Reflections on Verifying Software with Whiley

David J. Pearce(✉) and Lindsay Groves

Victoria University of Wellington, Wellington, New Zealand
{djp,lindsay}@ecs.vuw.ac.nz

Abstract. An ongoing challenge for computer science is the development of a tool which automatically verifies that programs meet their specifications, and are free from runtime errors such as divide-by-zero, array out-of-bounds and null dereferences. Several impressive systems have been developed to this end, such as ESC/Java and Spec#, which build on existing programming languages (e.g. Java, C#). Unfortunately, such languages were not designed for this purpose and this significantly hinders the development of practical verification tools for them. For example, soundness of verification in these tools is compromised (e.g. arithmetic overflow is ignored). We have developed a programming language specifically designed for verification, called Whiley, and an accompanying verifying compiler. In this paper, we reflect on a number of challenges we have encountered in developing a practical system.

1 Introduction

The idea of verifying that a program meets a given specification for all possible inputs has been studied for a long time. Hoare's Verifying Compiler Grand Challenge was an attempt to spur new efforts in this area to develop practical tools [1]. A verifying compiler "*uses automated mathematical and logical reasoning to check the correctness of the programs that it compiles*". Hoare's intention was that verifying compilers should fit into the existing development tool chain, "*to achieve any desired degree of confidence in the structural soundness of the system and the total correctness of its more critical components*". For example, commonly occurring errors could be automatically eliminated, such as: *division-by-zero*, *integer overflow*, *buffer overruns* and *null dereferences.*

The first systems that could be reasonably considered as verifying compilers were developed some time ago, and include that of King [2], Deutsch [3], the Gypsy Verification Environment [4] and the Stanford Pascal Verifier [5]. Following on from these, was the Extended Static Checker for Modula-3 [6]. Later, this became the Extended Static Checker for Java (ESC/Java) — a widely acclaimed and influential work in this area [7]. Building on this success was the Java Modeling Language (and its associated tooling) which provided a standard notation for specifying functions in Java [8,9]. More recently, the Spec# language [10–12] was developed on top of C#, whilst Dafny was developed from scratch to simplify verification [13,14].

Continuing this line of work, we are developing a verifying compiler for the Whiley programming language [15–18]. Whiley is an imperative language designed to simplify verification and to be suitable for safety-critical systems. For example, Whiley uses unbounded integer and rational arithmetic in place of e.g. IEEE 754 floating point (which is notoriously difficult to reason about [19]). Likewise, pure (i.e. mathematical) functions are distinguished from those which may have side-effects. Our goal is to develop a verifying compiler which can automatically establish a Whiley program as: *correct with respect to its declared specifications*; and, *free from runtime error* (e.g. divide-by-zero, array index-out-of-bounds, etc.). More complex properties, such as establishing termination, are not considered (although would be interesting future work). Finally, the Whiley verifying compiler is released under an open source license (BSD), can be downloaded from http://whiley.org and forked at http://github.com/DavePearce/Whiley/. Note that development of the language and compiler is ongoing and should be considered a work-in-progress.

Contribution. The seminal works by Floyd [20], Hoare [21], Dijkstra [22], and others provide a foundation upon which to develop tools for verifying software. However, in developing a verifying compiler for Whiley, we have encountered some gaps between theory and practice. In this paper, we reflect on our experiences using Whiley to verify programs and, in particular, highlight a number of challenges we encountered.

2 Language Overview

We begin by exploring the Whiley language and highlighting some of the choices made in its design. For now, we stick to the basic issues of syntax, semantics and typing and, in the following section, we will focus more specifically on using Whiley for verification. Perhaps one of our most important goals was to make the system as accessible as possible. To that end, the language was designed to superficially resemble modern imperative languages (e.g. Python), and this decision has significantly affected our choices.

Overview. Languages like Java and C# permit arbitrary side-effects within methods and statements. This presents a challenge when such methods may be used within specifications. Systems like JML and Spec# require that methods used in specifications are *pure* (i.e. side-effect free). An important challenge here is the process of checking that a function is indeed pure. A significant body of research exists on checking functional purity in object-oriented languages (e.g. [23,24]). Much of this relies on interprocedural analysis, which is too costly for a verifying compiler. To address this, Whiley is a hybrid object-oriented and functional language which divides into *a functional core* and an *imperative outer layer*. Everything in the functional core can be modularly checked as being side-effect free.

Value Semantics. The prevalence of pointers — or references — in modern programming languages (e.g. Java, C++, C#) has been a major hindrance in the development of verifying compilers. Indeed, Mycroft recently argued that (unrestricted) pointers should be "considered harmful" in the same way that Dijkstra considered goto harmful [25]. To address this, all compound structures in Whiley (e.g. lists, sets, and records) have *value semantics*. This means they are passed and returned by-value (as in Pascal, MATLAB or most functional languages). But, unlike functional languages (and like Pascal), values of compound types can be updated in place. Whilst this latter point may seem unimportant, it serves a critical purpose: to give Whiley the appearance of a modern *imperative* language when, in fact, the functional core of Whiley is pure. This goes towards our goal of making the language as accessible as possible.

Value semantics implies that updates to a variable only affect that variable, and that information can only flow out of a function through its return value. Consider:

```
int f([int] xs):
    ys = xs
    xs[0] = 1
    ...
```

Here, `[int]` represents a list of `int`s (i.e. a variable-length array). The semantics of Whiley dictate that, having assigned `xs` to `ys` as above, the subsequent update to `xs` does not affect `ys`. Arguments are also passed by value, hence `xs` is updated inside `f()` and this does not affect `f`'s caller. That is, `xs` is not a *reference* to a list of `int`; rather, it *is* a list of `int`s and assignments to it do not affect state visible outside of `f()`.

