FAR-Cubicle - A new reachability algorithm for Cubicle

Sylvain Conchon∗† Amit Goel‡ Sava Krstic§ Rupak Majumdar∗ Mattias Roux∗

∗LRI, Université Paris Sud CNRS, Orsay F-91405
†INRIA Saclay – Ile-de-France, Orsay cedex, F-91893
‡Apple
§Intel Corporation
¶Max Planck Institute for Software Systems

Abstract—We present a fully automatic algorithm for verifying safety properties of parameterized software systems. This algorithm is based on both IC3 and Lazy Annotation. We implemented it in Cubicle, a model checker for verifying safety properties of array-based systems. Cache-coherence protocols and mutual exclusion algorithms are known examples of such systems. Our algorithm iteratively builds an abstract reachability graph refining the set of reachable states from counter-examples. Refining is made through counter-example approximation. We show the effectiveness and limitations of this algorithm and tradeoffs that results from it.

1. Introduction

We describe FAR (Forward Abstracted Reachability), an algorithm for fully automatic verification of parameterized software systems. A parameterized system describes a family of programs such as cache coherence protocols where the number of processes involved can change but the algorithm handling their behaviour is the same for all of them. Thus, the parameter allows to talk about these algorithms without knowing the actual number of processes that will be involved and then to prove its safety regardless of this number.

Safety properties state that “nothing bad happens” in our parameterized system. Verifying them can be reduced to finding an invariant of it. Building an invariant can be hard (and even undecidable [1]). The standard approach to do so is to find a formula Φ such that Φ is an inductive invariant of the system (i.e., the initial state of the system satisfies Φ and taking a transition from a state satisfying Φ leads to another state satisfying Φ).

In this paper we describe an algorithm for the automatic construction of inductive invariants for array-based systems (Section 2). This algorithm, based on both IC3 [2] and Lazy Abstraction [3], builds an inductive invariant by unwinding a graph (Section 3) building a forward abstract reachability of our system. This unwinding is described as a set of non-deterministic rules. We then provide an implementation in Cubicle [4], [5], [6] (Section 5) of these rules and test its effectiveness on several cache coherence protocols (Section 6).

2. Array-based Systems

An array-based system is described in [7] as first-order logic formulas on arrays. Such a system can be described as a set of basic types, a set X of system variables associated to type (built as usual with basic types and standard constructions), a formula init representing the initial states and a set Δ of transition rules τi(X, X′) (X′ is the set X where all the variables are primed which represents the next state reached after the application of a transition). Since we work on parameterized programs, our arrays are indexed by an infinite type proc.

We describe the Dekker mutual exclusion algorithm as an array-based system. Each process has two boolean variables, want (stating that the process wants to enter in critical section or not) and crit (stating that the process is in critical section or not). There is a global variable turn of type proc that tracks which process can go into the critical section. Since we work in the array-based systems fragment, we represent the local variables as arrays indexed by processes and containing booleans. The set X contains two arrays, want[proc] : bool and crit[proc] : bool and the global variable turn : proc. Initially, no process is or wants to be in critical section. Three transitions can be triggered, one to require an access to the critical section, one to enter it and one to exit it. According to the previous description, we write this algorithm as in Figure 1.

Since we focus on safety problems (nothing bad happens), we need to define what is considered as bad states. For Dekker algorithm, these states would be defined with the following formula:

\[ U \equiv \exists p_1, p_2, p_1 \neq p_2 \land crit[p_1] \land crit[p_2] \]

Our goal is then to prove that no state represented by U is reachable from init (which can be formulated as: there exists no path init = X_0 \xrightarrow{p_1} X_1 \xrightarrow{p_2} \cdots \xrightarrow{p_n} X_n = U with p_i \in \{req, enter, exit\}). To do so on parameterized systems, one of the main algorithms came from Ghilardi et al. with MCMT [8]. It builds the set of all reachable states by backward reachability (starting, then, from the unsafe state) and checks if this set contains an initial state. In this paper,
we implement a different algorithm which offers a wider range of possibilities in terms of reachability construction.

3. Program unwinding

This algorithm also starts from $U$ but tries to build an invariant of the system that does not contain it. Before going into details, we give a brief explanation. This invariant is iteratively built as an inductive invariant $\Theta$ that does not contain $U$:

- if $\Theta \land \Delta \land \neg\Theta'$ is unsatisfiable then we found an inductive invariant
- if $\Theta \land \Delta \land \neg\Theta'$ is satisfiable, our candidate invariant is not inductive and we try to refine it until we either discover that there is no such refinement or we find some.

