
For details regarding the final published version please click on the following DOI link: http://dx.doi.org/10.1109/HASE.2007.56
When citing this source, please use the final published version as above.
Multiple Pre/Post Specifications for Heap-Manipulating Methods

Wei-Ngan Chin¹,²  Cristina David¹  Huu Hai Nguyen²  Shengchao Qin³
¹ Department of Computer Science, National University of Singapore
² Computer Science Programme, Singapore-MIT Alliance
³ Department of Computer Science, Durham University
{chinwn,davidcri,nguyenh2}@comp.nus.edu.sg  shengchao.qin@durham.ac.uk

Abstract

Automated verification plays an important role for high assurance software. This typically uses a pair of pre/post conditions as a formal (but possibly partial) specification of each method before it is systematically verified. In this paper, we advocate for multiple pairs of pre/post conditions to be associated with each method which provides a way for such specification to be used in more scenarios. Multiple pre/post specifications are important for heap-manipulating programs where they can be precisely expressed using separation logic. This work highlights the importance of multiple pre/post specifications, and a methodology to capture them via a set of states during proof search.

1 Introduction

In recent years, separation logic formalism has been successfully applied to analysing and verifying heap-manipulating programs. This formalism supports succinct description of shapely data structures, such as near-balanced AVL-trees [10], and has been used to analyse and verify various program properties. One important feature of separation logic is its support for accurate references into memory states which can be captured with the help of inductive shape predicate and separating conjunction. For example, a list segment of length \( n \) can be specified by the following predicate definition:

\[
\text{lseg}(\text{root}, p, n) \equiv \text{root} \rightarrow p \land n = 0 \lor
\exists q \cdot \text{root} \rightarrow \text{node}(\cdot, q) \land \text{lseg}(q, p, n-1) \text{ inv } n \geq 0
\]

where \text{node} is an object type declared as

\[
\text{data} \text{ node} \{ \text{int val}; \text{node next} \}
\]

In separation logic notation, \( x \rightarrow \text{node}(a, b) \) captures a distinct memory cell referenced from \( x \), while separating conjunction \( \Delta_1 \ast \Delta_2 \) captures two disjoint heaps described by \( \Delta_1 \) and \( \Delta_2 \), respectively. In contrast, \( \Delta_3 \land \Delta_4 \) captures two overlapping heaps, \( \Delta_3 \) and \( \Delta_4 \). The formula \( n \geq 0 \) after the \text{inv} keyword captures a heap-independent invariant for the predicate. This invariant is an approximation of the predicate, in the sense that whenever the predicate holds, the invariant also holds. By default, each shape predicate is expected to have a special (first) parameter, named \text{root}, that can transitively reach all its memory cells. This “root” parameter will be instantiated with an actual argument for each instance of its predicate that appears in formulas. Furthermore, we shall use a uniform notation where object \( x \rightarrow \text{node}(a, \ldots) \) is written as \( x::\text{node}(a, \ldots) \), and predicate \( \text{pred}(x, a, \ldots) \) is written as \( x::\text{pred}(a, \ldots) \). We refer to them collectively as heap nodes. With this new notation, we can omit the \text{root} parameter in the head of the predicate, and define a non-empty circular list, as follows:

\[
\text{clist}(n) \equiv \exists p \cdot \text{root}::\text{node}(\ldots, p) \ast p::\text{lseg}(\text{root}, n-1) \text{ inv } n \geq 1
\]

Shape predicates can also be used to describe more complex data structures with stronger properties. An example is the non-empty sorted list:

\[
\text{sortl}(n, s, b) \equiv \text{root}::\text{node}(\text{null}, s) \land s = b \land n = 1 \lor
\exists p, t \cdot \text{root}::\text{node}(s, p) \ast p::\text{sortl}(n-1, t, b) \land s \leq t
\text{ inv } n > 0 \land s \leq b
\]

Predicate \( \text{sortl}(n, s, b) \) ensures that all values in the list are sorted in ascending order, with \( s \) and \( b \) capturing its min and max values, respectively. The sortedness property is ensured by the presence of \( s \leq t \) in the above predicate. The parameters \( n, s, b \) actually capture some derived properties of the heap data structure, as stated in the predicate definition. These parameters play a role similar to “model fields” in some specification languages, such as Spec# [9, 1] and JML [3]. As a shorthand, we may omit the existential quantifiers without ambiguity.

