Automata addresses both aspects, however, a number of serious limitations and the realizability problem remain. We have to first take into account that the rest of the UML model and code builds a data model which is observed and modified with queries and updates, respectively. The provided synchronous events mechanism is also not sufficient in the long term and asynchronous events are additionally required. In practice different priorities rather than only urgent or non urgent transitions would permit to describe the precedence of transitions in a more appropriate manner.

5.1 Extensions

To overcome the presented limitations we propose a notion of an Extended Hierarchical Timed Automaton for modelling of complex real-time behavior by extending the Hierarchical Timed Automata model accordingly.

**Data Model**  To integrate the UML data model with queries and updates within the formalism we extend the guards and resets with more generic elements for manipulation of a data model. When parallel access to the same part of the model occurs inconsistencies due to shared access may be possible. Therefore, we have to assume that the data model modifications are split into smaller steps which each guarantees not interference with other ones.

The intention behind such a more general treatment of the data model is, that such a formalization enables the following two strategies to incorporate the effects of complex data models:

1. Sophisticated data modelling approaches such as graph grammars may be employed to describe the corresponding semantics.

2. We may abstract from data model effects using non-determinism.

While we could fully replace the integer variables $V$, we better keep them to permit the explicit state space encoding by means of integer or enum types variables when appropriate.
Asynchronous Events  To support asynchronous events in a systematic manner in addition to the synchronous channels also asynchronous ones are supported. We support the same style of scoping as for synchronous channels and realize a queue management that erases only events when explicitly consumed.

Priorities  The support for urgent transitions does not permit to specify more detailed precedence information in case of multiple possible conflicting transitions. Therefore we replace the binary distinction between urgent and non urgent transitions and instead permit to specify priorities with natural numbers ($\mathbb{N}$). A priority 0 is identical to a not urgent transition in the Hierarchical Timed Automata model and for the case of synchronizing transitions $t_1$ and $t_2$ with priorities $p_1$ and $p_2$, we use $\min(p_1, p_2)$. To ensure fairness parallel transitions should not be effected by higher priority transitions in a parallel substate and therefore an enabled transition $t_1$ should only rule out another one ($t_2$) if both are conflicting transitions and $p_1 > p_2$.

5.2 Syntax

These extensions result in the following definition for an Extended Hierarchical Timed Automata.

Definition 10  The syntax of an Extended Hierarchical Timed Automaton is defined by a tuple $(S, S_0, \delta, \sigma, D, V, C, Inv, Ch, T)$ where

- $S$ is a finite set of locations. $\text{root} \in S$ is the root.
- $S_0 \in S$ is a set of initial locations.
- $\delta : S \to 2^S$. $\delta$ maps $l$ to all possible substates of $l$. $\delta$ is required to give rise to a tree structure with root $\text{root}$. We readily extend $\delta$ to operate on sets of locations in the obvious way.
- $\sigma : S \to \{\text{AND}, \text{XOR}, \text{BASIC}, \text{ENTRY}, \text{EXIT}, \text{HISTORY}\}$ is a type function on locations.
- $D, V, C, Ch$ are a data model, sets of variables, clocks, and channels. They give rise to Guard, Reset, Sync, and Invariant as defined in the following.
- $\text{Inv} : S \to \text{Invariant}$ maps every location $l$ to an invariant.
- $T \subseteq S \times (\text{Guard} \times \text{Sync} \times \text{Reset} \times \mathbb{N}) \times S$ is the set of transitions. A transition connects two locations $l$ and $l'$, has a guard $g$, a synchronization $s$, an assignment $r$ (including clock resets), and an priority $p$. We use the notation $l \xrightarrow{\delta : \sigma : \alpha : p} l'$ for this and omit $g, s, r, p$ when they are necessarily absent (or 0, in the case of $p$).

The data components in guards, synchronizations, resets, and assignment expressions are defined:

- $D$ is a possibly infinite data model. A set of possible queries $[D \to \{\text{true}, \text{false}\}]$ and transformations $[D \to D]$ are assumed. To further take into account that data model updates or queries cannot be considered to be timeless and atomic their interference has to be controlled in an appropriate manner. Therefore, we assume a symmetric function $\text{conflict}$ that for an update $\text{up}$ and an update or query $\text{up}'$ determines whether their concurrent execution can result in such interference ($\text{conflict}(\text{up}, \text{up}') = \text{true}$). A conflict can be excluded between a guard and an update only, when the update cannot effect the outcome of evaluating the guard. For two updates in contrast, basic concurrency techniques such as monitors or semaphores may be employed to ensure interference free execution which is modelled by different forms of interleaving.
• $V$ is the finite set of integer variables. $V(l) \subseteq V$ is the set of integer variables local to a superstate $l$.

• Let $C$ be a finite set of clock variables. The set $C(l) \subseteq C$ denotes the clocks local to a superstate $l$. If $l$ has a history entry, $C(l)$ contains only clocks, that are explicitly declared as forgetful. Other locally declared clocks of $l$ belong to $C(root)$.

• Let $Ch$ be a finite set of synchronization channels with subsets $Ch_s$ and $Ch_a$ for synchronous and asynchronous behavior. $Ch(l) \subseteq Ch$ is the set of channels that are local to a superstate $l$, i.e., there cannot be synchronization along a channel $c \in Ch(l)$ between one transition inside $l$ and one outside $l$. We further restrict the asynchronous communication channels to be only global ones.

