IF: An Intermediate Representation for SDL and its Applications

Marius Bozga[*], Jean-Claude Fernandez[†], Lucian Ghirvu[‡†], Susanne Graf[*],
Jean-Pierre Krimm[*], Laurent Mounier[*], Joseph Sifakis[*]

Key-words: SDL, Static Analysis, Validation, Model-Checking, Test, Test Generation.

Abstract

We present work of a project for the improvement of a specification/validation toolbox integrating a commercial toolset *Object*GEODE and different validation tools such as the verification tool CADP and the test sequence generator TGV.

The intrinsic complexity of most protocol specifications lead us to study combination of other techniques such as static analysis and abstraction together with the classical model-checking techniques. Experimentation and validation of our results in this context motivated the development of an intermediate representation for SDL called IF. In this intermediate representation, a system is represented as a set of timed automata communicating asynchronously through a set of buffers or by rendez-vous through a set of synchronization gates. The advantage of the use of such a program level intermediate representation is that it is easier to interface with various existing tools, such as static analysis, abstraction and compositional state space generation. Moreover, it allows to define for SDL different, but mathematically sound, notions of time.

We illustrate the use of IF on a small example, a distributed leader election algorithm, on which we perform experimentation using *static analysis* and *compositional generation*.

1 Introduction

SDL and related formalisms such as MSC and TTCN are at the base of a technology for the specification and the validation of telecommunication systems. This technology will be developing fast due to many reasons, institutional, commercial and economical. SDL is promoted by ITU and other international standardization bodies. There exist commercially available tools and most importantly, there are increasing needs for description and validation tools covering as many aspects of system development as possible. These needs motivate the work for enhancement of the existing standards undertaken by ITU and ETSI, in particular.

Among the work directions for improvement of SDL, an important one is the description of non functional aspects of the behavior, such as performance and timing. Finding a "reasonable" notion of time is a central problem which admits many possible solutions depending on choices of semantic models. This is certainly a non trivial question and this is reflected by the variety of the existing proposals.

Choosing an appropriate timed extension for SDL should take into account not only technical considerations about the semantics of timed systems but also more pragmatic ones related to

[*]VERIMAG-Centre Equation, 2 avenue de Vignate, F-38610 Gières, e-mail: Marius.Bozga@imag.fr
[†]LSR/IMAG, BP 82, F-38402 Saint Martin d'Hères Cedex, e-mail: Jean-Claude.Fernandez@imag.fr
[‡]Work partially supported by Région Rhône-Alpes, France

the appropriateness for use in a system engineering context. We believe that the different ideas about extensions of the language must be validated experimentally before being adopted to avoid phenomena of rejection by the users. Furthermore, it is important to ensure as much as possible compatibility with the existing technology and provide evidence that the modified standard can be efficiently supported by tools.

Another challenge for the existing technology for SDL to face the demand for description and validation of systems of increasing size, is to provide environments that allow the user to master this complexity. The existing commercial tools are quite satisfactory in several respects and this is a recognized advantage of SDL over other formalisms poorly supported by tools. However, it is necessary to improve the existing technology to avoid failing to keep up. Mastering complexity requires a set of integrated tools supporting user driven analysis. Of course, the existing tools such as simulators, verifiers, automatic test generators can be improved. Our experience from real case studies shows that another family of tools is badly needed to break down complexity. All the methods for achieving such a goal are important ranging from the simplest and most "naive" to the most sophisticated.

In this paper we present work of a project for the improvement of a specification/validation toolbox interconnecting *Object*GEODE[Ver96] and different validation tools such as CADP[FGK+96] developed jointly with the VASY team of Inria Rhône-Alpes and TGV[FJJV97] developed jointly with the PAMPA team of IRISA. The project has two complementary work directions. The first is the study and the implementation of timed extensions for SDL; this work is carried out in cooperation with Verilog, Sema Group and CNET within a common project. The second is coping with complexity by using a combination of techniques based on static analysis, abstraction and compositional generation. Achieving these objectives requires both theoretical and experimental work. Experimentation and validation of our results in this context motivated the development of an intermediate representation for SDL called IF. IF is based on a simple, and semantically sound model for distributed timed systems which is asynchronously communicating timed automata (automata with clocks). A translator from a static subset of SDL to IF has been developed and IF has been connected to different tools of our toolbox. The use of such an intermediate representation confers many advantages.

- It is possible to implement and evaluate different semantics of time for SDL as the underlying model of IF is general enough to encompass a large variety of notions of urgency, time non determinism and different kinds of real-time constructs.
- IF allows a flattened description of the corresponding SDL specification with the possibility of direct manipulation, simplification and generally application of analysis algorithms which are not easy to perform using commercial tools which, in general, are closed.
- IF can be considered as a common representation model for other existing languages such as Promela or for the combination of languages adopting different description styles.

Related work

After its standardization in the eighties, a lot of work has been done concerning the mathematical foundations of SDL. The first complete semantics was given by the annex F to the recommendation Z.100 and is based on a combination of CSP and META-IV. Even if it is the reference semantics of SDL (about 500 pages), it is far from being complete and contains many inconsistencies and obscure points.

In [Bro91] is given a semantics for SDL based on *streams* and *stream processing functions*. It deals with a subset of SDL and the timing aspects are simplified. An *operational semantics* which covers SDL systems, processes, blocks and channels is given in [God91]. It defines a method to build labeled transition systems from SDL specifications. The approach is relatively complete, however in this case too, time is not handled in a satisfactory manner. An important work is done in [BM95, BMU98] which gives a semantics based on *process algebra* to a rather simple subset of SDL, called φ^-SDL. A method is given, for translating each SDL system into a term of PA^-_{drt}-ID which is a discrete time process algebra extended with propositional signals and conditions, counting process creation operator, and a state operator. Finally, we mention the work of [MGHS97] which proposes an axiomatic semantics based on *Duration Calculus* and the work of [GK97] which uses *abstract real time machines*.

