Why Just Boogie?
Translating Between Intermediate Verification Languages

Michael Ameri\textsuperscript{1} and Carlo A. Furia\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1} Chair of Software Engineering, Department of Computer Science, ETH Zurich, Switzerland, mameri@student.ethz.ch
\textsuperscript{2} Department of Computer Science and Engineering, Chalmers University of Technology, Sweden, furia@chalmers.se

Abstract. The verification systems Boogie and Why3 use their respective intermediate languages to generate verification conditions from high-level programs. Since the two systems support different back-end provers (such as Z3 and Alt-Ergo) and are used to encode different high-level languages (such as C# and Java), being able to translate between their intermediate languages would provide a way to reuse one system’s features to verify programs meant for the other. This paper describes a translation of Boogie into WhyML (Why3’s intermediate language) that preserves semantics, verifiability, and program structure to a large degree. We implemented the translation as a tool and applied it to 194 Boogie-verified programs of various sources and sizes; Why3 verified 83% of the translated programs with the same outcome as Boogie. These results indicate that the translation is often effective and practically applicable.

1 Introduction

Intermediate verification languages (IVLs) are intermediate representations used in verification technology. Just like compiler design has benefited from decoupling front-end and back-end, IVLs help write verifiers that are more modular: the front-end specializes in encoding the rich semantics of a high-level language (say, an object-oriented language such as C#) as a program in the IVL; the back-end generates verification conditions (VCs) from IVL programs in a form that caters to the peculiarities of a specific theorem prover (such as an SMT solver).

Boogie \cite{2} and WhyML \cite{8} are prime examples of popular IVLs with different, often complementary, features and supporting systems (respectively called Boogie and Why3). In this paper we describe a translation of Boogie programs into WhyML programs and its implementation as the tool \texttt{b2w}. As we illustrate with examples in Sec. 3 using \texttt{b2w} increases the versatility brought by IVLs: without having to design and implement a direct encoding into WhyML or even being familiar with its peculiarities, users can take advantage of some of the best features of Why3 when working with high-level languages that translate to Boogie.

\textsuperscript{*} Work done mainly while affiliated with ETH Zurich.
**Boogie vs. WhyML.** While the roles of Boogie and WhyML as IVLs are similar, the two languages have different characteristics that reflect a focus on complementary challenges in automated verification. Boogie is the more popular language in terms of front-ends that use it as IVL, which makes a translation from Boogie more practically useful than one into it; it has a finely tuned integration with the Z3 prover that results from the two tools having been developed by the same group (Microsoft Research’s RiSE); it combines a simple imperative language with an expressive typed logic, which is especially handy for encoding object-oriented or, more generally, heap-based imperative languages. In contrast, WhyML has a more flexible support for multiple back-end provers it translates to, including a variety of SMT solvers as well as interactive provers such as Coq; it can split VCs into independent goals and dispatch them to different provers; it offers limited imperative constructs within a functional language that belongs to the ML family, which brings the side benefit of being able to execute WhyML programs—a feature quite useful to debug and validate verification attempts.

**Goals and evaluation.** The overall goal of this paper is devising a translation $T$ from Boogie to WhyML programs. The translation, described in Sec. 4, should preserve correctness, verifiability, and readability as much as possible. Preserving correctness means that, given a Boogie program $p$, if its translation $T(p)$ is a correct WhyML program then $p$ is correct (soundness); the converse should also hold as much as possible: if $T(p)$ is incorrect then $p$ is too (precision). Preserving verifiability means that, given a Boogie program $p$ that verifies in Boogie, its translation $T(p)$ is a WhyML program that verifies in Why3. Preserving readability means that the translation should not introduce unnecessary changes in the structure of programs.

The differences, outlined above, between Boogie and WhyML and their supporting systems make achieving correctness, verifiability, and readability challenging. While we devised $T$ to cover the entire Boogie language, its current implementation $b2w$ does not fully support a limited number of features (branching, the most complex polymorphic features, and bitvectors) that make it hard to achieve verifiability in practice. In fact, while replacing branching (goto) with looping is always possible [11], a general translation scheme does not produce verifiable loops since one should also infer invariants (which are often cumbersome due to the transformation). Polymorphic maps are supported to the extent that their type parameters can be instantiated with concrete types; this is necessary since WhyML’s parametric polymorphism cannot directly express all usages in Boogie, but it may also introduce a combinatorial explosion in the translation; hence, $b2w$ fails on the most complex instances that would be unmanageable in Why3. Boogie’s bitvector support is much more flexible than what provided by Why3’s libraries; hence $b2w$ may render the semantics of bitvector operations incorrectly.

These current implementation limitations notwithstanding (see Sec. 4 for details), we experimentally demonstrate that $b2w$ is applicable and useful in practice. As Sec. 5 discusses, we applied $b2w$ to 194 Boogie programs of different size and sources; most of the programs have not been written by us and exercise Boogie in a variety of different ways. For 83% (161) of these programs, $b2w$ produces a WhyML translation that Why3 can verify as well as Boogie can verify the original, thus showing the feasibility of automating translation between IVLs.
Tool availability. The tool b2w is available as open source at:

https://bitbucket.org/michael_ameri/b2w/

2 Related Work

Translations and abstraction levels. Translation is a ubiquitous technique in computer science; however, the most common translation schemes bridge different abstraction levels, typically encoding a program written in a high-level language (such as Java) into a lower-level representation which is suitable for execution (such as byte or machine code). Reverse-engineering goes the opposite direction—from lower to higher level—for example to extract modular and structural information from C programs and encode it using object-oriented constructs [26]. This paper describes a translation between intermediate languages—Boogie and Why3—which belong to similar abstraction levels. In the context of model transformations [29], so-called bidirectional transformations [25] also target lossless transformations between notations at the same level of abstraction.

Intermediate verification languages. The Spec# project [3] introduced Boogie to add flexibility to the translation between an object-oriented language (a dialect of C#) and the verification conditions in the logic fragments supported by SMT solvers. An intermediate verification language embodies the idea of intermediate representation—a technique widespread in compiler construction—in the context of verification. Since its introduction for Spec#, Boogie has been adopted as intermediate verification language for numerous other front-ends such as Dafny [16], AutoProof [28], Viper [12], and Joogie [1]; its popularity demonstrates the advantages of using intermediate verification languages.

While Boogie retains some support for different back-end SMT solvers, Z3 [7] remains its fully supported primary target. By contrast, supporting multiple, different back-ends is one of the main design goals behind the Why3 system [8], which does not merely generate verification conditions in different formats but offers techniques to split them into independently verifiable units and to dispatch each unit to a different prover. Why3 also fully supports interactive provers, which provide a powerful means of discharging the most complex verification conditions that defy complete automation.

Another element that differentiates Boogie and Why3 is the support for executing programs; this is quite useful for debugging verification attempts and for applying testing-like techniques to the realm of verification. Boogaloo [21] supports symbolic execution of Boogie programs; Symbooglix is a more recent project with the same goal [19]. Thanks to it being a member of the ML family, Why3 directly supports symbolic execution as well as compilation of WhyML programs to OCaml.

In all, while the Boogie and WhyML languages belong to a similar abstraction level, they are part of systems with complementary features, which motivates this paper’s idea of translating one language into the other. Since Boogie is overall more popular, in terms of tools that use it as a back-end, the translation from Boogie to WhyML is more practically useful than the one in the opposite direction.

3 In comparison, Boogie’s support for HOL is restricted and not up-to-date [4].
Other intermediate languages for verification are Pilar [24], used in the Sireum framework for SPARK; Silver [12], an intermediate language with native support for permissions in the style of separation logic; and the flavor of dynamic logic for object-oriented languages [23] used in the KeY system. Another approach to generalizing and reusing different translations uses notions from model transformations to provide validated mappings for different high-level languages [5]. Future work may consider supporting some of these intermediate languages and approaches.

3 Motivating Examples

Verification technology has made great strides in the last decade or two, but a few dark corners remain where automated reasoning shows its practical limitations. Fig. 1 provides three examples of simple Boogie programs that trigger incorrect or otherwise unsatisfactory behavior. We argue that translating these programs to WhyML makes it possible to verify them using a different, somewhat complementary verification tool; overall, confidence in the results of verification is improved.

Procedure not_verify in Fig. 1 has a contradictory postcondition (notice \( N < N \), \( N \) is a nonnegative constant, and the loop immediately terminates). Nonetheless, recent versions of Boogie and Z3 successfully verify it. More generally, unless the complete toolchain has been formally verified (a monumental effort that has only been performed in few case studies [18,13,14]), there is the need to validate the successful runs of a verifier. Translating Boogie to Why3 provides an effective validation, since Why3 has been developed independent of Boogie and uses a variety of backends that Boogie does not support. Procedure not_verify translated to Why3 (Fig. 2) does not verify as it should.

Procedures lemma_yes and lemma_no in Fig. 1 demonstrate Boogie’s support for mathematical real numbers, which is limited in the way the power operator \( ^* \) is handled. Boogie vacuously verifies both properties \( 2^{3} > 0 \) and \( 2^{1} < 0 \), even though Z3 outputs some unfiltered errors that suggest the verification is spurious (the power operator \( ^* \) is not properly supported); indeed, only the inequality encoded by lemma_yes is correct. Why3 provides a more thorough support for real arithmetic, both by translating to backends such as Alt-Ergo and by providing a more effective encoding in Z3; in fact, it verifies the translated procedure lemma_yes but correctly fails to verify lemma_no.

