Colloque d'Automne du LIX 2007

CAL07

A mathematical programming model for computing fixed points in static program analysis

Fabrizio Marinelli and Leo Liberti

{marinelli,liberti}@lix.polytechnique.fr

LIX, Laboratorie d'Informatique

École Polytechnique

Paris, October 3-4, 2007

Outline of the talk

- Some words on static analysis and abstract interpretation
- A mathematical formulation to compute fixed points
- Some preliminary computational results
- Conclusions and future work

Static analysis

- **goal**: statically infer run-time properties of programs (e.g., variable values and dependencies).
- **purpose**: program correctness proofs (e.g., safety, termination, run-time errors) , code optimization (e.g., compile-time garbage collection).
- **basic assumption**: the answers can only be approximate since problems are either undecidable (e.g., termination for all input data) or computationally intractable.
- **tools**: Abstract interpretation, dataflow analysis, control flow analysis, model checking.

Abstract Interpretation

- Focuses on a class of properties of program executions and yields an over-approximations of invariants.

- Starting from a *concrete* semantic,
  1. an *abstract domain* and an *abstract semantic* are defined,
  2. a fixpoint of the abstract semantic, preferably the <u>least</u> one, is computed. Fixpoints are in general obtained by means of iterative procedures based on Kleene's fixed point iteration algorithm.

- **Aim of this work**

  We propose an alternative approach based on a mathematical programming language, i.e. a language for expressing optimization and decision problems by means of mathematical relations, that allows the solution of them using generic algorithms.

  main advantage: flexibility

Preliminaries

- $p_1, \ldots, p_n$ — control points of a program P which only performs additions, subtractions and products with a constant.
- x — a variable of P
- $I_k = [x_k^l, x_k^u]$ — values that x can take at p_k $(x_k^l, x_k^u \in \mathrm{IR} \cup \{\pm\infty\})$
- $(\Lambda, \subseteq)$ — complete lattice of the closed intervals of IR ordered by inclusion (lowest element = $\varnothing$ and greatest element = $]-\infty, +\infty[$)
- $S: \Lambda^n \to \Lambda^n$ — an abstract semantic of P in the abstract domain of intervals

- Each function $S_k : (I_1, \ldots, I_n) \to I_k$ is an arithmetic logical expression involving binary operators in the set $\otimes = \{+, -, *, \cup, \cap\}$
- A least fixed point of S is an invariant of P. it can be obtained by solving

$$\min_{I_1,\ldots,I_n} \{S_k(I_1,\ldots,I_n) = I_k \mid k = 1,\ldots,n\}$$

Example

Program P

```
    void main(){
     int x = 0;
p1   while (x < 100){
p2     x = x + 1;
p3   }
p4 }
```

Semantic S: $\Lambda^4 \rightarrow \Lambda^4$

$S_1 : [0,0]$

$S_2 :]-\infty,99] \cap (I_1 \cup I_3)$

$S_3 : I_2 + [1,1]$

$S_4 : [100,+\infty[\cap (I_1 \cup I_3)$

System of fixed point equations

$$
\begin{cases}
I_1 = [0,0] \\
I_2 =]-\infty,99] \cap (I_1 \cup I_3) \\
I_3 = I_2 + [1,1] \\
I_4 = [100,+\infty[\cap (I_1 \cup I_3)
\end{cases}
$$

a fixed point

$$I_1 = [x_1^l, x_1^u] = [0,0]$$
$$I_2 = [x_2^l, x_2^u] = [0,99]$$
$$I_3 = [x_3^l, x_3^u] = [1,100]$$
$$I_4 = [x_4^l, x_4^u] = [100,100]$$

