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

Viorel Preoteasa and Ralph-Johan Back

September 19, 2015

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  2

2 Address Graph  3

3 Marking Using a Set  5
   3.1 Transitions  6
   3.2 Invariants  6
   3.3 Diagram  7
   3.4 Correctness of the transitions  8
   3.5 Diagram correctness  8

4 Marking Using a Stack  9
   4.1 Transitions  9
   4.2 Invariants  10
   4.3 Data refinement relations  10
   4.4 Data refinement of the transitions  11
   4.5 Diagram data refinement  11
   4.6 Diagram correctness  11
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 \cite{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 \cite{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 \((x.\text{left} := y)\) are done by modifying the global pointer function \(\text{left} := \text{left}(x := y)\). 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 polymorphism is used for the state of the computation. In \cite{13} the state of computation was introduced as a type variable, or even more generally, 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 \textit{next} on nodes. This locale is first partially interpreted into a locale which replaces \textit{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, \textit{left} and \textit{right}.

\section{Address Graph}

\begin{verbatim}
theory Graph imports Main begin

This theory introduces the graph to be marked as a relation \textit{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 \textit{next} to any node and any node is not related by \textit{next} to nil.

locale node =

3
\end{verbatim}
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 returns 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**

\[
\text{head } S \equiv (if \ S = [] then \text{nil else } (hd \ S))
\]

**definition**

\[
\text{tail } (S::'a \text{ list}) \equiv (if \ S = [] then [] else (tl \ S))
\]

**lemma [simp]:** \((\text{nil}, x) \in \text{next}) = False\n\langle proof \rangle

**lemma [simp]:** \((x, \text{nil}) \in \text{next}) = False\n\langle proof \rangle

**theorem head-not-nil [simp]:**
\((\text{head } S \neq \text{nil}) = (\text{head } S = \text{hd } S \land \text{tail } S = \text{tl } S \land \text{hd } S \neq \text{nil} \land S \neq [])\n\langle proof \rangle

**theorem nonempty-head [simp]:**
\(\text{head } (x \neq S) = x\)
\langle proof \rangle

**theorem nonempty-tail [simp]:**
\(\text{tail } (x \neq S) = S\)
\langle proof \rangle

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

**theorem (in graph) reach-nil [simp]:** \(\text{reach } \text{nil} = \{\}\)
\langle proof \rangle

**theorem (in graph) reach-next: b \in \text{reach } a \implies (b, c) \in \text{next} \implies c \in \text{reach } a\n\langle proof \rangle

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

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 \text{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 successor of \(x\) we add \(y\) to \(X\). If \(x \in X\) has no unmarked successors it is removed from \(X\). The algorithm terminates when \(X\) is empty.

\textbf{datatype} \(I = \text{init} \mid \text{loop} \mid \text{final}\)

\textbf{declare} \(I\).split [split]

\textbf{instantiation} \(I :: \text{well-founded-transitive}\)

\textbf{begin}

\textbf{definition}
\(\text{less-I-def}: i < j \equiv (j = \text{init} \land (i = \text{loop} \lor i = \text{final})) \lor (j = \text{loop} \land i = \text{final})\)

\textbf{definition}
\(\text{less-eq-I-def}: (i::I) \leq (j::I) \equiv i = j \lor i < j\)

\textbf{instance}
\(\langle \text{proof} \rangle\)

\textbf{end}

The set \(\text{path } S \text{ mrk}\) contains all reachable nodes from \(S\) along paths with unmarked nodes.

\textbf{lemma} \(\text{trascl-less}: x \neq y \Rightarrow (a, x) \in R^* \Rightarrow \)
\[(a, x) \in (R \cap (-\{y\}) \times (-\{y\}))^* \lor (y, x) \in R \cap (R \cap (-\{y\}) \times (-\{y\}))^* \]

**lemma (in graph)** add-set [simp]: \(x \neq y \implies x \in \text{path } S \text{ mrk} \implies x \in \text{path } (\text{insert } y S)\) (insert y mrk)

**lemma (in graph)** add-set2: \(x \in \text{path } S \text{ mrk} \implies x \notin \text{path } (\text{insert } y S)\) (insert y mrk) \implies x = y

**lemma (in graph)** del-stack [simp]: \((\forall y . (t, y) \in \text{next} \implies y \in \text{mrk}) \implies x \notin \text{mrk} \implies x \in \text{path } S \text{ mrk} \implies x \in \text{path } (S - \{t\}) \text{ mrk} \)

**lemma (in graph)** init-set [simp]: \(x \in \text{reach root} \implies x \neq \text{root} \implies x \in \text{path } \{\text{root}\}\{\text{root}\} \implies x = \text{root}\)