The loop in procedure trivial_inv in Fig. 1 includes an invariant asserting that \( i \) takes only even values. Even if this is clearly true, Boogie fails to check it; pinning down the precise cause of this shortcoming requires knowledge of Boogie’s (and Z3’s) internals, although it likely is a manifestation of the “triggers” heuristics that handle (generally undecidable) quantified expressions. Based on this knowledge, there are specification patterns that try to work around such idiosyncrasies; in the example, one could introduce a “witness” ghost variable \( k \) such that \( i = 2\times k \) is an invariant. However, if we insist on verifying the program in its original form, Why3 can dispatch verification conditions to interactive provers, where the user provides the crucial proof steps. Cases such as the loop invariant of trivial_inv where a proof is “obvious” to a human user

---

4 https://github.com/boogie-org/boogie/issues/25

5 Why3 can also check the invariant automatically by relying on the CVC4 SMT solver.
but it clashes against the default strategies to handle quantifiers are prime candidate to exploit interactive provers. Thus, translating Boogie to Why3 provides another means of exploiting the latter’s versatile support for interactive provers and multiple backends.

\begin{verbatim}
const N : int;
axiom 0 ≤ N;

procedure not_verify()
ensures (∀ k, l : int • 0 ≤ k ≤ l < N ⇒ N < N);
{
  var x : int;
  x := -N;
  while (x ≠ x) { }
}

procedure lemma_yes()
ensures 2.0**3.0 > 0.0;
{
}

procedure lemma_no()
ensures 2.0**3.0 < 0.0;
{
}

procedure trivial_inv()
{
  var i : int;
  i := 0;
  while (i < 10)
  invariant 0 ≤ i ≤ 10;
  invariant (∃ j : int • i = 2*j);
  { i := i + 2; }
}
\end{verbatim}

Fig. 1. Three simple Boogie programs for which automated reasoning is limited.

4 Boogie-to-Why3 Translation

Intermediate languages for verification combine programming constructs and a logic language. When used to encode programs written in a high-level language, the programming constructs encode program behavior, and the logic constructs encode specifications, constrain the semantics to conform to the high-level language’s (typically through axioms), and support other kinds of annotations (such as triggers).

Both Boogie and WhyML provide, as logic language, a typed first-order logic with arithmetic. Boogie’s programming constructs are a simple imperative language with both structured (while loops, procedures) and unstructured (jumps, global variables) statements. WhyML’s programming constructs combine an ML-like functional language with a few structured imperative features such as mutable variables and loops.

Correspondingly, we define a translation \( T : \text{Boogie} \rightarrow \text{WhyML} \) of Boogie to WhyML as the composition \( E \circ D \) of two translations: \( D : \text{Boogie} \rightarrow \text{Boogie} \) is a desugaring\(^6\), which rewrites away the Boogie constructs, such as \textit{call-forall}, that have no similar construct in WhyML by expressing them using other features of Boogie. Then, \( E : \text{Boogie} \rightarrow \text{WhyML} \) encodes Boogie programs simplified by \( D \) as WhyML programs, while introducing constraints that ensure that the semantics in WhyML mirrors the one in Boogie. For simplicity, the presentation does not sharply separate the two translations \( D \) and \( E \) but defines either or both of them as needed to describe the translation of arbitrary Boogie constructs.

A single feature of the Boogie language significantly compounds the complexity of the translation: \textit{polymorphic maps}, which correspond to mappings between domains of generic type. Why3 does support polymorphic maps through a library, but its type system is more restrictive and does not allow the same degree of freedom as Boogie’s

---

6 This is unrelated to Boogie’s built-in desugaring mechanism (option /printDesugared).
constant N: int
axiom A0: 0 ≤ N;

val not_verify (): ()
ensures { ∀ k, l: int . 0 ≤ k ≤ l < N → N < N }

let not_verify_impl(): ()
ensures { ∀ k, l: int . 0 ≤ k ≤ l < N → N < N }
= (let x = ref (any int) in x.contents ← -N;
while (x.contents ≠ x.contents) do done;
end)

val lemma_yes (): ()
ensures { (pow 2.0 3.0) > 0.0 }

val lemma_no (): ()
ensures { (pow 2.0 3.0) < 0.0 }

let lemma_yes_impl (): ()
ensures { (pow 2.0 3.0) > 0.0 }
= ()

let lemma_no_impl (): ()
ensures { (pow 2.0 3.0) < 0.0 }
= ()

val trivial_inv (): ()
let trivial_inv_impl (): ()
= (let i = ref (any int) in
  invariant { 0 ≤ i.contents ≤ 10 }
  invariant { ∃ j: int . i.contents = 2*j }
  i.contents ← i.contents + 2;
  done;
)

Fig. 2. The translation to WhyML of the three Boogie programs in Fig. 1. (Boilerplate such as general declarations, imports, and frame condition checking are omitted for clarity.)

in using variables of polymorphic map types. For clarity, the presentation of the translation initially ignores polymorphic maps. Then, Sec. 4.10 discusses how the general translation scheme can be extended to support them.

As running examples, Fig. 2 shows how \( T \) translates the examples of Fig. 1.

4.1 Types

Boogie types include primitive types, instantiated type constructors, and map types.

Primitive types are int (mathematical integers), real (mathematical reals), bool (Booleans), and bv\(n\) (n-bit vectors). \( T \) translates primitive types into their Why3 analogues as shown in Tab. 3. Since Why3 offers primitive types and operations on them through libraries, \( T \) also generates import statements for the libraries that provide the same operations that are available in Boogie, such as integer to/from real conversion.

| Type  | Why3 libraries                                      |
|-------|-----------------------------------------------------|
| int   | int.Int, int.EuclideanDivision                      |
| real  | real.RealInfix, real.FromInt, real_TRUNCATE, real.PowerReal |
| bool  | bool.Bool                                           |
| bv\(n\) | bv.BitVector with constant size = n                |

Table 3. Translation of primitive types, and Why3 libraries supplying the necessary operations.

Type constructors. A Boogie type declaration using the type constructor syntax\footnote{\( T \) ignores the optional type modifier finite, since it does not seem fully supported in Boogie.} introduces a new parametric type \( T \) with parameters \( a_1, \ldots, a_m \). \( T \) translates it to an algebraic type with a parameter list.

6
braic type with constructor $T$: $T(\text{type } T\ a_1...a_m) = \text{type } 'a_1...' 'a_m$ for $m \geq 0$, where ticks ‘ identify type parameters in WhyML.

Map types. A Boogie map type $M$ declared as: $\text{type } M = [T_1, \ldots, T_n] \ U$ defines the type of a mapping from $T_1 \times \cdots \times T_n$ to $U$, for $n \geq 1$. Why3 supports maps through its library Map, hence, $T(M) = \text{map } (T(T_1), \ldots, T(T_n)) \ T(U)$, where an $n$-tuple encapsulates the $n$-type domain of $M$.

4.2 Constants

The translation of constant declarations is generally straightforward, following the scheme:

$T(\text{const } c: T) = \text{constant } c: T(T)$

Unique constants. All constants of a type $T$ declared with the modifier unique have values that are pairwise different. Thus, for $m$ constants $\text{const unique } c_1, \ldots, c_m : T$, $T$ encodes the uniqueness properties using $\binom{m}{2}$ axioms:

axiom unique_c_i_j: $c_i \neq c_j$, for $1 \leq i \neq j \leq m$.

Orders. Boogie provides the operator $<$: to express partial order over every type; $T$ introduces a polymorphic operator $<$: and axiomatizes its reflexive, antisymmetric, and transitive properties:

predicate ($<$) (x: 'a) (y: 'a)
axiom ReflexivePO: $\forall x: 'a . x <: x$
axiom AntisymmetricPO: $\forall x y: 'a . x <: y \land y <: x \rightarrow x = y$
axiom TransitivePO: $\forall x y z: 'a . x <: y \land y <: z \rightarrow x <: z$

Boogie supplies special syntax to describe a partial-order relations with a certain structure, which corresponds to a DAG where any two nodes $x$ and $y$ are connected by an edge $x \rightarrow y$ iff $x <: y$ and $y$ is a direct successor of $x$ in the order. Let $a, b, c, d, e, f$ be unique constants of the same type $T$. The Boogie syntax to specify ordering between them is in Fig. 4. D reconstructs the DAG of the order specification, and then formalizes it in axiomatic form. For example, the specifications in Fig. 4 determine the DAG in Fig. 5 which is axiomatized as in Fig. 6.

4.3 Variables

Why3 supports mutable variables through the reference type ref from theory Ref. Boogie global variable declarations become global value declarations of type ref; Boogie local variable declarations become let bindings with local scope. Thus, if $v$ is a global variable and $\_v$ is a local variable in Boogie:

global variable $T(\text{var } v: T) = \text{val } v: \text{ref } T(T)$
local variable $T(\text{var } \_v: T) = \text{let } \_v = \text{ref } (\text{any } T(T))$ in

The expression any $T$ provides a nondeterministic value of type $T$.

8 Why3’s maps, like Boogie’s, do not satisfy extensionality ([http://lists.gforge.inria.fr/pipermail/why3-club/2013-February/000572.html](http://lists.gforge.inria.fr/pipermail/why3-club/2013-February/000572.html)).

9 Uniqueness is not required but makes the order specification easier to present.
const c: T extends a, b; a and b are the only direct successors of c
const a: T extends; a has no (direct) successors
const d: T extends c complete; c has no direct predecessors other than d and any others that are explicitly specified
const e, f: T extends unique d; d is the only direct successor of both e and f, and the subgraphs that originate in e and f are disjoint

Fig. 4. Ordering specifications in Boogie (older versions of Boogie use <: instead of extends).

Fig. 5. DAG corresponding to the ordering specification of Fig. 4. Solid edges denote the successor relation; dotted edges denote allowed (but not specified) relations; the dashed line expresses disjointness of the two sub-graphs.

4.4 Functions

Boogie function declarations become WhyML function declarations:

\[
T(\text{function } f(x_1: T_1, \ldots, x_n: T_n) \text{ returns } (U)) = \text{function } f (x_1: T(T_1)) \cdots (x_n: T(T_n)): T(U) \quad (1)
\]

WhyML function definitions require, unlike Boogie’s, a variant to ensure that recursion is well-formed. Therefore, Boogie function definitions are not translated into WhyML function definitions but are axiomatized: if function f in (1) has body B, D replaces the body with the axiom \((\forall z_1: T_1, \ldots, z_n: T_n \cdot f(z_1, \ldots, z_n) = B)\) [10]

4.5 Expressions

The translation of Boogie expressions to WhyML expressions is mostly straightforward, given the translation of types described above. We describe the few cases that deserve some detail.

