On the Complexity of Pointer Arithmetic in Separation Logic

James Brotherston[1]   Max Kanovich[1,2]

[1]University College London, UK

[2]National Research University Higher School of Economics, Russian Federation

APLAS-16, Wellington (NZ), 5th Dec 2018

Overview

- Industrial separation logic (SL) analysis is usually based on the "symbolic heap" fragment over pointers and list segments, which is PTIME-decidable.
- Many other features have been studied...
  - user-defined inductive predicates;
  - fractional permissions;
  - separating implication ($-\!\!*$);
  - arrays;
  - reachability predicates;
  - arithmetic;

  ...but they typically come with a complexity cost.
- Our focus is on pointer arithmetic in SL.

Pointer arithmetic in program analysis

- Pointer arithmetic is usually disallowed or at least discouraged in modern programming practice.
- However, it still arises implicitly, e.g., in array indexing and structure / union member selection.

$$\texttt{ptr[i]} \quad \Rightarrow \quad \texttt{ptr + (sizeof(*ptr)*i)}$$

- Thus program analyses must deal with pointer arithmetic even when programmers don't!

Question: *How much pointer arithmetic can one add to separation logic and remain within polynomial time?*

Minimal fragment, $\mathsf{SL}_{\mathsf{MPA}}$

- Terms t, pure formulas Π and spatial formulas F given by:

$$\begin{array}{rcl} t & ::= & x \mid x + k \\ \Pi & ::= & x = t \mid x \leq t \mid \Pi \wedge \Pi \\ F & ::= & \mathsf{emp} \mid t \mapsto t \mid t \mapsto \mathsf{nil} \mid F * F \end{array}$$

  where $x \in \mathsf{Var}$, $k \in \mathbb{Z}$.

- Symbolic heaps given by $\exists \mathbf{x}.\ \Pi : F$.

Semantics

Stacks are maps $\mathsf{Var} \to \mathbb{N} \cup \{nil\}$. Heaps are maps $\mathbb{N} \rightharpoonup_{\text{fin}} \mathbb{N}$. We write $\circ$ for the composition of domain-disjoint heaps and e for the empty heap.

Semantics given as usual by $s, h \models A$:

$$
\begin{array}{lcl}
s, h \models x = t & \Leftrightarrow & s(x) = s(t) \\
s, h \models x \leq t & \Leftrightarrow & s(x) \leq s(t) \\
s, h \models \Pi_1 \wedge \Pi_2 & \Leftrightarrow & s, h \models \Pi_1 \text{ and } s, h \models \Pi_2 \\
s, h \models \mathsf{emp} & \Leftrightarrow & h = e \\
s, h \models t_1 \mapsto t_2 & \Leftrightarrow & \mathrm{dom}(h) = \{s(t_1)\} \text{ and } h(s(t_1)) = s(t_2) \\
s, h \models F_1 * F_2 & \Leftrightarrow & \exists h_1, h_2.\ h = h_1 \circ h_2 \text{ and } s, h_1 \models F_1 \\
 & & \quad \text{and } s, h_2 \models F_2
\end{array}
$$

Difference constraints

Our pure formulas are conjunctions of difference constraints

$$x \leq y + k \ ,$$

where x, y are pointer variables and $k \in \mathbb{Z}$ is an integer offset.

Note: the satisfiability of these formulas can be decided in polynomial time.

$$\left.\begin{array}{l} x_1 \leq x_2 + k_1, \\ \dots \\ x_{m-1} \leq x_m + k_{m-1}, \\ x_m \leq x_1 + k_m \end{array}\right\} \Rightarrow \ x_1 - x_1 \leq \sum_{i=1}^{m} k_i$$

Problems of interest

Satisfiability problem. *Given symbolic heap A, decide if there is a stack-heap pair (s, h) with $s, h \models A$.*

Entailment problem. *Given symbolic heaps A and B, decide whether $A \models B$.*

Small model property. *Given a satisfiable symbolic heap A, does A have a model using only addresses and values of size polynomial in A?*

(The latter fails if we allow pointer sums, $x \leq y + z$.)

Some known upper bounds

$\mathsf{SL}_{\mathsf{MPA}}$ is subsumed by the array separation logic in

> J. Brotherston, N. Gorogiannis, and M. Kanovich. Biabduction (and related problems) in array separation logic. In Proc. *CADE* 2017.

This gives some immediate upper bounds by encoding into Presburger arithmetic (PbA):

- Satisfiability is in NP.
- Quantifier-free entailment is in coNP.
- Quantified entailment is in Π_1^{EXP}.

