Proving a Petri net model-checker implementation
Lukasz Fronc, Franck Pommereau

To cite this version:
Lukasz Fronc, Franck Pommereau. Proving a Petri net model-checker implementation. [Research Report] IBISC, university of Evry / Paris-Saclay. 2011. hal-02315062

HAL Id: hal-02315062
https://hal.science/hal-02315062
Submitted on 14 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Proving a Petri net model-checker implementation
IBISC technical report, March 2011

Lukasz Fronc and Franck Pommereau
IBISC, University of Évry, Tour Évry 2
523 place des terrasses de l’Agora, 91000 Évry, France
{fronc,pommereau}@ibisc.univ-evry.fr

Abstract. Petri nets are a widely used tool in verification through model-checking. In this approach, a Petri Net model of the system of interest is produced and its reachable states are computed, searching for erroneous executions. Compilation of such a Petri net model is one way to accelerate its verification. It consists in generating code to explore the reachable states of the considered Petri net, which avoids the use of a fixed exploration tool involving an “interpretation” of the Petri net structure.

In this paper, we show how to compile Petri nets targeting the LLVM language (a high-level assembly language) and formally prove the correctness of the produced code. To this aim, we define a structural operational semantics for the fragment of LLVM we use. The acceleration obtained from the presented compilation techniques has been evaluated in [6].

Keywords: explicit model-checking, model compilation, LLVM, formal semantics, correctness

1 Introduction

Verification through model-checking [2] consists in defining a formal model of the system to be analysed and then using automated tools to check whether the expected properties are met or not. We consider here more particularly the widely adopted setting in which models are expressed in coloured Petri nets [7] and there states are explored using explicit model-checking that enumerates them all (contrasting with symbolic model-checking that handles directly sets of states).

Model compilation is one of the numerous techniques to speedup explicit model-checking, it relies on generating source code (then compiled into machine code) to produce a high-performance implementation of the state space exploration primitives. For instance, this approach is successfully used in the popular coloured Petri net model-checker Helena [15, 4] that generates C code.

In this paper, we present an approach for proving the correctness of such an approach. More precisely, we focus on the produced code and prove that the object computed by its execution is an actual representation of the state space of the compiled model. We consider the Low-Level Virtual Machine (LLVM)
language as our target language for the models compilation, which reconciles two otherwise contradictory objectives: on the one hand, this is a typed language with reasonably high-level operations allowing to express algorithms quite naturally; on the other hand, it is still a low-level language and can be equipped with formal semantics allowing to formally prove the correctness of the programs. To carry on these proofs, we define a structural operational semantics of the fragment of LLVM we use and use it in various ways to establish the properties of our programs.

To the best of our knowledge, this is the first attempt to provide a formal semantics for LLVM. Moreover, if model-checkers are widely used tools, there exists surprisingly few attempts to prove them [18], contrasting with the domain of proof assistants [3, 16] where the importance of “proving the prover” is widely acknowledged. Others approaches like certifying model-checkers [8] produce deductive proofs on success or failure, and thus only that proof needs to be checked (for each result). On the side of compiler proving, proofs address mostly semantic preservation [13, 14], i.e., ensure that the compilation process respects the semantics of the input language when producing the compiled code. Others, address specific optimisation phases in a compiler [1].

The rest of the paper is organised as follows. We first recall the main notions about coloured Petri nets. Then, we present the LLVM framework, in particular the syntax of the language and its intuitive semantics, and how it can be embedded LLVM into a Petri net as a concrete colour domain. In section 4, we present algorithms and data structures for state space exploration. We formally define the operational semantics for LLVM, including a memory model with explicit heap and stack. Finally we present our correctness results. Due to limited number of pages, some definitions and intermediary results have been omitted, as well as the detailed proves. This material can be found in [5]. Notice also that our compilation approach is evaluated from a performance point of view in [6].

2 Coloured Petri nets

A (coloured) Petri net involves objects defined by a colour domain that provides data values, variables, operators, a syntax for expressions, possibly typing rules, etc. Usually, elaborated colour domains are used to ease modelling: in particular, one may consider a functional programming language [7, 17] or the functional fragment (expressions) of an imperative programming language like in Helena [15]. In this paper we will consider LLVM as a concrete colour domain. All these can be seen as implementations of a more general abstract colour domain:

- $\mathbb{D}$ is the set of data values;
- $\mathbb{V}$ is the set of variables;
- $\mathbb{E}$ is the set of expressions. Let $e \in \mathbb{E}$, we denote by $\text{vars}(e)$ the set of variables from $\mathbb{V}$ involved in $e$. Moreover, variables or values may be considered as (simple) expressions, i.e., we assume $\mathbb{D} \cup \mathbb{V} \subseteq \mathbb{E}$. 

We do not make any assumption about the typing or syntactical correctness of expressions; instead, we assume that any expression can be evaluated, possibly to \( \perp \notin \mathbb{D} \) (undefined value) in case of any error. More precisely, a binding is a partial function \( \beta : \mathbb{V} \to \mathbb{D} \cup \{ \perp \} \). Let \( e \in \mathbb{E} \) and \( \beta \) be a binding, we denote by \( \beta(e) \) the evaluation of \( e \) under \( \beta \); if the domain of \( \beta \) does not include \( \text{vars}(e) \) then \( \beta(e) \equiv \perp \). The application of a binding to evaluate an expression is naturally extended to sets and multisets of expressions.

**Definition 1 (Petri nets).** A Petri net is a tuple \((S, T, \ell)\) where:

- \( S \) is the finite set of places;
- \( T \), disjoint from \( S \), is the finite set of transitions;
- \( \ell \) is a labelling function such that:
  - for all \( s \in S \), \( \ell(s) \subseteq \mathbb{D} \) is the type of \( s \), i.e., the set of values that \( s \) is allowed to carry,
  - for all \( t \in T \), \( \ell(t) \in \mathbb{E} \) is the guard of \( t \), i.e., a condition for its execution,
  - for all \( (x, y) \in (S \times T) \cup (T \times S) \), \( \ell(x, y) \) is a multiset over \( \mathbb{E} \) and defines the arc from \( x \) toward \( y \).

A marking of a Petri net is a map that associates to each place \( s \in S \) a multiset of values from \( \ell(s) \). From a marking \( M \), a transition \( t \) can be fired using a binding \( \beta \) and yielding a new marking \( M' \), which is denoted by \( M[t, \beta]M' \), iff:

- there are enough tokens: for all \( s \in S \), \( M(s) \geq \beta(\ell(s, t)) \);
- the guard is validated: \( \beta(\ell(t)) = \text{true} \);
- place types are respected: for all \( s \in S \), \( \beta(\ell(t, s)) \) is a multiset over \( \ell(s) \);
- \( M' \) is \( M \) with tokens consumed and produced according to the arcs: for all \( s \in S \), \( M'(s) = M(s) - \beta(\ell(s, t)) + \beta(\ell(t, s)) \).

Such a binding \( \beta \) is called a mode of \( t \) at marking \( M \).

For a Petri net node \( x \in S \cup T \), we define \( x^\bullet \equiv \{ y \in S \cup T \mid \ell(y, x) \neq \emptyset \} \) and \( x^\star \equiv \{ y \in S \cup T \mid \ell(x, y) \neq \emptyset \} \) where \( \emptyset \) is the empty multiset. Finally, we extend the notation \( \text{vars} \) to a transition by taking the union of the variable sets in its guard and connected arcs.

In this paper, we assume that the considered Petri nets are such that, for all place \( s \in S \) and all transition \( t \in T \), \( \ell(s, t) \) is either \( \emptyset \) or contains a single variable \( x \in \mathbb{V} \). We also assume that \( \text{vars}(t) = \bigcup_{s \in S} \text{vars}(\ell(s, t)) \), i.e., all the variables involved in a transition can be bound using the input arcs. The second assumption is a classical one that allows to simplify the discovery of modes. The first assumption is made without loss of generality to simplify the presentation.

### 3 LLVM

The **LLVM** project (Low Level Virtual Machine) [10] is a modern and modular toolkit for compiler development. It began as a research project at the University of Illinois and has grown to an umbrella project for number of sub-projects, many
of which being used by a wide variety of commercial and open source projects [12] as well as academic researches [11].

The **LLVM-IR (LLVM Intermediate Representation)** [9] is a part of the LLVM project and is a low-level platform-independent intermediate language. Every program written in this LLVM language can be run in a virtual machine or compiled to native code on all the platforms supported by the LLVM project. Importantly, the LLVM compiler runs a set of optimisation passes on the LLVM-IR code, which allows us to produce simple source code knowing it will be optimised by LLVM, and so lets us concentrate on the higher-level optimisations of algorithms and data structures.

