

Relational Symbolic Execution

Gian Pietro Farina
University at Buffalo, SUNY
gianpiet@buffalo.edu

Stephen Chong
Harvard University
chong@seas.harvard.edu

Marco Gaboardi
University at Buffalo, SUNY
gaboardi@buffalo.edu

ABSTRACT

Symbolic execution is a classical program analysis technique used to show that programs satisfy or violate given specifications. In this work we generalize symbolic execution to support program analysis for relational specifications in the form of relational properties - these are properties about two runs of two programs on related inputs, or about two executions of a single program on related inputs. Relational properties are useful to formalize notions in security and privacy, and to reason about program optimizations. We design a relational symbolic execution engine, named RelSym which supports interactive refutation, as well as proving of relational properties for programs written in a language with arrays and for-like loops.

1 INTRODUCTION

Relational properties capture the relations between the behavior of two programs when run on two inputs, and as a special case the behavior of one program on two different inputs. Several safety and security properties can be described as relational properties: non-interference [Goguen and Meseguer 1982, 1984], compiler optimizations [Benton 2004], sensitivity and continuity analysis [Chaudhuri et al. 2010, 2012; Reed and Pierce 2010], and relative cost [Çiçek et al. 2017] are just some examples.

In order to *prove* a relational property, one must ensure that all the *pairs of related executions* satisfy it, instead of just single executions. Similarly, for *finding violations* to relational properties, we need to find *pairs of related executions* that violate the property. A natural way to approach the verification and the testing of relational properties is through their reduction to standard (unary) properties through ideas like self-composition [Barthe et al. 2004; Butler and Schulte 2011; Terauchi and Aiken 2005] and product programs [Barthe et al. 2011; Eilers et al. 2018]. This approach permits to use standard program verification and bug-finding techniques [Hritcu et al. 2013; Milushev et al. 2012], and to reduce the problem to designing convenient and efficient self-compositions and product programs.

Another way to approach the verification and testing of relational properties is through relational extensions of standard, non-relational, techniques for these tasks. Several works have explored this approach for techniques such as type systems [Barthe et al.

2014b, 2015; Nanevski et al. 2013; Pottier and Simonet 2003], program logics [Barthe et al. 2012; Benton 2004; Sousa and Dillig 2016], program analysis [Kwon et al. 2017], and abstract interpretation [Assaf et al. 2017; Feret 2001; Giacobazzi and Mastroeni 2004]. In this approach, one often aims at giving the user the choice on *how to explore* the use of the *relational assumptions*, (i.e., relational preconditions, relational intermediate assumptions, and relational invariants) and a way to relate two programs in order to prove relational properties. Relational assumptions have a different flavor than non-relational ones, since they permit to consider only a subset of the product-relation between inputs, and so only a subset of the pairs of execution of a program. These are often the key ingredients for reasoning in a natural way about relational properties. In this paper, we follow this approach and we propose *relational symbolic execution* (RelSym): a foundational technique combining the idea of relational analysis of programs and symbolic execution.

RelSym is a relational symbolic execution engine for a language with arrays and for-loops. The target applications we have in mind are data analysis and statistics, so we focused on a core calculus which constitute the basis of languages like R [R Core Team 2013]. In fact, the design of RelSym was at an early stage informed by the work in [Morandat et al. 2012], on the subset of that language: Core R. For-loops and arrays provide interesting challenges to both the design of the operational semantics and to the representation of the different execution paths in constraints.

RelSym combines both proving and interactive refutation of relational properties, with the option of providing loop invariants to effectively prove or refute properties of programs containing loops. RelSym is built on a hierarchy of four languages (two relational and two unary — two concrete and two symbolic) whose operational semantics are built on each other in a well-founded manner. In particular, the two relational languages are based on their unary versions and the two symbolic languages are, as it usually happens in symbolic execution, the symbolic versions (i.e., extended with symbolic values) of the concrete ones. The symbolic operational semantics collect constraints about the execution of a program, or about pairs of executions of programs, that can be used to prove or refute relational properties. This gives the user the ability to experiment with different ways of proving and interactively refuting relational properties, e.g., both using a single symbolic relational execution or using a pair of unary symbolic executions.

We implemented RelSym as a prototype, and we used it for experimenting with different examples of interactive refutation and verification for several relational properties coming from different domains. The range of properties and examples we considered show the flexibility and the feasibility of our approach.

We also compare RelSym with other non-relational methods such as self-composition and product programs (which can also be defined using our tool) in their basic form with no optimization. We find that our approach, thanks to the use of relational assumptions,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PPDP '19, October 7–9, 2019, Porto, Portugal

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7249-7/19/10...\$15.00
<https://doi.org/10.1145/3354166.3354175>

improves in efficiency with respect to self-composition. Product programs give verification conditions that are often comparable to the one obtained using relational methods, and they can use standard symbolic execution tools, but a challenge in using this technique is the additional cost, in term of design, in building the product—even if recent developments considerably eased this task, e.g., [Eilers et al. 2018]. In relational symbolic execution, we do not need any pre-processing and we can directly analyze a program in a relational way. This shows a trade-off between the different techniques which can be exploited accordingly to the concrete target application. At the current stage, RelSym users need to provide invariants for loops with symbolic guards. We envision for the future to combine our approach with invariant synthesis techniques, especially relational ones, e.g., [Chen et al. 2017, 2011; Qin et al. 2013; Sigurbjarnarson et al. 2018].

Summarizing, the main contributions of our work are:

- The design of a relational symbolic execution technique, RelSym, for a language containing for-loops and arrays. This technique is based on relational and unary symbolic operational semantics that permits to explore the different execution paths of programs, maintaining constraints about pairs of executions that can be used to prove or refute relational properties.
- The extension of relational symbolic execution to support relational and unary invariants to completely explore a loop with symbolic guards.
- We have implemented RelSym in a prototype. The implementation uses an SMT solver to discharge the generated constraints. We show the effectiveness of our approach by analyzing several examples for different relational properties.

Outline. The paper is structured in the following way: in Section 2 we introduce the different design choices behind RelSym in an informal way. Using four running examples Section 3 shows at an high level how RelSym works and how relational assumptions help in cutting the search space for proofs and refutation witnesses. In Section 4, 5, and 6 we provide the main technical material describing the four languages behind RelSym and the meta theoretical results that connect them. Section 7 provides some details about the RelSym implementation. In Section 8 we provide an experimental comparison of the relational symbolic approach with other standard techniques for the verification and bug finding of relational properties such as self-composition and product programs. Finally, in Section 9 we discuss related works and in Section 10 we conclude by providing a summary of this work.

2 RELATIONAL SYMBOLIC EXECUTION: INFORMALLY

In this section, we will give an high-level introduction to the main characteristics of RelSym.

Relational semantics. RelSym is based on a relational operational semantics, which describes the execution of two, potentially different, programs in two, potentially different, memories. In this semantics a memory e.g., $\mathcal{M}$ can map a variable e.g., x , either to a single value, for instance $\mathcal{M}(x) = 4$, or to a pair of values, for

instance $\mathcal{M}(x) = (3, 4)$. In the first case, we know that in the two executions x will take the same value 4. In the second case, x will take two different values in the two executions that is 3 and 4. In assertions, when we refer to one of the two executions of the program we use indexed objects. For instance by writing x_1 we mean the variable x interpreted in the first (left) execution. When we instead have a precondition that implies that the variable has the same value in both run we will just avoid indexes and write, for example, just x . The relational character of memories is extended also to the operational semantics of commands and expressions thanks to a pairing construct $\langle \cdot \mid \cdot \rangle$. In the spirit of [Pottier and Simonet 2003], with $\langle c_1 \mid c_2 \rangle$ we denote a pair of commands that might differ in two runs. These are needed, for instance, when the guard e , of a conditional *if* e *then* c_1 *else* c_2 , evaluates to different values in the two executions, and so the two executions need to take different branches. For instance, when evaluating *if* e *then* c_1 *else* c_2 , if e evaluates to $\langle 1 \mid 0 \rangle$, the first execution needs to evaluate c_1 , while the second one needs to evaluate c_2 . This situation is resolved by using the command pair $\langle c_1 \mid c_2 \rangle$. To relationally execute a paired command $\langle c_1 \mid c_2 \rangle$ we execute both c_1 and c_2 in a unary fashion on two different memories independently and when they both terminate we merge the two final unary memories in one final relational memory.

Symbolic semantics. To enable symbolic execution, the RelSym engine also supports symbolic values $X, Y \dots$. As in standard symbolic execution, a symbolic value X represents a set of possible concrete values. However, in relational symbolic execution, symbolic values can appear also in pairs $\langle X \mid Y \rangle$. During the computation, symbolic values are refined through constraints coming from pre and postconditions, invariants, and conditionals. At each step, the constraints describe all the possible concrete values that symbolic values, and pairs of symbolic values, can assume. As a simple example, consider symbolic execution of the program *if* $x = 0$ *then* c_1 *else* c_2 starting with a memory $\mathcal{M}$ where $\mathcal{M}(x) = X$. Note that the symbolic value X represents an arbitrary concrete value, but the value is the same for both executions. Symbolic execution of the program would follow both the first branch (collecting the constraint $X = 0$) and the second branch (collecting the constraint $X \neq 0$). The two constraints restrict the set of concrete values that X can represent in the two branches, respectively. Consider instead executing the same program but with an initial memory where $\mathcal{M}(x) = \langle X_1 \mid X_2 \rangle$. Here, the two executions map the variable x to different symbolic values, meaning that the value of variable x may differ in the two executions. Symbolic execution of the conditional would generate four possible configurations, based on all possible combinations of the left and right executions taking the true and false branches. Using relational assumptions, we can cut the space of the branches to explore and still get an analysis relational in nature that allows us to exploit the naturality of this approach instead of reducing it to a unary approach.

Relational ghost variables. We will make use of (relational) ghost variables [Hofmann and Pavlova 2008] to annotate programs or to give specifications for them. Ghost variables are variables that don't correspond to real program entities but appear only in the specification of a program. For instance when we will reason about

relational cost we will use a relational variable γ which counts the cost of the two runs. Other ghost variables can be used to reason about other properties for instance covert channels or trace equivalence. The operational semantics of the languages does not cover ghost variables by itself, but it can easily be extended by adding conditions to the rule describing how they evolve during the computation. For instance when reasoning about cost we can select a (potentially proper) subset of rules of the semantics which cover the cost model we have in mind, and extend them with conditions describing how γ evolves. For simplicity in Section 3 we will measure the cost of a program by the number of assignments it performs.

Proving relational specifications. Throughout the whole paper we will use (relational) Hoare triples to denote specifications of programs. That is, we will say that a program satisfies (or doesn't) the triple $\{\Phi\}c\{\Psi\}$. Symbolic execution can be used to prove valid specifications. In general, if starting from a symbolic initial state that satisfies a precondition Φ we execute (relationally and) symbolically a program c and we only reach final states where the path constraints imply the postcondition Ψ we know that the triple $\{\Phi\}c\{\Psi\}$ is valid.

Interactive refutation and counterexample generation. The dual way of reasoning is what symbolic execution is mostly used for. Symbolic execution searches for final states whose associated path constraints don't imply the postcondition desired, if they are found it means that there is at least one state where the desired postcondition might not hold. Symbolic execution has been proved useful to generate concrete test cases that demonstrate violation of specifications. This is usually done by using constraint solvers to find substitutions for symbolic values that satisfy at the same time the negation of the postcondition on the final states (the violation of the specification) and some *path condition* (i.e., constraints over symbolic values based on the control flow of the symbolic execution) guaranteeing the reachability of the violation. RelSym can be used in the same way to find violations of relational properties.

