Shape analysis abstractions

MPRI — Cours 2.6 “Interprétation abstraite : application à la vérification et à l’analyse statique”

Xavier Rival

INRIA

Jan, 17th, 2022
Shape analysis

Shape analyses aim at discovering structural invariants of programs that manipulate complex unbounded data-structures.

Applications:

- establish **memory safety**
- verify the preservation of **structural properties** e.g., list, doubly-linked lists, trees, ...
- reason about programs that manipulate **unbounded** memory states

Previous course: **TVLA**, *i.e.*,

- **logical predicates** which evaluate in three valued logic
- **shape graphs**, described by the predicates

\[
\begin{array}{c}
\text{true} \\
\text{false} \\
\frac{1}{2}
\end{array} = \text{true} \land \text{false}
\]
Another family of shape analyses

Today: **systematically avoid weak updates**

- a logic to describe properties of memory states
- abstract domain
- static analysis algorithms
- combination with numerical domains
- widening operators...
Outline

1. Introduction

2. Separation Logic

3. A shape abstract domain relying on separation

4. Standard static analysis algorithms

5. Combining shape and value abstractions

6. Conclusion
Separation logic principle: avoid weak updates

How to deal with weak updates?

Avoid them!

- Always materialize exactly the cell that needs to be modified.
- Can be very costly to achieve, and not always feasible.

- Notion of property that holds over a memory region: special separating conjunction operator $\ast$

- Local reasoning:
  - Powerful principle, which allows to consider only part of the memory.

- Separation logic has been used in many contexts, including manual verification, static analysis, etc...
Separation Logic

Two kinds of formulas:
- **pure formulas** behave like formulas in first-order logic i.e., are not attached to a memory region
- **spatial formulas** describe properties attached to a memory region

**Pure formulas** denote value properties

\[ e ::= n \quad (n \in \mathbb{N}) \quad l \quad e_0 + e_1 \quad \ldots \]

\[ P ::= e_0 = e_1 \quad P' \lor P'' \quad P' \land P'' \ldots \]

Pure formulas semantics:
\[ \gamma(P) \subseteq E \times H \]

\[ E = \text{env} \quad e \land v \quad h_\text{pop} \quad H = A \rightarrow V \]
Separation logic: points-to predicates

The next slides introduce the main separation logic formulas $F ::= \ldots$

We start with the most basic predicate, that describes a single cell:

Points-to predicate

- **Predicate:**
  
  $$F ::= \ldots | a \mapsto v$$

  where $a$ is an address and $v$ is a value

- **Concretization:**
  
  $$(e, h) \in \gamma(x \mapsto v)$$ if and only if $$h = \llbracket x \rrbracket(e, h) \mapsto v$$

- **Example:**
  
  $$F = \&x \mapsto 18$$

  We also note $1 \mapsto e$, as an l-value $1$ denotes an address
Separation logic: separating conjunction

Merge of concrete heaps: let \( h_0, h_1 \in (\mathbb{V}_{\text{addr}} \to \mathbb{V}) \), such that \( \text{dom}(h_0) \cap \text{dom}(h_1) = \emptyset \); then, we let \( h_0 \otimes h_1 \) be defined by:

\[
\begin{align*}
\text{dom}(h_0) & \cup \text{dom}(h_1) & \rightarrow & \mathbb{V} \\
\ast & \in \text{dom}(h_0) & \mapsto & h_0(\ast) \\
\ast & \in \text{dom}(h_1) & \mapsto & h_1(\ast)
\end{align*}
\]

Separating conjunction

- **Predicate:**
  \[
  F ::= \ldots \mid F_0 \star F_1
  \]

- **Concretization:**
  \[
  \gamma(F_0 \star F_1) = \{(e, h_0 \otimes h_1) \mid (e, h_0) \in \gamma(F_0) \land (e, h_1) \in \gamma(F_1)\}
  \]
An example

Concrete memory layout
(pointer values underlined)

\[
\begin{align*}
\&x &= 300 & 64 \\
&    &= 304 & 312 \\
\&y &= 308 & 312 \\
\&z &= 312 & 88 \\
&    &= 316 & 0x0
\end{align*}
\]

A formula that abstracts away the addresses:

\[
\&x \mapsto \langle 64, \&z \rangle * \&y \mapsto \&z * \&z \mapsto \langle 88, 0 \rangle
\]
Separation logic: non separating conjunction

We can also add the **conventional conjunction operator**, with its **usual concretization**:

**Non separating conjunction**

- **Predicate:**
  \[
  F ::= \cdots | F_0 \land F_1
  \]

- **Concretization:**
  \[
  \gamma(F_0 \land F_1) = \gamma(F_0) \cap \gamma(F_1)
  \]

**Exercise:** **describe** and **compare** the concretizations of

- \( \&a \mapsto \&b \land \&b \mapsto \&a \)
- \( \&a \mapsto \&b \ast \&b \mapsto \&a \)
Separation Logic

Separating conjunction vs non separating conjunction

- **Classical conjunction**: properties for the same memory region
- **Separating conjunction**: properties for disjoint memory regions

\&a \rightarrow \&b \land \&b \rightarrow \&a

- the same heap verifies \&a \rightarrow \&b and \&b \rightarrow \&a
- there can be only one cell
- thus \( a = b \)

\&a \rightarrow \&b \ast \&b \rightarrow \&a

- two separate sub-heaps respectively satisfy \&a \rightarrow \&b and \&b \rightarrow \&a
- thus \( a \neq b \)

- Separating conjunction and non-separating conjunction have very different properties
- Both express very different properties e.g., no ambiguity on weak / strong updates
Separating and non separating conjunction

Logic rules of the two conjunction operators of SL:

- **Separating conjunction:**

\[
(e, h_0) \in \gamma(F_0) \quad (e, h_1) \in \gamma(F_1) \\
(e, h_0 \otimes h_1) \in \gamma(F_0 \star F_1)
\]

- **Non separating conjunction:**

\[
(e, h) \in \gamma(F_0) \quad (e, h) \in \gamma(F_1) \\
(e, h) \in \gamma(F_0 \land F_1)
\]

Reminiscent of Linear Logic [Girard87]: resource aware / non resource aware conjunction operators
Separation logic: empty store

Empty store

- **Predicate:**
  \[ F ::= \ldots \mid \text{emp} \]

- **Concretization:**
  \[ \gamma(\text{emp}) = \{(e, []) \mid e \in E\} = E \times \{[]\} \]

where [] denotes the empty store

- emp is the **neutral element for \(\ast\)** (monoid structure induced by \(\ast\))
- by contrast the **neutral element for \(\land\) is \text{TRUE}, with concretization:**
  \[ \gamma(\text{TRUE}) = E \times \mathbb{H} \]
Separation logic: other connectors

**Disjunction:**
- $F ::= \ldots | F_0 \lor F_1$
- concretization:
  $$\gamma(F_0 \lor F_1) = \gamma(F_0) \cup \gamma(F_1)$$

**Spatial implication (aka, magic wand):**
- $F ::= \ldots | F_0 \models F_1$
- concretization:
  $$\gamma(F_0 \models F_1) = \left\{ (e, h) \mid \forall h_0 \in \mathbb{H}, (e, h_0) \in \gamma(F_0) \implies (e, h \oplus h_0) \in \gamma(F_1) \right\}$$
- very powerful connector to describe **structure segments**, used in complex SL proofs
Separation logic

Summary of the main separation logic constructions seen so far:

Separation logic main connectors

\[
\begin{align*}
\gamma(\mathsf{emp}) &= \mathbb{E} \times \{[]\} \\
\gamma(\mathsf{TRUE}) &= \mathbb{E} \times \mathbb{H} \\
\gamma(1 \mapsto v) &= \{(e, [[1](e, h) \mapsto v)) \mid e \in \mathbb{E}\} \\
\gamma(F_0 \ast F_1) &= \{(e, h_0 \otimes h_1) \mid (e, h_0) \in \gamma(F_0) \land (e, h_1) \in \gamma(F_1)\} \\
\gamma(F_0 \land F_1) &= \gamma(F_0) \cap \gamma(F_1) \\
\gamma(F_0 \lor F_1) &= \gamma(F_0) \cup \gamma(F_1) \\
\gamma(F_0 \rightarrow F_1) &= \{(e, h) \mid \forall h_0 \in \mathbb{H}, (e, h_0) \in \gamma(F_0) \Rightarrow (e, h \otimes h_0) \in \gamma(F_1)\}
\end{align*}
\]

Concretization of pure formulas is standard

How does this help for program reasoning?
Separation logic triple

**Program proofs based on Hoare triples**

- **Notation:** \( \{F\} p \{F'\} \) if and only if:
  \[
  \forall s, s' \in \mathcal{S}, \ s \in \gamma(F) \land s' \in \llbracket p \rrbracket(s) \implies s' \in \gamma(F')
  \]

- **Application:** **formalize proofs of programs**

A few rules (straightforward proofs):

\[
\begin{align*}
\{&x \mapsto ?\} x := e &\{&x \mapsto e\} \quad \text{mutation} \\
F_0 \implies F'_0 & \quad \{F'_0\} b \{F'\} & F'_1 \implies F_1 \quad \text{consequence} \\
& \{F_0\} b \{F_1\} \\
\quad \quad \text{x does not appear in F} & \quad \{&x \mapsto ? \ast F\} x := e \{&x \mapsto e \ast F\} \quad \text{mutation-2}
\end{align*}
\]

(we assume that \( e \) does not allocate memory)
The frame rule

What about the resemblance between rules “mutation” and “mutation-2”?

