

# More Data Flow Analyses

Reading: NNH 2.1

17-654/17-765

Analysis of Software Artifacts

Jonathan Aldrich

# Announcements

- Assignment due at 11:59pm Thursday
  - Under Dean's door (Wean 8130)
  - Or via Blackboard
- Nicholas Sherman office hours
  - Tuesday at 4pm
- Reading for Thursday: NNH 2.2
- Questions on the homework?
  - Applications of sign analysis?

# Some Notation

- This will help us describe analyses in a more precise and general way
  - $init(S)$  – the label of the first statement in  $S$
  - $final(S)$  – the set of labels of the last statements in  $S$ 
    - the last statement on each branch of an if
    - the test of a while
  - $blocks(S)$  – the set of primitive statements and tests in  $S$
  - $labels(S)$  – the set of labels of blocks in  $S$
  - $flow(S) = \{(\ell, \ell') \mid \text{control may transfer from block } \ell \text{ to block } \ell'\}$ 
    - A pair for each edge in the control flow graph
- The text defines these formally

# General Data Flow Equations

## Available Expressions

$$\begin{aligned} AE_{\text{entry}}(\ell) &= \emptyset && \text{if } (\ell = \text{init}(S_*)) \\ &= \cap \{ AE_{\text{exit}}(\ell') \mid (\ell', \ell) \in \text{flow}(S_*) \} && \text{otherwise} \end{aligned}$$

$$AE_{\text{exit}}(\ell) = (AE_{\text{entry}}(\ell) \setminus \text{kill}_{AE}(B^\ell)) \cup \text{gen}_{AE}(B^\ell)$$

## Reaching Definitions

$$\begin{aligned} RD_{\text{entry}}(\ell) &= \{(x, ?) \mid x \in \text{FV}(S_*)\} && \text{if } \ell = \text{init}(S_*) \\ &= \cup \{ RD_{\text{exit}}(\ell') \mid (\ell', \ell) \in \text{flow}(S_*) \} && \text{otherwise} \end{aligned}$$

$$RD_{\text{exit}}(\ell) = (RD_{\text{entry}}(\ell) \setminus \text{kill}_{RD}(B^\ell)) \cup \text{gen}_{RD}(B^\ell)$$

# Safety and Precision: May vs. Must

```
[y := x]1;  
[z := 1]2;  
while [y > 1]3 do  
  [z := z * y]4;  
  [y := y - 1]5;  
[y := 0]6;
```

- What definitions **may** reach entry to 5?
  - Best answer?  $\{(y,1), (y,5), (z,4)\}$
  - More precise (but unsafe)?  $\{(y,1), (z,4)\}$
  - Safe but less precise?  $\{(y,1), (y,5), (z,1), (z,4)\}$
- What expressions **must** be available at entry to 5?
  - Best answer?  $\{y > 1\}$
  - More precise (but unsafe)?  $\{y > 1, z * y\}$
  - Safe but less precise?  $\emptyset$

# Safety and Precision: May vs. Must

```
[y := x]1;  
[z := 1]2;  
while [y > 1]3 do  
  [z := z * y]4;  
  [y := y - 1]5;  
[y := 0]6;
```

- What definitions **may** reach entry to 5?
  - Best answer?  $\{(y,1), (y,5), (z,4)\}$
  - More precise (but unsafe)?  $\{(y,1), (z,4)\}$
  - Safe but less precise?  $\{(y,1), (y,5), (z,1), (z,4)\}$
- What expressions **must** be available at entry to 5?
  - Best answer?  $\{y > 1\}$
  - More precise (but unsafe)?  $\{y > 1, z * y\}$
  - Safe but less precise?  $\emptyset$

# Safety and Precision: May vs. Must

```
[y := x]1;  
[z := 1]2;  
while [y > 1]3 do  
  [z := z * y]4;  
  [y := y - 1]5;  
[y := 0]6;
```

- What definitions **may** reach entry to 5?
  - Best answer?  $\{ (y,1), (y,5), (z,4) \}$
  - More precise (but unsafe)?  $\{ (y,1), (z,4) \}$
  - Safe but less precise?  $\{ (y,1), (y,5), (z,1), (z,4) \}$
- What expressions **must** be available at entry to 5?
  - Best answer?  $\{ y > 1 \}$
  - More precise (but unsafe)?  $\{ y > 1, z * y \}$
  - Safe but less precise?  $\emptyset$

# Reaching Defs. vs. Available Exp.

- Reaching Defs. **May analysis**
  - Initial dataflow values: *empty* sets
  - *Union* at control flow merge
  - Precision: want *least* fixed point
  - Safety: err on the side of *larger* sets
- Available Exp. **Must analysis**
  - Initial dataflow values: *universal* sets
  - *Intersection* at control flow merge
  - Precision: want *greatest* fixed point
  - Safety: err on the side of *smaller* sets

# Constant Propagation

- For each program point, what value *may* each variable hold?