To take advantage of Why3’s well-formedness checks, we plan to offer translations of Boogie functions to WhyML functions as a user option in future work.

axiom (c <: a ∧ c <: b ∧ ∀ x: T \cdot c <: x \implies c = x ∨ a <: x ∨ b <: x)
axiom (V x: T \cdot \neg\{a <: x\})
axiom (d <: c ∧ ∀ x: T \cdot x <: c \implies c = x ∨ x <: d)
axiom (e <: d ∧ ∀ x: T \cdot e <: x \implies e = x ∨ d <: x)
axiom (f <: d ∧ ∀ x: T \cdot f <: x \implies f = x ∨ d <: x)
axiom (V x: T \cdot z <: e \implies \neg\{x <: f\})
axiom (V x: T \cdot z <: f \implies \neg\{x <: e\})

Fig. 6. Axiomatization of the ordering specification in Fig. 4.
Nondeterministic choice. The special value $\ast$ represent a nondeterministic Boolean choice (used in loop exit flags and conditionals); we define $T(\ast) = \text{any bool}$, which provides a nondeterministic Boolean value.

Variables. Since a Boogie variable $v$ of type $T$ turns into a value $v$ of type $\text{ref } T(T)$, occurrences of $v$ in an expression translate to $v.\text{contents}$, which represents the value attached to reference $v$.

Map expressions. $T$ translates map selection and update using functions $\text{get}$ and $\text{set}$ from theory $\text{Map}$. If $m$ is a map of type $M$ defined in Sec. 4.1, then:

| $E$                                      | $T(E)$                                      |
|------------------------------------------|---------------------------------------------|
| selection $m[e_1, \ldots, e_n]$         | $T(m)(T(e_1), \ldots, T(e_n))$             |
| update $m[e_1, \ldots, e_n := f]$       | $T(m)(T(e_1), \ldots, T(e_n))\ T(f)$       |

Lambda expressions. Boogie recently introduced lambda expressions as syntactic sugar for maps. While WhyML has lambda abstractions, they are not allowed as first-order values in programs [6]. Instead, the translation desugars lambda expression into constant maps:

$$D(\lambda x_1 : T_1, \ldots, x_n : T_n \cdot e) = \text{const lmb},$$

where $\text{const lmb} : [T_1, \ldots, T_n] \tau(e)$ is axiomatized by

$$\text{axiom } (\forall x_1 : T_1, \ldots, x_n : T_n \cdot \text{const lmb}[x_1, \ldots, x_n] = e),$$

and $\tau(e)$ is $e$'s type.

Old expression. Within a procedure’s postcondition or body, the expression $\text{old}(e)$ refers to the value of $e$ in the prestate. WhyML offers a more general construct to refer to an expression’s value at any labeled point within a procedure’s body. Hence, every WhyML procedure implementation translating a Boogie procedure implementation includes a label "begin", so that $T(\text{old}(e))$ is just $\text{old } T(e)$ within postconditions, and is $\text{at } T(e)$, "begin" within bodies.

Bitvectors. Why3’s theory $\text{BitVectors}$ does not provide all operations that are supported by Boogie. In particular, it does not support extraction expressions $b[n : m]$ (drop the $m$ least significant bits and return the next $n - m$ least significant bits) and concatenation expressions $b++c$ (the bit vector obtained by concatenating $b$ and $c$). $T$ introduces functions $\text{extract } (b : \text{bv}) (n : \text{int}) (m : \text{int}) : \text{bv}$ and $\text{cat } (b : \text{bv}) (c : \text{bv}) : \text{bv}$, and uses them to translate applications of these bit vector operators, but leaves them uninterpreted in Why3. $T$’s implementation currently supports only the bitvectors operations available in Why3’s theory $\text{BitVectors}$.

4.6 Procedures

Boogie procedures have a declaration (signature and specification) and zero or more implementations. The latter follow the general syntax of Fig. 7 (left), where a procedure $p$ with input argument $t$ and output argument $u$ has one implementation with local variable $l$ and body $B$. For simplicity of presentation, $p$ has one input argument, one output argument, and one local variable, but generalizing the description to an arbitrary number of variables is straightforward.

\footnote{Despite its name, set returns a new map rather than changing its argument’s value.}
procedure \( p(t : T \text{ where } Wt) \) returns \((u : U \text{ where } Wu)\);
requires \( R \);
free requires \( fR \);
modifies \( M \);
ensures \( E \);
free ensures \( fE \);
implementation \( p(t : T) \) returns \((u : U)\) {
var \( l : L \) where \( Wl \);
\}
\( B \) \)

val \( p(t : T(t)) : T(U) \) requires \( T(R) \)
writes \( M \)
returns \( \{ u \rightarrow T(E) \} \)
returns \( \{ u \rightarrow T(fE) \} \)
returns \( \{ u \rightarrow T(Wu) \} \)

let \( p_{impl0} (t : T(t)) : T(U) \) requires \( T(R) \) requires \( T(M) \)
returns \( \{ u \rightarrow T(E) \} \)
 RETURNS \( \{ u \rightarrow T(fE) \} \)
returns \( \{ u \rightarrow T(Wu) \} \)

let \( p_{impl0} (t : T(t)) : T(U) \) requires \( T(R) \) requires \( T(M) \)
writes \( M \)
reads \( G \) -- all globals
returns \( \{ u \rightarrow true \} \)
\( \)

\( T(var u : U; var l : l;) \)
assume \( (T(Wu)) \) \-- where of inputs
assume \( (T(Wl)) \) \-- where of locals
assume \( (T(Wu)) \) \-- where of outputs
try \( (T(R)) \)
with \ Return \ assume \( \{ true \} \ end \ T(u) \)

let \( p_{impl0}.frame (t : T(t)) : T(U) \) requires \( T(R) \) requires \( T(M) \)
writes \( M \)
reads \( G \) -- all globals
returns \( \{ u \rightarrow true \} \)
\( \)
\( T(m := m), \text{for } m \in M \)
assume \( \{ yes(g) \}, \text{for } g \in G \)
\( T(u) \)

Fig. 7. Translation of a Boogie procedure (left) into WhyML (right).

The specification of procedure \( p \) consists of preconditions \textbf{requires}, frame specification \textbf{modifies}, and postconditions \textbf{ensures}. A precondition is an assertion that callers of \( p \) must satisfy upon calling, and that every implementation of \( p \) can assume; \textbf{free} preconditions need not be satisfied by callers. A postcondition is an assertion that every implementation of \( p \) must satisfy upon terminating, and that every caller of \( p \) can assume; \textbf{free} postconditions need not be satisfied by implementations. Every implementation of \( p \) may only modify the global variables listed in \( p \)'s frame specification.

\( T \) translates a generic procedure \( p \) as shown in Fig. 7(right). The declaration of \( p \) determines \textbf{val} \( p \), which defines the semantics of \( p \) for clients: the \textbf{free} precondition \( fR \) does not feature there because clients don’t have to satisfy it, whereas both \textbf{free} and \textbf{non-free} postconditions are encoded as \textbf{returns} conditions. The implementation of \( p \) determines \textbf{let} \( p_{impl0} \), which triggers the verification of the implementation against its specification: both \textbf{free} and \textbf{non-free} preconditions are encoded, whereas the \textbf{free} postcondition \( fE \) does not feature there because implementations don’t have to satisfy it. The body introduces \textbf{let} bindings for the local variable \( l \) and for a new local variable \( u \) which represents the returned value; these declarations are translated as discussed in Sec. 4.3. Then, a series of \textbf{assume} encode the semantics of Boogie’s \textbf{where} clauses, which constrain the nondeterministic values variables can take (\( Wg \) comes from any global variables, which are visible everywhere); \( p \)'s body \( B \) is translated and wrapped inside an exception-handling block \textbf{try}, which does not do anything other than allow-
ing abrupt termination of the body’s execution upon throwing a \texttt{Return} exception (see \cite{Sec. 4.7} for details). Regardless of whether the body terminates normally or exception-ally, the last computed value of \(u\) is returned in the last line, and checked against the postcondition in \texttt{returns}. Another implementation \texttt{let p_impl0\_frame} checks the frame condition (\texttt{modifies} clause)\textsuperscript{12} It relies on the same full precondition as \texttt{p_impl0} but has postcondition \texttt{true} since \(E\) has already been checked; it includes a \texttt{writes} clause and a \texttt{reads} clause. Why3 checks that a global variable is in the \texttt{writes} clause if and only if it is written by the implementation; since Boogie’s \texttt{modifies} clause only expresses variables that \texttt{may} be written, \texttt{p_impl0\_frame} includes an assignment of every variable in \(M\) to itself so that the requirement that every variable in \(M\) is written is vacuously satisfied. When a \texttt{writes} clause is present, Why3 also requires a \texttt{reads} clause and checks that every variable in it is written, read, or both. The translation builds a \texttt{reads} clause with all global variables \(G\), and vacuously reads all of them using function \texttt{yes ‘a: bool}, which identically returns \texttt{true} for any input; this makes the \texttt{reads} clause satisfied by any implementation. In all, the modular semantics of Boogie’s procedure \(p\) is preserved.

### 4.7 Statements

\textbf{Axioms and assertions.} Boogie’s \texttt{assert e}, \texttt{assume e}, and \texttt{axiom e} statements translate to \texttt{assert \{T(e)\}}, \texttt{assume \{T(e)\}}, and \texttt{axiom A: T(e)} in WhyML.

\textbf{Assignments.} Assignments involve variables (global or local), which become mutable references in WhyML: \(T(v := e) = v.contents \leftarrow T(e)\). Boogie parallel assignments become simple assignments using \texttt{let} bindings of limited scope:

\[
T(v_1, \ldots, v_m := e_1, \ldots, e_m) = \begin{cases} 
\text{let } e'_1 = T(e_1), \ldots, e'_m = T(e_m) \text{ in } \\
T(v_1 := e'_1); \cdots; T(v_m := e'_m)
\end{cases} \tag{2}
\]

\textbf{Havoc.} An abstract function \texttt{val havoc () : ‘a provides a fresh, nondeterministic\textsuperscript{13} value of any type ‘a. It translates Boogie’s \texttt{havoc} statements following the scheme\textsuperscript{14}

\[
T(\text{havoc } u, v) = T(u)\leftarrow \text{havoc}(); T(v)\leftarrow \text{havoc}(); \texttt{assume \{T(wu)\}; assume \{T(wv)\}}
\]

where \(wu\) and \(wv\) are the \texttt{where} clauses of \(u\)’s and \(v\)’s declarations; the generalization to an arbitrary number of variables is obvious. It is important that the \texttt{assume} statements follow all the calls to \texttt{havoc}: since \(wv\) may involve \(u\)’s value, \texttt{havoc }u, \(v\) is not in general equivalent to \texttt{havoc }u; \texttt{havoc }v; the translation reflects this behavior.