Loops. Traditionally, symbolic execution has been used more for bug finding and testing [Khurshid et al. 2003; King 1976] than for proving. One of the reasons for this is that conditionals and loops may create state explosion, and long (possibly infinite) traces of configurations. To improve this situation we extend relational symbolic execution with loop invariants [Hentschel et al. 2014] so that the symbolic execution of a loop can be performed by *jumping* over the loop in one step and by adding an invariant to the path condition. We design two rules for unary and relational invariants which allow one to reason in one step about loops both for proving and for finding counterexamples. We will see in Section 3 that using an invariant allows us to reason about arrays with symbolic length, proving in this way this program satisfies a relational property (Lipschitz continuity) for arrays of arbitrary length. When searching for counterexamples, the situation is a bit more delicate. Indeed, just providing an inductive invariant may lead to *unrealizable* counterexamples: satisfiable substitutions that are not produced by any concrete execution. This can happen when the invariants do not determine precisely enough the state that can be reached after the

loop. To avoid this situation in subsection 6.3 we formalize a notion of *strength* of an invariant. RelSym uses this notion to check whether the invariant provided is strong enough ensuring that if a counterexample is found, then indeed it corresponds to a concrete execution (or a pair of concrete executions) violating the (possibly relational) specification of the program. Using loop with invariants mitigates in part the state explosion problem but it does not solve it entirely. A lot of research has focused and still focuses on taming the state explosion in traditional symbolic execution. These techniques can also be used for relational symbolic execution in order to tame this complexity. Since RelSym is intended as a foundational work we won't concern ourselves here with integrating the framework with standard techniques for reducing space explosion, or loop invariant synthesization, as our goal is to present a different approach to the verification and interactive refutation of relational properties.

Comparison with self-composition and product programs. We already discussed how self-composition and product programs are standard approaches which reduce relational properties to unary properties, and which allow one to use standard program verification and bug-finding techniques. At the design level, we do not propose our approach in contrast with these techniques but as an alternative. Indeed, one can use RelSym also as a standard symbolic execution engine and use these techniques as a pre-processing phase transforming the program in its self-composition or product program. However, we believe that at the technical level, relational symbolic execution offers, in several situations, some key advantages that permit to maximize the relational reasoning. Indeed in the next section we will see that we don't need to reason about the functional correctness of the programs, to prove or disprove (even though to effectively find counterexamples strong invariants involving a functional description might be needed) relational properties. This property is very useful in relational reasoning since it does allow one to reduce the complexity of the constraints that one need to consider. Self-composition cannot directly support this for example for arrays with symbolic length, while product programs can support it but it requires more complex invariants than in the case of relational symbolic execution. To understand better this kind of trade-offs we perform an experimental evaluation comparing RelSym with self-composition and product programs in Section 8.

3 EXAMPLES

In this section we present a few examples for proving and disproving relational properties of programs. We will hide many details in order to not distract the reader from the main point of the section which is to provide a general understanding of the way RelSym works. For example, in the following we use assertions and constraints interchangeably but later on (i.e., Section 4 and 5) they will be treated differently.

Proving anti-monotonicity of the inverse of cumulative distribution function (c.d.f) - concrete bounds. As a first motivating example we consider the program in Figure 1. The program takes in input a real number $q \in [0, 1]$ and an array d of size $k \geq 1$ such that $\forall i. 1 \leq i \leq k. d[i] = P[X \leq i]$, where X is some unspecified random variable. That is, d represents the c.d.f of a random variable X whose

```

1)  cum ← 0;
2)  x ← 0;
3)  for(i in 1 : len(cdf)) do
4)    if(cum ≥ q) then
5)      if(x ≤ 0) then
6)        x ← i
7)      else
8)        cum ← cum + cdf[i]

```

Figure 1: Let CDF the set of c.d.f.s. The program implements $F_q^{-1} : \text{CDF} \rightarrow \mathbb{R}$. F_q^{-1} is monotonically decreasing where the order $\leq_{cdf}$ on CDF, encoded in finite arrays, is defined as: $d_1 \leq_{cdf} d_2 \iff \forall x. d_1[x] \leq d_2[x]$ and we consider the standard order on $\mathbb{R}$.

realizations lie in the set $\{1, \dots, k\}$. The program then proceeds to compute the smallest x such that $P[X \leq x] \geq q$. If we consider d as its input and x as its output then the program implements the function F_q^{-1} , i.e., the inverse c.d.f function. It is natural to consider the point wise order on c.d.f.s described in Figure 1. The function F_q^{-1} then, obeys the following relational property: $\forall d_1, d_2, q. d_1 \leq_{cdf} d_2 \implies F_q^{-1}(d_1) \geq F_q^{-1}(d_2)$. This property should hence be true for the program considered. Let's see how to see this using RelSym. RelSym will start executing the program in a relational memory with two arrays d_1 and d_2 with the same length, say 5 for instance. Every value in the arrays will be symbolic. These arrays will be related by the following relational assumption (the precondition) $\Phi \equiv \forall i. 1 \leq i \leq 5. \implies d_1[i] \leq d_2[i]$. What we want to show is that in every final state $x_1 \geq x_2$. At the i -th iteration (when $x_h = 0$, for $h \in \{1, 2\}$) the constraint set will have the following constraints $cum_h = d_h[1] + \dots + d_h[i-1]$ ¹. RelSym has now four possible paths to explore given by the outer if-then-else, and for three of these there are four others given by the inner one, for a total of 13. Instead of following a brute force approach and continuing exploring all the paths we can see that one of the paths is already unsatisfiable. This because Φ implies $cum_2 \geq cum_1$ and hence the path characterized by the constraint $cum_1 \geq q \wedge cum_2 < q$ is not satisfiable, and hence not reachable, so it can safely be pruned at every i -th iteration. This pruning was possible thanks to the relational assumption Φ . Similarly, at every i -th iteration, from the symbolic state characterized by $cum_1 \geq q \wedge cum_2 \geq q$ we can disregard the path with constraints $x_1 > 0$ and $x_2 \leq 0$. Relational reasoning allowed us to reduce the number of paths to follow at every iteration from 13 to 8. It is easy to see how, following the remaining paths, RelSym only reaches final states where $x_1 \geq x_2$ and hence proves the specification.

Proving k -Lipschitz continuity of sorting - symbolic bounds. In the second running example - code in Figure 2 - we will again prove a relational property of a program acting on arrays. The difference with the previous example is that we will do it for array of symbolic (arbitrary) size n . To achieve that we will use a very natural relational invariant. In general given a sorting algorithm, run on two arrays a_1, a_2 of integers with the same length n and related by the following relational precondition $\Phi \equiv \forall t. 1 \leq t \leq n \implies |a_1[t] - a_2[t]| \leq k$, we expect the sorted arrays to still satisfy the same condition. We

```

1)  for(i in 1 : len(a) - 1) do
2)    for(j in i + 1 : len(a)) do
3)      if(a[i] > a[j]) then
4)        z ← a[i]
5)        a[i] ← a[j]
6)        a[j] ← z

```

Figure 2: k -Lipschitz continuity of a sorting algorithm.

can see this property as k -Lipschitz continuity of a sorting algorithm with respect to ℓ -infinity norm in both the input and output space. In the program under scrutiny at every iteration of the inner loop we select the smallest element $a[j]$ in the sub array $[i+1 \dots n]$ and we swap it, if necessary, with $a[i]$. In order to make sense of this example it's important to understand that the three lines 4), 5), and 6) which implement the swapping are actually continuous. Indeed, when RelSym is executing the branching instruction, there are four possible ways the two executions can proceed, that is: both take the same branch, or they get different branches. When the two executions take the same branch then obviously Φ still holds. The following Observation 1 guarantees that this is the case also when the two executions follow different branches.

OBSERVATION 1. $\forall x, y, z, w, k. |x - y| \leq k, |z - w| \leq k, x > z, y \leq w \implies |z - y| \leq k, |x - w| \leq k.$

For instance, instantiating $x = a_1[i], y = a_2[i], z = a_1[i+1], w = a_2[i+1]$, ensures that Φ still holds when the left execution takes the true branch and the right execution takes the false branch. So, omitting synchronization of the loop variables, by using the invariant: $I \equiv \forall t. 1 \leq t < i \implies |a_1[t] - a_2[t]| \leq k$ for both the loops we can *jump* outside of the external loop to a unique state where $I[\text{len}(a) + 1/i]$ holds. This state implies trivially the postcondition. The important fact to notice here is the very natural invariant that relational reasoning allows us to specify. In a unary execution instead we would have to come up with non trivial invariants allowing us to prove the functional correctness of the program. We will need to prove not only that the program produces a sorted sequence but also that the output is a a

Refuting cost equivalence - concrete bounds. In the next example we will use RelSym to refute a property about a pair of programs c_1, c_2 . Let's consider the programs in Figure 3. As we mentioned, RelSym rules can be extended to use ghost variables that can be updated at every step of execution of the abstract machine. We can in this way reason about relational cost [Çiçek et al. 2017], by using the relational ghost variable γ which gets incremented at every assignment. Let's see this in an example where the two programs take both in input an array of non negative symbolic integers of size 5 for instance. The two programs would sum in the variable t all their elements up to some value and save in the variable o the first index in the array that made $t \geq k$ true. Obviously the first program has a higher cost in terms of assignments performed. We want to refute that the two programs have the same cost, that is our postcondition to falsify is $\gamma_1 = \gamma_2$, while our precondition would be $\forall i. 1 \leq i \leq 5 \implies a_1[i] = a_2[i]$. At every iteration of the body, for i ranging from 1 to 5, RelSym would perform, using a specific rule, one step on the left execution updating t and no

¹Actually it will contain the translation of this assertion in a constraint, but this is a technical detail.

steps in the right execution. So γ_1 would be incremented but γ_2 would not. Now the two runs are both about to execute a branching instruction. If on the left execution the guard is true we perform the assignment, and the same assignment is performed on the right. Hence the difference in cost is preserved. If the guard on the left is false we loop, performing another assignment, while on the second run we don't. RelSym would explore these paths finding an initial state, that is a set of concrete values for the array for which the execution of the two programs would lead to a final relational state where $\gamma_1 > \gamma_2$. We stress here how RelSym can, with specific rules, relationally analyze programs with different syntactical structures by looking for synchronization points, i.e., branching instructions, to maximise relational reasoning.

1) $t \leftarrow 0; o \leftarrow 0$	$t \leftarrow 0; o \leftarrow 0$
2) $\text{for}(i \text{ in } 1 : \text{len}(a)) \text{ do}$	$\text{for}(i \text{ in } 1 : \text{len}(a)) \text{ do}$
3) $t \leftarrow a[i] + t$	$\text{if } (t \geq k) \text{ then}$
4) $\text{if } (t \geq k \wedge o \leq 0) \text{ then}$	$\text{if } (o \leq 0) \text{ then}$
5) $o \leftarrow i$	$o \leftarrow i$
6)	$\text{else } t \leftarrow a[i] + t$
Version 1	Version 2

Figure 3: The two versions of the program are not cost equivalent.

Refuting non-interference - symbolic bounds with weak invariant. The next running example involves non-interference [Goguen and Meseguer 1982]. Non-interference was introduced as a strong confidentiality guarantee preventing information to flow from secret values to public observable values. Non-interference can be formally stated as a relational property of two executions of a single program with different inputs: a program c is non-interferent if given two input memories $\mathcal{M}_1$ and $\mathcal{M}_2$ that agree on public data and possibly differ on confidential data, the execution of c on $\mathcal{M}_1$ and $\mathcal{M}_2$ results in memories $\mathcal{M}'_1$ and $\mathcal{M}'_2$, respectively, that agree on public data. That is, secret variables don't interfere with observable public variables. Let's consider the program c in Figure (4). The program takes in input a secret vector of integers s and password vector of integers p of the same length. It then scans the arrays and checks whether they are point wise equal. If not it saves in o the index of the first difference. If we assume s to be an high level variable and p, o, t low level variables, this program is obviously interferent. Starting from two memories where $\text{len}(s) = \text{len}(p) \wedge p_1 = p_2^2$ we can very well reach a final state where $o_1 = o_2 \wedge t_1 = t_2$ does not hold. We can check this (i.e., refute non-interference) for arbitrary length arrays of size n . In particular by using the relational invariant $I_w \equiv (t_1 = t_2 \wedge o_1 = o_2) \Leftrightarrow s_1 = s_2$. Using RelSym with that

1) $t \leftarrow 0; o \leftarrow 0$
2) $\text{for}(i \text{ in } 1 : \text{len}(s)) \text{ do}$
3) $\text{if}(s[i] \neq p[i] \wedge o \leq 0) \text{ then}$
4) $o \leftarrow 1; t \leftarrow i$

Figure 4: Interferent program

invariant will allow to disprove the postcondition $\gamma_1 = \gamma_2$, but the

²Equality on arrays is point wise equality, and can be easily encoded in a first order logic formula with one universal quantifier.

initial memories that RelSym would find might not correspond to real counterexamples this because the relational invariant was not strong enough.

