More
Data Flow Analyses

Reading: NNH 2.1

17-654/17-765
Analysis of Software Artifacts
Jonathan Aldrich

Available Expressions

For each program point, which expressions *must* have already been computed, and not later modified, on all paths to the program point.

- Applications
  - Avoid re-computing expressions

Available Expressions Example

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

$AE_{exit}(1) = \{a+b\}$
$AE_{exit}(2) = \{a+b, a*b\}$
$AE_{exit}(3) = \{a+b\}$
$AE_{exit}(4) = \varnothing$
$AE_{exit}(5) = \{a+b\}$

Available Expressions Equations

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

$AE_{entry}(1) = \varnothing$
$AE_{entry}(2) = AE_{exit}(1)$
$AE_{entry}(3) = AE_{exit}(2) \cap AE_{exit}(5)$
$AE_{entry}(4) = AE_{exit}(3)$
$AE_{entry}(5) = AE_{exit}(4)$

$AE_{exit}(1) = AE_{entry}(1) \cup \{a+b\}$
$AE_{exit}(2) = AE_{entry}(2) \cup \{a*b\}$
$AE_{exit}(3) = AE_{entry}(3) \cup \{a+b\}$
$AE_{exit}(4) = AE_{entry}(4) \setminus \{a+b, a*b, a+1\}$
$AE_{exit}(5) = AE_{entry}(5) \cup \{a+b\}$

Another Example

Program	Dataflow Eqns	Solutions
[x := a+b]^1;	$AE_{entry}(1) = \varnothing$	$\varnothing$
while [true]^2 do	$AE_{entry}(2) = AE_{exit}(1) \cap AE_{exit}(3)$	$\varnothing$ or {a+b}
[skip]^3;	$AE_{entry}(3) = AE_{exit}(2)$	$\varnothing$ or {a+b}
	$AE_{exit}(1) = AE_{entry}(1) \cup \{a+b\}$	{a+b}
	$AE_{exit}(2) = AE_{entry}(2)$	$\varnothing$ or {a+b}
	$AE_{exit}(3) = AE_{entry}(3)$	$\varnothing$ or {a+b}

What's the fixed point of these equations?
- Which is most informative?

Available Exp. vs. Reaching Defs.

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

Kill and Gen Sets

- Useful for defining transfer functions
  - Kill: dataflow facts the statement removes
  - Gen: dataflow facts the statement adds
  - Kill always comes first

Kill and Gen for Available Expressions

$kill_{AE}([x := a]^{\ell}) = \{a' \in \mathbf{AExp}_{\star} \mid x \in FV(a')\}$
$kill_{AE}([skip]^{\ell}) = \varnothing$
$kill_{AE}([b]^{\ell}) = \varnothing$

$gen_{AE}([x := a]^{\ell}) = \{a' \in \mathbf{AExp}(a) \mid x \notin FV(a')\}$
$gen_{AE}([skip]^{\ell}) = \varnothing$
$gen_{AE}([b]^{\ell}) = \mathbf{AExp}(b)$

We should define this!

Definition of **AExp**(a), **AExp**(b)

$\mathbf{AExp}(x) = \varnothing$
$\mathbf{AExp}(n) = \varnothing$
$\mathbf{AExp}(a_1\ op_a\ a_2) = \{a_1\ op_a\ a_2\} \cup \mathbf{AExp}(a_1) \cup \mathbf{AExp}(a_2)$

$\mathbf{AExp}(true) = \varnothing$
$\mathbf{AExp}(false) = \varnothing$
$\mathbf{AExp}(not\ b) = \{not\ b\} \cup \mathbf{AExp}(b)$
$\mathbf{AExp}(b_1\ op_b\ b_2) = \{b_1\ op_b\ b_2\} \cup \mathbf{AExp}(b_1) \cup \mathbf{AExp}(b_2)$
$\mathbf{AExp}(a_1\ op_r\ a_2) = \{a_1\ op_r\ a_2\} \cup \mathbf{AExp}(a_1) \cup \mathbf{AExp}(a_2)$

You'll do something similar on the homework

Kill and Gen for Reaching Definitions

$kill_{AE}([x := a]^{\ell}) = \{(x,?)\} \cup \{(x,\ell) \mid \ell \textit{ is any label}\}$
$kill_{AE}([skip]^{\ell}) = \varnothing$
$kill_{AE}([b]^{\ell}) = \varnothing$

$gen_{AE}([x := a]^{\ell}) = \{(x,\ell)\}$
$gen_{AE}([skip]^{\ell}) = \varnothing$
$gen_{AE}([b]^{\ell}) = \varnothing$

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) – $\{(\ell,\ell') \mid$ control may transfer from block ℓ to block $\ell'\}$
    - A pair for each edge in the control flow graph
- The text defines these formally

General Data Flow Equations

Available Expressions

$AE_{entry}(\ell) = \varnothing$ if $(\ell = init(S_{\star}))$
$\quad = \cap\{AE_{exit}(\ell') \mid (\ell', \ell) \in flow(S_{\star})\}$ otherwise

$AE_{exit}(\ell) = (AE_{entry}(\ell) \setminus kill_{AE}(B^{\ell})) \cup gen_{AE}(B^{\ell})$

