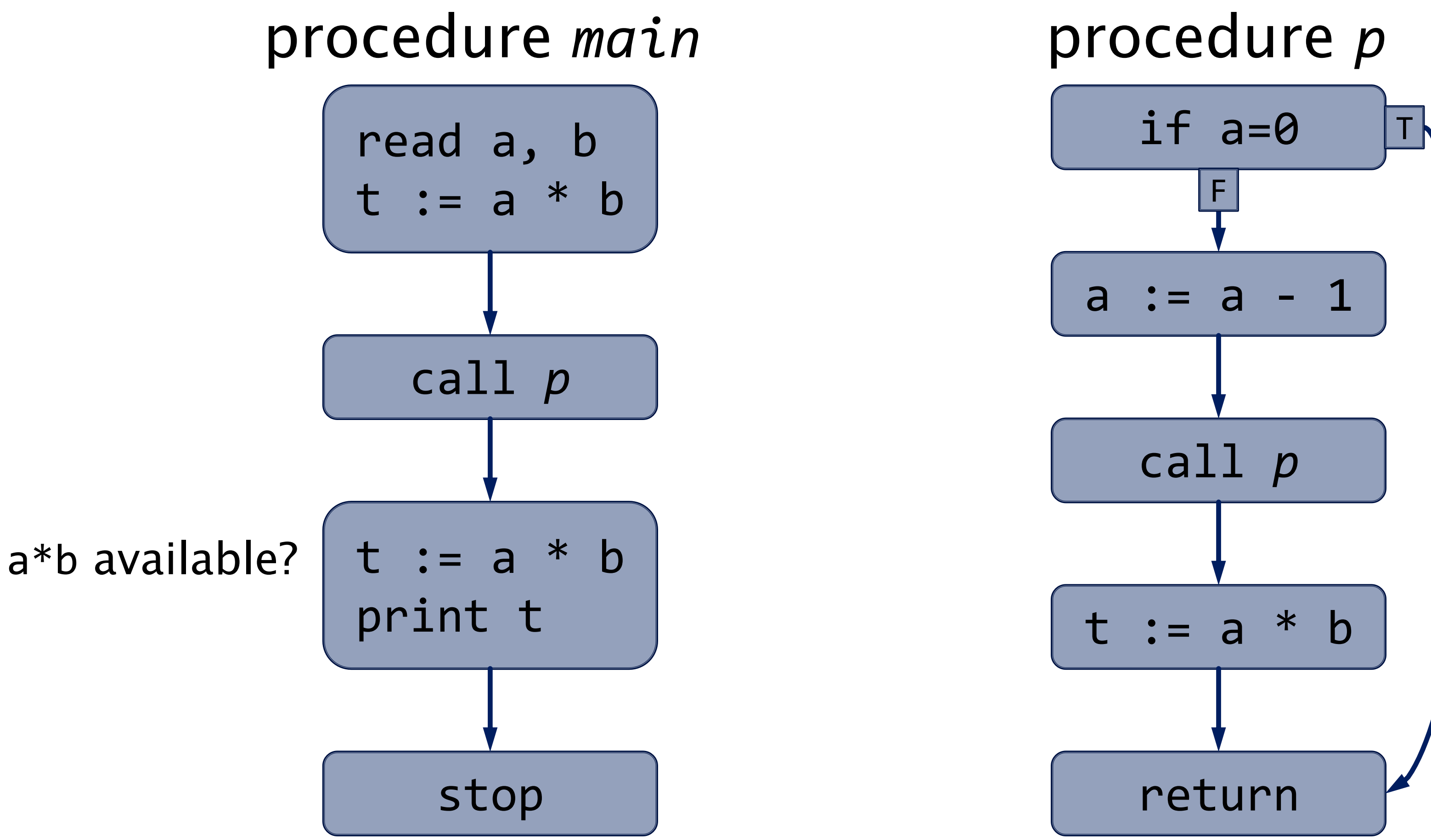
Two Approaches to Interprocedural Data Flow Analysis

Micha Sharir Amir Pnueli

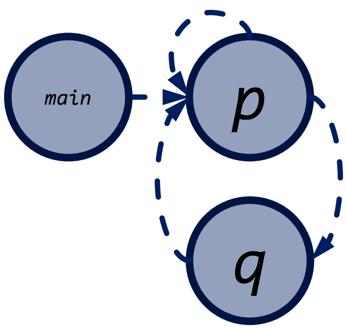
Part one: The Functional Approach

Intraprocedural analysis



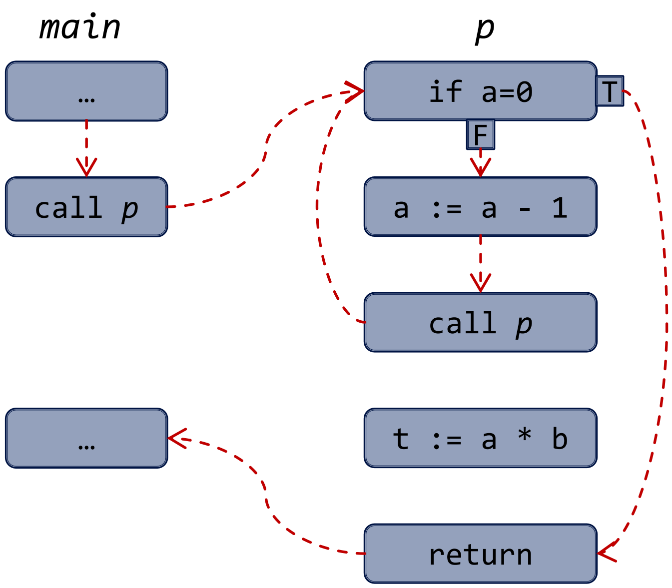
Interprocedural challenges

Recursion



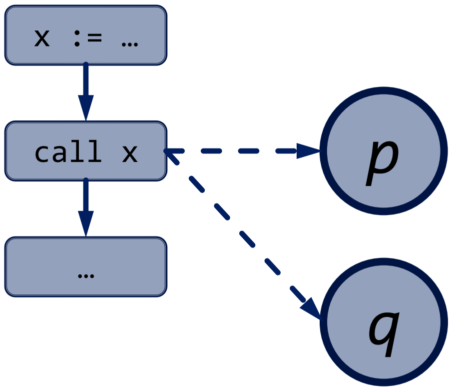
- Infinite paths
- Efficiency vs. Precision

Infeasible paths



- Filter invalid paths
- Precision and Efficiency

Function variables & Virtual functions



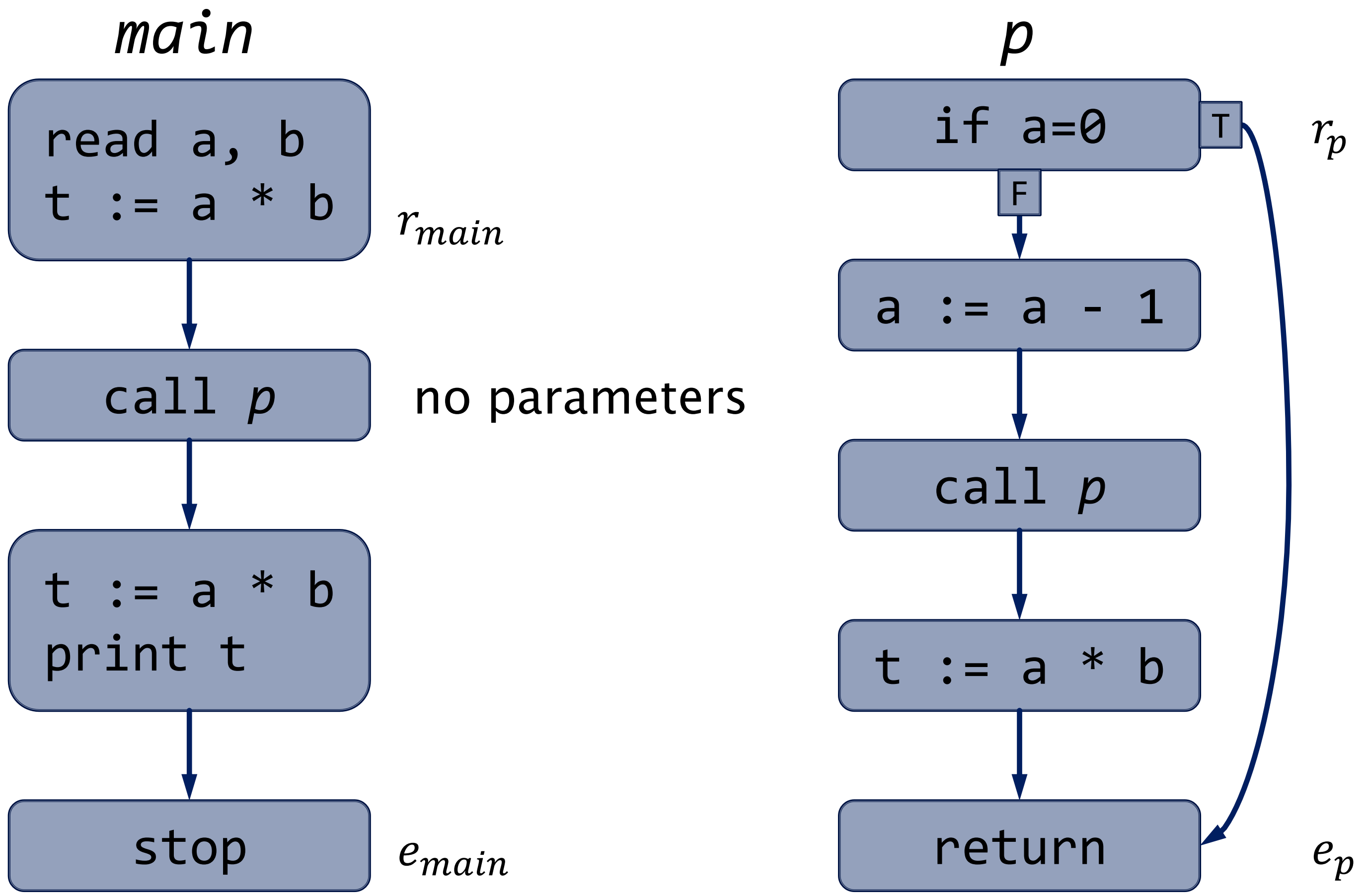
- No static call graph

Outline

- **Notation and Review**
- Functional Approach
- Interprocedural MOP
- Pragmatic Considerations

Notations

Control Flow Graphs

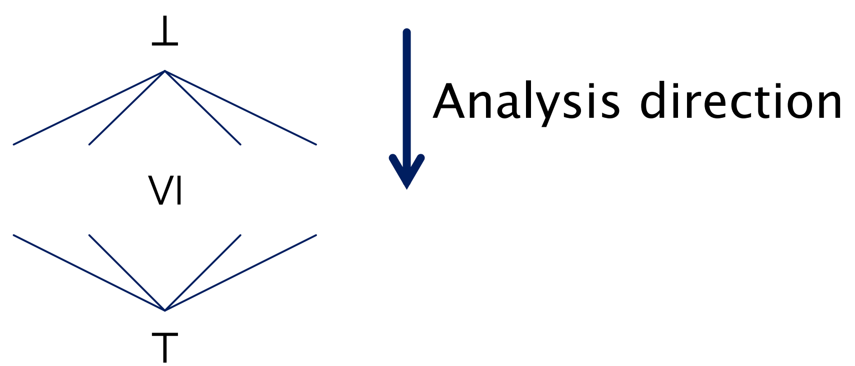


Data Flow Frameworks

- (L, F) is a data flow framework:

  L is a meet-semilattice

  - $\wedge$ = greatest lower bound
  - $\top$ = smallest element (no information)
  - $\perp$ = largest element ("undefined")
  - *bounded* – No infinite descending chain



Data Flow Frameworks

- (L, F) is a data flow framework:

  F is a monotone space of transfer functions

  1. Closed under composition and meet $(f \wedge g)(x) = f(x) \wedge g(x)$
  2. Contains $id_L(x) = x$ and $f_\perp(x) = \perp$

