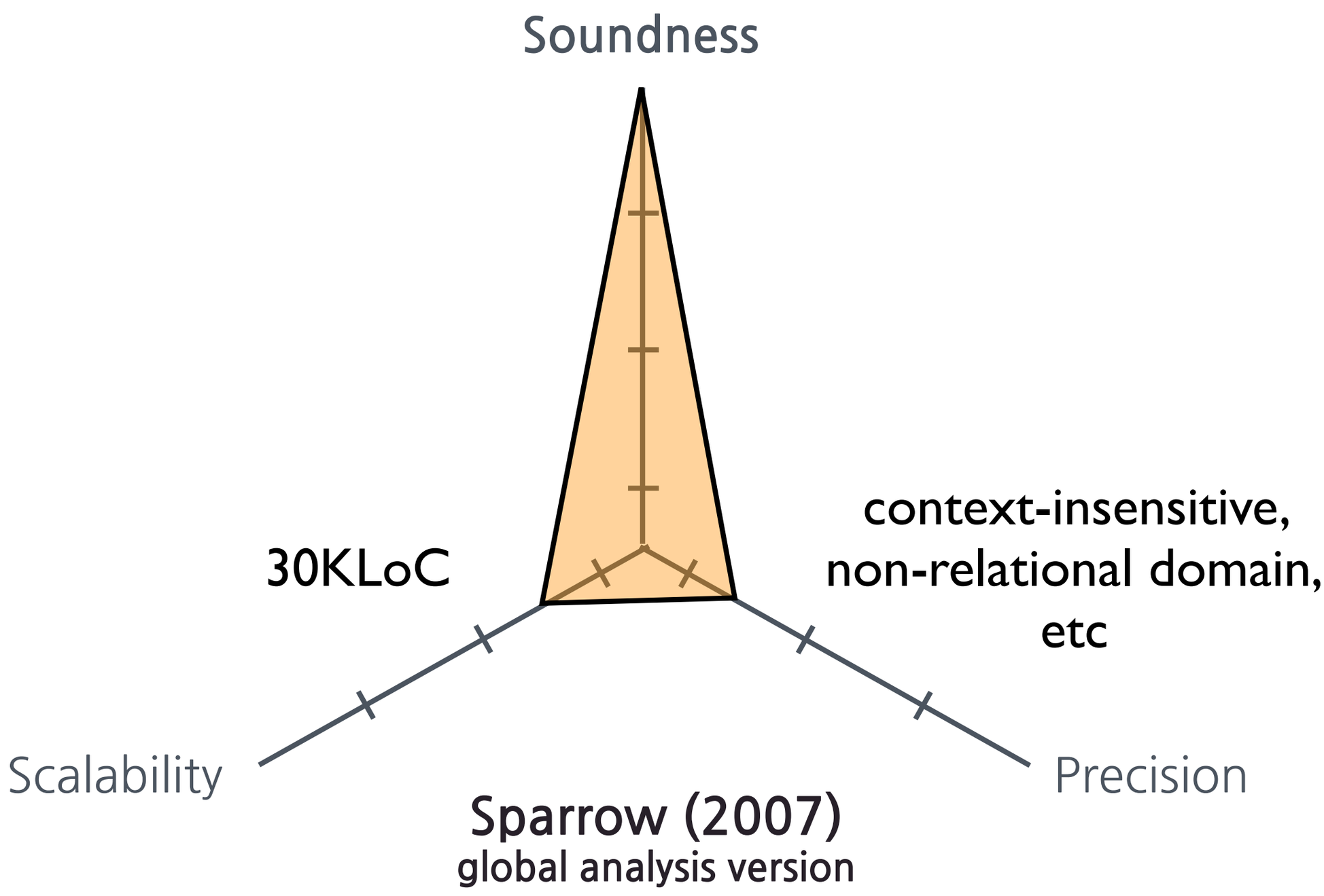
Selective-X Analysis Guided by Impact Pre-Analysis

Hakjoo Oh[1] Wonchan Lee[1] Kihong Heo[1]
Hongseok Yang[2] Kwangkeun Yi[1]

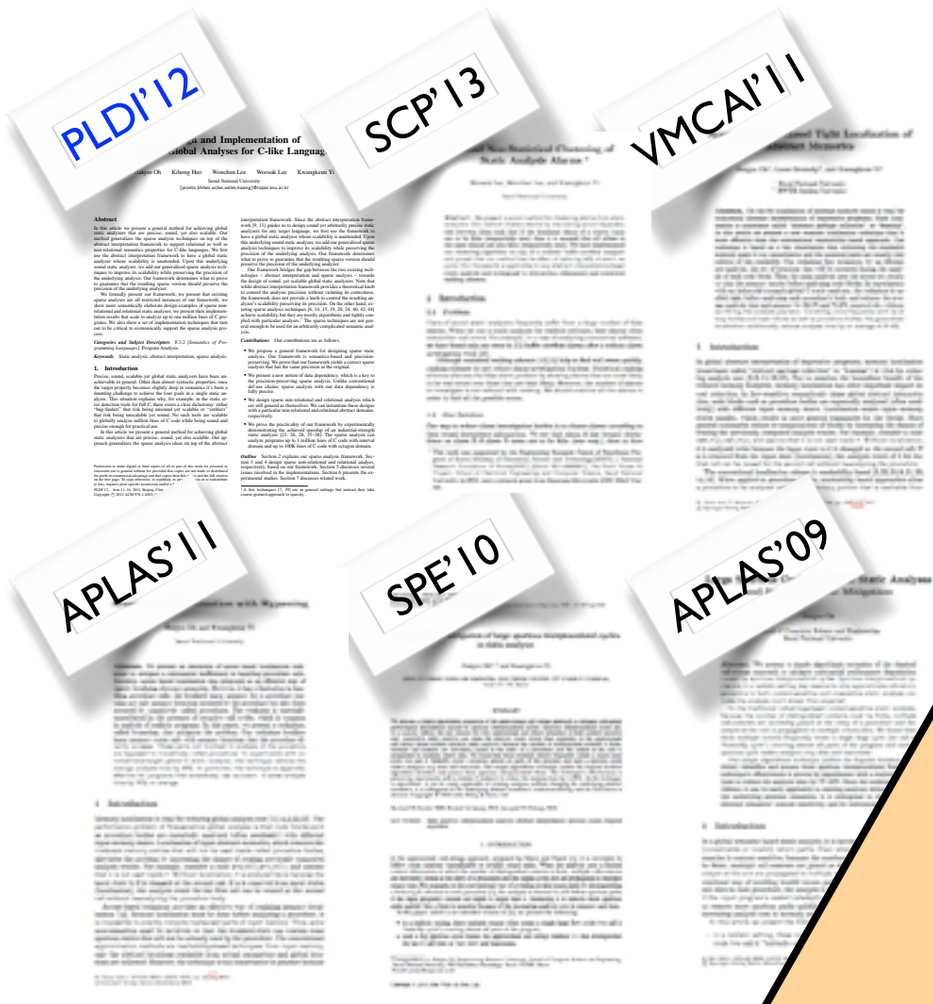
[1]Seoul National University
[2]University of Oxford

