From UML State-Machine Diagrams to Erlang∗

Ricardo J. Rodríguez, Lars-Åke Fredlund, Ángel Herranz

Babel Group, DLSIIS, Facultad de Informática
Universidad Politécnica de Madrid, Spain
{rjrodriguez, lfredlund, aherranz}@fi.upm.es

Abstract: The Unified Modelling Language (UML) is a semi-formal modelling language useful for representing architectural and behavioural aspects of concurrent and distributed systems. In this paper we propose a transformation from UML State-Machine diagrams to Erlang code. Erlang is a functional language, with strict evaluation, single assignment, and dynamic typing, and with good support for concurrency and distribution. The contribution of this transformation is twofold: it can reduce development time, and moreover it permits us to validate UML diagrams at an early development stage through the use of Erlang-based model checking techniques.

Keywords: UML, Erlang, automatic code generation

1 Introduction

The Software Development Life-Cycle (SDLC) [GJM02] is a process methodology that defines the phases involved for developing and maintaining software systems and software code. One of these phases is the deployment and maintenance phase, where the software is deployed into production and it is maintained to cope with possible faults or minor changes on current requirements. To verify the correctness of the developed software before it is deployed into production can save efforts as problems are detected before arise. A widely used approach for the formal verification of software is model checking [CGP99, BBF+10]. Model checking automatically provides complete proofs of correctness, and explains using counter-examples why a software system or software code is not correct.

In this paper, we provide an approach to be able to verify a software design during design phase through model-checking techniques (more precisely, by using Erlang-based model-checking techniques). This early verification step would improve the SDLC, as design problems can be detected during design phase and not at the end of the process, thus saving efforts and economic losses due to redesign and/or reimplementation. This approach is based in a transformation from UML diagrams (namely, UML State-Machine diagrams (UML-SMs)) to Erlang code. Since a number of effective verification and validation tools are available for Erlang, e.g., the Quviq QuickCheck random testing tool [AHJW06] and the McErlang model checker [FS07], we can use these tools to verify the transformed UML diagram. The well-known dining philosophers problem is used as a motivation example in this paper.

∗ This work was partially supported by ARTEMIS Joint Undertaking nSafeCer under grant agreement no. 295373 and from National funding.
The UML (Unified Modelling Language) [OMG05] is a standard *de facto* modelling language, commonly used by software engineers for high-level system design. UML provides several types of diagrams which allow to capture different aspects and views of the system. A UML model of a system consists of several diagrams which represent the functionality of the system, its static structure, the dynamic behaviour of each system component and the interactions among the system components.

Erlang [Arm97] is a functional and concurrent programming language commonly used for the development of distributed, fault-tolerant, soft-real-time, and non-stop applications [Wik94, AVWW96, Lar09].

The contribution of this transformation is twofold: firstly, it can minimise development time as it automatically generates Erlang code from UML-SMs, and secondly it enables us to apply Erlang-based model-checking techniques to UML specifications. This transformation can save efforts and future costs produced by a redesign and/or reimplementation of a designed software at the end of its SDLC. Besides, to the best of our knowledge, our proposal is the first one that provides an automatic generation of Erlang code from UML-SMs.

The remainder of this paper is as follows. Section 2 introduces some concepts in more detail, such as UML-SMs and the Erlang programming language. Then, Section 3 introduces our approach to transform a UML-SM into Erlang code. Section 4 discusses some approaches in the literature relevant to our work. Lastly, Section 5 provides concluding remarks and future research directions.

## 2 Background

In this section we provide some background needed to follow the rest of this paper. In first place, we introduce the UML State-Machine diagrams (UML-SMs). Finally, the Erlang programming language is introduced.

### 2.1 UML State-Machine Diagrams

The Unified Modelling Language (UML) [OMG05] provides several types of diagrams that allow to capture different aspects and views of the system. UML diagrams can be categorised in: static diagrams, which are intended to model the structure (logical and architectural) of the system; behavioural diagrams, which are intended to describe system dynamics (its dynamic behaviour and the interactions among the system components); and organisational diagrams, which allows to reduce the complexity of the system by organising it in a modular way.

