Static Program Analysis

Lecture 2: Dataflow Analysis I
(Introduction & Available Expressions Analysis)

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

Outline

Dataflow Analysis: the Approach

- Traditional form of program analysis
- Idea: describe how analysis information flows through program
- Distinctions:

direction of flow:
: forward vs. backward analyses

procedures:
: interprocedural vs. intraprocedural analyses

quantification over paths:
: may (union) vs. must (intersection) analyses

dependence on statement order:
: flow-sensitive vs. flow-insensitive analyses

Labeled Programs

- Goal: localization of analysis information
- Dataflow information will be associated with
  - `skip` statements
  - assignments
  - tests in conditionals (`if`) and loops (`while`)
- Assume set of labels L with meta variable $l \in L$ (usually $L = \mathbb{N}$)

Definition 2.1 (Labeled WHILE programs)

The syntax of labeled WHILE programs is defined by the following context-free grammar:

$$\begin{aligned} a &::= z \mid x \mid a_1 \texttt{+} a_2 \mid a_1 \texttt{-} a_2 \mid a_1 \texttt{*} a_2 \in AExp \\ b &::= t \mid a_1 \texttt{=} a_2 \mid a_1 \texttt{>} a_2 \mid \neg b \mid b_1 \wedge b_2 \mid b_1 \vee b_2 \in BExp \\ c &::= [\texttt{skip}]^l \mid [x \texttt{ := } a]^l \mid c_1 \texttt{;} c_2 \mid \\ &\quad\;\; \texttt{if } [b]^l \texttt{ then } c_1 \texttt{ else } c_2 \mid \texttt{while } [b]^l \texttt{ do } c \in Cmd \end{aligned}$$

- All labels in $c \in Cmd$ assumed distinct, denoted by L_c
- Labeled fragments of c called blocks, denoted by Blk_c

A WHILE Program with Labels

Example 2.2

```
x := 6;
y := 7;
z := 0;
while x > 0 do
  x := x - 1;
  v := y;
  while v > 0 do
    v := v - 1;
    z := z + 1;
```

Representing Control Flow I

Every (labeled) statement has a single entry (given by the initial label) and generally multiple exits (given by the final labels):

Definition 2.3 (Initial and final labels)

The mapping init : $Cmd \to L$ returns the initial label of a statement:

$$\begin{aligned}
\text{init}([\texttt{skip}]^l) &:= l\\
\text{init}([x \texttt{ := } a]^l) &:= l\\
\text{init}(c_1 \texttt{;} c_2) &:= \text{init}(c_1)\\
\text{init}(\texttt{if } [b]^l \texttt{ then } c_1 \texttt{ else } c_2) &:= l\\
\text{init}(\texttt{while } [b]^l \texttt{ do } c) &:= l
\end{aligned}$$

The mapping final : $Cmd \to 2^L$ returns the set of final labels of a statement:

$$\begin{aligned}
\text{final}([\texttt{skip}]^l) &:= \{l\}\\
\text{final}([x \texttt{ := } a]^l) &:= \{l\}\\
\text{final}(c_1 \texttt{;} c_2) &:= \text{final}(c_2)\\
\text{final}(\texttt{if } [b]^l \texttt{ then } c_1 \texttt{ else } c_2) &:= \text{final}(c_1) \cup \text{final}(c_2)\\
\text{final}(\texttt{while } [b]^l \texttt{ do } c) &:= \{l\}
\end{aligned}$$

Representing Control Flow II

Definition 2.4 (Flow relation)

Given a statement $c \in Cmd$, the (control) flow relation $\mathrm{flow}(c) \subseteq L \times L$ is defined by

$$
\begin{aligned}
\mathrm{flow}([\texttt{skip}]^l) &:= \emptyset \\
\mathrm{flow}([x \texttt{ := } a]^l) &:= \emptyset \\
\mathrm{flow}(c_1 \texttt{;} c_2) &:= \mathrm{flow}(c_1) \cup \mathrm{flow}(c_2) \cup \\
&\quad\ \{(l, \mathrm{init}(c_2)) \mid l \in \mathrm{final}(c_1)\} \\
\mathrm{flow}(\texttt{if } [b]^l \texttt{ then } c_1 \texttt{ else } c_2) &:= \mathrm{flow}(c_1) \cup \mathrm{flow}(c_2) \cup \\
&\quad\ \{(l, \mathrm{init}(c_1)), (l, \mathrm{init}(c_2))\} \\
\mathrm{flow}(\texttt{while } [b]^l \texttt{ do } c) &:= \mathrm{flow}(c) \cup \{(l, \mathrm{init}(c))\} \cup \\
&\quad\ \{(l', l) \mid l' \in \mathrm{final}(c)\}
\end{aligned}
$$