\textbf{Return.} The behavior of Boogie’s \texttt{return} statement, which determines the abrupt termination of a procedure’s execution, is translated to WhyML using exception handling. An exception handling block wraps each procedure’s body, as illustrated in \fig{7} and catches an \texttt{exception Return}; thus, \(T(\text{return}) = \texttt{raise Return}\).

\textsuperscript{12}The tool \texttt{b2w} does not currently implement frame condition checks.

\textsuperscript{13}\url{http://lists.gforge.inria.fr/pipermail/why3-club/2013-April/000615.html}

\textsuperscript{14}Alternatively, we could define \(T(\text{havoc } v) = \texttt{any } T(v)\), where \(T\) is \(v\)’s type.
Jumps (branching). In addition to structured while loops (discussed below), Boogie provides jump statements of the form goto $l_1, \ldots, l_n$, which nondeterministically jump to any of the locations labeled by $l_k$. The translation must remove jump statements in a way that preserves verifiability; this rules out “global” approaches using a program counter \[11,27\], since they would require new invariants about the counter. Instead, we introduce simple heuristics that replace jumps with structured code; since the usage of jumps in Boogie programs tend to follow well-defined patterns that can be traced back to structured loops, the heuristics may be sufficient in practice\[\textsuperscript{15}\].

Consider the control-flow graph $G$ of a procedure body; each node $N$ of $G$ is a simple block: a linear piece of code with a label $\ell_N$ on the first statement, no labels anywhere else in $N$, and a goto as last statement or no goto statements at all; arrows connect $N$ to the locations mentioned in $N$’s goto statement (if $N$ has no goto, we call it a terminal node). We apply three kinds of transformations on $G$ exhaustively.

Sequencing: if $N \to M$ is the only arrow out of $N$ and the only arrow into $M$, and $M \not\to N$, replace $N$ and $M$ with the single block $N; M$ with the goto at the end of $N$ and label $\ell_M$ removed.

Choosing: if $N \to \{M_1, \ldots, M_n\}$ are the only arrows out of $N$ and the only arrows into each $M_1, \ldots, M_n$, and every $M_k$, for $1 \leq k \leq n$, is a terminal node, replace $N, M_1, \ldots, M_n$ with the single block:

$$N; \text{if } (*)\{M_1\} \text{else } \{ \text{if } (*)\{M_2\} \text{else } \cdots \text{else } \{ M_n \}\cdots \}$$

with the goto at the end of $N$ and all labels other than $\ell_N$ removed\[\textsuperscript{16}\].

Looping: replace the subgraph of Fig. 8 (left) with the structured loop to its right.

Conditionals. The translation of conditionals is straightforward:

$$T(\text{if } b \text{ then } \{BT\} \text{ else } \{BE\}) = \text{if } T(b) \text{ then } \{ T(BT) \} \text{ else } \{ T(BE) \}$$

\[\textsuperscript{15}\] $T$’s implementation currently does not support this translation of goto statements.

\[\textsuperscript{16}\] This is after Dafny’s calculational proof approach \[17\].
Loops. Fig. 9 shows the translation of a Boogie loop into a WhyML loop. An invariant marked as free can be assumed but need not be checked; correspondingly, the translation adds assumptions that ensure it holds at loop entrance and after every iteration. The exception handling block surrounding the loop in WhyML emulates the semantics of the control-flow breaking statement break: $T(\text{break}) = \text{raise Break}$.

```whyml
while (b) invariant I; free invariant fI;
{ B }
assume { $T(fI)$ }
try while $T(b)$ do
invariant { $T(I)$ }
invariant { $T(fI)$ }
$T(B)$
assume { $T(fI)$ }
end
```

Fig. 9. Translation of a Boogie loop (left) into WhyML (right).

Procedure calls. The translation of procedure calls is straightforward; for Boogie procedure $p$ in Fig. 7: $T(\text{call } r := p(e)) = T(t) \leftarrow p(T(e))$. Since WhyML function calls translating Boogie procedures use the val style of declaration rather than the recursive function style (rec), the modular semantics of procedure calls (where the behavior is entirely determined by the specification) is correctly preserved.

Call-forall. $T$ translates call-forall statements (supported in older versions of Boogie [15]) by axiomatizing their semantics:

$$D(\text{call forall Lemma(*)}) = \text{assume } (\forall t : T \bullet R(t) \Rightarrow E(t))$$

where Lemma is declared as procedure Lemma(t : T) requires R(t); ensures E(t).

4.8 Attributes

$T$ translates triggers using WhyML’s syntax:

$$T(\forall x : X \bullet (\text{trig}) E(x)) = \forall x : T(X) [T(\text{trig})].T(E(x))$$

The translation discards other application-specific attributes, which have no equivalent in Why3.

4.9 Identifiers and Visibility

Boogie is more liberal than WhyML in the range of characters that are allowed in identifier names; therefore, the translation defines an injective renaming of identifiers when necessary.
Boogie allows local declarations to shadow global declarations of entities with the same name. Since WhyML does not allow shadowing, the translation introduces fresh names for local declarations when necessary to avoid name clashes with the shadowed declarations.

While the order of declarations is immaterial in Boogie, in WhyML reference must follow declaration. Thus, the translation reorders declarations to comply with WhyML’s requirements; it also introduces a canonical order of declarations: types, global variables, functions, axioms, procedure declarations (val), procedure definitions (let), other declarations.

4.10 Polymorphic Maps

We now consider polymorphic map types, declared in Boogie as:

\[ \text{type } pM = \langle \alpha \rangle [T_1, \ldots, T_n] U \]

where \( \alpha \) is a vector \( \alpha_1, \ldots, \alpha_m \) of \( m > 0 \) type parameters, and some of the types \( T_1, \ldots, T_n \), \( U \) in \( pM \)'s definition depend on \( \alpha \). In the next paragraph, we explain why polymorphic maps cannot be translated to WhyML directly. Instead, we replace them with several monomorphic maps based on a global analysis of the types that are actually used in the Boogie program being translated. The result of this rewrite is a Boogie program without polymorphic maps, which we can translate to Why3 following the rules we previously described. The shortcoming of this approach is that it gives up modularity: verification holds only for the concrete types that are used (closed-word assumption); this seems to be necessary to express Boogie’s extremely liberal polymorphism without resorting to intricate “semantic” translations, which would likely fail verifiability.

**Boogie vs. WhyML polymorphism.** While WhyML also supports generic polymorphism, like every functional language in the ML family to which it belongs, its usage is more restrictive than Boogie’s. The first difference is that mutable maps cannot be polymorphic in WhyML; therefore, Boogie variables of polymorphic map type require a special translation. The second difference is that, in some contexts, a variable of polymorphic map type in Boogie effectively corresponds to multiple maps, one for each possible concrete type, and the different maps can be combined in the same expression. Consider, for example, a type \( \text{Mix} = \langle \alpha \rangle [\alpha] \alpha \) of maps from generic type \( \alpha \) to \( \alpha \); Boogie accepts formulas such as

\[ \text{axiom } \forall m: \text{Mix} \bullet m[0] = 1 \land m[\text{true}] \]

where \( m \) acts as a map over int in the first conjunct and as a map over bool in the second conjunct. WhyML, in contrast, always makes the type parameters explicit; hence, a logic variable of type map ‘\( \alpha \)' ‘\( \alpha \)' denotes a single map of a generic type that can only feature in expressions which do not assume anything about the concrete type that will instantiate ‘\( \alpha \)’. Note that Boogie even allows expressions that introduce inconsistencies, such as

\[ \forall \langle \beta \rangle x: \beta, y: \text{Mix} \bullet y[x] = 3 \land y[x] = \text{true} \]

(where the quantification is also type-generic), which passes typechecking but allows one to derive false.
Besides type declarations and quantifications, polymorphic maps can appear within polymorphic functions and procedures, declared as:

\[
\text{function } pF(\alpha)(x_1: T_1, \ldots, x_n: T_n) \text{ returns } (U) \quad (4)
\]

\[
\text{procedure } pP(\alpha)(x_1: T_1, \ldots, x_n: T_n) \text{ returns } (u: U) \quad (5)
\]

Precisely, two kinds of polymorphic maps may feature within polymorphic functions and procedures: polymorphic maps generic with respect to explicitly declared function or procedure parameters are similar to Why3’s, and hence different from those generic with respect to implicit type parameters declared outside the function or procedure. For example, implementations of a procedure \( p(\beta)(m: \text{ Mix}, n: [\beta] \beta) \) can select elements of any concrete type from \( n \), but only elements of parametric type \( \beta \) from \( n \).

