Verification of the Deutsch-Schorr-Waite Graph Marking Algorithm using Data Refinement

Viorel Preoteasa and Ralph-Johan Back

July 1, 2010

Abstract

The verification of the Deutsch-Schorr-Waite graph marking algorithm is used as a benchmark in many formalizations of pointer programs. The main purpose of this mechanization is to show how data refinement of invariant based programs can be used in verifying practical algorithms. The verification starts with an abstract algorithm working on a graph given by a relation *next* on nodes. Gradually the abstract program is refined into Deutsch-Schorr-Waite graph marking algorithm where only one bit per graph node of additional memory is used for marking.

Contents

1 Introduction

The verification of the Deutsch-Schorr-Waite (DSW) [14, 10] graph marking algorithm is used as a benchmark in many formalizations of pointer programs [11, 1]. The main purpose of this mechanization is to show how data refinement [12] of invariant based programs [3, 4, 5, 6] can be used in verifying practical algorithms.

The DSW algorithm marks all nodes in a graph that are reachable from a *root* node. The marking is achieved using only one extra bit of memory for every node. The graph is given by two pointer functions, *left* and *right*, which for any given node return its left and right successors, respectively. While marking, the left and right functions are altered to represent a stack that describes the path from the root to the current node in the graph. On completion the original graph structure is restored. We construct the DSW algorithm by a sequence of three successive data refinement steps. One step in these refinements is a generalization of the DSW algorithm to an algorithm which marks a graph given by a family of pointer functions instead of left and right only.

Invariant based programming is an approach to construct correct programs where we start by identifying all basic situations (pre- and post-conditions, and loop invariants) that could arise during the execution of the algorithm. These situations are determined and described before any code is written. After that, we identify the transitions between the situations, which together determine the flow of control in the program. The transitions are verified at the same time as they are constructed. The correctness of the program is thus established as part of the construction process.

Data refinement [9, 2, 7, 8] is a technique of building correct programs working on concrete data structures as refinements of more abstract programs working on abstract data structures. The correctness of the final program follows from the correctness of the abstract program and from the correctness of the data refinement.

Both the semantics and the data refinement of invariant based programs were formalized in [13], and this verification is based on them.

We use a simple model of pointers where addresses (pointers, nodes) are the elements of a set and pointer fields are global pointer functions from addresses to addresses. Pointer updates ($x.left := y$) are done by modifying the global pointer function $left := left(x := y)$. Because of the nature of the marking algorithm where no allocation and disposal of memory are needed we do not treat these operations.

A number of Isabelle techniques are used here. The class mechanism is used for extending the complete lattice theories as well as for introducing well founded and transitive relations. The polimorphism is used for the state of the computation. In [13] the state of computation was introduced as a type variable, or even more generaly, state predicates were introduced as elements of a complete (boolean) lattice. Here the state of the computation is instantiated with various tuples ranging from the abstract data in the first algorithm to the concrete data in the final refinement. The locale mechanism of Isabelle is used to introduce the specification variables and their invariants. These specification variables are used for example to prove that the main variables are restored to their initial values when the algorithm terminates. The locale extension and partial instantiation mechanisms turn out to be also very useful in the data refinements of DSW. We start with a locale which fixes the abstract graph as a relation *next* on nodes. This locale is first partially interpreted into a locale which replaces *next* by a union of a family of pointer functions. In the final refinement step the locale of the pointer functions is interpreted into a locale with only two pointer functions, *left* and *right*.

2 Address Graph

theory *Graph*

imports *Main*

begin

This theory introduces the graph to be marked as a relation next on nodes (addresses). We assume that we have a special node nil (the null address). We have a node root from which we start marking the graph. We also assume that nil is not related by next to any node and any node is not related by next to nil.

locale *node* =
  fixes *nil* :: *'node*
  fixes *root* :: *'node*

locale *graph* = *node* +
  fixes *next* :: (*'node* × *'node*) *set*
  assumes *next-not-nil-left*: (!! *x* . (*nil*, *x*) ∉ *next*)
  assumes *next-not-nil-right*: (!! *x* . (*x*, *nil*) ∉ *next*)
begin

On lists of nodes we introduce two operations similar to existing hd and tl for getting the head and the tail of a list. The new function head applied to a nonempty list returns the head of the list, and it reurns nil when applied to the empty list. The function tail returns the tail of the list when applied to a non-empty list, and it returns the empty list otherwise.

  definition
    head S ≡ (*if S* = [] *then nil else* (*hd S*))

  definition
    tail (*S*::*'a list*) ≡ (*if S* = [] *then* [] *else* (*tl S*))

  lemma [*simp*]: ((*nil*, *x*) ∈ *next*) = *False*
    ⟨*proof*⟩

  lemma [*simp*]: ((*x*, *nil*) ∈ *next*) = *False*
    ⟨*proof*⟩

  theorem *head-not-nil* [*simp*]:
    (*head S* ≠ *nil*) = (*head S* = *hd S* ∧ *tail S* = *tl S* ∧ *hd S* ≠ *nil* ∧ *S* ≠ [])
    ⟨*proof*⟩

  theorem *nonempty-head* [*simp*]:
    head (*x* # *S*) = *x*
    ⟨*proof*⟩

  theorem *nonempty-tail* [*simp*]:
    tail (*x* # *S*) = *S*
    ⟨*proof*⟩

  end

end

3 Marking Using a Set

theory *SetMark*

imports *Graph ../DataRefinementIBP/DataRefinement*

begin

We construct in this theory a diagram which computes all reachable nodes from a given root node in a graph. The graph is defined in the theory Graph and is given by a relation *next* on the nodes of the graph.

The diagram has only three ordered situation (*init* > *loop* > *final*). The termination variant is a pair of a situation and a natural number with the lexicographic ordering. The idea of this ordering is that we can go from a bigger situation to a smaller one, however if we stay in the same situation the second component of the variant must decrease.