In this paper, we focus on UML State-Machine diagrams (UML-SMs) that model the dynamic behaviour of a system through finite state-transitions. We assume that such a system is composed of several components intercommunicated among them by sending asynchronous messages. The UML-SM semantics is the one proposed by Harel in [Har87] for *state-charts*. Graphically, a UML-SM is a directed graph in which nodes denote states and connectors denote state transitions. A state is represented by a rounded rectangle labelled with a state name, while transitions are represented by arrows labelled with the triggered events followed (optionally) by effects. A transition can optionally have a guard, indicated between brackets. The initial transition origi-
Figures 1 and 2 show UML-SMs describing the behavior of a philosopher and a fork of the “dining philosophers” example. We assume that the thinking time of each philosopher is equal to 1000 units of time. We also assume that each philosopher first grabs his left fork, and then he waits holding it until his right fork can be acquired. Note that this “eating protocol” leads to a deadlock, i.e., a situation where philosophers are waiting for some event which will never happen.

The initial transition of a philosopher in Figure 1 needs two parameters, \( f_L \) and \( f_R \), which represent respectively his left and right forks. Note that in the corresponding UML class diagram, the Philosopher class would have a double association with the Fork class, thus this initial parameters would be no longer needed. The first state of a philosopher is the Thinking state. This state has a do action, which sets a timeout of a duration equal to 1000 units of time. The single exit transition of this state has a guard that indicates the timeout has expired, and an effect of sending an asynchronous acquire(self) message to \( f_L \). After sending such a message, a philosopher moves to the Hungry state. From this state he moves to the Waiting state when an event acquired is received, and sends also an acquire(self) message to \( f_R \). Once the acquired event is received, a philosopher enters the Eating state, where starts the eatSpaghetti() do activity. When this activity has finished, the release() message is sent to both \( f_L \) and \( f_R \) before exiting the state and move to the Thinking state, starting the cycle again.

The behavior of a fork, depicted in Figure 2 is much simpler. It starts in the Idle state, and moves to the BeingUsed state when it receives an acquire(p) event. This parameter \( p \) is the philosopher who triggers the event. When entering in the BeingUsed state, it sends an acquired event to \( p \), as indicated by the entry activity. The statement acquire / defer in the BeingUsed indicates that all acquired events received in this state are deferred, i.e., stored for later handling. When the release event is received, a fork moves to the Idle state, starting its cycle again. Note that a deferred event could trigger immediately the transition to the BeingUsed state.
2.2 The Erlang Programming Language

Erlang [AVWW96] is a functional concurrent programming language created by the Ericsson company in the 1980s. Ericsson is still maintaining the main Erlang implementation, but it is available as open source since 1998. The chief strength of the language is that it provides excellent support for concurrency, distribution and fault tolerance on top of a dynamically typed and strictly evaluated functional programming language. Concurrency is achieved by lightweight processes communicating through asynchronous message passing. Although Erlang is not a new language, it has experienced considerable growth in users in recent years. This is due in most part to its focus on message passing instead of variable sharing as the main communication mechanism, which enables programmers to write robust and clean code for modern multiprocessor and distributed systems.

Today Erlang is used by Ericsson and many other companies (T-Mobile (UK), and many smaller start-up companies such as e.g. Interoud in Spain and Klarna in Sweden) to develop industrial applications, often implementing crucial internet server-side applications. Examples include a high-speed ATM switch developed at Ericsson with over a million lines of Erlang code which had to meet very challenging requirements on software reliability and overall system availability [BRA+99, WAB02], parts of Facebook chat, Apache CouchDB – a distributed, fault-tolerant and schema-free document-oriented database accessible via a RESTful HTTP/JSON API, etc.

3 A Transformation Approach

This section introduces the transformation approach from UML State-Machine diagrams (UML-SMs) to Erlang code. The aim of this section is to show how our approach works. Basically, for each UML-SM, an Erlang module is generated following the steps described in Algorithm 1.

The transformation algorithm of the approach is described in Algorithms 1 and 2, as pseudo-code. As input, it receives a UML-SM $\mathcal{D}$. As output, it returns an Erlang module $\mathcal{E}$ with the same behaviour as $\mathcal{D}$.

