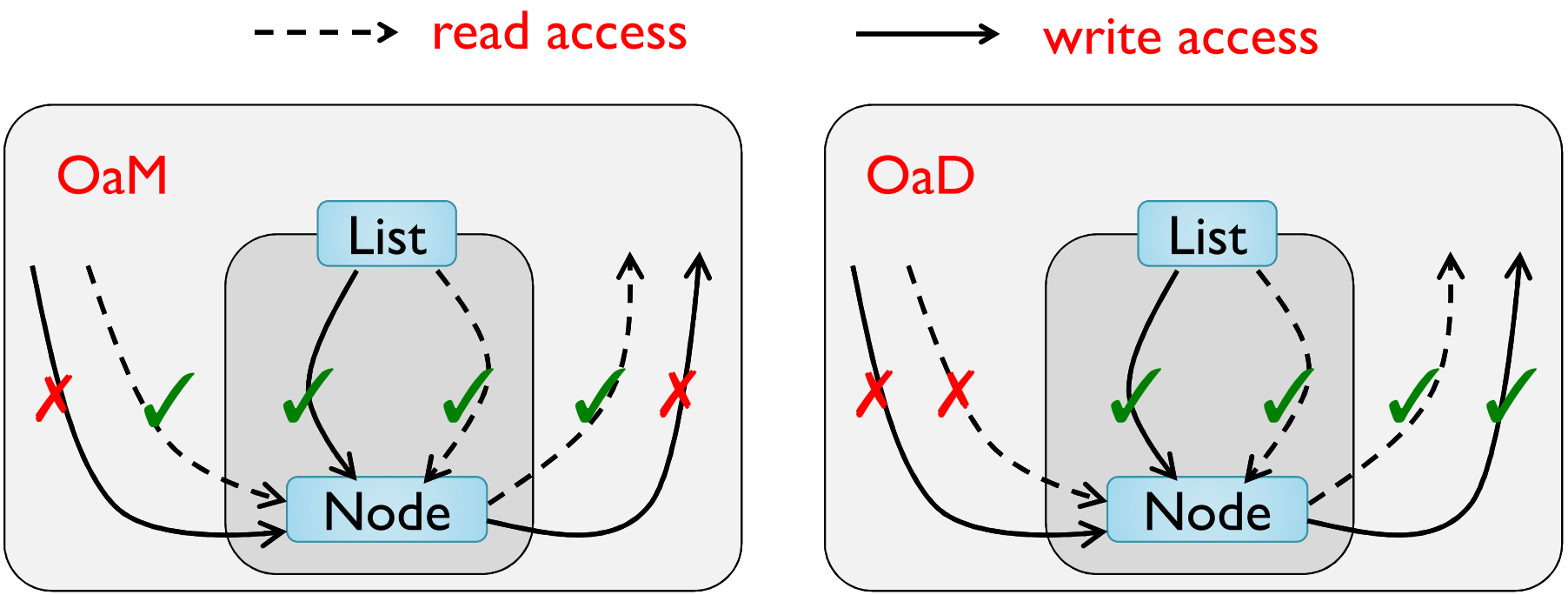
Inference and Checking of Object Ownership

Wei Huang[1], Werner Dietl[2],
Ana Milanova[1], Michael D. Ernst[2]

[1]Rensselaer Polytechnic Institute
[2]University of Washington

Ownership Types

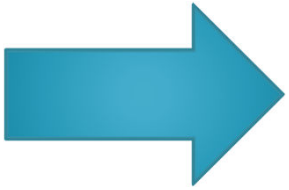
- Owner-as-Modifier (OaM)
  - Universe Types (UT)
- Owner-as-Dominator (OaD)
  - Ownership Types (OT)



Annotation Burden is High

```
1   class Link {
2     Link next;  X data;
3     Link(X inData) {
4       next = null;
5       data = inData;
6     }
7   }
8   class XStack {
9     Link top;
10    void push(X data) {
11      Link newTop;
12      newTop = new Link(data);
13      newTop.next = top;
14      top = newTop;
15    }
16    X pop() {
17      Link oldTop = top;
18      top = oldTop.next;
19      return oldTop.data;
20    }
21    boolean isEmpty() {
22      return top == null; }
23    public static void
    main(String[] args) {
24      XStack s;
25      s = new XStack();
26      X x = new X();
27      s.push(x);
28      x = s.pop();
29    }
30  }
```

13 annotations are used in this small program!

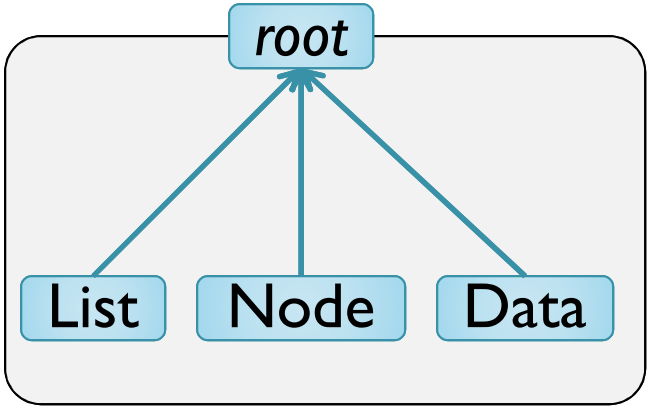


```
1   class Link {
2     <rep|p> Link next;  <p|p> X data;
3     Link(<p|p> X inData) {
4       next = null;
5       data = inData;
6     }
7   }
8   class XStack {
9     <rep|p> Link top;
10    void push(<p|p> X data) {
11      <rep|p> Link newTop;
12      newTop = new <rep|p> Link(data);
13      newTop.next = top;
14      top = newTop;
15    }
16    <p|p> X pop() {
17      <rep|p> Link oldTop = top;
18      top = oldTop.next;
19      return oldTop.data;
20    }
21    boolean isEmpty() {
22      return top == null; }
23    public static void main(String[]
    args) {
24      <rep|rep> XStack s;
25      s = new <rep|rep> XStack();
26      <rep|rep> X x = new <rep|rep> X();
27      s.push(x);
28      x = s.pop();
29    }
30  }
```

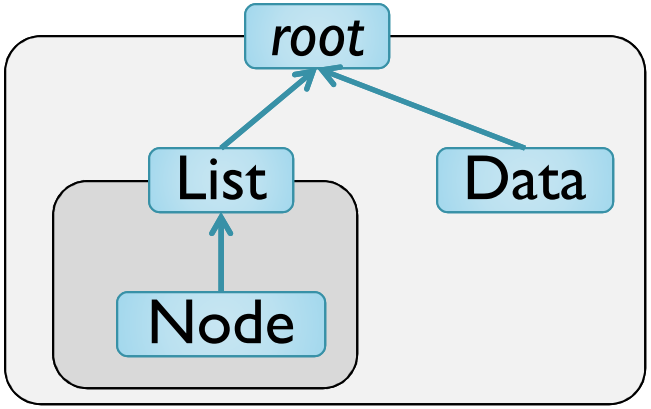
Ownership Type Inference

- Transforms un-annotated or partially-annotated programs into fully annotated ones
  - Facilitates practical adoption of ownership types
  - Reveals how ownership concepts are expressed in existing programs

Many Valid Typings!