Unbounded Arithmetic. Modern languages typically provide fixed-width numeric types, such as 32 bit twos-complement integers, or 64-bit IEEE 754 floating point numbers. Such data types are notoriously difficult for an automated theorem prover to reason about [19]. Systems like JML and Spec# assume (unsoundly) that numeric types do not overflow or suffer from rounding. To address this, Whiley employs *unbounded integers* and *rationals* in place of their fixed-width alternatives and, hence, does not suffer the limitations of soundness discussed above.

Flow Typing & Unions. An unusual feature of Whiley is the use of a *flow typing system* (see e.g. [18,26,27]) coupled with *union types* (see e.g. [28,29]). This gives Whiley the look-and-feel of a dynamically typed language (e.g. Python). For example, local variables are never explicitly declared; rather, they are declared by assignment. To illustrate, we consider null references. These have been a significant source of error in languages like Java and C#. The issue is that, in such languages, one can treat *nullable* references as though they are *non-null* references [30] (Hoare calls this his billion dollar mistake [31]). Although many approaches have been proposed (e.g. [32–36]), Whiley's type system provides an elegant solution:

```
int|null indexOf(string str, char c):
   ...

[string] split(string str, char c):
   idx = indexOf(str,c)
   // idx has type null|int
   if idx is int:
       // idx now has type int
       below = str[0..idx]
       above = str[idx..]
       return [below,above]
   else:
       // idx now has type null
       return [str]
```

Here, `indexOf()` returns the first index of a character in the string, or **`null`** if there is none. The type **`int|null`** is a union type, meaning it is either an **`int`** *or* **`null`**. After the assignment "`idx = indexOf(str,c)`" variable `idx` has type **`int|null`**. The system ensures **`null`** is never dereferenced because the type **`int|null`** cannot be treated as an **`int`**. Instead, one must first check it *is* an **`int`** using e.g. "`idx` **`is int`**" (similar to `instanceof` in Java). Furthermore, Whiley's flow type system automatically retypes variables through such conditionals. In the example above, the variable `idx` is automatically retyped by "`idx` **`is int`**" to have type **`int`** on the true branch, and type **`null`** on the false branch. This prevents the needs for explicit casts after a type test (as required in e.g. Java).

As another example, we consider unions of the same kind (e.g. a union of record types, or a union of list types). These expose commonality and are called *effective unions* (e.g. an effective record type). In the case of a union of records, fields common to all records are exposed:

```
define Circle as { int x, int y, int radius }
define Rectangle as { int x, int y, int width, int height }
define Shape as Circle | Rectangle
```

A `Shape` is either a `Rectangle` or a `Circle` (which are both record types). Any variable of type `Shape` exposes fields `x` and `y` *because these are common to all cases*. Finally, it's interesting to note that the notion of an effective record type is similar, in some ways, to that of the *common initial sequence* found in C [37].

Recursive Data Types. Whiley provides recursive types which are similar to the abstract data types found in functional languages (e.g. Haskell, ML, etc.). For example:

```
define LinkedList as null | {int data, LinkedList next}

int length (LinkedList l):
  if l is null:
     // l now has type null
     return 0
  else:
    // l now has type {int data, LinkedList next}
    return 1 + length (l.next)
```

Here, we again see how flow typing gives an elegant solution. More specifically, on the false branch of the type test "`l is null`", variable `l` is automatically retyped to `{int data, LinkedList next}` — thus ensuring the subsequent dereference of `l.next` is safe. No casts are required as would be needed for a conventional imperative language (e.g. Java). Finally, like all compound structures, the semantics of Whiley dictates that recursive data types are passed by value (or, at least, appear to be from the programmer's perspective).

Performance. Many of our choices (e.g. value semantics and unbounded arithmetic) have a potentially detrimental effect on performance. Whilst this is a trade-off we accept, there are existing techniques which can help. For example, we can use reference counting to minimise unnecessary cloning of compound structures (see e.g. [38]). Furthermore, we can exploit the specifications that are an integral part of Whiley programs. That is, when the compiler can prove an integer will remain within certain bounds, it is free to use a fixed-width type (e.g. a 32 bit **`int`**).

3 Verification

A key goal of the Whiley project is to develop an open framework for research in automated software verification. As such, we now explore verification in Whiley.

Example 1 — Preconditions and Postconditions. The following Whiley code defines a function accepting a positive integer and returning a non-negative integer (i.e. natural number):

```
int f(int x) requires x > 0, ensures $ >= 0 && $ != x:
    return x-1
```

Here, the function `f()` includes a **`requires`** and **`ensures`** clause which correspond (respectively) to its *precondition* and *postcondition*. In this context, $ represents the return value, and must only be used in the **`ensures`** clause. The Whiley compiler statically verifies that this function meets its specification.

A slightly more unusual example is the following:

```
int f(int x) requires x >= 0, ensures 2*$ >= $:
    return x
```

In this case, we have two alternative (and completely equivalent) definitions for a natural number. We can see that the precondition is equivalent to the postcondition by subtracting $ from both sides. The Whiley compiler is able to reason that these are equivalent and statically verifies that this function is correct.

