Abstract
Software applications often require the transformation of an input source program into a translated one for optimization. In this process, preserving the semantics across the transformation also called equivalence checking is essential. In this paper, we present ongoing work on a novel translation validation technique for handling loop transformations such as loop swapping and distribution, which cannot be handled by state-of-the-art equivalence checkers. The method makes use of a reduced size Petri net model integrating SMT solvers for validating arithmetic transformations. The approach is illustrated with two simple programs and further validated with a programs benchmark.

Keywords: Translation Validation, Equivalence Checking, Color Petri Net, Z3 Theorem Prover

1 INTRODUCTION
Software applications often require the transformation of an input source program into a translated version while preserving the semantics across the transformation. These translations are performed to efficiently utilize the intrinsic computer architecture, such as multiple cores and vector registers. For safety-critical systems, these translations need to be formally validated before they are used, to certify system reliability and accuracy. Checking the equivalence of the functional behaviors of source and translated programs is thus an important step. This process is called translation validation.

Instruction-level parallelism is one such translation that is widely used in high level synthesis during the scheduling phase. Petri nets are a popular modeling paradigm that can capture and express instruction-level parallelism.

Path-Based Equivalence Checking (PBEC) is a popular method for translation validation, which is based on graphical models/representations of code. Petri net PBEC methods have been proposed in [14, 16] but they are not able to validate code with complex arithmetic expressions. CDFG PBEC [9] methods are not able to validate parallelizing transformations either.

Satisfiability Modulo Theories (SMT) solvers are tools used to solve constraint satisfaction problems. They are used in verification as a means of analyzing the symbolic execution and semantics of programs. Z3 Theorem Prover [2] is an industry-standard SMT Solver developed by Microsoft Research to solve such problems.

In this paper, we propose an approach for translation validation of several loop-involving and parallelizing code transformations. The approach, which is a work-in-progress, has three major parts: a Petri net model constructor, a Petri net path constructor, and an equivalence checker which consists of a path analyzer and the Z3 Theorem Prover.

The major contributions of this paper are as follows:

- Approach to validate several transformations such as loop swapping and distribution, and parallelization which cannot be handled by state-of-the-art CDFG-based equivalence checkers.
- Refinement and reduction in size of Petri net model from that employed in [14], which enhances the efficiency of the equivalence checking mechanism and helps with scalability issues.
- Integration of SMT solvers in the approach to check equivalence between two programs.

This paper is organized as follows: Section 2 presents an overview of the entire workflow of the approach. Through a motivating example, the workflow is explained in Section 3. Through a small set of experimentation, we compare our approach with [14, 15] and other CDFG-based PBEC. Section 4 compares the experimental results with these equivalence checkers. Section 5 describes the state of the art. We conclude the paper in Section 6.

2 WORKFLOW
The workflow of the proposed approach is illustrated in Fig. 1. Initially, a source program $P_s$, is subjected to a series of transformations, that result in a translated program $P_t$. In our approach, we have used Color Petri Net (CPN) as an intermediary modeling paradigm. This task is performed by the Model Constructor module which outputs two CPNs: $N_0$ and $N_1$, corresponding to the source and translated programs respectively.

To formally check behavioral equivalence between programs, there is a necessity to characterise the computations. However, in the case of loop(s), the number of loop traversals is indeterminate. To overcome this computational barrier, we represent the CPN model computations as a finite set of paths. This task of extracting the set of paths is performed by Path Analyzer module, which gives the set $\pi_0$ from $N_0$ and $\pi_1$ from $N_1$.

Using the path-cover data, the process of equivalence checking is carried out by the Path-Based Equivalence Checking (PBEC) module that is composed of the Path Analyzer and Z3 Theorem Prover. The equivalence checking process is dynamically performed by the Path Analyzer module.

This establishment of equivalence (or non-equivalence) of the characteristics of the two programs (rather, their corresponding path covers) is facilitated by the Z3 Theorem Prover. To utilize Z3, the Path Analyzer module generates a set of Z3-compatible input expressions from the path cover data ($I_0$ from $\pi_0$ and $I_1$ from $\pi_1$), to check for equivalence between paths.

After all candidate paths have been checked, a ‘Yes’ answer from the Path Analyzer implies equivalence while a ‘No’ answer is interpreted as ‘Can’t Say’, since the proposed equivalence checking method is sound but not complete. In the case of ‘Yes’, the Path Analyzer also outputs the equivalent pairs of paths from $N_0$ and $N_1$.