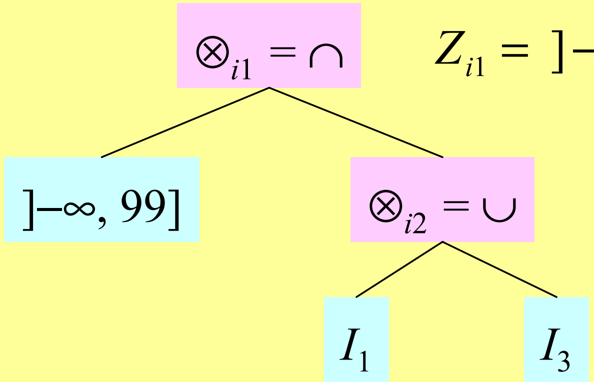
Mathematical model

- For each control point p_i let us define a pair of real variables $\left[x_i^l, x_i^u\right] = I_i$. Recall that $I_1, \ldots, I_n$ describe an invariant of P
- For each function S_i and for each operator $\otimes_{ij}$ in S_i let us define
  - a pair of real variables $\left[z_{ij}^l, z_{ij}^u\right] = Z_{ij}$
  - a set Ω_{ij} of variables and constraints that model the semantic of $\otimes_{ij}$ in the arithmetic of intervals

Mathematical model

- For each control point p_i let us define a pair of real variables $[x_i^l, x_i^u] = I_i$. Recall that $I_1,\ldots, I_n$ describe an invariant of P
- For each function S_i and for each operator $\otimes_{ij}$ in S_i let us define
  - a pair of real variables $[z_{ij}^l, z_{ij}^u] = Z_{ij}$
  - a set Ω_{ij} of variables and constraints that model the semantic of $\otimes_{ij}$ in the arithmetic of intervals

- Example: $I_i = \,]-\infty, 99] \cap (I_1 \cup I_3)$



$Z_{i1} = \,]-\infty, 99] \cap Z_{i2}$

$Z_{i2} = I_1 \cup I_3$

- Note

$\otimes_{i1},\ldots,\otimes_{im_i}$ are ranked according to the reverse order of evaluation;

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

$$x_i^l = z_{i1}^l \qquad i = 1, \ldots, n$$

$$x_i^u = z_{i1}^u \qquad i = 1, \ldots, n$$

$$\Omega_{ij} \qquad i = 1, \ldots, n,\ j = 1, \ldots, m_i$$

$$z_{ij}^l \le z_{ij}^u \qquad i = 1, \ldots, n,\ j = 1, \ldots, m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2 \right]$$

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

Fixed point of S

$$\boxed{\begin{array}{ll} x_i^l = z_{i1}^l & i = 1,\ldots,n \\ x_i^u = z_{i1}^u & i = 1,\ldots,n \end{array}}$$

$$\Omega_{ij} \qquad i = 1,\ldots,n,\ j = 1,\ldots,m_i$$

$$z_{ij}^l \le z_{ij}^u \qquad i = 1,\ldots,n,\ j = 1,\ldots,m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2\right]$$

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

Fixed point of S

$$\boxed{\begin{array}{ll} x_i^l = z_{i1}^l & i = 1,\ldots,n \\ x_i^u = z_{i1}^u & i = 1,\ldots,n \end{array}}$$

Semantic of operators

$$\boxed{\begin{array}{ll} \Omega_{ij} & i = 1,\ldots,n,\ j = 1,\ldots,m_i \end{array}}$$

$$z_{ij}^l \le z_{ij}^u \qquad i = 1,\ldots,n,\ j = 1,\ldots,m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2 \right]$$

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

Fixed point of S

$$x_i^l = z_{i1}^l \qquad i = 1,\ldots,n$$

$$x_i^u = z_{i1}^u \qquad i = 1,\ldots,n$$

Semantic of operators

$$\Omega_{ij} \qquad i = 1,\ldots,n,\ j = 1,\ldots,m_i$$

Proper definition of intervals

$$z_{ij}^l \le z_{ij}^u \qquad i = 1,\ldots,n,\ j = 1,\ldots,m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2 \right]$$

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

Minimum total length of $I_1, \dots, I_n$

Fixed point of S

$$x_i^l = z_{i1}^l \qquad i = 1, \dots, n$$

$$x_i^u = z_{i1}^u \qquad i = 1, \dots, n$$

Semantic of operators

$$\Omega_{ij} \qquad i = 1, \dots, n,\ j = 1, \dots, m_i$$

Proper definition of intervals

$$z_{ij}^l \leq z_{ij}^u \qquad i = 1, \dots, n,\ j = 1, \dots, m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2 \right]$$