The paper is organized as follows. In the next section, we present an example used throughout the paper to illustrate our work. Then, we describe the main features of the IF formalism used as an intermediate representation for SDL. Finally, we present an open validation environment for SDL specifications and illustrate its usefulness by means of some experimental results.

2 An example: a distributed leader election algorithm

We present a simple example used throughout the paper to illustrate the introduced formalisms and verification methods. We consider a *token ring*, that is a system of n stations $S_1, \ldots, S_n$, connected through a circular network, in which a station is allowed to access some shared resource R only when it "owns" a particular message, the *token*. If the network is unreliable, it is necessary to recover from token loss. This can be done using a *leader election algorithm* [Lan77, CR79] to designate a station responsible for generating a new token.

Formal specifications and verifications of these algorithms already exist and we consider here an SDL version of the one described in [GM96]. Figure 1 shows the system view of the specification. The signals `open` and `close` denote the access and the release of the shared resource (here a part of the environment). The signals `token` and `claim` are the messages circulating on the ring.

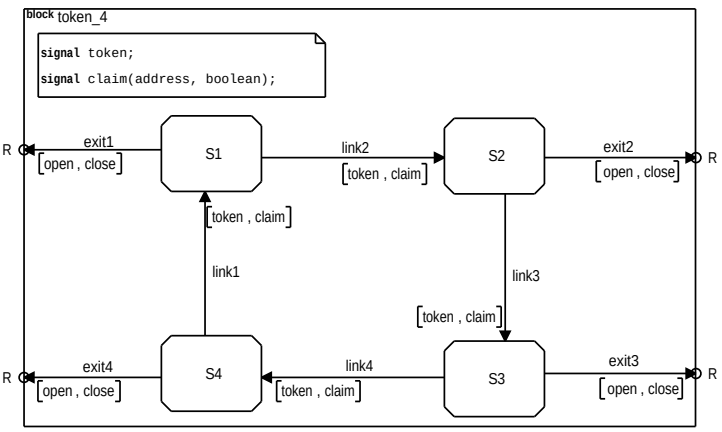

Figure 1: The *token-ring* architecture

All stations S_i are identical and modelled by the SDL process of Figure 2. On expiration of the timer `worried` token loss is assumed: this timer is set when the station waits for the token, and reset when it receives it. The "alternating bit" `round` is used to distinguish between valid claims (emitted during the current election phase) and old ones (cancelled by a token reception). In the `idle` state, a station may either receive the token from its neighbour (then it reaches the `critical` state and can access the resource) and receive the timer expiration signal (then it emits a claim stamped with its `address` and the current value of `round`) or receive a claim. A received claim is "filtered" if its associated `address` is smaller than its own address and transmitted unchanged if it is greater. If its own valid claim is received, then this station becomes elected and generates a new token.

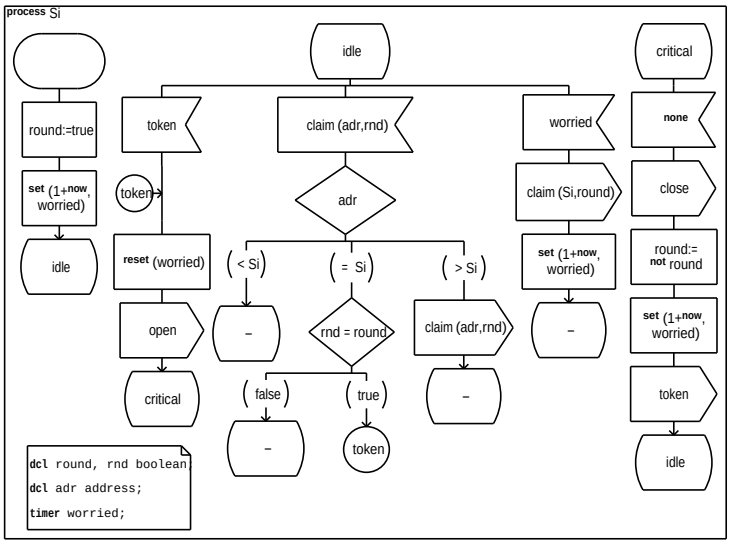

Figure 2: The behaviour of station S_i

To complete this specification, message loss should be modelled explicitly (for instance by introducing a non deterministic choice when a token or claim is emitted by a station). However, using the intermediate representation IF, message loss can be expressed directly using lossy buffers.

3 IF: An intermediate representation for SDL

In the following sections, we give a brief overview of the intermediate representation IF, its operational semantics in terms of labeled transition systems and the translation of a rather extended subset of SDL into IF. A more complete description of IF and its semantics can be found in [BFG+98]. In particular, we do not present the rendez-vous communication mechanism here.

3.1 An overview on IF

In IF, a system is a set of processes communicating asynchronously through a set of buffers (which may be lossy/reliable and bounded/unbounded). The timed behaviour of a system can

be controlled through clocks (like in timed automata [ACD93, HNSY94]) and timers (SDL timers, which can be set, reset and expire when they reach a value below 0).

3.1.1 IF system definition

A system is a tuple *Sys* = (*glob-def*,PROCS) where

- *glob-def* = (*type-def*,*sig-def*,*var-def*,*buf-def*) is a list of global definitions, where *type-def* is a list of type definitions, *sig-def* defines a list of parameterized signals (as in SDL), *var-def* is a list of global variable definitions, and finally, *buf-def* is a list of buffers through which the processes communicate by asynchronous signal exchange.
- PROCS defines a set of processes described in section 3.1.2.

