Homework 4

SNU 4541.664A

due: 5/01 in class

Kwangkeun Yi

- 이번 숙제의 목적은, D-nogoto가 D로 확장되면서 프로그램의 의미를 조립식으로 만들기 곤란해지는데, 이 경우 적절한 스타일의 의미구조로 정적분석기를 정의하고 안전성을 증명하는 것이다.
- 여기서 "정적분석기"는 일반화된(generic) 버전이다: 정적분석기(요약해석기)에서 사용하는 요약 의미공간(abstract domains)은 해당 의미공간과 갈로아 연결되었다고(Galois connection)만 가정한다. 요약 의미공간들을 실제로 정의할 필요는 없다.

D언어는 D-nogoto에서 `goto`문을 더한 것이다(프로젝트 설명자료와 동일):

$$
\begin{array}{rrcl}
\text{Cmd} & C & \to & x \texttt{ := } E \mid \texttt{*}x \texttt{ := } E \\
& & \mid & C \texttt{ ; } C \\
& & \mid & \texttt{if } E\ C\ C \\
& & \mid & \texttt{repeat } C\ E \\
& & \mid & \texttt{goto } E \\
\text{Exp} & E & \to & n \quad (n \in \mathbb{Z}) \\
& & \mid & E \texttt{ + } E \mid E \texttt{ * } E \mid \texttt{-} E \\
& & \mid & E \texttt{ < } E \\
& & \mid & x \mid \texttt{*}x \mid \texttt{\&}x \\
& & \mid & \texttt{readInt} \quad \text{(정수입력)}
\end{array}
$$

Exercise 1 D의 의미구조를 기계상태 전이과정(transitional style)으로 정의하고, 그 요약 버전을 교과서와 강의노트를 참고하여 정의한다.

분석기를 정의하는 여러분이 할 일을 다음과 같다:

1. 분석할 프로그램을 구성하는 Cmd 레이블 l 마다 $\texttt{next}(l), \texttt{nextTrue}(l), \texttt{nextFalse}(l)$ 관계를 정의한다. 이 관계들은 레이블 l에 해당하는 Cmd를 수행하고 난 후 다음으로 수행해야 할 Cmd를 정해준다. 이 관계들은 프로그램 텍스트만 보고 정해질 수 있는 한계안에서 정의한다.

   주어진 프로그램의 모든 Cmd마다 고유의 자연수 레이블이 붙어있다고 가정한다.

2. 각 Cmd마다 한걸음 관계 $\hookrightarrow$를

$$\hookrightarrow \quad \subseteq \quad \mathbb{S} \times \mathbb{S}$$

   정의한다. 여기서,

$$\begin{array}{rcl} \text{상태} \quad \mathbb{S} & = & \mathbb{L} \times \mathbb{M} \\ \text{레이블} \quad \mathbb{L}, & \text{메모리} & \quad \mathbb{M}. \end{array}$$

   프로그램 C의 모듬의미 $\underline{C}$(분석하고자하는 대상)은 다음이다: 관심있는 초기 메모리집합 M_0가 프로그램 시작점(레이블 l_0)에서부터

$$I = \{\langle l_0, m\rangle \mid m \in M_0\} \in 2^{\mathbb{S}}$$

   전이되는 모든 기계상태다. 즉, 아래 함수 F의 최소고정점

$$\underline{C} = \mathit{fix} F$$

   이다:

$$\begin{array}{rcl} F & : & 2^{\mathbb{S}} \to 2^{\mathbb{S}} \\ F(X) & = & I \cup \mathit{Step}(X) \\ \mathit{Step} & = & \breve{\wp}(\hookrightarrow) \end{array}$$

3. 각 Cmd마다 요약 한걸음 관계 $\hookrightarrow^\sharp$를

$$\hookrightarrow^\sharp \quad \subseteq \quad (\mathbb{L} \times \mathbb{M}^\sharp) \times (\mathbb{L} \times \mathbb{M}^\sharp)$$

   정의한다.

   다음을 가정한다: 모든 요약 의미공간들이 갈로아 연결되어 있다. 즉, 각 요약 의미공간 $X^\sharp$는

$$2^X \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} X^\sharp.$$

주어진 프로그램 C의 요약의미 $\underline{C}^\sharp$는 다음 함수 $F^\sharp$의 최소고정점

$$\underline{C}^\sharp = fix F^\sharp$$

로 정의된다:

$$\begin{array}{rcl} F^\sharp & : & (\mathbb{L} \to \mathbb{M}^\sharp) \to (\mathbb{L} \to \mathbb{M}^\sharp) \\ F^\sharp(X^\sharp) & = & \alpha(I) \ \cup^\sharp \ Step^\sharp(X^\sharp) \\ Step^\sharp & = & \wp(\mathrm{id}, \cup^\sharp) \circ \pi \circ \breve{\wp}(\hookrightarrow^\sharp) \end{array}$$

요약 한걸음 관계 $\hookrightarrow^\sharp$를 정의할 때 사용하는 모든 요약의미 연산자들(abstract semantic operators)들의 타입(입출력 요약 의미공간들)이 무엇인지 확인한다.