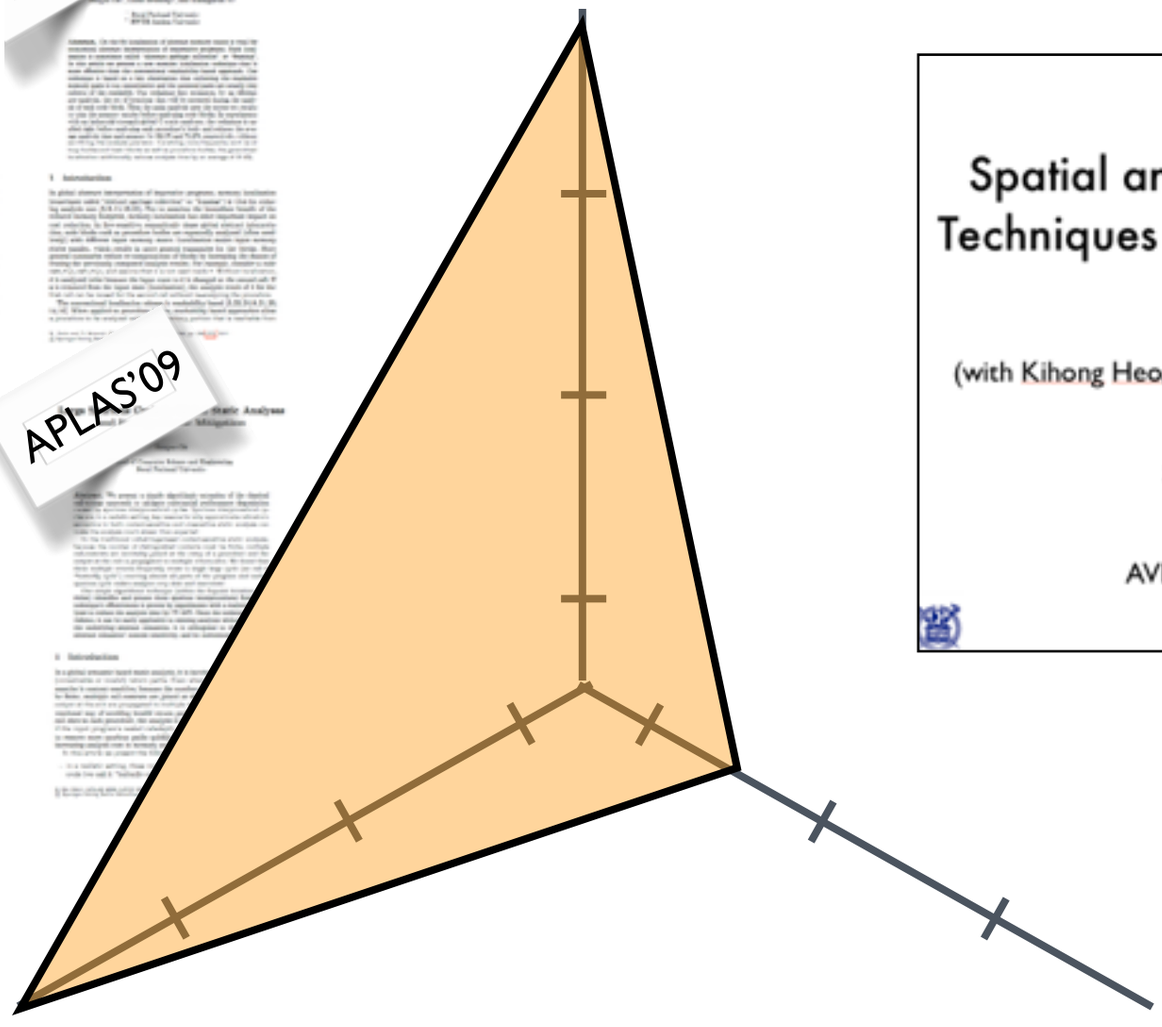
Dec. 23, 2013 @SNU

Motivation


Soundness
30KLoC
context-insensitive,
non-relational domain,
etc
Scalability
Precision
Sparrow (2007)
global analysis version