Though separation logic with inductive predicates can be highly expressive, current automated theorem provers hardly provide any support for this form of substructural logic. Several recent works, such as [2, 4], have attempted to address this shortcoming by building specialised solvers that work for a fixed set of predicates (e.g. \text{lseg} without the size parameter). Our recent work [10] has lifted one crucial limitation by supporting the automated reasoning of...
user-defined predicates. To this end, we have designed a general prover that uses unfold/fold reasoning on predicates to support entailment between two heap states, and also for computing its residual heap. This is essentially a prover with “frame inferring” capability. Given two heap states $\Delta_a$ and $\Delta_c$, our prover can check for an entailment of the form: $\Delta_a \vdash \Delta_c * \Delta_r$, where $\Delta_r$ is a residual heap from $\Delta_a$ after fully accounting for the memory heap state of $\Delta_c$. The residue is a consequence of the frame rule from separation logic that can be determined by an incremental matching algorithm in our prover. Our entailment also handles disjunctive formulae and existential quantifiers. While our prover is sound, terminating and automated, it is incomplete.

In this paper, we propose one approach to partially overcome this incompleteness shortcoming, by providing a mechanism for specifying and handling multiple pre/post conditions in separation logic. Traditionally, each method is given a single pair of pre/post conditions which describes the expected pre-state prior to a method call and its subsequent post-state. Furthermore, even when multiple pre/post conditions are allowed in some specification languages, such as JML [3], the standard technique for handling them is to re-combine into a single pre/post condition. As an example, consider a method with two (pairs of) pre/post conditions, $(\text{pre}_1, \text{post}_1)$ and $(\text{pre}_2, \text{post}_2)$, as shown:

- requires $\text{pre}_1$ ensures $\text{post}_1$
- also requires $\text{pre}_2$ ensures $\text{post}_2$

The proposed solution in JML [6] is to transform the two pre/post conditions into an integrated pre/post condition:

- requires $\text{pre}_1 \lor \text{pre}_2$
- ensures $\text{old}(\text{pre}_1) \implies \text{post}_1 \land \text{old}(\text{pre}_2) \implies \text{post}_2$

While this re-combination may work with traditional Hoare logic that is based on pure formulae, it cannot be used for an arbitrary heap state $\Delta$ of separation logic, since $\text{old}(\Delta)$ is not always determinable. As an example of multiple pre/post conditions in separation logic formula, consider the append method for joining two list together:

```
void append(node x, node y)
{if x.next $\neq$ null then { append(x.next, y) }
 else { x.next:=y } }
```

A simple specification for this method is for joining two disjoint lists into a single longer list, as illustrated by the first pre/post condition below:

- requires $x::\text{lseg}(\text{null}, n_1) + y::\text{lseg}(\text{null}, n_2) \land n_1 > 0$
- ensures $x::\text{lseg}(\text{null}, n_1 + n_2)$

The $n_1 > 0$ constraint ensures that the first input is non-empty. As a result, the pointer access operation $x.next \neq \text{null}$ can be proven to be safe. Another radically different view of this method is for the purpose of joining two sorted lists, whereby the largest value of the first sorted list is less than the smallest value of the second sorted list. This view is captured by a second pre/post condition below:

- requires $x::\text{sortl}(n_1, a, b) + y::\text{sortl}(n_2, c, d) \land b \leq c$
- ensures $x::\text{sortl}(n_1 + n_2, a, d)$