Mathematical model

$$\min \sum_{i=1}^{n} \left(x_i^u - x_i^l \right)$$

Minimum total length of $I_1, \ldots, I_n$

Fixed point of S

$$x_i^l = z_{i1}^l \qquad i = 1, \ldots, n$$

$$x_i^u = z_{i1}^u \qquad i = 1, \ldots, n$$

Semantic of operators

$$\Omega_{ij} \qquad i = 1, \ldots, n,\ j = 1, \ldots, m_i$$

Proper definition of intervals

$$z_{ij}^l \le z_{ij}^u \qquad i = 1, \ldots, n,\ j = 1, \ldots, m_i$$

$$z_{ij}^l, z_{ij}^u \in \left[-M/2, M/2 \right]$$

- for a suitable choice of M, the solution space coincides with the set of fixed points of S

$\Rightarrow$ an optimal solution of the model is a least fixed point of S

The key-role of M

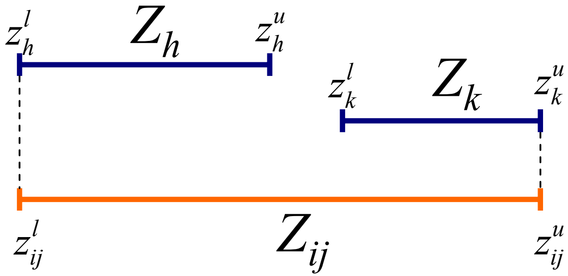
- numerical computation is performed by finite arithmetic
  $\Rightarrow$ the infinity value is represented by a suitable large number $M/2$
  $\Rightarrow$ the endpoints of intervals are limited to the range $[-M/2, M/2]$.
- The M parameter is also used to model implications between real and binary variables, e.g.,:

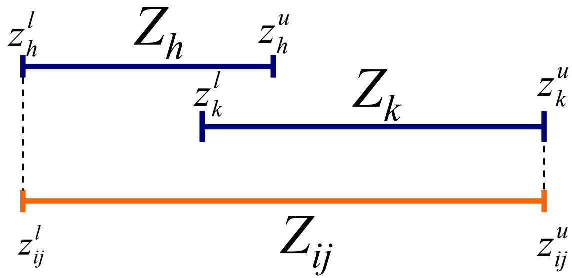
$$x \in [0, M], \quad y \in \{0,1\} \qquad x > 0 \Rightarrow y = 1 \quad \text{is modeled by} \quad x \le My$$

- A large value for M
  + allows the computation of better fixpoints, potentially a least one
  – makes the model ill-conditioned and harder to solve

$\otimes_{ij} = \cup$
$\otimes_{ij} = \cap$
$\otimes_{ij} = +$

Union operator: $Z_{ij} = Z_h \cup Z_k$





- Semantic

$$(i)\ \ z_{ij}^l = \min\left\{z_h^l, z_k^l\right\}$$

$$(ii)\ \ z_{ij}^u = \max\left\{z_h^u, z_k^u\right\}$$

- Constraints of Ω_{ij}

$$\begin{cases} z_{ij}^l \le z_h^l \\ z_{ij}^l \le z_k^l \end{cases}$$

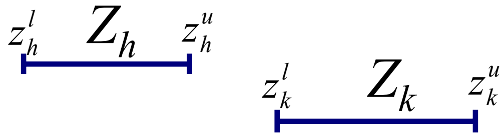
$$\begin{cases} z_{ij}^u \ge z_h^u \\ z_{ij}^u \ge z_k^u \end{cases}$$

$\otimes_{ij} = \cup$

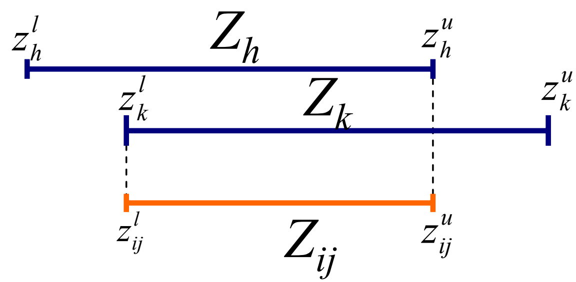
$\otimes_{ij} = \cap$

$\otimes_{ij} = +$

Intersection operator: $Z_{ij} = Z_h \cap Z_k$



$Z_{ij} = \varnothing$



- Semantic

$$(i)\ \ Z_{ij} = \varnothing \ \text{ if } \left(z_h^l > z_k^u\right) \vee \left(z_k^l > z_h^u\right)$$

$$(ii)\ \ z_{ij}^l = \max\left\{z_h^l, z_k^l\right\}$$

$$(iii)\ \ z_{ij}^u = \min\left\{z_h^l, z_k^l\right\}$$

- Empty intersection set must be considered
- *min* and *max* cannot be modeled by simple inequalities

