

Abstract Interpretation

Software Quality Assurance - Static Code Analysis, II | Florian Sihler | December 11, 2024

Outline

1. The Why

2. The How

3. Semantics

4. Soundness and Completeness

5. Outlook

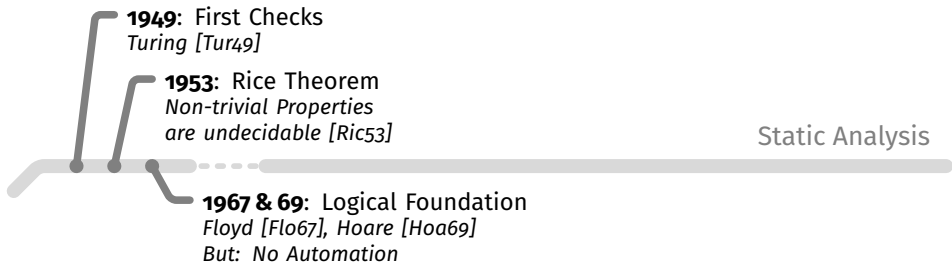
1. The Why

The Why

```
public static void main(String[] args) {  
    int a = 1;           { a ∈ {1} }  
    double r = Math.random() * 10; { r ∈ [0..10) }  
    if (r > 5) {  
        a = 2;           { a ∈ {2} }  
    }  
    System.out.println(a); { a ∈ {1,2} } → Valid? Ok? Safe?  
}
```

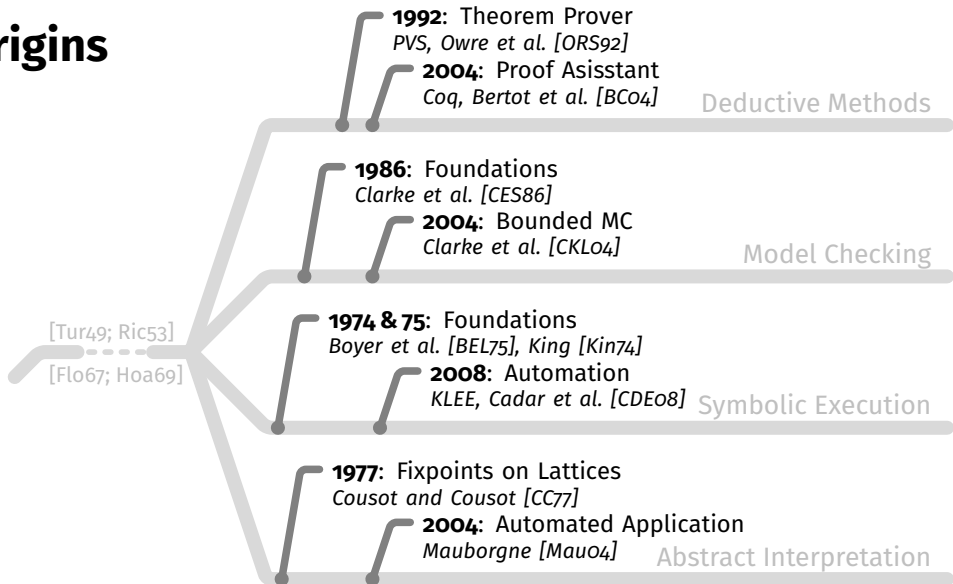
- We want to proof, that a program satisfies certain properties

Origins



Based on the amazing "Tutorial on Static Inference of Numeric Invariants by Abstract Interpretation" by Miné [Min17], <https://www.di.ens.fr/~cousot/AI/>, and [Bal+18; GR22]

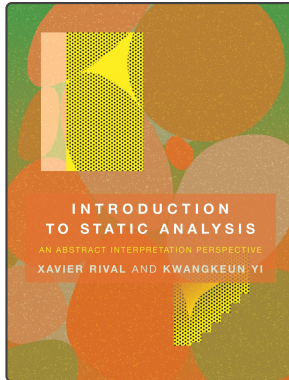
Origins



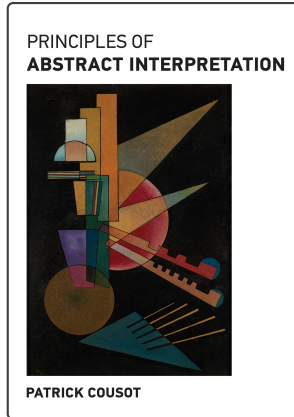
Based on the amazing "Tutorial on Static Inference of Numeric Invariants by Abstract Interpretation" by Miné [Min17], <https://www.di.ens.fr/~cousot/AI/>, and [Bal+18; GR22]

