Static Analysis Design Framework: Abstract Interpration

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

- 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

- without abstraction, can't capture all possible executions
- without abstraction, can't terminate
- Abstraction is not omission
 - \bullet reality: $\{2,4,6,8,\cdots\}$
 - "even number" (abstraction) vs "multiple of 4" (omission)

Abstract Interpration Framework

real execution abstract execution correctness implementation

$$\begin{split} \llbracket C \rrbracket &= fix F \in D \\ \llbracket \hat{C} \rrbracket &= \lim_{i \in \mathbb{N}} \hat{F}^i(\perp_{\hat{D}}) \in \hat{D} \\ \llbracket C \rrbracket &\approx \llbracket \hat{C} \rrbracket \\ \text{computation of } \llbracket \hat{C} \rrbracket \end{split}$$

The framework requires:

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

The framework guarantees:

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

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

- Define semantic domain CPO D.
- Define the real executions as the least fixed point $f\!i\!xF$ of continuous function $F\in D\to D$

$$fixF = \bigsqcup_{i \in \mathbb{N}} F^i(\bot_D)$$

Plan: define an abstraction that captures fixF

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 fixF by using \hat{F}

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

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

$$D \xrightarrow[]{\alpha}{\longrightarrow} \hat{D}.$$

• Galois connection:

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

• Galois connection captures our intention:

- bigger elements in \hat{D} means more.
- $\bullet \ \alpha$ abstracts .
- γ concretizes.

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

• \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 $igsqcup_{i\in\mathbb{N}}\hat{F}^i(\hat{ot})$

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

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

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

or,

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

$$\alpha(f) \sqsubseteq \widehat{f} \ \Rightarrow \ \alpha(F \, f) \sqsubseteq \widehat{F} \, \widehat{f}$$

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

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:

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

Theorem[fixpoint-transfer]

• Analysis result $\hat{\mathcal{A}}$ subsumes the real executions $f\!ixF$

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

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

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

The computation always terminates.

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

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

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

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

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

$$\begin{array}{rcl} \hat{X}_0 &=& \hat{\bot} \\ \hat{X}_{i+1} &=& \left\{ \begin{array}{cc} \hat{X}_i & \text{ if } \hat{F}(\hat{X}_i) \sqsubseteq \hat{X}_i \\ \hat{X}_i \bigtriangledown \hat{F}(\hat{X}_i) & \text{ o.w.} \end{array} \right. \end{array}$$

Conditions on Widening \bigtriangledown

Conditions

 $\bullet \ \forall a,b \in \hat{D}: (a \sqsubseteq a \bigtriangledown b) \ \land \ (b \sqsubseteq a \bigtriangledown b)$

• \forall increasing chain $\{a_i\}_i$: chain $x_0 = a_0, x_{i+1} = x_i \bigtriangledown a_{i+1}$ 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:

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

Theorem[widen's safety]

If \hat{F} is monotonic,

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

$$\hat{Y}_0 = \hat{\mathcal{A}} \hat{Y}_{i+1} = \hat{Y}_i \bigtriangleup \hat{F}(\hat{Y}_i)$$

Conditions

• $\forall a, b \in \hat{D} : a \sqsupseteq b \Rightarrow a \sqsupseteq (a \bigtriangleup b) \sqsupseteq b$

• \forall decreasing chain $\{a_i\}_i$: 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]

Fixpoint Transfer Theorem

Theorem (fixpoint transfer)

Let CPOs D and \hat{D} are Galois-connected. Function $F: D \to D$ is continuous. $\hat{F}: \hat{D} \to \hat{D}$ 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,

$$\alpha(fixF) \sqsubseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot}).$$

Widening/Narrowing Theorems

Theorem (widen's safety)

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

Theorem (narrow's safety)

Let $\hat{F}: \hat{D} \to \hat{D}$ be monotonic over CPO \hat{D} . Let narrowing operator $\Delta: \hat{D} \times \hat{D} \to \hat{D}$ 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 $\lim_{i \in \mathbb{N}} \hat{Y}_i \sqsupseteq \bigcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$. Abstract Interpretation Example (or, a Special Abstract Interpretation Framework)

Semantics as Trace

Program C's semantics $[\![C]\!]$ is the set of all execution traces

$$\begin{bmatrix} \mathbb{C} \end{bmatrix} \in 2^{Trace} \\ \tau, \tau_0 \tau_1 \cdots \tau_n \in Trace = State^* \\ State = Command \times Memory \times \cdots$$

Side:

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

$$2^{Trace} \xrightarrow{\gamma} Trace$$

 $\stackrel{\alpha_0}{\rightarrow}$ Trace of set of states: sequence of set of states appearing at a given time along at least one of the traces

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

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

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

 $\begin{array}{c} \overset{\alpha_2 \circ \alpha_1 \circ \alpha_0}{\rightarrow} \\ \text{Partitioned set of reachable states (local invariant): e.g.,} \\ \text{project along each control point} \in \Delta \text{ (a finite set)} \end{array}$

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

 $\stackrel{\alpha_3\circ\alpha_2\circ\alpha_1\circ\alpha_0}{\to} \text{ Abstracting the partitioned set of reachable states}$

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

where

$$2^{State} \Longrightarrow State$$

伺 ト く ヨ ト く ヨ ト

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

where

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

To show is $\alpha(fixF) \sqsubseteq fix\hat{F}$, i.e., $\alpha \circ F \sqsubseteq \hat{F} \circ \alpha$. A sufficient condition, if Trace and Trace are \sqcup -closed, is:

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

(easy to see, by Galois-connection.)

