

Semantics and Verification of Software

Lecture 20: Dataflow Analysis III (The Framework)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University
noll@cs.rwth-aachen.de

<http://www-i2.informatik.rwth-aachen.de/i2/svsw10/>

Summer Semester 2010

- 1 Repetition: Heading for a Dataflow Analysis Framework
- 2 Order-Theoretic Foundations
- 3 The Framework
- 4 Solving Dataflow Equation Systems

Similarities between Analysis Problems

- **Observation:** the analyses presented so far have some **similarities**
⇒ Look for underlying **framework**
- **Advantage:** possibility for designing (efficient) **generic algorithms** for solving **dataflow equations**
- **Overall pattern:** for $c \in Cmd$ and $l \in L_c$, the **analysis information** (AI) is described by **equations** of the form

$$AI_l = \begin{cases} \iota & \text{if } l \in E \\ \bigsqcup \{\varphi_{l'}(AI_{l'}) \mid (l', l) \in F\} & \text{otherwise} \end{cases}$$

where

- the set of extremal labels, E , is $\{\text{init}(c)\}$ or $\text{final}(c)$
- ι specifies the extremal analysis information
- the combination operator, \bigsqcup , is \cap or \cup
- $\varphi_{l'}$ denotes the transfer function of block $B^{l'}$
- the flow relation F is $\text{flow}(c)$ or $\text{flow}^R(c)$
($:= \{(l', l) \mid (l, l') \in \text{flow}(c)\}$)

- **Direction of information flow:**

- **forward:**

- $F = \text{flow}(c)$
 - Al_l concerns entry of B^l
 - c has isolated entry

- **backward:**

- $F = \text{flow}^R(c)$
 - Al_l concerns exit of B^l
 - c has isolated exits

- **Quantification over paths:**

- **may:**

- $\sqcup = \bigcup$
 - property satisfied by some path
 - interested in least solution (later)

- **must:**

- $\sqcup = \bigcap$
 - property satisfied by all paths
 - interested in greatest solution (later)

Goal: solve dataflow equation system by fixpoint iteration

- ① Introduce **partial order** for comparing analysis results
- ② Establish **least upper bound** as combination operator
- ③ Ensure **monotonicity** of transfer functions
- ④ Guarantee termination of fixpoint iteration (and continuity of functional) by **Ascending Chain Condition**
- ⑤ Optimize fixpoint iteration by **worklist algorithm**

- 1 Repetition: Heading for a Dataflow Analysis Framework
- 2 Order-Theoretic Foundations
- 3 The Framework
- 4 Solving Dataflow Equation Systems

Partial Orders

The domain of analysis information usually forms a partial order where the ordering relation compares the “precision” of information.

Definition 20.1 (Partial order; repetition of Def. 7.1)

A **partial order (PO)** (D, \sqsubseteq) consists of a set D , called **domain**, and of a relation $\sqsubseteq \subseteq D \times D$ such that, for every $d_1, d_2, d_3 \in D$,

reflexivity: $d_1 \sqsubseteq d_1$

transitivity: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_3 \implies d_1 \sqsubseteq d_3$

antisymmetry: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_1 \implies d_1 = d_2$

It is called **total** if, in addition, always $d_1 \sqsubseteq d_2$ or $d_2 \sqsubseteq d_1$.

Partial Orders

The domain of analysis information usually forms a partial order where the ordering relation compares the “precision” of information.

Definition 20.1 (Partial order; repetition of Def. 7.1)

A **partial order (PO)** (D, \sqsubseteq) consists of a set D , called **domain**, and of a relation $\sqsubseteq \subseteq D \times D$ such that, for every $d_1, d_2, d_3 \in D$,

reflexivity: $d_1 \sqsubseteq d_1$

transitivity: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_3 \implies d_1 \sqsubseteq d_3$

antisymmetry: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_1 \implies d_1 = d_2$

It is called **total** if, in addition, always $d_1 \sqsubseteq d_2$ or $d_2 \sqsubseteq d_1$.

