

New York University, CIMS, CS, Course CSCI-GA.3140-001, Spring 2024

“Abstract Interpretation”

Ch. 4, Syntax

Patrick Cousot

pcousot@cs.nyu.edu cs.nyu.edu/~pcousot

Class 2, Monday, January 29th 2024, 4:55-6:55 PM, WWH, Room CIWW 202

These slides are available at

<http://cs.nyu.edu/~pcousot/courses/spring24/CSCI-GA.3140-001/slides/02--2024-01-29-syntax-semantics-traces-oo/slides-04--syntax-AI.pdf>

What did we learned in

Ch. 3, Syntax, Semantics, Properties, and Static Analysis of Expressions?

- To **design a program analysis** (e.g. signs of expressions):
 - (1) Define the **syntax** of programs (e.g. expressions)
 - (2) Define the **semantics** of programs (e.g. what expressions compute)
 - (3) The **collecting** semantics of programs is the strongest property of the program semantics
 - (4) Define the **abstraction** of semantic program properties by a Galois connection (e.g. sign abstraction of value, environment, and semantic properties)
 - (5) Design the **abstract semantics** by calculational design of an abstraction of the collecting semantics

⇒ the abstract semantics specifies a **static analyzer** which is sound by construction
- These slides consider (1) for a mini-language (a subset of C).

Objective

The objective of this [chapter 4, “Syntax”](#) is to [introduce the syntax](#) of the small subset of the C programming language used throughout the book to exemplify abstract interpretation

[en.wikipedia.org/wiki/C_\(programming_language\)](https://en.wikipedia.org/wiki/C_(programming_language))
[en.wikipedia.org/wiki/Syntax_\(programming_languages\)](https://en.wikipedia.org/wiki/Syntax_(programming_languages))

Chapter 4

Ch. 4, Syntax

Syntax of programs

Syntax of statements

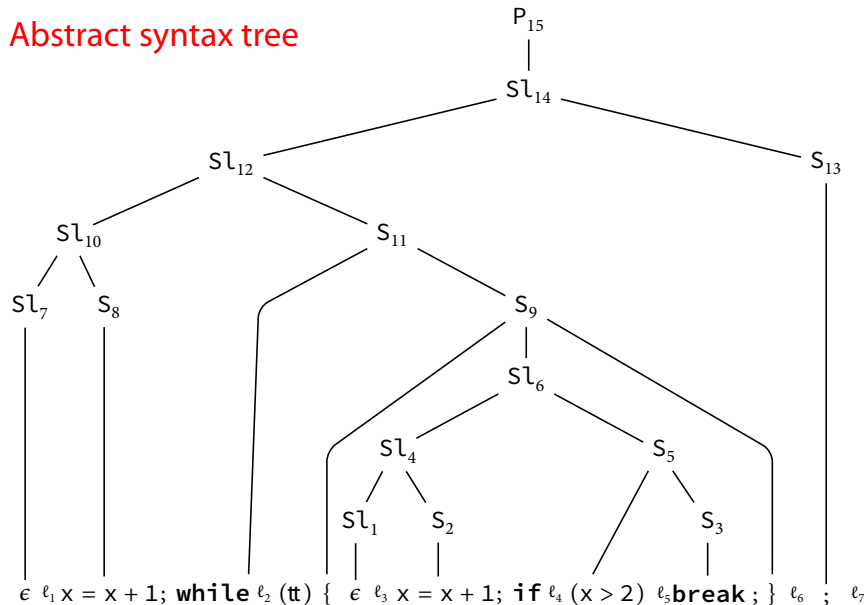
- A subset of C

$S ::=$	$\text{statement } S \in \mathcal{S}$
$x = A ;$	assignment
$;$	skip
$ \text{ if } (B) S$	conditional
$ \text{ if } (B) S \text{ else } S$	
$ \text{ while } (B) S$	iteration
$ \text{ break ;}$	iteration break (to inner enclosing loop exit)
$ \{ S_l \}$	compound statement
$S_l ::= S_l S \mid \epsilon$	$\text{statement list } S_l \in \mathcal{S}_l$