3 MOTIVATING EXAMPLE

The program $P_s$ takes five inputs $a$, $b$, $l$, $m$, and $n$, and computes the function:

$$k = (m \times 10^l) + (n \div 10^l) + (a - b)$$

The corresponding transformed program $P_t$ is obtained by loop distribution followed by thread level parallelizing transformation of $P_s$; the independent sub-expressions $m \times 10^l$ and $n \div 10^l$ are computed separately in two parallelized loops.

3.1 Model Formalism

A Petri net model $N$, is a bipartite directed graph; one subset $P$, say, of vertices comprises places and the other subset $T$, say, comprises transitions. If there is an arc $(p, t)$ from a place $p$ to a transition $t$, then $p$ is called a pre-place of $t$ and the arc is called in-coming arc of $t$. The set of all pre-places of $t$ is denoted as $\pi_t$. If there is an arc $(t, p')$ from a transition $t$ to a place $p'$, then $p'$ is called a post-place of $t$; the set of all post-places of $t$ is denoted as $\pi_t$. The arc is called an out-going arc of $t$.

The set $P_{in} \subset P$ is designated as the set of in-ports of the model. It comprises all places that are not post-places of any transition. Similarly, another set $P_{out} \subset P$ is called the set of out-ports, which comprises the places that are not pre-places of any transition. A place can hold an entity called token. A token is a set of variable-value pairs that can hold values for program variables. The marking of a net is a particular distribution of tokens over the net.

Each out-going arc is associated with a set of functions. This function-set $F$, say, is a set of arithmetic expressions over (a subset of) the program variables. Each transition $t$ is associated with a guard condition $g_t$, which is a Boolean function over (a subset of) the program variables. A transition $t$ is said to be enabled when all its pre-places have tokens and they hold values which satisfy $g_t$. Consequently to the firing of an enabled transition $t$, tokens are removed from all $p \in \pi_t$ and tokens are placed in all $p \in \pi'$. The value vector of the token(s) in the post-place(s) depends respectively, on the associated function-set $F$.

Each place $p \in P$ is associated with a vector of program variables $V_p$, say. For places that are in-ports, the vector consists of no variables. For places that are neither in-ports nor out-ports, there are two kinds of such variables: changed variables and unchanged variables. Changed variables are those variables whose values are changed from when the place was last marked. Similarly, unchanged variables are those whose values don’t change. The partition between changed and unchanged variables for each place, is defined dynamically during the computations of the Petri net and the same will be illustrated in the next subsection. Out-ports have no changed variables in the associated variable vector.

3.2 Model Construction

Using compiler internal infrastructure, the program is transformed into an intermediate representation. This representation is transformed to a Control Flow Graph (CFG) using the $fdump$ process of the GCC compiler. The elements of the CFG are mapped to the Petri net, for which, a rudimentary set of rules is described in Table 1.
3.3 Notion of Path on CPN Model

In a general program, the number of loop traversals is unbounded. Therefore, we cannot characterize the set of computations. From the classical program verification techniques, we introduce the concept of paths such that any computation can be represented in terms of a finite set of paths. To construct the path, we introduce the notion of cut-points. Using cut-points we ‘cut’ each loop. The notion of cut-points in our CPN model is as follows:

1. All in-ports, ∀p ∈ P_in, are cut-points.
2. All out-ports, ∀p ∈ P_out, are cut-points.
3. All places that have back-edges are cut-points.

A path is a sequence of out-going arcs from a set of cut-points to a cut-point, while having no cut-point in between. Through the backward cone of foci method, we construct the paths in the Petri net model. The detailed discussion of the path construction algorithm is given in [1]. It is to be noted that if an out-going arc is covered in one path, it need not be considered in another path.

3.4 Validity of PBEC

To prove the validity of the path-based equivalence checker, we show that any computation can be represented as a concatenation of parallel paths.

