Termination Analysis for Evolving Programs
An Incremental Approach by Reusing Certified Modules

FEI HE, Tsinghua University, China, Key Laboratory for Information System Security, MoE, China, and Beijing National Research Center for Information Science and Technology, China
JITAO HAN, Tsinghua University, China, Key Laboratory for Information System Security, MoE, China, and Beijing National Research Center for Information Science and Technology, China

Research on program termination has a long tradition. However, most of the existing techniques target a single program only. We propose in this paper an incremental termination analysis approach by reusing certified modules across different program versions. A transformation-based procedure is further developed to increase the reusability of certified modules. The proposed approach has wide applicability, applicable to various program changes. The proposed technique, to the best of our knowledge, represents a novel attempt to the termination analysis of evolving programs. We implemented the approach on top of Ultimate Automizer. Experimental results show dramatic improvement of our approach over the state-of-the-art tool.

CCS Concepts:
- Software and its engineering → Correctness; Software verification.

Additional Key Words and Phrases: Termination analysis, incremental analysis, Büchi automaton, ranking function

ACM Reference Format:

1 INTRODUCTION

Termination is a fundamental program property. There is a wide variety of work on proving program termination from different angles [Brockschmidt et al. 2013; Cook et al. 2006a,b, 2013; Gulwani et al. 2008; Harris et al. 2010; Heizmann et al. 2014; Kroening et al. 2008, 2010; Larraz et al. 2013; Podelski and Rybalchenko 2004; Podelski and Rybalchenko 2005, 2011; Ströder et al. 2017]. However, most of these techniques target a single program version only. Note that programs evolve throughout their life cycles. How to efficiently prove termination of evolving programs is the main problem addressed in this paper.

Incremental analysis is a methodology that attempts to reuse the intermediate results, computed in the previous round of analysis, in the current analysis. The incremental analysis has been successfully applied to model checking [Beyer et al. 2012; Conway et al. 2005; He et al. 2016; Henzinger et al. 2003; Lauterburg et al. 2008; Yang et al. 2009], program verification [Beyer et al. 2013; Fedyukovich et al. 2013; Rothenberg et al. 2018; Sery et al. 2012; Yang et al. 2014], etc. However,
to the best of our knowledge, we are not aware of any incremental technique for termination analysis of evolving programs.

Certified module, proposed by Heizmann, Hoenicke and Podelski [Heizmann et al. 2014], is a special case of Büchi automaton. A certified module guarantees that: 1) all accepted $\omega$-traces are terminating, and 2) they share the same termination argument. To prove the termination of a program, it suffices to find a set of certified modules such that all $\omega$-traces of the program are covered by these certified modules. The ULTIMATE AUTOMIZER tool that implements this approach won the 1st place in the Termination category of SV-COMP from 2017 – 2019.

We found that certified modules are reusable. Construction of a certified module is not guided by the program structure (see Section 2). Instead, a certified module may recognize traces that do not belong to the program, and can also reject traces of the program (i.e., left to other certified modules to cover). This freedom in representing the $\omega$-language gives them more possibilities to be reused across different program versions. Recall that a certified module represents a set of terminating $\omega$-traces. These traces can be subtracted from the program without proving their termination again. The whole verification efforts can thus be considerably reduced.

We propose an incremental termination analysis approach by reusing certified modules. Each certified module is decomposed into a number of Hoare triples. A Hoare triple $\{\phi\} \ b \ {\psi}$ is reusable as long as the block $b$ does not change. The certified module can guarantee that all decomposed Hoare triples are valid. So these triples can be reused without proving their validity again. Moreover, the precondition $\phi$ and the postcondition $\psi$ convey information about the ranking function. This information is also reusable when we consider to add new Hoare triples. In this way, we realize as much as possible reusing of certified modules.

Regarding the program as an automaton over the alphabet of blocks (or statements), the program changes can be categorized into two types: the block changes, that add or delete blocks to the program, or the control-flow changes, that add or delete control-flow edges to the program. We show that our approach is applicable to both types of changes. As long as the two program versions share some blocks (or statements), our incremental approach is beneficial.

We implemented our approach on top of ULTIMATE AUTOMIZER. We evaluated our approach on two sets of programs, where the first contains 90 manually-crafted program revisions, and the second contains 611 real-world program revisions. Compared to the state-of-the-art ULTIMATE AUTOMIZER, our approach is x4.16 faster for manually-crafted revisions, and x3.08 faster for real-world revisions.

The main technical contributions of this paper are summarized as follows:

- We proposed, to the best of our knowledge, the first approach for termination analysis of evolving programs.
- We developed a transformation-based procedure for increasing the reusability of certified modules. We showed that our approach are applicable to various program changes.
- We implemented the proposed approach. Experimental results showed dramatic improvement of our approach.

The remainder of this paper is organized as follows. Section 2 introduces the necessary preliminaries. Section 3 motivates our approach using a simple example. Section 2.5 reviews the existing termination analysis approach for a single program version. Section 4 presents our incremental termination analysis approach for evolving programs. Section 5 discusses the applicability of our approach. Section 6 reports evaluation results on our approach. Section 7 discusses related work and Section 8 concludes this paper.
2 PRELIMINARIES

2.1 Program, Alphabet and Trace

Let \( P \) be a program. A basic block (or shortly, block) \( b \) of \( P \) is a straight-line sequence of program statements in \( P \) without any program jumps. Let \( \Sigma \) be the set of blocks in \( P \). A program is a control-flow automaton \( P = (\text{Loc}, \delta, l_0) \), where \( \text{Loc} \) is the set of program locations, \( \delta \subseteq \text{Loc} \times \Sigma \times \text{Loc} \) is the set of control-flow edges labeled with blocks, and \( l_0 \) is the initial location.

We may treat \( \Sigma \) as the alphabet and each \( b \in \Sigma \) as a letter [Heizmann et al. 2013a]. Then a finite trace over \( \Sigma \) is a finite sequence of letters in \( \Sigma \), and an \( \omega \)-trace over \( \Sigma \) is an infinite sequence of letters in \( \Sigma \). Note that \( \Sigma \) is a finite set, the \( \omega \)-trace \( \sigma \) over \( \Sigma \) is always in the lasso shape, i.e., \( \sigma = b_0 \ldots b_{k-1} (b_k \ldots b_n)^\omega \), where the prefix \( b_0 \ldots b_{k-1} \) is called the stem of the lasso, and the periodic part \( b_k \ldots b_n \) is called the loop of the lasso. Denote \( \Sigma^* \) the set of all traces over \( \Sigma \), and \( \Sigma^\omega \) the set of all \( \omega \)-traces over \( \Sigma \).

A program is a special case of a Büchi automaton \( (\Sigma, \text{Loc}, \delta, l_0, \text{Loc}) \), where the alphabet is the set of program blocks, the set of nodes is the set of the program locations, and the set of accepting nodes contains all program locations. A path \( \pi \) of \( P \) is an alternating sequence of locations and blocks, i.e.,

\[
\pi = l_0 b_0 \rightarrow l_1 b_1 \rightarrow \ldots,
\]

such that \( (l_i, b_i, l_{i+1}) \in \delta \) for each \( i \geq 0 \). A trace (or an \( \omega \)-trace) over \( \Sigma \) is called a trace (or an \( \omega \)-trace) of \( P \) if it labels a path of \( P \). Denote \( \mathcal{L}(P) \) (or \( \mathcal{L}^\omega(P) \)) the set of all traces (or \( \omega \)-traces) recognized by \( P \). Obviously, \( \mathcal{L}(P) \subseteq \Sigma^* \) and \( \mathcal{L}^\omega(P) \subseteq \Sigma^\omega \).

2.2 Infeasible Trace

In the following, we use Hoare triple to define the feasibility of traces.

Let \( \varphi \) and \( \psi \) be two first-order predicates, and \( b \) be a program block. The Hoare-triple \( \{ \varphi \} b \{ \psi \} \) is valid iff \( \varphi \Rightarrow \wp(b, \psi) \), where \( \wp \) stands for the weakest precondition transformer [Bradley and Manna 2007]. Moreover, let \( b_0 b_1 \ldots b_n \) be a finite sequence of blocks, the Hoare triple \( \{ \varphi \} b_0 b_1 \ldots b_n \{ \psi \} \) is valid iff \( \varphi \Rightarrow \wp(b_0, \wp(b_1, \ldots, \wp(b_n, \psi))) \).

Let \( V \) be the set of variables in the program. A state of \( P \) is a pair \((l, s)\), where \( l \) is a program location, and \( s \) is a valuation to \( V \). Denote \( \phi_s \) the formula representation of \( s \). For instance, the valuation \( \{x_0 \mapsto 0, x_1 \mapsto 1\} \) can also be represented as \( x_0 = 0 \land x_1 = 1 \). Let \((l, s)\) and \((l', s')\) be two states of the program, we say \((l', s')\) is reachable from \((l, s)\), iff there is a trace \( \sigma \) such that \( l \xrightarrow{\sigma} l' \), and the Hoare triple \( \{ \phi_s \} \sigma \{ \phi_{s'} \} \) is valid.