Recommended Resources

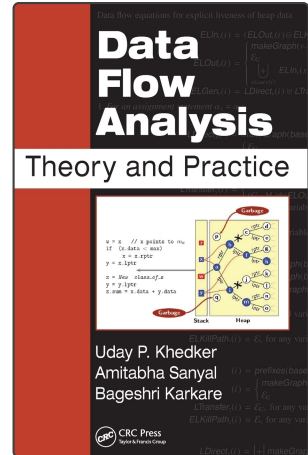
Using Analyses [RY20]



Formal Foundations [Cou21]



Dataflow Perspective [KSK09]



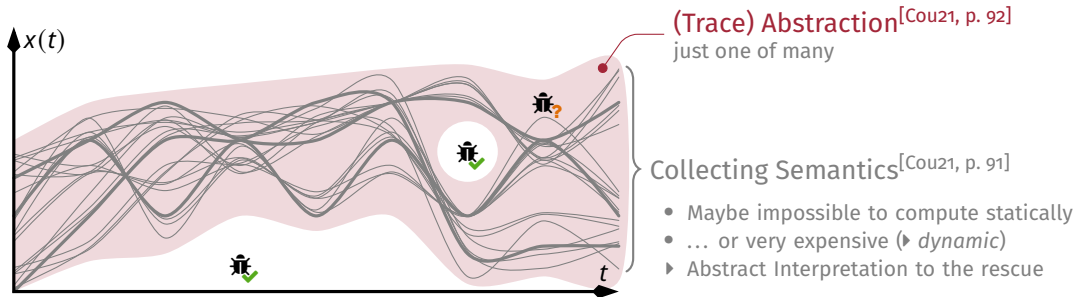
And for an overview: "Tutorial on Static Inference of Numeric Invariants by Abstract Interpretation" [Min17]

2. The How

Abstract Interpretation

- ```
public static void main(String[] args) {
 int a = 1; { a ∈ {1} }
 double r = Math.random() * 10; { r ∈ [0..10) }
 if (r > 5) {
 a = 2; { a ∈ {2} }
 }
 System.out.println(a); { a ∈ {1,2} }
}
```
- We want to proof interesting properties of programs
  - *Dataflow Properties*  
Liveness, Fainting, Reaching Definitions, ...
  - *Safety Properties*  
No Null Dereference, No Division by Zero, ...
  - *Numerical Properties*  
Signs, Intervals, Octagons, Polyhedra, ...
  - ...

# Abstract Interpretation



# Terminology

- **Property** — Set of states/traces that satisfy that property

Even integers:  $P = \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z} : z = 2k\} = \{0, 2, 4, 6, \dots\} \subseteq \mathcal{P}(\mathbb{Z})$

universe ( $\mathbb{U}$ )

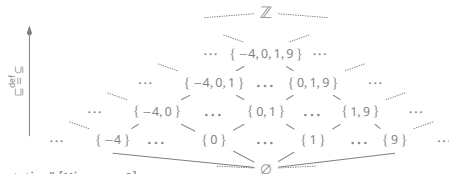


$$\forall x, y, z \in X : x \sqsubseteq y \wedge y \sqsubseteq z \implies x \sqsubseteq z$$