As an example, taking the translated model N1 in Fig. 3, we can express the computation as follows:

\[ \mu'_{p_4} = \langle \{t'_4, t'_5 \rangle, \{t'_3, t'_4 \} \rangle \]

We can express the same computation in terms of the sequence of transitions that are fired. :

\[ \mu'_{p_4} = \langle \{t'_3, t'_4 \} \rangle \]

We can now express the computation in terms of the out-going arcs. :

\[ \mu'_{p_4} = \langle \{(t'_4, p'_4), (t'_5, p'_5)\}, \{(t'_3, p'_3), (t'_4, p'_4)\} \rangle \]

The set of paths of N1, \( \pi_1 = \{p_1, p_2, p_3, p_4, p_5\} \). Initially, \( \mu'_{p_4} = \emptyset \). The last member of \( \mu'_{p_4} \) is \( (t'_4, p'_4) \). The path \( \beta_3 \) has \( (t'_3, p'_3) \) as its last member. So \( \beta_3 \) is prepended to \( \mu'_{p_4} \), and all the out-going arcs in \( \beta_3 \) (only \( (t'_4, p'_4) \)) are removed once from \( \mu'_{p_4} \).

Now, the last member of \( \mu'_{p_4} \) is \( \{t'_3, p'_3, t'_5, p'_5\} \). \( (t'_5, p'_5) \) is the last member of \( \beta_3 \) and \( (t'_3, p'_3) \) is the last member of \( \beta_5 \). So \( \beta_3 \) is prepended to \( \mu'_{p_4} \), and all the out-going arcs from \( \beta_3 \) and \( \beta_5 \) are removed once from \( \mu'_{p_4} \). This step will be repeated \( l - 1 \) times until the only element left in \( \mu'_{p_4} \) is \( \{(t'_4, p'_4), (t'_3, p'_3)\} \).

Since \( (t'_3, p'_3) \) is the last element of \( \beta_1 \) and \( (t'_4, p'_4) \) is the last element of \( \beta_2 \), \( \beta_1 \parallel \beta_2 \) is prepended to \( \mu'_{p_4} \). The algorithm is
now terminated since $\mu_p'$ is empty. Therefore
\[ p' = (\{\beta_1\}, \{\beta_3\}, \{\beta_4\}) \]

3.5 Equivalence Checking Mechanism

There are two entities associated with every path
1. Condition of execution, $R_\alpha$, which is associated with the guard conditions $g$.
2. Data transformation, $r_\alpha$, which is associated with the function-sets $F$.

Two paths $\alpha$ and $\beta$ are considered equivalent when $R_\alpha \approx R_\beta$ and $r_\alpha = r_\beta$. The equivalence checking mechanism is based on the principle: "$\forall \alpha \in \pi_0, \exists \beta \in \pi_1 \land \forall \alpha \in \pi_0 (\alpha \approx \beta \implies \pi_0 \approx \pi_1 \implies N_0 \approx N_1$". During checking, the algorithm constructs correspondence relationships between the places, variables, and transitions, respectively. To check two arithmetic or logical expressions, we integrate the Z3 Theorem Prover with the equivalence checker. Following are the informal algorithmic steps for checking equivalence between $N_0$ and $N_1$:

In our motivating example, the set of paths in $N_0$ and $N_1$ are \{e, f, g\} and \{d, e, f, g, h\} respectively. Also, $R_\alpha = g_1$, $R_\alpha = g_2$, $R_\alpha = g_3$, and $r_\alpha = F_1, r_\alpha = F_2, r_\alpha = F_3$. Similarly, $R_\beta = g_4$, $R_\beta = g_5$, $R_\beta = g_6$, $R_\beta = g_7$, $R_\beta = g_8$, and $r_\beta = F'_1, r_\beta = F'_2, r_\beta = F'_3, r_\beta = F'_4$.

**Step 1** Taking the first element of $\pi_0$, i.e. $\alpha_1$, we look at its pre-place $p_1$. Places $p_1$ and $p_1'$ correspond to each other since they are in-ports. Since $p_1'$ is a pre-place for paths $\beta_1$ and $\beta_2$, these two paths are candidate paths for $\alpha_1$. The SMT solver tells us that $R_{\alpha_1} \approx R_{\beta_1}$ (i.e. $g_{\alpha_1} = g_{\beta_1}$) and $R_{\alpha_1} \approx R_{\beta_1}$ (i.e. $g_{\alpha_1} = g_{\beta_1}$). The SMT solver also tells us that $r_{\alpha_1} = r_{\beta_1}$ (i.e. $F_1 = F'_1$) and $r_{\alpha_1} = r_{\beta_1}$ (i.e. $F_1 = F'_1$). Hence, $\alpha_1 \approx \beta_1$. From this information we also infer that the post-places of these paths correspond to each other, i.e. $p_2$ corresponds to $p_2'$ and $\beta_2$.

