

ABSTRACT INTERPRETATION: THEORY AND APPLICATIONS

P. COUSOT

Patrick.Cousot@ens.fr <http://www.di.ens.fr/~cousot>

Second International Summer School in Computational Logic, ISCL 2002

25th—30th August 2002, Acquafredda di Maratea (Basilicata, Italy)

© P. COUSOT, ALL RIGHTS RESERVED.

A Potpourri of Applications of Abstract Interpretation



Content of the Potpourri of Applications of Abstract Interpretation

1. Syntax	47
2. Semantics	51
3. Typing	62
4. Model Checking	76
5. Program Transformations	89
6. Static Program Analysis	93

Application to Typing

- P. Cousot, *Types as Abstract Interpretations*, ACM 24th POPL, 1997, pp. 316-331.



Syntax of the Eager Lambda Calculus

$x, f, \dots \in \mathbb{X}$:	variables
$e \in \mathbb{E}$:	expressions
$e ::= x$		variable
$\lambda x \cdot e$		abstraction
$e_1(e_2)$		application
$\mu f \cdot \lambda x \cdot e$		recursion
$\mathbf{1}$		one
$e_1 - e_2$		difference
$(e_1 ? e_2 : e_3)$		conditional

Semantic Domains

Ω	wrong/runtime error value
\perp	non-termination
$\mathbb{W} \stackrel{\text{def}}{=} \{\Omega\}$	wrong
$z \in \mathbb{Z}$	integers
$u, f, \varphi \in \mathbb{U} \cong \mathbb{W}_{\perp} \oplus \mathbb{Z}_{\perp} \oplus [\mathbb{U} \mapsto \mathbb{U}]_{\perp}^{19}$	values
$R \in \mathbb{R} \stackrel{\text{def}}{=} \mathbb{X} \mapsto \mathbb{U}$	environments
$\phi \in \mathbb{S} \stackrel{\text{def}}{=} \mathbb{R} \mapsto \mathbb{U}$	semantic domain

¹⁹ $[\mathbb{U} \mapsto \mathbb{U}]$: continuous, \perp -strict, Ω -strict functions from values \mathbb{U} to values \mathbb{U} .

Denotational Semantics with Run-Time Type Checking

$$\mathbf{S}[1]R \stackrel{\text{def}}{=} 1$$

$$\mathbf{S}[e_1 - e_2]R \stackrel{\text{def}}{=} (\mathbf{S}[e_1]R = \perp \vee \mathbf{S}[e_2]R = \perp ? \perp \\ | \mathbf{S}[e_1]R = z_1 \wedge \mathbf{S}[e_2]R = z_2 ? z_1 - z_2 \\ | \Omega)$$

$$\mathbf{S}[(e_1 ? e_2 : e_3)]R \stackrel{\text{def}}{=} (\mathbf{S}[e_1]R = \perp ? \perp \\ | \mathbf{S}[e_1]R = 0 ? \mathbf{S}[e_2]R \\ | \mathbf{S}[e_1]R = z \neq 0 ? \mathbf{S}[e_3]R \\ | \Omega)$$

$$\mathbf{S}[x]R \stackrel{\text{def}}{=} R(x)$$

$$\mathbf{S}[\lambda x \cdot e]R \stackrel{\text{def}}{=} \lambda u \cdot (u = \perp ? \perp \\ | u = \Omega ? \Omega \\ | \mathbf{S}[e]R[x \leftarrow u])$$

$$\mathbf{S}[e_1(e_2)]R \stackrel{\text{def}}{=} (\mathbf{S}[e_1]R = \perp \vee \mathbf{S}[e_2]R = \perp ? \perp \\ | \mathbf{S}[e_1]R = f \in [\mathbb{U} \mapsto \mathbb{U}] ? f(\mathbf{S}[e_2]R) \\ | \Omega)$$

$$\mathbf{S}[\mu f \cdot \lambda x \cdot e]R \stackrel{\text{def}}{=} \text{lfp}^{\sqsubseteq} \lambda \varphi \cdot \mathbf{S}[\lambda x \cdot e]R[f \leftarrow \varphi]$$