Example 2 — Conditionals. Variables in Whiley are described by their underlying type and those constraints which are shown to hold. As the automated theorem prover learns more about a variable, it automatically takes this into consideration when checking constraints are satisfied. For example:

```
int abs (int x) ensures $ >= 0:
    if x >= 0:
        return x
    else:
        return -x
```

The Whiley compiler statically verifies that this function always returns a non-negative integer. This relies on the compiler to reason correctly about the implicit constraints implied by the conditional. A similar, but slightly more complex example is that for computing the maximum of two integers:

```
int max(int x, int y)
        ensures $ >= x && $ >= y && ($==x || $==y):
    if x > y:
        return x
    else:
        return y
```

Again, the Whiley compiler statically verifies this function meets its specification. Here, the body of the function is almost completely determined by the specification — however, in general, this it not the case.

Example 3 — Bounds Checking. An interesting example which tests the automated theorem prover more thoroughly is the following:

```
null|int indexOf (string str, char c):
    for i in 0..|str|:
        if str[i] == c:
            return i
    return null
```

The access `str[i]` must be shown as within the bounds of the list `str`. Here, the range constructor `X..Y` returns a list of consecutive integers from `X` upto, but not including `Y` (and, futhermore, if `X >= Y` then the empty list is returned). Hence, this function cannot cause an out-of-bounds error and the Whiley compiler statically verifies this.

In fact, the specification for `indexOf()` could be made more precise as follows:

```
null|int indexOf (string str, char c)
    ensures $ == null || (0 <= $ && $ < |str|):
    ...
```

In this case, we are additionally requiring that, when the return value is an **`int`**, then it is a valid index into `str`. Again, the Whiley compiler statically verifies this is the case.

Example 4 — Loop Invariants. Another example illustrates the use of *loop invariants* in Whiley:

```
int sum([int] list)
        requires all { item in list | item >= 0 },
        ensures $ >= 0:
    r = 0
    for v in list where r >= 0:
        r = r + v
    return r
```

Here, a bounded quantifier is used to enforce that `sum()` accepts a list of natural numbers. Also, an explicit loop invariant has been given through a `where` clause. The key constraint is that summing a list of natural numbers yields a natural number (recall arithmetic is unbounded and does not overflow in Whiley). The Whiley compiler statically verifies that `sum()` does indeed meet this specification. The loop invariant is necessary to help the compiler generate a sufficiently powerful verification condition to prove the function meets the post condition (more on this later).

Example 5 — Recursive Structures. The Whiley language supports invariants over recursive structures, as the following illustrates:

```
define Tree as null | Node

define Node as { int data, Tree lhs, Tree rhs } where
            (lhs == null || lhs.data < data) &&
            (rhs == null || rhs.data > data)
```

This defines something approximating the notion of an unbalanced binary search tree. Unfortunately, the invariant permits e.g. `data < lhs.rhs.data` for a given tree node and, thus, is not sufficient to properly characterise binary search trees. Whilst our focus so far has been primarily on array programs and loop invariants, in the future we plan to place more emphasis on handling recursive structures, such as binary search trees.

4 Hoare Logic

We now briefly review Hoare logic [21] and Dijkstra's predicate transformers [22], before examining in Sect. 5 a number of challenges we encountered putting them into practice. Hoare logic provides some important background to understanding how the Whiley verifying compiler works, and why certain difficulties manifest themselves. Our discussion here is necessarily brief and we refer to Frade and Pinto for an excellent survey [39].

4.1 Overview

The rules of Hoare logic are presented as judgements involving triples of the form: $\{p\}$ **s** $\{q\}$. Here, **p** is the precondition, **s** the statement to be executed and q is the postcondition. Figure 1 presents the rules of Hoare Logic which, following Whiley, we have extended to include explicit loop invariants. To better understand these rules, consider the following example:

$$\left\{\mathtt{x} \geq \mathtt{0}\right\} \mathtt{x} = \mathtt{x} + \mathtt{1} \left\{\mathtt{x} > \mathtt{0}\right\}$$

Here we see that, if $\mathtt{x} \geq \mathtt{0}$ holds immediately before the assignment then, as expected, it follows that $\mathtt{x} > \mathtt{0}$ holds afterwards. However, whilst this is intuitively true, it is not so obvious how this triple satisfies the rules of Fig. 1. For example, as presented it does not immediately satisfy H-Assign. However, rewriting the triple is helpful here:

$$\left\{\mathtt{x} + \mathtt{1} > \mathtt{0}\right\} \mathtt{x} = \mathtt{x} + \mathtt{1} \left\{\mathtt{x} > \mathtt{0}\right\}$$

The above triple clearly satisfies H-Assign and, furthermore, we can obtain the original triple from it via H-Consequence (i.e. since $x+1 > 0 \implies x \geq 0$). The following illustrates a more complex example:

```
int f(int i) requires i >= 0, ensures $ >= 10:
    {$i \geq 0$}
    while i < 10 where i >= 0:
        {$i < 10 \wedge i \geq 0$}
        i = i + 1
        {$i \geq 0$}
    {$i \geq 10 \wedge i \geq 0$}
    return i
```

Here, we have provided the intermediate assertions which tie the Hoare triples together (note, these are not part of Whiley syntax). These assertions reflect the internal information a verifying compiler might use when establishing this function is correct.

4.2 Verification Condition Generation

Automatic program verification is normally done with a *verification condition generator* [7]. This converts the program source into a series of logical conditions — called *verification conditions* — to be checked by the automated theorem prover. There are two basic approaches: propagate *forward* from the precondition; or, propagate *backwards* from the postcondition. We now briefly examine these in more detail.