**lemma (in graph)** init-set2: \(x \in \text{reach root} \implies x \notin \text{path } \{\text{root}\}\{\text{root}\} \implies x = \text{root}\)

### 3.1 Transitions

**definition (in graph)**
\[Q1-a \equiv [: X, mrk \rightsquigarrow X', mrk', (\text{root}::'node) = \text{nil} \land X' = \{\} \land mrk' = mrk :]

**definition (in graph)**
\[Q2-a \equiv [: X, mrk \rightsquigarrow X', mrk', \text{root}::'node) \neq \text{nil} \land X' = \{\text{root}::'node\} \land mrk' = \{\text{root}::'node\} :]

**definition (in graph)**
\[Q3-a \equiv [: X, mrk \rightsquigarrow X', mrk', \exists x \in X . \exists y . (x, y) \in \text{next} \land y \notin \text{mrk} \land X' = X \cup \{y\} \land mrk' = mrk \cup \{y\})]

**definition (in graph)**
\[Q4-a \equiv [: X, mrk \rightsquigarrow X', mrk', \exists x \in X . (\forall y . (x, y) \in \text{next} \implies y \in \text{mrk}) \land X' = X - \{x\} \land mrk' = mrk ];]

**definition (in graph)**
\[Q5-a \equiv [: X, mrk \rightsquigarrow X', mrk', X = \{\} \land mrk = mrk ' :]

### 3.2 Invariants

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

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

definition
\[ term-eq \ t \ w = \{ s \ . \ t \ s = w \} \]

definition
\[ term-less \ t \ w = \{ s \ . \ t \ s < w \} \]

lemma union-term-eq [simp]: \( \bigcup \ w . \ term-eq \ t \ w = UNIV \)
(proof)

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

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

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

definition (in graph)
\[ SetMarkInv \ i = (case \ i \ of \)
\[ \quad \quad \quad \quad \ I . \ init \ \Rightarrow \ Init \mid \]
\[ \quad \quad \quad \quad \ I . \ loop \ \Rightarrow \ Loop \mid \]
\[ \quad \quad \quad \quad \ I . \ final \ \Rightarrow \ Final \) \]

definition (in graph)
\[ SetMarkInvFinal \ i = (case \ i \ of \)
\[ \quad \quad \quad \quad \ I . \ final \ \Rightarrow \ Final \mid \]
\[ \quad \quad \quad \quad \ \text{-} \ \Rightarrow \{\} \)

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

3.3 Diagram

definition (in graph)
\[ SetMark \equiv \lambda \ (i, j) . \ (case \ (i, j) \ of \)
\[ \quad \quad \quad \quad \ (I . \ init, \ I . \ loop) \ \Rightarrow \ Q1-a \cap Q2-a \mid \]
\[ \quad \quad \quad \quad \ (I . \ loop, \ I . \ loop) \ \Rightarrow \ Q3-a \cap Q4-a \mid \]
\[ \quad \quad \quad \quad \ (I . \ loop, \ I . \ final) \ \Rightarrow \ Q5-a \mid \]
\[ \quad \quad \quad \quad \ \text{-} \ \Rightarrow \ \text{top} \) \]
lemma (in graph) SetMark-dmono [simp]:
dmono SetMark
⟨proof⟩

3.4 Correctness of the transitions

lemma (in graph) init-loop-1-a [simp]: Init \{ Q1-a \} Loop
⟨proof⟩

lemma (in graph) init-loop-2-a [simp]: Init \{ Q2-a \} Loop
⟨proof⟩

lemma (in graph) loop-loop-1-a [simp]: (Loop \cap \{ s . \operator{trm}(s) = w \}) \{ Q3-a \}
⟨proof⟩

lemma (in graph) loop-loop-2-a [simp]: (Loop \cap \{ s . \operator{trm}(s) = w \}) \{ Q4-a \}
⟨proof⟩

lemma (in graph) loop-final-a [simp]: (Loop \cap \{ s . \operator{trm}(s) = w \}) \{ Q5-a \}
Final
⟨proof⟩

lemma union-term-w [simp]: (∪ w . \{ s . t(s) = w \}) = \operator{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 (range A) i = (∪ w . A w i)
⟨proof⟩

lemma forall-simp [simp]: (! a b . ∀ x ∈ A . (a = (t x)) → (h x) ∨ b ≠ u x) = (∀ x ∈ A . h x)
⟨proof⟩

lemma forall-simp2 [simp]: (! a b . ∀ x ∈ A . ! y . (a = t x y) → (h x y) → (g x y)) ∨ b ≠ u x y = (∀ x ∈ A . ! y . h x y → 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 ∈ I and n ∈ nat.
interpretation DiagramTermination λ (n::nat) (i :: I) . (i, n)
⟨proof⟩

theorem (in graph) SetMark-correct:
\[\vdash \text{SetMarkInv} \{\text{in SetMark}\} \text{SetMarkInvFinal} \]