$\otimes_{ij} = \cup$

$\otimes_{ij} = \cap$

$\otimes_{ij} = +$

Intersection operator: $Z_{ij} = Z_h \cap Z_k$

- Variables of Ω_{ij}

$$y_{ij}^{0} = \begin{cases} 1 \text{ if } Z_{ij} = \varnothing \\ 0 \text{ otherwise} \end{cases}$$

$$y_{ij}^{lt} = \begin{cases} 1 \text{ if } z_{ij}^{l} = z_{t}^{l} & t \in \{h, k\} \\ 0 \text{ otherwise} \end{cases}$$

$$y_{ij}^{ut} = \begin{cases} 1 \text{ if } z_{ij}^{u} = z_{t}^{u} & t \in \{h, k\} \\ 0 \text{ otherwise} \end{cases}$$

$\otimes_{ij} = \cup$

$\otimes_{ij} = \cap$

$\otimes_{ij} = +$

Intersection operator: $Z_{ij} = Z_h \cap Z_k$

- Variables of Ω_{ij}

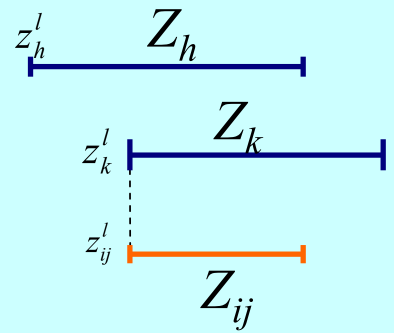
$$y_{ij}^{0} = \begin{cases} 1 & \text{if } Z_{ij} = \varnothing \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij}^{lt} = \begin{cases} 1 & \text{if } z_{ij}^{l} = z_{t}^{l} \qquad t \in \{h,k\} \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij}^{ut} = \begin{cases} 1 & \text{if } z_{ij}^{u} = z_{t}^{u} \qquad t \in \{h,k\} \\ 0 & \text{otherwise} \end{cases}$$

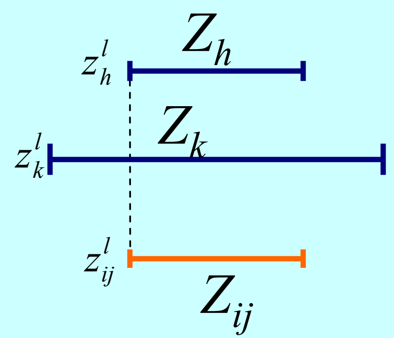
$y_{ij}^{lk} = 1$

$y_{ij}^{lh} = 0$



$y_{ij}^{lk} = 0$

$y_{ij}^{lh} = 1$



$\otimes_{ij} = \cup$

$\boldsymbol{\otimes_{ij} = \cap}$

$\otimes_{ij} = +$

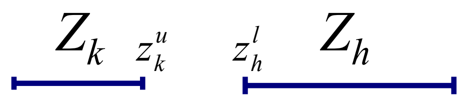
Intersection operator: $Z_{ij} = Z_h \cap Z_k$

- Semantic

$$(i)\;\; Z_{ij} = \varnothing \;\text{ if }\; \left(z_h^l > z_k^u\right) \vee \left(z_k^l > z_h^u\right)$$

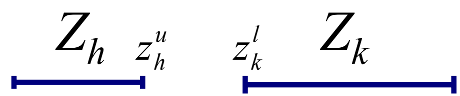
- Constraints of Ω_{ij}

$$\left.\begin{array}{l} z_h^l - z_k^u \le M y_{ij}^0 \\ \left|z_h^l - z_k^u\right| \ge \varepsilon y_{ij}^0 \end{array}\right\}$$



$\Rightarrow y_{ij}^0 = 1$

$$\left.\begin{array}{l} z_k^l - z_h^u \le M y_{ij}^0 \\ \left|z_k^l - z_h^u\right| \ge \varepsilon y_{ij}^0 \end{array}\right\}$$