3.1.2 IF process definition

Processes are defined by a set of local variables, a set of control states and a set of control transitions. A process P∈PROCS is a tuple P= (*var-def*, Q, CTRANS), where:

- *var-def* is a list of local variable definitions (including timers and clocks)
- Q is a set of control states on which the following attributes are defined:
  - *stable*(q) and *init*(q) are boolean attributes, where only stable states are visible on the semantic level.
  - for stable states, the *tpc*(q) attribute is a predicate on any variable "visible" in the process (global variables and local variables of P), defining when time is allowed to progress. In non stable states, time cannot progress.
  - *save*(q), *discard*(q) are lists of filters of the form
    signal-list [**in** buf] [**if** cond].
    save(q) is used to implement the **save** statement of SDL; its effect is to preserve all signals of the list in buf, whenever the condition cond holds.
    discard(q) is used to implement the implicit discarding of unconsumable signals of SDL. When reading the next input signal in buf, all signals to be discarded preceding it in buf are discarded in the same atomic transition.
- CTRANS is a set of control transitions, consisting of two types of transitions between two control states q,q'∈Q:
  - input transitions which are triggered by some signal read from one of the communication buffers as in SDL: $\mathtt{q} \xrightarrow[(\mathtt{u})]{\mathtt{g} \mapsto \mathtt{input}\ ;\ \mathtt{body}} \mathtt{q}'$
  - internal transitions depending not on communications: $\mathtt{q} \xrightarrow[(\mathtt{u})]{\mathtt{g} \mapsto \mathtt{body}} \mathtt{q}'$

Where in both cases:

- g is a predicate representing the *guard* of the transition which may depend on variables visible in the process (including timers, clocks and and buffers, where buffers are accessed through a set of primitives).
- body is a sequence of the following types of atomic actions:

- *outputs* of the form "**output** *sig*(*par_list*) **to** *buf*" have as effect to append a *signal* of the form "*sig*(*par_list*)" at the end of the buffer *buf*.
- usual *assignments*.
- *settings* of timers of the form "**set** *timer* := `exp`". This has the effect to activate `timer` and to set it to the value of `exp`. An active timer decreases with progress of time. SDL timers expire when they reach the value 0, but in IF any timer tests are allowed. Clocks are always active and they increase with progress of time.
- *resettings* of timers and clocks, which have the effect to inactivate timers and to assign the value 0 to clocks.

- The attribute `u`∈{**eager, delayable, lazy**} defines the urgency type of each transition. **eager** transitions have absolute priority over progress of time, **delayable** transitions may let time progress, but only as long as they remain enabled, whereas **lazy** transitions cannot prevent progress of time. These urgency types are introduced in [BST98].

- `input` is an input of the form "**input** `sig`(*reference_list*) **from** `buf` [**if** `cond`]" where
  - `sig` is a signal,
  - *reference_list* the list of references[1] in which the received parameters are stored,
  - `buf` is the name of the buffer from which the signal should be read
  - `cond` is a "post guard" defining the condition under which the received signal is accepted; `cond` may depend on received parameters.

3.2 Semantics of IF

3.2.1 Association of a model with a process

We show how with a process can be associated a labeled transition system, and then, how these process models can be composed to obtain a system model. Let `P`= (*var-def*, `Q`, cTRANS) be a process definition in the system `Sys` and:

- Let TIME be a set of environments for timers and clocks (for simplicity of the presentation, we suppose that these environments are global, that is, applicable to all timers and clocks occurring in `Sys`, if necessary, using renaming). An environment $\mathcal{T}$∈TIME defines for every clock a value in a time domain T (positive integers or reals), and for every timer either a value in T or the value *"inac"* (which can be represented by a negative value) meaning that the timer is not active. Setting or resetting a timer or a clock affects a valuation $\mathcal{T}$ in an obvious manner. Progress of time by an amount δ transforms the valuation $\mathcal{T}$ into the valuation $\mathcal{T} \boxplus \delta$ in which the values of all clocks are increased by δ, and the values of all timers are decreased by δ (where the minimal value is zero).
- Let BUF be a set of buffer environments $\mathcal{B}$, representing possible contents of the buffers of the system, on which all necessary primitives are defined: e.g. "get the first signal of a given buffer, taking into account the save and the discard attribute of a given control state", "append a signal at the end of a buffer",..."time progress by amount δ", denoted by $\mathcal{B} \boxplus \delta$, is necessary for buffers with delay.
- Let ENV be a set of environments $\mathcal{E}$ defining the set of valuations of all other variables defined in the system `Sys`.

[1] that is an "assignable" expression such as a variable or an element of an array

The semantics of P is the labeled transition system $[\mathtt{P}] = (\mathtt{Q}\times\text{VAL},\textsc{Trans},\textsc{Ttrans})$ where

- $\mathtt{Q}\times$VAL is the set of states and VAL$=$ ENV$\times$TIME$\times$BUF is the set of data states.
- Trans is the set of untimed transitions obtained from control transitions by the following rule: for any $(\mathcal{E},\mathcal{T},\mathcal{B}),(\mathcal{E}',\mathcal{T}',\mathcal{B}')\in$VAL and input transition (and simpler for an internal transition)

  $$\mathtt{q} \xrightarrow[(\mathtt{u})]{\mathtt{g} \mapsto (\mathtt{sig}(x_1 \ldots x_n),\mathtt{buf},\mathtt{cond})\ ;\ \mathtt{body}} \mathtt{q}' \in \text{cTrans} \qquad \text{implies}$$

  $$(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B})) \xrightarrow{\ell} (\mathtt{q}',(\mathcal{E}',\mathcal{T}',\mathcal{B}')) \in \textsc{Trans}, \qquad \text{if}$$

  - the guard g evaluates to `true` in the environment $(\mathcal{E},\mathcal{T},\mathcal{B})$
  - the first element of `buf` in the environment $\mathcal{B}$— after elimination of appropriate signals of the discard attribute and saving of the signals of the save attribute — is a signal $\mathtt{sig}(v_1 \ldots v_n)$, and the updated buffer environment, after getting $\mathtt{sig}(v_1 \ldots v_n)$, is $\mathcal{B}$"
  - $\mathcal{E}"=\mathcal{E}[v_1 \ldots v_n/x_1 \ldots x_n]$ and $\mathcal{T}"=\mathcal{T}[v_1 \ldots v_n/x_1 \ldots x_n]$ are obtained by assigning to x_i the value v_i of the received parameters,
  - the post guard `cond` evaluates to `true` in the environment $(\mathcal{E}",\mathcal{T}",\mathcal{B}")$
  - $\mathcal{E}'$ is obtained from $\mathcal{E}"$ by executing all the assignments of the body,
  - $\mathcal{T}'$ is obtained from $\mathcal{T}"$ by executing all the settings and resettings occurring in the body, without letting time progress,
  - $\mathcal{B}'$ is obtained from $\mathcal{B}"$ by appending all signals required by outputs in the body,
  - ℓ is an appropriate labeling function used for tracing.