With a more specialised pre-state, our specification is able to conclude that the resulting list is a longer sorted list. Furthermore, the $\text{sortl}$ predicate is non-empty which can prove that dereferencing by $x.next \neq \text{null}$ is safe. Such a specification captures a different view for the same append method, which would be needed towards the verification of the sortedness property for the quicksort algorithm. Our thesis is that it is often futile to combine these widely different pre/post conditions into a single pre/post condition, as the translations for $\text{old}(x::\text{lseg}(\text{null}, n_1) + y::\text{lseg}(\text{null}, n_2) \land n_1 > 0)$ and $\text{old}(x::\text{sortl}(n_1, a, b) + y::\text{sortl}(n_2, c, d) \land b \leq c)$ into heap (or pure) formulae cannot be systematically determined. For example, in JML, $\text{old}(e)$ is guaranteed to be safe to use if $e$ denotes a primitive value, or an immutable object [7] (sec 11.4.2). Otherwise, some ambiguity in specification is possible.

In this paper, we propose a new mechanism to support the direct handling of multiple pre/post conditions. Our mechanism is based on a set of states $\{\Delta_1, \ldots, \Delta_n\}$ which represent abstract states that may arise during proof search used by automated verification. As a first step, we propose to generalise the entailment procedure into a non-deterministic version that explicitly returns a set of residual states, namely: $\Delta_a \vdash \Delta_c \ast \{\Delta_1, \ldots, \Delta_n\}$, such that for all $1 \leq i \leq n$, $\Delta_a \vdash \Delta_c \ast \Delta_i$ holds. In the event of an entailment failure, the set of residual states is empty. Note that “states” here are formulas, representing the abstract states of the program.

The set of residual states is meant to contain syntactically different states. This does not imply that all the states in the set are semantically different. Our approach may indeed cause some redundancy in the search for proofs, but it does not affect soundness of the entailment prover. Explicit set of states allows easier integration between the prover and other components of our verification system. In particular, its use is critical for handling multiple pre/post specifications which can now be viewed as part of proof search.

This paper makes the following contributions:

- **Multiple pre/post**: This mechanism allows users to specify more properties of a procedure. Our system automatically verifies that each user-provided specification is correct, prior to using them for systematic proof search.
- **Set of states**: Provers for separation logic typically need to derive residues (or frames), which have to be communicated to program verifiers. Our direct support for non-deterministic proof search via set of states notation makes this task both explicit and exhaustive.
- **Implementation**: We have implemented a verification
2 Separation Logic Formalism Used

The syntax for separation logic formulae that we will use is given in Figure 1. Each shape predicate \( \text{spred} \) has a body \( \Phi \), which can be recursive as it can mention the predicate being defined. Each predicate can also be equipped with a heap-independent invariant \( \pi_0 \) that is user-supplied but machine-checked for its validity. The invariant is marked with a prior keyword \( \text{inv} \).

The separation constraints we use are in a disjunctive normal form \( \Phi \). Each disjunct consists of a \( \ast \)-separated heap constraint \( \kappa \), referred to as \( \text{heap part} \), and a heap-independent formula \( \pi \), referred to as \( \text{pure part} \). The pure part does not contain any heap nodes and is restricted to pointer equality/disequality (to facilitate precise aliasing and non-aliasing), Presburger arithmetic and set/bag constraints that can capture a collection of reachable values
addresses. Furthermore, \( \Delta \) denotes a composite formula that could always be normalised into the \( \Phi \) form (see Fig. 5 in Sec 4.1).

Separation constraints are used in pre/post conditions and shape definitions. We will also use another special variable \( \text{res} \) to denote the returned value of an expression (including the method body). In order to handle them correctly without running into unmatched residual heap nodes, we require all constraints used in specifications be well-formed.

**Definition 2.1 (Accessible)** A variable in a specification is said to be accessible if it is a method parameter or a special variable, i.e. root or res.