Counterexample generation for non-interference - symbolic bounds with strong invariant. In the above program we can get exact counterexamples by choosing the stronger relational invariant $I_s \equiv I_w \wedge t_1 = \min_h s_1[h] \neq p_1[h] \wedge t_2 = \min_h s_2[h] \neq p_2[h] \wedge o_1 \in \{0, 1\} \wedge o_2 \in \{0, 1\}^3$. As we can see we need to specify the functional (unary) behavior of the two programs in the relational invariant in order to strengthen it. RelSym would then disprove the specification by providing a relational initial memory $\mathcal{M}$ for which the precondition holds and a final relational memory $\mathcal{M}'$ related by the operational semantics of the program. For instance: $\mathcal{M}(p) = ([0], [0])$, $\mathcal{M}(s) = ([0], [1])$, $\mathcal{M}'(o) = (0, 1)$, $\mathcal{M}'(t) = (0, 1)$.

4 CONCRETE LANGUAGES: FOR, RFOR

As already mentioned RelSym is composed by four languages. That is, we extend the semantics of the simplest language FOR in two different directions: relationally (RFOR), and symbolically (SFOR). And then we extend them both to obtain RSFOR. In this section we describe the simplest language which is an imperative language (FOR) that contains for-loops and computes over integers and arrays of integers, and then extend it to a relational language (RFOR). We refer to these two languages as *concrete* to distinguish them from the *symbolic* languages that we will build on top of them in Section 5.

4.1 FOR

Programs in FOR have the following grammar, where $v \in \mathbb{Z}$ are values:

$$\begin{aligned}
 e &::= e \oplus e \mid a[e] \mid \text{len}(a) \mid x \mid v \\
 c &::= \text{skip} \mid c; c \mid x \leftarrow e \mid a[e] \leftarrow e \mid \text{if } e \text{ then } c \text{ else } c \mid \\
 &\quad \text{for } (x \text{ in } e:e) \text{ do } c
 \end{aligned}$$

A variable $x \in \mathbf{Var}$ denotes an integer while an array name $a \in \mathbf{Arrvar}$ denotes a function which maps the set of natural numbers $\{1, \dots, l\}$ to the set $\mathbb{Z}$, with l denoting the length of the array. The set of such functions is denoted by $\mathbf{Array}$. The symbol $\oplus$ denotes an arithmetic operation in $\{+, -, \dots\}$. Expressions are standardly evaluated using a big step judgment $\langle \mathcal{M}, e \rangle \Downarrow_F v$ whose defining rules we omit. Programs c are evaluated through a, mainly standard, small step judgment $\langle \mathcal{M}, c \rangle \xrightarrow{F} (\mathcal{M}', c')$, where memories $\mathcal{M}, \mathcal{M}' \in \mathbf{Mem}$ are partial functions with type $(\mathbf{Var} \rightarrow \mathbb{Z}) \cup (\mathbf{Arrvar} \rightarrow \mathbf{Array})$. We only show one rule for for-loop construct evaluation in Figure 5. Note that for-loops, and thus FOR programs, are always terminating.

4.2 Assertions, triples, validity

We state and validate program specifications using Hoare triples $\{\Phi\}c\{\Psi\}$, where c is a command in FOR and Φ and Ψ (respectively, the pre- and post-condition of the triple) are assertions. Assertions are first-order logical formulas with primitive predicates that

³Again, this invariant is expressible in the language, but it can be expressed easily in the language of our assertions.

$$\begin{array}{c}
\text{FOR-UNROLL} \\
\frac{\langle \mathcal{M}, e_1 \rangle \Downarrow_f v_1 \quad \langle \mathcal{M}, e_2 \rangle \Downarrow_f v_2 \quad v_1 \leq v_2}{(\mathcal{M}, \text{for } (x \text{ in } e_1 : e_2) \text{ do } c) \xrightarrow{f}} \\
\left(\mathcal{M}, x \leftarrow v_1; \text{if } v_2 - v_1 \text{ then for } (x \text{ in } v_1 + 1 : v_2) \text{ do } c \right. \\
\left. \text{else skip} \right)
\end{array}$$

Figure 5: A rule for loops in FOR

compare arithmetic expressions $aexp$. The latter are built from expressions in FOR extended with integer-valued logical variables ($i \in \mathbf{Lvar}$) and array expressions α . Array expressions include array names a , and array update expressions $\alpha[aexp_1 \mapsto aexp_2]$, which denotes the array α with the value at index $aexp_1$ updated to $aexp_2$. Array expressions allow us to express and reason about updates on arrays using the extensional theory of arrays [McCarthy 1961]. The truth of a unary assertion Φ is evaluated against a memory $\mathcal{M} \in \mathbf{Mem}$ and a logical interpretation $\mathcal{I} \in \mathbf{Intlog} \equiv \mathbb{Z}^{\mathbf{Lvar}}$. We write $\mathcal{M} \models_{\mathcal{I}} \Phi$ to denote that Φ holds in memory $\mathcal{M}$ with interpretation $\mathcal{I}$. The following definition, although standard, is given because it will later be extended to a relational setting.

DEFINITION 1. Let Φ and Ψ be unary assertions and c be a FOR command. We say that the triple $\{\Phi\}c\{\Psi\}$ is valid, and we write $\models \{\Phi\}c\{\Psi\}$, if and only if $\forall \mathcal{M}_1, \mathcal{M}_2 \in \mathbf{Mem}, \mathcal{I} \in \mathbf{Intlog}$, if $\mathcal{M}_1 \models_{\mathcal{I}} \Phi$ and $(\mathcal{M}_1, c) \xrightarrow{f}^* (\mathcal{M}_2, \text{skip})$ then $\mathcal{M}_2 \models_{\mathcal{I}} \Psi$.

4.3 RFOR

To enable relational reasoning we first build a *relational language* RFOR on top of FOR. Intuitively, execution of a single RFOR program represents the execution of two FOR programs. Inspired by the approach of Pottier and Simonet [2003], we extend the grammar of FOR with a pair constructor $\langle \cdot \mid \cdot \rangle$ which can be used at the level of values $\langle v_1 \mid v_2 \rangle$, expressions $\langle e_1 \mid e_2 \rangle$, or commands $\langle c_1 \mid c_2 \rangle$. Notice that c_i, e_i, v_i for $i \in \{1, 2\}$ are commands, expressions, and values in FOR, hence nested pairing is not allowed. This syntactic invariant is preserved by the rules handling the branching instruction. Pair constructs are used to indicate where commands, values, or expressions might be different in the two unary executions represented by a single RFOR execution. To define the semantics for RFOR, we first extend memories to allow program variables to map to pairs of integers, and array variables to map to pairs of arrays. That is, the type of memories for RFOR is $(\mathbf{Var} \rightarrow \mathbb{Z} \cup \mathbb{Z}^2) \cup (\mathbf{Array} \rightarrow \mathbf{Array} \cup \mathbf{Array}^2)$. The semantics of RFOR is defined as a big step judgment $\langle \mathcal{M}, e \rangle \Downarrow_{\text{RF}} v$ for expressions and a small step judgment $(\mathcal{M}, c) \xrightarrow{\text{RF}} (\mathcal{M}', c')$ for commands, where $\mathcal{M}, \mathcal{M}'$ are relational memories, c, c' are commands in RFOR, v ranges over $\mathbb{Z} \cup \mathbb{Z}^2$, and e is a relational expression. Figure 6 shows a selection of the inference rules for these judgments. The rules use auxiliary functions $[\cdot]_1$ and $[\cdot]_2$, which project, respectively, the first (left) and second (right) elements of a pair construct (i.e., $[\langle c_1 \mid c_2 \rangle]_i = c_i$, $[\langle e_1 \mid e_2 \rangle]_i = e_i$ with $[v]_i = v$ when $v \in \mathbb{Z}$), and are homomorphic for other constructs. For a relational memory $\mathcal{M}$, we write $[\mathcal{M}]_i$ for the (unary) memory that projects the co-domain appropriately: $\forall n \in \mathbf{dom}(\mathcal{M}). [\mathcal{M}]_i(n) = [\mathcal{M}(n)]_i$. Rule **R-LIFT** is the only evaluation rule for RFOR expressions. It evaluates the left and right projections of the memory and expression, and combines

the results into either a single value, if both projections produce the same result, or a pair value otherwise. Rule **R-IF-FALSE-FALSE** shows what happens if the left and right executions both agree on taking the false branch: the command $\text{if } e \text{ then } c_{\text{tt}} \text{ else } c_{\text{ff}}$ steps to command c_{ff} . However, if the left and right execution disagree on which branch to take, we need to introduce a command pair construct to indicate that the command being executed differs in the left and right executions. One instance of this is rule **R-IF-FALSE-TRUE**. We ensure well-formedness of the paired commands by projecting c_{tt} and c_{ff} before pairing them up. Rule **R-PAIR-STEP** evaluates a pair command by picking one projection, non nondeterministically, and evaluating it one step, using the semantics of FOR. The helper function $\text{merge}(\cdot, \cdot)$ merges two FOR memories $\mathcal{M}_1$ and $\mathcal{M}_2$ into a RFOR memory, using as few pair values as possible:

$$\text{merge}(\mathcal{M}_1, \mathcal{M}_2) = \lambda m. \begin{cases} \mathcal{M}_1(m) & \text{if } \mathcal{M}_1(m) = \mathcal{M}_2(m) \\ (\mathcal{M}_1(m), \mathcal{M}_2(m)) & \text{otherwise} \end{cases}$$

Another rule, not shown in the figure, reduces $(\mathcal{M}, \langle \text{skip} \mid \text{skip} \rangle)$ to $(\mathcal{M}, \text{skip})$. The rules regarding array assignments now have to take into account that arrays might differ in the two runs. In particular, given the command $a[e_l] \leftarrow e_h$ the two expressions e_l and e_h might evaluate differently in the left and right projections. In the case where $\mathcal{M}(a)$ is a unary array but the index expression evaluates to a pair value then the updated array will be a pair of arrays, as shown in **R-ARR-ASS-SPLIT**.

4.4 Relational assertions, relational triples, and relational validity

We again use Hoare triples to provide specifications of RFOR programs. However, assertions for RFOR must be able to express properties of both executions of a program, and the relationship between them. To achieve this, we extend expressions in the language to include indexed program variables and array variables, that is we equip an array name a or a program variable x with an index $i \in \{1, 2\}$ so that, for example a_1 denotes the array a in the left execution, or x_2 denotes the variable x in the right execution. We refer to the extended language as *relational assertions*. We extend relational operators ($=, \leq, <, \dots$) and binary operators ($+, -, \dots$) to work with two pairs of values in the obvious way, and adapt the definition of the truth of a relational assertion $(\mathcal{M} \models_{\mathcal{I}} \Phi)$ appropriately. Note that logical variables continue to range only over integers (and not over pairs of integers). Nonetheless, the logic allows us to express relational and unary properties easily. Validity of relational Hoare triples for RFOR is similar to Definition 1, except for the use of relational assertions, relational memories, and the semantic judgment of RFOR instead of FOR. In the same spirit of the consistency theorem in [Banerjee et al. 2016], the following lemma provides a semantical justification for RFOR with respect to FOR.

LEMMA 4.1. Let Φ and Ψ be relational assertions and c be a FOR command. If $\models \{\Phi\}c\{\Psi\}$ then for all unary memories $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}'_1, \mathcal{M}'_2$ and for all $\mathcal{I} \in \mathbf{Intlog}$ such that $\text{merge}(\mathcal{M}_1, \mathcal{M}_2) \models_{\mathcal{I}} \Phi$, $(\mathcal{M}_1, c) \xrightarrow{f}^* (\mathcal{M}'_1, \text{skip})$ and $(\mathcal{M}_2, c) \xrightarrow{f}^* (\mathcal{M}'_2, \text{skip})$ then $\text{merge}(\mathcal{M}'_1, \mathcal{M}'_2) \models_{\mathcal{I}} \Psi$.