### 3.1 Syntax and intuitive semantics.

A LLVM program is composed of a set of blocks identified by labels. Each block contains a sequence of instructions, some are entry points for subprograms (functions or procedures) and are identified by labels including the formal parameters. Entering or leaving a block is always explicit through branching instructions (jumps), subprograms calls or return instructions. Moreover, LLVM is a **SSA (Single Static Assignment)** based representation. It means that each variable is assigned exactly once, but since a block can be executed many times the value may be dynamically modified. This property is useful in some proofs because it can serve as an invariant.

To define the syntax, we consider the following pairwise disjoint sets:

- $\mathcal{P}$ is the set of pointers;
- $\mathcal{T}$ is the set of types, defined as the smallest set containing \{int, bool, object\} (i.e., integers, Boolean values and other values) and such that if $t_0, \ldots, t_n \in \mathcal{T}$ then $\text{struct}(t_0, \ldots, t_n) \in \mathcal{T}$, it represents a data structure with $n + 1$ fields of types $t_0$ to $t_n$;
- $\mathcal{L}$ is the set of labels, it contains some specific labels like $f(a_1, \ldots, a_n)$, where $a_i \in \mathcal{V}$ for $1 \leq i \leq n$, that correspond to subprograms entry points (including the formal parameters). We define a set $\mathcal{L}_{\bot} \equiv \mathcal{L} \cup \{\bot\}$ where $\bot$ is an undefined label.

A program is represented as a partial function $P$ with $\mathcal{L}$ as its domain and that associates each label in its domain to a sequence of instructions.

For our purpose, we need to consider a fragment of LLVM that is restricted to the types presented above and is formed by three main syntax classes: sequences ($seq$), commands ($cmd$) and expressions ($expr$). A sequence is a list of commands which may end with an expression, in which case it is considered as an expression.

We assume that programs are syntaxically correct and well typed, so that we can simplify the syntax by forgetting all types in LLVM source code. The resulting syntax is presented in figure 1. We introduce the sequencing operator ";", which corresponds to the line endings, in order to write one-line sequences. We introduce the skip command that denotes the empty sequence and does not exist in LLVM. It may be noted that $p\text{call}$ (procedure call) and $f\text{call}$ (function
call) do not exist in LLVM but are different instances of the call instruction. This distinction can be easily made in LLVM because the instruction contains the type of the subprogram (function or procedure). Instruction store (resp. load) is the action of storing (resp. loading) data into (resp. from) the memory through a pointer. Instruction icmp compares two integers. Instruction phi is used to access variables defined in previously executed blocks. Instruction gep corresponds to pointer arithmetic, we freeze the second argument to 0, which is enough to access fields in structures by numbers.

\[
\text{seq ::= cmd} \quad \mid \quad \text{expr} \\
\text{cmd ::= br label} \quad \mid \quad \text{expr} \\
\text{cmd; seq} \quad \mid \quad \text{cmd; expr} \\
\text{cmd ::= br rvalue, label, label} \quad \mid \quad \text{pcall label(rvalue, ..., rvalue)} \\
\text{ret} \quad \mid \quad \text{var = expr} \\
\text{store rvalue, rvalue} \quad \mid \quad \text{skip} \\
\text{expr ::= add rvalue, rvalue} \quad \mid \quad \text{load rvalue} \\
\text{expr ::= icmp op, rvalue, rvalue} \quad \mid \quad \text{phi (rvalue, label), ..., (rvalue, label)} \\
\text{expr ::= fcall label(rvalue, ..., rvalue)} \quad \mid \quad \text{alloc type} \\
\text{expr ::= ret rvalue} \quad \mid \quad \text{rvalue}
\]

Fig. 1. Our fragment of the LLVM syntax, where label ∈ L, rvalue ∈ D∪P∪V, var ∈ V, type ∈ T, nat ∈ N and op ∈ \{<,\leq,=,\neq,\geq,>\}.

3.2 LLVM-labelled Petri Nets.

To compile Petri nets as defined previously into LLVM programs, we need to consider a variant where the colour domain explicitly refers to a LLVM program.

**Definition 2 (LLVM labelled Petri nets).** A LLVM labelled Petri net is thus a tuple \(N \triangleq (S, T, \ell, P)\), where \(P\) is a LLVM program, and such that \((S, T, \ell)\) is a coloured Petri net with the following changes:

- for all place \(s \in S\), \(\ell(s)\) is a LLVM type in \(T\), which can be directly interpreted as a subset of \(D\);
- for all transition \(t \in T\), \(\ell(t)\) is a call to a Boolean function in \(P\) with the elements of \(\text{vars}(t)\) as parameters;
moreover, for all $s \in \mathcal{I}^*$, $\ell(t, s)$ is a singleton multiset whose unique element is a call to a $\ell(s)$-typed function in $P$ with the elements of $\text{vars}(t)$ as parameters.

With respect to the previous definition, we have concretized the place types and each expression is now implemented as a LLVM function called from the corresponding annotation. To simplify the presentation, we have also restricted the output arcs to be singleton multisets, but this can be easily generalised. Moreover, the definitions of binding and modes are extended to LLVM. A LLVM binding is a partial function $\beta : \mathcal{V} \mapsto \mathbb{D} \cup \mathbb{P}$, which maps each variable from its domain to a pointer or a value, and is widened to $\mathbb{D}$ by the identity function. $\mathbb{B}$ is the set of all LLVM bindings. Thus, A LLVM mode is a LLVM binding enabling a transition in a LLVM labelled Petri net.

4 State space exploration

Given an initial marking $M_0$, the state space we want to compute in this paper is the smallest set $R$ such that $M_0 \in R$ and, if $M \in R$ and $M[t, \beta]M'$ then $M' \in R$ also.

Interfaces. To implement algorithms, we need to define interfaces to the data structures we will use: multisets, places, markings, and (marking) sets. An interface is presented as a set of functions that manipulates a data structure through a pointer (C-like interfaces). Moreover, each of interface functions have a formal specification of their comportment, an example is given in section 4, and structure marking have to hold two properties defined in section 6. Below, for a given structure multiset of type $\text{Multiset}(d)$, what we call its domain is the set of values from $d$ having non-zero occurrences in the multiset. We explicit only the functions used in this document. The multiset interface contains the following elements:

- two functions $\text{addmset}(\text{p_mset, elt})$ and $\text{remmset}(\text{p_mset, elt})$ that respectively adds an element $\text{elt}$ into the multiset $\text{p_mset}$ and removes an element from the multiset;
- a function $\text{size_mset}(\text{p_mset})$ that returns the domain size of a multiset;
- a function $\text{nth_mset}(\text{p_mset, n})$ that returns the $n^{th}$ element from the multiset domain (for an arbitrary fixed order);

As a container of tokens, a place can be basically implemented as a multiset of tokens; so the place interface is exactly the multiset interface. The sole difference will be the annotation, a place interface function will be annotated by the place name, for instance $\text{add}_s$ is the function $\text{add_mset}$ specific to the place $s$. Alternative implementations are considered for optimisation purpose but omitted in this document.

The markings interface contains:
– a function \( \text{get}_{\ell}(p_{\text{mrk}}) \) for each place \( s \) that returns a pointer to the corresponding place structure;
– a function \( \text{copy}_{\text{mrk}}(p_{\text{mrk}}) \) that returns a copy of the structure marking.

Finally, the \( \text{set} \) interface contains:

– a function \( \text{cons}_{\text{set}}(\cdot) \) that builds a new empty set;
– a procedure \( \text{add}_{\text{set}}(p_{\text{set}}, \text{elt}) \) that adds an element \( \text{elt} \) to the set \( p_{\text{set}} \).