**Definition 2.2 (Reachable)** Given a heap constraint \( \kappa \) such that \( \kappa = p::c(v^*) \ast \kappa_1 \), node \( p::c(v^*) \) is reachable from a variable \( q \) iff they satisfy the following relation:

\[
\text{reach}(\kappa_1, q, p::c(v^*)) = \text{df} (p = q) \\
\forall (k_1 = q::c_q(. . ., r, ..) \ast \kappa_2 \land \text{reach}(\kappa_2, r, p::c(v^*)))
\]

**Definition 2.3 (Well-Formed Constraint)** A separation constraint \( \Phi \) is well-formed if (i) all objects and predicate instances in the constraint are reachable from accessible variables, (ii) \( \Phi \) is in a disjunctive normal form \( \bigvee \exists v^* \ast (\kappa \land \pi)^* \) where \( \kappa \) denotes heap constraints, \( \pi \) pointer, arithmetic and set/bag constraints.

The primary significance of the well-formed condition is that all heap nodes of a heap constraint are reachable from accessible variables. This allows the entailment proving procedure to correctly match nodes from the consequent with nodes from the antecedent of an entailment relation. Furthermore, arbitrary recursive shape relation can lead to non-termination in unfold/fold reasoning. We avoid this pitfall by using only well-founded shape predicates:

**Definition 2.4 (Well-Founded Predicate)** A shape predicate is said to be well-founded if its body satisfies three conditions, namely: (i) it is a well-formed constraint, (ii) the parameter root may not be bound to a predicate instance, (iii) there is at most one object in each disjunct.

3 Multiple Pre/Post Specifications

Our approach currently expects pre/post conditions to be specified for each method. With the rich variety of shapes that can be specified, there are often multiple ways of viewing a method’s intended behaviour. In this section, we explore the contexts in which multiple pre/post specifications are useful for heap-manipulating methods. For simplicity, we shall illustrate via our running append method, that is reproduced below.

```java
void append(node x, node y)
{ if x.next/=null then { append(x.next, y) }
  else { x.next:=y } }
```

Firstly, multiple pre/post specifications are helpful for capturing strong aliasing properties in separation logic. Our previous specification was for two disjoint input lists. We can also provide a pre/post specification in which \( x \) and \( y \) parameters are aliased, as follows:

```java
requires x::lseg(null,n)\land n>0 \land x=y 
ensures x::clist(n)
```

The aliasing by \( x=y \) causes the append function to return a circular list instead! To successfully verify the recursive call of append, we also require another pre/post specification that returns a list segment, as follows:

```java
requires x::lseg(null,n)\land n>0 
ensures x::lseg(n,y)
```
4 Our Approach

In our approach to verification, we expect users to supply two things that can assist in automatic verification of their code, namely shape predicate definitions and multiple pre/post specifications for each method. Our system has two major components, namely (i) Hoare-style verifier and (ii) entailment prover for separation logic, as shown in Figure 2.

![Figure 2. Our Approach to Verification](image)

4.1 Forward Verifier

Our Hoare-style verifier is defined for an imperative object-based language with syntax in Figure 3. A program $P$ consists of declarations $\text{tdecl}$ and methods $\text{meth}$. Declarations can be shape predicates $\text{spred}$ or object types $\text{objt}$. Each method is decorated with the specification $\{\Phi_i^{pr} \leftrightarrow \Phi_i^{po}\}_{i=1}^p$ that is made up of a collection of pre- and post-condition pairs. This more precise syntax for multiple pre/post pairs shall be used in our formalization. The intended meaning is that whenever the method is called in a program state satisfying precondition $\Phi_i^{pr}$ and if the method terminates, the resulting state will satisfy the corresponding postcondition $\Phi_i^{po}$. We handle while loop in a similar way. Other constructs are standard.