• $Ch$ gives rise to a finite set of channel synchronizations, called Sync. For synchronous channels $c \in Ch_s$, $c^? , c^! , \tau \in Sync_s$. For $s \in Sync_s - \{ \tau \}, \overline{\tau}$ denotes the matching complementary, i.e., $c^? = \overline{c^!}$ and $c^! = \overline{c^?}$. For events $c \in Ch_a$ we further can send and receive simultaneously on all asynchronous channels described by $Sync_a : [Ch_a \rightarrow \mathbb{N}] \times [Ch_a \rightarrow \mathbb{N}]$. We use the notation $(c^? , c^! , \ldots )$ respectively $(c^? , c^!, \ldots )$ to describe which events are received respectively send how often. A synchronization is then, for example, $(c^? , (c^? , c^! , \ldots ) , (c^? , c^!, \ldots ))$ and the overall set of synchronizations is build by $Sync := Sync_s \times Sync_a$ which permits multiple asynchronous send and receives.

• The data model constraint is any boolean query $q_D \in [D \rightarrow \{ true, false \}]$. A variable constraint is a boolean expressions of the form $A \sqcap A$, where $A$ is an arithmetic expression over $V$ and $\sqcap \in \{ <, >, =, \leq, \geq \}$. A clock constraint is an expression of the form $x \sqcap n$ or $x - y \sqcap n$, where $x, y \in C$ and $n \in \mathbb{N}$ with $\sqcap \in \{ <, >, =, \leq, \geq \}$. A clock constraint is downward closed, if $\sqcap \in \{ <, =, \leq \}$. A guard is a finite conjunction over data model constraints, variable constraints and clock constraints. An invariant is a finite conjunction over downward closed clock constraints. Guard is the set of guards, Invariant is the set of invariants. Both contain additionally the constants $true$ and $false$ and therefore when no additional guard is required simple the constant $true$ will be used.

• A clock reset is of the form $x := 0$, where $x \in C$. A data assignment is of the form $v := A$, where $v \in V$ and $A$ an arithmetic expression over $V$. Any $u \in [D \rightarrow D]$ is a data model update. Reset is the set of clock resets, data assignments, and data model updates. Reset is the set data assignments, and data model updates.

For $TYPE \in \{ AND, XOR, BASIC, ENTRY, EXIT, HISTORY \}$ we use the predicate notation $TYPE(l)$ for $l \in S$. E.g. $AND(l)$ is true, exactly if $\sigma(l) = AND$. The type $HISTORY$ is a special case of an entry. We use $ENTRY(l)$ to capture simple entry or history entry, i.e., $ENTRY(l) \equiv HISTORY(l)$.

For $l \neq root$ the parent function $\delta^{-1}(l)$ is defined as that unique $l'$ with $l \in \delta(l')$. To denote the set of all nested locations of a superstate $l$, including $l$, $\delta^*(l)$ is used. $\delta^{-*}$ is the set of ancestors of $l$, including $l$. Moreover we use $\overline{\delta}(l) = \delta^*(l) \setminus \{ l \}$. We introduce $\overline{\delta}$ to refer to the children, that are proper locations: $\overline{\delta}(l) = \{ n \in \delta(l) \mid BASIC(n) \lor XOR(n) \lor AND(n) \}$. We use $V^+(l)$ to denote the variables in the scope of location $l$: $V^+(l) = \bigcup_{n \in \delta^{-*}(l)} V(l)$. $C^+(l)$ and $Ch^+(l)$ are defined analogously.

We further require the following rules to ensure consistency of a given Hierarchical Timed Automaton.

**Definition 11** An Extended Hierarchical Timed Automaton is well-formed iff
1. The function $\delta$ has to give rise to a proper tree rooted at root, and $S = \delta^*(\text{root})$. Basic nodes have to be empty ($\text{BASIC}(l) \Leftrightarrow \delta(l) = \emptyset$), and invariants of pseudo-locations are trivial ($\text{ENTRY}(l) \lor \text{EXIT}(l) \Rightarrow \text{Inv}(l) = \text{true}$).

Additionally, to have a well-formed hierarchical state we require for all non BASIC locations to hold that all substates are valid refinement ($\neg\text{BASIC}(l) \Leftrightarrow \forall l' \in \delta(l) (\text{Inv}(l') \Rightarrow \text{Inv}(l))$). Note, that for $\text{Inv}(l') = \text{true}$ this condition does always hold.

2. $S_0$ corresponds to a consistent and proper control situation, i.e. root $\in S_0$ and for every $l \in S_0$ the following holds that it is a proper location ($\text{BASIC}(l) \lor \text{XOR}(l) \lor \text{AND}(l)$), $S_0$ does contain all parents ($l = \text{root} \lor \delta^{-1}(l) \in S_0$), only one alternative of an XOR-state is present ($\text{XOR}(l) \Rightarrow |\delta(l) \cap S_0| = 1$), and all substates of an AND-state are proper states ($\text{AND}(l) \Rightarrow \delta(l) \cap S_0 = \delta(l)$).

3. Conflicts in data model updates in synchronizing transitions and in general are disallowed by demanding that for any two locations $l_1$ and $l_2$ which can occur in parallel and any $l_1 \stackrel{g,c,r,u}{\longrightarrow} l'_1$ and $l_2 \stackrel{g',c',r',u'}{\longrightarrow} l'_2$ will hold $\neg\text{conflict}(g,u')$, $\neg\text{conflict}(u,g')$, and $\neg\text{conflict}(u,u')$.

