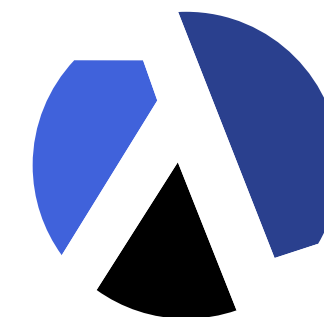


Debun: Detecting Bundled JavaScript Libraries on Web using Property-Order Graphs

Seojin Kim*, **Sungmin Park***, and Jihyeok Park




KOREA
UNIVERSITY



PLRG


Vulnerabilities in JavaScript Library



Snyk Vulnerability Database / npm / **lodash**

Prototype Pollution


Affecting [lodash](#) package, versions **<4.17.17**



Snyk Vulnerability Database / npm / **angular**

Arbitrary Code Execution

Affecting [angular](#) package, versions **<1.4.0-beta.6**



Snyk Vulnerability Database / npm / **vue**

Regular Expression Denial of Service (ReDoS)

Affecting [vue](#) package, versions **<2.5.14**

Prior Work

- **LDC** (Library Detector for Chrome)
 - Manually collected **property patterns** at runtime
 - Library Detection

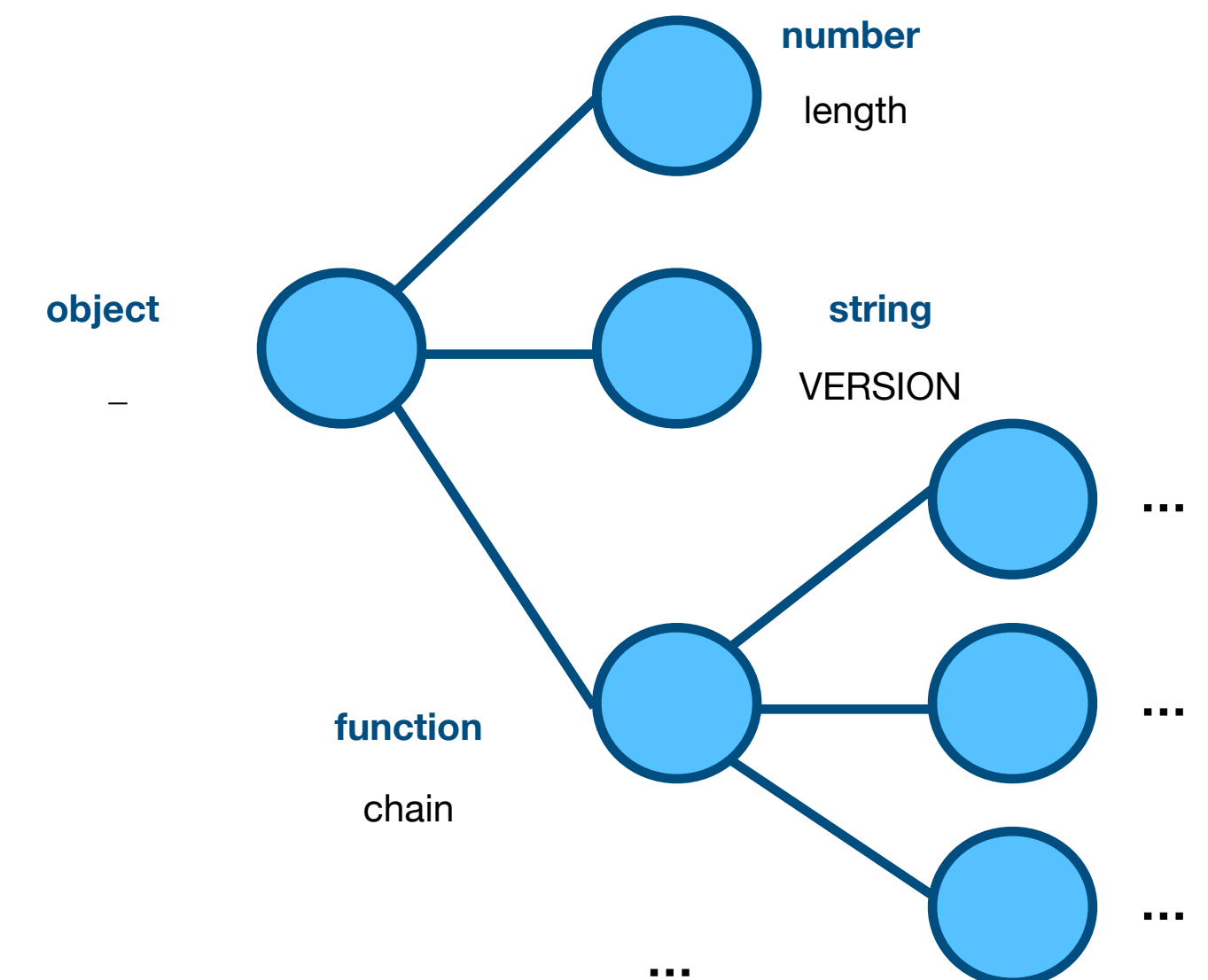
```
typeof (_ = window._) == 'function'
```

```
typeof (chain = _ && _.chain) == 'function'
```
 - Version Detection

```
return { version: _.VERSION || UNKNOWN_VERSION };
```
- **PTDetector** (ASE'23)
 - Automatic extraction of **property patterns** in tree form
- **Limitation**
 - Treating the **source code** as a **black box** can **miss libraries**

Lodash v4.17.21

```
6 // Define properties
7 lodash.chain = function(value) {
8   var result = lodash(value);
9   result.__wrapped__ = value;
10  return result;
11 }
12 lodash.differenceBy = ...
13 ...
14 lodash.VERSION = '4.17.21'
15 // Export lodash
16 window._ = lodash;
17 }.call(this);
```



Prior Work

- **LDC** (Library Detector for Chrome)
 - Manually collected **property patterns** at runtime

- Library Detection

```
typeof (_ = window._) == 'function'  
typeof (chain = _ && _.chain) == 'function'
```

- Version Detection

```
return { version: _.VERSION || UNKNOWN_VERSION };
```

- **PTDetector** (ASE'23)

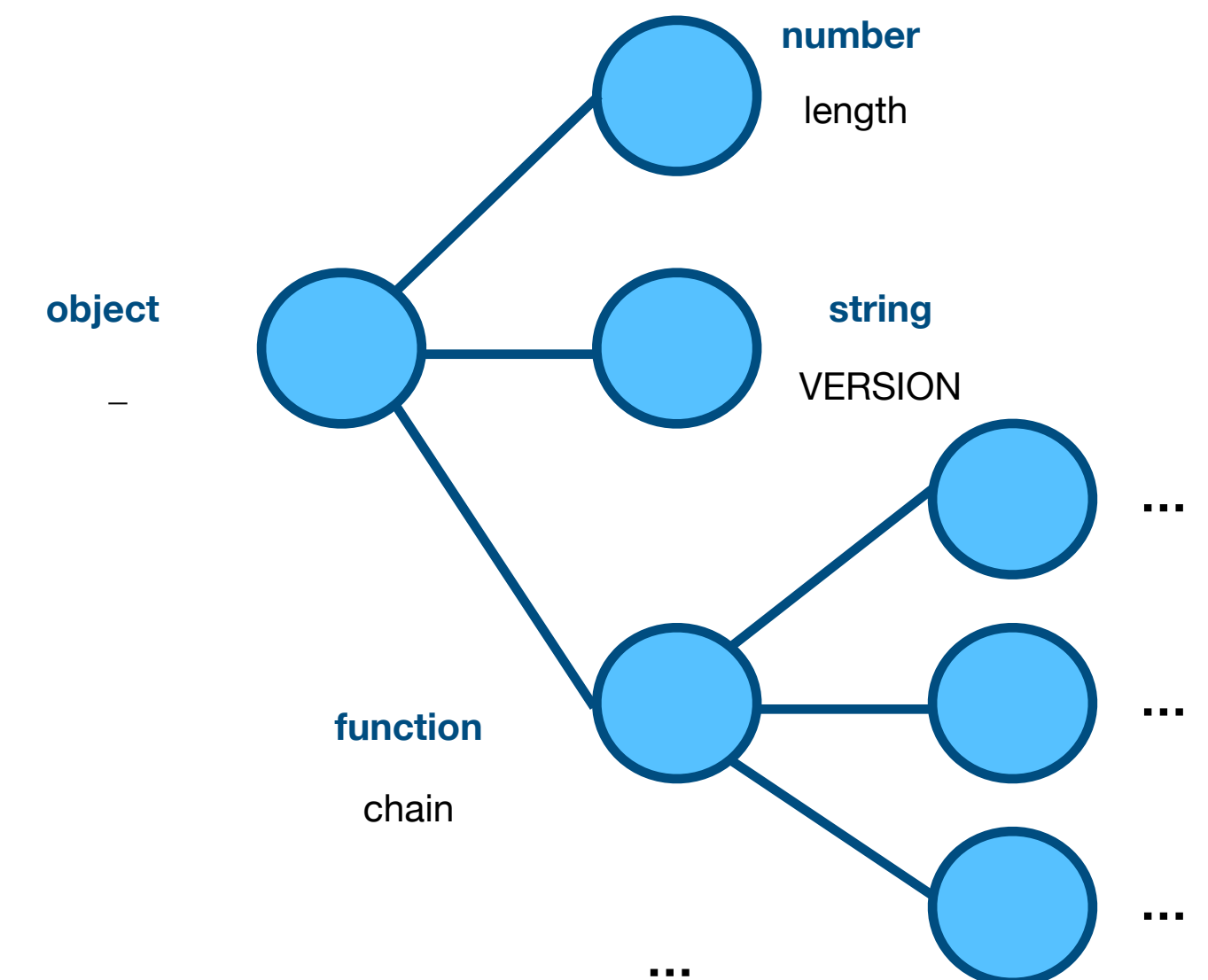
- Automatic extraction of **property patterns** in tree form

- **Limitation**

- Treating the **source code** as a **black box** can **miss libraries**

Lodash v4.17.21

```
6 // Define properties  
7 lodash.chain = function(value) {  
8   var result = lodash(value);  
9   result.__wrapped__ = value;  
10  return result;  
11 }  
12 lodash.differenceBy = ...  
13 ...  
14 lodash.VERSION = '4.17.21'  
15 // Export lodash  
16 window. = lodash;  
17 }.call(this));
```



Prior Work

- **LDC** (Library Detector for Chrome)
 - Manually collected **property patterns** at runtime

- Library Detection

```
typeof (_ = window._) == 'function'  
typeof (chain = _ && _.chain) == 'function'
```

- Version Detection

```
return { version: _.VERSION || UNKNOWN_VERSION };
```

- **PTDetector** (ASE'23)

