Lemmaless Induction in Trace Logic

Ahmed Bhayat[1(✉)], Pamina Georgiou[2], Clemens Eisenhofer[2],
Laura Kovács[2], and Giles Reger[1]

[1] University of Manchester, Manchester, UK
{ahmed.bhayat,giles.reger}@manchester.ac.uk
[2] TU Wien, Vienna, Austria
{pamina.georgiou,clemens.eisenhofer,laura.kovacs}@tuwien.ac.at

Abstract. We present a novel approach to automate the verification of first-order inductive program properties capturing the partial correctness of imperative program loops with branching, integers and arrays. We rely on trace logic, an instance of first-order logic with theories, to express first-order program semantics by quantifying over program execution timepoints. Program verification in trace logic is translated into a first-order theorem proving problem where, to date, effective reasoning has required the introduction of so-called trace lemmas to establish inductive properties. In this work, we extend trace logic with generic induction schemata over timepoints and loop counters, reducing reliance on trace lemmas. Inferring and proving loop invariants becomes an inductive inference step within superposition-based first-order theorem proving. We implemented our approach in the Rapid framework, using the first-order theorem prover Vampire. Our extensive experimental analysis shows that automating inductive verification in trace logic is an improvement compared to existing approaches.

1 Introduction

Automating the verification of programs containing loops and recursive data structures is an ongoing research effort of growing importance. While different techniques for proving the correctness of such programs are in place [5,6,10,13], most existing tools in this realm are heavily based on *satisfiability modulo theories* (SMT) backends [4,8] that come with strong theory reasoning but have limitations in quantified reasoning. In contrast, first-order theorem provers enable quantified reasoning modulo theories [19,24,25], such as linear integer arithmetic and arrays. First-order reasoning can thus complement the aforementioned verification efforts when it comes to proving program properties with complex quantification, as evidenced in our original work on the Rapid framework [11] which utilised the Vampire theorem prover [2,20].

At a high level, the Rapid framework [11] works by translating a program into *trace logic*, adding a number of ad hoc trace lemmas, asserting a desired property, and then running an automated theorem prover on the result. The effectiveness of this approach depends on the underlying trace lemmas. This

K. Buzzard and T. Kutsia (Eds.): CICM 2022, LNAI 13467, pp. 191–208, 2022.
https://doi.org/10.1007/978-3-031-16681-5_14

paper focuses on building induction support into the Vampire theorem prover to reduce reliance on these lemmas.

To understand the role of these trace lemmas (and therefore, what support must be added to the theorem prover) we briefly overview trace logic and the Rapid framework in a little more detail. Trace logic is an instance of first-order logic with theories, such that the program semantics of imperative programs with loops, branching, integers, and arrays can be directly encoded in trace logic. A key feature of this encoding is tracking program executions by quantifying over execution *timepoints* (rather than only over single states), which may themselves be parameterised by *loop iterations*. In principle, we can check whether a translated program entails the desired property in trace logic using an automated theorem prover for first-order logic. In our case, we make use of the saturation-based theorem prover Vampire which implements the superposition calculus [3]. However, a straightforward use of theorem proving often fails in establishing validity of program properties in trace logic, as the proof requires some specific induction, in general not supported by superposition-based reasoning.

In our previous work [11], we overcame this challenge by introducing so-called *trace lemmas* capturing common patterns of inductive loop properties over arrays and integers. Inductive loop reasoning in trace logic is then achieved by generating and adding trace lemma instances to the translated program. However, there are two significant limitations to using trace lemmas:

1. Trace lemmas capture inductive patterns/templates that need to be manually identified, as induction is not expressible in first-order logic. As such, they cannot be inferred by a first-order reasoner, implying that the effectiveness of trace logic reasoning depends on the expressiveness of manually supplied trace lemmas.
2. When instantiating trace lemmas with appropriate inductive program variables, a large number of inductive properties are generated, causing saturation-based proof search to diverge and fail to find program correctness proofs in reasonable time.

In this paper we address these limitations by reducing the need for trace lemmas. We achieve this by introducing a couple of novel induction inferences. Firstly, *multi-clause goal induction* which applies induction in a goal oriented fashion as many safety program assertions are structurally close to useful loop invariants. Secondly, *array mapping induction* which covers certain cases where the required loop invariant does not stem from the goal. Specifically, we make the following contributions:

Contribution 1. We introduce two new inference rules, *multi-clause goal* and *array mapping* induction, for *lemmaless induction* over loop iterations (Sects. 5–6). The inference rules are compatible with any saturation-based inference system used for first-order theorem proving and work by carrying out induction on terms corresponding to final loop iterations.

```
1   func main() {
2     const Int[] a;
3     const Int[] b;
4     Int[] c;
5     const Int length;
6     Int i = 0;
7
8     while (i < length) {
9       c[2*i] = a[i]
10      c[(2*i) + 1] = b[i]
11      i = i + 1;
12    }
13  }
14  assert (∀pos𝕀.∃l𝕀.((0 ≤ pos < (2 × length))
15      → c(main_end, pos) = a(l) ∨ c(main_end, pos) = b(l)
      ))
```

Fig. 1. Copying elements from arrays `a` and `b` to even/odd positions in array `c`.

Contribution 2. We implemented our approach in the first-order theorem prover Vampire [20]. Further, we extended the Rapid framework [11] to support inductive reasoning in the automated backend (Sect. 7). We carry out an extensive evaluation of the new method (Sect. 8) comparing against state-of-the-art approaches SeaHorn [12,13] and Vajra/Diffy [5,6].

2 Motivating Example

We motivate our work with the example program in Fig. 1. The program iterates over two arrays `a` and `b` of arbitrary, but fixed length `length` and copies array elements into a new array `c`. Each even position in `c` contains an element of `a`, while each odd position an element of `b`. Our task is to prove the safety assertion at line 14: at the end of the program, every element in `c` is an element from `a` or `b`. This property involves (i) alternation of quantifiers and (ii) is expressed in the first-order theories of linear integer arithmetic and arrays. Note that in the safety assertion, the program variable `length` is modeled as a logical constant of the same name of sort integer, whilst the constant arrays `a` and `b` are modeled as logical functions from integers to integers. The mutable array variable `c` is additionally equipped with a timepoint argument `main_end`, indicating that the assertion is referring to the value of the variable at the end of program execution.

Proving the correctness of this example program remains challenging for most state-of-the-art approaches, such as [5,6,10,12], mainly due to the complex quantified structure of our assertion. Moreover, it cannot be achieved in the current Rapid framework either, as existing trace lemmas do not relate the values of multiple program variables, notably equality over multiple array variables. In fact, to automatically prove the assertion, we need an inductive property/trace lemma formalizing that each element at an even position in c is an element of a or b at each valid loop iteration, thereby also restricting the bounds of the loop counter variable i. Naïvely adding such a trace lemma would be highly inefficient as automated generation of verification conditions would introduce many instances that are not required for the proof.

