CS577 Modern Language Processors
Spring 2006
Lecture 5
What about memory operations (globals, heap values)?
- Need to model dependences, changes in values.
- SSA doesn’t help directly.

Dependences (order of operations):

\[
g.a = 0; \\
g.b = 1; \\
g.a = g.b + 2; \\
g.b = g.a + 3;
\]

Last two statements must be done in order due to:
- **flow dependence** (write-before-read) on \(g.a\)
- **anti-dependence** (read-before-write) on \(g.b\)

First and third statements must be done in order due to:
- **output dependence** (write-before-write) on \(g.a\)
**Using Alias Information**

When can we do common subexpression elimination to save loads?

```c
    g.a = 10;
    h.a = 20;
    if (g.a == 10) // can only avoid if g and h don’t alias.
        ...
    g.b = 20;
    if (g.a == 10) // can always avoid load here given type info
        ...
```
Good example of complete system based on SSA. To cope with memory operations, they add explicit "threading" store variables.

```
int method (int a[], int b[]) {
    arr_store(a,0,10);
    arr_store(b,0,20);
    return arr_fetch(a,0);
}
```

becomes

```
(int,Store) method (int a[], int b[], Store S0) {
    S1 = arr_store(a,0,10,S0);
    S2 = arr_store(b,0,20,S1);
    return (arr_fetch(a,0,S2),S2);
}
```

where S0,S1,S2 are pseudo-values representing the global store.

Can now continue to use congruence testing to detect redundant computations.
Helps improve understanding of dependence between memory operations.

In last example, \( a \) and \( b \) might be the same array, e.g., called as method\((c,c)\).

Simplest form of alias analysis just uses types:

```c
int method (int a[], short b[]) {
    arr_store(a,0,10);
    arr_store(b,0,20);
    return arr_fetch(a,0); }
```

Now know \( a \) and \( b \) cannot be aliased to the same array.
A more sophisticated analysis (requiring dataflow analysis) tracks creation points:

```java
int method() {
    int a[] = new int[10];
    int b[] = new int[10];
    arr_store(a,0,10);
    arr_store(b,0,20);
    return arr_fetch(a,0); }
```

Once again, $a$ and $b$ cannot be aliased to the same array, even though they have the same type.
Can represent the results of this analysis by changing the store argument dependencies:

\[
\begin{align*}
S1 &= \text{arr\_store}(a, 0, 10, S0); \\
S2 &= \text{arr\_store}(b, 0, 20, S0); \quad // \text{not S1}! \\
S3 &= \text{phi}(S1, S2); \\
\text{return} \ (\text{arr\_fetch}(a, 0, S1), \quad // \text{not S2}! \\
\text{S3});
\end{align*}
\]

But this doesn’t scale well.
Aliasing problems arise:
- in heap for Java
- more broadly in CBR languages
- everywhere in C!

Divide memory pointers into alias classes that are guaranteed not to alias with each other.

Can use:
- types
- field names
- known objects

Alias analysis interacts with:
- class analysis (enhance type analysis to use knowledge about Java class hierarchy)
- escape analysis (determine which values can out-live the function that created them)
REGISTER ALLOCATION

Task: Manage scarce resources (registers) in environment with imperfect information (static program text) about dynamic program behavior.

General aim is to keep frequently-used values in registers as much as possible, to lower memory traffic. Can have a large effect on program performance.

Variety of approaches are possible, differing in sophistication and in scope of analysis used.

Allocator may be unable to keep every “live” variable in registers; must then “spill” variables to memory. Spilling adds new instructions, which often affects the allocation analysis, requiring a new iteration.
To determine how long to keep a given variable (or temporary) in a register, need to know the range of instructions for which the variable is live.

A variable or temporary is live immediately following an instruction if its current value will be needed in the future (i.e., it will be used again, and it won’t be changed before that use).

! live after instruction:

```
mov 3, %t2 ! %t2
mov %t2, %t3 ! %t2 %t3
add %t3, 4, %t4 | %t2 %t4
add %t2, %t4, %t4 | %t2 %t4
st %t4, [a] | (nothing)
```

It’s easy to calculate liveness for a consecutive series of instructions without branches, just by working backwards.
But if a value can stay in a register over a jump, things get harder.

```
1 mov 0, %t1
2 L1: add %t1, 1, %t2
3 add %t3, %t2, %t3
4 mul %t2, 2, %t1
5 cmp %t1, 1000
6 bl L1
7 return %t3
```

! live after instruction:

%t1 %t3
%t2 %t3
%t2 %t3
%t1 %t3
%t1 %t3
%t1 %t3
(none)

To calculate liveness in this case requires iterative flow analysis and the result is only conservative approximation to true liveness (more later).