Motivation





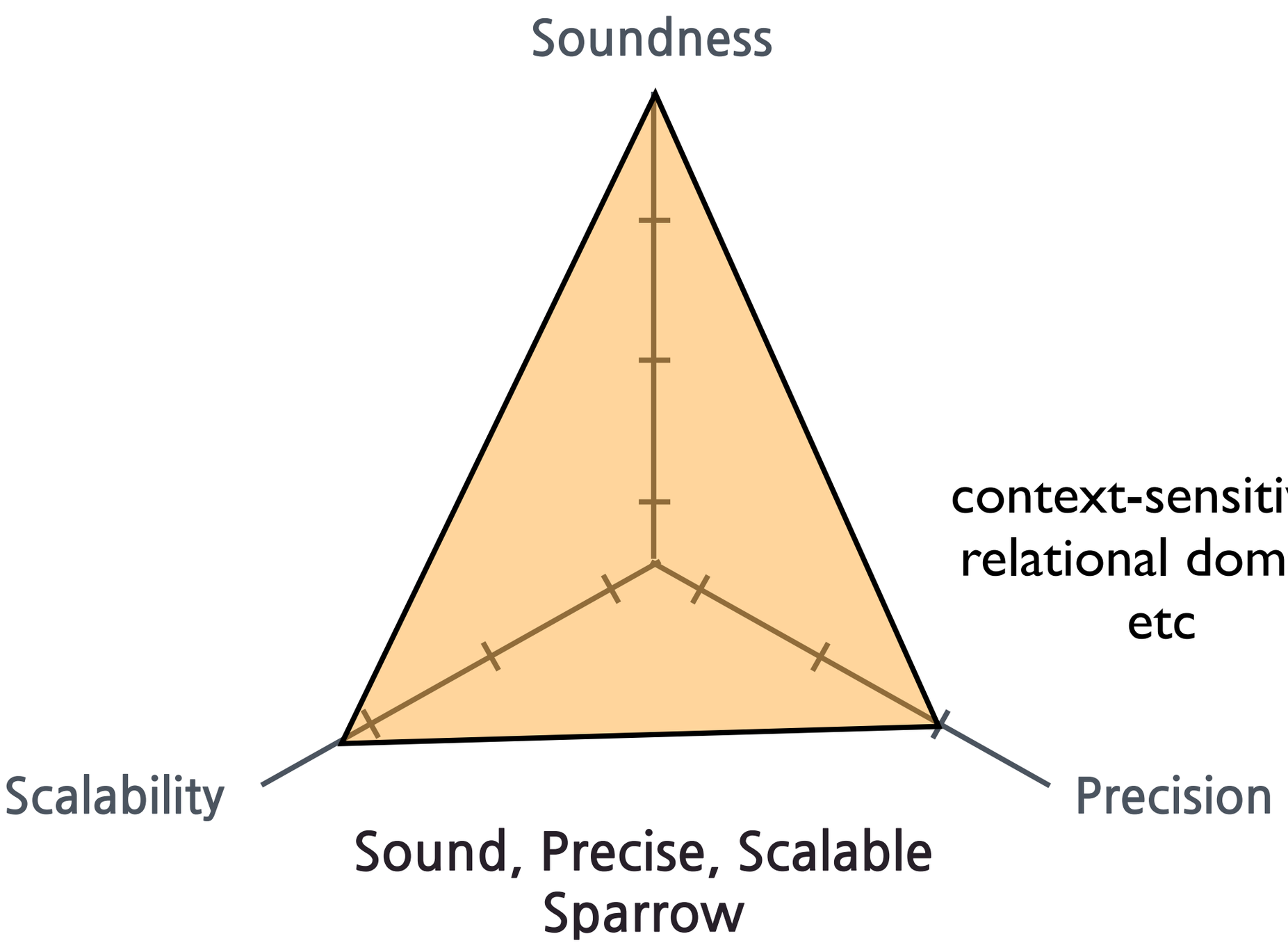
Spatial and Temporal Localization Techniques for Global Static Analysis

Hakjoo Oh
(with Kihong Heo, Wonchan Lee, Woosuk Lee, Kwangkeun Yi)

ROSAEC Center
Seoul National University

Oct. 26, 2012
AVDCPS 2012, Changsha, China

Motivation



context-sensitivity,
relational domain,
etc

Selective-X Analysis Guided by Impact Pre-Analysis

- Selectively apply a higher precision only when it is likely to benefit the final analysis results
- The selection is guided by an impact pre-analysis
- Two instances
  - selective context-sensitive analysis
  - selective relational analysis

Selective Context-Sensitive Analysis

Example Program

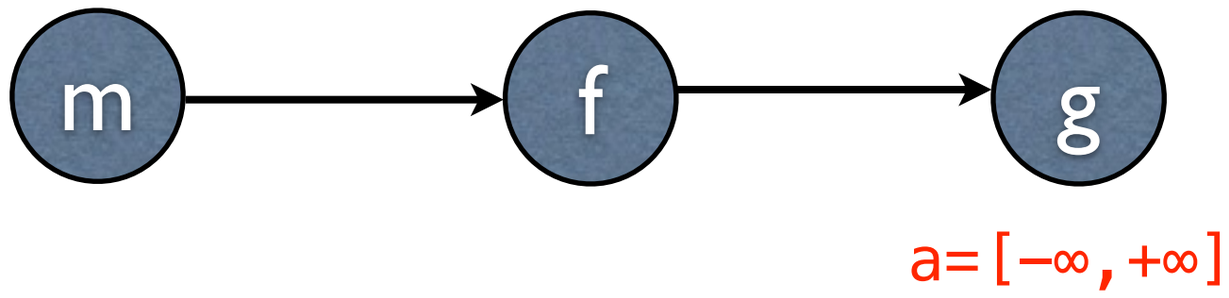
```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }
```

```
     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Query1
c5:    b = g(input());
       assert (b > 1); // Query2
     }

     void g(int a) {
       return a;
     }
```