4. Conflicts in assignments in synchronizing transitions are explicitly excluded by demanding that $l_1 \stackrel{g,c,r,u}{\longrightarrow} l'_1$, $l_2 \stackrel{g',c',r',u'}{\longrightarrow} l'_2 \in T$ follows $\text{vars}(r) \cap \text{vars}(r') = \emptyset$, where $\text{vars}(r)$ is the set of integer variables occurring in $r$. We require an analogous constraint to hold for the pseudo-transitions originating in the entry of an AND supertate. For $l \stackrel{g,c,r-u}{\longrightarrow} l' \in T$, $g,r$ has to be defined over $V^+(\delta^{-1}(l)) \cup C^+(\delta^{-1}(l))$ and $s$ is defined over $\text{Ch}^+(\delta^{-1}(l))$ to ensure a proper static scoping.

5. Let $e \in S, \text{ENTRY}(e)$. If $\text{XOR}(\delta^{-1}(l))$, then $T$ contains exactly one transition $e \longrightarrow l'$.

If $\text{AND}(\delta^{-1}(l))$, then $T$ contains exactly one transition $e \longrightarrow e_i$ for every proper substate $l_i \in \delta(\delta^{-1}(l))$, and $e_i \in \delta(l_i)$. In case of $\text{HISTORY}(e)$, outgoing transitions declare the default history locations. If a supertate $s$ has a history entry, then every substate $l$ of $s$ has to provide either a history entry or a default entry.

6. Transitions have to respect the structure given in $\delta$ and cannot cross levels in the hierarchy, except via connecting to entries or exits. Note that transitions cannot lead directly from entries to exits. Transitions $l \stackrel{g,c,r-u}{\longrightarrow} l'$ with $\text{ENTRY}(l)$ or $\text{EXIT}(l')$ are called pseudo-transitions. They are restricted in the sense, that they cannot carry synchronizations or urgency flags, and only either guards or assignments. For $\text{ENTRY}(l)$, only pseudo-transition of the form $l \stackrel{g,c,r-u}{\longrightarrow} l'$ are allowed. For $\text{EXIT}(l')$, only pseudo-transition of the form $l \stackrel{g,c,r-u}{\longrightarrow} l'$ are allowed. For $\text{EXIT}(l) \land \text{EXIT}(l')$, this is further restricted to be of the form $l \rightarrow l'$.

### 5.3 Semantics

The semantics of the Extended Hierarchical Timed Automaton model is further defined by using a configuration of the form $(\rho, \mu, \nu, \phi, \theta)$ which contains the active locations, the integer variable values, the clock values, asynchronous event queues, and the history.

- $\rho : S \rightarrow 2^S$ captures the control situation. $\rho$ can be understood as a partial, dynamic version of $\delta$, that maps every supertate $s$ to the set of active substates. If a supertate $s$ is not active, $\rho(s) = \emptyset$. We define $\text{Active}(l) = l \in \rho^\times(\text{root})$, where $\rho^\times(l)$ is the set of all active substates of $l$. Note that $\text{Active}(l) \Leftrightarrow l \in \rho(\delta^{-1}(l))$.

- $\mu : S \rightarrow (\mathbb{Z})^\star$. $\mu$ gives the valuation of the local integer variables of a supertate $l$ as a finite tuple of integer numbers. If $\neg\text{Active}(l)$ then $\mu(l) = \lambda$ (the empty tuple). If $\text{Active}(l)$ then we require that $|\mu(l)| = |V(l)|$ and $\mu$ are consistent with respect to the values of the
shared variables (i.e. always maps to the same value). We use \( \mu(l)(a) \) to denote the value of \( a \in V(l) \). When entering a non-basic location, local variables are added to \( \mu \) and set to an initial value (0 by default). We use the shorthand \( 0^{|V(l)|}l \) or the tuple \( (0, 0, ..., 0) \) with arity \( |V(l)| \).

- \( \nu : S \to (\mathbb{R}^+)^* \). \( \nu \) gives the real valuation of the clocks \( C(l) \) visible at location \( l \), thus \( |\nu(l)| = |C(l)| \). If \( \neg \text{Active}(l) \) then \( \nu(l) = \lambda \).

- \( \phi : [\text{Ch}_a \to \mathbb{N}] \) reflects the asynchronous event channels, that have to be stored. Channels are assumed to be persistent and thus local channels are not cleared when leaving or entering a location. By default all events are also preserved until explicitly consumed.

- \( \theta \) reflects the history, that might be restored by entering superstates via history entries. It is split up in the two functions \( \theta_{\text{state}} \) and \( \theta_{\text{var}} \), where \( \theta_{\text{state}}(l) \) returns the last visited substate of \( l \) – or an entry of the substates, in the case where the substate is not basic – (to restore \( \rho(l) \)), and \( \theta_{\text{var}}(l) \) returns a vector of values for the local integer variables. There is no history for clocks at the semantics level, all non-forgetful clocks belong to \( C(\text{root}) \).

To handle history entries the predicate \( \text{HasHistory}(l) = \exists n \in \delta(l)|\text{HISTORY}(n) \) is used. If \( \text{HasHistory}(l) \) holds, the term \( HEntry(l) \) denotes the unique history entry of \( l \). If \( \text{HasHistory}(l) \) does not hold, the term \( HEntry(l) \) denotes the default entry of \( l \). If \( l \) is basic \( HEntry(l) = l \). If none of the above is the case, then \( HEntry(l) \) is undefined. Initially, it holds \( \forall l \in S|\text{HasHistory}(l) \Rightarrow \theta_{\text{state}}(l) = HEntry(l) \land \theta_{\text{var}}(l) = 0^{|V(l)|} \).