Standard Denotational & Collecting Semantics

- The denotational semantics is:

$$\mathbf{S}[\bullet] \in \mathbf{E} \mapsto \mathbf{S}$$

- A concrete property P of a program is a set of possible program behaviors:

$$P \in \mathbb{P} \stackrel{\text{def}}{=} \wp(\mathbf{S})$$

- The standard collecting semantics is the strongest concrete property:

$$\mathbf{C}[\bullet] \in \mathbf{E} \mapsto \mathbb{P} \quad \mathbf{C}[e] \stackrel{\text{def}}{=} \{\mathbf{S}[e]\}$$



Church/Curry Monotypes

- Simple types are monomorphic:

$$m \in \mathbb{M}^c, \quad m ::= \text{int} \mid m_1 \rightarrow m_2 \quad \text{monotype}$$

- A type environment associates a type to free program variables:

$$H \in \mathbb{H}^c \stackrel{\text{def}}{=} \mathbb{X} \mapsto \mathbb{M}^c \quad \text{type environment}$$

Church/Curry Monotypes (continued)

- A **typing** $\langle H, m \rangle$ specifies a possible result type m in a given type environment H assigning types to free variables:

$$\theta \in \mathbb{I}^c \stackrel{\text{def}}{=} \mathbb{H}^c \times \mathbb{M}^c \quad \text{typing}$$

- An **abstract property** or **program type** is a set of typings;

$$T \in \mathbb{T}^c \stackrel{\text{def}}{=} \wp(\mathbb{I}^c) \quad \text{program type}$$

Concretization Function

The meaning of types is a program property, as defined by the concretization function γ^c :²⁰

- Monotypes $\gamma_1^c \in \mathbb{M}^c \mapsto \wp(\mathbb{U})$:

$$\gamma_1^c(\text{int}) \stackrel{\text{def}}{=} \mathbb{Z} \cup \{\perp\}$$

$$\gamma_1^c(m_1 \rightarrow m_2) \stackrel{\text{def}}{=} \{\varphi \in [\mathbb{U} \mapsto \mathbb{U}] \mid$$

$$\forall u \in \gamma_1^c(m_1) : \varphi(u) \in \gamma_1^c(m_2)\}$$

$$\cup \{\perp\}$$

²⁰ For short up/down lifting/injection are omitted.



- type environment $\gamma_2^c \in \mathbb{H}^c \mapsto \wp(\mathbb{R})$:

$$\gamma_2^c(H) \stackrel{\text{def}}{=} \{R \in \mathbb{R} \mid \forall \mathbf{x} \in \mathbb{X} : R(\mathbf{x}) \in \gamma_1^c(H(\mathbf{x}))\}$$



- type environment $\gamma_2^c \in \mathbb{H}^c \mapsto \wp(\mathbb{R})$:

$$\gamma_2^c(H) \stackrel{\text{def}}{=} \{R \in \mathbb{R} \mid \forall x \in \mathbb{X} : R(x) \in \gamma_1^c(H(x))\}$$

- typing $\gamma_3^c \in \mathbb{I}^c \mapsto \mathbb{P}$:

$$\gamma_3^c(\langle H, m \rangle) \stackrel{\text{def}}{=} \{\phi \in \mathbb{S} \mid \forall R \in \gamma_2^c(H) : \phi(R) \in \gamma_1^c(m)\}$$

- type environment $\gamma_2^c \in \mathbb{H}^c \mapsto \wp(\mathbb{R})$:

$$\gamma_2^c(H) \stackrel{\text{def}}{=} \{R \in \mathbb{R} \mid \forall x \in \mathbb{X} : R(x) \in \gamma_1^c(H(x))\}$$