- Ttrans is the set of *time progress transitions*, which are obtained by the following rule: in any state $(\mathrm{q},(\mathcal{E},\mathcal{T},\mathcal{B}))$, time can progress by the amount δ, that is

  $$(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B})) \xrightarrow{\delta} (\mathtt{q},(\mathcal{E},\mathcal{T}\boxplus\delta,\mathcal{B}\boxplus\delta)) \in \textsc{Ttrans} \qquad \text{if}$$

  1. q is *stable*
  2. time can progress in the state $(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B}))$, and,
  3. time can progress by steps until δ: whenever time has progressed by an amount δ' where $0 \le \delta' < \delta$, time can still progress in the reached state $(\mathtt{q},(\mathcal{E},\mathcal{T}\boxplus\delta',\mathcal{B}\boxplus\delta'))$.

  Time can progress in a state $(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B}))$ if and only if the following conditions hold:
  - the time progress attribute *tpc*(q) holds in $(\mathcal{E},\mathcal{T},\mathcal{B})$
  - *no* transition with urgency attribute **eager** is enabled in $(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B}))$
  - for each **delayable** transition `tr` enabled in $(\mathtt{q},(\mathcal{E},\mathcal{T},\mathcal{B}))$, there exists a positive amount of time ϵ, such that `tr` cannot be disabled while time progresses by ϵ.

3.2.2 Composition of models

The semantics of a system *Sys* $=$ (*glob-def*,procs) is obtained by composing the models of processes by means of an associative and commutative parallel operator $\|$.
Let $[\mathtt{P}_i] = (\mathtt{Q}_i\times\text{VAL},\textsc{Trans}_i,\textsc{Ttrans}_i)$ be the models associated with processes (or subsystems) of *Sys*. Then, $[\mathtt{P}_1] \,\|\, [\mathtt{P}_2] = (\mathtt{Q}\times\text{VAL},\textsc{Trans},\textsc{Ttrans})$ where

- $\mathtt{Q}=\mathtt{Q}_1\times\mathtt{Q}_2$ where
  $$\begin{array}{lcl} \mathit{init}((\mathtt{q}_1,\mathtt{q}_2)) & = & \mathit{init}(\mathtt{q}_1) \wedge \mathit{init}(\mathtt{q}_2) \\ \mathit{stable}((\mathtt{q}_1,\mathtt{q}_2)) & = & \mathit{stable}(\mathtt{q}_1) \wedge \mathit{stable}(\mathtt{q}_2) \end{array}$$

- TRANS is the smallest set of transitions obtained by the following rule and its symmetrical rule:

$$\frac{(q_1, \mathcal{V}) \xrightarrow{\ell} (q_1', \mathcal{V}') \in \textsc{Trans}_1 \quad \text{and} \quad \neg stable(q_1) \vee stable(q_2)}{((q_1, q_2), \mathcal{V}) \xrightarrow{\ell} ((q_1', q_2), \mathcal{V}') \in \textsc{Trans}}$$

- TTRANS is the smallest set of transitions obtained by the following rule

$$\frac{(q_1, \mathcal{V}) \xrightarrow{\delta} (q_1, \mathcal{V}') \in \textsc{Ttrans}_1 \quad \text{and} \quad (q_2, \mathcal{V}) \xrightarrow{\delta} (q_2, \mathcal{V}') \in \textsc{Ttrans}_2}{((q_1, q_2), \mathcal{V}) \xrightarrow{\delta} ((q_1, q_2), \mathcal{V}') \in \textsc{Ttrans}}$$

3.3 Translation from SDL to IF

3.3.1 Structure

SDL provides a complex structuring mechanism using blocks, substructures, processes, services, etc, whereas IF systems are *flat*, that is consisting of a single level of processes, communicating directly through buffers. Therefore, a structured SDL system is flattened by the translation into IF. Also, the structured communication mechanism of SDL using channels, signal routes, connection points, etc is transformed into point to point communication through buffers by computing for every output a statically defined unique receiver process (respectively its associated buffer).

All predefined SDL data types, arrays, records and enumerated types can be translated. For abstract data types, only the signatures are translated, and for simulation, the user must provide an appropriate implementation.

In SDL all signals are implicitly parameterized with the pid of the sender process, therefore in IF all signals have an additional first parameter of type `pid`.

3.3.2 Processes

Basically, for each instance of an SDL process, we generate an equivalent IF process and associate with it a default input queue. If the number of instances can vary in some interval, the maximal number of instances is created.

Variables: Each local variable/timer of an SDL process becomes a local variable/timer of the corresponding IF process. We define also variables `sender`, `offspring` and `parent` which are implicitly defined in SDL. Remote exported/imported variables declared inside an SDL processes become global variables, declared at IF system level.

States: All SDL states (including *start* and *stop*) are translated into *stable* IF control states. As IF transitions have a simpler structure than SDL transitions, we introduce also systematically auxiliary non stable states for each *decision* and each *label* (corresponding to a "join") within an SDL transition. For each *stable* IF state we define the *save* and *discard sets* to be the same as for the corresponding SDL state.

Transitions: For each *minimal path* between two IF control states, an IF transition is generated. It contains the triggers and actions defined on that path in the same order.