Flatter tree



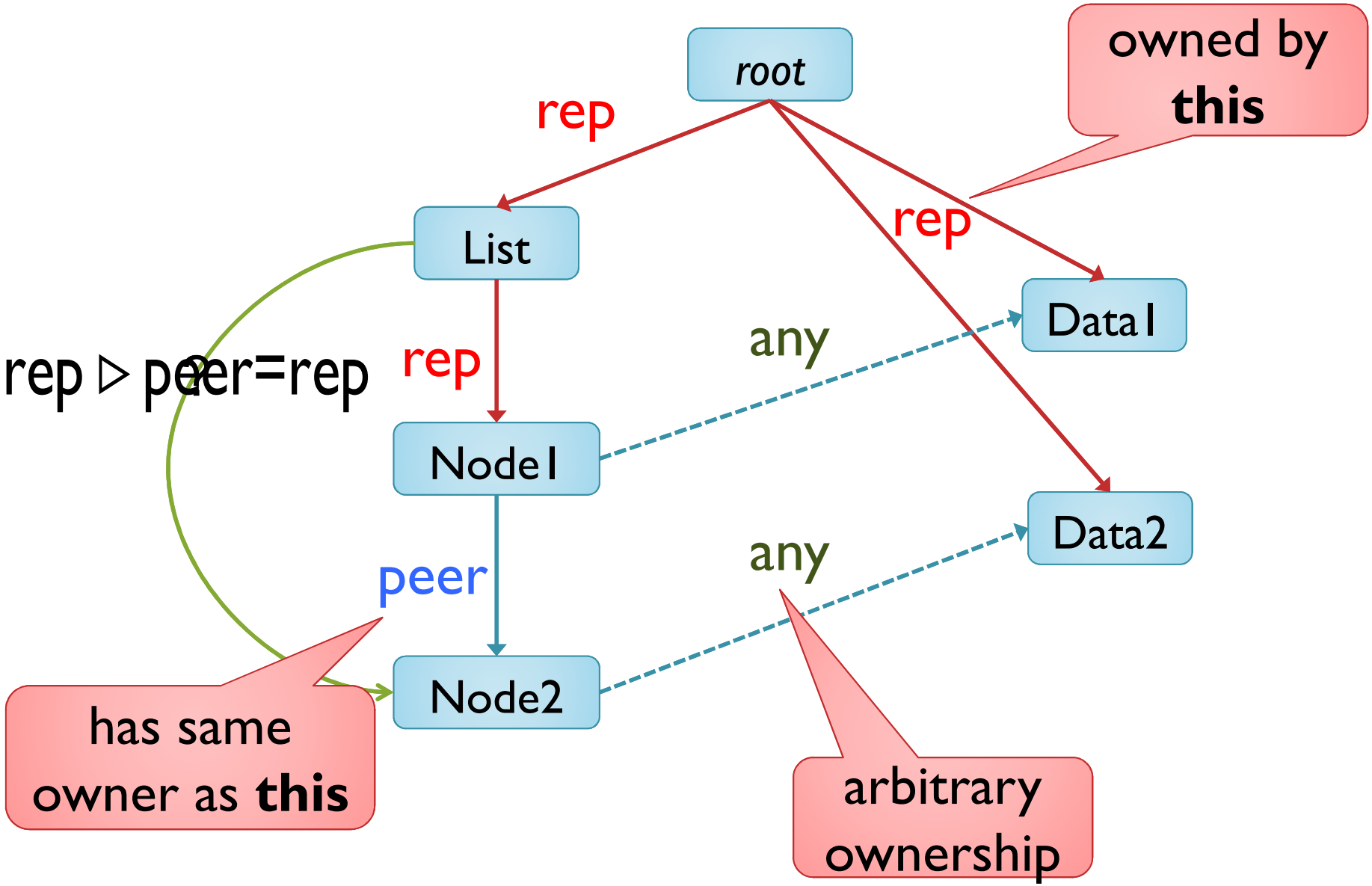
Deeper tree

- Goal: Infer the "best" typing
  - The typing that gives rise to the deepest tree

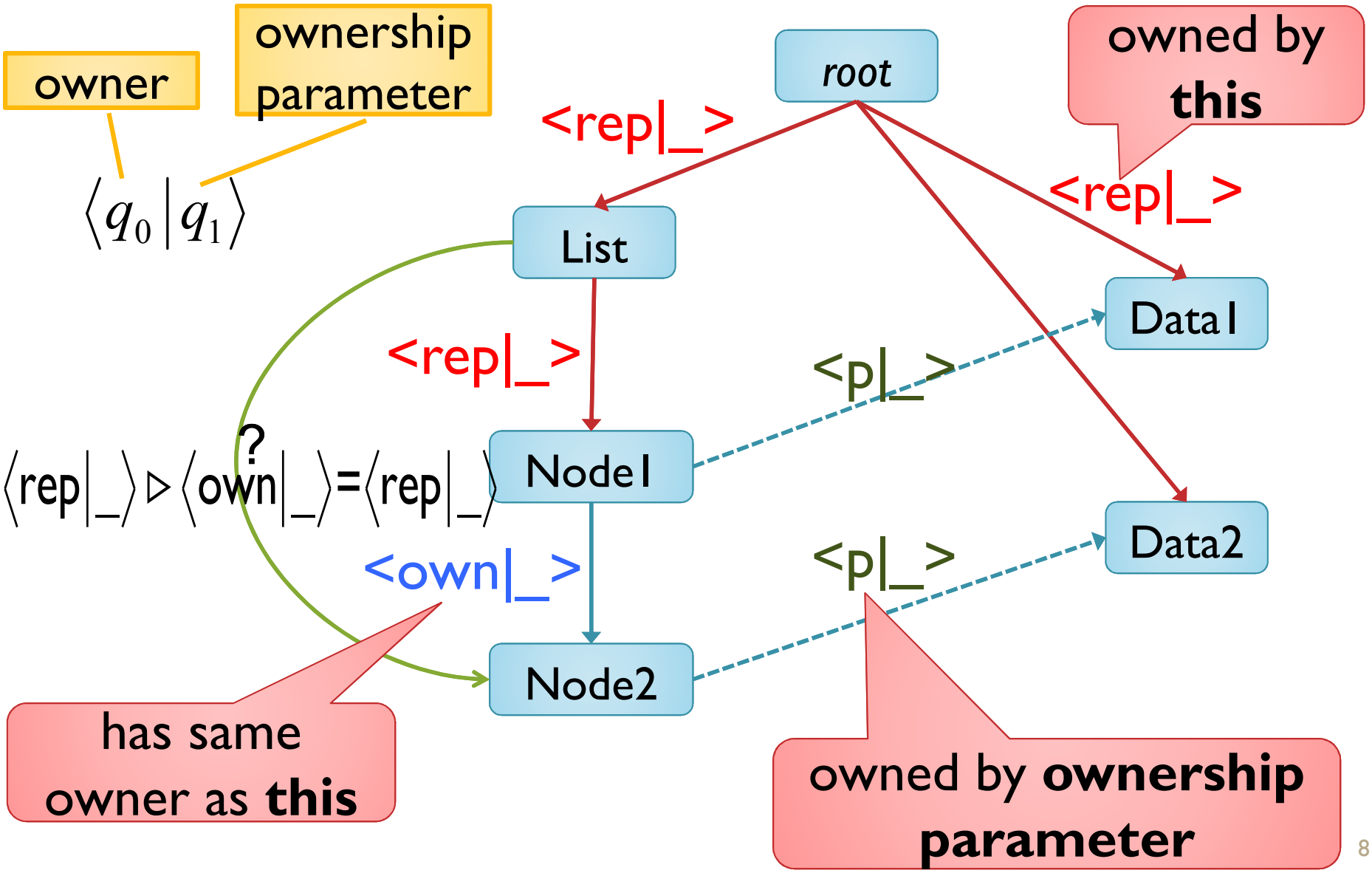
Contributions

- Unified typing rules
  - Universe Types (UT)
  - Ownership Types (OT)
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation
  - Results for UT and OT
  - Comparison of UT and OT

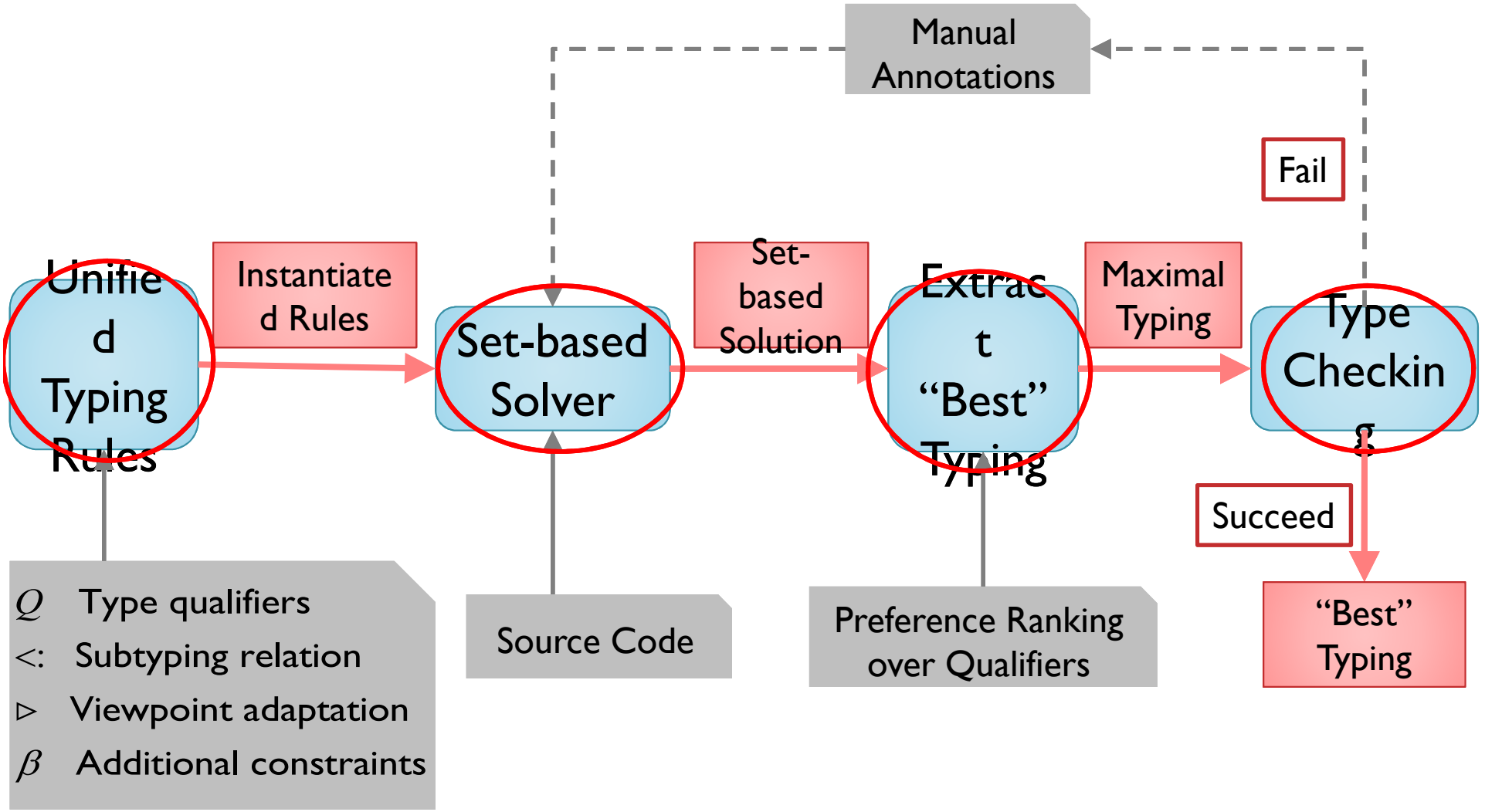
Universe Types [Dietl & Müller JOT'05]



