



Egyptian Mathematical Society
Journal of the Egyptian Mathematical Society

www.etms-eg.org
www.elsevier.com/locate/joems



ORIGINAL ARTICLE

Recognition of logically related regions based heap abstraction

Mohamed A. El-Zawawy

*College of Computer and Information Sciences, Al-Imam M.I.-S.I. University, Riyadh 11432, Saudi Arabia
Department of Mathematics, Faculty of Science, Cairo University, Giza 12613, Egypt*

Received 30 April 2011; revised 30 January 2012
Available online 28 September 2012

Abstract This paper presents a novel set of algorithms for heap abstraction, identifying logically related regions of the heap. The targeted regions include objects that are part of the same component structure (recursive data structure). The result of the technique outlined in this paper has the form of a compact normal form (an abstract model) that boosts the efficiency of the static analysis via speeding its convergence. The result of heap abstraction, together with some properties of data structures, can be used to enable program optimizations like static deallocation, pool allocation, region-based garbage collection, and object co-location.

More precisely, this paper proposes algorithms for abstracting heap components with the layout of a singly linked list, a binary tree, a cycle, and a directed acyclic graph. The termination and correctness of these algorithms are studied in the paper. Towards presenting the algorithms the paper also presents concrete and abstract models for heap representations.

© 2012 Egyptian Mathematical Society. Production and hosting by Elsevier B.V.
Open access under [CC BY-NC-ND license](#).

1. Introduction

This paper presents an efficient technique for heap abstraction which takes the form of identifying and grouping logically related regions of heaps. The result of heap abstraction is a normal form for the program heap. The normal form is necessary for abstractly modeling programs and it boosts the efficiency

of the static data flow analysis via assisting the analysis to converge faster. The information provided by the normal form can also be used by client optimization applications to achieve the analyses of object co-location, pool allocation [1], static deallocation [2], etc.

The concept of heap abstraction emerges naturally in the course of research on object allocation and memory layout where techniques like pool allocation and object co-location use the heap abstraction to improve object locality. The efficiency of garbage collection [3] is also boosted using heap abstraction by the means of other techniques. Applications that statically deallocate data-structures or regions use heap abstractions more directly than others. In the course of abstraction, by restricting the grouping process to regions of heap that are expected to contain dead objects, the abstracting information

E-mail address: maelzawawy@cu.edu.eg

Peer review under responsibility of Egyptian Mathematical Society.



Production and hosting by Elsevier

can be used to delay times of garbage collection. Parallel garbage collection can also be treated by other approaches that use heap write/read information to statically find regions of heap that can be securely grouped without burden for the mutator.

Various techniques for heap abstraction [4–6] are used by the approaches referred to above to get region information used later in the optimization stage. The simplest of these approaches groups the heap objects based on the result of a pointer analysis like Steensgaard analysis [7]. Most of these techniques do not conveniently model object-oriented properties of data structures because the techniques are based on pointer analysis or other analyses that are flow/context insensitive. This paper presents a technique which is much more precise than these techniques. Moreover, our technique can be used as a tool to optimize the heap, which results in boosting the efficiency of memory regions.

Additionally, our proposed technique is useful to improve the efficiency of a range of static analysis approaches. This is accomplished via using the heap abstraction to normalize abstract models [8] which in turn results in reducing the height of the abstract lattice. Therefore this normalization process can be realized as a widening operation that turn domains of infinite height into ones of finite height. The normalization mentioned here has two aspects. One aspect is the compactification of recursive structures of possibly infinite size into finite structures. The other aspect is the using of a similarity relation to group objects, making up composite data structures.

Although the idea of heap abstraction (normalization) has already appeared in existing research, the algorithms presented in this paper achieving heap abstraction are more general, precise, and reliable (their correctness are proved) than those that have been developed in previous works. Precisely, our approach applies to a variety of recursive data structures such as singly linked lists, binary trees, cycles, and acyclic directed graphs. The technique presented in this paper can appropriately handle multi-component structures which could not be handled by most existing works.

1.1. Motivating example

Figs. 1, 3, 5, and 7 present motivating examples for our work. Suppose that we have a heap whose cells have the shape of four components; the singly linked list, the binary tree, the cycle, and the DAG (abbreviation for directed acyclic graph) on the left-hand side of Figs. 1, 3, 5, and 7, respectively. We note that some cells like

1. the nodes pointed to by variables s and e in the linked list, the cycle, and the DAG.
2. the second last node of the linked list, and
3. the root of the binary tree and the third and fourth (from left to right) nodes of the third level of the binary tree

are interesting and contain additional information as compared to the remaining cells. Other cells of the heap, like the

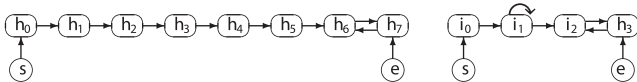


Figure 1 A concrete singly linked list together with its abstraction.

second and third nodes of the linked list, are ordinary and carries no extra information. It is wise and helpful to abstract such heap into one that consists of the four components on the right hand side of Figs. 1, 3, 5, and 7. The meant abstraction here is that of grouping logically related cells of the heap.