**Step 2** Taking the next element of $\pi_0$ i.e. $\alpha_2$. The pre-place of $\alpha_2$ is $p_2$, which corresponds to $p_2'$ and $p_2'$. Since $\beta_1$ and $\beta_2$ have the two places as their pre-place, they are candidate paths for $\alpha_2$. Checking for equivalence between these paths results in a ‘No’ answer from the SMT solver. So, we go for path extension. The paths $\beta_3$ and $\beta_3$ can be merged parallelly, due to place, variable, and transition correspondence. The SMT solver tells us that $R_{\alpha_2} \approx R_{\beta_3}$ and $R_{\alpha_2} = R_{\beta_3}$. Similarly, $R_{\alpha_2} = R_{\beta_3}$. Hence, $\alpha_2 \approx \beta_3$.

**Step 3** Finally, taking the path $\alpha_3$, it’s pre-place is $p_3$ which has correspondence to $p_3'$ and $p_3'$ which are the pre-places of $\beta_4$. Similarly, the post-place of $\alpha_3$ corresponds to the post-place of $\beta_4$ since they are out-ports in their respective nets. Hence, $\beta_4$ is a candidate path for $\alpha_3$. The SMT solver tells us that $R_{\alpha_3} \approx R_{\beta_3}$ and $R_{\alpha_3} = R_{\beta_3}$. Hence, $\alpha_3 \approx \beta_4$. So, $\alpha_3 \approx \beta_4$.

Since “$\forall \alpha \in \pi_0 \exists \beta \in \pi_1 | \alpha \approx \beta \implies \pi_0 \approx \pi_1 \implies N_0 \approx N_1$". That is, the programs in Listing 1 and Listing 2 are semantically equivalent.

3.5.1 Z3 Theorem Prover

For two candidate paths $\alpha$ and $\beta$, the Z3 Theorem Prover (Z3) receives the conditions of execution, $R_\alpha$ and $R_\beta$, and the data transformation, $r_\alpha$ and $r_\beta$, from the path analyzer. All the program statements are encoded as Static Single Assignments to preserve the order of execution. The sub-scripts ‘s’ and ‘t’ are appended for variables of $P_s$ and $P_t$ respectively. The input to Z3 consists of:

1. Variables and corresponding type declarations.
2. Functions in the form of assert statements
3. Test statements asserted as negations. Z3 returns a sat (true) answer if it finds even one case (from the entire model space) that satisfies equivalence. Using the negation, we can test that equality is satisfied over the entire model space. Mathematically: for $\xi$ (the model space) and $c$ (the cases), by De Morgan’s Law, $\neg(\bigcup_{c \in \xi} c) = \bigcap_{c \in \xi} \neg c$.

So, an unsat output from Z3 actually corresponds to equivalence and a sat output implies non-equivalence. Also, the test statements check for equality only between the common variables of $P_s$ and $P_t$. In case of multiple assignment of the same variable, only the last executed variable is considered (i.e. the variable with highest numerical suffix).

As an example, in **Step 3** for checking equivalence between $R_{\alpha_3}$ and $R_{\beta_4}$, the Z3 input is as follows:

```
(declare-const g3_s Bool) (declare-const g4_s Bool) 1
(declare-const i0_s Int) (declare-const g4_s Bool) 2
(declare-const i0_s Int) (declare-const j0_s Int) 3
(declare-const i0_s Int) 4
(assert (= g3_s (< i0_s l_s))) 5
(assert (= g4_s (and (>= i0_s l_s) (< j0_s l_s)))) 6
(assert (= l_s l_t)) (assert (= i0_s j0_t)) 7
(assert (= i0_s i0_t)) 8
(check-sat) 9
```

**Listing 3.** Checking equivalence of $R_{\alpha_3}$ and $R_{\beta_3}$

In Listing 3, lines 1-4 define the guard conditions as Boolean functions and define the associated variables. Lines 5-6 define $g_{\alpha_3} = i \geq l$ and $g_{\beta_4} = i \geq l \land j \geq l$. To facilitate equivalence checking, equivalence between variables is asserted in lines 7-8. $i_0 \land t = i_0 \land t$ is inferred from $F'_1$ and $F'_2$. Line 9 is the assert statement for equivalence checking defined as a negation. In the last line we check equivalence. Z3 returns unsat which implies $R_{\alpha_3} = R_{\beta_4}$.