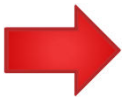
Ownership Types [Clark et al. OOPSLA'98]



Architecture



Outline



- Unified typing rules
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation

Typing Rule (TWRITE): x.f = y

UT: (TWRITE)

$$\frac{\Gamma(x) = q_x \quad \Gamma(y) = q_y \quad typeof(f) = q_f \quad q_y <: q_x \triangleright q_f \quad q_x \neq \text{any} \quad q_x \triangleright q_f \neq \text{lost}}{\Gamma \vdash x.f = y}$$

OT: (TWRITE)

$$\frac{\Gamma(x) = q_x \quad \Gamma(y) = q_y \quad typeof(f) = q_f \quad q_y <: q_x \triangleright q_f}{\Gamma \vdash x.f = y}$$

UT Adaptations:

rep ▷ peer=rep

peer ▷ peer=peer

…

OT Adaptations:

$\langle rep|p \rangle \triangleright \langle own|p \rangle = \langle rep|p \rangle$

$\langle own|p \rangle \triangleright \langle own|p \rangle = \langle own|p \rangle$

…

Unified: (TWRITE)

$$\frac{\Gamma(x) = q_x \quad \Gamma(y) = q_y \quad typeof(f) = q_f \quad q_y <: q_x \triangleright q_f \quad \beta_{(\text{TWRITE})}}{\Gamma \vdash x.f = y}$$

Outline

- Unified typing rules
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation

Set-based Solver

- Set Mapping *S*: variable → {possible qualifiers}
  - e.g. *S*(x) = {any, rep, peer}
- Iterates over statements s
  - Applies the function f_s
  - f_s removes infeasible qualifiers for each variable in s according to the instantiated rules
- Until
  - Reaches a fixpoint, or
  - Assigns the empty set to a variable

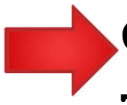
Example

```
1   class XStack {
2     {any, rep, peer} Link top;
3      void push( {any, rep, peer} X d)  {
4     {any, rep, peer} Link newTop;
5      newTop = new {any, rep, peer} Link();
6      newTop.init(d);
7      ...
8     }
9   }
10  class Link {
11     ...
12     void init( {any, rep, peer} X inData)  {
13       ...
14     }
15  }
```

First Iteration

```
1   class XStack {
2     {any, rep, peer} Link top;
3      void push( {any, rep, peer} X d) {
4     {any, rep, peer} Link newTop;
5      newTop = new {any, rep, peer} Link();
6      newTop.init(d);
7      ...
8    }
9   }
10  class Link {
11     ...
12     void init( {any, rep, peer} X inData) {
13       ...
14     }
15  }
```

First Iteration



```
1   class XStack {
2     {any, rep, peer} Link top;
3      void push( {any, rep, peer} X d) {
4     {any, rep, peer} Link newTop;
5      newTop = new {any, rep, peer} Link();
6      newTop.init(d);
7      ...
8    }
9   }
10  class Link {
11     ...
12     void init( {any, rep, peer} X inData) {
13       ...
14     }
15  }
```

Final Result: A Set-based Solution

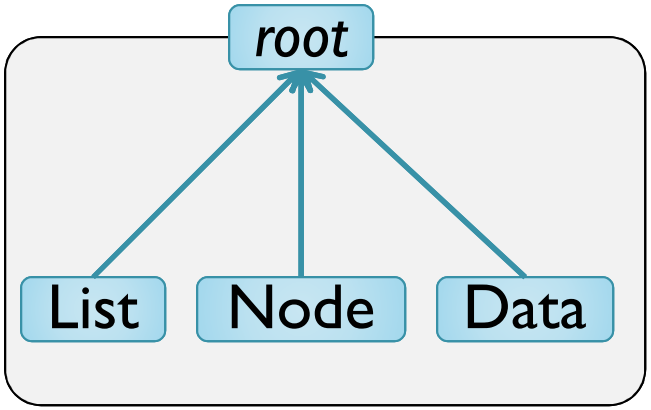
```
1   class XStack {
2     {any, rep, peer} Link top;
3      void push( {any, rep, peer} X d) {
4     {any, rep, peer} Link newTop;
5      newTop = new {any, rep, peer} Link();
6      newTop.init(d);
7      ...
8     }
9   }
10  class Link {
11     ...
12     void init( {any, rep, peer} X inData) {
13       ...
14     }
15  }
```

Outline

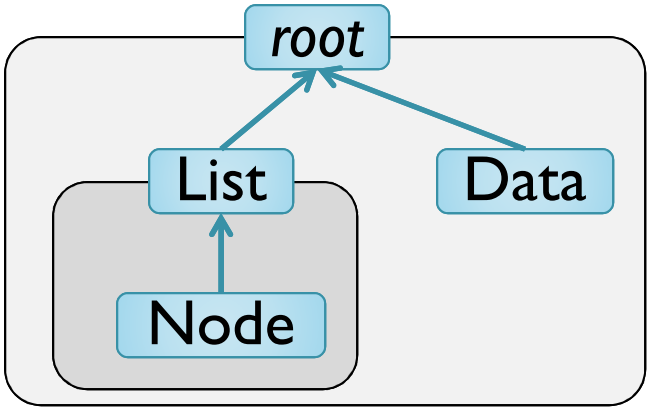
- Unified typing rules
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation

Set-based Solution

- Many valid typings can be extracted from the solution
- Which one is the "best"?
  - Deeper ownership tree has better encapsulation



Flatter tree



Deeper tree

Notion of "Best" Typing

- Objective functions rank valid typings
- T is a valid typing

$$o_{UT}(T) = (\left|T^{-1}(\text{any})\right|, \left|T^{-1}(\text{rep})\right|, \left|T^{-1}(\text{peer})\right|)$$

ranks UT typings; a proxy for deep UT tree

$$o_{OT}(T) = (\left|T^{-1}(\langle \text{rep} | \_ \rangle)\right|, \left|T^{-1}(\langle \text{own} | \_ \rangle)\right|, \left|T^{-1}(\langle \text{p} | \_ \rangle)\right|)$$

ranks OT typings; a proxy for deep OT tree

- "Best" typing maximizes objective function

Maximal Typing

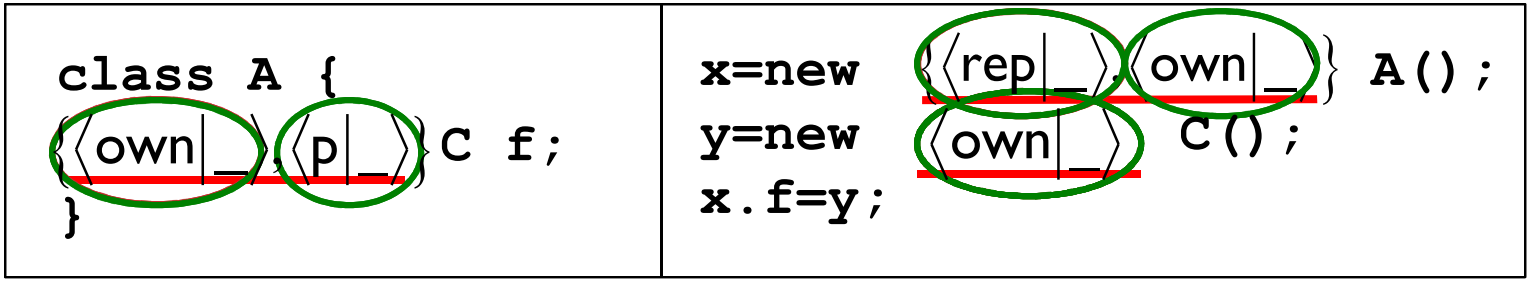
- Maximal typing assigns to each variable x the maximally preferred qualifier from *S*(x)
  - Preference ranking over qualifiers
    - UT: any > rep > peer
    - OT: $\langle \text{rep}|\_\rangle > \langle \text{own}|\_\rangle > \langle \text{p}|\_\rangle$
- *Theorem: If the maximal typing type-checks, then it maximizes the objective function*
  - UT: the maximal typing always type-checks
  - OT: it does not always type-check