- Example:

$x = x + 1 ; \text{while } (\text{tt}) \{ x = x + 1 ; \text{if } (x > 2) \text{ break ; } \};$ (4.1)

Abstract syntax tree



Syntax of programs

$P ::= S \mid$ program $P \in \mathcal{P}$

$\mathcal{P}_C \triangleq S \cup S \mid \cup \mathcal{P}$ program component $S \in \mathcal{P}_C$

Labels of programs

Labels

- Labels are unique and designate program points of statements
- Labels are not part of the program syntax, their syntax is free
- Example: labelling of program (4.1):

```
ℓ1 x = x + 1 ;  
    while ℓ2 (tt) {  
        ℓ3 x = x + 1 ;  
        if ℓ4 (x > 2) ℓ5 break ; } ℓ6;ℓ7
```

(4.5)

break ; statements are used to exit iteration statements only (so the **break ;** at ℓ₅ exits to ℓ₆).

en.wikipedia.org/wiki/Control_flow#Early_exit_from_loops
[en.wikipedia.org/wiki/Label_\(computer_science\)](https://en.wikipedia.org/wiki/Label_(computer_science))

Labels

$\text{at}[[S]]$	the program point at which execution of program component S starts;
$\text{after}[[S]]$	the program exit point after program component S , at which execution of S is supposed to normally terminate, if ever;
$\text{escape}[[S]]$	a boolean indicating whether or not the program component S contains a break ; statement escaping out of that component S ;
$\text{break-to}[[S]]$	the program point at which execution of the program component S goes to when a break ; statement escapes out of that component S ;
$\text{breaks-of}[[S]]$	the set of labels of all break ; statements that can escape out of S
$\text{in}[[S]]$	the set of program points inside program component S (including $\text{at}[[S]]$ but excluding $\text{after}[[S]]$ and $\text{break-to}[[S]]$);
$\text{labx}[[S]]$	the potentially reachable program points while executing program component S either at, in, or after the program component, or resulting from a break.

Label at a statement S

- $\text{at}[[S]]$: the program point at which execution of S starts;

$P ::= S_l$	$\text{at}[[P]] \triangleq \text{at}[[S_l]]$
$S_l ::= S_l' S$	$\text{at}[[S_l]] \triangleq \text{at}[[S_l']]$
$S_l ::= \epsilon$	$\text{at}[[S_l]] \triangleq \text{after}[[S_l]]$
$S ::= \{ S_l \}$	$\text{at}[[S]] \triangleq \text{at}[[S_l]]$

Label after a statement S

- $\text{after}[[S]]$: the program point at which execution will continue upon termination of the statement S (unless there is a **break** ;;);

$P ::= S_l$	$\text{after}[[P]] \triangleq \text{after}[[S_l]]$
$S_l ::= S_l' S$	$\text{after}[[S_l']] \triangleq \text{at}[[S]], \quad \text{after}[[S]] \triangleq \text{after}[[S_l]]$
$S ::= \text{if } (B) S_t$	$\text{after}[[S_t]] \triangleq \text{after}[[S]]$
$S ::= \text{if } (B) S_t \text{ else } S_f$	$\text{after}[[S_t]] \triangleq \text{after}[[S_f]] \triangleq \text{after}[[S]]$
$S ::= \text{while } (B) S_b$	$\text{after}[[S_b]] \triangleq \text{at}[[S]]$
$S ::= \{ S_l \}$	$\text{after}[[S_l]] \triangleq \text{after}[[S]]$