Theorem: the frame rule

\[
\begin{align*}
\begin{array}{c}
\{F_0\} b \{F_1\} \\
\text{freevar}(F) \cap \text{write}(b) = \emptyset
\end{array} \\
\text{frame}
\end{align*}
\]

- Proof by induction on the logical rules on program statements, i.e., essentially a large case analysis
  (see biblio for a more complete set of rules)
- Rules are proved by case analysis on the program syntax

The frame rule allows to reason locally about programs

Ref
A program with intermittent invariants, derived using the frame rule, since each step impacts a disjoint region:

```c
int i;
int * x;
int * y;
{i ←? * &x ←? * &y ←?}
x = &i;
{i ←? * &x ←&i * &y ←?}
y = &i;
{i ←? * &x ←&i * &y ←&i}
*x = 42;
{i ←42 * &x ←&i * &y ←&i}
```

Many other program proofs done using separation logic
e.g., verification of the Deutsch-Shorr-Waite algorithm (biblio)
Summarization and inductive definitions

What do we still miss?
So far, formulas denote **fixed sets of cells**
Thus, no summarization of unbounded regions...

- **Example** all lists pointed to by `x`, such as:

  - How to precisely abstract these stores with a **single formula**
    *i.e.*, no infinite disjunction?

```
&x 0x0
&x 0x0
&x 0x0
&x 0x0
```
Inductive definitions in separation logic

List definition

\[ \alpha \cdot \text{list} := \begin{align*} & \alpha = 0 \land \text{emp} \\ \lor & \alpha \neq 0 \land \alpha \cdot \text{next} \leftrightarrow \delta \ast \alpha \cdot \text{data} \leftrightarrow \beta \ast \delta \cdot \text{list} \end{align*} \]

- Formula abstracting our set of structures:
  \[ \&x \leftrightarrow \alpha \ast \alpha \cdot \text{list} \]
- **Summarization:**
  this formula is finite and describe infinitely many heaps
- **Concretization:** next slide...

Practical implementation in verification/analysis tools

- **Verification:** hand-written definitions
- **Analysis:** either built-in or user-supplied, or partly inferred
Concretization by unfolding

Intuitive semantics of inductive predicates

- Inductive predicates can be **unfolded**, by **unrolling their definitions**
  
  Syntactic unfolding is noted $\map U$

- A formula $F$ with inductive predicates describes all stores described by all formulas $F'$ such that $F \map U F'$

Example:

- Let us start with $x \mapsto \alpha_0 \ast \alpha_0 \cdot \text{list}$; we can unfold it as follows:
  
  $\&x \mapsto \alpha_0 \ast \alpha_0 \cdot \text{list}$
  
  $\map U \quad \&x \mapsto \alpha_0 \ast \alpha_0 \cdot \text{next} \mapsto \alpha_1 \ast \alpha_0 \cdot \text{data} \mapsto \beta_1 \ast \alpha_1 \cdot \text{list}$
  
  $\map U \quad \&x \mapsto \alpha_0 \ast \alpha_0 \cdot \text{next} \mapsto \alpha_1 \ast \alpha_0 \cdot \text{data} \mapsto \beta_1 \ast \text{emp} \land \alpha_1 = 0x0$

- We get the concrete state below:
Example: tree

- Example:

\[ \begin{align*}
\alpha \cdot \text{tree} & := \alpha = 0 \land \text{emp} \\
\vee \quad \alpha \neq 0 \land (\alpha \cdot \text{left} \rightarrow \beta) \land (\alpha \cdot \text{right} \rightarrow \delta) \\
\ast \beta \cdot \text{tree} \ast \delta \cdot \text{tree}
\end{align*} \]
**Example: doubly linked list**

- **Example:**

  ![Doubly Linked List Diagram]

  *Inductive definition*

  - We need to propagate the `prev` pointer as an additional parameter:

    \[
    \alpha \cdot \text{dll}(\delta) := \begin{align*}
    & \alpha = 0 \land \text{emp} \\
    & \lor \quad \alpha \neq 0 \land \alpha \cdot \text{next} \leftrightarrow \beta \ast \alpha \cdot \text{prev} \leftrightarrow \delta \\
    & \ast \beta \cdot \text{dll}(\alpha)
    \end{align*}
    \]
Example: sortedness

- **Example**: sorted list

\[
\begin{align*}
\&x & & \bullet & & \rightarrow & & \\
& & 8 & & \rightarrow & & 9 & & \rightarrow & & 0x0 \\
& & & & & & 33
\end{align*}
\]

**Inductive definition**

- Each element should be greater than the previous one
- The first element simply needs be greater than \(-\infty\)...
- We need to propagate the lower bound, using a scalar parameter

\[
\alpha \cdot \text{lsort}_{\text{aux}}(n) := \begin{cases} 
\alpha = 0 \land \text{emp} \\
\lor \alpha \neq 0 \land n \leq \beta \land \alpha \cdot \text{next} \leftrightarrow \delta \\
\ast \alpha \cdot \text{data} \leftrightarrow \beta \ast \delta \cdot \text{lsort}_{\text{aux}}(\beta)
\end{cases}
\]

\[
\alpha \cdot \text{lsort}() := \alpha \cdot \text{lsort}_{\text{aux}}(-\infty)
\]
Outline

1. Introduction
2. Separation Logic
3. A shape abstract domain relying on separation
4. Standard static analysis algorithms
5. Combining shape and value abstractions
6. Conclusion
Design of an abstract domain

A lot of things are missing to turn SL into an abstract domain

Set of logical predicates:
- separation logic formulas are very expressive
e.g., arbitrary alternations of $\land$ and $\ast$
- such expressiveness is not necessarily required in static analysis

Representation:
- unstructured formulas can be represented as ASTs,
  but this representation is not easy to manipulate efficiently
- intuition over memory states typically involves graphs

Analysis algorithms:
- inference of “optimal” invariants in SL, with numerical predicates obviously not computable
Basic abstraction: structures and their contents (1/2)

- **Concrete memory states**
  - very **low level** description
    - numeric offsets / field names
  - pointers, numeric values:
    - raw sequences of bits

\[
\begin{align*}
&(x \cdot n) = 0x...a0 & 17 \\
&(x \cdot d) = 0x...a4 & 0x...b0 \\
&(y \cdot n) = 0x...b0 & 17 \\
&(y \cdot d) = 0x...b4 & 0x0
\end{align*}
\]
Basic abstraction: structures and their contents (1/2)

- **Concrete memory states**

- **Abstraction of values** into **symbolic variables** (nodes)

  - characterized by **valuation** $\nu$
  - $\nu$ maps **symbolic variables** into **concrete addresses**

  $\nu(\alpha_0) = 0x\ldots a_0$
  $\nu(\alpha_1) = 17$
  $\nu(\alpha_2) = 0x\ldots b_0$
  $\nu(\alpha_3) = 17$
  $\nu(\alpha_4) = 0x0$
Basic abstraction: structures and their contents (1/2)

- **Concrete memory states**

- **Abstraction of values into symbolic variables / nodes**

- **Abstraction of regions into points-to edges**
Basic abstraction: structures and their contents (1/2)

- **Concrete memory states**

- **Abstraction of values into symbolic variables** / nodes

- **Abstraction of regions into points-to edges**

  ![Diagram](image)

  \[ \gamma_{sh}(G) = \{(h, \nu) \mid \ldots\} \]

  Valuation \( \nu \) plays an important role to combine abstraction...
Structure of shape graphs

Valuations bridge the gap between nodes and values

**Symbolic variables / nodes** and intuitively abstract concrete values:

**Symbolic variables**

We let $\mathbb{V}$ denote a countable set of **symbolic variables**; we usually let them be denoted by Greek letters in the following: $\mathbb{V} = \{\alpha, \beta, \delta, \ldots\}$

When concretizing a shape graph, we need to **characterize how the concrete instance evaluates each symbolic variable**, which is the purpose of the **valuation functions**:

**Valuations**

A **valuation** is a function from **symbolic variables** into **concrete values** (and is often denoted by $\nu$): $\text{Val} : \mathbb{V} \rightarrow \mathbb{V}$

Note that valuations treat **in the same way addresses and raw values**
A shape abstract domain relying on separation

Structure of shape graphs

Distinct edges describe separate regions

In particular, if we split a graph into two parts:

Separating conjunction

\[ \gamma_{\text{sh}}(S_0^\# \ast S_1^\#) = \{(h_0 \otimes h_1, \nu) \mid (h_0, \nu) \in \gamma_{\text{sh}}(S_0^\#) \land (h_1, \nu) \in \gamma_{\text{sh}}(S_1^\#)\} \]

