CSE211: Compiler Design Oct. 22, 2020

- **Topic**: Local value numbering continued and data flow analysis
- Questions:

Questions/comments about homework 1?

What are some difficult programs for local value numbering?



Announcements

- Homework 2 released! Have a look but don't panic
 - Remember, due dates pushed back 1 week
 - Part 1 should be possible after today's lecture
 - The theory for Part 2 is in lecture. We will go over code next lecture.



CSE211: Compiler Design Oct. 22, 2020

- **Topic**: Local value numbering continued and data flow analysis
- Questions:

Questions/comments about homework 1?

What are some difficult programs for local value numbering?



- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

$$\begin{array}{r} a2 = b0 + c1; \\ b4 = a2 - d3; \\ c5 = b4 + c1; \\ d6 = a2 - d3; \end{array}$$

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

$$\Rightarrow \begin{bmatrix} a2 &= b0 + c1; \\ b4 &= a2 - d3; \\ c5 &= b4 + c1; \\ d6 &= a2 - d3; \end{bmatrix}$$

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

$$a2 = b0 + c1;$$

$$b4 = a2 - d3;$$

$$c5 = b4 + c1;$$

$$d6 = a2 - d3;$$

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

$$a2 = b0 + c1;$$

$$b4 = a2 - d3;$$

$$c5 = b4 + c1;$$

$$d6 = a2 - d3;$$

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

$$a2 = b0 + c1;$$

$$b4 = a2 - d3;$$

$$c5 = b4 + c1;$$

$$d6 = a2 - d3;$$

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

- Algorithm: Now that variables are numbered
- Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.
- At each step, check to see if the rhs has already been computed.

Adding Commutativity

- Certain operators are commutative (e.g. ADD and MULTIPLY)
- In this case, the analysis should consider a deterministic order of operands.
- You can use variable numbers or lexigraphical order

 Algorithm optimization: for commutative operations, re-order operands into a deterministic order

cannot re-order because - is not commutative

• Algorithm optimization: for commutative operations, re-order operands into a deterministic order

re-ordered because a2 < d3 lexigraphically

$$a2 = c1 - b0;$$

$$f4 = d3 * a2;$$

$$c5 = b0 - c1;$$

$$d6 = a2 * d3;$$

$$a2 = c1 - b0;$$

f4 = d3 * a2;
c5 = b0 - c1;
d6 = a2 * d3;

a2	=	c1	_	b0;
f4	=	d3	*	a2;
c 5	=	b0	_	c1;
 d6	=	a2	*	d3;

$$a2 = c1 - b0;$$

f4 = d3 * a2;
c5 = b0 - c1;
d6 = f4;

- We've assumed we have access to an unlimited number of virtual registers.
- In some cases we may not be able to add virtual registers
 - If an expensive register allocation pass has already occurred.
- We need to give back a program such that variables without numbers is still valid.

• Example:



• Solutions?



a	=	X	+	у;
a	=	Z ;		
b	=	Х	+	у;
С	=	Χ	+	у;

• Keep another hash table to keep the current variable number



We cannot optimize the first line, but we can optimize the second

a	=	X	+	у;
a	=	Z ;		
b	=	X	+	у;
C	=	Χ	+	у;

```
Current_val = {
}
```

• Keep another hash table to keep the current variable number

 $\rightarrow \begin{vmatrix} a3 = x1 + y2; \\ a5 = z4; \end{vmatrix}$

b6 = x1 + y2; c7 = x1 + y2;

• Keep another hash table to keep the current variable number

→ a3 = x1 + y2; a5 = z4; b6 = x1 + y2; c7 = x1 + y2;

• Keep another hash table to keep the current variable number

→ a3 = x1 + y2; a5 = z4; b6 = x1 + y2; c7 = x1 + y2;

• Keep another hash table to keep the current variable number

a3 = x1 + y2;

 $\Rightarrow \begin{vmatrix} a5 &= z4; \\ b6 &= x1 + y2; \\ c7 &= x1 + y2; \end{vmatrix}$

• Keep another hash table to keep the current variable number

a3

a5

b6

c7
: 5, : <mark>6</mark>

• Keep another hash table to keep the current variable number

a3

a5

b6

• Keep another hash table to keep the current variable number

——**—**

a3

a5

b6

• Final heuristic: keep sets of possible values

• Final heuristic: keep sets of possible values

Current_val = {
}

a	=	x	+	у;
b	=	X	+	у;
а	=	Z ;		
С	=	Х	+	у;

• Final heuristic: keep sets of possible values

• Final heuristic: keep sets of possible values



• Final heuristic: keep sets of possible values



but we could have replaced it with b4!

