A Real-time Profile for UML and how to adapt it to SDL*

Susanne Graf and Ileana Ober

VERIMAG ** – Centre Equation – 2, avenue de Vignate – F-38610 Gières – France
http://www-verimag.imag.fr/{graf,iober}

Abstract. This paper presents work of the IST project OMEGA, where we have defined a UML profile for real-time compatible with the “Profile for Performance, Scheduling and Real-time” accepted recently at OMG.

In contrast to this OMG profile, we put emphasis on semantics and on its use in the context of timed analysis of real-time embedded systems. The defined profile is compatible with the time concepts existing in SDL, and we show how we can adapt also those notations to SDL and MSC which do not yet exist in these ITU languages.

1 Introduction

Today’s embedded applications, have often strong constraints with respect to deadlines, response time and other non-functional aspects¹. They may be distributed and run on different execution platforms and this influences strongly the non functional characteristics. In order to allow early analysis of timing properties, we propose to lift the choices and constraints coming from a given run-time system, as well as assumptions on the environment, to the abstract model level in terms of a set of annotations which can be interpreted by timed validation tools. In order to make this approach feasible, the annotation language must be flexible and orthogonal to the functional description. It must allow the expression of global constraints and their refinement into a set of local constraints.

The aim:

Our aim is not only to provide means for describing systems with time dependent choices, but also for defining timed models of systems, including time related assumptions on the environment, e.g. on the arrival time of inputs of the environment and (platform dependent) durations of tasks. Thus, the idea is to perform validation (simulation, symbolic test or verification) on a timed model, built from knowledge or assumptions on the environment and the platform including aspects like resource sharing and scheduling (which have to be confirmed later by testing, run-time monitoring or code analysis). At requirement level, this

---

* This work has been partially supported by the IST-2002-33522 OMEGA project
** VERIMAG is a research laboratory of CNRS, Université Joseph Fourier and Institut National Polytechnique de Grenoble
¹ We restrict ourselves here exclusively to timing aspects
approach can be used to validate the consistency of a set of requirements and to build the loosest timed model satisfying all requirements, which than can be further refined compositionally.²

Overview on existing approaches:

Modelling languages and tools used in the domain of embedded and real-time software, usually do have time related concepts, but in general, they are not sufficient for timed analysis at model level. The existence of analysis tools is crucial in this context, but presently only partially covered.

Languages using an asynchronous event driven approach, like SDL [IT00a], ROOM [SGW94], State charts [Har87] and others, use a notion of global and external time. In SDL and ROOM, global time is available within the system through a special expression, and can be stored in variables or sent as parameters of messages. The Unified Modeling Language UML [UML99] does not state its position concerning the semantics of time progress, with the aim of being flexible enough to accommodate all possible approaches.

The development tools existing for SDL, [Tel99a,Tel99b], allow code generation and some restricted form of analysis. Properties are expressed in terms of MSC [IT00b] or by means of a tool internal observer language. These tools define a particular time model and special profiles for limited performance and timing analysis based on simulation of scenarios. Most approaches for enriching SDL with time constraints and providing tools for performance or timing analysis are scenario based [Leu95,BAL97]. Also, an SDL methodology for the development of real-time systems [ADL01+01] and on the extension of SDL with time constraints for timing and performance analysis [MTMC99,BGM+01] have been studied.

Although standard UML does not include any particular time framework, a number of tool supported frameworks have been proposed. For example, Rose RT (a variant of the ROOM framework), the Real-Time Studio of Artisan [Art01], Rhapsody [Ilc] and, more recently, the tool TAU Generation 2 [Tel02] propose UML frameworks including real time aspects. These tools allow automatic or semi-automatic code generation and have facilities for simulation and restricted functional analysis but almost no support for timing analysis. For instance, for Rhapsody exist validation tools [BDW00], but only in the context of a rather deterministic, external time semantics and not for time dependent properties.

Recently, a UML Profile for Schedulability, Performance and Time Specifications [OMG02] has been defined. It integrates the ideas and concepts from most previously named approaches. It is very general, in order to be able to adapt to any possible real-time framework and for all kind of diagrams, and it defines essentially a vocabulary. It is very much tailored towards timed scenarios, and for the moment, it exists essentially on paper.

Timed automata [HNSY92,BST98b] have been used for modelling real-time aspects of systems, for defining semantics of modelling languages and for studying controller synthesis [MPS95,AMP95] and scheduling frameworks [BGSO00][HHIK01]. A number of validation and analysis tools, such as Hytech [HHWT97],

² the traffic light presented in section 4 is an instance of this type of use
Kronos/IF [Yov97,BGGM00,BGM02] or Uppaal/Time [LP Y97,FPY02], exist for the framework of timed automata (extended with data), but they are in general not closely coupled with development environments.

In the OMEGA IST project\(^3\), aiming at the definition of a development methodology for embedded and real-time systems in UML based on formal techniques, we have started to refine the UML Profile for Schedulability, Performance and Time in order to make it usable for efficient analysis. In this article, we report on this work and show how it can be adapted for the definition of a real-time profile for SDL which is compatible with the already existing time related features of SDL and MSC.

We start, in Section 2, with an overview of the concepts necessary in a real-time framework. In Section 3.1, we define the basic time concepts of the UML real-time profile of the Omega project and and their semantics. They are expressive enough to express all time related elements of a model. An increase in expressiveness can be obtained by an extension with probabilistic features, as used in the context of performance analysis. In Section 3.2, we define the semantics of the basic time concepts as an extension of any untimed formalism which can be interpreted as an event labeled transition system. In particular, we distinguish two different interpretations of Boolean expressions, either as predicates or as constraints. In Section 3.3, we define a set of derived notations, intended to give the user a convenient means for the expression of common constraint patterns. They can be expressed in terms of basic concepts, at least in the context of the expression language OCL [WK98], allowing powerful quantifications. The example in Section 4 illustrates some of the concepts introduced previously. Finally, in Section 5, we propose a way to adapt this UML profile to SDL.