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

Focus on:

$$2^{State} \stackrel{\gamma}{\underset{\alpha}{\longleftarrow}} (\Delta \to State)$$

that is,

• program's all executions = the collection of all the machine states occuring during the executions

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

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

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

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

For
$$f \in A \to B$$
,
• $\wp f \in 2^A \to 2^B$ is $(\wp f)X = \{fx \mid x \in X\}$.
• Abusely, $\wp f \in (\Delta \to A) \to 2^B$ is
 $(\wp f)X = \{fx \mid x \in rangeX\}$.

-2

イロト イヨト イヨト イヨト

A Sufficient Condition for $\alpha \circ Next \sqsubseteq Next \circ \alpha$ (2/4)

The Galois-connection

$$2^{State} \xrightarrow[\alpha]{\gamma} (\Delta \to State)$$

is

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

• α_1 abstracts sets of states into abstract states:

$$2^{State} \xrightarrow{\gamma_1} State.$$

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

$$\begin{array}{rcl} \pi & \in & 2^{State} \to 2^{2^{State}} \\ \hat{\pi} & \in & 2^{State} \to (\Delta \to 2^{State}) \end{array}$$

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

Define

$$\begin{array}{ll} Next = \wp next & \in 2^{State} \to 2^{State} \\ \hat{Next} = (\wp \sqcup) \circ \hat{\pi} \circ \cup \circ (\wp next) & \in (\Delta \to State) \to (\Delta \to State) \end{array}$$

where

• concrete transition *next*:

$$next \in State \rightarrow State$$

(transitions terminal state into itself)

• abstract transition \hat{next} :

$$\hat{next} \in \hat{State} \to 2^{\hat{State}}$$

(may transition one abstract state into multiple abstract states)

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

Theorem (Correctness)

Let Next and Next be:

 $\begin{array}{ll} \textit{Next} = \wp\textit{next} & \in 2^{\textit{State}} \rightarrow 2^{\textit{State}} \\ \hat{\textit{Next}} = (\wp \sqcup) \circ \hat{\pi} \circ \cup \circ (\wp \hat{\textit{next}}) & \in (\Delta \rightarrow S \hat{\textit{tate}}) \rightarrow (\Delta \rightarrow S \hat{\textit{tate}}) \end{array}$

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

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

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

2. Condition on abstract transition(next):

$$next \ x \in (\cup \circ (\wp \gamma) \circ next \circ \alpha_1) \ \{x\}$$
(2)

In Proof

Notation

•
$$\uparrow \in X \to 2^X$$
 is $\uparrow x = \{x\}$.
• For $f \in A \to B$, $\wp f \in 2^A \to 2^B$ is $(\wp f)X = \{fx \mid x \in X\}$.
• Abusely, $\wp f \in (\Delta \to A) \to 2^B$ is $(\wp f)X = \{fx \mid x \in rangeX\}$.
• For $f \in A \to 2^B$, $\wp_{\cup} f = \cup \circ \wp f$.

Facts

•
$$\wp_{\cup}(f \circ g) = (\wp_{\cup}f) \circ (\wp g).$$

• $\wp_{\cup}(\wp_{\cup}f) \circ (\wp g) = (\wp_{\cup}f) \circ (\wp_{\cup}g)$

• For
$$x \in A, X \in 2^A$$
, $fx \in gx$ implies $(\wp f)X \subseteq (\wp \cup g)X$.

-2

<ロ> <部> < 部> < き> < き> <</p>

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

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

Because,

$$\begin{array}{ll} \wp next & \sqsubseteq & \wp_{\cup}((\wp_{\cup}\gamma) \circ n\hat{ext} \circ \alpha_{1} \circ \uparrow) & (\text{cond. (2), } (fx \in gx \text{ then } (\wp f)X \subseteq (\wp_{\cup}g). \\ & = & \wp_{\cup}(\wp_{\cup}\gamma) \circ (\hat{ext} \circ \alpha_{1} \circ \uparrow) & (\wp_{\cup}(f \circ g) = (\wp_{\cup}f) \circ (\wp g)) \\ & = & (\wp_{\cup}\gamma) \circ (\wp_{\cup}n\hat{ext}) \circ (\wp\alpha_{1}) \circ (\wp \uparrow) & (\wp_{\cup}(\wp_{\cup}f) \circ (\wp g) = (\wp_{\cup}f) \circ (\wp_{\cup}g)) \\ & \sqsubseteq & (\wp_{\cup}\gamma) \circ (\wp_{\cup}n\hat{ext}) \circ (\wp\alpha_{1}) \circ \pi & (\gamma, n\hat{ext}, \alpha_{1} \text{ are all monotonic)} \\ & = & (\wp_{\cup}\gamma) \circ (\wp_{\cup}n\hat{ext}) \circ \alpha. \end{array}$$