**Type analysis.** We have seen that a Boogie polymorphic map may correspond to multiple monomorphic maps in certain contexts. The translation reifies this idea based on global type analysis: for every item (constant, program or logic variable, or formal argument) \( pm \) of polymorphic map type \( pM \) as in (3), it determines the set \( \text{types}(pm) \) of all actual types \( pm \) takes in expressions or assignments, as outlined in Tab. 10. This in turn determines the set \( \text{types}(pM) \) as the union of all sets \( \text{types}(p) \) for \( p \) of type \( pM \).

| expressions          | \( pm = [t_1, \ldots, t_n]u \) such that: |
|----------------------|------------------------------------------|
| read                 | \( pm[e_1, \ldots, e_n] \)                |
| select               | \( e_i :: ]t_1, \ldots, t_n[.pm[e_1, \ldots, e_n] :: u \) |
| update               | \( pm[e_1, \ldots, e_n := f] \)           |
| function reference   | \( f(it) \) \( : \) \( \text{it} :: ]t_1, \ldots, t_n[u \), where function \( f(pm: pM) \) |
| copy                 | \( pm := it \) \( : \) \( \text{it} :: ]t_1, \ldots, t_n[\) \( u \) |
| assignment           | \( pm[e_1, \ldots, e_n] := f \) \( : \) \( e_i :: ]t_1, \ldots, t_n[.f :: u \) |
| havoc                | \( \text{havoc } pm \) \( : \) \( \text{it} :: ]t_1, \ldots, t_n[\) \( u \) |
| procedure call in    | \( \text{call } p(it) \) \( : \) \( \text{it} :: ]t_1, \ldots, t_n[\) \( u \), where procedure \( p(pm: pM) \) |
| procedure call out   | \( \text{call } it := p() \) \( : \) \( \text{it} :: ]t_1, \ldots, t_n[\) \( u \), where procedure \( p() \) \( \text{returns}(pm: pM) \) |

**Table 10.** Each occurrence of an item \( pm \) of polymorphic map type \( pM \) determines the set \( \text{types}(pm) \) of actual types. \((x :: t \text{ denotes that } x \text{ has type } t)\)

The types in \( \text{types}(pM) \) include in general both concrete and parametric types. For example, the program of Fig. 11 (left) determines \( \text{types}(n) = \{[\text{int}]\text{int}, [\beta]\beta\} \), \( \text{types}(\alpha) = \{[\text{bool}]\text{bool}\} \), and \( \text{types}(M) = \text{types}(n) \cup \text{types}(\alpha) \), where \( \beta \) is the type parameter of the procedure \( p \)’s type signature (since \( p \) is not called anywhere, that’s the only known actual type of \( x \)). Let \( \text{conc}(pM) \) denote the set of all concrete types in \( \text{types}(pM) \).

**Desugaring polymorphic maps.** To describe how the translation replaces polymorphic maps by monomorphic maps, we introduce a pseudo-code notation that allows tuples (in round brackets) of program elements where normally only a single element is allowed. The semantics of this notation corresponds quite intuitively to multiple statements or

\footnote{A parameter’s actual type is ambiguous if the parameter appears in the map type’s codomain but not in its domain; in this case, Boogie defaults to type \text{int}.}
Declarations. For example, a variable declaration \( \text{var } (x, y): (\text{int}, \text{bool}) \) is a shorthand for declaring variables \( x: \text{int} \) and \( y: \text{bool} \); a formula \( (x, y) = (3, \text{true}) \) is a shorthand for \( x = 3 \land y \); and a procedure declaration using the tuple notation \( \text{procedure } (p.\text{int}, p.\text{bool}, p.\text{a})(x: (\text{int}, \text{bool})) \) is a shorthand for declaring two procedures \( p.\text{int}(x: \text{int}) \) and \( p.\text{bool}(x: \text{bool}) \).

We also use the following notation: given an \( n \)-vector \( \mathbf{a} = a_1, \ldots, a_n \) and a type expression \( T \) parametric with respect to \( \mathbf{a}, T_a \) denotes \( T \) with \( a_k \) substituted for \( \alpha_k \), for \( k = 1, \ldots, n \). If \( T \) is a set of types obtained from the same type expression \( T \), such as \( \text{types}(p\mathbf{M}) \) with respect to \( p\mathbf{M} \)'s definition, and \( \mathbf{i}d \) is an identifier, let \( \langle T \rangle \) denote \( T \) as a tuple, and \( \langle \mathbf{id}, T \rangle \) denote the tuple of identifiers \( \mathbf{id}, T \) such that \( T_i \) is the corresponding type in \( T \). In the example of Fig. 11 if \( T = [\alpha] \) then \( T_{\mathbf{i}d} = [\text{int}, \text{int}, \{\mathbf{i}d\} \times \mathbf{a}, (\mathbf{j} \text{.types}(\mathbf{m})) = [\text{int}, \mathbf{j} \beta] \) and \( \{\mathbf{j} \text{.types}(\mathbf{m})\} = ([\mathbf{j} \text{.int}, \mathbf{j} \beta]) \). Throughout, we also assume that an uninterpreted type \( \alpha_k \) is available for \( k = 1, \ldots, n \), that \( M_\mathbf{a} \) denotes the type expression \( [T_1, \ldots, T_n] \cup \{\mathbf{a}\} \) with each \( \alpha_k \) replaced by \( a_k \), and that \( \text{conc}^+(p\mathbf{M}) = \text{conc}(p\mathbf{M}) \cup \{M_\mathbf{a}\} \).

Declarations. Type declaration \( \langle \mathbf{3} \rangle \) desugars to several type declarations:

\[
\text{type } (p\mathbf{M}, \text{conc}^+(p\mathbf{M})) = (\text{conc}^+(p\mathbf{M}))
\] (6)

The declaration of an \( \text{item } p\mathbf{M} \), where \( p\mathbf{M} \) can be a constant, or a program or logic variable, desugars to a declaration \( (p\mathbf{M}, \text{conc}^+(p\mathbf{M})) \): \( (\text{conc}^+(p\mathbf{M})) \) of multiple items of the same kind. The declaration of a \( \text{procedure or function } g \) with an (input or output) argument \( \mathbf{x}: p\mathbf{M} \) desugars to a declaration of multiple procedures or functions \( (g, \text{conc}^+(p\mathbf{M}))(\mathbf{x}: (\text{conc}^+(p\mathbf{M}))) \) — multiple declarations each with one variant of \( \mathbf{x} \); if \( g \) has multiple arguments of this kind, the desugaring is applied recursively to each variant. Fig. 11(right) shows how the polymorphic map type \( \mathbf{M} \) and each of the items \( \mathbf{m} \) and \( \mathbf{n} \) of type \( \mathbf{M} \) become 3 monomorphic types and 3 items of these monomorphic types.

For every polymorphic function or procedure \( g \) with type parameters \( \mathbf{\beta} \), also consider any one of their arguments declared as \( \mathbf{x}: X \). If \( X \) is a type expression that depends on \( \beta \), and there exists a map type \( \{\mathbf{V}_1, \ldots, \mathbf{V}_n\} \) \( \mathbf{V}_0 \) in \( \text{types}(p\mathbf{M}) \) such that \( X = \mathbf{V}_k \) for some \( k = 0, \ldots, n \), then \( g \) becomes \( (g, \mathbf{V}_k)(\mathbf{x}: (\mathbf{V}_k)) \) — corresponding to multiple \( g \)’s each with one argument, where \( \mathbf{V}_k = \{\mathbf{V}_k \mid \{\mathbf{V}_1, \ldots, \mathbf{V}_n\} \mathbf{V}_0 \in \text{conc}^+(p\mathbf{M})\} \) is
the set of all concrete types that instantiate the \( k \)th type component. This transformation enables assigning arguments to polymorphic maps inside polymorphic functions or procedures that have become monomorphic. [Fig. 11](right) shows how argument \( x : \beta \) becomes an argument of concrete type \( \text{int}, \text{bool}, \) or \( a \), since \( \{ \beta \} \beta \in \text{types}(M) \). (As procedure \( p \) does not use \( \beta \) elsewhere, we drop it from the signature.)

**Expressions.** Every occurrence—in expressions, as l-values of assignments, and as targets of *havoc* statements—of an item \( w \) of polymorphic type \( \overline{W} \) whose declaration has been modified to remove polymorphic map types is replaced by one or more of the newly introduced monomorphic types as follows. If \( w \)'s actual type within its context is a concrete type \( C \), then we replace \( w \) with \( w_c \) such that \( \overline{W}_c = C \); otherwise, \( w \)'s actual type is a parametric type, and we replace \( w \) with the tuple \( (w_\_X) \), including all variants of \( w \) that have been introduced. In [Fig. 11](right), \( n[true] \) rewrites to just \( n_{\text{bool}}[true] \) since the concrete type is \( \text{bool} \); the assignment in \( p \)'s body, whose actual type is parametric with respect to \( \beta \), becomes an assignment involving each of the three variants of \( m \) corresponding to the three variants of \( p \) that have been introduced.

5 Implementation and Experiments

### 5.1 Implementation

We implemented the translation \( T \) described in [Sec. 4](#) as a command-line tool \( b2w \) implemented in Java 8. \( b2w \) works as a staged filter: 1) it parses and typechecks the input Boogie program, and creates a Boogie AST (abstract syntax tree); 2) it desugars the Boogie AST according to \( D \); 3) it transforms the Boogie AST into a WhyML AST according to \( E \); 4) it outputs the WhyML AST in the form of code.

Stage 1) relies on Schäf’s parsing and typechecking library Boogieamp\(^{18}\), which we modified to support access using the visitor pattern, AST in-place modifications, and the latest syntax of Boogie (e.g., for integer vs. real division\(^{19}\)). Stages 2) and 3) are implemented by multiple AST visitors, each taking care of a particular aspect of the translation, in the style of \( [27] \); the overhead of traversing the AST multiple times is negligible and improves modularity: handling a new construct (for example, in future versions of Boogie) or changing the translation of one feature only requires adding or modifying one feature-specific visitor class. A similar technique is also advocated in \( [22] \).

### 5.2 Experiments

The goal of the experiments is ascertaining that \( b2w \) can translate realistic Boogie programs producing WhyML programs that can be verified taking advantage of Why3’s multiple back-end support. The experiments are limited to fully-automated verification, and hence do not evaluate other possible practical benefits of translating programs to WhyML such as support for interactive provers and executability for testing purposes.