All the generated transitions are by default *eager* i.e. they have higher priority than the progress of time; this allows to be conform with the notion of time progress of the tool *Object*GEODE; more liberal notions of time progress can be obtained by using different translations from SDL to IF (see the example below).

- inputs: SDL signal inputs are translated directly into IF inputs, where the sender parameter must be handled explicitly: each signal receives the first parameter in the local variable `sender`.
  Spontaneous input none is translated by an assignment of the `sender` to the pid of the current process. No input part is generated in this case.
- timeouts expirations are *not* notified via timeout signals in IF: each timeout signal consumption in an SDL process is translated into a transition without input, which tests if the corresponding timer evaluates to *zero*, followed by the reset of that timer. The reset is needed to avoid multiple consumption of the same timeout expiration.
- priority inputs: are translated into normal inputs by enforcing the guards of all low priority inputs and the save set of the source state. The guard of each low priority input is conjuncted with a term saying that "*there is no higher priority signal in the buffer*". All low priority signals are explicitly saved if "*at least one input with higher priority exists in the buffer*". Such tests can effectively be expressed by predefined predicates on buffers.
- continuous signal: SDL transitions triggered by a continuous signal test, generate IF transitions without input. They are translated by an IF transition, whose guard is equivalent to the SDL continuous signal.
- enabling condition: an enabling condition following an SDL input signal is translated directly into a post guarded input where the received parameters can be tested.
- task: all SDL formal tasks are translated into IF assignments. Informal tasks become comments in the IF specification.
- set and reset: SDL timer sets become IF timer sets, where an *absolute* value "$now + T$" becomes in IF a *relative* value "T". SDL timer resets become IF timer resets.
- output: SDL outputs become IF outputs: if the *to pid-expression* clause is present in the SDL output, the same pid-expression is taken as destination for the IF output. Otherwise, according to *signal routes signature*, *via restrictions*, *connections*, etc. we compute **statically** the set of all possible destinations. If this set contains exactly one process instance, it become the IF destination, otherwise, this output is not translated. Every output contains as first parameter the `pid` of the sending process.
- decision: each alternative of an SDL formal decision is translated into a guard starting an IF-transition from the corresponding non stable state.
- create: the dynamic creation of processes is not yet handled. But we intend to translate this construction by using the rendez-vous mechanism of IF: a new instance is created (an "inactive" instance is activated) by synchronizing its first action with the process creating (activating) it. During this synchronization, parameters can be passed between the "creating" and the "created" processes, such as the the values of the `parent` and the `offspring` variables, etc.
- procedures: IF does not directly support procedures. But we handle a relatively large class of SDL programs containing procedures by *procedure inlining*, which consists in directly inserting the procedure graph, instead of its call, in the process graph.

Example: translation of the token ring to IF

To illustrate IF, we present the translation of the token ring introduced in Section 2. The translation of the structure is completely straightforward in this example. Figure **??** contains the IF version of the process S_1, where the additional non stable states are dotted.

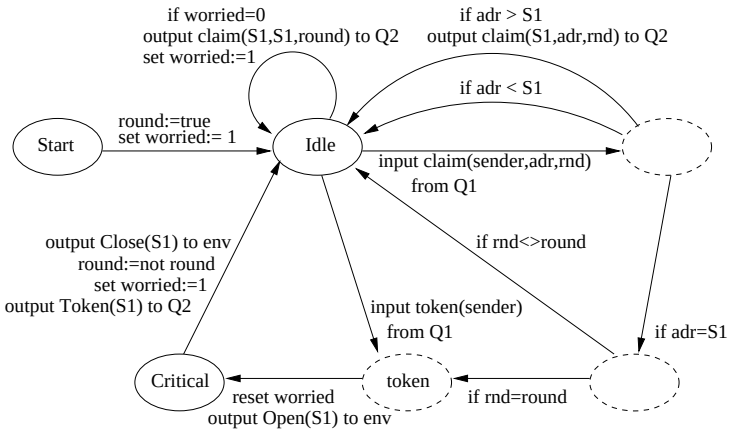

Figure 3: The "graphical" IF description of station S_1

By default, all transitions are **eager**, which leads to the same behaviour as in *Object*GEODE. Thus, time can only progress, and the timeout occur, if the token is really lost (that is, no transition is enabled), and therefore a leader election algorithm is only initiated if necessary. In IF, a different notion of time, closer to reality, can be modeled, e.g. by considering the transition from the `critical` state as **lazy**, thus allowing time to pass there by an arbitrary amount. In order to limit the time that a process can remain in `critical`, one can introduce a clock `cl_crit` which is reset when entering `critical`, add to the outgoing transition the guard `cl_crit`$\leq$*some_limit* and consider this transition as **delayable**.

4 An open validation environment based on IF

One of the main motivations for developing IF is to provide an intermediate representation between several tools in an "open" validation environment for SDL. Indeed, none of the existing tools provides all the validation facilities a user may expect. Therefore, we want to allow them to cooperate, as much as possible using program level connections. An important feature is the ability of the environment to be open: in particular connections with KRONOS [Yov97] (a model checker for timed automata) and INVEST [GS97, BLO98] (a tool computing abstractions) are envisaged.

In this section, we first present the architecture of this environment and its main components. Then, we describe in a more detailed manner two more recent modules concerning static analysis (section 4.2) and compositional generation (section 4.3) which are based on IF.