- typing $\gamma_3^c \in \mathbb{I}^c \mapsto \mathbb{P}$:

$$\gamma_3^c(\langle H, m \rangle) \stackrel{\text{def}}{=} \{\phi \in \mathbb{S} \mid \forall R \in \gamma_2^c(H) : \phi(R) \in \gamma_1^c(m)\}$$

- program type $\gamma^c \in \mathbb{T}^c \mapsto \mathbb{P}$:

$$\gamma^c(T) \stackrel{\text{def}}{=} \bigcap_{\theta \in T} \gamma_3^c(\theta)$$

$$\gamma^c(\emptyset) \stackrel{\text{def}}{=} \mathbb{S}$$

Program Types

- Galois connection:

$$\langle \mathbb{P}, \subseteq, \emptyset, \mathbb{S}, \cup, \cap \rangle \begin{array}{c} \xleftarrow{\gamma^c} \\ \xrightarrow{\alpha^c} \end{array} \langle \mathbb{T}^c, \supseteq, \mathbb{I}^c, \emptyset, \cap, \cup \rangle$$

Program Types

- Galois connection:

$$\langle \mathbb{P}, \subseteq, \emptyset, \mathcal{S}, \cup, \cap \rangle \begin{array}{c} \xleftarrow{\gamma^c} \\ \xrightarrow{\alpha^c} \end{array} \langle \mathbb{T}^c, \supseteq, \mathbb{I}^c, \emptyset, \cap, \cup \rangle$$

- Types $\mathbf{T}[e]$ of an expression e :

$$\mathbf{T}[e] \subseteq \alpha^c(\mathbf{C}[e]) = \alpha^c(\{\mathbf{S}[e]\})$$

Program Types

- Galois connection:

$$\langle \mathbb{P}, \subseteq, \emptyset, \mathcal{S}, \cup, \cap \rangle \begin{array}{c} \xleftarrow{\gamma^c} \\ \xrightarrow{\alpha^c} \end{array} \langle \mathbb{T}^c, \supseteq, \mathbb{I}^c, \emptyset, \cap, \cup \rangle$$

- Types $\mathbf{T}[e]$ of an expression e :

$$\mathbf{T}[e] \subseteq \alpha^c(\mathbf{C}[e]) = \alpha^c(\{\mathbf{S}[e]\})$$

Typable Programs Cannot Go Wrong

$$\Omega \in \gamma^c(\mathbf{T}[e]) \iff \mathbf{T}[e] = \emptyset$$



Church/Curry Monotype Abstract Semantics

$$\mathbf{T}[x] \stackrel{\text{def}}{=} \{\langle H, H(x) \rangle \mid H \in \mathbb{H}^c\} \quad (\text{VAR})$$

$$\mathbf{T}[\lambda x \cdot e] \stackrel{\text{def}}{=} \{\langle H, m_1 \rightarrow m_2 \rangle \mid \langle H[x \leftarrow m_1], m_2 \rangle \in \mathbf{T}[e]\} \quad (\text{ABS})$$

$$\mathbf{T}[e_1(e_2)] \stackrel{\text{def}}{=} \{\langle H, m_2 \rangle \mid \langle H, m_1 \rightarrow m_2 \rangle \in \mathbf{T}[e_1] \wedge \langle H, m_1 \rangle \in \mathbf{T}[e_2]\} \quad (\text{APP})$$

$$\mathbf{T}[\mathbf{1}] \stackrel{\text{def}}{=} \{\langle H, \text{int} \rangle \mid H \in \mathbb{H}^c\} \quad (\text{CST})$$

$$\mathbf{T}[e_1 - e_2] \stackrel{\text{def}}{=} \{\langle H, \text{int} \rangle \mid \langle H, \text{int} \rangle \in \mathbf{T}[e_1] \cap \mathbf{T}[e_2]\} \quad (\text{DIF})$$

$$\mathbf{T}[(e_1 ? e_2 : e_3)] \stackrel{\text{def}}{=} \{\langle H, m \rangle \mid \langle H, \text{int} \rangle \in \mathbf{T}[e_1] \wedge \langle H, m \rangle \in \mathbf{T}[e_2] \cap \mathbf{T}[e_3]\} \quad (\text{CND})$$

$$\mathbf{T}[\mu f \cdot \lambda x \cdot e] \stackrel{\text{def}}{=} \{\langle H, m \rangle \mid \langle H[f \leftarrow m], m \rangle \in \mathbf{T}[\lambda x \cdot e]\} \quad (\text{REC})^{21}$$