\(^{18}\) [https://github.com/martinschaef/boogieamp](https://github.com/martinschaef/boogieamp)

\(^{19}\) [http://boogie.codeplex.com/discussions/397357](http://boogie.codeplex.com/discussions/397357)
Programs. The experiments target a total of 194 Boogie programs from three groups according to their origin: group NAT (native) includes 29 programs that encode algorithmic verification problems directly in Boogie (as opposed to translating from a higher-level language); group OBJ (object-oriented) includes 6 programs that are based on a heap-based memory model; group TES (tests) includes 159 programs from Boogie’s test suite. Tab. 12 summarizes the sizes of the programs in each group.

| GROUP | # | LOC BOOGIE | LOC WHYML |
|-------|---|------------|-----------|
| NAT   | 29 | 20 73 253 2110 | 62 128 318 3716 |
| OBJ   | 6  | 44 146 365 878  | 90 208 446 1245  |
| TES   | 159 | 3 21 155 3272 | 36 64 290 10180 |
| **Total** | 194 | 3 34 385 6260 | 36 106 446 15141 |

Table 12. A summary of the Boogie programs used in the experiments, and their translation to WhyML using b2w. For each program GROUP, the table reports how many programs it includes (#), the minimum \(m\), mean \(\mu\), maximum \(M\), and total \(\Sigma\) length in non-comment non-blank lines of code \(\text{LOC}\) of those BOOGIE programs and of their WHYML translations.

The programs in NAT, which we developed in previous work [10,9], include several standard algorithms such as sorting and array rotation. The programs in OBJ include 2 simple examples in Java and 1 in Eiffel, encoded in Boogie by Joogie [1] and AutoProof [28] (we manually simplified AutoProof’s translation to avoid features b2w doesn’t support), and 3 algorithmic examples adapted from NAT to use a global heap in the style of object-oriented programs. Among the 515 programs that make up Boogie’s test suite, we retained in TES those that mainly exercise features supported by b2w. This meant excluding several groups of tests that exercise special options (Houdini, assertion inference, special Z3 encodings and directives, etc.), unsupported language features (bitvectors, gotos, etc.), and the correctness of typechecking (b2w assumes well-formed Boogie input). It also meant excluding 4 programs that triggered Boogie errors (a Boogie error means here a problem with the input such as a typechecking or parsing error due to a feature not activated; it is not a verification error, which just denotes a failed verification attempt and is fair game for evaluating the translation); and another 35 programs that b2w failed to translate because of unsupported features that we identified a posteriori.

Setup. Each experiment targets one Boogie program \(b\): it runs Boogie with command boogie \(b\) and a timeout of 180 seconds; it runs b2w to translate \(b\) to \(w\) in WhyML; for each SMT solver \(p\) among Alt-Ergo, CVC3, CVC4, and Z3, it runs Why3 with command why3 prove -P \(p\) \(w\), also with a timeout of 180 seconds [21]. For each run we collected the wall-clock running time, the total number of verification goals, and how many of such goals the tool verified successfully [22].

20 https://github.com/boogie-org/boogie/tree/master/Test
21 The timeouts were enforced using the Unix command timeout. We also set a 20-second timeout per procedure (option /timeLimit in Boogie) or goal (option -T in Why3).
22 The number of verification goals of each program is the same in Boogie and Why3: the number of procedure implementations.
All the experiments ran on a Ubuntu 14.04 LTS GNU/Linux box with 8-core Intel i7-4790 CPU at 3.6 GHz and 16 GB of RAM, with the following tools: Alt-Ergo 0.99.1, CVC3 2.4.1, CVC4 1.4, Z3 4.3.2, Mono 4.2.2, OCaml 4.02.3, Boogie 2.3.0.61016, and Why3 0.86.2. To account for noise, we repeated each verification three times and report the mean value of the 95th percentile of the running times.

Table 13. A summary of how Boogie performs in comparison with Why3. For each program GROUP, the table reports how many programs it includes (#), for how many of the programs Boogie verifies as many goals ($B = W$), more goals ($B > W$), or fewer goals ($B < W$) than Why3 with any of the SMT solvers; for how many of the programs both Boogie and Why3 verify none (0=0), some but not all (50=50), or all (100=100) of the goals; the last column (SPURIOUS) indicates that $b2w$’s translation never introduces spurious goals that are proved by Why3 (that is, if Boogie’s input has zero goals, so does WhyML’s translation).

| GROUP | # | $B = W$ | $B > W$ | $B < W$ | 0=0 | 50=50 | 100=100 | SPURIOUS |
|-------|---|---------|---------|---------|-----|-------|--------|----------|
| NAT   | 29 | 19      | 10      | 0       | 1   | 0     | 18     | 0        |
| OBJ   | 6  | 5       | 0       | 1       | 1   | 2     | 2      | 0        |
| TES   | 159| 137     | 21      | 1       | 71  | 21    | 45     | 0        |
| Total | 194| 161     | 31      | 2       | 73  | 23    | 65     | 0        |

Results. Tab. 13 shows a summary of the results where we compare Why3’s best performance, with any one of the four SMT solvers, against Boogie’s. The most significant result is that the WhyML translation produced by $b2w$ behaves like the Boogie original in 83% (161, $B = W$) of the experiments. This means that Boogie may fail to verify all goals (column 0=0), verify some goals and fail on others (column 50=50), or verify all goals (column 100=100); in each case, Why3 consistently verifies the same goals on $b2w$’s translation. Indeed, many programs in TES are tests that are supposed to fail verification; hence, the correct behavior of the translation is to fail as well. We also checked the failures of programs in NAT and OBJ to ascertain that $b2w$’s translation preserves correctness. Tab. 13 does not show this, but we also found another 2 programs in NAT ($inv\_survey/bst$ and $rotation/rotation\_reverse$) where Why3 proves the same goals as Boogie only by combining the results of multiple SMT solvers.

Boogie verifies more goals than Why3 in 16% (31, $B > W$) of the experiments, where it is more effective because of better features (default triggers, invariant inference, SMT encoding) or simply because of some language features that are not fully supported by $b2w$ (examples are Z3-specific annotations, which $b2w$ simply drops, and $\texttt{goto}$, which $b2w$ encodes as $\texttt{assert false}$ to ensure soundness). In 1% (2, $B < W$) of the experiments, Why3 even verifies more goals than Boogie. One program in OBJ ($\texttt{rotation\_by\_copy}$) is a genuine example where Why3’s Z3 encoding is more effective than Boogie’s: the one program in TES ($\texttt{test2/Quantifiers}$) should instead be considered spurious, as it deploys some trigger specifications that are Boogie-specific (negated triggers) or interact in a different way with the default triggers. (Procedures $S$, $U0$, and $U1$ use regular triggers whose translation to Why3 yields a different behavior, probably because of Why3’s default triggers differ from Boogie’s; procedures $W$ and $X2$ however, Boogie also succeeds given a longer timeout than the one used in the experiments.)
use negated triggers, which b2w ignore.) As this was the only program in our experiments that introduced clearly spurious behavior, the experiments provide convincing evidence that b2w’s translation preserves correctness and verifiability to a large degree.

Table 14. For each program GROUP the table reports how many programs it includes (#) and, for both Boogie and Why3 for each choice of SMT solver among ALT-Ergo, CVC3, and Z3: the mean percentage of goals verified in each program (OUTCOME \(\mu\)), how many programs were completely verified (OUTCOME \(\forall\)), and how many were not verified at all (OUTCOME \(\negexists\)), the mean \(\mu\) and total \(\Sigma\) verification TIME in seconds (including time outs), and how many programs timed out.

| GROUP | #  | Z3 | ALT-Ergo | Why3 | CVC3 | Why3 | CVC4 | Why3 | CVC4 | Why3 | Z3 | Why3 |
|-------|----|----|---------|------|------|------|------|------|------|------|----|------|
| NAT   | 29 | 0.4| 12.2    | 6    | 14   | 22.6 | 398  | 0    | 12   | 5    | 11  | 381  |
| OBJ   | 6  | 2  | 10.9    | 181  | 12   | 30.1 | 2    | 0    | 2    | 0    | 8   | 170  |
| YES   | 159| 55 | 71.3    | 2    | 45   | 25.8 | 4066 | 1    | 33   | 1    | 37  | 4360 |
| Total | 194| 82 | 71.7    | 2    | 45   | 25.8 | 4066 | 1    | 33   | 1    | 37  | 4360 |

Tab. 14 provides data about the experiments’ running times, and differentiates the performance of each SMT solver with Why3. Z3 is the most effective SMT solver in terms of programs it could completely verify (columns \(\forall\)), followed by Alt-Ergo. While CVC3 is generally the least effective, it has the advantage of returning very quickly (only 0.2 seconds of average running time), even more quickly than Z3 in Boogie. CVC4 falls somewhere in the middle, in terms both of effectiveness and of running time. Boogie’s responsiveness remains excellent if balanced against its effectiveness; a better time-effectiveness of Why3 with Alt-Ergo and Z3 could be achieved by setting tight per-goal timeouts (in most cases, verification attempts that last longer than a few seconds do not eventually succeed).

6 Discussion

The current implementation of the translation \(T\) has some limitations that somewhat restrict its applicability. As we already mentioned in the paper, some features of the Boogie language are not supported (bitvectors, gotos), or only partially supported (polymorphic mappings); and frame specifications are assumed. All of these are, however, limitations of the current prototype implementation only, and we see no fundamental hurdles to extending b2w along the lines of the definition of \(T\) in Sec. 4.

Since b2w also takes great care to confine the effect of translating Boogie programs that include unsupported features, and to fail when it cannot produce a correct translation, it still largely preserves correctness (soundness, in particular). For example, a goto statement is rendered as assert false; therefore, the translated program verifies only if the goto is never executed in the original program, which ensures soundness. On the other hand, our experiments also demonstrate that the translation \(T\), as implemented by b2w, largely meets the other goal of preserving verifiability: even if the experimental subjects all are idiomatic Boogie programs written independent of the translation effort, 83% of the translated programs behave in Why3 as they do in Boogie.
In future work, we will address the features of Boogie that are still not satisfactorily supported by b2w. We will also devise strategies to take advantage of Why3’s multi-prover support. Other possible directions include formalizing the translation to prove that it preserves correctness; and devising a reverse translation from WhyML to Boogie.