Weakest Preconditions. Perhaps the most common way to generated verification conditions is via the *weakest precondition transformer* [22]. This determines the weakest precondition (written *wp*(s,q)) that ensures a statement s meets a given postcondition q. Roughly speaking, this corresponds to propagating the postcondition backwards through the statement. For example, consider verifying this triple:

$$\frac{}{\left\{p[x/e]\right\}\ \mathtt{x = e}\ \left\{p\right\}}\ \text{(H-ASSIGN)} \qquad \frac{\left\{p\right\}\ \mathtt{s_1}\ \left\{r\right\}\ \left\{r\right\}\ \mathtt{s_2}\ \left\{q\right\}}{\left\{p\right\}\ \mathtt{s_1; s_2}\ \left\{q\right\}}\ \text{(H-SEQUENCE)}$$

$$\frac{\begin{array}{c}\left\{p_1\right\}\ \mathtt{s}\ \left\{q_1\right\}\\ \mathtt{p_2} \implies \mathtt{p_1}\ \ \mathtt{q_1} \implies \mathtt{q_2}\end{array}}{\left\{p_2\right\}\ \mathtt{s}\ \left\{q_2\right\}}\ \text{(H-CONSEQUENCE)} \qquad \frac{\begin{array}{c}\left\{p \land e_1\right\}\ \mathtt{s_1}\ \left\{q\right\}\\ \left\{p \land \neg e_1\right\}\ \mathtt{s_2}\ \left\{q\right\}\end{array}}{\left\{p\right\}\ \mathtt{if\ e_1:\ s_1\ else:\ s_2}\ \left\{q\right\}}\ \text{(H-IF)}$$

$$\frac{\left\{e_1 \land e_2\right\}\ \mathtt{s}\ \left\{e_2\right\}}{\left\{e_2\right\}\ \mathtt{while\ e_1\ where\ e_2:\ s}\ \left\{\neg e_1 \land e_2\right\}}\ \text{(H-WHILE)}$$

Fig. 1. Hoare logic.

$$\left\{\mathtt{x} \geq 0\right\}\ \mathtt{x} = \mathtt{x} + 1\ \left\{\mathtt{x} > 0\right\}$$

Propagating $\mathtt{x} > 0$ backwards through $\mathtt{x} = \mathtt{x} + 1$ gives $\mathtt{x} + 1 > 0$ via H-ASSIGN. From this, we can generate a verification condition to check that the given precondition implies this calculated weakest precondition (i.e. $\mathtt{x} \geq 0 \implies \mathtt{x} + 1 > 0$). To understand this process better, let's consider verifying a Whiley function:

```
int f(int x) requires x >= 0, ensures $ >= 0:
    x = x - 1
    return x
```

The implementation of this function does not satisfy its specification. Using weakest preconditions to determine this corresponds to the following chain of reasoning:

$$\begin{array}{l}\mathtt{x} \geq 0 \implies wp(\mathtt{x} = \mathtt{x} - 1, \mathtt{x} \geq 0)\\ \hookrightarrow \mathtt{x} \geq 0 \implies \mathtt{x} - 1 \geq 0\\ \hookrightarrow \mathtt{false}\end{array}$$

Here, the generated verification condition is $\mathtt{x} \geq 0 \implies wp(\mathtt{x} = \mathtt{x} - 1, \mathtt{x} \geq 0)$. This is then reduced to a contradiction (e.g. by the automated theorem prover) which indicates the original program did not meet its specification.

Strongest Postconditions. By exploiting Floyd's rule for assignment [20], an alternative formulation of Hoare logic can be developed which propagates in a forward direction and, thus, gives a *strongest postcondition transformer* [39,40]. This determines the strong postcondition (written $sp(\mathtt{p},\mathtt{s})$) that holds after a given statement $\mathtt{s}$ with pre-condition $\mathtt{p}$. For example, propagating $\mathtt{x} = 0$ forwards through $\mathtt{x} = \mathtt{x} + 1$ yields $\mathtt{x} = 1$. Using strongest postconditions to verify functions is similar to using weakest preconditions, except operating in the opposite direction. Thus, for a triple $\{\mathtt{p}\}\ \mathtt{s}\ \{\mathtt{q}\}$, we generate the verification condition $sp(\mathtt{p}, \mathtt{s}) \implies \mathtt{q}$. For example, consider:

$$\left\{\mathtt{x} = 0\right\}\ \mathtt{x} = \mathtt{x} + 1\ \left\{\mathtt{x} > 0\right\}$$

In this case, the generated verification condition will be $\texttt{x} = \texttt{1} \implies \texttt{x} > \texttt{0}$, which can be trivially established by an automated theorem prover.

5 Experiences

In the previous section, we outlined the process of automatic verification using Hoare logic and Dijkstra's predicate transformers. This was the starting point for developing our verifying compiler for Whiley. However, whilst Hoare logic provides an excellent foundation for reasoning about programs, there remain a number of hurdles to overcome in developing a practical tool. We now reflect on our experiences in this endeavour using examples based on those we have encountered in practice.

5.1 Loop Invariants

The general problem of automatically determining loop invariants is a hard algorithmic challenge (see e.g. [41–43]). However, we want to cover as many simple cases as possible to reduce programmer burden. We now examine a range of simple cases that, in our experience, appear to occur frequently.