²¹ The abstract fixpoint has been eliminated thanks to fixpoint induction: $\text{lfp}F \sqsubseteq P \Leftrightarrow \exists I : F(I) \sqsubseteq I \wedge I \sqsubseteq P$.



The Herbrand Abstraction to Get Hindley's Unification-Based Type Inference Algorithm

$\langle \emptyset(\text{ground}(T)), \subseteq, \emptyset, \text{ground}(T), \cup, \cap \rangle$

$\begin{array}{c} \xleftarrow{\text{ground}} \\ \xrightarrow{\text{lcg}} \end{array} \langle T/\equiv, \leq, \emptyset, ['a]_{\equiv}, \text{lcg}, \text{gci} \rangle$

where:

- T : set of terms with variables 'a, . . . ,
- lcg : least common generalization,
- ground : set of ground instances,
- \leq : instance preordering,
- gci : greatest common instance.



Application to Model Checking

- P. Cousot & R. Cousot, *Temporal Abstract Interpretation*, ACM 27th POPL, 2000, pp. 12-25.



Model Checking

- 1) Built a **model** M of the computer system;
- 2) **Check** (i.e. prove enumeratively) or **semi-check** (with semi-algorithms) that the model satisfies a specification given (as a set of traces φ) by a (linear) temporal formula: $M \subseteq \varphi$ or $M \cap \varphi \neq \emptyset$.

Objective of Model Checking

- 1) Built a **model** M of the computer system;
 - 2) **Check** (i.e. prove enumeratively) or **semi-check** (with semi-algorithms) that the model satisfies a specification given (as a set of traces φ) by a (linear) temporal formula: $M \subseteq \varphi$ or $M \cap \varphi \neq \emptyset$.
- The **model** and **specification** should be proved to be **correct abstractions** of the computer system (often taken for granted, could be done by abstract interpretation);

Abstractions in Model Checking

Main **abstractions in model checking**:

- **Implicit abstraction**: to informally design the model of reference;
- **Polyhedral abstraction** (with widening): synchronous, real-time & hybrid system verification;
- **Finitary abstraction** (without widening): hardware & protocols verification²²;

²² Abstracting concrete transition systems to abstract transition systems so as to reuse existing model checkers in the abstract.



Model-checking itself is an abstraction

- Universal abstraction:

$$\langle \wp(\Sigma^+ \cup \Sigma^\omega), \supseteq \rangle \begin{array}{c} \xleftarrow{\gamma_M^\forall} \\ \xrightarrow{\alpha_M^\forall} \end{array} \langle \wp(\Sigma), \supseteq \rangle$$

$$\alpha_M^\forall(\Phi) \stackrel{\text{def}}{=} \{s \mid \{\sigma \in M \mid \sigma_0 = s\} \subseteq \Phi\}$$

- Existential abstraction:

$$\langle \wp(\Sigma^+ \cup \Sigma^\omega), \subseteq \rangle \begin{array}{c} \xleftarrow{\gamma_M^\exists} \\ \xrightarrow{\alpha_M^\exists} \end{array} \langle \wp(\Sigma), \subseteq \rangle$$

$$\alpha_M^\exists(\Phi) \stackrel{\text{def}}{=} \{s \mid \{\sigma \in M \mid \sigma_0 = s\} \cap \Phi \neq \emptyset\}$$