Hoare-style code verifier is based on forward rules of the form $\{ \Delta_1 \}_{\text{code}} \{ \Delta_2 \}$, where $\Delta_1$ is a prestate in separation logic and $\Delta_2$ is a poststate that shall be computed by forward reasoning. To capture proof search, we generalize the forward rule to the form $\{ \Delta \}_{\text{code}} \{ S \}$ where $S$ is a set of heap states, discovered by a search-based

![Figure 3. A Core Imperative Language](image)

verification process. When $S$ is empty, the forward verification is said to have failed for $A$ as prestate. For convenience, we also provide lifted variant of the forward verifier to take a set of prestates. Verification in such a case succeeds if any of the prestates gives rise to a successful verification, that is if at least one of the $S_i$ is non-empty. This rule is useful when the forward verifier has processed at least one subexpression, potentially giving rise to a set of residual states.

$$\forall i \in 1..n \cdot \{ \Delta_i \} \text{ code } \{ S_i \}$$

$$\vdash \{ \{ \Delta_1, ..., \Delta_n \} \} \text{ code } \bigcup_{i=1}^n S_i$$

Verification of a method starts with each precondition, and proves that the corresponding postcondition is guaranteed at the end of the method. The verification is formalized in the following rule:

$$\text{[FV–METH]}$$

$$V = \{ v_1..v_n \} \quad W = \text{prime}(V)$$

$$\forall i \in 1, ..., p \cdot ( \vdash \{ \Phi_{po} \land \text{nochange}(V) \} \in \{ S_i \}$$

$$(3W \cdot S_i) + \Phi_{po} + S_2 \oplus \{ \} \quad \vdash t_0 mn(t_1 v_1, ..., t_n v_n) \text{ where } \{ \Phi_{po} \land \text{nochange} \}_{i=1}^p (e)$$

The function $\text{prime}(V)$ returns $\{ v' \mid v \in V \}$. The predicate $\text{nochange}(V)$ returns $\bigwedge_{v \in V} (v = v')$. If $V = \{ \}$, $\text{nochange}(V) = \text{true}$. $3W \cdot S$ returns $\{ 3W \cdot S \mid S \in S \}$. The entailment $\{ \Phi_{po} + S_2 \} \vdash \{ \} \oplus \{ \}$ is discharged by the entailment described in the next subsection.

At a method call, each of the method’s preconditions is checked. The combination of the residue $S_i$ and the postcondition is added to the poststate. If a precondition is not entailed by the program state $\Delta$, the corresponding residue is not added to the set of states. The test $S \neq \{ \} \implies \{ \}$ ensures that at least one precondition is satisfied.

$$\text{[FV–CALL]}$$

$$t_0 m(t_1 v_1, ..., t_n v_n) \text{ where } \{ \Phi_{po} \land \text{nochange} \}_{i=1}^p (e) \in P$$

$$\rho = [v'_i/v_i]_{i=1}^p \quad \Delta = r_{\rho \Phi_{po} + S_i} \quad \forall i \in 1, ..., p$$

$$S = \bigcup_{i=1}^n S_i \ast \Phi_{po} \quad S \neq \{ \}$$

$$\vdash \{ \Delta \} m(t_1 v_1, ..., t_n v_n) \{ S \}$$

Note that the verification rule also invokes the entailment prover to discharge $\Delta \ast r_{\rho \Phi_{po} + S_i}$, where $\rho$ represents a substitution of $v_j$ by $v'_j$, for all $j = 1, ..., n$. The lifted separation conjunction $\ast$ over a set (i.e., $S_i \ast \Phi_{po}^i$) is defined in Fig. 5 in the next subsection.

Our verifier also ensures that each field access is safe from null dereferencing. This is shown in the field access rules in Fig. 4 which also includes other forward verification rules for the language. The verification rules attempt to track heap states, as accurately as possible, with path-sensitivity captured by $\text{[FV–IF]}$ rule, flow-sensitivity by $\text{[FV–SEQ]}$ rule and context sensitivity by the $\text{[FV–CALL]}$ rule. In a nutshell, verification is carried out at three places. For each call site, the $\text{[FV–CALL]}$ rule (mentioned earlier) ensures that at least one of its method’s preconditions is satisfied. At each method definition, the $\text{[FV–METH]}$ rule checks that every postcondition holds for the method body assuming its respective precondition. At each shape definition, $\text{[FV–SPRED]}$ checks that its given invariant $\pi_{inv}$ is sound w.r.t. (i.e., semantic consequence of) the well-formed heap formula $\Phi$. (The rule for while loop is omitted but is essentially similar to the mechanics for handling tail-recursive methods.) The function $\text{XPure}_0(\Phi)$ generates a sound and heap-independent approximation of the heap constraint $\Phi$. For instance,