### Transitions firing

Let \( t \in T \) be a transition such that \( s_1, \ldots, s_n \) are its input places (i.e., \( \ell(t) = \{s_1, \ldots, s_n\} \)) and \( s_1, \ldots, s_m' \) are its output places (i.e., \( \ell(t') = \{s_1', \ldots, s_m'\} \)). Then, function \( \text{fire}_t \) can be written as shown in figure 2. This function simply creates a copy \( M' \) of the marking \( M \), removes from it the consumed tokens and adds the produced ones before to return \( M' \). One could remark that it avoids a loop over the Petri net places but instead it executes a sequence of statements. There are at least two advantages in this approach: it is more efficient (no branching penalties, no loop overhead, no array for the functions \( f_{t,s_1}, \ldots \)) and the resulting code is simple to produce. Let also \( x_{mrq} \) be a pointer to a structure marking implementing \( M \), and let \( x_1, \ldots, x_n \) be representations of the tokens as in figure 2. Then the firing algorithm can be implemented as shown in figure 4. The correctness and termination of this implementation can be formally proved and a sketch of the proof is given in section 6, in particular it is proved that the returned pointer actually represents \( M' \).

\[
\begin{align*}
\text{fire}_t : M : Mrk, & x_1 : \ell(s_1), \ldots, x_n : \ell(s_n) \rightarrow \\
M' & \leftarrow \text{copy}(M) \\
M'(s_1) & \leftarrow M(s_1) - \{x_1\} \\
\ldots & \\
M'(s_n) & \leftarrow M(s_n) - \{x_n\} \\
M'(s'_1) & \leftarrow M(s'_1) + f_{t,s'_1}(x_1, \ldots, x_n) \\
\ldots & \\
M'(s'_m) & \leftarrow M(s'_m) + f_{t,s'_m}(x_1, \ldots, x_n) \\
\text{return } M' & 
\end{align*}
\]

**Fig. 2.** Transition firing algorithm, where \( f_{t,s} \) is the function that evaluates \( \ell(t,s) \) and \( Mrk \) is the \( \text{Marking} \) type.

\[
P(\text{fire}_t(x_{mrq}, x_1, \ldots, x_n)) \equiv \\
\left\{
\begin{array}{ll}
// \text{copy the structure marking} \\
M_{\text{mrq}} & = \text{call copy}_{\text{mrk}}(x_{mrq}) \\
// \text{consume tokens} \\
x_{s_i} & = \text{call get}_{\text{mrk}}(x_{mrq}) \\
\text{pcall rem}_{s_i}(x_{s_i}, x_i) & \quad \text{for } 1 \leq i \leq n \\
// \text{produce tokens} \\
x_{s'_j} & = \text{call get}_{s'_j}(x'_{mrq}) \\
o_{s'_j} & = \text{call f}_{t,s'_j}(x_1, \ldots, x_n) \\
\text{pcall add}_{s'_j}(x_{s'_j}, o_{s'_j}) & \quad \text{for } 1 \leq j \leq m \\
// \text{return the new marking} \\
\text{ret } x'_{mrq}
\end{array}
\right.
\]

**Fig. 3.** LLVM translation of the firing algorithm, where \( x_i \) is the variable in \( \ell(s_i, t) \) for all \( 1 \leq i \leq n \), and \( f_{t,s'_j} \) is the function called in \( \ell(t, s'_j) \) for all \( 1 \leq j \leq m \).
Successors computation. To discover all the possible modes for a transition, the algorithm enumerates all the combinations of tokens from the input places. If a combination corresponds to a mode then the suitable transition firing function is called to produce a new marking. This algorithm is shown in figure 4. Note the nesting of loops that avoids an iteration on \( t \), which saves from querying the Petri net structure and avoids the explicit construction of a binding. Moreover, if \( g_t \) is written in the target language, we avoid an interpretation of the corresponding expression. For the LLVM version, let \( x_{mrq} \) be a pointer to a structure marking and \( x_{next} \) be a pointer to a marking set structure. Then, the algorithm from figure 4 can be implemented as shown in figure 5. Each loop at the level \( k \) is implemented as a set of blocks subscribed by \( t, k \) (for \( n \geq k \geq 1 \)), the blocks subscribed by \( t, 0 \) corresponding to the code inside the innermost loop. Note the phi instruction to update the value of index \( i_{s_i} \) allowing to enumerate the tokens in place \( s_i \): when the program enters block \( loop_i \) for the first time, it comes from block header, so we initialise the value of \( i_{s_i} \) to the maximal index value in the domain of \( s_i \) in the current marking; later, the program comes back to block \( loop_i \) from block footer, so it assigns \( i'_{s_i} \) to \( i_{s_i} \) which is exactly the value of \( i_{s_i} \) minus one (the previous index). The correctness and termination of this implementation can be formally proved and a sketch of the proof is given in section 6.

\[
\text{successor}_t : M : \text{Marking}, \text{next} : \text{MarkingSet} \rightarrow \\
\text{for } x_n \text{ in } M(s_n) \text{ do} \\
\text{...} \\
\text{for } x_1 \text{ in } M(s_1) \text{ do} \\
\text{if } g_t(x_1, \ldots, x_n) \text{ then} \\
\text{next } \leftarrow \text{next } \cup \{ \text{fire}_t(M, x_1, \ldots, x_n) \} \\
\text{endif} \\
\text{endfor} \\
\text{...} \\
\text{endfor}
\]

Fig. 4. Transition specific successors computation algorithm, where \( g_t \) is the function that evaluates the guard \( \ell(t) \).

The global successor function \( succ \) returns the set of all the successors of a marking by calling all the transition specific successor functions and accumulating the discovered markings into the same set. The algorithm is shown in figure 6 and its translation in LLVM is shown in figure 7.
\[
P(succ_t(x_{\text{mrq}}, x_{\text{next}})) \overset{\text{df}}{=} \begin{cases} 
\text{br header}_{t,n} \\
\{ x_{sk} = \text{fcall get}_{sk}(x_{\text{mrq}}) \\
\quad \text{br loop}_{t,k} \\
\} 
\end{cases}
\]

\[
P(header_{t,k}) \overset{\text{df}}{=} \begin{cases} 
\{ s_{sk} = \text{fcall size}_{sk}(x_{sk}) \\
\quad \text{br loop}_{t,k} \\
\} 
\end{cases}
\]

\[
P(loop_{t,k}) \overset{\text{df}}{=} \begin{cases} 
i_{sk} = \phi(i_{sk}, header_{t,k}, (i_{sk}', footer_{t,k}) \\
\quad c_{sk} = \text{icmp} >, i_{sk}, 0 \\
\quad \text{br header}_{t,k-1} \\
\} 
\end{cases}
\]

\[
P(body_{t,k}) \overset{\text{df}}{=} \begin{cases} 
x_{sk} = \text{fcall nth}_{sk}(x_{sk}, i_{sk}) \\
\quad \text{br header}_{t,k-1} \\
\} 
\end{cases}
\]

\[
P(footer_{t,k}) \overset{\text{df}}{=} \begin{cases} 
\text{br loop}_{t,k} \\
i_{sk} = \text{add} i_{sk}, -1 \\
\} 
\end{cases}
\]

\[
P(header_{t,0}) \overset{\text{df}}{=} \begin{cases} 
c_{sk} = \text{fcall g}_{sk}(x_{sk}, x_{sk}, \ldots, x_{sk}) \\
\quad \text{br c}_{sk}, body_{t,k}, \text{footer}_{t,k} \\
\} 
\end{cases}
\]

\[
P(body_{t,0}) \overset{\text{df}}{=} \begin{cases} 
x_{\text{mrq}} = \text{fcall fire}_{sk}(x_{\text{mrq}}, x_{sk}, x_{sk}, \ldots, x_{sk}) \\
\quad \text{pcall add}_{sk}(x_{\text{next}}, x_{\text{mrq}}) \\
\quad \text{br footer}_{t,k+1} \\
\} 
\end{cases}
\]

\[
P(footer_{t,n+1}) \overset{\text{df}}{=} \begin{cases} 
\text{ret} x_{\text{next}} \\
\} 
\end{cases}
\]

Fig. 5. LLVM successor function, for \( 1 \leq k \leq n \).

\[
P(succ]\{x_{\text{mrq}}\} \overset{\text{df}}{=} \begin{cases} 
x_{\text{next}} = \text{fcall cons}_{sk}() \\
p\text{call succ}_{t_1}(x_{\text{mrq}}, x_{\text{next}}) \\
p\text{call succ}_{t_2}(x_{\text{mrq}}, x_{\text{next}}) \\
\ldots \\
p\text{call succ}_{t_n}(x_{\text{mrq}}, x_{\text{next}}) \\
\text{ret} x_{\text{next}} \\
\} 
\end{cases}
\]

Fig. 7. LLVM implementation of the global successor function, where \( x_{\text{mrq}} \) is a pointer to a structure marking.

5 Formal semantics of LLVM

5.1 Memory model

Heap We use pointers, therefore we need a heap to access pointed data.

Definition 3 (Heaps). A heap is a partial function \( H : P \to T \times (\mathbb{D} \cup \mathbb{P})^* \) with a finite domain. Each heap maps a pointer to a pair formed of a type and a tuple of values or pointers. The set of all heaps is \( \mathbb{H} \).