Model-checking itself is an abstraction

- Universal abstraction:

$$\langle \wp(\Sigma^+ \cup \Sigma^\omega), \supseteq \rangle \xleftrightarrow[\alpha_M^\forall]{\gamma_M^\forall} \langle \wp(\Sigma), \supseteq \rangle$$

$$\alpha_M^\forall(\Phi) \stackrel{\text{def}}{=} \{s \mid \{\sigma \in M \mid \sigma_0 = s\} \subseteq \Phi\}$$

- Existential abstraction:

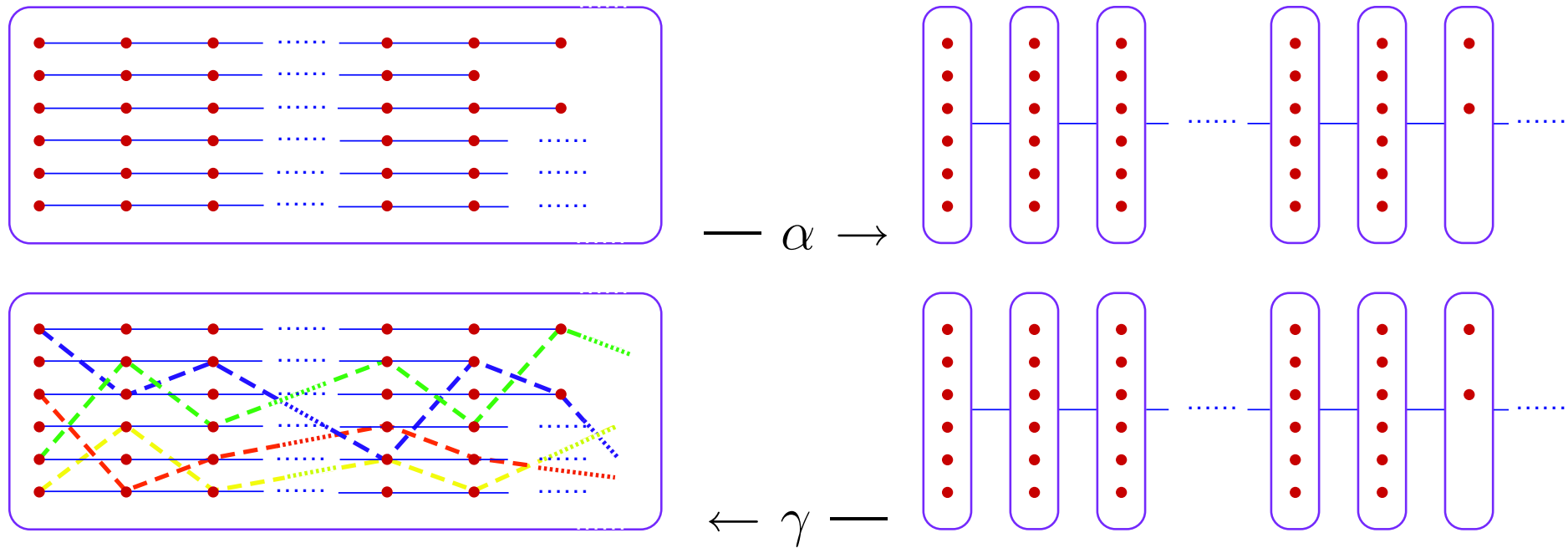
$$\langle \wp(\Sigma^+ \cup \Sigma^\omega), \subseteq \rangle \xleftrightarrow[\alpha_M^\exists]{\gamma_M^\exists} \langle \wp(\Sigma), \subseteq \rangle$$

$$\alpha_M^\exists(\Phi) \stackrel{\text{def}}{=} \{s \mid \{\sigma \in M \mid \sigma_0 = s\} \cap \Phi \neq \emptyset\}$$

These abstractions lead, by fixpoint approximation of the trace semantics, to the classical (finite-state or nonterminating) **model-checking algorithms**.



