Monitoring Distributed Systems with Distributed PolyLarva

Ian Cassar, Adrian Francalanza, Christian Colombo
Faculty of ICT, CS Department
University of Malta
Emails: {ian.cassar.10, adrian.francalanza, christian.colombo}@um.edu.mt

***Abstract*—PolyLarva is a language-agnostic RV tool, which converts a PolyLarvaScript into a monitor for a given system. While an implementation for PolyLarva exists, the language and its compilation have not been formalised. We therefore present a formal implementation-independent model which describes the behaviour of PolyLarvaScript, comprising of the μLarvaScript calculus and of a set of operational semantics. This allows us to prove important properties, such as determinism, and also enables us to reason about ways of re-designing the tool in a more scalable way. We also present a collection of denotational mappings for μLarvaScript converting the constructs of our calculus into constructs of a formal actor-based model [7], thus providing an Actor semantics for μLarvaScript. We are also able to prove certain correctness properties of the denotational translation such as that the denoted Actors behave in a way which corresponds to the behaviour described by our implementation-independent model. We finally present DistPolyLarva, a prototype implementation of the distributed PolyLarva tool, which implements the new actor-based semantics over a language that can natively handle distribution and concurrency called Erlang.**

I. Introduction

Runtime Verification (RV) [3] is a dynamic [3], [5] verification technique which invokes monitoring procedures at runtime so as to verify that the current execution, of the system being verified, is correct with respect to a given specification. It is therefore important that RV tools should be verified for correctness themselves, thus making users more confident in trusting and relying on such tools for verification. As RV tools weave additional monitoring code into the system being verified, an inevitable runtime overhead is imposed upon the system. Moreover, monitoring demands may quickly increase especially when monitoring distributed systems, as these systems are able to scale up rapidly. Such a drastic increase in monitoring load would impose a negative effect on the monitoring efficiency, thus also affecting the performance of the monitored system. For this reason, various ways are being explored by which this overhead can be minimized [6], [7]. Concurrency and parallelisation provide a way of decreasing these overheads by exploiting tightly-coupled, multi-core architectures. When dealing with high monitoring demands, distributed monitoring may also be a more scalable and feasible alternative for increasing monitoring efficiency as distribution also enables the exploitation of loosely-coupled processing units.

A. PolyLarva

PolyLarva [11], [6] is a language agnostic RV compiling tool, which when given an RV specification written in polyLS (short for poly-LarvaScript), creates the additional monitoring computation for a given system. polyLS language provides an event-driven monitoring framework by which one can identify and specify a number of monitoring requests, that each monitor can handle, in terms of *Events*. For each monitor, one can also specify a set of monitoring checks and handling procedures in terms of *Conditions* and *Actions*. These three components are then associated with one another in the monitor's list of *rules*.

Example 1.1.

BR1=ReqFunds(Usr,Sum)/ !IsUsrValid(Usr)→WarnUsr();
BR2=ReqFunds(Usr,Sum)/ !EnoughFunds(Sum)
→WarnUsr();
BR3=ReqFunds(Usr,Sum)→TransferFunds(Usr,Sum);

Example 1.1 shows a sample pseudo-script defining three rules all of which are related to the same *ReqFunds* event. Whenever the monitor receives an event e from the system, it starts by matching it with the event pattern of the first rule in the sequence, i.e., BR1. If e is for example of the form ReqFunds("usr1",9000), it would match the rule's pattern ReqFunds(Usr,Sum) and as a result replace every occurrence of variables Usr by "usr1" and Sum by 9000, else the event is matched to the event pattern of Rule BR2. When e matches the event pattern of BR1, the associated condition !IsUsrValid(Usr) would change into !IsUsrValid("usr1") and evaluate to either true or false. If true, the rule's action WarnUsr() would also execute, otherwise the rule is ignored and the event would be matched with the pattern of BR2.

B. Problem Definition

There are several problems with the original PolyLarva [11]:

(i) PolyLarva was developed using a compiler-driven[1] [6] approach, hence no formal language semantics exist for polyLS. This is not ideal as one would require a thorough understanding of how the PolyLarva compiler is implemented, in order to understand the behaviour of the language constructs. This also makes it hard to understand how the PolyLarva compiler interprets and converts the polyLS constructs into monitoring constructs and even harder to improve it.

(ii) Since no formal model exists for PolyLarva, there also does not exist any type of formal proof which substantiates the validity and the correctness of the PolyLarva compiler. This makes it hard for users to trust that our RV tool would correctly verify their system, as specified in their compiled script.

[1]The aim was to develop an actual compiler implementation.

(iii) Due to the shared-state, multi-threaded design of the synthesised monitor, POLYLARVA does not provide a foundation by which the compiled monitor could be easily scaled up in order to make use of distributed architectures. A distributed design would introduce more areas that can be explored in order to exploit the advantages of distributed architectures so as to be capable of handling higher monitoring demands.