Therefore,

$$\begin{array}{lll} \alpha \circ Next &=& (\wp\alpha_1) \circ \pi \circ (\wp next) \\ & \sqsubseteq & (\wp\alpha_1) \circ \pi \circ (\wp_{\cup}\gamma) \circ (\wp_{\cup} next) \circ \alpha & (\mathsf{cond.} \ (3)) \\ & \sqsubseteq & (\wp_{\sqcup}) \circ \pi \circ (\wp_{\cup} next) \circ \alpha & (\mathsf{cond.} \ (1)) \\ & = & Next \circ \alpha. \end{array}$$

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

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

- ∢ ≣ ▶

-

Trace Abstract Interpretaion's Algorithm (1/4)

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

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

where

$$\begin{array}{rcl} \hat{F} & \in & Trace \to Trace \\ Trace & = & \Delta \to State \\ Next & = & (\wp \sqcup) \circ \hat{\pi} \circ (\wp \sqcup n \hat{e}xt) \\ n \hat{e}xt & \in & State \to 2^{State}. \end{array}$$

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

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

Hence,

$$\begin{array}{l} T,T'\colon\Delta\to State;\\ \texttt{begin}\\ T:=T':=\alpha(T_0);\\ \texttt{repeat}\\ T':=T;\\ T:=\alpha(T_0)\sqcup\ ((\wp\sqcup)\circ\hat\pi)(\bigcup_{i\in\Delta}n\hat{ext}\ T[i]);\\ \texttt{until}\ T\sqsubseteq T';\ (\texttt{* no more increase }\texttt{*})\\ \texttt{return}\ T';\\ \texttt{end}\\ \hline\\ \texttt{Figure: Naive algorithm} \end{array}$$

Trace Abstract Interpretaion's Algorithm (2/4)

When widening(\bigtriangledown) and narrowing(\triangle) are necessary, we compute the folloing two things in sequence:

$$\begin{split} \textit{Widen}(\hat{F}) &= \lim_{i \in \mathbb{N}} \left\{ \begin{array}{ll} \hat{Y}_0 &= \alpha(T_0) \\ \hat{Y}_{i+1} &= \left\{ \begin{array}{ll} \hat{Y}_i & \text{if } \hat{F}(\hat{Y}_i) \sqsubseteq \hat{Y}_i \\ \hat{Y}_i \bigtriangledown \hat{F}(\hat{Y}_i) & \text{o.w.} \end{array} \right. \end{split} \right. \end{split}$$

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

Hence,

(日) (同) (三) (三)

Trace Abstract Interpreation'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

```
\begin{array}{l} T,T':\Delta\rightarrow State;\\ W:2^{\Delta};\ (\texttt{* worklist}\,\texttt{*})\\ \texttt{begin}\\ T:=T':=\alpha(T_0);\quad W:=\Delta;\\ \texttt{repeat}\\ T':=T;\\ T:=\alpha(T_0)\sqcup((\wp\sqcup)\circ\hat{\pi})(\bigcup_{i\in W}\ next\,T[i]);\\ W:=\{i\in\Delta\mid T[i]\not\sqsubseteq T'[i]\};\\ \texttt{until}\quad W=\{\};\ (\texttt{* no more increase}\,\texttt{*})\\ \texttt{return}\quad T';\\ \texttt{end}\\ \hline \texttt{Figure: Worklist algorithm} \end{array}
```

Trace Abstract Interpretaion's Algorithm (4/4)

```
T, T', Y: \Delta \rightarrow State:
   W: 2^{\Delta}; (* worklist *)
   begin
         T := T' := \alpha(T_0); \quad W := \Delta;
         repeat
               T' := T:
               Y := \alpha(T_0) \sqcup ((\wp \sqcup) \circ \hat{\pi}) (\bigcup_{i \in W} \hat{next} T[i]);
               T := \text{if } Y \sqsubseteq T' \text{ then } T' \text{ else } T' \bigtriangledown Y;
                W := \{i \in \Delta \mid T[i] \not \sqsubset T'[i]\};
         until W = \{\}; (* no more increase *)
         W := \Delta;
         repeat
               T := T':
               T' \ \triangle := \alpha(T_0) \sqcup ((\wp \sqcup) \circ \hat{\pi})(\bigcup_{i \in W} next \ T[i]);
                W := \{i \in \Delta \mid T[i] \not \sqsubset T'[i]\};
         until W = \{\}; (* \text{ no more decrease } *)
         return T:
   end
Figure: Worklist algorithm with widening and narrowing
```