$$\text{XPure}_0(x: \text{node}(\_ \_)) \equiv x > 0$$

$$\text{XPure}_0(x: \text{node}(\_ \_) \ast y: \text{node}(\_ \_)) \equiv x > 0 \land y > 0 \land x \neq y$$

$$\text{XPure}_0(x: \text{lsseg}(p, n)) \equiv n \geq 0$$

For the shape predicate case above, we can get a more precise approximation by unrolling the predicate definition once, for example:

$$\text{XPure}_1(x: \text{lsseg}(p, n)) \equiv (x = p \land n = 0 \lor x > 0 \land n > 0)$$

The definition for the general approximation procedure $\text{XPure}_n(\Phi)$ (also used in the entailment prover) can be found in [10] where $n$ denotes the number of unrollings done on the shape predicates.

The operators $\land_{\{v\}}$ (in assignment rule) and $*_{\text{w}}$ (in while rule) are composition with update operators. Given a state $\Delta_1$, a state change $\Delta_2$, and a set of variables to be updated $X = \{ x_1, ..., x_m \}$, the composition operator $\oplus_X$ is defined as:

$$\Delta_1 \oplus_X \Delta_2 = \forall y_1, ..., y_n \cdot \rho_1 \Delta_1 \oplus \rho_2 \Delta_2$$

where $r_1, ..., r_n$ are fresh variables;

$$\rho_1 = [r_1/x_1]_{i=1}^m : \rho_2 = [r_1/x_1]_{i=1}^n$$

Note that $\rho_1$ and $\rho_2$ are substitutions that link each latest value of $x'_i$ in $\Delta_1$ with the corresponding initial value $x_i$ in $\Delta_2$ via a fresh variable $r_i$. The binary operator $\oplus$ is either $\land$ or $\lor$.

Normal rules for separation constraints and lifted operators over sets of states are given in Fig. 5.

4.2 Entailment Prover

The other major component of our system is the entailment prover. This prover directly uses the definitions of shape predicates to reason about entailment between two heap formulas. A key novelty is the use of a set of heap states to support non-deterministic entailment. By non-determinism, we mean a search process that returns multiple answers, any one of which indicates a successful verification. Our entailment prover has the form $\Delta_a \vdash \Delta_c \ast S$ where $S$ is a set of possible residual poststates. The entailment succeeds when $S$ is non-empty, otherwise it is deemed to have failed. When $S$ captures multiple residual states, they signify different search outcomes during proving.

During entailment, each pair of aliased objects/predicates from $\Delta_a$ and $\Delta_c$ are matched up, whenever
...from the antecedent that have been used to match and free variables from the matched node in the consequent precise to cover all nodes from the consequent.

If no immediate match can be identified, an unfold/fold
... of a Set

Figure 5. Normalization Rules for Separation Constraints and with operators lifted to a Set

they are proven identical. The formal rule for matching is:

\[
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_1=p_2, \quad \rho=[v_1^*/v^*] \\
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_2=p_3, \quad \rho=[v_2^*/v^*] \\
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_3=p_4, \quad \rho=[v_3^*/v^*] \\
\]