Similarly, when considering the empty set of edges, we get the empty heap (where \( \mathbb{V}^\# \) is the set of nodes):

\[ \gamma_{\text{sh}}(\text{emp}) = \{((\emptyset, \nu) \mid \nu : \mathbb{V}^\# \rightarrow \mathbb{V}\} \]
A shape abstract domain relying on separation

Abstraction of contiguous regions

A single points-to edge represents one heap cell

A points-to edge encodes basic points to predicate in separation logic:

Points-to edges

- **Syntax**
  - Graph edge
  
  \[
  \alpha \xrightarrow{f} \beta
  \]
  
  - Separation logic formula
    
    \[
    \alpha \cdot f \leftrightarrow \beta
    \]
    
  - Concrete view
    
    \[
    \begin{array}{c}
    \nu(\alpha) \\
    \text{offset}(f) \\
    \nu(\beta)
    \end{array}
    \]

- **Concretization:**
  
  \[
  \gamma_{sh}(\alpha \cdot f \leftrightarrow \beta) = \{ ([\nu(\alpha) + \text{offset}(f) \leftrightarrow \nu(\beta)], \nu) \mid \nu : \{\alpha, \beta, \ldots\} \rightarrow \mathbb{N} \}
  \]
A shape abstract domain relying on separation

Abstraction of contiguous regions

Contiguous regions are described by adjacent points-to edges

To describe blocks containing series of cells (e.g., in a C structure), shape graphs utilize several outgoing edges from the node representing the base address of the block.

Field splitting model

- Separation impacts edges / fields, not pointers
  
  ![Diagram showing separation impacts edges / fields, not pointers](image)

- Shape graph accounts for both abstract states below:

  ![Shape graph diagram](image)

In other words, in a field splitting model, separation:

- asserts addresses are distinct
- says nothing about contents
Abstraction of the environment

\[ m :: = (e, h) \]

Environments bind variables to their (concrete / abstract) address

\&x = \& (x \cdot n) = 0x...a0
\& (x \cdot d) = 0x...a4
\&y = \& (y \cdot n) = 0x...b0
\& (y \cdot d) = 0x...b4

Abstract environments

- An abstract environment is a function \( e^\# \) from variables to symbolic nodes
- The concretization extends as follows:

\[ \gamma_{\text{mem}}(e^\#, S^\#) = \{(e, h, \nu) \mid (h, \nu) \in \gamma_{\text{sh}}(S^\#) \land e = \nu \circ e^\#\} \]
A shape abstract domain relying on separation

Basic abstraction: summarization

Set of all lists of any length:

Well-founded list inductive def.

\[
\begin{align*}
\alpha \cdot \text{list} := & \quad (\text{emp} \land \alpha = 0x0) \\
& \lor \quad (\alpha \cdot d \mapsto \beta_0 \land \alpha \cdot n \mapsto \beta_1 \\
& \land \beta_1 \cdot \text{list} \land \alpha \neq 0x0)
\end{align*}
\]

well-founded predicate

Inductive summary predicates

Concretization based on unfolding and least-fixpoint:

- \( \mathcal{U} \) replaces an \( \alpha \cdot \text{list} \) predicate with one of its premises

- \( \gamma(S', F) = \bigcup \{ \gamma(S'_u, F_u) \mid (S', F) \xrightarrow{\mathcal{U}} (S'_u, F_u) \} \)
Inductive structures: a few instances

As before, many interesting inductive predicates encode nicely into graph inductive definitions:

- **More complex shapes: trees**

  ![Tree Diagram]

- **Relations among pointers: doubly-linked lists**

  ![Doubly-Linked List Diagram]

- **Relations between pointers and numerical: sorted lists**

  ![Sorted List Diagram]
Inductive segments

A frequent pattern:

A first attempt:

- \(x\) points to a list, so \(\&x \mapsto \alpha_1 \cdot \alpha_2 \cdot \text{list}\) holds
- \(y\) points to a list, so \(\&y \mapsto \beta_1 \cdot \beta_2 \cdot \text{list}\) holds

However, the following **does not hold**

\[ \&x \mapsto \alpha_1 \cdot \alpha_2 \cdot \text{list} \land \&y \mapsto \beta_1 \cdot \beta_2 \cdot \text{list} \]

Why? **violation of separation!**

A second attempt:

\[ (\&x \mapsto \alpha_1 \cdot \alpha_2 \cdot \text{list} \land \text{TRUE}) \land (\&y \mapsto \beta_1 \cdot \beta_2 \cdot \text{list} \land \text{TRUE}) \]

Why is it still not all that good? **relation lost!**
Inductive segments

A frequent pattern:

Could be expressed directly as an inductive with a parameter:

\[
\alpha \cdot \text{list}_\text{endp}(\pi) \ := \ (\text{emp}, \alpha = \pi) \\
\ | \ (\alpha \cdot \text{next} \mapsto \beta_0 \ * \ \alpha \cdot \text{data} \mapsto \beta_1 \\
\ * \ \beta_0 \cdot \text{list}_\text{endp}(\pi), \alpha \neq 0)
\]

This definition straightforwardly derives from list
Thus, we make segments part of the fundamental predicates of the domain

Multi-segments: possible, but harder for analysis
Shape graphs and separation logic

Semantic preserving translation $\Pi$ of graphs into separation logic formulas:

<table>
<thead>
<tr>
<th>Graph $S^# \in \mathbb{D}_{sh}^#$</th>
<th>Translated formula $\Pi(S^#)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\langle \alpha \rightarrow \beta \rangle_f$</td>
<td>$\alpha \cdot f \leftrightarrow \beta$</td>
</tr>
<tr>
<td>$S_0^#$</td>
<td>$\Pi(S_0^#) \ast \Pi(S_1^#)$</td>
</tr>
<tr>
<td>$\langle \alpha \rightarrow \text{list} \rangle$</td>
<td>$\alpha \cdot \text{list}$</td>
</tr>
<tr>
<td>$\langle \alpha \rightarrow \text{list} \text{list} \rightarrow \delta \rangle$</td>
<td>$\alpha \cdot \text{list}_{-\text{endp}}(\delta)$</td>
</tr>
<tr>
<td>other inductives and segments</td>
<td>similar</td>
</tr>
</tbody>
</table>

Note that:
- shape graphs can be encoded into separation logic formula
- the opposite is usually not true

Value information:
- discussed in the next course
- intuitively, assume we maintain numerical information next to shape graphs
Outline

1 Introduction

2 Separation Logic

3 A shape abstract domain relying on separation

4 Standard static analysis algorithms
   - Overview of the analysis
   - Post-conditions and unfolding
   - Folding: widening and inclusion checking
   - Abstract interpretation framework: assumptions and results
   - Comparing Separation Logic and Three-Valued logic abstractions

5 Combining shape and value abstractions

6 Conclusion
A list insertion function:

```c
list * l assumed to point to a list
list * t assumed to point to a list element

list * c = l;
while (c != NULL && c -> next != NULL && (...)){
    c = c -> next;
}
t -> next = c -> next;
c -> next = t;
```

Result of the (interprocedural) analysis

- **Over-approximations** of reachable concrete states
  
  *e.g.*, after the insertion:

- list inductive structure def.
- Abstract precondition:
Transfer functions

Abstract interpreter design

- **Follows the semantics** of the language under consideration
- The abstract domain should provide **sound transfer functions**

Transfer functions:

- **Assignment**: $x \rightarrow f = y \rightarrow g$ or $x \rightarrow f = e_{\text{arith}}$
- **Test**: analysis of conditions (if, while)
- Variable **creation** and **removal**
- **Memory management**: `malloc`, `free`

Abstract operators:

- **Join** and **widening**: over-approximation
- **Inclusion checking**: check stabilization of abstract iterates

Should be **sound** *i.e.*, not forget any concrete behavior
Abstract operations

Denotational style abstract interpreter

- Concrete **denotational semantics** \( \llbracket b \rrbracket : S \rightarrow \mathcal{P}(S) \)
- **Abstract post-condition** \( \llbracket b \rrbracket^S(S) \), computed by the analysis:

\[
s \in \gamma(S) \implies \llbracket b \rrbracket(s) \subseteq \gamma(\llbracket b \rrbracket^S(S))
\]

**Analysis by induction on the syntax** using **domain operators**

\[
\begin{align*}
\llbracket b_0; b_1 \rrbracket^S(S) &= \llbracket b_1 \rrbracket^S \circ \llbracket b_0 \rrbracket^S(S) \\
\llbracket 1 = e \rrbracket^S(S) &= \text{assign}(1, e, S) \\
\llbracket 1 = \text{malloc}(n) \rrbracket^S(S) &= \text{alloc}(1, n, S) \\
\llbracket \text{free}(1) \rrbracket^S(S) &= \text{free}(1, n, S) \\
\llbracket \text{if}(e) \ b_t \ \text{else} \ b_f \rrbracket^S(S) &= \begin{cases} \\
\text{join}(\llbracket b_t \rrbracket^S(\text{test}(e, S))), & \\
\llbracket b_f \rrbracket^S(\text{test}(e = \text{false}, S))) & \\
\end{cases} \\
\llbracket \text{while}(e)b \rrbracket^S(S) &= \text{test}(e = \text{false}, \text{lfp}_SF^\#) \\
\text{where, } F^\# : S_0 &\mapsto \llbracket b \rrbracket^S(\text{test}(e, S_0))
\end{align*}
\]
The algorithms underlying the transfer functions