A finite trace \( \tau \) of \( P \) is infeasible if \{true\} \( \tau \{false\} \) is valid, i.e., there exists no possible execution of \( P \) for \( \tau \). An \( \omega \)-trace \( \tau \) is infeasible if either: 1) any of its finite prefix is infeasible, or 2) there exists no initial valuation to enable any infinite execution of \( \tau \).

2.3 Module

Termination analysis needs only to consider \( \omega \)-traces of the program. An \( \omega \)-trace is terminating if and only if it is infeasible. In the following, we introduce module [Heizmann et al. 2014] as a representation of a set of \( \omega \)-traces.

**Definition 1.** A module is a 5-tuple \( M = (\Sigma, Q, \delta, q_0, q_\ast) \), where \( \Sigma \) is the alphabet, \( Q \) is a finite set of nodes, \( \delta \subseteq Q \times \Sigma \times Q \) is a set of edges labeled with letters in \( \Sigma \), \( q_0 \in Q \) is the initial node, \( q_\ast \in Q \) is the accepting node, and \( Q \) can be partitioned into two sets \( Q^0 \) and \( Q^1 \), such that \( q_0 \in Q^0 \), \( q_\ast \in Q^1 \) and no node in \( Q^1 \) has a successor in \( Q^0 \).

Note that a node in a module is not necessarily a program location, and an edge of a module may not correspond to any control-flow edge of the program.
A module is a special case of a Büchi automaton where the set of states is the set of nodes and the set of accepting states contains \( q_\bullet \) only. An infinite path of \( M \) is fair (also said, accepted by \( M \)) if it visits the accepting node \( q_\bullet \) infinitely often. An \( \omega \)-trace of \( M \) is fair (also said, accepted by \( M \)) if it labels a fair path of \( M \). Denote \( L^\omega(M) \) the set of all fair \( \omega \)-traces accepted by \( M \). A module \( M \) is terminating if all traces in \( L^\omega(M) \) are terminating.

### 2.4 Ranking Function and Certified Module

Let \( f \) be a function over the program variables \( V \) that returns an element in a well-ordered set \( (S, \prec) \). The function \( f \) is called a ranking function for \( M \) [Heizmann et al. 2014] if the values of \( f(V) \) decrease every time the accepting node is visited. Note that this definition is a little different from the ranking function of a program, where the values of \( f(V) \) are required to decrease on every transition of the program.

In the following, we use an auxiliary variable \( \text{oldrnk} \) to refer to the value of \( f(V) \) at the previous visit of the accepting node \( q_\bullet \), and \( \infty \) to represent a big value that is greater than all other values from the well-ordered \( S \). The value of \( \text{oldrnk} \) needs to be updated every time \( q_\bullet \) is visited. To certify that \( f(V) \) is indeed a ranking function, we need to guarantee \( f(V) < \text{oldrnk} \) and \( \text{oldrnk} \geq 0 \) at each visit of \( q_\bullet \).

**Definition 2.** Given a module \( M = (\Sigma, Q, \delta, q_\circ, q_\bullet) \) and a function \( f(V) \) that returns an element in a well-ordered set \( (S, \prec) \), a rank certificate \( I \) for \( f \) and \( M \) is an annotation that maps each node of \( M \) to a predicate, such that

- the initial node \( q_\circ \) is mapped to a predicate where \( \text{oldrnk} \) has the value \( \infty \), i.e.,
  \[ I(q_\circ) \iff \text{oldrnk} = \infty; \]
- the accepting node \( q_\bullet \) is mapped to a predicate stating that the value of \( f(V) \) decreases since the last visit of \( q_\bullet \), i.e.,
  \[ I(q_\bullet) \implies (f(V) < \text{oldrnk} \land \text{oldrnk} \geq 0); \]
- for any edge \((q, b, q') \in \delta \) where \( q \neq q_\bullet \),
  \[ \{I(q)\} b \{I(q')\} \text{ is valid; and} \]
- for any edge \((q_\bullet, b, q') \in \delta \),
  \[ \{I(q_\bullet)\} \text{ oldrnk} := f(V); b \{I(q')\} \text{ is valid.} \]

The triple \((M, f, I)\) is called a certified module.

Throughout the paper, we often write a certified module as \( M \) when \( f \) and \( I \) are clear from the context.

**Lemma 1.** [Heizmann et al. 2014] A certified module is always terminating.

A certified module can be used as a representation of a set of fair \( \omega \)-traces that share the same reason (i.e., the ranking function and the rank certificate) for termination.

### 2.5 Termination Analysis Using Certified Modules

**Theorem 1.** [Heizmann et al. 2014] A program \( P \) is terminating if there exists a set of terminating modules \( M_0, M_1, \ldots, M_n \), such that

\[
L^\omega(P) \subseteq L^\omega(M_0) \cup L^\omega(M_1) \cup \cdots \cup L^\omega(M_n).
\]
Heizmann et al. [2014] proposed a decomposition-based approach for proving termination of a single program version. Figure 1 shows an overview of this approach. The basic idea is to iteratively construct a set of certified modules $M_0, M_1, \ldots, M_n$, such that $L^\omega(P) \subseteq \bigcup_{i=0}^{n} L^\omega(M_i)$.

Each iteration of this approach consists of two phases: the language inclusion checking and the certified module construction. During the language inclusion checking, we check if $L^\omega(P) \subseteq \bigcup_{i=0}^{n} L^\omega(M_i)$ or not. If the check succeeds, by Theorem 1, we immediately conclude that $P$ is terminating. Otherwise, the program termination can not be proved with the current set of certified modules, and the checker returns a counterexample $\sigma \in L^\omega(P) \setminus \bigcup_{i=0}^{n} L^\omega(M_i)$.

During the certified module construction phase, the $\omega$-trace $\sigma$ returned by the former phase is semantically analyzed to decide if it is terminating or not. Termination analysis for a single $\omega$-trace (or equivalently, for a lasso program that contains a single $\omega$-trace) has long been studied. Many specialized methods was developed for lasso programs [Ben-Amram and Genaim 2013, 2014, 2015, 2017; Bradley et al. 2005a; Cook et al. 2010; Heizmann et al. 2013b; Kroening et al. 2008; Podelski and Rybalchenko 2004]. Even so, termination of this kind of simple programs is undecidable [Ben-Amram and Genaim 2014]. If the termination of $\sigma$ cannot be decided, we report “unknown”.

Otherwise, with the termination argument (usually, a ranking function) generated by the existing methods, we proceed to construct a certified module.
3 A MOTIVATING EXAMPLE

Figure 2 shows a simple program loop$_0$ that contains a double nested while loop. Blocks in this program are: \( b_0 = \text{assume } x>0; y=1, b_1 = \text{assume } y<x; y=2*y, \) and \( b_3 = \text{assume } y=x; x=x-1. \)

3.1 Termination Analysis of loop$_0$

To prove the termination of loop$_0$, we need to show that all \( \omega \)-traces of loop$_0$ are terminating.

Let us first take an \( \omega \)-trace of loop$_0$:

\[
b_0b_1(b_2)^\omega
\]

that enters the outer while loop and then visits the inner while loop infinitely often. Termination of this trace can be certified by the module \( M_0 \) in Figure 3a together with the linear ranking function \( f(x, y) = x - y \) and the following rank certificate:

\[
I(q_1) \iff \text{oldrnk} = \infty, \quad \text{and}
I(q_2) \iff (\text{oldrnk} > x - y \land \text{oldrnk} \geq 0 \land y > 0).
\]

Note that the module \( M_0 \) accepts not only the input trace but also all traces that eventually always take the inner loop, i.e., \( (\Sigma_0)^\omega b_1(b_2)^\omega \), where \( \Sigma_0 = \{ b_0, b_1, b_2, b_3 \} \). The rank certificate for \( M_0 \) further guarantees that all \( \omega \)-traces accepted by \( M_0 \) are terminating.

Then, we check if \( L^\omega(\text{loop}_0) \subseteq L^\omega(M_0) \) or not. This check returns a counterexample. Let us take the following counterexample trace:

\[
b_0(b_1b_3)^\omega
\]

that visits the outer while loop infinitely often. The termination of this trace can be certified by the module \( M_1 \) in Figure 3b together with the ranking function \( f(x) = 2x + 1 \) and the following rank certificate:

\[
I(q_0) \iff \text{oldrnk} = \infty,
I(q_1) \iff (\text{oldrnk} > 2x + 1 \land \text{oldrnk} \geq 0), \quad \text{and}
I(q_2) \iff (\text{oldrnk} \geq 2x + 1 \land x \geq 0).
\]

Again, all \( \omega \)-traces accepted by \( M_1 \) are terminating.