References

1. S. Arlt and M. Schäf. Joogie: Infeasible code detection for Java. In Proceedings of CAV, volume 7358 of LNCS, pages 767–773. Springer, 2012.
2. M. Barnett, B. E. Chang, R. DeLine, B. Jacobs, and K. R. M. Leino. Boogie: A modular reusable verifier for object-oriented programs. In Proceedings of FMCO, volume 4111 of LNCS, pages 364–387. Springer, 2006.
3. M. Barnett, M. Fähndrich, K. R. M. Leino, P. Müller, W. Schulte, and H. Venter. Specification and verification: the Spec# experience. Commun. ACM, 54(6):81–91, 2011.
4. S. Böhme, K. R. M. Leino, and B. Wolff. HOL-Boogie – an interactive prover for the Boogie program-verifier. In Proceedings of TPHOLs, volume 5170, pages 150–166. Springer, 2008.
5. Z. Cheng, R. Monahan, and J. F. Power. A sound execution semantics for ATL via translation validation – research paper. In Proceedings of ICMT, volume 9152 of LNCS, pages 133–148. Springer, 2015.
6. M. Clochard, J. Filliâtre, C. Marché, and A. Paskevich. Formalizing semantics with an automatic program verifier. In Proceedings of VSTTE, volume 8471 of LNCS, pages 37–51, 2014.
7. L. M. de Moura and N. Bjørner. Z3: An efficient SMT solver. In Proceedings of TACAS, pages 337–340, 2008.
8. J. Filliâtre and A. Paskevich. Why3 – where programs meet provers. In Proceedings of ESOP, volume 7792 of LNCS, pages 125–128. Springer, 2013.
9. C. A. Furia. Rotation of sequences: Algorithms and proofs. http://arxiv.org/abs/1406.5453 June 2014.
10. C. A. Furia, B. Meyer, and S. Velder. Loop invariants: Analysis, classification, and examples. ACM Computing Surveys, 46(3):Article 34, 2014.
11. D. Harel. On folk theorems. Commun. ACM, 23(7):379–389, 1980.
12. S. Heule, I. T. Kassios, P. Müller, and A. J. Summers. Verification condition generation for permission logics with abstract predicates and abstraction functions. In Proceedings of ECOOP, volume 7920 of LNCS, pages 451–476. Springer, 2013.
13. G. Klein, J. Andronick, K. Elphinstone, G. Heiser, D. Cock, P. Derrin, D. Elkaduwe, K. Engelhardt, R. Kolanski, M. Norrish, T. Sewell, H. Tuch, and S. Winwood. seL4: formal verification of an operating-system kernel. Commun. ACM, 53(6):107–115, 2010.
14. R. Kumar, M. O. Myreen, M. Norrish, and S. Owens. CakeML: a verified implementation of ML. In Proceedings of POPL, pages 179–192. ACM, 2014.
15. K. R. M. Leino. This is Boogie 2. 2008. http://goo.gl/0sh6q
16. K. R. M. Leino. Developing verified programs with Dafny. In Proceedings of ICSE, pages 1488–1490. ACM, 2013.
17. K. R. M. Leino and N. Polikarpova. Verified calculations. In Proceedings of VSTTE, pages 170–190, 2013.
18. X. Leroy. Formal verification of a realistic compiler. Commun. ACM, 52(7):107–115, 2009.
19. S. Liew, C. Cadar, and A. Donaldson. Symbooglix: A symbolic execution engine for Boogie programs. In Proceedings of ICST. IEEE, 2016.
20. T. Mens and P. Van Gorp. A taxonomy of model transformation. Electr. Notes Theor. Comput. Sci., 152:125–142, 2006.
21. N. Polikarpova, C. A. Furia, and S. West. To run what no one has run before: Executing an intermediate verification language. In Proceedings of RV, volume 8174 of LNCS, pages 251–268. Springer, 2013.

22. D. Sarkar, O. Waddell, and R. K. Dybvig. Educational pearl: A nanopass framework for compiler education. J. Funct. Program., 15(5):653–667, 2005.

23. P. H. Schmitt, M. Ulbrich, and B. Weiβ. Dynamic frames in Java dynamic logic. In Proceedings of FoVeOOS, volume 6528 of LNCS, pages 138–152. Springer, 2011.

24. L. Segal and P. Chalin. A comparison of intermediate verification languages: Boogie and Sireum/Pilar. In Proceedings of VSTTE, volume 7152 of LNCS, pages 130–145. Springer, 2012.

25. P. Stevens. A landscape of bidirectional model transformations. In Proceedings of GTTSE, volume 5235 of LNCS, pages 408–424. Springer, 2008.

26. M. Trudel, C. A. Furia, M. Nordio, and B. Meyer. Really automatic scalable object-oriented reengineering. In Proceedings of ECOOP, volume 7920 of LNCS, pages 477–501. Springer, 2013.

27. M. Trudel, C. A. Furia, M. Nordio, B. Meyer, and M. Oriol. C to O-O translation: Beyond the easy stuff. In Proceedings of WCRE, pages 19–28. IEEE Computer Society, October 2012.

28. J. Tschannen, C. A. Furia, M. Nordio, and N. Polikarpova. AutoProof: Auto-active functional verification of object-oriented programs. In Proceedings of TACAS, volume 9035 of LNCS, pages 566–580. Springer, 2015.
| NAME | LOC | Z3 | ALT-ERGO | CVC3 | CVC4 | Z3 |
|------|-----|----|----------|------|------|----|
| inv_survey/array_partitioning_v1 | 42 | 100 | 0.5 | 100 | 50 | 20.1 | 50 | 20.3 |
| inv_survey/array_partitioning_v2 | 53 | 100 | 0.4 | 125 | 100 | 0.3 | 50 | 0.1 |
| inv_survey/bst | 125 | 100 | 0.5 | 204 | 100 | 0.3 | 86 | 0.2 |
| inv_survey/bubble_sort.basic | 49 | 100 | 0.5 | 113 | 100 | 0.2 | 50 | 0.1 |
| inv_survey/bubble_sort_improved | 53 | 100 | 0.5 | 118 | 100 | 0.6 | 50 | 0.2 |
| inv_survey/comb_sort | 56 | 100 | 0.4 | 124 | 100 | 0.3 | 50 | 0.2 |
| inv_survey/dutch_flag | 133 | 100 | 0.5 | 50 | 20.1 | 50 | 0.2 |
| inv_survey/insertion.sort | 97 | 100 | 0.3 | 97 | 100 | 0.7 | 50 | 0.1 |
| inv_survey/Levenshtein.distance | 100 | 0.4 | 91 | 100 | 0.4 | 0 | 100 | 0.9 |
| inv_survey/max.of.array.v1 | 20 | 100 | 0.4 | 66 | 100 | 0.1 | 0 | 21.2 |
| inv_survey/max_of_array_v2 | 20 | 100 | 0.4 | 66 | 100 | 0.2 | 0 | 21.2 |
| inv_survey/partition | 63 | 100 | 0.4 | 177 | 100 | 0.8 | 50 | 0.1 |
| inv_survey/planeau | 34 | 100 | 0.5 | 84 | 0 | 20.1 | 0 | 1 |
| inv_survey/reverse | 68 | 100 | 0.4 | 131 | 100 | 0.3 | 50 | 1 |
| inv_survey/selection.sort | 72 | 100 | 0.4 | 160 | 100 | 4.9 | 33 | 0.2 |
| inv_survey/sequential.search.v1 | 28 | 100 | 0.4 | 72 | 0 | 20.1 | 0 | 1 |
| inv_survey/sequential_search_v2 | 23 | 100 | 0.4 | 70 | 0 | 20.1 | 0 | 1 |
| inv_survey/sort_of_array | 21 | 100 | 0.3 | 62 | 100 | 0.1 | 0 | 21.2 |
| inv_survey/welfare_crook | 44 | 100 | 0.3 | 86 | 100 | 0.1 | 0 | 100 |
| rotation/rotation.copy | 57 | 100 | 0.4 | 128 | 33 | 40.1 | 33 | 0.2 |
| rotation/rotation_copy_plain | 41 | 100 | 0.3 | 80 | 0 | 20.1 | 0 | 1 |
| rotation/rotation_reverse | 201 | 90 | 0.4 | 318 | 40 | 120.5 | 100 | 0.4 |
| rotation/rotation_swap-1_3 | 48 | 0 | 0.3 | 88 | 0 | 20.1 | 0 | 1 |
| rotation/rotation_swap-2_3 | 175 | 60 | 0.4 | 201 | 20 | 80.2 | 200 | 2 |
| rotation/rotation_swap-3_3 | 47 | 100 | 0.3 | 96 | 67 | 20.1 | 67 | 0.2 |
| rotation/rotation_swap_iterative-1_2 | 152 | 100 | 0.5 | 184 | 33 | 40.2 | 33 | 0.2 |
| rotation/rotation_swap_iterative-2_2 | 253 | 60 | 0.4 | 224 | 20 | 80.2 | 200 | 2 |