4 EXPERIMENTAL RESULTS

We manually tested our equivalence checking algorithm on five examples, where parallelising transformations are applied using Pluto [18] and Par4All [11] compilers. The programs and their descriptions can be found in [1].

Table 2 presents a comparative study of the model size of our proposed approach with the models of two other Petri net-based equivalence checking tools ST-1 [15] and ST-2 [14]. It is to be noted that the model size of the current method is comparable with ST-2.

Table 3 presents transformation verification capabilities of the proposed approach, compared with ST-1, ST-2 and two CDFG (Control Data Flow Graph) based tools namely, FSM-DVP (FSMD with Value Propagation) [10] and FSM-DVP (FSMD with Extended Value Propagation) [13]. It is to be noted that
Table 2. Model size for different Petri-net PBEC

<table>
<thead>
<tr>
<th>Example</th>
<th>ST-1</th>
<th>ST-2</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCM</td>
<td>34</td>
<td>28</td>
<td>6  6  3  2</td>
</tr>
<tr>
<td>MINMAX</td>
<td>31</td>
<td>27</td>
<td>7  7  4  6</td>
</tr>
<tr>
<td>PETERSON</td>
<td>11</td>
<td>9</td>
<td>4  2  6  8</td>
</tr>
<tr>
<td>DEKKERS</td>
<td>19</td>
<td>14</td>
<td>6  6  8  8</td>
</tr>
<tr>
<td>LUP</td>
<td>28</td>
<td>21</td>
<td>6  4  10 16</td>
</tr>
</tbody>
</table>

Table 3. Capabilities of different PBEC

<table>
<thead>
<tr>
<th>Example</th>
<th>FSMD-VP</th>
<th>FSMD-EVP</th>
<th>ST-1</th>
<th>ST-2</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCM</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>MINMAX</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PETERSON</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>DEKKERS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>LUP</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

both FSMD based PBEC cannot handle the parallelizing transformations because FSMD is a sequential model of computation. ST-1 and ST-2 cannot handle arithmetic transformations. They have their own normalizer, which are their limitations. These limitations are overcome by Z3.

5 RELATED WORK

Translation validation was introduced in [4] and was demonstrated in [20] and [12]. The approach was enhanced in [17]. All these techniques are bisimulation based methods. A bisimulation method for parallel programs is reported in [19].

Another equivalence checking method is the inductive-inferencing based technique reported in [7]. The method only works for scalar handing programs.

A major limitation of these methods is that termination is not guaranteed. To alleviate this shortcoming, a path based equivalence checker for the FSMD model was proposed for uniform and non-uniform code motions, code motion across loop and loop invariant code optimizations in [6, 10, 13]. However, they cannot handle loop swapping transformations and many thread-level parallelizing transformations because FSMDs cannot capture parallel behaviors easily.

The literature records no significant attempts for devising formal equivalence checking methods using Petri net based models. Although, there are several works on property verification using Petri net modelling paradigm [3, 5, 8, 21]. In [16], the validation of loop swapping and thread level parallelising transformations using Petri nets was reported. The major limitation of this method is it cannot handle loop invariant code motion as well as polynomial arithmetic transformations. Also, the model size presents a scalability issue. To over come the limitations, a modification in the model construction an equivalence checking was reported in [14]. However, the method cannot handle polynomial arithmetic transformations.

6 CONCLUSION

In this paper we presented our ongoing work on developing an approach to check the equivalence of software programs using a novel translation validation technique for handling loops. In addition, our approach makes use of SMT solvers to validate arithmetic transformations. Such constructions cannot be handled by state-of-the-art equivalence checkers.

We presented an initial validation of the approach for a standard benchmark. Currently this validation was performed manually. Therefore, our future work is to implement a tool-chain supporting the approach and validate it on a larger benchmark. For this, we will reuse existing compiler front-ends (e.g. GCC) and automatically construct the Petri Net models from the generated intermediate code representation so that the approach can be tested on different programming languages, potentially including existing architecture description languages such as UML, SysML and AADL. This will also allow us to further characterize the domain of applicability of the approach; i.e. which language constructions and translations are handled by our approach and to evaluate scalability for large programs.

Acknowledgments

To Soumyadip and Dominique, for being amazing guides.

References

[18] Uday Bondhugula et al. 2008. PLuTo: A practical and fully automatic polyhedral program optimization system. In PLDI.