The live range of a variable is the set of instructions which leave it live. E.g. in 2nd example, live range of %t1 is \{1, 4, 5, 6\}.

Basic idea: If two variables have disjoint live ranges, they can occupy the same physical register.

So in both examples, 2 physical registers suffice to allocate all vars.
Using live ranges turns out to be computationally expensive (more later). A simple alternative is to approximate each live range by a **live interval**. This is the consecutive interval of instructions between the first and last use of each variable.

\[
\begin{align*}
1 & \text{ mov 0, } %t1 & \text{ live after instruction:} & \%t1 & \%t3 \\
2 & \text{ L1: add } %t1, 1, %t2 & \%t2 & \%t3 \\
3 & \text{ add } %t3, %t2, %t3 & \%t2 & \%t3 \\
4 & \text{ mul } %t2, 2, %t1 & \%t1 & \%t3 \\
5 & \text{ cmp } %t1, 1000 & \%t1 & \%t3 \\
6 & \text{ bl L1} & \%t1 & \%t3 \\
7 & \text{ return } %t3 & \text{(nothing)}
\end{align*}
\]

Live ranges: %t1: 1,4,5,6 %t2:2,3 %t3:1,2,3,4,5,6
Live intervals: %t1: [1,6] %t2: [2,3] %t3: [1,6]

(Revised) Basic idea: if two temporaries have non-overlapping live intervals, they can occupy the same physical register.
LINEAR SCAN ALLOCATION ALGORITHM DETAILS

1. Compute \( \text{startpoint}[i] \) and \( \text{endpoint}[i] \) of live interval \( i \) for each variable. Store the intervals in a list in order of increasing start point.

2. Initialize set \( \text{active} := \emptyset \) and pool of free registers = all usable registers.

3. For each live interval \( i \) in order of increasing start point:
   
   (a) For each interval \( j \) in \( \text{active} \), in order of increasing end point
   
   • If \( \text{endpoint}[j] \geq \text{startpoint}[i] \) break to step 3b.
   • Remove \( j \) from \( \text{active} \).
   • Add \( \text{register}[j] \) to pool of free registers.

   (b) Set \( \text{register}[i] := \) next register from pool of free registers, and remove it from pool. (If pool is already empty, need to spill.)

   (c) Add \( i \) to \( \text{active} \), sorted by increasing end point.

A more complicated form of linear scan does take advantage of knowing precise live ranges (expressed as "holes" in live intervals).
Liveness is another property that is computed using dataflow analysis. Classic algorithm is defined on standard CFG (not necessarily in SSA).

\[
\begin{align*}
\text{gen}[n] &= \text{set of variables used by node } n \\
\text{kill}[n] &= \text{set of variables defined by node } n
\end{align*}
\]

\[
\begin{align*}
s & \quad \text{gen}[s] \quad \text{kill}[s] \\
t & \leftarrow b \quad bop \quad c & \{b,c\} & \{t\} \\
t & \leftarrow M[b] & \{b\} & \{t\} \\
M[a] & \leftarrow b & \{a,b\} & {} \\
\text{if } a \text{ relop } b & \text{ then } L & \{a,b\} & {} \\
t & \leftarrow f(a_1, \ldots an) & \{a_1, \ldots an\} & \{t\}
\end{align*}
\]

Note that flow equations for this problem are backwards, i.e., data flows in reverse direction from control flow.

\[
\begin{align*}
\text{in}[n] &= \text{gen}[n] \cup (\text{out}[n] - \text{kill}[n]) \\
\text{out}[n] &= \bigcup_{s \in \text{succ}[n]} \text{in}[s]
\end{align*}
\]

In this case we want the fixed point solution having the smallest sets.
LIVENESS EXAMPLE

1  a <- 0
2  L:  b <- a + 1
3  c <- c + b
4  a <- b * 2
5  if a < N goto L
6  f(c)

Assume that a, b, c are local variables not used after the termination of this code fragment.

We can extract the problem parameters:

<table>
<thead>
<tr>
<th>n</th>
<th>succ[n]</th>
<th>gen[n]</th>
<th>kill[n]</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2,6</td>
<td>a</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>b,c</td>
<td>c</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-</td>
<td>a</td>
</tr>
</tbody>
</table>
Here’s a solution. The control flow equations are solved as usual, but it is more efficient (takes fewer iterations) to fill them in in (roughly) reverse execution order, computing \texttt{out[]} before \texttt{in[]}.

