

COSE419: Software Verification

Lecture 14 – Pointer Analysis

Hakjoo Oh
2024 Spring

Need for Pointer Analysis

- E.g., detecting memory errors in C programs

```
int main() {
    int a[10], int *p;
    int x, y;

    x = get_external();
    y = get_external();

    if (x >= 0) {
        if (x < 16) {
            if (y) {
                if (x >= 10)
                    return 0;
                a[x] = 1;
            }
            p = a;
            p[x] = 1;
        }
    }
}
```

Pointer Analysis

- Pointer analysis computes the set of memory locations (objects) that a pointer variable may point to at runtime.
- One of the most important static analyses: all interesting questions about program properties need pointer analysis.
 - E.g., control-flows, data-flows, types, numeric values, etc

Abstraction of Memory Objects

- Memory locations are unbounded:

```
def id (p): return p
```

```
def f():  
    x = A()    // 11  
    y = id(x)
```

```
def g():  
    a = B()    // 12  
    b = id(a)
```

```
while True: {f(); g() }
```

- In a typical pointer analysis, objects are abstracted into their **allocation-sites**. Pointer analysis result:

$$x \mapsto \{l_1\}, y \mapsto \{l_1\}, a \mapsto \{l_2\}, b \mapsto \{l_2\}, p \mapsto \{l_1, l_2\}$$

cf) Flow Sensitivity

- A flow-sensitive analysis maintains abstract states separately for each program point: e.g.,

```
x = A ()  
y = id (x)  
x = B ()  
y = id (x)
```

- Pointer analysis is often defined flow-insensitively

Pointer Analysis in Datalog

- Pointer analysis is expressed as subset constraints. The analysis is to compute the smallest solution of the constraints. E.g.,

$$\begin{array}{l} x = A() \quad // \quad ll \\ y = x \end{array} \quad \Longrightarrow \quad \begin{array}{l} \{l_1\} \subseteq pts(x) \\ pts(x) \subseteq pts(y) \end{array}$$

- We use the Datalog language to express such constraints
- Datalog is a declarative logic programming language, which has application in database, information extraction, networking, program analysis, security, etc.

Input and Output Relations

- A program is represented by a set of "facts" (relations):

$\text{Alloc}(var : V, heap : H)$

$\text{Move}(to : V, from : V)$

$\text{Load}(to : V, base : V, fld : F)$

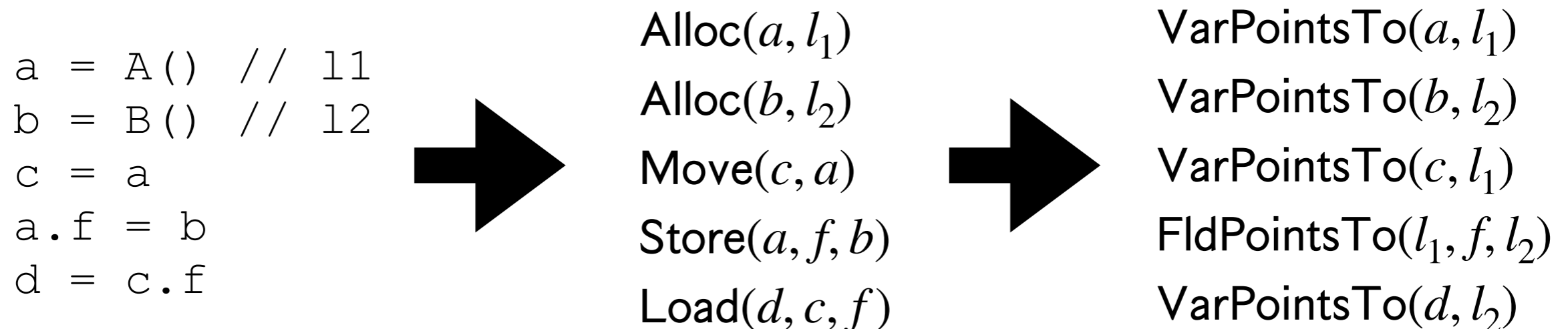
$\text{Store}(base : V, fld : F, from : V)$