II. THE HIGH-LEVEL MODEL

The main focus of this model is that of providing a formal, implementation-independent description of the runtime behaviour of polyLS. In fact, this model formally describes the behaviour of the most essential constructs of POLYLARVA's polyLS. It consists of the μLarvaScript calculus, derived from the original polyLS language, and from a series of *operational semantics* which provide a formal implementation-independent description of the runtime behaviour of the constructs in our calculus.

A. The μLarvaScript Calculus

The following μLarvaScript calculus is made from *abstract syntax*, meaning, that the language is treated as if it has already been *parsed* and hence assumed to be syntactically correct. It assumes denumerable sets of values $v \in Val$, variables $x \in Var$, and identifiers $i \in Id = Val \cup Var$, within its other constructs. It also assumes the inclusion of predicate functions, which are used in conditions so as to perform checks on the monitor's state. The entire μLarvaScript calculus is defined below.

Table 3.1. $M \in Mons ::= \langle s,d \rangle \mid M_0 \| M_1$
$d \in RulesList ::= r;\ d \mid \varepsilon$
$r \in Rule ::= ((q,c) \mapsto a)$
$n \in EventName \supseteq \{mthdInvoked,\ exThrown,\ mthdRet,\ internal\}$
$s \in State : Var^* ::= \{\ x_0\ ,\ x_1\ ,\ \dots\}$
$e \in Event ::= n(v_0 \dots v_k)$
$t \in EventStream ::= e;t \mid \varepsilon$
$q \in Query ::= n(i_0 \dots i_k)$
$b \in Boolean ::= true \mid false$
$c \in Conditions ::= b \mid !(c) \mid c_1\ \&\&\ c_2 \mid p(v_0 \in Val, \dots, v_k \in Val)$
$a \in Actions : (State \rightarrow State) ::= stop \mid fail \mid noOp \mid a_1,a_2 \mid update(S,F) \mid load(M)$

A monitoring system consists of a collection of concurrent monitors, $M_0 \| M_1$, where each individual monitor, $\langle s, d\rangle$, possesses its own current *local state* "*s*" and its own *rule list* "*d*". Monitors are able to process sequences of events "*t*" which are forwarded to the monitor by the system. The state of a monitor, "*s*", comprises a set of local variables, $\{x_0, ..., x_n\}$, while a rule list, "*d*" consists of a sequence of *rules*. Each individual rule, of the form $((q, c) \mapsto a)$, binds an event query "q", and a condition "c", with an action "a". Although an event query, "q", has a very similar structure to an event, "e", the latter describes an actual event which originates from the system being monitored. Conversely, the former is used to describe a *pattern* which states that the host monitor is able to handle system events which match the pattern denoted by the query. A condition "c", can be a boolean formula or a predicate which performs checks on the monitor's current state and on the values passed as its arguments, so as to yield a boolean result. Similarly, an action "a" is a deterministic function which processes a sequence of operations which can possibly modify the monitor's current state. The following example script shows the same rules defined in Example 1.1, written in μLarvaScript syntax:

Example 3.1.

$\langle\{usr1, funds\},$
$((ReqFunds(Usr, Sum), !IsUsrValid(Usr)) \mapsto WarnUsr());$
$((ReqFunds(Usr, Sum), !EnoughFunds(Sum)) \mapsto WarnUsr());$
$((ReqFunds(Usr, Sum), true) \mapsto TransferFunds(Usr,Sum);\rangle$

B. Operational Semantics

The operational semantics for polyLS consists of a group of reduction rules. These rules, defined below, are segmented into high level monitoring rules, denoted by the high-level relation ($\longmapsto$), and into the low-level monitoring rules, denoted by the low-level relation ($\rightarrow$) relation. These rules serve to indicate how a collection of monitors would behave when they receive a system event. In fact, they describe how an event is *ignored* when no monitor in the collection is able to handle the event. They also describe how an event is *consumed* and removed from the event stream if there exists a single monitor which is capable of consuming that event.

μLarvaScript High-Level Monitoring rules:

$$\text{rHlMon1}\ \frac{t \rhd M \rightarrow t' \rhd M'}{t \rhd M \longmapsto t' \rhd M'} \qquad \text{rHlMon2}\ \frac{e;t \rhd M \not\rightarrow}{e;t \rhd M \longmapsto t \rhd M}$$

μLarvaScript Low-Level Monitoring rules:

$$\text{rParMon}\ \frac{t \rhd M_0 \rightarrow t' \rhd M_0'}{t \rhd M_0 \| M_1 \rightarrow t' \rhd M_0' \| M_1}$$

$$\text{rMonEvtHandling}\ \frac{e, s, d \Downarrow s'}{e;t \rhd \langle s, d\rangle \rightarrow t \rhd \langle s', d\rangle}$$

μLarvaScript Event Consumption rules:

$$\text{rConsAx}\ \frac{\text{matches(q,e)}=\sigma \qquad s, c\sigma \Downarrow^c \text{true}}{e, s, ((q,c) \mapsto a); d \Downarrow a\sigma(s)}$$

$$\text{rConsInd1}\ \frac{\text{matches(q,e)}\neq\sigma \qquad e, s, d \Downarrow s'}{e, s, ((q,c) \mapsto a); d \Downarrow s'}$$

$$\text{rConsInd2}\ \frac{\text{matches(q,e)}=\sigma \qquad s, c\sigma \Downarrow^c \text{false} \qquad e, s, d \Downarrow s'}{e, s, ((q,c) \mapsto a); d \Downarrow s'}$$

The high-level monitoring rules state that a high-level reduction is only possible if $t \rhd M$ is able to reduce into $t' \rhd M'$ through some low-level reduction. However, if a low-level reduction is unable to reduce $e; t \rhd M$ into some other form, then it means that event "e" will be *ignored*, thus reducing $e; t \rhd M$ into $t \rhd M$ where "t" is the tail of "$e; t$" and "M" remained unmodified by the reduction.