  F is *distributive* iff $\forall f, x, y$: $f(x) \wedge f(y) = f(x \wedge y)$

- Restrict F to graph $G = (N, E)$:

  Smallest $S \subseteq F$ such that $\{f_{(m,n)} \mid (m, n) \in E\} \subseteq S$ and 1. and 2. hold

Intraprocedural example

- Available expression framework for the single expression a * b:

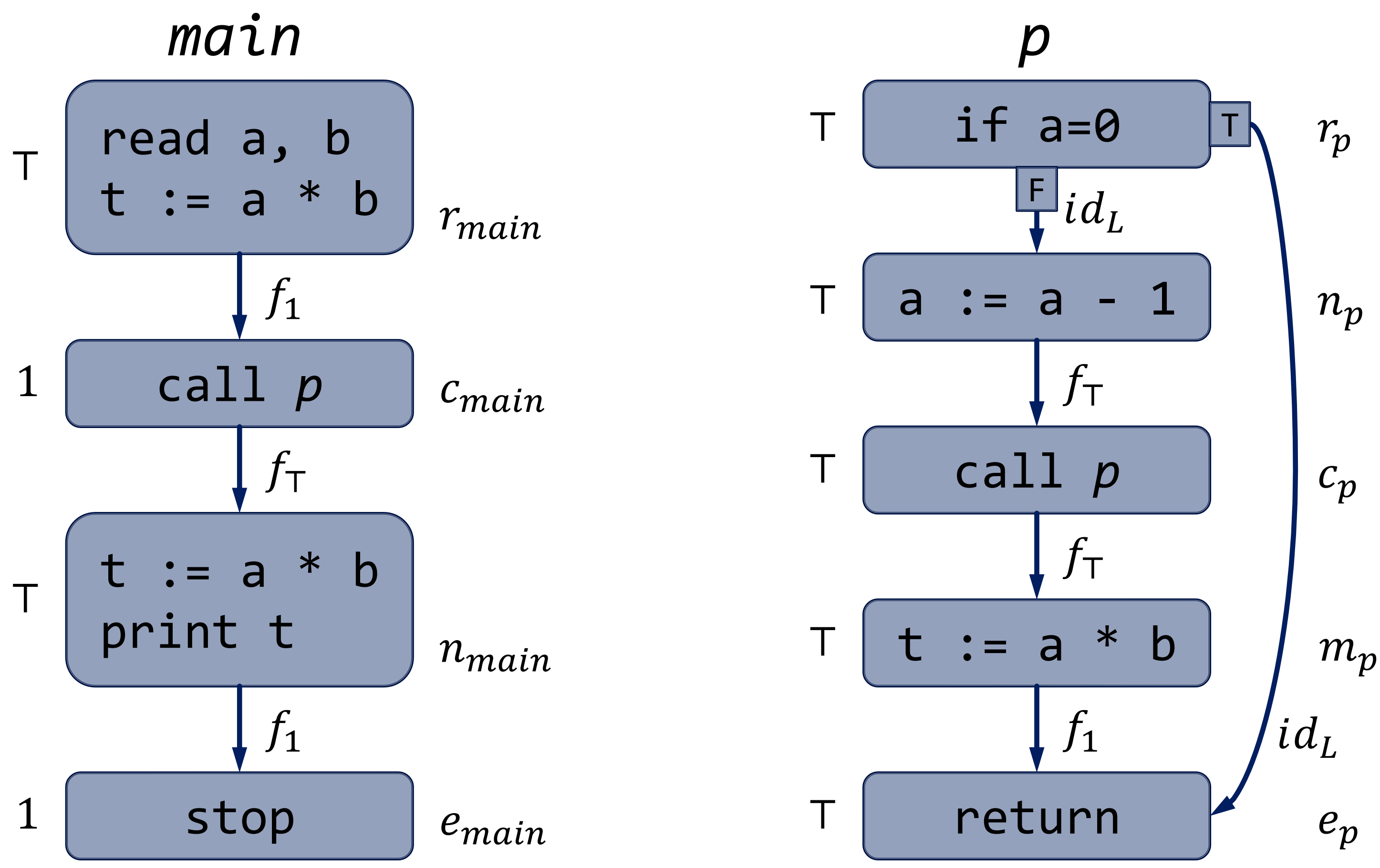
$L = \{\top, 1, \bot\} \quad F = \{f_\top, f_1, id_L, f_\bot\}$

⊤: a * b not available

1: a * b available

$f_\top(x) = \top,\ f_1(x) = 1$

Intraprocedural example



Intraprocedural equations

The data flow equations

$$x_r = \top$$

$$x_n = \bigwedge_{(m,n)\in E} f_{(m,n)}(x_m) \qquad n \in N - \{r\}$$

approximate the meet-over-all paths (MOP) solution

$$y_n = \bigwedge \{f_p(\top) | p \in path_G(r,n)\} \qquad n \in N$$

where $f_{p=(n_1,\ldots,n_k)} = f_{(n_{k-1},n_k)} \circ \cdots \circ f_{(n_1,n_2)}$

Intraprocedural solutions

F is *distributive* $\Longrightarrow$

The maximum fixed point solution $x_n^* = y_n$

F is *monotone* $\Longrightarrow x_n^* \leq y_n$

Outline

- Notation and Review
- **Functional Approach**
- Interprocedural MOP
- Pragmatic Considerations

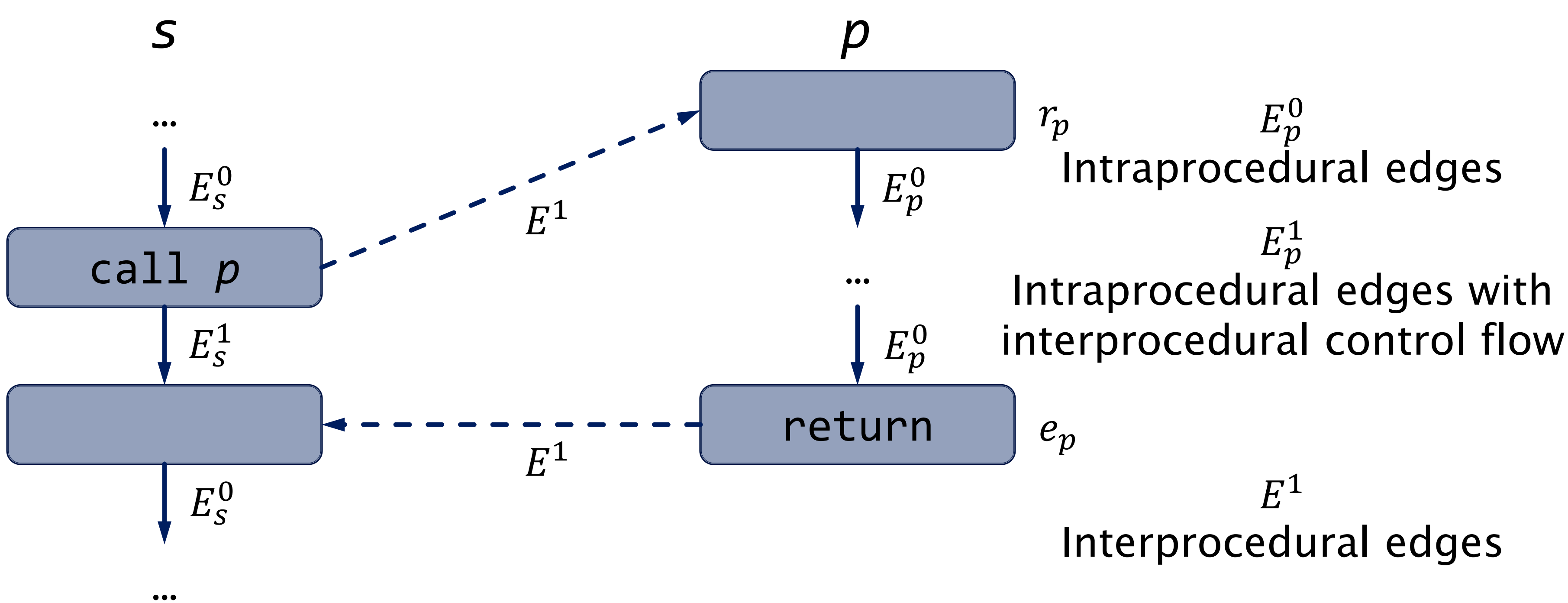
Interprocedural Graphs

Two representations:

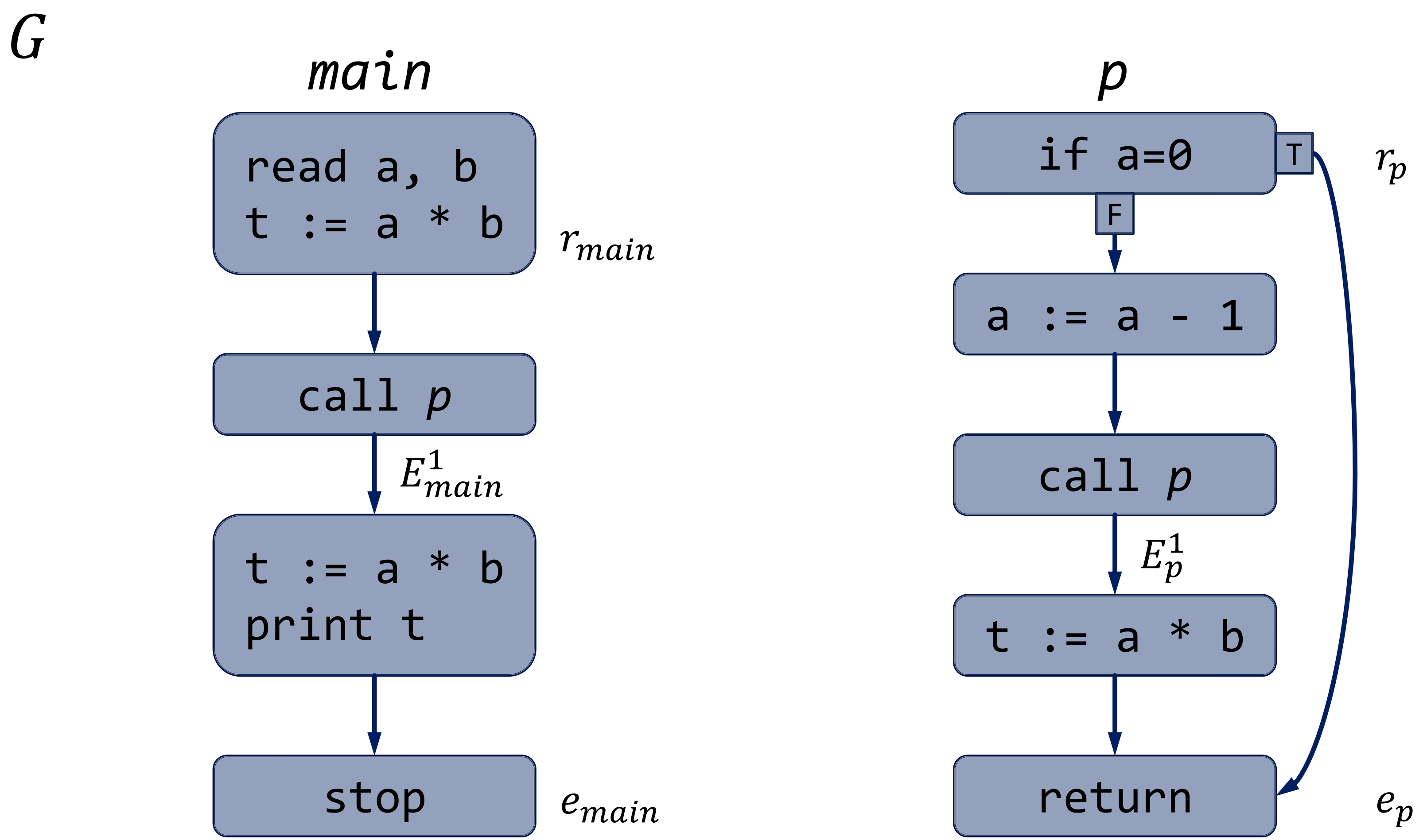
1. $G = \left(\bigcup_p N_p, \bigcup_p E_p\right)$
2. $G^* = \left(\bigcup_p N_p, E^*\right)$

