

Static Analysis Design Framework: Abstract Interpretation

Kwangkeun Yi

Seoul National University, Korea
<http://ropas.snu.ac.kr/~kwang>

2/26/2012 – 3/2/2012

17th Estonian Winter School in Computer Science, Palmse,
Estonia

Abstract Interpretation

A powerful framework for designing correct static analysis

- “framework”: correct static analysis comes out, reusable
- “powerful”: all static analyses are understood in this framework
- “simple”: prescription is simple
- “eye-opening”: any static analysis is an abstract interpretation

Why Abstraction?

- without abstraction, can't capture all possible executions
- without abstraction, can't terminate

Abstraction is not omission

- reality: $\{2, 4, 6, 8, \dots\}$
- “even number” (abstraction) vs “multiple of 4” (omission)

Abstract Interpretation Framework

real execution	$\llbracket C \rrbracket = \text{fix } F \in D$
abstract execution	$\llbracket \hat{C} \rrbracket = \lim_{i \in \mathbb{N}} \hat{F}^i(\perp_{\hat{D}}) \in \hat{D}$
correctness	$\llbracket C \rrbracket \approx \llbracket \hat{C} \rrbracket$
implementation	computation of $\llbracket \hat{C} \rrbracket$

The framework requires:

- a relation between D and \hat{D}
- a relation between $F \in D \rightarrow D$ and $\hat{F} \in \hat{D} \rightarrow \hat{D}$

The framework guarantees:

- correctness: $\llbracket C \rrbracket \approx \llbracket \hat{C} \rrbracket$
- implementation: computation of $\llbracket \hat{C} \rrbracket$
- freedom: any such \hat{F} and \hat{D} are fine

Static Analysis Design: step 1

Define the input program's real executions(concrete semantics)

- Define semantic domain CPO D .
- Define the real executions as the least fixed point $fix F$ of continuous function $F \in D \rightarrow D$

$$fix F = \bigsqcup_{i \in \mathbb{N}} F^i(\perp_D)$$

Plan: define an abstraction that captures $fix F$

Define the input program's abstract semantics

- Define abstract domain CPO \hat{D} .
 - Establish a Galois connection between D and \hat{D}
- Define an abstract semantic function $\hat{F} \in \hat{D} \rightarrow \hat{D}$
 - \hat{F} must be monotonic or extensive

Plan: define an abstraction that captures $fix F$ by using \hat{F}

Requirement 1: about \hat{D} in relation with D

\hat{D} must be Galois-connected with D

$$D \underset{\alpha}{\overset{\gamma}{\rightleftarrows}} \hat{D}.$$

- Galois connection:

$$\forall x \in D, \hat{x} \in \hat{D} : \alpha(x) \sqsubseteq \hat{x} \iff x \sqsubseteq \gamma(\hat{x}).$$

- Galois connection captures our intention:
 - bigger elements in \hat{D} means more.
 - α abstracts .
 - γ concretizes.

Plan: static analysis is computing an upper bound of $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$

Requirement 2: about \hat{F}

- \hat{F} must be monotonic:

$$\forall x, y \in \hat{D} : x \sqsubseteq y \Rightarrow \hat{F}(x) \sqsubseteq \hat{F}(y)$$

or extensive:

$$\forall x \in \hat{D} : x \sqsubseteq \hat{F}(x).$$

Plan: static analysis is computing an upper bound of $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$

Requirement 3: \hat{F} in relation with F

- Concrete semantic ftn F and its abstract version \hat{F} must satisfy

$$\alpha \circ F \sqsubseteq \hat{F} \circ \alpha, \quad \text{i.e.,} \quad F \circ \gamma \sqsubseteq \gamma \circ \hat{F}$$

or,

- Concrete semantic ftn F and its abstract version \hat{F} must satisfy

$$\alpha(f) \sqsubseteq \hat{f} \Rightarrow \alpha(F f) \sqsubseteq \hat{F} \hat{f}$$

Plan: static analysis is computing an upper bound of $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$

Then: a Correct Static Analysis

static analysis = computing an upper bound of $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$.