RPARMON is a low-level inductive rule which determines whether $t \rhd M_0 \| M_1$, consisting of a sequence of events "t" and monitor collection "$M_0 \| M_1$", is capable of reducing into $t' \rhd M_0' \| M_1$, where "t'" is a modified stream of events while "$M_0' \| M_1$" represents a modified monitor collection. It states that such a reduction is only allowed if *there exists* some sub-monitor collection "M_0", which when given the same event stream, "t", reduces it into event stream "t' " and "M_0'", i.e., a modified version of collection "M_0". rMONEVTHANDLING is an axiom which specifies that a monitor, of the form "$\langle s, d \rangle$" which is provided with a sequence of events "$e; t$", changes its state to "s'". It also specifies that this reduction is allowed if the event "e", together with the current monitor's state "s" and rule list "d", are able to evaluate into the next state "s'" by using the *Event Consumption Evaluation rules*.

These rules describe how an individual monitor, consisting of state "s" and rule list "d", reacts and behaves in order to handle the received event "e". In fact they indicate that a *successive* state "s'" is derived once the event has been handled by the monitor and removed from the event stream. Hence, the above rules, describe the operational behaviour by which a μLarvaScript monitor consumes a system event. Particularly, these rules define that a modified state "s'" is only produced when the received system event "e" *matches* a query "q" of one of the monitor's rules, which causes condition "c" to evaluate to true, thus invoking an action "a" which modifies state "s" into some "s'".

C. The Single Receiver Property

One of the most prominent properties observed in POLYLARVA was that no matter how many monitors are specified, only a maximum of *one* monitor ends up receiving and handling an event. For this reason we assume that a *sound monitoring specification* is one which coincides with the Single Receiver Property defined by Definition 3.1. Moreover, we will base our arguments and evaluation proofs upon this important property, meaning that any guarantees offered by our models, only apply for sound specifications.

Definition 3.1. The Single Receiver Property.

$$t \rhd M_0 \| M_1 \to t' \rhd M' \quad implies \\ t \rhd M_0 \to t' \rhd M_0' \;\; and \;\; t \rhd M_1 \not\to$$

III. THE DISTRIBUTED-STATE MODEL AND ITS TRANSLATION

This model aims to provide a formal description of the behaviour of the μLarvaScript constructs in a way which is closely related to an actual, distributed-state implementation. In fact, this distributed-state model consists in a formal translation from μLarvaScript constructs to constructs of a formal Actor model for Erlang adapted from [7] by Seychell et al. In this way, the meaning of the μLarvaScript constructs is given in terms of a highly scalable [10], distributed state model, which produces a monitoring system capable of handling larger monitoring demands with the same or better performance. This claim is supported by Gustafson's Law [9].

A. Concurrency, the Actor Model & Erlang

The Actor Model [8] is a highly scalable paradigm [10] which offers a level of abstraction by which both data and procedures can be encapsulated into a single construct.

Actors differ from objects since actors are also concurrent units of execution, each of which executes independently and asynchronously. This fusion of data abstraction and concurrency relieves the developer from having to recur to the explicit concept of a thread in order to make use of concurrency. Moreover, since Actors communicate through Message Passing [8], the developer does not need to develop explicit synchronization mechanisms to prohibit dangerous concurrent access to the data, shared amongst the communicating threads.

Additionally, message passing between these actors is performed asynchronously [8], which means, that an Actor is able to send a message without having to wait for the receiver's response. Conversely, the receiver does not need to be listening for incoming messages in order to receive them since the messages are deposited in the Actor's mailbox.

In order for an actor to retrieve the received data, it must issue a receive command to recover a message from its mailbox. An important factor is that message passing in the Actor model normally assumes fairness, that is, any message sent by an actor to another existing actor, is *guaranteed* to eventually be deposited inside the target actor's mailbox. In addition to this merger between data, functions and concurrency, an actor is also assigned a unique and persistent identifier, which is essential to identify the target destination actor of the message being sent. A case in point is Erlang [13], [2], a programming language which natively implements this model.

Although forms of concurrency are employed in the monitors synthesised by POLYLARVA, this is done through multi-threading and shared state communication [11] using explicit locking mechanisms. As these concurrent monitors do not use a distributed state[2], they can only be executed concurrently on the same machine. This implies that unlike a distributed multi-processing design, a multi-threaded monitor side cannot exploit the full processing capabilities of loosely coupled distributed architectures, making it less scalable [1].

