Using SMT engine to generate Symbolic Automata*

Xudong Qin[2,3] Eric Madelaine[1,2] Min Zhang[3]

[1] Univ. of Nice Sophia Antipolis, CNRS, UMR 7271, 06900 Sophia Antipolis, France
[2] INRIA Sophia Antipolis Méditérannée, BP 93, 06902 Sophia Antipolis, France
[3] Shanghai Key Laboratory of Trustworthy Computing, ECNU, China

Abstract. **[TODO:not good for a FM, (to be rewriten by Eric)]** Open pNets are used to model the behavior of open synchronous or asynchronous systems expressed in various calculi or languages. They are endowed with a symbolic operational semantics in terms of so-called "Open Automata". We implement an algorithm computing this semantics, and building predicates expressing the synchronization conditions allowing some combination of events in the pNet system. Checking such predicates requires symbolic reasoning over first order logics, and specific functions of each action algebra. To reduce the complexity of the generated open automata, we use the Z3 SMT engine to check satisfiability of the predicates, and minimize the number of significant symbolic transitions.

1 Introduction

In the nineties, several works extended the basic behavioral models based on labelled transition systems to address value-passing or parameterized systems, using various symbolic encodings of the transitions [?,?,?,?]. In [?], H.M. Lin addressed value-passing calculi, for which he developed a symbolic behavioral semantics, and proved algebraic properties. Separately J. Rathke [?] defined another symbolic semantics for a parameterized broadcast calculus, together with strong and weak bisimulation equivalences, and developed a symbolic model-checker based on a tableau method for these processes. Thirty years later, no practical verification approach and no verification platform are using this kind of approaches to provide proof methods for value-passing processes or open process expressions.

Parameterized Networks of Synchronized Automata (pNets) were proposed to give a behavioral specification formalism for distributed systems, synchronous, asynchronous, or heterogeneous. It is used in VerCors, a platform for designing and verifying distributed systems, as the intermediate language for various high-level languages. The high-level languages in VerCors formalize each component of the distributed system and gives out the composition of these components.

* This work was partially funded by the Associated Team FM4CPS between INRIA and ECNU, Shanghai

pNets provides the core low-level semantic formalism for VerCors, and is made of a hierarchical composition of (value-passing) automata, called parameterized labelled transition systems (pLTS), where each hierarchical level defines the possible synchronization of the lower levels. Traditionally, pNets have been used to formalize fully defined systems or softwares. But we want also to define and reason about incompletely defined systems, like program skeletons, operators, or open expressions of process calculi. The open pNet is proposed to solve the problem of the undefined components contained in the systems using "holes" as the process parameter dealing with these components. The hole acts as the placeholder for the uncertain component with a set of possible behaviors the component might conduct, presenting the "open" property. The pNet model was developed in a series of papers [?,?] in which many examples have been introduced showing its ability to encode the operators from some other algebras or program skeletons. The operational semantic of an (open) pNet is defined as an Open Automaton in which Open Transitions contain logical predicates expressing the relations between the behavior of the holes, and the global behavior of the system. In the previous publication, only a sketch of a procedure allowing to compute this semantics was presented, together with a proof of finiteness of the open automaton, under reasonable hypotheses on the pNet structure.

Implementing this semantics raised several challenges, in order:

- to get a tool that could be applied to pNets representing various languages, in particular various actions algebras, with their specific decision theories,
- to separate clearly on one hand the algorithmic part, as an algorithm generating the transitions of the open automaton from combination of all possible (symbolic) behaviors; on the other hand the symbolic reasonning part, specificaly here using an SMT engine to check the satisfiability of the predicates generated by our algorithm,
- to build a prototype and validate the approach on our basic case-studies, and understand the efficiency of the interaction with the SMT solver.

In the long goal, we want to be able to check the equivalence between open systems encoded as pNets. The equivalence between pNets is "FH-bisimulation" taking the predicate of the open transitions into account each time matching such open transitions. We foresee that the interplay with the SMT solver that we use here for satisfiability of open transitions will be similar with what we need to prove (symbolic) equivalence between open transitions.

Contribution In the article we show how:

- We define the open automaton generation algorithm, and we implemented a full working prototype, within the VerCors platform. In the process, we improved the semantics rules from [?], and add features in the algorithm to deal the full model, including management of variables and assignments.
- We implement the interaction between our algorithm and the Z3 SMT solver, for checking satisfiability of the transitions generated by the algorithm.

- We show the interest of this approach on an industrial-inspired use-case, namely one architectural pattern extracted from the BIP specification of a nano satellite on-bord software.

Related works There is not much research work towards trying to develop symbolic bisimulation approaches for the value-passing process algebra and languages, as our long-term goals, especially no algorithm treatment of the symbolic systems developed by interacting with automatic theorem provers. The closest work is the one already mentioned from J. Rathke [**?**], who developed the symbolic bisimulation for a calculus of broadcasting system (CBS). CBS is similar with classic process calculi such as CCS and CSP, but communicating by broadcasting values, one-to-many communication instead of one-to-one communication and transmitting values without blocking. That makes the definition of the symbolic semantic and bisimulation equivalence different from the classic works.

For other applications, e.g. for the analyses of programming languages, there exist dedicated platforms making use of external automatic theorem provers (ATP), or even some automatic tactics from interactive theorem provers (ITP), to perform symbolic reasoning, and for example to discharge some subgoals in the proofs. Tools like Rodin [**?**,**?**,**?**] have already integrated several provers, like Z3, as modules for proving the proof obligations generated from the model inputed by user. The prover we used, which also happens to be Z3, is developed by Microsoft Research based on the satisfiability modulo theories (SMT), mainly applied in extended static checking, test case generation, and predicate abstraction. In a similar way, there are several ATPs/ITPs we could consider to use for the result pruning and bisimulation checking in our algorithm, as an alternative to Z3, such as CVC4 [**?**], Coq [**?**], Isabelle [**?**] or others.