Implicit Abstraction in Model Checking



Spurious traces: - - -, - - -, - - -, - - -, ... ;

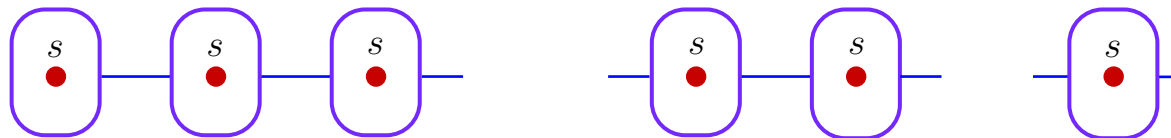
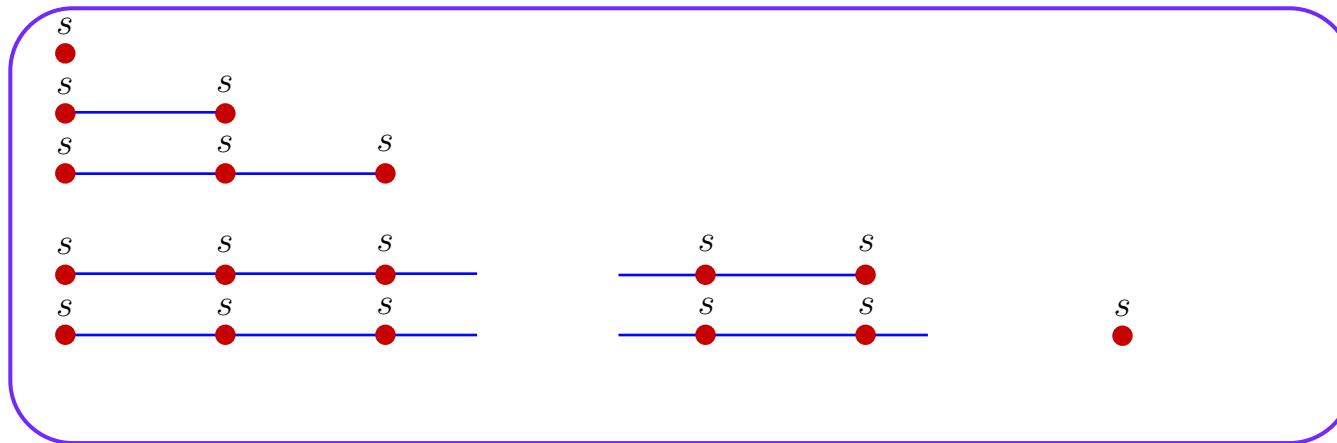
The semantics of the μ -calculus is closed under this abstraction.

Soundness

For a *given class* of properties, **soundness** means that:

Any property (in the *given class*) of the abstract world must hold in the concrete world;

Example for Unsoundness



All abstract traces are infinite but not the concrete ones!