B. Alternative Semantics for μLarvaScript

The denotations in Figure 4.1 convert μLarvaScript constructs into constructs of the formal Actor model for Erlang [7], thus giving Actor semantics to μLarvaScript. Also one must distinguish between the constructs which are declared *within* the denotations and those declared without any denotation. The constructs declared in a denotation are μLarvaScript constructs, for example, abc in $[\![abc]\!]^m$ refer to a μLarvaScript construct, while if abc is not declared in a denotation, then it is a construct of the Erlang model [7].

$[\![t \rhd M]\!]^m$ presents the root denotational function which takes an event stream t and a μLarvaScript monitor specification "M". It then invokes another denotational function $[\![t]\!]^m_{es}$, which creates a coordinating Actor that executes in parallel with the monitoring actors returned by $\mathsf{fst}(\ [\![M]\!]^m_{par}\ )$. Moreover, in order for the denotation $[\![t]\!]^m_{es}$ to keep on reducing, it requires a list of process identifiers[3] (PIDS) returned by $\mathsf{snd}(\ [\![M]\!]^m_{par}\ )$.

The translation $[\![t]\!]^m_{es}$ converts an event stream into a *coordinating actor*, when given a list of PIDS. This special Actor is required to interface with the monitored system and to make sure that the synthesized monitor is behaving in accordance with

[2]"Distributed state" means that each monitor has its own local state and communicate through message passing.

[3]A PID uniquely identifies an Actor.

the Single Receiver Property. In fact, $\llbracket t \rrbracket^m_{es}$ creates an actor with $\llbracket t \rrbracket^m_{mb}$ as its mailbox, meaning that the system events will be delivered to the coordinator's mailbox. Moreover, the coordinator consists of a recursive function which takes a list of PIDS and listens for messages in its mailbox via a recv command. Whenever the coordinator receives the message $\{new, Pid\}$, it signifies that one of the concurrent monitors has issued a $\llbracket load(M) \rrbracket^m_a$ action, so as to dynamically create a new concurrent monitor. For this reason, the coordinator adds the PID of the new monitor to its PID-list and issues a recursive call, to restart listening for other messages. Conversely, when the coordinator reads a system event message, $\{evt, E\}$, it broadcasts the message[4] $e_{msg} \equiv \{\mathsf{self}, E\}$ to all monitors executing concurrently, by using the "$bcast$" function. The coordinator then awaits feedback from the monitors by calling "$await(count)$", where "$count$" is initially set to be the length of the coordinator's PID-list. Moreover, the "$await$" function makes use of a selective receive so as to only retrieve feedback messages, of the form "*ok*" or "*nok*", from all the monitors in its PID-list. This makes sure that only a maximum of *one* monitor has indeed handled the broadcasted event. In fact it issues an error if more than one monitor handles the event, thus signifying that the Single Receiver Property has been violated by the translated monitoring specification.

$\llbracket - \rrbracket^m_{par}$ is a function that converts a μLarvaScript monitor into a meta-level tuple containing a list of monitoring actors together with another list with their PIDS. The meta-functions fst and snd are then invoked at compile-time so as to extract the two separate lists from the denoted meta-tuple. Each actor denoted by $\llbracket \langle s, d \rangle \rrbracket^m_{par}$ is *always* associated with a unique PID, "i", and is initialized with an empty mailbox "ε" so as to wait for event messages of the form {CoordPid, e}, by issuing a "recv" command so as to listen for messages from the coordinator. This command is followed by $\llbracket d \rrbracket^m_d$ which converts a μLarvaScript rule list into an Erlang list of guarded rules. An empty μLarvaScript rule list, is converted by $\llbracket \varepsilon \rrbracket^m_d$ into a guarded rule which matches *any* broadcasted event message. This is required since when a message matches its pattern, the monitor sends a rejection feedback to the coordinator by using "$Coord!\,nok$" and leaves the monitor's current state unmodified.

Each μLarvaScript rule, in a non-empty rule list, is translated through $\llbracket ((q, c) \mapsto a) \rrbracket^m_r$ into an Erlang guarded command. Whenever the guarded rule's tuple query, of the form $\{\text{Coord}, \llbracket q \rrbracket^m_q\}$, pattern matches the structure of the received event in a way which causes condition $\llbracket c \rrbracket^m_c$ to return $true$, the rule sends an "*ok*" feedback message to the coordinator, which signifies that the event has been handled. It then executes the function denoted by $\llbracket a \rrbracket^m_a$ on the monitor's current state, thus generating the next state.

The denotation $\llbracket - \rrbracket^m_s$, for the monitor's state, dictates that the monitor's state variables are converted into a list of Erlang variables. The translation $\llbracket - \rrbracket^m_e$, states that a μLarvaScript event is translated into an Erlang tuple containing the event name and a *tuple of values* created by the system, while the query denotation, $\llbracket - \rrbracket^m_q$, returns an Erlang tuple containing the event name and a *tuple of identifiers*, where each identifier can be either a value or a variable. The condition denotation $\llbracket - \rrbracket^m_c$, converts μLarvaScript conditions into Erlang functions

[4] Where self refers to the coordinator's PID and E is the actual system event received.