Context-Insensitive Analysis

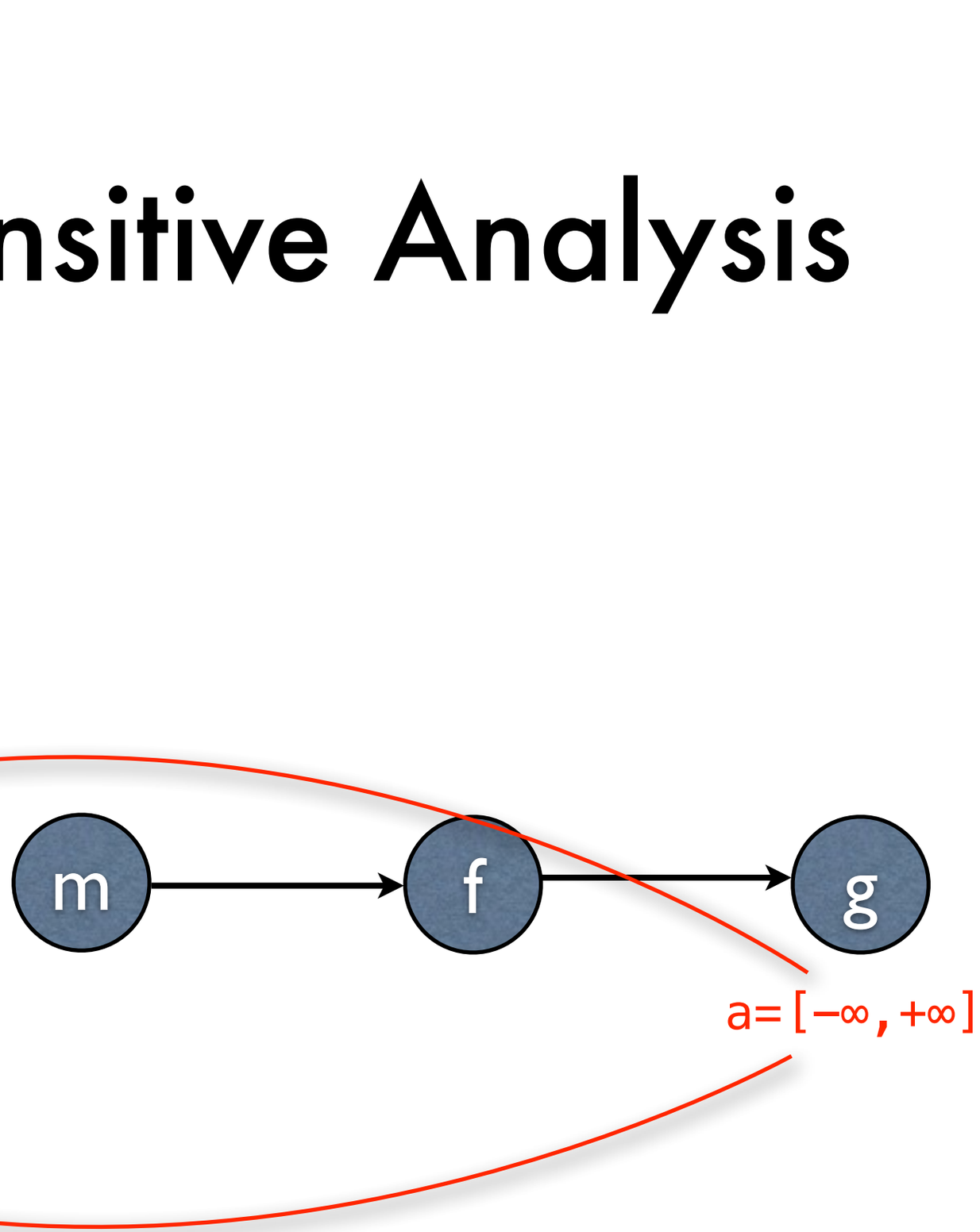
```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }
```

```
     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    b = g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



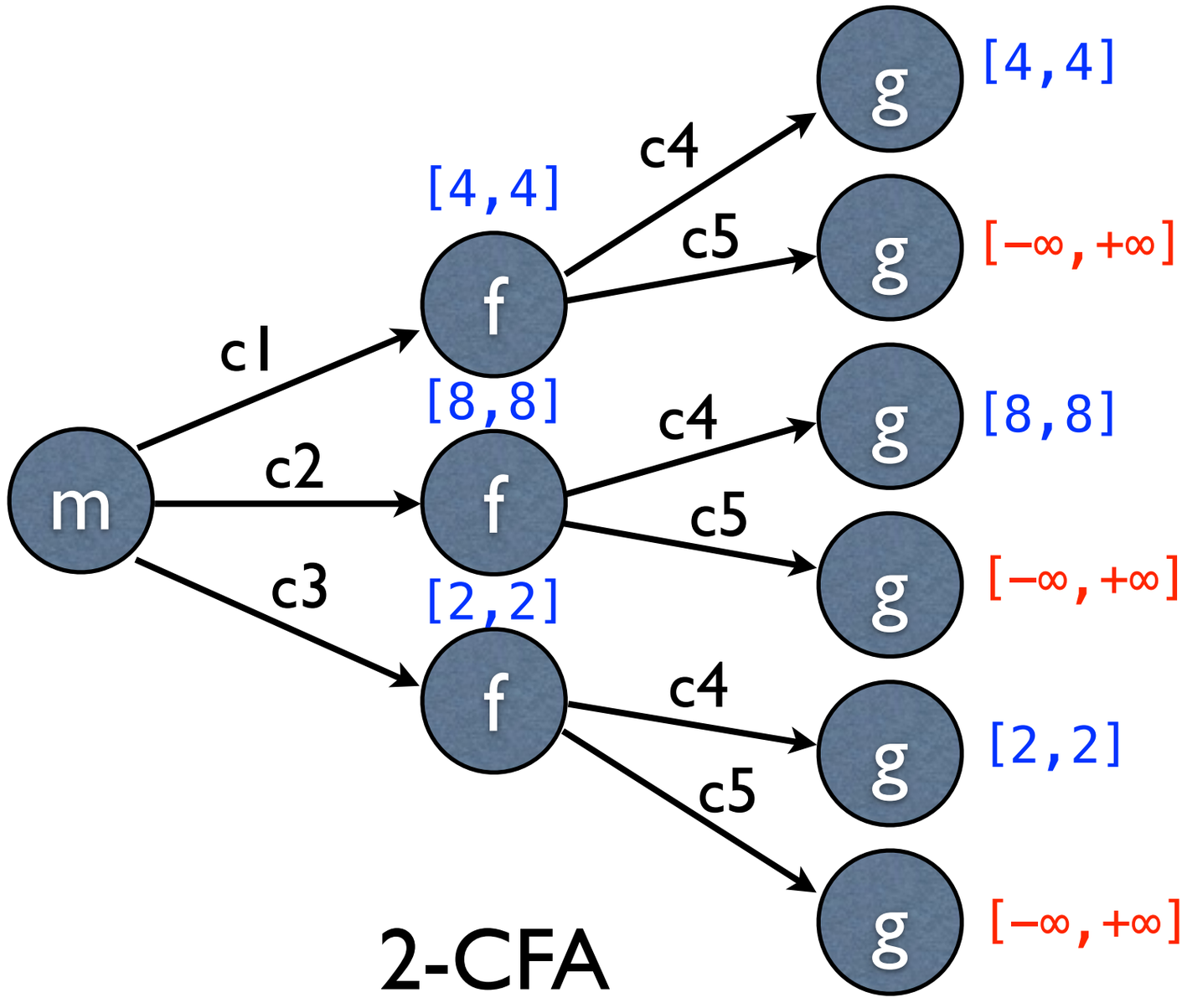
Context-Insensitive Analysis

```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    b = g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