Definition 4 (Well formed heaps). A heap \( H \) is well formed for a pair \( (t,p) \) where \( t \) is a type and \( p \) a pointer in \( \text{dom}(H) \), if the value pointed by \( p \) in \( H \) is sound with \( t \), more precisely:

- if \( H(p) = (\text{int},d) \), then \( d \) is an integer;
– if \( H(p) = (\text{bool}, d) \), then \( d \) is a Boolean;
– if \( H(p) = (\text{struct}(t_0, \ldots, t_n), d) \), then \( d \) is a tuple \((p_0, \ldots, p_n)\) and \( H \) is well formed for each \((t_i, p_i)\) such that \( 0 \leq i \leq n \);

A heap \( H \) is well formed if \( H \) is well formed for each \((t, p)\) in \( \text{Im}(H) \).

\[ \text{Definition 5 (Accessibility).} \quad \text{The set of all accessible pointers from a pointer } p \text{ in a heap } H \text{ is denoted as } p \downarrow H. \text{ This operation is defined for each } p \in \mathbb{P} \text{ as follows:} \]

\[ p \downarrow H \overset{df}{=} \begin{cases} \{ \} & \text{if } p \notin \text{dom}(H) \\ \{ p \} & \text{if } H(p) = (t, v) \text{ and } t \in \{ \text{int, bool} \} \\ p \downarrow H \cup p_0 \downarrow H \cup \cdots \cup p_n \downarrow H & \text{if } H(p) = (\text{struct}(t_0, \ldots, t_n), (p_0, \ldots, p_n)) \text{ and } t_0, \ldots, t_n \in T \end{cases} \]

One could show that if a heap \( H \) is well formed then \( p \downarrow H \subseteq \text{dom}(H) \) for \( p \in \text{dom}(H) \), and even that \( \text{dom}(H) = \bigcup_{p \in \mathbb{P}} p \downarrow H \).

**Property 1.** Let \( H \) be a well formed heap. If \( p \in \text{dom}(H) \) then \( p \downarrow H \subseteq \text{dom}(H) \).

**Corollary 1.** For each well formed heap \( H \), we have \( \text{dom}(H) = \bigcup_{p \in \mathbb{P}} p \downarrow H \).

\[ \text{Definition 6 (Data structure traversal).} \quad \text{For each heap } H, \text{ we define a data structure traversal function } [\cdot]_H : \mathbb{P} \times \mathbb{N} \to \mathbb{P} \cup \mathbb{D} \text{ as:} \]

\[ p[i]_H \overset{df}{=} \begin{cases} \text{undefined} & \text{if } p \notin \text{dom}(H) \\ \text{undefined} & \text{if } H(p) = (t, v) \text{ and } t \in \{ \text{int, bool} \} \\ p_i & \text{if } H(p) = (\text{struct}(t_0, \ldots, t_n), (p_0, \ldots, p_n)) \text{ and } 0 \leq i \leq n \end{cases} \]

\[ \text{Definition 7 (Heaps overwriting).} \quad \text{The overwriting function } \oplus : \mathbb{H} \times \mathbb{H} \to \mathbb{H}, \text{ which represents the writing into memory, is defined for each } p \in \mathbb{P} \text{ as:} \]

\[ (H \oplus H')(p) \overset{df}{=} \begin{cases} H'(p) & \text{if } p \in \text{dom}(H') \\ H(p) & \text{if } p \notin \text{dom}(H') \land p \in \text{dom}(H) \\ \text{undefined otherwise} \end{cases} \]

**Property 2.** The heap overwriting operation \( \oplus \) is associative.
Definition 8 (Structural equivalence). Let $H$ and $H'$ be two well formed heaps, $v$ and $v'$ two values or pointers. We have $(H, v) =_{st} (H', v')$ if one of the following conditions is true:

- $v \notin \text{dom}(H)$, $v' \notin \text{dom}(H')$ and $v = v'$;
- $H(v) = ((t_0, \ldots, t_n), (p_0, \ldots, p_n))$, $H'(v') = ((t_0', \ldots, t_n'), (p'_0, \ldots, p'_n))$ and $(H, p_i) =_{st} (H', p'_i)$ where $0 \leq i \leq n$.

Property 3. $=_{st}$ is an equivalence relation.

We need also to define an operation to build new heaps. The first step is to define a helper function $\text{alloc}$ then the function $\text{new}$ that we seek.

Definition 9 (Alloc and new). The $\text{new} : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{H} \times \mathbb{P}$ and $\text{alloc} : 2^\mathbb{P} \times \mathbb{P} \times \mathbb{T} \rightarrow \mathbb{H}$ functions are defined as follows:

$$\text{new}(H, t) \overset{\text{df}}{=} (\text{alloc}(\text{dom}(H) \cup \{p\}, p, t), p)$$

for $p \notin \text{dom}(H)$ a “fresh” pointer

$$\text{alloc}(d, p, t) \overset{\text{df}}{=} \{ p \mapsto (t, \bot) \}$$

for $t \in \{\text{int}, \text{bool}\}$

$$\text{alloc}(d, p, \text{struct}(t_0, \ldots, t_n)) \overset{\text{df}}{=} \{ p \mapsto (\text{struct}(t_0, \ldots, t_n), (p_0, \ldots, p_n)) \}$$

$\oplus \text{alloc}(d \cup \{p_0, \ldots, p_n\}, p_0, t_0)$

$\oplus \ldots$

$\oplus \text{alloc}(d \cup \{p_0, \ldots, p_n\}, p_n, t_n)$

for $p_0, \ldots, p_n \notin d$ “fresh” pointers

One could show that $\text{new}$ returns a well formed heap, and two calls of $\text{new}$, with equivalent heaps, return two equivalent heaps.

Property 4. Let $H$ be a well formed heap, and let $t$ be a type. If $(H', p) = \text{new}(H, t)$ then $H'$ is a well formed heap.

Property 5. Let $H_1$, $H_2$ be two heaps and $t$ a type. If $\text{new}(H_1, t) = (H, p)$ and $\text{new}(H_2, t) = (H', p')$ then $(H, p) =_{st} (H', p')$.

Corollary 2. Let $H_1$, $H_2$ be two well formed heaps and $t$ a type. If $\text{new}(H_1, t) = (H, p)$ and $\text{new}(H_2, t) = (H', p')$ then $H$ and $H'$ are well formed.

Stack Our memory model uses also a stack for subprogram calls. This stack is manipulated implicitly by the inference rules and only one part called frame is manipulated explicitly.

Definition 10 (Frames). A frame is a tuple in $\mathbb{F} \overset{\text{df}}{=} \mathbb{L} \times \mathbb{L} \times \mathbb{B}$, thus $\mathbb{F}$ is the set of all frames. For each frame $F \overset{\text{df}}{=} (l_p, F, l_{c,F}, \beta_F)$, we have:
– $l_{p,F}$ the label of a previous block (where we come from), or undefined;
– $l_{c,F}$ the label of the current block (where we are);
– $\beta_F$ a LLVM binding, representing the current evaluation context.

We widen the binding functional notation to the frames in order to access the binding, we note $F(x)$ the binding $\beta_F(x)$ of $x$ by $\beta_F$.

In matter of simplification, we always note for a frame $F$:
– $l_{p,F}$ its first component,
– $l_{c,F}$ its second component and
– $\beta_F$ its third component. As for heaps we need operations to handle frames.

**Definition 11 (Binding and frame binding overwriting).** The binding overwriting operation $\oplus : \mathbb{B} \rightarrow \mathbb{B}$ and the frame binding overwriting operation $\oplus : F \times \mathbb{B} \rightarrow F$ are defined as follows:

$$(\beta \oplus \beta')(p) \overset{\text{df}}{=} \begin{cases} 
\beta'(p) & \text{if } p \in \text{dom}(\beta') \\
\beta(p) & \text{if } p \notin \text{dom}(\beta') \land p \in \text{dom}(\beta) \\
\text{undefined otherwise}
\end{cases}$$

$$(l, l', \beta) \oplus \beta' \overset{\text{df}}{=} (l, l', \beta \oplus \beta')$$

The operator $\oplus$ is used as well for the heap overwriting as for the (frame) binding overwriting because of the similar nature of the performed operation.

**Property 6.** The binding overwriting operation $\oplus$ is associative.

**Definition 12 (Structural equivalence 2).** Let $H$, $H'$ be two heaps and $F$, $F'$ two frames then the structural equivalence for heap-frame pairs is defined as:

$$(H,F) =_{st} (H',F') \iff \forall x \in \text{dom}(\beta_F) \cup \text{dom}(\beta_{F'}), \ (H,F(x)) =_{st} (H',F'(x))$$

The structural equivalence has been widened to pairs of heaps and frames noted $(H,F) =_{st} (H',F')$ for $H,H'$ two heaps and $F,F'$ two frames. Intuitively, we check that all data accessible from the frame bindings are structurally equivalent. This holds also for values in bindings since the heap equivalence has been defined on $\mathbb{B}$ as the identity check.

### 5.2 Inference rules

The operational semantics is defined for a fixed and immutable program $P$, it means that any new function, nor block cannot be created or modified during the execution. We introduce a new notation for computations, we note $\overline{\text{result of a computation}}$ for example $5 =_{\text{df}} 2 + 3$. The main object used in our rewriting rules are **configurations** which represent a state of the program during the execution.
Definition 13 (Configurations). A configuration is a tuple \((\text{seq}, H, F)\), where:

- \(\text{seq}\) is a sequence of instructions;
- \(H\) is a heap;
- \(F\) is a frame.