- Such an upper bound $\hat{\mathcal{A}}$ is correct:

$$\begin{aligned} \alpha(\text{fix } F) &\sqsubseteq \hat{\mathcal{A}}, & \text{that is,} \\ \text{fix } F &\sqsubseteq \gamma \hat{\mathcal{A}} \end{aligned}$$

Theorem[fixpoint-transfer]

- Analysis result $\hat{\mathcal{A}}$ subsumes the real executions $\text{fix } F$

How to Compute an Upper Bound of $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$

- If abstract semantic domain \hat{D} 's height is finite then, we can directly compute

$$\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp}).$$

The computation always terminates.

- Otherwise, we compute a finite chain $\{\hat{X}_i\}_i$ such that

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\perp})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i).$$

Finite chain $\{\hat{X}_i\}_i$ such that

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\perp})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i)$$

- If \hat{F} is monotonic, a chain by an widening operator ∇ :

$$\begin{aligned} \hat{X}_0 &= \hat{\perp} \\ \hat{X}_{i+1} &= \begin{cases} \hat{X}_i & \text{if } \hat{F}(\hat{X}_i) \sqsubseteq \hat{X}_i \\ \hat{X}_i \nabla \hat{F}(\hat{X}_i) & \text{o.w.} \end{cases} \end{aligned}$$

Conditions

- $\forall a, b \in \hat{D} : (a \sqsubseteq a \nabla b) \wedge (b \sqsubseteq a \nabla b)$
- \forall increasing chain $\{a_i\}_i : \text{chain } x_0 = a_0, x_{i+1} = x_i \nabla a_{i+1}$ is finite

Then

- $\{\hat{X}_i\}_i$ is a finite chain.
- Its limit (\hat{X} such that $\hat{F}(\hat{X}) \sqsubseteq \hat{X}$) is correct:

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\perp)) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i).$$

Theorem[widen's safety]

Refining the Widening Result

If \hat{F} is monotonic,

- We can refine the widened result $\hat{\mathcal{A}} \stackrel{\text{let}}{=} \lim_{i \in \mathbb{N}} (\hat{X}_i)$ by a narrowing operator Δ .
- Compute chain $\{\hat{Y}_i\}_i$

$$\begin{aligned}\hat{Y}_0 &= \hat{\mathcal{A}} \\ \hat{Y}_{i+1} &= \hat{Y}_i \Delta \hat{F}(\hat{Y}_i)\end{aligned}$$

Conditions on Narrowing \triangle

Conditions

- $\forall a, b \in \hat{D} : a \sqsupseteq b \Rightarrow a \sqsupseteq (a \triangle b) \sqsupseteq b$
- \forall decreasing chain $\{a_i\}_i : \text{chain } y_0 = a_0, y_{i+1} = y_i \triangle a_{i+1}$ is finite

Then

- $\{\hat{Y}_i\}_i$ is a finite chain.
- Its limit $\lim_{i \in \mathbb{N}}(\hat{Y}_i)$ is still correct:

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\perp})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{Y}_i).$$

Theorem[narrow's safety]

Why Above Prescription Is Correct? (1/2)

Fixpoint Transfer Theorem

Theorem (fixpoint transfer)

Let CPOs D and \hat{D} are Galois-connected. Function $F : D \rightarrow D$ is continuous. $\hat{F} : \hat{D} \rightarrow \hat{D}$ is either monotonic or extensive. Either $\alpha \circ F \sqsubseteq \hat{F} \circ \alpha$ or $\alpha f \sqsubseteq \hat{f}$ implies $\alpha(F f) \sqsubseteq \hat{F} \hat{f}$. Then,

$$\alpha(\text{fix } F) \sqsubseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp}).$$

Why Above Prescription Is Correct? (2/2)

Widening/Narrowing Theorems

Theorem (widen's safety)

Let $\hat{F} : \hat{D} \rightarrow \hat{D}$ be monotonic over CPO \hat{D} . Let widening operator $\nabla : \hat{D} \times \hat{D} \rightarrow \hat{D}$ satisfies the widening conditions. Then the widened chain $\{\hat{X}_i\}_i$ is finite and its limit satisfies $\lim_{i \in \mathbb{N}} \hat{X}_i \sqsupseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$.