Example 20.2

① (Live Variables) $(2^{Var_c}, \subseteq)$ is a (non-total) partial order

Partial Orders

The domain of analysis information usually forms a partial order where the ordering relation compares the “precision” of information.

Definition 20.1 (Partial order; repetition of Def. 7.1)

A **partial order (PO)** (D, \sqsubseteq) consists of a set D , called **domain**, and of a relation $\sqsubseteq \subseteq D \times D$ such that, for every $d_1, d_2, d_3 \in D$,

reflexivity: $d_1 \sqsubseteq d_1$

transitivity: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_3 \implies d_1 \sqsubseteq d_3$

antisymmetry: $d_1 \sqsubseteq d_2$ and $d_2 \sqsubseteq d_1 \implies d_1 = d_2$

It is called **total** if, in addition, always $d_1 \sqsubseteq d_2$ or $d_2 \sqsubseteq d_1$.

Example 20.2

- ❶ (Live Variables) $(2^{Var_c}, \subseteq)$ is a (non-total) partial order
- ❷ (Available Expressions) $(2^{AExp_c}, \supseteq)$ is a (non-total) partial order

Upper Bounds

In the dataflow equation system, analysis information from several predecessors is combined by taking the least upper bound.

Definition 20.3 ((Least) upper bound; repetition of Def. 7.4)

Let (D, \sqsubseteq) be a partial order and $S \subseteq D$.

- ① An element $d \in D$ is called an **upper bound** of S if $s \sqsubseteq d$ for every $s \in S$ (notation: $S \sqsubseteq d$).

Upper Bounds

In the dataflow equation system, analysis information from several predecessors is combined by taking the least upper bound.

Definition 20.3 ((Least) upper bound; repetition of Def. 7.4)

Let (D, \sqsubseteq) be a partial order and $S \subseteq D$.

- ① An element $d \in D$ is called an **upper bound** of S if $s \sqsubseteq d$ for every $s \in S$ (notation: $S \sqsubseteq d$).
- ② An upper bound d of S is called **least upper bound (LUB)** or **supremum** of S if $d \sqsubseteq d'$ for every upper bound d' of S (notation: $d = \sqcup S$).

Upper Bounds

In the dataflow equation system, analysis information from several predecessors is combined by taking the least upper bound.

Definition 20.3 ((Least) upper bound; repetition of Def. 7.4)

Let (D, \sqsubseteq) be a partial order and $S \subseteq D$.

- ① An element $d \in D$ is called an **upper bound** of S if $s \sqsubseteq d$ for every $s \in S$ (notation: $S \sqsubseteq d$).
- ② An upper bound d of S is called **least upper bound (LUB)** or **supremum** of S if $d \sqsubseteq d'$ for every upper bound d' of S (notation: $d = \sqcup S$).

Example 20.4

- ① (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$. Given $V_1, \dots, V_n \subseteq Var_c$,
$$\sqcup\{V_1, \dots, V_n\} = \bigcup\{V_1, \dots, V_n\}$$

Upper Bounds

In the dataflow equation system, analysis information from several predecessors is combined by taking the least upper bound.

Definition 20.3 ((Least) upper bound; repetition of Def. 7.4)

Let (D, \sqsubseteq) be a partial order and $S \subseteq D$.

- ① An element $d \in D$ is called an **upper bound** of S if $s \sqsubseteq d$ for every $s \in S$ (notation: $S \sqsubseteq d$).
- ② An upper bound d of S is called **least upper bound (LUB)** or **supremum** of S if $d \sqsubseteq d'$ for every upper bound d' of S (notation: $d = \sqcup S$).