The idea of the algorithm is that it starts with a set X containing the root element and the root is marked. As long as X is not empty, if $x \in X$ and y is an unmarked sucessor of x we add y to X. If $x \in X$ has no unmarked sucessors it is removed from X. The algorithm terminates when X is empty.

datatype I = *init* | *loop* | *final*

declare *I.split* [*split*]

instantiation I :: *well-founded-transitive*
begin

definition
less-I-def: $i < j \equiv (j = init \wedge (i = loop \vee i = final)) \vee (j = loop \wedge i = final)$

definition
less-eq-I-def: $(i::I) \leq (j::I) \equiv i = j \vee i < j$

instance ⟨*proof*⟩

end

definition (**in** *graph*)
reach $x \equiv \{y\ .\ (x, y) \in next^* \wedge y \neq nil\}$

theorem (**in** *graph*) *reach-nil* [*simp*]: *reach nil* = {}
⟨*proof*⟩

theorem (**in** *graph*) *reach-next*: $b \in reach\ a \Longrightarrow (b, c) \in next \Longrightarrow c \in reach\ a$
⟨*proof*⟩

definition (**in** *graph*)
path S mrk $\equiv \{x\ .\ (\exists\ s\ .\ s \in S \wedge (s, x) \in next\ O\ (next \cap ((-mrk) \times (-mrk)))^*$
$)\}$

The set *path S mrk* contains all reachable nodes from S along paths with unmarked nodes.

lemma (**in** *graph*) *trascl-less*: $x \neq y \Longrightarrow (a,\ x) \in R^* \Longrightarrow$
$((a,x) \in (R \cap (-\{y\})\times(-\{y\}))^* \vee\ (y,x) \in R\ O\ (R \cap (-\{y\})\times(-\{y\}))^*\ )$
⟨*proof*⟩

lemma (**in** *graph*) *add-set* [*simp*]: $x \neq y \Longrightarrow x \in path\ S\ mrk \Longrightarrow x \in path\ (insert\ y\ S)\ (insert\ y\ mrk)$
⟨*proof*⟩

lemma (**in** *graph*) *add-set2*: $x \in path\ S\ mrk \Longrightarrow x \notin path\ (insert\ y\ S)\ (insert\ y\ mrk) \Longrightarrow x = y$
⟨*proof*⟩

lemma (**in** *graph*) *del-stack* [*simp*]: $(\forall\ y\ .\ (t,\ y) \in next \longrightarrow y \in mrk) \Longrightarrow x \notin mrk \Longrightarrow x \in path\ S\ mrk \Longrightarrow x \in path\ (S - \{t\})\ mrk$
⟨*proof*⟩

lemma (**in** *graph*) *init-set* [*simp*]: $x \in reach\ root \Longrightarrow x \neq root \Longrightarrow x \in path\ \{root\}\ \{root\}$
⟨*proof*⟩

lemma (**in** *graph*) *init-set2*: $x \in reach\ root \Longrightarrow x \notin path\ \{root\}\ \{root\} \Longrightarrow x = root$
⟨*proof*⟩

3.1 Transitions

definition (**in** *graph*)
$Q1 \equiv \lambda\ (X::('node\ set),\ mrk::('node\ set))\ .\ \{\ (X'::('node\ set),\ mrk')\ .\ (root::'node) = nil \wedge X' = \{\} \wedge mrk' = mrk\}$

definition (**in** *graph*)
$Q2 \equiv \lambda\ (X::('node\ set),\ mrk::('node\ set))\ .\ \{\ (X',\ mrk')\ .\ (root::'node) \neq nil \wedge X' = \{root::'node\} \wedge mrk' = \{root::'node\}\}$

definition (**in** *graph*)
$Q3 \equiv \lambda\ (X,\ mrk)\ .\ \{\ (X',\ mrk')\ .$
$(\exists\ x \in X\ .\ \exists\ y\ .\ (x,\ y) \in next \wedge y \notin mrk \wedge X' = X \cup \{y\} \wedge mrk' = mrk \cup \{y\})\}$

definition (**in** *graph*)
$Q4 \equiv \lambda\ (X,\ mrk)\ .\ \{\ (X',\ mrk')\ .$
$(\exists\ x \in X\ .\ (\forall\ y\ .\ (x,\ y) \in next \longrightarrow y \in mrk) \wedge X' = X - \{x\} \wedge mrk' = mrk)\}$

definition (**in** *graph*)
$Q5 \equiv \lambda\ (X::('node\ set),\ mrk::('node\ set))\ .\ \{\ (X'::('node\ set),\ mrk')\ .\ X = \{\} \wedge mrk = mrk'\}$

3.2 Invariants

definition (**in** *graph*)
$Loop \equiv \{ (X, mrk) .$
$finite\ (-mrk) \land finite\ X \land X \subseteq mrk \land$
$mrk \subseteq reach\ root \land reach\ root \cap -mrk \subseteq path\ X\ mrk\}$

definition
$trm \equiv \lambda\ (X, mrk) .\ 2 * card\ (-mrk) + card\ X$

definition
$term\text{-}eq\ t\ w = \{s .\ t\ s = w\}$

definition
$term\text{-}less\ t\ w = \{s .\ t\ s < w\}$

lemma *union-term-eq*[*simp*]: $(\bigcup\ w .\ term\text{-}eq\ t\ w) = UNIV$
⟨*proof*⟩

lemma *union-less-term-eq*[*simp*]: $(\bigcup v{\in}\{v.\ v < w\}.\ term\text{-}eq\ t\ v) = term\text{-}less\ t\ w$
⟨*proof*⟩

definition (**in** *graph*)
$Init \equiv \{ (X{::}('node\ set), mrk{::}('node\ set)) .\ finite\ (-mrk) \land mrk = \{\}\}$

definition (**in** *graph*)
$Final \equiv \{ (X{::}('node\ set), mrk{::}('node\ set)) .\ mrk = reach\ root\}$

definition (**in** *graph*)
$SetMarkInv\ i = (case\ i\ of$
$I.init \Rightarrow Init\ |$
$I.loop \Rightarrow Loop\ |$
$I.final \Rightarrow Final)$

definition (**in** *graph*)
$SetMarkInvFinal\ i = (case\ i\ of$
$I.final \Rightarrow Final\ |$
$\text{-} \Rightarrow \{\})$

definition (**in** *graph*) [*simp*]:
$SetMarkInvTerm\ w\ i = (case\ i\ of$
$I.init \Rightarrow Init\ |$
$I.loop \Rightarrow Loop \cap \{s .\ trm\ s = w\}\ |$
$I.final \Rightarrow Final)$

definition (**in** *graph*)
$SetMark\text{-}rel \equiv \lambda\ (i, j) .\ (case\ (i, j)\ of$
$(I.init, I.loop) \Rightarrow Q1 \sqcup Q2\ |$
$(I.loop, I.loop) \Rightarrow Q3 \sqcup Q4\ |$
$(I.loop, I.final) \Rightarrow Q5\ |$