$$E_p = E_p^0 \cup E_p^1 \qquad E^0 = \bigcup_p E_p^0$$

$$E^* = E^0 \cup E^1$$

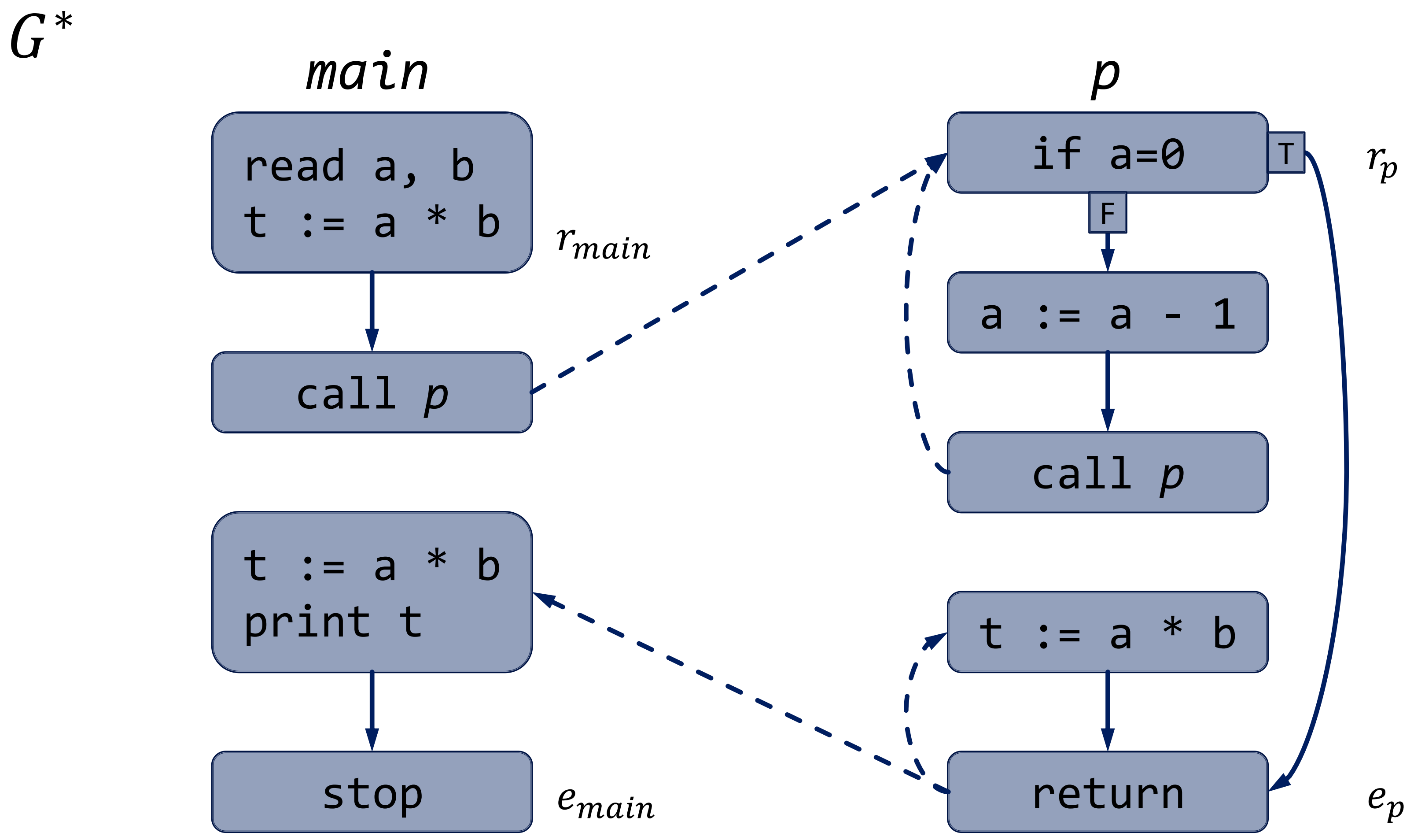


Example



G
$main$
read a, b
t := a * b
r_{main}
call p
E^1_{main}
t := a * b
print t
stop
e_{main}
p
if a=0
T
F
r_p
a := a - 1
call p
E^1_p
t := a * b
return
e_p

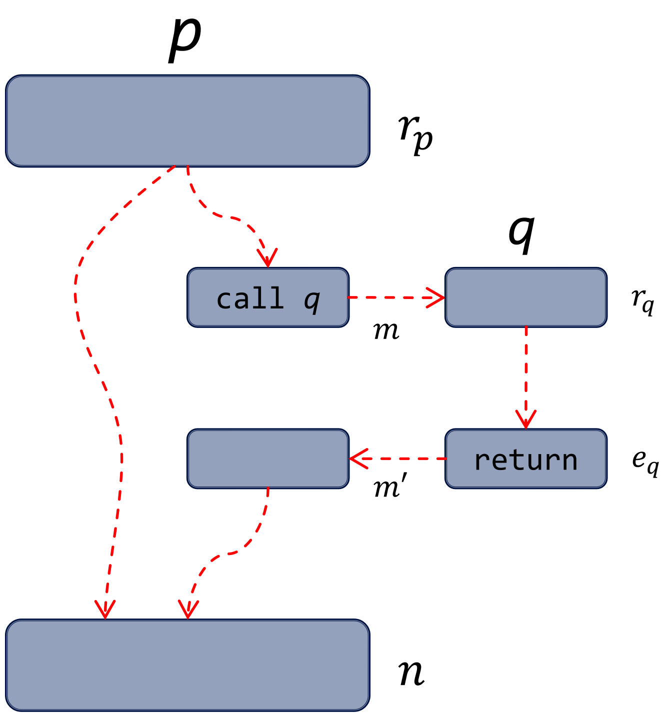
Example



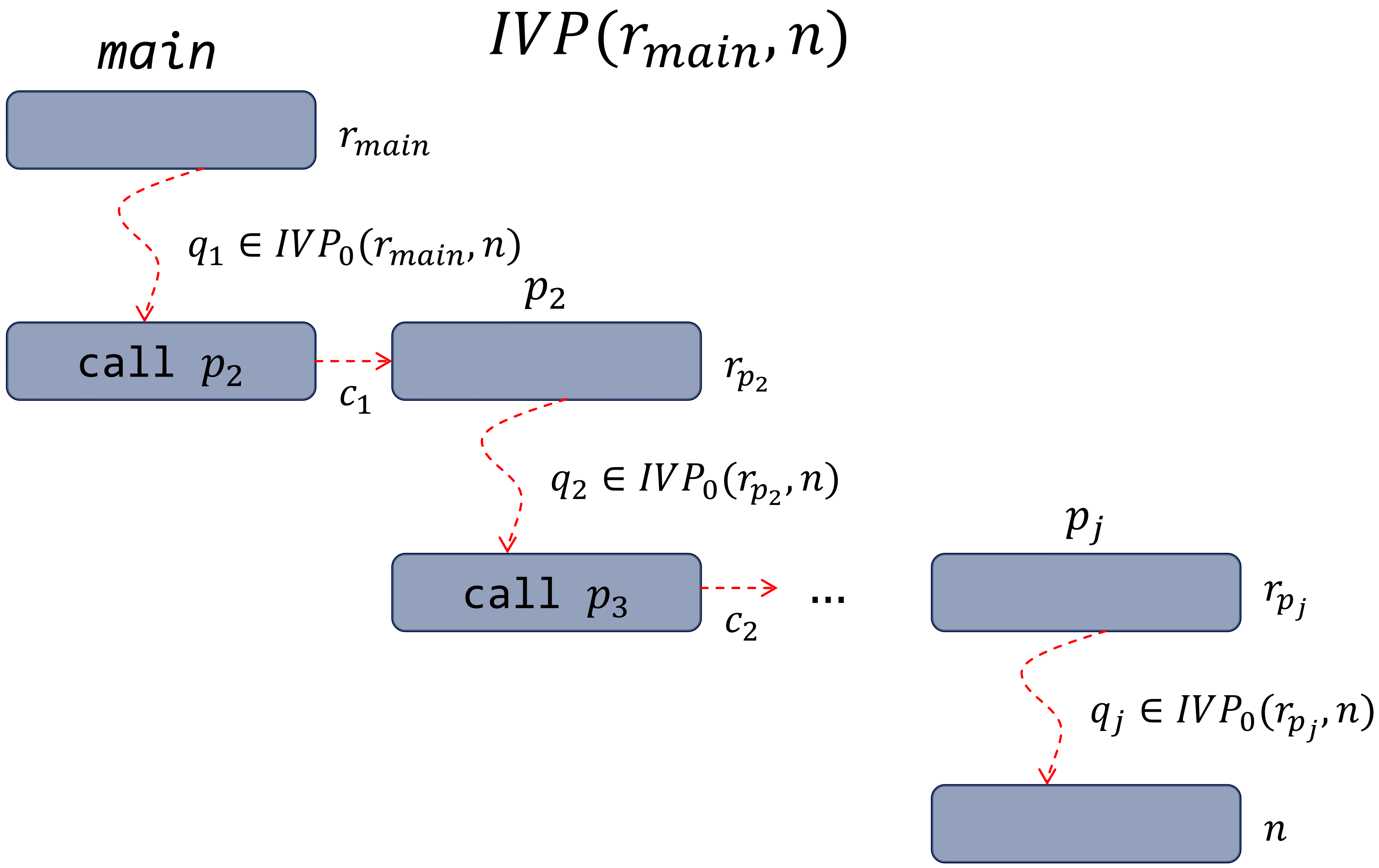
G*
main
p
read a, b
t := a * b
r_main
call p
t := a * b
print t
stop
e_main
if a=0
T
F
r_p
a := a - 1
call p
t := a * b
return
e_p

Interprocedurally Valid paths

$$IVP_0(r_p, n)$$



Interprocedurally Valid paths



Path notations

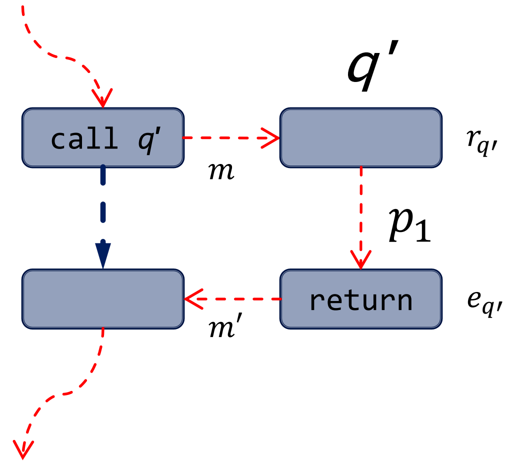
- $p_1, p_2 \in path_{G^*}(r_q, n)$
- $p_1|_{E^1} \coloneqq$ Sequence of call & return edges in p_1
- $p = p_1 \parallel p_2 \coloneqq$ Concatenation of p_1, p_2

Interprocedurally Valid paths

- $p \in path_{G^*}(r_q, n)$ *is in* $IVP_0(r_q, n) \Leftrightarrow$

  $p|_{E^1}$ *is complete*

  defined as:



1. $p|_{E^1} = \varepsilon$
2. $p|_{E^1} = p_1 \parallel p_2$ and p_1, p_2 *are complete*
3. $p|_{E^1} = (m, r_{q'}) \parallel p_1 \parallel (e_{q'}, m')$ and p_1 *is complete*