3 Related Work

Most of recent research in verifying inductive properties of array-manipulating programs focuses on quantified invariant generation and/or is mostly restricted to proving universally quantified program properties. The works [10,13] generate universally quantified inductive invariants by iteratively inferring and strengthening candidate invariants. These methods use SMT solving and as such are restricted to first-order theories with a finite model property. Similar logical restrictions also apply to [23], where linear recurrence solving is used in combination with array-specific proof tactics to prove quantified program properties. A related approach is described in [6], where relational invariants instead of recurrence equations are used to handle universal and quantifier-free inductive properties. Unlike these, our work is not limited to universal invariants but can infer and prove inductive program properties with alternations of quantifiers.

With the use of extended expressions and induction schemata, our work shares some similarity with template-based approaches [16,21,26]. These works infer and prove universal inductive properties based on Craig interpolation, formula slicing and/or SMT generalizations over quantifier-free formulas. Unlike these works, we do not require any assumptions on the syntactic shape of the first-order invariants. Moreover, our invariants are not restricted to the shape of our induction schemata. Rather, we treat inductive (invariant) inferences as additional rules of first-order theorem provers, maintaining thus the efficient handling of arbitrary first-order quantifiers. Our framework can be used in arbitrary first-order theories, even with theories that have no interpolation property and/or a finite axiomatization, as exemplified by our experimental results using inductive reasoning over arrays and integers.

Inductive theorem provers, such as ACL2 [17] and HipSpec [7], implement powerful induction schemata and heuristics. However these provers, to the best of our knowledge, automate inductive reasoning for only universally quantified inductive formulas using a goal/subgoal architecture, for which user-guidance is needed to split conjectures into subgoals. In contrast, our work can prove formulas of full first-order logic by integrating and fully automating induction in saturation-based proof search. By combining induction with saturation, we allow these techniques to interleave and complement each other, something that pure induction provers cannot do. Unlike tools such as Dafny [22], our approach is fully automated requiring no user annotations.

First-order theorem proving has been used to derive invariants with alternations of quantifiers in our previous work [11]. Our current work generalizes the inductive capabilities of [11] by reducing the expert knowledge of [11] in introducing inductive lemmas to guide the process of proving inductive properties.

4 Preliminaries

Many-Sorted First-Order Logic. We consider standard many-sorted first-order logic with built-in equality, denoted by $\simeq$. By $s = F[u]$ we indicate that the term u is a subterm of s surrounded by (a possibly empty) context F.

We use x, y to denote variables, l, r, s, t for terms and sk for Skolem symbols. A *literal* is an atom A or its negation $\neg A$. A *clause* is a disjunction of literals $L_1 \vee ... \vee L_n$, for $n \geq 0$. Given a formula F, we denote by $\text{CNF}(F)$ the clausal normal form of F.

For a logical variable x of sort S we write x_S. A *first-order theory* denotes the set of all valid formulas on a class of first-order structures. Any symbol in the signature of a theory is considered *interpreted.* All other symbols are *uninterpreted.* In particular, we use the theory of linear integer arithmetic denoted by $\mathbb{I}$ and the boolean sort $\mathbb{B}$. We consider natural numbers as the term algebra $\mathbb{N}$ with four symbols in the signature: the constructors `0` and successor `suc`, as well as `pred` and $<$ respectively interpreted as the predecessor function and less-than relation. Note that we do not define any arithmetic on naturals. We assume familiarity with the basics of saturation theorem proving.

4.1 Trace Logic $\mathcal{L}$

Trace logic, denoted as $\mathcal{L}$, is an instance of many-sorted first-order logic with theories. Its signature is $\Sigma(\mathcal{L}) := S_{\mathbb{N}} \cup S_{\mathbb{I}} \cup S_{\mathbb{L}} \cup S_V \cup S_n$, includes respectively the signatures of the theory of natural numbers $\mathbb{N}$ (as a term algebra), the inbuilt integer theory $\mathbb{I}$, a set $S_{\mathbb{L}}$ of timepoints (also referred to as *locations*), a set of symbols representing program variables S_V, as well as a set of symbols representing last iteration symbols S_n. For more details on trace logic, refer to [11].

4.2 Programming Model $\mathcal{W}$

We consider programs written in a WHILE-like programming language $\mathcal{W}$, as given in the (partial) language grammar of Fig. 2. Programs in $\mathcal{W}$ contain mutable and immutable integer as well as integer-array program variables and consist of a single top-level function `main` comprising arbitrary nestings of while-loops and if-then-else branching. We consider expressions over booleans and integers without side effects.

4.3 Translating Expressions to Trace Logic

Locations and Timepoints. We consider programs as sets of locations over time: given a program statement `s`, we denote its location by l_s of type $\mathbb{L}$, the location/timepoint sort, corresponding to the line of the program where the statement appears. When `s` is a while-loop the corresponding location is revisited at multiple timepoints of the execution. Thus, we model such locations

```
   program ::= function
  function ::= func main(){ subprogram }
subprogram ::= statement | context
   context ::= statement; ... ; statement
 statement ::= atomicStatement
             | if( condition ){ context } else { context }
             | while( condition ){ context }
```

Fig. 2. Grammar of $\mathcal{W}$.

as functions over *loop iterations* $l_s : \mathbb{N} \mapsto \mathbb{L}$, where the argument of sort $\mathbb{N}$ intuitively corresponds to the number of loop iterations. Further, for each loop statement `s` we model the last loop iteration by a symbol $nl_s \in S_n$ of target sort $\mathbb{N}$. Let `p` be a program statement or context. We use $start_{\texttt{p}}$ to denote the location at which the execution of `p` has started and $end_{\texttt{p}}$ to denote the location that occurs just after the execution of `p`. We use *main_end* to denote the location at the end of the main function.

Example 1. Consider line 6 of our running example in Fig. 1. Term l_6 corresponds to the timepoint of the first assignment of 0 to program variables `i` while $l_8(\texttt{0})$ and $l_8(nl_8)$ denote the timepoints of the loop at the first and last loop iteration respectively. Further, we can quantify over all executions of the loops by quantifying over all iterations smaller than the last e.g. $\forall it_{\mathbb{N}}.it < nl_8 \rightarrow F[l_8(it)]$ where $F[l_8(it)]$ is some first-order formula.

Program Variables. Program variable are expressed as functions over timepoints. We express an integer variable `v` as a function $v : \mathbb{L} \mapsto \mathbb{I}$, where $v \in S_V$. Let tp be a term of sort $\mathbb{L}$. Then, $v(tp)$ denotes the value of `v` at timepoint tp. We model numeric array variables `v` with an additional argument of sort $\mathbb{I}$ to denote the position of an array access. We obtain $v : \mathbb{L} \times \mathbb{I} \mapsto \mathbb{I}$. Immutable variables are modelled as per their mutable counterparts, but without the timepoint argument.