Let us recall how this transformation algorithm works by means of the running example as defined by the UML-SMs depicted in Figures 1 and 2. The Erlang code generated by Algorithms 1 and 2 is listed in Code 1 and 2.

First, let us consider the UML-SM in Figure 1 that describes the behaviour of a philosopher. In the translation, the two initial parameters $fL, fR$ of the initial transition are stored in a set $\mathcal{P}$. The set $\mathcal{P}$ is passed along as a parameter to $\text{start}$ and of all other Erlang functions. In the set $\mathcal{E}$ we collect all the events of the UML-SM that is being transformed. This is needed because, unlike Erlang messages, UML events sent between different UML-SMs are lost if they are not handled in the current state of a UML-SM. Then, the $\text{module}$ and $\text{export}$ Erlang compiler attributes are output. Each UML-SM transformed to Erlang has a dedicated function $\text{start}$ which is used to create a new Erlang process executing the (translated) state machine.

The first state is the Thinking state. An Erlang function is defined with that name, and receiving the parameters contained in $\mathcal{P}$. The do activity $\text{setTimeout}$ and the conditional exit trigger $\text{trigger TOUT expired}$ are assumed to represent a timeout activation and expiration in UML-SMs, and generate a $\text{receive}$ clause $\text{after}$ with a timeout corresponding to the parameter
Input: A UML-SM \( D \)

Output: An Erlang source file \( E \)

1. Store in \( D \) the parameters of the initial transition in \( D \)
2. Create a module named as \( D \) and an export declaration of start function
3. Create a start function with parameters \( D \) and invoke the first state \( S \) of \( D \)
4. Create a set of events \( E \) containing all trigger events in \( D \)
5. \textbf{forall the state} \( S \) of \( D \) \textbf{do}
   6. Create a function \( F \) with parameters \( D \) named as state \( S \)
   7. Execute Algorithm 2 with entry activity as input (if any), and invoke the result
   8. Execute Algorithm 2 with do activity (if any)
   9. \textbf{if state} \( S \) \textbf{has not a do activity called setTimeout} \textbf{then}
      10. Execute Algorithm 2 with do activity, and invoke the result
   11. \textbf{end}
   12. \textbf{if state} \( S \) \textbf{has output transitions then}
      13. \textbf{if some output transition has a guard and/or a triggered event then}
      14. Create an Erlang receive skeleton
      15. \textbf{forall the output transition} \( O \) of \( S \) \textbf{triggered by an event do}
      16. Map the event of \( O \) to an Erlang message reception
      17. Create a when clause if output transition \( O \) has a condition guard
      18. Execute Algorithm 2 with output transition \( O \), and invoke the result
      19. Invoke the Erlang function that represents the exit action (if any)
      20. Invoke the function \( F' \) representing the destination of \( O \)
      21. \textbf{end}
      22. \textbf{if exists some event in} \( E \) \textbf{not deferred in} \( S \) \textbf{and not handled then}
      23. Create an Erlang code to ignore such an event by invoking \( F \)
      24. \textbf{end}
      25. \textbf{if state} \( S \) \textbf{has a do activity called setTimeout and an output transition} \( O \) \textbf{is triggered when the timeout has expired then}
      26. Create an after clause
      27. Execute Algorithm 2 with output transition \( O \), and invoke the result
      28. Invoke the Erlang function that represents the exit action (if any)
      29. Invoke the function \( F' \) representing the destination of \( O \)
      30. \textbf{end}
      31. \textbf{end}
      32. Invoke the Erlang function that represents the exit action (if any)
      33. Invoke the function \( F' \) representing the destination of \( O \)
      34. else
      35. Invoke the Erlang function that represents the exit action (if any)
      36. end
   12. \textbf{end}

\textbf{Algorithm 1}: A transformation from UML State-Machine diagrams to Erlang code.
Input: A state $S$ and an activity $A$
Output: A piece of Erlang code

1 if $A$ concerns another state machine $S M'$ then
2 Send a message with the mapped $A$ (and parameters) to the Erlang process corresponding to state machine $S M'$
3 else
4 Map $A$ to Erlang code
5 end

Algorithm 2: Transformation of any activity to Erlang code.