Context-Sensitive Analysis

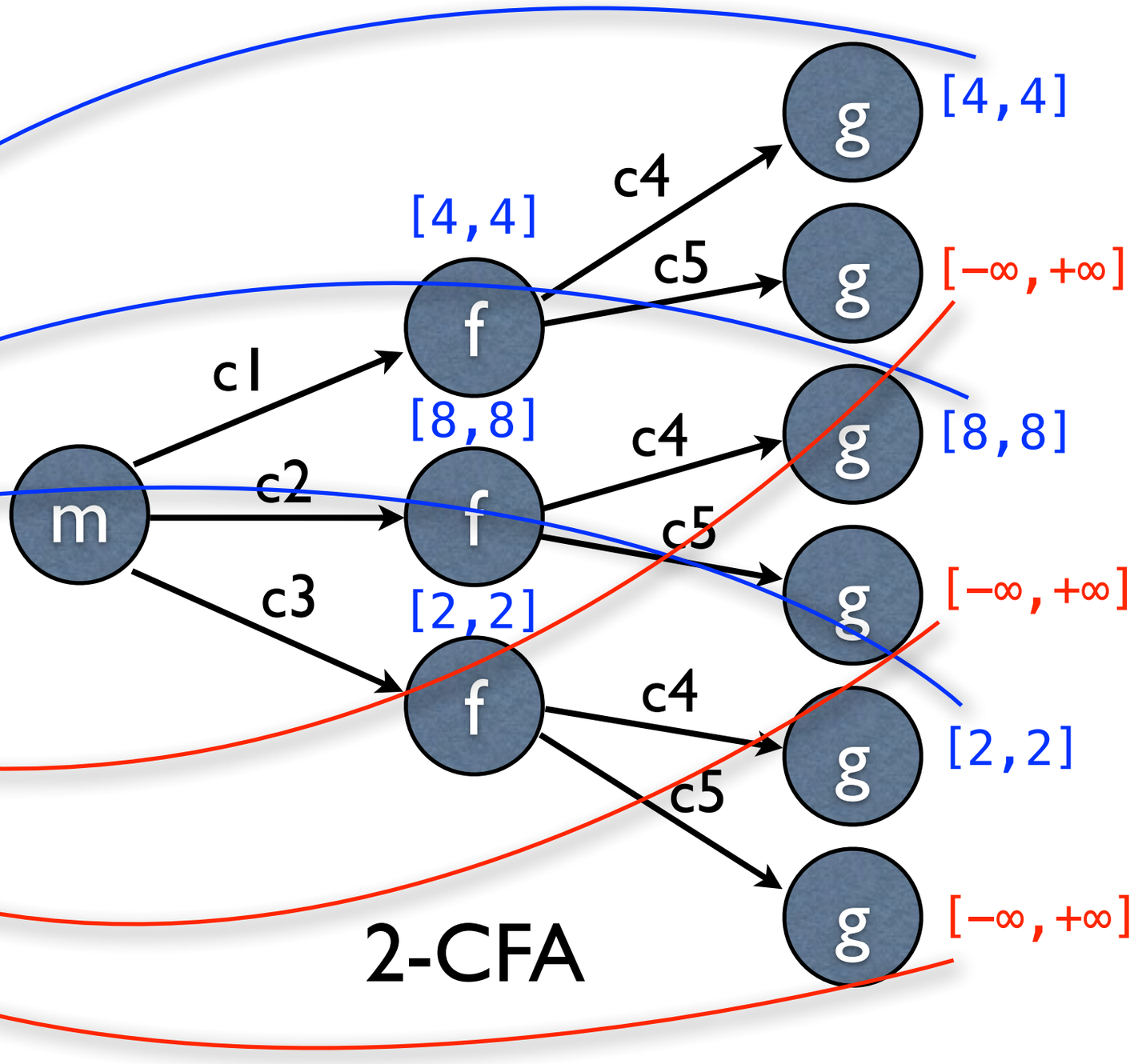
```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



2-CFA

Context-Sensitive Analysis

```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```

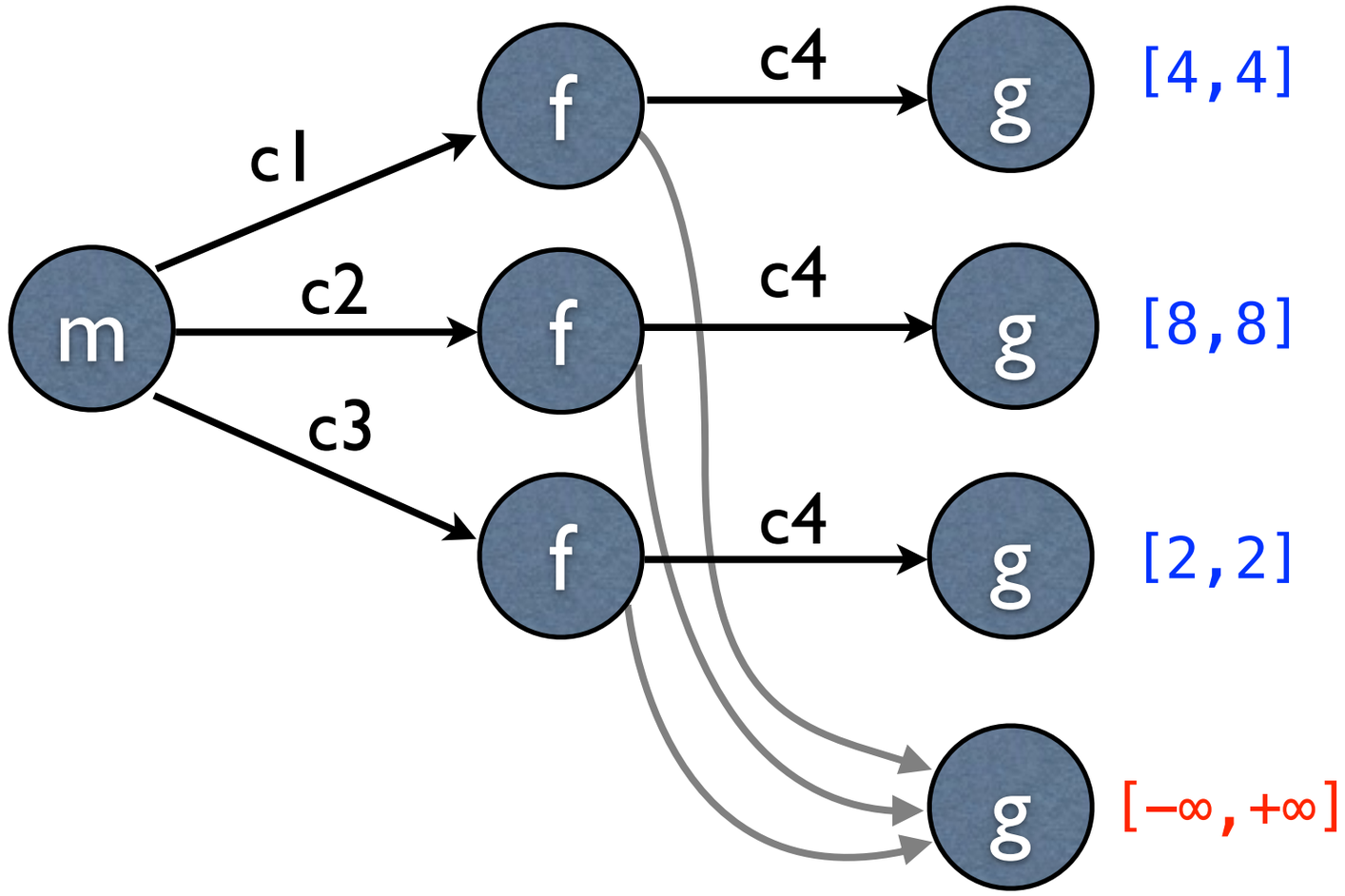


Unnecessarily precise

- Thus, too expensive
  - ex) Sparrow with 3-CFA does not stop after 30min for 10K programs