Explicit labelling

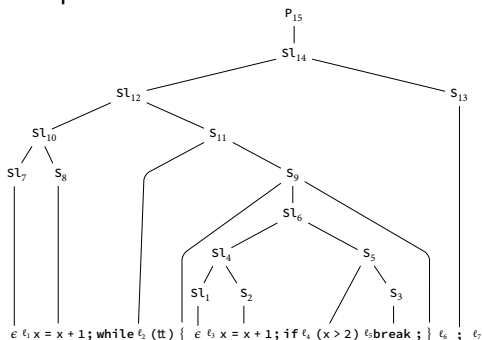
- When explicitly decorating programs with labels, we should have

$S ::= \ell \ x = A ;$	$\text{at}[\![S]\!] \triangleq \ell$
$S ::= \ell ;$	$\text{at}[\![S]\!] \triangleq \ell$
$S ::= \text{if } \ell \ (B) \ S_t$	$\text{at}[\![S]\!] \triangleq \ell$
$S ::= \text{if } \ell \ (B) \ S_t \ \text{else} \ S_f$	$\text{at}[\![S]\!] \triangleq \ell$
$S ::= \text{while } \ell \ (B) \ S_b$	$\text{at}[\![S]\!] \triangleq \ell$
$S ::= \ell \ \text{break} ;$	$\text{at}[\![S]\!] \triangleq \ell$
$P ::= S \ell$	$\text{after}[\![P]\!] \triangleq \text{after}[\![S \ell]\!] \triangleq \ell$

- “In $\ell_1 \ x = A ; \ell_2$, execution starts at ℓ_1 and continues at ℓ_2 ” is a shorthand for “In $x = A ;$, execution starts at $\text{at}[\![x = A ;]\!]$ and continues at $\text{after}[\![x = A ;]\!]$ ”
- The program labelling is arbitrary and can change and be generated automatically

Another choice for labels

- A label can be a path in the abstract syntax tree
- Example:



$$l_3 = P_{15} SL_{14} SL_{12} S_{11} S_9 SL_6 SL_4 S_2$$

Escaping a statement S

- $\text{escape}[[S]]$: true (tt) if and only if the statement S contains a **break** ; that can escape out of that S;

$P ::= S_l$	$\text{escape}[[P]] \triangleq \text{escape}[[S_l]]$	$\text{escape}[[P]] = \text{ff}$
$S_l ::= S_l' S$	$\text{escape}[[S_l]] \triangleq \text{escape}[[S_l']] \vee \text{escape}[[S]]$	
$S_l ::= \epsilon$	$\text{escape}[[S_l]] \triangleq \text{ff}$	
$S ::= x = A ;$	$\text{escape}[[S]] \triangleq \text{ff}$	
$S ::= ;$	$\text{escape}[[S]] \triangleq \text{ff}$	
$S ::= \text{if } (B) S_t$	$\text{escape}[[S]] \triangleq \text{escape}[[S_t]]$	
$S ::= \text{if } (B) S_t \text{ else } S_f$	$\text{escape}[[S]] \triangleq \text{escape}[[S_t]] \vee \text{escape}[[S_f]]$	
$S ::= \text{while } (B) S_b$	$\text{escape}[[S]] \triangleq \text{ff}$	
$S ::= \text{break} ;$	$\text{escape}[[S]] \triangleq \text{tt}$	
$S ::= \{ S_l \}$	$\text{escape}[[S]] \triangleq \text{escape}[[S_l]]$	

Break label of a statement S

- **break-to**[[S]]: the program point at which execution of S goes to when a **break** ; statement is executed while executing S;

$S_l ::= S_l' S$	$\text{break-to}[[S_l']] \triangleq \text{break-to}[[S]] \triangleq \text{break-to}[[S_l]]$
$S ::= \text{if } (B) S_t$	$\text{break-to}[[S_t]] \triangleq \text{break-to}[[S]]$
$S ::= \text{if } (B) S_t \text{ else } S_f$	$\text{break-to}[[S_t]] \triangleq \text{break-to}[[S_f]] \triangleq \text{break-to}[[S]]$
$S ::= \text{while } (B) S_b$	$\text{break-to}[[S_b]] \triangleq \text{after}[[S]]$
$S ::= \{ S_l \}$	$\text{break-to}[[S_l]] \triangleq \text{break-to}[[S]]$