Finally, we check whether \( L^\omega(\text{loop}_0) \subseteq L^\omega(M_0) \cup L^\omega(M_1) \). We see that this check succeeds, i.e., all \( \omega \)-traces of loop$_0$ are covered by \( M_0 \) and \( M_1 \). We thus conclude that loop$_0$ is terminating (by Theorem 1).

3.2 Termination Analysis of loop$_1$

Now, let us assume that the program loop$_0$ evolves into a new version loop$_1$, where the statement \( y=2*y \) (at line 6) is replaced by the statement \( y=y+1 \). In terms of alphabet, loop$_1$ differs loop$_0$ only in block \( b_2 \). Denote \( b_2' = \text{assume } y<x; y=y+1 \) as the altered version of \( b_2 \). Note that this change has side effects to the ranking function \( f(x, y) = x - y \) and also the inner loop condition.

Observe that the certified modules \( M_0 \) and \( M_1 \) are intermediate results generated in the previous analysis, conveying important information about the termination of loop$_0$. Note that these two modules are defined over the alphabet \( \Sigma_0 \), but not \( \Sigma_1 \). They cannot directly be applied to the termination analysis of loop$_1$. However, if we look deep into the nodes and the edges of these two modules, reusable information can be identified.

---

1Before that, we will try to apply the non-termination analysis [Leike and Heizmann 2018] to get a conclusive “non-terminating” result.
Let us consider the edge \((q_1, b_1, q_2)\) of \(M_0\). According to the definition of a certified module (Definition 2), the Hoare triple \(\{I(q_1)\} b_1 \{I(q_2)\}\) is valid, where

\[
I(q_1) \iff \text{oldrnk} = \infty \\
I(q_2) \iff (\text{oldrnk} > x - y \land \text{oldrnk} \geq 0 \land y > 0)
\]

are two predicates mapped by the rank certificate \(I\) to \(q_1\) and \(q_2\), respectively.

The above Hoare triple is reusable, because \(b_1\) is not changed in \(\text{loop}_0\). With this valid Hoare triple, we know that if \(b_1\) is executed from a state satisfying \(I(q_1)\), its final state must satisfy \(I(q_2)\). The validity of this Hoare triple just tells a fact about the semantics of \(b_1\), which always hold as long as \(b_1\) does not change. By reusing this Hoare triple, we need not to check its validity again. Given that the validity checking of Hoare triples needs to invoke an SMT solver, and in many cases is quite time-consuming. Reusing this information significantly reduces redundant computation.

Moreover, let us look at the postcondition \(I(q_2)\) of this Hoare triple. The ranking function \(f(x, y) = x - y\) is embraced in this predicate. In other words, reusing this Hoare triple can save us the efforts for re-computing the ranking function (if this ranking function still takes effect). Recall that the synthesis of ranking function is among the most time-consuming operations for termination analysis, reusing this information can save a great deal of effort.

## 4 Incremental Termination Analysis

This section proposes our incremental termination analysis approach. The basic idea is to reuse the previously-computed information in the current analysis.

In the sequel of this section, we first present the framework of our approach, then explain our Reuse technique, and our on-demand construction technique, and finally prove the correctness of our approach.

### 4.1 The Whole Procedure

Let \(P^-\) and \(P\) be two versions of a program, and \(\Sigma^-\) and \(\Sigma\) be their alphabets, respectively. Assume the termination of \(P^-\) has already been analyzed, generating a set of certified modules, say \(M^-_0, M^-_1, \ldots, M^-_m\). Figure 4 shows an overview of our incremental termination analysis using certified modules. Before the new round of analysis starts, an initialization procedure is performed, which attempts to reuse the information stored in the previously-constructed certified modules.
The reuse procedure is applied to each of the input certified modules, and outputs a reuse module. A reuse module is a certified module over Σ, and can be used for proving termination of \( P \). Due to the diversity of program changes, the constructed reuse module may recognize an empty \( \omega \)-language. For this case, this module covers no \( \omega \)-traces of \( P \), and is directly abandoned. Therefore, the number \( k \) of constructed modules is less than or equal to the number \( m \) of the input certified modules, i.e., \( k \leq m \).

In each iteration, the language inclusion checking

\[
L^\omega(P) \subseteq \bigcup_{i=0}^{n} L^\omega(M_i)
\]

is equivalent to check if \( P \cap M_0 \cap \cdots \cap M_n \) recognizes an empty language or not. The latter check can be implemented in an incremental way, i.e., \( P(0) = P, P(1) = P(0) \cap M_0, \ldots, P(n+1) = P(n) \cap M_n \). Each automaton \( P(l) \) \((0 \leq l \leq n + 1)\) is called a remainder automaton. We iteratively check if \( P(l) \) recognizes an empty \( \omega \)-language or not.

### 4.2 The Reuse Procedure

Given a certified module \( M^- \) of the previous program version, we construct a new module \( M \), called the reuse module, in the following steps.

**Step 1. Extract predicates and Hoare triples.** Recall that the previously-constructed certified module \( M^- \) and the current program version \( P \) are over different alphabets. The \( M^- \) may not be entirely reusable for the current analysis. However, we may decompose the certified module into a set of Hoare triples, many of which are reusable for the current analysis.

According to the certified module definition (Definition 2), on each edge of \( M^- \), there is a valid Hoare triple. Particularly, for any edge \((q, b, q')\) of \( M^- \), where \( q \neq q^- \), we get a Hoare triple

\[
\{I(q)\} b \{I(q')\};
\]

and for any edge \((q^-, b, q')\), we get a Hoare triple

\[
\{I(q^-)\} \text{oldrank:=f(V); } b \{I(q')\}.
\]

Denote \( \mathcal{H}_1 \) as the set of Hoare triples obtained in this step.

---

Fig. 4. Incremental termination analysis using certified modules

---

We also collect the set of predicates used in the rank certificate of $M^-$, i.e.

$$U_1 = \bigcup_{\forall q \in Q^-} \{I(q)\}.$$  

Especially, we use $u_o = I(q_o^-)$ and $u_* = I(q_*^-)$ to mark the predicates that are mapped to the initial and accepting nodes of $M^-$, respectively. Note that $U_1$ can also be considered as the set of predicates (precondition or postcondition) of all Hoare triples in $H_1$.

**Example 1.** Consider the certified module $M_0$ in Figure 3a, the predicates mapped to $q_1$ and $q_2$ by the rank certificate are respectively

$$\varphi_0 \iff \text{oldrnk} = \infty, \quad \text{and}$$

$$\varphi_1 \iff (\text{oldrnk} > x - y \land \text{oldrnk} \geq 0 \land y > 0).$$

The set of predicates of $M_0$ is thus $\{\varphi_0, \varphi_1\}$. The Hoare triple for edges $(q_1, b_1, q_2)$ and $(q_2, b_2, q_2)$ are respectively

$$\{\varphi_0\} b_1 \{\varphi_1\}, \quad \text{and}$$

$$\{\varphi_1\} \odlr = f(V); b_2 \{\varphi_1\}.$$  

The set of all Hoare triples of $M_0$ are listed in Figure 5a.

**Step 2. Delete inapplicable predicates and Hoare triples.** The extracted predicates and Hoare triples record information about the previous program version. They may not all be applicable to the current analysis.

Let $V$ be the set of program variables in $P$. For any predicate $\varphi \in U$, define $FV(\varphi)$ as its set of free variables. Basically, $\varphi$ can only use program variables in $V$ and a special variable oldrnk as its free variables. The predicate $\varphi$ is inapplicable to $P$ if

$$FV(\varphi) \setminus (V \cup \{\text{oldrnk}\}) \neq \emptyset.$$  

A Hoare triple $\{\varphi\} b \{\varphi'\}$ is inapplicable to $P$ if either (1) $\varphi$ or $\varphi'$ is inapplicable to $P$, or (2) $b$ does not exist in $P$.

The inapplicable predicates and Hoare triples need be removed from $U$ and $H$, respectively. Note that $u_o$ and $u_*$ are necessary for constructing the reuse module. If $u_*$ is deleted from $U$ in this step, we stop the reuse procedure for $M^-$, and proceed to process the next certified module. Note that $u_o$ will never be deleted, since it always represents oldrnk = $\infty$, containing a single free variable oldrnk. Denote the resulting sets of predicates and Hoare triples after elimination as $U_2$, $H_2$, respectively.

**Example 2.** Consider the predicate set $U$ obtained in Example 1, except of oldrnk, other two variables $x$ and $y$ are declared in loop$_1$, thus no predicate needs to be eliminated from $U$. Consider the Hoare triple set $H_1$ obtained in Example 1 (in Figure 5a). Note that the only changed block in loop$_1$ against loop$_0$ is $b_2$. Therefore, the third and sixth Hoare triples in $H_1$ are inapplicable to loop$_1$ and need to be removed from this set.