V : the set of program variables

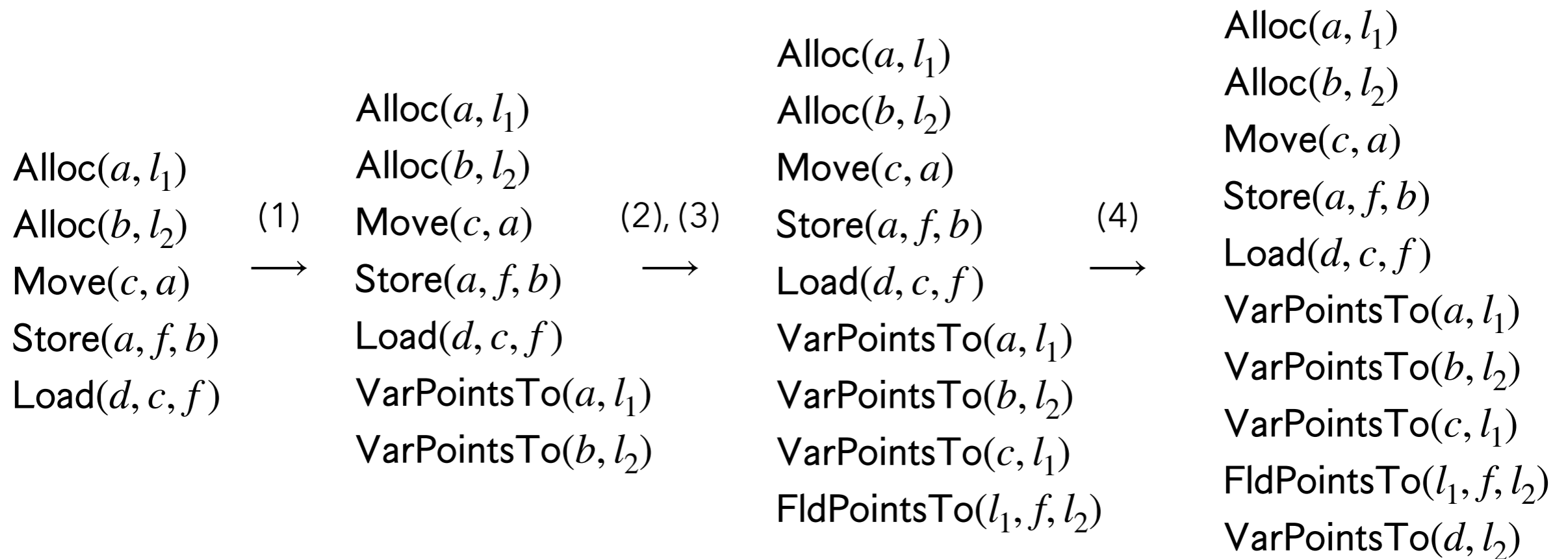
H : the set of allocation sites

F : the set of field names

- Output relations: $\text{VarPointsTo}(var : V, heap : H)$
 $\text{FldPointsTo}(baseH : H, fld : F, heap : H)$



Fixed Point Computation

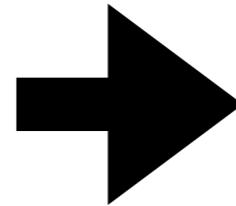


Pointer Analysis Rules

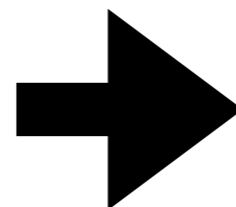
- (1) $\text{VarPointsTo}(var, heap) \leftarrow \text{Alloc}(var, heap)$
- (2) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Move}(to, from), \text{VarPointsTo}(from, heap)$
- (3) $\text{FldPointsTo}(baseH, fld, heap) \leftarrow$
 $\text{Store}(base, fld, from), \text{VarPointsTo}(from, heap),$
 $\text{VarPointsTo}(base, baseH)$
- (4) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Load}(to, base, fld), \text{VarPointsTo}(base, baseH),$
 $\text{FldPointsTo}(baseH, fld, heap)$

Interprocedural Analysis (First-Order)

```
def f(p): // m1
    return p
a = A() // l1
b = f(a) // l2
```



FormalArg($m_1, 0, p$)
FormalReturn(m_1, p)
Alloc($a, l_1, global$)
CallGraph(l_2, m_1)
Reachable($global$)
Reachable(m_1)
ActualArg($l_2, 0, a$)
ActualReturn(l_2, b)



InterProcAssign(p, a)
InterProcAssign(b, p)
VarPointsTo(a, l_1)
VarPointsTo(p, l_1)
VarPointsTo(b, l_1)

Input and Output Relations

- Input relations (program representation)