Theorem (narrow's safety)

Let $\hat{F} : \hat{D} \rightarrow \hat{D}$ be monotonic over CPO \hat{D} . Let narrowing operator $\Delta : \hat{D} \times \hat{D} \rightarrow \hat{D}$ satisfies the narrowing conditions. If $\hat{F}(\hat{A}) \sqsubseteq \hat{A}$ then the narrowed chain $\{\hat{Y}_i\}_i$ is finite and its limit satisfies $\lim_{i \in \mathbb{N}} \hat{Y}_i \sqsupseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\perp})$.

Abstract Interpretation Example

(or, a Special Abstract Interpretation Framework)

Semantics as Trace

Program C 's semantics $\llbracket C \rrbracket$ is the set of all execution traces

$$\begin{aligned}\llbracket C \rrbracket &\in 2^{\text{Trace}} \\ \tau, \tau_0 \tau_1 \cdots \tau_n &\in \text{Trace} = \text{State}^* \\ \text{State} &= \text{Command} \times \text{Memory} \times \cdots\end{aligned}$$

Side:

$\text{Trace} = \text{State}^\omega$	v.s.	State^*
liveness analysis		safety analysis
prop. after infinite traces		prop. within finite traces

$$2^{Trace} \xleftrightarrow[\alpha]{\gamma} \hat{Trace}$$

α_0 Trace of set of states: sequence of set of states appearing at a given time along at least one of the traces

$$\alpha_0(X) = \lambda i. \{\tau_i \mid \tau \in X, 0 \leq i < |\tau|\} \in \hat{Trace} = \mathbb{N} \xrightarrow{\text{fin}} 2^{State}$$

$\alpha_1 \circ \alpha_0$ Set of reachable states (global invariant): set of states appearing at least once along a trace

$$\alpha_1(Y) = \bigcup \{Y(i) \mid i \in \text{Dom } Y\} \in \hat{Trace} = 2^{State}$$

$\alpha_2 \circ \alpha_1 \circ \alpha_0$ Partitioned set of reachable states (local invariant): e.g., project along each control point $\in \Delta$ (a finite set)

$$\alpha_2(Z) = \lambda c. \{s_i \mid \langle c_i, s_i \rangle \in Z, c_i = c \in \Delta\} \in \hat{Trace} = \Delta \rightarrow 2^{State}$$

$\alpha_3 \circ \alpha_2 \circ \alpha_1 \circ \alpha_0$ Abstracting the partitioned set of reachable states

$$\alpha_3(\Phi) = \lambda c. \alpha(\Phi c) \in \hat{Trace} = \Delta \rightarrow \hat{State}$$

where

$$2^{State} \xleftrightarrow{\alpha} \hat{State}$$

Trace Abstract Interpretation's Correctness Condition

$$\text{fix}(F \stackrel{\text{let}}{=} \lambda T. T_0 \sqcup \text{Next } T) \quad \text{and} \quad \text{fix}(\hat{F} \stackrel{\text{let}}{=} \lambda \hat{T}. \alpha(T_0) \sqcup \text{Next } \hat{T})$$

where

$$F \in 2^{\text{Trace}} \rightarrow 2^{\text{Trace}} \quad \text{and} \quad \hat{F} \in \text{Trace} \rightarrow \text{Trace}.$$

To show is $\alpha(\text{fix } F) \sqsubseteq \text{fix } \hat{F}$, i.e., $\alpha \circ F \sqsubseteq \hat{F} \circ \alpha$.

A sufficient condition, if Trace and Trace are \sqcup -closed, is:

$$\alpha \circ \text{Next} \sqsubseteq \text{Next} \circ \alpha.$$

(easy to see, by Galois-connection.)