2 Needs for timed specifications

A minimal set of real-time related concepts necessary for a modelling language with the aim to support all the stages of the development of a real-time system, are the following ones:

A global notion of time. If necessary, local time can be defined by means of local clocks which have a well-defined relationship to global time (which might be defined in terms of drift (maximal deviation of speed), offset (max. deviation of value), ...).

Explicit functional use of time by means of access to time through timers, clocks and a construct like now (allowing access to global time), allowing to make control- or dataflow explicitly time dependent. It is desirable that the necessary discretization can be decided as late as possible in order to avoid unnecessary complexity.

Means for the specification of time constraints representing requirements or assumptions on the environment; for the underlying execution platform assumptions (or knowledge) on the execution time of activities must be expressible and

\(^3\) see http://www-omega.imag.fr
taken into account at model level. We propose constraints on durations between events as the basic means for expressing time constraints, as it is the case in Sequence Charts and in Timed Automata. This is also in line with the UML RT profile.

Timing issues are strongly influenced by the execution mode and the type of concurrency between different parts of the system. Concerning parallelism one can distinguish simulated parallelism (concurrent entities sharing computing resources) and distributed parallelism. This distinction is generally not made in modelling languages emphasizing on functional aspects, but it is important when assumptions on execution times and effects of resource sharing should be taken into account.

For entities composed by simulated parallelism, one may define restrictions on the level of granularity at which computations of concurrent objects are interleaved: in the case of functionally independent behaviours, the choice of this granularity has no influence on the functional properties; at the level of the implementation, this granularity is defined by the preemptibility or non-preemptibility of steps of computations. When interleaving a set of functionally dependent parallel behaviours, the choice of the granularity influences not only the timed but also the functional behaviour. It defines when stimuli from outside can be taken into account. Extreme choices consist in:

- making no hypothesis on the granularity (e.g. in Java): any parallel behaviour of the set can accept stimuli at any time, independently of all others. This oblige the designer to handle mutual exclusion and access of shared variables explicitly,
- run-to-completion semantics: stimuli from the outside of each entity are only accepted when all the parallel behaviours of the set are stable (waiting for a stimulus) and no stimulus from within the entity is present.

In the context of the second solution, activities are insensitive to their environment before they are terminated, that is atomic as seen from outside. In this context, it is important to have an interrupt concept, allowing for example to terminate useless activities without cutting all the atomic steps into smaller pieces (and adding corresponding locks, etc.). In this article, we focus only on purely time related aspects, and we consider the execution model to be part of the functional model. In the subset of UML we consider in Omega, a notion of activity group (basically an active object and several passive objects, executing like an SDL process in a run-to-completion fashion) defines how activities within a group can interleave.

Scheduling. For schedulability analysis, one needs to distinguish between simulated and distributed parallelism between activity groups, as this determines which execution times add up and which ones don’t. We introduce the following notations\(^4\)

- a notion of resource and a deployment mapping from activity groups to resources

\(^4\) partly described in Section 3.3, but still to be worked out in detail
the distinction between *preemptible* and *non preemptible* resources, actions
and objects for the definition of a notion of atomicity.

- the distinction between *execution delay* and *execution time* of actions or
  transitions, where the first one refers to the time elapsing between the start
time and the end time of an action, and the second one refers to pure execution
time\(^5\) which is necessary for an analysis taking into account scheduling
issues.

- *priorities* within or between concurrent behaviours, defined in a hierarchical
  fashion, are a powerful means to eliminate non determinism and can be used
at model level to define any scheduling policy. At implementation level, pri-
orities might either be reflected by a run-time scheduler (possibly a hierarchy
of explicit event handlers) or by sequential code generated according to the
defined priority rules.

3 Ingredients for a UML profile for timed verification

We start by defining a set of basic time related concepts, that give expressiveness
to the framework, in Section 3.2 we define their semantics, and finally in
Section 3.3, derived concepts are introduced to make the use of the framework
easier.

3.1 Basic time related concepts

The only time related concept existing in the standard UML is the *Time* data
type. UML defines no mechanisms for representing time progress nor operations
on *Time*, *profile for schedulability, performance and time* [OMG02] defines a large
vocabulary of time related concepts, which is not completely worked out and
which is more or less syntax so far. We consider here only the subset concerned
with real-time, add some missing concepts, and propose a semantic framework.

*Primitives for modelling time.* As in SDL, the time model is based on two data
types: *Time*, relating to time instances, and *Duration*, relating to the time elaps-
ing between two instances of time. These data types can be used, like any UML
data type, in attributes, parameters, etc.

A particular instance of time is *now*. It always holds the *current time* and is
visible in all parts of the model. Using the vocabulary of ASM [Gur95], *now* is a
monitored variable, i.e. some external mechanism changes its value. A general
constraint imposes its values to monotonically grow, and model dependent *time
constraints* can introduce additional restrictions on time progress as compared
with system progress.

*Timing mechanisms.* In the UML RT Profile, two related timing mechanisms are
introduced, informing the system on time progress: *timer* and *clock*. A *timer* is
an extension of its homonym in SDL (as it can be *set*, *reset* and sends *timeout
signals). Additionally, timers can be periodic, can be paused or restarted. *Clocks*
are similar to periodic timers and emit *Ticks*.

\(^5\) obtained by measurements on the target platform or by static timing analysis as
done, for example, in [SRW02]
Events play an important role in the UML RT profile. They are defined in UML as a “specification of a type of observable occurrence”, that is as an (observable) state change. There exists no explicit notion of event in SDL, but MSC define behaviours in terms of events. We use TimedEvents with a time stamp holding their occurrence time. We define a rich set of events allowing to refer to all relevant time points of a behaviour.