Example 2. To denote program variable `i` at the location of the assignment in line 6, we use the equation $i(l_6) \simeq 0$. For the first assignment of `c` within the loop, we write $c(l_8(it), 2 \times i(l_8(it))) \simeq a(i(l_8(it)))$ for some iteration it. As `a` is a constant array, the timepoint argument is omitted.

Program Expressions. Let `e` be an arbitrary program expression. We write $[\![\texttt{e}]\!](tp)$ to denote the logical denotation of `e` at timepoint tp. We do not provide the full inductive definition of the denotation function $[\![\ ]\!](tp)$ here, just a few of its cases. If `e` is an integer variable `v`, then $[\![\texttt{e}]\!](tp) = v(tp)$. If `e` is an integer array access of the form $\mathtt{v}[\mathtt{e}_1]$, then $[\![\mathtt{e}]\!](tp) = v(tp, [\![\mathtt{e}_1]\!](tp))$. If $\mathtt{e}$ is an expression of the form $\mathtt{e}_1 + \mathtt{e}_2$, then $[\![\mathtt{e}]\!](tp) = [\![\mathtt{e}_1]\!](tp) + [\![\mathtt{e}_2]\!](tp)$.

Common Abbreviations. Let $\mathtt{e}, \mathtt{e}_1, \mathtt{e}_2$ be program expressions, tp_1, tp_2 be two timepoints and $v \in S_V$ denote the functional representation of a program variable. The trace logic formula $v(tp_1) \simeq v(tp_2)$ asserts that the variable v has the same value at timepoints tp_1 and tp_2. We introduce definitions for two formulas that are widely used in defining the axiomatic semantics of $\mathcal{W}$ in the next section. To ease the notational burden, we ignore array variables in the definitions provided. Firstly, we introduce a definition for the formula that expresses that the value of a variable v changes between timepoints tp_1 and tp_2 whilst the values of all other variables remain the same.

$$Update(v, e, tp_1, tp_2) \quad := \quad v(tp_2) \simeq [\![\mathtt{e}]\!](tp_1) \wedge \textstyle\bigwedge_{v' \in S_V \setminus \{v\}} v'(tp_1) \simeq v'(tp_2),$$

Secondly, we introduce a definition for the formula that expresses that the value of all variables stays the same between timepoints tp_1 and tp_2

$$EqAll(tp_1, tp_2) := \bigwedge_{v \in S_V} v(tp_1) \simeq v(tp_2)$$

4.4 Axiomatic Semantics of $\mathcal{W}$ in $\mathcal{L}$

The semantics of a program in $\mathcal{W}$ is given by the conjunction of the respective axiomatic semantics of each program statement of $\mathcal{W}$ occurring in the program. In general, we define reachability of program statements over timepoints rather than program states. We briefly recall the axiomatic semantics of assignments and while-loops respectively, again ignoring the array variable case.

Assignments. Let `s` be an assignment `v = e`, where `v` is an integer-valued program variable and `e` is an expression. The evaluation of `s` is performed in one step such that, after the evaluation, the variable `v` has the same value as `e` before the evaluation while all other variables remain unchanged. We obtain

$$[\![\mathtt{s}]\!] := Update(v, e, start_{\mathtt{s}}, end_{\mathtt{s}}) \tag{1}$$

While-Loops. Let `s` be the while-statement `while(Cond){c}` where `Cond` is the *loop condition.* The semantics of `s` is given by the conjunction of the following properties: (2a) the iteration nl_s is the first iteration where `Cond` does not hold anymore, (2b) jumping into the loop body does not change the values of the variables, (2c) the values of the variables at the end of evaluating the loop `s` are equal to the values at the loop condition location in iteration nl_s. As such, we have

$$\begin{array}{llr}
[\![\mathtt{s}]\!] := & \forall it^s_{\mathbb{N}}.\ (it^s < nl_{\mathtt{s}} \rightarrow [\![\mathtt{Cond}]\!](tp_{\mathtt{s}}(it^s))) & \\
\wedge & \neg[\![\mathtt{Cond}]\!](tp(nl_{\mathtt{s}})) & (2a) \\
\wedge & \forall it_{\mathbb{N}}.\ (it < nl_{\mathtt{s}} \rightarrow EqAll(start_{\mathtt{c}}, tp_{\mathtt{s}}(it)) & (2b) \\
\wedge & EqAll(end_{\mathtt{s}}, tp_s(nl_{\mathtt{s}})) & (2c)
\end{array}$$

4.5 Trace Lemma Reasoning

Trace logic $\mathcal{L}$ allows one to naturally express common program behavior over timepoints. Specifically, it allows us to reason about (i) all iterations of a loop, and (ii) the existence of specific timepoints. In [11], we leveraged such reasoning with the use of so-called *trace lemmas*, capturing common inductive properties of program loops. Trace lemmas are instances of the schema of bounded induction for natural numbers

$$\begin{aligned}&\Big(P(bl) \wedge \forall x_{\mathbb{N}}.\big((bl \leq x < br \wedge P(x)) \rightarrow P(\mathtt{suc}(x))\big)\Big) \rightarrow \\ &\quad \forall x_{\mathbb{N}}.\big(bl \leq x < br \wedge P(x)\big)\end{aligned} \tag{3}$$

An example of a trace lemma would be the statement formalising that a certain program variable's value remains unchanged from a specific iteration to the end of loop execution. In this work, instead of adding instances of (3) statically to strengthen loop semantics, we move induction into the first-order prover. The advantage of adding instances of (3) dynamically is that during proof search we have more information available and can thus perform induction in a more controlled and goal oriented fashion.

Nonetheless, due to some limitations in our first-order prover, we are unable to completely do away with additional lemmas. Specifically, we need to nudge the prover to deduce that a loop counter expression will, at the end of loop execution, have the value of the expression it is compared against in the loop condition.

