

Gradually Verified Language with Recursive Predicates

Henry Blanchette

Contents

1	Weakest Liberal Predonditions	2
1.1	Concrete Weakest Liberal Precondition (WLP) Rules	2
1.1.1	Handling Method Calls	3
1.1.2	Handling While Loops	4
1.2	Assumed and Tainted Logic	4
1.2.1	Assumptions	5
1.2.2	Taints	5
1.3	Utility Functions	6
1.4	Gradual Weakest Liberal Precondition (WLP) Rules	8
1.4.1	Handling Gradual Method Calls	9
1.4.2	Handling Gradual While Loops	10
1.5	Gradual Utility Functions	10

1 Weakest Liberal Predonditions

1.1 Concrete Weakest Liberal Precondition (WLP) Rules

$\text{WLP} : \text{STATEMENT} \times \text{FRMSATFORMULA} \rightarrow \text{FRMSATFORMULA}$

$\text{WLP}(s, \phi) := \text{match } s \text{ with}$

skip	$\mapsto \phi$
$s_1; s_2$	$\mapsto \text{WLP}(s_1, \text{WLP}(s_2, \phi))$
$T \ x$	$\mapsto \text{assert } x \text{ does not appear in } \phi; \phi$
$x := e$	$\mapsto \text{required}(e) \wedge [e/x]\phi$
$x := \text{new } C$	$\mapsto [\text{new}(C)/x]\phi$
$x.f := y$	$\mapsto \text{required}(x.f) \wedge [y/x.f]\phi$
$y := z.m_C(\bar{e})$	$\mapsto \text{required}(\bar{e}) \wedge z \neq \text{null} \wedge \text{pre}(z.m_C(\bar{e})) * \text{handleMethodCall}(z.m_C(\bar{e}), \phi)$
if $(e) \{s_{\text{the}}\} \text{ else } \{s_{\text{els}}\}$	$\mapsto \text{required}(e) \wedge$ $\text{if } (e) \text{ then } \text{WLP}(s_{\text{the}}, \phi) \text{ else } \text{WLP}(s_{\text{els}}, \phi)$
while $(e) \text{ invariant } \phi_{\text{inv}} \{s_{\text{bod}}\}$	$\mapsto \text{required}(e) \wedge \phi_{\text{inv}} \wedge$ $(\text{if } (e) \text{ then } \text{WLP}(s_{\text{bod}}, \phi_{\text{inv}}) \text{ else true}) * \text{handleWhileLoop}(\phi_{\text{inv}}, s_{\text{bod}}, \phi)$
assert ϕ_{ass}	$\mapsto \text{required}(\phi_{\text{ass}}) \wedge \phi_{\text{ass}} \wedge \phi$
hold $\phi_{\text{hol}} \{s_{\text{bod}}\}$	$\mapsto (\text{unimplemented})$
release ϕ_{rel}	$\mapsto (\text{unimplemented})$
unfold $\alpha_C(\bar{e})$	$\mapsto \text{required}(\alpha_C(\bar{e})) \wedge$ $[\text{body}(\alpha_C(\bar{e}))/\alpha_C(\bar{e}), \phi'/\text{unfolding } \alpha_C(\bar{e}) \text{ in } \phi']\phi$
fold $\alpha_C(\bar{e})$	$\mapsto \text{required}(\bar{e}) \wedge [\alpha_C(\bar{e})/\text{body}(\alpha_C(\bar{e}))]\phi$

Since WLP takes a framed, satisfiable formula and yields a framed, satisfiable formula, there is an implicit check that asserts these properties before and after WLP is computed. Note that the substitutions in the above rules do not substitute instances that appear inside of accesses (i.e. of the form $\text{acc}(e.f)$) or meta-predicates such as **tainted**, etc.

Note the following syntax rules:

- The OCaml-inspired syntax of the form $a; s$ for side-effects in evaluation is defined as “execute side-effect a , then evaluate as s .”
- The meta-function **assert** \cdot is executed imperitively, raising an error if the argument is false.