$\text{Alloc}(var : V, heap : H, inMeth : M)$

$\text{Move}(to : V, from : V)$

$\text{Load}(to : V, base : V, fld : F)$

$\text{Store}(base : V, fld : F, from : V)$

$\text{CallGraph}(invo : I, meth : M)$

$\text{Reachable}(meth : M)$

$\text{FormalArg}(meth : M, i : \mathbb{N}, arg : V)$

$\text{ActualArg}(invo : I, i : \mathbb{N}, arg : V)$

$\text{FormalReturn}(meth : M, ret : V)$

$\text{ActualReturn}(invo : I, var : V)$

V : the set of program variables

H : the set of allocation sites

F : the set of field names

M : the set of method identifiers

S : the set of method signatures

I : the set of instructions

T : the set of class types

\mathbb{N} : the set of natural numbers

- Output relations

$\text{VarPointsTo}(var : V, heap : H)$

$\text{FldPointsTo}(baseH : H, fld : F, heap : H)$

$\text{InterProcAssign}(to : V, from : V)$

Fixed Point Computation

FormalArg($m_1, 0, p$)		FormalArg($m_1, 0, p$)		FormalArg($m_1, 0, p$)
FormalReturn(m_1, p)		FormalReturn(m_1, p)		FormalReturn(m_1, p)
Alloc($a, l_1, global$)		Alloc($a, l_1, global$)		Alloc($a, l_1, global$)
CallGraph(l_2, m_1)		CallGraph(l_2, m_1)		CallGraph(l_2, m_1)
Reachable($global$)	(1), (5), (6)	Reachable($global$)	(7)	Reachable($global$)
Reachable(m_1)	→	Reachable(m_1)	→*	Reachable(m_1)
ActualArg($l_2, 0, a$)		ActualArg($l_2, 0, a$)		ActualArg($l_2, 0, a$)
ActualReturn(l_2, b)		ActualReturn(l_2, b)		ActualReturn(l_2, b)
		VarPointsTo(a, l_1)		VarPointsTo(a, l_1)
		InterProcAssign(p, a)		InterProcAssign(p, a)
		InterProcAssign(b, p)		InterProcAssign(b, p)
				VarPointsTo(p, l_1)
				VarPointsTo(b, l_1)

Analysis Rules

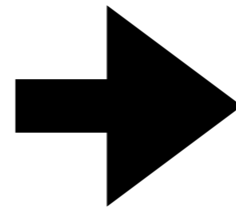
- (1) $\text{VarPointsTo}(var, heap) \leftarrow \text{Reachable}(meth), \text{Alloc}(var, heap, meth)$
- (2) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Move}(to, from), \text{VarPointsTo}(from, heap)$
- (3) $\text{FldPointsTo}(baseH, fld, heap) \leftarrow$
 $\text{Store}(base, fld, from), \text{VarPointsTo}(from, heap), \text{VarPointsTo}(base, baseH)$
- (4) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Load}(to, base, fld), \text{VarPointsTo}(base, baseH), \text{FldPointsTo}(baseH, fld, heap)$
- (5) $\text{InterProcAssign}(to, from) \leftarrow$
 $\text{CallGraph}(invo, meth), \text{FormalArg}(meth, n, to), \text{ActualArg}(invo, n, from)$
- (6) $\text{InterProcAssign}(to, from) \leftarrow$
 $\text{CallGraph}(invo, meth), \text{FormalReturn}(meth, from), \text{ActualReturn}(invo, to)$
- (7) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{InterProcAssign}(to, from), \text{VarPointsTo}(from, heap)$

Interprocedural Analysis (Higher-Order)

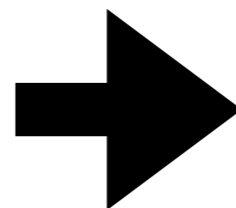
```
class C:
    def id(self, v): // m1
        return v
```

```
class B:
    def g(self): // m2
        c = C() // l1
        s = D() // l2
        t = E() // l3
        d = c.id(s) // l4
        e = c.id(t) // l5
```

```
class A:
    def f(self): // m3
        b = B() // l6
        b.g() // l7
        b.g() // l8
```