\textbf{Theorem (in graph) SetMark-correct1 [simp]:}

\[\text{Hoare-dgr SetMarkInv SetMark } (\text{SetMarkInv } \cap (\neg \text{ grd } \text{ step SetMark})) \]

\textbf{Theorem (in graph) stack-not-nil [simp]:}

\[(\text{mrk}, S) \in \text{Loop} \Rightarrow x \in S \Rightarrow x \neq \text{nil} \]

\textbf{end}

\section{Marking Using a Stack}

\textbf{theory StackMark}

\textbf{imports SetMark ../DataRefinementIBP/DataRefinement}

\textbf{begin}

In this theory we refine the set marking diagram to a diagram in which the set is replaced by a list (stack). Initially 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 successors, 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.

\subsection{Transitions}

\textbf{definition (in graph)}

\[Q1'\alpha \equiv [\lambda (\text{stk}::\text{('node list)}, \text{mrk}::\text{('node set)}) . ((\text{stk}'::\text{('node list)}, \text{mrk}') . \text{root} = \text{nil} \wedge \text{stk}' = [] \wedge \text{mrk}' = \text{mrk})]]\]

\textbf{definition (in graph)}

\[Q2'\alpha \equiv [\lambda (\text{stk}::\text{('node list)}, \text{mrk}::\text{('node set)}) . ((\text{stk}', \text{mrk}') . \text{root} \neq \text{nil} \wedge \text{stk}' = [\text{root}] \wedge \text{mrk}' = \text{mrk} \cup \{\text{root}\})]]\]

\textbf{definition (in graph)}

\[Q3'\alpha \equiv [\lambda (\text{stk}, \text{mrk}) . ((\text{stk}', \text{mrk}') . \text{stk} \neq [] \wedge (\exists y . (\text{hd stk}, y) \in \text{next} \wedge y \notin \text{mrk} \wedge \text{stk}' = y \# \text{stk} \wedge \text{mrk}' = \text{mrk} \cup \{y\})]]\]

\textbf{definition (in graph)}

\[Q4'\alpha \equiv [\lambda (\text{stk}, \text{mrk}) . ((\text{stk}', \text{mrk}') . \text{stk} \neq [] \wedge (\forall y . (\text{hd stk}, y) \in \text{next} \Rightarrow y \in \text{mrk}) \wedge \text{stk}' = \text{tl stk} \wedge \text{mrk}' = \text{mrk} \cup \{y\})]]\]

\textbf{definition}

\[Q5'\alpha \equiv [\lambda (\text{stk}, \text{mrk}) . ((\text{stk}', \text{mrk}') . \text{stk} = [] \wedge \text{mrk}' = \text{mrk})] \]

9
4.2 Invariants

**definition**

\[ \text{Init}' \equiv \text{UNIV} \]

**definition**

\[ \text{Loop}' \equiv \{ (\text{stk}, \text{mrk}) . \text{distinct stk} \} \]

**definition**

\[ \text{Final}' \equiv \text{UNIV} \]

**definition [simp]:**

\[ \text{StackMarkInv } i = (\text{case } i \text{ of } \]
\[ \text{I.init } \Rightarrow \text{Init}' \]
\[ \text{I.loop } \Rightarrow \text{Loop}' \]
\[ \text{I.final } \Rightarrow \text{Final}') \]

4.3 Data refinement relations

**definition**

\[ \text{R1-a} \equiv \{ : \text{stk}, \text{mrk} \mapsto X, \text{mrk}' . \text{mrk}' = \text{mrk} :\} \]

**definition**

\[ \text{R2-a} \equiv \{ : \text{stk}, \text{mrk} \mapsto X, \text{mrk}' . X = \text{set stk} \land (\text{stk}, \text{mrk}) \in \text{Loop}' \land \text{mrk}' = \text{mrk} :\} \]

**lemma [simp]:** \( \text{R1-a} \in \text{Apply.Disjunctive} \)

**lemma [simp]:** \( \text{R2-a} \in \text{Apply.Disjunctive} \)

**definition [simp]:**

\[ \text{R-a } i = (\text{case } i \text{ of } \]
\[ \text{I.init } \Rightarrow \text{R1-a } | \]
\[ \text{I.loop } \Rightarrow \text{R2-a } | \]
\[ \text{I.final } \Rightarrow \text{R1-a} ) \]

**lemma [simp]:** \( \text{Disjunctive-fun R-a} \)

**definition**

\[ \text{angelic-fun } r = (\lambda i . \{ : r i \}) \]