**Step 3. Add new Hoare triples.** The new program version may contain new blocks, which should be tested for addition of new Hoare triples.

Let $b \in \Sigma \setminus \Sigma^-$ be a new block. In this step, we limit the precondition and postcondition of the candidate Hoare triple to among the predicates in $U_2$. For any two predicates $\varphi, \varphi' \in U_2$, if $\varphi \neq u_*$, we test the validity of $\{\varphi\} b \{\varphi'\}$; if $\varphi = u_*$, we test the validity of $\{\varphi\} \odlr = f(V); b \{\varphi'\}$; we then add the passed Hoare triples to $U_2$. In this way, the previously-computed rank certificate are adapted to the new program version. Let $H_3$ be the triple set after addition of new Hoare triples.
EXAMPLE 3. Consider the Hoare triple set $\mathcal{H}_2$ obtained in Example 2. Note that $b'_2$ is the only new block in loop $p_1$ compared to loop $p_0$. After validity testing, two new Hoare triples (colored red in Figure 5b) are added into this triple set.

**Step 4. Construct the reuse module.** Given $\mathcal{U}_2$ and $\mathcal{H}_3$, we now construct a module $M$, called a reuse module, and a mapping $I$, in the following steps:

- create a node $q$ for each predicate $\varphi$ in $\mathcal{U}_2$, and set $I(q) = \varphi$;
- add an edge $(q,b,q')$, if $q \neq q_*$ and there is a Hoare triple
  \[ \{I(q)\} b \{I(q')\} \in \mathcal{H}; \]
- add an edge $(q,b,q')$, if $q = q_*$ and there is a Hoare triple
  \[ \{I(q)\} \text{oldrnk} := f(V); b \{I(q')\} \in \mathcal{H}; \] and
- the node $q$ is made the initial node (or the accepting node) if $I(q) = u_0$ (or $I(q) = u_*$).

Recall that the extraction procedure in Step 1 extracts a set of valid Hoare triples from a certified module. The above construction procedure works in the reverse direction, i.e., constructs a certified module from a set of valid Hoare triples.

**Example 4.** Consider the predicate set $\mathcal{U}_2 = \{\varphi_0, \varphi_1\}$ and the Hoare triple set $\mathcal{H}_3$ in Figure 5b. According to the above module construction rules, two nodes are created for the reuse module, named $q_0$ and $q_1$, and mapped to $\varphi_0$ and $\varphi_1$, respectively. Additionally, six edges are connected in the reuse module, each standing for one Hoare triple in Figure 5b. The reuse module after construction is shown in Figure 5c.

**Theorem 2.** The reuse module constructed in the reuse procedure is terminating.

**Proof.** Let $M = (Q, \delta, q_0, q_*)$ be a reuse module constructed from a certified module $M^- = (Q^-, \delta^-, q_0^-, q_*^*)$ by the reuse procedure. By Step 1 and Step 4, $I(q_0) = I(q_0^-)$ and $I(q_*) = I(q_*^-)$. Moreover, $M^-$ is a certified module, the first two conditions of Definition 2 are thus satisfied. For any edge $(q,b,q') \in \delta$, it is either inherited from $M^-$, or added by Step 3 of the reuse procedure. For the former case, the validity of the corresponding Hoare triple is ensured by $M^-$ (a certified module); for the latter case, the processing of Step 3 guarantees the validity of the corresponding Hoare triple. Therefore, the last two conditions in Definition 2 are also satisfied. Therefore, $M$ is a certified module (by Definition 2), and $M$ is terminating (by Lemma 1). \qed

### 4.3 On-Demand Construction

The reuse procedure indeed enforces an eager construction of the reuse module – the whole module is constructed before it participates in the verification. In practice, the eager construction can be very expensive. Note that in Step 3 of the reuse procedure, we need to check the validity of all candidate Hoare triples. In the worst case, the number of validity checking is $O(\delta \cdot |\mathcal{U}|^2)$, where $\delta$ is the number of new blocks in $P$, and $|\mathcal{U}|$ is the number of predicates in $\mathcal{U}$. The validity checking (in Step 3) is the most computationally expensive step in the reuse procedure.

Recall that the constructed certified modules are devoted to cover $\omega$-traces of $P$. Not all edges of the reuse module are related to the automaton of $P$. Validation checking of Hoare triples for these irrelevant edges are redundant. To avoid such redundant validation checking, we apply an on-demand approach to construct the reuse module on-the-fly. The basic idea is to defer the validity checking to the verification phase, and add the corresponding edges only on demand.

More specifically, let $P^{(l)}$ be the current remainder automaton, and $M^-$ the certified module to be processed, in thelanguage inclusion checking, we compute a new remainder automaton $P^{(l+1)}$, and
We prove the correctness of our incremental termination analysis in this subsection. Let us start with the soundness.

**Theorem 4.** If our procedure in Figure 4 returns “terminating”, the input program is terminating.
Algorithm 1: On-demand construction

Input: A remainder automaton $P^{(l)} = (Σ, Q^{(l)}, δ^{(l)}, q_0^{(l)}, q_*^{(l)})$ and a certified module $M^-$. 
Output: A new remainder automaton $P^{(l+1)} = (Σ, Q^{(l+1)}, δ^{(l+1)}, q_0^{(l+1)}, q_*^{(l+1)})$ and a reuse module $M = (Σ, Q, δ, q_0, q_*).$

/* Certified module transformation */
1. $(U, H) ← Extract(M^-);$ /* step 1: extract Hoare triples */
2. $(U, H) ← Filter(U, H);$ /* step 2: delete inapplicable triples */
3. $(U, H) ← Construct(U, H);$ /* step 4: construct reuse module */
/* Language inclusion checking */
4. Initialize $P^{(l+1)}$ to be an empty automaton;
5. $Q^{(l+1)} ← \{(q_0^{(l)}, q_0)\}$, waitlist $← \{(q_0^{(l)}, q_0)\};$
6. while waitlist $≠ \emptyset$ do
7. Pop $(q_1^{(l)}, q_1)$ from waitlist;
8. foreach $(q_1^{(l)}, b, q_2^{(l)}) ∈ δ^{(l)}$ do /* following an edge of $P^{(l)} */
9. foreach $q_2 ∈ Q$ do /* edge with same label */
10. if $(q_1, b, q_2) ≠ δ$ then
11. if Triple$(I(q_1), b, I(q_2))$ is valid then
12. $δ ← δ ∪ \{(q_1, b, q_2)\};$ /* add the edge to $M */
13. else
14. Continue;
15. end
16. end
17. if $(q_2^{(l)}, q_2) ∉ Q^{(l+1)}$ then /* a new state of $P^{(l+1)} */
18. $Q^{(l+1)} ← Q^{(l+1)} ∪ \{(q_2^{(l)}, q_2)\};$
19. waitlist $←$ waitlist $∪ \{(q_2^{(l)}, q_2)\};$
20. end
21. $δ^{(l+1)} ← δ^{(l+1)} ∪ \{(q_1^{(l)}, q_1), b, (q_2^{(l)}, q_2)\};$
22. end
23. end
24. return $(P^{(l+1)}, M);$ 

Proof. This procedure returns “terminating” only when the language inclusion checking succeeds. In other words, there exists a set of modules $M_0, M_1, \ldots, M_n$, each of which is either constructed from an $ω$-trace, or transformed from a previously-constructed certified module, such that $L^ω(P) ⊆ L^ω(M_0) \cup L^ω(M_1) \cup \cdots \cup L^ω(M_n)$. The modules constructed from $ω$-traces [Heizmann et al. 2014] are terminating. By Theorem 2 and Theorem 3, the reuse modules are terminating. Thus all modules are terminating. According to Theorem 1, the program is terminating. □

The completeness of our approach cannot be directly guaranteed, since our approach requires to synthesize a ranking function for any lasso program, which in some cases maybe impossible [Ben-Amram and Genaim 2014]. Note that a terminating program can always be decomposed into a
finite set of certified modules $M_0, M_1, \ldots, M_n$ [Heizmann et al. 2014], such that

$$L^\omega(P) \subseteq L^\omega(M_0) \cup L^\omega(M_1) \cup \cdots \cup L^\omega(M_n).$$

Then, under the assumption of an oracle that synthesizes a ranking function for any lasso program, we prove the relative completeness of our approach – if the input program is terminating, the procedure in Figure 4 always deduces this result and returns "terminating".

5 APPLICABILITY TO VARIOUS PROGRAM CHANGES

This section discusses the applicability of our approach to various kinds of program changes.

A program is a control-flow automaton over the alphabet of blocks. In this respect, the program changes can be categorized into two types:

- **Type I (Block changes):** the changes that add or delete basic blocks to the program.
- **Type II (Structure changes):** the changes that add or delete control-flow edges to the program.