Remark 1. Self edges in abstracted components of the heap of our example have different meanings depending on the component layout.

1.2. Contributions

Contributions of this paper are the following:

1. A new technique for heap abstraction; novel algorithms for identifying and grouping logically related cells in singly linked lists, binary trees, cycles, and directed acyclic graphs. The termination and correctness of these algorithms are studied.
2. New concrete and abstract models for heap representations; a formal concept (valid abstraction) capturing the relationship between a concrete model and its abstraction.

1.3. Organization

The rest of the paper is organized as follows: Section 2 presents the parametrically labeled storage shape graph models (concrete and abstract) that we use to describe our new technique for heap abstraction. Sections 3–6 present new algorithms for identifying and grouping logically related cells in singly linked lists, binary trees, cycles, and directed acyclic graphs, respectively. The algorithm that abstracts heaps and that calls other introduced algorithms is presented in Section 7. Related work is briefly reviewed in Section 8.

2. Concrete and abstract heap models

This section introduces heap models that the work presented herein builds on. Graphs are basic components of our models. Similar models are used in related work [6,9] towards shape and sharing analysis of Java programs. It is worth mentioning that concepts of this paper are applicable in other techniques based on separation logic [10–12].

As usual, our semantics of memory is defined using pairs of stacks and heaps. The stack assigns values to variables and the heap assigns values to memory addresses. Each pair of a stack and a heap is called a *concrete component*. The concept of *concrete heap* denotes a finite set of concrete components. A concrete component is represented by a labeled directed graph which has a layout attribute that captures the layout of the memory cells represented by the component. The precise definition is the following:

Definition 1. A *concrete heap* is a finite set of disjoint labeled directed graphs (called *concrete components*) C_1, \dots, C_n each of which has a *layout* attribute that can have one of four values; singly linked list (*SLL*), tree (*T*), cycle (*C*), and directed acyclic graph (*DAG*). The layout of a component, C_i , is denoted by $C_i.layout$. More precisely, $C_i = (V_i, A_i, P_i)$, where:

1. V_i is a finite set of variables; $V_i \subseteq Var$.
2. A_i is a finite set of memory addresses; $A_i \subseteq Addr_s$.
3. When $C_i.layout = SLL, DAG$, or C , P_i is a set of pointers defined by $P_i \subseteq (V_i \times A_i) \cup (A_i \times A_i)$.
4. When $C_i.layout = T$, $P_i \subseteq (V_i \times A_i) \cup (A_i \times A_i \times \{l, r\})$.

Regions in heaps, edges of regions, edges entering regions, and edges leaving regions are defined as follows:

Definition 2. A set R is said to be a *region* in a concrete component $C = (V, A, P)$ if $R \subseteq A$. Moreover,

1. $P(R) = \{(a_1, a_2, nx), (a_1, a_2) \in P \mid a_1, a_2 \in R\}$,
2. $P_{in}(R) = \{(a_1, a_2, nx), (a_1, a_2) \in P \mid a_1 \in A \setminus R, a_2 \in R\}$, and
3. $P_{out}(R) = \{(a_1, a_2, nx), (a_1, a_2) \in P \mid a_1 \in R, a_2 \in A \setminus R\}$.

Our concept of abstract heap is inspired by the technique of storage shape graph presented in [13,6]. The concept of concrete component is abstracted by that of abstract component which is a labeled directed graph $(\widehat{V}, \widehat{N}, \widehat{P})$, where (a) \widehat{V} is a set of nodes correspond to variables, (b) \widehat{N} is a set of nodes each of which corresponds to (abstracts) a region of a concrete component, and (c) \widehat{P} is a set of graph edges, each of which corresponds to (abstracts) a set of pointers. Analogously to concrete component, each abstract component has a layout attribute. More precisely abstract heaps and abstract components are defined as follows:

Definition 3. An *abstract heap* is a finite set of disjoint labeled directed graphs (called *abstract components*) $\widehat{C}_1, \dots, \widehat{C}_n$ each of which has a *layout* attribute that is SLL, T, C , or DAG . More precisely, $\widehat{C}_i = (\widehat{V}_i, \widehat{N}_i, \widehat{P}_i)$ where:

1. $\widehat{V}_i \subseteq Var$.
2. \widehat{N}_i is a finite set of node identifiers (each represents a region of the heap).
3. When $\widehat{C}_i.layout = SLL, DAG$, or C , P_i is a set of pointers defined by $\widehat{P}_i \subseteq (\widehat{V}_i \times \widehat{N}_i) \cup (\widehat{N}_i \times \widehat{N}_i)$.
4. When $\widehat{C}_i.layout = T$, $P_i \subseteq (\widehat{V}_i \times \widehat{N}_i) \cup (\widehat{N}_i \times \widehat{N}_i \times \{l, r\})$.

Regions in abstract components are defined analogously as sets of nodes identifiers.

Remark 2. Every concrete heap is an abstract one.