Interprocedurally Valid paths

- $q \in IVP(r_{main}, n) :\Leftrightarrow$

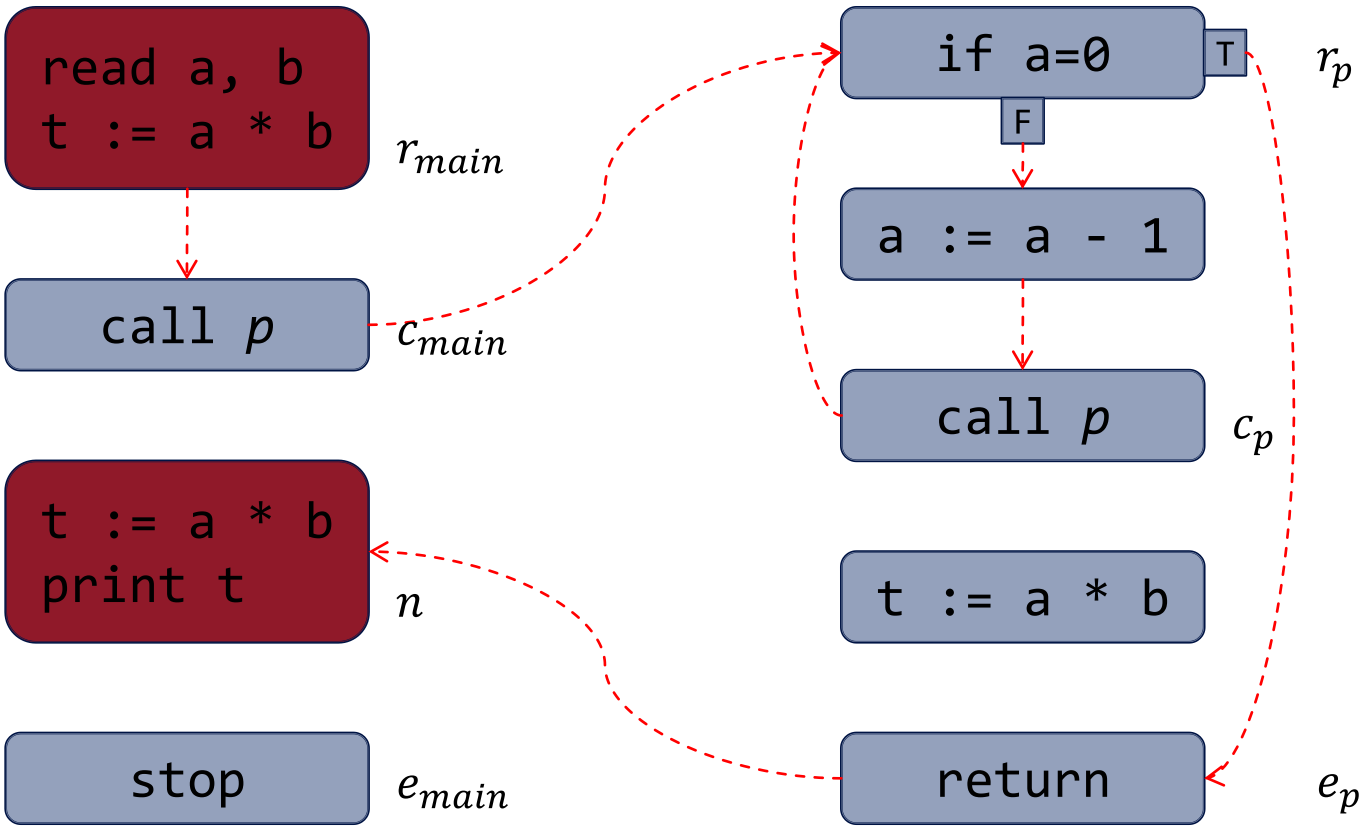
  $q = q_1 \parallel (c_1, r_{p_2}) \parallel q_2 \parallel \cdots \parallel \left(c_{j-1}, r_{p_j}\right) \parallel q_j$

  $\forall i < j: q_i \in IVP_0(r_{p_i}, c_i)$ and $q_j \in IVP_0\left(r_{p_j}, n\right)$
- Also called Path Decomposition

Examples

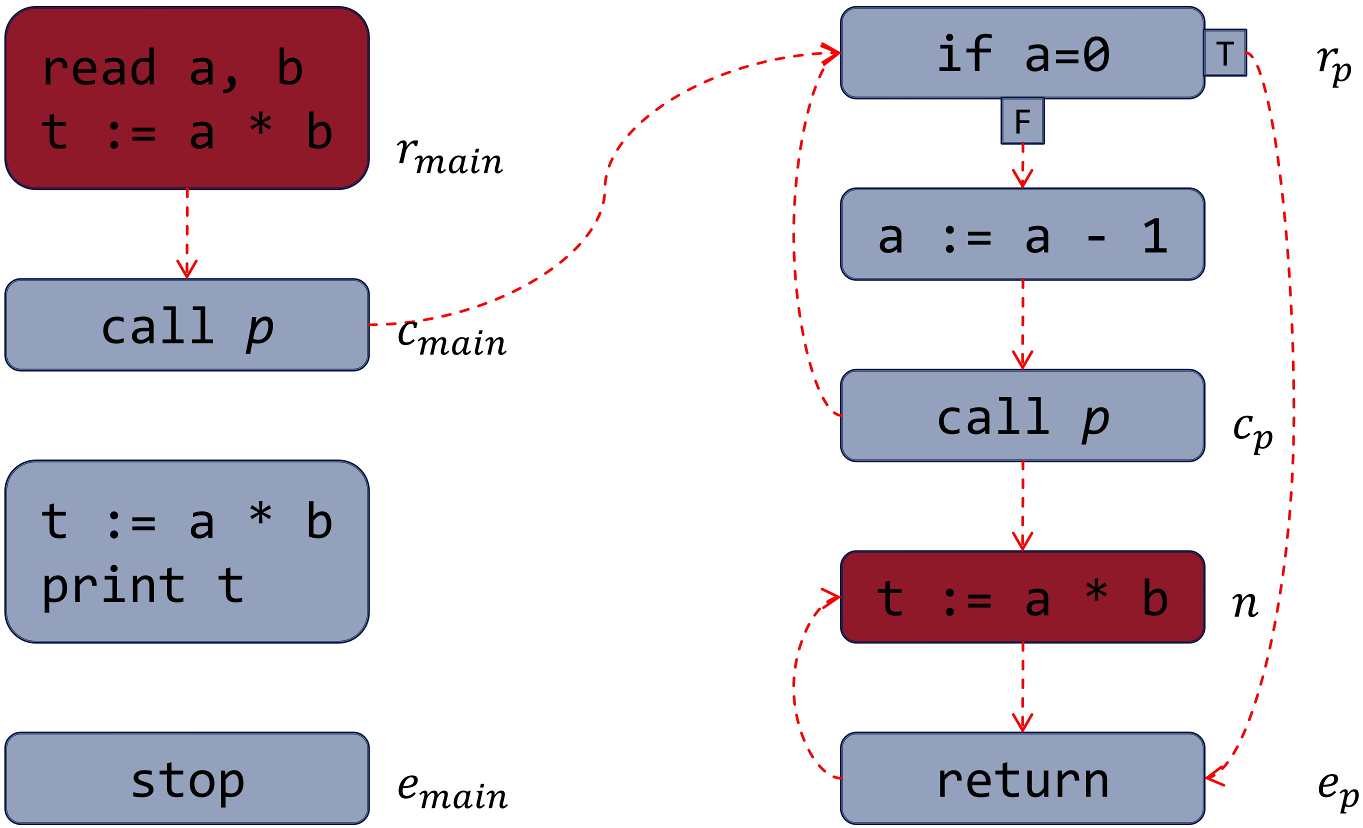
G^*

$$(c_{main}, r_p), (c_p, r_p), (e_p, n)$$

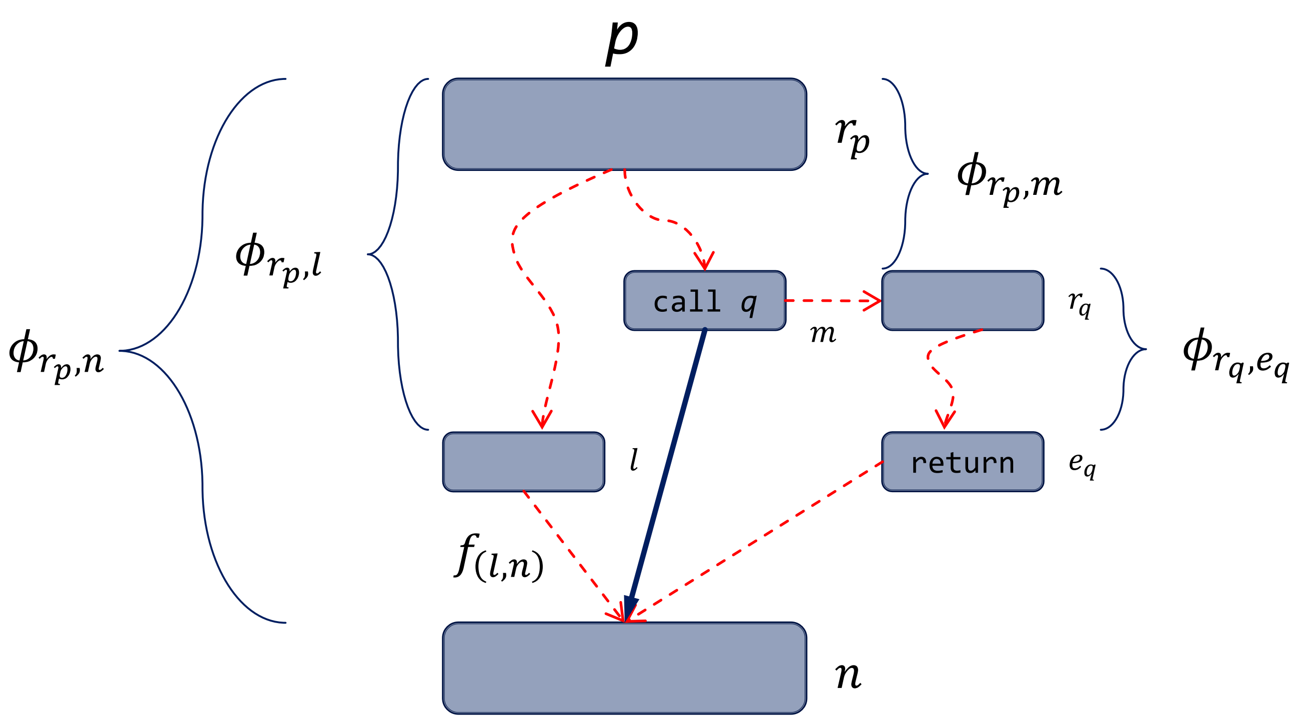


Examples

$$G^* \quad (c_{main}, r_p), (c_p, r_p), (c_p, r_p), (e_p, n), (e_p, n)$$



Functional approach



Information x at r_p

is transformed to $\phi_{r_p,n}(x)$ at n

Functional Approach equations

$$\phi_{r_p,r_p} = id_L$$

$$\phi_{r_p,n} = \bigwedge_{(m,n) \in E_p} \left(h_{(m,n)} \circ \phi_{r_p,m}\right) \qquad n \in N_p - \{r_p\}$$

$$h_{(m,n)} = \begin{cases} f_{(m,n)} & (m,n) \in E_p^0 \\ \phi_{r_q,e_q} & (m,n) \in E_p^1, m\ calls\ q \end{cases}$$