That’s because liveness is a "backwards" flow problem.

\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
\textbf{iteration 0} & \textbf{iteration 1} & \textbf{iteration 2} & \textbf{iteration 3} \\
\hline
\textbf{n} & \textbf{out[n]} & \textbf{in[n]} & \textbf{out[n]} & \textbf{in[n]} & \textbf{out[n]} & \textbf{in[n]} & \textbf{out[n]} & \textbf{in[n]} \\
\hline
6 & - & - & - & c & - & c & - & c \\
5 & - & - & c & a,c & a,c & a,c & a,c & a,c \\
4 & - & - & a,c & b,c & a,c & b,c & a,c & b,c \\
3 & - & - & b,c & b,c & b,c & b,c & b,c & b,c \\
2 & - & - & b,c & a,c & b,c & a,c & b,c & a,c \\
1 & - & - & a,c & c & a,c & c & a,c & c \\
\hline
\end{tabular}

Note that this in example, no more than two of \{a, b, c\} are ever simultaneously live, so two registers will suffice to hold these variables at all times.
VISUALIZING RESULTS

1  a <- 0
2  L:  b <- a + 1
3     c <- c + b
4  a <- b * 2
5  if a < N goto L
6     f(c)

<table>
<thead>
<tr>
<th>Live ranges</th>
<th>Live Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
</tbody>
</table>
STATIC VS. DYNAMIC LIVENESS EXAMPLE
1 V
|------------------|
a = b*b
|------------------|

2 V
|------------------|
c = a + b
|------------------|

3 V
|------------------|
c >= b ?
|------------------|

4 V
|------------------|
return a
|------------------|

5 V
|------------------|
return c
|------------------|
Is a live-out at node 2? It depends on whether control flow ever reaches node 4.

A smart compiler could answer no.

A smarter compiler could answer similar questions about more complicated programs.

But no compiler can ever always answer such questions correctly. This is a consequence of the uncomputability of the Halting Problem.

So we must be content with static liveness, which talks about paths of control-flow edges, and is just a conservative approximation of dynamic liveness, which talks about actual execution paths.
Coloring Register Interference Graphs

More general mechanism for doing register allocation is by translation to graph coloring.

- Build a **register interference graph**, which has
  - a node for each logical register.
  - an edge between two nodes if the corresponding registers are simultaneously live.
- Attempt to **color** the nodes of the graph so that no two nodes connected by an edge have the same color.
  (Like coloring a map, where nodes=countries and edges connect countries with a common border.)
- If we have \( k \) physical registers, we try to color with \( k \) colors.
- If this fails, we must spill and try again. (Nasty in practice.)
EXAMPLE

Live after instr.

```
ld a,t0 ; a:t0 t0
ld b,t1 ; b:t1 t0 t1
sub t0,t1,t2 ; t:t2 t0 t2
ld c,t3 ; c:t3 t0 t2 t3
sub t0,t3,t4 ; u:t4 t0 t2 t4
add t2,t4,t5 ; v:t5 t4 t5
add t5,t4,t6 ; d:t6 t6
st t6,d
```

```
/_____________________
|                     |
t0 ------ t1 t2      
|                     |
|                     |
|                     |
|                     |
|                     |
t3 t4------- t5 t6   
```
COLORING HEURISTICS

In general case, determining whether a graph can be $k$-colored is hard (N.P. Complete, and hence probably exponential).

But a simple heuristic will **usually** find a $k$-coloring if there is one.

1. Choose a node with fewer than $k$ neighbors.

2. Remove that node. Note that if we can color the resulting graph with $k$ colors, we can also color the original graph, by giving the deleted node a color different from all its neighbors.

3. Repeat until **either**
   
   - there are no nodes with fewer than $k$ neighbors, in which case we must spill; **or**
   
   - the graph is gone, in which case we can color the original graph by adding the deleted nodes back in one at a time and coloring them.

For our example, this heuristic finds a 3-coloring, which is the best we can do.
LIVENESS AND COLORING FOR SSA GRAPH

Computing liveness is easier for SSA graphs:

- Don’t need kill sets!
- Must be careful about $\phi$-nodes:
  - Assignments must be done in parallel.
  - Arguments are not really live simultaneously.

(Very) recent work indicates that interference graphs derived from SSA code can be colored in polynomial time

- Should allow considerably simpler coloring-based allocation algorithms.