- **Unfolding**: cases analysis on summaries

  \[
  \text{list} \quad \xrightarrow{y} \quad \text{list} \\
  \]

- **Abstract postconditions**, on “exact” regions, e.g. insertion

  \[
  \text{next} \quad 0x0 \quad \xrightarrow{y} \quad \text{next} \\
  \]

- **Widening**: builds summaries and ensures termination

  \[
  \text{list} \quad \xrightarrow{\nabla} \quad \text{list} \\
  \]

\[
\text{next} \quad \text{data} \quad \xrightarrow{y} \quad \text{next} \\
\]

\[
\text{next} \quad \text{data} \quad \xrightarrow{y} \quad \text{next} \\
\]

\[
\text{next} \quad \text{data} \quad \xrightarrow{y} \quad \text{next} \\
\]

Xavier Rival (INRIA)  
Shape analysis abstractions  
Jan, 17th, 2022  
42 / 94
Outline

1. Introduction
2. Separation Logic
3. A shape abstract domain relying on separation
4. Standard static analysis algorithms
   - Overview of the analysis
   - Post-conditions and unfolding
     - Folding: widening and inclusion checking
     - Abstract interpretation framework: assumptions and results
     - Comparing Separation Logic and Three-Valued logic abstractions
5. Combining shape and value abstractions
6. Conclusion
Analysis of an assignment in the graph domain

Steps for analyzing \( x = y \rightarrow \text{next} \) (local reasoning)

1. Evaluate \( l\text{-value} \) \( x \) into points-to edge \( \alpha \mapsto \beta \)
2. Evaluate \( r\text{-value} \) \( y \rightarrow \text{next} \) into node \( \beta' \)
3. Replace points-to edge \( \alpha \mapsto \beta \) with points-to edge \( \alpha \mapsto \beta' \)

With pre-condition:

- Step 1 produces \( \alpha_0 \mapsto \beta_0 \)
- Step 2 produces \( \beta_2 \)
- End result:

With pre-condition:

- Step 1 produces \( \alpha_0 \mapsto \beta_0 \)
- Step 2 fails
- Abstract state too abstract
- We need to refine it
Unfolding as a local case analysis

Unfolding principle

- **Case analysis**, based on the inductive definition
- Generates *symbolic disjunctions* (analysis performed in a *disjunction domain*, e.g., trace partitioning)

- Example, for lists:

  \[
  \begin{align*}
  \gamma_{sh}(S^\#) \subseteq \bigcup \{ \gamma_{sh}(S_0^\#) \mid S^\# \xrightarrow{\cup} S_0^\# \}
  \end{align*}
  \]

- **Numeric predicates**: next course on shape + value abstraction

Soundness: by definition of the concretization of inductive structures
Analysis of an assignment, with unfolding

**Principle**
- We have $\gamma_{sh}(\alpha \cdot \iota) = \bigcup \{\gamma_{sh}(S^\#) \mid \alpha \cdot \iota \xrightarrow{U} S^\#\}$
- Replace $\alpha \cdot \iota$ with a finite number of disjuncts and continue

**Disjunct 1:**
- $x = y \rightarrow \text{next}$
- Step 1 produces $\alpha_0 \mapsto \beta_0$
- **Step 2 fails:** Null pointer!
- In a **correct** program, would be ruled out by a **condition** $y \neq 0$
  - i.e., $\beta_1 \neq 0$ in $\mathbb{D}^\#_{\text{num}}$
  - $\text{if}(y \neq 0)$ ...

**Disjunct 2:**
- Step 1 produces $\alpha_0 \mapsto \beta_0$
- Step 2 produces $\beta_2$
- **End result:**
Unfolding and degenerated cases

```plaintext
assume(1 points to a dll)
c = 1;
① while(c ≠ NULL && condition)
c = c -> next;
② if(c ≠ 0 && c -> prev ≠ 0)
c = c -> prev → prev;
```

- at ①: \( \alpha_0 \xrightarrow{dll(\delta_1)} l, c \)
- at ②: \( \alpha_0 \xrightarrow{dll(\delta_0)} \alpha_1 \xrightarrow{dll(\delta_1)} c \)

⇒ non trivial unfolding

Materialization of \( c \rightarrow prev \):

Segment splitting lemma: basis for segment unfolding

Materialization of \( c \rightarrow prev \rightarrow prev \):

Implementation issue: discover which inductive edge to unfold very hard!
Outline

1 Introduction

2 Separation Logic

3 A shape abstract domain relying on separation

4 Standard static analysis algorithms
   - Overview of the analysis
   - Post-conditions and unfolding
   - Folding: widening and inclusion checking
   - Abstract interpretation framework: assumptions and results
   - Comparing Separation Logic and Three-Valued logic abstractions

5 Combining shape and value abstractions

6 Conclusion
Need for a folding operation

Back to the **list traversal** example:

**First iterates** in the loop:

- **at iteration 0** (before entering the loop):
  
  \[
  \begin{align*}
  \alpha_0 & \rightarrow \text{list} \\
  l & \rightarrow \alpha_1 \\
  c & \rightarrow \alpha_0
  \end{align*}
  \]

- **at iteration 1**:
  
  \[
  \begin{align*}
  \alpha_0 & \rightarrow \text{list} \\
  \text{next} & \rightarrow \alpha_1 \\
  c & \rightarrow \alpha_0 \\
  \text{data} & \rightarrow \alpha_1
  \end{align*}
  \]

- **at iteration 2**:
  
  \[
  \begin{align*}
  \alpha_0 & \rightarrow \text{list} \\
  \text{next} & \rightarrow \alpha_1 \\
  \text{next} & \rightarrow \alpha_2 \\
  \text{data} & \rightarrow \alpha_1, \alpha_2
  \end{align*}
  \]

The analysis **unfolds**, but **never folds**:

- **assume**(l points to a list)
  
  \[
  \begin{align*}
  c & = l; \\
  \text{while}(c \neq \text{NULL}) \{ \\
  c & = c \rightarrow \text{next}; \\
  \}
  \end{align*}
  \]

- **How to guarantee** **termination** of the analysis?
- **How to introduce segment edges** / perform **abstraction**?
Widening

- The lattice of shape abstract values has infinite height
- Thus iteration sequences may not terminate

Definition of a widening operator \( \bigtriangleup \)

- **Over-approximates join:**
  \[
  \begin{align*}
  \gamma(X^\#) & \subseteq \gamma(X^\# \bigtriangleup Y^\#) \\
  \gamma(Y^\#) & \subseteq \gamma(X^\# \bigtriangleup Y^\#)
  \end{align*}
  \]

- **Enforces termination:** for all sequence \((X_n^\#)_{n \in \mathbb{N}}\), the sequence \((Y_n^\#)_{n \in \mathbb{N}}\) defined below is ultimately stationary
  \[
  \begin{align*}
  \forall n \in \mathbb{N}, \quad Y_{n+1}^\# & = Y_n^\# \bigtriangleup X_{n+1}^\#
  \end{align*}
  \]
Canonicalization

Upper closure operator

\[ \rho : D^\# \rightarrow D^\#_{can} \subseteq D^\# \] is an **upper closure operator** (uco) iff it is monotone, extensive and idempotent.

Canonicalization

- **Disjunctive completion:** \( D^\#_{\land} = \) finite disjunctions over \( D^\# \)
- **Canonicalization operator** \( \rho \lor \) defined by \( \rho \lor : D^\#_{\land} \rightarrow D^\#_{can \lor} \) and
  \[ \rho \lor (X^\#) = \{ \rho(x^\#) \mid x^\# \in X^\# \} \] where \( \rho \) is an uco and \( D^\#_{can} \) is finite

- Canonicalization is used in **many shape analysis tools**
- **Easier to compute** but **less powerful** than widening: does not exploit history of computation
Weakening: definition

To design **inclusion test**, **join** and **widening** algorithms, we first study a more general notion of **weakening**:

**Weakening**

We say that \( S_0 \) **can be weakened into** \( S_1 \) if and only if

\[
\forall (\mathcal{h}, \nu) \in \gamma_{sh}(S_0), \ \exists \nu' \in \text{Val}, \ (\mathcal{h}, \nu') \in \gamma_{sh}(S_1)
\]

We then note \( S_0 \preceq S_1 \)

**Applications:**

- **inclusion test** (comparison) inputs \( S_0, S_1 \); if returns true \( S_0 \preceq S_1 \)
- **canonicalization** (unary weakening) inputs \( S_0 \) and returns \( \rho(S_0) \) such that \( S_0 \preceq \rho(S_0) \)
- **widening / join** (binary weakening ensuring termination or not) inputs \( S_0, S_1 \) and returns \( S_{up} \) such that \( S_i \preceq S_{up} \)
Weakening: example

We consider $S_0$ defined by:

$$
\begin{align*}
\alpha_0 & \rightarrow \alpha_1 \\
& \rightarrow \alpha_2
\end{align*}
$$

and $S_1$ defined by:

$$
\begin{align*}
\beta_0 & \rightarrow \beta_1 \\
& \rightarrow \beta_2
\end{align*}
$$