4.1 Architecture

The environment is based on two validation toolsets, *Object*GEODE and CADP, connected through the intermediate representation IF. There exists already a connection between these toolsets at

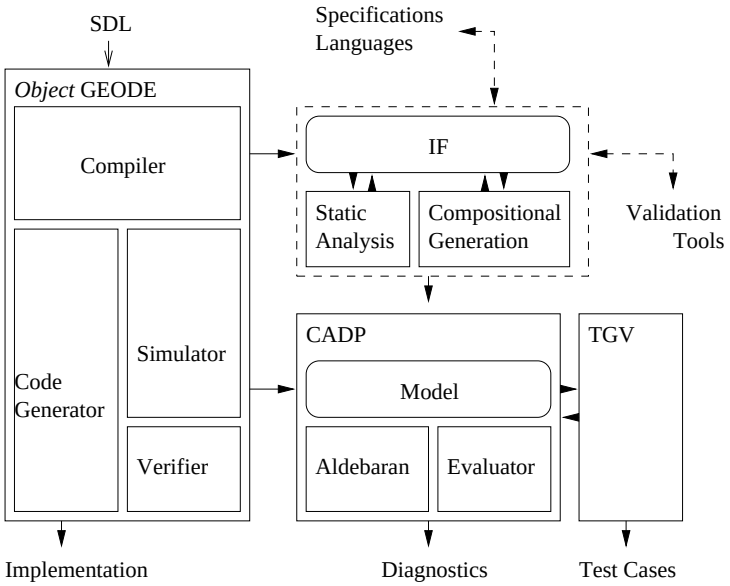

Figure 4: An open validation environment for SDL

the simulator level [KRL97], however using IF offers two main advantages:

- The architecture still allows connections with many other specification languages or tools. Thus, even specifications combining several formalisms could be translated into a single IF intermediate code and globally verified.
- The use of an intermediate program representation where all the variables, timers, buffers and the communication structure are still explicit, allows to apply methods such as static analysis, abstraction, compositional generation. These methods are crucial for the applicability of the model checking algorithms.

*Object*GEODE

*Object*GEODE is a toolset developed by VERILOG supporting the use of SDL, MSC and OMT. It includes graphical editors and compilers for each of these formalisms. It also provides a C code generator and a simulator to help the user to interactively debug an SDL specification. The *Object*GEODE simulator also offers some verification facilities since it allows to perform automatic simulation (either randomly or exhaustively), and behavioral comparison of the specification with special state machines called observers [ALH95].

CADP and TGV

We have been developing for more than ten years a set of tools dedicated to the design and verification of critical systems. Some of them are distributed in collaboration with the VASY team of INRIA Rhône-Alpes as part of the CADP toolset [FGK+96, BFKM97]. We briefly present here two verifiers integrated in CADP (ALDEBARAN and EVALUATOR) and the test sequence

generator TGV [FJJV97] built upon CADP jointly with the PAMPA project of IRISA. These tools apply model-checking on behavioral models of the system in the form of labeled transition systems (LTS). ALDEBARAN allows to compare and to minimize finite LTS with respect to various *simulation* or *bisimulation* relations. This allows the comparison between the observable behavior of a given specification with its expected one, expressed at a more abstract level. EVALUATOR is a model-checker for temporal logic formulas expressed on finite LTS. The temporal logic considered is the alternating-free μ-calculus. TGV aims to automatically generate test cases for conformance testing of distributed systems. Test cases are computed during the exploration of the model and they are selected by means of *test purposes*. Test purposes characterize some abstract properties that the system should have and one wants to test. They are formalized in terms of LTS, labeled with some interactions of the specification. Finally, an important feature of CADP is to offer several representations of LTS, enumerative and symbolic ones based on BDD, each of them being handled using well-defined interfaces such as OPEN-CAESAR [Gar98] and SMI [Boz97].

SDL2IF and IF2C

To implement the language level connection through the IF intermediate representation we take advantage of a well-defined API provided by the *Object*GEODE compiler. This API offers a set of functions and data structures to access the abstract tree generated from an SDL specification. SDL2IF uses this abstract tree to generates an IF specification operationally equivalent to the SDL one.

IF is currently connected to CADP via the *implicit model representation* feature supported by CADP. IF programs are compiled using IF2C into a set of C primitives providing a full basis to simulate their execution. An exhaustive simulator built upon these primitives is also implemented to obtain the explicit LTS representation on which all CADP verifiers can be applied.

4.2 Static analysis

The purpose of static analysis is to provide global informations about how a program manipulates data without executing it. Generally, static analysis is used to perform global optimizations on programs [ASU86, WZ91, Muc97]. Our goal is quite different: we use static analysis in order to perform model reductions before or during its generation or validation. The expected results are the reduction of the state space of the model or of the state vector.

We want to perform two types of static analysis: *property independent* and *property dependent* analysis. In the first case, we use classic analysis methods such as live variable analysis or constant propagation, without regarding any particular property or test purpose we are interesting to validate. In the second case, we take into account informations on data involved in the property and propagate them over the static control structure of the program. Presently, only analysis of the first type is implemented but, we are also investigating constraint propagation and more general abstraction techniques. For instance, through the connection with INVEST we will be able to compute abstract IF programs using general and powerful abstraction techniques.

Live variables analysis

A variable is *live* in a control state if there is a path from this state along which its value can be used before it is redefined. An important reduction of the state space of the model can be

obtained by taking into account in each state only the values of the live variables.

More formally, the reduction considered is based on the relation $\sim_{live}$ defined over model states: two states are related if and only if they have the same values for all the live variables. It can be easily proved that $\sim_{live}$ is an equivalence relation and furthermore, that it is a bisimulation over the model states. This result can be exploited in several ways. Due to the local nature of $\sim_{live}$ it is possible to directly generate the quotient model w.r.t. $\sim_{live}$ instead of the whole model without any extra computation. Exactly the same reduction is obtained when one modifies the initial program by introducing systematic assignments of non-live variables to a particular value. This second approach is presently implemented for IF programs.

Consider now the token ring protocol example. In the `idle` state the live variables are `round` and `worried`, in the `critical` state only `round` is live, while variables `sender`, `adr` and `rnd` are never live. The reduction obtained by the live reduction is shown in Table 1 (line 3).