Structure. In section 2 we give a description and a formal definition of the pNet model, as found in previous publications. Then in **??** we present a simple example that will illustrate the rest of the paper. Section 5 recalls the operational semantics of pNet, including its structural rules. Section 6 explains in details our implementation within the Vercors platform, and shows the full result of the semantic computation on the running example. Finally we conclude and discuss perspectives in section 7.

2 Background: pNets definition

[TODO:To be reduced, moving more material to the appendix]

This section introduces pNets and the notations we will use in this paper. Then it gives the formal definition of pNet structures, together with an operational semantics for open pNets.

pNets are tree-like structures, where the leaves are either *parameterized labelled transition systems (pLTSs)*, expressing the behavior of basic processes, or *holes*, used as placeholders for unknown processes, of which we only specify their set of possible actions, named *sort*. Nodes of the tree (pNet nodes) are synchronizing artifacts, using a set of *synchronization vectors* that express the possible synchronization between the parameterized actions of a subset of the sub-trees.

Notations. We extensively use indexed structures over some countable indexed sets, which are equivalent to mappings over the countable set. $a_i^{i \in I}$ denotes a family of elements a_i indexed over the set I. $a_i^{i \in I}$ defines both I the set over which the family is indexed (called *range*), and a_i the elements of the family. E.g., $a^{i \in \{3\}}$ is the mapping with a single entry a at index 3 ; abbreviated $(3 \mapsto a)$ in the following. When this is not ambiguous, we shall use notations for sets, and typically write "indexed set over I" when formally we should speak of multisets, and write $x \in a_i^{i \in I}$ to mean $\exists i \in I. x = a_i$. An empty family is denoted $\emptyset$. We denote classically $\overline{a}$ a family when the indexing set is not meaningful. $\uplus$ is the disjoint union on indexed sets.

Term algebra. Our models rely on a notion of parameterized actions, that are symbolic expressions using data types and variables. As our model aims at encoding the low-level behavior of possibly very different programming languages, we do not want to impose one specific algebra for denoting actions, nor any specific communication mechanism. So we leave unspecified the constructors of the algebra that will allow building expressions and actions. Moreover, we use a generic *action interaction* mechanism, based on unification between two or more action expressions. This will be used in the semantics of synchronization vectors to express various kinds of communication or synchronization mechanisms.

Formally, we assume the existence of a term algebra $\mathcal{T}_{\Sigma,\mathcal{P}}$, where Σ is the signature of the data and action constructors, and $\mathcal{P}$ a set of variables. Within $\mathcal{T}_{\Sigma,\mathcal{P}}$, we distinguish a set of data expressions $\mathcal{E}_{\mathcal{P}}$, including a set of boolean expressions $\mathcal{B}_{\mathcal{P}}$ ($\mathcal{B}_{\mathcal{P}} \subseteq \mathcal{E}_{\mathcal{P}}$). On top of $\mathcal{E}_{\mathcal{P}}$ we build the action algebra $\mathcal{A}_{\mathcal{P}}$, with $\mathcal{A}_P \subseteq \mathcal{T}_{\mathcal{P}}, \mathcal{E}_P \cap \mathcal{A}_P = \emptyset$; naturally action terms will use data expressions as sub-terms. The function $vars(t)$ identifies the set of variables in a term $t \in \mathcal{T}$.

pNets can encode naturally the notion of input actions as found e.g. in value-passing CCS [**?**] or of usual point-to-point message passing calculi, but it also allows for more general mechanisms, like gate negotiation in Lotos, or broadcast communications.

2.1 The (open) pNets Core Model

A pLTS is a labelled transition system with variables; variables can be manipulated, defined, or accessed inside states, actions, guards, and assignments.

Each state has its set of variables called *State variables*, which can only be modified by the assignment in the transitions targeting its owner state. A global state variable of a pLTS is a variable x satisfying $\forall s \in S. x \in vars(s)$.

Note that we make no assumption on finiteness of the set of states nor on finite branching of the transition relation.

We first define the set of actions a pLTS can use, let a range over action labels, op are operators, and x_i range over variable names. Action terms are:

$$\begin{array}{rll} \alpha \in \mathcal{A} ::= & a(p_1, \ldots, p_n) & \text{action terms} \\ p_i ::= & \mathit{Expr} & \text{parameters} \\ \mathit{Expr} ::= & \mathit{Value} \mid x \mid op(\mathit{Expr}_1, .., \mathit{Expr}_n) & \text{Expressions} \end{array}$$

Definition 1 (pLTS). *A pLTS is a tuple* $pLTS \triangleq \langle\!\langle S, s_0, \to \rangle\!\rangle$ *where:*

- *S is a set of states.*
- *$s_0 \in S$ is the initial state.*
- *$\to \subseteq S \times L \times S$ is the transition relation and L is the set of labels of the form $\langle \alpha,\ e_b,\ (x_j := e_j)^{j\in J} \rangle$, where $\alpha \in \mathcal{A}$ is a parameterized action, $e_b \in \mathcal{B}$ is a guard, and the variables $x_j \in P$ are assigned the expressions $e_j \in \mathcal{E}$. If $s \xrightarrow{\langle \alpha,\ e_b,\ (x_j := e_j)^{j\in J} \rangle} s' \in \to$ then $vars(\alpha) \subseteq vars(s)$, $vars(e_b) \subseteq vars(s')$, and $\forall j \in J.\ vars(e_j) \subseteq vars(s) \land x_j \in vars(s')$.*
- *The x_j here are state variables of state s'.*

Now we define pNet nodes, as constructors for hierarchical behavioral structures. A pNet node has a set of sub-pNets that can be either pNets or pLTSs, and a set of Holes, playing the role of process parameters.