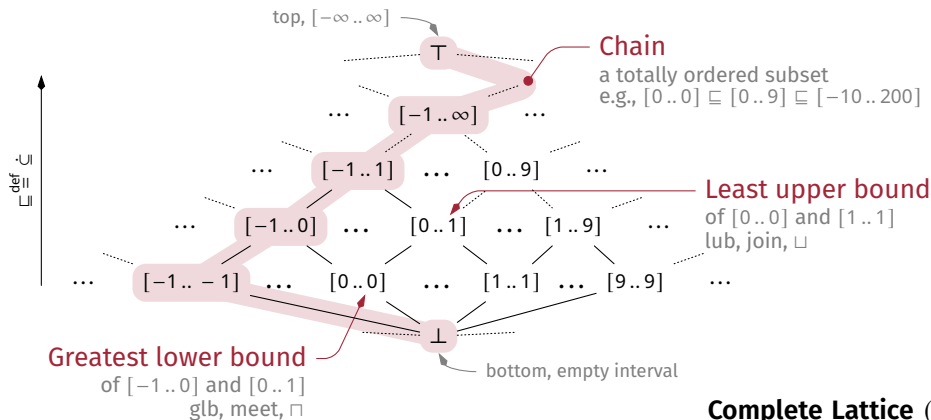
$$\forall x \in X : x \sqsubseteq x \quad \downarrow \quad \forall x, y \in X : x \sqsubseteq y \wedge y \sqsubseteq x \implies x = y$$

- **Partial Order** — A reflexive, transitive, antisymmetric relation on a set

$(\mathbb{Z}, \leq), (\mathcal{P}(\mathbb{Z}), \subseteq), \dots$



# Chains and Lattices

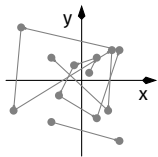


**Complete Lattice**  $(X, \subseteq, \cup, \cap, \perp, \top)$

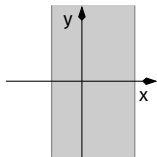
- $(X, \subseteq)$  is a partial order
- $\forall A \subseteq X : \cup A$  and  $\cap A$  exist
- $\perp / \top$  as smallest/largest element

# Abstract Domains

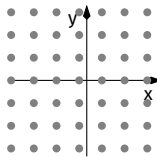
# Numerical



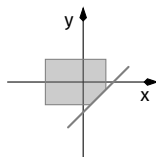
Collecting Semantics



Intervals  $x \in [a..b]$



Simple Congruences



Pentagons

- Octagons

- Ellipses

- Exponentials

- Signs

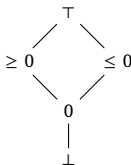
# Sign Analysis

## Simple Sign Domain

- We still have no program semantics, but we can try...

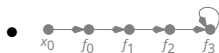
|                             |                                          |
|-----------------------------|------------------------------------------|
| <code>int a = 0;</code>     | $\{ a = 0 \}$                            |
| <code>int b = 12;</code>    | $\{ b \geq 0 \}$                         |
| <code>int c = a + b;</code> | $\{ c \geq 0 \quad (= 0 + \geq 0) \}$    |
| <code>int d = c - b;</code> | $\{ d = \top \quad (\geq 0 - \geq 0) \}$ |

- But how to handle control flow? Loops? Recursion?



# Fixpoints

- For operators  $f : X \rightarrow X$  a **fixpoint** is a  $x \in X$  such that  $f(x) = x$
- If we iterate  $f$  starting from some  $x_0 \in X$ :



reach a fixpoint,  $f^p = f(f^p)$



reach a cycle,  $f^{p+\ell} = f^p, \ell > 0$



iterate forever,  $\forall p \neq q \in \mathbb{N} : f^p \neq f^q$

$f : \mathbb{N} \rightarrow \mathbb{N}, f(x) = x + 1$

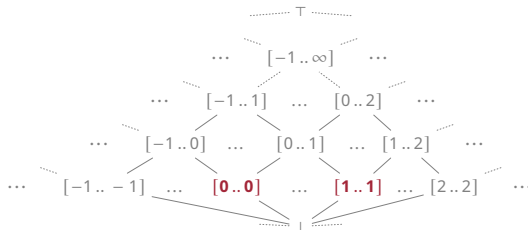
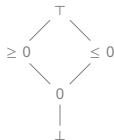
- If our function is monotonic, we can always find a fixpoint<sup>[Tar55]</sup>