(A) Equal Lengths Trace Lemma We define a common property of loop counter expressions. We call a program expression `e` *dense* at loop `w` if:

$$Dense_{w,e} := \forall it_{\mathbb{N}}.\Big(it < nl_{\mathtt{w}} \rightarrow \begin{pmatrix} [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{suc}(it))) \simeq [\![\mathtt{e}]\!](tp_{\mathtt{w}}(it)) \vee \\ [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{suc}(it))) \simeq [\![\mathtt{e}]\!](tp_{\mathtt{w}}(it)) + 1 \end{pmatrix}\Big).$$

Let `w` be a while-statement, $C_{\mathtt{w}} := \mathtt{e} < \mathtt{e'}$ be the loop condition where `e'` is a program expression that remains constant during iterations of `w`. The *equal lengths trace lemma of* `w`*,* `e` *and* `e'` is defined as

$$\begin{aligned}&\big(Dense_{w,e} \wedge [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{0})) \leq [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(\mathtt{0}))\big) \rightarrow \\ &\qquad [\![\mathtt{e}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}})) \simeq [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}})).\end{aligned} \tag{A}$$

Trace lemma A states that a dense expression `e` smaller than or equal to some expression `e'` that does not change in the loop, will eventually, specifically in the last iteration, reach the same value as `e'`. This follows from the fact that we assume termination of a loop, hence we assume the existence of a timepoint $nl_{\mathtt{w}}$ where the loop condition does not hold anymore. As a consequence, given that the loop condition held at the beginning of the execution, we can derive that the loop counter value immediately after the loop execution $[\![\mathtt{e}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}}))$ will necessarily equate to $[\![\mathtt{e'}]\!](tp_{\mathtt{w}}(\mathtt{0})) = [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}}))$. Note that a similar lemma can just as easily be added for dense but decreasing loop counters.

5 Multi-Clause Goal Induction for Lemmaless Induction

As mentioned above, the main focus of our work is moving induction into the saturation prover. We achieve this by adding inference rules that apply induction to loop counter terms. We leverage recent theorem proving effort on *bounded (integer) induction* in saturation [14,15]. However, as illustrated in the following, these recent efforts cannot be directly used in trace logic reasoning since we need to (i) adjust bounded induction for the setting of natural numbers, and (ii) generalise to multi-clause induction. We discuss these steps using Fig. 1. Verifying the safety assertion of Fig. 1 requires proving the trace logic formula:

$$\begin{aligned} \forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times length)) \\ \rightarrow (c(main\_end, pos) \simeq a(j) \vee c(main\_end, pos) \simeq b(j)) \end{aligned} \tag{4}$$

For proving (4), it suffices to prove that the following, slightly modified statement is a loop invariant of Fig. 1:

$$\begin{aligned} \forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \rightarrow \forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times i(tp_{\mathtt{w}}(it)))) \\ \rightarrow (c(tp_{\mathtt{w}}(it), pos) \simeq a(j) \vee c(tp_{\mathtt{w}}(it), pos) \simeq b(j)) \end{aligned} \tag{5}$$

where `w` refers to the loop statement in Fig. 1. As part of the program semantics in trace logic, we have formula (6) which links the value of c at the end of the loop to its value at the end of the program. Moreover, using the trace lemma A, we also derive formula (7) in trace logic:

$$\forall pos_{\mathbb{I}}. c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), pos) \simeq c(main\_end, pos) \tag{6}$$

$$i(tp_{\mathtt{w}}(nl_{\mathtt{w}})) \simeq length \tag{7}$$

It is tempting to think that in the presence of these clauses (6)–(7), a saturation-based prover would rewrite the negated conjecture (4) to

$$\begin{aligned} \neg(\forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times i(tp_{\mathtt{w}}(nl_{\mathtt{w}})))) \\ \rightarrow (c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), pos) \simeq a(j) \vee c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), pos) \simeq b(j))) \end{aligned}$$

from which a bounded natural number induction inference (similar to the $\mathtt{IntInd}_{<}$ rule of [15]) would quickly introduce an induction hypothesis with (5) as the conclusion, by induction over $nl_{\mathtt{w}}$. However, this is not the case, as most saturation provers work by first *clausifying* their input. The negated conjecture (4) would not remain a single formula, but be split into the following clauses where sk is a Skolem symbol:

$$\begin{array}{cc} a(x) \not\simeq c(main\_end, sk) & b(x) \not\simeq c(main\_end, sk) \\ \neg(sk \leq 0) & sk \leq 2 \times length \end{array}$$

These clauses can be rewritten using (6)–(7). For example, the first clause can be rewritten to $a(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk))$. However, attempting to prove the negation of any of the rewritten clauses individually via induction would merely result in the addition of useless induction formulas to the search space. For example, attempting to prove $\forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \rightarrow (\exists x_{\mathbb{I}}.\, a(x) \simeq c(tp_{\mathtt{w}}(it), sk))$, is pointless as it is clearly false. *The solution we propose in this work is to use multi-clause induction*, whereby we attempt to prove the negation of the conjunction of multiple clauses via a single induction inference. For our running example Fig. 1, we can use the following rewritten versions of clauses from the negated conjecture $a(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk))$, $b(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk))$, and $sk \leq 2 \times i(tp_{\mathtt{w}}(nl_{\mathtt{w}}))$, with induction term $nl_{\mathtt{w}}$, to obtain the induction formula:

$$\begin{array}{lcl}
\begin{array}{l}
\neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(0)), sk) \\
\quad \wedge \;\; \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(0), sk)) \\
\quad \wedge \;\; sk \leq 2 \times i(tp_{\mathtt{w}}(0))\Big) \\
\wedge \;\; \mathit{StepCase}
\end{array}
& \rightarrow &
\begin{array}{l}
\forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \; \rightarrow \\
\quad \neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(it), sk) \\
\qquad \wedge \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(it), sk) \\
\qquad \wedge \;\; sk \leq 2 \times i(tp_{\mathtt{w}}(it))\Big)
\end{array}
\end{array} \tag{8}$$

where *StepCase* is the formula:

$$\begin{array}{lcl}
\begin{array}{l}
\forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \wedge \\
\neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(it)), sk) \\
\quad \wedge \;\; \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(it)), sk) \\
\quad \wedge \;\; sk \leq i(tp_{\mathtt{w}}(y))\Big)
\end{array}
& \rightarrow &
\begin{array}{l}
\neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(\mathtt{suc}(it))), sk) \\
\quad \wedge \, \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(\mathtt{suc}(it))), sk) \\
\quad \wedge \;\; sk \leq 2 \times i(tp_{\mathtt{w}}(\mathtt{suc}(it)))\Big)
\end{array}
\end{array}$$

Using the induction formula (8), a contradiction can then easily be derived, establishing validity of (4). In what follows, we formalize the multi-clause induction principle we used above. To this end, we introduce a generic inference rule, called *multi-clause goal induction* and denoted as `MCGLoopInd`.