**definition (in graph)**

\[ \text{StackMark-a } = (\lambda (i,j) . \text{case } (i,j) \text{ of } \]
\[ (\text{I.init}, \text{I.loop}) \Rightarrow Q1'-a \cap Q2'-a | \]
\[ (\text{I.loop}, \text{I.loop}) \Rightarrow Q3'-a \cap Q4'-a | \]
\[ (\text{I.loop}, \text{I.final}) \Rightarrow Q5'-a | \]
\[ - \Rightarrow T) ) \]
4.4 Data refinement of the transitions

**Theorem (in graph) init-nil [simp]:**
\[\text{DataRefinement} (\{\text{Init}\} \circ Q1-a) R1-a R2-a Q1'-a\]
\[\langle \text{proof} \rangle\]

**Theorem (in graph) init-root [simp]:**
\[\text{DataRefinement} (\{\text{Init}\} \circ Q2-a) R1-a R2-a Q2'-a\]
\[\langle \text{proof} \rangle\]

**Theorem (in graph) step1 [simp]:**
\[\text{DataRefinement} (\{\text{Loop}\} \circ Q3-a) R2-a R2-a Q3'-a\]
\[\langle \text{proof} \rangle\]

**Theorem (in graph) step2 [simp]:**
\[\text{DataRefinement} (\{\text{Loop}\} \circ Q4-a) R2-a R2-a Q4'-a\]
\[\langle \text{proof} \rangle\]

**Theorem (in graph) final [simp]:**
\[\text{DataRefinement} (\{\text{Loop}\} \circ Q5-a) R2-a R1-a Q5'-a\]
\[\langle \text{proof} \rangle\]

4.5 Diagram data refinement

**Lemma assert-comp-choice:**
\[\{.p.\} \circ (S \sqcap T) = (\{.p.\} \circ S) \sqcap (\{.p.\} \circ T)\]
\[\langle \text{proof} \rangle\]

**Theorem (in graph) StackMark-DataRefinement [simp]:**
\[\text{DgrDataRefinement2 SetMarkInv SetMark R-a StackMark-a}\]
\[\langle \text{proof} \rangle\]

4.6 Diagram correctness

**Theorem (in graph) StackMark-correct:**
\[\text{Hoare-dgr} (R-a \ldots \text{SetMarkInv}) \text{StackMark-a} ((R-a \ldots \text{SetMarkInv}) \sqcap (\neg \text{grd} \ (\text{step} (\text{StackMark-a}))))\]
\[\langle \text{proof} \rangle\]

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 \( label : \text{node} \rightarrow \text{index} \) associates a label to every node and the variable \( link : \text{index} \rightarrow \text{node} \rightarrow \text{node} \) is a collection of pointer functions indexed by the set \( \text{index} \) of labels. For \( x \in \text{node} \), \( link i x \) 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 \( \text{nil} \) and they are labeled by a special label \( \text{none} \in \text{index} \).

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 \Rightarrow 'node \Rightarrow 'node
  fixes label0 :: 'node \Rightarrow 'index

  assumes (nil::'node) = nil
begin
  definition next = \{ (a, b) . (\\exists i . link0 i a = b) \& a \neq \text{nil} \& b \neq \text{nil} \& label0 a = \text{none} \}
end
```

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

The next two recursive functions, \( \text{label}_0 \), \( \text{link}_0 \) are used to compute the initial values of the variables \( label \) and \( link \) from their current values.

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

lemma label-cong [cong]: \( f = g \Rightarrow xs = ys \Rightarrow \text{pointer.label}_0 n f xs = \text{pointer.label}_0 n g ys \)
(proof)
```
The function *stack* defined below 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 ∈ set l ∧ stack lnk lbl (lnk (lbl x) x) l)

**lemma** label-out-range0 [simp]:

¬ x ∈ set S ⇒ label-0 lbl S x = lbl x

(proof)

**lemma** link-out-range0 [simp]:

¬ x ∈ set S ⇒ link-0 link label p S i x = link i x

(proof)

**lemma** link-out-range [simp]: ¬ x ∈ set S ⇒ 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]: ¬ p ∈ set S ⇒ stack (link(i := (link i)(p := q))) label x S = stack link label x S

(proof)

**lemma** stack-out-label-range [simp]: ¬ p ∈ set S ⇒ stack link (label(p := q)) x S = stack link label x S

(proof)

**definition**

g mrk lbl ptr x ≡ ptr x ≠ nil ∧ ptr x ∉ mrk ∧ lbl x = none

**lemma** g-cong [cong]: mrk = mrk' ⇒ lbl = lbl' ⇒ ptr = ptr' ⇒ x = x'

(proof)
5.1 Transitions

**Definition Q1.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ \text{root} = \text{nil} \land p' = \text{nil} \land t' = \text{nil} \land \text{link}' = \text{link} \land \text{label}' = \text{label} \land \text{mark}' = \text{mark}. \]