FormalArg($m_1, 0, v$)	VCall(c, id, l_4, m_2)
FormalReturn(m_1, v)	VCall(c, id, l_5, m_2)
ThisVar($m_1, self$)	ActualArg($l_4, 0, s$)
LookUp(C, id, m_1)	ActualArg($l_5, 0, t$)
ThisVar($m_2, self$)	ActualReturn(l_4, d)
LookUp(B, g, m_2)	ActualReturn(l_5, e)
Alloc(c, l_1, m_2)	ThisVar($m_3, self$)
Alloc(s, l_2, m_2)	LookUp(A, f, m_3)
Alloc(t, l_3, m_2)	Alloc(b, l_6, m_3)
HeapType(l_1, C)	HeapType(l_6, B)
HeapType(l_2, D)	VCall(b, g, l_7, m_3)
HeapType(l_3, E)	VCall(b, g, l_8, m_3)
	Reachable(m_3)
VarPointsTo(b, l_6)	
Reachable(m_2)	InterProcAssign(v, s)
VarPointsTo($self, l_6$)	InterProcAssign(v, t)
CallGraph(l_7, m_2)	InterProcAssign(d, v)
CallGraph(l_8, m_2)	InterProcAssign(e, v)
VarPointsTo(c, l_1)	VarPointsTo(v, l_2)
VarPointsTo(s, l_2)	VarPointsTo(v, l_3)
VarPointsTo(t, l_3)	VarPointsTo(d, l_2)
Reachable(m_1)	VarPointsTo(d, l_3)
VarPointsTo($self, l_1$)	VarPointsTo(e, l_2)
CallGraph(l_4, m_1)	VarPointsTo(e, l_3)
CallGraph(l_5, m_1)	



Input and Output Relations

- Input relations

$\text{Alloc}(var : V, heap : H, inMeth : M)$

$\text{Move}(to : V, from : V)$

$\text{Load}(to : V, base : V, fld : F)$

$\text{Store}(base : V, fld : F, from : V)$

$\text{VCall}(base : V, sig : S, invo : I, inMeth : M)$

$\text{FormalArg}(meth : M, i : \mathbb{N}, arg : V)$

$\text{ActualArg}(invo : I, i : \mathbb{N}, arg : V)$

$\text{FormalReturn}(meth : M, ret : V)$

$\text{ActualReturn}(invo : I, var : V)$

$\text{ThisVar}(meth : M, this : V)$

$\text{HeapType}(heap : H, type : T)$

$\text{LookUp}(type : T, sig : S, meth : M)$

- Output relations

$\text{VarPointsTo}(var : V, heap : H)$

$\text{FldPointsTo}(baseH : H, fld : F, heap : H)$

$\text{InterProcAssign}(to : V, from : V)$

$\text{CallGraph}(invo : I, meth : M)$

$\text{Reachable}(meth : M)$

Analysis Rules

- (1) $\text{VarPointsTo}(var, heap) \leftarrow \text{Reachable}(meth), \text{Alloc}(var, heap, meth)$
- (2) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Move}(to, from), \text{VarPointsTo}(from, heap)$
- (3) $\text{FldPointsTo}(baseH, fld, heap) \leftarrow$
 $\text{Store}(base, fld, from), \text{VarPointsTo}(from, heap), \text{VarPointsTo}(base, baseH)$
- (4) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{Load}(to, base, fld), \text{VarPointsTo}(base, baseH), \text{FldPointsTo}(baseH, fld, heap)$
- (5) $\text{InterProcAssign}(to, from) \leftarrow$
 $\text{CallGraph}(invo, meth), \text{FormalArg}(meth, n, to), \text{ActualArg}(invo, n, from)$
- (6) $\text{InterProcAssign}(to, from) \leftarrow$
 $\text{CallGraph}(invo, meth), \text{FormalReturn}(meth, from), \text{ActualReturn}(invo, to)$
- (7) $\text{VarPointsTo}(to, heap) \leftarrow$
 $\text{InterProcAssign}(to, from), \text{VarPointsTo}(from, heap)$

Analysis Rules

(8) $\text{Reachable}(toMeth),$

$\text{VarPointsTo}(this, heap),$

$\text{CallGraph}(invo, toMeth) \leftarrow$

$\text{VCall}(base, sig, invo, inMeth), \text{Reachable}(inMeth),$

$\text{VarPointsTo}(base, heap),$

$\text{HeapType}(heap, heapT), \text{LookUp}(heapT, sig, toMeth),$

$\text{ThisVar}(toMeth, this)$

- This analysis performs **on-the-fly call-graph construction**. Pointer analysis and call-graph construction are closely inter-connected in object-oriented and higher-order languages. For example, to resolve call $obj.fun()$, we need pointer analysis. To compute points-to set of a in $f(\text{Object } a) \{ \dots \}$, we need call-graph.