For Dekker algorithm, for example, let’s take $\Theta = \neg U = \forall p. \neg want[p] \land \neg crit[p]$:

- $\Theta \land \Delta \land \neg\Theta'$ is satisfiable (if we take, for example, the following state: $\varphi_1 = crit[p_1] \land want[p_2] \land \neg crit[p_2]$, $\varphi_1 = \Theta$ but if we apply $enter$ to it we obtain the state $\varphi_2 = crit[p_1] \land crit[p_2] \land \ldots$ and $\varphi_2 \not\models \Theta$).
- We need to create $\Theta'' = \Theta \land \rho$ which is a refinement of $\Theta$ that does not contain $\varphi_1$.

To do so, we build an unwinding of the algorithm as a quadruple $(V, E, W, B)$, where:

- $(V, E)$ is a rooted graph with edges labeled by transitions from $\Delta$;
- $W$ associates a formula (called world of the vertex) to each vertex;
- $B$ associates a formula (called bad part of the vertex) to each vertex.

This graph contains three initial vertices:

- $\epsilon$ : the root vertex, $W(\epsilon) = init$ and $B(\epsilon) = \perp$;
- $\beta$ : the unsafe vertex, $W(\beta) = \top$ and $B(\beta) = U$;
- $\omega$ : the sink vertex, $W(\omega) = \perp$ and $B(\omega) = \perp$.

We define $V^\epsilon = \{ v \in V, \epsilon \xrightarrow{*} v \in E \}$ (i.e. the set of vertices that are linked to the root) and $(F \models f G) \equiv (F \land f \models G)$

The idea behind this unwinding it that if we manage to create a graph $G$ of a system $S = \langle init, \Delta \rangle$ where every vertex in $V^\epsilon$ does not contain a bad part and from which no more transitions can be taken, then the disjunction of their worlds ($\Theta = \bigvee_{v \in V^\epsilon} W(v)$) is an invariant of the system ($init \models \Theta$ and $\Theta \models \Delta \Theta$).

We now propose a set of non-deterministic rules for building this unwinding. Let $(X, init, \Delta, U)$ be an array-based system. Initially, $G$ is defined as follows:

- $V = \{ \epsilon, \omega, \beta \}$
- $E = \emptyset$

The unwinding works by the non-deterministic application of the following rules:

**Rule 1 (Extend).** If $\exists v, \tau \in \Delta. W(v) \models \top$ and $\exists v', v \xrightarrow{\tau} v' \in E$ then $E = E \cup \{ v \xrightarrow{\tau} \beta \}$

**Rule 2 (Refine).** If $\exists v, v' \in V, \tau \in \Delta. v \xrightarrow{\tau} v' \in E$, $B(v') \not\models \perp$, $\exists \varphi. W(v) \models \varphi$ and $\varphi \models \neg B(v')$ then we create a new vertex $v''$ such that $W(v'') = W(v') \land \varphi$ and $E = E \cup \{ v \xrightarrow{\tau} v'' \} \setminus \{ v \xrightarrow{\tau} v' \}$

**Rule 3 (Propagate).** If $\exists v, v' \in V, \tau \in \Delta. v \xrightarrow{\tau} v' \in E$, $B(v') \not\models \perp$, $\exists \gamma. W(v), \varphi \models \gamma$ and $\gamma \models B(v')$ then $B(v) \leftarrow \gamma$

**Rule 4 (Cover).** If $\exists v, v' \in V, \tau \in \Delta. v \xrightarrow{\tau} v' \in E, v'' \in V$ such that $W(v'') = W(v') \land W(v)$ then $E = E \cup \{ v \xrightarrow{\tau} v'' \} \setminus \{ v \xrightarrow{\tau} v' \}$

**Rule 5 (Sink).** If $\exists v, \tau \in \Delta. W(v) \models \perp$ and $\exists v', v \xrightarrow{\tau} v' \in E$ then $E = E \cup \{ v \xrightarrow{\tau} \omega \}$

$S$ is safe if and only if no rule can be applied to $G$, an unwinding of it and $B(\epsilon) = \perp$. Intuitively, since no more transitions can be taken and all the vertices connected to the root are not bad, $init$ will never be able to lead to $U$.