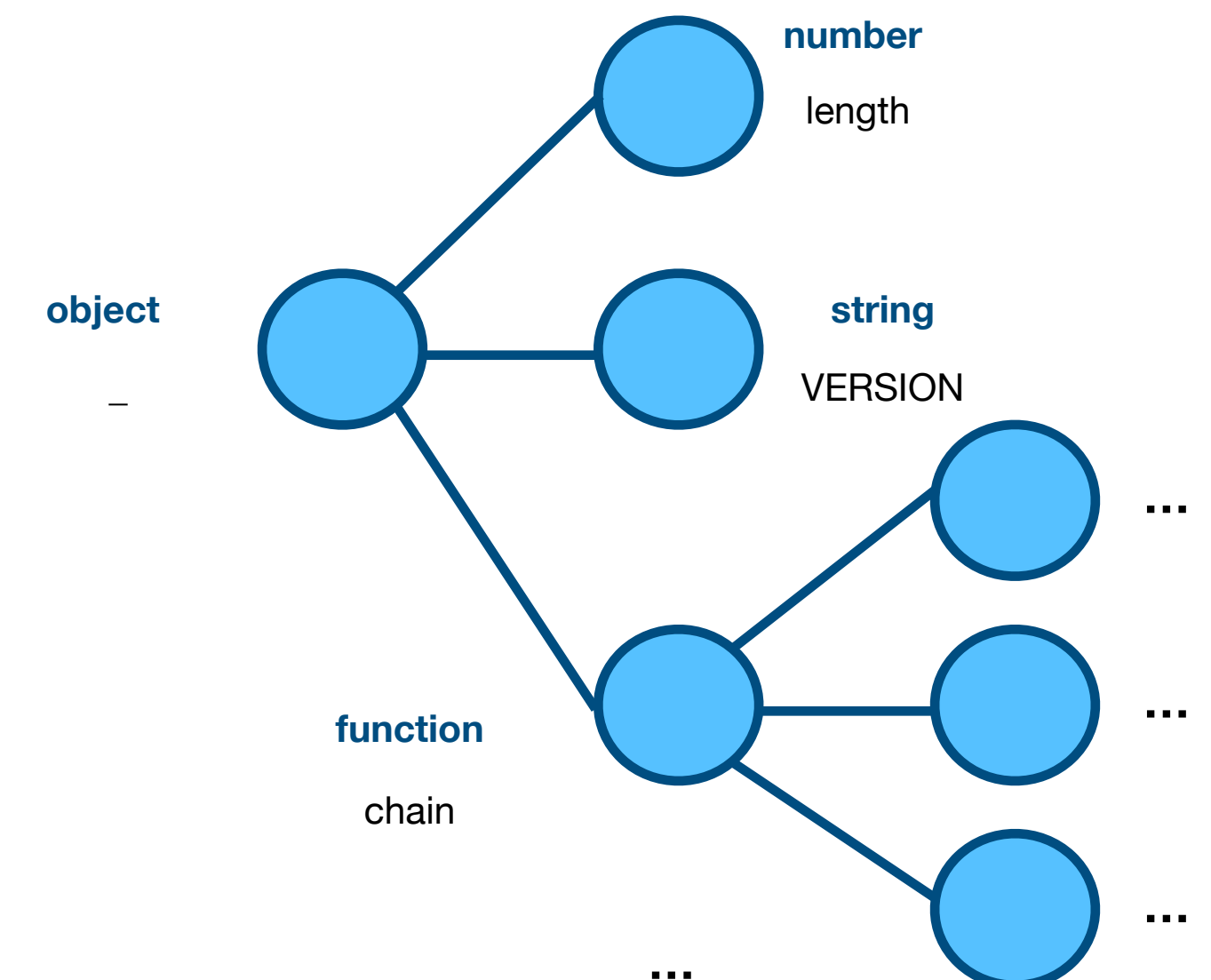
- Automatic extraction of **property patterns** in tree form

- **Limitation**

- Treating the **source code** as a **black box** can **miss libraries**

Lodash v4.17.21

```
6 // Define properties  
7 lodash.chain = function(value) {  
8   var result = lodash(value);  
9   result.__wrapped__ = value;  
10  return result;  
11 }  
12 lodash.differenceBy = ...  
13 ...  
14 lodash.VERSION = '4.17.21'  
15 // Export lodash  
16 window. = lodash;  
17 }.call(this);
```



Prior Work

- **LDC** (Library Detector for Chrome)
 - Manually collected **property patterns** at runtime

- Library Detection

```
typeof (_ = window._) == 'function'  
typeof (chain = _ && _.chain) == 'function'
```

- Version Detection

```
return { version: _.VERSION || UNKNOWN_VERSION };
```

- **PTDetector** (ASE'23)

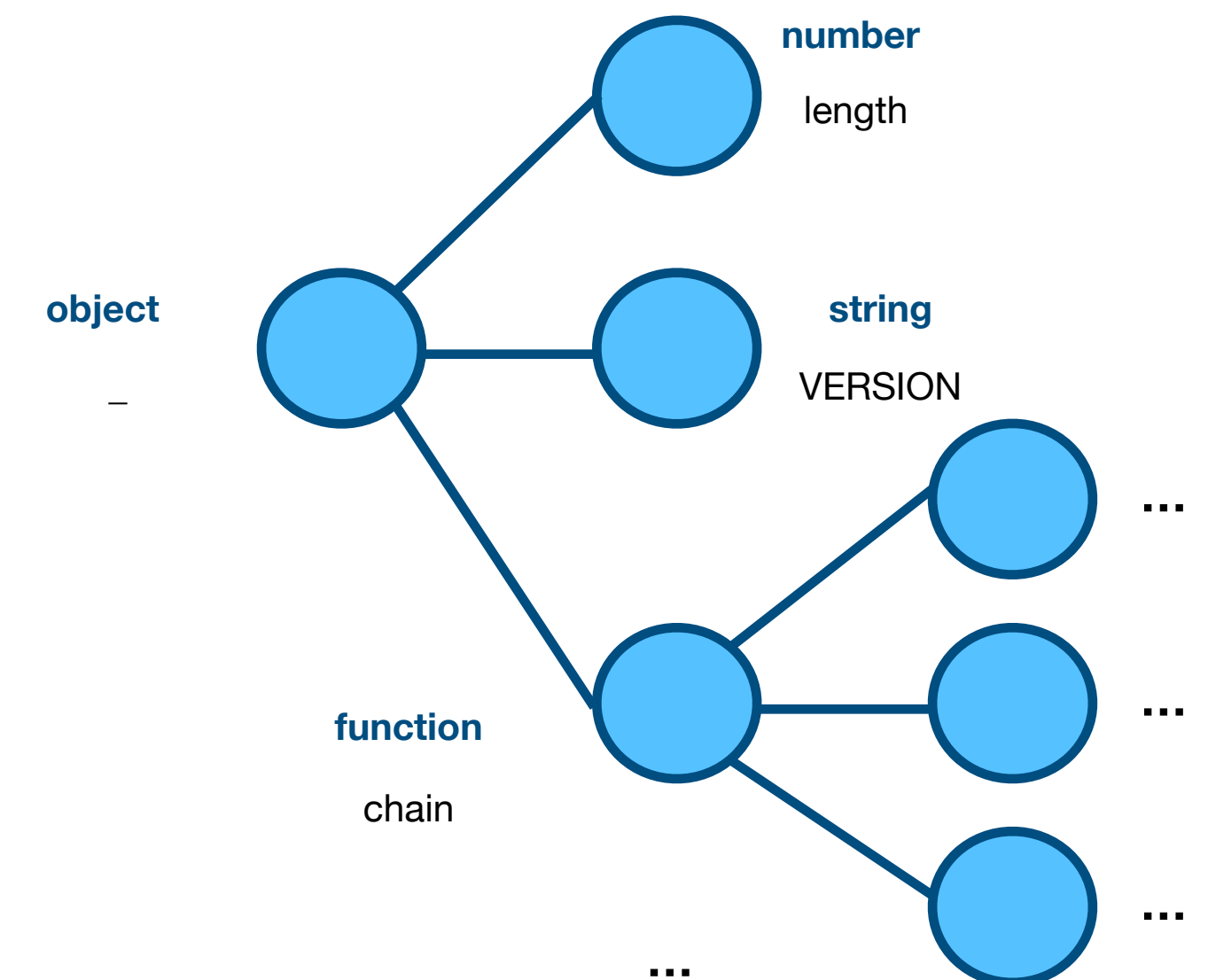
- Automatic extraction of **property patterns** in tree form

- **Limitation**

- Treating the **source code** as a **black box** can **miss libraries**

Lodash v4.17.21

```
6 // Define properties  
7 lodash.chain = function(value) {  
8   var result = lodash(value);  
9   result.__wrapped__ = value;  
10  return result;  
11 }  
12 lodash.differenceBy = ...  
13 ...  
14 lodash.VERSION = '4.17.21'  
15 // Export lodash  
16 window. = lodash;  
17 }.call(this);
```



Why does prior work treat **source code** as **blackbox**

- Bundlers *modify* and *compress* **user code** and **libraries** together using diverse **transpilers**
- It makes **difficult to detect libraries** in transpiled code in web applications

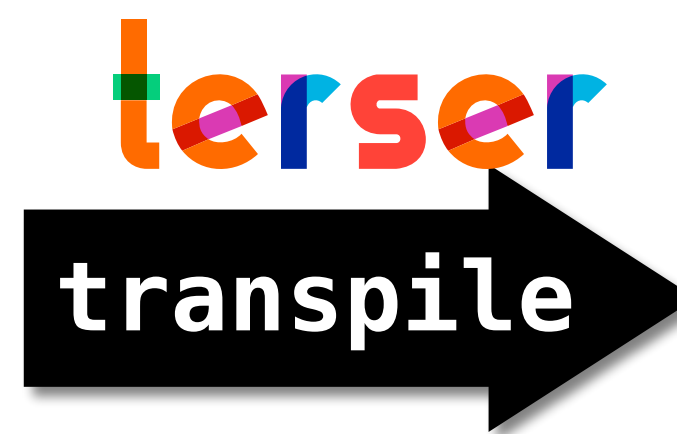


Challenges of Library Detection in Transpiled Code

```
function remove(array, predicate) {
  var result = [];
  if (!(array && array.length)) {
    return result;
  }
  var index = -1,
      indexes = [],
      length = array.length;

  predicate = getIteratee(predicate, 3);
  while (++index < length) {
    var value = array[index];
    if (predicate(value, index, array)) {
      result.push(value);
      indexes.push(index);
    }
  }
  basePullAt(array, indexes);
  return result;
}
```

"remove" function in **Lodash.js v4.17.21**



```
function f(e,r){var t=[];if(!e||!e.length)
return t;var n=-1,u=[],a=e.length;
for(r=an(r,3);++n<a;){var h=e[n];
r(h,n,e)&&(t.push(h),u.push(n))}
return bf(e,u),t}
```

Transpiled Code

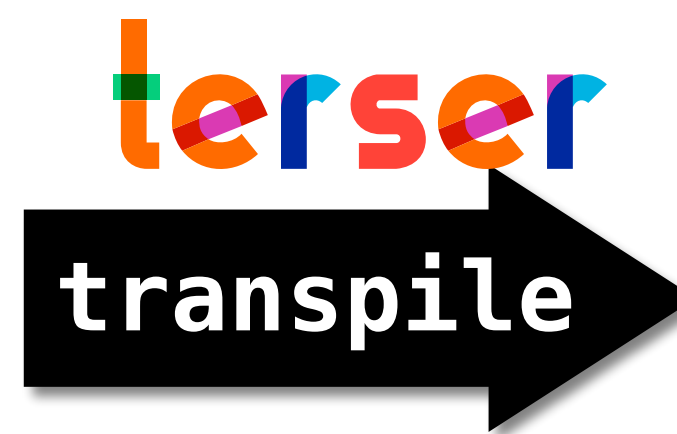
Challenges of Library Detection in Transpiled Code

```
function remove(array, predicate) {
  var result = [];
  if (!(array && array.length)) {
    return result;
  }
  var index = -1,
    indexes = [],
    length = array.length;

  predicate = getIteratee(predicate, 3);
  while (++index < length) {
    var value = array[index];
    if (predicate(value, index, array)) {
      result.push(value);
      indexes.push(index);
    }
  }
  basePullAt(array, indexes);
  return result;
}
```

"remove" function in **Lodash.js v4.17.21**

1. Mangled Names of Variables/Functions



```
function f(e,r){var t=[];if(!e||!e.length)
return t;var n=-1,u=[],a=e.length;
for(r=an(r,3);++n<a;){var h=e[n];
r(h,n,e)&&(t.push(h),u.push(n))}
return bf(e,u),t}
```

Transpiled Code

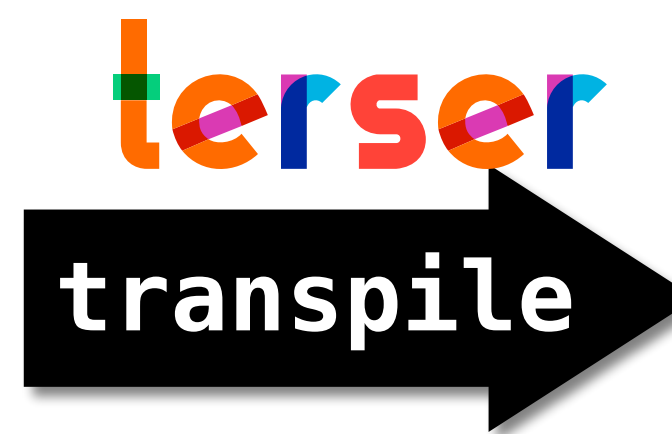
Challenges of Library Detection in Transpiled Code

```
function remove(array, predicate) {
  var result = [];
  if (!(array && array.length)) {
    return result;
  }
  var index = -1,
      indexes = [],
      length = array.length;

  predicate = getIteratee(predicate, 3);
  while (++index < length) {
    var value = array[index];
    if (predicate(value, index, array)) {
      result.push(value);
      indexes.push(index);
    }
  }
  basePullAt(array, indexes);
  return result;
}
```

"remove" function in **Lodash.js v4.17.21**

1. Mangled Names of Variables/Functions
2. Changed Control Statements



```
function f(e,r){var t=[];if(!e||!e.length)
return t;var n=-1,u=[],a=e.length;
for(r=an(r,3);++n<a;){var h=e[n];
r(h,n,e)&&(t.push(h),u.push(n))}
return bf(e,u),t}
```

Transpiled Code

What is Preserved after Transpilation?