- ⇒ ⊥)

3.3 Diagram

definition (**in** *graph*)
SetMark ≡ λ (*i*, *j*) . (*case* (*i*, *j*) *of*
(*I.init*, *I.loop*) ⇒ (*demonic Q1*) ⊓ (*demonic Q2*) |
(*I.loop*, *I.loop*) ⇒ (*demonic Q3*) ⊓ (*demonic Q4*) |
(*I.loop*, *I.final*) ⇒ *demonic Q5* |
- ⇒ *top*)

lemma (**in** *graph*) *dgr-dmonic-SetMark* [*simp*]:
dgr-demonic SetMark-rel = *SetMark*
⟨*proof*⟩

lemma (**in** *graph*) *SetMark-dmono* [*simp*]:
dmono SetMark
⟨*proof*⟩

3.4 Correctness of the transitions

lemma (**in** *graph*) *init-loop-1*[*simp*]: ⊨ *Init* {| *demonic Q1* |} *Loop*
⟨*proof*⟩

lemma (**in** *graph*) *init-loop-2*[*simp*]: ⊨ *Init* {| *demonic Q2* |} *Loop*
⟨*proof*⟩

lemma (**in** *graph*) *loop-loop-1*[*simp*]: ⊨ (*Loop* ∩ {*s* . *trm s* = *w*}) {| *demonic Q3* |} (*Loop* ∩ {*s*. *trm s* < *w*})
⟨*proof*⟩

lemma (**in** *graph*) *loop-loop-2*[*simp*]: ⊨ (*Loop* ∩ {*s* . *trm s* = *w*}) {| *demonic Q4* |} (*Loop* ∩ {*s*. *trm s* < *w*})
⟨*proof*⟩

lemma (**in** *graph*) *loop-final*[*simp*]: ⊨ (*Loop* ∩ {*s* . *trm s* = *w*}) {| *demonic Q5* |} *Final*
⟨*proof*⟩

lemma *union-term-w*[*simp*]: (⋃ *w*. {*s*. *t s* = *w*}) = *UNIV*
⟨*proof*⟩

lemma *union-less-term-w*[*simp*]: (⋃ *v*∈{*v*. *v* < *w*}. {*s*. *t s* = *v*}) = {*s* . *t s* < *w*}
⟨*proof*⟩

lemma *sup-union*[*simp*]: *SUP A i* = (⋃ *w* . *A w i*)
⟨*proof*⟩

lemma *empty-pred-false*[*simp*]: $\{\}\ a = \mathit{False}$
⟨*proof*⟩

lemma *forall-simp*[*simp*]: $(!a\ b.\ \forall\ x\in A\ .\ (a = (t\ x)) \longrightarrow (h\ x) \lor b \neq u\ x) = (\forall\ x\in A\ .\ h\ x)$
⟨*proof*⟩

lemma *forall-simp2*[*simp*]: $(!a\ b.\ \forall\ x\in A\ .\ !y\ .\ (a\ = t\ x\ y) \longrightarrow (h\ x\ y) \longrightarrow (g\ x\ y) \lor b \neq u\ x\ y) = (\forall\ x\in A\ .\ !\ y\ .\ h\ x\ y \longrightarrow g\ x\ y\ )$
⟨*proof*⟩

3.5 Diagram correctness

The termination ordering for the *SetMark* diagram is the lexicographic ordering on pairs (i, n) where $i \in I$ and $n \in nat$.

interpretation *DiagramTermination* $\lambda\ (n{::}nat)\ (i :: I)\ .\ (i, n)$
⟨*proof*⟩

theorem (**in** *graph*) *SetMark-correct*:
$\models$ *SetMarkInv* {|*pt SetMark*|} *SetMarkInvFinal*
⟨*proof*⟩

theorem (**in** *graph*) *SetMark-correct1* [*simp*]:
Hoare-dgr SetMarkInv SetMark (*SetMarkInv* $\sqcap$ ($-$ *grd* (*step SetMark*)))
⟨*proof*⟩

theorem (**in** *graph*) *stack-not-nil* [*simp*]:
$(mrk, S) \in Loop \Longrightarrow x \in S \Longrightarrow x \neq nil$
⟨*proof*⟩

end

4 Marking Using a Stack

theory *StackMark*

imports *SetMark DataRefinement*

begin

In this theory we refine the set marking diagram to a diagram in which the set is replaced by a list (stack). Iniatially the list contains the root element and as long as the list is nonempty and the top of the list has an unmarked successor y, then y is added to the top of the list. If the top does not have unmarked sucessors, it is removed from the list. The diagram terminates when the list is empty.

The data refinement relation of the two diagrams is true if the list has distinct elements and the elements of the list and the set are the same.

consts
  $dist :: 'a\ list \Rightarrow bool$

primrec
  $dist\ [] \quad = True$
  $dist\ (a \# L) = (\neg\ a\ mem\ L \wedge dist\ L)$

4.1 Transitions

definition (**in** *graph*)
  $Q1'\ s \equiv let\ (stk::('node\ list),\ mrk::('node\ set)) = s\ in\ \{(stk'::('node\ list),\ mrk')$
.
  $root = nil \wedge stk' = [] \wedge mrk' = mrk\}$

definition (**in** *graph*)
  $Q2'\ s \equiv let\ (stk::('node\ list),\ mrk::('node\ set)) = s\ in\ \{(stk',\ mrk')\ .\ root \neq nil$
$\wedge\ stk' = [root] \wedge mrk' = mrk \cup \{root\}\}$

definition (**in** *graph*)
  $Q3'\ s \equiv let\ (stk,\ mrk) = s\ in\ \{(stk',\ mrk')\ .\ \ stk \neq [] \wedge (\exists\ y\ .\ (hd\ stk,\ y) \in$
$next\ \wedge$
  $y \notin mrk \wedge stk' = y \# stk \wedge mrk' = mrk \cup \{y\})\}$

definition (**in** *graph*)
  $Q4'\ s \equiv let\ (stk,\ mrk) = s\ in\ \{(stk',\ mrk')\ .\ stk \neq [] \wedge$
  $(\forall\ y\ .\ (hd\ stk,\ y) \in next \longrightarrow y \in mrk) \wedge stk' = tl\ stk \wedge mrk' = mrk\}$

definition
  $Q5'\ s \equiv let\ (stk,\ mrk) = s\ in\ \{(stk',\ mrk')\ .\ stk = [] \wedge mrk' = mrk\}$