Notice that concrete relational semantics is incomplete with respect to the unary semantics with respect to traces in the sense

$$\begin{array}{c}
\textbf{R-IF-FALSE-FALSE} \\
\frac{\langle \mathcal{M}, e \rangle \Downarrow_{\text{RF}} \langle v_1 \mid v_2 \rangle \quad v_1 \leq 0 \quad v_2 \leq 0}{(\mathcal{M}, \text{if } e \text{ then } c_{tt} \text{ else } c_{ff}) \xrightarrow{\text{RF}} (\mathcal{M}, c_{ff})} \\
\\
\textbf{R-IF-FALSE-TRUE} \\
\frac{\langle \mathcal{M}, e \rangle \Downarrow_{\text{RF}} \langle v_1 \mid v_2 \rangle \quad v_1 \leq 0 \quad v_2 > 0}{(\mathcal{M}, \text{if } e \text{ then } c_{tt} \text{ else } c_{ff}) \xrightarrow{\text{RF}} (\mathcal{M}, \langle \lfloor c_{ff} \rfloor_1 \mid \lfloor c_{tt} \rfloor_2 \rangle)} \\
\\
\textbf{R-PAIR-STEP} \\
\frac{\{i, j\} = \{1, 2\} \quad \langle \lfloor \mathcal{M} \rfloor_i, c_i \rangle \xrightarrow{E} (\mathcal{M}'_i, c'_i) \quad c'_j = c_j \quad \mathcal{M}'_j = \lfloor \mathcal{M} \rfloor_j \quad \mathcal{M}' = \text{merge}(\mathcal{M}'_i, \mathcal{M}'_j)}{(\mathcal{M}, \langle c_1 \mid c_2 \rangle) \xrightarrow{\text{RF}} (\mathcal{M}', \langle c'_1 \mid c'_2 \rangle)} \\
\\
\textbf{R-ARR-ASS-SPLIT} \\
\frac{\langle \mathcal{M}, e_l \rangle \Downarrow_{\text{RF}} v_l \quad \langle \mathcal{M}, e_h \rangle \Downarrow_{\text{RF}} v_h \quad \mathcal{M}(a) = f \quad \lfloor v_l \rfloor_1, \lfloor v_l \rfloor_2 \in \text{dom}(f) \quad \lfloor v_l \rfloor_1 \neq \lfloor v_l \rfloor_2 \quad f_1 = f[\lfloor v_l \rfloor_1 \mapsto \lfloor v_h \rfloor_1] \quad f_2 = f[\lfloor v_l \rfloor_2 \mapsto \lfloor v_h \rfloor_2]}{(\mathcal{M}, a[e_l] \leftarrow e_h) \xrightarrow{\text{RF}} (\mathcal{M}[a \mapsto \langle f_1 \mid f_2 \rangle], \text{skip})} \\
\\
\textbf{R-LIFT} \\
\frac{\langle \lfloor \mathcal{M} \rfloor_1, \lfloor e \rfloor_1 \rangle \Downarrow_F v_1 \quad \langle \lfloor \mathcal{M} \rfloor_2, \lfloor e \rfloor_2 \rangle \Downarrow_F v_2 \quad v = \begin{cases} v_1 & \text{if } v_1 = v_2 \\ (v_1, v_2) & \text{otherwise} \end{cases}}{\langle \mathcal{M}, e \rangle \Downarrow_{\text{RF}} v}
\end{array}$$

Figure 6: Semantics of RFOR (selected rules).

that the iterations of a loop go in lockstep until at least one side terminates (after which the other side may continue). In fact, in order to keep the design of the language simple we only allow pair commands to be introduced by a branching instruction. In general this causes RFOR to not be complete with respect to FOR. So it is not possible to use invariants that hold between different iterations by using rule such as the *dissonant loop rule* in [Beringer 2011]. Indeed in RelSym the following two programs cannot not be proved equivalent, for arbitrary positive n : for $(i \text{ in } 1:2 * n)$ do $x \leftarrow x + 1$ and for $(i \text{ in } 1:n)$ do $x \leftarrow x + 1$; for $(i \text{ in } 1:n)$ do $x \leftarrow x + 1$.

5 SYMBOLIC LANGUAGES: SFOR, RSFOR

Symbolic execution [King 1976] extends a language with *symbolic values* that represent unknown or undetermined concrete values. Symbolic execution uses symbolic values in logical formulas that track the conditions under which a particular execution path is taken. By exploring different execution paths and finding satisfying assignments to these logical formulas (i.e., finding concrete values to substitute for symbolic values such that the formulas will be satisfied), symbolic execution of a program can be used to find concrete test cases that demonstrate an assertion violation in a program. Conversely, if all execution paths of a program are explored and no violation is found, then symbolic execution shows that a program is guaranteed to meet its specification. In this section, we extend the FOR and RFOR languages with symbolic execution, giving us, respectively, the languages SFOR and RSFOR. In particular, RSFOR allows us to reason symbolically about two executions of a FOR program, and thus enables us to look for violations of relational assertions of FOR programs. However, we need to define SFOR in order to fully specify the semantics of RSFOR, indeed, similarly to how the semantics of RFOR relies on the semantics of FOR, the semantics of RSFOR relies on the semantics of SFOR.

The main insight of symbolic execution is to represent sets of concrete values (in this case integers) and sets of concrete runs of a program with symbolic values drawn from a set **Symval**. Symbolic values can be refined during the computation using constraints expressed as formulas in some formal theory. For instance, when the guard X of an if construct is symbolic, we might choose to symbolically execute the true branch and refine the set of possible concrete values that X denotes by adding the constraint $X > 0$ to the *path condition*. The connection between symbolic languages and concrete languages is given by ground substitutions $\sigma \in \Sigma \equiv$

Symval $\rightarrow \mathbb{Z} \cup \text{Array}$. We say that a constraint ϕ is satisfiable if there exists a $\sigma \in \Sigma$ that makes it true. That is, if substituting all the symbolic values X appearing in ϕ with $\sigma(X)$ gives us a true statement. If that's the case we write $\sigma \models \phi$. When we are only interested in expressing the satisfiability of ϕ with no interest in specifying the actual substitutions we will write **SAT**(ϕ). Given a set of constraints $\mathcal{S}$, abusing notation, we denote by $\mathcal{S}$ the constraint $\bigwedge_{s \in \mathcal{S}} s$. Satisfiable path conditions denote actual concrete executions.

That is, all those concrete executions which assign to the symbolic values concrete values that make the path condition true. If a path condition is unsatisfiable then it does not represent any concrete execution. A set of constraints is *valid* if it is true under every possible substitution. We denote the validity of a constraint ϕ by $\models \phi$. Building on the previous section we can now define the two symbolic languages SFOR and RSFOR.

5.1 SFOR

We extend the syntax of FOR expressions by adding to its values elements $X \in \text{Symval}$, denoting symbolic values. Now memories in SFOR map program variables to either integers or symbolic values. We also represent symbolic arrays in memory as pairs (X, v) , where v is a (concrete or symbolic) integer value representing the length of the array, and X is a symbolic value representing the array contents, as in the standard theory of arrays [McCarthy 1961]. The content of the arrays can be refined in a set of constraints described below. Thus, memories in SFOR have the type $(\text{Var} \rightarrow \mathbf{V}_s) \cup (\text{Arrvar} \rightarrow \text{Array}_s)$, where $\mathbf{V}_s \equiv \mathbb{Z} \cup \text{Symval}$ and $\text{Array}_s \equiv \text{Symval} \times \mathbf{V}_s$. Configurations in SFOR are triples $(\mathcal{M}, c, \mathcal{S})$ where $\mathcal{M}$ is a memory, c is a SFOR command, and $\mathcal{S}$ is a set of *constraints*. Constraints are first-order logical formulas with primitive predicates that compare expressions (e) over concrete ($n \in \mathbb{Z}$), symbolic values ($X \in \text{Symval}$) and logical variables ($i \in \text{Lvar}$). Constraint expression $\text{select}(e_1, e_2)$ represents the (integer) result of reading the array denoted by e_1 at the index denoted by e_2 , while $\text{store}(e_1, e_2, e_3)$ represents the (array) result of updating the array denoted by e_1 at index e_2 with value e_3 . A set of constraints $\mathcal{S}$ is used to record restrictions on symbolic values that must hold in order for program execution to reach a specific configuration.

Note that although both assertions and constraints are logical formulas that include comparisons of expressions, they differ because assertions may contain program variables and array names but may

not contain symbolic values; constraints on the other hand may contain symbolic values (including $\text{select}(\cdot, \cdot)$ and $\text{store}(\cdot, \cdot, \cdot)$ expressions) and may not contain program variables or array variables. Given a memory M , we can translate assertions to constraints, using M to replace program variables and array names with the (symbolic or concrete) values M maps them to. We write $\llbracket \cdot \rrbracket_M$ for this translation function defined inductively on the shape of the expression. Symbolic values can now appear in expressions, so a for loop executed by unrolling might not terminate. For this reason we extend the category of commands to also contain commands of this form: $\text{for } (x \text{ in } e_1:e_2) \text{ do } c$. Where I is an assertion intended to be a loop invariant. The two kinds (with and without invariant) of for-loops are treated as distinct syntactic forms.

The semantics of SFOR is defined through a big-step judgment, $\langle M, e, S \rangle \Downarrow_{\text{SF}} \langle v, S' \rangle$, for expressions, and a small-step judgment $\langle M, c, S \rangle \xrightarrow{\text{SF}} \langle M', c', S' \rangle$ for commands. Figure 7 shows some selected rules defining the judgments. Notice that evaluating an expression might generate new symbolic values, and this is why also $\Downarrow_{\text{SF}}$ returns an updated set of constraints S' . In rules for conditionals, like the rule **S-IF-TRUE**, we record in the constraint set the information about the control flow path. Rules handling the conditionals make the small-step operational semantics non-deterministic, since we have to consider both the case when the guard reduces to a value greater than 0 and when it reduces to a value less or equal than 0. In rules for arrays, we record in the constraint set the description of arrays. For example, the rule **S-ARR-READ** records the selection in the constraint using a fresh symbol Y which has never occurred in the computation before that point. Rule **S-ARR-WRITE** evaluates the index of the array to update and the right hand side of the assignment after updating the memory it records the array update in the set of constraints. As already mentioned we allow the user to specify invariant for loops and use the rule **S-FOR-INV**. This rule allows to skip in one step the whole unrolling of the for-loop provided that the user has specified an actual inductive invariant. Specifically, the semantic judgment $\models \{I \wedge e_1 \leq x \wedge x \leq e_2\} c_b \{I[x + 1/x]\}$ imposes that I holds before and after every iteration of the body of the loop provided that the counter variable x is between the bounds. Checking that $e_1, e_2 \in \text{aexp}$ makes sure that the premise of the triple is actually an assertion and does not contain symbolic values, as it could be the case since e_1, e_2 are expressions in SFOR. The additional check, $\models S'' \implies \llbracket I[e_1/x] \wedge e_1 \leq e_2 \rrbracket_M$, imposes that the constraints collected before executing the loop are strong enough to imply the invariant right before the start of the loop. The configuration to which the for-loop with invariant steps to has a set of constraint S_f which records the fact that the for-loop has terminated and so includes the constraint $\llbracket I[e_2 + 1/x] \rrbracket_{M_f}$. The final memory M_f maps to fresh symbolic values all the variables, or array names which might have been updated in the body c_b (**Upd**($\cdot$) performs a syntactic check on c_b , soundly approximating the set of variables updated by c_b). Notice that we don't update the length of the arrays, because we consider only arrays of fixed (static or concrete) length. At the exit of the loop the counter variable has to map to the value to which the second guard of the for-loop was reduced to.