Recursion implicitly encoded in equations

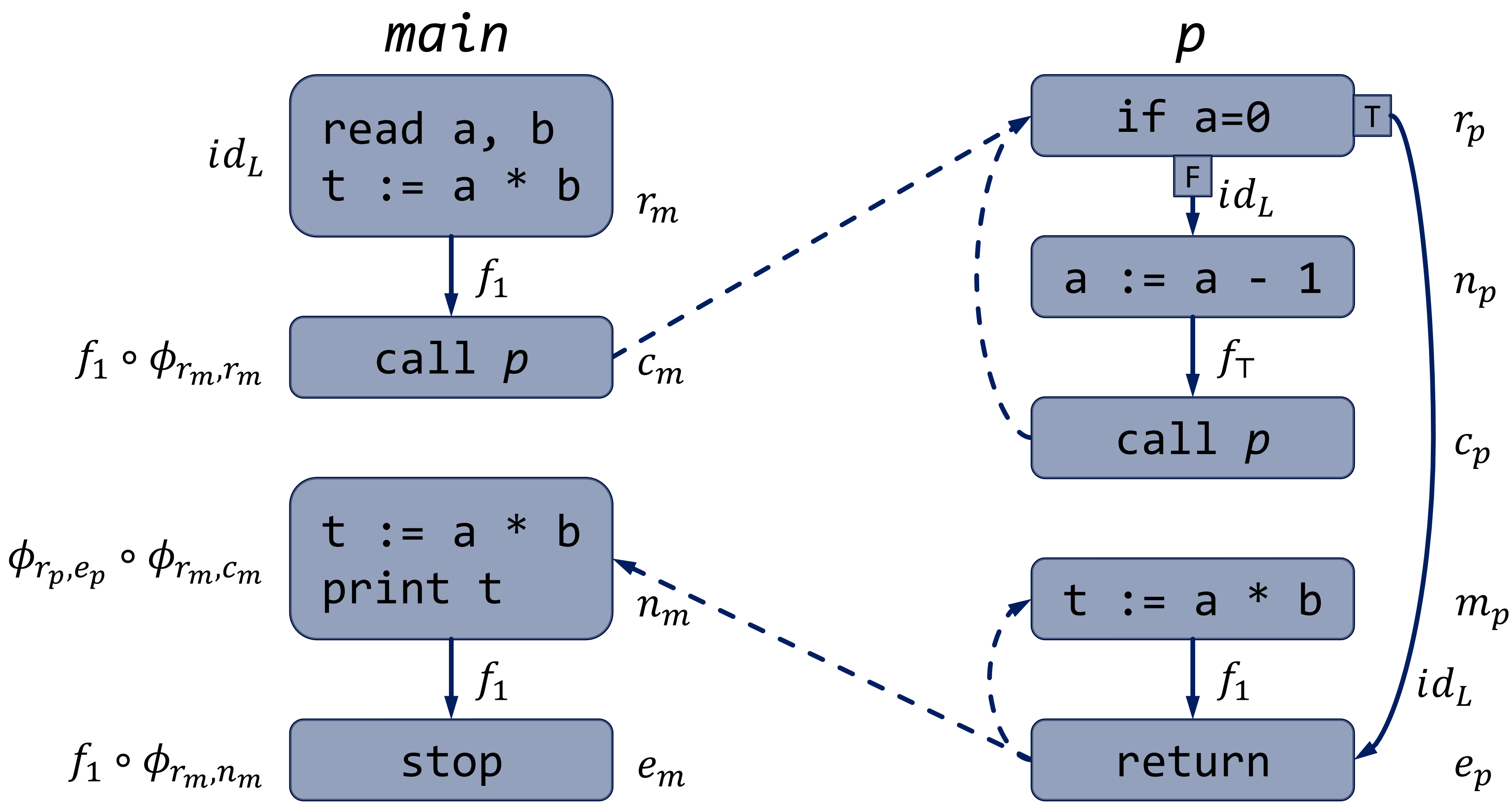
- Initialize the equations with

$$\phi_{r_p,r_p} = id_L$$

$$\phi_{r_p,n} = f_\perp \qquad n \in N_p - \{r_p\}$$

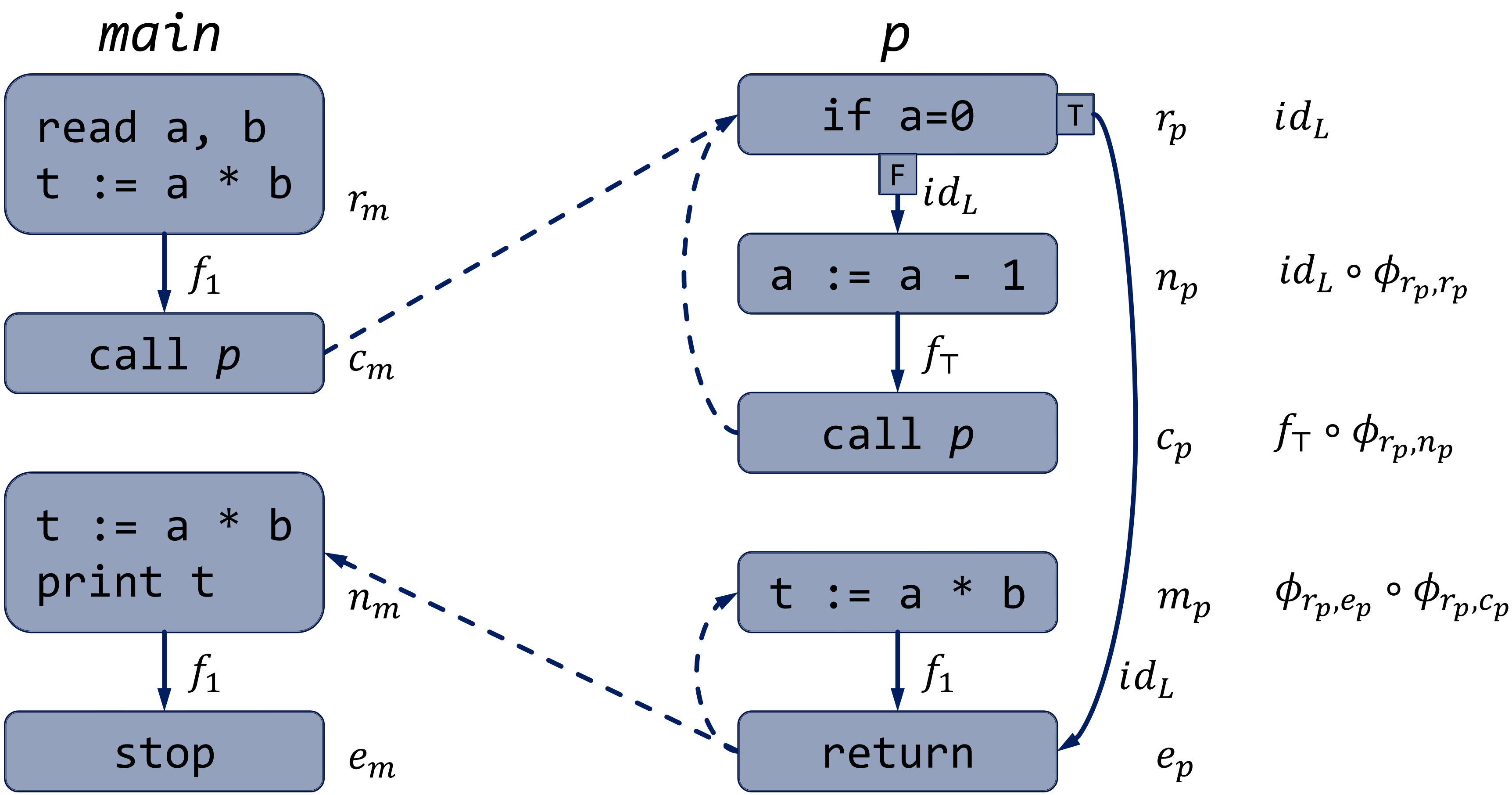
- Compute the maximal fixed point

Example



main
p
read a, b
t := a * b
id_L
r_m
f_1
call p
$f_1 \circ \phi_{r_m,r_m}$
c_m
t := a * b
print t
$\phi_{r_p,e_p} \circ \phi_{r_m,c_m}$
n_m
f_1
stop
$f_1 \circ \phi_{r_m,n_m}$
e_m
if a=0
T
F
r_p
id_L
a := a - 1
n_p
f_T
call p
c_p
t := a * b
m_p
f_1
id_L
return
e_p

Example



$$\left(id_L \circ \phi_{r_p,r_p}\right) \wedge \left(f_1 \circ \phi_{r_p,m_p}\right)$$

Example

$\phi_{r_m,r_m} = id_L$

$\phi_{r_m,c_m} = f_1 \circ \phi_{r_m,r_m}$

$\phi_{r_m,n_m} = \phi_{r_p,e_p} \circ \phi_{r_m,c_m}$

$\phi_{r_m,e_m} = f_1 \circ \phi_{r_m,n_m}$

$\phi_{r_p,r_p} = id_L$

$\phi_{r_p,n_p} = id_L \circ \phi_{r_p,r_p}$

$\phi_{r_p,c_p} = f_\top \circ \phi_{r_p,n_p}$

$\phi_{r_p,m_p} = \phi_{r_p,e_p} \circ \phi_{r_p,c_p}$

$\phi_{r_p,e_p} = \left(id_L \circ \phi_{r_p,r_p}\right) \wedge \left(f_1 \circ \phi_{r_p,m_p}\right)$

Function	Initial value	Iteration 1	2	3
ϕ_{r_m,r_m}	id_L	id_L	id_L	id_L
ϕ_{r_m,c_m}	$f_\perp$	f_1	f_1	f_1
ϕ_{r_m,n_m}	$f_\perp$	$f_\perp$	f_1	f_1
ϕ_{r_m,e_m}	$f_\perp$	f_1	f_1	f_1
ϕ_{r_p,r_p}	id_L	id_L	id_L	id_L
ϕ_{r_p,n_p}	$f_\perp$	id_L	id_L	id_L
ϕ_{r_p,c_p}	$f_\perp$	$f_\top$	$f_\top$	$f_\top$
ϕ_{r_p,m_p}	$f_\perp$	$f_\perp$	$f_\top$	$f_\top$
ϕ_{r_p,e_p}	$f_\perp$	id_L	id_L	id_L

Solution

$$x_{r_{main}} = \top$$

$$x_{r_p} = \bigwedge \left\{ \phi_{r_q,c}\left(x_{r_q}\right) \mid c\ calls\ p\ in\ q \right\}$$

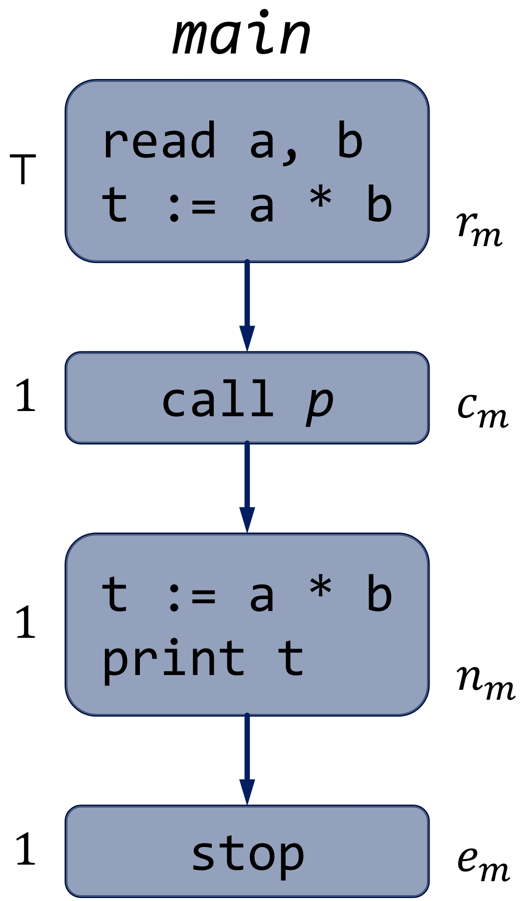
- Compute the maximal fixed point iteratively

$$x_n = \phi_{r_p,n}\left(x_{r_p}\right)$$

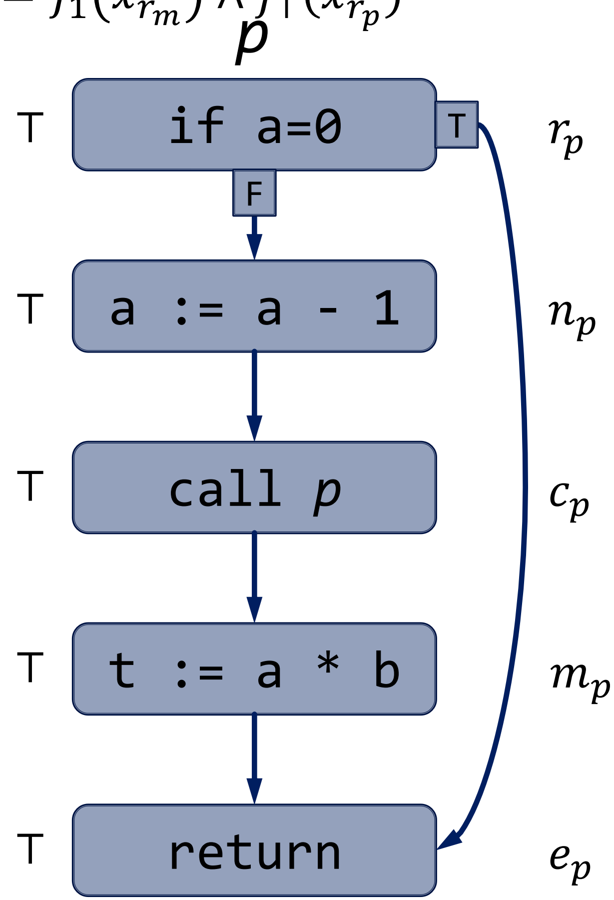
- Computes the solution for all other nodes