A composite pNet consists of a set of sub-pNets exposing a set of actions, each of them triggering internal actions in each of the sub-pNets. The synchronization between global actions and internal actions is given by *synchronization vectors*: a synchronization vector synchronizes one or several internal actions, and exposes a single resulting global action.

Definition 2 (pNets). *A pNet is a hierarchical structure which leaves are pLTSs and holes:*
$pNet \triangleq pLTS \mid \langle\!\langle pNet_i^{i\in I}, S_j^{j\in J}, SV_k^{k\in K} \rangle\!\rangle$ *where*

- *$I \in \mathcal{I}$ is the set over which sub-pNets are indexed.*
- *$pNet_i^{i\in I}$ is the family of sub-pNets.*
- *$J \in \mathcal{I}_\mathcal{P}$ is the set over which holes are indexed. I and J are* disjoint*: $I \cap J = \emptyset$, $I \cup J \neq \emptyset$*
- *$S_j \subseteq \mathcal{A}$ is a set of action terms, denoting the* Sort[4] *of hole j.*
- *$SV_k^{k\in K}$ is a set of synchronization vectors ($K \in \mathcal{I}_\mathcal{P}$). $\forall k \in K, SV_k = \alpha_l^{l \in I_k \uplus J_k} \to \alpha'_k$ where $\alpha'_k \in \mathcal{A}_\mathcal{P}$, $I_k \subseteq I$, $J_k \subseteq J$, $\forall i \in I_k.\ \alpha_i \in \mathrm{Sort}(pNet_i)$, $\forall j \in J_k.\ \alpha_j \in S_j$, and $vars(\alpha'_k) \subseteq \bigcup_{l \in I_k \uplus J_k} vars(\alpha_l)$. The global action of a vector SV_k is* $\mathrm{Label}(SV_k) = \alpha'_k$.

[4] The formal definition of *Sorts* (set of actions of a Hole or pNet), *Leaves* and *Holes* (all pLTSs (resp holes) in a pNet hierarchical system. These can be found in [?].

3 BIP architectures, and their encodings into pNets

4 Running example

Several publications [**?**,**?**] already have introduced many examples of pNets, encoding operators of various classical process algebras, or more complex synchronization structures in distributed or parallel languages.

5 Operational Semantics for Open pNets

The semantics of open pNets will be defined as an open automaton. An open automaton is an automaton where each transition composes transitions of several LTSs with the actions of some holes; the transition occurs if some predicates hold, and can involve a set of state modifications.

Definition 3 (Open transitions). *An* open transition *over a set* $(S_i, s_{0i}, \rightarrow_i)^{i\in I}$ *of LTSs, a set* J *of holes with sorts* $Sort_j^{j\in J}$*, and a set of states* $\mathcal{S}$ *is a structure of the form:*

$$\frac{\{s_i \xrightarrow{a_i}_i s'_i\}^{i\in I}, \{\xrightarrow{b_j}_j\}^{j\in J}, Pred, Post}{s \xrightarrow{v} s'}$$

Where $s, s' \in \mathcal{S}$ *and for all* $i \in I$, $s_i \xrightarrow{a_i}_i s'_i$ *is a transition of the LTS* $(S_i, s_{0i}, \rightarrow_i)$*, and* $\xrightarrow{b_j}_j$ *is a transition of the hole* j*, for any action* b_j *in the sort* $Sort_j$*. Pred is a predicate over the variables of the terms, labels, and states* s_i*,* b_j*,* s*,* v*. Post is a set of equations that hold* after the open transition*, represented as a substitution* $\{x_k \leftarrow e_k\}^{k\in K}$ *where* x_k *are variables of* s'*,* s'_i*, and* e_k *are expressions over the other variables of the open transition.*

Definition 4 (Open automaton). *An* open automaton *is a structure*
$A =< LTS_i^{i\in I}, J, \mathcal{S}, s_0, \mathcal{T} >$ *where:*

- I *and* J *are sets of indices,*
- $LTS_i^{i\in I}$ *is a family of LTSs,*
- $\mathcal{S}$ *is a set of states and* s_0 *an initial state among* $\mathcal{S}$*,*
- $\mathcal{T}$ *is a set of open transitions and for each* $t \in \mathcal{T}$ *there exist* I'*,* J' *with* $I' \subseteq I$*,* $J' \subseteq J$*, such that* t *is an open transition over* $LTS_i^{i\in I'}$*,* J'*, and* $\mathcal{S}$*.*

When building an Open Automaton as the semantics of a pNet, its *states*, and the shape of the *predicates* in its transitions have a specific structure:

States of open pNets: A state of an open pNet is a tuple of the states of its leaves (in which we denote tuples in structured states as $\triangleleft \ldots \triangleright$). For any pNet p, let $\overline{Leaves} = \langle\!\langle S_i, s_{i0}, \rightarrow_i \rangle\!\rangle^{i\in L}$ be the set of pLTS at its leaves, then $States(p) = \{\triangleleft s_i^{i\in L} \triangleright | \forall i \in L. s_i \in S_i\}$. A pLTS being its own single leave: $States(\langle\!\langle S, s_0, \rightarrow \rangle\!\rangle) = \{\triangleleft s \triangleright | s \in S\}$. The initial state is defined as: $InitState(p) = \triangleleft {s_{i0}}^{i\in L} \triangleright$.

Predicates: Let $\langle\!\langle \overline{pNet}, \overline{S}, SV_k^{k\in K} \rangle\!\rangle$ be a pNet. Consider a synchronization vector SV_k, for $k \in K$. We build a predicate *MkPred* relating the actions of the involved sub-pNets and the resulting actions. This predicate verifies:

$$MkPred(SV_k, a_i^{i\in I}, b_j^{j\in J}, v) \Leftrightarrow \begin{array}{c} \exists (a'_i)^{i\in I}, (b'_j)^{j\in J}, v'.\, SV_k = (a'_i)^{i\in I}, (b'_j)^{j\in J} \to v' \\ \wedge \forall i \in I.\, a_i = a'_i \wedge \forall j \in J.\, b_j = b'_j \wedge v = v' \end{array}$$

Structural Semantic Rules: Now we build the semantics of an open pNet as an open automaton where LTSs are the pLTSs at the pNet leaves, and the states are structured as in the previous section. To build an open transition one first projects the global state into states of the leaves, then applies pLTS transitions on these states, and compose them with actions of holes using synchronisation vectors.

