Program Analysis (70020) While Language

Herbert Wiklicky

Department of Computing Imperial College London

herbert@doc.ic.ac.uk
h.wiklicky@imperial.ac.uk

Autumn 2024

Syntactic Constructs

We use the following syntactic categories:

 $a \in AExp$ arithmetic expressions

 $b \in \mathbf{BExp}$ boolean expressions

 $S \in$ **Stmt** statements

The syntax of the language WHILE is given by the following abstract syntax:

а

b

S

The syntax of the language WHILE is given by the following abstract syntax:

```
a ::= x
```

b

 \mathcal{S}

The syntax of the language WHILE is given by the following abstract syntax:

$$a ::= x \mid n$$

b

 \mathcal{S}

$$a ::= x \mid n \mid a_1 \text{ op}_a a_2$$

$$b$$

$$S$$

```
a ::= x \mid n \mid a_1 op_a a_2
b ::= true
S
```

```
a ::= x \mid n \mid a_1 \text{ op}_a a_2
b ::= true \mid false
S
```

```
a ::= x \mid n \mid a_1 \text{ op}_a a_2
b ::= true \mid false \mid not b
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S ::= x := a
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip | S_1; S_2
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip | S_1; S_2 | if b then S_1 else S_2
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a
| skip | S_1; S_2
| if b then S_1 else S_2
| while b do S
```

Syntactical Categories

We assume some countable/finite set of variables is given;

```
x, y, z, \dots \in Var variables n, m, \dots \in Num numerals
```

Syntactical Categories

We assume some countable/finite set of variables is given;

```
x, y, z, \dots \in Var variables n, m, \dots \in Num numerals \ell, \dots \in Lab labels
```

Numerals (integer constants) will not be further defined and neither will the operators:

```
egin{array}{lll} op_a & \in & \mathbf{Op}_a & 	ext{arithmetic} & 	ext{operators, e.g. } +,-,	imes,\dots \\ op_b & \in & \mathbf{Op}_b & 	ext{boolean} & 	ext{operators, e.g. } \wedge,ee,\dots \\ op_r & \in & \mathbf{Op}_r & 	ext{relational} & 	ext{operators, e.g. } =,<,\leq,\dots \\ \end{array}
```

The labelled syntax of the language WHILE is given by the following abstract syntax:

а

b

S

The labelled syntax of the language WHILE is given by the following abstract syntax:

```
a ::= x
```

b

 \mathcal{S}

The labelled syntax of the language WHILE is given by the following abstract syntax:

```
a ::= x \mid n
```

b

 \mathcal{S}

```
a ::= x \mid n \mid a_1 op_a a_2

b

S
```

```
a ::= x \mid n \mid a_1 op_a a_2
b ::= true
S
```

```
a ::= x \mid n \mid a_1 \text{ op}_a a_2
b ::= true \mid false
S
```

```
a ::= x \mid n \mid a_1 \text{ op}_a a_2
b ::= true \mid false \mid not b
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S ::= [x := a]^{\ell}
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= [x := a]^{\ell} | [skip]^{\ell}
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= [x := a]^{\ell} | [\mathbf{skip}]^{\ell} | S_1; S_2
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= [x := a]^{\ell} | [skip]^{\ell} | S_1; S_2 | if [b]^{\ell} then S_1 else S_2
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= [x := a]^{\ell} | [skip]^{\ell} | S_1; S_2 | if [b]^{\ell} then S_1 else S_2 | while [b]^{\ell} do S
```

An Example in WHILE

An example of a program written in this WHILE language is the following one which computes the factorial of the number stored in x and leaves the result in z:

An Example in WHILE

An example of a program written in this WHILE language is the following one which computes the factorial of the number stored in x and leaves the result in z:

```
[y := x]^1;

[z := 1]^2;

while [y > 1]^3 do (

[z := z * y]^4;

[y := y - 1]^5);

[y := 0]^6
```

An Example in WHILE

An example of a program written in this WHILE language is the following one which computes the factorial of the number stored in x and leaves the result in z:

```
[y := x]^1;

[z := 1]^2;

while [y > 1]^3 do (

[z := z * y]^4;

[y := y - 1]^5);

[y := 0]^6
```

Note the use of meta-symbols, brackets, to group statements.