A configuration is noted \((\text{seq})_{H,F}\). ♦

The set of inference rules for expressions is shown in figure 8; expressions are evaluated to values in a frame context. The set of inference rules for sequence and commands are shown in figure 9; sequences and commands are evaluated to other sequences in a heap-frame context.

One can remark the stacking up of frames in pcall and fcall rules, it corresponds to the frame replacement in the subsumption of these rules for evaluating the body of a subprogram.

This semantics mixes up small-step and big-step reductions. Indeed, the most of rules are small-step except for pcall and fcall rules. In these two rules, we link the computation to its result and we make a sequence of reductions in the rule subsumption.

6 Results

6.1 Interpreting data structures.

The link between Petri Nets and their implementation is formalised with a family of interpretation functions for all data structures.

Definition 14 (Interpretations). An interpretation is a partial function which maps a pair formed of a heap and a pointer to a marking, a set of markings, a multiset of tokens or a single token depending on the interpreted object. The interpretations are noted \([H,p]^*\) with \(H \in \mathbb{H}, p \in \mathbb{P} \cup \mathbb{D}\) and where * is an annotation describing the interpreted object. In case where p is a pointer, we suppose that the result depends only on the accessible data from p, i.e., \(p \downarrow_H\). Moreover every interpretation function have to hold the following requirement. ♦

Requirement. Let \(H, H'\) be two heaps, \([\cdot, \cdot]^*\) the interpretation function, and \(p, p'\) two pointers or values. If \((H, p) =_{st} (H', p')\) then \([H, p]^* = [H', p']^*\). ♦

As presented in section 4, we use data structures and functions in order to reach our goal, they are supposed given or produced by the Petri net compilation process. Each of these functions and data structures are specified or axiomatized by a formal interface. Mainly, it helps to ensure independence and modularity between components in a programmatic and formal way. Specifying an interface leads to define a set of primitives that hold certain set of derivations and interpretations. For example, let \(H\) be a heap and \(F\) a frame such that, \(F(x_{\text{mset}}) = p_{\text{mset}}\) and is a pointer on a structure multiset of elements of type
\[
F(x) = p \quad H(p) = (t, v) \\
(\text{load } x)_{H,F} \leadsto (v)_H \\
F(x) = p \quad p[i] \text{ défini} \\
(\text{gep } x, 0, i)_{H,F} \leadsto (p[i])_{H,F} \\
\]

\[
F(x_1) + F(x_2) = v \\
(\text{add } x_1, x_2)_{H,F} \leadsto (v)_H \\
add \\
(H', p) = \text{new}(H, t) \\
(\text{alloc } t)_{H,F} \leadsto (p)_{H \uplus H'} \\
alloc \\
F(x_1) \text{ op } F(x_2) = v \\
(\text{icmp } \text{op}, x_1, x_2)_{H,F} \leadsto (v)_H \\
\]

\[
1 \leq i \leq n \\
(\phi (x_1, l_1), \ldots, (x_n, l_n))_{H,(l_1, l_2, \ldots)} \leadsto (F(x_i))_H \\
\phi \\
f(a_1, \ldots, a_n) \in \text{dom}(P) \\
F_0 = (\perp, f(a_1, \ldots, a_n), \{a_1 \mapsto F(r_1), \ldots, a_n \mapsto F(r_n)\}) \\
(P(f(a_1, \ldots, a_n)))_{H,F_0} \leadsto^* (v)_{H'} \\
\text{fcall} \\
\]

\[
\text{(fcall } f(r_1, \ldots, r_n))_{H,F} \leadsto (v)_{H'} \\
\text{ret} \\
\]

Fig. 8. Rules for expressions.

\[
\text{(cmd)}_{H,F} \leadsto (\text{seq'})_{H',F'} \text{ seq} \\
\text{(cmd; seq)}_{H,F} \leadsto (\text{seq'; seq})_{H',F'} \text{ seqskip} \\
\text{(skip; seq)}_{H,F} \leadsto (\text{seq})_{H,F} \text{ seqskip} \\
\]

\[
(\text{br } l)_{H,(l_p, l_c, \beta)} \leadsto (P(l))_{H,(l_c, l, \beta)} \text{ branch} \\
\]

\[
(\beta(r) = \text{true} \land l = l_1) \lor (\beta(r) = \text{false} \land l = l_2) \\
(\text{br } r, l_1, l_2)_{H,(l_p, l_c, \beta)} \leadsto (P(l))_{H,(l_c, l, \beta)} \text{ branch} \\
\]

\[
f(a_1, \ldots, a_n) \in \text{dom}(P) \\
F_0 = (\perp, f(a_1, \ldots, a_n), \{a_1 \mapsto F(r_1), \ldots, a_n \mapsto F(r_n)\}) \\
(P(f(a_1, \ldots, a_n)\})_{H,F_0} \leadsto^* (\text{ret})_{H',F'} \\
\text{pcall} \\
\]

\[
\text{(pcall } f(r_1, \ldots, r_n))_{H,F} \leadsto (\text{skip})_{H',F} \\
\text{assign} \\
\]

\[
(x = \text{expr})_{H,F} \leadsto (\text{skip})_{H',F \oplus \{x \mapsto v\}} \text{ assign} \\
\]

\[
F(r_p) = p \quad H(p) = (t, d) \\
H' = \{p \mapsto t, F(r_{new})\} \\
\text{store} \\
\]

Fig. 9. Rules for sequences and commands.
In these conditions, the function \( \text{add}_{\text{mset}} \) that adds an element into a multiset containing objects of type \( t \) is specified by:

\[
\left( \text{pcall } \text{add}_{\text{mset}}(x_{\text{mset}}, x) \right)_{H,F} \rightsquigarrow \left( \text{skip} \right)_{H\oplus H',F}
\]

\[
\text{dom}(H) \cap \text{dom}(H') \subseteq \text{pmset}\downarrow_H
\]

\[
\left[ H \oplus H', \text{pmset}\downarrow_H \right]^{\text{mset}(t)} = \left[ H, \text{pmset}\downarrow_H \right]^{\text{mset}(t)} + \{(H, F(x)) \}
\]

The condition (1) describes the reduction, the condition (2) localises the modifications in the heap and the condition (3) explains the comportment in terms of Petri nets.

Moreover, the structures marking, i.e., every implementation, have to hold two properties:

**Requirement. (soundness)** Let \( H, H' \) be two heaps, \( F, F' \) two frames, \( p_{\text{mrq}} \) a pointer in \( \text{dom}(H) \) on a structure marking, and \( p_s \) a pointer on a structure place in the pointed marking. If \( p_s \in p_{\text{mrq}} \downarrow_H, \left[ H, p_{\text{mrq}}\right]^{\text{mrq}}(s) = \left[ H, p_s\right]^s, (\text{seq})_{H,F} \rightsquigarrow (\text{seq}')_{H\oplus H',F'} \) et \( p_{\text{mrq}} \notin \text{dom}(H') \) then

\[
\left[ H \oplus H', p_{\text{mrq}}\right]^{\text{mrq}}(s) = \left[ H \oplus H', p_s\right]^s
\]

**Requirement. (separation)** Let \( p_{\text{mrq}} \) be a pointer on a structure marking in a heap \( H \). If \( p_s \) and \( p_{s'} \) are pointers on two distinct places in this structure then we have \( p_s \downarrow_H \cap p_{s'} \downarrow_H = \emptyset \).

The soundness property ensures that any modification made on a place with a pointer returned by \( \text{get}_s \) is reflected on the marking (not on a copy). The separation property ensures that places do not share memory, i.e., when one modifies a place in the marking these modifications have no side effects on other places.

### 6.2 Insulation theorem.

One of the main results on the semantics is the insulation theorem. It allows to consider reductions in minimal contexts, then generalize them to more complex ones. The exposition of the theorem is long because we must address commands and expressions in the same time.