UT: Maximal Typing Always Type Checks

```
1   class XStack {
2     {any, rep, peer} Link top;
3      void push( {any, rep, peer} X d) {
4     {any, rep, peer} Link newTop;
5      newTop = new {any, rep, peer} Link();
6      newTop.init(d);
7      ...
8     }
9   }
10  class Link {
11     ...
12     void init( {any, rep, peer} X inData) {
13       ...
14     }
15  }
```

OT: Maximal Typing Does Not Always Type Check

- Conflict: picking the maximal qualifiers doesn't type-check
- Prompts user for manual annotations



$$\underset{\texttt{x}}{\langle rep|\_\rangle} \triangleright \underset{\texttt{f}}{\langle own|\_\rangle} = \underset{\texttt{x}}{\langle rep|\_\rangle} \neq \underset{\texttt{y}}{\langle own|\_\rangle}$$

Outline

- Unified typing rules
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation

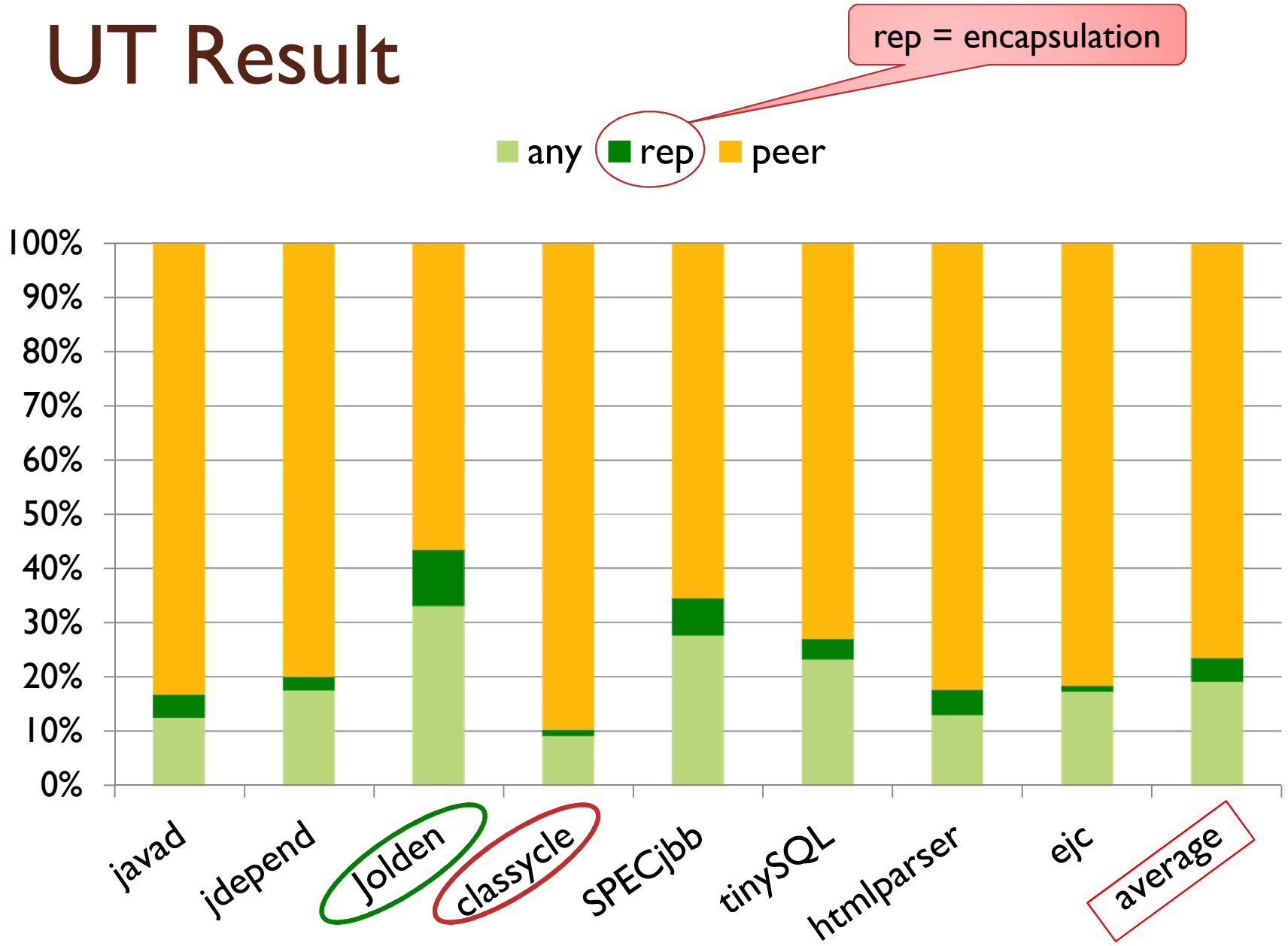
Implementation

- Built on top of the Checker Framework (CF) [Papi et al. ISSTA'08, Dietl et al. ICSE'11]
- Extends the CF to specify:
  - Preference ranking over qualifiers
  - Viewpoint adaptation function
  - Additional constraints
- Publicly available at
  - http://www.cs.rpi.edu/~huangw5/cf-inference

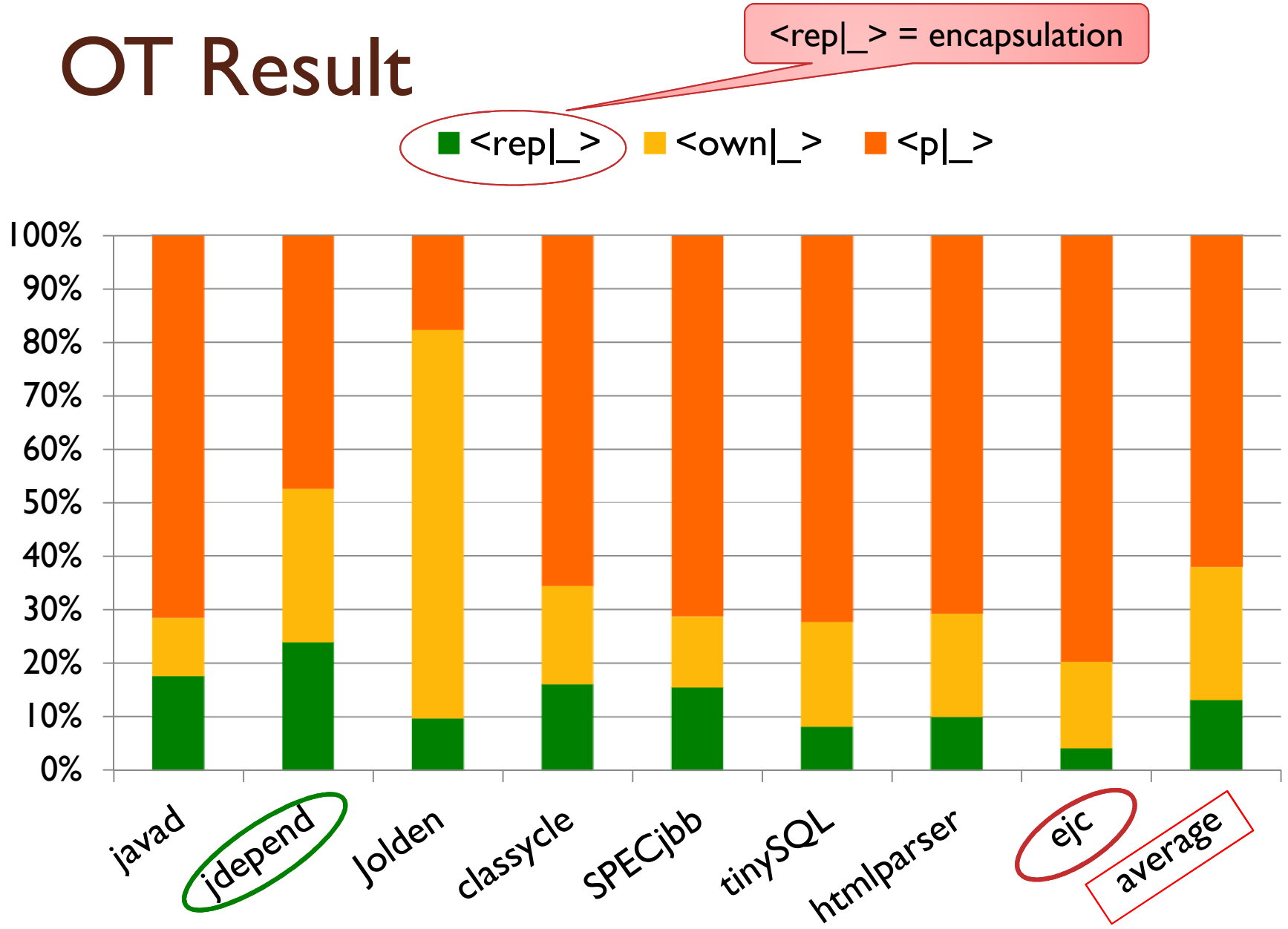
Benchmarks

Benchmark	#Line	Description
javad	4,207	Java class file disassembler
jdepend	4,351	Java package dependency analyzer
JOlden	6,223	Benchmark suite of 10 small programs
classycle	8,972	Java class and package dependency analyzer
SPECjbb	12,076	SPEC's benchmark for evaluating server side Java
tinySQL	31,980	Database engine
htmlparser	62,627	HTML parser
ejc	110,822	Java compiler of the Eclipse IDE