5.1 Type I Changes

This type of changes may alter the alphabet of the program. Let $\Sigma^-$ and $\Sigma$ be the alphabet before and after the program changes, respectively.

*Type I* changes can be efficiently handled by our approach. For any block $b \in \Sigma \cap \Sigma^-$, all previously-established Hoare triples on $b$ are reusable. For any new block $b' \in \Sigma \setminus \Sigma^-$, we process as the Step 3 in Section 4 to prove the validity of new Hoare triples on $b'$. Finally, a certified module is constructed from the set of valid Hoare triples.

**Example 5.** The change of loop$_1$ over loop$_0$ (in Figure 2) is a Type I change. It modifies the value of $y$, and has side effects to the inner loop condition. As discussed before (see Section 4), our incremental approach can efficiently handle this change.

5.2 Type II Changes

This type of changes does not alter the program alphabet. As a result, all previously-constructed certified modules can be directly reused in the current analysis. There is even no need to apply the Reuse procedure. The certified module $M^-$ can be directly used as the reuse module. The only consequence of this type of changes is that some traces of $L^\omega(M^-)$ may no longer belong to $L^\omega(P)$.

**Example 6.** The following two versions of a program show a Type II change.

1. $x = \text{nondet\_int}();$
2. $y = \text{nondet\_int}();$
3. while ($y < x$) {
4. $y = 2 \times y$;
5. }
6. $x = x - 1;$

1. $x = \text{nondet\_int}();$
2. $y = \text{nondet\_int}();$
3. while ($y \geq x$) {
4. $x = x - 1;$
5. }
6. $y = 2 \times y;$

All certified modules of the first program can be directly applied to the analysis of the second program.

5.3 Mixed Changes

A real-world program revision often relates to both types of program changes.

**Example 7.** The following program can also be considered as a changed version of loop$_0$ (in Figure 2), although they looks very different:
We find two common blocks between the above program and loop, i.e., \(\text{assume } x\geq 0; \ x=x-1\) and \(\ x=x\). The previously-established results on these two blocks are reusable. We still benefit from the incremental approach.

In conclusion, our incremental approach has a wide applicability. No matter what types of program changes, it is applicable, and can lead a significant saving in computational efforts.

### 5.4 More Reuse Possibilities

Throughout the paper, programs are assumed to be composed of blocks. Our approach can be naturally adapted to the statement-based program model.

If we consider the termination analysis under the statement-based program model, more information is reusable. Let \(b = s_{t_1}s_{t_2}...s_{t_l}\) be a block, where each \(s_{t_i}\) (\(1 \leq i \leq l\)) is a statement. With the block-based program model, any statement change leads the previously-established results on the whole block be abandoned. On the contrary, under the statement-based program model, only Hoare triples on the changed statements need to be abandoned, previously-established results on other unchanged statements can be kept for further analysis.

### 6 EVALUATION

We implemented our incremental termination analysis algorithm on top of Ultimate Automizer, a verification tool that implemented the decomposition-based approach [Heizmann et al. 2014]. In the following, we refer to the underlying Ultimate Automizer and our enhanced implementation as Baseline and Reuse, respectively.

We used two benchmarks to evaluate our approach. The first benchmark contains 90 artificial programs and the second benchmark contains 611 real-world programs. In our experiments, a run set consists of a series of revisions of the same program. The first revision is analyzed from scratch, and the subsequent revisions are verified with/without reuse. In the following, we call a run of termination analysis on any program revision an analysis task, and a run of termination analysis on not-first revision an regression analysis task. We compare the performance of Baseline and Reuse on regression analysis tasks.

All experiments were conducted on a machine with an Intel(R) Core(TM) i7-8700 CPU of 3.2GHz and 16GB RAM. We used Ultimate Automizer of version 0.1.24-57e1ae7 with the default configuration. We took Z3 as the default SMT solver and set 500 seconds as the time limit for all analysis tasks.

#### 6.1 Experiment on Artificial Benchmark

**Benchmark.** The first benchmark consists of 90 revisions of 15 distinct programs. The 15 original programs were obtained from [Heizmann et al. 2014]. We then asked five students (2 undergraduate students and 3 first-year graduated students, none participated in this project) to manually create more revisions for these programs. Every student was assigned three programs and was required to
craft five new revisions for each of them. They can alter the program as they want, disregarding the original meanings of the program; especially, they are encouraged to make loop-relevant changes; the only requirement is that the program after modification must be compilable.

In total, this benchmark contains 90 analysis tasks, among which 75 are regression analysis tasks.

Results. Experimental results on artificial benchmark are listed in Table 1. The first column assembly (“Run set”) lists information about the run set, including the program name (“Programs”), the number of terminating/non-terminating (“T/N”) cases in this set, the total number of lines of code (“LoC”) for all revisions in this set, and the consumed CPU time (“$T_r$”) for verifying the first revision of this set. The following two column assemblies report the experimental results of Baseline and Reuse, respectively. For each approach, we report the total consumed CPU time (“Time”) and the total number (“#Iter.”) of iterations for this run set. The “#CertModule” column assembly lists information about the certified modules, including the total number (“Avai.”) of available certified modules that were generated in the analysis of the previous revision, and the total number (“Reuse”) of (non-empty) reuse modules that were transformed from the previous certified modules. The last column (“Speedup”) shows the speedup of Reuse over Baseline, computed by Baseline.time/Reuse.time. Note that only statistics for regression analysis tasks are counted. The table is sorted by the “speedup” column. The last two rows (“Sum” and “Average”) report the total and average amounts of all rows in the table, respectively.

From Table 1, we observed that our approach outperforms baseline in all run sets, with a x4.16 average speedup. Looking at the edn.t2 run set, the maximum speedup is x9.99. Even in the worst case ud.t2, reuse outperforms Baseline by x1.95 of speedup.

The “#Iter.” columns can reveal the main reason of the efficiency of our technique. Observing these columns, we found that each run set has a dramatic reduction of iterations. On average, our approach decreases the number of iterations by 89.8%. From the “#CertModule” column assembly, we conclude that our algorithm is practical on these artificial revisions. The average proportion of the reuse modules among all previously-constructed certified modules (“Reuse” / “Avai.”) is 89.6%.

Results for a single run set. To address the detailed mechanism of our approach and to demonstrate the acceleration of our approach on three types of program changes, we present the results for bubbleSort.t2 in Table 2.

The first column (“Rev.”) indicates the number of revisions, where the first row with Rev. = 0 represents the original revision of bubbleSort.t2. The second column (“Type”) presents the type of program changes applied to this revision (see Section 5). The column (“Rst.”) presents the verification result (T for terminating and N for non-terminating). All other columns have the same meanings as in Table 1.

During the preparation of this run set, we deleted an irrelevant variable and the corresponding statements (Type I changes) from Rev. 0 to Rev. 1. Among the 8 certified modules, 7 can be reused with the iterations decreased from 7 to 2, achieving a speedup of x12.86. In Rev. 2, we modify several variables via adding some statements (Type I changes). By reusing all 8 previously-constructed certified modules successfully, reuse reduces iterations from 5 to 0 and gets a x4.42 speedup. Rev. 3 changes control-flow edges (Type II changes) by inserting goto statements into the program causing the program to be non-terminating. Baseline spends some time on the last iteration for finding a non-terminating arguments. By reusing 5 out of 8 certified modules, our approach reduces 80% of iterations on Rev. 3. Although the last iteration is a little bit time-consuming, we still get a x1.83 speedup. Rev. 4 deletes some basic blocks directly and removes two goto statements (Mixed changes). The program turns to be much simpler, and Baseline only needs 4 iterations to get the result. Only 2 certified modules are successfully reused from 8 previously-constructed certified modules. Reuse decreases 2 iterations and gets a x1.37 speedup. The last revision changes
Table 1. Experimental results on artificial benchmark