Now we introduce the concept of abstraction. An abstract component \widehat{C} is described as a valid abstraction of another one \widehat{C}' , if (a) they have the same layout and same sets of variables and (b) there are two maps; a map from nodes of \widehat{C} to nodes of \widehat{C}' and a map from edges of \widehat{C} to edges of \widehat{C}' such that these maps preserve the connectivity of the components.

Definition 4. An abstract component $\widehat{C}_1 = (\widehat{V}_1, \widehat{N}_1, \widehat{P}_1)$ is a *valid abstraction* of another abstract component $\widehat{C}_2 = (\widehat{V}_2, \widehat{N}_2, \widehat{P}_2)$ if $\widehat{C}_1.layout = \widehat{C}_2.layout$, $\widehat{V}_1 = \widehat{V}_2$, and there are two onto maps $f_N : \widehat{N}_1 \rightarrow \widehat{N}_2$ and $f_P : \widehat{P}_1 \rightarrow \widehat{P}_2$ such that:

1. $\forall (v, n_2) \in \widehat{P}_2. f_P^{-1}((v, n_2)) \subseteq \{(v, n_1) \in \widehat{P}_1 \mid n_1 \in f_N^{-1}(n_2)\}$.

2. $\forall (n_2, n'_2) \in \widehat{P}_2. f_P^{-1}((n_2, n'_2)) \subseteq \{(n_1, n'_1) \in \widehat{P}_1 \mid n_1 \in f_N^{-1}(n_2) \wedge n'_1 \in f_N^{-1}(n'_2)\}$.
3. $\forall (n_2, n'_2, nx) \in \widehat{P}_2. f_P^{-1}((n_2, n'_2, nx)) \subseteq \{(n_1, n'_1, nx) \in \widehat{P}_1 \mid n_1 \in f_N^{-1}(n_2) \wedge n'_1 \in f_N^{-1}(n'_2)\}$.

The pair (f_N, f_P) is called the *witness* of the valid abstraction.

Lemma 1. *The valid-abstraction relation on abstract components is transitive.*

Proof. Suppose \widehat{C}_2 is a valid abstraction of \widehat{C}_1 with witness (f_N, f_P) and \widehat{C}_3 is a valid abstraction of \widehat{C}_2 with witness (f'_N, f'_P) . Then, it is easy to see that $(f'_N \circ f_N, f'_P \circ f_P)$ witnesses that \widehat{C}_3 is a valid abstraction of \widehat{C}_1 . \square

Definition 5. An abstract heap $(\widehat{C}_1, \dots, \widehat{C}_n)$ is a valid abstraction of a concrete heap (C_1, \dots, C_n) if for every $1 \leq i \leq n$, \widehat{C}_i is a valid abstraction of C_i .

Remark 3. Every concrete heap is a valid abstraction of itself.

3. Abstracting singly linked lists

This section presents a novel algorithm for abstracting heap components whose layouts are singly linked list. Fig. 1 presents a linked list of length 8 (left) together with its abstracted representation (right) which is a list of length 4. Clearly, some nodes of the concrete list are grouped into regions. We note that nodes h_0 and h_7 are special nodes as they are pointed to by two variables s and e , respectively. The node h_6 is also special as it shares a back edge with the node h_7 . These special nodes are not grouped in the abstracted version of the list. The nodes h_1 up to the node h_5 are ordinary nodes; there is nothing special about them. These ordinary nodes are grouped into the node (region) i_1 in the abstracted version. The self-edge of i_1 captures the phenomenon that i_1 represents a region of the concrete component.

Various properties of lists are captured by partitioning the list nodes into two classes; ordinary and special nodes. First, the compressed representation of the list in the abstracted version substantially boosts the efficiency of the analysis. Next, the first and last nodes, pointed to by variables s and e respectively, and the second-to-last node are kept separate. This supports the analysis to conveniently simulate the semantics of later program commands. Although ordinary nodes of each list in the program are grouped into a single node in the abstracted version of the list, unrelated lists of the program are kept separate. In other words, separate lists in the concrete heap are kept separate in the abstract model while nodes in the same lists are grouped together. This helps in preserving the information required by many optimization techniques. The formal definition of special and ordinary nodes of singly linked lists is as follows:

Definition 6. Let $\widehat{C} = (\widehat{V}, \widehat{N}, \widehat{P})$ be an abstract component whose layout is SLL . Then, a node $n \in \widehat{N}$ is *special* if either:

1. for some $\hat{v} \in \widehat{V}$, $(\hat{v}, n) \in \widehat{P}$, or
2. there exists $(a, b) \in \widehat{P}$ such that $a, b \in \widehat{N}$, $depth(a) > depth(b)$, and $n \in \{a, b\}$; i.e., n contributes to a back edge.

A node is *ordinary* if it is not special.

Fig. 2 outlines a novel algorithm for abstracting singly linked lists. The algorithm first collects ordinary nodes of the input linked list in a set M . Then, the algorithm merges any pair of nodes in M that shares an edge. The merging process includes adding self-edges. The algorithm supposes that there exists a function *remove-node* that removes a node from a linked list.