• Final heuristic: keep sets of possible values

rewind to this point a3 = x1 + y2; b4 = x1 + y2; a6 = z5; c7 = x1 + y2;

• Final heuristic: keep sets of possible values

• Final heuristic: keep sets of possible values



fast forward again

• Final heuristic: keep sets of possible values

again

fast forward
again

$$fast forward$$

 $again$
 \rightarrow
 $again$
 $Current_val = {
"a" : 6,
"b" : 4
}
H = {
"x1 + y2" : ["a3", "b4"],
}$

• Consider a 3 address code that allows memory accesses



- How to number:
 - Number each pointer/index pair

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

(a[i],3) = (x[j],1) + (y[k],2);(b[i],6) = (x[j],4) + (y[k],5);

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

$$(a[i],3) = (x[j],1) + (y[k],2);$$

 $(b[i],6) = (x[j],4) + (y[k],5);$

compiler analysis:

can we trace a, x, y to
a = malloc(...);
x = malloc(...);
y = malloc(...);

// a, x, y are never overwritten

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

in this case we do not have to update the number

compiler analysis:

can we trace a, x, y to
a = malloc(...);
x = malloc(...);
y = malloc(...);

// a, x, y are never overwritten

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

$$(a[i],3) = (x[j],1) + (y[k],2);$$

 $(b[i],6) = (x[j],4) + (y[k],5);$

programmer annotations can also tell the compiler that no other pointer can access the memory pointed to by a

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

$$(a[i],3) = (x[j],1) + (y[k],2);$$

 $(b[i],6) = (x[j],4) + (y[k],5);$

in this case we do not have to update the number

restrict a

programmer annotations can also tell the compiler that no other pointer can access the memory pointed to by a

- How to number:
 - Number each pointer/index pair
 - Any pointer/index pair that are not proven not to alias must be incremented at each instruction

(a[i],3) = (x[j],1) + (y[k],2); (b[i],6) = (a[i],3);

Optimizing over wider regions

- Local value numbering operated over just one basic block.
- We want optimizations that operate over several basic blocks (a region), or across an entire procedure (global)
- For this, we need Control Flow Graphs and Flow Analysis

A graph where:

- nodes are basic blocks
- edges mean that it is possible for one block to branch to another

stä	art	:		
r0	=	• • •	;	
r1	=	• • •	;	
br	r0	, i.	f,	else;
<i>if:</i> r2 br	; = en	 d_i.	; f;	
els	se:			

r3 = ...;

end_if: r4 = ...;

A graph where:

- nodes are basic blocks
- edges mean that it is possible for one block to branch to another

sta	art:		
r0	= .	••;	
r1	= .	••;	
br	r0,	if,	else;





A graph where:

- nodes are basic blocks
- edges mean that it is possible for one block to branch to another



Simple analysis:

What property did we rely on for local value numbering?



Simple analysis:

What property did we rely on for local value numbering?

we say that a node b_x "dominates" another node b_y iff:

every path from the start to b_y goes through b_x



Simple analysis:

What property did we rely on for local value numbering?

we say that a node b_x "dominates" another node b_y iff:

every path from the start to b_y goes through b_x



b0 dominates b3

Other examples

- The PyCFG tool draws CFGs for simple python code:
 - Single statement basic blocks

Dominance

• Given a CFG, determine for each node b_x , the set of nodes that dominate b_x



Node	Dominators
B0	во
B1	B0, B1
B2	B0, B1, B2
B3	
B4	
B5	
B6	B0,B1,B5,B6
B7	
B8	



Node	Dominators
B0	ВО
B1	B0, B1
B2	B0, B1, B2
B3	B0, B1, B3
B4	B0, B1, B3, B4
B5	B0, B1, B5
B6	B0, B1, B5, B6
B7	B0, B1, B5, B7
B8	B0, B1, B5, B8