Constant propagation

A variable is *constant* in a control state if its value can be statically determined in the state. Two reductions are possible. The first one consists in modifying the source program by replacing constant variables with their value. Thus, it is possible to identify and then to eliminate parts of dead code of the program e.g. guarded by expressions which always evaluates to `false`, therefore to increase the overall efficiency of the program. The second reduction concerns the size of the state vector: for a control state we store only the values of the non-constant variables. The constant values do not need to be stored, they can always be retrieved by looking at the control state.

Note that, both of the proposed reductions do not concern the size of the model, they only allow to improve the state space exploration (time and space). However, this kind of analysis may be particularly useful when considering extra information about the values assigned to variables, extracted from the property to be checked.

4.3 Compositional generation

As shown in the previous section, efficient reductions are obtained by replacing a model M by its quotient w.r.t an equivalence relation like $\sim_{live}$. However, stronger reductions can be obtained by taking into account the properties under verification. In particular, it is interesting to consider a weaker equivalence R — which should be a congruence for parallel composition —, able to abstract away non observable actions. The main difficulty is to obtain the quotient M/R without generating M first.

A possible approach is based on the "divide and conquer" paradigm: it consists in splitting the program description into several pieces (i.e., processes or process sets), generating the model M_i associated to each of them, and then composing the quotients M_i/R. Thus, the initial program is never considered as a whole and the generated models can be kept small.

This compositional generation method has already been applied for specification formalisms based on *rendez-vous* communication between processes, and has been shown efficient in practice [GLS96, Val96, KM97]. Surprisingly, to our knowledge it has not been investigated within an SDL framework, may be, because buffers raise several difficulties or due to lack of suitable tools.

To illustrate the benefit of a compositional approach we briefly describe here its application to the token ring protocol:

1. We split the IF description into two parts, the first one contains processes S_1 and S_2 and the second one contains processes S_3 and S_4. For each of these descriptions the internal buffer between the two processes is *a priori* bounded to two places. Note that, when a bounded buffer overflows during simulation, a special *overflow* transition occurs in the corresponding execution sequence.
2. The LTS associated with each of these two descriptions are generated considering the "most general" environment, able to provide any potential input. Therefore, the *overflow* transitions appear in these LTS (`claim` and `token` can be transmitted at any time).
3. In each LTS the input and output transitions relative to the internal buffers (Q_2 and Q_4) are hidden (i.e., renamed to the special τ action); then these LTS are reduced w.r.t an equivalence relation preserving the properties under verification. For the sake of efficiency we have chosen the branching bisimulation [vGW89], also preserving all the safety properties (e.g. mutual exclusion).
4. Each reduced LTS is translated back into an IF process, and these two processes are combined into a single IF description, including the two remaining buffers (Q_1 and Q_3). It turns out that the LTS generated from this new description contains no *overflow* transitions (they have been cut off during this last composition, which confirms the hypothesis on the maximal size of the internal buffers).

The final LTS is branching bisimilar to the one obtained from the initial IF description. The gain, obtained by using compositional generation in addition to static analysis, can be found in Table 1 (line 4).

Results

We summarize in Table 1 the size of the LTS obtained from the token-ring protocol using several generation strategies.

	Generation method	*Number of states*	*Number of transitions*
1	*Object*GEODE	3018145	7119043
2	IF	537891	2298348
3	IF + live reduction	4943	19664
4	IF + compositional generation	1184	4788

Table 1: LTS obtained for the token ring example

The difference between the model generated by *Object*GEODE (line 1) and the one obtained from IF (line 2) are due to the following reasons:

- the handling of timer expirations in *Object*GEODE involves two steps: *first* the timeout signal is appended to the input buffer of the process, and *later* it is consumed, whereas in IF these two steps are collapsed into a single one, bypassing the buffer.
- *Object*GEODE introduces "visible" states for each informal decision, whereas these states do not appear in the model obtained from IF.

However, the abstraction from these extra states in IF, preserves all relevant properties.

The most spectacular reduction is obtained by the live-reduction: the reduced model is about 100 times smaller than the one obtained by direct generation, preserving all properties (models 2 and 3 are strongly bisimilar).

Finally, when considering as visible only the `open` and `close` signals all four LTS are branching bisimilar to the one shown in Figure 4, which proves, in particular, the mutual exclusion property of the protocol.

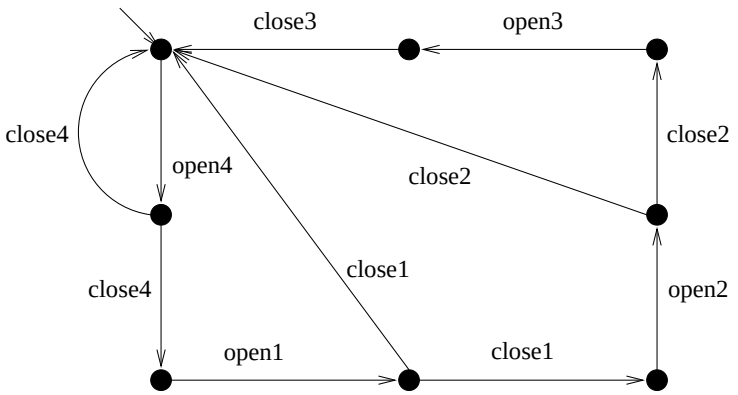

Figure 5: The reduced behavior of the token ring.

5 Conclusion and perspectives

We have presented the formalism IF which is an intermediate representation for SDL allowing the interconnection of tools such as *Object*GEODE, CADP, TGV and specific designed for IF. The use of IF offers several advantages:

- IF has a clear, formal semantics of time, and an explicit notion of atomicity. Moreover, it has some powerful concepts which are interesting for specification, such as different urgency types of transitions, synchronous communication, various buffer types (bounded, unbounded, lossy, ...).

- IF makes verification easier from several points of view:
  - the models generated by state space enumeration are less complex, though equivalent, due to the elimination of certain transient states.
  - IF is a "program level" connection between the toolsets. This allows to combine program analysis techniques (e.g., *static analysis* or *compositional generation*) and classical model checking techniques. We envisage to implement more sophisticated static analysis, such as constraints propagation and more general abstraction techniques.