The termination and correctness of *Abstract-SLL* are proved as follow:

Theorem 1. *The algorithm Abstract-SLL always terminates.*

Proof. We note that M is finite because $M \subseteq \hat{N}$ and \hat{N} is finite. If the cardinality of M is m , then the *while* loop in the second step iterates at most $m - 1$ times. \square

Theorem 2. *Suppose that $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ is an abstract component whose layout is SLL and $\hat{C}' = \text{Abstract-SLL}(\hat{C})$. Then, \hat{C}' is a valid abstraction of \hat{C} .*

Proof. We note that there two kinds of operations that occur throughout the algorithm; removing nodes and adding self-edges. Since both of these actions do not affect the layout of the component, the layout of the output component is guaranteed to remain SLL. By induction on the cardinality of M , we complete the proof that \hat{C}' is a valid abstraction of \hat{C} . For the induction base, for $|M| = 0$ and for $|M| = 1$, the required result is trivial. For the inductive hypothesis, we assume that the required result is true for any finite set N with $|N| = n$ for some positive integer n . For the inductive step, we prove the required result holds for a finite set M with $|M| = n + 1$ as follows. We assume that (a, b) is the edge picked at the first iteration of the loop (if there are none, then the algorithm terminates and the output is clearly correct). Clearly \hat{C} is a valid abstraction of itself with the identity witness (I_N, I_P) . Now the component obtained after the first iteration of the loop, denoted by \hat{C}'' , is a valid abstraction of \hat{C} with witness $w'' = (I_N[a \mapsto b], I_P[(a, b) \mapsto (a, a)])$. The running of the rest of the algorithm on \hat{C} is equivalent of that on \hat{C}'' . Clearly $|M| = n$ for the run of \hat{C}'' . Therefore, by induction hypothesis \hat{C}' is a valid abstraction of \hat{C}'' with some witness w' . By Lemma 1, \hat{C}' is a valid abstraction of \hat{C} with witness $w = w' \circ w''$. \square

Algorithm : *Abstract-SLL*

- Input : An abstract component $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ such that $\hat{C}.layout = SLL$;
- Output : An abstract component $\hat{C}' = (\hat{V}', \hat{N}', \hat{P}')$ such that \hat{C}' is a valid abstraction for \hat{C} ;
- Method :
 1. $M \leftarrow$ ordinary nodes of \hat{N} ;
 2. While there are $a, b \in M$ such that $a \neq b$ and $(a, b) \in \hat{P}$
 - (a) $(\hat{V}, \hat{N}, \hat{P}) = \text{remove-node}(b, \hat{V}, \hat{N}, \hat{P})$;
 - (b) $\hat{P} \leftarrow \hat{P} \cup \{(a, a)\}$;
 - (c) $M \leftarrow M \setminus \{b\}$;
 3. Return $(\hat{V}, \hat{N}, \hat{P})$;

Figure 2 The algorithm *Abstract-SLL*.

4. Trees abstraction

In this section, we present a new algorithm for abstracting heap components whose layout are tree. Fig. 3 presents a binary tree (left) of size 15 and height 3 together with its abstracted representation (right) of size 8 and height 3. We note that node h_0 is special because it is pointed to by the variable R . Also nodes h_5 and h_6 are special as they share a horizontal edge of the tree. The special nodes are not grouped in the abstracted version of the tree. We also note that there is nothing special about the left subtree. Therefore, the left subtree is grouped in the node i_1 of the compact tree. The self-edges of i_1 model the fact that i_1 represents a full binary subtree.

Definition 7. Let $(\hat{V}, \hat{N}, \hat{P})$ be an abstract component whose layout is T . Then, a node $n \in \hat{N}$ is *special* if

1. for some $\hat{v} \in \hat{V}$, $(\hat{v}, n) \in \hat{P}$, or
2. there exists $(a, b, -) \in \hat{P}$ such that $a \neq b$, $a, b \in \hat{N}$, $depth(a) \geq depth(b)$, and $n \in \{a, b\}$; i.e., n contributes to a horizontal or a back edge.

A node is *ordinary* if it is not special.

Fig. 4 presents a new algorithm for abstracting trees. The algorithm first collects ordinary nodes of the input tree in a set M . Then, the algorithm traverses the tree bottom-up, merging ordinary nodes. The merging process includes adding self-edges. The algorithm supposes that there is a function *remove-nodes* that removes a couple of nodes from a tree.

The proofs of the following two theorems run along similar lines as those of Theorems 1 and 2, respectively.

Theorem 3. *The algorithm Abstract-T always terminates.*

Theorem 4. *Suppose that $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ is an abstract component whose layout is T and $\hat{C}' = \text{Abstract-T}(\hat{C})$. Then, \hat{C}' is a valid abstraction of \hat{C} .*

5. Cycles abstraction

The new algorithm presented in this section takes care of abstracting heap components whose layout is a cycle. Fig. 5 presents a cycle (left) of size 8 together with its abstracted representation (right) of size 4. We note that node h_0 is special be-

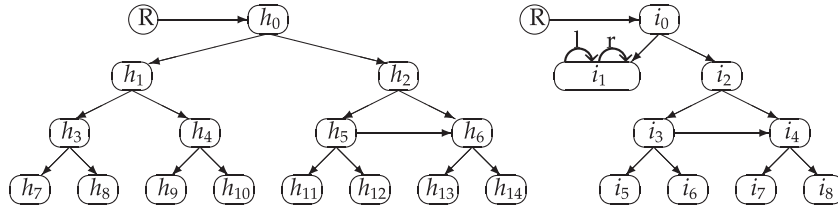


Figure 3 A concrete binary tree together with its abstraction.

