Intra-Procedural Dataflow Analysis

Backward Analyses

Markus Schordan

Institut für Computersprachen
Technische Universität Wien

Live Variable Analysis

A variable is live at the exit from a label if there is a path from the label to a use of the variable that does not re-define the variable.

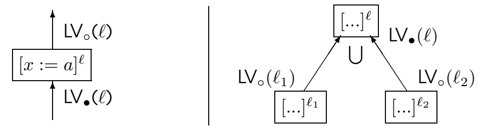
The aim of the Live Variables Analysis is to determine

For each program point, which variables may be live at the exit from the point.

[y := 0]^0; [u := a+b]^1; [y := a*u]^2; while [y $>$ u]^3 do [a := a + 1]^4; [u := a + b]^5; [x := u]^6 od

dead dead

Basic Idea



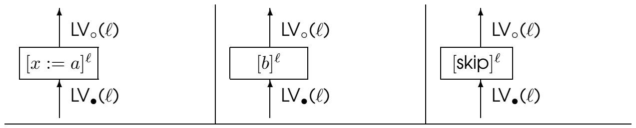
Analysis information: $LV_\circ(\ell), LV_\bullet(\ell) : Lab_\star \to \mathcal{P}(Var_\star)$

- $LV_\circ(\ell)$: the variables that are live at entry of block ℓ.
- $LV_\bullet(\ell)$: the variables that are live at exit of block ℓ.

Analysis properties:

- Direction: backward
- May analysis with combination operator $\bigcup$

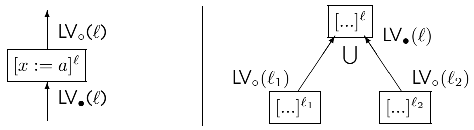
Analysis of Elementary Blocks



$$
\begin{aligned}
kill_{LV}([x := a]^\ell) &= \{x\} \\
kill_{LV}([skip]^\ell) &= \emptyset \\
kill_{LV}([b]^\ell) &= \emptyset \\
gen_{LV}([x := a]^\ell) &= FV(a) \\
gen_{LV}([skip]^\ell) &= \emptyset \\
gen_{LV}([b]^\ell) &= FV(b)
\end{aligned}
$$

$$LV_\circ(\ell) = (LV_\bullet(\ell) \backslash kill_{LV}(B^\ell)) \cup gen_{LV}(B^\ell) \quad \text{where } B^\ell \in blocks(S_\star)$$

Analysis of the Program



$$LV_\circ(\ell) = (LV_\bullet(\ell) \backslash kill_{LV}(B^\ell)) \cup gen_{LV}(B^\ell) \quad \text{where } B^\ell \in blocks(S_\star)$$

$$LV_\bullet(\ell) = \begin{cases} \emptyset & : \text{ if } \ell = final(S_\star) \\ \bigcup\{LV_\circ(\ell') \mid (\ell', \ell) \in flow^R(S_\star)\} & : \text{ otherwise} \end{cases}$$

Example

Program	$LV_\bullet(\ell)$	$LV_\circ(\ell)$	ℓ	$kill_{LV}(\ell)$	$gen_{LV}(\ell)$
[y := 0]^0;	{a, b}	{a, b}	0	{y}	$\emptyset$
[u := a+b]^1;	{u, a, b}	{a, b}	1	{u}	{a,b}
[y := a*u]^2;	{u, a, b, y}	{u, a, b}	2	{y}	{a,u}
while[y $>$ u]^{3}do	{a, b, y}	{u, a, b, y}	3	$\emptyset$	{y,u}
[a := a + 1]^4;	{a, b, y}	{a, b, y}	4	{a}	{a}
[u := a + b]^5;	{u, a, b, y}	{a, b, y}	5	{u}	{a,b}
[x := u]^6 od	{u, a, b, y}	{u, a, b, y}	6	{x}	{u}
[skip]^7	$\emptyset$	$\emptyset$	7	$\emptyset$	$\emptyset$

Dead Code Elimination (DCE)

An assignment $[x := a]^\ell$ is dead if the value of x is not used before it is redefined. Dead assignments can be eliminated.

Analysis: Live Variables Analysis

Transformation: For each $[x := a]^\ell$ in $S_\star$ with $x \notin \mathsf{LV}_\bullet(\ell)$ (i.e. dead) eliminate $[x := a]^\ell$ from the program.