**Definition Q2.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ \text{root} \neq \text{nil} \land p' = \text{root} \land t' = \text{nil} \land \text{link}' = \text{link} \land \text{label}' = \text{label} \land \text{mark}' = \text{mark} \cup \{\text{root}\}. \]

**Definition Q3.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ p \neq \text{nil} \land \exists i. \neg g \text{mark} \text{label}(\text{link} i) p \land p' = \text{link} i p \land t' = \text{link}(i := (\text{link} i)(p := t)) \land \text{label}' = \text{label}(p := i) \land \text{mark}' = \text{mark} \cup \{\text{link} i p\}. \]

**Definition Q4.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ p \neq \text{nil} \land \forall i. \neg g \text{mark} \text{label}(\text{link} i) p \land t = \text{nil} \land p' = \text{nil} \land t' = \text{link} \text{label}(t := \text{none}) \land \text{mark}' = \text{mark}. \]

**Definition Q5.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ p = \text{nil} \land p' = p \land t' = t \land \text{link}' = \text{link} \land \text{label}' = \text{label} \land \text{mark}' = \text{mark}. \]

**Definition Q6.** \[ p, t', \text{link}', \text{label}', \text{mark}' \]

\[ p = p \land t' = t \land \text{link}' = \text{link} \land \text{label}' = \text{label} \land \text{mark}' = \text{mark}. \]

5.2 Invariants

**Definition Init.** \( \mathcal{I} = \{ (p, t, \text{link}, \text{label}, \text{mark}) \mid \text{link} = \text{link}_0 \land \text{label} = \text{label}_0 \} \)

**Definition Loop.** \( \mathcal{L} = \text{UNIV} \)

**Definition Final.** \( \mathcal{F} = \mathcal{I} \)

5.3 Data refinement relations
\[
R'1-a \equiv \{ \colon p, t, \text{lnk}, \text{lbl}, \text{mrk} \rightharpoonup \text{stk}, \text{mrk}' . (p, t, \text{lnk}, \text{lbl}, \text{mrk}) \in \text{Init}'' \land \text{mrk}' = \text{mrk} : \}
\]

**definition**

\[
R'2-a \equiv \{ \colon p, t, \text{lnk}, \text{lbl}, \text{mrk} \rightharpoonup \text{stk}, \text{mrk}' .
\]

\[
p = \text{head} \text{stk} \land
\]

\[
t = \text{head} (\text{tail} \text{stk}) \land
\]

\[
\text{stack} \text{lnk} \text{lbl} t (\text{tail} \text{stk}) \land
\]

\[
\text{link}0 = \text{link}-0 \text{lnk} \text{lbl} (\text{tail} \text{stk}) \land
\]

\[
\text{label}0 = \text{label}-0 \text{lbl} (\text{tail} \text{stk}) \land
\]

\[
\neg \text{nil} \in \text{set} \text{stk} \land
\]

\[
\text{mrk}' = \text{mrk} : \}
\]

**lemma [simp]:** \(R'1-a \in \text{Apply} . \text{Disjunctive} \langle \text{proof} \rangle\)

**lemma [simp]:** \(R'2-a \in \text{Apply} . \text{Disjunctive} \langle \text{proof} \rangle\)

**definition [simp]:**

\[
R'\cdot-a \ i = \text{case} \ i \ \text{of}
\]

\[
1.\text{init} \Rightarrow R'1-a \ \mid
\]

\[
1.\text{loop} \Rightarrow R'2-a \ \mid
\]

\[
1.\text{final} \Rightarrow R'1-a
\]

**lemma [simp]:** \(\text{Disjunctive-fun} \ R'\cdot-a \langle \text{proof} \rangle\)

### 5.4 Diagram

**definition**

\[
\text{LinkMark} = (\lambda \ (i, j) . \text{case} \ (i, j) \ \text{of}
\]

\[
(1.\text{init}, 1.\text{loop}) \Rightarrow Q1''-a \cap Q2''-a \mid
\]

\[
(1.\text{loop}, 1.\text{loop}) \Rightarrow Q3''-a \cap (Q4''-a \cap Q5''-a) \mid
\]

\[
(1.\text{loop}, 1.\text{final}) \Rightarrow Q6''-a \mid
\]

\[
- \Rightarrow \top \})
\]

**definition [simp]:**

\[
\text{LinkMarkInv} \ i = \text{case} \ i \ \text{of}
\]

\[
1.\text{init} \Rightarrow \text{Init}'' \mid
\]

\[
1.\text{loop} \Rightarrow \text{Loop}'' \mid
\]

\[
1.\text{final} \Rightarrow \text{Final}''
\]