$\Rightarrow y_{ij}^0 = 1$

$$\left.\begin{array}{l} z_{ij}^l + M y_{ij}^0 \le M/2 \\ z_{ij}^u - M y_{ij}^0 \ge -M/2 \end{array}\right\} \quad y_{ij}^0 = 1 \Rightarrow \left[z_{ij}^l, z_{ij}^u\right] = \left[-M/2, M/2\right]$$

$\otimes_{ij} = \cup$

$\otimes_{ij} = \cap$

$\otimes_{ij} = +$

Intersection operator: $Z_{ij} = Z_h \cap Z_k$

- Semantic

$$(ii)\quad z_{ij}^{l} = \max\left\{z_h^l, z_k^l\right\}$$

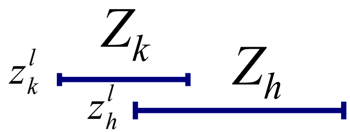
- Constraints of Ω_{ij}

$$z_h^l - z_k^l \le M\left(y_{ij}^{lh} + y_{ij}^{0}\right)$$

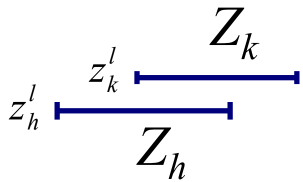
$$z_k^l - z_h^l \le M\left(y_{ij}^{lk} + y_{ij}^{0}\right)$$

$$y_{ij}^{lh} + y_{ij}^{lk} + y_{ij}^{0} = 1$$

$$y_{ij}^{lt}\left(z_{ij}^l - z_t^l\right) = 0 \quad t \in \{h, k\}$$



$\Rightarrow y_{ij}^{lh} = 1$ or $y_{ij}^{0} = 1$



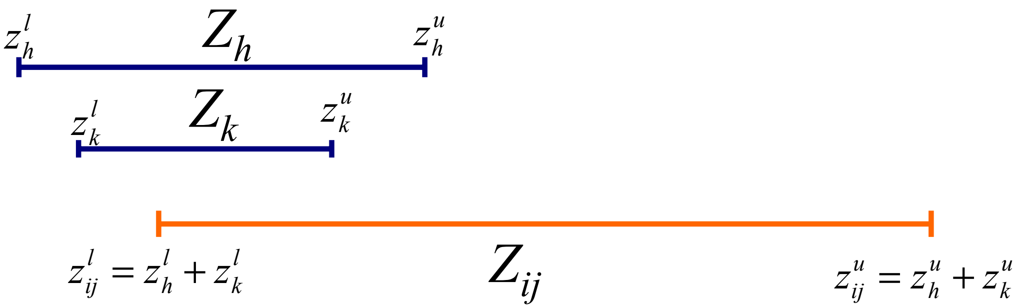
$\Rightarrow y_{ij}^{lk} = 1$ or $y_{ij}^{0} = 1$

$$\begin{cases} y_{ij}^{lh} = 1 \Rightarrow z_{ij}^l = z_h^l \\ y_{ij}^{lk} = 1 \Rightarrow z_{ij}^l = z_k^l \end{cases}$$

- Similar constraints are defined to model $z_{ij}^{u} = \min\left\{z_h^u, z_k^u\right\}$

$\otimes_{ij} = \cup$
$\otimes_{ij} = \cap$
$\otimes_{ij} = +$

Plus operator: $Z_{ij} = Z_h + Z_k$



- Semantic

$$
\begin{array}{lll}
(i) & z_{ij}^l = -\infty & \text{if } \left(z_h^l = -\infty\right) \vee \left(z_k^l = -\infty\right) \\
(ii) & z_{ij}^l = z_h^l + z_k^l & \text{if } z_h^l, z_k^l \neq -\infty \\
(iii) & z_{ij}^u = +\infty & \text{if } \left(z_h^u = +\infty\right) \vee \left(z_k^u = +\infty\right) \\
(iv) & z_{ij}^u = z_h^u + z_k^u & \text{if } z_h^u, z_k^u \neq +\infty
\end{array}
$$