4.2 Invariants

definition
  $Init' \equiv UNIV$

definition
  $Loop' \equiv \{\ (stk,\ mrk)\ .\ dist\ stk\}$

definition
  $Final' \equiv UNIV$

definition [*simp*]:
  $StackMarkInv\ i = (case\ i\ of$
    $I.init\ \Rightarrow Init'\ |$
    $I.loop\ \Rightarrow Loop'\ |$

$I.final \Rightarrow Final'$)

4.3 Data refinement relations

definition
$R1 \equiv \lambda\ (stk,\ mrk)\ .\ \{(X,\ mrk')\ .\ mrk' = mrk\}$

definition
$R2 \equiv \lambda\ (stk,\ mrk)\ .\ \{(X,\ mrk')\ .\ X = \{x\ .\ x\ mem\ stk\} \wedge (stk,\ mrk) \in Loop' \wedge mrk' = mrk\}$

definition [*simp*]:
$R\ i = (case\ i\ of$
$I.init \ \Rightarrow R1\ |$
$I.loop \ \Rightarrow R2\ |$
$I.final \Rightarrow R1)$

definition (**in** *graph*)
$StackMark\text{-}rel = (\lambda\ (i,\ j)\ .\ (case\ (i,\ j)\ of$
$(I.init,\ I.loop) \ \Rightarrow Q1' \sqcup Q2'\ |$
$(I.loop,\ I.loop) \ \Rightarrow Q3' \sqcup Q4'\ |$
$(I.loop,\ I.final) \Rightarrow Q5'\ |$
$\text{-} \Rightarrow \bot))$

4.4 Data refinement of the transitions

theorem (**in** *graph*) *init-nil* [*simp*]:
DataRefinement Init Q1 R1 R2 (demonic Q1')
⟨*proof*⟩

theorem (**in** *graph*) *init-root* [*simp*]:
DataRefinement Init Q2 R1 R2 (demonic Q2')
⟨*proof*⟩

theorem (**in** *graph*) *step1* [*simp*]:
DataRefinement Loop Q3 R2 R2 (demonic Q3')
⟨*proof*⟩

theorem (**in** *graph*) *step2* [*simp*]:
DataRefinement Loop Q4 R2 R2 (demonic Q4')
⟨*proof*⟩

theorem (**in** *graph*) *final* [*simp*]:
DataRefinement Loop Q5 R2 R1 (demonic Q5')
⟨*proof*⟩

4.5 Diagram data refinement

theorem (**in** *graph*) *StackMark-DataRefinement* [*simp*]:

DgrDataRefinement SetMarkInv SetMark-rel R (*dgr-demonic StackMark-rel*)
⟨*proof*⟩

4.6 Diagram correctness

theorem (**in** *graph*) *StackMark-correct*:
Hoare-dgr (*dangelic R SetMarkInv*) (*dgr-demonic StackMark-rel*) ((*dangelic R SetMarkInv*) $\sqcap$ ($-$ *grd* (*step* ((*dgr-demonic StackMark-rel*)))))
⟨*proof*⟩

end

5 Generalization of Deutsch-Schorr-Waite Algorithm

theory *LinkMark*

imports *StackMark*

begin

In the third step the stack diagram is refined to a diagram where no extra memory is used. The relation *next* is replaced by two new variables *link* and *label*. The variable $label : node \to index$ associates a label to every node and the variable $link : index \to node \to node$ is a collection of pointer functions indexed by the set *index* of labels. For $x \in node$, *link i x* is the successor node of x along the function *link i*. In this context a node x is reachable if there exists a path from the root to x along the links *link i* such that all nodes in this path are not *nil* and they are labeled by a special label $none \in index$.

The stack variable S is replaced by two new variables p and t ranging over nodes. Variable p stores the head of S, t stores the head of the tail of S, and the rest of S is stored by temporarily modifying the variables *link* and *label*.

This algorithm is a generalization of the Deutsch-Schorr-Waite graph marking algorithm because we have a collection of pointer functions instead of left and right only.

locale *pointer* = *node* +
fixes *none* :: *'index*
fixes *link0*::*'index* ⇒ *'node* ⇒ *'node*
fixes *label0* :: *'node* ⇒ *'index*

assumes (*nil*::*'node*) = *nil*
begin
definition *next* = $\{(a, b) . (\exists\ i . link0\ i\ a = b) \land a \neq nil \land b \neq nil \land label0\ a = none\}$

end

sublocale *pointer* $\subseteq$ *graph nil root next*
⟨*proof*⟩

The locale pointer fixes the initial values for the variables *link* and *label* and it defines the relation next as the union of all *link i* functions, excluding the mappings to *nil*, the mappings from *nil* as well as the mappings from elements which are not labeled by *none*.