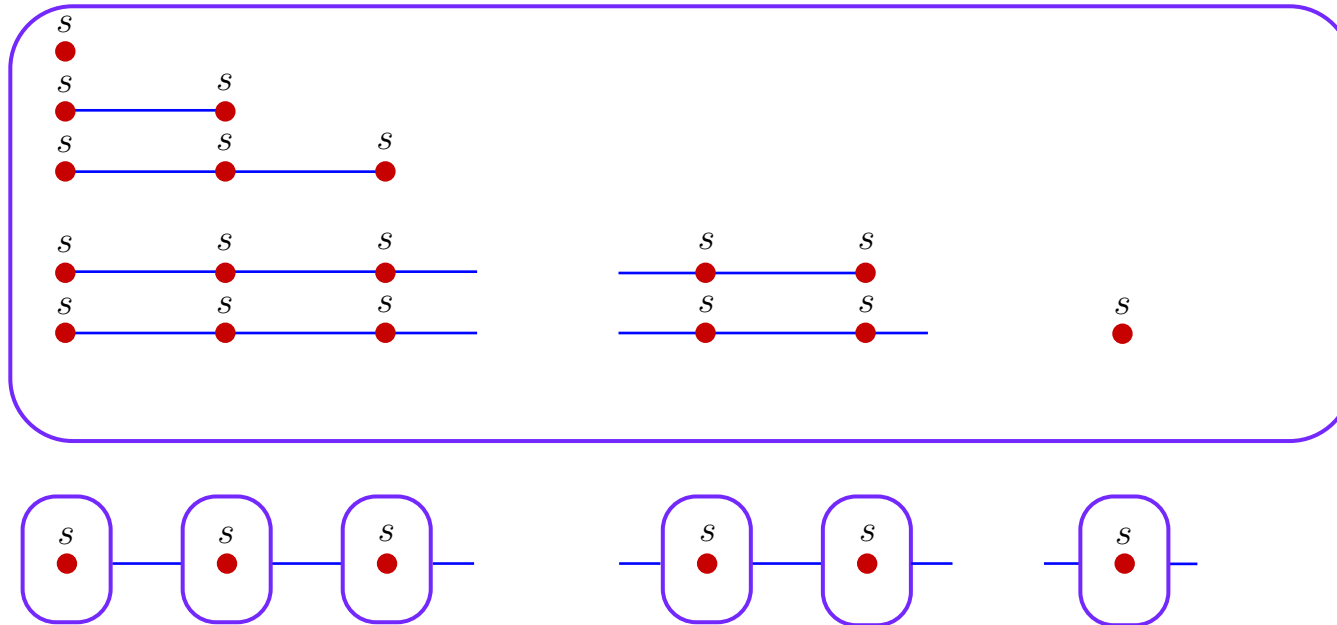


Completeness

For a *given class* of properties, **completeness** means that:

Any property (in the *given class*) of the concrete world must hold in the abstract world;

Example for Incompleteness



All concrete traces are finite but not the abstract ones!

On the Completeness of Model-Checking

- Contrary to program analysis, model checking is **complete**;

On the Completeness of Model-Checking

- Contrary to program analysis, model checking is **complete**;
- Completeness is relative to the **model**, **not** the **program semantics**;



On the Completeness of Model-Checking

- Contrary to program analysis, model checking is **complete**;
- Completeness is relative to the **model**, **not** the **program semantics**;
- Completeness follows from **restrictions** on the models and specifications (e.g. closure under the implicit abstraction);

On the Completeness of Model-Checking

- Contrary to program analysis, model checking is **complete**;
- Completeness is relative to the **model**, **not** the **program semantics**;
- Completeness follows from **restrictions** on the models and specifications (e.g. closure under the implicit abstraction);
- There are models/specifications (such as the μ^* -calculus using **bidirectional traces**) for which:
 - The implicit abstraction is **incomplete** (POPL'00),
 - **Any** abstraction is **incomplete** (Ranzato, ESOP'01).



On the Completeness of Model-Checking

- Contrary to program analysis, model checking is **complete**;
- Completeness is relative to the **model**, **not** the **program semantics**;
- Completeness follows from **restrictions** on the models and specifications (e.g. closure under the implicit abstraction);
- There are models/specifications (such as the μ^* -calculus using **bidirectional traces**) for which:
 - The implicit abstraction is **incomplete** (POPL'00),
 - **Any** abstraction is **incomplete** (Ranzato, ESOP'01).in both cases, even for **finite** transition systems.