- **Property names** are *preserved* to support JavaScript's dynamic property access

```
array['len' + 'gth'] // array.length
```

- **Execution order** between **property reads/writes** is *preserved* to maintain side effects

```
obj = {  
  get p() { console.log(1); }  
  set q() { console.log(2); }  
}
```

```
obj.p; // print 1  
obj.q = 42; // print 2
```

≠

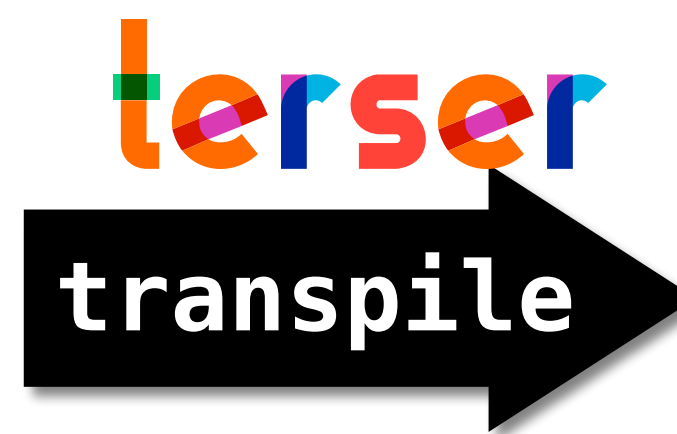
```
obj.q = 42; // print 2  
obj.p; // print 1
```

What is Preserved after Transpilation?

```
function remove(array, predicate) {
  var result = [];
  if (!(array && array.length)) {
    return result;
  }
  var index = -1,
      indexes = [],
      length = array.length;

  predicate = getIteratee(predicate, 3);
  while (++index < length) {
    var value = array[index];
    if (predicate(value, index, array)) {
      result.push(value);
      indexes.push(index);
    }
  }
  basePullAt(array, indexes);
  return result;
}
```

"remove" function in **Lodash.js v4.17.21**



```
function f(e,r){var t=[];if(!e||!e.length)
return t;var n=-1,u=[],a=e.length;
for(r=an(r,3);++n<a;){var h=e[n];
r(h,n,e)&&(t.push(h),u.push(n))}
return bf(e,u),t}
```

Transpiled Code

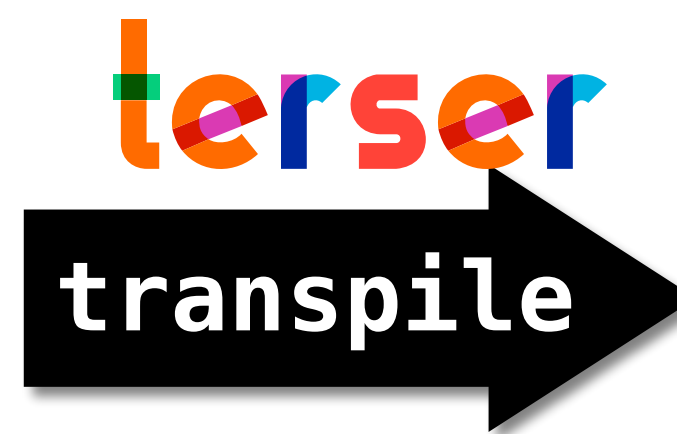
What is Preserved after Transpilation?

```
function remove(array, predicate) {
  var result = [];
  if (!(array && array.length)) {
    return result;
  }
  var index = -1,
      indexes = [],
      length = array.length;

  predicate = getIteratee(predicate, 3);
  while (++index < length) {
    var value = array[index];
    if (predicate(value, index, array)) {
      result.push(value);
      indexes.push(index);
    }
  }
  basePullAt(array, indexes);
  return result;
}
```

"remove" function in **Lodash.js v4.17.21**

1. Property Names



```
function f(e,r){var t=[];if(!e||!e.length)
return t;var n=-1,u=[],a=e.length;
for(r=an(r,3);++n<a;){var h=e[n];
r(h,n,e)&&(t.push(h),u.push(n))}
return bf(e,u),t}
```

Transpiled Code

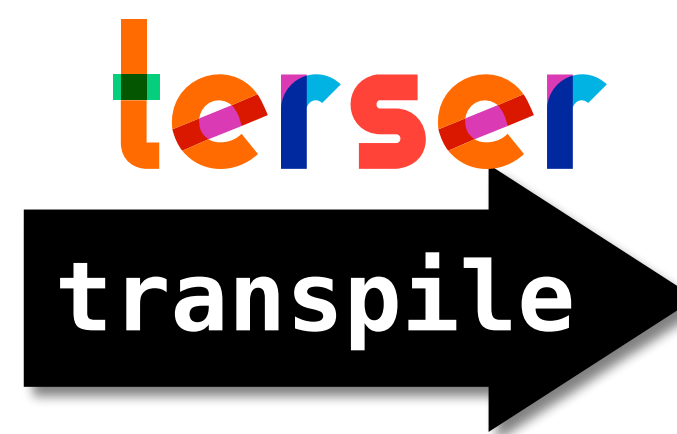
What is Preserved after Transpilation?

```
function remove(array, predicate) {  
  var result = [];  
  if (!(array && array.length)) {  
    return result;  
  }  
  var index = -1,  
      indexes = [],  
      length = array.length;  
  
  predicate = getIteratee(predicate, 3);  
  while (++index < length) {  
    var value = array[index];  
    if (predicate(value, index, array)) {  
      result.push(value);  
      indexes.push(index);  
    }  
  }  
  basePullAt(array, indexes);  
  return result;  
}
```

"remove" function in **Lodash.js v4.17.21**

1. Property Names

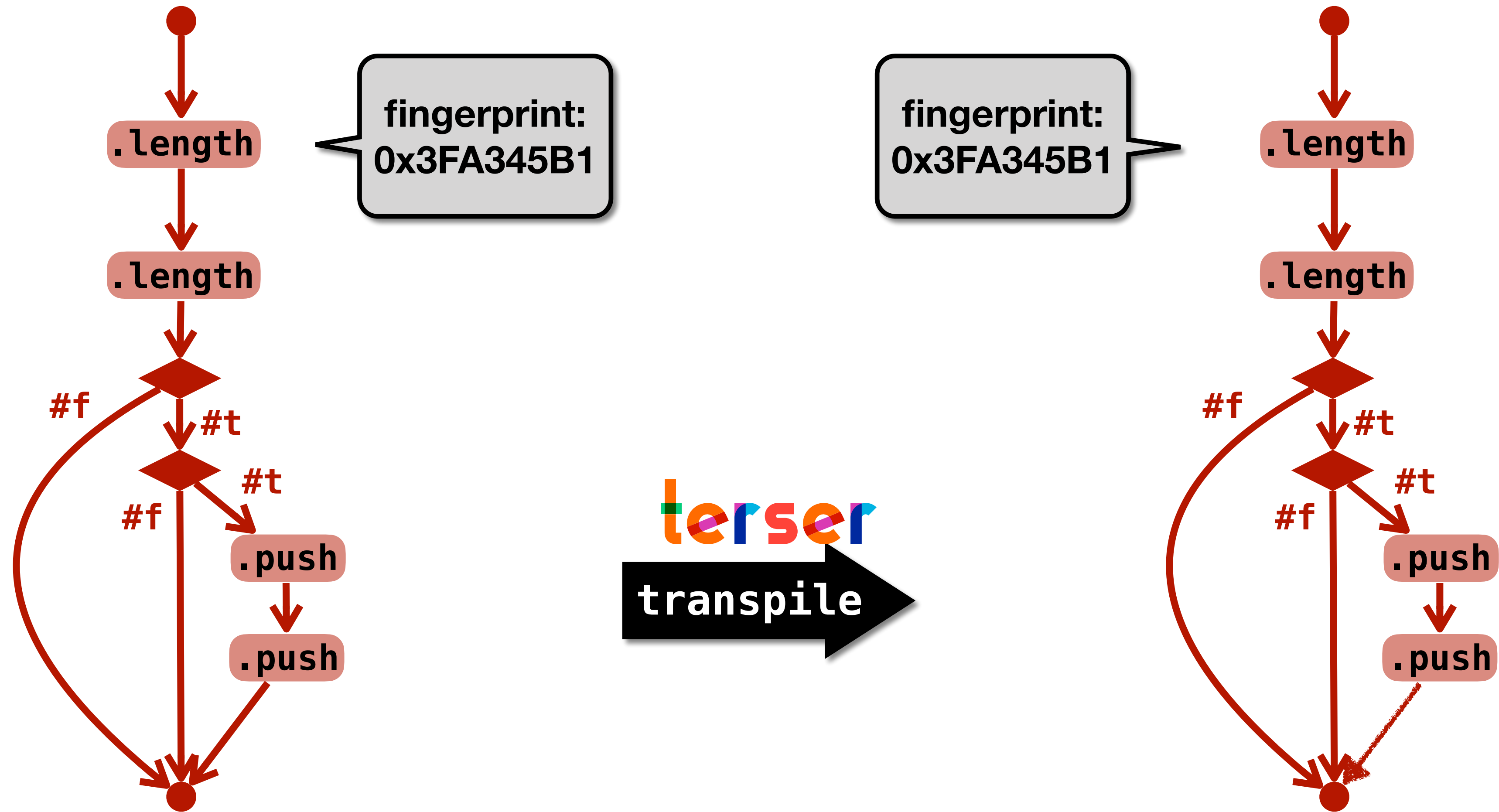
2. Execution Order between **Property Reads/Writes**



```
function f(e,r){var t=[];if(!e||e.length)  
return t;var n=-1,u=[],a=e.length;  
for(r=an(r,3);++n<a;){var h=e[n];  
r(h,n,e)&&t.push(h),u.push(n)}}  
#f  
return bf(e,u);}
```