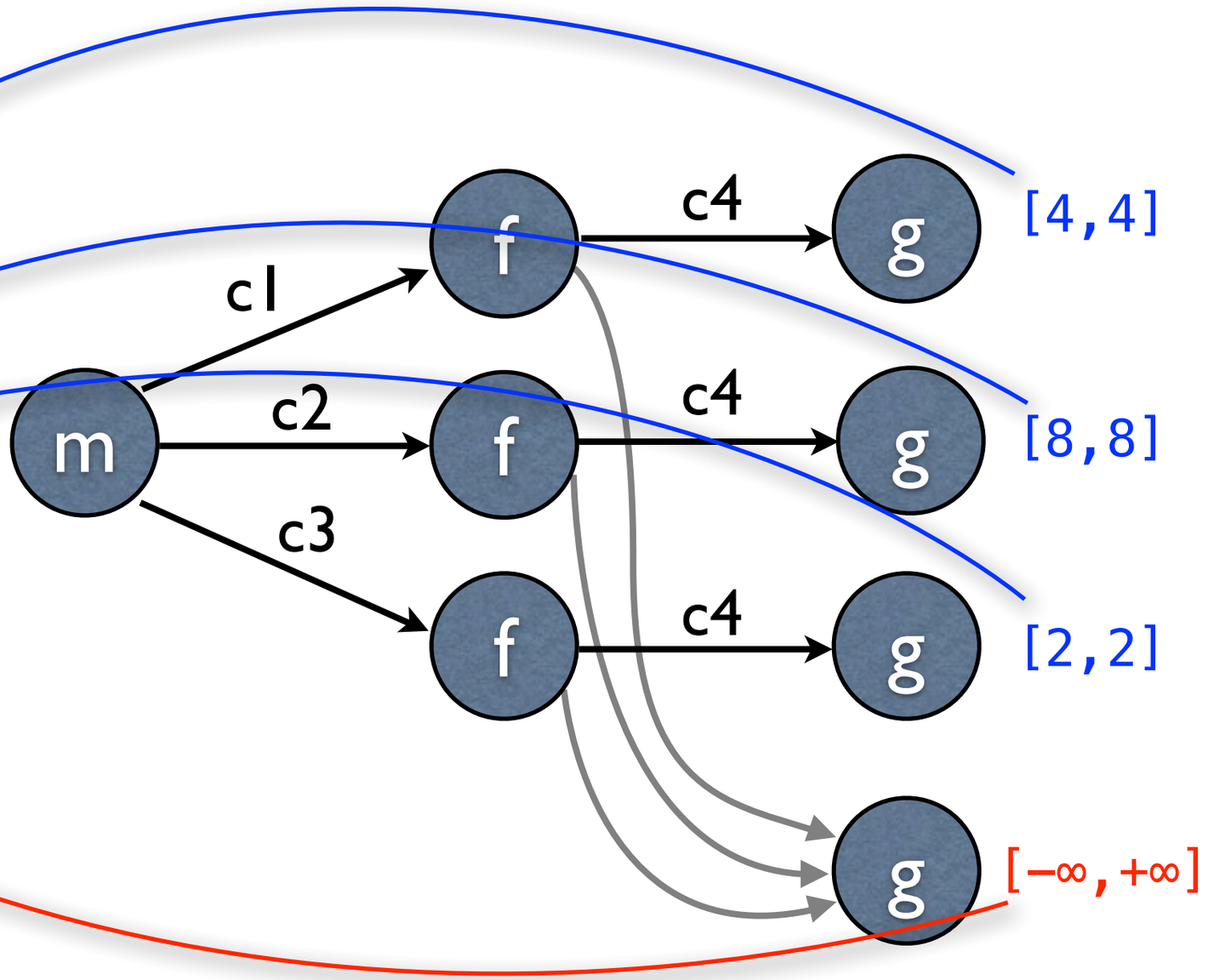
Selective Context-Sensitive Analysis

```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



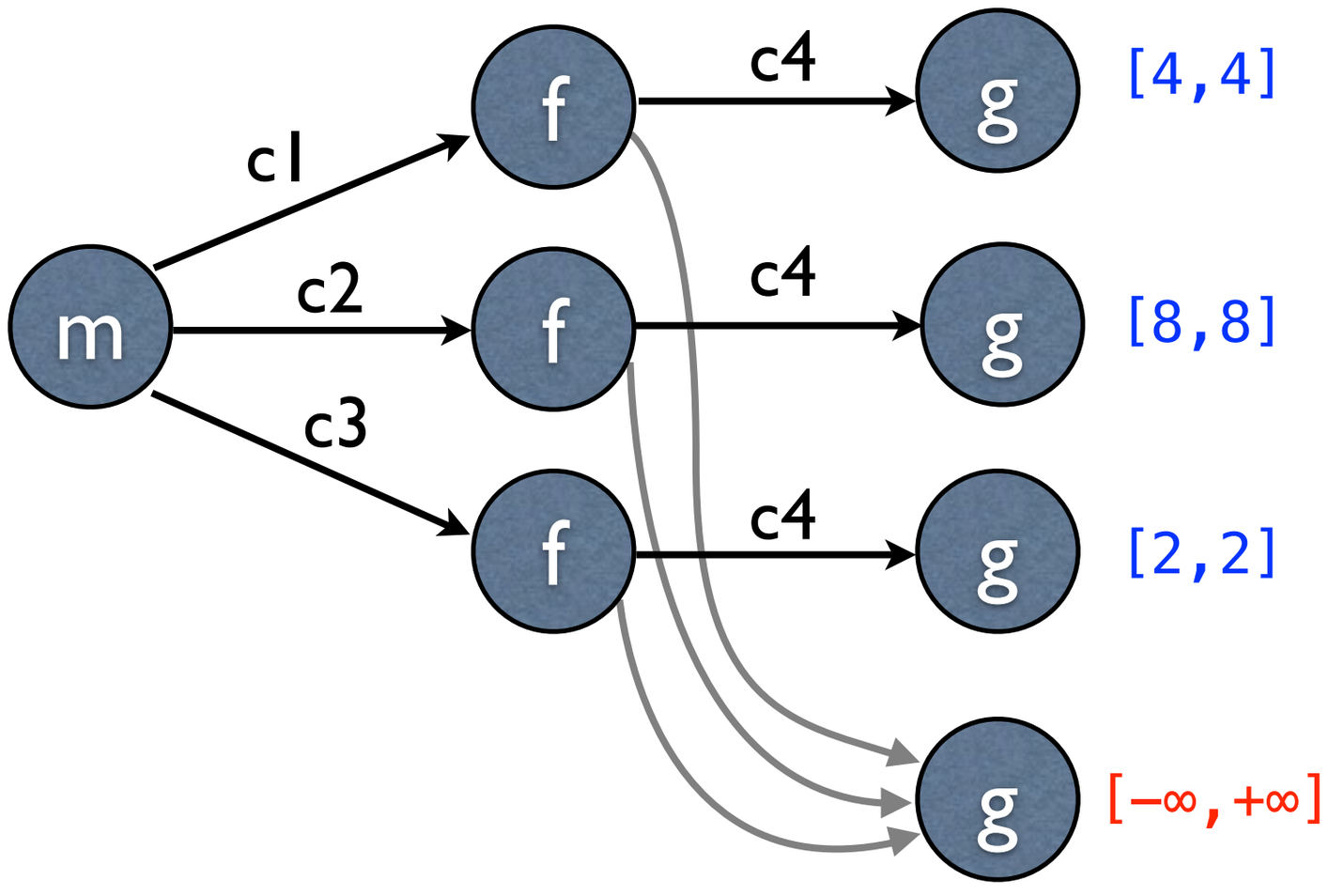
Selective Context-Sensitive Analysis