Then, we have the weakening $S_0 \preceq S_1$ up-to a renaming in $S_1$:

$$
\Psi : \begin{align*}
\beta_0 & \mapsto \alpha_0 \\
\beta_1 & \mapsto \alpha_1
\end{align*}
$$

- weakening up-to renaming is generally required as graphs do not have the same name space
- formalized a bit later...
\[
\begin{array}{c}
\Sigma_0 \\
\beta_3 \xrightarrow{+0} \beta_1 \\
\xrightarrow{+4} \beta_2 \\
\Sigma_1 \\
\beta_1 \xrightarrow{\psi} \alpha_1 \\
\beta_2 \xrightarrow{\psi} \alpha_2
\end{array}
\]
Local weakening: separating conjunction rule

We can apply the local reasoning principle to weakening

If \( S_0 \preceq S_{0,\text{weak}} \) and \( S_1 \preceq S_{1,\text{weak}} \) then:

\[
\begin{array}{ccc}
\alpha_0 & S_0 & \alpha_1 \\
\alpha_2 & S_1 & \alpha_3
\end{array}
\preceq
\begin{array}{ccc}
\alpha_0 & S_{0,\text{weak}} & \alpha_1 \\
\alpha_2 & S_{1,\text{weak}} & \alpha_3
\end{array}
\]

Separating conjunction rule (\( \preceq \ast \))

Let us assume that

- \( S_0 \) and \( S_1 \) have distinct set of source nodes
- we can weaken \( S_0 \) into \( S_{0,\text{weak}} \)
- we can weaken \( S_1 \) into \( S_{1,\text{weak}} \)

then:

\[
\text{we can weaken } S_0 \ast S_1 \text{ into } S_{0,\text{weak}} \ast S_{1,\text{weak}}
\]
Local weakening: unfolding rule, identity rule

**Weakening unfolded region (≲_{U})**

Let us assume that \( S_0 \xrightarrow{U} S_1 \). Then, by definition of the concretization of unfolding

we can weaken \( S_1 \) into \( S_0 \)

- the proof follows from the definition of unfolding
- it can be applied locally, on graph regions that differ due to unfolding of inductive definitions

**Identity weakening (≲_{Id})**

we can weaken \( S^\# \) into \( S^\# \)

- the proof is trivial:
  \[
  \gamma_{sh}(S^\#) \subseteq \gamma_{sh}(S^\#)
  \]
- on itself, this principle is not very useful, but it can be applied locally, and combined with (≲_{U}) on graph regions that are not equal
Local weakening: example

By rule \((<_{\text{Id}})\):

Thus, by rule \((<_{\mathcal{U}})\):

Additionally, by rule \((<_{\text{Id}})\):

Thus, by rule \((<_{\ast})\):
Inclusion checking rules in the shape domain

Graphs to compare have distinct sets of nodes, thus inclusion check should carry out a valuation transformer $\Psi : \forall^\#(S_1^\#) \rightarrow \forall^\#(S_0^\#)$ (important when dealing also with content values)

Using (and extending) the weakening principles, we obtain the following rules (considering only inductive definition list, though these rules would extend to other definitions straightforwardly):

- **Identity rules:**
  \[
  \forall i, \Psi(\beta_i) = \alpha_i \implies \forall i, \Psi(\beta) = \alpha \implies \alpha_0 \cdot \text{list} \ni \beta_0 \cdot \text{list} \ni \beta_0 \cdot \text{list_endp}^\#(\beta_1)
  \]

- **Rules on inductives:**
  \[
  \forall i, \Psi(\beta_i) = \alpha \implies \text{emp} \ni \beta_0 \cdot \text{list_endp}^\#(\beta_1)
  \]
  \[
  \text{if } \beta_1 \text{ fresh, } \Psi' = \Psi[\beta_1 \mapsto \alpha_1] \text{ and } \Psi(\beta_0) = \alpha_0 \text{ then, } \\
  S_0 \ni \beta_1 \cdot \text{list} \implies \alpha_0 \cdot \text{list_endp}^\#(\alpha_1) \ni S_0 \ni \beta_0 \cdot \text{i}
  \]
Inclusion checking algorithm

Comparison of \((e^0_0, S^0_0)\) and \((e^1_1, S^1_1)\)

1. Start with \(\Psi\) defined by \(\Psi(\beta) = \alpha\) if and only if there exists a variable \(x\) such that \(e^0_0(x) = \alpha \land e^1_1(x) = \beta\)

2. Iteratively apply local rules, and extend \(\Psi\) when needed

3. Return true when both shape graphs become empty

- The first step ensures both environments are consistent

This algorithm is sound:

Soundness

\((e^0_0, S^0_0) \sqsubseteq^\# (e^1_1, S^1_1) \implies \gamma(e^0_0, S^0_0) \subseteq \gamma(e^1_1, S^1_1)\)
Over-approximation of union

The principle of join and widening algorithm is similar to that of $\subseteq^*$:

- It can be computed \textbf{region by region}, as for weakening in general:
  \begin{equation}
  \text{If } \forall i \in \{0, 1\}, \forall s \in \{\text{lft, rgh}\}, \quad S_{i,s}^\# \leq S_s^\#,
  \end{equation}

\begin{center}
\begin{tikzpicture}[scale=0.8, every node/.style={scale=0.8}]
  \node[shape=rectangle, draw] (S0) at (0,0) {$S_{0,\text{lft}}^\#$};
  \node[shape=rectangle, draw] (S1) at (1,0) {$S_{1,\text{lft}}^\#$};
  \node[shape=rectangle, draw] (S0rgh) at (0,-1) {$S_{0,\text{rgh}}^\#$};
  \node[shape=rectangle, draw] (S1rgh) at (1,-1) {$S_{1,\text{rgh}}^\#$};
  \draw[->] (S0) -- (S1) node[midway, above] {$\psi$};
  \draw[->] (S0) -- (S0rgh) node[midway, left] {$\alpha_0$};
  \draw[->] (S1) -- (S1rgh) node[midway, left] {$\alpha_1$};
  \draw[->] (S0rgh) -- (S1rgh) node[midway, left] {$\beta_0$};
  \draw[->] (S1rgh) -- (S0) node[midway, left] {$\beta_1$};
  \end{tikzpicture}
\end{center}

\text{The partitioning of inputs / different nodes sets requires a \textbf{node correspondence function}}

\begin{equation}
\psi : \forall^\#(S_{\text{lft}}^\#) \times \forall^\#(S_{\text{rgh}}^\#) \rightarrow \forall^\#(S^\#)
\end{equation}

- The computation of the shape join progresses by the application of \textbf{local join rules}, that produce a \textbf{new (output) shape graph, that weakens both inputs}

Over-approximation of union: syntactic identity rules

In the next few slides, we focus on $\nabla$
though the abstract union would be defined similarly in the shape domain.

Several rules derive from $(\preceq_{\text{Id}})$:

- If $S_{\text{left}}^\# = \alpha_0 \cdot f \mapsto \alpha_1$
  and $S_{\text{left}}^\# = \beta_0 \cdot f \mapsto \beta_1$
  and $\Psi(\alpha_0, \beta_0) = \delta_0$, $\Psi(\alpha_1, \beta_1) = \delta_1$, then:

  $S_{\text{left}}^\# \nabla S_{\text{right}}^\# = \delta_0 \cdot f \mapsto \delta_1$

- If $S_{\text{left}}^\# = \alpha_0 \cdot \text{list}$
  and $S_{\text{left}}^\# = \beta_0 \cdot \text{list}_1$
  and $\Psi(\alpha_0, \beta_0) = \delta_0$, then:

  $S_{\text{left}}^\# \nabla S_{\text{right}}^\# = \delta_0 \cdot \text{list}$
Over-approximation of union: segment introduction rule

Rule

\[
\begin{align*}
\text{if } & \quad S_{\text{left}} \subseteq S_{\text{right}} \\
\text{then } & \quad \begin{cases} 
S_{\text{left}} \Delta S_{\text{right}} = \delta_0 \\
(\alpha, \beta_0) \leftrightarrow \delta_0 \\
(\alpha, \beta_1) \leftrightarrow \delta_1
\end{cases}
\end{align*}
\]

Application to list traversal, at the end of iteration 1:

- before iteration 0:

- end of iteration 0:

- join, before iteration 1:
Over-approximation of union: segment extension rule

Rule

\[
\text{if } S^\#_{\text{left}} \leq S^\#_{\text{right}} \quad \text{then} \quad \begin{cases}
S^\#_{\text{left}} \bigtriangledown S^\#_{\text{right}} = \delta_0 \\
(\alpha_0, \beta_0) \leftrightarrow \delta_0 \\
(\alpha_1, \beta_1) \leftrightarrow \delta_1
\end{cases}
\]

Application to list traversal, at the end of iteration 1:

- previous invariant before iteration 1:

- end of iteration 1:

- join, before iteration 1:
Over-approximation of union: rewrite system properties

- Comparison, canonicalization and widening algorithms can be considered **rewriting systems over tuples of graphs**