Algorithm : *Abstract-T*

- Input : An abstract component $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ whose layout is T and whose height is denoted by h ;
- Output : An abstract component $\hat{C}' = (\hat{V}', \hat{N}', \hat{P}')$ which is a valid abstraction of \hat{C} ;
- Method :
 1. $M \leftarrow$ ordinary nodes of \hat{N} ;
 2. For $(i = h - 1; i > 0; i --)$
 - (a) While M has distinct elements a, b, c such that $depth(a) = i$ and $(a, b, l), (a, c, r) \in \hat{P}$
 - i. $(\hat{V}, \hat{N}, \hat{P}) = remove_nodes(b, c, \hat{V}, \hat{N}, \hat{P})$;
 - ii. $\hat{P} \leftarrow \hat{P} \cup \{(a, a, l), (a, a, r)\}$;
 - iii. $M \leftarrow M \setminus \{b, c\}$;
 3. Return $(\hat{V}, \hat{N}, \hat{P})$;

Figure 4 The algorithm *Abstract-T*.

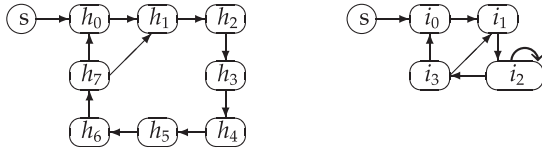


Figure 5 A concrete cycle together with its abstraction.

cause it is pointed to by the variable S . Also nodes h_7 and h_1 are special because there are more than one edge leaving and entering, respectively, the nodes. As it should happen, the special nodes are not grouped in the abstracted version of the cycle. We also note that there is nothing special about nodes h_2 up to node h_6 . Therefore, these nodes are grouped in the node i_2 of the compressed cycle. The self-edge of i_2 models the fact that i_2 represents a sequence of arbitrary length of the cycle. Special and ordinary nodes of cycles are defined as follows:

Definition 8. Let $(\hat{V}, \hat{N}, \hat{P})$ be an abstract component whose layout is C . Then, a node $n \in \hat{N}$ is *special* if

- for some $\hat{v} \in \hat{V}$, $(\hat{v}, n) \in \hat{P}$, or
- $|P_{in}(\{n\})| > 1$, or
- $|P_{out}(\{n\})| > 1$.

A node is *ordinary* if it is not special.

Fig. 6 presents an original algorithm, *Abstract-C*, for cycle abstraction. Similar to the algorithms presented so far, the first step of the algorithm is to collect ordinary nodes of the cycle. The algorithm then repeatedly picks a pair of ordinary nodes that share a direct edge. The algorithm removes one of the two nodes with its edges and adds a self-node to the remaining node.

The proofs of the following two theorems, which address termination and correctness of the algorithm *Abstract-C*, are similar to proofs of Theorems 1 and 2, respectively.

Theorem 5. The algorithm *Abstract-C* always terminates.

Theorem 6. Suppose that $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ is an abstract component whose layout is C and $\hat{C}' = Abstract - C(\hat{C})$. Then, \hat{C}' is a valid abstraction of \hat{C} .

6. DAG abstraction

This section presents a novel way to abstract heap components whose layout are *DAG*. Fig. 7 presents a *DAG* (left) of size 8 together with its abstracted representation (right) of size 4. Node h_0 is special because it is pointed to by the variable S . This special node is kept separate in the abstracted version of the *DAG*. There is nothing special about nodes h_2 up to node h_6 . Therefore, these nodes are grouped in the node i_2 of the compressed *DAG*. The self-edge of i_2 models the fact that i_2 represents a set of nodes that is reference similar. Two distinct nodes of a *DAG* are reference similar if they are not connected by an edge, point to the same set of nodes, and are pointed to by the same set of nodes. A set of nodes is reference similar if every distinct pair of its elements are reference similar. The following definitions formally introduce concepts of special nodes, ordinary nodes, and reference similarity.

Definition 9. Let $(\hat{V}, \hat{N}, \hat{P})$ be an abstract component whose layout is *DAG*. Then, a node $n \in \hat{N}$ is *special* if for some $\hat{v} \in \hat{V}$, $(\hat{v}, n) \in \hat{P}$. A node is *ordinary* if it is not special.