- Addition between intervals must be extended to deal with infinity values

$\otimes_{ij} = \cup$

$\otimes_{ij} = \cap$

$\otimes_{ij} = +$

Plus operator: $Z_{ij} = Z_h + Z_k$

- Variables of Ω_{ij}

$$w_t^l = \begin{cases} 1 \text{ if } z_t^l \neq -\infty & t \in \{h, k\} \\ 0 \text{ otherwise} \end{cases}$$

$$r_{hk}^l = \begin{cases} 1 \text{ if } \left(z_h^l = -\infty\right) \vee \left(z_k^l = -\infty\right) \\ 0 \text{ otherwise} \end{cases}$$

- $w^l{}_h = 1(w^l{}_k = 1)$ indicates that the lower limit of Z_h (Z_k) is finite

- $r^l{}_{hk} = 1$ indicates that one of the lower limits is infinite

$$w_t^u = \begin{cases} 1 \text{ if } z_t^u \neq +\infty & t \in \{h, k\} \\ 0 \text{ otherwise} \end{cases}$$

$$r_{hk}^u = \begin{cases} 1 \text{ if } \left(z_h^u = +\infty\right) \vee \left(z_k^u = +\infty\right) \\ 0 \text{ otherwise} \end{cases}$$

Plus operator: $Z_{ij} = Z_h + Z_k$

$\otimes_{ij} = \cup$
$\otimes_{ij} = \cap$
$\otimes_{ij} = +$

- Constraints of Ω_{ij}

$$
\left.
\begin{array}{ll}
z_t^l - Mw_t^l \le -M/2 & t \in \{h,k\} \\
z_t^l + M(1 - w_t^l) \ge -\varepsilon - M/2 & t \in \{h,k\}
\end{array}
\right\} \text{Proper definition of } w^l{}_h \text{ and } w^l{}_k
$$

$$r_{hk}^l + w_t^l \ge 1 \qquad t \in \{h,k\} \qquad\qquad w^l{}_h = 0 \text{ or } w^l{}_k = 0 \Rightarrow r^l{}_{hk} = 1$$

$$z_{ij}^l = (z_h^l + z_k^l)(1 - r_{hk}^l) - \frac{M}{2} r_{hk}^l \qquad\qquad z^l{}_{ij} = \begin{cases} z^l{}_h + z^l{}_k & \text{for “finite” values} \\ -M/2 & \text{otherwise} \end{cases}$$

- Similar constraints are defined for the upper limit of Z_{ij}
- Minus operator $Z_{ij} = Z_h - Z_k$ can be easily transformed into plus operator by setting $Z_k = [-z^u{}_k, -z^l{}_k]$

Solution of the model

- All the non-linear constraints can be easily linearized.
- The model is a Mixed Integer Linear Program which can be solved by branch-and-bound algorithms coded in standard tools (e.g., Cplex, Xpress-MP, Lp-solve).

- **branch-and-bound**:
  - decomposes the problem in sub-problems easier and easier;
  - the process is represented with an enumeration tree where the root node is the original problem and the leaves are solutions;
  - bounding of sub-problems is performed by comparing lower and upper bounds.

- Branch-and-bound is an exponential algorithm in the worst case.

Computational validation

- Instance set: 40 toy examples in C language

- 62.5% of the problems are solved at root node (10% just by preprocessing)
- The average size of the enumeration tree is 7,275 nodes
- 93,87 simplex iterations are performed on the average
- Computational times are negligible (< 0.01 sec.)

- Model solver: Cplex 10.1
- Machine: AMD Athlon 64 1.8GHz

Conclusions and future work

Conclusions

- A mathematical programming approach to compute fixpoints in the abstract domain of intervals has been proposed
- The model
  - can be used together with existing methods to derive better approximations of invariants and
  - can be useful for parameterized fixpoint computation (e.g., optimization of the fixed point formats of numbers)
- The model has been validated on small examples in C language

Future work

- Testing on real case instances
- Numerical problems and weakness of lower bound due to the large constant M
- Extension to relational domains such as octagons and polyedra

Colloque d'Automne du LIX 2007

CAL07

A mathematical programming model for computing fixed points in static program analysis

Leo Liberti and Fabrizio Marinelli

{liberti,marinelli}@lix.polytechnique.fr

LIX, Laboratorie d’Informatique

École Polytechnique

Paris, October 3-4, 2007