Finally, the idiom “ a appears in b ” is defined as follows:

e appears in e'	\iff	$\exists e'_{\text{sub}}$ a sub-expression of $e' : e = e'_{\text{sub}}$
e appears in acc (e')	\iff	false
e appears in if e' then ϕ_{the} else ϕ_{els}	\iff	e appears in at least one of $e', \phi_{\text{the}}, \phi_{\text{els}}$
e appears in $\alpha_C(\bar{e})$	\iff	e appears in at least one of \bar{e}

1.1.1 Handling Method Calls

The `handleMethodCall` helper function, for a given method call $z.m_C(\bar{e})$ and post-condition ϕ , does the following:

- assert that permissions in **required**(ϕ) and granted by **pre**($z.m_C(\bar{e})$) are also granted by **post**($z.m_C(\bar{e})$)
- assume taint-substituted **pre**($z.m_C(\bar{e})$)
- return taint-substituted ϕ

The following definition reflects the above descriptions, in order:

```

handleMethodCall( $z.m_C(\bar{e})$ ,  $\phi$ ) :=
  assert granted(post( $z.m_C(\bar{e})$ ))  $\impl$   $\pi$ ,
     $\forall \pi : \text{required}(\phi), \text{granted}(\text{pre}(z.m_C(\bar{e}))) \impl \pi$ ;
  assume [tainteduid( $z.m_C(\bar{e})$ )( $r$ )/ $r : r$  isTaintedBy  $z.m_C(\bar{e})$ ]pre( $z.m_C(\bar{e})$ );
  [tainteduid( $z.m_C(\bar{e})$ )( $r$ )/ $r : r$  isTaintedBy  $z.m_C(\bar{e})$ ] $\phi$ 

```

1.1.2 Handling While Loops

The `handleWhileLoop` helper function, for a given while loop with condition e , invariant ϕ_{inv} , and post-condition ϕ , does the following:

- assume taint-substituted ϕ_{inv}
- return taint-substituted ϕ

The following definition reflects the above descriptions, in order:

$$\begin{aligned} \text{handleWhileLoop}(\phi_{\text{inv}}, s_{\text{bod}}, \phi) &:= \\ &\text{assume } [\text{tainted}_{\text{uid}(\text{while}(e, \phi_{\text{inv}}))}(r)/r : r \text{ isTaintedBy } s_{\text{bod}}] \phi_{\text{inv}}; \\ &[\text{tainted}_{\text{uid}(\text{while}(e, \phi_{\text{inv}}))}(r)/r : r \text{ isTaintedBy } s_{\text{bod}}] \phi \end{aligned}$$

1.2 Assumed and Tainted Logic

Assumed logic concerns assumed formulas that do not result directly from statically verifying the visible code. *Tainted logic* concerns how references (variables and field references) may have their referenced values changed by sources external to the visible code. These logics are handled in the following cases:

- Method calls — The specification of a called method is visible, but the body is not visible due to the (intended) modular structure of verification. So, the validity of the called method's implementation is assumed. Additionally, a method call taints references that it requires access to.
- While loops — the actual execution of a while loop's body is statically invisible since the number of times the while loop's body will execute is not statically calculated. So, references that are set inside the while loop's body are tainted.

Define a *reference*, r , to be an instance of x (a variable), $e.f$ or $\alpha_C(\bar{e})$. Then *access to a reference* is defined as follows:

$$\text{access}(r) := \begin{cases} \text{false} & \text{if } r = x \\ \text{acc}(e.f) & \text{if } r = e.f \\ \alpha_C(\bar{e}) & \text{if } r = \alpha_C(\bar{e}) \end{cases}$$

1.2.1 Assumptions

The *assumed* formula, local to the encompassing highest-level $\text{WLP}(s, \phi)$ calculation, represents the truths that are assumed via references external to the direct implications of s . For example, the post-condition of a method call appearing in s may yield truths that are accepted as assumptions due to the modular structure of verification — the method call is assumed to be verified separately (modularly).

These truths must be kept separate from $\phi_{\text{WLP}} := \text{WLP}(s, \phi)$ because they do not need to be implied by the pre-condition concerning ϕ_{WLP} . The $\text{assume}(\phi)$ function is how these truths are accumulated during the WLP computation.

$\text{assume } \phi := \text{ set the } \textit{assumed} \text{ formula, } \phi_{\text{ass}}, \text{ to } \phi \wedge \phi_{\text{ass}}$

1.2.2 Taints

The *tainted* meta-predicate indicates that the wrapped reference has been *tainted* by a source identified by the given unique identifier. A *tainted* reference is one that relies on the values of parts of the heap that may have been changed externally. For example, if a method call requires access to $x.f$, then $x.f$ is tainted because the method call could have changed the value of $x.f$.