- The semantics of time in IF is general enough to implement and experiment different time semantics for SDL.

References

[ACD93] R. Alur, C. Courcoubetis, and D.L. Dill. Model Checking in Dense Real Time. *Information and Computation*, 104(1), 1993.

[ALH95] B. Algayres, Y. Lejeune, and F. Hugonnet. GOAL: Observing SDL Behaviors with GEODE. In *Proceedings of SDL Forum'95*. Elsevier Science, 1995.

[ASU86] A. Aho, R. Sethi, and J.D. Ullman. *Compilers: Principles, Techniques and Tools*. Addison-Wesley, Readings, MA, 1986.

[BFG+98] M. Bozga, J-C. Fernandez, L. Ghirvu, S. Graf, L. Mounier, J.P. Krimm, and J. Sifakis. The Intermediate Representation IF. Technical report, Vérimag, 1998.

[BFKM97] M. Bozga, J.-C. Fernandez, A. Kerbrat, and L. Mounier. Protocol Verification with the ALDEBARAN Toolset. *Software Tools for Technology Transfer*, 1, December 1997.

[BLO98] S. Bensalem, Y. Lakhnech, and S. Owre. Computing Abstractions of Infinite State Systems Compositionally and Automatically. In *Proceedings of CAV'98*, volume 1427 of *LNCS*, June 1998.

[BM95] J.A. Bergstra and C.A. Middelburg. Process Algebra Semantics of φSDL. In *2nd Workshop on ACP*, 1995.

[BMU98] J.A. Bergstra, C.A. Middelburg, and Y.S. Usenko. Discrete Time Process Algebra and the Semantics of SDL. Technical Report SEN-R9809, CWI, June 1998.

[Boz97] M. Bozga. SMI: An Open Toolbox for Symbolic Protocol Verification. Technical Report 97-10, Vérimag, September 1997.

[Bro91] M. Broy. Towards a Formal Foundation of the Specification and Description Language SDL. *Formal Aspects on Computing*, 1991.

[BST98] S. Bornot, J. Sifakis, and S. Tripakis. Modeling Urgency in Timed Systems. In *International Symposium: Compositionality - The Significant Difference, Malente (Holstein, Germany)*, 1998. to appear in LNCS.

[CR79] E. Chang and R. Roberts. An Improved Algorithm for Decentralized Extrema-Finding in Circular Configurations of Processes. *Communications of ACM*, 22(5), May 1979.

[FGK+96] J.-C. Fernandez, H. Garavel, A. Kerbrat, R. Mateescu, L. Mounier, and M. Sighireanu. CADP: A Protocol Validation and Verification Toolbox. In *Proceedings of CAV'96 (New Brunswick, USA)*, volume 1102 of *LNCS*, August 1996.

[FJJV97] J.-C. Fernandez, C. Jard, T. Jéron, and C. Viho. An Experiment in Automatic Generation of Test Suites for Protocoles with Verification Technology. *Science of Computer Programming*, 29, 1997.

[Gar98] H. Garavel. OPEN/CÆSAR: An Open Software Architecture for Verification, Simulation, and Testing. In *Proceedings of TACAS'98*, volume 1384 of *LNCS*, March 1998.

[GK97] U. Gläser and R. Karges. Abstract State Machine Semantics of SDL. *Journal of Universal Computer Science*, 3(12), 1997.

[GLS96] S. Graf, G. Lüttgen, and B. Steffen. Compositional Minimisation of Finite State Systems using Interface Specifications. *Formal Aspects of Computation*, 3, 1996.

[GM96] H. Garavel and L. Mounier. Specification and Verification of Distributed Leader Election Algorithms for Unidirectional Ring Networks. *Science of Computer Programming*, 1996.

[God91] J.C. Godskesen. An Operational Semantic Model for Basic SDL. Technical Report TFL RR 1991-2, Tele Danmark Research, 1991.

[GS97] S. Graf and H. Saidi. Construction of Abstract State Graphs with PVS. In *Proceedings of CAV'97, Haifa*, volume 1254 of *LNCS*, June 1997.

[HNSY94] T.A. Henzinger, X. Nicollin, J. Sifakis, and S. Yovine. Symbolic Model Checking for Real-Time Systems. *Information and Computation*, 111(2), 1994.

[KM97] J.P. Krimm and L. Mounier. Compositional State Space Generation from Lotos Programs. In *Proceedings of TACAS'97*, Enschede, The Netherlands, 1997.

[KRL97] A. Kerbrat, C. Rodriguez, and Y. Lejeune. Interconnecting the *Object*GEODE and CADP Toolsets. In *Proceedings of SDL Forum'97*. Elsevier Science, 1997.

[Lan77] G. Le Lann. Distributed Systems – Towards a Formal Approach. In *Information Processing 77*. IFIP, North Holland, 1977.

[MGHS97] S. Mork, J.C. Godskesen, M.R. Hansen, and R. Sharp. A Timed Semantics for SDL. In *FORTE IX: Theory, Applications and Tools*, 1997.

[Muc97] S. Muchnick. *Advanced Compiler Design Implementation*. Morgan Kaufmann Publishers, San Francisco, CA, 1997.

[Val96] A. Valmari. *Compositionality in State Space Verification*, volume 1091 of *LNCS*. 1996.

[Ver96] Verilog. Object*GEODE SDL Simulator - Reference Manual*, 1996.

[vGW89] R.J. van Glabbeek and W.P. Weijland. Branching-Time and Abstraction in Bisimulation Semantics. CS R8911, CWI, 1989.

[WZ91] M.N. Wegman and F.K. Zadeck. Constant Propagation with Conditional Branches. *ACM Transactions on Programming Languages and Systems*, 13(2), April 1991.

[Yov97] S. Yovine. KRONOS: A Verification Tool for Real-Time Systems. *Software Tools for Technology Transfer*, 1(1-2), December 1997.