In order to denote the set of locations reached by following a fork, the function \( \text{Targets}_\theta : 2^S \to 2^S \) relative to \( \theta \) is used.

\[
\text{Targets}_\theta(l) = L \cup \bigcup_{l \leq l'} \left( \{ n | n \in \theta_{\text{state}}(l) \land \text{HISTORY}(l) \} \cup \{ n | l \xrightarrow{r} n \land \text{ENTRY}(l) \} \right)
\]

The notation \( \text{Targets}_\theta(l) \) for \( \text{Targets}_\theta(\{l\}) \) is used, if the argument is a singleton. \( \text{Targets}_\theta^* \) is the reflexive transitive closure of \( \text{Targets}_\theta \).

Firing a transition \( t : l \xrightarrow{g \cdot s \cdot r \cdot n} l' \) entails in general 1. executing a join to exit \( l \), 2. taking the proper transition \( t \) itself, and 3. executing a fork at \( l \). If \( l \) (respectively \( l' \)) is a basic location, part 1. (respectively 3.) is trivial. Together, this defines a run-to-completion step. We represent a run-to-completion step formally by a transformation function \( \Gamma_t \), which depends on a particular transition \( t \). The three parts of this step are described as follows.

1. join: \( (\rho, \mu, \nu, \phi, \theta) \) is transformed to \( (\rho^1, \mu^1, \nu^1, \phi^1, \theta^1) \) as follows: \( \rho \) is updated to \( \rho^1 \xleftarrow{\text{join}} \rho|n \in \rho^x(l) \land \rho^x(l)[n \rightarrow \emptyset] \), \( \mu \) is updated to \( \mu^1 \xleftarrow{\text{join}} \mu|\forall n \in \rho^x(l) \land \rho^x(l)[n \rightarrow \lambda] \), \( \nu \) is updated to \( \nu^1 \xleftarrow{\text{join}} \nu|\forall n \in \rho^x(l) \land \nu^1(n) \rightarrow \lambda \). The asynchronous channels are simply preserved (\( \phi^1 = \phi \)). If \( \text{EXIT}(l) \), the history is recorded. Let \( H \) be the set of superstates \( h \in \rho^x(\delta^{-1}(l)) \), where \( \text{HasHistory}(h) \) holds. Then \( \theta^1_{\text{state}} := \theta_{\text{state}} \land \forall h \in H \Rightarrow HEntry(\rho(h)) \) and \( \theta^1_{\text{var}} := \theta_{\text{var}} \land \forall h \in H \Rightarrow \mu(h) \). If \( \neg \text{EXIT}(l) \) or \( H = \emptyset \), then \( \theta^1 := \theta \).

2. \( (\rho^2, \mu^2, \nu^2, \phi^2, \theta^2) := (\rho^1[l'/l], r(\mu^1), r(\nu^1), \phi^1 - s_2 + s_3, \theta^1) \) is computed with \( s = (s_1, s_2, s_3) \). \( r(\mu^1) \) denotes the updated values of the integers after the assignments, \( r(\nu^1) \) the updated clocks after the resets, and \( \phi^1 - s_2 + s_3 \) is computed by multi-set subtraction of the asynchronous receive part of \( s \) and addition of the asynchronous send part of \( s \).

3. fork: \( (\rho^3, \mu^3, \nu^3, \phi^3, \theta^3) \) is transformed to \( (\rho^3, \mu^3, \nu^3, \phi^3, \theta^3) \) by moving the control to all proper locations reached by the fork, i.e. those in \( \text{Targets}_{\text{fork}}(l') \). Note that \( \rho^3(n) = \emptyset \) for all \( n \in \delta^x(l') \). Thus we can compute \( \rho^3 \) from \( \rho^2 \) by modifying it iteratively for all \( n \in \text{Targets}_{\text{fork}}(l') \) for \( \text{ENTRY}(n) \) such that \( \rho^3(\delta^{-2}(n)) := \rho^3(\delta^{-2}(n)) \cup \{ \delta^{-1}(n) \} \) or
else \( \rho^3(\delta^{-1}(n)) := \{n\} \). \( \phi^3 \) is identical to \( \phi^2 \) and \( \mu^3 \) is derived from \( \mu^2 \) by first initializing all local variables of the superstates \( s \) in \( \text{Targets}_{\mu^2}(l') \), i.e. \( \mu^3(V(s)) := 0^V(s) \). If \( \text{HasHistory}(s), \theta_{\text{var}}(s) \) is used instead of \( 0^V(s) \). Then all variable assignments and clock-resets along the pseudo-transitions belonging to this fork are executed to update \( \mu^3 \) and \( \nu^3 \). The history does not change \( \theta^3 \) identical to \( \theta^3 \).

Note that parts 1. and 3. correspond to the identity transformation, if \( l \) and \( l' \) are basic locations. The resulting configuration vector transformation \( \Gamma_t \) for a transition \( t : l \stackrel{g,s,r,m}{\longrightarrow} l' \) is defined accordingly as the result of these three steps as \( \Gamma_t(\rho, \mu, \nu, \theta) = (\rho^3, \mu^3, \nu^3, \phi^3, \theta^3) \). If the context is unambiguous, we use \( \rho^{l'} \) and \( \nu^{l'} \) for the parts \( \rho^3 \) respectively \( \nu^3 \) of the transformed configuration corresponding to transition \( t \).