Example (continued)

$$x_{r_p} = \phi_{r_m,c_m}(x_{r_m}) \wedge \phi_{r_p,c_p}(x_{r_p})$$
$$= f_1(x_{r_m}) \wedge f_\top(x_{r_p})$$



Functions
$\phi_{r_m,r_m} = id_L$
$\phi_{r_m,c_m} = f_1$
$\phi_{r_m,n_m} = f_1$
$\phi_{r_m,e_m} = f_1$
$\phi_{r_p,r_p} = id_L$
$\phi_{r_p,n_p} = id_L$
$\phi_{r_p,c_p} = f_\top$
$\phi_{r_p,m_p} = f_\top$
$\phi_{r_p,e_p} = id_L$



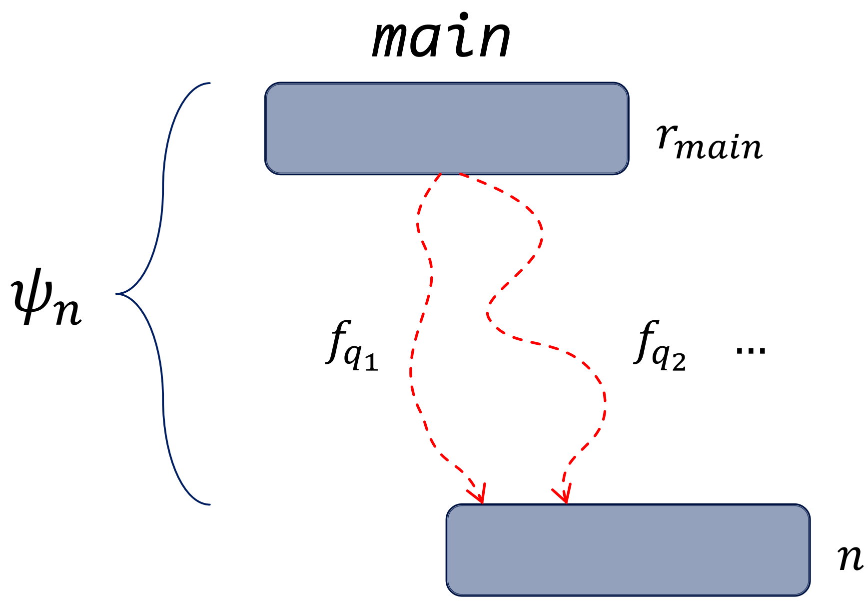
Outline

- Notation and Review
- Functional Approach
- **Interprocedural MOP**
- Pragmatic Considerations

Interprocedural MOP solution

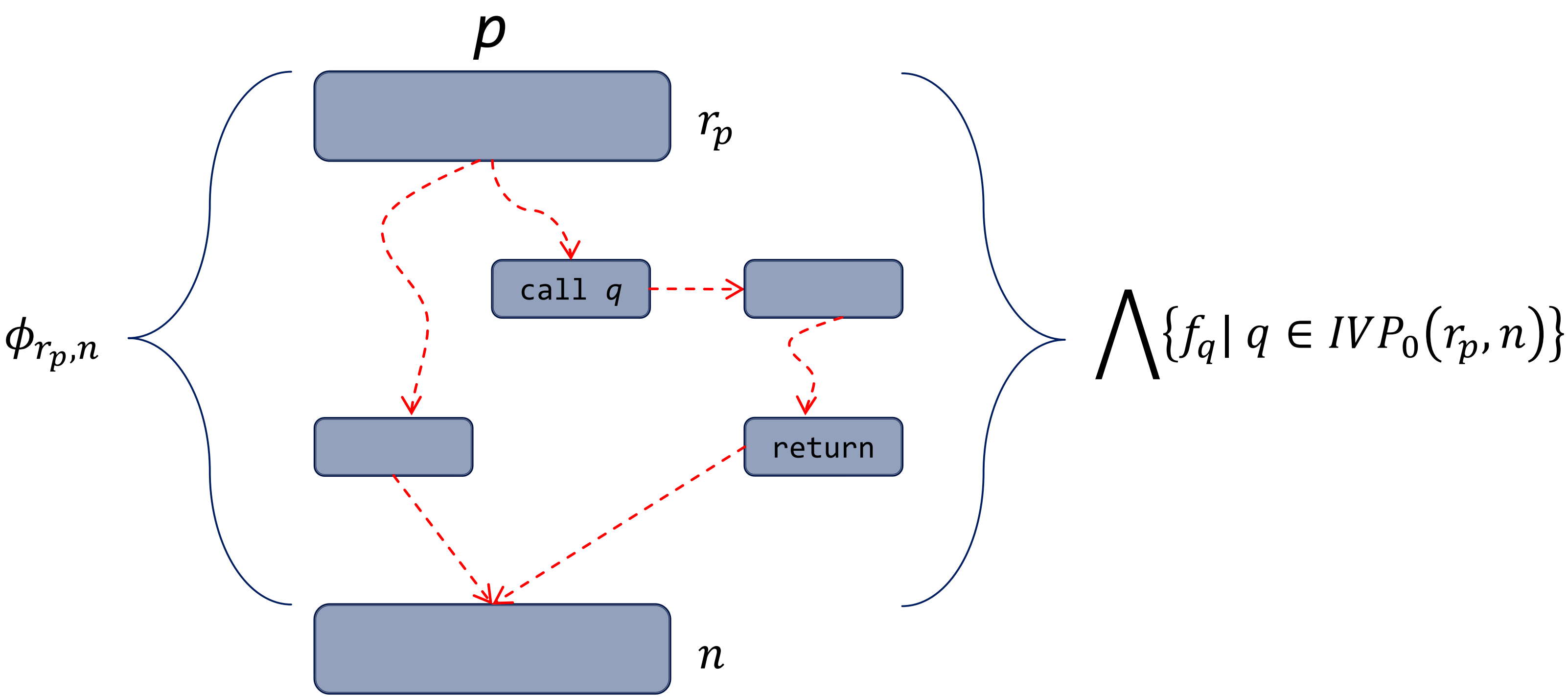
$$\psi_n = \bigwedge \{ f_q \mid q \in IVP(r_{main}, n) \}$$

$$y_n = \psi_n(\top)$$



IVP_0 Lemma

- $\phi_{r_p,n} = \bigwedge\{f_q \mid q \in IVP_0(r_p, n)\}$



IVP_0 Lemma: Proof

- $\phi_{r_p,n} \geq \bigwedge\{f_q \mid q \in IVP_0(r_p, n)\}$
- By induction on $\phi^i_{r_p,n}$, the i-th approximation

For $\mathrm{i} = 0$:

  - If $n = r_p$ then $\phi^0_{r_p,n} = id_L = f_\varepsilon \geq \bigwedge\{f_q \mid q \in IVP_0(r_p, r_p)\}$
  - If $n \neq r_p$ then $\phi^0_{r_p,n} = f_\perp \geq f \in F$

IH: $\phi^i_{r_p,n} \geq \bigwedge\{f_q \mid q \in IVP_0(r_p, n)\}$

For $\mathrm{i} + 1$: see blackboard

IVP_0 Lemma: Proof

- $\bigwedge\{f_q \mid q \in IVP_0(r_p, n)\} \geq \phi_{r_p,n}$
- Show: $\forall q \in IVP_0(r_p, n)$: $f_q \geq \phi_{r_p,n}$ by induction over the length k of q

For $\mathrm{k} = 0$:

  - Then $f_q = \phi_{r_p,r_p} = id_L$

IH: $\forall q \in IVP_0(r_p, n), |q| \leq k$: $f_q \geq \phi_{r_p,n}$

For $\mathrm{k} + 1$: see blackboard

MOP Rewritten

- $\psi_n = \wedge\{f_q |\, q \in IVP(r_{main}, n)\}$
- $\chi_n =$

  $\wedge\left\{\phi_{r_{p_j},n} \circ \phi_{r_{p_{j-1}},c_{j-1}} \circ \cdots \circ \phi_{r_{main},c_1} \middle|\, c_i\ calls\ p_{i+1}\right\}$
- Then $\psi_n = \chi_n$ with IVP_0 Lemma and Path decomposition
- Thus $y_n = \psi_n(\top) = \chi_n(\top)$

Equivalence to MOP

- $F\ distributive \implies y_n = x_n$
- Proof: Show that $x_{r_p} = y_{r_p}$ then with

$$x_n = \phi_{r_p,n}\left(x_{r_p}\right) = \phi_{r_p,n}\left(y_{r_p}\right),\ y_n = \chi_n(\top) =$$

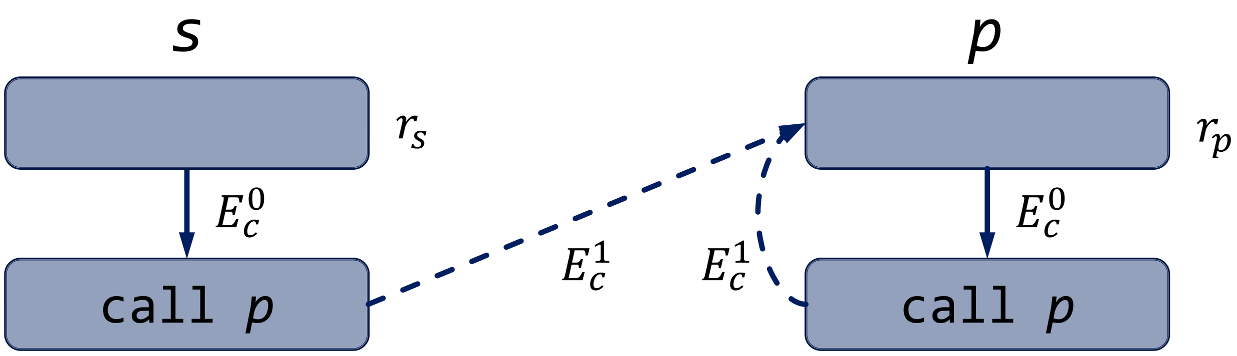
$$\bigwedge\left\{\phi_{r_{p_j},n} \circ \phi_{r_{p_{j-1}},c_{j-1}} \circ \cdots \circ \phi_{r_{main},c_1} \mid c_i\ calls\ p_{i+1}\right\}(\top) =$$

$$\left(\phi_{r_p,n} \circ \chi_{r_p}\right)(\top) = \phi_{r_p,n}\left(y_{r_p}\right)$$

it follows that $x_n = y_n\ \forall n \in N^*$

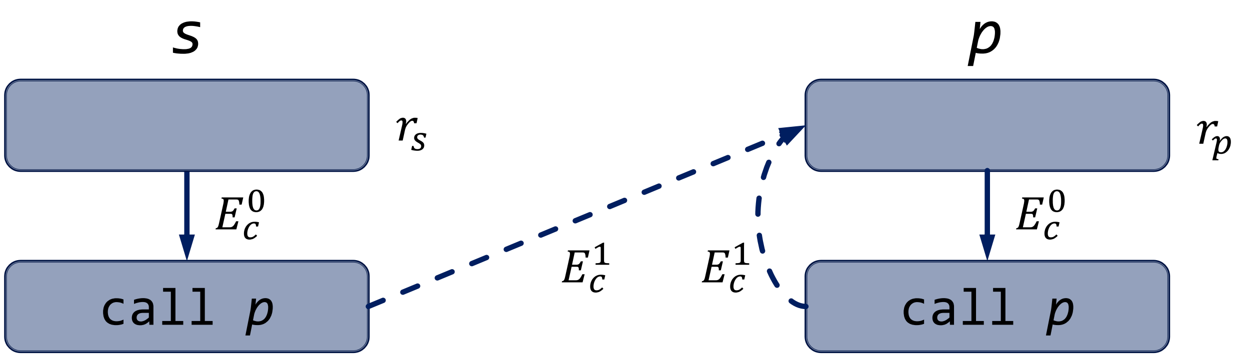
Equivalence to MOP

- New "Intraprocedural" Data Flow Problem:
  - $G_c = (N_c, E_c)$ $E_c = E_c^0 \cup E_c^1$