4. Example

The example shown in Figure 2 describes the first four runs of the unwinding on the Dekker algorithm (we decided not to show $\omega$ since it just serves as a sink for the transitions that can not be taken from a vertex):

(a) initially, the only rule that can be applied is **Extend** from $\epsilon$ (with $W(\epsilon) \equiv init \equiv \forall p. \neg want[p] \land \neg crit[p]$) with $req$;
(b) then we can only apply **Refine** because $init \not\models \forall req U \equiv B(\beta)$. We create a new vertex called $v_1$;
(c) we can chose here to apply **Extend** from the new vertex with any transition. We chose to take the transition $req$;
(d) $W(v_1) \not\models req B(\beta)$ and $W(v_1) \models req W(v_1)$ so we can apply the cover rule;
(e) if we keep applying these rules, we reach a fixpoint at the third created vertex.
5. Implementation

We implemented the unwinding algorithm in Cubicle 1. To do so we had to choose a deterministic strategy depending on multiple parameters:

- the order in which the rules are applied;
- which vertex and transition should be taken for the Extend rule;
- which formula \( \varphi \) should we take for the Refine rule;
- which formula \( \gamma \) should we take for the Propagate rule;
- which vertex \( v'' \) should we take for the Cover rule.

We came out with the following algorithm where we write \( v = (W, B) \) to denote the fact that \( W(v) = W \) and \( B(v) = B \):

Algorithm 1 Graph unwinding - main loop

```
procedure FAR-CUBICLE(\( S = (\text{init}, \Delta, U) \))
    \( \epsilon \leftarrow (\text{init}, \bot) \)
    \( \beta \leftarrow (\top, U) \)
    \( \omega \leftarrow (\bot, \bot) \)
    \( V \leftarrow \{ \epsilon, \beta, \omega \} \)
    \( E = \emptyset \)
    \( \text{PUSH}(Q, \epsilon) \) \>
    \( \text{Q is a priority queue} \)
    while NOT_EMPTY(\( Q \)) do
        \( v \leftarrow \text{POP}(Q) \)
        for all \( \tau \in \Delta \) do
            if \( W(v) \models \tau \top \) then
                \( E = E \cup \{ v \xrightarrow{\tau} \beta \} \)
                \( \text{UNWIND}(v \xrightarrow{\tau} \beta) \)
            else
                \( E = E \cup \{ v \xrightarrow{\tau} \omega \} \)
        return safe
```

Algorithm 2 Graph unwinding - unwinding procedure

```
procedure UNWIND(\( v \xrightarrow{\tau} v' \))
    if \( B(v) = \bot \land B(v') \neq \bot \) then
        switch CLOSE(\( v \xrightarrow{\tau} v' \)) do \>
        case Covered \( v'' \)
            \( E = E \cup \{ v \xrightarrow{\tau} v'' \} \)
            UNWIND(\( v \xrightarrow{\tau} v'' \))
        case Bad \( \varphi \)
            \( B(v) \leftarrow \varphi \)
            for all \( u \xrightarrow{\tau} v \) do
                UNWIND(\( u \xrightarrow{\tau} v \))
        case Refined \( v'' \)
            \( E = E \cup \{ v \xrightarrow{\tau} v'' \} \)
            \( \text{PUSH}(Q, v'' \downarrow) \)
```

- \( \mathcal{W}(v) \models \tau \top \) \( B(v'). \) In this case, all the edges coming to it must be unwound again;

- there exists another vertex \( v'' \) such that \( v \models v'' \) and \( \mathcal{W}(v'') \models \mathcal{W}(v') \). In this case, \( \tau \) is deleted and a new one from \( v \) to \( v'' \) is created and unwound;

If all the edges are closed and the queue is empty, the system is unsafe. If the propagation of bad parts reaches the root vertex, the system is unsafe.