Tainted references can only be asserted in some specific ways. For example, the previously mentioned method call could ensure that $x.f = v$, where v is some value, and this would yield the *assumption* that $\text{tainted}_{\text{uid}(z.m_C(\bar{e}))}(x.f) = v$. The following rules define the `isTaintedBy` relation between references (left) and statements or statement-fragments (right).

$$\begin{array}{ll}
r \text{ isTaintedBy } r := e & \Leftarrow \text{ true} \\
r \text{ isTaintedBy } y := z.m_C(\bar{e}) & \iff r \text{ isTaintedBy } z.m_C(\bar{e}) \\
r \text{ isTaintedBy } z.m_C(\bar{e}) & \iff \text{required}(\text{pre}(z.m_C(\bar{e}))) \implies \text{access}(r) \\
r \text{ isTaintedBy } s_1 ; s_2 & \iff r \text{ isTaintedBy } s_1 \vee r \text{ isTaintedBy } s_2
\end{array}$$

The $\text{uid}(\cdot)$ function generates a unique identifier for the given instance. This is needed because instances that contain the same arguments but appear in different parts of a program (where heap state may be different) must be treated as unique. The following function gathers all the references tainted via the arguments:

1.3 Utility Functions

The implementations of the functions in this section can be made much more efficient than the naive definition here in mathematical notation. For example, calculating the footprint of expressions and formulas can avoid redundancy by not generating permission-subformulas that are already satisfied. This can be implemented as implicit in \wedge by a wrapper \wedge_{wrap} operation in some way similar to this:

$$\phi \wedge_{\text{wrap}} \phi' := \begin{cases} \phi & \text{if } \phi \implies \phi' \\ \phi \wedge \phi' & \text{otherwise} \end{cases}$$

The following functions are useful abbreviations for common constructs.

$\text{new}(C)$	$:=$	an object that is a new instance of class C , where all fields are assigned to their default values
$\text{pre}(z.m_C(\bar{e}))$	$:=$	$[z/\mathbf{this}, \bar{e}/x]\text{pre}(m_C)$
$\text{pre}(m_C)$	$:=$	the static-contract pre-condition of m_C
$\text{post}(z.m_C(\bar{e}))$	$:=$	$[z/\mathbf{this}, \overline{e/\text{old}(x)}]\text{post}(m_C)$
$\text{post}(m_C)$	$:=$	the static-contract post-condition of m_C
$\text{body}(\alpha_C)$	$:=$	the body formula of α_C
$\text{body}(\alpha_C(\bar{e}))$	$:=$	$[\bar{e}/x]\text{body}(\alpha_C)$

The footprint function, $\text{required}(\cdot)$, generates a formula containing all the permissions necessary to frame its argument. With efficient implementations of a wrapped \wedge , this can result in the smallest such formula.

$$\begin{aligned}
\text{required}(e) &:= \text{match } e \text{ with} \\
&\quad \left| \begin{array}{ll} e.f & \mapsto \text{required}(e') \wedge e' \neq \text{null} \wedge \text{acc}(e'.f) \\ e_1 \oplus e_2 & \mapsto \text{required}(e_1) \wedge \text{required}(e_2) \\ e_1 \odot e_2 & \mapsto \text{required}(e_1) \wedge \text{required}(e_2) \\ e & \mapsto \text{true} \end{array} \right. \\
\text{required}(\bar{e}) &:= \bigwedge \text{required}(e) \\
\text{required}(\phi) &:= \bigwedge \{ \text{required}(e) : e \text{ appears in } \phi \} \wedge \\
&\quad \bigwedge \{ \alpha_C(\bar{e}) : \text{unfolding } \alpha_C(\bar{e}) \text{ in } \phi' \text{ appears in } \phi \}
\end{aligned}$$

1.4 Gradual Weakest Liberal Precondition (WLP) Rules

$\widetilde{\text{WLP}} : \text{STATEMENT} \times \text{FRMSATFORMULA} \rightarrow \text{FRMSATFORMULA}$