The semantics regularly instantiates *fresh* variables, and uses a *clone* operator that clones a term replacing each variable with a fresh one. The variables in each synchronization vector are considered local: for a given pNet expression, we must have fresh local variables for each occurrence of a vector (= each time we instantiate rule Tr2). Similarly the state variables of each copy of a given pLTS in the system, must be distinct, and those created for each application of Tr2 have to be fresh and all distinct.

Definition 5 (Operational semantics of open pNets). *The semantics of a pNet p is an open automaton $A =< Leaves(p), J, \mathcal{S}, s_0, \mathcal{T} >$ where:*

- *J is the indices of the holes: $Holes(p) = H_j^{j\in J}$.*
- *$\overline{\mathcal{S}} = States(p)$ and $s_0 = InitState(p)$*
- *$\mathcal{T}$ is the smallest set of open transitions satisfying the rules below:*

*The rule (**Tr1**) for a pLTS p checks that the guard is verified and transforms assignments into post-conditions:*

$$\textbf{Tr1:}\quad \dfrac{s \xrightarrow{\langle \alpha,\ e_b,\ (x_j{:=}e_j)^{j\in J}\rangle} s' \in\to \qquad \texttt{fresh}(v) \qquad Pred =\ e_b \wedge (v = \alpha)}{p = \langle\!\langle S, s_0, \to \rangle\!\rangle \models \dfrac{\{s \xrightarrow{\alpha}_p s'\}, \emptyset, Pred, \{x_j \leftarrow e_j\}^{j\in J}}{\triangleleft s \triangleright \xrightarrow{v} \triangleleft s' \triangleright}}$$

*The second rule (**Tr2**) deals with pNet nodes: for each possible synchronization vector applicable to the rule subject, the premisses include one* open transition *for each sub-pNet involved, one possible* action *for each Hole involved, and the predicate relating these with the resulting action of the vector. A key to understand this rule is that the open transitions are expressed in terms of the leaves and holes of the pNet structure, i.e. a flatten view of the pNet: e.g. L is the index set of the Leaves, L_k the index set of the leaves of one subnet, so all L_k are disjoint subsets of L.*

Tr2:

$$\frac{\begin{array}{c} k\in K \qquad SV=clone(SV_k)=\alpha_m^{m\in I_k\uplus J_k}\rightarrow\alpha'_k, G_k \qquad Leaves(p)=pLTS_l^{l\in L} \\ \forall m\in I_k. pNet_m \models \frac{\{s_i \xrightarrow{a_i}_i s'_i\}^{i\in I'_m}, \{\xrightarrow{b_j}_j\}^{j\in J'_m}, Pred_m, Post_m}{\triangleleft s_i^{i\in L_m}\triangleright \xrightarrow{v_m} \triangleleft {s'_i}^{i\in L_m}\triangleright} \qquad I'=\biguplus_{m\in I_k} I'_m \\ J'=\biguplus_{m\in I_k} J'_m \uplus J_k \qquad Pred=\bigwedge_{m\in I_k} Pred_m \wedge MkPred(SV, v_m^{m\in I_k}, b_j^{j\in J_k}, v) \\ \forall j\in J_k.\texttt{fresh}(b_j) \qquad \texttt{fresh}(v) \qquad \forall i\in L\backslash I'. s'_i = s_i \end{array}}{p = \langle\!\langle pNet_i^{i\in I}, S_j^{j\in J}, SV_k^{k\in K}\rangle\!\rangle \models \frac{\{s_i \xrightarrow{a_i}_i s'_i\}^{i\in I'}, \{\xrightarrow{b_j}_j\}^{j\in J'}, Pred, \uplus_{m\in I_k} Post_m}{\triangleleft s_i^{i\in L}\triangleright \xrightarrow{v} \triangleleft {s'_i}^{i\in L}\triangleright}}$$

Example 1. Using the operational rules to compute open-transitions: in Fig. 3 we show the deduction tree used to construct and prove the open transition ot_2.

The proof tree uses TR1 twice, for the δ transition of C_1 and for the $acc(x)$ transition of C_2, then uses an action hb_{12} of hole P, and combines the results using the first vector of the PN2 sub-pNet, and the second vector of the top node according to TR2. This yields a final $\underline{\delta(x)}$ transition. The deduction tree in the figure shows how the predicates are generated in this process.

6 Implementation

The VerCors platform uses pNets as the intermediate language for some high-level language or graphical formalism to be translated into both input for a model-checker and for generating executable code automatically [?]. We have extended the pNet API in VerCors to deal with open pNets, and also to specify the structure of action algebras.

In this section we describe the algorithm implementing the pNet semantics, the interaction with the Z3 SMT solver, and we show the result on our example.

Algorithm 1 shows the main structure of the algorithm: the computation repeats applying Tr1 and Tr2 to build all possible combinations of behaviors of pLTS and Holes, through all possible synchronisation vectors. The resulting open transitions are checked for satisfiability by the SMT solver.

Our implementation does not build explicitly the proof trees for each open transition, but instead : [Combining] enumerate all possible combinations of pLTS transition (using TR1) and each synchronization vector choice (TR2); [Matching] build the corresponding transition, and in particular its predicate, in a hierarchical manner. Each time we apply Tr2, we build a predicate as a conjunction of two parts. The first part, which is the simpler one, composes the predicate from the subnets and the guard from transitions in sub-pLTS or synchronization vectors. The second part is the conjunction of equations generated during matching the subnets' behaviors with the synchronization vectors.