**Theorem 7. (insulation).** Let \( \text{seq} \) be a sequence, \( H \) a heap and \( F \) a frame. In this context, if

\[
(\text{seq})_{H,F} \rightsquigarrow^{*} (v')_{H\oplus H'}, \text{ or, resp., } (\text{seq})_{H,F} \rightsquigarrow^{*} (\text{seq}')_{H\oplus H',F'}
\]

then for all heap \( H_0 \) and all binding \( \beta_0 \) such that

(i) \( \text{dom}(H) \cap \text{dom}(H_0) = \emptyset \)
(ii) \( \text{dom}(\beta) \cap \text{dom}(\beta_0) = \emptyset \)
(iii) any assigned variable in seq (or one of its redex) is not in \( \text{dom}(\beta_0) \)
then one can make the same reduction in an environments widen by the heap \( H_0 \) and the binding \( \beta_0 \), i.e.:

\[
(\text{seq})_{H \oplus H_0, F \oplus \beta_0} \leadsto^* (v'')_{H \oplus H_0 \oplus H''}
\]

or, resp.,

\[
(\text{seq})_{H \oplus H_0, F \oplus \beta_0} \leadsto^* (\text{seq'})_{H \oplus H_0 \oplus H'', F'' \oplus \beta_0}
\]

with \( H'' \) a heap and \( F'' \) a frame such that:

\[
\text{dom}(H_0) \cap \text{dom}(H'') = \emptyset \quad (4)
\]

\[
\forall p \in \text{dom}(H), \quad (H', p) =_{st} (H'', p) \quad (5)
\]

\[
(H \oplus H', v') =_{st} (H \oplus H'', v'') \quad (6)
\]

or, respectively, (4), (5) and

\[
(F', H') =_{st} (F'', H'') \quad (7)
\]

\[
\text{dom}(\beta_0) \cap \text{dom}(\beta_{F''}) = \emptyset \quad (8)
\]

\[
\text{dom}(\beta_{F'}) = \text{dom}(\beta_{F''}) \quad (9)
\]

Intuitively, the result (4) ensures that one does not modify the heap \( H_0 \), and so the evaluation of the sequence have no side effects. The result (5) ensures that \( H' \) is sound with \( H'' \), i.e. that every structure reachable from \( H \) which was modified during the reduction have been modified in \( H'' \) in the same way. The results (6) and (7) ensures that the reduction produces the same computation. The result (8) ensure that one does not modify a variable in \( \text{dom}(\beta_0) \). The result (9) ensures that the same variables were added into the frames.

**Proof.** The proof is made by induction on derivation trees ordered by a lexicographic order with respect to the maximal height of the derivation trees during the reduction and the step count in the reduction.

- Initial case:

\[
(\text{seq})_{H,F} \leadsto^* (\text{seq'})_{H \oplus H', F'}
\]

we have:

\[
(\text{seq})_{H \oplus H_0, F \oplus \beta_0} \leadsto^* (\text{seq'})_{H \oplus H_0 \oplus H'', F'' \oplus \beta_0}
\]

where \( \text{seq} = \text{seq'} \), \( H' = H'' = \emptyset \) and \( F = F' = F'' \). Structural equivalence consequences and equalities on domains are trivial.

- Inductive cases: we consider the first step during the reduction sequence. Indeed, if one prove that the property holds for the first step then we can conclude by the induction hypothesis since the remaining reduction is smaller (the remaining reduction sequence maximal height is less or equal but a step shorter).
1. **seq case**: If the first step is an application of the *seq* rule:

\[
\begin{align*}
& (\text{cmd})_{H,F} \leadsto (\text{seq'})_{H \oplus H', F'} \quad \text{seq} \\
& (\text{cmd}; \text{seq})_{H,F} \leadsto (\text{seq'}; \text{seq})_{H \oplus H', F'}
\end{align*}
\]

We apply the induction hypothesis on the rule subsumption, we have:

\[
(\text{cmd})_{H \oplus H_0 \oplus \beta_0} \leadsto (\text{seq'})_{H \oplus H_0 \oplus H', F' \oplus \beta_0}
\]

with a heap \(H''\) and a frame \(F''\) such that:

\[
\begin{align*}
& \text{dom}(H_0) \cap \text{dom}(H'') = \emptyset \quad (10) \\
& \forall p \in \text{dom}(H), \ (H', p) =_{st} (H'', p) \quad (11) \\
& \text{dom}(\beta_0) \cap \text{dom}(\beta_{F''}) = \emptyset \quad (12) \\
& \text{dom}(\beta_{F'}) = \text{dom}(\beta_{F''}) \quad (13) \\
& (F', H') =_{st} (F'', H'') \quad (14)
\end{align*}
\]

By reinjecting into the rule, we have:

\[
(\text{cmd}; \text{seq})_{H \oplus H_0 \oplus F \oplus \beta_0} \leadsto (\text{seq'}; \text{seq})_{H \oplus H_0 \oplus H', F' \oplus \beta_0}
\]

Structural equality and domain constraints are the same.

2. **seq skip**: similar to the initial case.

3. **branch1** and **branch2** cases:

\[
(\text{br l})_{H,(l_p,l_c,l,\beta)} \leadsto (\text{P(l)})_{H \oplus H', (l_c,l,\beta)}
\]

where \(H' = \emptyset\). We have

\[
(\text{br l})_{H \oplus H_0,(l_p,l_c,l,\beta) \oplus \beta_0} \leadsto (\text{P(l)})_{H \oplus H_0 \oplus H' \oplus (l_c,l,\beta) \oplus \beta_0}
\]

where \(H'' = \emptyset\), because

\[
(l_p,l_c,\beta) \oplus \beta_0 = (l_p,l_c,\beta \oplus \beta_0)
\]

\[
(l_c,l,\beta) \oplus \beta_0 = (l_c,l,\beta \oplus \beta_0)
\]

Structural equality and domain constraints are trivial since \(H' = H'' = \emptyset\) and \(\text{dom}(\beta') = \text{dom}(\beta'') = \text{dom}(\beta)\).

The **branch2** case is similar.

4. **pcall case**: If the first reduction step is **pcall**, we know that

\[
\begin{align*}
& f(a_1, \ldots, a_n) \in \text{dom}(P) \\
& F_0 = (\bot, f(a_1, \ldots, a_n), \{a_1 \mapsto F(r_1), \ldots, a_n \mapsto F(r_n)\}) \\
& (P(f(a_1, \ldots, a_n)))_{H,F_0} \leadsto^* (\text{ret})_{H \oplus H', F_1} \\
& (\text{pcall } f(r_1, \ldots, r_n))_{H,F} \leadsto (\text{skip})_{H \oplus H', F}
\end{align*}
\]

**pcall**
where \( H' = \emptyset \). We build the conclusion by applying induction hypothesis on the subsumption (smaller derivation tree).

\[
\begin{align*}
  f\langle a_1, \ldots, a_n \rangle &\in \text{dom}(P) \\
  F_0 = (\bot, f\langle a_1, \ldots, a_n \rangle, \{ a_1 \mapsto F(r_1), \ldots, a_n \mapsto F(r_n) \}) \\
  (P(f\langle a_1, \ldots, a_n \rangle))_{H\oplus H_0, F_0, H'\oplus H''} \leadsto^* (\text{ret})_{H\oplus H_0, H'\oplus H''} \\
  (\text{pcall } f(r_1, \ldots, r_n))_{H\oplus H_0, F\oplus \beta_0} \leadsto (\text{skip})_{H\oplus H_0, H'', F\oplus \beta_0} \\
\end{align*}
\]

Structural equivalence and domain equalities are true because they were for the subsumption (induction hypothesis) and \( F' = F'' = F \oplus \beta_0 \).

The \text{fcall} case is similar, the induction hypothesis provides conclusions for the expressions.

5. \text{assign} case: If the first reduction step is \text{assign}, we know that

\[
\begin{align*}
  (\text{expr})_{H,F} \leadsto (v')_{H\oplus H'} \\
  (x = \text{expr})_{H,F} \leadsto (\text{skip})_{H\oplus H', F\oplus \{x \mapsto v'\}} \\
\end{align*}
\]

By applying the induction hypothesis (smaller derivation tree)

\[
(\text{expr})_{H\oplus H_0, F\oplus \beta_0} \leadsto (v'')_{H\oplus H_0, H''}
\]

where \( H'' \) is a heap such that:

\[
\begin{align*}
  \text{dom}(H_0) \cap \text{dom}(H'') &= \emptyset \quad (15) \\
  \forall p \in H, \ (H', p) &=_{st} (H'', p) \quad (16) \\
  (H', v') &=_{st} (H'', v'') \quad (17)
\end{align*}
\]

By reinjecting in the derivation we have:

\[
(x = \text{expr})_{H\oplus H_0, F\oplus \beta_0} \leadsto (\text{skip})_{H\oplus H_0, H'', F\oplus \beta_0 \oplus \{x \mapsto v''\}}
\]

The consequences (4) and (5) are true because of the induction hypothesis consequences (15) and (16). Since \( x \) is not an assigned variable \( x \notin \text{dom}(\beta_0) \), so \( \beta_0 \) and \( \{x \mapsto v''\} \) commutes, and so \( F \oplus \beta_0 \oplus \{x \mapsto v''\} = F \oplus \{x \mapsto v''\} \oplus \beta_0 \) and \( \text{dom}(F \oplus \{x \mapsto v''\}) \cap \text{dom}(\beta_0) = \emptyset \) which gives us the consequence (8). The consequence (9) is trivial because \( \text{dom}(\beta') = \text{dom}(\beta'') = \{x\} \). It remains to show that \( (F', H \oplus H') =_{st} (F'', H \oplus H'') \), i.e., \( (F \oplus \{x \mapsto v'\}, H \oplus H') =_{st} (F \oplus \{x \mapsto v''\}, H \oplus H'') \), showing that \( \{x \mapsto v'\}, H'') =_{st} (x \mapsto v'') \), \( H'' \) is sufficient since (16) implies the result for \( F \), but this result is a immediate consequence of (17).