For instance, with any transmission of a signal sig, are associated three events: send(sig) - the instant at which the signal is sent by the sender, receive(sig) - the point of time at which the signal is received by the receiver (in its input queue), and consume(sig) - the instant at which the signal is consumed and a transition is triggered (or the signal is discarded). Not all three events associated with a signal must necessarily be distinct. For example, if receive(sig) and the corresponding consume(sig) occur at the same time, they represent a single event, and this corresponds to non buffered communication. With any action are associated events start(action) and end(action). In case of an instantaneous action, the start and end event are always simultaneous and denoted by the event start-end(action). The complete list of predefined events (associated with operation calls, signal exchange, actions, transitions, timers, etc.) is given in the Appendix.

As an event is an instant of state change, it is defined by a triple of the form (occurrence time, current state, next state) or equivalently by a triple of the form (occurrence time, current state, action), where the next state is defined by the current state and the action defines the rules according to which the state after the event can be obtained. We chose the second type of representation, by encoding the relevant part of the current state and the action in an event name, thus allowing constraints on an event only to depend on its occurrence time, current state and the parameters of its action.

For instance, a send event associated with a call is identified by the relevant part of the state in which the call is issued, that is: the object initiating the call, the target object, the operation that is called its parameters, and possibly also the place in the control flow of the state machine in which the call takes place as well as some condition on the local or global state.

Example: The expression cond send(o#tr@lab:target(sig)) identifies all the events in which the signal sig is sent by an object identified by o to an object target as a result of the execution of an action labeled lab on a transition named tr (of the state machine associated with o); moreover the condition cond must hold on the global state when the event occurs. Events can be specified partially, for example send(o:sig) represents any event in which the signal sig is sent by an object o, regardless of target, action that generated it, parameters or the global state.

Note that, even if all the parameters are identified, the event may have multiple occurrences in a given execution. Thus, for any given execution, an event specifications defines a sequence of event occurrences. In order to be able to distinguish different occurrences of an events and to talk about the history of a given event, previous occurrences of an event can be referred to by means of expressions of the form pre(event), pre(pre(event)),...
Time expressions evaluate to Time instances. A particular time expression, evaluated only in events, is now. In any state, the expression \( \text{time}(ev) \) evaluates to the time point of the last occurrence of an event matching \( ev \), whereas \( \text{time}(\text{pre}(ev)) \) evaluates to the point of time of the occurrence before the last one. ... Simple arithmetic expressions, as in SDL, are other examples of time expressions.

*Example:* \( t : \text{Time} \pm d : \text{Duration}, t : \text{Time} \pm \text{real} \times d : \text{Duration} \).

Duration expressions evaluate to Duration instances. All evaluations of features of type Duration are duration expressions. Other duration expressions are, as in SDL, simple arithmetic expressions, such as time instance subtraction.

*Examples:*

\[
\begin{align*}
& t_1 : \text{Time} - t_2 : \text{Time} \quad \text{(under the condition that } t_1 \geq t_2) \\
& \text{pos} \times d : \text{Duration} \quad \text{(a scalar product)} \\
& \text{now - time}(ev) \quad \text{(time elapsed since the last occurrence of event } ev) \\
& \text{time}(ev) - \text{time}(\text{pre}(ev)) \quad \text{(time elapsed between the very last and the previous occurrence of event } ev)
\end{align*}
\]

**Predicates on Time and Duration:** any Boolean expression containing duration or time instances is a time dependent predicate. It can be used (just as any Boolean expression) as guard or test within an action, or as a predicate in a property to be checked. Although in principle arbitrary time dependent Boolean expression can be used, in practice only simple forms of predicates are really useful. Indeed, the type of predicates used in properties or guards determine the possibility of analysis (see Section 5).

*Examples:*

\[
\begin{align*}
& t_1 : \text{Time} - t_2 : \text{Time} \leq d : \text{Duration} \quad \text{(simple duration constraint)} \\
& \text{time}(ev_1) - \text{time}(ev_2) \leq d \quad \text{(simple duration constraint)} \\
& \text{now - time}(\text{event}) \leq \text{time}(\text{pre}(\text{event})) - \text{time}(\text{pre}(\text{pre}(\text{event}))) \quad \text{(constraint on difference of durations, as used in auto adaptive algorithms)}
\end{align*}
\]

Time dependent predicates are evaluated in events, and the fact that an event satisfies some predicate amounts to the evaluation of an assertion of the form \( (s,t,a) \models p^0 \) which is equivalent to the assertion \( (s,a) \models p[t/\text{now}] \) or \( s \models p[t/\text{now}] \) if \( p \) is independent of the parameters of the action. The interpretation of time dependent predicates can be extended to states:

\[
s \models p \text{ iff } \forall t \in [\text{time}(\text{enter}(s)), \text{time}(\text{exit}(s))] . (s,t) \models p\]

that is if \( p \) holds in all time points in which the system stays in state \( s \).

We distinguish between predicates, which are evaluated to true or false in individual events (or states), and invariants - of the form \( \text{invariant}(p) \) (where \( p \) is a predicate) - which hold if \( p \) has been true ever since the initial state. Interesting invariants are often of the form \( \text{invariant}(\text{in}(s) \Rightarrow p) \) or \( \text{invariant}(\text{ad}(e) \Rightarrow p) \), requiring \( p \) to hold at each occurrence of event \( e \), respectively in all time points in which the system is in state \( s \). In order to ease the expression of such invariants, we allow to “attach” \( \text{invariant}(p) \) with states \( s \) or events \( e \) as a short hand.

\( ^0 \) where \( s \) is a state, \( t \) a time value, and \( a \) represents the parameters of the action represented by the event.
3.2 Semantics

The semantics of time related concepts is defined independently of the semantics of the formalism used for expressing functional behavior. We only suppose that the semantics of the functional behavior of systems can be viewed as a labeled transition system, where transitions represent events. An additional requirement is that all the events referred to, are identifiable in the transition system defining the functional semantics.

We define the semantics of the time related concepts by a set of timed automata with urgency [BST98b] constraining the occurrence time (and only on the occurrence time) of all the events whose possible occurrence ordering are given by the untimed semantics.

Thus, the timed behaviour of a system can be represented by the synchronized product of the possibly infinite event labeled transition system defining its behaviours in terms of state changes and (partial) order of events, and a set of time automata defining allowed occurrence times of events.