Concrete Syntax of WHILE

To avoid using brackets (as meta-symbols) we could also use the **concrete syntax** of the language WHILE as follows:

а

b

S

To avoid using brackets (as meta-symbols) we could also use the **concrete syntax** of the language WHILE as follows:

```
a ::= x
```

b

S

To avoid using brackets (as meta-symbols) we could also use the **concrete syntax** of the language WHILE as follows:

```
a ::= x \mid n
```

b

S

$$a ::= x \mid n \mid a_1 op_a a_2$$
 b
 S

```
a ::= x \mid n \mid a_1 op_a a_2
b ::= true
S
```

```
a ::= x \mid n \mid a_1 \text{ op}_a a_2
b ::= true \mid false
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S
```

```
a ::= x \mid n \mid a_1 \ op_a \ a_2
b ::= true \mid false \mid not \ b \mid b_1 \ op_b \ b_2 \mid a_1 \ op_r \ a_2
S ::= x := a
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip | S_1; S_2
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a | skip | S_1; S_2 | if b then S_1 else S_2 fi
```

```
a ::= x | n | a_1 op_a a_2
b ::= true | false | not b | b_1 op_b b_2 | a_1 op_r a_2
S ::= x := a
| skip | S_1; S_2
| if b then S_1 else S_2 fi | while b do S od
```

Initial Label

When presenting examples of Data Flow Analyses we will use a number of operations on programs and labels. The first of these is

which returns the initial label of a statement:

$$init([x := a]^{\ell}) = \ell$$
 $init([\mathbf{skip}]^{\ell}) = \ell$
 $init(S_1; S_2) = init(S_1)$
 $init(\mathbf{if}[b]^{\ell} \mathbf{then} S_1 \mathbf{else} S_2) = \ell$
 $init(\mathbf{while}[b]^{\ell} \mathbf{do} S) = \ell$

Final Labels

We will also need a function which returns the set of final labels in a statement; whereas a sequence of statements has a single entry, it may have multiple exits (e.g. in the conditional):

```
\mathit{final}: \mathbf{Stmt} 	o \mathcal{P}(\mathbf{Lab}) \mathit{final}([\ x := a\ ]^\ell) \ = \ \{\ell\} \mathit{final}([\ \mathbf{skip}\ ]^\ell) \ = \ \{\ell\} \mathit{final}(S_1; S_2) \ = \ \mathit{final}(S_2) \mathit{final}(\mathbf{if}\ [b]^\ell \ \mathbf{then}\ S_1 \ \mathbf{else}\ S_2) \ = \ \mathit{final}(S_1) \cup \mathit{final}(S_2) \mathit{final}(\mathbf{while}\ [b]^\ell \ \mathbf{do}\ S) \ = \ \{\ell\}
```

The **while**-loop terminates immediately after the test fails.

$$ightharpoonup [x := a]^{\ell}$$
, or

- \triangleright [x := a] $^{\ell}$, or
- ► [skip]^ℓ, as well as

- \triangleright [x := a] $^{\ell}$, or
- ▶ [skip]^ℓ, as well as
- tests of the form $[b]^{\ell}$.