Example 2.5

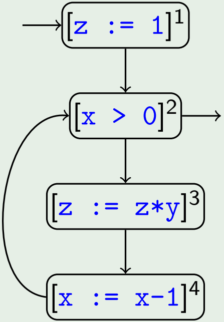
```
c = [z := 1]^1;
    while [x > 0]^2 do
      [z := z*y]^3;
      [x := x-1]^4
```

$\text{init}(c) = 1$

$\text{final}(c) = \{2\}$

$\text{flow}(c) = \{(1,2),(2,3),(3,4),(4,2)\}$

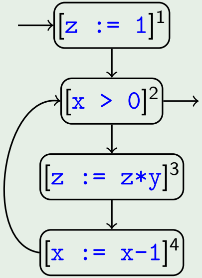
Visualization by (control) flow graph:



Representing Control Flow IV

- To simplify the presentation we will often assume that the program $c \in Cmd$ under consideration has an isolated entry, meaning that
$$\{l \in L \mid (l, \mathrm{init}(c)) \in \mathrm{flow}(c)\} = \emptyset$$
(which is the case when c does not start with a `while` loop)
- Similarly: $c \in Cmd$ has isolated exits if
$$\{l' \in L \mid (l, l') \in \mathrm{flow}(c) \text{ for some } l \in \mathrm{final}(c)\} = \emptyset$$
(which is the case when no final label identifies a loop header)

Example 2.6



has an isolated entry but not isolated exits

Outline

Goal of Available Expressions Analysis

Available Expressions Analysis

The goal of Available Expressions Analysis is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- Can be used for Common Subexpression Elimination: replace subexpression by variable that contains up-to-date value
- Only interesting for non-trivial (i.e., complex) arithmetic expressions

Example 2.7 (Available Expressions Analysis)

```
[x := a+b]^1;
[y := a*b]^2;
while [y > a+b]^3 do
  [a := a+1]^4;
  [x := a+b]^5
```

- a+b available at label 3
- a+b not available at label 5
- possible optimization: `while [y > x]^3 do`

Formalizing Available Expressions Analysis I

- Given $a \in AExp$, $b \in BExp$, $c \in Cmd$
  - $Var_a / Var_b / Var_c$ denotes the set of all variables occurring in $a/b/c$
  - $AExp_b / AExp_c$ denote the sets of all complex arithmetic expressions occurring in b/c
- An expression a is killed in a block B if any of the variables in a is modified in B
- Formally: $\text{kill}_{\text{AE}} : Blk_c \to 2^{AExp_c}$ is defined by

$$\begin{aligned}
\text{kill}_{\text{AE}}([\texttt{skip}]^l) &:= \emptyset \\
\text{kill}_{\text{AE}}([x \texttt{ := } a]^l) &:= \{a' \in AExp_c \mid x \in Var_{a'}\} \\
\text{kill}_{\text{AE}}([b]^l) &:= \emptyset
\end{aligned}$$

- An expression a is generated in a block B if it is evaluated in and none of its variables are modified by B
- Formally: $\text{gen}_{\text{AE}} : Blk_c \to 2^{AExp_c}$ is defined by

$$\begin{aligned}
\text{gen}_{\text{AE}}([\texttt{skip}]^l) &:= \emptyset \\
\text{gen}_{\text{AE}}([x \texttt{ := } a]^l) &:= \{a \mid x \notin Var_a\} \\
\text{gen}_{\text{AE}}([b]^l) &:= AExp_b
\end{aligned}$$

Formalizing Available Expressions Analysis II

Example 2.8 ($\text{kill}_{\text{AE}}/\text{gen}_{\text{AE}}$ functions)

```
c = [x := a+b]^1;
    [y := a*b]^2;
    while [y > a+b]^3 do
      [a := a+1]^4;
      [x := a+b]^5
```