- **Success configuration**: weakening applies on all components, *i.e.*, the inputs are fully "consumed" in the weakening process

- **Failure configuration**: some components **cannot be weakened** *i.e.*, the algorithm should return the conservative answer (*i.e.*, $\top$)

**Termination**

- The systems are **terminating**

- This ensures comparison, canonicalization, widening are **computable**

**Non confluence!**

- The results depends on the order of application of the rules

- Implementation requires the choice of an **adequate strategy**
Over-approximation of union in the combined domain

Widening of \((e_0^\#, S_0^\#)\) and \((e_1^\#, S_1^\#)\)

1. define \(\Psi, e\) by \(\Psi(\alpha, \beta) = e(x) = \delta\) (where \(\delta\) is a fresh node) if and only if \(e_0^\#(x) = \alpha \land e_1^\#(x) = \beta\)

2. iteratively apply join local rules, and extend \(\Psi\) when new relations are inferred (for instance for points-to edges)

3. return the result obtained when all regions of both inputs are approximated in the output graph

This algorithm is sound:

**Soundness**

\[ \gamma(e_0^\#, S_0^\#) \cup \gamma(e_1^\#, S_1^\#) \subseteq \gamma(e^\#, S^\#) \]

Widening also enforces **termination** (it only introduces segments, and the growth induced by the introduction of segments is bounded)
Outline

1 Introduction

2 Separation Logic

3 A shape abstract domain relying on separation

4 Standard static analysis algorithms
   - Overview of the analysis
   - Post-conditions and unfolding
   - Folding: widening and inclusion checking
   - Abstract interpretation framework: assumptions and results
   - Comparing Separation Logic and Three-Valued logic abstractions

5 Combining shape and value abstractions

6 Conclusion
Assumptions

**What assumptions do we make?**
**How do we prove soundness of the analysis of a loop?**

- **Assumptions in the concrete level**, and for block b:
  
  \[ (\mathcal{P}(\mathcal{M}), \subseteq) \] is a complete lattice, hence a CPO
  
  \[ F : \mathcal{P}(\mathcal{M}) \rightarrow \mathcal{P}(\mathcal{M}) \] is the concrete semantic ("post") function of \( b \)

  thus, the concrete semantics writes down as \([b] = \text{lfp}_0 F\)

- **Assumptions in the abstract level:**

  \( M^\# \) set of abstract elements, no order a priori
  
  \( m^\# ::= (e^\#, S^\#) \) concretization
  
  \( \gamma_{\text{mem}} : M^\# \rightarrow \mathcal{P}(\mathcal{M}) \)
  
  \( F^\# : M^\# \rightarrow M^\# \) sound abstract semantic function

  \( i.e., \) such that \( F \circ \gamma_{\text{mem}} \subseteq \gamma_{\text{mem}} \circ F^\# \)

  \( \triangledown : M^\# \times M^\# \rightarrow M^\# \) widening operator, terminates, and such that

  \( \gamma_{\text{mem}}(m_0^\#) \cup \gamma_{\text{mem}}(m_1^\#) \subseteq \gamma_{\text{mem}}(m_0^\# \triangledown m_1^\#) \)
Computing a loop abstract post-condition

Loop abstract semantics

The abstract semantics of loop $\textbf{while}(\texttt{rand}())\{b\}$ is calculated as the limit of the sequence of abstract iterates below:

$$
\left\{
\begin{array}{c}
m^0 = \perp \\
m^k_{n+1} = m^k_n \sqcup F^k(m^k_n)
\end{array}
\right.
$$

Soundness proof:

- by induction over $n$, $\bigcup_{k \leq n} F^k(\emptyset) \subseteq \gamma_{\text{mem}}(m^k_n)$
- by the property of widening, the abstract sequence converges at a rank $N$: $\forall k \geq N$, $m^k_k = m^k_N$, thus

$$
\text{lfp}_\emptyset F = \bigcup_k F^k(\emptyset) \subseteq \gamma_{\text{mem}}(m^k_N)
$$
Discussion on the abstract ordering

How about the abstract ordering? We assumed **NONE** so far...

- **Logical ordering**, induced by concretization, used for **proofs**
  \[
  m_0^\# \sqsubseteq m_1^\# \quad ::= \quad "\gamma_{\text{mem}}(m_0^\#) \subseteq \gamma_{\text{mem}}(m_1^\#)"
  \]

- **Approximation of the logical ordering**, implemented as a function
  \[
  \text{is\_le}: M^\# \times M^\# \to \{\text{true}, \top\}, \text{used to test the convergence of abstract iterates}
  \]
  \[
  \text{is\_le}(m_0^\#, m_1^\#) = \text{true} \quad \Rightarrow \quad \gamma_{\text{mem}}(m_0^\#) \subseteq \gamma_{\text{mem}}(m_1^\#)
  \]

Abstract semantics is not assumed (and is actually most likely NOT) monotone with respect to either of these orders...

- Also, **computational ordering** would be used for **proving widening termination**
Outline

1 Introduction

2 Separation Logic

3 A shape abstract domain relying on separation

4 Standard static analysis algorithms
   - Overview of the analysis
   - Post-conditions and unfolding
   - Folding: widening and inclusion checking
   - Abstract interpretation framework: assumptions and results
   - Comparing Separation Logic and Three-Valued logic abstractions

5 Combining shape and value abstractions

6 Conclusion
Separation logic

Separation logic formulas (main connectors only)

\[ F ::= \begin{align*}
& \text{emp} \\
| & \text{TRUE} \\
| & l \mapsto l \\
| & F_0 \ast F_1 \\
| & F_0 \land F_1 \\
| & F_0 \rightarrow F_1
\end{align*} \]

Concretization:

\[ \gamma(\text{emp}) = E \times \{[]\} \]
\[ \gamma(\text{TRUE}) = E \times \mathbb{H} \]
\[ \gamma(l \mapsto v) = \{ (e, [l](e, h) \mapsto v) \mid e \in E \} \]
\[ \gamma(F_0 \ast F_1) = \{ (e, h_0 \otimes h_1) \mid (e, h_0) \in \gamma(F_0) \land (e, h_1) \in \gamma(F_1) \} \]
\[ \gamma(F_0 \land F_1) = \gamma(F_0) \cap \gamma(F_1) \]
\[ \gamma(F_0 \rightarrow F_1) = \text{exercise} \]

Program reasoning: frame rule and strong updates
## Shape graphs and separation logic

**Shape graphs**: provide an efficient data-structure to describe a *subset* of separation logic predicates, and do static analysis with them.

**Important addition**: inductive predicates.

**Semantic preserving translation** $\Pi$ of graphs into separation logic formulas:

<table>
<thead>
<tr>
<th>Graph $S^# \in \mathbb{D}^#_{sh}$</th>
<th>Translated formula $\Pi(S^#)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Shape Graph" /></td>
<td>$\alpha \cdot f \leftrightarrow \beta$</td>
</tr>
<tr>
<td><img src="image" alt="Translated Formula" /></td>
<td>$\Pi(S^#_0) \ast \Pi(S^#_1)$</td>
</tr>
<tr>
<td><img src="image" alt="List Predicate" /></td>
<td>$\alpha \cdot \text{list}$</td>
</tr>
<tr>
<td><img src="image" alt="List Segment" /></td>
<td>$\alpha \cdot \text{list_endp}(\delta)$</td>
</tr>
<tr>
<td>other inductives and segments</td>
<td>similar</td>
</tr>
</tbody>
</table>

Note that:
- shape graphs can be encoded into separation logic formula
- the opposite is usually not true
## Comparing the structure of abstract formulae

### Separation logic:

\[ F_0 \ast F_1 \ast \ldots \ast F_n \]

- first the heap is partitioned
- each region is described separately
- some of the \( F_i \) components may be summary predicates, describing unbounded regions
- reachability is implicit
- allows local reasoning

### Three valued logic:

\[ p_0 \land p_1 \land \ldots \land p_n \]

- first a conjunction of properties
- each predicate \( p_i \) may talk about any heap region
- no direct heap partitioning
- reachability can be expressed (natively)
- no local reasoning

---

**Two very different sets of predicates**

- one allows local reasoning, the other not
- the other way for reachability predicates
Summarization: one abstract cell, many concrete cells

Large / unbounded numbers of concrete cells need to be abstracted

- **Dynamic structures** (lists, trees) have an unknown and unbounded number of cells, hence require summarization
- We also needed summaries to deal with **arrays**

Summary

A **summary predicate** allows to describe an **unbounded number** of memory locations using a fixed, finite set of predicates

Principles underlying summarization:

- **in separation logic**: using inductive definitions for lists, trees... unbounded size of the summarized region is hidden in the **recursion**
- **in three-valued logic**: summary nodes + high level predicates (such as reachability) one summary node **carries the properties** of an unbounded number of cells
Concretize partially, update, abstract

For precise analysis, summaries need to be (temporarily) refined

Separation logic:

Local (partial) concretization

For materialization:

\[ S_{\text{pre}}^{\#} \xrightarrow{\text{ref}} S_{\text{post}}^{\#} \]

Global abstraction: widening

In both cases, two mechanisms are needed:
1. refine summaries
2. synthesize summaries

TVLA:

Focus, analyze, canonicalize

\[ S_{\text{pre}}^{\#} \xrightarrow{\text{abstract}} S_{\text{post}}^{\#} \]
Combining shape and value abstractions

Outline

1. Introduction

2. Separation Logic

3. A shape abstract domain relying on separation

4. Standard static analysis algorithms

5. Combining shape and value abstractions
   - Shape and value properties
   - Combined abstraction with cofibered abstract domain
   - Combined analysis algorithms

6. Conclusion
Common data-structures require to reason both about shape and data:

- **hybrid stores:** data stored next to inductive structures
- **list of even elements:**
  
  ![List of even elements diagram]
  
- **sorted list:**
  
  ![Sorted list diagram]
  
- **list with a length constraint**
- **tries:** binary trees with paths labelled with sequences of “0” and “1”
- **balanced trees:** red-black, AVL...

