

Static Program Analysis

Lecture 19: Extensions I

(Interprocedural Dataflow Analysis – MVP Solution)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University

`noll@cs.rwth-aachen.de`

`http://www-i2.informatik.rwth-aachen.de/i2/spa11/`

Summer Semester 2011

- 1 Interprocedural Dataflow Analysis
- 2 Intraprocedural vs. Interprocedural Analysis
- 3 The MVP Solution

- **So far:** only **intraprocedural analyses**
(i.e., without user-defined functions or procedures)
- **Now:** **interprocedural dataflow analysis**
- **Complications:**
 - correct **matching** between calls and returns
 - **parameter passing** (aliasing effects)
- **Here:** simple setting
 - only **top-level declarations**, no blocks or nested declarations
 - mutual **recursion**
 - one call-by-value and one call-by-result **parameter**
(extension to multiple and call-by-value-result parameters straightforward)

Extending the Syntax

Syntactic categories:

Category	Domain	Meta variable
Procedure identifiers	$PVar = \{P, Q, \dots\}$	P
Procedure declarations	$PDec$	p
Commands (statements)	Cmd	c

Context-free grammar:

$$\begin{aligned} p &::= \text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}; p \mid \varepsilon \in PDec \\ c &::= [\text{skip}]^l \mid [x := a]^l \mid c_1; c_2 \mid \text{if } [b]^l \text{ then } c_1 \text{ else } c_2 \mid \\ &\quad \text{while } [b]^l \text{ do } c \mid [\text{call } P(a, x)]_{l_r}^{l_c} \in Cmd \end{aligned}$$

- All labels and procedure names in **program** $p\ c$ distinct
- In $\text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}$, l_n (l_x) refers to the **entry** (**exit**) of P
- In $[\text{call } P(a, x)]_{l_r}^{l_c}$, l_c (l_r) refers to the **call** of (**return** from) P
- First parameter **call-by-value**, second **call-by-result**

Example 19.1 (Fibonacci numbers)

(with extension by multiple call-by-value parameters)

```
proc [Fib(val x, y, res z)]1 is
  if [x<2]2 then
    [z := y+1]3
  else
    [call Fib(x-1, y, z)]45;
    [call Fib(x-2, z, z)]67;
  [end]8;
[call Fib(5, 0, v)]910
```

Definition 19.2 (Procedure flow graphs)

The auxiliary functions **init**, **final**, and **flow** are extended as follows:

$$\begin{aligned}\text{init}(\text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}) &:= l_n \\ \text{final}(\text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}) &:= \{l_x\} \\ \text{flow}(\text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}) &:= \{(l_n, \text{init}(c))\} \\ &\quad \cup \text{flow}(c) \\ &\quad \cup \{(l, l_x) \mid l \in \text{final}(c)\} \\ \text{init}([\text{call } P(a, x)]_{l_r}^{l_c}) &:= l_c \\ \text{final}([\text{call } P(a, x)]_{l_r}^{l_c}) &:= \{l_r\} \\ \text{flow}([\text{call } P(a, x)]_{l_r}^{l_c}) &:= \{(l_c; l_n), (l_x; l_r)\}\end{aligned}$$

if $\text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x}$ is in p .

Moreover the **interprocedural flow** of a program p is defined by

$$\begin{aligned}\text{iflow} &:= \{(l_c, l_n, l_x, l_r) \mid p \text{ contains } [\text{call } P(a, x)]_{l_r}^{l_c} \text{ and} \\ &\quad \text{proc } [P(\text{val } x, \text{res } y)]^{l_n} \text{ is } c \text{ [end]}^{l_x} \subseteq L^4\end{aligned}$$

Example 19.3 (Fibonacci numbers)

Flow graph of

```
proc [Fib(val x, y, res z)]1 is
  if [x<2]2 then
    [z := y+1]3
  else
    [call Fib(x-1, y, z)]45;
    [call Fib(x-2, z, z)]67;
  [end]8;
  [call Fib(5, 0, v)]910
```

(on the board)

Here iflow = {(9, 1, 8, 10), (4, 1, 8, 5), (6, 1, 8, 7)}

- 1 Interprocedural Dataflow Analysis
- 2 Intraprocedural vs. Interprocedural Analysis
- 3 The MVP Solution

- **Attempt:** directly transfer **techniques from intraprocedural analysis**
 $\implies$ treat $(l_c; l_n)$ like (l_c, l_n) and $(l_x; l_r)$ like (l_x, l_r)
- Given: dataflow system $S = (L, E, F, (D, \sqsubseteq), \iota, \varphi)$
- For each procedure call $[\text{call } P(a, x)]_{l_r}^{l_c}$:
transfer functions $\varphi_{l_c}, \varphi_{l_r} : D \rightarrow D$ (definition later)
- For each procedure declaration $\text{proc } [P(\text{val } x, \text{res } y)]_{l_n}^{\text{is } c} [\text{end}]_{l_x}$:
transfer functions $\varphi_{l_n}, \varphi_{l_x} : D \rightarrow D$ (definition later)
- Induces **equation system**
$$AI_l = \begin{cases} \iota & \text{if } l \in E \\ \bigsqcup \{ \varphi_{l'}(AI_{l'}) \mid (l', l) \in F \text{ or } (l'; l) \in F \} & \text{otherwise} \end{cases}$$
- **Problem:** procedure calls $(l_c; l_n)$ and procedure returns $(l_x; l_r)$ treated like goto's
 - $\implies$ **nesting** of calls and returns ignored
 - $\implies$ too many **paths**
 - $\implies$ analysis information **imprecise** (but still correct)

Example 19.4 (Fibonacci numbers)

```
proc [Fib(val x, y, res z)]1 is
  if [x<2]2 then
    [z := y+1]3
  else
    [call Fib(x-1, y, z)]45;
    [call Fib(x-2, z, z)]67;
  [end]8;
[call Fib(5, 0, v)]910
```