Example 20.4

- ① (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$. Given $V_1, \dots, V_n \subseteq Var_c$,
$$\sqcup\{V_1, \dots, V_n\} = \bigcup\{V_1, \dots, V_n\}$$
- ② (Avail. Expr.) $(D, \sqsubseteq) = (2^{AExp_c}, \supseteq)$. Given $A_1, \dots, A_n \subseteq AExp_c$,
$$\sqcup\{A_1, \dots, A_n\} = \bigcap\{A_1, \dots, A_n\}$$

Complete Lattices

Since $\{\varphi_{l'}(\text{Al}_{l'}) \mid (l', l) \in F\}$ is not necessarily a chain (Def. 7.4), chain completeness (Def. 7.6) is not sufficient for guaranteeing the well-definedness of the equation system. A stronger property is required:

Definition 20.5 (Complete lattice)

A **complete lattice** is a partial order (D, \sqsubseteq) such that all subsets of D have least upper bounds. In this case,

$$\perp := \bigsqcup \emptyset$$

denotes the **least element** of D .

Complete Lattices

Since $\{\varphi_{l'}(\text{Al}_{l'}) \mid (l', l) \in F\}$ is not necessarily a chain (Def. 7.4), chain completeness (Def. 7.6) is not sufficient for guaranteeing the well-definedness of the equation system. A stronger property is required:

Definition 20.5 (Complete lattice)

A **complete lattice** is a partial order (D, \sqsubseteq) such that all subsets of D have least upper bounds. In this case,

$$\perp := \bigsqcup \emptyset$$

denotes the **least element** of D .

Example 20.6

① (Live Variables)

$(D, \sqsubseteq) = (2^{Var_c}, \sqsubseteq)$ is a complete lattice with $\perp = \emptyset$

Complete Lattices

Since $\{\varphi_{l'}(\text{Al}_{l'}) \mid (l', l) \in F\}$ is not necessarily a chain (Def. 7.4), chain completeness (Def. 7.6) is not sufficient for guaranteeing the well-definedness of the equation system. A stronger property is required:

Definition 20.5 (Complete lattice)

A **complete lattice** is a partial order (D, \sqsubseteq) such that all subsets of D have least upper bounds. In this case,

$$\perp := \bigsqcup \emptyset$$

denotes the **least element** of D .

Example 20.6

① (Live Variables)

$(D, \sqsubseteq) = (2^{Var_c}, \sqsubseteq)$ is a complete lattice with $\perp = \emptyset$

② (Available Expressions)

$(D, \sqsubseteq) = (2^{AExp_c}, \sqsupseteq)$ is a complete lattice with $\perp = AExp_c$

- Dual concept of least upper bound: greatest lower bound
- **Definitions:**
 - An element $d \in D$ is called a **lower bound** of $S \subseteq D$ if $d \sqsubseteq s$ for every $s \in S$ (notation: $d \sqsubseteq S$).
 - A lower bound d is called **greatest lower bound (GLB)** or **infimum** of S if $d' \sqsubseteq d$ for every lower bound d' of S (notation: $d = \sqcap S$).

- Dual concept of least upper bound: greatest lower bound
- **Definitions:**
 - An element $d \in D$ is called a **lower bound** of $S \subseteq D$ if $d \sqsubseteq s$ for every $s \in S$ (notation: $d \sqsubseteq S$).
 - A lower bound d is called **greatest lower bound (GLB)** or **infimum** of S if $d' \sqsubseteq d$ for every lower bound d' of S (notation: $d = \sqcap S$).
- **Examples:**
 - (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \sqsubseteq)$,
 $\sqcap\{V_1, \dots, V_n\} = \bigcap\{V_1, \dots, V_n\}$
 - (Available Expressions) $(D, \sqsubseteq) = (2^{Exp_c}, \supseteq)$,
 $\sqcap\{A_1, \dots, A_n\} = \bigcup\{A_1, \dots, A_n\}$