Combining: Algorithm 2 shows the combining algorithm for enumerating all possible working status combinations from subnets. Dealing with a list TR containing sets of open transitions from different subnets, the algorithm choses one

Algorithm 1 Open Automaton Generation

```
Initialize sets $U$, $E$ for unexplored/explored global states $U$, $L$ for result OTs;
while !isEmpty(U) do
    Chose a global state $S$ in $U$; recursively applying Tr1 or Tr2 on the root node;
    if Subnet is a pLTS then
        Get the local state of the pLTS $s$ from $S$;
        for each transition $tr$ starting from $s$ do Apply Tr1
            Build the corresponding $OT$; store $OT$ in $L$;
        end for
    end if
    if Subnet is a pNet then
        for each $sv \in SV$ do Apply Tr2
            Invoke $Combining()$ to generate the combination of sub-transitions, $Matching()$ to match the combination with the $sv$; Store the result OTs in $L$;
        end for
    end if
    for each target global state $T$ from $OT \in L$ do
        if (!$U.contains(T)$) && (!$E.contains(T)$) then Add $T$ into the $U$;
    end for
end while
Prune the open transitions in $L$ using the SMT solver;
return $L$;
```

Algorithm 2 Combining

```
Initialize an empty combination list $L_C$;
Extract a list of open transitions $L_{ot}$ from $TR$;
Add $null$ to $L_{ot}$;
Get the combinations $L'_C$ of the $TR$;
for each $ot \in L_{ot}$ do
    for each $C' \in L'_C$ do
        Add $ot$ into $C'$ to get a new tuple $C$;
        Add $C$ into $L_C$;
    end for
end for
return $L_C$;
```

of sets L_{ot} and does combining on the remaining part of TR recursively. The case when a subnet is not working is managed by adding a *null* into the L_{ot}.

Algorithm 3 Matching

```
Initial an empty result list $L_{ot}$;
for each $C \in L_C$ do
    for each $sv \in SV$ do
        Clone the $sv$, the result is $sv'$;
        Generate the fresh result action $ra$;
        Combine $C, B, ra$ as a new vector $v$;
        for each $(e_1 \in v)\&\&(e_2 \in sv')$ do
            if $(e_1$ is null $\&\&$ $e_2$ is not null$)$ || $(e_1$ is not null $\&\&$ $e_2$ is null$)$ then
                Skip;
            else Add $(e_1 = e_2)$ into the $Pred$;
            end if
        end for
        Generate the result open transition $ot$ with $Pred$;
        Add the $ot$ into $L_{ot}$;
    end for
end for
return $L_{ot}$;
```

Matching: Algorithm 3 shows the algorithm matches combination L_C, hole behaviors B with synchronization vectors SV according to the definition of the predicate declared before. We clone the sv using the fresh function to generate a new sv' with fresh variables.The mismatch on working status is easy to detect so we filter them in the algorithm through checking the matching elements e_1 and e_2. Then we build the resulting open transition. Satisfiability of the predicate is not checked at this level, but only later, at the toplevel of the pNet.

6.1 Fresh variables

Application of the semantic rules require generating a lot of fresh names, for different kind of variables. We could use a global name generator to guarantee unicity, but at the same time, we also hope them still readable. Here we rename the variables with a regular format using the fresh function. More precisely, the fresh function generates a new name adding a suffix after the original name. The suffix contains three parts combined by a colon.

Definition 6 (Fresh variable). *The format of* fresh variable *(renaming) is defined as:* "prefix : tree index : counter*".*

- *$prefix$ identifies the kind of the variable. current kinds are: sva (SV action), ra (result action), hb (hole behavior).*

- *tree index is the index of the node containing the variable in the tree-like structure of the pNet.*
- *counter is the current value of the corresponding counter.*

For example, in the running example, the first possible behavior of the hole, it doesn't have a name, only have the prefix "hb". The hole is the second subnets of the root node so its "tree index" is "12". The fresh variable name is "`:hb:12:1`".

6.2 Management of state variable assignments

In a pLTS, there may be several incoming transitions of some states that assign potentially different values to a state variable. To handle such cases, the algorithm manages the variables together with a list for each pLTS. The list we used contains several triples $<v, S, AssignRH>$ where v is the variable in pLTS, S is its owner state and $AssignRH$ is a list of expressions over other variables in the right hand side of assignments of v. As a pragmatic extension to the formal definition, we also manage "global variables", defined in all states of the pLTS.

6.3 Pruning the unsatisfiable results

We use a brute-force method to generate all the possible open transitions in the open automaton, using all possible combinations of synchronization vectors. Naturally this builds some transitions where the predicates express incompatible constraints. Even if having an unsatisfiable (symbolic) transition in the open automaton would not be incorrect, we want to check them for satisfiability, and reduce the number of transitions and states in the automaton. In Fig. 4, we display an example of an unsatisfiable open transition from the result of our example. It shows the case where the controller C1 wants to move the control from P to $(acc(x);Q)$, conducting a δ transition. However, a synchronization vector is chosen that does not match with this action, and we can easily find the contradiction in the generated predicate term "$\texttt{:ra:11:1} = l \wedge \delta = \texttt{:ra:11:1}$".

Checking satisfiability requires some symbolic computation on the action expressions and the predicates, and this may depend on the specific theory of the action algebra. We use the SMT solver Z3 to check the predicate of the result open transitions, it will return whether the predicate is satisfiable or not. The "Modulo Theory" part of SMT solvers is important here, so that the solver can use specific properties of each action algebra.

Presentation of algebra **[TODO:Improve this, inserting the main definition from the "Checker" document]** As seen in section 2, the *term algebra* used to express the action expressions and data parameters is left open in the pNet definition.

So, for a given language, we now need to specify its data and action domains, giving them an abstract syntax (sorts, constructors, operators, and predicates). We call this a *Presentation*. We also define a concrete syntax, that will be used for pretty-printing, but also for translating the presentation, and the predicates, into Z3 syntax (see the “declare-datatypes” statement in Fig 5).