Challenge 1 — Loop Invariant Variables. From the perspective of a practical verification tool, the rule H-WHILE from Fig. 1 presents something of a hurdle. This is because it relies on the programmer to completely specify the loop invariant *even in cases where this appears unnecessary.* For example, consider the following Whiley program:

```
int f(int x) requires x > 0, ensures $ >= 10:
    i = 0
    while i < 10 where i >= 0:
        i = i + x
    return i
```

Intuitively, we can see this program satisfies its specification. Unfortunately, this program cannot be shown as correct under the rules of Fig. 1 because the loop invariant is too weak. Unfortunately, rule H-WHILE only considers those facts given in the loop condition and the declared loop invariant — hence, all information about x is discarded. Thus, under H-WHILE, the verifier must assume that x could be negative within the loop body — which may seem surprising because x is not modified by the loop!

We refer to x in the example above as a *loop invariant variable.* To verify this program under rule H-WHILE, the loop invariant must be strengthened as follows:

```
int f(int x) requires x > 0, ensures $ >= 10:
    i = 0
    while i < 10 where i >= 0 && x >= 0:
        i = i + x
    return i
```

Now, one may say the programmer made a mistake here in not specifying the loop invariant well enough; however, our goal in developing a practical tool is to reduce programmer effort as much as possible. Therefore, in the Whiley verifying compiler, *loop invariant variables are identified automatically so that the programmer does not need to respecify their invariants.*

Challenge 2 — Simple Synthesis. As mentioned above, generating loop invariants in the general case is hard. However, there are situations where loop invariants can easily be determined. The following illustrates an interesting example:

```
int sum([int] xs)
      requires all { x in xs | x >= 0 }, ensures $ >= 0:
   i = 0
   r = 0
   while i < |xs| where r >= 0:
       r = r + xs[i]
       i = i + 1
   return r
```

This function computes the sum of a list of natural numbers, and returns a natural number. The question to consider is: *did the programmer specify the loop invariant properly?* Unfortunately, the answer again is: *no.* In fact, the loop invariant needs to be strengthened as follows:

```
   ...
   while i < |xs| where r >= 0 && i >= 0:
       r = r + xs[i]
       i = i + 1
   return r
```

The need for this is frustrating as, intuitively, it is trivial to see that `i >= 0` holds throughout. In the future, we aim to automatically synthesize simple loop invariants such as this.

Observation. The Whiley language also supports the notion of a *constrained type* as follows:

```
define nat as int where $ >= 0
```

Here, the **`define`** statement includes a **`where`** clause constraining the permissible values for the type ($ represents the variable whose type this will be). Thus, `nat` defines the type of non-negative integers (i.e. the natural numbers).

An interesting aspect of Whiley's design is that local variables are not explicitly declared. This gives Whiley the look-and-feel of a dynamically typed language and goes towards our goal of making the language accessible. In fact, permitting variable declarations would provide an alternative solution to the above issue with `sum()`:

```
int sum([int] xs)
      requires all { x in xs | x >= 0 }, ensures $ >= 0:
   nat i = 0
   nat r = 0
```

```
    while i < |xs|:
        r = r + xs[i]
        i = i + 1
    return r
```

Here, variable declarations are used to restrict the permitted values of variables `i` and `r` throughout the function. Unfortunately, Whiley does currently not permit local variable declarations and, hence, the above is invalid. In the future, we plan to support them for this purpose, although care is needed to integrate them with flow typing.

Challenge 3 — Loop Invariant Properties. Whilst our verifying compiler easily handles loop invariant variables, there remain situations when invariants need to be needlessly respecified. Consider the following:

```
[int] add([int] v1, [int] v2)
      requires |v1| == |v2|, ensures |$| == |v1|:
    i = 0
    while i < |v1| where i >= 0:
        v1[i] = v1[i] + v2[i]
        i = i + 1
    return v1
```

This example adds two vectors of equal size. Unfortunately, this again does not verify under the rule H-While because the loop invariant is too weak. The key problem is that `v1` is modified in the loop and, hence, our above solution for loop invariant variables does not apply. Following rule H-While, the verifying compiler can only reason about what is specified in the loop condition and invariant. Hence, it knows nothing about the size of `v1` after the loop. This means, for example, it cannot establish that `|v1| == |v2|` holds after the loop. Likewise (and more importantly in this case), it cannot establish that the size of `v1` is unchanged by the loop (which we refer to as a *loop invariant property*). Thus, it cannot establish that the size of the returned vector equals that held in `v1` on entry, and reports the function does not meet its postcondition.

In fact, it is possible to specify a loop invariant which allows the above function to be verified by our compiler. Since `v2` is a loop invariant variable and `|v1| == |v2|` held on entry, we can use `i >= 0 && |v1| == |v2|` as the loop invariant.

Observation. The example above presents an interesting challenge that, by coincidence, can be resolved by exploiting a loop invariant variable. However, it raises a more general question: *how can we specify that the size of a list is loop invariant?* Unfortunately, this is impossible in the Whiley language developed thus far because it requires some notion of a variable's value *before* and *after* the loop body is executed. To illustrate, consider the following hypothetical syntax in Whiley:

```
    ...
    while i < |v1| where i >= 0 && |v1'| == |v1|:
        v1[i] = v1[i] + v2[i]
        i = i + 1
    return v1
```

Here, `v1'` represents the value of `v1` on the previous iteration. Unfortunately, this syntax is not yet supported in Whiley and, furthermore, its semantics are unclear. For example, on entry to the loop it's unclear how `|v1'| == |v1|` should be interpreted.