```
    void main() {
c1:   f(4);
c2:   f(8);
c3:   f(2);
    }

    void f(int a) {
c4:   b = g(a);
      assert (b > 1); // Q1
c5:   g(input());
      assert (b > 1); // Q2
    }

    void g(int a) {
      return a;
    }
```



Problem

- How to select the contexts?



Our Solution: Impact Pre-Analysis

- An over-approximation of the fully context-sensitive main analysis
- An impact pre-analysis for interval analysis
  - abstract domain: approximation of intervals

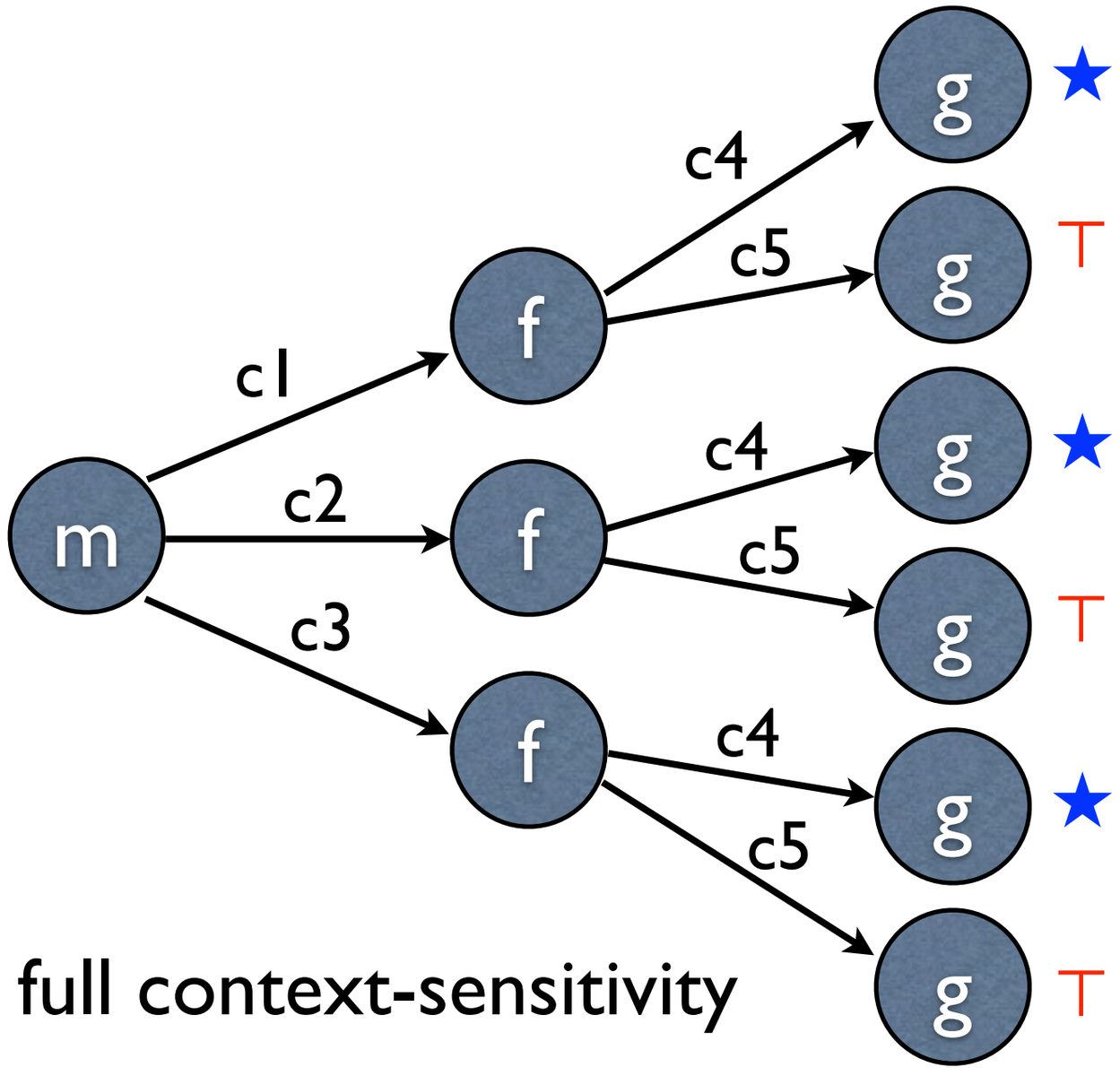
$$\mathbb{V} = \{\bot_v, \bigstar, \top_v\}$$

$$\gamma_v(\bigstar) = \{[a, b] \in \mathbb{I} \mid 0 \leq a\},$$

  - instead, fully context-sensitive

Example

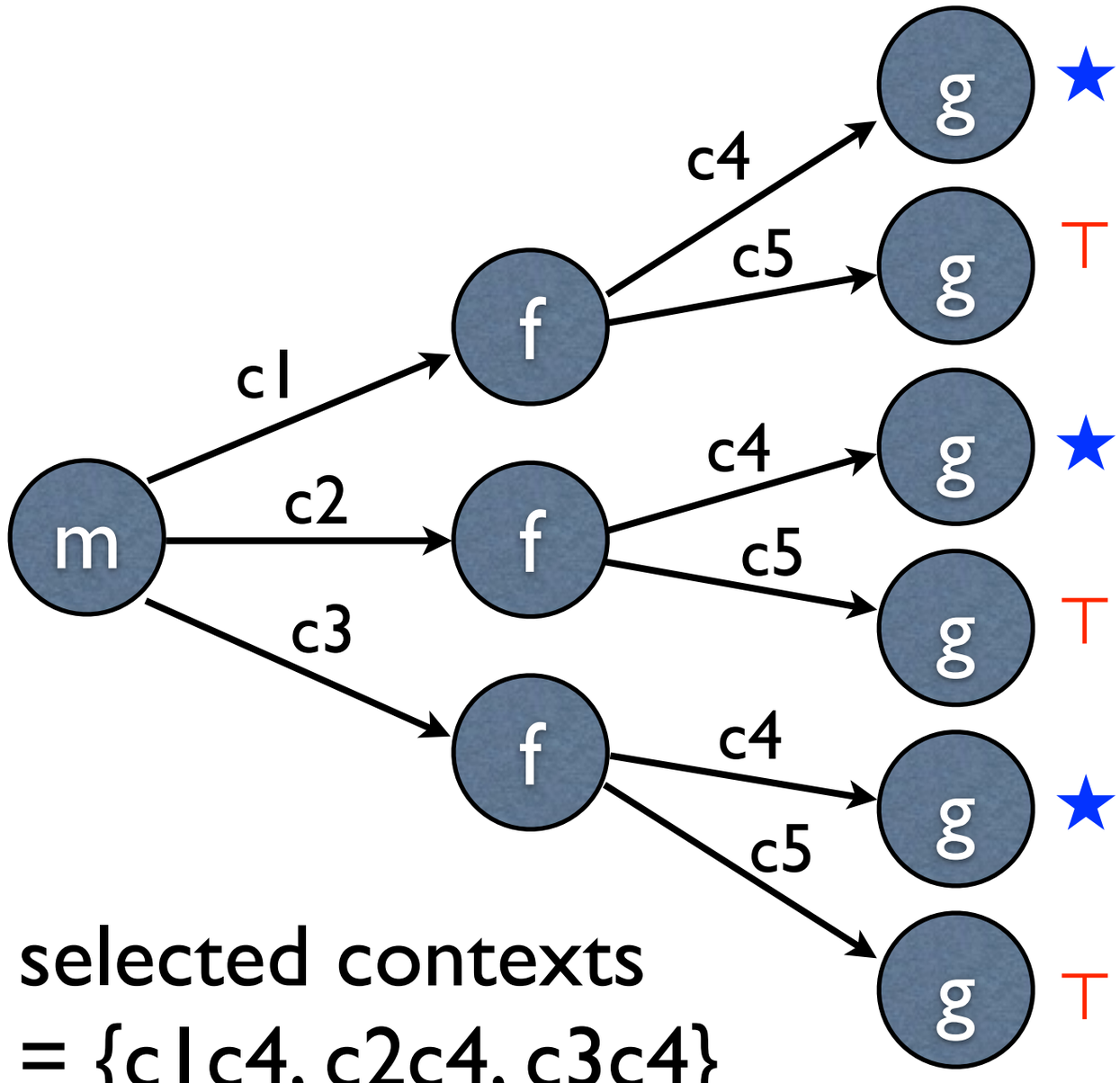
```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