### 5.5 Data refinement of the transitions

**theorem init\{1\}-a [simp]:**

\[
\text{DataRefinement} \ \{(\text{init}'). \ o \ Q1'-a \} \ R1'-a \ R2'-a \ Q1''-a \langle \text{proof} \rangle
\]

**theorem init\{2\}-a [simp]:**

\[
\text{DataRefinement} \ \{(\text{init}'). \ o \ Q2'-a \} \ R1'-a \ R2'-a \ Q2''-a \langle \text{proof} \rangle
\]
**Theorem step1-a [simp]:**
\[
\text{DataRefinement } (\{.\text{Loop}'.} \circ Q3'\text{-a} ) \ R2'\text{-a} \ R2''\text{-a} \ Q3''\text{-a}
\]
⟨proof⟩

**Lemma neqif [simp]:** 
\[
x \neq y \implies (\text{if } y = x \text{ then } a \text{ else } b) = b
\]
⟨proof⟩

**Theorem step2-a [simp]:**
\[
\text{DataRefinement } (\{.\text{Loop}'.} \circ Q4'\text{-a} ) \ R2'\text{-a} \ R2''\text{-a} \ Q4''\text{-a}
\]
⟨proof⟩

**Lemma setsimp: a = c \implies (x \in a) = (x \in c) **  
⟨proof⟩

**Theorem step3-a [simp]:**
\[
\text{DataRefinement } (\{.\text{Loop}'.} \circ Q4'\text{-a} ) \ R2'\text{-a} \ R2''\text{-a} \ Q5''\text{-a}
\]
⟨proof⟩

**Theorem final-a [simp]:**
\[
\text{DataRefinement } (\{.\text{Loop}'.} \circ Q5'\text{-a} ) \ R2'\text{-a} \ R1'\text{-a} \ Q6''\text{-a}
\]
⟨proof⟩

**5.6 Diagram data refinement**

**Lemma apply-fun-index [simp]:** 
\[
(r \ldots P) i = (r i) (P i)
\]
⟨proof⟩

**Lemma [simp]: Disjunctive-fun \((r::(\forall c \Rightarrow 'a::complete-lattice \Rightarrow 'b::complete-lattice))\) \implies \text{mono-fun } r **  
⟨proof⟩

**Theorem LinkMark-DataRefinement-a [simp]:**
\[
\text{DgrDataRefinement2 } (R-a \ldots \text{SetMarkInv}) \text{ StackMark-a } R'-a \text{ LinkMark}
\]
⟨proof⟩

**Lemma [simp]: \text{mono } Q1'-a** ⟨proof⟩
**Lemma [simp]: \text{mono } Q2'-a** ⟨proof⟩
**Lemma [simp]: \text{mono } Q3'-a** ⟨proof⟩
**Lemma [simp]: \text{mono } Q4'-a** ⟨proof⟩
**Lemma [simp]: \text{mono } Q5'-a** ⟨proof⟩

**Lemma [simp]: \text{dmono } StackMark-a** ⟨proof⟩

**5.7 Diagram correctness**

**Theorem LinkMark-correct:**
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 (\(\text{left}, \text{right}\)) from every node. There is also a new variable, \(\text{atom} : \text{node} \rightarrow \text{bool}\) which associates to every node a Boolean value. The data invariant of this refinement step requires that \(\text{index}\) has exactly two distinct elements \(\text{none}\) and \(\text{some}\), \(\text{left} = \text{link}\) \(\text{none}\), \(\text{right} = \text{link}\) \(\text{some}\), and \(\text{atom}\) \(x\) is true if and only if \(\text{label}\) \(x\) = \(\text{some}\).

We use a new locale which fixes the initial values of the variables \(\text{left}, \text{right},\) and \(\text{atom}\) in \(\text{left0}, \text{right0},\) and \(\text{atom0}\) respectively.

datatype \(\text{Index} = \text{none} \mid \text{some}\)

locale classical = node +
  fixes left0 :: 'node ⇒ 'node
  fixes right0 :: 'node ⇒ 'node
  fixes atom0 :: 'node ⇒ bool
  assumes (nil::'node) = nil
begin
  definition
  \(\text{link0}\) \(i\) = (if \(i\) = (\(\text{none}::\text{Index}\)) then \(\text{left0}\) else \(\text{right0}\))

  definition
  \(\text{label0}\) \(x\) = (if \(\text{atom0}\) \(x\) then (\(\text{some}::\text{Index}\)) else \(\text{none}\))

end

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

context classical
begin

lemma [simp]:
(\(\text{label0}\) = (\(\lambda\) \(x\). if \(\text{atom}\) \(x\) then \(\text{some}\) else \(\text{none}\))) = (\(\text{atom0}\) = \(\text{atom}\))
⟨proof⟩

definition
\(\text{gg mrk}\) \(\text{atom}\) \(\text{ptr}\) \(x\) ≡ \(\text{ptr}\) \(x\) \(\neq\) nil ∧ \(\text{ptr}\) \(x\) \(\notin\) \(\text{mrk}\) ∧ \(\neg\) \(\text{atom}\) \(x\)