FormalArg($m_1, 0, v$)			Reachable(m_2)			
FormalReturn(m_1, v)			VarPointsTo($self, l_6$)	(1)	VarPointsTo(c, l_1)	
ThisVar($m_1, self$)	(1)		CallGraph(l_7, m_2)		VarPointsTo(s, l_2)	
LookUp(C, id, m_1)	→	VarPointsTo(b, l_6)	→	CallGraph(l_8, m_2)	→	VarPointsTo(t, l_3)
ThisVar($m_2, self$)						
LookUp(B, g, m_2)						
Alloc(c, l_1, m_2)		Reachable(m_1)		InterProcAssign(v, s)		
Alloc(s, l_2, m_2)	(8)	VarPointsTo($self, l_1$)	(5), (6)	InterProcAssign(v, t)	(7)	VarPointsTo(v, l_2)
Alloc(t, l_3, m_2)	→	CallGraph(l_4, m_1)	→	InterProcAssign(d, v)	→	VarPointsTo(v, l_3)
HeapType(l_1, C)		CallGraph(l_5, m_1)		InterProcAssign(e, v)		
HeapType(l_2, D)						
HeapType(l_3, E)						
VCall(c, id, l_4, m_2)		VarPointsTo(d, l_2)				
VCall(c, id, l_5, m_2)	(7)	VarPointsTo(d, l_3)				
ActualArg($l_4, 0, s$)	→	VarPointsTo(e, l_2)				
ActualArg($l_5, 0, t$)		VarPointsTo(e, l_3)				
ActualReturn(l_4, d)						
ActualReturn(l_5, e)						
ThisVar($m_3, self$)						
LookUp(A, f, m_3)						
Alloc(b, l_6, m_3)						
HeapType(l_6, B)						
VCall(b, g, l_7, m_3)						
VCall(b, g, l_8, m_3)						
Reachable(m_3)						


```

class C:
    def id(self, v): // m1
        return v

class B:
    def g(self): // m2
        c = C() // 11
        s = D() // 12
        t = E() // 13
        d = c.id(s) // 14
        e = c.id(t) // 15

class A:
    def f(self): // m3
        b = B() // 16
        b.g() // 17
        b.g() // 18

```

Context Sensitivity

```
class C:  
    def id(self, v): // m1  
        return v
```

```
class B:  
    def g(self): // m2  
        c = C() // 11  
        s = D() // 12  
        t = E() // 13  
        d = c.id(s) // 14  
        e = c.id(t) // 15
```

```
class A:  
    def f(self): // m3  
        b = B() // 16  
        b.g() // 17  
        b.g() // 18
```

```
VarPointsTo(b, l6)  
VarPointsTo(self, l6)  
VarPointsTo(c, l1)  
VarPointsTo(s, l2)  
VarPointsTo(t, l3)  
VarPointsTo(self, l1)  
VarPointsTo(v, l2)  
VarPointsTo(v, l3)  
VarPointsTo(d, l2)  
VarPointsTo(d, l3)  
VarPointsTo(e, l2)  
VarPointsTo(e, l3)
```

```
VarPointsTo(b, ★, l6, ★)  
VarPointsTo(self, l7, l6, ★)  
VarPointsTo(self, l8, l6, ★)  
VarPointsTo(c, l7, l1, ★)  
VarPointsTo(s, l7, l2, ★)  
VarPointsTo(t, l7, l3, ★)  
VarPointsTo(c, l8, l1, ★)  
VarPointsTo(s, l8, l2, ★)  
VarPointsTo(t, l8, l3, ★)  
VarPointsTo(self, l4, l1, ★)  
VarPointsTo(self, l5, l1, ★)  
VarPointsTo(v, l4, l2, ★)  
VarPointsTo(v, l5, l3, ★)  
VarPointsTo(d, l7, l2, ★)  
VarPointsTo(d, l8, l2, ★)  
VarPointsTo(e, l7, l3, ★)  
VarPointsTo(e, l8, l3, ★)
```

context-insensitive

context-sensitive

Domains

V : the set of program variables

H : the set of allocation sites

F : the set of field names

M : the set of method identifiers

S : the set of method signatures

I : the set of instructions

T : the set of class types

\mathbb{N} : the set of natural numbers

C : a set of calling contexts

HC : a set of heap contexts

Output Relations

- The output relations are modified to add contexts:

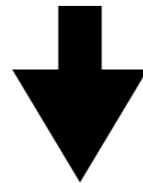
$\text{VarPointsTo}(var : V, heap : H)$

$\text{FldPointsTo}(baseH : H, fld : F, heap : H)$

$\text{InterProcAssign}(to : V, from : V)$

$\text{CallGraph}(invo : I, meth : M)$

$\text{Reachable}(meth : M)$



$\text{VarPointsTo}(var : V, ctx : C, heap : H, hctx : HC)$

$\text{FldPointsTo}(baseH : H, baseHCtx : HC, fld : F, heap : H, hctx : HC)$

$\text{InterProcAssign}(to : V, toCtx : C, from : V, fromCtx : C)$

$\text{CallGraph}(invo : I, callerCtx : C, meth : M, calleeCtx : C)$

$\text{Reachable}(meth : M, ctx : C)$

Context Constructors

- Different choices of constructors yield different context-sensitivity flavors

Record(*heap* : *H*, *ctx* : *C*) = *newHCtx* : *HC*

Merge(*heap* : *H*, *hctx* : *HC*, *invo* : *I*, *ctx* : *C*) = *newCtx* : *C*

- **Record** generates heap contexts
- **Merge** generates calling contexts

Analysis Rules

$\mathbf{Record}(heap, ctx) = hctx,$

$\mathbf{VarPointsTo}(var, ctx, heap, hctx) \leftarrow$

$\mathbf{Reachable}(meth, ctx), \mathbf{Alloc}(var, heap, meth)$

$\mathbf{VarPointsTo}(to, ctx, heap, hctx) \leftarrow$

$\mathbf{Move}(to, from), \mathbf{VarPointsTo}(from, ctx, heap, hctx)$

$\mathbf{FldPointsTo}(baseH, baseHCtx, fld, heap, hctx) \leftarrow$

$\mathbf{Store}(base, fld, from), \mathbf{VarPointsTo}(from, ctx, heap, hctx),$

$\mathbf{VarPointsTo}(base, ctx, baseH, baseHCtx)$

$\mathbf{VarPointsTo}(to, ctx, heap, hctx) \leftarrow$

$\mathbf{Load}(to, base, fld), \mathbf{VarPointsTo}(base, ctx, baseH, baseHCtx),$

$\mathbf{FldPointsTo}(baseH, baseHCtx, fld, heap, hctx)$

Analysis Rules

Merge(*heap, hctx, invo, callerCtx*) = *calleeCtx*,
Reachable(*toMeth, calleeCtx*),
VarPointsTo(*this, calleeCtx, heap, hctx*),
CallGraph(*invo, callerCtx, toMeth, calleeCtx*) ←
 VCall(*base, sig, invo, inMeth*), Reachable(*inMeth, callerCtx*),
 VarPointsTo(*base, callerCtx, heap, hctx*),
 HeapType(*heap, heapT*), LookUp(*heapT, sig, toMeth*),
 ThisVar(*toMeth, this*)

Analysis Rules

$\text{InterProcAssign}(to, calleeCtx, from, callerCtx) \leftarrow$
 $\text{CallGraph}(invo, callerCtx, meth, calleeCtx),$
 $\text{FormalArg}(meth, n, to), \text{ActualArg}(invo, n, from)$

$\text{InterProcAssign}(to, callerCtx, from, calleeCtx) \leftarrow$
 $\text{CallGraph}(invo, callerCtx, meth, calleeCtx),$
 $\text{FormalReturn}(meth, from), \text{ActualReturn}(invo, to)$

$\text{VarPointsTo}(to, toCtx, heap, hctx) \leftarrow$
 $\text{InterProcAssign}(to, toCtx, from, fromCtx),$
 $\text{VarPointsTo}(from, fromCtx, heap, hctx)$

Call-Site Sensitivity

- The best-known flavor of context sensitivity, which uses call-sites as contexts.
- A method is analyzed under the context that is a sequence of the last k call-sites
- The current call-site of the method, the call-site of the caller method, the call-site of the caller method's caller, ..., up to a pre-defined depth (k)

Call-Site Sensitivity

- 1-call-site sensitivity with context-insensitive heap:

$$C = I, \quad HC = \{ \star \}$$

$$\mathbf{Record}(heap, ctx) = \star$$

$$\mathbf{Merge}(heap, hctx, invo, ctx) = invo$$

- 1-call-site sensitivity with context-sensitive heap:

$$C = I, \quad HC = I$$

$$\mathbf{Record}(heap, ctx) = ctx$$

$$\mathbf{Merge}(heap, hctx, invo, ctx) = invo$$

- 2-call-site sensitivity with 1-call-site sensitive heap:

$$C = I \times I, \quad HC = I$$

$$\mathbf{Record}(heap, ctx) = first(ctx)$$

$$\mathbf{Merge}(heap, hctx, invo, ctx) = pair(invo, first(ctx))$$

Object Sensitivity

- The dominant flavor of context sensitivity for object-oriented languages
- Object abstractions (i.e., allocation sites) are used as contexts, qualifying a method's local variables with the allocation site of the receiver object of the method call.

```
class A:  
    def m(self):  
        return
```

```
a = A()    // 11  
a.m()     // 12
```