How time related concepts can be represented as timed automata is relatively straightforward for most constructs. We present here the timed automaton associated with a timer to demonstrate that this way of defining time extension has the expected effect, and we discuss the interpretation of time dependent Boolean expressions, as for them several interpretations coexist.

The time automaton associated with any timer instance is given in Figure 1a. It does not express any constraint on the occurrence times of the actions set and reset, but we suppose that they occur immediately after the preceding action (in the same sequential behaviour), which in the timed automaton is expressed by an eager transition. Also, given an occurrence of set, this automaton constrains the occurrence time of the timeout (i.e. the occur event) to exactly the defined timeout time \(\text{time(set(timer,delay))} + \text{delay}\) which is also expressed by an eager transition.

The point of time of the consumption of the timeout, however, is not restricted by the timer itself, but possibly by some additional constraint, for example "always the time between the timeout and the corresponding consumption is smaller than \(\Delta\)\(^a\), which can be expressed by an invariant of the form

\[
\text{invariant}(\ \text{at(consume(t))} \Rightarrow \text{now} \cdot \text{time(occur(t))} \leq \text{consume-d delay})
\]

or, as we will see in the next section, by the shorthand

\[
\text{invariant}(\ \text{duration(occur(t), consume(t))} \leq \text{consume-d delay})
\]

Figure 1b shows the timed automaton corresponding to the composition of the timed automaton and the timed automata associated with this constraint, where the transition associated with the consume event is delayable, meaning that it will occur somewhere within the specified delay.

\(^a\) Alternatively, we could have another formalism, such as ASM, but the advantage of timed automata is that they have as predefined concepts, all the primitives making the expression of the semantics easy.

\(^b\) under the condition on the functional model that the timer is only set again after the timeout has been consumed.
Interpretations of time dependent Boolean expressions

Boolean expressions can be interpreted in different ways, depending on their use:

**Predicates:** Time dependent Boolean expressions used as guards of transition triggers or in decisions are predicates, which take at each occurrence of the transition either a value *true* or *false* depending on the value of time at the instance the trigger event occurs. The value of complex predicates can be combined from basic ones and can be used in other event based property formalisms: for example in sequence charts, predicates are used as *conditions*.

A timed automaton representing a time dependent predicate, is an observer which, depending on the point of time at which the concerned events are observed, provides a Boolean value. In such a timed automata all transitions are *lazy*, meaning that no constraint on possible time progress is expressed, but the Boolean value (the value of the predicate) depends on time progress. As an example, see the timed automaton representing the predicate $\text{duration}(\text{start}, \text{end})$ in Figure 2.

**Invariants:** The value of a (time dependent) invariant of the form $\text{invariant}(p)$ is *true* only if $p$ holds on the entire execution leading to the current state or event. Invariants can be used as properties, meaning that they should be derivable from a given system description. Live Sequence Charts [DH99, HM02], as they are used in Omega, provide an intuitive means to express complex invariants as well as liveness properties. The timed automaton associated with $\text{invariant}(p)$ is that of $p$, except that, once it takes the value *false*, its value remains *false* forever.

**Constraints:** Invariants can also be used as constraints or (assumed) facts on the system under consideration, its environment or the underlying execution platform. An assignment $x := 0$ represents the invariant *whenever this step is executed, in the state just after the execution the variable $x$ has the value 0* without necessarily saying how this is achieved exactly. Analogously, an invariant of the form $\text{duration(action)} \leq 5^9$ can be used to restrict the model to the set $\vdots$

---

Fig. 1. timed automaton for a timer (a) with constraint consumption delay (b)
of executions satisfying the functional constraints in which not more than 5 time units pass between the event \textit{start}(action) and the corresponding event \textit{end}(action).

Such a constraint on the time elapsing between two events, \textit{forces} time progress or waiting in accordance with the constraint. In timed automata, this is obtained by means of \textit{eager} or \textit{delayable} transitions. Whenever, it is impossible to force time progress in such a way as to satisfy some constraint, this means that the set of time constraints is either inconsistent or incomplete and can be represented in the timed automaton implicitly by a timelock or explicitly by an error condition (see Figure 2b for an example).

\textit{Constraints} restrict the model on which \textit{properties} are verified. If constraints are \textit{non overlapping}, that is, in every possible execution, any event is under the scope of at most one time constraint, they are consistent. In this case, the set of associated timed automata define indeed an executable model, in the sense that any computation satisfying the constraints up to some point of time, can be continued by satisfying all constraints. When constraints are overlapping, they might be inconsistent or incomplete. Inconsistency must lead to some redesign. Figure 3 shows two incomplete sets of (overlapping) constraints. Both sets of constraints are satisfiable, but must be completed by a constraint on the occurrence time of the event \textit{e}3 (relative to the occurrence time of \textit{e}1) in order to guarantee that all other constraints remain satisfiable. Most existing tools, such as the simulation tools for SDL or the tool for executing timed live sequence charts of [HM02], avoid this problem by a priori choosing the earliest point of time at which an event is possible. This corresponds often to a possible execution (at least when the set of constraints is consistent), but not always, as shows the second example of Figure 3. Moreover, our aim is to keep constraints as loose as possible, and thus to characterize the largest set of timed behaviours of (sequential) subsystems and to enable compositional analysis.

In [BS00] is described a framework for more flexible composition of duration constraints than conjunction which could be used to further extend the real-time profile presented here.

\subsection*{3.3 Derived duration expressions}

Our approach is based on duration constraints between the occurrences of an event \textit{e}v1 and a subsequent event \textit{e}v2. In a sequence of occurrences of \textit{e}v1 and \textit{e}v2,
several options exist to identify the “matching” pairs between which to impose the constraint. We propose expressions for three different ways of identifying such matching pairs.