Challenge 4 — Overriding Invariants. In most cases, the loop condition and invariant are used independently to increase knowledge. However, in some cases, they need to be used in concert. The following illustrates:

```
[int] create(int count, int value)
     requires count >= 0, ensures |$| == count:
   r = []
   i = 0
   while i < count:
      r = r + [value]
      i = i + 1
   return r
```

This example uses the list append operator (i.e. `r + [value]`) and is surprisingly challenging. An obvious approach is to connect the size of `r` with `i` as follows:

```
   ...
   while i < count where |r| == i:
      r = r + [value]
      i = i + 1
   return r
```

Unfortunately, this is insufficient under the rule H-While from Fig. 1. This is because, after the loop is complete, the rule establishes the invariant and the *negated* condition. Thus, after the loop, we have $\texttt{i} \geq \texttt{count} \wedge |\texttt{r}| == \texttt{i}$, but this is insufficient to establish that $|\texttt{r}| == \texttt{count}$. In fact, we can resolve this by using an *overriding loop invariant* as follows:

```
   ...
   while i < count where i <= count && |r| == i:
      r = r + [value]
      i = i + 1
   return r
```

In this case, $\texttt{i} \geq \texttt{count} \wedge \texttt{i} \leq \texttt{count} \wedge |\texttt{r}| == \texttt{i}$ holds after the loop, and the automated theorem prover will trivially establish that $|\texttt{r}| == \texttt{count}$. We say that the loop invariant *overrides* the loop condition because `i <= count` implies `i < count`.

5.2 Error Reporting

Error reporting is an important practical consideration for any verification tool, as we want error messages which are as meaningful, and precise, as possible. We now consider how the two approaches to verification condition generation affect this.

Weakest Preconditions. An unfortunate side-effect of operating in a backwards direction, as $wp(\mathtt{s}, \mathtt{q})$ does, is that reporting useful errors in the source program is more difficult. For example, consider this example which performs an integer division:

```
int f(int x) requires x > 0, ensures $ > 0:
    x = 1 / (x - 1)
    return x
```

This function contains a bug which can cause a division-by-zero failure (i.e. if `x==1` on entry). Using $wp(\mathtt{s}, \mathtt{q})$, a single verification condition is generated for this example:

$$x > 0 \implies (x - 1 \neq 0 \land \frac{1}{x-1} > 0) \tag{1}$$

A modern automated theorem prover (e.g. [44,45]) will quickly establish this condition does not hold. At this point, the verifying compiler should report a helpful error message. Unfortunately, during the weakest precondition transform, information about where exactly the error arose was lost. To identify where the error occurred, there are two intrinsic questions we need to answer: *where exactly in the program code does the error arise?* and, *which execution path(s) give rise to the error?* The $wp(\mathtt{s}, \mathtt{q})$ transform fails to answer both because it generates a single verification condition for the entire function which is either shown to hold, or not [46,47]. One strategy for resolving this issue is to embed attributes in the verification condition identifying where in the original source program particular components originated [7]. Unfortunately, this requires specific support from the automated theorem prover (which is not always available).

Strongest Postconditions. Instead of operating in a backwards direction, our experience suggests it is inherently more practical to generate verification conditions in a forwards direction (and there is anecdotal evidence to support this [39]). Recall that this corresponds to generating strongest postconditions, rather than weakest preconditions. The key advantage is that verification conditions can be emitted at the specific points where failures may occur. In the above example, there are two potential failures: (1) `1/(x-1)` should not cause division-by-zero; (2) the postcondition `$ > 0` must be met. A forward propagating verification condition generator can generate separate conditions for each potential failure. For example, it can emit the following verification conditions:

$$\mathtt{x} > 0 \implies \mathtt{x} - 1 \neq 0 \tag{2}$$

$$\mathtt{x} > 0 \implies \frac{1}{\mathtt{x} - 1} > 0 \tag{3}$$

Each of these can be associated with the specific program point where it originated and, in the case it cannot be shown, an error can be reported at that point. For example, since the first verification condition above does not hold, an error can be reported for the statement $\texttt{x} = \texttt{1}/(\texttt{x} - \texttt{1})$. When generating verification conditions based on $wp(\texttt{s}, \texttt{q})$, it is hard to report errors at the specific point they arise because, at each point, only the weakest precondition for *subsequent* statements is known.

6 Related Work

Hoare provided the foundation for formalising work in this area with his seminal paper introducing *Hoare Logic* [21]. This provides a framework for proving that a sequence of statements meets its postcondition given its precondition. Unfortunately Hoare logic does not tell us how to *construct* such a proof; rather, it gives a mechanism for *checking* a proof is correct. Therefore, to actually verify a program is correct, we need to construct proofs which satisfy the rules of Hoare logic.

The most common way to automate the process of verifying a program is with a verification condition generator. As discussed in Sect. 4.2, such algorithms propagate information in either a forwards or backwards direction. However, the rules of Hoare logic lend themselves more naturally to the latter [39]. Perhaps for this reason, many tools choose to use the weakest precondition transformer. For example, the widely acclaimed ESC/Java tool computes weakest preconditions [7], as does the Why platform [48], Spec# [49], LOOP [50], JACK [51] and SnuggleBug [52]. This is surprising given that it leads to fewer verification conditions and, hence, makes it harder to generate useful error messages (recall our discusion from Sect. 4.2). To workaround this, Burdy *et al.* embed path information in verification conditions to improve error reporting [51]. A similar approach is taken in ESC/Java, but requires support from the underlying automated theorem prover [45]. Denney and Fischer extend Hoare logic to formalise the embedding of information within verification conditions [53]. Again, their objective is to provide useful error messages.