Note that the complete form for the entailment relation is \(A_1 \land \pi \implies \phi \equiv \Box \) which denotes \(k \equiv \pi \implies \Box \). As mentioned earlier, the purpose of the entailment prover is to check that heap nodes in the antecedent \(A_1 \land \pi \) are sufficient to cover all nodes from the consequent \(\phi \), and to compute a set of possible residual heap states \(S\) (which is empty if the entailment search fails). \(k\) is the history of nodes from the antecedent that have been used to match nodes from the consequent, \(V\) is the list of existentially quantified variables from the consequent. Note that \(k\) and \(V\) are derived. The entailment prover is invoked with \(k=\emptyset\) and \(V=\emptyset\). When a match occurs, the bindings between free variables from the matched node in the consequent and the corresponding variables from the antecedent are generated and kept in the antecedent via \(\text{freeEqn}(\rho, V)\) which is defined as follows:

\[
\text{freeEqn}(\{u_i/v_i\}_{i=1}^m, V) = v_i = 0
\]

If no immediate match can be identified, an unfold/fold operation may be invoked. The rule for unfolding below is to unfold a predicate in the antecedent \(A_1 \land \pi\) so as to match up with an object in the consequent \(A_2\):

\[
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_1=p_2 \\
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_2=p_3 \\
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_3=p_4
\]

The test \(\text{IsPred}(c)\) (resp. \(\text{IsObj}(c)\)) returns \text{true} if \(c\) is defined as a shape predicate (resp. an object).

Alternatively, a predicate in the consequent that is aliased with an object in the antecedent is handled by folding. The folding rule given next is a recursive invocation of the entailment procedure for a predicate from the consequent, in order to identify a set of heap nodes that match with that predicate’s definition. It is different from unfolding a predicate in the antecedent as we allow bindings on free variables to transfer into the antecedent at the end of folding. This is critical for capturing free variables from preconditions that are used in postconditions.

\[
(\Delta_1 \land \pi \land \Delta_2) \land \pi \implies (\Delta_1 \land \pi \land \Delta_2) \\
(\Delta_1 \land \pi \land \Delta_2) \land \pi \implies (\Delta_1 \land \pi \land \Delta_2) \\
(\Delta_1 \land \pi \land \Delta_2) \land \pi \implies (\Delta_1 \land \pi \land \Delta_2) \\
(\Delta_1 \land \pi \land \Delta_2) \land \pi \implies (\Delta_1 \land \pi \land \Delta_2)
\]

\[
\Delta_1 \land \pi \land \Delta_2 \land \pi \implies (\exists x \cdot \Delta_1 \land \pi \land \Delta_2) \\
\Delta_1 \land \pi \land \Delta_2 \land \pi \implies (\exists y \cdot \{x/y\} \Delta_1 \land \pi \land \Delta_2)
\]

\[
\text{IsPred}(c_2) \land \text{IsObj}(c_1) \\
\text{XPure}_n(p_1::c(v_1^*)\land k_1\land \pi_1) \implies p_1=p_2 \\
(\pi_1^*,\pi_2^*)=\text{split}(\pi_1^*) \\
\Delta_1 \land \pi_1^* \land \Delta_2 \land \pi_2^* \implies S_i
\]

When a fold to a predicate \(p_2::c_2(v_2^*)\) is performed, the constraints related to variables \(v_2^*\) are important. The \text{split} function projects these constraints out and differentiates them based on free variables. For instance, let us consider that the parameters of the folded predicate are \(A\) and \(B\), and the bindings introduced by the folding process are \(n=0\land B=\{1\}\). If \(B\) is a free variable and \(n\) is bound, then the \text{split} function differentiates the given constraints as follows:
The folding operation requires a special version of entailment that returns three extra things: (i) consumed heap nodes, $\kappa_i$, (ii) existential variables used, $V_i$, and (iii) final consequent, $\pi_i$. The final consequent is used to return a constraint for $\{v^s\}$ via $\exists W_i \pi_i$. A set of answers is returned by the fold step as we allow it to explore multiple ways of matching up its disjunctive definition:

$$
\begin{align*}
\text{fold}^\nu (\kappa \land \pi, p : c(v^s)) = \\
\text{def } \left\{ \left( \Delta_i, \kappa_i, \exists W_i \pi_i \right) \right\}_i
\end{align*}
$$

## 5 Implementation

We have built a prototype system using Objective Caml. The proof obligations generated by our verification are discharged by our entailment proving procedure with the help of Omega Calculator [12] for arithmetic constraints of Presburger form and MONA [5] for set constraints.