An algebra presentation contains:

- Sorts: Constants sets of the algebra or types of the data and actions (integer and boolean sort are implicitly declared). There *must* be one sort for actions, and eventually others for the types of data parameters. In the LOTOS example, we have the *Action* sort: `Action =` $\{l, \delta, r, p, q, acc\}$.
- Operators: constructors and predicates for data and action sorts. For instance, in our LOTOS example, we have a generic action constructor “*ACT*”, taking as argument an action litteral and a data parameter, declared as: `ACT:< Action, Data >` $\rightarrow$ `Action.`,
- Constants: The satisfiability solver need to know which actions in the pLTS behaviors are constants (like the l, δ, r of the controllers in running example). For a given pNet system, the sort *Action* will include all these constants.

Interaction with Z3 From inside our algorithm code, we submit satisfiability requests to Z3 using its JAVA API. Here for readability, we show the Z3 code using its SMT-LIB input language. As an example in Fig 5, we show the input for checking the transition of Fig. 4. It contains the declaration of the LOTOS action algebra types and constructors, then the declaration of variables, and finally the predicate to be checked, encoded as a set of assertions. The result "unsat" in the output is just what we expected.

To build the input submitted to Z3 for each OT, we translate the algebra *presentation*, the predicates and the variable assignments (the Post part) into Z3 (Java-API) syntax.

Translation of Action algebra presentation **[TODO:Exerpt of the formalisation]** The *datatypes* are declared according to the *sorts* the users declared in the *presentation*. If the return type of the *functions* is one of the *datatypes*, we add this *function* to the *datatypes* declaration. So when we declare the operator ACT we illustrated in section 4.1, it will be just declared together with the constants like :

```
(ACT (action3 Action)(data Int))
```

The *functions* can also be declared independently, it usually applies to functions returning integer or boolean, in the format like:

```
(declare-fun MAX (Int Int) Int)
```

Translating assignments into predicate terms **[TODO:Exerpt of the formalisation]**

State-variable assignments are also treated as a part of the predicates when checking the satisfiability. For each assignement in a *Post* predicate, such as $\{s_0 \leftarrow 1\}$ in section 5, we translate it into an equation $s_0 = 1$. For several assignments of the same variable in the same state, we generate the disjunction of these equations. Correspondingly, we generate a conjunction for the assignments from different variables.

Checking satisfiability Here we declare first all the variables in the predicates, with their sorts. An example of syntax is::

```
(declare-const |:ra:1:1| Action)
```

Then each term of the predicate is an assertion in Z3, as an example, in the former unsatisfiable result in Fig 4, the contradiction is occurred in terms:

```
(:ra:11:1=l) ∧ (delta=:ra:11:1)
```

from which we generate the following two assertions:

```
(assert (= |:ra:11:1| l))   (assert (= delta |:ra:11:1|))
```

The predicate submitted to the Z3 also contains the assignments encodings. In the case of ot_2, it gives the assertion:

```
assert (or (= var_s0 0) (= var_s0 01)))
```

6.4 Result of the running example

[TODO:Update: we have 11 satisfiable OTs here] Our tool builds 30 open transitions, out of which 27 are detected unsatisfiable by Z3, we get only 3 results left; the resulting open automaton is shown in Fig. 7, together with its open transitions.

The resulting open transitions represent all the possible movements of the LOTOS example. What can be observed is any actions of the process P from the beginning, until P performs a $\delta(x)$ and the value of x is transmitted to the $acc(x); Q$ giving out a silent action. Such behavior is performed by the ot_1 then ot_2. Then it keeps presenting the behavior of Q, as seen in ot_3.

7 Conclusion and Discussion

In this paper we presented the algorithm for generating an open automaton representing the semantics of an open pNet, and we described the implementation of the algorithm. Our implementation includes two main parts. First, we compute all the possible open transitions. The actions in pLTS and hole behaviors are composed in the algorithm while the generation of predicate need to match the combinations of the subnets with the synchronization vectors. Some of the open transitions obtained at this intermediate step, constructed in a structural manner from all possible combinations of possible logical predicates, may contain predicates which do not represent any possible concrete instantiations. To get rid of these useless transitions, we use the SMT solver Z3 for checking the satisfiability of the predicate in each open transition. In order to do that, we encode into Z3 the action algebra presentation, the a representation of the predicates, before submitting them to the Z3 solver. We implemented our algorithm in the VerCors platform and use some running examples, encoding expressions from various process algebras, to test the algorithm. Our use-cases show that our encoding identifies successfully all unsatisfiable open transitions and that the resulting automaton reflects correctly the expected movements of the encoded process expressions.

Among the questions arising during this work, an important one is the efficiency of Z3 for our needs. Z3 is famous for being very fast for solving very large sets of assertions, and this could be important for us. But we encountered some problems during the implementation the checking method using Z3. For example, if the *functions* and the *datatypes* part of the algebra presentation we submit to the solver are independent with each other, then checking the *assertions* is a simple work. However, *functions* on recursive *datatypes* make it more complex, some special rules might be defined by user for the induction. Also, we need more evidence that dealing with more complex action algebras, that would involve axiomatizing their data structures into Z3 theories, will indeed allow us to decide validity of predicates. We may also try other automatic theorem provers, depending on the structure of the proofs we need.

[TODO:Change perspective: we are working on the formal extensino of BIB architectures, and their semantics in terms of pNets. Then bisimulation and model-checking principles adapted to open pNet systems are under development]

Naturally, our next goals after the generation of the open automata will be to model-check, and to check equivalence of pNets. While model-checking open automata (with a suitable logic including predicates on data) seems easy to define, equivalence checking is more challenging. In previous paper, we have already found the FH-bisimulation, inspired by [?] as a suitable definition. But usual approaches of bisimulation by partition refinement does not work easily with symbolic transitions. In a first step we may want instead to define and implement an algorithm for checking that a given equivalence relation between open automata states is indeed a FH-Bisimulation. Finding automatically a bisimulation relation or the most general bisimulation will be more challenging. It appears also that when comparing different pNets with different structure, strong bisimulation will be too strong a notion, and that we will have to define proper notions of weak equivalence or refinements.