1. \( \text{duration}(e1, e2) \) represents the duration between an occurrence of \( e1 \) and the next occurrence of \( e2 \) such that there is no other occurrence of \( e1 \) in between. If evaluating \( \text{duration}(e1, e2) \) only at the occurrence of the first \( e2 \) after a (series of) \( e1 \), the value of \( \text{duration}(e1, e2) \) is that of now - time(\( e1 \)) that is the time passed since the last event \( e1 \). At other points of time, its value is difficult to express only in terms of occurrence times of events: time(pre\(^n\) (\( e2 \))) - time(\( e1 \)) where \( n \) is chosen such that time(pre\(^{n-1}\) (\( e2 \))) \leq time(\( e1 \)) \leq time(pre\(^n\) (\( e2 \))). Using additional time stamp variables or a timed automaton makes the expression of this constraint easy (see Figure 2 for the expression of \( \text{duration}(\text{start, end}) \leq d \). A particular instance is \( \text{duration}(e1, \text{now}) \) which represents at any point of time the duration since the last occurrence of \( e1 \), that is now - time(\( e1 \)). This expression corresponds thus to an implicit duration counter associated with event \( e1 \), and it can alternatively be expressed by means of a time stamp at each occurrence of \( e1 \).

2. When several request events (each corresponding to \( e1 \)), may all be answered by a single effect event (corresponding to \( e2 \)), and this should happen within a limited time starting from the first request, the time elapsed between the first event of a series of occurrences of \( e1 \) and the first consecutive \( e2 \) is relevant. The expression \( \text{duration}\_\text{first}(e1, e2) \) can be used in this case (the timed automaton associated with \( \text{duration}\_\text{first}(\text{start, end}) \leq d \) is obtained by eliminating in the timed automaton of Figure 2 the set(\( x \)) action in case of repeated start events).

3. Finally, the case where several observations on pairs (\( e1,e2 \) are “active” simultaneously, must be considered. In most cases, one can find a parameter \( x \) (or combination of parameters) identifying the matching pairs, and express the required constraint by a constraint of the form

\[
\text{for any } x: \text{duration}(e1(x,x), e2(x,x)) \leq d.
\]

Nevertheless, due to the implicit use of FIFO buffers for storing signals in objects, it might be impossible to always find the matching parameters, and for this purpose \( \text{pipelineDuration}(e1, e2) \) is introduced, where the matching pairs may be overlapping, observed in a pipelined fashion. In any case, such a set of constraints cannot be expressed by a single timed automaton, but by multiple instances of a parameterized timed automaton, such that at every occurrence of \( e1 \), a new instance (with appropriate values
of parameters) is instantiated, constraining only its corresponding \texttt{ev2} event; each instance is killed when the constrained event \texttt{ev2} occurs.

Moreover, we propose some convenient shorthands

1. Instead of \texttt{duration(enter(state),exit(state))} or \texttt{duration(trued(cond),falsed(cond))},
   the shorthands \texttt{duration(in(state))}, respectively \texttt{duration(in(cond))} can be used, and analogously for any type of “duration”.

2. We allow also expressions of the form \texttt{duration(SE1,SE2)}, where \texttt{SE1}, \texttt{SE2} are lists of events\textsuperscript{10}, representing the distance between any occurrence of an event in \texttt{SE} and the next instance of any event in \texttt{SE}.

Additionally, duration expressions are defined which correspond to frequently used duration patterns. We define their semantics by a mapping to basic duration expressions:

\begin{itemize}
  \item \textit{Client response time} attached with operation calls and signals with some specified response, is defined as \texttt{duration:invoke(op),consume-return(op)}, that is, the time elapsed between the moment at which the operation call (respectively the initiating signal) is emitted and the moment at which the response (return or defined response signal) is consumed by the client.

  \textit{Example: Client response time} \texttt{(sourceObj:sig1,sig2)\leq 7}, means that each time the object \texttt{o} sends \texttt{sig1}, then, if it receives a response \texttt{sig2} some time later, this will happen within less than 7 time units.

  Similarly, a \textit{Server response time} attached with operation calls or signals with specified response, is defined by \texttt{duration:receive(op),invoke-return(op)}, that is, the time between the reception of a request (in the input queue) and the moment at which the response (or return) is sent back (from the server point of view).

  A \textit{Duration} can be associated with some behaviour (action, transition, operation, or signal with specified response), as a shorthand for the duration between the start and the termination time of the behaviour. That is, one can simply write \texttt{duration:act} for \texttt{duration:start(act),end(act)}. It does not include waiting time in the event queue when the execution refers the treatment of a signal, but it does include potential preemption time during the execution of the behaviour. \textit{Execution time} associated with a behaviour, is similar to its \textit{duration}, except that it also takes deployment issues into account and accounts only for the time during which the object is executing and neither preempted nor waiting for some response from external objects.

  A \textit{Period}, attached with an event, is defined as the duration between successive occurrences of the event, that is \texttt{duration:pre(event),event}.

  \textit{Reactivity} is attached with objects or groups of objects and is defined as the maximum delay between any event of the form \texttt{Receive(req)} and the corresponding event \texttt{Consume(Req)}, that is the maximal delay which may elapse between the moment at which a request reaches the object and the moment at which it starts to treat it. This feature is useful, when the size of input queue can be statically bounded.

  A \textit{Transmission delay} attached with associations stereotyped as communication paths (similar to SDL channels), defines a communication delay.
\end{itemize}

\textsuperscript{10} where a list is used to represent a set of events
The list of derived duration expressions has the status of an initial proposal. It contains a number of concepts likely to be interesting and serves to illustrate the general idea. Nevertheless, a larger discussion and more feedback from users is necessary. Some of the concepts may need adaptation; for instance, the more restricted notion of Worst case execution time might be more useful than Execution time, it might be interesting to distinguish different kinds of transmission delays, and predefined notions of jitter might be introduced explicitly, instead of asking the user to write down his preferred formula relating the occurrence times of \( n \) consecutive events \( ev, \) \( pre(ev), \ldots, \) \( pre^n (ev) \).