<table>
<thead>
<tr>
<th>Programs</th>
<th>Run Set</th>
<th>Baseline</th>
<th></th>
<th>Reuse</th>
<th></th>
<th>#CertModule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Programs</td>
<td>T/N</td>
<td>Loc</td>
<td>$T_{r0}$</td>
<td>Time #Iter</td>
<td>Time #Iter</td>
</tr>
<tr>
<td>edn.t2</td>
<td>5/0</td>
<td>1471</td>
<td>223.73</td>
<td>1059.02</td>
<td>666</td>
<td>106.01</td>
</tr>
<tr>
<td>firewire.t2</td>
<td>5/0</td>
<td>897</td>
<td>14.02</td>
<td>66.97</td>
<td>77</td>
<td>10.89</td>
</tr>
<tr>
<td>eric.t2</td>
<td>2/3</td>
<td>266</td>
<td>3.71</td>
<td>25.23</td>
<td>28</td>
<td>4.91</td>
</tr>
<tr>
<td>a10.c.t2</td>
<td>2/3</td>
<td>922</td>
<td>87.80</td>
<td>320.86</td>
<td>39</td>
<td>72.61</td>
</tr>
<tr>
<td>traverse_twice</td>
<td>5/0</td>
<td>7132</td>
<td>4.74</td>
<td>19.54</td>
<td>28</td>
<td>6.01</td>
</tr>
<tr>
<td>sas2.t2</td>
<td>2/3</td>
<td>1364</td>
<td>17.02</td>
<td>53.55</td>
<td>42</td>
<td>17.44</td>
</tr>
<tr>
<td>reverse.t2</td>
<td>4/1</td>
<td>9683</td>
<td>15.71</td>
<td>54.62</td>
<td>21</td>
<td>19.92</td>
</tr>
<tr>
<td>b20.t2</td>
<td>5/0</td>
<td>772</td>
<td>3.00</td>
<td>13.60</td>
<td>31</td>
<td>4.98</td>
</tr>
<tr>
<td>consts1.t2.c</td>
<td>4/1</td>
<td>204</td>
<td>1.31</td>
<td>6.54</td>
<td>18</td>
<td>2.40</td>
</tr>
<tr>
<td>s3-work.t2</td>
<td>4/1</td>
<td>23137</td>
<td>11.90</td>
<td>68.71</td>
<td>46</td>
<td>27.20</td>
</tr>
<tr>
<td>sumit.t2</td>
<td>3/2</td>
<td>420</td>
<td>4.64</td>
<td>15.99</td>
<td>29</td>
<td>6.86</td>
</tr>
<tr>
<td>spiral.t2</td>
<td>5/0</td>
<td>333</td>
<td>4.70</td>
<td>26.09</td>
<td>81</td>
<td>11.60</td>
</tr>
<tr>
<td>bubbleSort.t2</td>
<td>4/1</td>
<td>547</td>
<td>9.78</td>
<td>22.93</td>
<td>28</td>
<td>10.66</td>
</tr>
<tr>
<td>mc91.t2</td>
<td>4/1</td>
<td>237</td>
<td>4.49</td>
<td>23.54</td>
<td>53</td>
<td>11.50</td>
</tr>
<tr>
<td>ud.t2</td>
<td>5/0</td>
<td>1402</td>
<td>158.92</td>
<td>421.27</td>
<td>206</td>
<td>215.72</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>59/16</td>
<td>48787</td>
<td>565.48</td>
<td>2198.45</td>
<td>1393</td>
<td>528.74</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>-</td>
<td>650</td>
<td>7.54</td>
<td>29.31</td>
<td>18.57</td>
<td>7.05</td>
</tr>
</tbody>
</table>

Table 2. Results for run set bubbleSort.t2

<table>
<thead>
<tr>
<th>Rev.</th>
<th>Type</th>
<th>Rst.</th>
<th>Baseline</th>
<th>Reuse</th>
<th>#CertModule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time</td>
<td>#Iter.</td>
<td>Time #Iter</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>T</td>
<td>9.78</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>Type I</td>
<td>T</td>
<td>9.87</td>
<td>7</td>
<td>0.77</td>
</tr>
<tr>
<td>2</td>
<td>Type I</td>
<td>T</td>
<td>1.60</td>
<td>5</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>Type II</td>
<td>N</td>
<td>1.48</td>
<td>5</td>
<td>0.81</td>
</tr>
<tr>
<td>4</td>
<td>Mixed</td>
<td>T</td>
<td>0.98</td>
<td>4</td>
<td>0.72</td>
</tr>
<tr>
<td>5</td>
<td>Type I</td>
<td>T</td>
<td>9.00</td>
<td>7</td>
<td>8.01</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>-</td>
<td>-</td>
<td>22.93</td>
<td>28</td>
<td>10.66</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>-</td>
<td>-</td>
<td>4.59</td>
<td>5.60</td>
<td>2.13</td>
</tr>
</tbody>
</table>

a loop condition (Type I changes) which affects the corresponding ranking function. To prove the termination of Rev. 5, our algorithm needs to recompute new ranking functions. Even so, Reuse can get a fair acceleration (x1.12) against Baseline by reducing 4 iterations.

On average, Reuse gets x2.15 speedup on run set bubbleSort.t2 by reusing 70% of certified modules. This experiment reveals the relevance of the performance of our algorithm to the adopted program changes (refer to Section 5): no matter which type of program changes, our algorithm is always applicable and can often lead to significant computational saving.
6.2 Experiment on Real-World Benchmark

*Benchmark.* This set of programs were collected from [Codeforces](https://codeforces.com), a website that hosts competitive programming contests. The program collection was limited to the first 100 contests hosted on this website, and filtered by the following restrictions:

- C programs;
- compilable by gcc;
- contains at least 2 loops; and
- ULTIMATE AUTOMIZER can prove termination/non-termination in 500 seconds.

The initial collection contains some trivial, identical or isolated revisions. We took the following steps to remove them:

- We grouped the programs submitted by some contestant for a certain problem into a run set;
- We removed trivially non-terminating programs that contains statements like `while (scanf("%d", &n));`;
- We compared each pair of adjacent revisions using the following command:
  
  ```
  $ git diff --ignore-cr-at-eol --ignore-all-space
  --ignore-blank-lines --ignore-space-change --no-index -U0
  ```
  and removed one if they are identical;
- We removed the run sets with only one program revision.

Finally, we get 276 run sets of 611 program revisions. Therefore, this benchmark contains 611 analysis tasks, among which 335 are regression analysis tasks.

*Results.* Experimental results on real-world benchmark are listed in Table 3. Due to page limitation, we restrict this table to the 20 best and 20 worst run sets, sorted by the “speedup” column.

Looking at the 20 best run sets, our approach gets noteworthy acceleration. The reason is that for almost of these run sets, the previously-constructed certified modules can all be reused and the transformed reuse modules can cover all $\omega$-traces of the program. Thus, the termination is proved immediately.

For the worst several run sets, our approach gets deceleration. The reason is multiple-folds. One is that the program changes make the subsequent revisions much simpler and their termination becomes easy to determine (even for Baseline), such as 59A-usr206, 6C-usr199 and 59A-usr047. These cases are not friendly to our approach, since there are a lot of previously-constructed certified modules, while most of them are useless for proving termination of the subsequent revisions, leading worthless computation of our approach. For example, when analyzing the second revision of 59A-usr206, our approach only reduces 5 iterations by reusing 16 out of 54 certified modules generated by the first version.

Another reason is that the program changes may have impact to the previous termination arguments, causing the previously-constructed certified modules invalid for the current analysis, such as 15C-usr072, 4A-usr082.

The third reason is that the number of the previously-constructed certificate modules is too small (e.g. $\leq 2$), such as 1B-usr035, 14D-usr103 and 58A-usr154. As a result, there is little chance to get a non-empty reuse module, so as to accelerate the current analysis. This can be avoided by setting a minimal number of certified modules. Only when the available number of certified modules exceeds that number, can we apply the incremental approach.

**Example 8.** We present two revisions of 4A-usr082 in the following. The right version heavily changes the nested loop (lines 4 – 5) of the left version, leading the termination arguments for the left

---

2https://codeforces.com
Table 3. Experimental results on real-world benchmark

<table>
<thead>
<tr>
<th>Run set</th>
<th>Baseline</th>
<th>Reuse</th>
<th>#CertModule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programs</td>
<td>T/N</td>
<td>Loc</td>
<td>$T_{r_0}$</td>
</tr>
<tr>
<td>96A-usr146</td>
<td>2/0</td>
<td>121</td>
<td>39.99</td>
</tr>
<tr>
<td>94A-usr112</td>
<td>2/0</td>
<td>122</td>
<td>178.52</td>
</tr>
<tr>
<td>66A-usr023</td>
<td>3/0</td>
<td>80</td>
<td>12.07</td>
</tr>
<tr>
<td>36A-usr085</td>
<td>3/0</td>
<td>120</td>
<td>22.38</td>
</tr>
<tr>
<td>71A-usr056</td>
<td>2/0</td>
<td>84</td>
<td>22.38</td>
</tr>
<tr>
<td>58A-usr237</td>
<td>2/0</td>
<td>98</td>
<td>16.66</td>
</tr>
<tr>
<td>94A-usr167</td>
<td>5/0</td>
<td>203</td>
<td>26.19</td>
</tr>
<tr>
<td>58A-usr234</td>
<td>2/0</td>
<td>114</td>
<td>8.22</td>
</tr>
<tr>
<td>16A-usr055</td>
<td>2/0</td>
<td>69</td>
<td>41.44</td>
</tr>
<tr>
<td>58A-usr184</td>
<td>2/0</td>
<td>82</td>
<td>41.86</td>
</tr>
<tr>
<td>2A-usr133</td>
<td>0/5</td>
<td>904</td>
<td>15.33</td>
</tr>
<tr>
<td>71A-usr041</td>
<td>2/0</td>
<td>58</td>
<td>19.85</td>
</tr>
<tr>
<td>58A-usr150</td>
<td>2/0</td>
<td>72</td>
<td>3.11</td>
</tr>
<tr>
<td>58A-usr002</td>
<td>2/0</td>
<td>68</td>
<td>3.55</td>
</tr>
<tr>
<td>71A-usr058</td>
<td>2/0</td>
<td>48</td>
<td>5.66</td>
</tr>
<tr>
<td>9A-usr068</td>
<td>2/0</td>
<td>55</td>
<td>9.87</td>
</tr>
<tr>
<td>78A-usr217</td>
<td>2/0</td>
<td>68</td>
<td>5.36</td>
</tr>
<tr>
<td>25A-usr227</td>
<td>4/0</td>
<td>142</td>
<td>6.52</td>
</tr>
<tr>
<td>11A-usr091</td>
<td>2/0</td>
<td>60</td>
<td>95.97</td>
</tr>
<tr>
<td>38A-usr049</td>
<td>2/0</td>
<td>33</td>
<td>3.79</td>
</tr>
</tbody>
</table>