Before that work, we need to reconsider the result open transitions from our algorithm. The ideal result that we showed in the bottom part of Fig. 7 is not the one generated by the algorithm, because the original one contains many intermediate variables, such as fresh local variables of synchronization vectors, and fresh result variables of intermediate open transitions. Such intermediate variables come from the way predicates are generated, and depends on the particular structure of the pNet. But they are not significant in the resulting behavior, and should be eliminated before comparing such behavior with another pNet. To eliminate the intermediate variables in the predicate, we want to define the set of the significant variables that we won't eliminate: the local variables of pLTS transitions, the hole behaviors and the result action at the top level. The result of the elimination will be simpler predicates, that can be successfully compared (e.g. by Z3) within the bisimulation checking algorithm.

A More details on the generation algorithm

A.1 Algebra specification

A.2 Term algebra: type system and static semantics

A.3 Variable managment

A.4 Filtering the behaviour combinations

A.5 Translation to SMTlib input language

A.6 Elimination of intermediate variables

B Full examples

[TODO:May be we use two architecture examples here, a simpler one, before the Failure Timer ?]

B.1 A very simple architecture

B.2 Full input for the FailureTimer architecture

B.3 Full output for the FailureTimer open automaton

B.4 Scaling up: several subsystems on the same bus

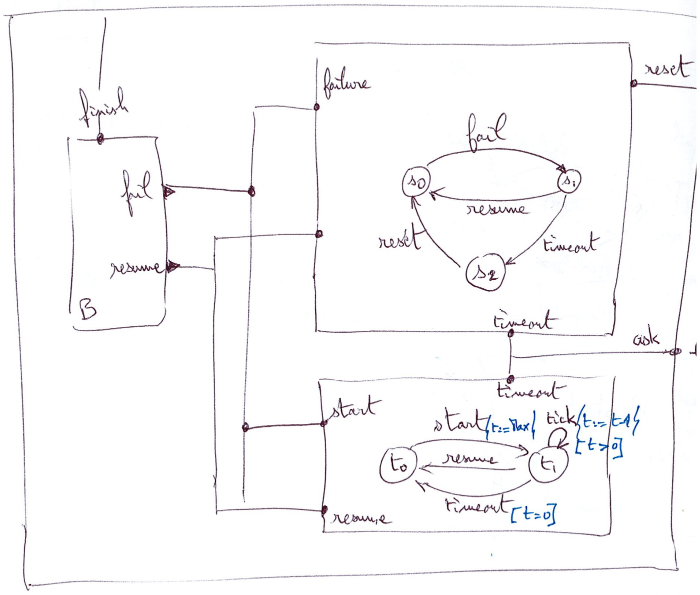
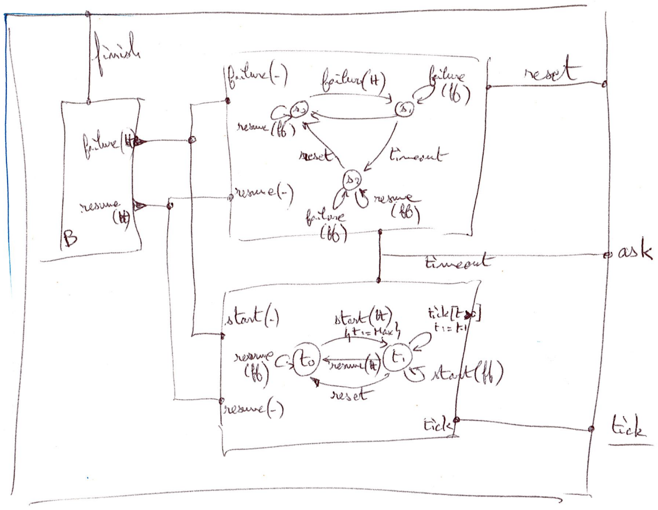


Fig. 1. FailureTimer Architecture in BIP graphical syntax



Pnet structure = < Failure, Timer, B >

SV_1 < failure(b_1), ~~failure~~ start(b_2), failure(b_0)> fail [$b_1 = b_2 \wedge$ $b_1 \vee b_2 \Rightarrow b_0$]

SV_2 < resume(b_1), resome(b_2), resume (b_0)> resume [$b_1 = b_2 \wedge$ $b_1 \vee b_2 \Rightarrow b_0$]