$\widetilde{\text{WLP}}(s, \tilde{\phi}) := \text{match } s \text{ with}$

$s_1; s_2$	$\mapsto \widetilde{\text{WLP}}(s_1, \widetilde{\text{WLP}}(s_2, \tilde{\phi}))$
$y := z.m_C(\bar{e})$	$\mapsto \text{handleMethodCall}(z.m_C(\bar{e}), \tilde{\phi})$
$\text{if } (e) \{s_{\text{the}}\} \text{ else } \{s_{\text{els}}\}$	$\mapsto \text{required}(e) \wedge \text{imprecision}(\widetilde{\text{WLP}}(s_{\text{the}}, \tilde{\phi}) \wedge \widetilde{\text{WLP}}(s_{\text{els}}, \tilde{\phi})) \wedge$ $(\text{if } (e) \text{ then concrete}(\widetilde{\text{WLP}}(s_{\text{the}}, \tilde{\phi})) \text{ else concrete}(\widetilde{\text{WLP}}(s_{\text{els}}, \tilde{\phi})))$
$\text{while } (e) \text{ invariant } \tilde{\phi}_{\text{inv}} \{s_{\text{bod}}\}$	$\mapsto \text{handleWhileLoop}(\tilde{\phi}_{\text{inv}}, s_{\text{bod}}, \tilde{\phi})$
$\text{unfold } \alpha_C(\bar{e})$	$\mapsto \text{required}(\alpha_C(\bar{e})) \wedge [\text{body}(\alpha_C(\bar{e}))/\alpha_C(\bar{e}), \phi'/\text{unfolding } \alpha_C(\bar{e}) \text{ in } \phi']\tilde{\phi}$
$\text{fold } \alpha_C(\bar{e})$	$\mapsto \text{required}(\bar{e}) \wedge [\alpha_C(\bar{e})/\text{body}(\alpha_C(\bar{e}))]\tilde{\phi}$
otherwise	$\mapsto \text{WLP}(s, \text{concrete}(\phi)) * \text{imprecision}(\phi)$

Note the following syntactical conventions:

- $*$ binds tighter than \wedge .
- Since $* ?$ is syntactical sugar, $(\phi_1 * ?) \circledast (\phi_2 * ?) := \phi_1 \circledast \phi_2 * ?$, preserving $* ?$ at the top level of the formula. This is because $* ?$ cannot and should not nest. Due to this, it would make more sense to write $\wedge ?$ actually.
- The meta-function $\text{imprecision}(\tilde{\phi})$ is defined as follows:

$$\phi' \circledast \text{imprecision}(\tilde{\phi}) := \begin{cases} \phi' \circledast \tilde{\phi} & \text{if } \tilde{\phi} \text{ is imprecise} \\ \phi' & \text{if } \tilde{\phi} \text{ is concrete} \end{cases}$$

1.4.1 Handling Gradual Method Calls

This handler builds upon the structure of the handler defined in section 1.1.1. However it also handle's certain cases of imprecision specially. Overall, $\text{handleMethodCall}(z.m_C(\bar{e}), \tilde{\phi})$ satisfies the following, where $\phi_{\text{pre}} := \text{pre}(z.m_C(\bar{e}))$ and $\phi_{\text{post}} := \text{post}(z.m_C(\bar{e}))$.

- If each of $\tilde{\phi}_{\text{pre}}, \tilde{\phi}_{\text{post}}, \tilde{\phi}$ are concrete, then $\widetilde{\text{WLP}}$ is exactly WLP.
- The imprecision of $\tilde{\phi}_{\text{pre}}, \tilde{\phi}_{\text{post}}, \tilde{\phi}$ are propogated through to the result of $\widetilde{\text{WLP}}$. In other words, if at least one of them are imprecise, then the result of $\widetilde{\text{WLP}}$ is imprecise.
- $\text{assume concrete}(\tilde{\phi}_{\text{post}})$. Only the concrete part is assumed because the imprecise cases are handled by the points below.
- $\text{required}(\tilde{\phi}_{\text{pre}})$ is included in the result of $\widetilde{\text{WLP}}$. Note that, via the definitions in the next section, this only considers the concrete part of $\tilde{\phi}_{\text{pre}}$.
- For each permission required by $\tilde{\phi}$ and granted by $\tilde{\phi}_{\text{pre}}$ and required by $\tilde{\phi}$, the permission must also be granted by $\tilde{\phi}_{\text{post}}$. This is the same concept behind **assert** in concrete handleMethodCall , but its interactions with imprecision are special in the following cases:
 - If $\tilde{\phi}_{\text{pre}}$ is concrete and $\tilde{\phi}_{\text{post}}$ is imprecise, then $\text{required}(\tilde{\phi}) \setminus \text{granted}(\tilde{\phi}_{\text{pre}})$ is included in the result of $\widetilde{\text{WLP}}$. This is because even though $\tilde{\phi}_{\text{post}}$ is imprecise, it is impossible for it to grant permissions not granted by $\tilde{\phi}_{\text{pre}}$.
 - If $\tilde{\phi}_{\text{pre}}$ is imprecise and $\tilde{\phi}_{\text{post}}$ is concrete, then $\text{assume } \tilde{\phi}_{\text{post}}$ and nothing from $\tilde{\phi}$ is propogated to the result of $\widetilde{\text{WLP}}$. This is because the method call “consumes” an unknown set of permissions that may not be given back via $\tilde{\phi}_{\text{post}}$. This is reflected by the imprecision in the result of $\widetilde{\text{WLP}}$ propogated to it by the imprecision of $\tilde{\phi}_{\text{pre}}$.
 - If both of $\tilde{\phi}_{\text{pre}}, \tilde{\phi}_{\text{post}}$ are imprecise, then nothing from $\tilde{\phi}$ is propogated to the result of $\widetilde{\text{WLP}}$. This is because the method call could, since it potentially has permission to access anything in its imprecise pre-condition, imply anything ranging from none of $\tilde{\phi}$ to all of $\tilde{\phi}$. This unknown is reflected by the imprecision of the resulting $\widetilde{\text{WLP}}$ by propogating to it the imprecision of $\tilde{\phi}_{\text{pre}}$ and $\tilde{\phi}_{\text{post}}$.