Transpiled Code

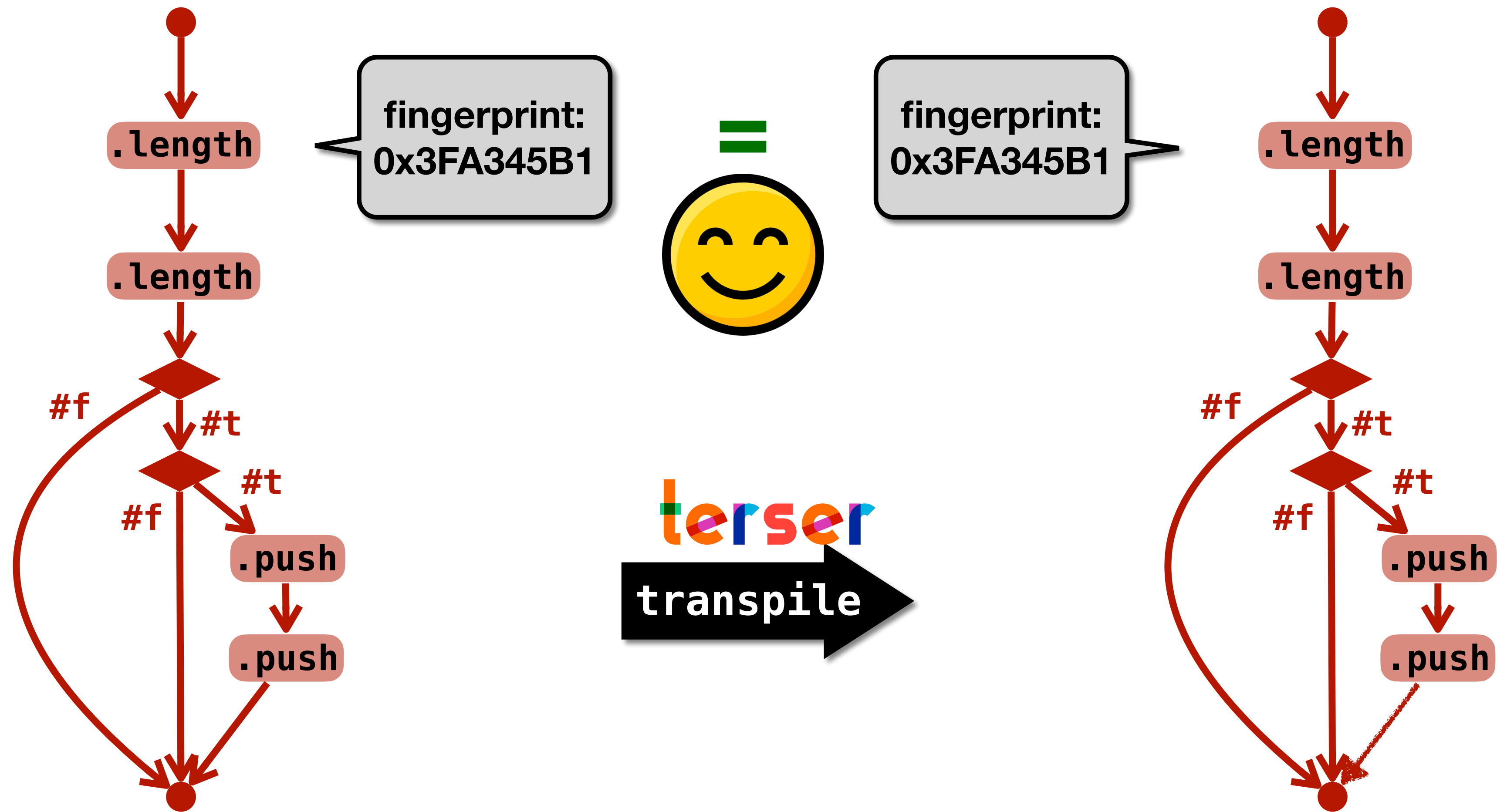
Property-Order Graph (POG)



"remove" function in Lodash.js v4.17.21

Transpiled Code

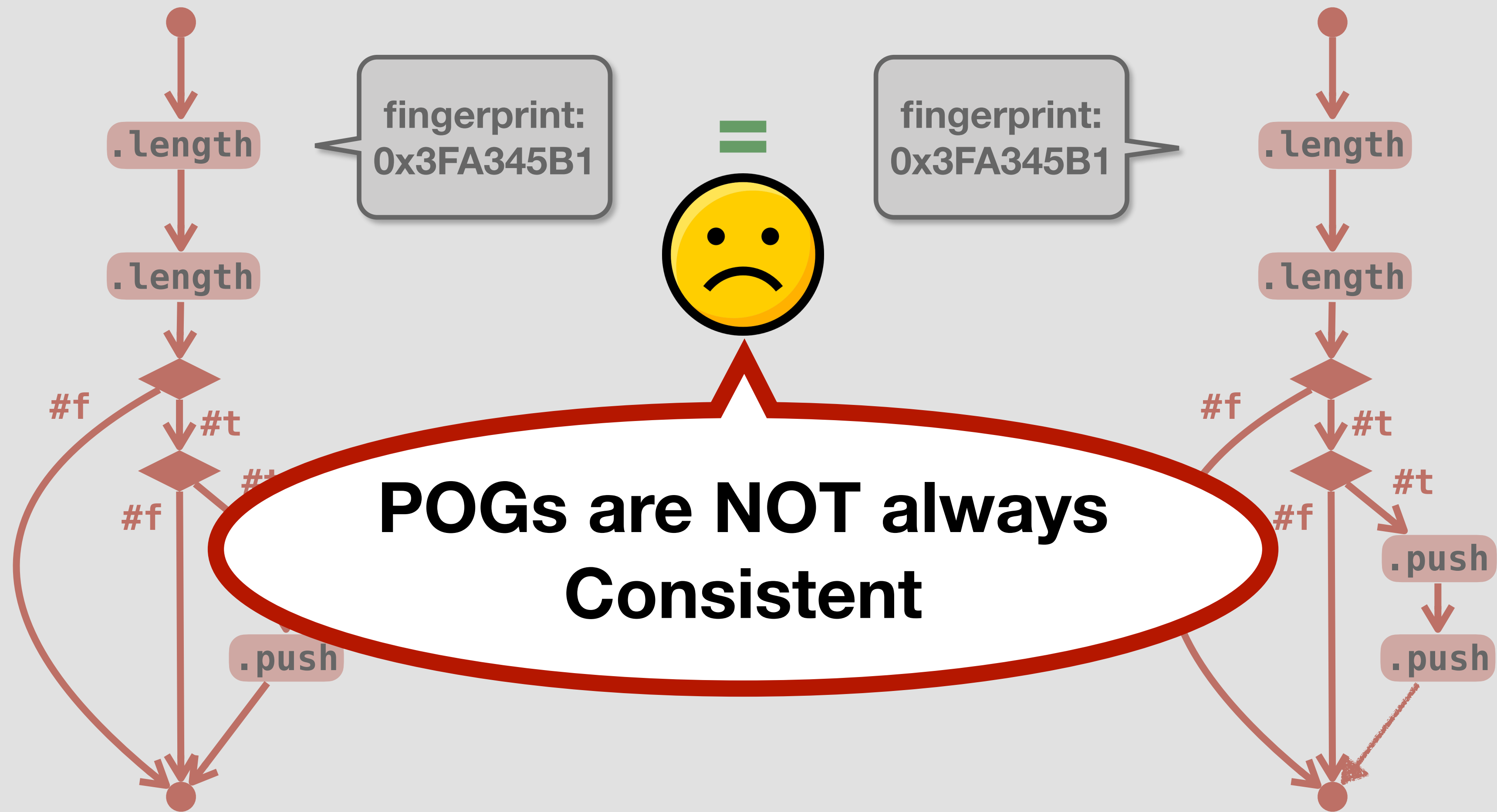
Property-Order Graph (POG)



"remove" function in Lodash.js v4.17.21

Transpiled Code

Property-Order Graph (POG)



"remove" function in Lodash.js v4.17.21

Transpiled Code

Problem - POGs are NOT always Consistent

1. Flipped branch

```
if (!y) x.p; else x.q;
```

transpile

```
if (y) x.q; else x.p;
```

2. Merged branch

```
while (x.p) {  
    if (x.q) break;  
}
```

transpile

```
for (; x.p && !x.q; );
```

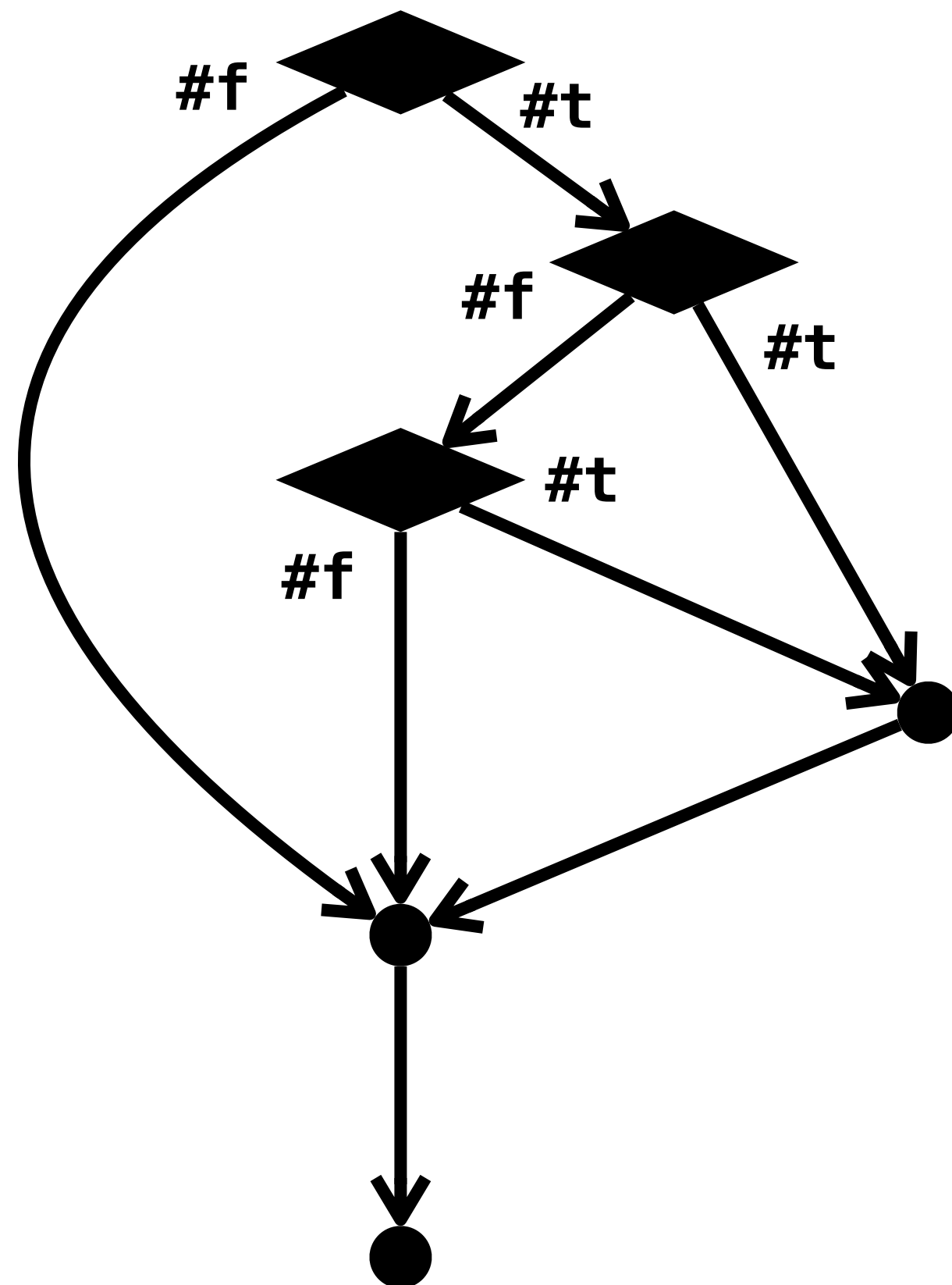
3. Combined duplicate code

```
if (y) x.p = x.q = e1;  
else x.p = x.q = e2;
```