for complete, nonempty lattices  
Tarski's Theorem

- Analyzing, e.g. loops, we “go up” the lattice until we reach a least fixpoint

# Interval Analysis, I

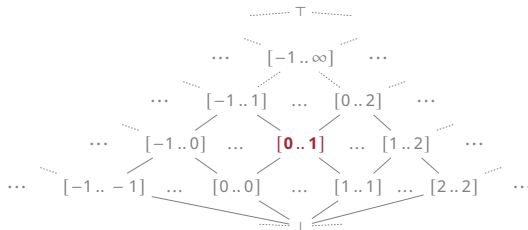
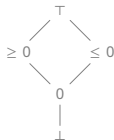
```
int x = 0;
while(x < 2) {
 x = x + 1;
}
```

$$\begin{aligned} &\{ x_0 \in [0..0] \} \\ &\{ [\text{pre}] x_1 \in [0..0] \} \\ &\{ [\text{in}] x_2 \in [0..0] \quad ([0..0] \cap (-\infty..1]) \} \\ &\{ x_3 \in [1..1] \quad ([0..0] \oplus [1..1]) \} \end{aligned}$$




# Interval Analysis, I

```
int x = 0;
while(x < 2) {
 x = x + 1;
}
```

$$\begin{aligned} & \{ x_0 \in [0..0] \} \\ & \{ \text{[pre]} \ x_1 \in [0..1] \quad ([0..0] \cup [1..1]) \} \\ & \{ \text{[in]} \ x_2 \in [0..1] \quad ([0..1] \cap (-\infty..1]) \} \\ & \{ x_3 \in [1..2] \quad ([0..1] \oplus [1..1]) \} \end{aligned}$$


# Interval Analysis, I

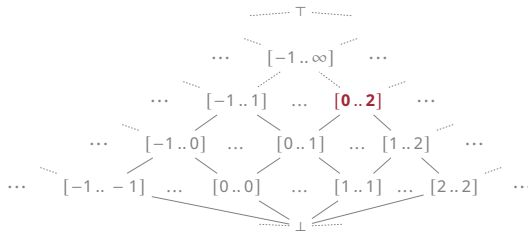
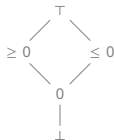
```
int x = 0;
while(x < 2) {
 x = x + 1;
}
```

Intervals

$\{ x_0 \in [0..0] \}$   
 $\{ [\text{pre}] x_1 \in [0..2] \quad ([0..1] \cup [1..2]) \}$   
 $\{ [\text{in}] x_2 \in \}$   
 $\{ x_3 \in \}$   
 $\{ [\text{post}] x_4 \in [2..2] \quad ([0..2] \cap [2..\infty)) \}$

Signs

$\{ x_0 = 0 \}$   
 $\{ [\text{pre}] x_1 \geq 0 \}$   
 $\{ [\text{in}] x_2 \geq 0 \}$   
 $\{ x_3 \geq 0 \}$   
 $\{ [\text{post}] x_4 \geq 0 \}$



# 3. Semantics

# Semantics

# Program Syntax (simplified)

Variable  $v \in \mathbb{V}$   
Assignment  
Sequence  
Numeric Constant  $c \in \mathbb{I}$   
Loop  
Comparison  $\bowtie \in \{\leq, <, \dots\}$   
Binary Expression

```
int x = 0;
while(x < 2) {
 x = x + 1;
}
```

|             |       |                                             |                                                 |
|-------------|-------|---------------------------------------------|-------------------------------------------------|
| <i>stm</i>  | $::=$ | $V \leftarrow \text{expr}$                  | (assignment, $V \in \mathbb{V}$ )               |
|             | $ $   | $\text{stm}_1; \text{stm}_2$                | (sequence)                                      |
|             | $ $   | <b>while</b> ( <i>cond</i> ) { <i>stm</i> } | (loop)                                          |
| <i>expr</i> | $::=$ | $V$                                         | (variable, $V \in \mathbb{V}$ )                 |
|             | $ $   | $c$                                         | (constant, $c \in \mathbb{I}$ )                 |
|             | $ $   | $\text{expr}_1 \diamond \text{expr}_2$      | (bin. expr., $\diamond \in \{+, -, \dots\}$ )   |
| <i>cond</i> | $::=$ | $b$                                         | (boolean, $b \in \mathbb{B}$ )                  |
|             | $ $   | $\text{expr}_1 \bowtie \text{expr}_2$       | (comparison, $\bowtie \in \{\leq, <, \dots\}$ ) |