The next two recursive functions, *label*_0, *link*_0 are used to compute the initial values of the variables *label* and *link* from their current values.

context *pointer*
begin
primrec
label-0:: ('*node* ⇒ '*index*) ⇒ ('*node list*) ⇒ ('*node* ⇒ '*index*) **where**
label-0 lbl [] = *lbl* |
label-0 lbl (*x* # *l*) = *label-0* (*lbl*(*x* := *none*)) *l*

lemma *label-cong* [*cong*]: *f* = *g* ⟹ *xs* = *ys* ⟹ *pointer.label-0 n f xs* = *pointer.label-0 n g ys*
⟨*proof*⟩

primrec
link-0:: ('*index* ⇒ '*node* ⇒ '*node*) ⇒ ('*node* ⇒ '*index*) ⇒ '*node* ⇒ ('*node list*) ⇒ ('*index* ⇒ '*node* ⇒ '*node*) **where**
link-0 lnk lbl p [] = *lnk* |
link-0 lnk lbl p (*x* # *l*) = *link-0* (*lnk*((*lbl x*) := ((*lnk* (*lbl x*))(*x* := *p*)))) *lbl x l*

The function *stack* defined bellow is the main data refinement relation connecting the stack from the abstract algorithm to its concrete representation by temporarily modifying the variable *link* and *label*.

primrec
stack:: ('*index* ⇒ '*node* ⇒ '*node*) ⇒ ('*node* ⇒ '*index*) ⇒ '*node* ⇒ ('*node list*) ⇒ *bool* **where**
stack lnk lbl x [] = (*x* = *nil*) |
stack lnk lbl x (*y* # *l*) =
(*x* ≠ *nil* ∧ *x* = *y* ∧ ¬ *x mem l* ∧ *stack lnk lbl* (*lnk* (*lbl x*) *x*) *l*)

lemma *label-out-range0* [*simp*]:
¬ *x mem S* ⟹ *label-0 lbl S x* = *lbl x*
⟨*proof*⟩

lemma *link-out-range0* [*simp*]:
¬ *x mem S* ⟹ *link-0 link label p S i x* = *link i x*
⟨*proof*⟩

lemma *link-out-range* [*simp*]: $\neg$ *x mem S* $\Longrightarrow$ *link-0 link* (*label*(*x* := *y*)) *p S* = *link-0 link label p S*
⟨*proof*⟩

lemma *empty-stack* [*simp*]: *stack link label nil S* = (*S* = [])
⟨*proof*⟩

lemma *stack-out-link-range* [*simp*]: $\neg$ *p mem S* $\Longrightarrow$ *stack* (*link*(*i* := (*link i*)(*p* := *q*))) *label x S* = *stack link label x S*
⟨*proof*⟩

lemma *stack-out-label-range* [*simp*]: $\neg$ *p mem S* $\Longrightarrow$ *stack link* (*label*(*p* := *q*)) *x S* = *stack link label x S*
⟨*proof*⟩

definition
g mrk lbl ptr x $\equiv$ *ptr x* $\neq$ *nil* $\wedge$ *ptr x* $\notin$ *mrk* $\wedge$ *lbl x* = *none*

lemma *g-cong* [*cong*]: *mrk* = *mrk1* $\Longrightarrow$ *lbl* = *lbl1* $\Longrightarrow$ *ptr* = *ptr1* $\Longrightarrow$ *x* = *x1* ==>
pointer.g n m mrk lbl ptr x = *pointer.g n m mrk1 lbl1 ptr1 x1*
⟨*proof*⟩

5.1 Transitions

definition
Q1'' *s* $\equiv$ *let* (*p*, *t*, *lnk*, *lbl*, *mrk*) = *s in* { (*p'*, *t'*, *lnk'*, *lbl'*, *mrk'*) .
root = *nil* $\wedge$ *p'* = *nil* $\wedge$ *t'* = *nil* $\wedge$ *lnk'* = *lnk* $\wedge$ *lbl'* = *lbl* $\wedge$ *mrk'* = *mrk*}

definition
Q2'' *s* $\equiv$ *let* (*p*, *t*, *lnk*, *lbl*, *mrk*) = *s in* { (*p'*, *t'*, *lnk'*, *lbl'*, *mrk'*) .
root $\neq$ *nil* $\wedge$ *p'* = *root* $\wedge$ *t'* = *nil* $\wedge$ *lnk'* = *lnk* $\wedge$ *lbl'* = *lbl* $\wedge$ *mrk'* = *mrk* $\cup$ {*root*}}

definition
Q3'' *s* $\equiv$ *let* (*p*, *t*, *lnk*, *lbl*, *mrk*) = *s in* { (*p'*, *t'*, *lnk'*, *lbl'*, *mrk'*) .
p $\neq$ *nil* $\wedge$
($\exists$ *i* . *g mrk lbl* (*lnk i*) *p* $\wedge$
p' = *lnk i p* $\wedge$ *t'* = *p* $\wedge$ *lnk'* = *lnk*(*i* := (*lnk i*)(*p* := *t*)) $\wedge$ *lbl'* = *lbl*(*p* := *i*) $\wedge$
mrk' = *mrk* $\cup$ {*lnk i p*})}

definition
Q4'' *s* $\equiv$ *let* (*p*, *t*, *lnk*, *lbl*, *mrk*) = *s in* { (*p'*, *t'*, *lnk'*, *lbl'*, *mrk'*) .
p $\neq$ *nil* $\wedge$
($\forall$ *i* . $\neg$ *g mrk lbl* (*lnk i*) *p*) $\wedge$ *t* $\neq$ *nil* $\wedge$
p' = *t* $\wedge$ *t'* = *lnk* (*lbl t*) *t* $\wedge$ *lnk'* = *lnk*(*lbl t* := (*lnk* (*lbl t*))(*t* := *p*)) $\wedge$ *lbl'*

$= lbl(t := none) \wedge$
$mrk' = mrk\}$

definition
$Q5''\ s \equiv let\ (p,\ t,\ lnk,\ lbl,\ mrk) = s\ in\ \{\ (p',\ t',\ lnk',\ lbl',\ mrk')\ .$
$p \neq nil \wedge$
$(\forall\ i\ .\ \neg\ g\ mrk\ lbl\ (lnk\ i)\ p) \wedge t = nil \wedge$
$p' = nil \wedge t' = t \wedge lnk' = lnk \wedge lbl' = lbl \wedge mrk' = mrk\}$

definition
$Q6''\ s \equiv let\ (p,\ t,\ lnk,\ lbl,\ mrk) = s\ in\ \{(p',\ t',\ lnk',\ lbl',\ mrk')\ .\ p = nil \wedge$
$p' = p \wedge t' = t \wedge lnk' = lnk \wedge lbl' = lbl \wedge mrk' = mrk\}$