Node	Dominators
B0	BO
B1	B0, B1
B2	B0, B1, B2
B3	B0, B1, B3
B4	B0, B1, B3, B4
B5	B0, B1, B5
B6	B0, B1, B5, B6
B7	B0, B1, B5, B7
B8	<mark>B0, B1, B5, B8</mark>

Can treat this sequence as region, i.e. and perform local value numbering over it



Computing Dominance

- Iterative fixed point algorithm
- Initial state, all nodes start with all other nodes are dominators:
 - *Dom(n)* = *N*
 - Dom(start) = {start}

iteratively compute:

$$Dom(n) = \{n\} \cup (\bigcap_{\min preds(n)} Dom(m))$$

initial conditions

Node	Initial
B0	B0
B1	N
B2	N
B3	N
B4	N
B5	N
B6	N
B7	N
B8	N



$Dom(n) = \{n\} \cup (\bigcap_{\min preds(n)} Dom(m))$

Node	Initial	11
ВО	ВО	
B1	Ν	BO, B1
B2	Ν	B0, B1, B2
B3	Ν	B0, B1, B2, B3
B4	Ν	
B5	Ν	
B6	N	
В7	Ν	
B8	Ν	



$Dom(n) = \{n\} \cup (\bigcap_{\min preds(n)} Dom(m))$

Node	Initial	11
ВО	BO	ВО
B1	Ν	B0,B1
B2	Ν	B0,B1,B2
B3	Ν	B0,B1,B2,B3
B4	Ν	B0,B1,B2,B3,B4
B5	Ν	B0,B1,B5
B6	Ν	B0,B1,B5,B6
B7	Ν	B0,B1,B5,B6,B7
B8	Ν	B0,B1,B5,B8


Node	Initial	11	12
ВО	B0	ВО	
B1	Ν	B0,B1	
B2	N	B0,B1,B2	
B3	Ν	B0,B1,B2,B3	
B4	Ν	B0,B1,B2,B3,B4	
B5	Ν	B0,B1,B5	
B6	Ν	B0,B1,B5,B6	
В7	Ν	B0,B1,B5,B6,B7	B0, B1, B5
B8	Ν	B0,B1,B5,B8	



Node	Initial	11	12
ВО	ВО	ВО	
B1	Ν	B0,B1	
B2	Ν	B0,B1,B2	
B3	Ν	B0,B1,B2,B3	B0,B1,B3
B4	Ν	B0,B1,B2,B3,B4	B0,B1,B3,B4
B5	Ν	B0,B1,B5	
B6	Ν	B0,B1,B5,B6	
В7	Ν	B0,B1,B5,B6,B7	B0,B1,B5,B7
B8	Ν	B0,B1,B5,B8	



Node	Initial	11	12	13
B0	B0	ВО		
B1	N	B0,B1		
B2	N	B0,B1,B2		
B3	N	B0,B1,B2,B3	B0,B1,B3	
B4	N	B0,B1,B2,B3,B4	B0,B1,B3,B4	
B5	N	B0,B1,B5		
B6	N	B0,B1,B5,B6		
B7	N	B0,B1,B5,B6,B7	B0,B1,B5,B7	
B8	N	B0,B1,B5,B8		



Node	Initial	11	12	13
ВО	B0	B0		
B1	N	B0,B1		
B2	N	B0,B1,B2		
B3	N	B0,B1,B2,B3	B0,B1,B3	
B4	N	B0,B1,B2,B3,B4	B0,B1,B3,B4	
B5	N	B0,B1,B5		
B6	N	B0,B1,B5,B6		
B7	N	B0,B1,B5,B6,B7	B0,B1,B5,B7	
B8	N	B0,B1,B5,B8		



No change so algorithm terminates!

Next week

- Flow analysis examples continued:
 - node traversal order for faster convergence
 - Live variable analysis
- Generalized framework for flow analysis
- SSA form