UT Result

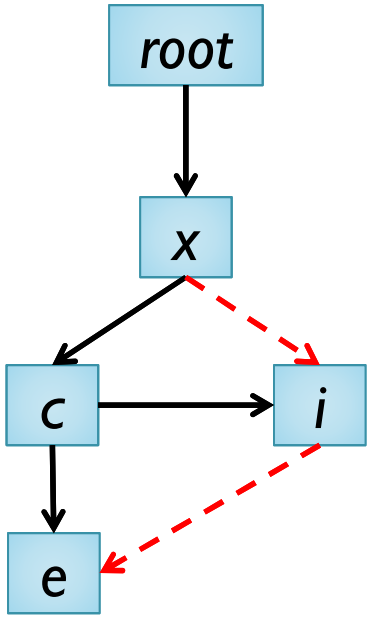

rep = encapsulation
any
rep
peer
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
javad
jdepend
Jolden
classycle
SPECjbb
tinySQL
htmlparser
ejc
average

OT Result

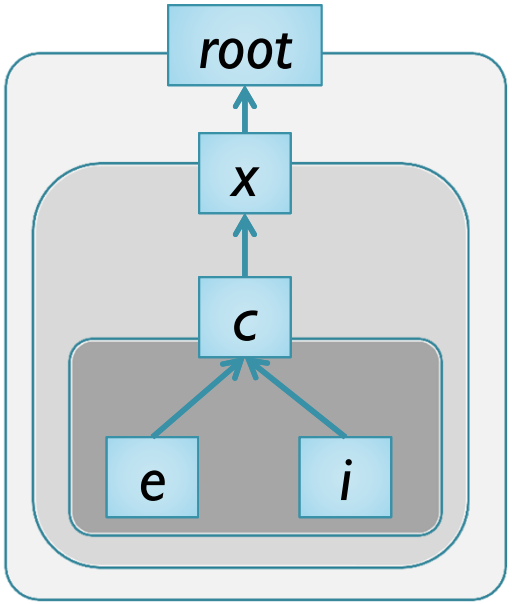

<rep|_> = encapsulation
<rep|_>
<own|_>
<p|_>
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
javad
jdepend
JoIden
classycle
SPECjbb
tinySQL
htmlparser
ejc
average

Owner-as-Modifier vs Owner-as-Dominator

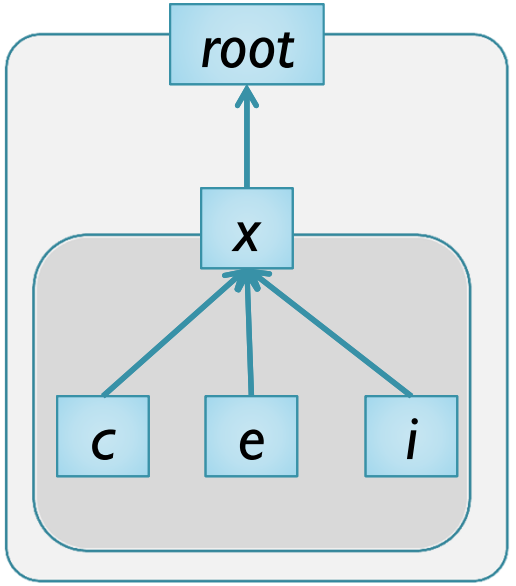
- UT gives rise to a deeper tree
  when access to object e from *x* is readonly