6. \text{store} case:

\[
\begin{align*}
  F(r_p) = p \quad H(p) = (t, d) \quad H' = \{ p \mapsto t, F(r_{\text{new}}) \} \\
  (\text{store } r_{\text{new}}, r_p)_{H,F} \leadsto (\text{skip})_{H\oplus H', F'} \\
\end{align*}
\]
where $F = F'$, we build the conclusion:

$$
(F \oplus \beta_0)(r_p) = F(r_p) = p \quad (H \oplus H_0)(p) = H(p) = (t, d)
$$

$$
H'' = \{ p \mapsto t, F(r_{\text{new}}) \}
$$

$$
(\text{store } r_{\text{new}}, r_p)_{H \oplus H_0, F \oplus \beta_0} \leadsto \{ \text{skip} \}_{H \oplus H_0 \oplus H'', F'' \oplus \beta_0}
$$

where $F = F''$, because $\text{dom}(F) \cap \text{dom}(\beta_0) = \emptyset$ and $\text{dom}(H) \cap \text{dom}(H_0) = \emptyset$. Since $H' = H''$ and $F = F' = F''$, structural equalities and domain constraints are trivially held.

7. load case:

$$
F(x) = p \quad H(p) = (t, v)
$$

$$
(\text{load } x)_{H', F} \leadsto (v)_{H \oplus H'}
$$

where $H' = \emptyset$, immediate

$$
(F \oplus \beta_0)(x) = F(x) = p \quad (H \oplus H_0)(p) = H(p) = (t, v)
$$

$$
(\text{load } x)_{H \oplus H_0, F \oplus \beta_0} \leadsto (v)_{H \oplus H_0 \oplus H''}
$$

where $H'' = \emptyset$, because $\text{dom}(F) \cap \text{dom}(\beta_0) = \emptyset$ and $\text{dom}(H) \cap \text{dom}(H_0) = \emptyset$. Structural equalities and domain constraints are trivial.

8. $\text{gep}_0$ case:

$$
F(x) = p \quad p[i]_H \text{ défini}
$$

$$
(\text{gep } x, 0, i)_{H, F} \leadsto (p[i]_H)_{H \oplus H'}
$$

where $H' = \emptyset$. Since $p[i]_H$ is defined, we have $p$, $p[i]_H \in H$, and we conclude with

$$
(F \oplus \beta_0)(x) = F(x) = p \quad p[i]_{H \oplus H_0} = p[i]_H \text{ défini}
$$

$$
(\text{gep } x, 0, i)_{H \oplus H_0, F \oplus \beta_0} \leadsto (p[i]_{H \oplus H_0})_{H \oplus H_0 \oplus H''}
$$

where $H'' = \emptyset$, because $\text{dom}(H) \cap \text{dom}(H_0) = \emptyset$. The structural equalities and domain constraints are trivially held.

9. $\text{icmp}$ and $\text{add}$ cases:

$$
\frac{F(x_1)_{\text{op}} F(x_2) = v}{\text{icmp op, x_1, x_2}}_{H \oplus H'}
$$

where $H' = \emptyset$, we conclude

$$
\frac{(F \oplus \beta_0)(x_1)_{\text{op}} (F \oplus \beta_0)(x_2) = v}{\text{icmp op, x_1, x_2}}_{H \oplus H_0 \oplus H''}
$$

where $H'' = \emptyset$ because $(F \oplus \beta_0)(x_1) = F(x_1)$ et $(F \oplus \beta_0)(x_2) = F(x_2)$.

Structural equalities and domain constraints are trivial.

The $\text{add}$ case is similar.
10. phi case:

\[ 1 \leq i \leq n \]

\[
\left( \phi(x_1, l_1), \ldots, (x_n, l_n) \right)_{H, (l_1, l_c, \beta)} \rightsquigarrow \left( \beta(x_i) \right)_{H \oplus H'}
\]

where \( H' = \emptyset \), we have

\[ 1 \leq i \leq n \]

\[
\left( \phi(x_1, l_1), \ldots, (x_n, l_n) \right)_{H \oplus H_0, (l_1, l_c, \beta) \oplus \beta_0} \rightsquigarrow \left( (\beta \oplus \beta_0)(x_i) \right)_{H \oplus H_0 \oplus H''}
\]

where \( H'' = \emptyset \). By hypothesis \( \text{dom}(\beta) \cap \text{dom}(\beta_0) = \emptyset \), which implies that \((\beta \oplus \beta_0)(x_i) = \beta(x_i)\), thus we have the same computation result (6).

\[ H' = H'' = \emptyset \] and so (4) and (5) are trivial.

11. alloc case:

\[
\left( H', p' \right) = \text{new}(H, t) \]

\[
(\text{alloc } t)_{H,F} \rightsquigarrow \left( p' \right)_{H \oplus H'}
\]

We build the conclusion:

\[
\left( H'', p'' \right) = \text{new}(H \oplus H_0, t) \]

\[
(\text{alloc } t)_{H \oplus H_0, F} \rightsquigarrow \left( p'' \right)_{H \oplus H_0 \oplus H''}
\]

by definition of \( \text{new} \) (section 5.1, property 5.), we have \((H', p') =_{st} (H'', p'')\), moreover \( \text{dom}(H'') \cap \text{dom}(H \oplus H_0) = \emptyset \), and so \( \text{dom}(H'') \cap \text{dom}(H_0) = \emptyset \). (4), (5) and (6) are true.

12. ret case: trivial because \((\beta \oplus \beta_0)(r) = \beta(r)\).

### 6.3 Correction of fireₜ

**Theorem 8. (Correction of fireₜ).** Let \( M \) be a marking, \( H \) a heap and \( p_{mrq} \) a pointer on a structure marking such that \( \text{dom}(H) = p_{mrq} \downarrow H \) and \( \llbracket H, p_{mrq} \rrbracket^{mrq} = M \). Let \( \beta = \{ x_1 \mapsto v_1, \ldots, x_n \mapsto v_n \} \) be a LLVM mode of the transition \( t \), which implies that each \( v_i \) is a value or a pointer encoding a token (for the place \( s_i \)) and \( \llbracket H, v_i \rrbracket^{(s_i)} \in M(s_i) \). Let \( F \) be a frame such that \( \beta_F = \beta \oplus \{ x_{mrq} \mapsto p_{mrq} \} \).

If

\[
M[t, \beta]M' \quad \text{and} \quad \left( \text{fcall } \text{fire}_{t}(x_{mrq}, x_1, \ldots, x_n) \right)_{H,F} \rightsquigarrow \left( p'_{mrq} \right)_{H \oplus H'}
\]

then

\[
\llbracket H \oplus H', p'_{mrq} \rrbracket^{mrq} = M' \quad \text{and} \quad \text{dom}(H) \cap \text{dom}(H') = \emptyset
\]

In order to achieve the proof we need two auxiliary lemmas: one for the token consumption and one for the token production.
Lemma 1. Let $H$ be a heap, $p'_m$ a pointer on a structure marking such that $dom(H) = p'_m \downarrow H$ and $\left[H, p'_m\right]_{m} = M$. Let $F$ be a frame such that $\beta_F \equiv \{x'_m \mapsto p'_m, x_1 \mapsto v_1, \ldots, x_n \mapsto v_n\}$ where all $v_i$ are pointers or values of $s_i$ places types and $\left[H, v_i\right]_{M(s_i)} \in M(s_i)$ (for $1 \leq i \leq n$). We consider the seq sequence corresponding to the consumption of tokens in the $f_i$ function:

$$seq = \begin{cases} x_{s_1} = \text{fcall get}_{s_1}(x'_m) \\
\text{pcall rem}_{s_1}(x_{s_1}, x_1) \\
\ldots \\
x_{s_n} = \text{fcall get}_{s_n}(x'_m) \\
\text{pcall rem}_{s_n}(x_{s_n}, x_n) \end{cases}$$

If one reduces the sequence totally, i.e., if

$$(seq)_{H,F} \rightarrow^* (\text{skip})_{H \oplus H', F \oplus \beta}$$

then

$$\beta \equiv \{x_{s_1} \mapsto p_{s_1}, \ldots, x_{s_n} \mapsto p_{s_n}\} \quad (18)$$

$$dom(H) \cap dom(H') \subseteq p_{s_1} \downarrow H \cup \cdots \cup p_{s_n} \downarrow H \quad (19)$$

$$\left[H \oplus H', p'_m\right]_{m} (s_i) = M(s_i) - \left[H, v_i\right]_{M(s_i)} \text{ pour } i \in \{1, \ldots, n\} \quad (20)$$