Algorithm 3 Graph unwinding - closing edge procedure

```
1: procedure CLOSE(\( v \xrightarrow{\tau} v' \))
2: if \( \exists v'', \mathcal{W}(v'') \models \mathcal{W}(v') \land \mathcal{W}(v) \models \tau \mathcal{W}(v'') \) then
3: return Covered \( v'' \)
4: else if \( \mathcal{W}(v) \models \neq B(v') \) then
5: return Bad \( \text{PRE}(B(v'), \tau) \)
6: else
7: \( v'' \leftarrow (\mathcal{W}(v') \land \text{GENERALIZE}(\neg B(v')), \bot) \)
8: return Refined \( v'' \)
```

As we can see on line 5 of the CLOSE procedure, the formula \( \gamma \) chosen for the Propagate rule is the pre image of the bad formula of the vertex \( v' \). Also, on line 7 of the CLOSE procedure, the formula \( \varphi \) chosen for the Refine rule

1. cubicle.lri.fr/far
is a generalization of the negation of the bad part of the vertex \(v\). These are, of course, implementation choices. Other implementation could involve model finding, interpolants ... In our case, the generalization is a naive one consisting in taking the smallest part of the resulting formula that was not already taken and that still satisfies the conditions of the \textit{Refine} rule.

6. Benchmarks

We compared our implementation to the backward reachability algorithm already implemented in Cubicle (without the invariants inference implemented with BRAB [4], [6]) and obtained the following results (the timeout was set to 5 minutes and the \(\alpha\) version uses an abstraction engine related to the approximation implemented in BRAB to get better refinement):

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Cubicle</th>
<th>FAR</th>
<th>FAR-(\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dekker</td>
<td>0.04s</td>
<td>0.04s</td>
<td>0.03s</td>
</tr>
<tr>
<td>mux_sem</td>
<td>0.04s</td>
<td>0.05s</td>
<td>0.03s</td>
</tr>
<tr>
<td>german-ish</td>
<td>0.06s</td>
<td>0.1s</td>
<td>0.55s</td>
</tr>
<tr>
<td>german-ish2</td>
<td>0.13s</td>
<td>0.11s</td>
<td>0.65s</td>
</tr>
<tr>
<td>german-ish3</td>
<td>1.2s</td>
<td>8.3s</td>
<td>0.65s</td>
</tr>
<tr>
<td>german-ish4</td>
<td>3.5s</td>
<td>2.5s</td>
<td>0.75s</td>
</tr>
<tr>
<td>german-ish5</td>
<td>1.9s</td>
<td>8.2s</td>
<td>0.60s</td>
</tr>
<tr>
<td>german</td>
<td>18s</td>
<td>5.8s</td>
<td>4.25s</td>
</tr>
<tr>
<td>szymanski_at</td>
<td>TO</td>
<td>13s</td>
<td>2.60s</td>
</tr>
<tr>
<td>szymanski_na</td>
<td>TO</td>
<td>TO</td>
<td>16s</td>
</tr>
</tbody>
</table>

As we can see in this table, this algorithm is competitive and even better when good refinements can be found.

7. Related Works

There has been a lot of research in software model checking and Property-Driven Reachability. This type of algorithm was first introduced by Bradley in [2] and McMillan revisited his Lazy Annotation (which shares similarities with PDR algorithms) in [9] or the recent approach from Cimatti et al. [10] and Z3 with a PDR approach in [11] and [12]. Even though some of these tools are supposed to work on parameterized systems, we were either not able to find them or they were not able to prove our examples.

8. Conclusion

We presented the problem of parameterized protocol verification and gave an algorithm to automatically do it. This new algorithm was implemented in Cubicle and successfully applied to many cache coherence protocols.

This algorithm could be improved with a better generalisation engine (allowing to explore less vertices), an incremental approach (the parameterized aspect of our language makes it hard to remember the state of our SMT solver). Other optimizations could involve a novel way of refining our formulas (it is clear that the best refinements are inductive invariants but it is still an open problem as how to find these).

Some optimizations were not documented in this article such as

- \textit{Set-theoretic test} : some formulas are trivially unsatisfiable and don’t require call to the SMT solver;
- \textit{relevant instantiations} : handling universally quantified formulas can lead to multiple useless instantiations that are trivially unsatisfiable or valid and do not help the SMT solver to solve the whole formula. This optimization allows to gain a significant time in the SMT solver.
- \textit{selecting good bads} : handling bad parts from the ones with less processes involved allows to control the number of processes that have to be instantiated when checking the satisfiability of formulas. It is mandatory, if we want to have a competitive algorithm, that we handle the bad parts cleverly (this can be done in the priority queue).

References