$$\frac{C_1[nl_{\mathtt{w}}] \qquad C_2[nl_{\mathtt{w}}] \quad \dots \quad C_n[nl_{\mathtt{w}}]}{\mathrm{CNF}\left(\begin{array}{c}
\left(\begin{array}{c}
\neg(C_1[0] \;\wedge\; C_2[0] \wedge \ldots \wedge C_n[0]) \;\wedge \\
\forall it_{\mathbb{N}}.\, \left(\begin{array}{c} ((it < nl_{\mathtt{w}}) \wedge \neg(C_1[it] \;\wedge\; C_2[it] \wedge \ldots \wedge C_n[it])) \rightarrow \\ \neg(C_1[\mathtt{suc}(it)] \;\wedge\; C_2[\mathtt{suc}(it)] \wedge \ldots \wedge C_n[\mathtt{suc}(it)]) \end{array}\right)
\end{array}\right) \\
\rightarrow (\forall it_{\mathbb{N}}.\, (it < nl_{\mathtt{w}}) \rightarrow \neg(C_1[it] \;\wedge\; C_2[it] \wedge \ldots \wedge C_n[it]))
\end{array}\right)}$$

For performance reasons, we mandate that the premises $C_1 \ldots C_n$ be derived from trace logic formulas expressing safety assertions and not from formulas encoding the program semantics. The `MCGLoopInd` rule is formalised only as an induction inference over last loop iteration symbols. While restricting to $nl_{\mathtt{w}}$ terms is of purely heuristic nature, our experiments justify the necessity and usefulness of this condition (Sect. 8).

6 Array Mapping Induction for Lemmaless Induction

Multi-clause goal induction neatly captures goal-oriented application of induction. Nevertheless, there are verification challenges where `MCGLoopInd` fails to prove inductive loop properties. This is particularly the case for benchmarks

```
 func main(){
  const Int alength;
  Int[] a;
  Int i = 0;
  const Int n;

  while(i < alength){
   a[i] = a[i] + n;
   i = i + 1;
  }

  Int j = 0;
  while(j < alength){
   a[j] = a[j] - n;
   j = j + 1;
  }
 }
assert (∀pos𝕀.((0 ≤ pos < alength)
   → a(main_end, pos) = a(main_start, pos)))
```

Fig. 3. Adding and subtracting `n` to every element of array `a`.

containing multiple loops, such as in Fig. 3. We first discuss the limitations of `MCGLoopInd` using Fig. 3, after which we present our solution, the *array mapping induction* inference.

Let $\mathtt{w}_1$ be the first loop statement of Fig. 3 and $\mathtt{w}_2$ be the second loop. Using `MCGLoopInd`, we would attempt to prove

$$\begin{aligned}&\forall it_{\mathbb{N}}.\, it \leq nl_{\mathtt{w}_2} \rightarrow \\ &\quad \forall pos_{\mathbb{I}}.\, (0 \leq pos < j(tp_{\mathtt{w}_2}(it))) \rightarrow (a(tp_{\mathtt{w}_2}(it), pos) \simeq a(main\_start, pos)\end{aligned} \tag{9}$$

However, formula (9) is not a useful invariant for proving the assertion. Rather, for $\mathtt{w}_2$ we need a loop invariant similar to

$$\begin{aligned}\forall it_{\mathbb{N}}.\, it \leq nl_{\mathtt{w}_2} \rightarrow \forall pos_{\mathbb{I}}.\, (0 \leq pos < j(tp_{\mathtt{w}_2}(it))) \\ \rightarrow (a(tp_{\mathtt{w}_2}(it), pos) \simeq a(tp_{\mathtt{w}_2}(0), pos) - n\end{aligned} \tag{10}$$

and a similar loop invariant for loop $\mathtt{w}_1$. The loop invariant (10) is however not linked to the safety assertion of Fig. 3, and thus multi-clause goal induction is unable to infer and prove with it. To aid with the verification of benchmarks such as Fig. 3, we introduce another induction inference which we call *array mapping induction*. In this case, we trigger induction not on clauses and terms coming from the goal, but on clauses and terms appearing in the program semantics.

The *array mapping induction* inference rule, denoted as `AMLoopInd` is given below. Essentially, `AMLoopInd` involves analysing a clause set to heuristically devise a suitable loop invariant. Guessing a candidate loop invariant is a difficult problem. The `AMLoopInd` inference is triggered if clauses of the shapes of C_1 and

C_2 defined below are present in the clause set. Intuitively, C_2 can be read as saying that on each round of some loop $\mathtt{w}$, some array a at position i is set to some function F of its previous value at that position. Clause C_1 states that i increases by m in each round of the loop. Together the two clauses suggest that the loop is mapping the function F to each mth location of the array starting from the array cell located at $i(tp_{\mathtt{w}}(0))$. This is precisely what the induction formula attempts to prove. Note that for ease of notation, we present the inference for the case where the indexing variable is *increasing*. It is straightforward to generalise to the decreasing case. The `AMLoopInd` rule is[1]

$$\frac{\begin{array}{c} C_1 = i(tp_{\mathtt{w}}(\mathtt{suc}(x))) \simeq i(tp_{\mathtt{w}}(x)) + m \ \vee\ \neg(x < nl_{\mathtt{w}}) \\ C_2 = a\big(tp_{\mathtt{w}}(\mathtt{suc}(x)), i(tp_{\mathtt{w}}(x))\big) \simeq F[a\big(tp_{\mathtt{w}}(x), i(tp_{\mathtt{w}}(x))\big)] \ \vee\ \neg(x < nl_{\mathtt{w}}) \end{array}}{\mathrm{CNF}(\mathit{StepCase} \rightarrow \mathit{Conclusion})}$$

where $\mathtt{w}$ is some loop and F an arbitrary non-empty context. Let i_0 be an abbreviation for $i(tp_{\mathtt{w}}(0))$. Then:

$$\begin{aligned}
\mathit{StepCase}: \quad & \forall it_{\mathbb{N}}.\ \big(\forall y_{\mathbb{I}}.\ it < nl_{\mathtt{w}} \wedge \\
& \qquad y < i(tp_{\mathtt{w}}(it)) - i_0 \wedge y \geq 0 \wedge y \ \mathtt{mod}\ m = 0 \\
& \qquad\qquad \rightarrow a(tp_{\mathtt{w}}(it), i_0 + y) \simeq F[a(tp_{\mathtt{w}}(0), i_0 + y)]\big) \rightarrow \\
& (\forall y_{\mathbb{I}}.\ y < i(tp_{\mathtt{w}}(\mathtt{suc}(it))) - i_0 \wedge y \geq 0 \wedge y \ \mathtt{mod}\ m = 0 \\
& \qquad\qquad \rightarrow a(tp_{\mathtt{w}}(\mathtt{suc}(it)), i_0 + y) \simeq F[a(tp_{\mathtt{w}}(0), i_0 + y)]) \\
\mathit{Conclusion}: \quad & \forall x_{\mathbb{I}}.\ x < i(tp_{\mathtt{w}}(nl_{\mathtt{w}})) - i_0 \wedge x \geq 0 \wedge x \ \mathtt{mod}\ m = 0 \\
& \qquad\qquad \rightarrow a(tp_{\mathtt{w}}(nl_{\mathtt{w}}), i_0 + x) \simeq F[a(tp_{\mathtt{w}}(0), i_0 + x)]
\end{aligned}$$

To prove *StepCase*, it is necessary to be able to reason that positions in the array a remain unchanged until visited by the indexing variable. This can be achieved via the addition of another induction to the conclusion of the inference. We do not provide details of this induction formula here, but it is added to the conclusion by our implementation which we present in Sect. 7. The `AMLoopInd` inference is thus sufficient to prove the assertion of Fig. 3. While `AMLoopInd` is a limited approach for guessing inductive loop invariants, we believe it can be extended towards further, more generic methods to guess invariants, as discussed in Sect. 9. We conclude this section by noting that our induction rules are sound, based on trace logic semantics. Since both rules merely add instances of the bounded induction schema for natural numbers (3) to the search space, soundness is trivial and we do not provide a proof.