A Sufficient Condition for $\alpha \circ Next \sqsubseteq \widehat{Next} \circ \alpha$ (1/4)

Focus on:

$$2^{State} \xrightleftharpoons[\alpha]{\gamma} (\Delta \rightarrow \hat{State})$$

that is,

- program's all executions = the collection of all the machine states occurring during the executions

$$[[C]] \in 2^{State}$$

- program's abstract semantics = partition and abstract the collection:

$$[[\hat{C}]] \in \Delta \rightarrow \hat{State}$$

- Δ : a finite set of partitioning indices
- e.g.) $\Delta =$ the set of program points

A Sufficient Condition for $\alpha \circ \text{Next} \sqsubseteq \widehat{\text{Next}} \circ \alpha$ (2/4)

The Galois-connection

$$2^{\text{State}} \underset{\alpha}{\overset{\gamma}{\rightleftarrows}} (\Delta \rightarrow \widehat{\text{State}})$$

is

$$\alpha = (\wp \alpha_1) \circ \pi.$$

- α_1 abstracts sets of states into abstract states:

$$2^{\text{State}} \underset{\alpha_1}{\overset{\gamma_1}{\rightleftarrows}} \widehat{\text{State}}.$$

- π and $\hat{\pi}$ are partition functions:

$$\begin{aligned} \pi &\in 2^{\text{State}} \rightarrow 2^{2^{\text{State}}} \\ \hat{\pi} &\in 2^{\widehat{\text{State}}} \rightarrow (\Delta \rightarrow 2^{\widehat{\text{State}}}) \end{aligned}$$

A Sufficient Condition for $\alpha \circ Next \sqsubseteq \hat{Next} \circ \alpha$ (3/4)

Define

$$\begin{aligned} Next &= \cup \circ (\wp next) && \in 2^{State} \rightarrow 2^{State} \\ \hat{Next} &= (\wp \perp) \circ \hat{\pi} \circ \cup \circ (\wp n\hat{e}xt) && \in (\Delta \rightarrow State) \rightarrow (\Delta \rightarrow State) \end{aligned}$$

where

- concrete transition $next$:

$$next \in State \rightarrow State$$

(transitions terminal state into itself)

- abstract transition $n\hat{e}xt$:

$$n\hat{e}xt \in State \rightarrow 2^{State}$$

(may transition one abstract state into multiple abstract states)

A Sufficient Condition for $\alpha \circ Next \sqsubseteq \hat{Next} \circ \alpha$ (4/4)

Theorem (Correctness)

Let $Next$ and \hat{Next} be:

$$Next = \cup \circ (\wp next) \in 2^{State} \rightarrow 2^{State}$$

$$\hat{Next} = (\wp \sqcup) \circ \hat{\pi} \circ \cup \circ (\wp n\hat{ext}) \in (\Delta \rightarrow State) \rightarrow (\Delta \rightarrow State)$$

If the below two conditions hold then $\alpha \circ Next \sqsubseteq \hat{Next} \circ \alpha$.

1. Condition on abstract partitioning($\hat{\pi}$):

$$(\wp \alpha_1) \circ \pi \circ \cup \circ (\wp \gamma) \sqsubseteq (\wp \sqcup) \circ \hat{\pi} \quad (1)$$

2. Condition on abstract transition($n\hat{ext}$):

$$next\ x \in (\cup \circ (\wp \gamma) \circ n\hat{ext} \circ \alpha_1) \{x\} \quad (2)$$

Notation

- $\uparrow \in X \rightarrow 2^X \quad \uparrow x = \{x\}$.
- $f \in A \rightarrow B \quad \wp f \in 2^A \rightarrow 2^B \quad (\wp f)X = \{fx \mid x \in X\}$.
- $f \in A \rightarrow 2^B \quad \wp_{\cup} f = \cup \circ \wp f$.