A superstate \( s \) can only be exited, if all its parallel substates can synchronize on this exit. For an exit \( l \in \delta(s) \) we recursively define the family of sets of exits \( \text{PreExitSets}(l) \). Each element \( X \) of \( \text{PreExitSets}(l) \) is itself a set of exits. If transitions are enabled to all exits in \( X \), then all substates can synchronize.

\[
\text{PreExitSets}(l) = \begin{cases} 
\bigcup_{n_1, \ldots, n_k \in \text{Basic}(l)} \text{PreExitSets}(n_i), & \text{if EXIT}(l) \land \text{AND} (\delta(l)) \\
\{l\} \cup \bigcup_{m \in \delta(l) \land \text{PreExitSets}(m)} \text{PreExitSets}(m), & \text{if EXIT}(l) \land \text{XOR} (\delta(l)) \\
\{\}, & \text{if BASIC}(l)
\end{cases}
\]

Here, the operator \( \oplus \) is a product over families of sets, i.e. it maps \( \{A_1, \ldots, A_a\}, \{B_1, \ldots, B_b\} \) to \( \{A_1 \cup B_1, A_1 \cup B_2, \ldots, A_a \cup B_b\} \) and is extended to operate on an arbitrary finite number of arguments in the obvious way.

To give the transition system rules for \( \longrightarrow \), predicates that evaluate conditions on the dynamic \( \rho \) are further required. The set of active leaves (in the tree described by \( \rho \)), which are the innermost active states in a superstate \( l \), are defined by: \( \text{Leaves}(\rho, l) = \{n \in \rho^X(l) | \rho(n) = \emptyset\} \).

The predicate expressing that all the substates of a state \( l \) can synchronize on a join is:

\[
\text{JoinEnabled}(\rho, \mu, \nu, l) = \text{BASIC}(l) \lor \exists X \in \text{PreExitSets}(l) : \forall n \in \text{Leaves}(\rho, l) : \exists n' \in X, n \longrightarrow n' \land g(\mu, \nu)
\]

Note that \( \text{JoinEnabled} \) is trivially true for a basic location \( l \).

The invariants of a location are denoted by \( \text{Inv}_\nu : S \rightarrow \{\text{true, false}\} \), which evaluates the invariant of a given location with respect to a clock evaluation \( \nu \) and the predicate \( \text{Inv}(\rho, \nu) \) defined by \( \text{Inv}(\rho, \nu) = \bigwedge_{n \in \rho^X(\text{root})} \text{Inv}_\nu(n) \) is used to express, that for control situation \( \rho \) and clock valuation \( \nu \) all invariants are satisfied. The predicate \( \text{Enabled} \) over transitions \( t : l \stackrel{g,s,r,m}{\longrightarrow} l' \) is used to describe whether \( t \) is enabled:

\[
\text{Enabled}(t : l \stackrel{g,s_1,s_2,s_3,r,u}{\longrightarrow} l', \mu, \nu, \phi) := \neg \text{EXIT}(l') \land g(\mu, \nu, \phi) \land (s_2 \leq \phi) \land \text{JoinEnabled}(\rho, \mu, \nu, l) \land \text{Inv}(\rho^{l'}, \nu^{l'})
\]

Since urgency has precedence over delay, we have to capture the global situation, where an urgent transition is enabled. We do this via the predicate \( \text{UrgentEnabled} \) over a configuration without
taking the history into account.

\[
\text{UrgentEnabled}(\rho, \mu, \nu, \phi) = \left( \exists t : l. \text{Enabled}(t, \rho, \mu, \nu, \phi) \land l' \land p > 0 \land \text{Enabled}(t, \rho, \mu, \nu, \phi) \right) \lor \\
\left( \exists t_1 : l_1. \text{Enabled}(t_1, \rho, \mu, \nu, \phi) \land l_1' \land p_1 > 0 \land \text{Enabled}(t_1, \rho, \mu, \nu, \phi) \right)
\]

We further define for each priority \( p \) its own transition relation \( \xrightarrow{\text{trans}_p} \). For a transition \( t \in T \) (action), the elapsing of time (delay), and synchronization (sync) we have:

\[
\frac{\text{Enabled}(t : l. s.(r, s_1, s_2), l', \rho, \mu, \nu, \phi)}{(\rho, \mu, \nu, \phi, \theta) \xrightarrow{\text{action}} l'}
\]

Here \( g \) is the guard of the transition and \( r \) the set of resets and assignments. The urgency flag \( u \) has no effect here. This rule applies for action transitions between basic locations as well as superstates. In the latter case, this includes the appropriate joins and/or fork operations.

\[
\forall d' 0 \leq d' < d : \neg \text{Inv(l)}(\rho, \nu + d') \xrightarrow{\text{delay}} \neg \text{UrgentEnabled}(\rho, \mu, \nu + d', \phi)
\]

In the delay transition rule (delay) \( \nu + d \) stands for the current clock assignment plus the delay for all the clocks. Time elapses in a configuration only when all invariants are satisfied and there is no urgent transition enabled. Note, that we also have to take possible time deltas \( d' \) into account, to ensure that no urgent transition has become enabled before the time delta \( d' \) has been elapsed.

\[
\frac{\text{Enabled}(t_1 : l_1. s.(r, s_1, s_2), l_1', \rho, \mu, \nu, \phi) \quad l_1' \notin \delta^+ (l_2) 
\text{AND} \quad c \in \text{Ch}_s}{(\rho, \mu, \nu, \phi, \theta) \xrightarrow{\text{sync}} l_1' \land l_2' \land \text{Enabled}(t_2 : l_2. g.(r, s_3, s_4), l_2', \rho, \mu, \nu, \phi) 
\text{AND} \quad l_2' \notin \delta^+ (l_1)}}
\]