Break label of a statement S

- **breaks-of** $\llbracket S \rrbracket$: collects the labels of all **break** ; statements that can escape out of S (so excluding **break** ; statements inside an iteration statement within S).
- The definition checks that **break** ; statements can only appear within loops;

$P ::= S_l$	$\text{breaks-of} \llbracket P \rrbracket \triangleq \text{breaks-of} \llbracket S_l \rrbracket$	$\text{breaks-of} \llbracket P \rrbracket = \emptyset$
$S_l ::= S_l' S$	$\text{breaks-of} \llbracket S_l \rrbracket \triangleq \text{breaks-of} \llbracket S_l' \rrbracket \cup \text{breaks-of} \llbracket S \rrbracket$	
$S_l ::= \epsilon$	$\text{breaks-of} \llbracket S_l \rrbracket \triangleq \emptyset$	
$S ::= x = A ;$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \emptyset$	
$S ::= ;$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \emptyset$	
$S ::= \text{if } (B) S_t$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \text{breaks-of} \llbracket S_t \rrbracket$	
$S ::= \text{if } (B) S_t \text{ else } S_f$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \text{breaks-of} \llbracket S_t \rrbracket \cup \text{breaks-of} \llbracket S_f \rrbracket$	
$S ::= \text{while } (B) S_b$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \emptyset$	
$S ::= \ell \text{ break ;}$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \{\text{at} \llbracket S \rrbracket\}$	
$S ::= \{ S_l \}$	$\text{breaks-of} \llbracket S \rrbracket \triangleq \text{breaks-of} \llbracket S_l \rrbracket$	

Labels in a statement S

- $\text{in}[[S]]$: the program points inside S (including $\text{at}[[S]]$ but excluding $\text{after}[[S]]$ and $\text{break-to}[[S]]$)
- The definition checks that program labels are unique.

$P ::= S_l$	$\text{in}[[P]] \triangleq \text{in}[[S_l]]$	$\text{after}[[S_l]] \notin \text{in}[[S_l]]$
$S_l ::= S_l' S$	$\text{in}[[S_l]] \triangleq \text{in}[[S_l']] \cup \text{in}[[S]]$	$\text{in}[[S_l']] \cap \text{in}[[S]] = \emptyset$ when $S_l' \neq \{ \dots \{ \epsilon \} \dots \}$
$S_l ::= \epsilon$	$\text{in}[[S_l]] \triangleq \{ \text{at}[[S_l]] \}$	
$S ::= x = A ;$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \}$	
$S ::= ;$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \}$	
$S ::= \text{if } (B) S_t$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \} \cup \text{in}[[S_t]]$	$\text{at}[[S]] \notin \text{in}[[S_t]]$
$S ::= \text{if } (B) S_t \text{ else } S_f$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \} \cup \text{in}[[S_t]] \cup \text{in}[[S_f]]$	$\text{at}[[S]] \notin \text{in}[[S_t]] \cup \text{in}[[S_f]]$ $\text{in}[[S_t]] \cap \text{in}[[S_f]] = \emptyset$
$S ::= \text{while } (B) S_b$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \} \cup \text{in}[[S_b]]$	$\text{at}[[S]] \notin \text{in}[[S_b]]$
$S ::= \text{break} ;$	$\text{in}[[S]] \triangleq \{ \text{at}[[S]] \}$	
$S ::= \{ S_l \}$	$\text{in}[[S]] \triangleq \text{in}[[S_l]]$	