- “Valid” path:
[9, 1, 2, 3, 8, 10]
- “Invalid” path:
[9, 1, 2, 4, 1, 2, 3, 8, 10]

Example 19.5 (Impreciseness of constant propagation analysis)

```
proc [P(val x, res y)]1 is
  [y := x]2
[end]3;
if [y=0]4 then
  [call P(1, y)]56;
  [y := y-1]7
else
  [call P(2, y)]89;
  [y := y-2]10;
[skip]11
```

Two “valid” and two “invalid” paths:

- Valid: [4, 5, 1, 2, 3, 6, 7, 11]
⇒ $y = 0$ at label 11
- Valid: [4, 8, 1, 2, 3, 9, 10, 11]
⇒ $y = 0$ at label 11
- Invalid: [4, 5, 1, 2, 3, 9, 10, 11]
⇒ $y = -1$ at label 11
- Invalid: [4, 8, 1, 2, 3, 6, 7, 11]
⇒ $y = 1$ at label 11

⇒ actually $y = 0$ at 11, but naive method yields $y = \top$

- 1 Interprocedural Dataflow Analysis
- 2 Intraprocedural vs. Interprocedural Analysis
- 3 The MVP Solution

- Consider only paths with **correct nesting** of procedure calls and returns
- Will yield **MVP** solution (**Meet over all Valid Paths**)

Definition 19.6 (Valid paths I)

Given a dataflow system $S = (L, E, F, (D, \sqsubseteq), \iota, \varphi)$ and $l_1, l_2 \in L$, the set of **valid paths from l_1 to l_2** is generated by the nonterminal symbol $P[l_1, l_2]$ according to the following productions:

$$\begin{array}{ll} P[l_1, l_2] \rightarrow l_1 & \text{whenever } l_1 = l_2 \\ P[l_1, l_3] \rightarrow l_1, P[l_2, l_3] & \text{whenever } (l_1, l_2) \in F \\ P[l_c, l] \rightarrow l_c, P[l_n, l_x], P[l_r, l] & \text{whenever } (l_c, l_n, l_x, l_r) \in \text{iflow} \end{array}$$

Example 19.7 (Fibonacci numbers (cf. Example 19.4))

proc [Fib(val x, y, res z)]¹ is Valid paths from 9 to 10:

```
  if [x<2]2 then
    [z := y+1]3
  else
    [call Fib(x-1, y, z)]45;
    [call Fib(x-2, z, z)]67;
  [end]8;
  [call Fib(5, 0, v)]910
```

$P[9, 10] \rightarrow 9, P[1, 8], P[10, 10]$
 $P[1, 8] \rightarrow 1, P[2, 8]$
 $P[2, 8] \rightarrow 2, P[3, 8]$
 $P[2, 8] \rightarrow 2, P[4, 8]$
 $P[3, 8] \rightarrow 3, P[8, 8]$
 $P[4, 8] \rightarrow 4, P[1, 8], P[5, 8]$
 $P[5, 8] \rightarrow 5, P[6, 8]$
 $P[6, 8] \rightarrow 6, P[1, 8], P[7, 8]$
 $P[7, 8] \rightarrow 7, P[8, 8]$
 $P[8, 8] \rightarrow 8$
 $P[10, 10] \rightarrow 10$

Reminder:

$P[l_1, l_2] \rightarrow l_1$
 for $l_1 = l_2$
 $P[l_1, l_3] \rightarrow l_1, P[l_2, l_3]$
 for $(l_1, l_2) \in F$
 $P[l_c, l] \rightarrow l_c, P[l_n, l_x], P[l_r, l]$
 for $(l_c, l_n, l_x, l_r) \in \text{iflow}$

Thus $[9, 1, 2, 3, 8, 10] \in L(P[9, 10])$,
 $[9, 1, 2, 4, 1, 2, 3, 8, 10] \notin L(P[9, 10])$

Definition 19.8 (Valid paths II)

Let $S = (L, E, F, (D, \sqsubseteq), \iota, \varphi)$ be a dataflow system. For every $l \in L$, the set of **valid paths up to l** is given by

$$VPath(l) := \{[l_1, \dots, l_{k-1}] \mid k \geq 1, l_1 \in E, l_k = l, \\ [l_1, \dots, l_k] \text{ valid path from } l_1 \text{ to } l_k\}.$$

For a path $p = [l_1, \dots, l_{k-1}] \in VPath(l)$, we define the **transfer function** $\varphi_p : D \rightarrow D$ by

$$\varphi_p := \varphi_{l_{k-1}} \circ \dots \circ \varphi_{l_1} \circ \text{id}_D$$

(so that $\varphi_{[]} = \text{id}_D$).

Definition 19.9 (MVP solution)

Let $S = (L, E, F, (D, \sqsubseteq), \iota, \varphi)$ be a dataflow system where $L = \{l_1, \dots, l_n\}$. The **MVP solution** for S is determined by

$$\text{mvp}(S) := (\text{mvp}(l_1), \dots, \text{mvp}(l_n)) \in D^n$$

where, for every $l \in L$,

$$\text{mvp}(l) := \bigsqcup \{\varphi_p(\iota) \mid p \in V\text{Path}(l)\}.$$

Corollary 19.10

- ① $\text{mvp}(S) \sqsubseteq \text{mop}(S)$
- ② *The MVP solution is undecidable.*

Proof.

- ① since $V\text{Path}(l) \subseteq \text{Path}(l)$ for every $l \in L$
- ② since $\text{mvp}(S) = \text{mop}(S)$ in intraprocedural case, and by undecidability of MOP solution (cf. Theorem 6.2)