The Dafny language has been developed with similar goals in mind to Whiley [14]. In particular, Dafny was designed to simplify verification and, to this end, makes similar choices to Whiley. For example, all arithmetic is unbounded and a strong division is made between functional and imperative constructs. Here, pure functions are supported for use in specifications and directly as code, whilst methods may have side-effects and can describe pointer-based algorithms. These two aspects are comparable (respectively) to Whiley's functional core and imperativer outer layer. Finally, Dafny supports explicit pre- and post-conditions for functions and methods which are discharged using Z3 [44].

7 Conclusion

In this paper, we reflected on our experiences using the Whiley verifying compiler. In particular, we identified a number of practical considerations for any verifying compiler which are not immediately obvious from the underlying theoretical foundations.

Acknowledgements. This work is supported by the Marsden Fund, administered by the Royal Society of New Zealand.

References

1. Hoare, C.A.R.: The verifying compiler: a grand challenge for computing research. JACM **50**(1), 63–69 (2003)
2. King, S.: A program verifier. Ph.D. thesis, Carnegie-Mellon University (1969)
3. Peter Deutsch, L.: An interactive program verifier. Ph.D. thesis, University of California (1973)
4. Good, D.I.: Mechanical proofs about computer programs. In: Hoare, C.A.R., Shepherdson, J.C. (eds.) Mathematical Logic and Programming Languages, pp. 55–75. Prentice Hall, Englewood Cliffs (1985)
5. Luckham, D.C., German, S.M., von Henke, F.W., Karp, R.A., Milne, P.W., Oppen, D.C., Polak, W., Scherlis, W.L.: Stanford pascal verifier user manual. Technical report CS-TR-79-731, Department of Computer Science, Stanford University (1979)
6. Detlefs, D.L., Leino, K.R.M., Nelson, G., Saxe, J.B.: Extended static checking. SRC Research report 159, Compaq Systems Research Center (1998)
7. Flanagan, C., Leino, K.R.M., Lillibridge, M., Nelson, G., Saxe, J.B., Stata, R.: Extended static checking for Java. In: Proceedings of PLDI, pp. 234–245 (2002)
8. Leavens, G.T., Cheon, Y., Clifton, C., Ruby, C., Cok, D.R.: How the design of JML accommodates both runtime assertion checking and formal verification. Sci. Comput. Program. **55**(1–3), 185–208 (2005)
9. Cok, D.R., Kiniry, J.R.: ESC/Java2: uniting ESC/Java and JML. In: Barthe, G., Burdy, L., Huisman, M., Lanet, J.-L., Muntean, T. (eds.) CASSIS 2004. LNCS, vol. 3362, pp. 108–128. Springer, Heidelberg (2005)
10. Barnett, M., Rustan, K., Leino, M., Schulte, W.: The Spec# programming system: an overview. Technical report, Microsoft Research (2004)
11. Barnett, M., DeLine, R., Fähndrich, M., Leino, K.R.M., Schulte, W.: Verification of object-oriented programs with invariants. J. Object Technol. **3**(6), 27–56 (2004)
12. Barnett, M., Evan Chang, B.-Y., DeLine, R., Jacobs, B., Leino, K.R.M.: Boogie: a modular reusable verifier for object-oriented programs. In: de Boer, F.S., Bonsangue, M.M., Graf, S., de Roever, W.-P. (eds.) FMCO 2006. LNCS, vol. 4111, pp. 364–387. Springer, Heidelberg (2006)
13. Leino, K.R.M.: Dafny: an automatic program verifier for functional correctness. In: Clarke, E.M., Voronkov, A. (eds.) LPAR-16 2010. LNCS, vol. 6355, pp. 348–370. Springer, Heidelberg (2010)
14. Rustan, K., Leino, M.: Developing verified programs with Dafny. In: Joshi, R., Müller, P., Podelski, A. (eds.) VSTTE 2012. LNCS, vol. 7152, p. 82. Springer, Heidelberg (2012)
15. The whiley programming language. http://whiley.org
16. Pearce, D.J., Groves, L.: Whiley: a platform for research in software verification. In: Erwig, M., Paige, R.F., Van Wyk, E. (eds.) SLE 2013. LNCS, vol. 8225, pp. 238–248. Springer, Heidelberg (2013)
17. Pearce, D., Noble, J.: Implementing a language with flow-sensitive and structural typing on the JVM. Electron. Notes Theoret. Comput. Sci. **279**(1), 47–59 (2011)
18. Pearce, D.J.: Sound and complete flow typing with unions, intersections and negations. In: Giacobazzi, R., Berdine, J., Mastroeni, I. (eds.) VMCAI 2013. LNCS, vol. 7737, pp. 335–354. Springer, Heidelberg (2013)