17
6.1 Transitions

definition
QQ1-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{root } = \text{nil} \land p' = \text{nil} \land t' = \text{nil} \land \text{mrk'} = \text{mrk} \land \text{left'} = \text{left} \\
\land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}

definition
QQ2-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{root } \neq \text{nil} \land p' = \text{root} \land t' = \text{nil} \land \text{mrk'} = \text{mrk} \cup \{\text{root}\} \\
\land \text{left'} = \text{left} \land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}

definition
QQ3-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{p } \neq \text{nil} \land \text{gg mrk atom left } p \land \\
p' = \text{left } p \land t' = p \land \text{mrk'} = \text{mrk} \cup \{\text{left } p\} \land \\
\text{left'} = \text{left}(p := t) \land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}

definition
QQ4-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{p } \neq \text{nil} \land \text{gg mrk atom right } p \land \\
p' = \text{right } p \land t' = p \land \text{mrk'} = \text{mrk} \cup \{\text{right } p\} \land \\
\text{left'} = \text{left}(p := t) \land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}

definition
QQ5-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{p } \neq \text{nil} \land \text{gg mrk atom left } p \land \\
\lnot \text{gg mrk atom right } p \land \\
\lnot \text{atom } t \land \\
p' = t \land t' = \text{left } t \land \text{mrk'} = \text{mrk} \land \\
\text{left'} = \text{left}(t := p) \land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}

definition
QQ6-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{p } \neq \text{nil} \land \text{gg mrk atom left } p \land \\
\lnot \text{gg mrk atom right } p \land \\
\lnot \text{atom } t \land \\
p' = t \land t' = \text{right } t \land \text{mrk'} = \text{mrk} \land \\
\text{left'} = \text{left} \land \text{right'} = \text{right}(t := p) \land \text{atom'} = \text{atom}(t := \text{False}). \\
\end{array}

definition
QQ7-a \equiv \begin{array}{l}
\text{p, t, left, right, atom, mrk } \rightsquigarrow \ p', t', \text{ left'}, right', \text{ atom'}, \text{ mrk'}. \\
\text{p } \neq \text{nil} \land \\
\lnot \text{gg mrk atom left } p \land \lnot \text{gg mrk atom right } p \land \\
t = \text{nil} \land \\
p' = \text{nil} \land t' = t \land \text{mrk'} = \text{mrk} \land \\
\text{left'} = \text{left} \land \text{right'} = \text{right} \land \text{atom'} = \text{atom}. \\
\end{array}
QQ8-a \equiv \{ p, t, \text{left}, \text{right}, \text{atom}, \text{mrk} \leadsto p', t', \text{left}', \text{right}', \text{atom}', \text{mrk}' \}

p = \text{nil} \land p' = p \land t' = t \land \text{mrk}' = \text{mrk} \land \text{left}' = \text{left} \land \text{right}' = \text{right} \land \text{atom}' = \text{atom} ::

7 Data refinement relation

\text{definition}

\text{RR}-a \equiv \{ p, t, \text{left}, \text{right}, \text{atom}, \text{mrk} \leadsto p', t', \text{lnk}, \text{lbl}, \text{mrk}' \}

\text{lnk none} = \text{left} \land \text{lnk some} = \text{right} \land
\text{lbl} = (\lambda x . \text{if atom} x \text{ then some else none}) \land
p' = p \land t' = t \land \text{mrk}' = \text{mrk} ::

\text{definition [simp]}

R''-a i = \text{RR}-a

\text{definition}

\text{ClassicMark} = (\lambda (i, j) . (\text{case} (i, j) \text{ of})
(I.\text{init}, I.\text{loop}) \Rightarrow \text{QQ1-a} \sqcap \text{QQ2-a} |
(I.\text{loop}, I.\text{loop}) \Rightarrow (\text{QQ3-a} \sqcap \text{QQ4-a}) \sqcap ((\text{QQ5-a} \sqcap \text{QQ6-a}) \sqcap \text{QQ7-a}) |
(I.\text{loop}, I.\text{final}) \Rightarrow \text{QQ8-a} |
- \Rightarrow \top)