- Dual concept of least upper bound: greatest lower bound
- **Definitions:**
 - An element $d \in D$ is called a **lower bound** of $S \subseteq D$ if $d \sqsubseteq s$ for every $s \in S$ (notation: $d \sqsubseteq S$).
 - A lower bound d is called **greatest lower bound (GLB)** or **infimum** of S if $d' \sqsubseteq d$ for every lower bound d' of S (notation: $d = \sqcap S$).
- **Examples:**
 - (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \sqsubseteq)$,
 $\sqcap\{V_1, \dots, V_n\} = \bigcap\{V_1, \dots, V_n\}$
 - (Available Expressions) $(D, \sqsubseteq) = (2^{Exp_c}, \supseteq)$,
 $\sqcap\{A_1, \dots, A_n\} = \bigcup\{A_1, \dots, A_n\}$
- **Lemma:** the following are equivalent:
 - (D, \sqsubseteq) is a complete lattice
(i.e., every subset of D has a least upper bound)
 - Every subset of D has a greatest lower bound

- Dual concept of least upper bound: greatest lower bound
- **Definitions:**
 - An element $d \in D$ is called a **lower bound** of $S \subseteq D$ if $d \sqsubseteq s$ for every $s \in S$ (notation: $d \sqsubseteq S$).
 - A lower bound d is called **greatest lower bound (GLB)** or **infimum** of S if $d' \sqsubseteq d$ for every lower bound d' of S (notation: $d = \sqcap S$).
- **Examples:**
 - (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \sqsubseteq)$,
 $\sqcap\{V_1, \dots, V_n\} = \bigcap\{V_1, \dots, V_n\}$
 - (Available Expressions) $(D, \sqsubseteq) = (2^{Exp_c}, \supseteq)$,
 $\sqcap\{A_1, \dots, A_n\} = \bigcup\{A_1, \dots, A_n\}$
- **Lemma:** the following are equivalent:
 - (D, \sqsubseteq) is a complete lattice
(i.e., every subset of D has a least upper bound)
 - Every subset of D has a greatest lower bound
- **Corollary:** every complete lattice has a greatest element $\top := \sqcap \emptyset$

Chains are generated by the approximation of the analysis information in the fixpoint iteration.

Definition 20.7 (Chain; repetition of Def. 7.4 and 7.6)

Let (D, \sqsubseteq) be a partial order.

- ① A subset $S \subseteq D$ is called a **chain** in D if, for every $s_1, s_2 \in S$,

$$s_1 \sqsubseteq s_2 \text{ or } s_2 \sqsubseteq s_1$$

(that is, S is a totally ordered subset of D).

Chains are generated by the approximation of the analysis information in the fixpoint iteration.

Definition 20.7 (Chain; repetition of Def. 7.4 and 7.6)

Let (D, \sqsubseteq) be a partial order.

- ① A subset $S \subseteq D$ is called a **chain** in D if, for every $s_1, s_2 \in S$,
$$s_1 \sqsubseteq s_2 \text{ or } s_2 \sqsubseteq s_1$$
(that is, S is a totally ordered subset of D).
- ② (D, \sqsubseteq) is called **chain complete (CCPO)** if each of its chains has a least upper bound.

Chains are generated by the approximation of the analysis information in the fixpoint iteration.

Definition 20.7 (Chain; repetition of Def. 7.4 and 7.6)

Let (D, \sqsubseteq) be a partial order.

- ① A subset $S \subseteq D$ is called a **chain** in D if, for every $s_1, s_2 \in S$,
$$s_1 \sqsubseteq s_2 \text{ or } s_2 \sqsubseteq s_1$$
(that is, S is a totally ordered subset of D).
- ② (D, \sqsubseteq) is called **chain complete (CCPO)** if each of its chains has a least upper bound.

Corollary 20.8

Every complete lattice is a CCPO.

Monotonicity of Functions

The monotonicity of transfer functions (which formalize the impact of a block in the program on the analysis information) excludes “oscillating behavior” in fixpoint iteration.

Definition 20.9 (Monotonicity; repetition of Def. 8.1)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be partial orders, and let $F : D \rightarrow D'$. F is called **monotonic (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq'))** if, for every $d_1, d_2 \in D$,

$$d_1 \sqsubseteq d_2 \implies F(d_1) \sqsubseteq' F(d_2).$$

Monotonicity of Functions

The monotonicity of transfer functions (which formalize the impact of a block in the program on the analysis information) excludes “oscillating behavior” in fixpoint iteration.