19. Bryant, R.E., Kroening, D., Ouaknine, J., Seshia, S.A., Strichman, O., Brady, B.A.: Deciding bit-vector arithmetic with abstraction. In: Grumberg, O., Huth, M. (eds.) TACAS 2007. LNCS, vol. 4424, pp. 358–372. Springer, Heidelberg (2007)
20. Floyd, R.W.: Assigning meaning to programs. In: Proceedings AMS, vol. 19, pp. 19–31. American Mathematical Society (1967)
21. Hoare, C.A.R.: An axiomatic basis for computer programming. CACM **12**, 576–580 (1969)
22. Dijkstra, E.W.: Guarded commands, nondeterminancy and formal derivation of programs. CACM **18**, 453–457 (1975)
23. Rountev, A.: Precise identification of side-effect-free methods in Java. In: Proceedings of ICSM, pp. 82–91. IEEE Computer Society (2004)
24. Sălcianu, A., Rinard, M.: Purity and side effect analysis for Java programs. In: Cousot, R. (ed.) VMCAI 2005. LNCS, vol. 3385, pp. 199–215. Springer, Heidelberg (2005)
25. Mycroft, A.: Programming language design and analysis motivated by hardware evolution. In: Riis Nielson, H., Filé, G. (eds.) SAS 2007. LNCS, vol. 4634, pp. 18–33. Springer, Heidelberg (2007)
26. Tobin-Hochstadt, S., Felleisen, M.: Logical types for untyped languages. In: Proceedings of ICFP, pp. 117–128 (2010)
27. Guha, A., Saftoiu, C., Krishnamurthi, S.: Typing local control and state using flow analysis. In: Barthe, G. (ed.) ESOP 2011. LNCS, vol. 6602, pp. 256–275. Springer, Heidelberg (2011)
28. Barbanera, F., Dezani-Cian Caglini, M.: Intersection and union types. In: Proceedings of the TACS, pp. 651–674 (1991)
29. Igarashi, A., Nagira, H.: Union types for object-oriented programming. J. Object Technol. **6**(2), 31–52 (2007)
30. Pierce, B.C.: Types and Programming Languages. MIT Press, Cambridge (2002)
31. Hoare, T.: Null references: The billion dollar mistake, presentation at QCon (2009)
32. Fähndrich, M., Leino, K.R.M.: Declaring and checking non-null types in an object-oriented language. In: Proceedings of the OOPSLA, pp. 302–312. ACM Press (2003)
33. Ekman, T., Hedin, G.: Pluggable checking and inferencing of non-null types for Java. J. Object Technol. **6**(9), 455–475 (2007)
34. Chalin, P., James, P.R.: Non-null references by default in Java: alleviating the nullity annotation burden. In: Ernst, E. (ed.) ECOOP 2007. LNCS, vol. 4609, pp. 227–247. Springer, Heidelberg (2007)
35. Male, C., Pearce, D.J., Potanin, A., Dymnikov, C.: Java bytecode verification for @NonNull types. In: Hendren, L. (ed.) CC 2008. LNCS, vol. 4959, pp. 229–244. Springer, Heidelberg (2008)
36. Hubert, L.: A non-null annotation inference for Java bytecode. In: Proceedings of the PASTE, pp. 36–42. ACM (2008)
37. ISO/IEC. international standard ISO/IEC 9899, programming languages – C (1990)
38. Lameed, N., Hendren, L.: Staged static techniques to efficiently implement array copy semantics in a MATLAB JIT compiler. In: Knoop, J. (ed.) CC 2011. LNCS, vol. 6601, pp. 22–41. Springer, Heidelberg (2011)
39. Frade, M.J., Pinto, J.S.: Verification conditions for source-level imperative programs. Comput. Sci. Rev. **5**(3), 252–277 (2011)
40. Gordon, M., Collavizza, H.: Forward with Hoare. In: Roscoe, A.W., Jones, C.B., Wood, K.R. (eds.) Reflections on the Work of C.A.R. Hoare, History of Computing, pp. 101–121. Springer, London (2010)

41. Chadha, R., Plaisted, D.A.: On the mechanical derivation of loop invariants. J. Symbolic Comput. **15**(5 & 6), 705–744 (1993)
42. Leino, K.R.M., Logozzo, F.: Loop invariants on demand. In: Yi, K. (ed.) APLAS 2005. LNCS, vol. 3780, pp. 119–134. Springer, Heidelberg (2005)
43. Furia, C.A., Meyer, B.: Inferring loop invariants using postconditions. CoRR, abs/0909.0884 (2009)
44. de Moura, L., Bjørner, N.: Z3: An efficient SMT solver. In: Proceedings of the TACAS, pp. 337–340, (2008)
45. Detlefs, D.L., Nelson, G., Saxe, J.B.: Simplify: a theorem prover for program checking. JACM **52**, 365–473 (2005)
46. Leino, K.R.M., Millstein, T.D., Saxe, J.B.: Generating error traces from verification-condition counterexamples. Sci. Comput. Program. **55**(1–3), 209–226 (2005)
47. Jager, I., Brumley, D.: Efficient directionless weakest preconditions. Technical Report CMU-CyLab-10-002, Carnegie Mellon University (2010)
48. Filliâtre, J.-C., Marché, C.: The Why/Krakatoa/Caduceus platform for deductive program verification. In: Damm, W., Hermanns, H. (eds.) CAV 2007. LNCS, vol. 4590, pp. 173–177. Springer, Heidelberg (2007)
49. Barnett, M., Rustan M. Leino, K.: Weakest-precondition of unstructured programs. In: Proceedings of the PASTE, pp. 82–87. ACM Press (2005)
50. Jacobs, B.: Weakest pre-condition reasoning for Java programs with JML annotations. JLAP **58**(1–2), 61–88 (2004)
51. Burdy, L., Requet, A., Lanet, J.-L.: Java applet correctness: a developer-oriented approach. In: Araki, K., Gnesi, S., Mandrioli, D. (eds.) FME 2003. LNCS, vol. 2805, pp. 422–439. Springer, Heidelberg (2003)
52. Chandra, S., Fink, S.J., Sridharan, M.: Snugglebug: a powerful approach to weakest preconditions. In: Proceedings of the PLDI, pp. 363–374. ACM Press (2009)
53. Denney, E., Fischer, B.: Explaining verification conditions. In: Meseguer, J., Roşu, G. (eds.) AMAST 2008. LNCS, vol. 5140, pp. 145–159. Springer, Heidelberg (2008)