Interprocedural Data flow Analysis

Aditya V. Nori
MSR India

Functional Approach

- Compute a function for each procedure describing the “abstract” effect of the procedure.
- These functions are then used in a standard (intraprocedural) algorithm.
- The solution computed is an MOP solution only if the underlying DFF is distributive.

Functional Approach - Snags

- Computation of values from $L \rightarrow L$. This space must be finite for termination.
- Guaranteed to terminate only for finite lattices.
- Computationally expensive.

Call String Approach

Basic Idea:

- Consider procedure calls and returns as ordinary transfers of control.
- Avoid data propagation along interprocedurally invalid paths.
- Tag propagated data with a call string.

Definitions

A call string γ is a tuple of call blocks $c_1, c_2, \ldots, c_j$ in N^* for which there exists an execution path $q \in IVP(r_1,n)$ terminating at some $n \in N^*$, such that the path decomposition of q has the form

$$q_1||(c_1,r_{p_2})||q_2||\cdots||(c_j,r_{p_{j+1}})||q_{j+1}$$

where $q_i \in IVP_0(r_{p_i},c_i)$ for each $i \leq j$, $q_{j+1} \in IVP_0(r_{j+1},n)$.

Define $CM:IVP(r_1,n) \rightarrow \Gamma$, such that $CM(q) = \gamma$.

Example

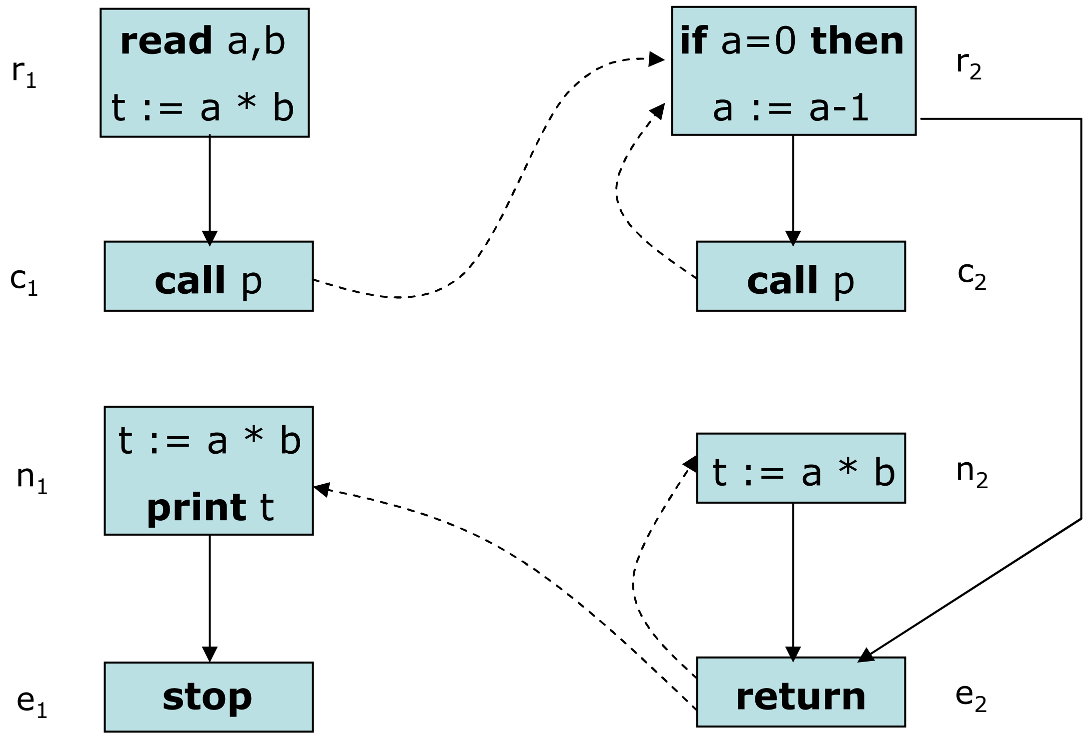
main program

```
read a, b;
t := a * b;
call p;
t := a * b;
print t;
stop;
end
```

procedure p

```
if a = 0 then return;
else
   a := a – 1;
   call p;
   t := a * b;
endif;
return;
end
```

Interprocedural CFG G*



The following call strings are possible: (λ), (c_1), (c_1c_2), $(c_1c_2c_2)$ …

Let Γ denote the set of all call strings γ corresponding to IVPs in G^*.

If G^* is nonrecursive $\Rightarrow \Gamma$ is finite else it is infinite.