# Atomic Expression Semantics

```
int x = 0;
while(x < 2) {
 x = x + 1;
}
```

$expr ::= V$  (variable,  $V \in \mathbb{V}$ )  
 $| c$  (constant,  $c \in \mathbb{I}$ )  
 $| expr_1 \diamond expr_2$  (bin. expr.,  $\diamond \in \{+, -, \dots\}$ )

- We use an environment  $\mathcal{E} \stackrel{\text{def}}{=} \mathbb{V} \rightarrow \mathbb{I}$  to represent the current program state  
 Usually written as  $\mathbb{E} \llbracket expr \rrbracket \rho$
- Now we can define  $\text{evalExpr}(expr, env)$  for an environment  $env \in \mathcal{E}$

| $\mathbb{V}$ | $\mathbb{I}$ |
|--------------|--------------|
| x            | 0            |
| c            | 5            |

$\text{evalExpr}(V, env) \stackrel{\text{def}}{=} env(V)$  ← Value of  $V \in \mathbb{V}$  in Environment  $env$   
 $\text{evalExpr}(c, env) \stackrel{\text{def}}{=} c$   
 $\text{evalExpr}(expr_1 + expr_2, env) \stackrel{\text{def}}{=} \text{evalExpr}(expr_1, env) + \text{evalExpr}(expr_2, env)$   
 $\vdots$

- Additionally we can define  $\text{evalCond}(cond, envs)$  and  $\text{evalStm}(stm, envs)$

$\mathbb{C} \llbracket cond \rrbracket \mathcal{D}$

$\mathbb{S} \llbracket stm \rrbracket \mathcal{D}$

# Denotational Semantics

## while loops

Suppose we start the loop with states *Start*

$$\text{while}(\text{cond}) \{ \text{stm} \} \quad \begin{array}{c} \downarrow \\ \text{iterate to find the least fixpoint [Min17, p. 52]} \end{array} \quad F(X) \stackrel{\text{def}}{=} \text{Start} \cup \text{evalStm}(\text{stm}, \text{evalCond}(\text{cond}, X))$$

Keep only states *S* with  $\text{evalCond}(\neg \text{cond}, S)$

There are alternatives (e.g., equation systems, [Cou21, part 7])

We achieve their abstract counterpart using the same principles but for abstract domains!

Usually written as  $S^\#, C^\#, E^\#, \dots$

# Interval Analysis, II

|                                     |                                                                                                    |
|-------------------------------------|----------------------------------------------------------------------------------------------------|
| <code>int x = 0;</code>             | $\{ x_0 \in [0..0] \}$                                                                             |
| <code>while(x &lt; 999999) {</code> | $\{ [pre] x_1 \in [0..2] \quad ([0..1] \cup [1..2]) \} \quad \nabla \implies x_1 \in [0.. \infty)$ |
| <code>x = x + 1;</code>             | $\{ [in] x_2 \in [0..1] \quad ([0..1] \cap (-\infty..1]) \}$                                       |
| <code>x = x + 1;</code>             | $\{ x_3 \in [1..2] \quad ([0..1] \oplus [1..1]) \}$                                                |
| <code>}</code>                      | $\{ [post] x_4 \in [999999.. \infty) \quad ([0.. \infty) \cap [999999.. \infty)) \}$               |

- Fixpoint iteration can be very expensive, and may not stabilize
- *Widening* ( $\nabla$ ) is crucial, computing an upper bound

# **4. Soundness and Completeness**

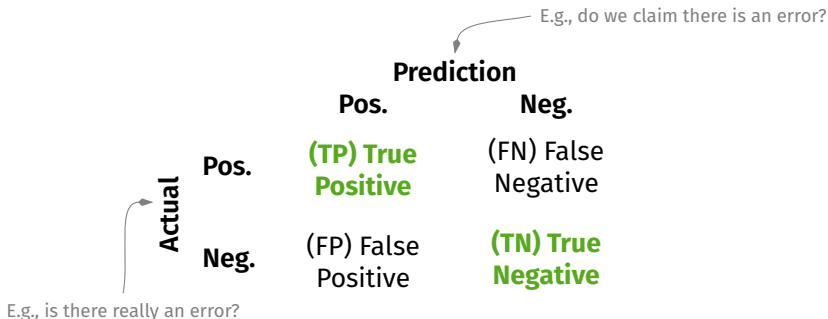


# Rice's Theorem

- We want to prove properties of programs (e.g., no overflow, shapes, ...)
- However, thanks to Rice [Ric53] we know:  
*Rice's theorem states that all nontrivial semantic properties of programs are undecidable. [Cou21, p. 100]*
- We can not solve the halting problem
- We have to approximate the reality