Since priority can result in precedence between conflicting transitions, we have to formalize what conflict means. Therefore, the structural relation between the source states reflected by their common root are important.

\[
\text{CommonRoot}(l_1, l_2) := l_c \text{ with } l_c \in S' \land \delta(l_c) \notin S' \text{ for } S' = \delta^S(l_1) \cap \delta^S(l_2).
\]

For all location pairs such a common root exists and the two related transitions are in conflict iff it is not a parallel AND-state that permits to execute them independently.

\[
\text{ConflictingTransition}(t_1 : l_1. \text{Enabled}(t_1, \rho, \mu, \nu, \phi, \theta) \quad l_1' \notin \delta^+ (l_2) 
\text{AND} \quad c \in \text{Ch}_s}{(\rho, \mu, \nu, \phi, \theta) \xrightarrow{\text{sync}} l_1' \land l_2' \land \text{Enabled}(t_2 : l_2. g.(r, s_3, s_4), l_2', \rho, \mu, \nu, \phi) 
\text{AND} \quad l_2' \notin \delta^+ (l_1)}}
\]

The notion of conflicting transitions can be further extended to all possible priority specific transition labels \( \alpha, \beta \in T \cup T^2 \cup R^+ \). For \( \text{Trans}(\alpha) \) defined as \( \emptyset \) for \( \alpha \in N \), \( \alpha \) for \( \alpha \in T \), and \{ \( t_1, t_2 \) \} for \( \alpha = (t_1, t_2) \in T^2 \) we have

\[
\text{Conflicting}(\alpha, \beta) := \bigvee_{t_\alpha \in \text{Trans}(\alpha), t_\beta \in \text{Trans}(\beta)} \text{ConflictingTransition}(t_\alpha, t_\beta).
\]

The resulting rule for the transition system steps \( \xrightarrow{} \) for \( \alpha, \beta \) elements of \( T \cup T^2 \cup R^+ \) is then:

\[
\frac{\text{Enabled}(t : l. s.(r, s_1, s_2), l', \rho, \mu, \nu, \phi, \theta) \quad l' \notin \delta^+ (l) \text{ AND} \quad \forall \beta, \Delta > 0 \left( \text{Enabled}(t, \rho, \mu, \nu, \phi, \theta) \xrightarrow{\text{prio}} \neg \text{Conflicting}(\alpha, \beta) \right)}{(\rho, \mu, \nu, \phi, \theta) \xrightarrow{\text{prio}} (\rho', \mu', \nu, \phi', \theta')}\]
5.4 Summary

The model of the Extended Hierarchical Timed Automata covers –like the Hierarchical Timed Automata– the requirements temporal behavior, predictability and expressiveness. Furthermore the data model has been extended, so that it is possible to model complex data operations. In addition Extended Hierarchical Timed Automata support synchronous events as well as asynchronous and associate priorities with transitions in order to achieve determinism on the modelling level.

Thus Extended Hierarchical Timed Automata allow the modeling of different kinds of behaviors. So it is possible to specify limited behavior as well as unlimited behavior (cf. Section 2.1). To ensure, that a specification is only limited and thus can be implemented on a real physical controller, another model is required. Such a model is presented in the next section.

6 Real-Time Statecharts

In [Bur02] a Statechart notion named Real-Time Statecharts has been developed, which permits to model only limited behavior as defined in Section 2.1. Also code generation for the Java Real-Time platform has been realized as an extension of the Fujaba UML CASE tool.\(^4\)

6.1 Extension

In order to specify complex temporal behavior, the advantages of modelling complex behavior with Statecharts is combined with the advantage of specifying temporal behavior with ExHTA, resulting in the real-time extension of Statecharts.

Unit: msec

<table>
<thead>
<tr>
<th>State S₁</th>
<th>t₀ ≤ 5</th>
<th>e [x ≤ 2] [1 ≤ t₀] /action() w = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>entry:</td>
<td>entryS₁(); w = 1 {t₀}</td>
<td></td>
</tr>
<tr>
<td>do:</td>
<td>doS₁(); w = 1; p ∈ [2; 3]</td>
<td></td>
</tr>
<tr>
<td>exit:</td>
<td>exitS₁(); w = 1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State S₂</th>
<th>t₀ ≤ 20 ∧ t₁ ≤ 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>t₁ ∈ [3; 6]</td>
<td>[0; 10]</td>
</tr>
<tr>
<td>exit:</td>
<td>{t₀, t₁}</td>
</tr>
</tbody>
</table>

Figure 1: Real-Time Statechart

Figure 1 depicts a (part of a) Real-Time Statechart, consisting of 2 states and 1 transition. The state S₁ has the time invariant t₀ ≤ 5msec, S₂’s invariant is t₀ ≤ 20msec ∧ t₁ ≤ 13msec (t₀ and t₁ denote clocks). The transition is triggered by the event e, the guard [x ≤ 2] and by the time guard [1msec ≤ t₀]. When firing t₂ is reset and action(), which has the WCET w = 2msec, is executed. The firing of the transition has to be finished at the latest 10msec after being triggered and at t₁ == 6. The earliest point in time, when S₂ may be entered is at t₁ == 3. Worst-case execution times are attached to the entry()-, exit()- and do()-operations, too, clock can be reset when entering and exiting the states, too. As the do()-operation is executed periodically, an interval is specified from where a fixed period is chosen.