Object Graph



UT Tree

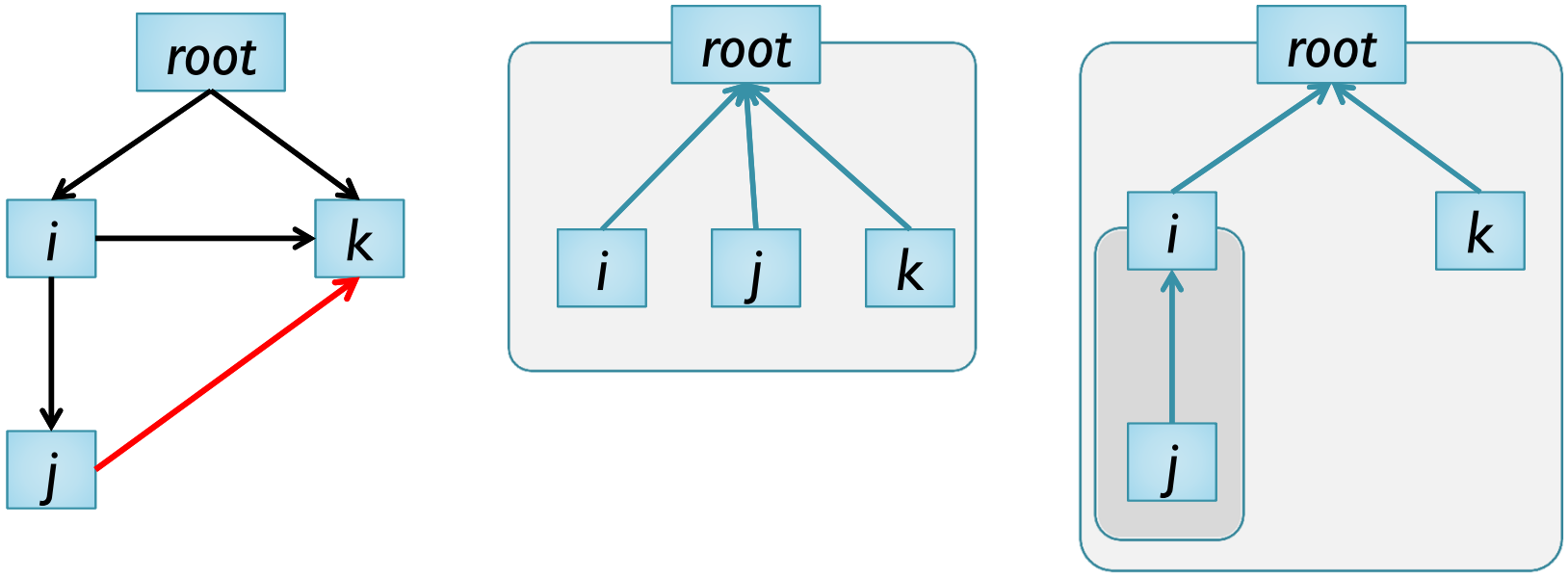


OT Tree

Owner-as-Modifier vs Owner-as-Dominator

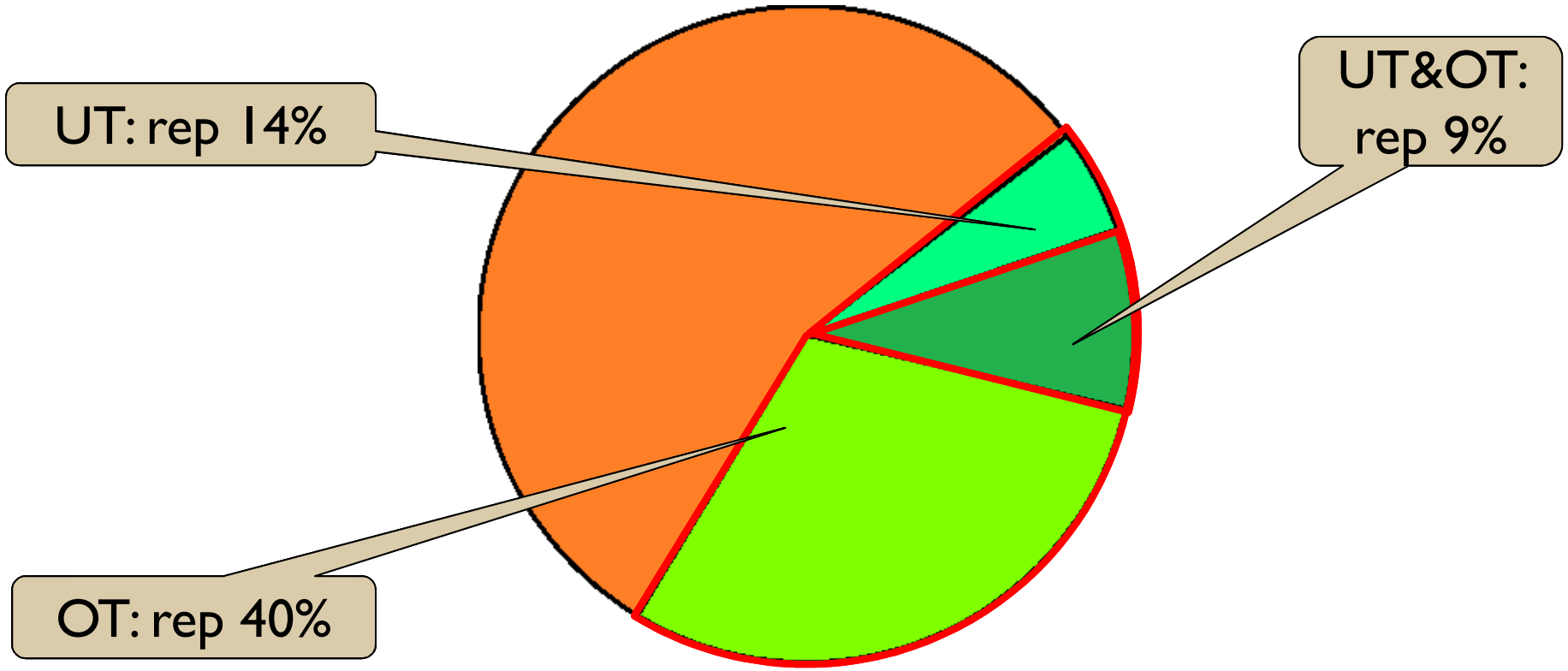
- OT gives rise to a deeper tree

  when object *j* modifies object *k* from an enclosing context



Object Graph UT Tree OT Tree

Allocation Sites



OT has deeper tree: Modification to objects

UT and OT give rise to different ownership trees

Summary of Results

- Manual annotations
  - UT: 0 annotations
  - OT: 6 annotations per 1 kLOC
- Programs can be refactored to have better OaM or OaD structure
- UT requires no manual annotations; annotations are easy to interpret
- OT requires manual annotations; annotations are hard to interpret

Related Work

- Tip et al. [TOPLAS'11]
  - Similar algorithm: starts with all possible answers and iteratively removes infeasible elements
  - We also use qualifier preference ranking
- Dietl et al. [ECOOP'11]
  - Tunable Inference for Generic Universe Types
  - Encodes type constraints and solved by Max-SAT solver
- Sergey & Clark [ESOP'12]
  - Gradual Ownership Types
  - Requires both static and dynamic analyses
  - Analyzes 8,200 lines of code in total

Conclusions

- An inference framework for ownership-like type systems
- Definition of "best" typing
- Evaluation on 241 kLOC
- Publicly available at
  - http://www.cs.rpi.edu/~huangw5/cf-inference

Conclusions

- An inference framework for ownership-like type systems
- Definition of "best" typing
- Evaluation on 241 kLOC
- Publicly available at
  - http://www.cs.rpi.edu/~huangw5/cf-inference

Typing Rule (TCALL): x = y.m(z)

UT: (TCALL)

$$\frac{typeof(\mathrm{m}) = q_p \rightarrow q_{ret} \quad \Gamma(\mathrm{x}) = q_x \quad \Gamma(\mathrm{y}) = q_y \quad typeof(\mathrm{z}) = q_z \quad q_z <: q_y \triangleright q_p \quad q_y \triangleright q_{ret} <: q_x \quad q_y \triangleright q_p \neq \mathrm{lost} \quad impure(\mathrm{m}) \Rightarrow q_y \neq \mathrm{any}}{\Gamma \mapsto \mathrm{x} = \mathrm{y.m(z)}}$$

OT: (TWRITE)

$$\frac{typeof(\mathrm{m}) = q_p \rightarrow q_{ret} \quad \Gamma(\mathrm{x}) = q_x \quad \Gamma(\mathrm{y}) = q_y \quad typeof(\mathrm{z}) = q_z \quad q_z <: q_y \triangleright q_p \quad q_y \triangleright q_{ret} <: q_x \quad q_p \neq \langle \mathrm{rep} | \_ \rangle}{\Gamma \mapsto \mathrm{x} = \mathrm{y.m(z)}}$$

Unified: (TWRITE)

$$\frac{typeof(\mathrm{m}) = q_p \rightarrow q_{ret} \quad \Gamma(\mathrm{x}) = q_x \quad \Gamma(\mathrm{y}) = q_y \quad typeof(\mathrm{z}) = q_z \quad q_z <: q_y \triangleright q_p \quad q_y \triangleright q_{ret} <: q_x \quad \beta_{(\mathrm{TCALL})}}{\Gamma \mapsto \mathrm{x} = \mathrm{y.m(z)}}$$

Delelte

UT Result

Benchmark	TotalVar	any	rep	peer	#Manual	Time(s)
JOlden	685	227	71	387	0	11.3
tinySQL	2711	630	104	1977	0	18.2
htmlparser	3269	426	153	2690	0	22.9
ejc	10957	1897	122	8938	0	119.7
javad	249	31	11	207	0	4.1
SPECjbb	1066	295	74	697	0	13.6
jdepend	542	95	14	433	0	7.2
classycle	946	87	11	848	0	9.9

- Running times range from 4 sec. to 120 sec.
- Zero manual annotations are required

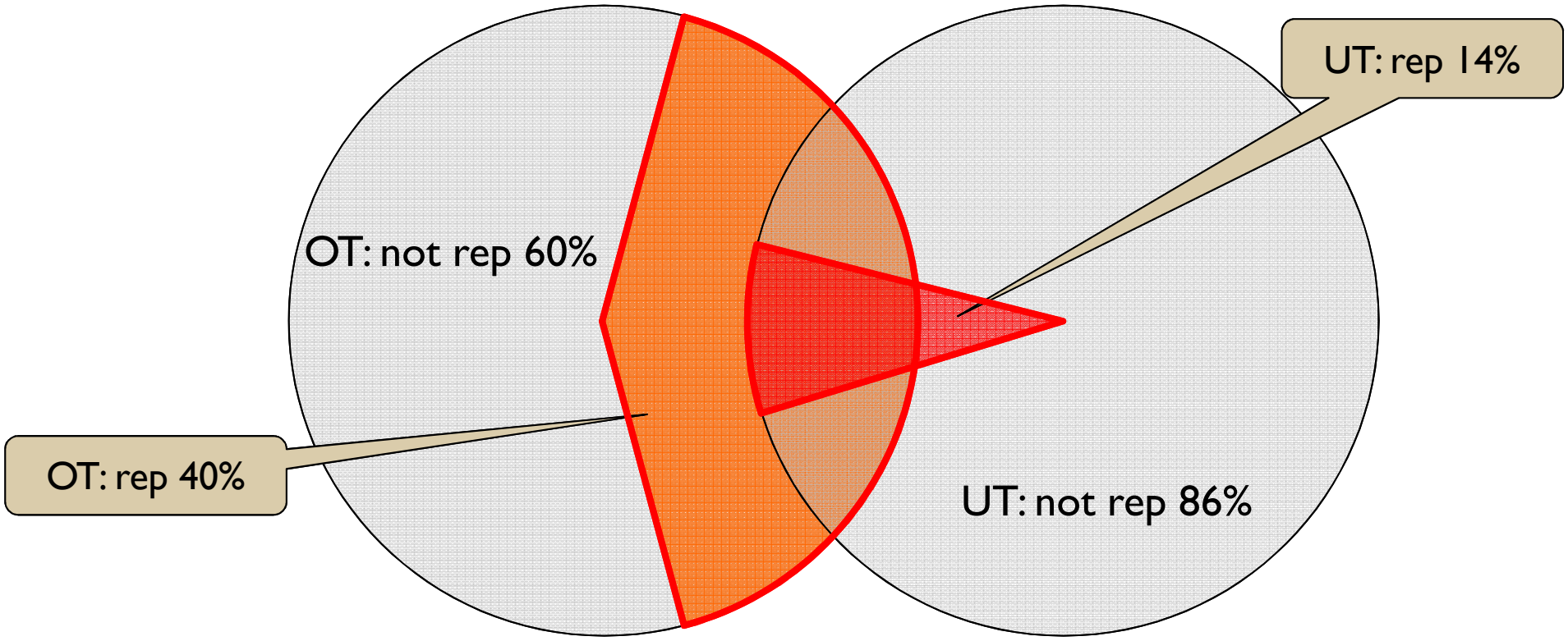
Delelte

OT Result

Benchmark	TotalVar					#Manual	Time(s)
JOlden	685	#⟨rep|_⟩ 67	#⟨own|_⟩ 497	#⟨p|_⟩ 24	#⟨norep|_⟩ 97	13(2/KLOC)	10.3
tinySQL	2711	224	530	5	1952	215(7/KLOC)	18.4
htmlparser	3269	330	629	36	2274	200(3/KLOC)	33.6
ejc	10957	467	1768	50	8672	592(5/KLOC)	122.4
javad	249	44	27	74	104	46(10/KLOC)	5.5
SPECjbb	1066	166	141	71	688	73(6/KLOC)	17.1
jdepend	542	130	156	128	128	26(6/KLOC)	13.7
classycle	946	153	173	28	592	90(10/KLOC)	11.7