Blocks

To access the statements or test associated with a label in a program we use the function

```
blocks: \mathbf{Stmt} 	o \mathcal{P}(\mathbf{Block})
blocks([\ x := a\ ]^\ell) = \{[\ x := a\ ]^\ell\}
blocks([\ \mathbf{skip}\ ]^\ell) = \{[\ \mathbf{skip}\ ]^\ell\}
blocks(S_1;S_2) = blocks(S_1) \cup blocks(S_2)
blocks(\mathbf{if}\ [b]^\ell\ \mathbf{then}\ S_1\ \mathbf{else}\ S_2) = \{[b]^\ell\} \cup \\ blocks(S_1) \cup blocks(S_2)
blocks(\mathbf{while}\ [b]^\ell\ \mathbf{do}\ S) = \{[b]^\ell\} \cup blocks(S)
```

Labels

Then the set of labels occurring in a program is given by

$$\textit{labels}: \textbf{Stmt} \rightarrow \mathcal{P}(\textbf{Lab})$$

where

$$labels(S) = \{\ell \mid [B]^{\ell} \in blocks(S)\}$$

Clearly $init(S) \in labels(S)$ and $final(S) \subseteq labels(S)$.

Flow

$$flow : \mathbf{Stmt} \to \mathcal{P}(\mathbf{Lab} \times \mathbf{Lab})$$

which maps statements to sets of flows:

$$\begin{array}{rcl} \mathit{flow}([\ x := a\]^\ell) &= \emptyset \\ \mathit{flow}([\ \mathsf{skip}\]^\ell) &= \emptyset \\ \mathit{flow}(S_1;S_2) &= \mathit{flow}(S_1) \cup \mathit{flow}(S_2) \cup \\ & \{(\ell,\mathit{init}(S_2)) \mid \ell \in \mathit{final}(S_1)\} \\ \mathit{flow}(\mathsf{if}\ [b]^\ell \ \mathsf{then}\ S_1 \ \mathsf{else}\ S_2) &= \mathit{flow}(S_1) \cup \mathit{flow}(S_2) \cup \\ & \{(\ell,\mathit{init}(S_1)), (\ell,\mathit{init}(S_2))\} \\ \mathit{flow}(\mathsf{while}\ [b]^\ell \ \mathsf{do}\ S) &= \mathit{flow}(S) \cup \{(\ell,\mathit{init}(S))\} \cup \\ & \{(\ell',\ell) \mid \ell' \in \mathit{final}(S)\} \end{array}$$

An Example Flow

Consider the following program, power, computing the x-th power of the number stored in y:

```
[z := 1]^1;

while [x > 1]^2 do (

[z := z * y]^3;

[x := x - 1]^4)
```

An Example Flow

Consider the following program, power, computing the x-th power of the number stored in y:

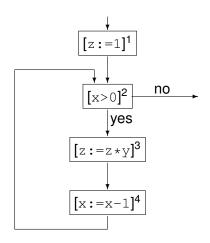
$$[z := 1]^1;$$

while $[x > 1]^2$ do (
 $[z := z * y]^3;$
 $[x := x - 1]^4)$

We have $labels(power) = \{1, 2, 3, 4\}$, init(power) = 1, and $final(power) = \{2\}$. The function flow produces the set:

$$flow(power) = \{(1,2), (2,3), (3,4), (4,2)\}$$

Flow Graph



Forward Analysis

The function *flow* is used in the formulation of *forward analyses*. Clearly init(S) is the (unique) entry node for the flow graph with nodes labels(S) and edges flow(S). Also

$$egin{array}{ll} ext{labels}(\mathcal{S}) &=& \{ ext{init}(\mathcal{S})\} \ \ & \{\ell \mid (\ell,\ell') \in ext{flow}(\mathcal{S})\} \ \ \ & \{\ell' \mid (\ell,\ell') \in ext{flow}(\mathcal{S})\} \end{array}$$

and for composite statements (meaning those not simply of the form $[B]^\ell$) the equation remains true when removing the $\{init(S)\}$ component.