Proof. All three consequences are proved simultaneously by induction on $n$ using the specifications of get and rem functions as well as the SSA property. The SSA property is needed in the inductive case for building a suitable binding from the bidding provided by the induction hypothesis. \hfill $\square$

Lemma 2. Let $H$ be a heap and $p'_m$ a pointer on a structure marking such that $dom(H) = p'_m \downarrow H$ and $\left[H, p'_m\right]_{m} = M$. Let $F$ be a frame such that $\beta_F \equiv \{x'_m \mapsto p'_m, x_1 \mapsto v_1, \ldots, x_n \mapsto v_n\}$ where all $v_i$ are pointers or values of $s_i$ places types (for $1 \leq i \leq n$). We consider the seq sequence that corresponds to token production in the function $f_i$:

$$seq = \begin{cases} x_{s'_1} = \text{fcall get}_{s'_1}(x'_m) \\
o_{s'_1} = \text{fcall f}_{s'_1}(x_1, \ldots, x_n) \\
\text{pcall add}_{s'_1}(x_{s'_1}, o_{s'_1}) \\
\ldots \\
x_{s'_m} = \text{fcall get}_{s'_m}(x'_m) \\
o_{s'_m} = \text{fcall f}_{s'_m}(x_1, \ldots, x_n) \\
\text{pcall add}_{s'_m}(x_{s'_m}, o_{s'_m}) \end{cases}$$

If one reduces the sequence totally, i.e., if $(seq)_{H,F} \rightarrow^* (\text{skip})_{F \oplus \beta}$, then

$$\beta \equiv \begin{cases} x_{s'_1} \mapsto p_{s'_1}, \quad o_{s'_1} \mapsto f_{s'_1}(v_1, \ldots, v_n), \\
\ldots \\
x_{s'_m} \mapsto p_{s'_m}, \quad o_{s'_m} \mapsto f_{s'_m}(v_1, \ldots, v_n) \end{cases} \quad (21)$$
\[ \text{dom}(H) \cap \text{dom}(H') \subseteq p_{s_1} \downarrow_H \cup \cdots \cup p_{s_m} \downarrow_H \]  \hspace{1cm} (22)

\[
\left[H \oplus H', p_{mrq}^{\prime}\right] \left[H \oplus H', J_H^{\prime}(v_1, \ldots, v_n)\right] \left(H(s_i)\right) \wedge 1 \leq i \leq m \]  \hspace{1cm} (23)

Proof (Correction fire_t). The proof of the function fire_t uses the two lemmas: one for the token consumption, and one for the token production. Both proved by induction on the number of places in the structure marking. The final proof consists in applying the two previous lemmas.

Corollary 3. Under the same hypothesis, the call of function fire_t terminates.

Proof. Immediate consequence of the correction theorem.

6.4 Correction of succ_t

Theorem 9. (Correction of succ_t). Let F be a frame and p_{mrq} a pointer in a heap H such that p_{mrq} \downarrow_H = \text{dom}(H), \left[H, p_{mrq}\right] = M, \beta_F(x_{mrq}) = p_{mrq} and \beta_F(x_{set}) = p_{set}. If

\[
\left(jcall\ succ_t(x_{mrq})\right)_{H,F} \rightsquigarrow^* \left(p_{set}\right)_{H \oplus H',F} \; \text{and} \; \left[H, p_{set}\right] = E
\]

then

\[ \text{dom}(H) \cap \text{dom}(H') \]  \hspace{1cm} [H \oplus H', p_{set}]^\text{set} = E \cup \{M' | \exists \beta, M[t,\beta]M'\}

In order to achieve the proof we introduce a property that will serve as an invariant.

Property 10. A pair (H, F) where H is a heap and F a frame respects the property P if the following conditions are true:

- p_{mrq} \in \text{dom}(H) and \left[H, p_{mrq}\right] = M;
- p_{set} \in \text{dom}(H) and \left[H, p_{set}\right] is a set of markings;
- F(x_{mrq}) = p_{mrq};
- F(x_{set}) = p_{set}.

The first lemma ensures that if the execution flow passes through a header_t,k block then it will reach a footer_t,k+1 block. Moreover, it guarantees that we will execute any block annotated by an index greater than k.

Lemma 3. Let n be the number of input places of t, let H be a heap and F a frame such that (H, F) satisfies P, then we have:

\[
\left(P(\text{header}_{t,n})\right)_{H,F} \rightsquigarrow^* \left(P(\text{footer}_{t,n+1})\right)_{H',F'}
\]
where \((H', F')\) satisfies \(P\) and without executing any block with an index greater than \(n\), i.e.,
\[
(P(\text{header}_{t,n}))(H, F) \leadsto^* (P(\text{block}_{t,i}))(H', F')
\]
for \(i > n\) and block \(\in \{\text{header, loop, body, footer}\}\) is not a strict prefix of the previous reduction.

**Proof.** We proceed by induction on \(n\) with a second induction on the number of tokens remaining in the \(n^{\text{th}}\) place during the inductive case.

**Corollary 4.** Under the same hypothesis, the call of function \(\text{succ}_t\) terminates.

**Proof.** The proof proceeds by reducing the body of the \(\text{succ}_t\) function, checking that the heap and the frame verify the property \(P\) and applying the previous lemma.

The second lemma is used to show that we enumerate all combinations of tokens for the input places. It implies that we enumerate all modes of \(t\).

**Lemma 4.** Let \(H\) be a heap and \(F\) a frame such that \((H, F)\) satisfies \(P\). Let \(\beta = \{x_n \mapsto v_n, \ldots, x_1 \mapsto v_1\}\) be a LLVM binding such that \([H, v_i]^{(s_i)}(s_i) \in M(s_i)\) for \(1 \leq i \leq n\). If
\[
(P(\text{header}_{t,n}))(H, F) \leadsto^* (P(\text{header}_{t,0}))(H', F')
\]
then \(\beta \subseteq \beta F'\) and \((H', F')\) satisfies \(P\). Moreover, the execution flow did not reach any block indexed by \(i\) such that \(i > n\), i.e.,
\[
(P(\text{header}_{t,n}))(H, F) \leadsto^* (P(\text{block}_{t,i}))(H'', F'')
\]
for \(i > n\) and block \(\in \{\text{header, loop, body, footer}\}\) is not a strict prefix of the previous reduction.

**Proof.** The proof proceeds exactly like the previous lemma.

**Proof (correction \(\text{succ}_t\)).** The first result is showed by remarking that any pointer is used explicitly, thus it can be showed by using the specifications of called functions. The second result is showed in two steps:

1. This part is proved using two auxiliary lemmas:
   - one that if the execution flow passes through a \(\text{header}_{t,k}\) block then it will reach a \(\text{footer}_{t,k+1}\) block. Moreover, it guaranties that we will execute any block annotated by an index greater than \(k\);
   - the second lemma is used to show that we enumerate all combinations of tokens for the input places, which implies that we enumerate all modes of \(t\).

2. next we show that we only add the successor markings into the set, which gives the other inclusion. This part is proved by remarking that if a marking is added into the set, then it is added from the \(\text{body}_{t,0}\) block. So we show that this block is executed if and only if the binding contains a mode of \(t\).

**Corollary 5.** Under the same hypothesis, the call of function \(\text{succ}_t\) terminates. \(\square\)
7 Conclusion

We have shown how a Petri Net can be compiled, targeting a fragment of the LLVM language. This compilation process produces code that provides the primitives to compute the state space of the compiled Petri net model. Then we have defined formal semantics for the fragment of the LLVM language we use. To produce a readable and usable system of inference rules, we have defined a memory model based on heaps and stacks. Finally we presented two theorems proving the correction of the code generated by our compiler. The detailed proofs provided in [5] are quite long because they are very much detailed to improve our confidence into their correctness, but they are at the same time easy to follow.

Current and future work will address a generalisation of the presented approach to compile a wider variety of coloured Petri nets, in particular nets embedding annotation languages easier to use for the modeller than LLVM. We are also interested in particular in exploiting remarkable structures of Petri net models that allow to optimise the code generated by the compiler. Such optimisations also need to be formally proved; preliminary results about this can be found in [5]. A complementary aspect is to evaluate the performance of the state space generation, that is of course another important motivation, see [6].

References

3. ADT Coq/INRIA. The Coq proof assistant. ⟨http://coq.inria.fr⟩.
11. C. Lattner and al. LLVM related publications. ⟨http://llvm.org/pubs⟩.