full context-sensitivity

Example

```
     void main() {
c1:    f(4);
c2:    f(8);
c3:    f(2);
     }

     void f(int a) {
c4:    b = g(a);
       assert (b > 1); // Q1
c5:    g(input());
       assert (b > 1); // Q2
     }

     void g(int a) {
       return a;
     }
```



selected contexts
= {c1c4, c2c4, c3c4}

Experiments

Program	LOC	Baseline		Our Selectively Context-Sensitive Analysis						imprvd	overhead$_1$	overhead$_2$
		#alarm	time	#alarm	pre	main	total	#selected call-sites	depth			
spell-1.0	2,213	58	0.6	30	0.1	0.8	0.9	25 / 124 (20.2 %)	1.08 (3)	48.3%	16.7%	33.3%
bc-1.06	13,093	606	14.0	483	1.9	14.3	16.2	29 / 777 (3.7 %)	1.16 (2)	20.3%	13.6%	2.1%
tar-1.17	20,258	940	42.1	799	5.4	41.8	47.2	51 / 1213 (4.2 %)	1.02 (3)	15.0%	12.8%	−0.7%
less-382	23,822	654	123.0	562	3.3	163.1	166.4	51 / 1,522 (3.4 %)	1.71 (4)	14.1%	2.7%	32.6%
sed-4.0.8	26,807	1,325	107.5	1,238	7.4	110.2	117.6	25 / 868 (2.9 %)	1.4 (3)	6.6%	6.9%	2.5%
make-3.76.1	27,304	1,500	84.4	1,028	7.1	99.1	106.2	67 / 1,050 (6.4 %)	1.20 (2)	31.5%	8.4%	17.4%
grep-2.5	31,495	735	12.1	653	2.4	13.5	15.9	33 / 530 (6.2 %)	1.16 (3)	11.2%	19.8%	11.6%
wget-1.9	35,018	1,307	69.0	942	12.5	69.6	82.1	79 / 1,973 (4.0 %)	1.39 (5)	27.9%	18.1%	0.9%
a2ps-4.14	64,590	3,682	118.1	2,121	29.5	148.2	177.7	237 / 2,450 (9.7 %)	2.20 (9)	42.4%	25.0%	25.5%
bison-2.5	101,807	1,894	136.3	1,742	34.6	138.8	173.4	173 / 2,038 (8.5 %)	1.54 (4)	8.0%	25.4%	1.8%
Total	346,407	12,701	707.1	9,598	104.2	799.4	903.6	770 / 12,545 (6.1 %)		**24.4%**	**14.7%**	**13.1%**

24.4% reduction with 27.9% overhead

Summary

- Selective context-sensitivity guided by impact pre-analysis
- General idea for other selective analyses
  - selective relational analysis with octagons
  - selective flow-sensitive analysis
  - etc