7 Implementation

Our approach is implemented as an extension of the Rapid framework, using the first-order theorem prover Vampire.

[1] In the conclusion we ignore the base case of the induction formula as it is trivially true.

***Extensions to* Rapid.** Rapid takes as an input a $\mathcal{W}$ program along with a property expressed in $\mathcal{L}$. It outputs the semantics of the program expressed in $\mathcal{L}$ using SMT-LIB syntax along with the property to be proven. For our "lemmaless induction" framework, we have extended Rapid as follows. Firstly, we prevent the output of all trace lemmas other than trace lemma A (Sect. 4.5). We added custom extensions to the SMT-LIB language to identify trace logic symbols, such as loop iteration symbols, program variables, within the Rapid encodings. This way, trace logic symbols to be used for induction inferences are easily identified and can also be used for various proving heuristics. We refer to this version (available online[2]) as Rapid^{l-}.

***Extensions to* Vampire.** We implemented the `MCGLoopInd` inference rule and a slightly simplified version of the `AMLoopInd` rule in a new branch of Vampire[3]. The main issue with the induction inferences `MCGLoopInd` and `AMLoopInd` is their explosiveness which can cause proof search to diverge. We have, therefore, introduced various heuristics in the implementation to try and control them. For `MCGLoopInd` we not only necessitate that the premises are derived from the conjecture, but that their derivation length from the conjecture is below a certain distance controlled by an option. The premises must be unit clauses unless another option `multi_literal_ clauses` is toggled on. The option `induct_all_loop_counts` allows `MCGLoopInd` induction to take place on all loop counter terms, not just final loop iterators. In order for the `MCGLoopInd` and `AMLoopInd` inferences to be applicable, we need to rewrite terms not containing final loop counters to terms that do. However, rewriting in Vampire is based on superposition, which is parameterised by a term order preventing smaller terms to be rewritten into larger ones. In this case, the term order may work against us and prevent such rewrites from happening. We implemented a number of heuristics to handle this problem. One such heuristic is to give terms representing constant program variables a large weight in the ordering. Then, equations such as $alength \simeq i(tp_{\mathtt{w}}(nl_{\mathtt{w}}))$ will be oriented left to right as desired. We combined these options with others to form a portfolio of strategies[4] that contains 13 strategies each of which runs in under 10s.

8 Experimental Results

Benchmarks. For our experiments, we use a total of 111 examples whose verification involved proving safety assertions of different logical complexity (quantifier-free, only universally/existentially quantified, and with quantifier alternations). Our benchmarks are divided into four groups, as indicated in Table 1: (i) the first 13 problems have quantifier-free proof obligations; (ii) the majority of benchmarks, in total 68 examples, contain universally quantified

[2] See commit `285e54b7e` of https://github.com/vprover/rapid/tree/ahmed-induction-support.

[3] See commit `4a0f319f` of https://github.com/vprover/vampire/tree/ahmed-rapid.

[4] `--mode portfolio --schedule rapid_induction.`.

Table 1. Experimental results.

Benchmark	(1)	(2)	(3)	(4)
atleast_one_iteration_0	✓	✓	✓	✓
atleast_one_iteration_1	✓	✓	✓	✓
count_down	✓	-	-	-
eq	✓	-	✓	-
find_sentinel	✓	✓	-	-
find1_0	✓	✓	✓	-
find1_1	✓	✓	✓	-
find2_0	✓	✓	✓	-
find2_1	✓	✓	✓	-
indexn_is_arraylength_0	✓	✓	✓	-
indexn_is_arraylength_1	✓	✓	✓	-
set_to_one	✓	✓	✓	✓
str_cpy_3	✓	✓	✓	-
add_and_subtract	✓	-	-	✓
both_or_none	✓	✓	-	✓
check_equal_set_flag_1	✓	✓	-	✓
collect_indices_eq_val_0	✓	✓	-	✓
collect_indices_eq_val_1	✓	✓	-	-
copy	✓	✓	-	✓
copy_absolute_0	✓	✓	-	✓
copy_absolute_1	✓	✓	-	✓
copy_and_add	✓	-	-	✓
copy_nonzero_0	✓	✓	-	✓
copy_partial	✓	✓	-	✓
copy_positive_0	✓	✓	-	✓
copy_two_indices	✓	✓	-	-
find_max_0	✓	✓	-	✓
find_max_2	✓	✓	-	✓
find_max_from_second_0	✓	-	-	✓
find_max_local_2	-	-	-	-
find_max_up_to_0	-	-	-	-
find_max_up_to_2	-	-	-	-
find_min_0	✓	✓	-	✓
find_min_2	✓	✓	-	-
find_min_local_2	-	-	-	-
find_min_up_to_0	-	-	-	-
find_min_up_to_2	-	-	-	-
find1_4	-	✓	-	-
find2_4	✓	✓	-	-
in_place_max	✓	✓	-	✓
inc_by_one_0	✓	✓	-	✓
inc_by_one_1	✓	✓	-	✓
inc_by_one_harder_0	✓	✓	-	✓
inc_by_one_harder_1	✓	✓	-	✓
init	✓	✓	-	-
init_conditionally_0	✓	✓	-	-
init_conditionally_1	✓	✓	-	✓
init_non_constant_0	✓	✓	-	-
init_non_constant_1	✓	✓	-	✓
init_non_constant_2	✓	✓	-	✓
init_non_constant_3	✓	✓	-	✓
init_non_constant_easy_0	✓	✓	-	-
init_non_constant_easy_1	✓	✓	-	✓
init_non_constant_easy_2	✓	✓	-	✓
init_non_constant_easy_3	✓	✓	-	✓
init_partial	✓	✓	-	✓
init_prev_plus_one_0	✓	✓	-	-
init_prev_plus_one_1	✓	✓	-	-
init_prev_plus_one_alt_0	✓	✓	-	-
init_prev_plus_one_alt_1	✓	✓	-	-
insertion_sort	-	-	-	-
max_prop_0	✓	✓	-	✓
max_prop_1	✓	✓	-	✓
merge_interleave_0	✓	-	-	✓
merge_interleave_1	✓	-	-	✓
min_prop_0	✓	✓	-	✓
min_prop_1	✓	✓	-	✓
partition_0	✓	✓	-	✓
partition_1	✓	✓	-	✓
push_back	✓	✓	-	✓
reverse	✓	✓	-	-
rewnifrev	✓	-	-	✓
rewrev	✓	-	-	✓
skipped	✓	-	-	✓
str_cpy_0	✓	✓	-	-
str_cpy_1	✓	✓	-	-
str_cpy_2	✓	✓	-	-
swap_0	-	✓	✓	✓
swap_1	-	✓	✓	✓
vector_addition	✓	✓	-	✓
vector_subtraction	✓	✓	-	✓
check_equal_set_flag_0	✓	✓	-	-
find_max_1	-	-	-	-
find_max_from_second_1	✓	-	-	-
find1_2	✓	✓	-	-
find1_3	✓	✓	-	-
find2_2	✓	✓	-	-
find2_3	✓	✓	-	-
collect_indices_eq_val_2	-	✓	-	-
collect_indices_eq_val_3	✓	-	-	-
copy_nonzero_1	✓	✓	-	-
copy_positive_1	✓	✓	-	-
find_max_local_0	-	-	-	-
find_max_local_1	✓	-	-	-
find_max_up_to_1	-	-	-	-
find_min_1	-	-	-	-
find_min_local_0	-	-	-	-
find_min_local_1	✓	-	-	-
find_min_up_to_1	-	-	-	-
merge_interleave_2	✓	-	-	-
partition_2	✓	✓	-	-
partition_3	✓	✓	-	-
partition_4	-	-	-	-
partition_5	-	✓	-	-
partition_6	-	-	-	-
partition-harder_0	✓	✓	-	-
partition-harder_1	✓	✓	-	-
partition-harder_2	✓	-	-	-
partition-harder_3	✓	-	-	-
partition-harder_4	✓	-	-	-
str_len	✓	✓	-	-
Total solved	**93**	**78**	**13**	**47**