| Sum              | 586/25   | 24627 | 4054.47     | 4555.25| 2962   | 1476.76| 984 | 2677 | 1902 |
| Average           | -        | 40.31 | 14.69       | 13.60  | 8.84   | 4.41   | 2.94| 7.99 | 5.68 | 3.08   |

version inapplicable to the right version. As a result, the right version can only reuse 1 out of the 10 previously-constructed certified modules. In the end, Reuse takes more time than Baseline.

```c
int w, i, j, a;
scanf("%d", &w);
if(w%2==1)
  for(i=2; i<w; i=i+2)
    for(j=w-1; j>0; j=j-2) {
      a = i + j;
      if(w==a)  
        return 0;
    }

...  
```

On average, Baseline requires 8.84 iterations and 13.60 seconds to complete an analysis task. In contrast, Reuse needs 2.94 iterations and 4.41 seconds. Our approach achieves an overall x3.08 speedup over Baseline, reusing 5.68 out of 7.99 certified modules.

6.2.1 Degree of Program Changes. To further analyze the results in Table 3, we use DiffRate to measure the degree of program changes, and then analyze the impact of DiffRate to the efficiency of our approach. The DiffRate, i.e., the rate of different lines of codes, is calculated by

$$\text{DiffRate} = \frac{\#\text{deleted_lines} + \#\text{added_lines}}{\#\text{LoC}_1 + \#\text{LoC}_2},$$

where \#\text{LoC}_1 and \#\text{LoC}_2 are lines of codes of the two comparative programs, respectively.

We first divided DiffRate into 10 ranges: 0%–10%, 10%–20%, ..., 90%–100%. Then, we counted the number of programs in each range, and summarized the average speedup of our approach among the programs in each region. The statistics results are depicted in Figure 6. The first two histograms indicate there are 183 programs with DiffRate less than 10%, and 41 programs with DiffRate between 10% and 20%. These programs have relatively small changes; the speedup of our approach on these programs is significant. It is worth noting that the cases with DiffRate between 50% and 60% also show significant speedup, which suggest the capability of our approach on large changes. However,
Table 4. Comparison on loop-relevant and loop-irrelevant changes

| Relevant? | #Prog. | Baseline | | Reuse | | #CertModule | | Time | | #Iter. | | Time | | #Iter. | | Avai. | | Reuse | | Speedup |
|----------|--------|----------||-------||-------||-------||-------||-------||-------||-------||-------||-------||-------||-------|
| No       | 66     | 449.64   | 558 | 95.77 | 64 | 495 | 459 | 4.69 |
| Yes      | 269    | 4105.61  | 2404 | 1380.99 | 920 | 2182 | 1443 | 2.97 |
| Sum      | 335    | 4555.25  | 2962 | 1476.76 | 984 | 2677 | 1902 | 3.08 |

from Figure 6, we did not observe a clear relation between the degree of program changes and the efficiency of our approach. This explains that our approach is not very sensitive to the degree of program changes (also witnessed in Section 5).

Example 9. Consider the following two revisions of 94A-usr005. Although the right version changes a lot against the left version (even altering the number of loops), our approach can still reuse 25 out of 26 certified modules, reduces 24 iterations and achieves $x_{14.33}$ acceleration.

```c
int i,j,t[8];
for(i=0;i<8;i++){
    for(j=0;j<10;j++)
    scanf("%c ",&x[i][j]);
    x[i][10] = '\0';
}
for(i=0;i<10;i++)
    gets(y[i]);
for(i=0;i<8;i++)
    for(j=0;j<10;j++)
        if(strcmp(x[i],y[j])==0)
            t[i]=j;
for(i=0;i<8;i++){
    printf("%d",t[i]);
}
```

6.2.2 Loop-Relevant vs. Loop-Irrelevant Changes. To further analyze the results in Table 3, we applied Frama-C [Cuq et al. 2012] (more specifically, its engine of change impact analysis) to distinguish loop-relevant and loop-irrelevant changes. A loop-relevant change has impact to loops (either loop conditions or loop bodies) of the program. We then compared the efficiency of our approach with these two kinds of changes.

The statistics results are listed in Table 4, where the first column ("Relevant?") indicates if the change is loop-relevant or not. The second column ("# Prog.") presents the number of programs in each type of changes. All other columns have the same meanings as in Table 1.

From Table 4, we observed that our approach get a significant acceleration ($x_{4.69}$) for 66 loop-irrelevant changes. This is consistent with our intuition. Program termination is closely related to the loops. If a change has no impact to loops, we have much bigger chance (93% in the table) to keep all previously-constructed certified modules. We also observed that our approach gets considerable acceleration ($x_{2.97}$) for 269 loop-relevant changes. This result is more meaningful, witnessing the wide applicability of our approach for various program changes.
Example 10. The following two revisions of 12A-usr172 show a Type II change. The left version is obviously non-terminating. The right version deletes the innermost while loop and becomes terminating. Although the terminating results of these two programs are different, our approach still benefits from the certified modules generated by the left program, and gets a 4.91x speedup.

```c
char x[3][3];
int i, j;
for (i = 0; i < 3; i++) {
    for (j = 0; j < 3; j++) {
        do {
            scanf("%c", &x[i][j]);
        } while (x[i][j] == '\n');
    }
}
```

```c
char x[3][3];
int i, j;
for (i = 0; i < 3; i++) {
    for (j = 0; j < 3; j++) {
        while (x[i][j] == '\n');
    }
}
```

Example 11. The following two revisions of 96A-usr208 show a Mixed change. The right version deletes a statement at line 4 and adds a statement at line 9. Among the 7 certified modules generated by the left version, only 2 can be reused, due to the impact of program changes to the termination arguments. Nevertheless, our method still achieves a non-trivial acceleration of x1.79.

```c
char a[150];
int i, j, flag = 0;
scanf("%s", a);
for (i = 0; a[i] != '\0'; i++) {
    for (j = 0; j < 7; j++) {
        if (a[i] != a[i + j])
            break;
    }
    if (j == 7) {
        flag = 1; break;
    }
}
```

```c
char a[150];
int i, j, flag = 0;
scanf("%s", a);
for (i = 0; a[i] != '\0'; ) {
    for (j = 0; j < 7; j++) {
        if (a[i] != a[i + j])
            break;
    }
    i++; break;
}
```

In summary, by reusing the previously-constructed certified modules, our incremental approach can significantly reduce the time consumption for termination verification, and has good applicability to various types of real-world program changes.

7 RELATED WORK
7.1 Termination Analysis
Termination analysis was investigated mainly in two directions, proving termination and proving non-termination. Our paper was focused on termination proving only. But for the sake of completeness, we give a brief overview on both directions.

Proving Termination. A wide range of methods utilize iterative reasoning and counterexample driven method for proving termination [Brockschmidt et al. 2013; Cook et al. 2006a,b, 2013; Gulwani et al. 2008; Harris et al. 2010; Heizmann et al. 2014; Kroening et al. 2008, 2010; Larraz et al. 2013; Podelski and Rybalchenko 2004; Podelski and Rybalchenko 2005, 2011; Ströder et al. 2017]. These methods usually rely on termination arguments (e.g., ranking function) to show that their transition models cannot be executed forever. Most methods find termination arguments using constraint-based synthesis, among which techniques [Bagnara et al. 2012; Ben-Amram and Genaim 2013; Colón and Sipma 2001; Cook et al. 2010; Podelski and Rybalchenko 2004] are based on linear...
arithmetic constraint solving, and techniques [Bradley et al. 2005a,b] are based on non-linear arithmetic constraint solving. In some methods [Brockschmidt et al. 2013; Cook et al. 2006a; Urban et al. 2016], the termination proving is reduced to a safety verification problem. For example, Urban et al. [Urban et al. 2016] used a safety verifier to build the termination arguments. Bound analysis [Gulwani et al. 2009; Nori and Sharma 2013] goes another direction by computing a symbolic upper bound on the number of iterations for each loop. It often works in the infer-and-validate framework – first infers a candidate bound and then validate it. If we get a valid bound for each loop, we prove the termination of the program.