Giving a compact overview, Real-Time Statecharts offer the following constructs for states: Hi-

\(^4\)www.fujaba.de
erarchy, parallelism, and time invariants. Transitions make use of the constructs: Events, guards, time guards, resets, priorities, and synchronization.

It is shown, that they can be mapped to a corresponding ExHTA. The semantics of the Real-Time Statecharts is defined by these mapping rules.

### 6.2 Semantics

Before outlining the semantics by showing the mapping rules, we will introduce an abbreviation. Figure 2 a) shows an ExHTA with an urgent transition from location $S_1$ to $S_2$ and a non-urgent transition from $S_2$ to $S_3$. $S_2$ is a hierarchical state with the invariant $t_0 \leq 10$ and the substates $S_2^1$, $S_2^2$, and $S_2^3$ which have the same invariant (in this example); $\text{action}_1()$ and $\text{action}_2()$ are resets. Such a hierarchical state can be abbreviated as shown in Figure 2 b). There, the sequential executed resets $\text{action}_1()$ and $\text{action}_2()$ are combined to $\text{action}()$ and the substates disappear.

Unit: msec

![Diagram](attachment:figure2.png)

Figure 2: Extended Hierarchical Timed Automata

Using this abbreviation, a Real-Time Statechart can be mapped to an ExHTA –as depicted in Figures 3 and 4. Figure 3 shows how non-hierarchical states are mapped to locations: The name as well as the invariant are held. The $\text{entry}()$- and $\text{exit}()$-operations and the sets of clock-resets will be rolled out of the states. To model the periodic execution of the $\text{do}()$-operation, a transition with its specific state as source and target will be created. This transition resets the new introduced clock $t_{int}$ and is only triggered at $t_{int} = p$, where $p \in [p_{low}, p_{up}]$.

When firing a transition, the $\text{exit}()$-operation of the source state, the data assignment and the target state’s $\text{entry}()$-method are executed sequentially. This is modelled in the ExHTA by putting
Realtime Statechart

\[
\begin{align*}
S_1 & \quad \text{inv} \\
& \quad \text{entry()} \ CR_{\text{entry}} \\
& \quad \text{do()} \ p \in [p_{\text{low}}; p_{\text{up}}] \\
& \quad \text{exit()} \ CR_{\text{exit}} \\
\end{align*}
\]

Extended Hierarchical Timed Automata

\[
\begin{align*}
S_1 & \quad \text{inv} \\
& \quad \{t_{\text{int}}\} \\
& \quad \text{do()} \\
& \quad p \leq t_{\text{int}} \leq p \\
\end{align*}
\]

Figure 3: Mapping of non-hierarchical states

Realtime Statechart

\[
\begin{align*}
& \quad e \ g \ tg \ \text{prio} \\
& \quad \text{exit}_{\text{src}}(\text{a}) \ CR \ [d_{\text{low}}; d_{\text{up}}] \\
& \quad \prod_i (t_i \in [d_{\text{low}}^i; d_{\text{up}}^i]) \\
\end{align*}
\]

ExHTA

\[
\begin{align*}
& \quad e \ g \ \text{prio} \\
& \quad \text{tg} \ \text{CR} \cup \{t_{\text{int}}\} \cup CR_{\text{exit}_{\text{src}}} \\
& \quad w_{\text{exit}_{\text{src}}} + w_a + w_{\text{entry}_{\text{tgt}}} \\
& \quad (t_{\text{int}} \leq d_{\text{ap}}) \land \prod_i (t_i \leq d_{\text{up}}^i) \\
\end{align*}
\]

Figure 4: Mapping of transitions
them in a hierarchical state –similar to Figure 2– and using the abbreviation, introduced above. This leads to 2 different kinds of locations: *State locations*, representing states and *action locations* in which operations are executed. The WCET of these 3 actions is the sum of their single WCETs. In Figure 4 is shown, that a transition is mapped to an actionlocation and 2 transitions (an urgent and a non-urgent one), connecting it with the source and the target state. The event, guard, time-guard and priority from the Real-Time Statechart is recovered in the urgent transition as these attributes shall trigger the actions. The deadline is recovered at multiple places. On the one hand it is used to extend the invariant of the actionlocation, on the other hand it is used to prevent the actionlocation to be left before it may be left, specified by the lower bound of the deadline. As the last point of time, when the actionlocation may be left, is specified by the upper bound of the deadline, it is used for the upper bound of the timeguard of the non-urgent transition.

The constructs of parallelity, hierarchie and history can be adopted without changes, because Real-Time Statecharts offer them as well as ExHTA.

### 6.3 Realizability

To achieve a mapping from the model to real physical threads, deadlines and WCETs are required for every transition. These deadlines specify the point in time, when the execution has to be ended relative to the moment of triggering or relative to one or multiple clocks. Thus, the modeller has to respect, that the firing of a transition (respectively the execution of resets and data assignments) is not infinitely fast and needs time to be executed. Usually, the WCETs are dependant on the target platform. Note that the WCET of an operation needs to be less than its deadline. The timedifference between WCET and deadline respects that other tasks require resources simultaneously and that on real physical devices a triggered transition is usually recognized with a delay. This delay is compensated by that timedifference. Thus, the designation of WCETs and deadlines permits the mapping of the model to real physical threads that can be executed on a (limited) physical digital controller. These threads can be implemented in every programming language, supporting real-time functionality. The CASE-Tool Fujaba\(^5\) delivers an implementation of Real-Time Statecharts for Real-Time Java.