Definition 20.9 (Monotonicity; repetition of Def. 8.1)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be partial orders, and let $F : D \rightarrow D'$. F is called **monotonic (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq'))** if, for every $d_1, d_2 \in D$,

$$d_1 \sqsubseteq d_2 \implies F(d_1) \sqsubseteq' F(d_2).$$

Example 20.10

① (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$

Each transfer function $\varphi_{l'}(V) := (V \setminus \text{kill}_{\text{LV}}(B^{l'})) \cup \text{gen}_{\text{LV}}(B^{l'})$ is obviously monotonic

Monotonicity of Functions

The monotonicity of transfer functions (which formalize the impact of a block in the program on the analysis information) excludes “oscillating behavior” in fixpoint iteration.

Definition 20.9 (Monotonicity; repetition of Def. 8.1)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be partial orders, and let $F : D \rightarrow D'$. F is called **monotonic (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq'))** if, for every $d_1, d_2 \in D$,

$$d_1 \sqsubseteq d_2 \implies F(d_1) \sqsubseteq' F(d_2).$$

Example 20.10

- ① (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$

Each transfer function $\varphi_{l'}(V) := (V \setminus \text{kill}_{\text{LV}}(B^{l'})) \cup \text{gen}_{\text{LV}}(B^{l'})$ is obviously monotonic

- ② (Available Expressions) $(D, \sqsubseteq) = (2^{AExp_c}, \supseteq)$

Each transfer function $\varphi_{l'}(A) := (A \setminus \text{kill}_{\text{AE}}(B^{l'})) \cup \text{gen}_{\text{AE}}(B^{l'})$ is obviously monotonic

Termination of fixpoint iteration is guaranteed by the Ascending Chain Condition.

Definition 20.11 (Ascending Chain Condition)

A partial order (D, \sqsubseteq) satisfies the **Ascending Chain Condition (ACC)** if each ascending chain $d_1 \sqsubseteq d_2 \sqsubseteq \dots$ eventually stabilizes, i.e., there exists $n \in \mathbb{N}$ such that $d_n = d_{n+1} = \dots$

The Ascending Chain Condition

Termination of fixpoint iteration is guaranteed by the Ascending Chain Condition.

Definition 20.11 (Ascending Chain Condition)

A partial order (D, \sqsubseteq) satisfies the **Ascending Chain Condition (ACC)** if each ascending chain $d_1 \sqsubseteq d_2 \sqsubseteq \dots$ eventually stabilizes, i.e., there exists $n \in \mathbb{N}$ such that $d_n = d_{n+1} = \dots$

Example 20.12

- ① (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$ satisfies ACC since Var_c (unlike Var) is finite

Termination of fixpoint iteration is guaranteed by the Ascending Chain Condition.

Definition 20.11 (Ascending Chain Condition)

A partial order (D, \sqsubseteq) satisfies the **Ascending Chain Condition (ACC)** if each ascending chain $d_1 \sqsubseteq d_2 \sqsubseteq \dots$ eventually stabilizes, i.e., there exists $n \in \mathbb{N}$ such that $d_n = d_{n+1} = \dots$

Example 20.12

- ❶ (Live Variables) $(D, \sqsubseteq) = (2^{Var_c}, \subseteq)$ satisfies ACC since Var_c (unlike Var) is finite
- ❷ (Available Expressions) $(D, \sqsubseteq) = (2^{AExp_c}, \supseteq)$ satisfies ACC since $AExp_c$ (unlike $AExp$) is finite

Theorem 20.13 (Fixpoint Theorem; repetition of Thm. 8.7)

Let (D, \sqsubseteq) be a CCPo and $F : D \rightarrow D$ continuous. Then

$$\text{fix}(F) := \bigsqcup \{F^n(\bigsqcup \emptyset) \mid n \in \mathbb{N}\}$$

is the least fixpoint of F .