This part of the course:

- how to **express both shape and numerical properties**?
- how to **extend shape analysis algorithms**
Description of a sorted list

- **Example:** sorted list

![Sorted list diagram]

Inductive definition

- Each element should be greater than the previous one
- The first element simply needs be greater than $-\infty$...
- We need to propagate the lower bound, using a scalar parameter

\[
\alpha \cdot \text{lsort}_{\text{aux}}(n) := \begin{cases} 
\alpha = 0 \land \text{emp} \\
\lor \alpha \neq 0 \land n \leq \beta \land \alpha \cdot \text{next} \rightarrow \delta \\
\star \alpha \cdot \text{data} \rightarrow \beta \star \delta \cdot \text{lsort}_{\text{aux}}(\beta)
\end{cases}
\]

\[
\alpha \cdot \text{lsort}() := \alpha \cdot \text{lsort}_{\text{aux}}(-\infty)
\]
Combining shape and value abstractions

Adding value information (here, numeric)

Concrete numeric values appear in the valuation thus the abstracting contents boils down to abstracting $\nu$!

**Example**: all lists of length 2, sorted in the increasing order of data fields

**Memory abstraction**:

\[
\begin{array}{c}
\alpha_0 \\
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\end{array}
\begin{array}{cc}
+0 & +0 \\
+4 & +4 \\
\end{array}
\]

\[
\begin{array}{c}
0x\ldots a0 \\
0x\ldots a4 \\
0x\ldots b0 \\
0x\ldots b4 \\
\end{array}
\begin{array}{c}
7 \\
0x\ldots b0 \\
9 \\
0x0 \\
\end{array}
\]

$\nu : \alpha_1 \mapsto 7$

$\alpha_3 \mapsto 9$

$\ldots \mapsto \ldots$

\[
\begin{array}{c}
0x\ldots a0 \\
0x\ldots a4 \\
0x\ldots b0 \\
0x\ldots b4 \\
\end{array}
\begin{array}{c}
8 \\
0x\ldots b0 \\
12 \\
0x0 \\
\end{array}
\]

$\nu : \alpha_1 \mapsto 8$

$\alpha_3 \mapsto 12$

$\ldots \mapsto \ldots$

**Abstraction of valuations**: $\nu(\alpha_1) < \nu(\alpha_3)$, can be described by the constraint $\alpha_1 < \alpha_3$
A first step towards a combined domain

Domains and their concretization:

- **shape abstract domain** $D^*_{sh}$ of graphs
  abstract stores together with a **physical mapping** of nodes
  
  $$\gamma_{sh} : D^*_{sh} \rightarrow \mathcal{P}((D^*_{sh} \rightarrow M) \times (V^* \rightarrow V))$$

- **numerical abstract domain** $D^*_{num}$, abstracts physical mapping of nodes
  
  $$\gamma_{num} : D^*_{num} \rightarrow \mathcal{P}(V^* \rightarrow V)$$

**Combined domain [CR]**

- **Set of abstract values:** $D^* = D^*_{sh} \times D^*_{num}$
- **Concretization:**
  
  $$\gamma(S^#, N^#) = \{(h, \nu) \in M | \nu \in \gamma_{num}(N^#) \land (h, \nu) \in \gamma_{sh}(S^#)\}$$

**Can it be described as a reduced product?**

- **product abstraction:** $D^* = D^*_{0} \times D^*_{1}$ (componentwise ordering)
- **concretization:** $\gamma(x_0, x_1) = \gamma(x_0) \cap \gamma(x_1)$
- **reduction:** $D^*_r$ is the quotient of $D^*$ by the equivalence relation $\equiv$ defined by
  
  $$(x_0, x_1) \equiv (x'_0, x'_1) \iff \gamma(x_0, x_1) = \gamma(x'_0, x'_1)$$
Formalizing the product domain

The use of a simple reduced product raises several issues

Elements without a clear meaning:

- this element exists in the reduced product domain (independent components)
- but, ... what is $\alpha_3$?

Unclear comparison:
How can we compare the two elements below?

- in the reduced product domain, they are not comparable:
  nodes do not match, so componentwise comparison does not make sense
- when concretizing them, there is clear inclusion
Towards a more adapted combination operator

Reason why the reduced product construction does not work well:

- the set of nodes / symbolic variables is not fixed
- the set of dimensions in the numerical domain depends on the shape abstraction

⇒ thus the product is not symmetric
however, the reduced product construction is symmetric

Intuitions

- Graphs form a **shape domain** $D_{sh}^\#$
- For each graph $S^\# \in D_{sh}^\#$, we have a **numerical lattice** $D_{num}^\langle S^\# \rangle$
  - example: if graph $S^\#$ contains nodes $\alpha_0, \alpha_1, \alpha_2$, $D_{num}^\langle S^\# \rangle$ should abstract \{\alpha_0, \alpha_1, \alpha_2\} → $\forall$
- An abstract value is a pair $(S^\#, N^\#)$, such that $N^\# \in D_{num}^\langle N^\# \rangle$
Outline

1. Introduction
2. Separation Logic
3. A shape abstract domain relying on separation
4. Standard static analysis algorithms
5. Combining shape and value abstractions
   - Shape and value properties
   - Combined abstraction with cofibered abstract domain
   - Combined analysis algorithms
6. Conclusion
Cofibered domain

Definition, for shape + num

- **Basis:** abstract domain \((\mathbb{D}^\#_{sh}, \sqsubseteq^\#_{sh})\), with concretization \(\gamma_{sh} : \mathbb{D}^\#_{sh} \rightarrow \mathbb{D}\)

- **Function:** \(\phi : \mathbb{D}^\#_{sh} \rightarrow \mathbb{D}\), where each element of \(\mathbb{D}\) is an abstract domain instance \((\mathbb{D}^\#_{num}, \sqsubseteq^\#_{num})\), with a concretization \(\gamma_{num} : \mathbb{D}^\#_{num} \rightarrow \mathbb{D}\) *(tied to a shape graph)*

- **Domain** \(\mathbb{D}^\#:\) set of pairs \((S^\#, N^\#)\) where \(N^\# \in \phi(S^\#)\)

- **Concretization:** \(\gamma(S^\#, N^\#) = \gamma(S^\#) \cap \gamma(N^\#)\)

- **Lift functions:** \(\forall S^\#_0, S^\#_1 \in \mathbb{D}^\#_{sh}, \text{ such that } S^\#_0 \sqsubseteq^\# S^\#_1, \text{ there exists a function } \prod_{S^\#_0, S^\#_1} : \phi(S^\#_0) \rightarrow \phi(S^\#_1), \text{ that is monotone for } \gamma_{S^\#_0} \text{ and } \gamma_{S^\#_1}\)

- General construction presented in [AV](Arnaud Venet)

- Intuition: a dependent domain product
**Overall abstract domain structure**

**Implementation exploiting the modular structure**
- Each layer accounts for one aspect of the concrete states
- Each layer boils down to a module or functor in ML

State abstract domain $D_{\text{mem}}^\#$
$(e^\#, S^\#, N^\#)$ abstracts sets of $(e, h)$

Combined shape-value abstract domain $D_{\text{cof}}^\#$
$(S^\#, N^\#)$ abstracts sets of $(h, \nu)$

Shape abstract domain $D_{\text{sh}}^\#$
$S^\#$ abstracts sets of $(h, \nu)$

Value abstract domain $D_{\text{num}}^\#$
$N^\#$ abstracts sets of $\nu$

How about operations, transfer functions? Also to be modularly defined
Domain operations

The cofibered structure allows to define **standard domain operations**:
- **ift functions** allow to **switch domain when needed**
- computations first done in the basis, then in the numerical domains, after lifting, when needed

### Comparison of \((S_0^\#, N_0^\#)\) and \((S_1^\#, N_1^\#)\)

1. First, **compare** \(S_0^\#\) and \(S_1^\#\) in \(D_{sh}^\#\)
2. If \(S_0^\# \sqsubseteq S_1^\#\), **compare** \(\Pi_{S_0^\#, S_1^\#}(N_0^\#)\) and \(N_1^\#\)