- Running times range from 4 sec. to 120 sec.
- 6/KLOC manual annotations on average

Allocation Sites in All Benchmarks



Modification of objects from enclosing context happens more often than readonly exposure

Universe Types

- Owner-as-Modifier encapsulation (OaM)
- Type qualifiers:
  - **rep**: owned by **this**
  - **peer**: has same owner as **this**
  - **any**: arbitrary ownership

Classical Ownership Types

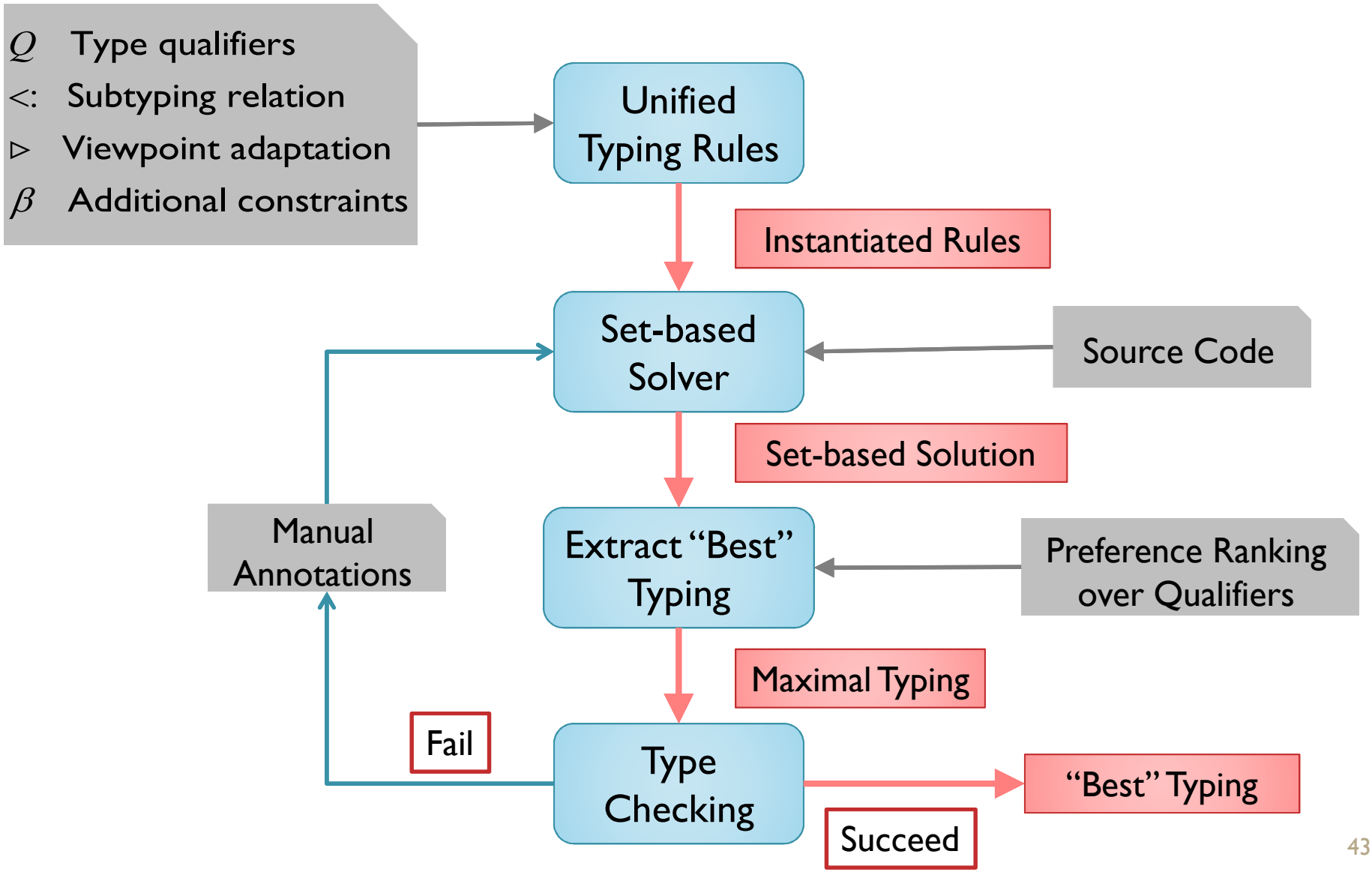
- Owner-as-Dominator encapsulation (OaD)
- Type qualifier $\langle q_0 | q_1 \rangle$
  - q_0 is the owner of the object
  - q_1 is the ownership parameter
  - **rep**: owned by **this**
  - **own**: has same owner as **this**
  - **p**: owned by the ownership parameter

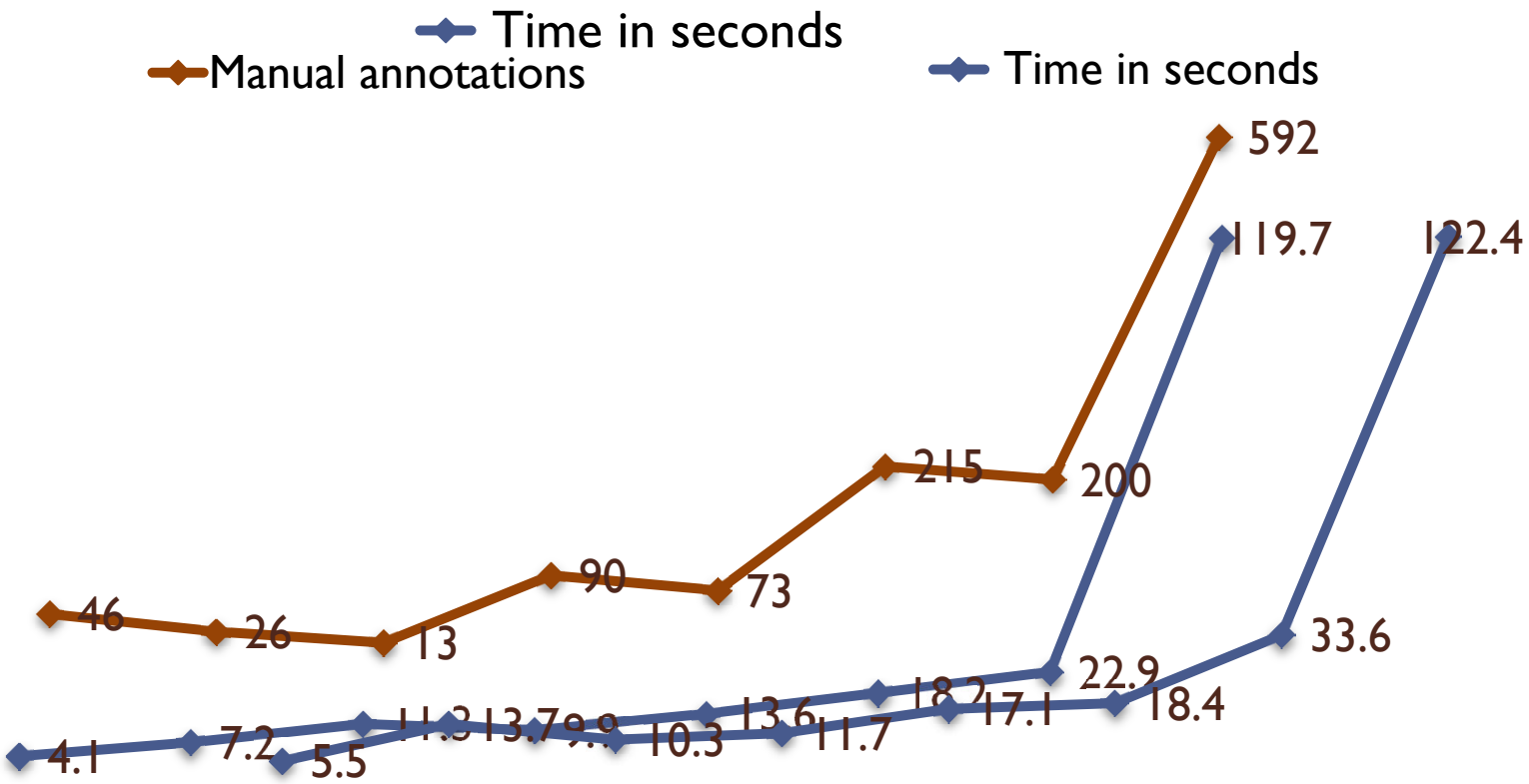
Owner-as-Modifier vs Owner-as-Dominator

- Goal: compare UT (OaM) to OT (OaD)
- In certain cases, UT gives rise to a deeper tree than OT
- In other cases, OT gives rise to a deeper tree

- Does UT or OT has deeper trees?
- Do UT and OT give rise to different trees?