Theorem 20.13 (Fixpoint Theorem; repetition of Thm. 8.7)

Let (D, \sqsubseteq) be a CCPO and $F : D \rightarrow D$ continuous. Then

$$\text{fix}(F) := \bigsqcup \{F^n(\bigsqcup \emptyset) \mid n \in \mathbb{N}\}$$

is the least fixpoint of F .

Definition 20.14 (Continuity; repetition of Def. 8.5)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be CCPDs and $F : D \rightarrow D'$ monotonic. Then F is called **continuous** (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq')) if, for every non-empty chain $S \subseteq D$,

$$F(\bigsqcup S) = \bigsqcup F(S).$$

Fixpoints

Theorem 20.13 (Fixpoint Theorem; repetition of Thm. 8.7)

Let (D, \sqsubseteq) be a CCPO and $F : D \rightarrow D$ continuous. Then

$$\text{fix}(F) := \bigsqcup \{F^n(\bigsqcup \emptyset) \mid n \in \mathbb{N}\}$$

is the least fixpoint of F .

Definition 20.14 (Continuity; repetition of Def. 8.5)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be CCPs and $F : D \rightarrow D'$ monotonic. Then F is called **continuous** (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq')) if, for every non-empty chain $S \subseteq D$,

$$F(\bigsqcup S) = \bigsqcup F(S).$$

Corollary 20.15

Monotonic functions on partial orders that satisfy ACC are continuous.

Fixpoints

Theorem 20.13 (Fixpoint Theorem; repetition of Thm. 8.7)

Let (D, \sqsubseteq) be a CCPo and $F : D \rightarrow D$ continuous. Then

$$\text{fix}(F) := \bigsqcup \{F^n(\bigsqcup \emptyset) \mid n \in \mathbb{N}\}$$

is the least fixpoint of F .

Definition 20.14 (Continuity; repetition of Def. 8.5)

Let (D, \sqsubseteq) and (D', \sqsubseteq') be CCPOs and $F : D \rightarrow D'$ monotonic. Then F is called **continuous** (w.r.t. (D, \sqsubseteq) and (D', \sqsubseteq')) if, for every non-empty chain $S \subseteq D$,

$$F(\bigsqcup S) = \bigsqcup F(S).$$

Corollary 20.15

Monotonic functions on partial orders that satisfy ACC are continuous.

Proof.

on the board



- 1 Repetition: Heading for a Dataflow Analysis Framework
- 2 Order-Theoretic Foundations
- 3 The Framework
- 4 Solving Dataflow Equation Systems

Definition 20.16 (Dataflow system)

A **dataflow system** $S = (L, E, F, (D, \sqsubseteq), \iota, \varphi)$ consists of

- a finite set of (program) **labels** L (here: L_c),
- a set of **extremal labels** $E \subseteq L$ (here: $\{\text{init}(c)\}$ or $\text{final}(c)$),
- a **flow relation** $F \subseteq L \times L$ (here: $\text{flow}(c)$ or $\text{flow}^R(c)$),
- a **complete lattice** (D, \sqsubseteq) that satisfies ACC
(with LUB operator \sqcup and least element \perp),
- an **extremal value** $\iota \in D$ (for the extremal labels), and
- a collection of **monotonic transfer functions** $\{\varphi_l \mid l \in L\}$ of type $\varphi_l : D \rightarrow D$.

Example 20.17

Problem	Available Expressions	Live Variables
E	$\{\text{init}(c)\}$	$\text{final}(c)$
F	$\text{flow}(c)$	$\text{flow}^R(c)$
D	2^{AExp_c}	2^{Var_c}
\sqsubseteq	\supseteq	\subseteq
\sqcup	\bigcap	\bigcup
\perp	$AExp_c$	\emptyset
ι	\emptyset	Var_c
φ_l	$\varphi_l(d) = (d \setminus \text{kill}(B^l)) \cup \text{gen}(B^l)$	