transpile

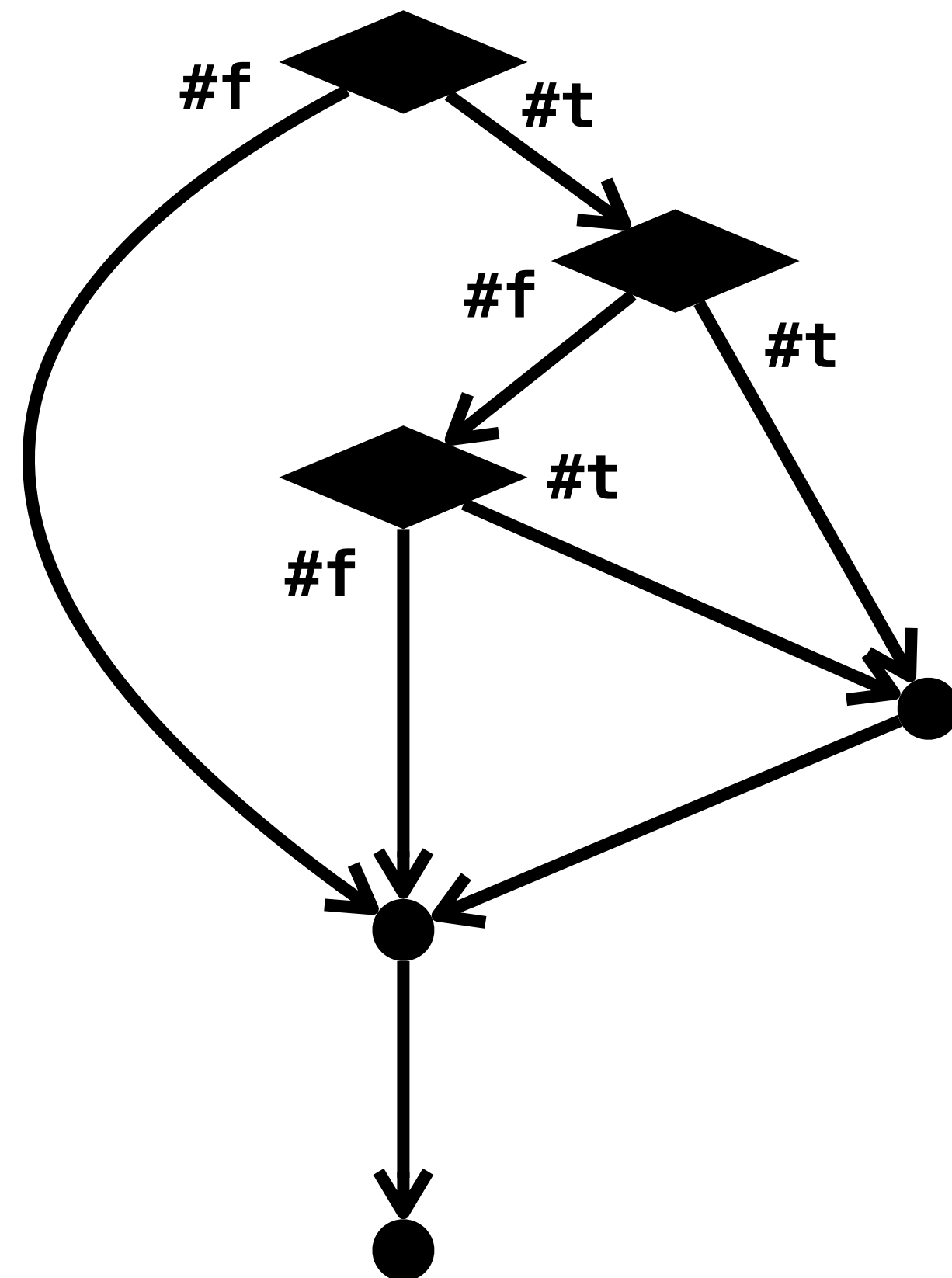
```
x.p = x.q = y ? e1 : e2;
```

Solution - Refine POGs using Path-Sensitive Truthy Analysis

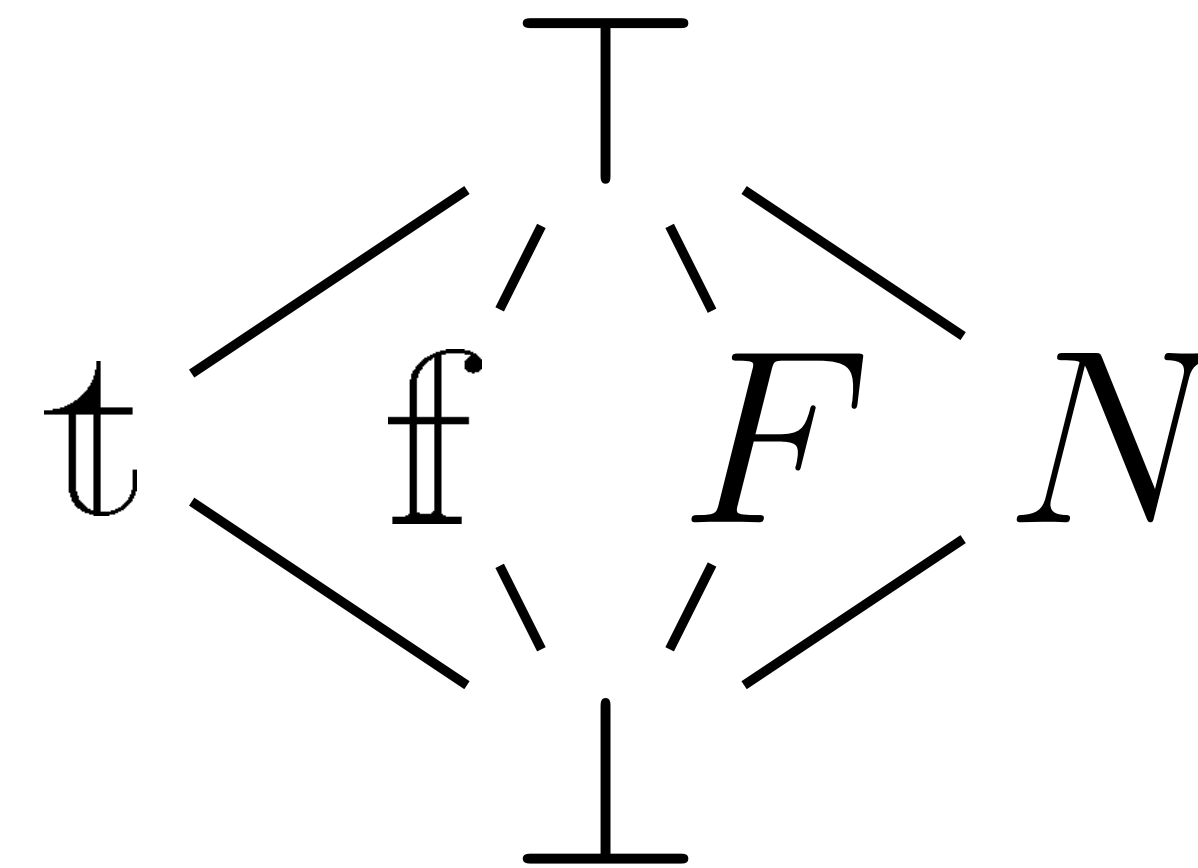


Track truthiness of variables along execution paths!
(Path = Control flow from each branch)

Solution - Refine POGs using Path-Sensitive Truthy Analysis

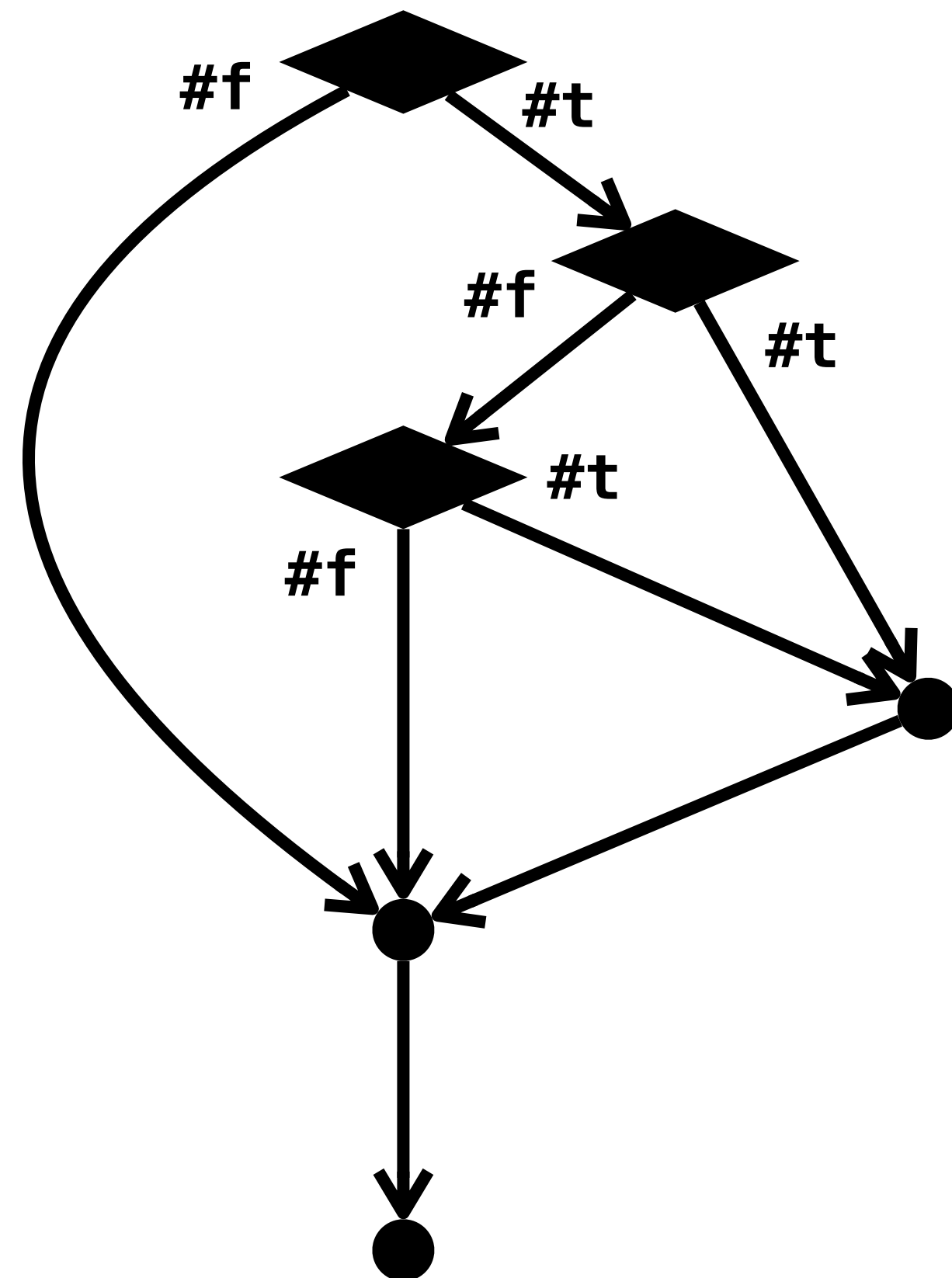


Track truthiness of variables along execution paths!
(Path = Control flow from each branch)