Object Sensitivity

- 1-object sensitivity with context-insensitive heap:

$$C = H, \quad HC = \{ \star \}$$

$$\mathbf{Record}(heap, ctx) = \star$$

$$\mathbf{Merge}(heap, hctx, invo, ctx) = heap$$

- 2-object sensitivity with 1-call-site sensitive heap:

$$C = H \times H, \quad HC = H$$

$$\mathbf{Record}(heap, ctx) = first(ctx)$$

$$\mathbf{Merge}(heap, hctx, invo, ctx) = pair(heap, hctx)$$

Example

- 2-object sensitivity with 1-call-site sensitive heap:

```
class C:
    def h(self):
        return

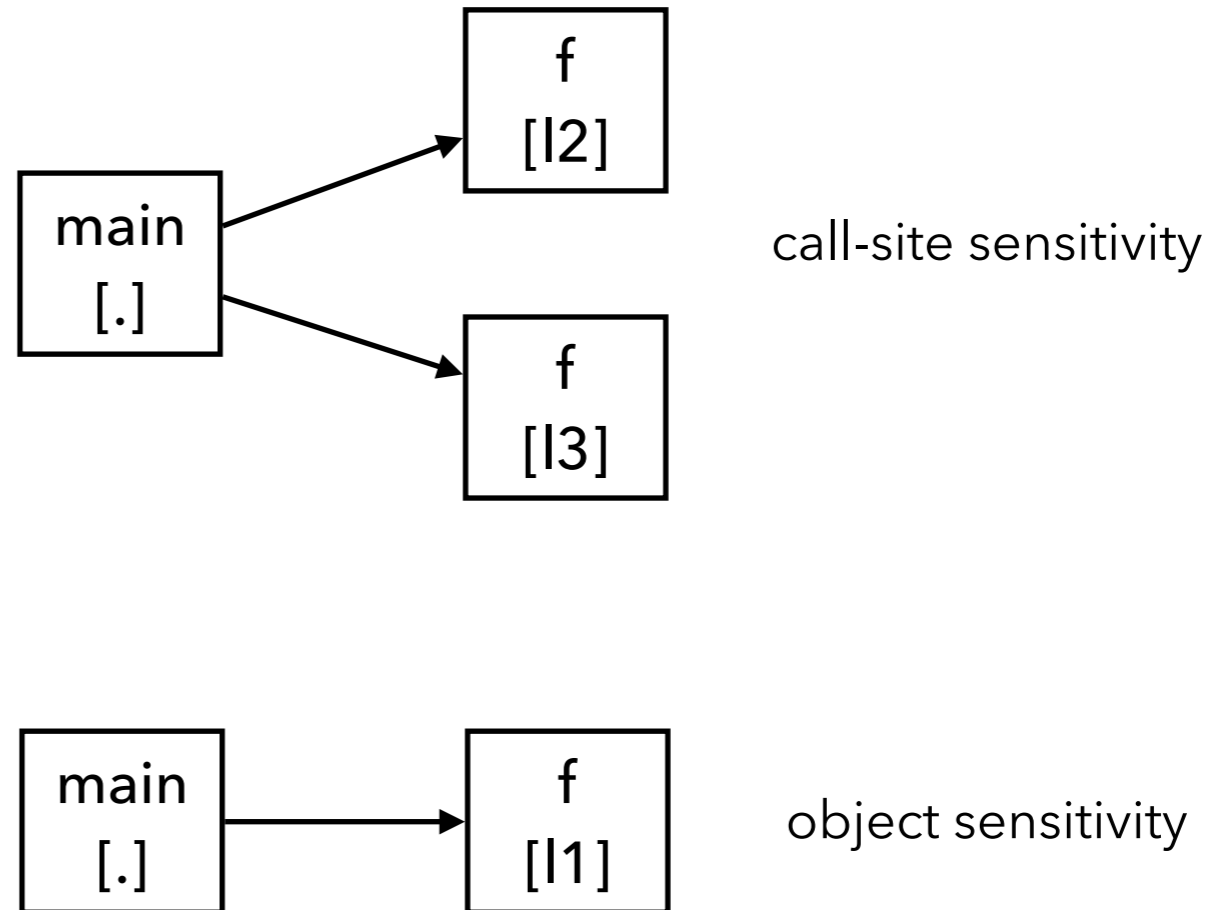
class B:
    def g(self):
        c = C()           // l3, heap objects: (l3, [l1]), (l3, [l2])
        c.h()             // contexts: (l3, l1), (l3, l2)

class A:
    def f(self):
        b1 = B()         // l1
        b2 = B()         // l2
        b1.g()           // context: l1
        b2.g()           // context: l2
```