5.2 Invariants

definition
$Init'' \equiv \{\ (p,\ t,\ lnk,\ lbl,\ mrk)\ .\ lnk = link0 \wedge lbl = label0\}$

definition
$Loop'' \equiv UNIV$

definition
$Final'' \equiv Init''$

5.3 Data refinement relations

definition
$R1' \equiv (\lambda\ (p,\ t,\ lnk,\ lbl,\ mrk)\ .\ \{(stk,\ mrk')\ .\ (p,\ t,\ lnk,\ lbl,\ mrk) \in Init'' \wedge mrk'$
$= mrk\})$

definition
$R2' \equiv (\lambda\ (p,\ t,\ lnk,\ lbl,\ mrk)\ .\ \{(stk,\ mrk')\ .$
$p = head\ stk \wedge$
$t = head\ (tail\ stk) \wedge$
$stack\ lnk\ lbl\ t\ (tail\ stk) \wedge$
$link0 = link\text{-}0\ lnk\ lbl\ p\ (tail\ stk) \wedge$
$label0 = label\text{-}0\ lbl\ (tail\ stk) \wedge$
$\neg\ nil\ mem\ stk \wedge$
$mrk' = mrk\})$

definition [*simp*]:
$R'\ i = (case\ i\ of$
$I.init \Rightarrow R1'\ |$
$I.loop \Rightarrow R2'\ |$
$I.final \Rightarrow R1')$

5.4 Diagram

definition
$LinkMark\text{-}rel = (\lambda\ (i,\ j)\ .\ (case\ (i,\ j)\ of$

$(I.init,\ I.loop) \ \Rightarrow\ Q1'' \sqcup Q2''\ |$
$(I.loop,\ I.loop) \ \Rightarrow\ Q3'' \sqcup (Q4'' \sqcup Q5'')\ |$
$(I.loop,\ I.final) \Rightarrow\ Q6''\ |$
$\text{-} \Rightarrow \bot))$

definition [*simp*]:
LinkMarkInv i = (*case* i *of*
$I.init \ \Rightarrow\ Init''\ |$
$I.loop \ \Rightarrow\ Loop''\ |$
$I.final \Rightarrow\ Final'')$

5.5 Data refinement of the transitions

theorem *init1* [*simp*]:
DataRefinement $Init'\ Q1'\ R1'\ R2'$ (*demonic* $Q1''$)
⟨*proof*⟩

theorem *init2* [*simp*]:
DataRefinement $Init'\ Q2'\ R1'\ R2'$ (*demonic* $Q2''$)
⟨*proof*⟩

theorem *step1* [*simp*]:
DataRefinement $Loop'\ Q3'\ R2'\ R2'$ (*demonic* $Q3''$)
⟨*proof*⟩

lemma *neqif* [*simp*]: $x \neq y \Longrightarrow (\textit{if } y = x \textit{ then } a \textit{ else } b) = b$
⟨*proof*⟩

theorem *step2* [*simp*]:
DataRefinement $Loop'\ Q4'\ R2'\ R2'$ (*demonic* $Q4''$)
⟨*proof*⟩

lemma *setsimp*: $a = c \Longrightarrow (x \in a) = (x \in c)$
⟨*proof*⟩

theorem *step3* [*simp*]:
DataRefinement $Loop'\ Q4'\ R2'\ R2'$ (*demonic* $Q5''$)
⟨*proof*⟩

theorem *final* [*simp*]:
DataRefinement $Loop'\ Q5'\ R2'\ R1'$ (*demonic* $Q6''$)
⟨*proof*⟩

5.6 Diagram data refinement

theorem *LinkMark-DataRefinement* [*simp*]:
DgrDataRefinement (*dangelic R SetMarkInv*) *StackMark-rel* R' (*dgr-demonic LinkMark-rel*)
⟨*proof*⟩

5.7 Diagram correctness

theorem *LinkMark-correct*:
  Hoare-dgr (*dangelic R′* (*dangelic R SetMarkInv*)) (*dgr-demonic LinkMark-rel*)
((*dangelic R′* (*dangelic R SetMarkInv*)) ⊓ (− *grd* (*step* ((*dgr-demonic LinkMark-rel*)))))
  ⟨*proof*⟩

end

end

6 Deutsch-Schorr-Waite Marking Algorithm

theory *DSWMark*

imports *LinkMark*

begin

Finally, we construct the Deutsch-Schorr-Waite marking algorithm by assuming that there are only two pointers (*left*, *right*) from every node. There is also a new variable, *atom* : *node* → *bool* which associates to every node a Boolean value. The data invariant of this refinement step requires that *index* has exactly two distinct elements *none* and *some*, *left* = *link none*, *right* = *link some*, and *atom x* is true if and only if *label x* = *some*.

We use a new locale which fixes the iniatial values of the variables *left*, *right*, and *atom* in *left0*, *right0*, and *atom0* respectively.

datatype *Index* = *none* | *some*

locale *classical* = *node* +
  fixes *left0* :: ′*node* ⇒ ′*node*
  fixes *right0* :: ′*node* ⇒ ′*node*
  fixes *atom0* :: ′*node* ⇒ *bool*
  assumes (*nil*::′*node*) = *nil*
  begin
    definition
      link0 i = (*if i* = (*none*::*Index*) *then left0 else right0*)

    definition
      label0 x = (*if atom0 x then* (*some*::*Index*) *else none*)
  end

sublocale *classical* ⊆ *pointer nil root none*::*Index link0 label0*
⟨*proof*⟩

context *classical*
begin

lemma [*simp*]:
$(label0 = (\lambda\ x\ .\ if\ atom\ x\ then\ some\ else\ none)) = (atom0 = atom)$
$\langle proof \rangle$

definition
$gg\ mrk\ atom\ ptr\ x \equiv ptr\ x \neq nil \wedge ptr\ x \notin mrk \wedge \neg\ atom\ x$