Fig 4.1 The formal translation.

$$\llbracket t \rhd M \rrbracket^m \stackrel{def}{=} \llbracket t \rrbracket^m_{es}(\mathsf{snd}(\ \llbracket M \rrbracket^m_{par}\ )) \| \mathsf{fst}(\ \llbracket M \rrbracket^m_{par}\ )$$

$$\begin{array}{l}
\llbracket t \rrbracket^m_{es}(\text{PidList}) \stackrel{def}{=} coord\ [(\mu\ y_{rec}\ \cdot\ \lambda\ X_{lst} \cdot (\\
\quad \text{recv } \{\text{evt,E}\}: \rightarrow\\
\qquad \text{bcast}(\{\ \text{E,self()}\ \}, \text{X}_{lst}),\\
\qquad \text{case await(len(X}_{lst}\text{)-1) of}\\
\qquad\quad 0 \rightarrow y_{rec}(\text{X}_{lst});\\
\qquad\quad 1 \rightarrow y_{rec}(\text{X}_{lst});\\
\qquad\quad _ \rightarrow \text{error}\\
\qquad \text{end}\\
\quad \{\text{new,Pid}\} \rightarrow\\
\qquad y_{rec}(\text{X}_{lst}\text{:Pid});\\
\quad \text{end.)(PidList)} \quad \lhd \quad \llbracket t \rrbracket^m_{mb}\ ]
\end{array}$$

$$\begin{array}{l}
\llbracket M_0 \| M_1 \rrbracket^m_{par} \stackrel{def}{=} (\mathsf{fst}(\ \llbracket M_0 \rrbracket^m_{par}\ ) \| \mathsf{fst}(\ \llbracket M_1 \rrbracket^m_{par}\ )\ ,\\
\qquad \mathsf{snd}(\ \llbracket M_0 \rrbracket^m_{par}\ ) : \mathsf{snd}(\ \llbracket M_1 \rrbracket^m_{par}\ )\ )
\end{array}$$

$$\begin{array}{l}
\llbracket \langle s, d \rangle \rrbracket^m_{par} \stackrel{def}{=} (i[(\mu y_{rec} \cdot \lambda X_{state} \cdot X_{new} {=} \mathsf{recv}\ (\llbracket d \rrbracket^m_d\\
\qquad (X_{state}))\ \mathsf{end}, y_{rec}(X_{new}).)(\llbracket s \rrbracket^m_s))) \lhd \varepsilon]\ ,\ i\ )
\end{array}$$

$$\llbracket \varepsilon \rrbracket^m_d \stackrel{def}{=} \lambda X_{state} \cdot \{\ \mathsf{Coord}\ , \_\} \rightarrow \mathsf{Coord}\ !\ \mathsf{nok}, (X_{state});$$

$$\llbracket r_1; d_1 \rrbracket^m_d \stackrel{def}{=} \lambda X_{state} \cdot \llbracket r_1 \rrbracket^m_r(X_{state}); \llbracket d_1 \rrbracket^m_d(X_{state})$$

$$\begin{array}{l}
\llbracket ((q, c) \mapsto a) \rrbracket^m_r \stackrel{def}{=} \lambda X_{state} \cdot \{\ \mathsf{Coord}, \llbracket q \rrbracket^m_q\} \mathsf{when}\\
\qquad (\llbracket c \rrbracket^m_c(X_{state})) \mapsto (\ \mathsf{Coord}\ !\ \mathsf{ok}, \llbracket a \rrbracket^m_a(X_{state}))
\end{array}$$

$$\llbracket \{x_0, x_1, \ldots, x_k\} \rrbracket^m_s \stackrel{def}{=} \llbracket x_0 \rrbracket^m_i : \llbracket x_1 \rrbracket^m_i : \ldots : \llbracket x_k \rrbracket^m_i$$

$$\llbracket \emptyset \rrbracket^m_s \stackrel{def}{=} \varepsilon$$

$$\llbracket n(v_0, \ldots, v_k) \rrbracket^m_e \stackrel{def}{=} \{'n', \{\llbracket v_0 \rrbracket^m_i : \llbracket v_1 \rrbracket^m_i : \ldots : \llbracket v_k \rrbracket^m_i\}\}$$

$$\llbracket n(i_0, \ldots, i_k) \rrbracket^m_q \stackrel{def}{=} \{'n', \{\llbracket i_0 \rrbracket^m_i : \llbracket i_1 \rrbracket^m_i : \ldots : \llbracket i_k \rrbracket^m_i\}\}$$

$$\llbracket true \rrbracket^m_c \stackrel{def}{=} \lambda X_{state} \cdot \mathsf{true}$$

$$\llbracket !\,(C) \rrbracket^m_c \stackrel{def}{=} \lambda X_{state} \cdot \mathsf{not} \llbracket C \rrbracket^m_c$$