safety assertions; (iii) 7 problems come with the task of verifying existentially quantified assertions; (iv) and the last 23 programs contain assertions with alternation of quantifiers. The examples from (i)-(ii), a total of 81 programs, come from the array verification benchmarks of SV-COMP repository [1], with most of these examples originating from [9,13].[5] These examples correspond to the set of those SV-COMP benchmarks which use the C fragment supported by Rapid; specifically, when selecting examples (i)-(ii) from SV-COMP, we omitted examples containing pointers or memory management. All SV-COMP examples from (i)-(ii) are adapted to our input format, as for example arrays in trace logic are treated as unbounded data structures. Further, the examples (iii)-(iv) are new examples crafted by us, in total 30 new examples. They contain existential and alternating quantification in safety assertions. We intend to submit these 30 examples from (iii)-(iv) to SV-COMP.

Experimental Setting. We used two versions of Rapid in our experiments. First, `(1)` Rapid^{l-} denotes our Rapid approach, using lemmaless induction `MCGLoopInd` and `AMLoopInd` in Vampire. Further, `(2) Rapid`$^{l+}$ uses trace lemmas for inductive reasoning, as described in [11]. We also compared Rapid^{l-} with other verification tools. In particular, we considered `(3)` SeaHorn and `(4)` Vajra (and its extension Diffy that produced for us exactly the same results as Vajra). SeaHorn converts the program into a constrained horn clause (CHC) problem and uses the SMT solver Z3 for solving. Vajra and Diffy implement inductive reasoning and recurrence solving over loop counters; in the background, they also use Z3.

Rapid ***Experiments.*** Table 1 shows that Rapid^{l-} is superior to Rapid^{l+}, as it solves a total of 93 problems, while Rapid^{l+} only proved 78 assertions correct. Particularly, Rapid^{l-} can solve benchmark `merge_interleave_2` corresponding to our motivating example 1, and other challenging problems such as `find_max_local_1` also containing quantifier alternations.

While Rapid^{l-} can solve a total of ten problems more than Rapid^{l+}, it is interesting to look into which problems can now be solved. Many of the newly solved problems are structurally very close to the loop invariants needed to prove them. This is where multi-clause goal-oriented induction `MCGoalInd` makes the biggest impact. For instance, this allows Rapid^{l-} to prove the partial correctness of `find_max_ from_second_0` and `find_max_from_second_1`.

On the other hand, Rapid^{l-} also lost two challenging benchmarks that were previously solved by Rapid^{l+}, namely `swap_0` and `partition_5`. This could be for two reasons: (1) the strategies in the induction schedule of Rapid^{l-} are too restrictive for such benchmarks, or (2) the step case of the induction axiom introduced by our two rules are too difficult for Vampire to prove. Strengthening lemmaless induction with additional trace lemmas from Rapid^{l+} is an interesting line of further work.

[5] Artifact evaluation: in order to reproduce the results reported in this section, please follow the instructions at https://github.com/vprover/vampire_publications/tree/master/experimental_data/CICM-2022-RAPID-INDUCTION.

Comparing with other tools. Both, SeaHorn and Vajra/Diffy require C code as input, whereas Rapid uses its own syntax. We translated our benchmarks to C code expressing the same problem. However, a direct comparison of Rapid, and in particular Rapid^{l-}, with most other verifiers requiring standard C code as an input is not possible as we consider slightly different semantics. In contrast to SeaHorn and Vajra/Diffy, we assume that integers and arrays are unbounded and that all array positions are initialized by arbitrary data. Further, we can read/write at any array position without allocating the accessed memory beforehand. Apart from semantic differences, Rapid can directly express assertions and assumptions containing quantifiers and put variable contents from different points in time into relation. In order to deal with the latter, we introduced history variables in the code provided to SeaHorn and Vajra/Diffy. Quantification was simulated by non-deterministically assigned variables and by loops. As a result, SeaHorn verified 13 examples, whereas Vajra/Diffy 47 of our benchmarks. As Vajra/Diffy restrict their input programs to contain only loops having very specific loop-conditions, several of our benchmarks failed. For example, $i < length$ is permitted, whereas $a[i] \neq 0$ is not. Vajra/Diffy could prove correctness for nearly all the programs satisfying these restrictions. SeaHorn, on the other hand, has problems with the complexity introduced by the arrays. It could solve especially those benchmarks whose correctness do not depend on the arrays' content.