- Algorithm: *Abstract-C*
- Input : An abstract component $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ such that $\hat{C}.layout = C$;
 - Output : An abstract component $\hat{C}' = (\hat{V}', \hat{N}', \hat{P}')$ such that \hat{C}' is a valid abstraction for \hat{C} ;
 - Method :
 1. $M \leftarrow$ ordinary nodes of \hat{N} ;
 2. While there are $a, b \in M$ such that $a \neq b$ and $(a, b) \in \hat{P}$
 - While there exists $(b, c) \in \hat{P}$
 - (a) $\hat{P} \leftarrow (\hat{P} \setminus \{(b, c)\}) \cup \{(a, c)\}$;
 - $\hat{N} \leftarrow \hat{N} \setminus \{b\}$;
 - $\hat{P} \leftarrow (\hat{P} \setminus \{(a, b)\}) \cup \{(a, a)\}$;
 - $M \leftarrow M \setminus \{b\}$;
 3. Return $(\hat{V}, \hat{N}, \hat{P})$;

Figure 6 The algorithm *Abstract-C*.

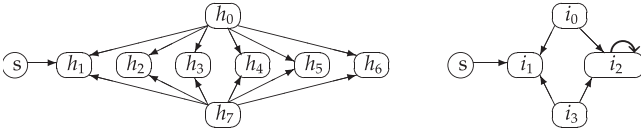


Figure 7 A concrete DAG together with its abstraction.

Definition 10. Let $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ be an abstract component such that $\hat{C}.layout = DAG$. Two distinct nodes $a, b \in \hat{N}$ are *reference similar* with respect to \hat{C} if

1. $(a, b) \notin \hat{P}$ and $(b, a) \notin \hat{P}$,
2. $\{c \in \hat{N} \mid (c, a) \in \hat{P}\} = \{c \in \hat{N} \mid (c, b) \in \hat{P}\}$, and
3. $\{c \in \hat{N} \mid (a, c) \in \hat{P}\} = \{c \in \hat{N} \mid (b, c) \in \hat{P}\}$.

A set of nodes $A \subseteq \hat{N}$ is *reference similar* with respect to \hat{C} if every pair of distinct elements in A is reference similar with respect to \hat{C} .

Fig. 8 presents the algorithm *Abstract-DAG* that abstracts heap components with *DAG* layout. The algorithm calls the algorithm *Ref-similar-DAG*, Fig. 9, that for a given heap component calculates a set of reference similar sets.

The first step of the algorithm *Abstract-DAG* is to calculate a set, G' , of reference similar sets. The singleton elements of G' are filtered out to obtain the set G . For every set A in G , the algorithm groups elements of A into a single node of the abstracted *DAG* with a self-edge. The algorithm *Ref-similar-DAG* first initializes G to the empty set. Secondly, the algorithm stores the ordinary nodes of the input component in the set M . The third step is to partition the set of ordinary ele-

- Algorithm : *Abstract-DAG*
- Input : An abstract component $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ such that $\hat{C}.layout = DAG$;
 - Output : An abstract component $\hat{C}' = (\hat{V}', \hat{N}', \hat{P}')$ such that \hat{C}' is a valid abstraction for \hat{C} ;
 - Method :
 1. $G' \leftarrow Ref-similar-DAG(\hat{C})$;
 2. $G \leftarrow \{A \mid A \in G' \text{ and } |A| > 1\}$;
 3. For every $A = \{a_1, a_2, \dots, a_n\} \in G$
 - (a) $\hat{N} \leftarrow \hat{N} \setminus \{a_2, \dots, a_n\}$;
 - (b) $\hat{P} \leftarrow \hat{P} \setminus \{(a, b) \mid \{a, b\} \cap \{a_2, \dots, a_n\} \neq \emptyset\}$;
 - (c) $\hat{P} \leftarrow \hat{P} \cup \{(a_1, a_1)\}$;
 4. Return $(\hat{V}, \hat{N}, \hat{P})$;

Figure 8 The algorithm *Abstract-DAG*.

ments, M , into reference similar sets. This is done via picking an element $a \in M$ and adding all elements that are reference similar to a to the partition of a .

The proof of the following theorem is similar to that of Theorem 1.

Theorem 7. *The algorithm Ref-similar-DAG always terminates.*

The proof of the following theorem is by induction on the cardinality of M .

Theorem 8. *Suppose that $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ is an abstract component and $G = Ref-similar-DAG(\hat{C})$. Then, every element of G is reference similar with respect to \hat{C} .*

Theorem 9. *The algorithm Abstract-DAG always terminates.*

Theorem 10. *Suppose that $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ is an abstract component whose layout is *DAG* and $\hat{C}' = Abstract-DAG(\hat{C})$. Then, \hat{C}' is a valid abstraction of \hat{C} .*

7. Heap abstraction

This section presents our basic algorithm, *Heap-Abstract*, for heap abstraction. For a given abstract heap of n components, the algorithm checks the layout of each component and calls the appropriate algorithm for abstracting the component in hand. The algorithm is outlined in Fig. 10. The termination and correctness of the algorithm are inherited from those of algorithms presented so far.