Let (L,F) be a DFF. Define a new DFF (L^*,F^*) as follows:

- $L^* = L^\Gamma$.
- F^* will be defined later.

If $\xi \in L^*$ and $\gamma \in \Gamma$, then intuitively,

$\xi(\gamma)$ = data propagated along paths in $CM^{-1}(\gamma)$.

L^* is a semilattice

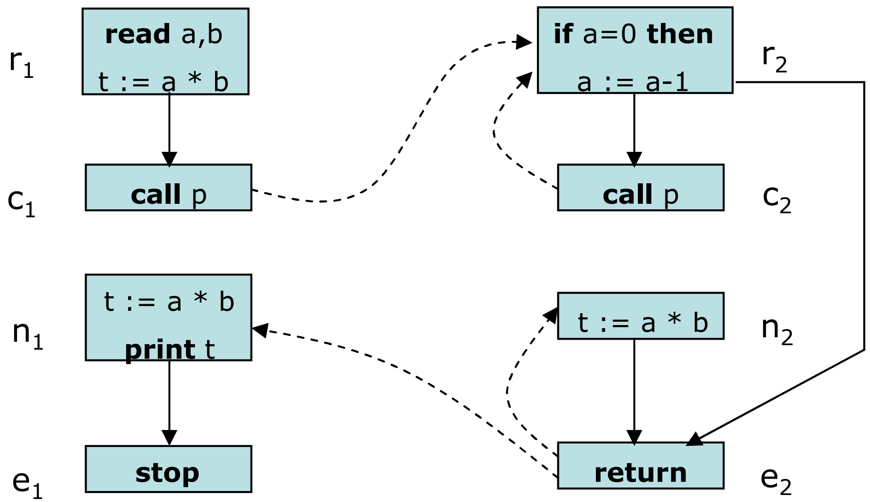
- Meet operation in L^* is a pointwise meet on Γ. That is, for $\xi_1, \xi_2 \in L^*$, $\gamma \in \Gamma$, $(\xi_1 \wedge \xi_2)(\gamma) = \xi_1(\gamma) \wedge \xi_2(\gamma)$.
- The smallest element in L^* is 0^*, where $0^*(\gamma) = 0$ for each $\gamma \in \Gamma$.
- The largest element in L^* is Ω^*, where $\Omega^*(\gamma) = \Omega$ for each $\gamma \in \Gamma$.

Definition: $^\circ : \Gamma \times E^* \to \Gamma$ is a partially defined function such that for each $\gamma \in \Gamma$ and $(m,n) \in E^*$ s.t. $CM^{-1}(\gamma) \cap IVP(r_1,m) \neq \emptyset$, we have:

$\gamma^\circ (m,n) =$

- γ if $(m,n) \in E^0$
- $\gamma || [m]$ if (m,n) is a call edge in E^1
- $\gamma(1:\#\gamma-1)$ if (m,n) is a return edge st. $\gamma(\#\gamma)$ is its corresponding call edge.

Lemma: Let $\gamma \in \Gamma$, $(m,n) \in E^*$, $q \in IVP(r_1,m)$ s.t. $CM(q)=\gamma$. Then $\gamma_1=\gamma^\circ(m,n)$ is defined iff $q_1=q||(m,n) \in IVP(r_1,n)$, in which case $CM(q_1) = \gamma_1$.



$$\lambda \circ (c_1, r_2) = (c_1)$$

$$(c_1) \circ (c_2, r_2) = (c_1 c_2)$$

$$(c_1 c_2) \circ (e_2, n_2) = (c_1)$$

$$(c_1 c_2) \circ (e_2, n_1) = \perp$$

Definition of F^*

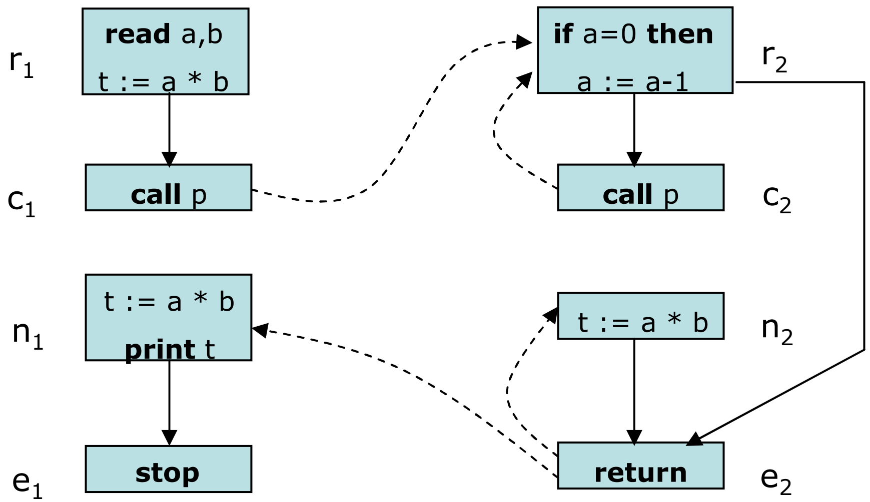
Let $(m,n) \in E^*$ and let $f_{(m,n)}$ be the data propagation map associated with (m,n). Define $f^*_{(m,n)}: L^* \to L^*$ as follows:

For each $\xi \in L^*$, $\gamma \in \Gamma$,

$f^*_{(m,n)}(\xi(\gamma)) =$

- $f_{(m,n)}(\xi(\gamma_1))$ if there exists a γ_1 st. $\gamma_1 \circ (m,n) = \gamma$
- Ω otherwise.

Intuitive interpretation: $f^*_{(m,n)}(\xi)$ represents information at the start of n which is obtained by propagation of the information ξ at the start of m along the edge (m,n).



Let $\xi_0 = \{(\lambda,1)\} \in L^*$. Then

$\xi_1 = f^*_{(c_1,r_2)}(\xi_0) = \{(c_1,1)\}$, since $\lambda \circ (c_1,r_2) = c_1$.

$\xi_2 = f^*_{(c_2,r_2)}(\xi_2) = \{(c_1,f_{(r_2,c_2)}(1))\} = \{(c_1,0)\}$.

$\xi_3 = f^*_{(c_2,r_2)}(\xi_2) = \{(c_1c_2,0)\}$.

$\xi_4 = f^*_{(r_2,e_2)}(\xi_3) = \{(c_1c_2,0)\}$.

$\xi_5 = f^*_{(e_2,n_2)}(\xi_4) = \{(c_1,0)\}$.

$\xi_6 = f^*_{(n_2,e_2)}(\xi_5) = \{(c_1,1)\}$.

$\xi_7 = f^*_{(e_2,n_1)}(\xi_6) = \{(\lambda,1)\}$.

Definition of F^*

F^* is the smallest set of maps $L^* \to L^*$ which contains $\{f^*_{(m,n)} : (m,n) \in E^*\}$, id_{L^*}, and is closed under functional composition and meet.

Lemma:

- If F is monotone in L, then F^* is monotone in L^*.
- If F is distributive in L, then F^* is distributive in L^*.
- If F is distributive in L, then for each $(m,n) \in E^*$, $f^*_{(m,n)}$ is continuous in L^*, i.e., $f^*_{(m,n)}(\wedge_k \xi_k) = \wedge_k f^*_{(m,n)}(\xi_k)$, for every collection $\{\xi_k\}_{k \geq 1} \subseteq L^*$.

DFP for G^*

Find the MFP solution of the following equations:

$x^*_{r_1} = \{(\lambda,0)\}$, where λ is the empty call string.

$x^*_n = \wedge_{(m,n) \in E^*} f^*_{(m,n)}(x^*_m)$, $n \in N^*-\{r_1\}$.

Existence of a solution?

Simple induction on iteration number.

Convert this solution to values in L:

For each $n \in N^*$, $x'_n = \wedge_{\gamma \in \Gamma} x_n^*(\gamma)$.

MFP vs. MOP?

Definition: Let $path_{G^*}(r_1,n)$ denote the set of all execution paths leading from r_1 to $n \in N^*$. For each $p = (r_1,s_2,\dots,s_k,n) \in path_{G^*}(r_1,n)$, define $f_p^* = f_{(s_k,n)} \circ f_{(s_{k-1},s_k)} \circ \cdots \circ f_{(r_1,s_2)}$. For each $n \in N^*$, define (MOP)

$$y^*_n = \wedge \{f_p^*(x_{r_1}^*) : p \in path_{G^*}(r_1,n)\}.$$

Theorem: If (L,F) is a distributive DFF, then, for each $n \in N^*$, $x_n^* = y_n^*$.

Summary - Call String Approach

- If Γ is infinite (G^* has recursive procedures) $\Rightarrow$ not a feasible solution.
- Practical variant of this algorithm is given in the paper.
- Further extensions by keeping track of semantic restrictions (call-return being a special case) that control flow paths.