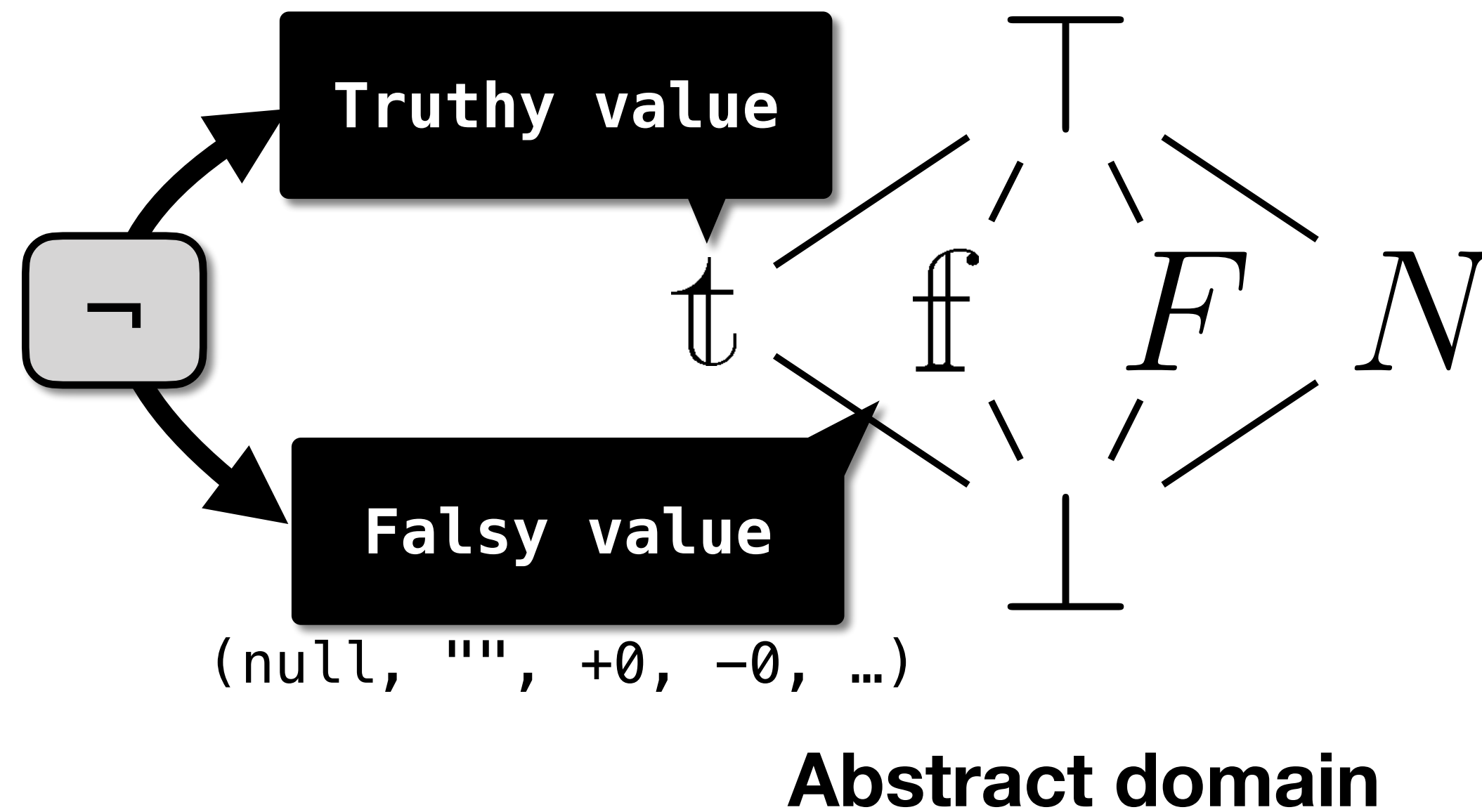


Abstract domain

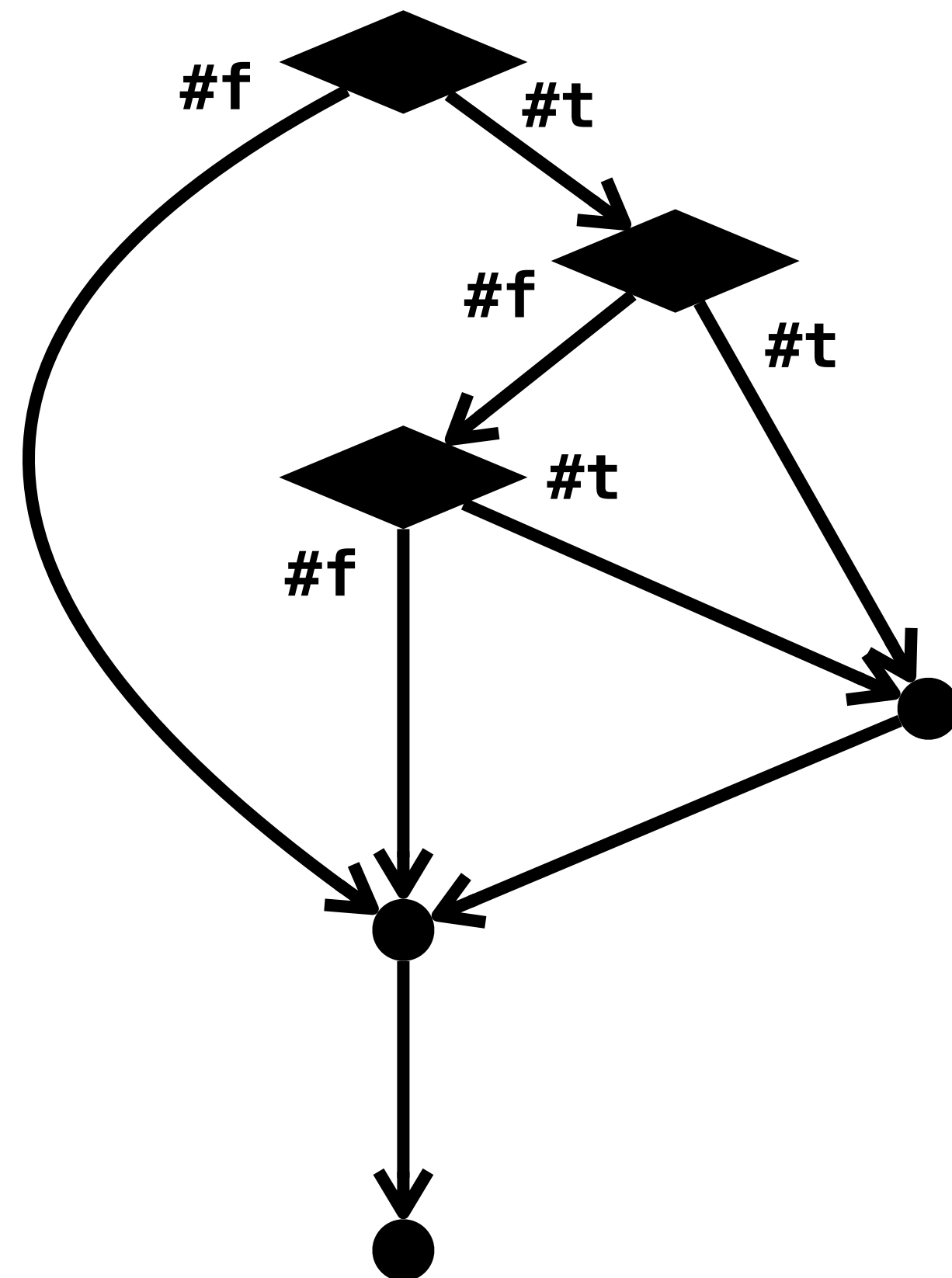
Solution - Refine POGs using Path-Sensitive Truthy Analysis



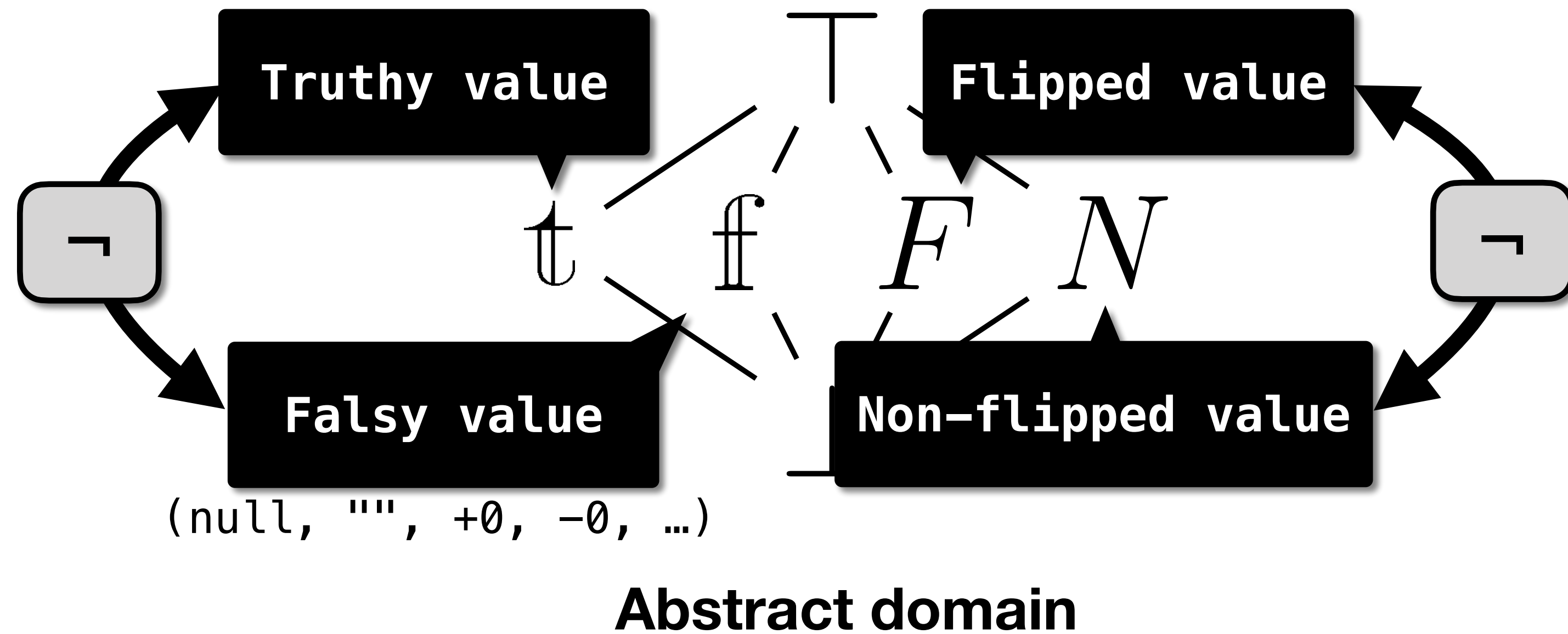
Track truthiness of variables along execution paths!
(Path = Control flow from each branch)



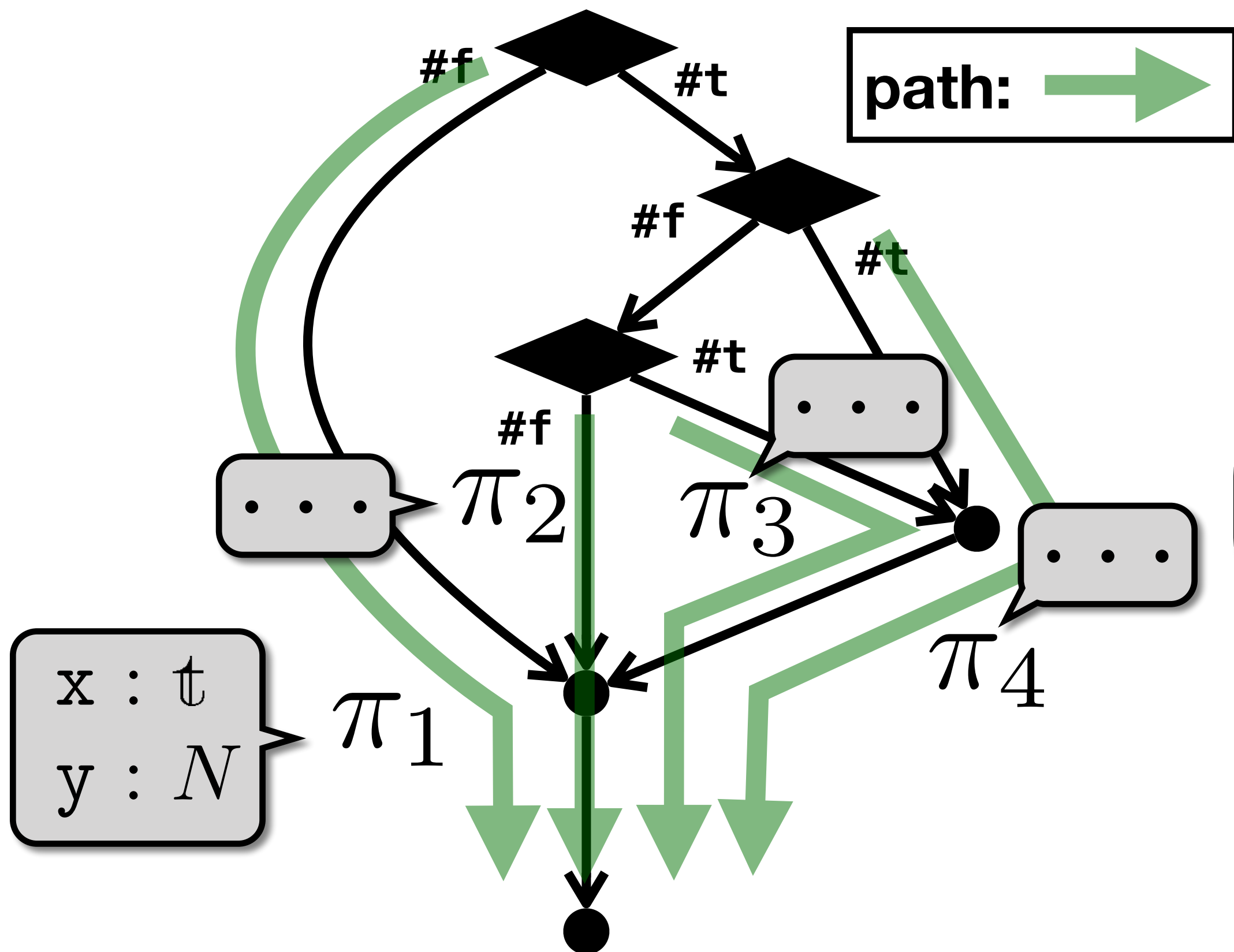
Solution - Refine POGs using Path-Sensitive Truthy Analysis