Architecture



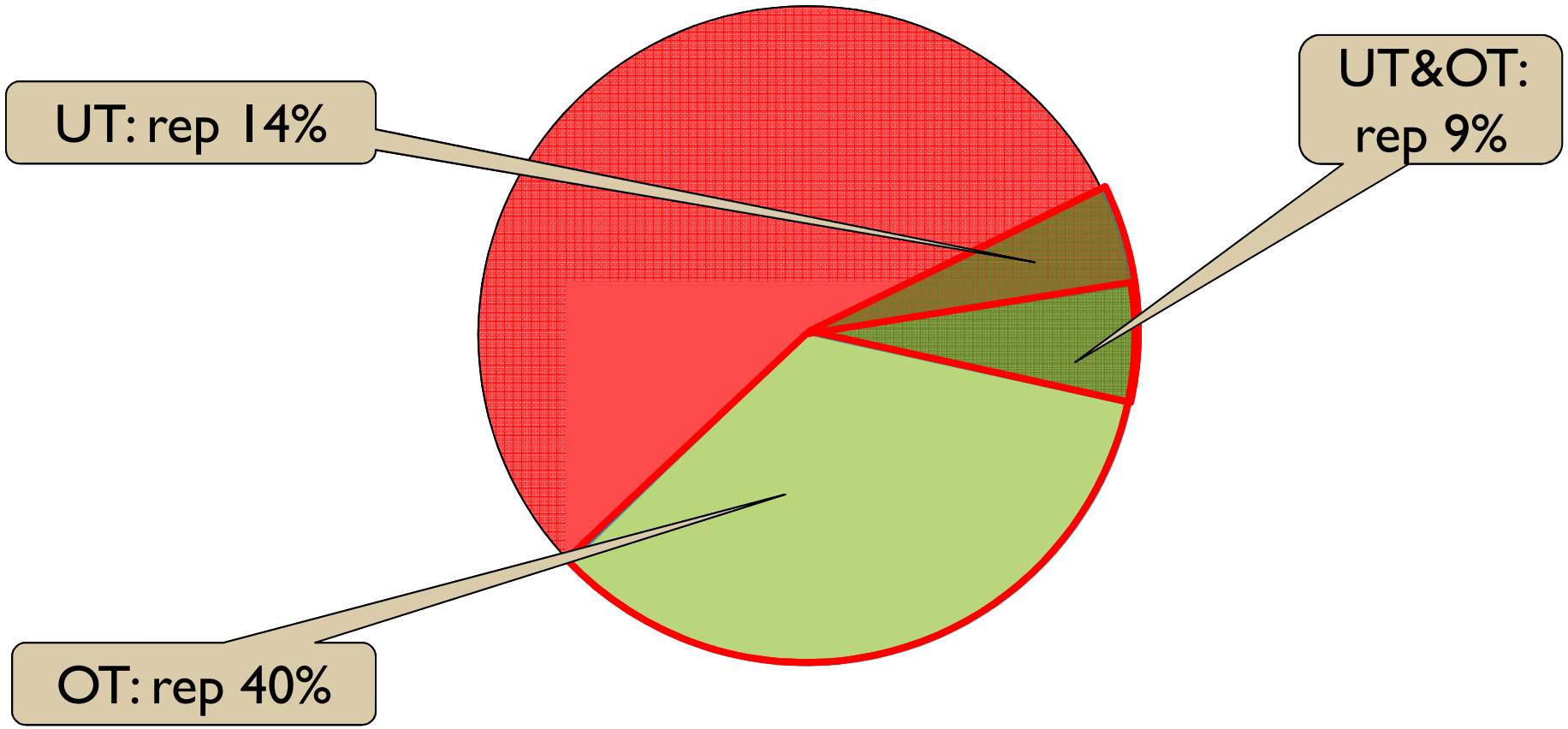

Time in seconds
Manual annotations
Time in seconds
592
119.7
122.4
46
26
13
90
73
215
200
4.1
7.2
5.5
10.3
11.7
13.6
18.2
22.9
17.1
18.4
33.6

Summary of Results

- Many objects are owned (encapsulated)
  - UT: 14% of allocation sites are rep (upper bound!)
  - OT: 40% of allocation sites are rep (close to upper bound!)

- UT requires no manual annotations
  - Programs can be refactored to have better OaM structure
- OT requires manual annotations
  - Annotations are hard to understand

Allocation Sites



Modification from external context happens more o

UT and OT give rise to different ownership trees

Running Time and Manual Annotation

Benchmark	#Line	Running Time (s)		Manual Annotations	
		UT	OT	UT	OT
javad	4,207	4.1	5.5	0	46
jdepend	4,351	7.2	13.7	0	26
JOlden	6,223	11.3	10.3	0	13
classycle	8,972	9.9	11.7	0	90
SPECjbb	12,076	13.6	17.1	0	73
tinySQL	31,980	18.2	18.4	0	215
htmlparser	62,627	22.9	33.6	0	200
ejc	110,822	119.7	122.4	0	592

- Zero manual annotation for UT
- 6 manual annotations per kLOC on average

Notion of "Best" Typing

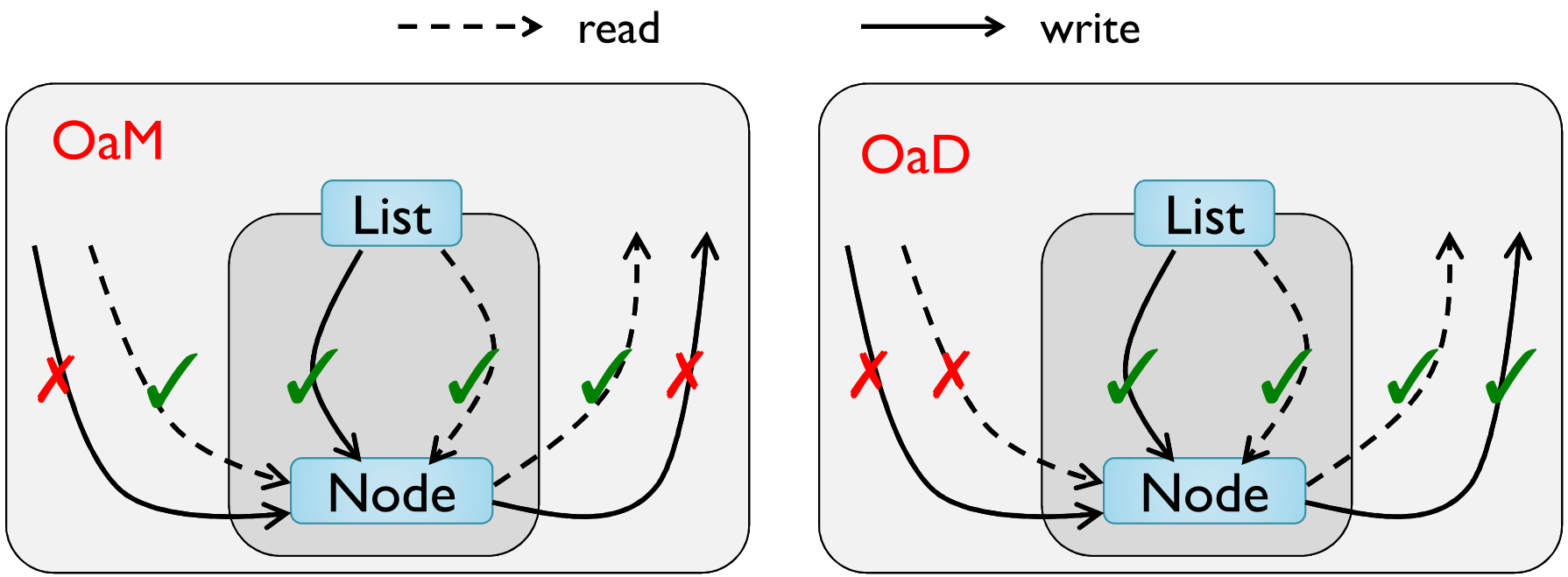
- Objective functions rank valid typings
- T is a valid typing
- $o_{UT}(T)$ ranks UT typings
  - Maximizes number of allocation sites typed rep
- $o_{OT}(T)$ ranks OT typings
  - Maximizes number of object graph edges typed with owner rep

Outline

- Unified typing rules
- Unified inference approach
- Notion of "best" typing
- Implementation and evaluation

Ownership Types

- Owner-as-Modifier
- Owner-as-Dominator



Summary of Results

- Many objects are owned (encapsulated)
  - UT: 14% of allocation sites are rep (upper bound!)
  - OT: 40% of allocation sites are rep (close to upper bound!)
- UT requires no manual annotations; annotations are easy to interpret
- OT requires manual annotations; annotations are hard to interpret