Bidirectional Traces

- $\langle i, \sigma \rangle$

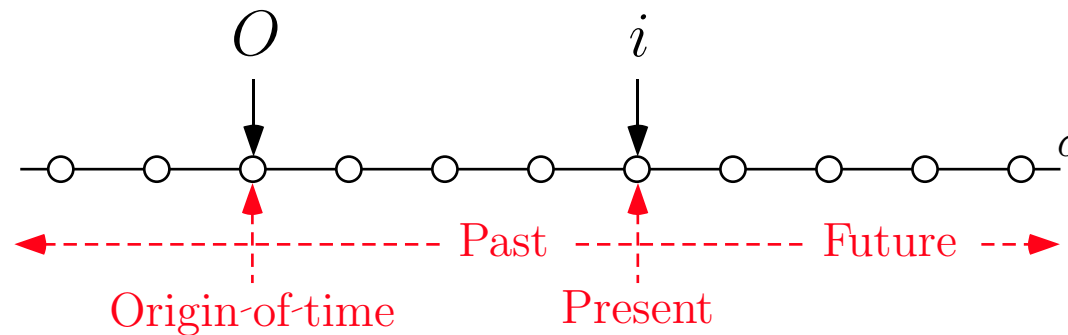
$$\sigma \in \mathbb{Z} \mapsto \Sigma$$

$$i \in \mathbb{Z}$$

bidirectional trace

trace

present time



The reversible μ^* -calculus

$\varphi ::= \sigma_S$ ²³	$\llbracket \sigma_S \rrbracket \rho \stackrel{\text{def}}{=} \{ \langle i, \sigma \rangle \mid \sigma_i \in S \}$
π_t ²⁴	$\llbracket \pi_t \rrbracket \rho \stackrel{\text{def}}{=} \{ \langle i, \sigma \rangle \mid \langle \sigma_i, \sigma_{i+1} \rangle \in t \}$
$\oplus \varphi_1$ ²⁵	$\llbracket \oplus \varphi_1 \rrbracket \rho \stackrel{\text{def}}{=} \{ \langle i, \sigma \rangle \mid \langle i+1, \sigma \rangle \in \llbracket \varphi_1 \rrbracket \rho \}$
$\varphi_1^{\curvearrowright}$	$\llbracket \varphi_1^{\curvearrowright} \rrbracket \rho \stackrel{\text{def}}{=} \{ \langle i, \sigma \rangle \mid \langle -i, \lambda j. \sigma_{-j} \rangle \in \llbracket \varphi_1 \rrbracket \rho \}$
$\varphi_1 \vee \varphi_2$	$\llbracket \varphi_1 \vee \varphi_2 \rrbracket \rho \stackrel{\text{def}}{=} \llbracket \varphi_1 \rrbracket \rho \cup \llbracket \varphi_2 \rrbracket \rho$
$\neg \varphi_1$	$\llbracket \neg \varphi_1 \rrbracket \rho \stackrel{\text{def}}{=} \neg \llbracket \varphi_1 \rrbracket \rho$

²³ $S \in \wp(\Sigma)$.

²⁴ $t \in \wp(\Sigma \times \Sigma)$.

²⁵ \oplus is next time.



The reversible μ^* -calculus (cont'd)

| ...

| X ²⁶ $\llbracket X \rrbracket \rho \stackrel{\text{def}}{=} \rho(X)$

| $\mu X \cdot \varphi_1$ $\llbracket \mu X \cdot \varphi_1 \rrbracket \rho \stackrel{\text{def}}{=} \text{lfp}^{\subseteq} \lambda x \cdot \llbracket \varphi_1 \rrbracket \rho X x$

| $\nu X \cdot \varphi_1$ $\llbracket \nu X \cdot \varphi_1 \rrbracket \rho \stackrel{\text{def}}{=} \text{gfp}^{\subseteq} \lambda x \cdot \llbracket \varphi_1 \rrbracket \rho X x$

| $\forall \varphi_1 : \varphi_2$ ²⁷ $\llbracket \forall \varphi_1 : \varphi_2 \rrbracket \rho \stackrel{\text{def}}{=} \{ \langle i, \sigma \rangle \in \llbracket \varphi_1 \rrbracket \rho \mid \{ \langle i, \sigma' \rangle \in \llbracket \varphi_1 \rrbracket \rho \mid \sigma'_i = \sigma_i \} \subseteq \llbracket \varphi_2 \rrbracket \rho \}$

²⁶ variable.

²⁷ The traces of φ_1 such that all traces of φ_1 with same present state satisfy φ_2 .