Call-site vs. Object Sensitivity

- Typical example that benefits from call-site sensitivity:

```
class A:  
    def f(self): return  
  
def main():  
    a = A()    // 11  
    a.f()     // 12  
    a.f()     // 13
```

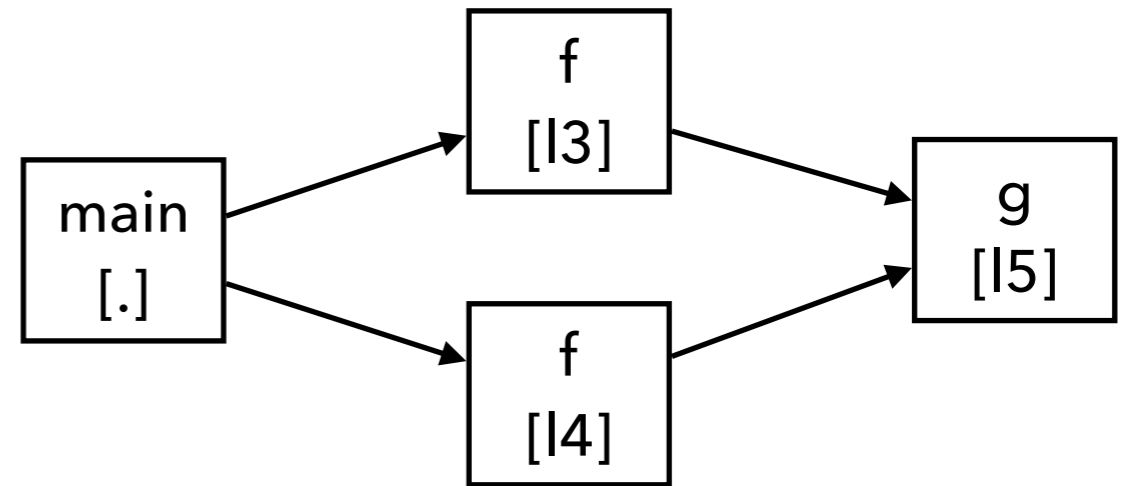


Call-site vs. Object Sensitivity

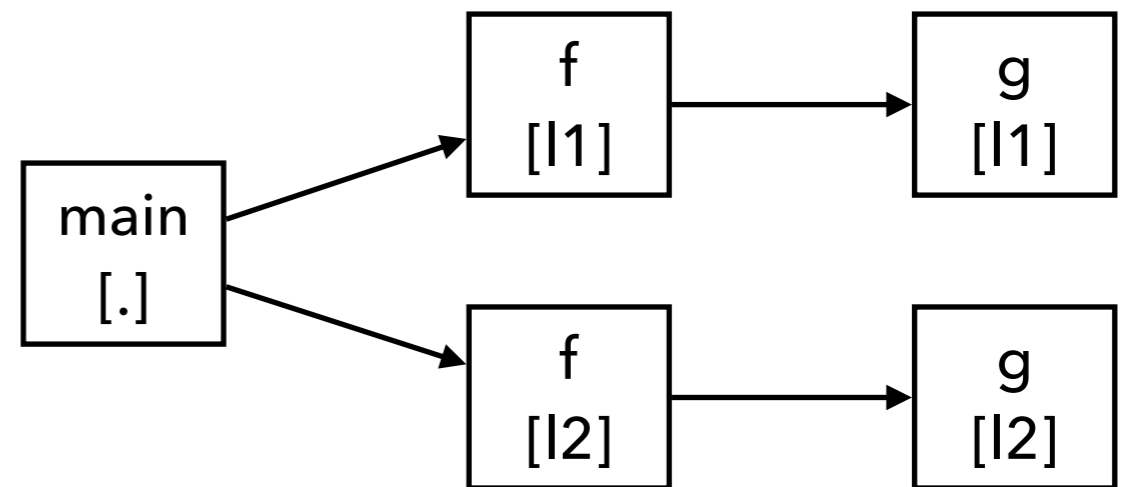
- Typical example that benefits from object sensitivity:

```
class A:  
    def g(self):  
        return  
    def f(self):  
        return self.g() // 15
```

```
def main():  
    a = A() // 11  
    b = A() // 12  
    a.f() // 13  
    b.f() // 14
```



1-call-site sensitivity



1-object sensitivity

Summary

- Covered a number of key concepts in static analysis
 - Pointer analysis
 - Constraint-based analysis
 - Interprocedural analysis
 - Analysis of higher-order programs
 - Context sensitivity