  - $g_{(m,n)} = \begin{cases} \phi_{m,n}, (m,n) \in E_c^0 \\ id_L, \quad (m,n) \in E_c^1 \end{cases}$

Equivalence to MOP



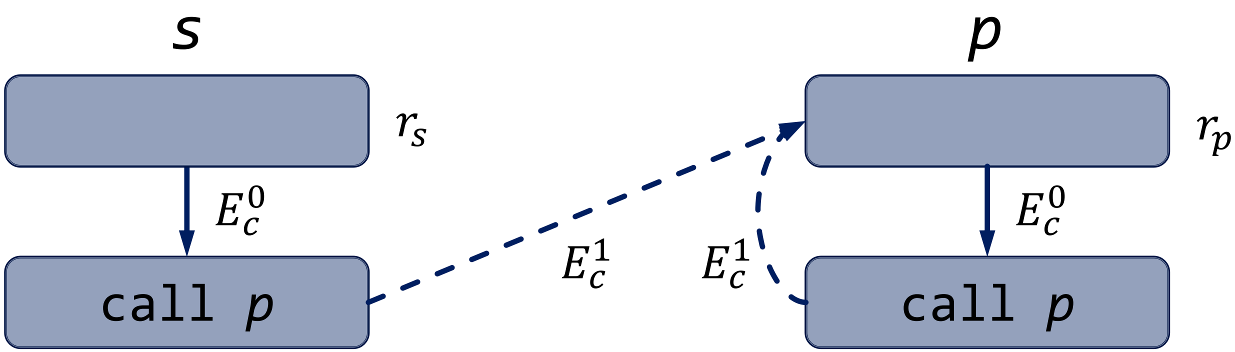
- $g_{(m,n)} = \begin{cases} \phi_{m,n}, (m,n) \in E_c^0 \\ id_L, \quad (m,n) \in E_c^1 \end{cases}$

interprocedural	intraprocedural
$x_{r_{main}} = \top$ $x_{r_p} = \bigwedge \left\{\phi_{r_q,c}\left(x_{r_q}\right) \mid c \text{ calls } p \text{ in } q\right\}$ $x_n = \phi_{r_p,n}\left(x_{r_p}\right)$	$x_r = \top$ $x_n = \bigwedge_{(m,n)\in E} g_{(m,n)}(x_m)$

Equivalence to MOP

interprocedural	intraprocedural
$x_{r_{main}} = \top$	$x_{r_{main}} = \top$
$x_{r_p} = \bigwedge \left\{ \phi_{r_q,c}\left(x_{r_q}\right) \mid c \textit{ calls } p \textit{ in } q \right\}$	$x_{r_p} = \bigwedge_{(m,r_p)\in E_c^1} id_L(x_m) = \bigwedge_{(m,r_p)\in E_c^1} x_m = \bigwedge \left\{ \phi_{r_q,c}\left(x_{r_q}\right) \mid c \textit{ calls } p \textit{ in } q \right\}$
$x_n = \phi_{r_p,n}\left(x_{r_p}\right)$	$x_c = \bigwedge_{(m,c)\in E_c^0} \phi_{m,c}(x_m) = \phi_{r_p,c}\left(x_{r_p}\right)$

Equivalence to MOP



- $g_{(m,n)} = \begin{cases} \phi_{m,n}, & (m,n) \in E_c^0 \\ id_L, & (m,n) \in E_c^1 \end{cases}$

interprocedural	intraprocedural
$y_n = \bigwedge \{f_q \mid q \in IVP(r_{main}, n)\}(\top)$	$y_n = \bigwedge \{f_p(\top) \mid p \in path_G(r,n)\}$

- Follows from $y_n = \wedge \left\{ \phi_{r_{p_j},n} \circ \cdots \circ \phi_{r_{main},c_1} \mid c_i \; calls \; p_{i+1} \right\}(\top)$ and by construction of G_c

Equivalence to MOP

F is *distributive* $\Longrightarrow$

The maximum fixed point solution x_n^* of the functional approach = the interprocedural MOP solution y_n

F is *monotone* $\Longrightarrow x_n^* \le y_n$

Outline

- Notation and Review
- Functional Approach
- Interprocedural MOP
- **Pragmatic Considerations**

Pragmatic problems

- Representation of ϕ:
  - Symbolic not always possible
  - Explicit representation maybe not finite if $L\ infinite$:

```
main:

A := 0
call P
print A
```

```
p:

if cond then
   A := A + 1
   call P
   A := A - 1
endif
return
```

$$\phi_{r_p,e_p}(\{(A,0)\})$$

$$\phi_{r_p,e_p}(\{(A,1)\})$$

$$\phi_{r_p,e_p}(\{(A,2)\})$$

…

Pragmatic problems

- Approach with symbolic representation may not converge when $L\ infinite\ and\ F\ unbounded$:

  $\forall k \geq 0$: Need k iterations to show $\phi_{r_p,e_p}(\{(A,-k)\}) = \emptyset$

```
main:              p:

A := 0             if cond then
call P                A := A + 2 + (A<0)?-1:1
print A               call P
                      A := A - 1
                   endif
                   return
```

Practical result

- If (L, F) $distributive$ and L $finite$ then the iterative solution of

$$\phi_{r_p,r_p} = id_L$$
$$\phi_{r_p,n} = \bigwedge_{(m,n)\,\in\, E_p} \left(h_{(m,n)} \circ \phi_{r_p,m}\right)$$

converges and

$$x_{r_{main}} = \top$$
$$x_{r_p} = \bigwedge \left\{\phi_{r_q,c}\left(x_{r_q}\right) \mid c\ calls\ p\ in\ q\right\}$$
$$x_n = \phi_{r_p,n}\left(x_{r_p}\right)$$

results in the interprocedural MOP solution

Algorithm

- Assume $L\ finite$ and no symbolic representation available
- Compute $\phi_{r_p,n}$ only for necessary values

$$W\ := \{(r_{main}, \top)\};\ \ PHI(r_{main}, \top) = \top$$

$$while\ W \neq \emptyset$$

$$\quad remove\ some\ (n, x)\ from\ W$$

$$\quad update\ PHIs\ and\ W$$

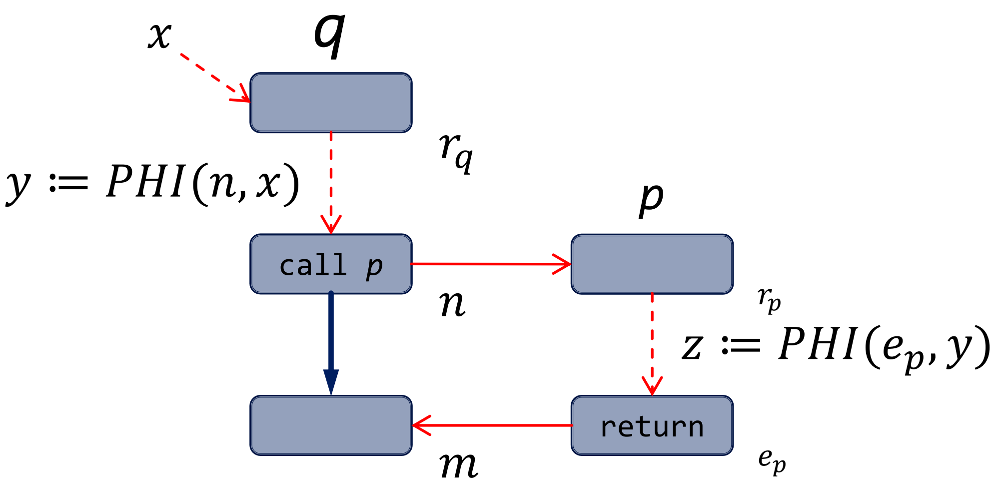
$$x_n =\ \wedge_{a \in L}\ PHI(n, a)$$

Algorithm: updating $PHIs$ and W

$$
\begin{array}{l}
propagate(x, z, m): \\
\quad PHI(m, x) \coloneqq PHI(m, x) \land z; \\
\quad if\ PHI\ changed: W \coloneqq W \cup \{(m, x)\}
\end{array}
$$

Algorithm: updating $PHIs$ and W

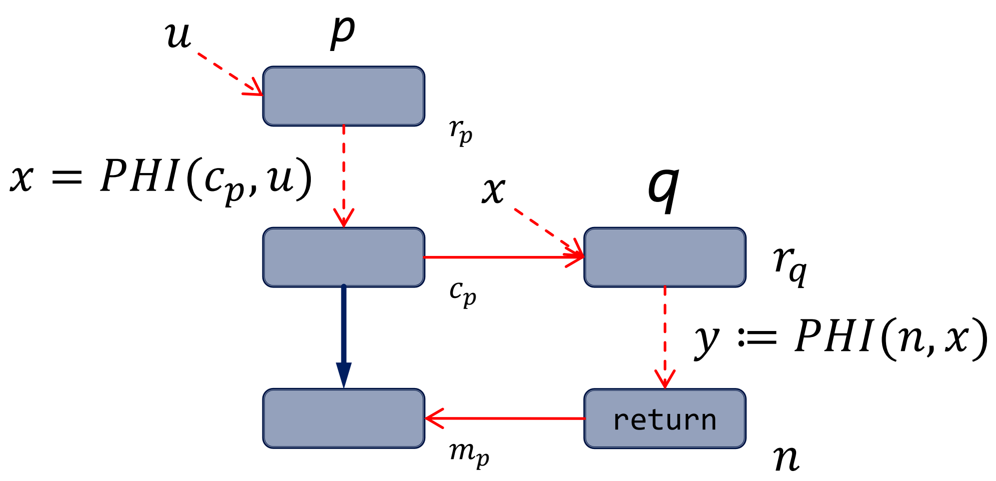
Case 1



$if\ z\ undefined$:
  $propagate(y, y, r_p)$
$else$:
  $propagate(x, z, m)$

Algorithm: updating $PHIs$ and W

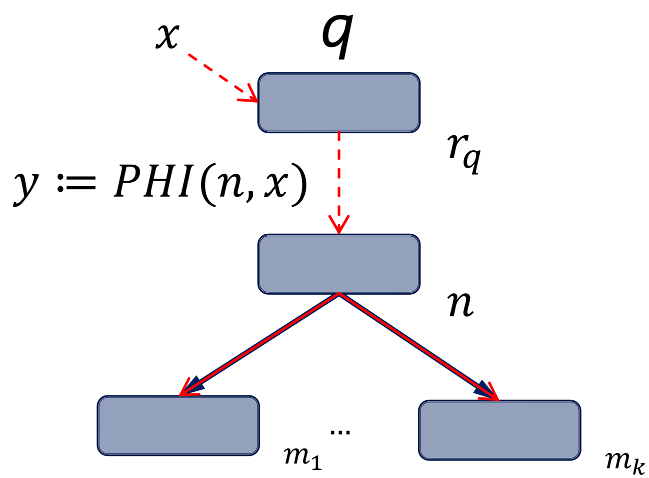
Case 2



$$
\boxed{\begin{array}{l} for\ each\ call\ c_p\ to\ q\ and \\ \quad u \in L\ with\ x = PHI(c_p, u) : \\ \quad\quad propagate(u, y, m_p) \end{array}}
$$