$$\begin{aligned}
& \text{handleMethodCall}(z.m_C(\bar{e}), \tilde{\phi}) := \\
& \quad \text{assume}(\text{concrete}(\tilde{\phi}_{\text{post}})); \\
& \quad \text{imprecision}(\tilde{\phi}_{\text{pre}}, \tilde{\phi}_{\text{post}}, \tilde{\phi}) * \text{required}(\bar{e}) \wedge z \neq \text{null} \wedge \text{required}(\tilde{\phi}_{\text{pre}}) * \\
& \quad \text{match } \tilde{\phi}_{\text{pre}}, \tilde{\phi}_{\text{post}} \text{ with} \\
& \quad \left| \begin{array}{ll} \phi_{\text{pre}}, & \phi_{\text{post}} \quad \mapsto \quad \text{handleMethodCall}(z.m_C(\bar{e}), \tilde{\phi}) \\ \phi_{\text{pre}}, & \phi_{\text{post}} * ? \quad \mapsto \quad \text{required}(\tilde{\phi}) \setminus \text{granted}(\phi_{\text{pre}}) \\ \phi_{\text{pre}} * ?, & \phi_{\text{post}} \quad \mapsto \quad \text{assume } \phi_{\text{post}}; \text{true} \\ \phi_{\text{pre}} * ?, & \phi_{\text{post}} * ? \quad \mapsto \quad \text{true} \end{array} \right.
\end{aligned}$$

1.4.2 Handling Gradual While Loops

$$\begin{aligned}
& \text{handleWhileLoop}(\tilde{\phi}_{\text{inv}}, s_{\text{bod}}, \tilde{\phi}) := \\
& \quad \text{match } \tilde{\phi}_{\text{inv}} \text{ with} \\
& \quad \left| \begin{array}{ll} \phi_{\text{inv}} & \mapsto \quad \text{handleWhileLoop}(\phi_{\text{inv}}, s_{\text{bod}}, \text{concrete}(\tilde{\phi})) * \text{imprecision}(\tilde{\phi}) \\ \phi_{\text{inv}} * ? & \mapsto \quad \text{required}(e) \wedge \phi_{\text{inv}} * \widetilde{\text{WLP}}(\text{if } (e) \{s_{\text{bod}}\} \text{ else } \{\text{skip}\}, \phi_{\text{inv}} * ?) * ? \end{array} \right.
\end{aligned}$$

1.5 Gradual Utility Functions

$\text{granted}(\tilde{\phi})$ preserves the imprecision of $\tilde{\phi}$ in order to reflect the fact that imprecision grants *at least* the permissions of $\text{concrete}(\tilde{\phi})$, and additionally may grant other permissions as well if the $\tilde{\phi}$ is imprecise. $\text{required}(\tilde{\phi})$, on the other hand, only considers $\text{concrete}(\tilde{\phi})$ because permissions required by sub-formulas introduced via imprecision will be checked for framing at the time they are introduced.

$$\begin{aligned}
\text{required}(\tilde{\phi}) &:= \text{required}(\text{concrete}(\tilde{\phi})) \\
\text{granted}(\tilde{\phi}) &:= \text{granted}(\text{concrete}(\tilde{\phi})) * \text{imprecision}(\tilde{\phi})
\end{aligned}$$

Gradual substitution introduces instances not found in the given imprecise formula into the resulting imprecise formula.

$$[y/x](\phi * ?) \quad := \quad \begin{cases} ([y/x]\phi) * ? & \text{if } x \text{ appears in } \phi \\ \phi \wedge y * ? & \text{otherwise} \end{cases}$$