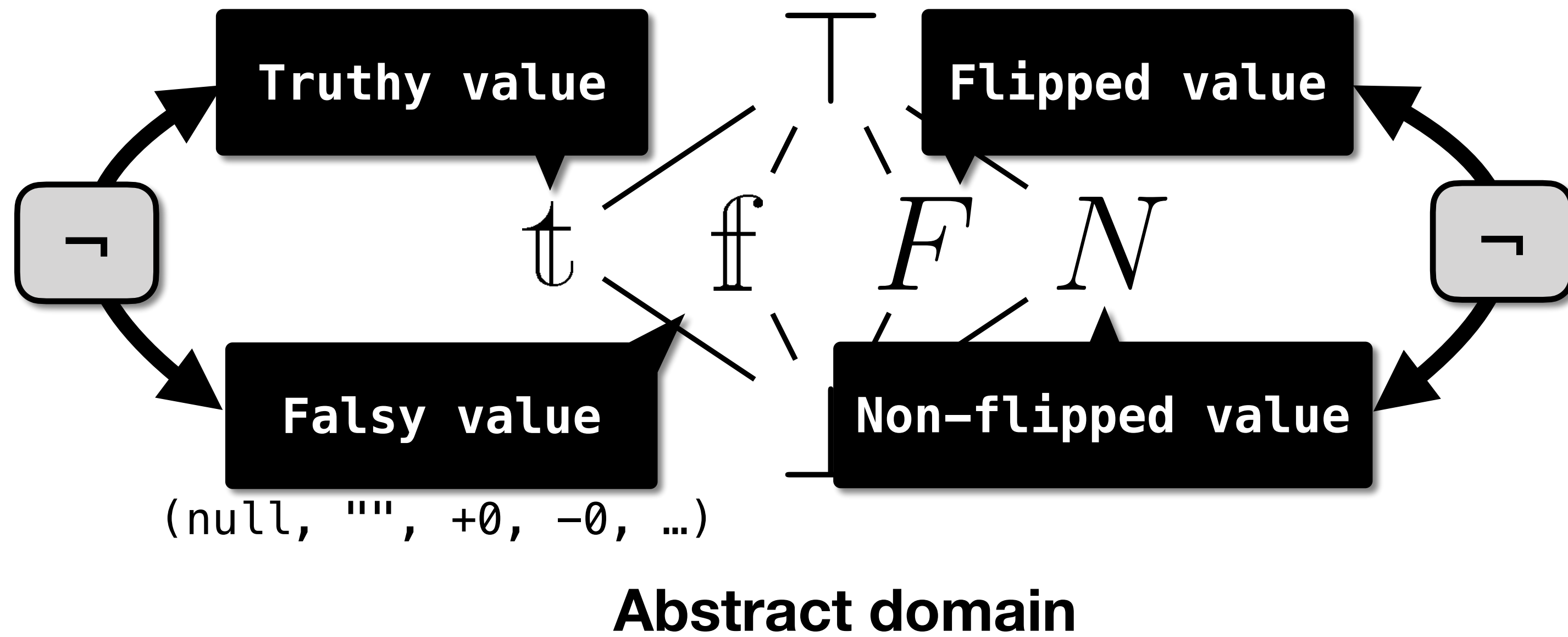
Track truthiness of variables along execution paths!
(Path = Control flow from each branch)



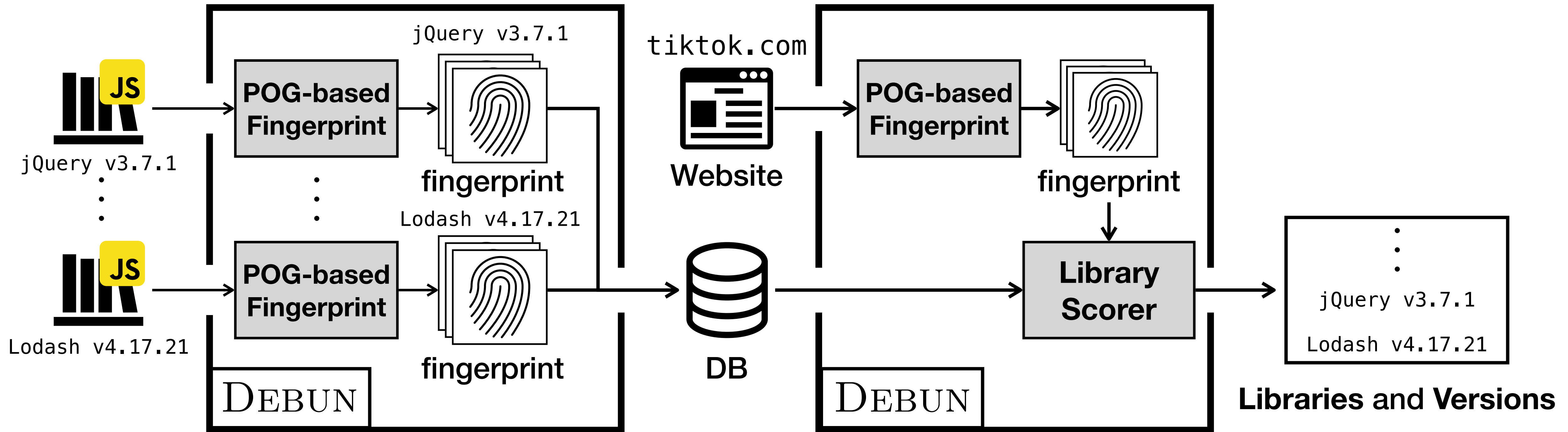
Solution - Refine POGs using Path-Sensitive Truthy Analysis



Track truthiness of variables along execution paths!
 (Path = Control flow from each branch)



Debun - A POG-based Library Detector



1. Library Fingerprint Database Construction

2. Library and Version Detection

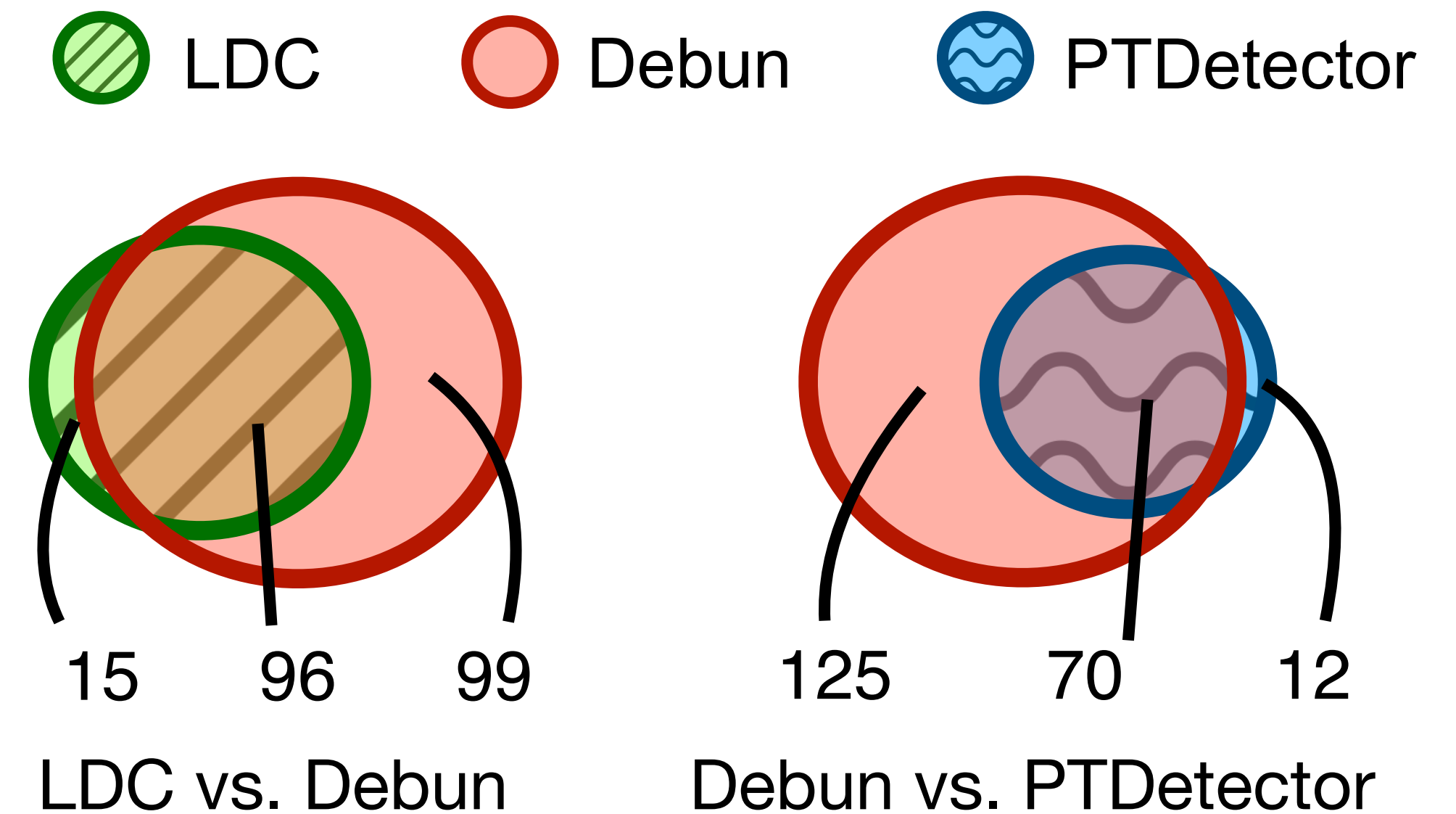
Evaluation

- Constructed fingerprint database from **8,256 versions** of **78 libraries** registered on cdnjs
- Evaluated on **68 crawlable websites** from the top 100 high-traffic websites
- **RQ1) Library detection** performance compared to existing techniques
- **RQ2) Library version detection** performance compared to existing techniques
- **RQ3) Ablation study** - effectiveness of analysis-based graph refinement

RQ1) Library Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	PTDETECTOR	DEBUN
TP	111	82	195
FP	3	9	7
FN	112	141	28
Precision	97.37%	90.11%	96.53%
Recall	49.78%	36.77%	87.44%
F1-score	65.88%	52.23%	91.76%

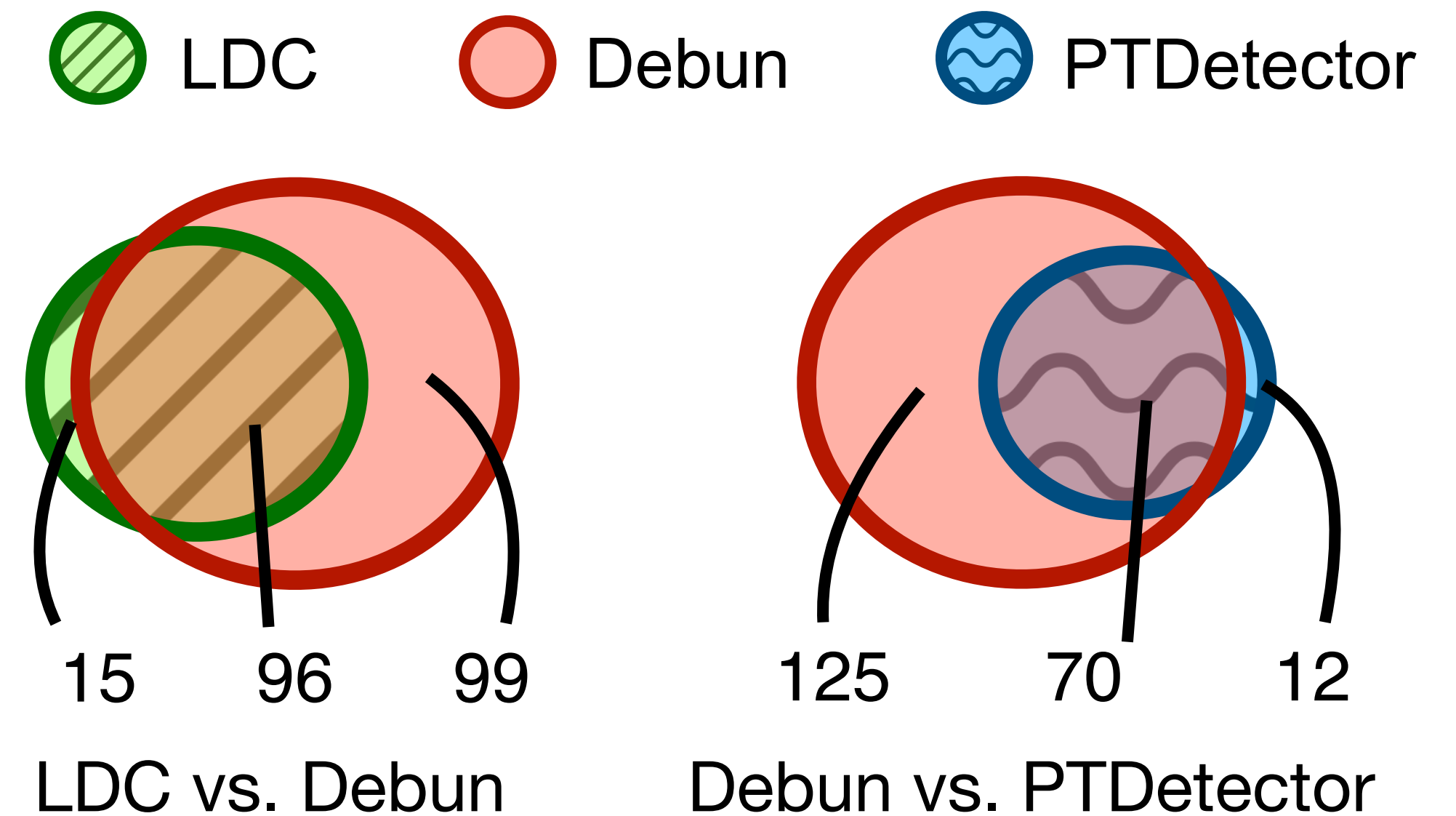


RQ1) Library Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	PTDETECTOR	DEBUN
TP	111	82	195
FP	3	9	7
FN	112	141	28
Precision	97.37%	90.11%	96.53%
Recall	49.78%	36.77%	87.44%
F1-score	65.88%	52.23%	91.76%