6.1 Transitions

definition
$QQ1 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$root = nil \wedge\ p' = nil \wedge t' = nil \wedge mrk' = mrk \wedge left' = left \wedge right' = right \wedge atom' = atom\}$

definition
$QQ2 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$root \neq nil \wedge\ p' = root \wedge t' = nil \wedge mrk' = mrk \cup \{root\} \wedge left' = left \wedge right' = right \wedge atom' = atom\}$

definition
$QQ3 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$p \neq nil \wedge gg\ mrk\ atom\ left\ p \wedge$
$p' = left\ p \wedge t' = p \wedge mrk' = mrk \cup \{left\ p\} \wedge$
$left' = left(p := t) \wedge right' = right \wedge atom' = atom\}$

definition
$QQ4 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$p \neq nil \wedge gg\ mrk\ atom\ right\ p \wedge$
$p' = right\ p \wedge t' = p \wedge mrk' = mrk \cup \{right\ p\} \wedge$
$left' = left \wedge right' = right(p := t) \wedge atom' = atom(p := True)\}$

definition
$QQ5 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$p \neq nil \wedge (*not\ needed\ in\ the\ proof*)$
$\neg\ gg\ mrk\ atom\ left\ p \wedge \neg\ gg\ mrk\ atom\ right\ p \wedge$
$t \neq nil \wedge \neg\ atom\ t \wedge$
$p' = t \wedge t' = left\ t \wedge mrk' = mrk \wedge$
$left' = left(t := p) \wedge right' = right \wedge atom' = atom\}$

definition
$QQ6 \equiv \lambda\ (p, t, left, right, atom, mrk)\ .\ \{(p', t', left', right', atom', mrk')\ .$
$p \neq nil \wedge (*not\ needed\ in\ the\ proof*)$
$\neg\ gg\ mrk\ atom\ left\ p \wedge \neg\ gg\ mrk\ atom\ right\ p \wedge$
$t \neq nil \wedge atom\ t \wedge$
$p' = t \wedge t' = right\ t \wedge mrk' = mrk \wedge$
$left' = left \wedge right' = right(t := p) \wedge atom' = atom(t := False)\}$

definition
$QQ7 \equiv \lambda\ (p,\ t,\ left,\ right,\ atom,\ mrk)\ .\ \{(p',\ t',\ left',\ right',\ atom',\ mrk')\ .$
$p \neq nil\ \wedge$
$\neg\ gg\ mrk\ atom\ left\ p\ \wedge\ \neg\ gg\ mrk\ atom\ right\ p\ \wedge$
$t = nil\ \wedge$
$p' = nil\ \wedge\ t' = t\ \wedge\ mrk' = mrk\ \wedge$
$left' = left\ \wedge\ right' = right\ \wedge\ atom' = atom\}$

definition
$QQ8 \equiv \lambda\ (p,\ t,\ left,\ right,\ atom,\ mrk)\ .\ \{(p',\ t',\ left',\ right',\ atom',\ mrk')\ .$
$p = nil\ \wedge\ p' = p\ \wedge\ t' = t\ \wedge\ mrk' = mrk\ \wedge\ left' = left\ \wedge\ right' = right\ \wedge$
$atom' = atom\}$

7 Data refinement relation

definition
$RR \equiv \lambda\ (p,\ t,\ left,\ right,\ atom,\ mrk)\ .\ \{(p',\ t',\ lnk,\ lbl,\ mrk')\ .$
$lnk\ none = left\ \wedge\ lnk\ some = right\ \wedge$
$lbl = (\lambda\ x\ .\ if\ atom\ x\ then\ some\ else\ none)\ \wedge$
$p' = p\ \wedge\ t' = t\ \wedge\ mrk' = mrk\}$

definition [*simp*]:
$R''\ i = RR$

definition
$ClassicMark\text{-}rel = (\lambda\ (i,\ j)\ .\ (case\ (i,\ j)\ of$
$(I.init,\ I.loop)\ \Rightarrow\ QQ1 \sqcup QQ2\ |$
$(I.loop,\ I.loop)\ \Rightarrow\ (QQ3 \sqcup QQ4) \sqcup ((QQ5 \sqcup QQ6) \sqcup QQ7)\ |$
$(I.loop,\ I.final) \Rightarrow\ QQ8\ |$
$\text{-} \Rightarrow \bot))$

7.1 Data refinement of the transitions

theorem *init1* [*simp*]:
$DataRefinement\ Init''\ Q1''\ RR\ RR\ (demonic\ QQ1)$
⟨*proof*⟩

theorem *init2* [*simp*]:
$DataRefinement\ Init''\ Q2''\ RR\ RR\ (demonic\ QQ2)$
⟨*proof*⟩

lemma *index-simp*:
$(u = v) = (u\ none = v\ none\ \wedge\ u\ some = v\ some)$
⟨*proof*⟩

theorem *step1* [*simp*]:
$DataRefinement\ Loop''\ Q3''\ RR\ RR\ (demonic\ QQ3)$

⟨*proof*⟩

theorem *step2* [*simp*]:
DataRefinement Loop'' Q3'' RR RR (*demonic QQ4*)
⟨*proof*⟩

theorem *step3* [*simp*]:
DataRefinement Loop'' Q4'' RR RR (*demonic QQ5*)
⟨*proof*⟩

lemma *if-set-elim*: $(x \in (\textit{if } b \textit{ then } A \textit{ else } B)) = ((b \wedge x \in A) \vee (\neg\, b \wedge x \in B))$
⟨*proof*⟩

theorem *step4* [*simp*]:
DataRefinement Loop'' Q4'' RR RR (*demonic QQ6*)
⟨*proof*⟩

theorem *step5* [*simp*]:
DataRefinement Loop'' Q5'' RR RR (*demonic QQ7*)
⟨*proof*⟩

theorem *final-step* [*simp*]:
DataRefinement Loop'' Q6'' RR RR (*demonic QQ8*)
⟨*proof*⟩

7.2 Diagram data refinement

theorem *ClassicMark-DataRefinement* [*simp*]:
DgrDataRefinement (*dangelic R'* (*dangelic R SetMarkInv*)) *LinkMark-rel R''* (*dgr-demonic ClassicMark-rel*)
⟨*proof*⟩

7.3 Diagram corectness

theorem *ClassicMark-correct* [*simp*]:
Hoare-dgr (*dangelic R''* (*dangelic R'* (*dangelic R SetMarkInv*))) (*dgr-demonic ClassicMark-rel*)
((*dangelic R''* (*dangelic R'* (*dangelic R SetMarkInv*))) ⊓ (− *grd* (*step* ((*dgr-demonic ClassicMark-rel*)))))
⟨*proof*⟩

We have proved the correctness of the final algorithm, but the pre and the post conditions involve the angelic choice operator and they depend on all data refinement steps we have used to prove the final diagram. We simplify these conditions and we show that we obtained indead the corectness of the marking algorithm.