SV_3 < -, tick, -> tick

SV_4 <timout, timeout, -> timeout

SV_5 <reset, -, -> reset

SV_6 < - - finish> finish

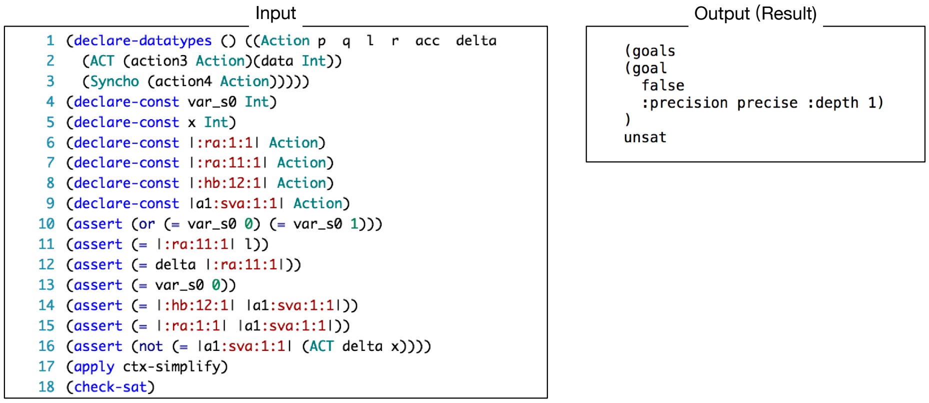
Fig. 2. FailureTimer pNet encoding

$$\dfrac{\dfrac{0 \xrightarrow{\delta}_{C_1} 0 \quad [s_0 = 0]}{C_1 \models \dfrac{0 \xrightarrow{\delta}_{C_1} 0 \quad [s_0 = 0] \quad \{s_0 \leftarrow 1\}}{\lhd 0 \rhd \xrightarrow{\delta} \lhd 0 \rhd}} \qquad \dfrac{\dfrac{0 \xrightarrow{acc(x)}_{C_2} 1}{C_2 \models \dfrac{0 \xrightarrow{acc(x)}_{C_2} 1}{\lhd 0 \rhd \xrightarrow{acc(x)} \lhd 1 \rhd}}}{\text{acc(x);Q} \models \dfrac{0 \xrightarrow{acc(x)}_{C_2} 1}{\lhd 0 \rhd \xrightarrow{acc(x)} \lhd 1 \rhd}}}{\text{PN1} \models \dfrac{0 \xrightarrow{\delta}_{C_1} 0, 0 \xrightarrow{acc(x)}_{C_2} 1 \quad \{\xrightarrow{hb_{12}}_{P}\} \quad [s_0 = 0 \wedge hb_{12} = \delta(x)] \quad \{s_0 \leftarrow 1\}}{\lhd 00 \rhd \xrightarrow{\tau} \lhd 01 \rhd}}$$

Fig. 3. Proof of ot_2 (interaction of process P and action $acc(x)$) for “P»(acc(x);Q)”

$$ot = \dfrac{\begin{array}{c}\{0 \xrightarrow{\delta} 0\}, \quad \{\xrightarrow{:hb:12:1}\}, \quad \texttt{:ra:11:1} = l \ \wedge \ \delta = \texttt{:ra:11:1} \ \wedge \ \texttt{s0} = 0 \\ \wedge \ \texttt{:hb:12:1} = \texttt{a1:sva:1:1} \ \wedge \ \texttt{:ra:1:1} = \texttt{a1:sva:1:1} \ \wedge \ \texttt{a1:sva:1:1} \neq \delta(x), \quad \{\texttt{s0} \leftarrow 1\}\end{array}}{\lhd 00 \rhd \xrightarrow{\texttt{:ra:1:1}} \lhd 00 \rhd}$$

Fig. 4. One of the unsatisfiable open transitions in LOTOS running example



Input

```
(declare-datatypes () ((Action p  q  l  r  acc  delta
  (ACT (action3 Action)(data Int))
  (Syncho (action4 Action)))))
(declare-const var_s0 Int)
(declare-const x Int)
(declare-const |:ra:1:1| Action)
(declare-const |:ra:11:1| Action)
(declare-const |:hb:12:1| Action)
(declare-const |a1:sva:1:1| Action)
(assert (or (= var_s0 0) (= var_s0 1)))
(assert (= |:ra:11:1| l))
(assert (= delta |:ra:11:1|))
(assert (= var_s0 0))
(assert (= |:hb:12:1| |a1:sva:1:1|))
(assert (= |:ra:1:1| |a1:sva:1:1|))
(assert (not (= |a1:sva:1:1| (ACT delta x))))
(apply ctx-simplify)
(check-sat)
```

Output (Result)

```
(goals
(goal
  false
  :precision precise :depth 1)
)
unsat
```

Fig. 5. The input of the Z3 solver in SMT-LIB language and the output result

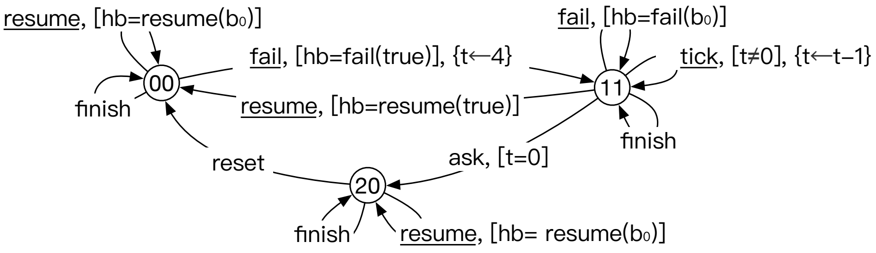

resume, [hb=resume(b_0)]
fail, [hb=fail(true)], {t←4}
fail, [hb=fail(b_0)]
tick, [t≠0], {t←t–1}
00
11
finish
resume, [hb=resume(true)]
finish
reset
ask, [t=0]
20
finish
resume, [hb= resume(b_0)]

Fig. 6. Open Automaton**[TODO:syntax could be better...]**

[TODO:Detaisl of some OTs of the FailureTimer]

$$ot_1 = \frac{\{0 \xrightarrow{l}_{C_1} 0\},\quad \{\xrightarrow{hb_1}_{P}\},\quad [s_0 = 0 \wedge hb_1 \neq \delta(x) \wedge v_1 = hb_1]}{\triangleleft 00 \triangleright \xrightarrow{v_1} \triangleleft 00 \triangleright}$$

$$ot_2 = \frac{\{0 \xrightarrow{\delta}_{C_1} 0, 0 \xrightarrow{acc(x)}_{C_2} 1\},\quad \{\xrightarrow{hb_2}_{P}\},\quad [s_0 = 0 \wedge hb_2 = \delta(x) \wedge v_2 = \underline{\delta(x)}],\quad \{s_0 \leftarrow 1\}}{\triangleleft 00 \triangleright \xrightarrow{v_2} \triangleleft 01 \triangleright}$$

Fig. 7. The open automaton of the LOTOS formula