$$\begin{array}{c}
\textbf{S-ARR-READ} \\
\frac{\langle M, e, S \rangle \Downarrow_{\text{SF}} \langle v_s, S' \rangle \quad M(a) = (X, v'_s) \quad Y \text{ fresh}}{\langle M, a[e], S \rangle \Downarrow_{\text{SF}} \langle Y, S' \cup \{Y = \text{select}(X, v_s), v_s > 0, v_s \leq v'_s\} \rangle} \\
\\
\textbf{S-ARR-WRITE} \\
\frac{\langle M, e_1, S \rangle \Downarrow_{\text{SF}} \langle v_1, S' \rangle \quad \langle M, e_2, S' \rangle \Downarrow_{\text{SF}} \langle v_2, S'' \rangle \quad M(a) = (X, v_l) \quad Y \text{ fresh} \quad M' = M[a \mapsto (Y, l)] \quad S''' \equiv S'' \cup \{Y = \text{store}(X, v_1, v_2), v_1 > 0, v_1 \leq l\}}{(\langle M, a[e_1] \leftarrow e_2, S \rangle \xrightarrow{\text{SF}} \langle M', \text{skip}, S''' \rangle)} \\
\\
\textbf{S-IF-TRUE} \\
\frac{\langle M, e, S \rangle \Downarrow_{\text{SF}} \langle v_s, S' \rangle}{(\langle M, \text{if } e \text{ then } c_{tt} \text{ else } c_{ff}, S \rangle \xrightarrow{\text{SF}} \langle c_{tt}, S' \cup \{v_s > 0\} \rangle)} \\
\\
\textbf{S-FOR-INV} \\
\frac{\langle M, e_1, S \rangle \Downarrow_{\text{SF}} \langle v_1, S' \rangle \quad \langle M, e_2, S' \rangle \Downarrow_{\text{SF}} \langle v_2, S'' \rangle \quad e_1, e_2 \in \text{aexp} \quad \models \{I \wedge e_1 \leq x \wedge x \leq e_2\} c_b \{I[x + 1/x]\} \quad M_f = \lambda n. \begin{cases} v_2, & \text{if } n = x \\ X, & \text{if } n \in \text{Upd}(c_b), n \in \text{Var}, X \text{ fresh} \\ (X, l), & \text{if } n \in \text{Upd}(c_b), M(n) = (Z, l), X \text{ fresh} \\ M(n) & \text{otherwise} \end{cases} \quad \models S'' \implies \llbracket I[e_1/x] \wedge e_1 \leq e_2 \rrbracket_M \quad S_f = S'' \cup \{\llbracket I[e_2 + 1/x] \rrbracket_{M_f}\}}{(\langle M, \text{for } (x \text{ in } e_1:e_2) \text{ do } c_b, S \rangle \xrightarrow{\text{SF}} \langle M_f, \text{skip}, S_f \rangle)}
\end{array}$$

Figure 7: Semantics of SFOR (selected rules).

5.2 RSFOR

Similarly to what we did in the previous section, we now extend the language RFOR to RSFOR using symbolic values X . The symbolic extension of the relational language follows the same steps as the unary with the difference that now symbolic values can also appear in pairs of expressions $\langle e_1 \mid e_2 \rangle$ and pairs of commands $\langle c_1 \mid c_2 \rangle$ and pairs of values in a memory $\langle v_1, v_2 \rangle$. As in the case of the previous languages we give the semantics to RSFOR by means of a big step semantics for symbolic relational expressions proving judgments of the shape $\langle M, e, S \rangle \Downarrow_{\text{RSF}} \langle v, S' \rangle$, and a small step semantics for symbolic relational commands proving judgments of the shape $\langle M, c, S \rangle \xrightarrow{\text{RSF}} \langle M', c', S' \rangle$. We provide a selection of the rules to prove those judgments in Figure 8. Projection functions need now to be smartly extended to relational assertions, this would be particularly useful for example when a for-loop with invariant I appears in one of the branches of an if construct with a guard which evaluates to a relational value $\langle v_1 \mid v_2 \rangle$, since both cases $v > 0, v \leq 0$ have to be considered. For this reason we extend projection functions for basic relational assertions in the following way (where $\{p, q\} = \{a, b\}$, and where the function $\text{Idx}(\cdot)$ returns the set (potentially empty) of indices $i \in \{1, 2\}$ appearing in a relational expression):

$$\begin{aligned}
\lfloor e_a \otimes e_b \rfloor_i &= \lfloor e_a \rfloor_i \otimes \lfloor e_b \rfloor_i && \text{if } \text{Idx}(e_q) \subseteq \text{Idx}(e_p) = \{i\} \text{ or } \text{Idx}(e_q) = \text{Idx}(e_p) = \emptyset \\
\lfloor e_a \otimes e_b \rfloor_i &= \text{true} && \text{otherwise} \\
\lfloor x_i \rfloor_i &= x \\
\lfloor x \rfloor_i &= x
\end{aligned}$$

For other forms of assertions projection functions behave homomorphically. So for $i \in \{1, 2\}$ we can now define $\lfloor \text{for } (x \text{ in } e_1:e_2) \text{ do } c_b \rfloor_i \equiv$

for $(x \text{ in } [e_1]_i : [e_2]_i) \text{ do } [c]_i$. Also, **merge-s**($\cdot, \cdot$) plays a similar role in the relational symbolic semantics to what **merge**($\cdot, \cdot$) does in the concrete one. The rule **R-S-LIFT** relies on SFOR. It evaluates a relational symbolic expression and returns a single symbolic value if the two unary symbolic execution reduce to the same integer value, otherwise it splits. Rule **R-S-ARR-ASS-SPLIT** takes care of an array assignment when the array is symbolic unary but the right hand side of the assignment is different, and hence the array needs to be split. Rule **R-S-IF-FALSE-TRUE** is similar to the analogous rule for the concrete semantics we presented in Figure 6, the main difference is that now the path conditions are recorded in the constraint set. Rule **R-S-IF-RIGHT** takes care of a pair command with a branching instruction on the right and a different command on the left. This rule, and a similar one for the left execution, helps synchronization of the two runs. In rule **R-S-PAIR-STEP** takes care of the general case, where $c_1 \equiv c_2$ means structural equality, for instance c_1 and c_2 are both assignments. Similarly to the analogous concrete rule, one side of the two is chosen non-nondeterministically, and one step on that side is performed using the unary symbolic semantics. Finally, the rule **R-S-FOR-INV** allows the user to specify a relational invariant for a for-loop which might diverge because one of the guards evaluates to a value containing a symbolic value. The rule **R-S-FOR-INV** behaves similarly to **S-FOR-INV** but in a relational setting.

5.3 Unary and relational collecting semantics

Building on the $\xrightarrow{\text{SF}}$ and $\xrightarrow{\text{RF}}$ semantics, we define now two collecting semantics which consider only reachable configurations, namely those whose set of constraints is satisfiable. Overloading the symbol $\Rightarrow$ we will denote by it both the unary and relational collecting semantics. Both semantics are defined through only one rule presented in Figure 9. In rule **SET-STEP** we remove from the set of configurations taken in consideration the current configuration and we add to it all the configurations reachable in one step that are satisfiable.

$$\begin{array}{c} \text{SET-STEP} \\ \frac{\mathcal{F}_t = \{(M', c', S') \mid (M, c, S) \xrightarrow{\dagger} (M', c', S') \wedge \text{SAT}(S')\} \quad (M, c, S) \in \mathcal{F} \quad \mathcal{F}' = \left(\mathcal{F} \setminus \{(M, c, S)\} \right) \cup \mathcal{F}_t}{\mathcal{F} \Rightarrow \mathcal{F}'} \end{array}$$

Figure 9: Unary and relational collecting semantics rule schema.

$\dagger \in \{\text{SF}, \text{RSF}\}$

6 META THEORY

In this section we will make more precise the connection between concrete and symbolic languages. In order to do this, we need to reason about *ground substitutions* turning object containing symbolic values into concrete objects. Given a command c or an expression e in SFOR (or in RSFOR) and a ground substitution $\sigma \in \Sigma$ we write $\sigma(c)$ (and $\sigma(e)$) for the application of σ to c (and e). We can also apply a substitution to a unary symbolic memory:

DEFINITION 2. Given a ground substitution $\sigma \in \Sigma$ we define its application to a unary symbolic memory as

$$\begin{array}{c} \text{R-S-LIFT} \\ \frac{\langle [M]_1, [e]_1, S \rangle \Downarrow_{\text{SF}} \langle v_1, S' \rangle \quad \langle [M]_2, [e]_2, S' \rangle \Downarrow_{\text{SF}} \langle v_2, S'' \rangle}{\langle v, S''' \rangle = \begin{cases} \langle v_1, S''' \rangle & \text{if } (v_1, v_2) \in \mathbb{Z}^2 \wedge v_1 = v_2 \\ \langle (v_1, v_2), S''' \rangle & \text{otherwise} \end{cases}} \\ \langle M, e, S \rangle \Downarrow_{\text{RSF}} \langle v, S''' \rangle \\ \\ \text{R-S-ARR-ASS-SPLIT} \\ \frac{\begin{array}{l} M(a) = (X, l) \in \text{Symval} \times \mathbf{V}_s \quad Z \text{ fresh} \quad W \text{ fresh} \\ \langle M, e_i, S \rangle \Downarrow_{\text{RSF}} \langle v_i, S' \rangle \quad \langle M, e_h, S' \rangle \Downarrow_{\text{RSF}} \langle v_h, S'' \rangle \\ S''' = S' \cup \{[v_h]_1 \neq [v_h]_2, Z = \text{store}(X, [v_i]_1, [v_h]_1)\} \\ S'''' = S''' \cup \{W = \text{store}(X, [v_i]_2, [v_h]_2), 0 < [v_i]_1 \leq l, 0 < [v_i]_2 \leq l\} \end{array}}{(M, a[e_i] \leftarrow e_h, S) \xrightarrow{\text{RSF}} (M[a \mapsto ((Z, l), (W, l))], \text{skip}, S'''')} \\ \\ \text{R-S-IF-FALSE-TRUE} \\ \frac{\langle M, e, S \rangle \Downarrow_{\text{RSF}} \langle v, S' \rangle \quad S'' = S' \cup \{[v]_1 \leq 0, [v]_2 > 0\}}{(M, \text{if } e \text{ then } c_{\text{tt}} \text{ else } c_{\text{ff}}, S) \xrightarrow{\text{RSF}} (M, \langle [c_{\text{ff}}]_1 \mid [c_{\text{tt}}]_2 \rangle, S'') \\ \\ \text{R-S-PAIR-STEP} \\ \frac{(\langle [M]_i, c_i, S \rangle \xrightarrow{\text{SF}} (M'_i, c'_i, S'')) \quad \left(\begin{array}{l} \text{if } \cdot \text{ then } \cdot \text{ else } \cdot \neq c_j = c'_j \quad \text{or} \quad c_1 \equiv c_2 \\ \{1, 2\} = \{i, j\} \quad M'_j = [M]_j \quad M' = \text{merge-s}(M'_1, M'_2) \end{array} \right)}{(M, \langle c_1 \mid c_2 \rangle, S) \xrightarrow{\text{RSF}} (M', \langle c'_1 \mid c'_2 \rangle, S'') \\ \\ \text{R-S-IF-RIGHT} \\ \frac{c_1 \equiv \text{if } \cdot \text{ then } \cdot \text{ else } \cdot \quad c_2 \notin \{\text{if } \cdot \text{ then } \cdot \text{ else } \cdot, \text{skip}\} \quad (\langle [M]_2, c_2, S \rangle \xrightarrow{\text{SF}} (M'_2, c'_2, S'')) \quad M' = \text{merge-s}([M]_1, M'_2)}{(M, \langle c_1 \mid c_2 \rangle, S) \xrightarrow{\text{RSF}} (M', \langle c_1 \mid c'_2 \rangle, S'') \\ \\ \text{R-S-FOR-INV} \\ \frac{\begin{array}{l} \langle M, e_a, S \rangle \Downarrow_{\text{RSF}} \langle v_a, S' \rangle \quad \langle M, e_b, S' \rangle \Downarrow_{\text{RSF}} \langle v_b, S'' \rangle \\ \models \{I \wedge e_a \leq x \wedge x \leq e_b\} c \{I[x_1 + 1/x_1][x_2 + 1/x_2]\} \\ \models S'' \Rightarrow \llbracket I[e_a]_1/x_1 \rrbracket \llbracket [e_a]_2/x_2 \rrbracket \wedge e_a \leq e_b \rrbracket_M \\ S_f = S'' \cup \{ \llbracket I[v_b]_1 + 1/x_1 \rrbracket \llbracket [v_b]_2 + 1/x_2 \rrbracket \rrbracket_{M_f} \} \end{array}}{M_f = \lambda n. \begin{cases} v_b, & \text{if } n = x \\ (X, Y), & \text{if } n \in \text{Updr}(c), M(n) \in \mathbf{V}_s \cup \mathbf{V}_s^2 \\ & X \text{ fresh}, Y \text{ fresh} \\ ((X, l), (Y, l)), & \text{if } n \in \text{Updr}(c), M(n) \in \text{Array}_s \\ & \pi_2(M(n)) = l \\ ((X, l), (Y, l)), & \text{if } n \in \text{Updr}(c), M(n) \in \text{Array}_s^2 \\ & \pi_2(\pi_2(M(n))) = l \\ M(n), & \text{otherwise} \end{cases}} \\ \\ (M, \text{for } (x \text{ in } e_a : e_b) \text{ do } c, S) \xrightarrow{\text{RSF}} (M_f, \text{skip}, S_f) \end{array}$$

Figure 8: Semantics of RSFOR (simplified selected rules).

$$\sigma(M) = \lambda m. \begin{cases} \sigma(M(m)) & \text{if } M(m) \in \text{Symval} \\ \sigma(\pi_1(M(m))) & \text{if } M(m) \in \text{Symval} \times \mathbf{V}_s \\ M(m) & \text{otherwise} \end{cases}$$

where m ranges over $\text{Var} \cup \text{Arrvar}$.