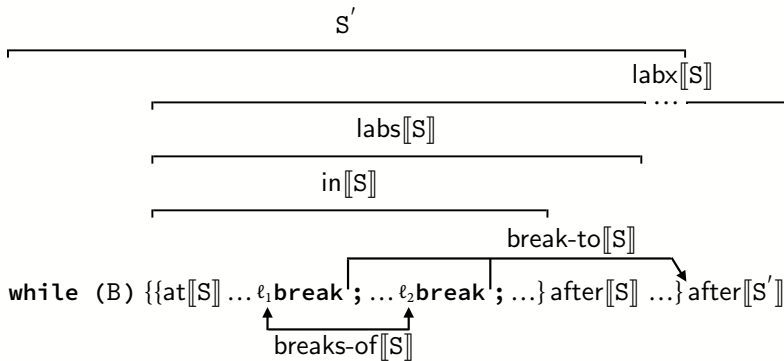
Labels of a statement S

- $\text{labs}[[S]]$: the potentially reachable program points while executing S either in or after the statement;
- $\text{labx}[[S]]$: the potentially reachable program points while executing S either in or after the statement, or resulting from a break.

$$\text{labs}[[S]] \triangleq \text{in}[[S]] \cup \{\text{after}[[S]]\}$$

$$\text{labx}[[S]] \triangleq \text{labs}[[S]] \cup (\text{escape}[[S]] \text{ ? } \{\text{break-to}[[S]]\} \text{ : } \emptyset)$$

Informal illustration of the definitions



Example of program labelling (home work)

The program $P = \text{while } \ell_1 (\text{tt}) \ell_2 \text{ break ; } \ell_3 x = 7 ; \ell_4$ has grammatical structure

$$P \left[\begin{array}{c} S\ell_1 \left[\begin{array}{c} S\ell_2 \left[\begin{array}{c} S\ell_3 \left[\begin{array}{c} \epsilon \\ \text{while } \ell_1 (\text{tt}) \\ S_5 \left[\begin{array}{c} \ell_2 \text{ break ;} \\ \ell_3 x = 7 ; \end{array} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \ell_4$$

The labelling is

S	at[S]	in[S]	escape[S]	break-to[S]	after[S]	labs[S]	labx[S]
P	ℓ_1	ℓ_1, ℓ_2, ℓ_3	ff	—	ℓ_4	$\ell_1, \ell_2, \ell_3, \ell_4$	$\ell_1, \ell_2, \ell_3, \ell_4$
$S\ell_1$	ℓ_1	ℓ_1, ℓ_2, ℓ_3	ff	—	ℓ_4	$\ell_1, \ell_2, \ell_3, \ell_4$	$\ell_1, \ell_2, \ell_3, \ell_4$
$S\ell_2$	ℓ_1	ℓ_1, ℓ_2	ff	—	ℓ_3	ℓ_1, ℓ_2, ℓ_3	ℓ_1, ℓ_2, ℓ_3
$S\ell_3$	ℓ_1	ℓ_1	ff	—	ℓ_1	ℓ_1	ℓ_1
S_4	ℓ_1	ℓ_1, ℓ_2	ff	—	ℓ_3	ℓ_1, ℓ_2, ℓ_3	ℓ_1, ℓ_2, ℓ_3
S_5	ℓ_2	ℓ_2	tt	ℓ_3	ℓ_1	ℓ_1, ℓ_2	ℓ_1, ℓ_2, ℓ_3
S_6	ℓ_3	ℓ_3	ff	—	ℓ_4	ℓ_3, ℓ_4	ℓ_3, ℓ_4

Properties of the program labelling

Lemma (4.15) For all program components S of a program P , $\text{at}[[S]] \in \text{in}[[S]]$.

Lemma (4.16) For all program non-empty components $S \neq \{ \dots \{ \epsilon \} \dots \}$ of a program P , $\text{after}[[S]] \notin \text{in}[[S]]$.

Lemma (4.17) For all program components S of a program P , $\text{escape}[[S]] \Rightarrow (\text{break-to}[[S]] \notin \text{in}[[S]])$.

Lemma (4.18) For all program components S of a program P , $\text{escape}[[S]] \Rightarrow (\text{break-to}[[S]] \neq \text{after}[[S]])$.

Proofs in the book.

The End, Thank you