4 Example

We illustrate a possible use of time annotations by modeling a part of the traffic light system of the SDL design contest. Given the level of description of this example, execution times, response times, etc play no role. At control level, only expressions of type \( \text{duration}[\] and the distinction of different types of invariants are relevant. The main idea is to illustrate the possibility of complete separation of functional and time dependent specifications and of the expression of time constraints by means of a set of relative simple constraints, each one depending on not more than a few events.

A traffic light controller consists of a set of traffic light controller proxies (TLCP)\(^1\). Each TLCP controls the light switches of all its associated traffic lights and gets information about presence or absence of traffic on its controlled lanes from a number of sensors (at different distances from the lights on all the lanes) by means of signals traffic and nottraffic from a traffic sensor handler associated with each TLCP, not described here. The information about the status is stored in a Boolean variable traffic. The state machine of TLCP described using SDL (see Figure 4) has only two states, red and green, and it can pass from green to red without any condition, and from red to green when traffic has been detected and all other traffic lights are red (this requires a global variable AllRed updated by all the TCLPs at every transition). Each traffic light has an additional orange phase of some fixed length, but this is handled in the traffic lights themselves (not described here).

The state machine of Figure 4 describing the functional behaviour is totally time independent. In addition, a number of time constraints need to be satisfied. We show that also for this type of “functional” time constraints, it can be advantageous to use a constraint based approach, instead of starting immediately with an implementation in terms of timers or some cyclic environment observation.

The first constraint prevents livelocks: whenever there is traffic, the light will be turned green within some delay \( \text{max} \). This is expressed by an invariant of the following form, where we omit the context \( \text{TLCP}(i) \) to simplify the expressions:

\(^1\)In the contest only two TLCP corresponding to two “directions” (that is sets of coupled traffic lights) are considered, but in a more general setting a larger number of directions may exist (the approach could also be extended to dynamically formed sets of lights with common green phases).
Fig. 4. TLCP functional behaviour

\[
duration\{\text{enter}(\text{red}), \text{true}(\text{traffic})\}, \text{enter}(\text{red}) \leq \max_t, \quad (1)
\]

Due to the existence of an orange phase, and in order to avoid unstable behaviour, a minimal duration of both the green and the red phase must be fixed:

\[
duration(\text{in(green)}) \geq \min_g \quad (2)
\]
\[
duration(\text{in(red)}) \geq \min_r \quad (3)
\]

Constraint (1) implies a constraint on the “nominal” length of the green phase, depending on the number \( N \) of “directions”. Moreover, once the minimal delay passed, as soon as there is no traffic anymore on the lanes of \( \text{TLCP}(i) \), the green phase should be terminated within a delay \( \epsilon \) (except if within \( \epsilon \) new traffic arrives). When there is traffic all the time (with the exception of some durations not longer than \( \epsilon \)), the green phase takes its nominal length \( \max_g \leq \max_r / (N - 1) \). We express this as an invariant\(^\text{12}\) attached with the green-to-red transition:

\[
\begin{align*}
\text{(duration(\text{in(green)}) = \min_g)} & \implies \text{duration(\text{true(nottraffic)}, \text{now})} \geq \epsilon \\
\land \ (\min_g \leq \text{duration(\text{in(green)})} \leq \max_g) & \implies \text{duration(\text{true(nottraffic)}, \text{now})} = \epsilon \\
\land \ (\text{duration(\text{in(green)})} \geq \max_g) & \implies \text{duration(\text{true(nottraffic)}, \text{now})} \leq \epsilon
\end{align*}
\quad (4)
\]

The composition of the timed automata associated with all constraints\(^\text{13}\) is given in Figure 5. The so defined traffic light never turns green without traffic and stays green a limited amount of time. In order to make sure that even with permanent traffic on all directions each light turns green after at most \( \max_g \) time, the order in which the lights turn green cannot be chosen non deterministically\(^\text{14}\). Instead of choosing some fixed order, we propose a more flexible solution based on a dynamic priority, initialized arbitrarily to

\[
\text{init(priority)} = \forall j \ (\text{priority}(j) := j) \quad (5)
\]

and updated each time \( \text{TLCP}(i) \) changes to green by

\[
\text{update}(i, \text{priority}) = \forall j \ (\text{priority}(j) := (\text{if } j = i \text{ then } 0 \text{ else } \text{priority}(j) + 1)) \quad (6)
\]

\(^\text{12}\) using the “min-synchronization” of the framework of \([\text{BS00}]\) instead of conjunction with constraint (2), would allow to simplify a lot the expression of constraint (4)

\(^\text{13}\) taking into account that the green and the red phase alternate, and simplified by the assumption that traffic cannot go away in the red phase

\(^\text{14}\) in order to allow a non deterministic choice, the condition for turn green must also depend on the waiting time of all the TLCP(i)
that is, the longer a light remains red, even in absence of traffic on its lanes, the higher gets its priority, allowing occasional traffic to pass quickly without disturbing lanes with heavy traffic.

Figure 6 shows a possible run for 3 directions, depending on traffic. If permanent traffic is present on all directions, each traffic light spends the maximum time in red and in green (the first period in Figure 6). If there is no traffic on direction 2 before the end of the nominal green phase, then its light turns red earlier. If no traffic is present on some direction (here 2), even if it has the highest priority, it passes its turn keeping its priority.

![Fig. 5. TLCP behaviour: time constraints](image)

![Fig. 6. A possible evolution of traffic lights over time](image)

The timed traffic light behaviour is defined by the conjunction of the behaviours defined by Fig. 4 and Fig. 5. It can not be directly translated into SDL by replacing clocks by time stamps, for two reasons:

- the use of urgency is essential to obtain the expected behaviour
- SDL transitions represent actions or activities which might take time, whereas timed automata transitions represent events that is instantaneous state changes.

An actual traffic light satisfying all the constraints can finally be implemented in different ways. One consists in introducing signals `true/false(traffic)` and
timers for the $\epsilon$ waiting period. A second solution consists in a process scanning every $\epsilon$ for changes of the traffic variable, where the length $\epsilon$ of the cycle is induced by the required minimal reaction time. Notice that the cycle length is not necessarily constant: during the red phase, it depends on the precision with which one wants to determine the time of traffic occurrence, and, as long as the minimal waiting time ($\min_g$, $\min_w$) has not been reached, there is no need to scrutinize for changes of traffic at all.