When multiple Real-Time Statecharts are executed in parallel on the same processor, WCETs and deadlines are used to perform a schedulability analysis before code is generated.

The specification of invariants, deadlines, guards and WCETs potentiates to model time inconsistencies, e.g. an invariant of a state expires before any transition, leaving from this state, is triggered. By analysing the structure of the Real-Time Statechart and the coherency of the temporal annotations, some conflicts and inconsistencies can be detected with an appropriate complexity (e.g. no state space exploration, usually caused by model checking tools). For cases where this analysis is not capable to determine whether a timing fault is present or not, model checking should be employed.

### 6.4 Summary

With the model of Real-Time Statecharts it is possible to specify complex behavior by the use of hierarchy and parallelism as well as temporal behavior. Additionally, it is possible to specify operations executed at the moment of entrance or exit of a state (*entry()*/*exit()*-operation) or being periodically invoked during residence in a state (*do()*-operation). Real-Time Statecharts fulfill all requirements pointed out in Section 2 and ensure that only limited behavior is specified that is realizable.

\(^5\)www.fujaba.de
7 Related Work

Harel introduced in [Har98] 2 extensions of Statecharts to bring in time constraints: Using the \emph{timeout-construct} an event $e$ and a time $t$ is associated with a transition. The transition is triggered $t$ time-steps after the occurrence of the event $e$. If a reset or data assignment is marked as \emph{scheduled action}, it is not executed during firing a transition, but a specified time later (although the state is changed immediately). In [KP92] an extension is described, that allows to specify a kind of delay interval for a transition. The bounds of this interval define how much time elapses at least and at most between triggering and firing the transition. In [PAS94] a similar approach is used. There, the interval describes the points in time, when the source state can be left earliest and when it has to be left at the latest. In addition, every event obtains a duration, indicating how long it can be consumed after creation. All the constructs, described above, are useful to specify temporal behavior for unlimited systems. To map them on man made controllers, more informations are needed (e.g. WCETs, deadlines, etc.).

In [SKW00] the model of an automatom is mapped to a so called \emph{multi-tasking model}. The multitasking model describes how the actions of the model are placed in physical threads. The annotation of WCETs and deadlines in the model makes this mapping possible. The automaton offers communication via synchronous and asynchronous events. Compared to high-level automata, it is restricted, as it is not possible to use hierarchy, parallelism or to trigger transitions dependant on clocks or variables.

[LQV01] deals with validation of real-time models. UML (in particular class-diagrams and extended Statecharts) is used to specify the structure and behavior, based on time. As no WCETs and deadlines are required and time comes in just by the use of the after-construct, the problems mentioned above, occure here, too. The claimed invariants of the system behavior are expressed by OCL constraints. The specified model can be translated into the first order temporal logic language \emph{TRIO}, that can be used as input for the modelchecker TRIO-Matic in order to validate the properties of the model. To formalize the semantics of the extended Statechart model, it is mapped to an extension of Timed Statecharts in [BLM02].

In [LC96] \emph{Duration Calculus} – an extended first order logic language – is used to extend Statecharts with duration. Thus, there exist a couple of functions determining if and for which time states are active within a specific timeinterval. This framework, expressing temporal requirements as functions, was regarded as breakthrough in validation of not only functional but also temporal correctness. Another approach for verifying real-time systems is given in [KMR02], where behavior is specified by UML Timed State Machines, that are mapped to Timed Automata, serving as input for the Modelchecker \emph{UPPAAL} [DMY02]. Also in this work, it is assumed that the Timed State Machine runs on an unlimited, infinite fast device. Anyhow, it is assumed, that the Statechart is distributed and it is possible to specify an upper bound for the delay of communication. In addition to the Timed State Machine, so called \emph{collaborations} are specified by the use of sequence diagrams. These sequence diagrams are transformed to the UPPAAL conformant Observer Timed Automata. After that UPPAAL is used to verify the Timed Automaton against the Observer Timed Automaton.

As model checking requires a formal semantics, that is not provided by UML, [EW00] delivers another formal semantics for Statecharts at the requirements-level, orientated at the StateMate semantics. In the paper, a comparison is given of structured and object oriented Statechart semantics as well as a comparison of Statecharts at the requirements- and at the implementation-level. The main aspect describes the communication between several Statecharts.

All this work and the extensions are useful for verification and validation of systems, whose behavior is based on time. As none of the models have been designed with the aim of generating code, they do not provide solutions appearing in this domain.
8 Conclusion

For the proposed notion of Extended Hierarchical Timed Automata the required modeling notation properties expressiveness, temporal behavior and predictability follows from the extension of the hierarchical Timed Automata concepts. While the proposed extension of priorities limit at least the ability to employ model checking the resulting extended hierarchical state machines, integrating an arbitrary complex data, permit to even check an finite abstraction of real models.

The presented formal extension of the hierarchical Timed Automata has been employed in order to define the semantics for Real-Time Statecharts [Bur02]. Each Real-Time Statechart is mapped to a limited Extended Hierarchical Timed Automaton which is bi-partite with respect to do and activity states. The realizability of consistent models of this limited notion of Extended Hierarchical Timed Automata has been shown in Section 6. The presented notion therefore supports all four identified requirements and thus better serves the required modeling of complex reactive real-time systems.

Acknowledgements

The authors thank Martin Hirsch and Matthias Meyer for comments on earlier versions of the report.

References


REFERENCES