Algorithm: updating $PHIs$ and W

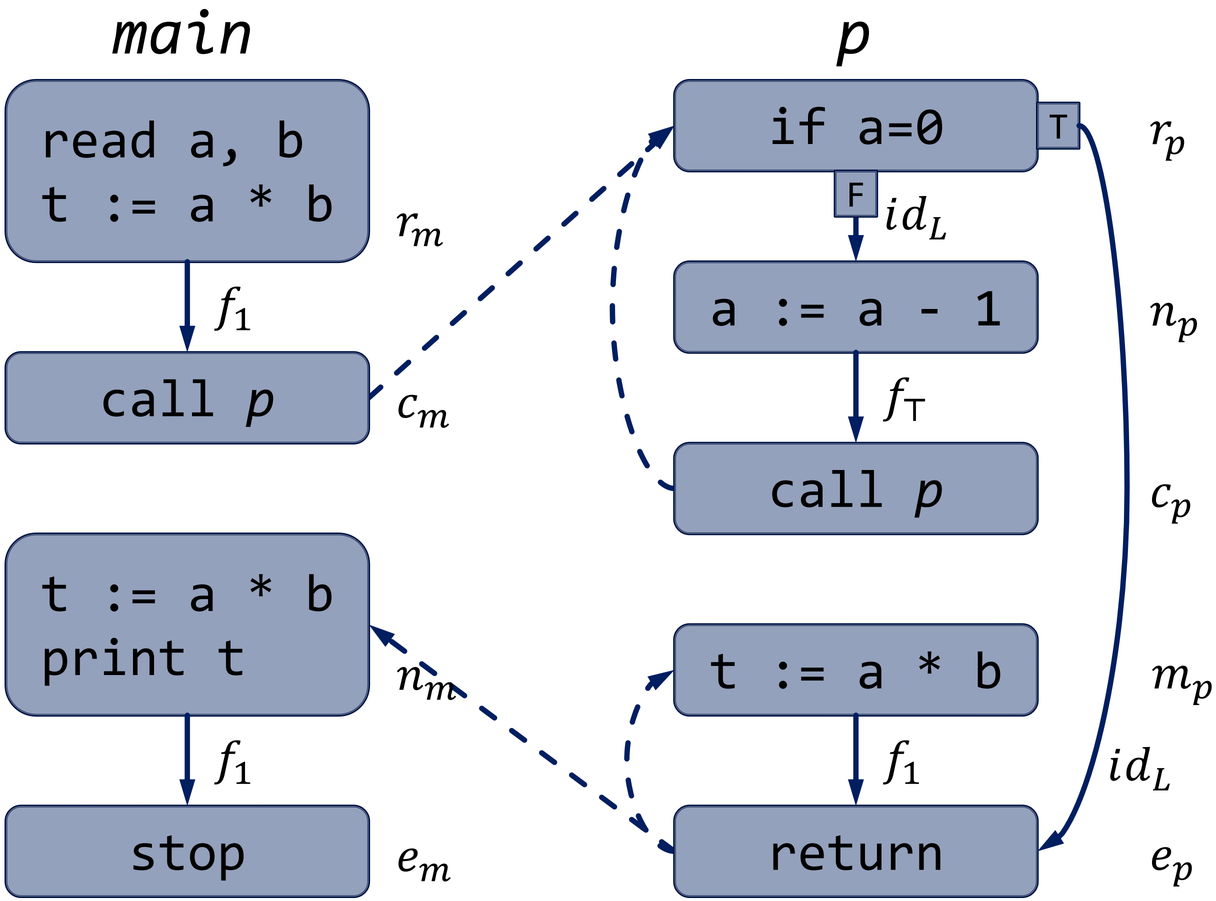
Case 3



$$\boxed{\begin{array}{l} for\ each\ (n,m) \in E_q^0: \\ \quad propagate\big(x, f_{(n,m)}(y), m\big) \end{array}}$$

Example

$$W := \{(r_m, \top) \quad (c_m, \top) \quad (r_p, 1) \quad \ldots \quad (e_p, 1) \quad \ldots\}$$



$PHI(r_m, \top) = \top$

$propagate(\top, f_1(\top), c_m)$

$PHI(c_m, \top) = 1$

$propagate(1, 1, r_p)$

$PHI(r_p, 1) = 1$

$propagate(1, id_L(1), n_p)$

$propagate(1, id_L(1), e_p)$

$PHI(n_p, 1) = 1$

$PHI(e_p, 1) = 1$

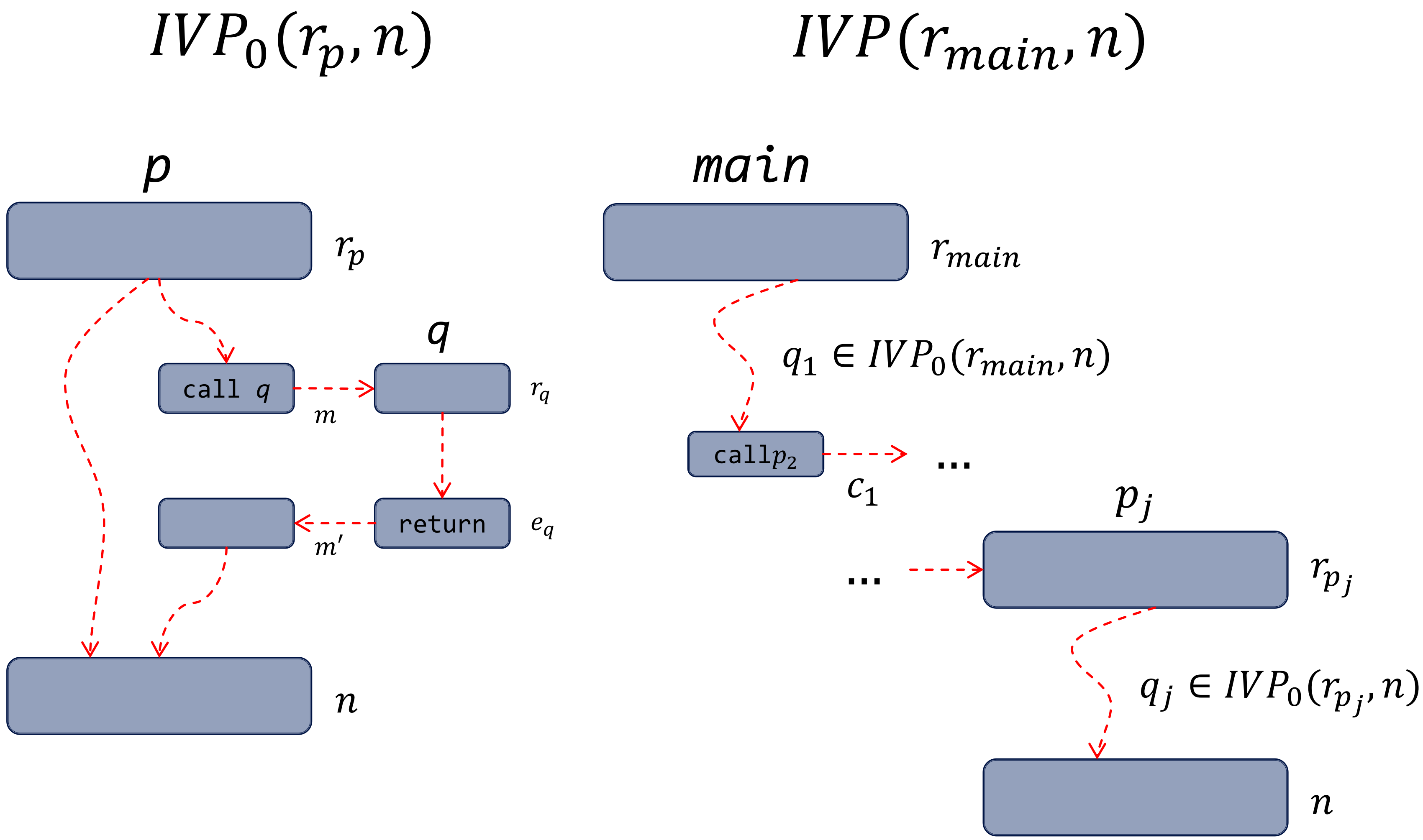
…

$propagate(\top, 1, n_m)$

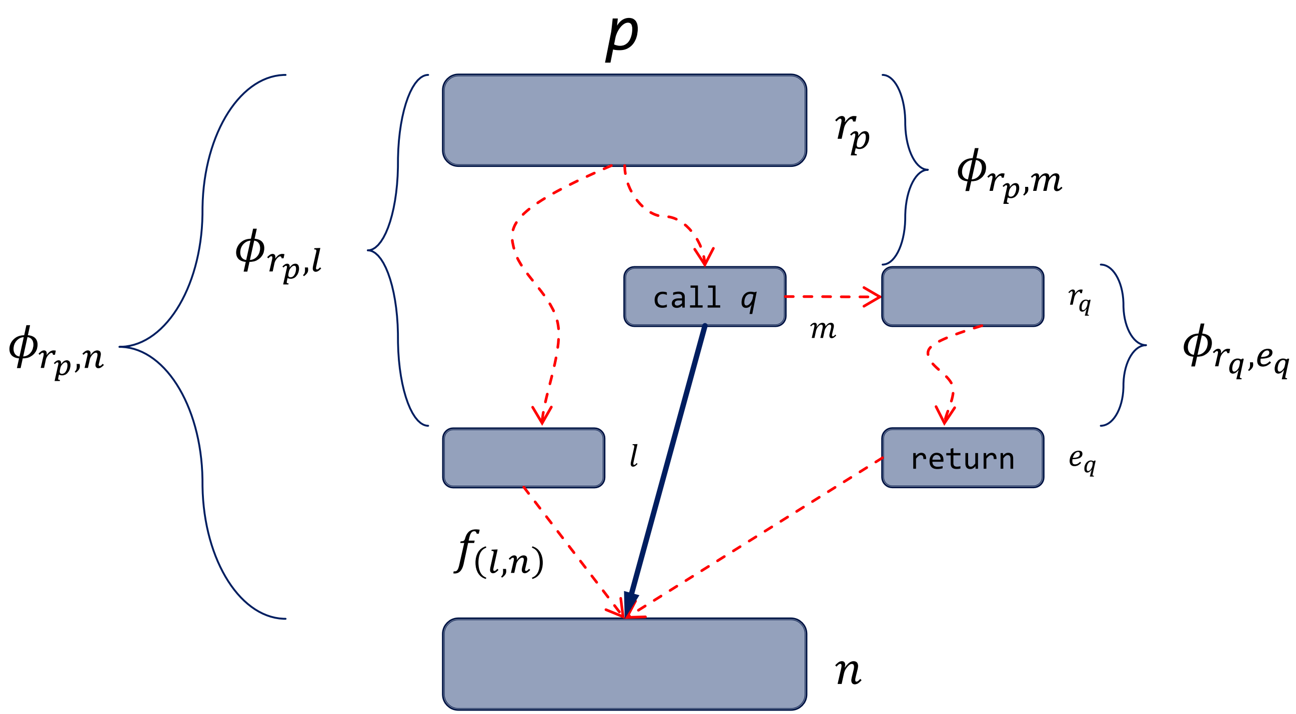
$PHI(n_m, \top) = 1$

…

Summary: Define valid paths



Summary: Functional Approach



Information x at r_p

is transformed to $\phi_{r_p,n}(x)$ at n

Summary: Functional Result

F is $\mathit{distributive} \Longrightarrow$

The maximum fixed point solution x_n^* of the functional approach $=$ the interprocedural MOP solution y_n

Summary: Practical Aspects

- Represent $\phi_{r_p,n}$ symbolically or by explicit relation
- Explicit approach may require much space
- $L\ infinite\ \Longrightarrow$ Explicit approach may diverge
- $L\ infinite, F\ unbounded\ \Longrightarrow$ Symbolic approach may diverge