Choosing $\max_g = \max_w / (N - 1)$ allows to respect constraint (1). Nevertheless, in case of high traffic on all lanes, no flexibility is possible: the only solution satisfying all the constraints consists in turning the lights green in a round robin fashion. A smaller value of $\max_g$, would allow to modify the priority rules and to provide additional green phases to those lanes with the highest traffic.

Another flexibility could be obtained by not a priori fixing the maximal length of the green phase; there is no need to quit the green phase if there is no traffic on any other line. This can be done by relaxing constraint (4) when there is no traffic on any other line.

Notice also that the proposed solution making use of a global variable priority would be made easy by the introduction of a flexible notion of global priority in SDL.

5 Discussion: Adapt this profile to SDL

SDL already contains most primitives for dealing with time. Nevertheless, SDL suffers from some deficiencies which are discussed in this section.

SDL does not offer the possibility to define local time, timers are less powerful as they cannot be periodic and they cannot express clocks which “tick” periodically, where ticks are consumed instantaneously or lost. The extension of SDL with more powerful timers and UML clocks has already been proposed previously [BGM+01, MSDH01]. Also the introduction of local time is straightforward by having a different now attached with each “locality”. The values of different local times need not to evolve strictly synchronously, but by respecting some specified drift, offset,... Nevertheless, explicit modelling of local clocks should be avoided whenever possible, and the uncertainty on the relative speed or values of local clocks expressed in terms of uncertainties on the occurrence times or durations.

In SDL, it is left open how concurrent activities are sequentialized (when they do not run in a distributed fashion). It is however specified that activities within a process agent are executed in a run-to-completion fashion. This makes the definition of “interrupts” in principle impossible. It has been proposed (e.g. in [BGM+01]) to use the exception mechanism of SDL, which can be considered as a kind of exception to the run-to-completion principle, for the definition of interrupts.

Our main concern is the expression of time constraints on execution times and durations between events. The SDL semantics says that time passes in actions
and not in states, but no means is provided to specify constraints on the time
passed in actions. Waiting can be forced by means of timers and time guards. Ex-
isting tools implement a different interpretation: Like in MSC or timed automata,
"atomic" parts of transitions are viewed as instantaneous events. Moreover, time
times only where explicitly required by waiting conditions. This view allows the
expression of time constraints by means of time guards, decisions and explicit
error states. Nevertheless, such error states do not exist in SDL, and they make
the models cumbersome. Moreover, some of the identified events, such as "re-
ception of a signal in the input queue" has no syntactic representation and can
therefore not be constraint.

Alternatives for the expression of time constraints exist, which can be ex-
tended in such a way as to serve our purpose.

MSC: are an event based formalism allowing the expressing of constraints on
time elapsing between events. By introducing also instances of agents which
are implicit in SDL, such as the input queue of a process or the agent asso-
ciated with a process set, all introduced events can be represented. MSC
are indeed used in practice to express time constraints of SDL specifications.
But MSC are not expressive enough for this purpose. For example, they do
not allow the expression of constraints of the form "if the entire scenario in
between two constraint events occurs, then the time constraint must hold,
otherwise no constraint is imposed". Live Sequence Charts [DH99] are more
powerful as they allow to express such implications by means of cold event
occurrences and hot time constraints.

An extension of MSC, with the concepts of LSC would be very useful and
allow the expression of timed scenarios in practice. Notice however, that
scenario based formalisms can be very cumbersome when the number of
possible alternative scenarios is high or when only events at the interface
of a single agent are constraint, and they should not be the only means for
expressing time constraints.

Annotations: Some tools allow the expression of duration constraints on ac-
tions in the form of special comments. The inconvenience is that each tool
defines its own notation and semantics.

For simple constraints associated with processes tasks, signals, ... specifying
execution times of tasks or durations like response times, process execution
periods, ... as they occur for example in contracts specifying interfaces,
annotations, using the event names and duration expressions defined in Sec-
tion 3.3, are a convenient means. These annotations can take the form of
special comments, but their syntax and semantics shouldn't be left open as
a tool issue.

Observers: Some SDL tools provide the facility to define observers as a means
for defining constraints and properties. Observers allow to impose constraints
on occurrences of events like MSC, but they can also observe and constrain
the global state of the observed system, and look almost like SDL processes.
Observers are a very powerful means for expressing properties and con-
straints, but they exist only in particular tools, SDL does not provide any
standard notation for them.
Triggers of observers are any of the defined events, and they define constraints on their occurrence time may using either time stamps (e.g. \( \text{now} - t \leq 3 \)) or the duration expressions defined in Section 3.3 (e.g. \( \text{duration} ( \text{ev}, \text{now} ) \leq 3 \)). In order to define properties and constraints, they may use the above mentioned error states or an explicit urgency like in timed automata.

In order modelling languages to be used in practice, it is very important that the user has the freedom to express time constraints in the most convenient manner, depending on each situation. For this reason all three above mentioned approaches are useful. The interest to separate functional specification and timing information, can be motivated by the fact that timing information is mostly platform or implementation dependent. Separation makes it easier to adapt the time related specifications to a different platform. In all three types of formalisms, we distinguish between constraints which are part of the definition of the system, and properties which must be derivable from the system definition and constraints.

In order to take into account scheduling and deployment related information, new notations have to be introduced in SDL. The simplest way is almost the one proposed in QSDL [MTMC99]. At the architecture level, a list of resources with an attribute defining their preemptibility is defined. For block and process agents as well as tasks, it is specified on which resource they are executed, where inner agents and tasks run on the same resource as long not otherwise stated. Scheduling policies are defined by keywords, such as RMA or EDF, or by priority rules. They are attached with resources or with agents, in an hierarchical fashion. Within a process, priorities are defined between transitions or their triggers, and within a block, they are defined between sub-agents. Priorities may be dynamic, where dynamic priorities can be specified in a declarative way, that is depending on some precondition, or by means of an observer (attached with the concerned agent), which explicitly updates priorities depending on the observed states and events.