The predicate *ClassicInit* which is true for the *init* situation states that initially the variables *left*, *right*, and *atom* are equal to their initial values and also that no node is marked.

The predicate *ClassicFinal* which is true for the *final* situation states that at the end the values of the variables *left*, *right*, and *atom* are again equal to their initial values and the variable *mrk* records all reachable nodes. The reachable nodes are defined using our initial *next* relation, however if we unfold all locale interpretations and definitions we see easeily that a node x is reachable if there is a path from *root* to x along *left* and *right* functions, and all nodes in this path have the atom bit false.

definition
ClassicInit = {(*p*, *t*, *left*, *right*, *atom*, *mrk*) .
atom = *atom0* ∧ *left* = *left0* ∧ *right* = *right0* ∧
finite (− *mrk*) ∧ *mrk* = {}}

definition
ClassicFinal = {(*p*, *t*, *left*, *right*, *atom*, *mrk*) .
atom = *atom0* ∧ *left* = *left0* ∧ *right* = *right0* ∧
mrk = *reach root*}

theorem [*simp*]:
ClassicInit ⊆ (*angelic RR* (*angelic R1′* (*angelic R1* (*SetMarkInv init*))))
⟨*proof*⟩

theorem [*simp*]:
ClassicInit ⊆ (*angelic* (*R″ init*) (*angelic* (*R′ init*) (*angelic* (*R init*) (*SetMarkInv init*))))
⟨*proof*⟩

theorem [*simp*]:
(*angelic RR* (*angelic R1′* (*angelic R1* (*SetMarkInv final*)))) ≤ *ClassicFinal*
⟨*proof*⟩

theorem [*simp*]:
(*angelic* (*R″ final*) (*angelic* (*R′ final*) (*angelic* (*R final*) (*SetMarkInv final*)))) ≤ *ClassicFinal*
⟨*proof*⟩

The indexed predicate *ClassicPre* is the precondition of the diagram, and since we are only interested in starting the marking diagram in the *init* situation we set *ClassicPre loop* = *ClassicPre final* = ∅.

definition [*simp*]:
ClassicPre i = (*case i of*
I.init ⇒ *ClassicInit* |
I.loop ⇒ {} |
I.final ⇒ {})

We are interested on the other hand that the marking diagram terminates only in the *final* situation. In order to achieve this we define the postcondition of the diagram as the indexed predicate *ClassicPost* which is empty on every situation except $final$.

definition [*simp*]:
ClassicPost i = (*case i of*
I.init ⇒ {} |
I.loop ⇒ {} |
I.final ⇒ *ClassicFinal*)

definition [*simp*]:
ClassicMark = *dgr-demonic ClassicMark-rel*

lemma *exists-or*:
$(\exists\ x\ .\ p\ x \lor q\ x) = ((\exists\ x\ .\ p\ x) \lor (\exists\ x\ .\ q\ x))$
⟨*proof*⟩

lemma [*simp*]:
(− *grd* (*step* (*dgr-demonic ClassicMark-rel*))) *init* = {}
⟨*proof*⟩

lemma [*simp*]:
(− *grd* (*step* (*dgr-demonic ClassicMark-rel*))) *loop* = {}
⟨*proof*⟩

The final theorem states the correctness of the marking diagram with respect to the precondition *ClassicPre* and the postcondition *ClassicPost*, that is, if the diagram starts in the initial situation, then it will terminate in the final situation, and it will mark all reachable nodes.

theorem ⊨ *ClassicPre* {| *pt ClassicMark* |} *ClassicPost*
⟨*proof*⟩
end

end

References

[1] J.-R. Abrial. Event based sequential program development: Application to constructing a pointer program. In K. Araki, S. Gnesi, and D. Mandrioli, editors, *FME*, volume 2805 of *Lecture Notes in Computer Science*, pages 51–74. Springer, 2003.

[2] R. J. Back. *Correctness preserving program refinements: proof theory and applications*, volume 131 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1980.

[3] R.-J. Back. Semantic correctness of invariant based programs. In *International Workshop on Program Construction*, Chateau de Bonas, France, 1980.

[4] R.-J. Back. Invariant based programs and their correctness. In W. Biermann, G. Guiho, and Y. Kodratoff, editors, *Automatic Program Construction Techniques*, pages 223–242. MacMillan Publishing Company, 1983.

[5] R.-J. Back. Invariant based programming: Basic approach and teaching experience. *Formal Aspects of Computing*, 2008.

[6] R.-J. Back and V. Preoteasa. Semantics and proof rules of invariant based programs. Technical Report 903, TUCS, Jul 2008.

[7] R. J. Back and J. von Wright. Encoding, decoding and data refinement. *Formal Aspects of Computing*, 12:313–349, 2000.

[8] W. DeRoever and K. Engelhardt. *Data Refinement: Model-Oriented Proof Methods and Their Comparison.* Cambridge University Press, New York, NY, USA, 1999.

[9] C. A. R. Hoare. Proof of correctness of data representations. *Acta Informatica*, 1(4), Dec. 1972.

[10] D. E. Knuth. *The art of computer programming, volume 1 (3rd ed.): fundamental algorithms.* Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1997.

[11] F. Mehta and T. Nipkow. Proving pointer programs in higher-order logic. *Information and Computation*, 199:200–227, 2005.

[12] V. Preoteasa and R.-J. Back. Data refinement of invariant based programs. *Electronic Notes in Theoretical Computer Science*, 259:143 – 163, 2009. Proceedings of the 14th BCS-FACS Refinement Workshop (REFINE 2009).

[13] V. Preoteasa and R.-J. Back. Semantics and data refinement of invariant based programs. In G. Klein, T. Nipkow, and L. Paulson, editors, *The Archive of Formal Proofs*. http://afp.sf.net/entries/DataRefinementIBP.shtml, May 2010. Formal proof development. Submitted.

[14] H. Schorr and W. M. Waite. An efficient machine-independent procedure for garbage collection in various list structures. *Commun. ACM*, 10(8):501–506, 1967.