- $AExp_c = \{\texttt{a+b}, \texttt{a*b}, \texttt{a+1}\}$
-

L_c	$\text{kill}_{\text{AE}}(B^l)$	$\text{gen}_{\text{AE}}(B^l)$
1	$\emptyset$	$\{\texttt{a+b}\}$
2	$\emptyset$	$\{\texttt{a*b}\}$
3	$\emptyset$	$\{\texttt{a+b}\}$
4	$\{\texttt{a+b}, \texttt{a*b}, \texttt{a+1}\}$	$\emptyset$
5	$\emptyset$	$\{\texttt{a+b}\}$

The Equation System I

- Analysis itself defined by setting up an equation system
- For each $l \in L_c$, $\mathsf{AE}_l \subseteq AExp_c$ represents the set of available expressions at the entry of block B^l
- Formally, for $c \in Cmd$ with isolated entry:

$$\mathsf{AE}_l = \begin{cases} \emptyset & \text{if } l = \mathsf{init}(c) \\ \bigcap\{\varphi_{l'}(\mathsf{AE}_{l'}) \mid (l', l) \in \mathsf{flow}(c)\} & \text{otherwise} \end{cases}$$

  where $\varphi_{l'} : 2^{AExp_c} \to 2^{AExp_c}$ denotes the transfer function of block $B^{l'}$, given by

$$\varphi_{l'}(A) := (A \setminus \mathsf{kill}_{\mathsf{AE}}(B^{l'})) \cup \mathsf{gen}_{\mathsf{AE}}(B^{l'})$$

- Characterization of analysis:

  forward: starts in $\mathsf{init}(c)$ and proceeds downwards

  must: $\bigcap$ in equation for AE_l

  flow-sensitive: results depending on order of assignments

- Later: solution not necessarily unique

  $\Longrightarrow$ choose greatest one

The Equation System II

Reminder:

$$\mathsf{AE}_l = \begin{cases} \emptyset & \text{if } l = \mathsf{init}(c) \\ \bigcap\{\varphi_{l'}(\mathsf{AE}_{l'}) \mid (l', l) \in \mathsf{flow}(c)\} & \text{otherwise} \end{cases}$$

$$\varphi_{l'}(E) = (E \setminus \mathsf{kill}_{\mathsf{AE}}(B^{l'})) \cup \mathsf{gen}_{\mathsf{AE}}(B^{l'})$$

Example 2.9 (AE equation system)

$c =$
```
[x := a+b]1;
[y := a*b]2;
while [y > a+b]3 do
  [a := a+1]4;
  [x := a+b]5
```

$l \in L_c$	$\mathsf{kill}_{\mathsf{AE}}(B^l)$	$\mathsf{gen}_{\mathsf{AE}}(B^l)$
1	$\emptyset$	{a+b}
2	$\emptyset$	{a*b}
3	$\emptyset$	{a+b}
4	{a+b, a*b, a+1}	$\emptyset$
5	$\emptyset$	{a+b}

Equations:

$$\begin{aligned} \mathsf{AE}_1 &= \emptyset \\ \mathsf{AE}_2 &= \varphi_1(\mathsf{AE}_1) = \mathsf{AE}_1 \cup \{\texttt{a+b}\} \\ \mathsf{AE}_3 &= \varphi_2(\mathsf{AE}_2) \cap \varphi_5(\mathsf{AE}_5) \\ &= (\mathsf{AE}_2 \cup \{\texttt{a*b}\}) \cap (\mathsf{AE}_5 \cup \{\texttt{a+b}\}) \\ \mathsf{AE}_4 &= \varphi_3(\mathsf{AE}_3) = \mathsf{AE}_3 \cup \{\texttt{a+b}\} \\ \mathsf{AE}_5 &= \varphi_4(\mathsf{AE}_4) = \mathsf{AE}_4 \setminus \{\texttt{a+b}, \texttt{a*b}, \texttt{a+1}\} \end{aligned}$$

Solution:

$$\begin{aligned} \mathsf{AE}_1 &= \emptyset \\ \mathsf{AE}_2 &= \{\texttt{a+b}\} \\ \mathsf{AE}_3 &= \{\texttt{a+b}\} \\ \mathsf{AE}_4 &= \{\texttt{a+b}\} \\ \mathsf{AE}_5 &= \emptyset \end{aligned}$$