Code 1: Erlang code generated from basic UML-SM in Figure 1.

```erlang
-module(philosopher).
-export([start/2]).

start(FL, FR) ->
    spawn(fun() -> thinking(FL, FR) end).

thinking(FL, FR) ->
    receive
    X -> thinking(FL, FR)
    after 1000 -> FL!{acquire, self()}, hungry(FL, FR)
    end.

hungry(FL, FR) ->
    receive
    acquired -> FR!{acquire, self()}, waiting(FL, FR)
    end.

waiting(FL, FR) ->
    receive
    acquired -> eating(FL, FR)
    end.

eating(FL, FR) ->
    eatSpaghetti(),
    FL!release,
    FR!release,
    thinking(FL, FR).
```
Code 2: Erlang code generated from basic UML-SM in Figure 2.

```erlang
-module(fork).
-export([start/0]).

start() ->
    spawn(fun() -> idle() end).

idle() ->
    receive
        {acquire, P} -> beingUsed(P);
        X -> idle()
    end.

beingUsed(P) ->
    P!acquired(),
    receive
        release -> idle()
    end.
```

of the setTimeout activity. As there are events that are not handled in this state and it has output transitions triggered by events, a receive clause receives any value and recurses. The effect $\text{fl.acquire(self)}$ is transformed to a signal $\text{acquire, self()acquire, self()}$ sent to $\text{fl}$ in Erlang code. Finally the function that represents the Hungry state is called. This state has only an exit transition triggered by the acquired event; the construct is translated into a receive clause receiving acquired. The effect of this event sends a signal $\text{acquire, self()}$ to $\text{fr}$ and calls the function that represents the Waiting state. As previously, a receive clause is output where receiving acquired results in a call to the function corresponding to the Eating state. Finally, the do and exit activities are transformed to their equivalent Erlang code. More precisely, the exit activity that involves invocations of release() method in $\text{fl}$ and $\text{fr}$ are transformed to $\text{fl!release}$ and $\text{fr!release}$ in Erlang.

Let us consider now the UML-SM of the behaviour of a fork (Figure 2). The set of parameters of the initial transition (none) is stored in $\mathcal{P}$. Then, module(fork) and export([start/0]) compiler attributes are created in Erlang code. After this, the start function spawns a new process which behaves as the initial state of the state machine (the Idle state). As before, all the events of this UML-SM are stored in a set $\mathcal{E}$ (in this case, acquire and release).

The UML-SM depicted in Figure 2 has two states, Idle and BeingUsed. Let us focus first on the Idle state. A function idle() is created, without parameters, representing this state. This state does not have an entry activity, nor a do activity. It has just an output transition triggered by an event acquired(p), that is mapped to the reception of a message with a parameter P. When received, it invokes the function beingUsed(), which represents the other state of $\mathcal{D}$. As the event release is not handled in this state, an clause that handles any message is output invoking itself.
The entry action of BeingUsed is mapped to a sending of an acquired message to P, as identified by Algorithm 2). As before, it has an output transition triggered by the release event. When received, it invokes the idle() function, returning to the initial state. In this case, as the acquire event is deferred, there is no need to include a clause receiving any other message.

Finally, note how the absence of an acquire / defer statement in idle gives rise to an Erlang receive clause X -> idle() in Idle, whereas the occurrence of an acquire / defer statement in the state BeingUsed results in the absence of the corresponding receive clause X -> beingUsed(). Thus a signal cquire, P’ acquire, P’ that arrives in the state BeingUsed is not thrown away but deferred, i.e., stored and later handled in the Idle state.

The Erlang code resulting from the transformation can be checked using the McErlang model checker [FS07] without modifications. We can for example compose a system of N philosophers in Erlang using the function call run(N) below (located in the module dining):

```erlang
run(N) ->
    Forks = lists:map (fun (_) -> fork:start() end, lists:seq(1,N)),
    lists:foreach (fun ({L,R}) -> philosopher:start(L, R) end, adjacent(Forks)).
```

Invoking McErlang to check whether a configuration of two philosophers is deadlock free has the following result:

```
$ erl
> mce:start
(#mce_opts{program=fun () -> dining:run(2) end, monitor=mce_mon_deadlock}).
...
*** Monitor failed
monitor error:
deadlock
```

As we commented in Section 1, the dining philosophers problem modelled as illustrated in Figures 1 and 2 has a deadlock problem which evidently was detected by the McErlang model checker.