Facts

- $\wp_{\cup}(f \circ g) = (\wp_{\cup} f) \circ (\wp g)$.
- $\wp_{\cup}(\wp_{\cup} f) \circ (\wp g) = (\wp_{\cup} f) \circ (\wp_{\cup} g)$.
- $x \in A, X \in 2^A, fx \in gx \quad (\wp f)X \subseteq (\wp_{\cup} g)X$.

Proof. First, from condition (2) the following holds:

$$\wp next \sqsubseteq (\wp \cup \gamma) \circ (\wp \cup \hat{next}) \circ \alpha \quad (3)$$

Because,

$$\begin{aligned} \wp next &\sqsubseteq \wp \cup ((\wp \cup \gamma) \circ \hat{next} \circ \alpha_1 \circ \uparrow) && \text{(cond. (2), } (fx \in gx \text{ then } (\wp f)X \sqsubseteq (\wp g)X) \\ &= \wp \cup (\wp \cup \gamma) \circ \wp (\hat{next} \circ \alpha_1 \circ \uparrow) && (\wp \cup (f \circ g)) = (\wp \cup f) \circ (\wp g) \\ &= (\wp \cup \gamma) \circ (\wp \cup \hat{next}) \circ (\wp \alpha_1) \circ (\wp \uparrow) && (\wp \cup (\wp \cup f)) \circ (\wp g) = (\wp \cup f) \circ (\wp \cup g) \\ &\sqsubseteq (\wp \cup \gamma) \circ (\wp \cup \hat{next}) \circ (\wp \alpha_1) \circ \pi && (\gamma, \hat{next}, \alpha_1 \text{ are all monotonic)} \\ &= (\wp \cup \gamma) \circ (\wp \cup \hat{next}) \circ \alpha. \end{aligned}$$

Therefore,

$$\begin{aligned} \alpha \circ Next &= (\wp \alpha_1) \circ \pi \circ (\wp next) \\ &\sqsubseteq (\wp \alpha_1) \circ \pi \circ (\wp \cup \gamma) \circ (\wp \cup \hat{next}) \circ \alpha \quad \text{(cond. (3))} \\ &\sqsubseteq (\wp \sqcup) \circ \hat{\pi} \circ (\wp \cup \hat{next}) \circ \alpha \quad \text{(cond. (1))} \\ &= \hat{Next} \circ \alpha. \end{aligned}$$

That is, from condition (1) and condition (2), $\alpha \circ Next \sqsubseteq \hat{Next} \circ \alpha$ holds. Hence by the Fixpoint Transfer Theorem,

$$\alpha(\text{fix}(\lambda T. T_0 \cup Next T)) \sqsubseteq \text{fix}(\lambda \hat{T}. \alpha(T_0) \sqcup \hat{Next} \hat{T}).$$

□

Trace Abstract Interpretation's Algorithm (1/4)

Static analysis is to compute $[[\hat{C}]]$, which is

$$\text{fix}(\hat{F} \stackrel{\text{let}}{=} \lambda \hat{T}. \alpha(T_0) \sqcup \text{Next } \hat{T})$$

where

$$\begin{aligned}\hat{F} &\in \text{Trace} \rightarrow \text{Trace} \\ \text{Trace} &= \Delta \rightarrow \text{State} \\ \text{Next} &= (\wp \sqcup) \circ \hat{\pi} \circ (\wp \sqcup \hat{n}ext) \\ \hat{n}ext &\in \text{State} \rightarrow 2^{\text{State}}.\end{aligned}$$

Computing $\text{fix} \hat{F}$ is to compute Y_i until no change:

$$Y_0 = \alpha(T_0), \quad Y_{n+1} = \alpha(T_0) \sqcup \text{Next}(Y_n)$$

Hence,

```
T, T': Δ → State;
begin
  T := T' := α(T0);
  repeat
    T' := T;
    T := α(T0) ⊔ ((⊔) ∘ π̂)(⋃i ∈ Δ next T[i]);
  until T ⊆ T'; (* no more increase *)
return T';
end
```

Figure: Naive algorithm

Trace Abstract Interpretation's Algorithm (2/4)