# The Confusion Matrix



The diagram shows a 2x2 confusion matrix. The columns are labeled 'Prediction' with 'Pos.' and 'Neg.' headers. The rows are labeled 'Actual' with 'Pos.' and 'Neg.' headers. The cells contain: (TP) True Positive, (FN) False Negative, (FP) False Positive, and (TN) True Negative. Green text highlights the TP and TN values. An arrow points from the text 'E.g., do we claim there is an error?' to the 'Prediction' header. Another arrow points from the text 'E.g., is there really an error?' to the 'Actual' header.

|        |      | Prediction          |                     |
|--------|------|---------------------|---------------------|
|        |      | Pos.                | Neg.                |
| Actual | Pos. | (TP) True Positive  | (FN) False Negative |
|        | Neg. | (FP) False Positive | (TN) True Negative  |

E.g., do we claim there is an error?

E.g., is there really an error?

- **Precision:**  $\text{TP} / (\text{TP} + \text{FP})$  (“how many false alarms”)
- **Recall:**  $\text{TP} / (\text{TP} + \text{FN})$  (“how many errors did we find”)

# Soundness and Completeness

## Soundness

- All properties we derive are true (but we may miss some)
- If we report bugs for violated properties, we produce no false negative

## Completeness

- We are able to infer all interesting properties in the program
- If we report bugs for violated properties, we produce no false positive

Abstract interpretation soundly over-approximates the program semantics

# 5. Outlook

# Outlook

- Domain transformers  
combine abstract domains<sup>[Min17, p. 149]</sup>
- Galois connections  
define the relationship between concrete and abstract domains<sup>[Cou21, p. 110]</sup>
- Corresponding to widening, narrowing  
refines approximations<sup>[Cou21, p. 395]</sup>
- Function calls  
require special handling<sup>[MJ12]</sup>
- Existing libraries allow for easy implementation  
LiSA<sup>[Fer+21]</sup>, MOPSA<sup>[Jou+19]</sup>, Apron<sup>[JM09]</sup>

# References I

- [Bal+18] Roberto Baldoni et al. "A Survey of Symbolic Execution Techniques". In: *ACM Comput. Surv.* 51.3 (2018), 50:1–50:39. DOI: 10.1145/3182657. URL: <https://doi.org/10.1145/3182657>.
- [BCo4] Yves Bertot and Pierre Castéran. *Interactive Theorem Proving and Program Development - Coq'Art: The Calculus of Inductive Constructions*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2004. ISBN: 978-3-642-05880-6. DOI: 10.1007/978-3-662-07964-5. URL: <https://doi.org/10.1007/978-3-662-07964-5>.
- [BEL75] Robert S. Boyer, Bernard Elspas, and Karl N. Levitt. "SELECT - a formal system for testing and debugging programs by symbolic execution". In: *Proceedings of the International Conference on Reliable Software 1975, Los Angeles, California, USA, April 21-23, 1975*. Ed. by Martin L. Shooman and Raymond T. Yeh. ACM, 1975, pp. 234–245. DOI: 10.1145/800027.808445. URL: <https://doi.org/10.1145/800027.808445>.
- [Bir67] Garrett Birkhoff. "Lattice theory". In: *Publications of AMS* (1967).
- [CC77] Patrick Cousot and Radhia Cousot. "Abstract Interpretation: A Unified Lattice Model for Static Analysis of Programs by Construction or Approximation of Fixpoints". In: *Conference Record of the Fourth ACM Symposium on Principles of Programming Languages, Los Angeles, California, USA, January 1977*. Ed. by Robert M. Graham, Michael A. Harrison, and Ravi Sethi. ACM, 1977, pp. 238–252. DOI: 10.1145/512950.512973. URL: <https://doi.org/10.1145/512950.512973>.
- [CDE08] Cristian Cadar, Daniel Dunbar, and Dawson R. Engler. "KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs". In: *8th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2008, December 8-10, 2008, San Diego, California, USA, Proceedings*. Ed. by Richard Draves and Robbert van Renesse. USENIX Association, 2008, pp. 209–224. URL: [http://www.usenix.org/events/osdi08/tech/full%5C\\_papers/cadar/cadar.pdf](http://www.usenix.org/events/osdi08/tech/full%5C_papers/cadar/cadar.pdf).
- [CES86] Edmund M. Clarke, E. Allen Emerson, and A. Prasad Sistla. "Automatic Verification of Finite-State Concurrent Systems Using Temporal Logic Specifications". In: *ACM Trans. Program. Lang. Syst.* 8.2 (1986), pp. 244–263. DOI: 10.1145/5397.5399. URL: <https://doi.org/10.1145/5397.5399>.
- [CKLo4] Edmund Clarke, Daniel Kroening, and Flavio Lerda. "A tool for checking ANSI-C programs". In: *Tools and Algorithms for the Construction and Analysis of Systems: 10th International Conference, TACAS 2004, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004, Barcelona, Spain, March 29-April 2, 2004. Proceedings 10*. Springer, 2004, pp. 168–176.
- [Cou12] Patrick Cousot. "A casual introduction to Abstract Interpretation". In: *CMACS Workshop on Systems Biology and Formals Methods (SBFM'12)* (2012). URL: <https://pcousot.github.io/talks/PCousot-SBFM-2012-1.pdf> (visited on 12/09/2024).
- [Cou21] Patrick Cousot. "Principles of Abstract Interpretation". In: (2021).