We have a similar definition for relational symbolic memories which we omit here. From now on, we consider only substitutions σ which respect the type of the program variables and array names appearing in a symbolic expression or command. That is, given an

expressing e (or command c) we consider substitutions σ for which $\sigma(e)$ ($\sigma(c)$) is an expression (command) in FOR (RFOR) whenever e (c) is an expression (command) in SFOR (RSFOR). We also want to consider only substitutions mapping symbolic values to objects of their type. This is characterized by the following definition.

DEFINITION 3. We say that a ground substitution $\sigma \in \Sigma$ validates a configuration (M, c, S) and we write $\sigma \models (M, c, S)$ iff $\sigma \models S$, $\forall a \in \mathbf{Arrvar}. M(a) = (X, v) \Rightarrow \sigma(X) \in \{1, \dots, \sigma(v)\} \rightarrow \mathbb{Z}$, $\forall x \in \mathbf{Var}. M(x) = X \Rightarrow \sigma(X) \in \mathbb{Z}$, and σ respects the type of array names and program variables in c .

We also consider the natural partial order, $\leq$ over Σ given by the relation $\{(\sigma_1, \sigma_2) \in \Sigma^2 \mid \forall X \in \mathbf{dom}(\sigma_1). \sigma_1(X) = \sigma_2(X)\}$.

6.1 Coverage

We now want to formalize the idea that a run of the set semantics can capture (cover) many concrete runs. To do this we formalize what a final configuration ($\mathcal{F}$) of the $\Rightarrow$ semantics (Figure 9) is.

DEFINITION 4. A unary (or relational) configuration s is final, and we write $\mathbf{Final}(s)$, when $s = (M, \text{skip}, S)$. A set of configurations $\mathcal{F}$ is final, denoted $\mathbf{Final}(\mathcal{F})$, if and only if for all $s \in \mathcal{F}. \mathbf{Final}(s)$.

The following lemma states that any concrete execution can be covered by a symbolic path. This symbolic path will have a satisfiable set of constraints which will make it possible to map back symbolic final configurations to the concrete final configuration of the concrete path.

LEMMA 6.1. If $\mathcal{F} \Rightarrow^* \mathcal{F}'$, $(M_1, c_1, S_1) \in \mathcal{F}$, and $\sigma_1 \models (M_1, c_1, S_1)$ then $\exists k_{c_1}, \exists (M_2, c_2, S_2) \in \mathcal{F}', \exists \sigma_2 \in \Sigma$ such that $(\sigma_1(M_1), c_1) \xrightarrow{k_{c_1}} (\sigma_2(M_2), c_2)$ (or $(\sigma_1(M_1), c_1) \xrightarrow{\text{RE}}^{k_{c_1}} (\sigma_2(M_2), c_2)$), $\sigma_2 \models (M_2, c_2, S_2)$, and $\sigma_1 \leq \sigma_2$.

6.2 Proving and soundness

In symbolic execution we want to execute symbolically a program in order to reason about multiple concrete executions. In order to do this we need to specify an initial memory from which the symbolic execution can start. Without loss of generality we choose as initial memory the most abstract. This leads to the following definition:

DEFINITION 5. Let Φ be a unary assertion, and c a command in FOR. Define the following symbolic memory:

$$\text{Mem}_{\Phi, c} \equiv \lambda n \in \mathbf{VarOf}(\Phi) \cup \mathbf{VarOf}(c). \begin{cases} X, & \text{if } n \in \mathbf{Var} \\ (X, L), & \text{if } n \in \mathbf{Arrvar} \end{cases}$$

where all the variables X, L are meant to be distinct and fresh, and the function $\mathbf{VarOf}(\cdot)$ returns the set of program variables and array names appearing in the argument.

The previous definition can be easily extended to relational memories, assertions, and commands. As we already discussed, we are interested in using RelSym for proving valid specifications of programs. If we want to prove that a triple $\{\Phi\}c\{\Psi\}$ is valid, we can execute symbolically c starting from an initial symbolic configuration which satisfies the precondition Φ . If we reach only final configurations whose set of constraints imply the postcondition Ψ , then the triple is valid. Formally:

DEFINITION 6. Let c be a command in FOR (or RFOR) and Φ and Ψ unary (or relational) assertions. We say that c symbolically proves Ψ from Φ , and we write $c : \Phi \Rightarrow \Psi$ iff there exists $\mathcal{F}$ such that

- $\{(\text{Mem}_{\Phi, c}, c, \{\llbracket \Phi \rrbracket_{\text{Mem}_{\Phi, c}}\})\} \Rightarrow_{\text{set}}^* \mathcal{F}$
- $\mathbf{Final}(\mathcal{F})$
- $\forall (M, \text{skip}, S) \in \mathcal{F}. \models S \Rightarrow \llbracket \Psi \rrbracket_M$

It now makes sense to formulate the following soundness theorem:

THEOREM 6.2 (SOUNDNESS OF VERIFICATION). Let Φ and Ψ be unary (or relational) assertions and let c be a command in FOR (or RFOR). Then, if $c : \Phi \Rightarrow \Psi$ then $\models \{\Phi\}c\{\Psi\}$.

PROOF. By structural induction on c , using Lemma 6.1. $\square$

6.3 Finding counterexamples: strength of invariants and soundness

We now want to formalize the fact that we can use RelSym for finding counterexamples. Let us consider a program c , a precondition Φ and a postcondition Ψ . If starting to evaluate c from an initial symbolic configuration and a set of constraints that satisfy the precondition Φ , we arrive in a final configuration whose set of constraint is consistent with the negation of the postcondition Ψ (interpreted in the memory of the final configuration), then we know that the post-condition does not hold. This argument motivates the following definition.

DEFINITION 7. Let c be a command in FOR (or RFOR) and Φ, Ψ unary (or relational) assertions. We say that, c symbolically disproves Ψ from Φ and we write $c : \Phi \not\Rightarrow \Psi$ if and only if exists $\mathcal{F}$ such that

- $\{(\text{Mem}_{\Phi, c}, c, \{\llbracket \Phi \rrbracket_{\text{Mem}_{\Phi, c}}\})\} \Rightarrow_{\text{set}}^* \mathcal{F}$
- $\exists (M, \text{skip}, S) \in \mathcal{F}. \mathbf{SAT}(S \cup \{\llbracket \neg \Psi \rrbracket_M\})$

A counterexample to the validity of a unary triple $\{\Phi\}c\{\Psi\}$ consists of a pair of concrete memories M_1, M_2 and $I \in \mathbf{Intlog}$ such that $M_1 \models_I \Phi$ and $(M_1, c) \xrightarrow{*} (M_2, \text{skip})$ but $M_2 \not\models_I \Psi$.

We would like to be able to extract, from an execution showing $c : \Phi \not\Rightarrow \Psi$, a counterexample for $\{\Phi\}c\{\Psi\}$. Unfortunately, this cannot always be done.

Indeed because in presence of loops, invariants might just approximate the state after the loop has terminated. That is the invariants might not specify precisely enough the state after the loop body has been executed n times for arbitrary n . For instance: $\{z = 0 \wedge x > 0\} \text{for } (i \text{ in } 1:x) \text{ do } \mathbf{true} \ z \leftarrow z + 1 \{\mathbf{false}\}$ is obviously an invalid triple but the invariant does not say much about the value of z after the loop has been executed x times. The invariant $I_s \equiv z = i$ would instead do the job, specifying exactly the final state. When invariants have this property we say they are strong. With Definition 8 we capture the notion of strength of an invariant.

DEFINITION 8. Given a command $c \equiv \text{for } (x \text{ in } e_1:e_2) \text{ do } I \ c_b \text{ in FOR (or in RFOR)}$, we say that the invariant I is strong iff $\forall \sigma_1, \sigma_2 \in \Sigma$, if $\sigma_1 \models (M_f, \text{skip}, S_f)$, $\sigma_2 \models (M_f, \text{skip}, S_f)$, and $\sigma_1(M) =_R \sigma_2(M)$ then $\sigma_1(M_f) =_U \sigma_2(M_f)$.

Where $R = \left((\mathbf{Var} \cup \mathbf{Arrvar}) \setminus U \right) \cup \{x\}$, $U = \mathbf{Upd}(c_b)$ (or $U = \mathbf{Updr}(c_b)$), and M, M_f, S_f are respectively the memory right before

the execution of the for-loop, and the memory and the set of constraints after the application of the rule **S-FOR-INV** (or **R-S-FOR-INV**).

The following theorem allows to avoid false positives in interactive refutation.

THEOREM 6.3 (SOUNDNESS OF COUNTEREXAMPLE FINDING). *Let Φ, Ψ be unary (or relational) assertions and c a command in FOR (or RFOR). Then, if $c : \Phi \not\Rightarrow \Psi$ and all the invariants in c (if any) are strong then $\models \{\Phi\}c\{\Psi\}$.*

Theorem 6.3 is a soundness result for counterexample finding which implies (relative) completeness of the proving system w.r.t to the semantics of FOR (and RFOR). Indeed, provided the program c is annotated with strong enough invariants, if RelSym cannot derive $c : \Phi \not\Rightarrow \Psi$ then it has to be the case that $\models \{\Phi\}c\{\Psi\}$. The completeness just mentioned concerns the proving system and has nothing to do with the semantic completeness of RFOR w.r.t to FOR which has been already ruled out in Section 4.4.

7 IMPLEMENTATION

RelSym has been implemented in OCaml 4.06 in about 4k LOC. The queries on satisfiability of set of constraints are discharged using the SMT solver Z3 [De Moura and Bjørner 2008]. The implementation is not fully optimized.

7.1 Checking the semantic judgment

Rules **S-FOR-INV** and **R-S-FOR-INV**, include among the premises a Hoare triple validity judgment, which ensures that the assertion provided is an inductive invariant of the loop. By using semantic validity we allow other potential implementations to use different analysis techniques for the verification of that triple, e.g., a sound Hoare logic for FOR (or RFOR). Since we want RelSym to be a self-contained tool, in the implementation we prove this judgment by recursively calling RelSym. In particular, while executing the rule **S-FOR-INV** (or **R-S-FOR-INV**) on the command `for ( $x$  in  $e_1:e_2$ ) doI  $c$` we use recursively RelSym to prove $\models \{I \wedge e_1 \leq x \wedge x \leq e_2\}c\{I[x + 1/x]\}$, by checking that indeed $c : I \wedge e_1 \leq x \wedge x \leq e_2 \Rightarrow I[x + 1/x]$. This can also help in practice in finding the *right* invariant by giving the user prompt feedback on why the assertion used at the moment is not an inductive invariant.