$$\llbracket C_1 \&\& C_2 \rrbracket^m_c \stackrel{def}{=} \lambda X_{state} \cdot \llbracket C_1 \rrbracket^m_c \mathsf{and} \llbracket C_2 \rrbracket^m_c$$

$$\llbracket p(v_0, \ldots, v_k) \rrbracket^m_c \stackrel{def}{=} \lambda X_{state} \cdot \lambda v_0, \ldots, v_k \cdot P(\{v_0, \ldots, v_k\}, X_{state})$$

$$\llbracket \mathit{stop} \rrbracket^m_a \stackrel{def}{=} \lambda X_{state} \cdot \mathsf{exit.}$$

$$\llbracket \mathit{fail} \rrbracket^m_a \stackrel{def}{=} \lambda X_{state} \cdot \mathsf{Coord\ !\ error.}$$

$$\llbracket \mathit{noOp} \rrbracket^m_a \stackrel{def}{=} \lambda X_{state} \cdot X_{state}$$

$$\llbracket \mathit{update(S,F)} \rrbracket^m_a \stackrel{def}{=} \lambda F \cdot \lambda S \cdot F(S)$$

$$\begin{array}{l}
\llbracket \mathit{load(M)} \rrbracket^m_a \stackrel{def}{=} \lambda X_{state} \cdot (\mathsf{Coord}\ !\ \{new,\\
\qquad \mathsf{spw}(\ \mathsf{fst}(\ \llbracket M \rrbracket^m_{par}\ ))\ \}), X_{state}
\end{array}$$

$$\llbracket a_0, a_1 \rrbracket^m_a \stackrel{def}{=} \lambda X_{state} \cdot \llbracket a_1 \rrbracket^m_a(\llbracket a_0 \rrbracket^m_a(X_{state}))$$

which return a boolean value after performing a check on the monitor state passed as its argument. The action denotation $[\![-]\!]^m_a$, translates μLarvaScript actions into Erlang functions which take the monitor's current state and return an updated state accordingly.

Example 6.1. This example outlines how a monitor containing only the first rule used in Example 3.1, can be formally translated into Erlang code by applying the denotational functions provided.

$$[\![\langle\{usr1, funds\}, ((ReqFunds(Usr, Sum), \textit{!IsUsrValid(Usr)})\mapsto \textit{WarnUsr}()); \rangle]\!]^m$$

$$\stackrel{def}{=}\{ \textbf{\textit{By applying the root denotation }} [\![-]\!]^m\}$$

$$[\![t]\!]^m_{es}(\textit{snd(} [\![\langle\{usr1, funds\}, ((ReqFunds(Usr, Sum), \textit{!IsUsrValid(Usr)})\mapsto \textit{WarnUsr}()); \rangle]\!]^m_{par} \textit{))} \| \textit{fst(} [\![\langle\{usr1, funds\}, ((ReqFunds(Usr, Sum), \textit{!IsUsrValid(Usr)})\mapsto \textit{WarnUsr}()); \rangle]\!]^m_{par} \textit{)}$$

$$\stackrel{def}{=}\{ \textbf{\textit{Applying }} [\![-]\!]^m_{par} \textbf{\textit{, and extracting pidList "[i]" with the snd meta function and the actor expression with fst.}} \}$$

$$[\![t]\!]^m_{es}([i])\|i[(\mu y_{rec}\cdot\lambda X_{state}\cdot X_{new}{=}\textit{recv (} [\![((ReqFunds(Usr, Sum), \textit{!IsUsrValid(Usr)})\mapsto \textit{WarnUsr}())]\!]^m_d (X_{state})\textit{) end}, y_{rec}(X_{new}).)([\![\{usr1, funds\}]\!]^m_s)))\lhd\varepsilon]$$

$$\bullet\;\bullet\;\bullet$$

$$\stackrel{def}{=}\{ \textbf{\textit{After applying the necessary denotations}} \}$$

$$[\![t]\!]^m_{es}([i])\|i[(\mu y_{rec}\cdot\lambda X_{state}\cdot X_{new}{=}\textit{recv (}\lambda X_{state}\cdot \{\textit{Coord}, \{\textit{'ReqFunds',Usr,Sum}\}\}\textit{ when } (\textit{!IsUsrValid(Usr)}) (X_{state})\mapsto(Coord!\,ok, (WarnUsr()(X_{state}))); \{ \textit{ Coord} ,\_\}\to \textit{ Coord ! nok}, (X_{state})\textit{) end})\lhd\varepsilon]$$

$$\stackrel{def}{=}\{ \textbf{\textit{Applying}}[\![t]\!]^m_{es}\textbf{\textit{to create the coordinator}}\}$$

$$\textit{coord[(}\mu y_{rec}\cdot\lambda X_{lst}\cdot(\textit{recv }\{evt, E\} \to bcast(\{E, self()\}, X_{lst}), \textit{case await}(len(X_{lst})-1)\textit{ of }0\to y_{rec}(X_{lst}); 1\to y_{rec}(X_{lst}); \_\to \textit{error end;} \{new, Pid\}\to y_{rec}(X_{lst}{:}Pid)\textit{end.)([i])}\lhd [\![t]\!]^m_{mb}\textit{]}$$

$$\| \quad i[(\mu y_{rec}\cdot\lambda X_{state}\cdot X_{new}{=}\textit{recv (}\lambda X_{state}\cdot \{\textit{Coord}, \{\textit{'ReqFunds',Usr,Sum}\}\}\textit{ when } (\textit{!IsUsrValid(Usr)}) (X_{state})\mapsto(Coord!\,ok, (WarnUsr()(X_{state}))); \{ \textit{ Coord} ,\_\}\to \textit{ Coord ! nok}, (X_{state})\textit{) end})\lhd\varepsilon]$$