# References II

- [CZ11] Agostino Cortesi and Matteo Zanioli. "Widening and narrowing operators for abstract interpretation". In: *Comput. Lang. Syst. Struct.* 37.1 (2011), pp. 24–42. doi: 10.1016/J.CL.2010.09.001. URL: <https://doi.org/10.1016/j.cl.2010.09.001>.
- [Fer+21] Pietro Ferrara et al. "Static analysis for dummies: experiencing LiSA". In: *SOAP@PLDI 2021: Proceedings of the 10th ACM SIGPLAN International Workshop on the State Of the Art in Program Analysis, Virtual Event, Canada, 22 June, 2021*. Ed. by Lisa Nguyen Quang Do and Caterina Urban. ACM, 2021, pp. 1–6. doi: 10.1145/3460946.3464316. URL: <https://doi.org/10.1145/3460946.3464316>.
- [Flo67] Robert W. Floyd. "Assigning Meanings to Programs". In: *Proc. of the American Mathematical Society Symposia on Applied Mathematics*. Vol. 19. 1967, pp. 19–32.
- [GR22] Roberto Giacobazzi and Francesco Ranzato. "History of Abstract Interpretation". In: *IEEE Ann. Hist. Comput.* 44.2 (2022), pp. 33–43. doi: 10.1109/MAHC.2021.3133136. URL: <https://doi.org/10.1109/MAHC.2021.3133136>.
- [Hoa69] C. A. R. Hoare. "An Axiomatic Basis for Computer Programming". In: *Commun. ACM* 12.10 (1969), pp. 576–580. doi: 10.1145/363235.363259. URL: <https://doi.org/10.1145/363235.363259>.
- [JM09] Bertrand Jeannet and Antoine Miné. "Apron: A Library of Numerical Abstract Domains for Static Analysis". In: *Computer Aided Verification, 21st International Conference, CAV 2009, Grenoble, France, June 26 - July 2, 2009. Proceedings*. Ed. by Ahmed Bouajjani and Oded Maler. Vol. 5643. Lecture Notes in Computer Science. Springer, 2009, pp. 661–667. doi: 10.1007/978-3-642-02658-4\_52. URL: [https://doi.org/10.1007/978-3-642-02658-4\\_52](https://doi.org/10.1007/978-3-642-02658-4_52).
- [Jou+19] Matthieu Jounault et al. "Combinations of Reusable Abstract Domains for a Multilingual Static Analyzer". In: *Verified Software. Theories, Tools, and Experiments - 11th International Conference, VSTTE 2019, New York City, NY, USA, July 13-14, 2019, Revised Selected Papers*. Ed. by Supratik Chakraborty and Jorge A. Navas. Vol. 12031. Lecture Notes in Computer Science. Springer, 2019, pp. 1–18. doi: 10.1007/978-3-030-41600-3\_1. URL: [https://doi.org/10.1007/978-3-030-41600-3\\_1](https://doi.org/10.1007/978-3-030-41600-3_1).
- [Kin74] James C. King. "A New Approach to Program Testing". In: *Programming Methodology, 4th Informatik Symposium, IBM Germany, Wildbad, September 25-27, 1974*. Ed. by Clemens Hackl. Vol. 23. Lecture Notes in Computer Science. Springer, 1974, pp. 278–290. doi: 10.1007/3-540-07131-8\_30. URL: [https://doi.org/10.1007/3-540-07131-8\\_30](https://doi.org/10.1007/3-540-07131-8_30).
- [Kle52] Stephen Cole Kleene. "Introduction to metamathematics". In: (1952).
- [KSK09] Uday Khedker, Amitabha Sanyal, and Bageshri Karkare. *Data Flow Analysis: Theory and Practice*. 1st. USA: CRC Press, Inc., 2009. ISBN: 0849328802.
- [LF08] Francesco Logozzo and Manuel Fähndrich. "Pentagons: a weakly relational abstract domain for the efficient validation of array accesses". In: *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Fortaleza, Ceara, Brazil, March 16-20, 2008*. Ed. by Roger L. Wainwright and Hisham Haddad. ACM, 2008, pp. 184–188. doi: 10.1145/1363686.1363736. URL: <https://doi.org/10.1145/1363686.1363736>.