7.2 Checking the strength of the invariant

If we want to use RelSym for finding counterexamples to specifications, we might need to check that the invariant is strong as in Definition (8), so that by Theorem 6.3 we can be sure that the ground substitution provided (if any) by the SMT is indeed a counterexample. In particular, this ensures that, if the SMT returns a σ such that $\sigma \models \mathcal{S}_f \cup \{\neg \llbracket \Psi \rrbracket_{\mathcal{M}_f}^F\}$, then indeed:

- $\sigma(\text{Mem}_{\Phi,c}) \models \Phi$
- $(\sigma(\text{Mem}_{\Phi,c}), \sigma(c)) \xrightarrow{F^*} (\sigma(\mathcal{M}_f), \text{skip})$
(or $(\sigma_1(\mathcal{M}_1), c_1) \xrightarrow{RF^*} (\sigma_2(\mathcal{M}_2), c_2)$),
- $\sigma(\mathcal{M}_f) \models \neg \Psi$

A way to check this property is to check for unsatisfiability the following set of constraints:

$$\mathcal{S}_f \cup \{\llbracket I[v_2 + 1/x] \rrbracket_{\mathcal{M}_f}^F\} \cup \left\{ \bigvee_{\{X \in F\}} X \neq X' \right\}$$

where: F is the set of fresh symbols generated during the execution of the rule **S-FOR-INV** (or **R-S-FOR-INV**), and $\llbracket I[v_2 + 1/x] \rrbracket_{\mathcal{M}_f}^F$ is the result of taking the invariant where x has been substituted with $v_2 + 1$, interpreted as a constraint through $\mathcal{M}_f$, with all the symbols in F substituted with their primed versions. If it is not the case that $\text{SAT}(\mathcal{S}_f \cup \{\llbracket I[v_2 + 1/x] \rrbracket_{\mathcal{M}_f}^F\} \cup \{\bigvee_{\{X \in F\}} X \neq X'\})$ then there is only a possible way to satisfy $\mathcal{S}_f$ once the symbols generated before the loop have been fixed, that is given a ground substitution σ for which $\sigma \models \mathcal{S}_f$ then there is only one possible σ' such that $\sigma \leq \sigma'$ and $\sigma' \models \mathcal{S}_f$. This implies the strength of the invariant I .

8 EXPERIMENTAL RESULTS

We compared our relational symbolic semantics with other techniques used to prove or finding counterexamples to relational properties. In particular with naive self-composition, simple product programs and the product programs construction of [Eilers et al. 2018]. Since our implementation does not use any heuristics to try to improve efficiency it makes sense to compare it with vanilla versions of these techniques. Also, notice that product programs and self-composition can be easily embedded in our framework by just executing self-composed programs and product programs in SFOR, that is by just using unary symbolic semantics. In this section we can see some experimental results that show that relational symbolic execution is comparable in terms of execution time, calls to the solver, and number of steps with respect to self-composition and product programs. The results in Table (1) are about proving relational properties, while in Table (2) the results are about finding counterexamples to relational properties. Some of the examples are taken from standard literature (sometimes adapting them to our language). In the table an *R* (Relational) means that relational symbolic execution was used, while *U* denotes that the self composed program was analyzed with unary symbolic semantics, a *P* denotes a product program symbolically executed with unary semantics. Because of space reasons we only show information which showed discernible differences in resource usage. An $\uparrow$ denotes that the symbolic execution had to be terminated because it was running for too long, while an $\times$ means that the SMT solver was not able to discharge a query and so the result is unknown. Finally, a $?$ denotes absence of information, necessary when RelSym ran out of time limits. The results regarding execution time are an average over 50 runs executed on an Intel CPU, 2.80GHz with 16 GB of RAM memory.

The examples concern properties such as non-interference, e.g., *n-inter. example* and *inter. example* series, *ni-array* example, or execution time independence e.g., [Antonopoulos et al. 2017], or continuity e.g., *sum-k-lip-cont*, or *sort-k-lip-cont*. On this benchmark overall relational symbolic execution performs better with respect to standard unary self composition and comparably to product programs, in terms of execution time. Besides execution times (unary and relational semantics) we can consider as measures also other information such as the number of steps of the semantics (small-step #SS, big-steps #BS) performed, calls to the solver (#SMT) and number of final states reached (#S). Using these metrics shows more clearly how a relational approach can, at times, outperform other approaches for the verification or interactive refutation of relational properties. [Eilers et al. 2018] construction for product

Example	R/U/P	#BS	#SS	#SMT	#S	tm (s)
Darvas	R	8	5	3	2	0.39
Darvas	U	15	18	5	3	0.04
Costanzo	R	30032	42315	10921	4096	139
Costanzo	U	?	?	?	?	↑
Antonopoulos	R	68	101	20	10	0.15
Antonopoulos	U	70	94	22	10	0.16
Terauchi[1]	R	34	63	24	4	0.22
Terauchi[1]	P	309	2472	179	9	1.49
Terauchi[1]	U	46	141	22	4	0.22
Terauchi[2]	R	55	91	34	9	0.36
Terauchi[2]	U	31	155	10	3	✗
n-inter. example 1	R	5	4	1	1	0.01
n-inter. example 1	P	33	56	26	4	0.01
n-inter. example 1	U	9	16	1	1	0.01
n-inter. example 2	R	5	4	1	1	0.01
n-inter. example 2	U	9	16	1	1	0.01
n-inter. example 3	R	7	9	1	1	0.01
n-inter. example 3	P	30	87	24	2	0.16
n-inter. example 3	U	13	36	1	1	0.02
n-inter. example 4	R	14	13	8	4	0.08
n-inter. example 4	P	51	109	35	9	0.2
n-inter. example 4	U	16	14	10	4	0.1
sum-k-lip-cont	R	8	5	3	2	0.04
sum-k-lip-cont	U	8	11	3	2	✗
sort-k-lip-cont	U	55	72	23	12	0.12
sort-k-lip-cont	R	31	45	12	6	0.11
sort-k-lip-cont	P	50	66	15	12	0.12

Table 1: Experimental results of proving relational properties. Where Darvas stands for [Darvas et al. 2005], Costanzo stands for [Costanzo and Shao 2014], Antonopoulos stands for [Antonopoulos et al. 2017], Terauchi stands for [Terauchi and Aiken 2005]

Example	R/P/U	#BS	#SS	#SMT	#S	tm (s)
ni-array	R	291	380	132	37	1.08
ni-array	U	342	674	90	16	0.7
[Eilers et al. 2018]	R	9	7	3	2	0.03
[Eilers et al. 2018]	U	13	25	1	1	0.01
inter. example1	R	3	1	1	1	0.01
inter. example1	P	13	10	10	4	0.08
inter. example1	U	21	33	5	3	0.04
inter. example2	R	3	1	1	1	0.01
inter. example2	P	13	10	10	4	0.07
inter. example2	U	5	4	1	1	0.02
inter-password	R	485	714	169	64	1.40
inter-password	U	703	960	190	64	1.78

Table 2: Experimental results for finding counterexamples relational properties.

programs introduces new variables and new branching instructions. This is the main reasons why the number of SMT calls increases. More generally: consider the base product program construction in

[Butler and Schulte 2011] and the number of basic instructions performed (e.g. assignments) as a measure: commands are duplicated even when it doesn't help. Product self-composition is a generic syntactic technique. E.g.: take $c \equiv p \leftarrow p + 1$, and suppose we want to show that: $\models \{p1 = p2\}c\{p1 = p2\}$. Under product programs we could reduce the problem to verifying: $\models \{p1 = p2\}c1 \times c2\{p1 = p2\}$ that is $\models \{p1 = p2\}p1 \leftarrow p1 + 1; p2 \leftarrow p2 + 1\{p1 = p2\}$. In the unary symbolic execution of the product program *necessarily* two assignments will be performed. While executing relationally $p \leftarrow p + 1$ *might* only execute one assignment.

This evaluation shows that although we have trade-offs between the different techniques and none of them is always better, in several situations relational symbolic execution brings clear improvements.

9 RELATED WORKS

The works most closely related to ours are the ones that have used symbolic execution for relational properties. [Milushev et al. 2012] use symbolic execution to check non-interference by means of an analysis based on a type directed transformation of the program first presented in [Terauchi and Aiken 2005]. The analysis targets programs written in a subset of C which includes procedures calls, and dynamically allocated data structures modeled through a heap. A main difference with our work is that they focus only on non-interference while we focus on arbitrary relational properties. Additionally, they use self-composition while we focus on the design of a formal relational semantics. Finally, they use a generic approach based on heaps, instead, we focus on arrays as concrete data structures and we leverage their properties in the design of our semantics.

In [Person et al. 2008] symbolic execution is used to check differences between program versions. The property they analyze, although relational can be easily described with two separate execution of the two programs. Indeed, in their work symbolic execution is used separately for the two programs.

Relational properties have also been studied through many other techniques. We already mentioned different works that reduce the verification of relational properties to the one of properties through self-composition [Barthe et al. 2004; Terauchi and Aiken 2005] and product programs [Asada et al. 2017; Butler and Schulte 2011; Eilers et al. 2018]. Several works have studied relational versions of Hoare logics. For example, Benton [Benton 2004] studies relational Hoare logics for noninterference and program equivalence, and Barthe et al. [Barthe et al. 2014b, 2012] study relational Hoare logics for relational probabilistic properties, such as differential privacy. Their work is based on a denotational semantics based on couplings and probabilistic liftings, while ours is operational in nature. Other works such as [Banerjee et al. 2016] have focused on a relational Hoare logics with frame rules to deal with heap based semantics, and on situations where keeping the traces not aligned might be beneficial in the same spirit of dissonant loop rules introduced in [Beringer 2011]. Other works instead tried to maximize the amount of synchronicity between the two runs [Pick et al. 2018]. Several works have studied type systems for the verification of different relational properties, some examples are noninterference [Nanevski et al. 2013; Pottier and Simonet 2003; Volpano et al. 1996], security of cryptographic implementations [Barthe et al. 2014a], differential

privacy and mechanism design [Barthe et al. 2015], and relational cost [Çiçek et al. 2017]. These approaches are quite different from ours. For instance [Çiçek et al. 2017] focuses on functional programs, and uses a type discipline which requires a lot of domain expertise. Other works have applied abstract interpretation techniques to noninterference [Assaf et al. 2017; Feret 2001; Giacobazzi and Mastroeni 2004]. While symbolic execution and abstract interpretations share several similarities, the techniques that the approaches rely on are quite different. In [Austin and Flanagan 2012] authors introduce faceted values, that resemble our paired values. They do this to simulate simultaneous runs of the same program on different security levels, in order to provide information flow security with a dynamic approach as opposed to a static one as we do in this work. Cartesian Hoare Logic [Sousa and Dillig 2016] and its quantitative extension [Chen et al. 2017] can be used for reasoning about generic k-safety properties, and their quantitative analogues. The language that Cartesian Hoare Logic considers includes arrays and while loops with breaks. The class of properties they consider goes beyond relational properties and their analysis is automated. The main difference between their approach and ours is that we perform symbolic execution which can also be used to finding bugs while they only focus, at least on the theoretical part, on proving correctness via Hoare Logic. Kwon et al. [Kwon et al. 2017] recently proposed a program analysis for checking information flow policies over streams based on a technique for synthesizing relational invariants. This analysis is not based on symbolic execution, but we plan to explore if their algorithm for synthesizing relational invariants can be used in our setting.

Similar to our work their semantics is based on couplings and the probabilistic lifting of relations. Close to our work is also [Albarghouti and Hsu 2018] where a proof technique, casting differential privacy proofs as a strategy in a game encoded as a set of constraints, is presented. In that work authors focus again in finding proof and not in finding counter examples to differential privacy.

10 CONCLUSIONS

In this work we presented RelSym, a foundational framework for relational symbolic execution. The framework supports interactive refutation as well as proving of relational properties for a language with arrays and loops. We provided some meta theoretical results about symbolic execution for its use with respect to proving validity of triples and disproving them and we provided necessary conditions for which disproving is actually sound. We have shown the flexibility of this approach by analyzing examples for a range of different relational properties. We compared the analysis of this properties using different approaches, i.e., self-composition, product programs and relational approach. We have implemented the tool and in the future we plan to address more complex features like functions, promises and closures, as well as exploring the generation of relational loop invariants [Chen et al. 2011; Hoder et al. 2011; Khurshid et al. 2003; Kwon et al. 2017; Qin et al. 2013], limiting in this way the need for annotations provided by the user.