Satisfiability, lower bound

In fact, the lower bound for satisfiability is also NP.

3-colourability problem (NP-hard)

Given an undirected graph, decide whether there is a "perfect" 3-colouring of the vertices, such that no two adjacent vertices share the same colour.

First, choose numbers e_{ij} for each edge (v_i, v_j) such that $|e_{i'j'} - e_{ij}| \geq 4$ for any two distinct edges.

Next take a variable c_i for each vertex v_i.

Then encode in $\mathsf{SL}_{\mathsf{MPA}}$ as (slightly simplified)

$$\bigwedge_{i=1}^{n} 1 \leq c_i \leq 3 \colon \mathop{\Huge *}_{(v_i,v_j)\in E} (c_i + e_{ij} \mapsto \mathsf{nil} * c_j + e_{ij} \mapsto \mathsf{nil})$$

Small model property

Suppose A is satisfiable: $s, h \models A$.

There is an equisatisfiable PbA formula γ_A with $s \models \gamma_A$.

The formula γ_A can be written as a Boolean combination of difference constraints $x \leq y + k$.

Thus s can be viewed as a solution to the equation system

$$(x_1 \leq y_1 + k_1) \equiv \zeta_1, \ldots, (x_m \leq y_m + k_m) \equiv \zeta_m$$

where each $\zeta_i \in \{\top, \bot\}$.

Note that $(x \leq y + k) \equiv \bot$ means $y \leq x - k - 1$.

Small model property (2)

View the difference equations as a constraint graph, as follows:

$$\begin{array}{rcl} (x_i \leq y_i + k_i) \equiv \top & \sim & y_i \xrightarrow{k_i} x_i \\ (x_i \leq y_i + k_i) \equiv \bot & \sim & x_i \xrightarrow{-k_i-1} y_i \\ \text{all } x_i\text{:} & & x_0 \xrightarrow{0} x_i \end{array}$$

where x_0 is a new “maximum node”.

FACT: This graph cannot have a negative-weight cycle (else the equation system would have no solutions and thus $s \not\models \gamma_A$).

We construct a new model s' of γ_A with all values bounded by

$$M = \Sigma_{i=1}^{m} \; |k_i| + 1 \; .$$

Small model property (3)

Define a new, small model s' of γ_A as follows:

$$\begin{array}{rcl} d_i & = & \text{minimal path weight from } x_0 \text{ to } x_i \\ s'(x_i) & = & M + d_i \end{array}$$

Note $d_i \leq 0$ (a 0-weight path exists by construction), and d_i is always well defined (no negative-weight cycles). So s' is small.

Why is it a model of γ_A?

Consider constraint $x \leq y + k \equiv \top$. There is an edge $y \xrightarrow{k} x$. Thus $d_x \leq d_y + k$ and so $s'(x) \leq s'(y) + k$.

So s' satisfies our difference equation system, and thus $s' \models \gamma_A$. Then we can create a suitable h' with $s', h' \models A$.

Quantifier-free entailment

By adapting the tricks for satisfiability, we have the following for quantifier-free entailments $A \models B$:

1. a lower bound of coNP;

2. the small model property (any invalid entailment has a small countermodel).

Quantified entailment, lower bound

Lower bound is Π_2^P in the polynomial-time hierarchy.

Proof is by reduction from:

2-round 3-colourability problem (Π_2^P-hard)

Given an undirected graph, decide whether every 3-colouring of the leaves can be extended to a perfect 3-colouring of the graph.

Proof builds on the ideas for the satisfiability case, using a more sophisticated encoding.

Quantified entailment, upper bound

We have an encoding into Π_2^0 PbA, an upper bound of Π_1^{EXP} (in the exponential-time hierarchy).

In fact this upper bound is exponentially overstated! The upper bound is also Π_2^P.

The key difference between Π_2^0 PbA and Π_2^P is that in the latter all variables must be polynomially bounded. This follows from the small model property for quantified entailment.

Construction uses similar ideas to satisfiability case, but is (quite a bit) more complex.

Conclusions

- NP-hardness or worse is an inevitable consequence of adding pointer arithmetic to SL,
  - even for pointer data only, and
  - even for pointer-offset comparisons, $x \leq y + k$.
- Satisfiability is NP-complete.
- Quantifier-free entailment is coNP-complete.
- Quantified entailment is Π_2^P-complete.
- The small model property holds.

Thanks for listening!