7.1 Data refinement of the transitions

\text{theorem init1-a [simp]}
\text{DataRefinement} ((\text{.Init}'') \circ \text{Q1''-a}) \text{ RR-a RR-a QQ1-a}
(proof)

\text{theorem init2-a [simp]}
\text{DataRefinement} ((\text{.Init}'') \circ \text{Q2''-a}) \text{ RR-a RR-a QQ2-a}
(proof)

\text{lemma index-simp:}
(u = v) = (u \text{ none} = v \text{ none} \land u \text{ some} = v \text{ some})
(proof)

\text{theorem step1-a [simp]}
\text{DataRefinement} ((\text{.Loop''}) \circ \text{Q3''-a}) \text{ RR-a RR-a QQ3-a}
(proof)

\text{theorem step2-a [simp]}
\text{DataRefinement} ((\text{.Loop''}) \circ \text{Q4''-a}) \text{ RR-a RR-a QQ4-a}
(proof)

\text{theorem step3-a [simp]}
\text{DataRefinement} ((\text{.Loop''}) \circ \text{Q5''-a}) \text{ RR-a RR-a QQ5-a}
(proof)
lemma if-set-elim: \( (x \in (\text{if } b \text{ then } A \text{ else } B)) = (b \land x \in A) \lor (\neg b \land x \in B) \)
(proof)

theorem step4-a [simp]:
\[
\text{DataRefinement } \{\text{Loop}''\} \circ Q4''-a \text{ RR-a RR-a QQ6-a}
\]
(proof)

theorem step5-a [simp]:
\[
\text{DataRefinement } \{\text{Loop}''\} \circ Q5''-a \text{ RR-a RR-a QQ7-a}
\]
(proof)

theorem final-step-a [simp]:
\[
\text{DataRefinement } \{\text{Loop}''\} \circ Q6''-a \text{ RR-a RR-a QQ8-a}
\]
(proof)

7.2 Diagram data refinement

lemma [simp]: mono RR-a (proof)
lemma [simp]: RR-a \in Apply.Disjunctive (proof)
lemma [simp]: Disjunctive-fun R''-a (proof)

lemma [simp]: mono-fun R''-a (proof)

lemma [simp]: mono Q1''-a (proof)
lemma [simp]: mono Q2''-a (proof)
lemma [simp]: mono Q3''-a (proof)
lemma [simp]: mono Q4''-a (proof)
lemma [simp]: mono Q5''-a (proof)
lemma [simp]: mono Q6''-a (proof)

lemma [simp]: dmono LinkMark (proof)

theorem ClassicMark-DataRefinement-a [simp]:
\[
\text{DgrDataRefinement2 } (R''-a \ldots \text{(R-a \ldots \text{SetMarkInv}))} \text{ LinkMark } R''-a \text{ ClassicMark}
\]
(proof)

7.3 Diagram correctness

theorem ClassicMark-correct-a [simp]:
\[
\text{Hoare-dgr } (R''-a \ldots \text{(R-a \ldots \text{SetMarkInv))))} \text{ ClassicMark}
\]
\[
(R''-a \ldots \text{(R-a \ldots \text{SetMarkInv))))} \cap (\neg \text{ grd (step ClassicMark))})
\]
(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 indeed 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 easily 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 \land left = left0 \land right = right0 \land
finite (¬ mrk) \land mrk = \{\}\}

**definition**

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

**theorem** [simp]:

*ClassicInit* ⊆ (RR-a (R1'-a (R1-a (SetMarkInv init))))
(proof)

**theorem** [simp]:

(RR-a (R1'-a (R1-a (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*)
lemma exists-or:
(∃ x . p x ∨ q x) = ((∃ x . p x) ∨ (∃ x . q x))

lemma [simp]:
(- grd (step ClassicMark)) init = {}

lemma [simp]: grd ⊤ = ⊥

lemma [simp]: (- grd (step ClassicMark)) loop = {}

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.

lemma [simp]: mono QQ1-a (proof)
lemma [simp]: mono QQ2-a (proof)
lemma [simp]: mono QQ3-a (proof)
lemma [simp]: mono QQ4-a (proof)
lemma [simp]: mono QQ5-a (proof)
lemma [simp]: mono QQ6-a (proof)
lemma [simp]: mono QQ7-a (proof)
lemma [simp]: mono QQ8-a (proof)

lemma [simp]: dmono ClassicMark (proof)

theorem |= ClassicPre { | pt ClassicMark | } ClassicPost (proof)

end

end

References

to constructing a pointer program. In K. Araki, S. Gnesi, and
D. Mandrioli, editors, FME, volume 2805 of Lecture Notes in Computer

and applications, volume 131 of Mathematical Centre Tracts. Mathe-
matisch Centrum, Amsterdam, 1980.