Theorem 11. *The algorithm Heap-Abstract always terminates.*

Algorithm : *Ref-similar-DAG*

- Input : An abstract component $\hat{C} = (\hat{V}, \hat{N}, \hat{P})$ such that $\hat{C}.layout = DAG$;
- Output : A set G of finite subsets of \hat{N} such that every set of G is reference similar;
- Method :
 1. $G \leftarrow \emptyset$;
 2. $M \leftarrow$ ordinary elements of \hat{N} ;
 3. While $M \neq \emptyset$
 - (a) Pick a from M .
 - (b) $A \leftarrow \{a\}$
 - (c) For every $b \in M$,
If $A \cup \{b\}$ is reference similar, then $A \leftarrow A \cup \{b\}$;
 - (d) $G \leftarrow G \cup \{A\}$;
 - (e) $M \leftarrow M \setminus A$;
 4. Return G ;

Figure 9 The algorithm *Ref-similar-DAG*.

Algorithm : *Heap-Abstract*

- Input : A concrete heap $h = (C_1, \dots, C_n)$;
- Output : An abstract heap $\hat{h} = (\hat{C}_1, \dots, \hat{C}_n)$ such that \hat{h} is a valid abstraction for h ;
- Method :
 1. For ($i = 1$; $i++$; $i \leq n$)
 - (a) If $(C_i.layout = SLL)$, then $\hat{C}_i \leftarrow Abstract-SLL(\hat{C}_i)$;
 - (b) If $(C_i.layout = T)$, then $\hat{C}_i \leftarrow Abstract-T(\hat{C}_i)$;
 - (c) If $(C_i.layout = C)$, then $\hat{C}_i \leftarrow Abstract-C(\hat{C}_i)$;
 - (d) If $(C_i.layout = DAG)$, then $\hat{C}_i \leftarrow Abstract-DAG(\hat{C}_i)$;
 2. Return $(\hat{C}_1, \dots, \hat{C}_n)$;

Figure 10 The algorithm *Heap-Abstract*.

Theorem 12. *Suppose that h is a concrete heap and $\hat{h} = Heap - Abstract(h)$. Then, \hat{h} is a valid abstraction of h .*

8. Related work

The area of statically improving heap allocation, abstraction, and layout for object oriented programs is rich in literature [4,14,5,6,15,16,1]. These techniques are conveniently applicable to large programs and use results of pointer analysis to calculate static partitions that are required to compute region information. However, there are common drawbacks to these techniques; (a) they have a limited capability to conveniently analyze programs that rearrange regions and (b) they have a limited capability to conveniently explore components of large complex structures. These deficiencies are caused by imprecision of determined partitioning and flow insensitivity. Our algorithms for heap optimization presented in this paper overcome these drawbacks.

Other techniques that are based on separation logic [10,17] simulate destructive updates of heaps and how these updates modify heap layout [18,12,19,11,20–25]. These techniques precisely simulate complicated heap operations but the limitations imposed by them render these techniques inappropriate for region analysis. This drawback is witnessed by the fact that most of these approaches are formulated to analyze programs that handle only lists or trees. A future direction of research is to extend the techniques of these papers in the spirit of our present paper. This is huge potential in this direction by virtue of generality of separation logic as a general-purpose framework.

Mathematical domains and maps between domains can be used to mathematically represent programs and data

structures. This representation is called denotational semantics of programs. One of our directions for future research is to translate heap concepts to the side of denotational semantics [26,27]. Doing so provide a good tool to mathematically study in deep heap concepts. Then obtained results can be translated back to the side of programs and data structures.

References

- [1] Zhenjiang Wang, Chenggang Wu, Pen-Chung Yew, On improving heap memory layout by dynamic pool allocation, in: Andreas Moshovos, J. Gregory Steffan, Kim M. Hazelwood, David R. Kaeli (Eds.), CGO, ACM, 2010, pp. 92–100.
- [2] Javier de Dios, Manuel Montenegro, Ricardo Pena, Certified absence of dangling pointers in a language with explicit deallocation, in: Dominique Mery, Stephan Merz (Eds.), IFM, Lecture Notes in Computer Science, vol. 6396, Springer, 2010, pp. 305–319.
- [3] Martin Schoeberl, Wolfgang Puffitsch, Nonblocking real-time garbage collection, ACM Transactions on Embedded Computation Systems 10 (1) (2010) 91–101.
- [4] Sigmund Cherm, Radu Rugina, Compile-time deallocation of individual objects, in: Erez Petrank, J. Eliot, B. Moss (Eds.), ISMM, ACM, 2006, pp. 138–149.
- [5] Stephen Magill, Ming-Hsien Tsai, Peter Lee, Yih-Kuen Tsay, Automatic numeric abstractions for heap-manipulating programs, in: Manuel V. Hermenegildo, Jens Palsberg (Eds.), POPL, ACM, 2010, pp. 211–222.
- [6] Mark Marron, Deepak Kapur, Manuel V. Hermenegildo, Identification of logically related heap regions, in: Hillel Kolodner, Guy L. Steele Jr., (Eds.), ISMM, ACM, 2009, pp. 89–98.
- [7] Bjarne Steensgaard, Points-to analysis in almost linear time, in: POPL, 1996, pp. 32–41.