4. 모든 요약의미 연산자들이 안전하게 정의되었다고 가정한다. 즉, 의미 연산자

$$f : 2^A \to 2^B$$

와 그에 해당하는 요약의미 연산자

$$f^\sharp : A^\sharp \to B^\sharp$$

사이에 항상 다음의 조건이 만족된다고 가정한다:

$$f \circ \gamma_A \sqsubseteq \gamma_B \circ f^\sharp.$$

(위에서 A와 B는 의미 연사자들에 따라서 다른 집합들이다)

5. 이러한 가정아래서, 다음을 증명하면 충분하다:

$$\breve{\wp}(\hookrightarrow) \circ \gamma \subseteq \gamma \circ \breve{\wp}(\hookrightarrow^\sharp).$$

Exercise 2 위와 같이 정의한 분석기의 구현 알고리즘을 할일만하기(worklist algorithm) 버전으로 디자인한다. 강의 슬라이드에서 소개한 알고리즘은 상위 버전이다. 실행기함수(definitional interpreter)를 이용한 분석기 디자인 틀에서 소개한 (slides 8-1) 할일만하기 알고리즘 수준에서 고안한다.

Exercise 3 D의 의미구조를 실행기함수(definitional interpreter)를 이용해서 정의한다. 실행기는 모듬 실행기다. 메모리 집합을 받아서 메모리 집합을 내놓는다. 일반 실행기를 메모리 집합으로 자연스레 확장한 것이다(natural powerset exten-

sion)

$$\mathcal{I} : \text{Cmd} \to 2^{\mathbb{M}} \to 2^{\mathbb{M}}$$

재귀 함수(방정식으)로 표현될 위의 $\mathcal{I}$는 $\mathcal{I}$방정식을 정의하는 해당 연속함수

$$\mathcal{F} : (\text{Cmd} \to 2^{\mathbb{M}} \to 2^{\mathbb{M}}) \to (\text{Cmd} \to 2^{\mathbb{M}} \to 2^{\mathbb{M}})$$

의 최소고정점이다. 즉, 프로그램 C의 모듬의미(collecting semantics) $\underline{C}$는

$$\underline{C} = (\mathit{fix}\mathcal{F})(C)$$

이다.

정적분석은 관심있는 입력 메모리집합 M_0에 대한 $\underline{C}(M_0)$을 안전하게 어림잡는것이다(safe upper-approximation).

분석기를 정의하는 여러분이 할 일을 다음과 같다:

1. 모듬 실행기 $\mathcal{I}$

   $$\mathcal{I} : \text{Cmd} \to 2^{\mathbb{M}} \to 2^{\mathbb{M}}$$

   를 정의한다. 이때 사용하는 모든 의미 연산자들(semantic operators)을 정의하고 그 타입(입출력 의미공간들)이 무엇인지 확인한다.

2. 요약 실행기 $\mathcal{I}^\sharp$

   $$\mathcal{I}^\sharp : \text{Cmd} \to \mathbb{M}^\sharp \to \mathbb{M}^\sharp$$

   를 모듬 실행기와 같은 모양이 되도록 정의한다(homomorphic definition). 오직 다른 점은 의미 연산자들의 요약버전을 사용하는 것이다.

   이때 사용하는 모든 요약의미 연산자들(abstract semantic operators)들의 타입(입출력 요약 의미공간들)이 무엇인지 확인한다.

3. 재귀 함수(방정식으)로 표현될 위의 $\mathcal{I}^\sharp$는 $\mathcal{I}^\sharp$방정식을 정의하는 해당 연속함수

   $$\mathcal{F}^\sharp : (\text{Cmd} \to \mathbb{M}^\sharp \to \mathbb{M}^\sharp) \to (\text{Cmd} \to \mathbb{M}^\sharp \to \mathbb{M}^\sharp)$$

   의 최소고정점이다. 즉, 프로그램 C의 요약 의미 $\underline{C}^\sharp$는

   $$\underline{C}^\sharp = (\mathit{fix}\mathcal{F}^\sharp)(C)$$

   이다.

4. 모든 요약 의미공간들이 갈로아 연결되어 있다고 가정한다. 즉, 각 요약 의미공간 $X^\sharp$는

$$2^X \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} X^\sharp.$$

그리고, 모든 요약의미 연산자들이 안전하게 정의되었다고 가정한다. 즉, 의미 연산자

$$f : 2^A \to 2^B$$

와 그에 해당하는 요약의미 연산자

$$f^\sharp : A^\sharp \to B^\sharp$$

사이에 항상 다음의 조건이 만족된다고 가정한다:

$$f \circ \gamma_A \sqsubseteq \gamma_B \circ f^\sharp.$$

(위에서 A와 B는 의미 연사자들에 따라서 다른 집합들이다.)

5. 이러한 가정아래서 다음을 증명한다:

$$\forall C : \underline{C} \circ \gamma \sqsubseteq \gamma \circ \underline{C}^\sharp.$$