IV. THE DISTPOLYLARVA PROTOTYPE

DISTPOLYLARVA is prototype implementation based on our new actor-based design. This prototype seeks to re-implement POLYLARVA's *monitor compiler* in a way which conforms to the denotational translations provided in our distributed-state model. This ensures that any guarantees offered by the formal models would also apply for our prototype compiler.

Also, DISTPOLYLARVA parses a variant of polyLS, called Pseudo-polyLS, into a parse tree which, resembles the μLarvaScript abstract syntax, together with additional parsed constructs. Although our prototype compiler is able to recognize all polyLS keywords and synthesise additional monitoring features, which are not formalized in our models, it only guarantees correct behaviour for specifications which only use constructs from the formalized subset which forms μLarvaScript. The parsed constructs are then converted into Erlang actor expressions in a similar way as in our formal translation. Furthermore, this prototype was developed with the aim to demonstrate that our translation is implementable.

A. The Compilation Phases

DISTPOLYLARVA passes a given Pseudo-polyLS specification from four subsequent stages so as to synthesise the required monitoring Erlang code.

Lexical and Parsing Phases: The Lexical phase uses a *regular grammar* which defines a number of patterns that a character sequence, in the given Pseudo-polyLS script, must match in order to be translated into an abstract token. The generated token sequence is passed to the Parsing phase which checks that the structure of the script being compiled, is correct with respect to the production rules defined by the *context free grammar* of our language defined in **Table 3.1**. If the entire token sequence obeys the rules of the grammar, it is converted into an unambiguous *parse tree* which conforms to the abstract syntax of μLarvaScript. DISTPOLYLARVA's lexer was implemented using a lexer generator called LEEX while its parser was implemented using a parser generator called YECC [12].

Semantic Analysis and Code Generation Phase: This phase is essentially an Erlang implementation of our formal denotations in Figure 4.1. It starts by invoking the initial denotational function which inspects the initial node of the parse tree and invokes other denotational functions which inspect the semantics of its child nodes, from left to right. The compiler also checks that any event, condition and action referred by the rules of a specific monitor, is actually declared within the same monitor, so as to preserve scoping. The generated Erlang source modules (.erl) are then written in a directory specified by the user and are compiled into executable Beam files via the Erlang compiler.

V. EVALUATION

The high level and distributed-state models were evaluated by proving certain theorems about the runtime behaviour they describe. The guarantees obtained from proving these theorems are also inherited by DISTPOLYLARVA, as this was developed with a close relation to the formal denotational translation. Moreover, the prototype was further evaluated through a series of tests.

A. Evaluating the High-level Model

In order to evaluate the behaviour described by this model we proved a theorem which guarantees that any monitoring system, specified in μLarvaScript, will operate deterministically. This property is important since it ensures that whenever any collection of μLarvaScript monitors is in a particular collective state[5], and it receives a specific system event, it will *always*

[5] By "collective state" we refer to the local states of all monitors in the specified monitor collection.

handle the event in the *same* manner, thus transitioning to the same successive collective state. This means that no matter how many times the monitoring system is executed, depending on the current state, it will always handle a specific event in the *same* way, and so transition to *same* consecutive state. Hence, this guarantees that a monitoring system will operate consistently.

Theorem 6.1. μLarvaScript Determinism.
$t \triangleright M \longmapsto t' \triangleright M' \;\wedge\; t \triangleright M \longmapsto t'' \triangleright M'' \;\; implies \;\; t' = t'' \;\wedge\; M' \equiv M''$

Specifically, Theorem 6.1 states [4] that if M reduces to both $t \triangleright M'$ and $t \triangleright M''$, by a using *high-level* reduction ($\longmapsto$), then it implies that $t \triangleright M'$ and $t \triangleright M''$ are *equal* to each other. The proof of this theorem was divided into separate lemmas, each of which were proved accordingly by using various inductive techniques.

B. Evaluating the Formal Translation

The evaluation of our denotational semantics consisted in proving that our formal translation is in some sense correct. We showed that the behaviour of *any* actor-based monitoring system, derived using our denotational conversion, *corresponds* to the behaviour described by the high-level model. These proofs not only help to increase the user's confidence but also state that any property proved on our high-level model, such as determinism in Theorem 6.1, would also transitively apply to our synthesised monitoring system. In our proofs we assume that all μLarvaScript specifications observe the Single Receiver Property. This implies that the denotational translation is only guaranteed to provide a correctly-behaving actor implementation when the specification script being translated observes the Single Receiver Property.