Our system uses the set-of-states technique to implement multiple pre/post specification, but this technique is orthogonal to multiple pre/post. Set-of-states may also arise from disjunction in the consequent from entailment proving, even when single pre/post specifications are used. Our first set of experiments, in Figure 6, was designed to evaluate the cost of supporting set-of-states. We provide the time taken (in seconds) to verify each of those programs. Most of our examples can be verified within 5 seconds despite the use of a Presburger solver. Verification time of a function includes the time to verify all functions that it calls.

The examples show that proof search via set of states does not incur much overhead, since most of the time there is only one state. The average overhead introduced by the set of states for our examples is around 0.05 seconds. However, set of states is crucial to verify the count or delete methods of circular list, which is based on the $\text{isog}$ predicate, and in those examples that have been marked with a failed (to verify). From the experiments presented in Figure 6, we may conclude that the cost of supporting set of states is quite low. However, set of states may be crucial for some examples, as it may be required when handling disjunctive consequent or multiple pre/post.

Even if multiple pre/post conditions are not critically needed, there may be occasions when they are desirable due to the possibility of relying on smaller and simpler constraints during verification. In order to highlight the performance gains due to the use of multiple pre/post, Figure 7 compares the timings obtained for some programs with single pre/post (capturing the reachability property) that have been translated to equivalent programs with multiple pre/post. This translation exploits the use of smaller and simpler predicates, where possible, which resulted in faster overall verification. As the proof obligations generated for the examples in Figure 7 contain set constraints, we used MONA to discharge them when needed. However, for some examples with single pre/post, the complexity of the constraints caused an out-of-memory error.

As an example of breaking a single specification into multiple pre/post, let us consider the following predicate which describes a non-empty list sorted in ascending order.

$\text{insertion\_sort}$

$$
\begin{align*}
\exists p, \text{root}: \text{node}(m, \text{null}) \land B = \{m\} \land n = 1 \\
\lor \exists p, \text{root}: \text{node}(m, p) \land p:\text{insert}(n-1, m, B) \\
\land m \leq m_1 \land B = B_1 \cup \{m\} \land \text{inv } n > 0
\end{align*}
$$

In order to transform an example using the above predicate into multiple pre/post, $\text{insert}$ can be split into three smaller predicates, each of them specialized on a certain property: $\text{llR}$ which captures the set/bag of values and was already defined in Section 3, $\text{lllM}$ which captures the sortness and $\text{ll}$ which captures the size.

$\text{lllM}(m) \equiv \text{root}: \text{node}(m, \text{null})$

\begin{align*}
\lor \exists p, \text{root}: \text{node}(m, p) \land p:\text{insert}(m_i) \land m \leq m_1 \\
\land m = m \land B = B_1 \lor \{m\} \land \text{inv } n > 0
\end{align*}

By applying the above splitting to the insertion sort algorithm to obtain three smaller pairs of pre/post specification, we have managed to verify it in 3.96 seconds, while with a
We have introduced set of states as a way to support multiple pre/post specifications in separation logic. This approach allows proof search to be captured explicitly, permitting a simple automated search strategy to be developed. Our current search strategy is exhaustive but directed and is guaranteed to terminate.

Multiple pre/post method specifications enhances proof search in separation logic. This feature puts creative control back into users’ hands. Nevertheless, we provide machine support for automatically checking and then applying each high-level specification from the user. We believe that multiple pre/post specifications can greatly enhance both the expressivity and performance of automated verification via separation logic. As shown by our initial experiments, it allows more verifications to be carried out successfully and with potential improvement to performance.

Acknowledgement

This work is supported by A*STAR research grant R-252-000-233-205, and Shengchao Qin is supported in part by EPSRC grant EP/E021948/1.

References