# Constant Propagation

$$\begin{aligned} \text{CP}_{\text{entry}}(\ell) &= \{(x,n) \mid x \in \text{FV}(S_*), n \in \mathbb{N}\} && \text{if } (\ell = \text{init}(S_*)) \\ &= \cup \{ \text{CP}_{\text{exit}}(\ell') \mid (\ell', \ell) \in \text{flow}(S_*) \} && \text{otherwise} \end{aligned}$$

$$\text{CP}_{\text{exit}}(\ell) = (\text{CP}_{\text{entry}}(\ell) \setminus \text{kill}_{\text{CP}}(B^\ell)) \cup \text{gen}_{\text{CP}}(B^\ell)$$

$$\text{kill}_{\text{CP}}([x := a]^\ell) =$$

$$\text{kill}_{\text{CP}}([\text{skip}]^\ell) =$$

$$\text{kill}_{\text{CP}}([b]^\ell) =$$

$$\text{gen}_{\text{CP}}([x := a]^\ell) =$$

$$\text{gen}_{\text{CP}}([\text{skip}]^\ell) =$$

$$\text{gen}_{\text{CP}}([b]^\ell) =$$

# Constant Propagation

$$\begin{aligned} \text{CP}_{\text{entry}}(\ell) &= \{(x,n) \mid x \in \text{FV}(S_*), n \in \mathbb{N}\} && \text{if } (\ell = \text{init}(S_*)) \\ &= \cup \{ \text{CP}_{\text{exit}}(\ell') \mid (\ell', \ell) \in \text{flow}(S_*) \} && \text{otherwise} \end{aligned}$$

$$\text{CP}_{\text{exit}}(\ell) = (\text{CP}_{\text{entry}}(\ell) \setminus \text{kill}_{\text{CP}}(B^\ell)) \cup \text{gen}_{\text{CP}}(B^\ell)$$

$$\text{kill}_{\text{CP}}([x := a]^\ell) = \{(x,n) \mid n \in \mathbb{N}\}$$

$$\text{kill}_{\text{CP}}([\text{skip}]^\ell) = \emptyset$$

$$\text{kill}_{\text{CP}}([b]^\ell) = \emptyset$$

$$\text{gen}_{\text{CP}}([x := a]^\ell) = \{(x, n) \mid n \in \mathbf{CP}^\ell(a)\}$$

$$\text{gen}_{\text{CP}}([\text{skip}]^\ell) = \emptyset$$

$$\text{gen}_{\text{CP}}([b]^\ell) = \emptyset$$

# Constant Propagation

$$\mathbf{CP}^\ell(x) = \{ \mathbf{CP}_{\text{entry}(\ell)}(x) \}$$

$$\mathbf{CP}^\ell(n) = \{ n \}$$

$$\mathbf{CP}^\ell(a_1 \ op_a \ a_2) = \mathbf{CP}^\ell(a_1) \ \widehat{op}_a \ \mathbf{CP}^\ell(a_2)$$

$$\text{set}_1 \ \widehat{op}_a \ \text{set}_2 = \{ n_1 \ op_a \ n_2 \mid n_1 \in \text{set}_1, n_2 \in \text{set}_2 \}$$

# Constant Propagation Example

		<u>x</u>	<u>y</u>	<u>z</u>	<u>w</u>
$[y := 5]^1;$					
$[z := 1+y]^2;$	$AE_{\text{exit}}(1) =$	?	5	?	?
if $[x = 5]^3$	$AE_{\text{exit}}(2) =$	?	5	6	?
then $[w := x+1]^4;$	$AE_{\text{exit}}(3) =$	?	5	6	?
else $[w := y+1]^5$	$AE_{\text{exit}}(4) =$	?	5	6	?
$[skip]^6$	$AE_{\text{exit}}(5) =$	?	5	6	6
	$AE_{\text{exit}}(6) =$	?	5	6	?

Here ? is a shorthand for the set of all integers

# Constant Propagation Example

		<u>x</u>	<u>y</u>	<u>z</u>	<u>w</u>
$[y := 5]^1;$					
$[z := 1+y]^2;$	$AE_{\text{exit}}(1) =$	?	5	?	?
if $[x = 5]^3$	$AE_{\text{exit}}(2) =$	?	5	6	?
then $[w := x+1]^4;$	$AE_{\text{exit}}(3) =$	?	5	6	?
else $[w := y+1]^5$	$AE_{\text{exit}}(4) =$	?	5	6	?
$[\text{skip}]^6$	$AE_{\text{exit}}(5) =$	?	5	6	6
	$AE_{\text{exit}}(6) =$	?	5	6	?