## 4 Related Work

Automatic code generation from formal specifications is considered one of the silver bullet of software engineering. In this respect, state machines is one of the most successful formalism since its semantics is directly executable. UML state-charts is a variation of state machines so it is not strange that countless automatic code generation approaches can be found in the literature.

The single-threaded approach is simple and the inherent synchronous view of operations is relatively straightforward to compile into ordinary programming language constructs: case statements and function calls [HG96].
Here we restrict the consideration of related works to the more complex problem of automatic code generation for state-charts with multi-threaded behaviour and asynchronous communication (signals).

Most of such works rely on the implementation of a general interpreter that captures the executable semantics of a particular specification language. Given a set of concrete state machines written in a specification language, a translator compiles them into structures to represent states, nodes and transitions that the general interpreter operates on. Relevant works following this road are for example [PM03, AFL+10] (C code from extended finite state machines with a synchronous semantics for the moment), [NT03, KNNZ00] (asynchronous semantics with a state table to react to events), [TKUY01] (based on design pattern forms implementing state machines), [KM02] (with a Java thread per state-chart and Java objects to represent event queues).

All these works relies on the ability of the interpreter to manage multi-threading and event queues, two aspects where the Erlang programming language is particularly strong. As a consequence of the excellent support available in an Erlang run-time system for multiprogramming, in our approach no such interpreter is needed. Instead state-machine concepts such as machines, state, nodes and transitions are directly compiled into Erlang efficient constructs: processes, call by value function calls, efficient message passing, and event queues. As a result we expect our implementation of UML state-machines in Erlang to be very high-performing with regards to performance criteria such as the maximum number of messages sent per seconds, the maximum number of concurrent state machines supported, etc. We aim at performing some experiments for comparing performance results as future work.

Other works that worth to be mentioned transform state-charts to other models that are verifiable, such as [GLM99, LMM99, CEC12]. In [GLM99], a branching time model-checking approach is presented for automatically verify the correctness of UML-SMs using JACK as verification environment. The work in [LMM99] introduces an approach where linear-time model checking of UML-SMs is addressed by translating them to PROMELA, the modelling language of the SPIN model-checker. In [CEC12], UML class diagrams, UML-SMs and UML Communication diagrams are considered and automatically transformed to a model that can be verified by means of Maude LTL model checker.

5 Conclusions

The Unified Modelling Language (UML), a de facto standard as semi-formal modelling language, provides several types of diagrams that can capture different aspects and views of a modelled system. For instance, UML is useful for representing architectural and behavioural aspects of concurrent and distributed systems.

This paper explores the idea of automatic transformation from UML State-Machines diagrams (UML-SMs), which describe the dynamic behaviour of a system, to the Erlang programming language. Erlang is a functional language, with strict evaluation, single assignment, and dynamic typing, and with good support for concurrency and distribution. The contribution of this transformation can be summarized:

- It can reduce development time by automatically generating Erlang skeleton code from UML diagrams; and
• it enables validation of UML-SMs at an early development stage through the use of Erlang-based model checking and testing techniques.

Besides, this transformation also provides an alternative behavioural code pattern complementing frequently used Erlang code patterns (libraries) such as e.g. the generic server library \texttt{gen_server} and the finite-state machine library \texttt{gen_fsm}.

As future work, we plan also to significantly extend the scope of the transformation to provide support for additional UML-SM constructs. For example, we would like to provide support for preemptive UML-SM activities in Erlang. Moreover, supporting UML-SM constructs such as e.g. sub-states, entry, exit or alternative points, and call actions would be highly beneficial. Lastly, we aim at developing a plugin for some UML CASE tool (such as Papyrus, MagicDraw or ArgoUML) to package the transformation.

Bibliography