REFERENCES

Aws Albarghouti and Justin Hsu. 2018. Synthesizing coupling proofs of differential privacy. *PACMPL* 2, POPL (2018), 58:1–58:30. <https://doi.org/10.1145/3158146>

- Timos Antonopoulos, Paul Gazzillo, Michael Hicks, Eric Koskinen, Tachio Terauchi, and Shiyi Wei. 2017. Decomposition Instead of Self-composition for Proving the Absence of Timing Channels. In *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2017)*. ACM, New York, NY, USA, 362–375. <https://doi.org/10.1145/3062341.3062378>
- Kazuyuki Asada, Ryosuke Sato, and Naoki Kobayashi. 2017. Verifying Relational Properties of Functional Programs by First-order Refinement. *Sci. Comput. Program.* 137, C (April 2017), 2–62.
- Mounir Assaf, David A. Naumann, Julien Signoles, Eric Totel, and Frédéric Tronel. 2017. Hypercollecting semantics and its application to static analysis of information flow. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017, Paris, France, January 18–20, 2017*. 874–887.
- Thomas H. Austin and Cormac Flanagan. 2012. Multiple Facets for Dynamic Information Flow. In *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '12)*. ACM, New York, NY, USA, 165–178. <https://doi.org/10.1145/2103656.2103677>
- Anindya Banerjee, David A. Naumann, and Mohammad Nikouei. 2016. Relational Logic with Framing and Hypotheses. In *36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2016, December 13–15, 2016, Chennai, India*. 11:1–11:16.
- Gilles Barthe, Juan Manuel Crespo, and César Kunz. 2011. Relational Verification Using Product Programs. In *Proceedings of the 17th International Conference on Formal Methods (FM'11)*. Springer-Verlag, Berlin, Heidelberg, 200–214. <http://dl.acm.org/citation.cfm?id=2021296.2021319>
- Gilles Barthe, Pedro R. D'Argenio, and Tamara Rezk. 2004. Secure Information Flow by Self-Composition. In *Proceedings of the 17th IEEE Workshop on Computer Security Foundations (CSFW '04)*.
- Gilles Barthe, Cédric Fournet, Benjamin Grégoire, Pierre-Yves Strub, Nikhil Swamy, and Santiago Zanella Béguelin. 2014a. Probabilistic relational verification for cryptographic implementations. In *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14, San Diego, CA, USA, January 20–21, 2014*. 193–206.
- Gilles Barthe, Cédric Fournet, Benjamin Grégoire, Pierre-Yves Strub, Nikhil Swamy, and Santiago Zanella Béguelin. 2014b. Probabilistic Relational Verification for Cryptographic Implementations. *SIGPLAN Not.* 49, 1 (Jan. 2014), 193–205.
- Gilles Barthe, Marco Gaboardi, Emilio Jesús Gallego Arias, Justin Hsu, Aaron Roth, and Pierre-Yves Strub. 2015. Higher-Order Approximate Relational Refinement Types for Mechanism Design and Differential Privacy. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2015, Mumbai, India, January 15–17, 2015*. 55–68. <https://doi.org/10.1145/2676726.2677000>
- Gilles Barthe, Boris Köpf, Federico Olmedo, and Santiago Zanella Béguelin. 2012. Probabilistic relational reasoning for differential privacy. (2012), 97–110. <https://doi.org/10.1145/2103656.2103670>
- Nick Benton. 2004. Simple Relational Correctness Proofs for Static Analyses and Program Transformations. In *Proceedings of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '04)*. ACM, New York, NY, USA, 14–25.
- Lennart Berlinger. 2011. Relational Decomposition. In *Interactive Theorem Proving*. Springer Berlin Heidelberg, Berlin, Heidelberg, 39–54.
- Michael J. Butler and Wolfram Schulte (Eds.). 2011. *FM 2011: Formal Methods - 17th International Symposium on Formal Methods, Limerick, Ireland, June 20–24, 2011. Proceedings*. Lecture Notes in Computer Science, Vol. 6664. Springer.
- Swarat Chaudhuri, Sumit Gulwani, and Roberto Lubliner. 2010. Continuity Analysis of Programs. In *Proceedings of the 37th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '10)*. ACM, New York, NY, USA, 57–70. <https://doi.org/10.1145/1706299.1706308>
- Swarat Chaudhuri, Sumit Gulwani, and Roberto Lubliner. 2012. Continuity and Robustness of Programs. *Commun. ACM* 55, 8 (Aug. 2012), 107–115.
- Jia Chen, Yu Feng, and Isil Dillig. 2017. Precise Detection of Side-Channel Vulnerabilities using Quantitative Cartesian Hoare Logic. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017*. 875–890.
- Shikun Chen, Zhoujun Li, Xiaoyu Song, and Mengjun Li. 2011. An Iterative Method for Generating Loop Invariants. In *Proceedings of the 5th Joint International Frontiers in Algorithms, and 7th International Conference on Algorithmic Aspects in Information and Management (FAW-AAIM'11)*.
- Ezgi Çiçek, Gilles Barthe, Marco Gaboardi, Deepak Garg, and Jan Hoffmann. 2017. Relational cost analysis. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017, Paris, France, January 18–20, 2017*. 316–329.
- David Costanzo and Zhong Shao. 2014. A Separation Logic for Enforcing Declarative Information Flow Control Policies. In *Principles of Security and Trust - Third International Conference, POST 2014, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2014, Grenoble, France, April 5–13, 2014. Proceedings*. 179–198. https://doi.org/10.1007/978-3-642-54792-8_10

- Ádám Darvas, Reiner Hähnle, and David Sands. 2005. A Theorem Proving Approach to Analysis of Secure Information Flow. In *Proceedings of the Second International Conference on Security in Pervasive Computing (SPC'05)*. Springer-Verlag, Berlin, Heidelberg, 193–209. https://doi.org/10.1007/978-3-540-32004-3_20
- Leonardo De Moura and Nikolaj Bjørner. 2008. Z3: An Efficient SMT Solver. In *Proceedings of the Theory and Practice of Software, 14th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'08/ETAPS'08)*. 337–340.
- Marco Eilers, Peter Müller, and Samuel Hitz. 2018. Modular Product Programs. In *Programming Languages and Systems*, Amal Ahmed (Ed.). Springer International Publishing, Cham, 502–529.
- Jérôme Feret. 2001. Abstract Interpretation-Based Static Analysis of Mobile Ambients. In *Eighth International Static Analysis Symposium (SAS'01) (LNCS)*. Springer-Verlag.
- Roberto Giacobazzi and Isabella Mastroeni. 2004. Abstract non-interference: parameterizing non-interference by abstract interpretation. In *Proceedings of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2004, Venice, Italy, January 14–16, 2004*. 186–197.
- Joseph A. Goguen and José Meseguer. 1982. Security Policies and Security Models. In *1982 IEEE Symposium on Security and Privacy, Oakland, CA, USA, April 26–28, 1982*. 11–20.
- Joseph A. Goguen and José Meseguer. 1984. Unwinding and Inference Control. In *Proceedings of the 1984 IEEE Symposium on Security and Privacy, Oakland, California, USA, April 29 - May 2, 1984*. 75–87.
- Martin Hentschel, Richard Bubel, and Reiner Hähnle. 2014. Symbolic Execution Debugger (SED). In *Proceedings of Runtime Verification 2014 (2014-01-01) (LNCS)*, Borzoo Bonakdarpour and Scott A. Smolka (Eds.). Springer, 255–262.
- Krystof Hoder, Laura Kovács, and Andrei Voronkov. 2011. Invariant Generation in Vampire. In *Tools and Algorithms for the Construction and Analysis of Systems - 17th International Conference, TACAS 2011, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2011, Saarbrücken, Germany, March 26–April 3, 2011. Proceedings*. 60–64.
- Martin Hofmann and Mariela Pavlova. 2008. Elimination of Ghost Variables in Program Logics. In *Trustworthy Global Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Catalin Hritcu, John Hughes, Benjamin C. Pierce, Antal Spector-Zabusky, Dimitrios Vytiniotis, Arthur Azevedo de Amorim, and Leonidas Lampropoulos. 2013. Testing Noninterference, Quickly. In *Proceedings of the 18th ACM SIGPLAN International Conference on Functional Programming (ICFP '13)*. 455–468.
- Sarfraz Khurshid, Corina S. Păsăreanu, and Willem Visser. 2003. Generalized Symbolic Execution for Model Checking and Testing. In *Proceedings of the 9th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'03)*. 553–568.
- James C. King. 1976. Symbolic Execution and Program Testing. *Commun. ACM* 19, 7 (July 1976), 385–394.
- Hyoukjun Kwon, William Harris, and Hadi Esmaeilzadeh. 2017. Proving Flow Security of Sequential Logic via Automatically-Synthesized Relational Invariants. In *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21–25, 2017*. 420–435.
- John McCarthy. 1961. A Basis for a Mathematical Theory of Computation (preliminary report). In *Proceedings of the Western Joint Computer Conference*, Cicely M. Popplewell (Ed.). IRE, AIEE, ACM, 225–238.
- Dimitar Milushev, Wim Beck, and Dave Clarke. 2012. Noninterference via Symbolic Execution. In *Proceedings of the IFIP WG 6.1 International Conference on Formal Techniques for Distributed Systems (FMOODS'12/FORTE'12)*. 152–168.
- Floralé Morandat, Brandon Hill, Leo Osvald, and Jan Vitek. 2012. Evaluating the Design of the R Language - Objects and Functions for Data Analysis. In *ECOOP 2012 - Object-Oriented Programming - 26th European Conference, Beijing, China, June 11–16, 2012. Proceedings*. 104–131.
- Aleksandar Nanevski, Anindya Banerjee, and Deepak Garg. 2013. Dependent Type Theory for Verification of Information Flow and Access Control Policies. *ACM Trans. Program. Lang. Syst.* 35 (2013), 6:1–6:41.
- Suzette Person, Matthew B. Dwyer, Sebastian Elbaum, and Corina S. Păsăreanu. 2008. Differential Symbolic Execution. In *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering (SIGSOFT '08/FSE-16)*. 226–237.
- Lauren Pick, Grigory Fedyukovich, and Aarti Gupta. 2018. Exploiting Synchrony and Symmetry in Relational Verification. In *Computer Aided Verification*.
- François Pottier and Vincent Simonet. 2003. Information Flow Inference for ML. *ACM Trans. Program. Lang. Syst.* 25, 1 (Jan. 2003), 117–158.
- Shengchao Qin, Guanhua He, Chenguang Luo, Wei-Ngan Chin, and Xin Chen. 2013. Loop invariant synthesis in a combined abstract domain. *J. Symb. Comput.* 50 (2013), 386–408.
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Jason Reed and Benjamin C. Pierce. 2010. Distance makes the types grow stronger: a calculus for differential privacy. In *Proceeding of the 15th ACM SIGPLAN international conference on Functional programming, ICFP 2010, Baltimore, Maryland, USA, September 27–29, 2010*. 157–168.
- Helgi Sigurbjarnarson, Luke Nelson, Bruno Castro-Karney, James Bornholt, Emina Torlak, and Xi Wang. 2018. Nickel: A Framework for Design and Verification of Information Flow Control Systems. In *13th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2018, Carlsbad, CA, USA, October 8–10, 2018*.
- Marcelo Sousa and Isil Dillig. 2016. Cartesian Hoare Logic for Verifying K-safety Properties. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI '16)*. 57–69.
- Tachio Terauchi and Alex Aiken. 2005. Secure Information Flow As a Safety Problem. In *Proceedings of the 12th International Conference on Static Analysis (SAS'05)*. 352–367.
- Dennis M. Volpano, Cynthia E. Irvine, and Geoffrey Smith. 1996. A Sound Type System for Secure Flow Analysis. *Journal of Computer Security* 4, 2/3 (1996), 167–188.