When widening(∇) and narrowing(Δ) are necessary, we compute the following two things in sequence:

$$\text{Widen}(\hat{F}) = \lim_{i \in \mathbb{N}} \begin{cases} \hat{Y}_0 & = \alpha(T_0) \\ \hat{Y}_{i+1} & = \begin{cases} \hat{Y}_i & \text{if } \hat{F}(\hat{Y}_i) \sqsubseteq \hat{Y}_i \\ \hat{Y}_i \nabla \hat{F}(\hat{Y}_i) & \text{o.w.} \end{cases} \end{cases}$$

$$\text{Narrow}(\hat{m}) = \lim_{i \in \mathbb{N}} \begin{cases} \hat{Z}_0 & = \hat{m} \\ \hat{Z}_{i+1} & = \hat{Z}_i \Delta \hat{F}(\hat{Z}_i) \end{cases}$$

Hence,

```
T, T', Y:  $\Delta \rightarrow \text{State}$ ;  
begin  
  T := T' :=  $\alpha(T_0)$ ;  
  repeat  
    T' := T;  
    Y :=  $\alpha(T_0) \sqcup ((\wp \sqcup) \circ \hat{\pi})(\bigcup_{i \in \Delta} \text{next } T[i])$ ;  
    T := if Y  $\sqsubseteq$  T' then T' else T'  $\nabla$  Y;  
  until T  $\sqsubseteq$  T'; (* no more increase *)  
  
  repeat  
    T := T'  
    T' \Delta :=  $\alpha(T_0) \sqcup ((\wp \sqcup) \circ \hat{\pi})(\bigcup_{i \in \Delta} \text{next } T[i])$ ;  
  until T  $\sqsupseteq$  T'; (* no more decrease *)  
  return T;  
end
```

Figure: Naive algorithm with widening and narrowing

Trace Abstract Interpretation's Algorithm (3/4)

Worklist method:

- wasteful at each iteration to compute

$$\bigcup_{i \in \Delta} \hat{next} T[i]$$

for every index in Δ .

- enough to compute those affected from the previous iteration

```
T, T':  $\Delta \rightarrow State$ ;  
W:  $2^\Delta$ ; (* worklist *)  
begin  
  T := T' :=  $\alpha(T_0)$ ;  W :=  $\Delta$ ;  
  repeat  
    T' := T;  
    T :=  $\alpha(T_0) \sqcup ((\emptyset \sqcup) \circ \hat{\pi})(\bigcup_{i \in W} \hat{next} T[i])$ ;  
    W :=  $\{i \in \Delta \mid T[i] \not\sqsubseteq T'[i]\}$ ;  
  until W =  $\{\}$ ; (* no more increase *)  
return T';  
end
```

Figure: Worklist algorithm

Trace Abstract Interpretation's Algorithm (4/4)

```
 $T, T', Y: \Delta \rightarrow \hat{State};$   
 $W: 2^\Delta; (* \text{worklist} *)$   
begin  
   $T := T' := \alpha(T_0); \quad W := \Delta;$   
  repeat  
     $T' := T;$   
     $Y := \alpha(T_0) \sqcup ((\emptyset \sqcup) \circ \hat{\pi})(\bigcup_{i \in W} \text{next } T[i]);$   
     $T := \text{if } Y \sqsubseteq T' \text{ then } T' \text{ else } T' \nabla Y;$   
     $W := \{i \in \Delta \mid T[i] \not\sqsubseteq T'[i]\};$   
  until  $W = \{\}; (* \text{no more increase} *)$   
  
   $W := \Delta;$   
  repeat  
     $T := T';$   
     $T' \Delta := \alpha(T_0) \sqcup ((\emptyset \sqcup) \circ \hat{\pi})(\bigcup_{i \in W} \text{next } T[i]);$   
     $W := \{i \in \Delta \mid T[i] \not\sqsubseteq T'[i]\};$   
  until  $W = \{\}; (* \text{no more decrease} *)$   
  return  $T;$   
end
```

Figure: Worklist algorithm with widening and narrowing