Table 15. Results for the programs in groups NAT (above the horizontal line) and OBJ (below it) in the experiments. For each program (NAME) the Boogie program length in non-comment non-empty lines of code (LOC) and the length of its WHY3 translation; and, for both Boogie and WHY3, for each choice of SMT solver among ALT-ERGO, CVC3, and Z3: the percentage of goals verified in each program (% V.) and the verification time (T) in seconds (with a timeout of 180 seconds).
| NAME                     | LOC | Z3 | ALT-ERGO | CVC3 | CVC4 | Z3 |
|-------------------------|-----|----|----------|------|------|----|
| doomed/doored           | 73  | 43.3 | 185 | 43.0 | 80.2 | 43 | 84.7 | 43 | 80.7 |
| doomed/notdoored        | 43  | 50 | 0.3 | 107 | 50 | 40.1 | 50 | 0.1 | 50 | 42.4 | 50 | 40.4 |
| doomed/saske0           | 61  | 67.3 | 148 | 67 | 40.1 | 67 | 0.2 | 67 | 42.4 | 67 | 40.4 |
| lock/Lock               | 86  | 100.0 | 143 | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| lock/LockIncorrect      | 34  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed             | 41  | 100.0 | 3 | 108.0 | 100 | 0.1 | 100 | 0.2 | 100 | 0.1 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| lock/LockIncorrect      | 34  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/saske0             | 41  | 100.0 | 3 | 108.0 | 100 | 0.1 | 100 | 0.2 | 100 | 0.1 |
| snap/saske0/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/saske0/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/saske0/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/3           | 38  | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| snap/doomed/notdoomed   | 12  | 67.3 | 60  | 67 | 20.1 | 67 | 0.1 | 67 | 21.3 | 67 | 20.2 |
| snap/doomed/smoke0      | 10  | 50 | 0.3 | 64  | 0.0 | 20.1 | 0.0 | 0.1 | 21.2 | 0.0 | 20.2 |
| NAME                        | LOC | Z3 | ALT-ERGO | CVC3 | CVC4 | Z3 |
|-----------------------------|-----|----|----------|------|------|----|
| snapshots/Snapshots31.v1    | 11  | 0  | 0.4      | 43   | 0.1  | 0  |
| snapshots/Snapshots32.v8    | 12  | 0  | 0.4      | 44   | 100  | 0.1|
| snapshots/Snapshots32.v1    | 9   | 0  | 0.3      | 41   | 0.1  | 0  |
| snapshots/Snapshots32.v8    | 12  | 0  | 0.3      | 44   | 100  | 0.1|
| snapshots/Snapshots33.v1    | 6   | 0  | 0.4      | 37   | 100  | 0.1|
| snapshots/Snapshots33.v8    | 6   | 0  | 0.3      | 38   | 100  | 0.1|
| snapshots/Snapshots34.v1    | 5   | 0  | 0.3      | 36   | 0.1  | 0  |
| snapshots/Snapshots34.v8    | 6   | 0  | 0.3      | 38   | 100  | 0.1|
| snapshots/Snapshots35.v1    | 5   | 0  | 0.3      | 36   | 0.1  | 0  |
| snapshots/Snapshots35.v8    | 11  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots36.v1    | 11  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots36.v8    | 7   | 0  | 0.3      | 42   | 100  | 0.1|
| snapshots/Snapshots37.v1    | 10  | 0  | 0.4      | 43   | 100  | 0.1|
| snapshots/Snapshots38.v1    | 11  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots38.v2    | 11  | 0  | 0.3      | 44   | 100  | 0.1|
| snapshots/Snapshots39.v1    | 10  | 0  | 0.3      | 43   | 100  | 0.1|
| snapshots/Snapshots39.v8    | 11  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots40.v1    | 11  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots40.v8    | 12  | 0  | 0.4      | 45   | 0.1  | 0  |
| snapshots/Snapshots41.v1    | 31  | 0  | 0.3      | 39   | 100  | 0.1|
| snapshots/Snapshots41.v8    | 9   | 0  | 0.3      | 39   | 0.1  | 0  |
| snapshots/Snapshots42.v1    | 12  | 0  | 0.3      | 43   | 0.1  | 0  |
| snapshots/Snapshots42.v8    | 14  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots43.v1    | 14  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots43.v8    | 11  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots44.v1    | 13  | 0  | 0.3      | 44   | 0.1  | 0  |
| snapshots/Snapshots44.v8    | 11  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots45.v1    | 37  | 0  | 0.3      | 38   | 0.1  | 0  |
| snapshots/Snapshots45.v8    | 23  | 0  | 0.3      | 39   | 0.1  | 0  |
| snapshots/Snapshots46.v1    | 27  | 0  | 0.3      | 39   | 0.1  | 0  |
| snapshots/Snapshots46.v8    | 31  | 0  | 0.3      | 40   | 0.1  | 0  |
| snapshots/Snapshots47.v1    | 44  | 0  | 0.3      | 40   | 0.1  | 0  |
| snapshots/Snapshots47.v8    | 12  | 0  | 0.3      | 43   | 0.1  | 0  |
| snapshots/Snapshots48.v1    | 14  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots48.v8    | 14  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots49.v1    | 11  | 0  | 0.3      | 45   | 0.1  | 0  |
| snapshots/Snapshots49.v8    | 11  | 0  | 0.3      | 45   | 0.1  | 0  |
| test13/InTrsf2/LoopInvasionBPL | 23  | 0  | 0.3      | 46   | 0.1  | 0  |
| test13/LoopUnroll          | 15  | 0  | 0.3      | 69   | 0.1  | 0  |
| test15/IntTest             | 3   | 0  | 0.3      | 36   | 0.1  | 0  |
| test15/ModelTest           | 10  | 0  | 0.3      | 49   | 0.1  | 0  |
| test15/NullModel           | 5   | 0  | 0.3      | 39   | 0.1  | 0  |
| test16/LoopEnforce         | 63  | 0  | 0.3      | 124  | 0.1  | 0  |
| test17/ContractInfer       | 21  | 0  | 0.3      | 68   | 0.1  | 0  |
| test2/AssertVerifiedUnder0 | 26  | 0  | 0.3      | 100  | 0.1  | 0  |
| test2/AssumeEnsures        | 53  | 0  | 0.3      | 124  | 0.1  | 0  |
| test2/AssumptionVariables0 | 44  | 0  | 0.3      | 137  | 0.1  | 0  |
| test2/Axioms               | 24  | 0  | 0.3      | 73   | 0.1  | 0  |
| test2/B                    | 65  | 0  | 0.3      | 112  | 0.1  | 0  |
| test2/Call                 | 49  | 0  | 0.3      | 117  | 0.1  | 0  |
| test2/ContractEvaluationOrder | 26  | 0  | 0.3      | 105  | 0.1  | 0  |
| test2/CutBackEdge          | 35  | 0  | 0.3      | 96   | 0.1  | 0  |
| test2/Ensure               | 61  | 0  | 0.3      | 168  | 0.1  | 0  |
| test2/False                | 14  | 0  | 0.3      | 54   | 0.1  | 0  |
| test2/FormalTerm2          | 36  | 0  | 0.3      | 104  | 0.1  | 0  |
| test2/FreeCall             | 59  | 0  | 0.3      | 185  | 0.1  | 0  |
| test2/Implies              | 28  | 0  | 0.3      | 97   | 0.1  | 0  |
| test2/InvariantUnder0      | 42  | 0  | 0.3      | 146  | 0.1  | 0  |
| test2/LoopInvasionAssume   | 15  | 0  | 0.3      | 44   | 0.1  | 0  |
| test2/Passification        | 155 | 0  | 0.3      | 200  | 0.1  | 0  |
| test2/Quantifiers          | 122 | 0  | 0.3      | 254  | 0.1  | 0  |
| test2/SelectiveChecking    | 31  | 0  | 0.3      | 121  | 0.1  | 0  |
| test2/1k.hack              | 17  | 0  | 0.3      | 44   | 0.1  | 0  |
| test2/TimesOut             | 71  | 0  | 0.3      | 156  | 0.1  | 0  |
| test2/TypeEncodingM        | 19  | 0  | 0.3      | 60   | 0.1  | 0  |
| NAME | LOC | % V. | T | Z3 | ALT-ERGO | CVC3 | CVC4 | Z3 | T |
|------|-----|------|---|----|----------|------|------|----|---|
| test21/BooleanQuantification2 | 9 | 0.3 | 46 | 0 | 20.1 | 0 | 0.0 | 0 | 21.2 | 0 | 20.2 |
| test21/Boxing | 15 | 0.5 | 49 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/Colors | 7 | 0.3 | 48 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/DisjointDomains | 21 | 0.4 | 81 | 0 | 60.1 | 0 | 0.2 | 0 | 63.5 | 0 | 60.5 |
| test21/EmptySetBug | 18 | 0.4 | 54 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/FunAxioms | 24 | 50.5 | 81 | 50 | 20.1 | 50 | 0.1 | 50 | 21.3 | 50 | 20.2 |
| test21/FunAxioms2 | 13 | 0.3 | 49 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/InterestingExamples3 | 17 | 67.4 | 71 | 33 | 40.1 | 33 | 0.2 | 33 | 42.4 | 33 | 40.4 |
| test21/InterestingExamples5 | 9 | 100.3 | 45 | 100 | 0.1 | 0 | 100 | 0.1 | 100 | 0.1 |
| test21/Keywords | 5 | 100.3 | 38 | 100 | 0.1 | 100 | 0.1 | 100 | 0.1 |
| test21/LargeLiterals0 | 12 | 0.3 | 46 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/LetSorting | 11 | 100.3 | 43 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/Maps2 | 14 | 100.3 | 52 | 100 | 0.1 | 0 | 100 | 0.2 | 0 | 20.2 |
| test21/Orderings | 13 | 50.4 | 59 | 0 | 40.1 | 0 | 0.1 | 0 | 42.4 | 0 | 40.4 |
| test21/Orderings2 | 11 | 0.4 | 49 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/Orderings3 | 22 | 0.4 | 78 | 0 | 40.1 | 0 | 0.1 | 0 | 42.4 | 0 | 40.4 |
| test21/Orderings4 | 7 | 0.4 | 47 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test21/PolycList | 35 | 0.4 | 91 | 0 | 40.1 | 0 | 0.2 | 0 | 42.4 | 0 | 40.4 |
| test21/Triggers0 | 34 | 50.4 | 92 | 50 | 20.1 | 50 | 0.1 | 42.4 | 50 | 20.2 |
| test21/Triggers1 | 12 | 0.4 | 50 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test7/MultipleErrors | 14 | 0.3 | 42 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| test7/NestedVC | 20 | 50.3 | 61 | 0 | 40.1 | 0 | 0.1 | 0 | 42.4 | 0 | 40.3 |
| test7/UnreachableBlocks | 34 | 100.3 | 79 | 50 | 40.1 | 50 | 0.1 | 50 | 42.4 | 50 | 40.4 |
| textbook/Bubble | 47 | 100.4 | 110 | 0 | 20.1 | 0 | 0.1 | 0 | 21.3 | 0 | 20.2 |
| textbook/DutchFlag | 47 | 100.3 | 92 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |
| textbook/Find | 27 | 100.3 | 72 | 50 | 20.1 | 50 | 0.1 | 50 | 21.3 | 50 | 20.3 |
| textbook/McCarthy-91 | 11 | 100.3 | 47 | 100 | 0.1 | 100 | 0.1 | 100 | 0.1 |
| textbook/TuringFactorial | 27 | 100.3 | 81 | 0 | 20.1 | 0 | 0.1 | 0 | 21.2 | 0 | 20.2 |

Table 16: Results for the programs in group TES in the experiments. The measures are the same as in Tab. 15.