**Perspectives of simulation and validation**

In section 3.2 we have already discussed the the simulation issue: any SDL model with time constraints can be simulated. This is an important property, as this means that such a model can be used for model based testing.

Obviously, any interesting property is in general undecidable on an SDL specification due to infiniteness of data domains, and unbounded message buffers and agent creation. Nevertheless, the verification of time related properties can often be done on a finite control abstraction, that is a system with finite data domains, bounded message buffers and bounded agent creation. A number of interesting verification problems are decidable on such a finite control abstraction under the condition that in the timed automata, obtained by translation, clocks can only be reset to zero, stopped and restarted, and the only allowed tests are comparisons of clock values or differences of clock values with constants. This means that the only allowed constraints imposed on events are boolean combinations of comparisons of the type “duration since the occurrence of some
event lays within an interval" or "the difference of occurrence times between two past events lays within an interval".

MSC which impose only interval constraints on occurrence times of events or on durations between pairs of events and timers, are in this restricted set.

Observers satisfy this constraint if they use only constraints which are Boolean combinations of interval constraints on expressions of the form \( \text{duration}(e_1,e_2) \). An alternative, is using time stamps of the form "\( t := \text{now} \)" and comparisons of the form "\( \text{now} - t \in [2,3] \)" or "\( t_2 - t_1 \in [2,3] \)". The general time stamping mechanism - which allow the comparison of time distances with different end points (e.g., "\( (\text{now} - t_2) \leq (t_2 - t_1) \)"), as they are used in self-adaptive algorithms - is more expressive and outside the decidable set.

Schedulers make use of integrators, that is, they imply clocks used for counting execution time which are stopped when the process is suspended and restarted when it becomes executing again. As long as only interval constraints on such integrators are tested, which is typically the case when worst-case execution times and deadlines are specified, decidability is preserved.

There exist several tools based on timed automata allowing some verification under the above mentioned restrictions, such as Hytech [HHWT97], the Kronos/IF tool [Yov97,BGGM00] and Uppaal [LPY97].

Let us consider a number of relevant time related verification problems:

- Consistency checking of a time constraint system and verification of properties expressing requirements of the same kind, that is interval constraints on distances between events are very similar in nature, and can be done with the same tools.
  Notice that, when execution time constraints and deadlines are specified, and several concurrent behaviours are executed on the same resource, consistency checking and verification includes schedulability checking.

- Incomplete specifications, as in Figure 3, are problematic, as they may require backtracking during simulation. Under the before mentioned restrictions, the verification algorithm does synthesize for any transition the weakest constraint guaranteeing the satisfaction of all future constraints. When constraints are not "cyclically overlapping", completing an incomplete sequential specification can be done in linear time, even when interval bounds have parameters. This is useful for simplifying simulation, even in the case where the overall system is infinite.

Nevertheless, constraint propagation can be done automatically with a reasonable effort only within a sequential behaviour (defined by an agent or a small set of agents). In general, budgeting over concurrent agents, must be provided by the user.

Notice that in the case where the intervals defining constraints are closed, there always exists an exact discretization. Thus which time model is used is not important. In practise, dense time leads often to more tractable models.

\(^{15}\) or \( \text{first-duration}(e_1,e_2) \) or any of the predefined duration expression. Notice that \( \text{pipeline-duration} \) poses a problem when no bound on the maximal number of concurrently active constraints can be given a priori.
What to do when the system does not satisfy the requirements making the verification problem decidable? As long as all constraints are expressible by linear inequalities on time points or durations, the above mentioned verification problems are "semi-decidable", that is, successor sets can still be computed and termination detected, but there is no guarantee that the underlying model is finite and the verification procedure terminates. For these systems, one can still hope that the actually generated model is finite and small enough to be verified. Other remedies consist in using approximations of fixpoints or approximations based on reformulations of the constraints so that they can be analyzed by the above mentioned tools which use a more restricted internal representation than general linear inequalities.

References


[Ile] Ilogix. Rhapsody development environment.


6 Appendix: Predefined events

The set of predefined events which can be used in time-constraints are the following ones. Moreover, the user can define additional events by means of "named" instantaneous skip actions (corresponding to SDL informal tasks).

- **Invoke**: emission of the call request;
- **Receive**: reception of the call by the provider;
- **Accept**: start of the actual processing of the call by the provider;
- **Invokereturn**: emission of the return reply;
- **Receivreturn**: reception of the return by the caller;
- **Acceptreturn**: consumption of the return\(^\text{16}\).

Events associated with a *signal exchange* are:

1. **Send**: sending a signal;
2. **Receive**: reception of the signal by the target (i.e. when the signal is added in the queue);
3. **Consume**: start of treatment of the signal (triggering of transition or the moment the signal is discarded).

The events associated with an *action* are:

1. **Start**: starting time of the action;
2. **End**: termination time of the action;
3. **Startend**: simultaneous start and an termination of an instantaneous action.

The events associated with a *transition* of a state machine are:

1. **Start**: starting time of a transition;
2. **End**: termination of a transition (and entering the next state);

The events associated with a *state* of a state machine are:

1. **Entry**: time at which a state is entered;
2. **Exit**: time at which a state is exited;

The *change events* associated with *Boolean conditions* are:

1. **Trueed** time at which the condition becomes true;
2. **Falseed** time at which the condition becomes false.

The events associated with *timers and clocks* are (some of can be seen as synonyms for some other kinds of events described above):

1. **set/reset/start/stop** (a timer or clock): events corresponding to the instantaneous actions defined for timers and clocks;
2. **Occur**: reaching of timeout time and notification of timeout;
3. **Tick**: equivalent to occur, but related to clocks;
4. **Consume**: timeout or tick consumption.

\(^{16}\) in blocking call semantics Acceptreturn is the same as Receivereturn or Invokereturn.