Reaching Definitions

$RD_{entry}(\ell) = \{(x,?) \mid x \in FV(S_{\star})\}$ if $\ell = init(S_{\star})$
$\quad = \cup\{RD_{exit}(\ell') \mid (\ell', \ell) \in flow(S_{\star})\}$ otherwise

$RD_{exit}(\ell) = (RD_{entry}(\ell) \setminus kill_{RD}(B^{\ell})) \cup gen_{RD}(B^{\ell})$

Constant Propagation

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

Constant Propagation

$CP_{entry}(\ell) = \{(x,n) \mid x \in FV(S_*),\ n \in N\}$ if $(\ell = init(S_*))$
$\phantom{CP_{entry}(\ell)} = \cup \{ CP_{exit}(\ell') \mid (\ell', \ell) \in flow(S_*) \}$ otherwise

$CP_{exit}(\ell) = (CP_{entry}(\ell) \setminus kill_{CP}(B^\ell)) \cup gen_{CP}(B^\ell)$

$kill_{CP}([x := a]^\ell) = \{ (x,n) \mid n \in N \}$
$kill_{CP}([skip]^\ell) = \emptyset$
$kill_{CP}([b]^\ell) = \emptyset$

$gen_{CP}([x := a]^\ell) = \{ (x, n) \mid n \in \mathbf{CP}^\ell(a)) \}$
$gen_{CP}([skip]^\ell) = \emptyset$
$gen_{CP}([b]^\ell) = \emptyset$

Constant Propagation

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

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

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

$set_1\ \widehat{op}_a\ set_2 = \{ n_1\ op_a\ n_2 \mid n_1 \in set_1,\ n_2 \in set_2 \}$

Constant Propagation Example

Program		x	y	z	w
[y := 5][1];					
[z := 1+y][2];	$AE_{exit}(1)$ =	?	5	?	?
if [x = 5][3]	$AE_{exit}(2)$ =	?	5	6	?
then [w := x+1][4];	$AE_{exit}(3)$ =	?	5	6	?
else [w := y+1][5]	$AE_{exit}(4)$ =	?	5	6	?
[skip][6]	$AE_{exit}(5)$ =	?	5	6	6
	$AE_{exit}(6)$ =	?	5	6	?

Here ? is a shorthand for the set of all integers

Constant Propagation Example

Program		x	y	z	w
[y := 5][1];					
[z := 1+y][2];	$AE_{exit}(1)$ =	?	5	?	?
if [x = 5][3]	$AE_{exit}(2)$ =	?	5	6	?
then [w := x+1][4];	$AE_{exit}(3)$ =	?	5	6	?
else [w := y+1][5]	$AE_{exit}(4)$ =	?	5	6	?
[skip][6]	$AE_{exit}(5)$ =	?	5	6	6
	$AE_{exit}(6)$ =	?	5	6	?

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

Constant Propagation, Take 2

$CP_{entry}(\ell) = \{(x,n) \mid x \in FV(S_*),\ n \in N\}$ if $(\ell = init(S_*))$
$\phantom{CP_{entry}(\ell)} = \cup \{ CP_{exit}(\ell') \mid (\ell', \ell) \in flow(S_*) \}$ otherwise

$CP_T(\ell) = (CP_{entry}(\ell) \setminus kill_{CP,T}(B^\ell)) \cup gen_{CP,T}(B^\ell)$
$CP_F(\ell) = (CP_{entry}(\ell) \setminus kill_{CP,F}(B^\ell)) \cup gen_{CP,F}(B^\ell)$

$kill_{CP,T}([x := a]^\ell) = \{ (x, n) \mid n \in N \}$
$kill_{CP,T}([skip]^\ell) = \emptyset$
$kill_{CP,T}([b]^\ell) = \{ (x, n) \mid n \in N \text{ and } n \neq m\}$ when $b = (x=a)$ and $\mathbf{CP}^\ell(a)=\{m\}$
$\phantom{kill_{CP,T}([b]^\ell)} = \emptyset$ otherwise
$kill_{CP,F}([b]^\ell) = \{ (x, m) \}$ when $b = (x=a)$ and $\mathbf{CP}^\ell(a)=\{m\}$
$\phantom{kill_{CP,F}([b]^\ell)} = \emptyset$ otherwise

$gen_{CP,T}([x := a]^\ell) = \{ (x, n) \mid n \in \mathbf{CP}^\ell(a)) \}$
$gen_{CP,T}([skip]^\ell) = \emptyset$
$gen_{CP,T}([b]^\ell) = \emptyset$
$gen_{CP,F}([b]^\ell) = \emptyset$

Constant Propagation Example

```
[y := 5]1;
[z := 1+y]2;
if [x = 5]3
   then [w := x+1]4;
   else [w := y+1]5
[skip]6
```

	x	y	z	w
$AE_{exit}(1)$ =	?	5	?	?
$AE_{exit}(2)$ =	?	5	6	?
$AE_{T}(3)$ =	5	5	6	?
$AE_{F}(3)$ =	?\5	5	6	?
$AE_{exit}(4)$ =	5	5	6	6
$AE_{exit}(5)$ =	?\5	5	6	6
$AE_{exit}(6)$ =	?	5	6	6

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