Example:

Before:

$[\mathsf{y} := 0]^0$; $[\mathsf{u} := \mathsf{a{+}b}]^1$; $[\mathsf{y} := \mathsf{a{*}u}]^2$; while $[\mathsf{y} > \mathsf{u}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{u} := \mathsf{a} + \mathsf{b}]^5$; $[\mathsf{x} := \mathsf{u}]^6$ od

After:

$[\mathsf{u} := \mathsf{a{+}b}]^1$; $[\mathsf{y} := \mathsf{a{*}u}]^2$; while $[\mathsf{y} > \mathsf{u}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{u} := \mathsf{a} + \mathsf{b}]^5$; od

Example: Combining Optimizations

Example:
$[\mathsf{x} := \mathsf{a{+}b}]^1$; $[\mathsf{y} := \mathsf{a{*}x}]^2$; while $[\mathsf{y} > \mathsf{a{+}b}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{x} := \mathsf{a} + \mathsf{b}]^5$ od

Common Subexpression Elimination gives

$[\mathsf{u} := \mathsf{a{+}b}]^{1'}$; $[\mathsf{x} := \mathsf{u}]^1$; $[\mathsf{y} := \mathsf{a{*}x}]^2$; while $[\mathsf{y} > \mathsf{u}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{u} := \mathsf{a} + \mathsf{b}]^{5'}$; $[\mathsf{x} := \mathsf{u}]^5$ od

Copy Propagation gives

$[\mathsf{u} := \mathsf{a{+}b}]^{1'}$; $[\mathsf{y} := \mathsf{a{*}u}]^2$; while $[\mathsf{y} > \mathsf{u}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{u} := \mathsf{a} + \mathsf{b}]^{5'}$; $[\mathsf{x} := \mathsf{u}]^5$ od

Dead Code Elimination gives

$[\mathsf{u} := \mathsf{a{+}b}]^1$; $[\mathsf{y} := \mathsf{a{*}u}]^2$; while $[\mathsf{y} > \mathsf{u}]^3$ do $[\mathsf{a} := \mathsf{a} + 1]^4$; $[\mathsf{u} := \mathsf{a} + \mathsf{b}]^5$; od

What are the results for other optimization sequences?

Faint Variables

Consider the following program consisting of three statements:

$[\mathsf{x} := 1]^1$; $[\mathsf{x} := 2]^2$; $[\mathsf{y} := \mathsf{x}]^3$;

Clearly x is dead at the exit from 1 and y is dead at the exit of 3. But x is live at the exit of 2 although it is only used to calculate a new value for y that turns out to be dead.

We shall say that a variable is a faint variable if it is dead or if it is only used to calculate new values for faint variables; otherwise it is strongly live.

Very Busy Expressions Analysis

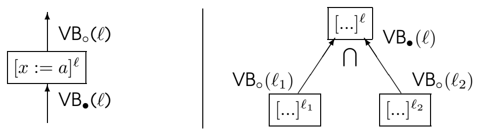
An expression is very busy at the exit from a label if, no matter what path is taken from the label, the expression is always used before any of the variables occurring in it are redefined.

The aim of the Very Busy Expression Analysis is to determine

For each program point, which expressions *must* be very busy at the exit from the point.

if $[\mathsf{a} > \mathsf{b}]^1$ then $([\mathsf{x} := \mathsf{b{-}a}]^2$; $[\mathsf{y} := \mathsf{a{-}b}]^3)$ else $([\mathsf{y} := \mathsf{b{-}a}]^4$; $[\mathsf{x} := \mathsf{a{-}b}]^5)$

Basic Idea



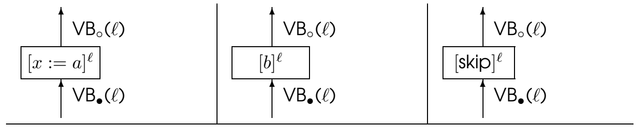
Analysis information: $\mathsf{VB}_\circ(\ell)$,$\mathsf{VB}_\bullet(\ell)$: $\mathsf{Lab}_\star \to \mathcal{P}(\mathsf{AExp}_\star)$

- $\mathsf{VB}_\circ(\ell)$: the expressions that are very busy at entry of block ℓ.
- $\mathsf{VB}_\bullet(\ell)$: the expressions that are very busy at exit of block ℓ.

Analysis properties:

- Direction: backward
- Must analysis with combination operator $\bigcap$

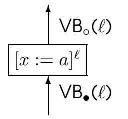
Analysis of Elementary Blocks

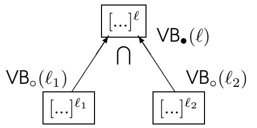


$$
\begin{aligned}
\mathsf{kill}_{\mathsf{VB}}([x := a]^\ell) &= \{a' \in \mathsf{AExp}_\star \mid x \in FV(a')\}\\
\mathsf{kill}_{\mathsf{VB}}([\mathsf{skip}]^\ell) &= \emptyset\\
\mathsf{kill}_{\mathsf{VB}}([b]^\ell) &= \emptyset\\
\mathsf{gen}_{\mathsf{VB}}([x := a]^\ell) &= \mathsf{AExp}(a)\\
\mathsf{gen}_{\mathsf{VB}}([\mathsf{skip}]^\ell) &= \emptyset\\
\mathsf{gen}_{\mathsf{VB}}([b]^\ell) &= \mathsf{AExp}(b)
\end{aligned}
$$

$$\mathsf{VB}_\circ(\ell) = (\mathsf{VB}_\bullet(\ell) \backslash \mathsf{kill}_{\mathsf{VB}}(B^\ell)) \cup \mathsf{gen}_{\mathsf{VB}}(B^\ell) \quad \text{where } B^\ell \in \mathsf{blocks}(S_\star)$$