Termination arguments are usually harder to solve for the general programs with branches and nesting. Consequently, many methods were focused on the lasso programs, that is a simpler type of programs containing only one infinite path. The termination argument for this type of programs can be solved rather efficiently [Ben-Amram and Genaim 2013, 2014, 2015, 2017; Bradley et al. 2005a; Cook et al. 2010; Heizmann et al. 2013b; Kroening et al. 2008; Podelski and Rybalchenko 2004]. In particular, Leike et al. [Leike and Heizmann 2014] use linear ranking templates for the constraint-based synthesis of termination arguments for linear loop programs. For the general program, some methods [Borralleras et al. 2017; Heizmann et al. 2014; Le et al. 2015; Urban et al. 2016] proposed to decompose the original program into components such that termination arguments for each component are quite simple. For example, HipTNT [Gupta et al. 2008] decomposed the state space of the program and infer ranking functions for each component separately. ULTIMATE AUTOMIZER found lasso traces in the control flow automaton iteratively, and converts them to lasso programs and finds termination arguments only for lasso-shaped programs.

All these research works were focused on a single program version. In this paper, we address the termination proving for evolving programs. We proposed an incremental termination proving approach. Experimental results show very promising performance of our approach.

Proving Non-termination. Most techniques for proving non-termination [Bakhirkin and Piterman 2016; Brockschmidt et al. 2012; Cook et al. 2014; Gupta et al. 2008; Leike and Heizmann 2018] are based on the searching for lasso-shaped traces or the discovering of recurrence sets. For example, TNT [Gupta et al. 2008] first enumerates lasso-shaped candidate paths for counterexamples dynamically, and then proves their feasibility statically. Leike et al. [Leike and Heizmann 2018] proved the non-terminating linear lasso programs by deciding the existence of a geometric non-termination argument, using a nonlinear algebraic $\exists$-constraint. ULTIMATE AUTOMIZER proved non-termination of programs by deciding if the selected lasso trace is non-terminating or not. If it is non-terminating, this lasso trace is returned as a counterexample. A minor range of tools reduce termination analysis to safety attracts significant attention. In particular, Velroyen et al. [Velroyen and Rümmer 2008] showed that terminating states of a program are unreachable from certain initial states via invariant generation. Gulwani et al. [Gulwani et al. 2008] used weakest precondition inference to discover the most-general characterization of the inputs under which the original program is non-terminating. Chen et al. [Chen et al. 2014] iteratively eliminated terminating traces through a loop by adding extra assumptions. ULTIMATE AUTOMIZER eliminates terminating lasso traces via the language-theoretic difference (complementation and intersection) of Büchi automata.

7.2 Incremental Verification
Incremental verification was also investigated mainly in two directions, the verification of differences and the reuse of intermediate results.

Verification of Differences. In this line of research, one attempts to establish the correctness of the new program by proving its (conditional) equivalence to an old and verified program. Many techniques [Backes et al. 2013; Beyer et al. 2012; Böhme et al. 2013; Chaki et al. 2012; Fedyukovich
et al. 2016; Felsing et al. 2014; Godlin and Strichman 2009; Mora et al. 2018; Rungta et al. 2012; Trostanetski et al. 2017; Yang et al. 2014] have been proposed in this area of research. For instance, Golden et al. [Godlin and Strichman 2009] proposed a technique for proving conditional equivalence of two programs by abstraction and decomposition of procedures. Backes et al. [Backes et al. 2013] proposed to distinguish the program behaviors that are impacted by the changes. Only the impacted program behaviors needed to be considered during the regression verification. Beyer et al. [Beyer et al. 2012] proposed the conditional model checking, which outputs a condition such that the program satisfies the specification under this condition. Felsing et al. [Felsing et al. 2014] reduced the equivalence proving of two related imperative integer programs to Horn constraints over uninterpreted predicates, and then solved the constraints using an SMT solver. Moreover, Rungta et al. [Rungta et al. 2012] presented a technique for interprocedural change impact analysis. Yang et al. [Yang et al. 2014] introduced an incremental approach for checking the conformance of code against different properties. Trostanetski et al. [Trostanetski et al. 2017] analyzed the semantic difference between successive revisions. Fedyukovich et al. [Fedyukovich et al. 2016] established the property directed equivalence of two programs by discovering a fresh property that actually holds using simulation.

**Reuse of Intermediate Results.** This area studies the reuse of previously-generated results to the current verification. A variety of information has been proposed for reuse.

Some researchers [Alt et al. 2017; Conway et al. 2005; Henzinger et al. 2003; Lauterburg et al. 2008; Sery et al. 2012; Yang et al. 2009] proposed to keep the reached state space and reuse them in the further verification runs. The rationale of these techniques is that state spaces of consecutive versions tend to be similar. For instance, Visser et al. [Visser et al. 2012] noticed the importance of constraint solving for symbolic execution. They proposed to cache and reuse the results of constraint solving. This approach was further improved in [Aquino et al. 2015; Jia et al. 2015] from different aspects. Beyer et al. [Beyer et al. 2013] proposed to use abstract precision as the intermediate result. [Fedyukovich et al. 2013; Sery et al. 2012] proposed a regression verification technique by means of interpolation-based procedure summaries. Pastore et al. [Pastore et al. 2014] proposed a method to validate that an already tested code has not been broken by an upgrade, it maintains a test suite that can be used to revalidate the software as it evolves.

Rothenberg et al. [Rothenberg et al. 2018] proposed to reuse the sequence of Floyd-Hoare automata computed in the trace abstraction. Their approach differs from ours in the following aspects. Firstly, the targeted problems are different. Their approach mainly considers the safety verification, while our paper is focused on the termination proving. To the best of our knowledge, our approach is the first attempt to incremental termination analysis. Secondly, the reused intermediate results are different. The Floyd-Hoare automaton was applied in [Rothenberg et al. 2018] to recognize a set of finite traces, while in our paper, we use certificate module to recognize a set of infinite traces. Last but not least, a so-called lazy reuse was proposed in [Rothenberg et al. 2018], which looks similar but is indeed different to our on-demand construction — the former is a scheme of module reuse, while the latter is a scheme of module construction. More specifically, considering the basic scheme of our approach in Figure 4, the on-demand construction will lead the full construction of reuse modules be deferred to the language inclusion checking phase, while the lazy reuse in [Rothenberg et al. 2018] will add a new “reuse” phase after the language inclusion checking and defer the subtraction of reuse modules to this new phase.

8 **CONCLUSION**

We proposed in this paper an incremental technique for termination analysis of evolving programs. A transformation procedure was developed to reuse certified modules across different program
The proposed approach is applicable to various types of program changes. Its soundness and relative completeness were formally proved. We implemented the approach in ULTIMATE AUTOMIZER. Experimental results show dramatic improvement of our approach.

ACKNOWLEDGMENTS
This work was partially funded by the National Key R&D Program of China (No. 2018YFB1308601), the NSF of China (No. 61672310 and No. 62072267), the CDZ project CAP (No. GZ 1023) and the Guangdong Science and Technology Department (No. 2018B010107004).

REFERENCES


Termination Analysis for Evolving Programs: An Incremental Approach by Reusing Certified Modules

Proceedings of the 46th Annual Design Automation

Benny Godlin and Ofer Strichman. 2009. Regression Verification. In

Grigory Fedyukovich, Ondrej Sery, and Natasha Sharygina. 2013. eVolCheck: Incremental Upgrade Checker for C. In

Pascal Cuoq, Florent Kirchner, Nikolai Kosmatov, Virgile Prevosto, Julien Signoles, and Julien Signoles. 2012. Frama-C.

Byron Cook, Andreas Podelski, and Andrey Rybalchenko. 2006b. Terminator: Beyond Safety. In


Inter-procedural Analysis of Safety Properties. In

Byron Cook, Carsten Fuhs, Kaustubh Nimkar, and Peter O’Hearn. 2014. Proving Nontermination via Safety. In

Byron Cook, Daniel Kroening, Philipp Rümmer, and Christoph M. Wintersteiger. 2010. Ranking Function Synthesis for
Bit-Vector Relations. In


Byron Cook, Abigail See, and Florian Zuleger. 2013. Ramsey vs. Lexicographic Termination Proving. In


Computer Aided Verification, Swarat Chaudhuri and Azadeh Farzan (Eds.). Springer International Publishing, Cham, 433–453. https://doi.org/10.1007/978-3-319-41540-6_24


Benny Godlin and Ofer Strichman. 2009. Regression Verification. In


Termination Analysis for Evolving Programs: An Incremental Approach by Reusing Certified Modules