But we know that  $x=5$  at statement 4. Can we do better?

# Constant Propagation, Take 2

$$\begin{aligned} \text{CP}_{\text{entry}}(\ell) &= \{ (x, n) \mid x \in \text{FV}(S_*), n \in \mathbb{N} \} && \text{if } (\ell = \text{init}(S_*)) \\ &= \cup \{ \text{CP}_{\text{exit}}(\ell') \mid (\ell', \ell) \in \text{flow}(S_*) \} && \text{otherwise} \end{aligned}$$

$$\begin{aligned} \text{CP}_T(\ell) &= (\text{CP}_{\text{entry}}(\ell) \setminus \text{kill}_{\text{CP},T}(B^\ell)) \cup \text{gen}_{\text{CP}}(B^\ell) \\ \text{CP}_F(\ell) &= (\text{CP}_{\text{entry}}(\ell) \setminus \text{kill}_{\text{CP},F}(B^\ell)) \cup \text{gen}_{\text{CP}}(B^\ell) \end{aligned}$$

$$\text{kill}_{\text{CP},T}([x := a]^\ell) = \{ (x, n) \mid n \in \mathbb{N} \}$$

$$\text{kill}_{\text{CP},T}([\text{skip}]^\ell) = \emptyset$$

$$\begin{aligned} \text{kill}_{\text{CP},T}([b]^\ell) &= \{ (x, n) \mid n \in \mathbb{N} \text{ and } n \neq m \} \\ &= \emptyset \end{aligned}$$

if  $b = (x=a)$  and  $\mathbf{CP}^\ell(a) = \{m\}$   
otherwise

$$\begin{aligned} \text{kill}_{\text{CP},F}([b]^\ell) &= \{ (x, m) \} \\ &= \emptyset \end{aligned}$$

if  $b = (x=a)$  and  $\mathbf{CP}^\ell(a) = \{m\}$   
otherwise

$$\text{gen}_{\text{CP}}([x := a]^\ell) = \{ (x, n) \mid n \in \mathbf{CP}^\ell(a) \}$$

$$\text{gen}_{\text{CP}}([\text{skip}]^\ell) = \emptyset$$

$$\text{gen}_{\text{CP}}([b]^\ell) = \emptyset$$

# Constant Propagation Example

		<u>x</u>	<u>y</u>	<u>z</u>	<u>w</u>
$[y := 5]^1;$					
$[z := 1+y]^2;$	$AE_{\text{exit}}(1) =$	?	5	?	?
if $[x = 5]^3$	$AE_{\text{exit}}(2) =$	?	5	6	?
then $[w := x+1]^4;$	$AE_{\top}(3) =$	5	5	6	?
else $[w := y+1]^5$	$AE_{\text{F}}(3) =$	?\5	5	6	?
$[\text{skip}]^6$	$AE_{\text{exit}}(4) =$	5	5	6	6
	$AE_{\text{exit}}(5) =$	?\5	5	6	6
	$AE_{\text{exit}}(6) =$	?	5	6	6

*Keeping track of data flow values separately on each branch supports a more precise final result.*

# General Monotonicity Proofs

- We proved RD was monotone for data flow equations for *a specific program*
- Here's a more general proof, for the assignment flow function:
  - To show: If  $RD_{\text{entry}}(\ell) \subseteq RD_{\text{entry}}'(\ell)$  then  $RD_{\text{exit}}(\ell) \subseteq RD_{\text{exit}}'(\ell)$ 
    - case:  $B^\ell = [x := a]^\ell$
  - Assume  $RD_{\text{entry}}(\ell) \subseteq RD_{\text{entry}}'(\ell)$
  - Now  $\text{kill}_{RD}([x := a]^\ell) = \{ (x, *) \}$  (where \* is any label or ?)
  - Thus  $RD_{\text{entry}}(\ell) \setminus \text{kill}_{RD}(B^\ell) \subseteq RD_{\text{entry}}'(\ell) \setminus \text{kill}_{RD}(B^\ell)$
  - And  $\text{gen}_{RD}([x := a]^\ell) = \{ (x, \ell) \}$
  - Therefore  $(RD_{\text{entry}}(\ell) \setminus \text{kill}_{RD}(B^\ell)) \cup \text{gen}_{RD}(B^\ell) \subseteq (RD_{\text{entry}}'(\ell) \setminus \text{kill}_{RD}(B^\ell)) \cup \text{gen}_{RD}(B^\ell)$ 
    - And we are done with the case for  $[x := a]^\ell$