Analysis of the Program





$$\mathsf{VB}_\circ(\ell) = (\mathsf{VB}_\bullet(\ell) \backslash \mathsf{kill}_{\mathsf{VB}}(B^\ell)) \cup \mathsf{gen}_{\mathsf{VB}}(B^\ell) \quad \text{where } B^\ell \in \mathsf{blocks}(S_\star)$$

$$\mathsf{VB}_\bullet(\ell) = \begin{cases} \emptyset & : \text{ if } \ell = \mathsf{final}(S_\star) \\ \bigcap\{\mathsf{VB}_\circ(\ell') \mid (\ell',\ell) \in \mathsf{flow}^R(S_\star)\} & : \text{ otherwise} \end{cases}$$

Example

if $[a > b]^1$ then ($[x := b{-}a]^2$; $[y := a{-}b]^3$) else ($[y := b{-}a]^4$; $[x := a{-}b]^5$)

ℓ	$\mathsf{VB}_\bullet(\ell)$	$\mathsf{VB}_\circ(\ell)$
1	{a−b, b−a}	{a−b, b−a}
2	{a−b}	{a−b, b−a}
3	$\emptyset$	{a−b}
4	{a−b}	{a−b, b−a}
5	$\emptyset$	{a−b}

ℓ	$\mathsf{kill}_{\mathsf{VB}}(\ell)$	$\mathsf{gen}_{\mathsf{VB}}(\ell)$
1	$\emptyset$	$\emptyset$
2	$\emptyset$	{b−a}
3	$\emptyset$	{a−b}
4	$\emptyset$	{b−a}
5	$\emptyset$	{a−b}

Code Hoisting

Code hoisting finds expressions that are always evaluated following some point in the program regardless of the execution path – and moves them to the earliest point (in execution order) beyond which they would always be executed.

Before:
if $[a > b]^1$ then ($[x := b{-}a]^2$; $[y := a{-}b]^3$) else ($[y := b{-}a]^4$; $[x := a{-}b]^5$)

After:
$[t1 := a{-}b]^0$; $[t2 := b{-}a]^{0'}$;
if $[a > b]^1$ then ($[x := t2]^2$; $[y := t1]^3$) else ($[y := t2]^4$; $[x := t1]^5$)

Summary of Classical Analyses

Analysis	may	must
Forward	Reaching Definitions	Available Expressions
Backward	Live Variables	Very Busy Expressions

Analysis	may	must
Combination Op.	$\cup$	$\cap$
Solution of equ.	smallest	largest

Analysis	Extremal labels set	Abstract flow graph
Forward	$\{\mathsf{init}(S_\star)\}$	$\mathsf{flow}(S_\star)$
Backward	$\mathsf{final}(S_\star)$	$\mathsf{flow}^R(S_\star)$

Bit Vectors

The classical analyses operate over elements of $\mathcal{P}(D)$ where D is a finite set.

The elements can be represented as bit vectors. Each element of D can be assigned a unique bit position i $(1 \leq i \leq n)$. A subset S of D is then represented by a vector of n bits:

- if the i'th element of D is in S then the i'th bit is 1.
- if the i'th element of D is not in S then the i'th bit is 0.

Then we have efficient implementations of

- set union as logical or
- set intersection as logical and

More Bit Vector Framework Examples

- Dual available expressions determines for each program point which expressions may not be available when execution reaches that point (forward may analysis)
- Copy analysis determines whether there on every execution path from a copy statement $x := y$ to a use of x there are no assignments to y (forward must analysis).
- Dominators determines for each program point which program points are guaranteed to have been executed before the current one is reached (forward must analysis).
- Upwards exposed uses determines for a program point, what uses of a variable are reached by a particular definition (assignment) (backward may analysis).

Non-Bit Vector Frameworks

- Constant propagation determines for each program point whether or not a variable has a constant value whenever execution reaches that point.
- Detection of signs analysis determines for each program point the possible signs that the values of the variables may have whenever execution reaches that point.
- Faint variables determines for each program point which variables are faint: a variable is faint if it is dead or it is only used to compute new values of faint variables.
- May be unitialized determines for each program point which variables have dubious values: a variable has a dubious value if either it is not initialized or its value depends on variables with dubious values.

References

- Material for this 3rd lecture

  `www.complang.tuwien.ac.at/markus/optub.html`
- Book

  Flemming Nielson, Hanne Riis Nielson, Chris Hankin:

  Principles of Program Analysis.

  Springer, (2nd edition, 452 pages, ISBN 3-540-65410-0), 2005.
  - Chapter 2 (Data Flow Analysis)