- [8] Maurice H. Ter Beek, Alessandro Fantechi, Stefania Gnesi, Franco Mazzanti, A state/event-based model-checking approach for the analysis of abstract system properties, *Science of Computer Programming* 76 (2) (2011) 119–135.
- [9] Mark Marron, Mario Mendez-Lojo, Manuel V. Hermenegildo, Darko Stefanovic, Deepak Kapur, Sharing analysis of arrays, collections, and recursive structures. in: Shriram Krishnamurthi, Michal Young (Eds.), *PASTE*, ACM, 2008, pp. 43–49.
- [10] Josh Berdine, Cristiano Calcagno, Byron Cook, Dino Distefano, Peter W. O'Hearn, Thomas Wies, Hongseok Yang, Shape analysis for composite data structures, in: Werner Damm, Holger Hermanns (Eds.), *CAV*, Lecture Notes in Computer Science, vol. 4590, Springer, 2007, pp. 178–192.
- [11] Bertrand Jeannot, Alexey Loginov, Thomas W. Reps, Mooly Sagiv, A relational approach to interprocedural shape analysis, *ACM Transactions on Programming Languages and Systems* 32 (2) (2010) 72–83.
- [12] Hongseok Yang, Oukseh Lee, Josh Berdine, Cristiano Calcagno, Byron Cook, Dino Distefano, Peter W. O'Hearn, Scalable shape analysis for systems code, in: Aarti Gupta, Sharad Malik (Eds.), *CAV*, Lecture Notes in Computer Science, vol. 5123, Springer, 2008, pp. 385–398.
- [13] David R. Chase, Mark N. Wegman, F. Kenneth Zadeck, Analysis of pointers and structures, in: *PLDI*, 1990, pp. 296–310.
- [14] Isil Dillig, Thomas Dillig, Alex Aiken, Symbolic heap abstraction with demand-driven axiomatization of memory invariants, in: William R. Cook, Siobhan Clarke, Martin C. Rinard (Eds.), *OOPSLA*, ACM, 2010, pp. 397–410.
- [15] Wolfgang Puffitsch, Benedikt Huber, Martin Schoeberl, Worst-case analysis of heap allocations, in: Tiziana Margaria, Bernhard Steffen (Eds.), *ISoLA* (2), Lecture Notes in Computer Science, vol. 6416, Springer, 2010, pp. 464–478.
- [16] Mohsen Vakilian, Danny Dig, Robert L. Bocchino Jr., Jeffrey Overbey, Vikram S. Adve, Ralph E. Johnson, Inferring method effect summaries for nested heap regions, in: *ASE*, IEEE Computer Society, 2009, pp. 421–432.
- [17] Bolei Guo, Neil Vachharajani, David I. August, Shape analysis with inductive recursion synthesis, in: Jeanne Ferrante, Kathryn S. McKinley (Eds.), *PLDI*, ACM, 2007, pp. 256–265.
- [18] Reinhard Wilhelm, Shmuel Sagiv, Thomas W. Reps, Shape analysis, in: David A. Watt (Ed.), *CC*, Lecture Notes in Computer Science, vol. 1781, Springer, 2000, pp. 1–17.
- [19] Dirk Beyer, Thomas A. Henzinger, Gregory Theoduloz, Damien Zufferey, Shape refinement through explicit heap analysis, in: David S. Rosenblum, Gabriele Taentzer (Eds.), *FASE*, Lecture Notes in Computer Science, vol. 6013, Springer, 2010, pp. 263–277.
- [20] Mohamed El-Zawawy, Nagwan Daoud, New error-recovery techniques for faulty-calls of functions, *Computer and Information Science* 4 (3) (2012).
- [21] Mohamed A. El-Zawawy, Program optimization based pointer analysis and live stack-heap analysis, *International Journal of Computer Science Issues* 8 (2) (2011) 98–107.
- [22] Mohamed A. El-Zawawy, Flow sensitive-insensitive pointer analysis based memory safety for multithreaded programs, in: Beniamino Murgante, Osvaldo Gervasi, Andres Iglesias, David Taniar, Bernady O. Apduhan (Eds.), *ICCSA* (5), Lecture Notes in Computer Science, vol. 6786, Springer, 2011, pp. 355–369.
- [23] Mohamed A. El-Zawawy, Probabilistic pointer analysis for multithreaded programs, *ScienceAsia* 37 (4) (2011).
- [24] Mohamed A. El-Zawawy, Dead code elimination based pointer analysis for multithreaded programs, *Journal of the Egyptian Mathematical Society* (2012), <http://dx.doi.org/10.1016/j.joems.2011.12.011>.
- [25] Mohamed A. El-Zawawy, Hamada A. Nayel, Partial redundancy elimination for multi-threaded programs, *IJCSNS International Journal of Computer Science and Network Security* 11 (10) (2011).
- [26] Mohamed A. El-Zawawy, Achim Jung, Priestley duality for strong proximity lattices, *Electronic Notes in Theoretical Computer Science* 158 (2006) 199–217.
- [27] Mohamed A. El-Zawawy, *Semantic Spaces in Priestley Form*. PhD thesis, University of Birmingham, UK, January 2007.