# References III

- [Mau04] Laurent Mauborgne. “Astrée: verification of absence of run-time error”. In: *Building the Information Society, IFIP 18th World Computer Congress, Topical Sessions, 22-27 August 2004, Toulouse, France*. Ed. by René Jacquart. Vol. 156. IFIP. Kluwer/Springer, 2004, pp. 385–392. DOI: 10.1007/978-1-4020-8157-6\_30. URL: [https://doi.org/10.1007/978-1-4020-8157-6%5C\\_30](https://doi.org/10.1007/978-1-4020-8157-6%5C_30).
- [Min17] Antoine Miné. “Tutorial on Static Inference of Numeric Invariants by Abstract Interpretation”. In: *Found. Trends Program. Lang.* 4:3-4 (2017), pp. 120–372. DOI: 10.1561/25000000034. URL: <https://doi.org/10.1561/25000000034>.
- [MJ12] Jan Midtgaard and Thomas P. Jensen. “Control-flow analysis of function calls and returns by abstract interpretation”. In: *Inf. Comput.* 211 (2012), pp. 49–76. DOI: 10.1016/J.IC.2011.11.005. URL: <https://doi.org/10.1016/j.ic.2011.11.005>.
- [ORS92] Sam Owre, John M. Rushby, and Natarajan Shankar. “PVS: A Prototype Verification System”. In: *Automated Deduction - CADE-11, 11th International Conference on Automated Deduction, Saratoga Springs, NY, USA, June 15-18, 1992, Proceedings*. Ed. by Deepak Kapur. Vol. 607. Lecture Notes in Computer Science. Springer, 1992, pp. 748–752. DOI: 10.1007/3-540-55602-8\_217. URL: [https://doi.org/10.1007/3-540-55602-8%5C\\_217](https://doi.org/10.1007/3-540-55602-8%5C_217).
- [Ric53] Henry Gordon Rice. “Classes of recursively enumerable sets and their decision problems”. In: *Transactions of the American Mathematical society* 74.2 (1953), pp. 358–366.
- [RY20] Xavier Rival and Kwangkeun Yi. “Introduction to Static Analysis: An Abstract Interpretation Perspective”. In: (2020).
- [Tar55] Alfred Tarski. “A lattice-theoretical fixpoint theorem and its applications.”. In: (1955).
- [Tur49] Alan Turing. “Checking a large routine”. In: *Report of a Conference on High Speed Automatic Calculating Machines*. 1949, pp. 67–69.