### Widening of \((S_0^\#, N_0^\#)\) and \((S_1^\#, N_1^\#)\)

1. First, compute the **widening in the basis** \(S^\# = S_0^\# \vee S_1^\#\)
2. Then **move to** \(\phi(S^\#)\), by computing \(N_{0c}^\# = \Pi_{S_0^\#, S_1^\#}(N_0^\#)\) and \(N_{1c}^\# = \Pi_{S_1^\#, S_1^\#}(N_1^\#)\)
3. Last **widen in** \(\phi(S^\#)\): \(N^\# = N_{0c}^\# \vee S^\# \vee N_{1c}^\#\)
4. Return \((S_0^\#, N_0^\#) \vee (S_A^\#, N_1^\#) = (S^\#, N^\#)\)
Outline

1. Introduction

2. Separation Logic

3. A shape abstract domain relying on separation

4. Standard static analysis algorithms

5. Combining shape and value abstractions
   - Shape and value properties
   - Combined abstraction with cofibered abstract domain
   - Combined analysis algorithms

6. Conclusion
Domain operations and transfer functions

Abstract assignments, condition tests:
- need to modify both the shape abstraction and the value abstraction
- both modifications are interdependent

Typical process to compute abstract post-conditions
1. compute the post in the shape abstract domain and update the basis
2. update the value abstraction (numerics) to model dimensions additions and removals
3. compute the post in the value abstract domain

Proofs of soundness of transfer functions rely on:
- the soundness of the lift functions
- the soundness of both domain transfer functions
Analysis of an assignment in the combined domain

\[ N^\# = \alpha_1 \geq 0 \land \alpha_3 \neq 0x0 \]

\[ y \rightarrow d = x + 1 \]

Abstract post-condition?
Analysis of an assignment in the combined domain

\[ N^\# = \alpha_1 \geq 0 \land \alpha_3 \neq 0 \times 0 \]

\[ y \rightarrow d = x + 1 \Rightarrow (\star \alpha_2) \cdot d = (\star \alpha_0) + 1 \]

Abstract post-condition?

Stage 1: environment resolution

- replaces \( x \) with \( \star e^\#(x) \)
Analysis of an assignment in the combined domain

Stage 2: propagate into the shape + numerics domain

- only symbolic nodes appear
Combining shape and value abstractions
Combined analysis algorithms

Analysis of an assignment in the combined domain

\[
\begin{align*}
&x \alpha_0 \rightarrow \alpha_1 \\
&y \alpha_2 \rightarrow \alpha_3 \\
N^\# &= \alpha_1 \geq 0 \land \alpha_3 \neq 0x0 \\
(*\alpha_2) \cdot d &= (*\alpha_0) + 1
\end{align*}
\]

Abstract post-condition?

Stage 3: resolve cells in the shape graph abstract domain

- \(*\alpha_0\) evaluates to \(\alpha_1\); \(*\alpha_2\) evaluates to \(\alpha_3\)
- \((*\alpha_2) \cdot d\) fails to evaluate: no points-to out of \(\alpha_3\)
Analysis of an assignment in the combined domain

\[ N^\# = \alpha_1 \geq 0 \land \alpha_3 \neq 0x0 \]

\[ (\ast\alpha_2) \cdot d = (\ast\alpha_0) + 1 \]

Abstract post-condition?

Stage 4 (a): unfolding triggered
- the analysis needs to locally materialize \( \alpha_3 \cdot \text{lpos} \)...
- thus, unfolding starts at symbolic variable \( \alpha_3 \)
Analysis of an assignment in the combined domain

Stage 4 (b): unfolding, shape part
- unfolding of the memory predicate part
- numerical predicates still need be taken into account
Analysis of an assignment in the combined domain

\[
N^\# = \alpha_1 \geq 0 \land \alpha_3 \neq 0x0 \land \alpha_4 \geq 0
\]

\[
(*\alpha_2) \cdot d = (*\alpha_0) + 1
\]

Abstract post-condition?

Stage 4 (c): unfolding, numeric part
- numerical predicates taken into account
- l-value \( \alpha_3 \cdot d \) now evaluates into edge \( \alpha_3 \cdot d \mapsto \alpha_4 \)

Xavier Rival (INRIA)
Analysis of an assignment in the combined domain

Stage 5: create a new node
- new node $\alpha_6$ denotes a new value
  will store the new value
Analysis of an assignment in the combined domain

Stage 6: perform numeric assignment

- numeric assignment *completely ignores* pointer structures to the new node
Analysis of an assignment in the combined domain

Stage 7: perform the update in the graph

- classic **strong update** in a pointer aware domain
- symbolic node $\alpha_4$ becomes redundant and can be removed
Widening / join in the combined domain

environment layer
shape + num + env

cofibered layer
shape + num

shape domain
numeric domain

\( \alpha_0 \rightarrow \alpha_1 \rightarrow \alpha_2 \rightarrow \text{lpos} \)
\( \&x \ \alpha_0 \rightarrow \alpha_1 \rightarrow \alpha_3 \)
\( \&y \ \alpha_4 \rightarrow \alpha_5 \)
\( N_{\text{lft}}^\# = \alpha_2 \geq \alpha_5 \geq 2 \)

\( \beta_0 \rightarrow \beta_1 \rightarrow \text{lpos} \)
\( \&x \ \beta_0 \rightarrow \beta_1 \)
\( \&y \ \beta_2 \rightarrow \beta_3 \)
\( N_{\text{rg}}^\# = \beta_3 \geq 1 \)
Widening / join in the combined domain

Stage 1: abstract environment

- compute new abstract environment and initial node relation
  e.g., $\alpha_0, \beta_0$ both denote $\&x$
Combining shape and value abstractions

Combined analysis algorithms

Widening / join in the combined domain

Stage 2: join in the “cofibered” layer

operations to perform:

1. compute the join in the graph
2. convert value abstractions, and join the resulting lattice
Combining shape and value abstractions  Combined analysis algorithms

Widening / join in the combined domain

Stage 2: graph join

- apply local join rules
  - ex: points-to matching, weakening to inductive ...
- incremental algorithm
Widening / join in the combined domain

- environment layer: shape + num + env
- cofibered layer: shape + num
- shape domain
- numeric domain

Stage 2: graph join
- apply local join rules
  - ex: points-to matching, weakening to inductive...
- incremental algorithm
Widening / join in the combined domain

Stage 2: graph join
- apply local join rules
  - ex: points-to matching, weakening to inductive...
- incremental algorithm

\[ \mathcal{N}_{\text{left}}^l = \alpha_2 \geq \alpha_5 \geq 2 \]
\[ \mathcal{N}_{\text{right}}^r = \beta_3 \geq 1 \]

\[ \psi(\alpha_0, \beta_0) = \delta_0 \]
\[ \psi(\alpha_4, \beta_2) = \delta_1 \]
\[ \psi(\alpha_1, \beta_1) = \delta_2 \]
\[ \psi(\alpha_5, \beta_3) = \delta_3 \]
Widening / join in the combined domain

Stage 3: conversion function application in numerics
- remove nodes that were abstracted away
- rename other nodes
Widening / join in the combined domain

Stage 4: join in the numeric domain

- apply $\sqcap$ for regular join, $\triangledown$ for a widening
Conclusion

Outline

1 Introduction

2 Separation Logic

3 A shape abstract domain relying on separation

4 Standard static analysis algorithms

5 Combining shape and value abstractions

6 Conclusion
Updates and summarization

Weak updates cause significant precision loss...
Separation logic makes updates strong

Separation logic

Separating conjunction combines properties on disjoint stores

- Fundamental idea: *forces to identify what is modified*
- Before an update (or a read) takes place, memory cells need to be materialized
- **Local reasoning**: properties on unmodified cells pertain

Summaries

**Inductive predicates** describe unbounded memory regions

- Last lecture: array segments and transitive closure (TVLA)
Bibliography

- **[JR]**: *Separation Logic: A Logic for Shared Mutable Data Structures.*  
  John C. Reynolds.  
  In LICS’02, pages 55–74, 2002.

- **[DHY]**: *A Local Shape Analysis Based on Separation Logic.*  
  Dino Distefano, Peter W. O’Hearn and Hongseok Yang.  
  In TACAS’06, pages 287–302.

- **[CR]**: *Relational inductive shape analysis.*  
  Bor-Yuh Evan Chang and Xavier Rival.  
Assignment and paper reading

The Frame rule:

- formalize the Hoare logic rules for a language with pointer assignments and condition tests
- prove the Frame rule by induction over the syntax of programs

Reading:


Formalizes the Frame rule, among others
Assignment: a simple analysis in Separation Logic (after TVLA)

1, k assumed to be disjoint lists

\[
\begin{align*}
&\text{while}(l \neq 0)\{ \\
&t = l \rightarrow n; \\
&l \rightarrow n = k; \\
&k = 1; \\
&l = t;
\end{align*}
\]
\( x \rightarrow x_0 \)

\( x = \text{mellor (4)} \)

\( x = \alpha_1 \ast \alpha_1 \rightarrow \alpha_2 \)

\[ \begin{array}{c}
\alpha. \text{list} = \ldots \ldots \\
\downarrow \\
\alpha. \text{list}(\exists) \\
symb. \text{set}
\end{array} \]

Huisong Li
SAS, '16