9 Future Directions and Conclusion

We introduced lemmaless induction to fully automate the verification of inductive properties of program loops with unbounded arrays and integers. We introduced goal-oriented and array mapping induction inferences, triggered by loop counters, in superposition-based theorem proving. Our results show that lemmaless induction in trace logic outperforms other state-of-the-art approaches in the area. There are various ways to further develop lemmaless induction in trace logic. On larger benchmarks, particularly those containing multiple loops, our approach struggles. For loops where the required invariant is not connected to the conjecture, we introduced array mapping induction. However, the array mapping induction inference is limited in the form of invariants it can generate. We would like to investigate other methods, such as machine learning for synthesising loop invariants that are not too prolific. A completely different line of research that we are currently working on, is updating the trace logic syntax and semantics of $\mathcal{W}$ to deal with memory and memory allocation, aiming to efficiently reason about loop operations over the memory.

As shown in [18], the validity problem for first-order formulas of linear arithmetic extended with non-theory function symbols is Π_1^1-complete. Therefore, we do not expect any completeness result for inductive theorem proving. Proving relative completeness results for our verification framework is an interesting question.

Acknowledgements. This research was partially supported by the ERC consolidator grant ARTIST 101002685, the FWF research project LogiCS W1255-N23, the TU Wien SecInt doctoral program, and the EUProofNet Cost Action CA20111. Our research was partially funded by the Digital Security by Design (DSbD) Programme delivered by UKRI to support the DSbD ecosystem.

References

1. SV-comp repository. https://gitlab.com/sosy-lab/benchmarking/sv-benchmarks
2. Vampire website. https://vprover.github.io/
3. Bachmair, L., Ganzinger, H.: Resolution theorem proving. In: Robinson, A., Voronkov, A. (Eds.) Handbook of Automated Reasoning, vol. I, chap. 2, pp. 19–99. Elsevier Science (2001)
4. Barrett, C., et al.: CVC4. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 171–177. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22110-1_14
5. Chakraborty, S., Gupta, A., Unadkat, D.: Verifying array manipulating programs with full-program induction. In: TACAS 2020. LNCS, vol. 12078, pp. 22–39. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45190-5_2
6. Chakraborty, S., Gupta, A., Unadkat, D.: Diffy: inductive reasoning of array programs using difference invariants. In: Silva, A., Leino, K.R.M. (eds.) CAV 2021. LNCS, vol. 12760, pp. 911–935. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-81688-9_42
7. Claessen, K., Johansson, M., Rosén, D., Smallbone, N.: Automating inductive proofs using theory exploration. In: Bonacina, M.P. (ed.) CADE 2013. LNCS (LNAI), vol. 7898, pp. 392–406. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38574-2_27
8. de Moura, L., Bjørner, N.: Z3: an efficient SMT solver. In: Ramakrishnan, C.R., Rehof, J. (eds.) TACAS 2008. LNCS, vol. 4963, pp. 337–340. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78800-3_24
9. Dillig, I., Dillig, T., Aiken, A.: Fluid updates: beyond strong vs. weak updates. In: Gordon, A.D. (ed.) ESOP 2010. LNCS, vol. 6012, pp. 246–266. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-11957-6_14
10. Fedyukovich, G., Prabhu, S., Madhukar, K., Gupta, A.: Quantified invariants via syntax-guided synthesis. In: Dillig, I., Tasiran, S. (eds.) CAV 2019. LNCS, vol. 11561, pp. 259–277. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25540-4_14
11. Georgiou, P., Gleiss, B., Kovács, L.: Trace logic for inductive loop reasoning. In: 2020 Formal Methods in Computer Aided Design (FMCAD), pp. 255–263. IEEE (2020)
12. Gurfinkel, A., Kahsai, T., Komuravelli, A., Navas, J.A.: The SeaHorn verification framework. In: Kroening, D., Păsăreanu, C.S. (eds.) CAV 2015. LNCS, vol. 9206, pp. 343–361. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21690-4_20
13. Gurfinkel, A., Shoham, S., Vizel, Y.: Quantifiers on demand. In: Lahiri, S.K., Wang, C. (eds.) ATVA 2018. LNCS, vol. 11138, pp. 248–266. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01090-4_15
14. Hajdú, M., Hozzová, P., Kovács, L., Schoisswohl, J., Voronkov, A.: Induction with generalization in superposition reasoning. In: Benzmüller, C., Miller, B. (eds.)

CICM 2020. LNCS (LNAI), vol. 12236, pp. 123–137. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-53518-6_8

15. Hozzová, P., Kovács, L., Voronkov, A.: Integer induction in saturation. In: Platzer, A., Sutcliffe, G. (eds.) CADE 2021. LNCS (LNAI), vol. 12699, pp. 361–377. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79876-5_21
16. Karpenkov, E.G., Monniaux, D.: Formula slicing: inductive invariants from preconditions. In: Bloem, R., Arbel, E. (eds.) HVC 2016. LNCS, vol. 10028, pp. 169–185. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49052-6_11
17. Kaufmann, M., Moore, J.S.: An industrial strength theorem prover for a logic based on common Lisp. In: IEEE Transactions on Software Engineering, pp. 203–213 (1997)
18. Korovin, K., Voronkov, A.: Integrating linear arithmetic into superposition calculus. In: Duparc, J., Henzinger, T.A. (eds.) CSL 2007. LNCS, vol. 4646, pp. 223–237. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74915-8_19
19. Kovács, L., Robillard, S., Voronkov, A.: Coming to terms with quantified reasoning. In: POPL, pp. 260–270 (2017)
20. Kovács, L., Voronkov, A.: First-order theorem proving and Vampire. In: Sharygina, N., Veith, H. (eds.) CAV 2013. LNCS, vol. 8044, pp. 1–35. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39799-8_1
21. Larraz, D., Rodríguez-Carbonell, E., Rubio, A.: SMT-based array invariant generation. In: Giacobazzi, R., Berdine, J., Mastroeni, I. (eds.) VMCAI 2013. LNCS, vol. 7737, pp. 169–188. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-35873-9_12
22. Leino, K.R.M.: Dafny: an automatic program verifier for functional correctness. In: Clarke, E.M., Voronkov, A. (eds.) LPAR 2010. LNCS (LNAI), vol. 6355, pp. 348–370. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17511-4_20
23. Rajkhowa, P., Lin, F.: Extending VIAP to handle array programs. In: Piskac, R., Rümmer, P. (eds.) VSTTE 2018. LNCS, vol. 11294, pp. 38–49. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03592-1_3
24. Bjoner, N., Reger, G., Suda, M., Voronkov, A.: AVATAR modulo theories. In: GCAI, pp. 39–52 (2016)
25. Reger, G., Schoisswohl, J., Voronkov, A.: Making theory reasoning simpler. In: TACAS 2021. LNCS, vol. 12652, pp. 164–180. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72013-1_9
26. Srivastava, S., Gulwani, S.: Program verification using templates over predicate abstraction. In: PLDI, pp. 223–234 (2009)