Reverse Flow

In order to formulate *backward analyses* we require a function that computes reverse flows:

$$flow^R : \mathbf{Stmt} \to \mathcal{P}(\mathbf{Lab} \times \mathbf{Lab})$$

$$flow^R(S) = \{(\ell, \ell') \mid (\ell', \ell) \in flow(S)\}$$

For the power program, flow^R produces

$$\{(2,1),(2,4),(3,2),(4,3)\}$$

Backward Analysis

In case final(S) contains just one element that will be the unique entry node for the flow graph with nodes labels(S) and edges $flow^R(S)$. Also

$$\begin{array}{lll} \textit{labels}(\mathcal{S}) & = & \textit{final}(\mathcal{S}) \ \cup \\ & & \{\ell \mid (\ell,\ell') \in \textit{flow}^R(\mathcal{S})\} \ \cup \\ & & \{\ell' \mid (\ell,\ell') \in \textit{flow}^R(\mathcal{S})\} \end{array}$$

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement)

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

▶ Lab_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

- ▶ Lab_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,
- ▶ Var_{\star} to represent the variables $(FV(S_{\star}))$ appearing in S_{\star} ,

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

- ▶ **Lab**_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,
- ▶ **Var**_{*} to represent the variables ($FV(S_*)$) appearing in S_* ,
- ▶ **Block**_{*} to represent the elementary blocks ($blocks(S_*)$) occurring in S_* , and

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

- ▶ **Lab**_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,
- ▶ **Var**_{*} to represent the variables ($FV(S_*)$) appearing in S_* ,
- ▶ **Block**_{*} to represent the elementary blocks ($blocks(S_*)$) occurring in S_* , and
- ▶ **AExp**_{*} to represent the set of *non-trivial* arithmetic subexpressions in S_*

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

- ▶ Lab_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,
- **Var**_{*} to represent the variables $(FV(S_*))$ appearing in S_* ,
- ▶ **Block**_{*} to represent the elementary blocks ($blocks(S_*)$) occurring in S_* , and
- ▶ **AExp** $_{\star}$ to represent the set of *non-trivial* arithmetic subexpressions in S_{\star}

An expression is trivial if it is a single variable or constant.

We will use the notation S_{\star} to represent the program we are analysing (the "top-level" statement) and furthermore:

- ▶ Lab_{*} to represent the labels (*labels*(S_*)) appearing in S_* ,
- **Var**_{*} to represent the variables $(FV(S_*))$ appearing in S_* ,
- ▶ **Block**_{*} to represent the elementary blocks ($blocks(S_*)$) occurring in S_* , and
- ▶ AExp_{*} to represent the set of non-trivial arithmetic subexpressions in S_{*} as well as
- ► **AExp**(*a*) and **AExp**(*b*) to refer to the set of non-trivial arithmetic subexpressions of a given arithmetic, respectively boolean, expression.

An expression is trivial if it is a single variable or constant.

Isolated Entries & Exits

Program S_{\star} has *isolated entries* if:

$$\forall \ell \in \mathsf{Lab} : (\ell, \mathit{init}(S_\star)) \notin \mathit{flow}(S_\star)$$

This is the case whenever S_{\star} does not start with a **while**-loop.

Isolated Entries & Exits

Program S_{\star} has *isolated entries* if:

$$\forall \ell \in \mathsf{Lab} : (\ell, \mathit{init}(S_{\star})) \notin \mathit{flow}(S_{\star})$$

This is the case whenever S_* does not start with a **while**-loop.

Similarly, we shall frequently assume that the program S_{\star} has isolated exits; this means that:

$$\forall \ell_1 \in \mathit{final}(S_\star) \ \forall \ell_2 \in \mathsf{Lab} : (\ell_1, \ell_2) \notin \mathit{flow}(S_\star)$$

Label Consistency

A statement, S, is label consistent if and only if:

$$[B_1]^\ell, [B_2]^\ell \in \mathit{blocks}(S) \text{ implies } B_1 = B_2$$

Label Consistency

A statement, S, is label consistent if and only if:

$$[B_1]^\ell, [B_2]^\ell \in \mathit{blocks}(S) \text{ implies } B_1 = B_2$$

Clearly, if all blocks in *S* are uniquely labelled (meaning that each label occurs only once), then *S* is label consistent.

When S is label consistent the statement or clause "where $[B]^{\ell} \in blocks(S)$ " is unambiguous in defining a partial function from labels to elementary blocks; we shall then say that ℓ labels the block B.