The evaluation consisted in proving the following two theorems:

Lemma 6.1. Single-Step Correspondence.
$t \triangleright M \longmapsto t' \triangleright M' \;\; implies \;\; [\![t \triangleright M]\!]^m \rightarrow^* [\![t' \triangleright M']\!]^m$

Lemma 6.2. Multi-Step Correspondence.
$t \triangleright M \longmapsto^* t' \triangleright M' \;\; implies \;\; [\![t \triangleright M]\!]^m \rightarrow^* [\![t' \triangleright M']\!]^m$

Lemma 6.1 guarantees that for *one* high-level reduction, i.e., $t \triangleright M \longmapsto t' \triangleright M'$, there exists a corresponding translation, $[\![t \triangleright M]\!]^m$, which reduces in 0 or more Erlang reduction steps into $[\![t' \triangleright M']\!]^m$. The proof for Lemma 6.2 relies on Lemma 6.1 so as to guarantee that for *0 or more* high level reductions, we can find a denotational translation which reduces $[\![t \triangleright M]\!]^m$ in 0 or more Erlang steps into $[\![t' \triangleright M']\!]^m$.

VI. Future Work

As part of our future work we propose to extend our μLarvaScript calculus so as to formalize other polyLS constructs such as timers. This extension requires modifications to our formal models, as well as, additional formal results. The new results would guarantee that the extended high-level model still operates deterministically and that its behaviour still corresponds to the behaviour of an extended version of our distributed-state model. The additional features in our DistPolyLarva compiler could then be properly implemented in a way which guarantees correct operation.

Moreover, as we were more concerned with the mathematical aspect of our designs and since our prototype implementation was only intended to demonstrate our actor-based concept, the DistPolyLarva compiler was rapidly developed. Hence we propose to provide a more thorough implementation based on our prototype and on our formal models. In fact we propose that the code of the prototype should be properly structured so as to be more maintainable in the future. Moreover, the synthesised monitoring code can be further optimized so as to reduce the tool's monitoring overhead as much as possible. Additionally, the finalized compiler should also provide better error reporting and error recovery mechanisms which would further aid users to debug their Pseudo-polyLS specification scripts. We also suggest that the proper implementation should also be tested for efficiency and compared with the original PolyLarva implementation.

VII. Conclusion

We have sought to increase the understandability and reliability of PolyLarva with the aim of elevating the user's level of confidence in our RV tool. This was done by providing a high-level operational model which describes the runtime behaviour of the core constructs of polyLS. The evaluation for this model consisted in proving that the model describes a deterministic monitoring behaviour. We also created denotational semantics which convert μLarvaScript specifications into Erlang actor expressions. The evaluation of this model consisted in proving the correctness of the formal translation, which permits that any property proved for the high-level model would also apply for the denoted monitoring Actors. This also helps in increasing the user's level of confidence in our tool. This formal translation was then implemented as the DistPolyLarva prototype compiler which guarantees a correct translation for Pseudo-polyLS specifications which only include constructs that are formalized in μLarvaScript.

References

[1] M. K. A and K. P. Distributed computing approaches for scalability and high performance, 2010.

[2] J. Armstrong. *Programming Erlang: Software for a Concurrent World*. Pragmatic Bookshelf, 2007.

[3] A. Bauer, M. Leucker, and C. Schallhart. Runtime verification for ltl and tltl. Technical report, 2006.

[4] I. Cassar. Monitoring distributed systems with distributed polylarva. Technical report, University of Malta, 2013. Final Year Project.

[5] C. Colombo. Practical runtime monitoring with impact guarantees of java programs with real-time constraints. Master's thesis, University of Malta, 2008.

[6] C. Colombo, A. Francalanza, R. Mizzi, and G. J. Pace. polylarva: Runtime verification with configurable resource-aware monitoring boundaries. In *Software Engineering and Formal Methods - 10th International Conference, SEFM 2012*, volume 7504 of *Lecture Notes in Computer Science*, pages 218–232. Springer, 2012.

[7] A. Francalanza and A. Seychell. Synthesising correct concurrent runtime monitors in erlang. Technical Report CS2013-01, University of Malta, 2013. Available from www.um.edu.mt/ict/cs/research/technical_reports.

[8] A. Gul A., T. Prasannaa, and Z. Reza. Actors: A model for reasoning about open distributed systems. Technical report, University of Illinois at Urbana USA, 2001.

[9] J. L. Gustafson. Reevaluating amdahl's law. *Communications of the ACM*, 31:532–533, 1988.

[10] P. Haller and F. Sommers. *Actors in Scala*. Artima Incorporation, USA, 2012.

[11] R. Mizzi. An extensible and configurable runtime verification framework. Master's thesis, University of Malta, 2012.

[12] Ericsson AB. Parse tools reference manual, Feb. 2013.

[13] R. Vermeersch. Concurrency in erlang and scala. Jan 2009.