+25.88%p



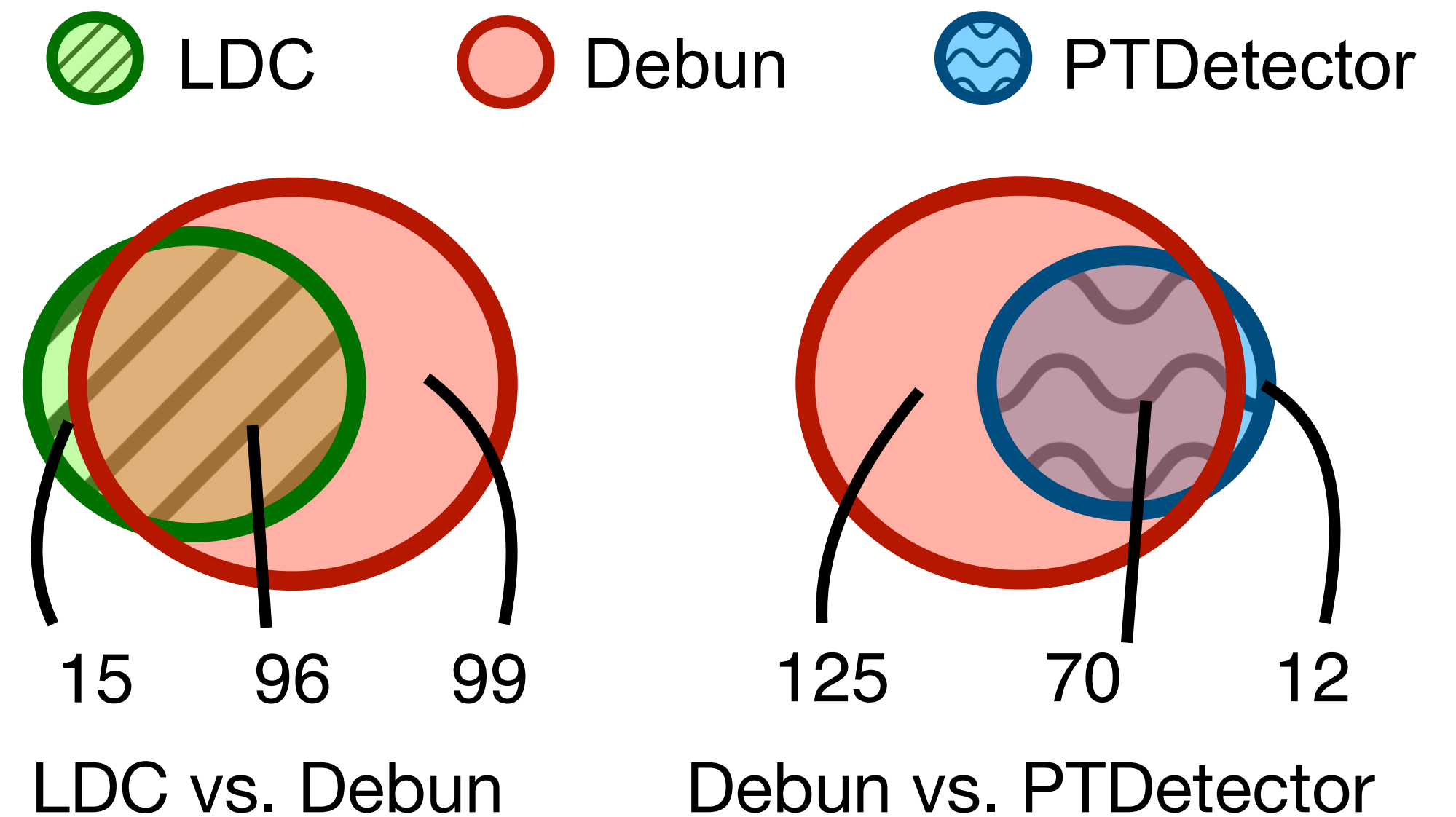
RQ1) Library Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	PTDETECTOR	DEBUN
TP	111	82	195
FP	3	9	7
FN	112	141	28
Precision	97.37%	90.11%	95.53%
Recall	49.78%	36.77%	87.44%
F1-score	65.88%	52.23%	91.76%

+25.88%p

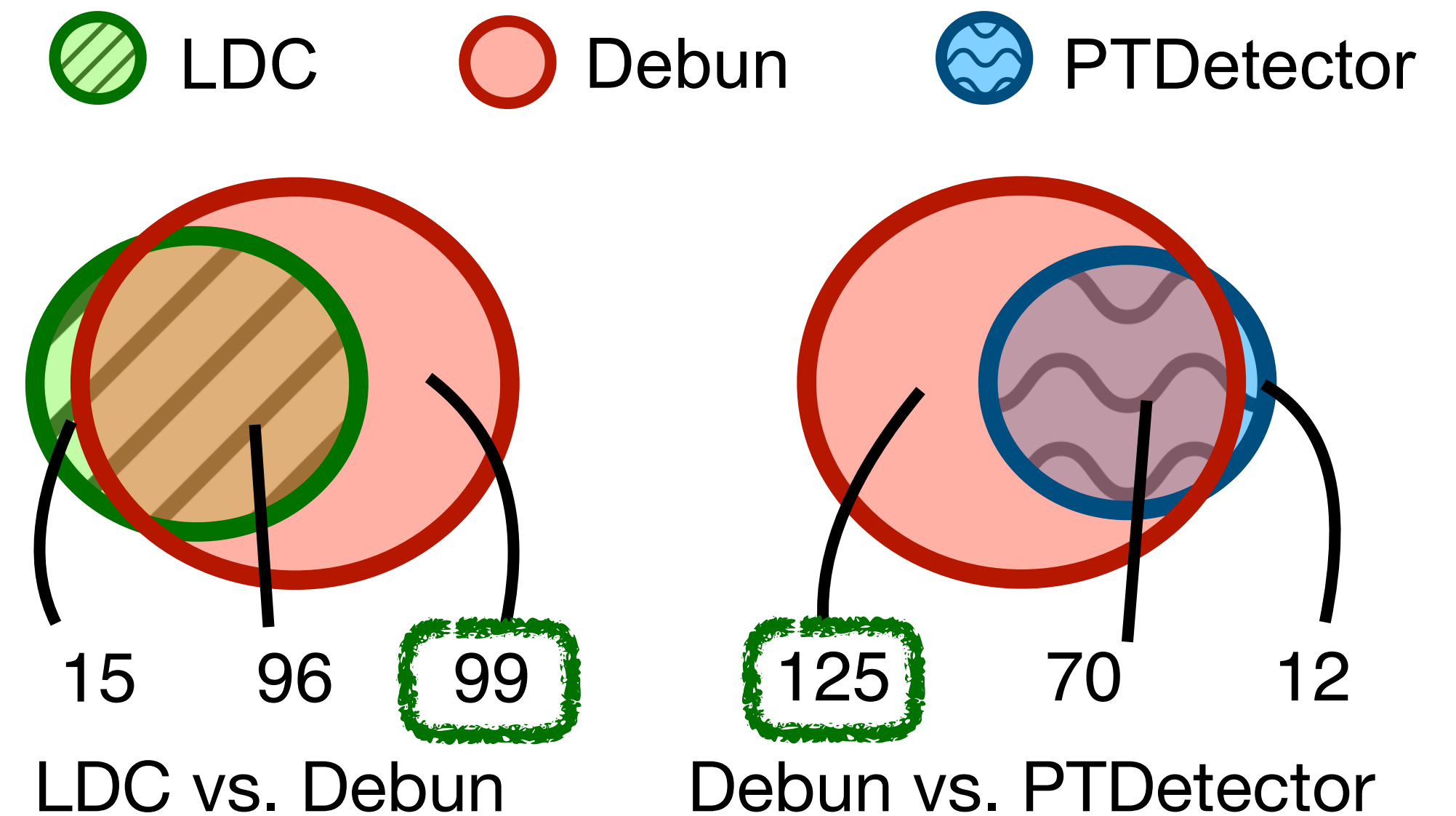
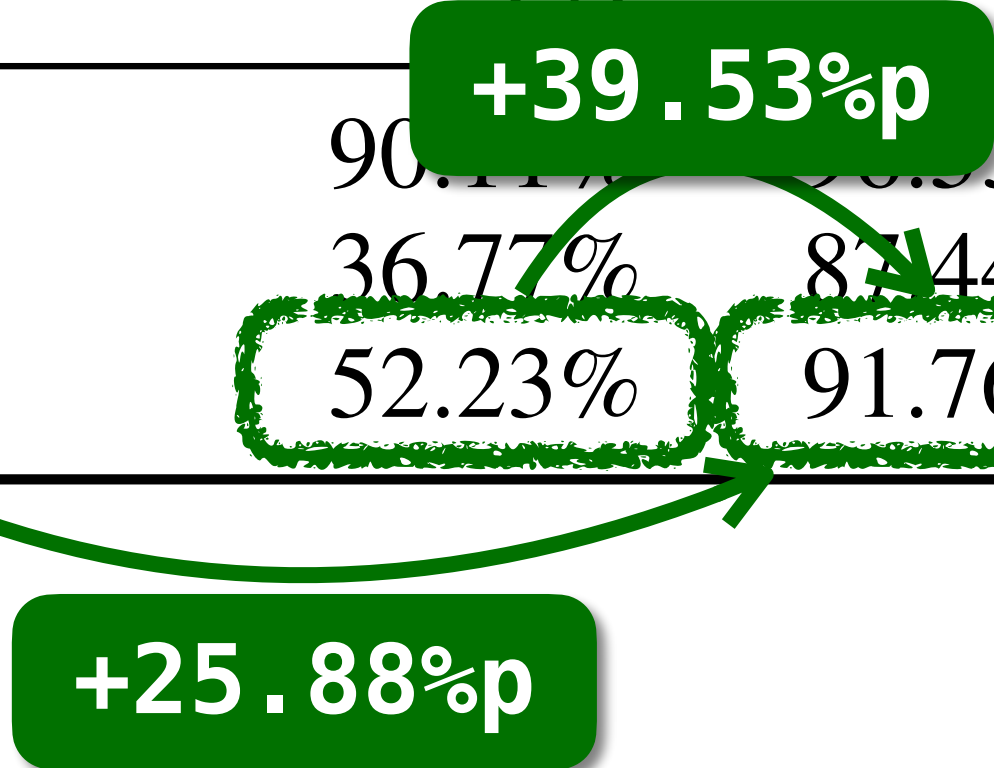
+39.53%p



RQ1) Library Detection

- Average detection time: **1,009 ms per website**

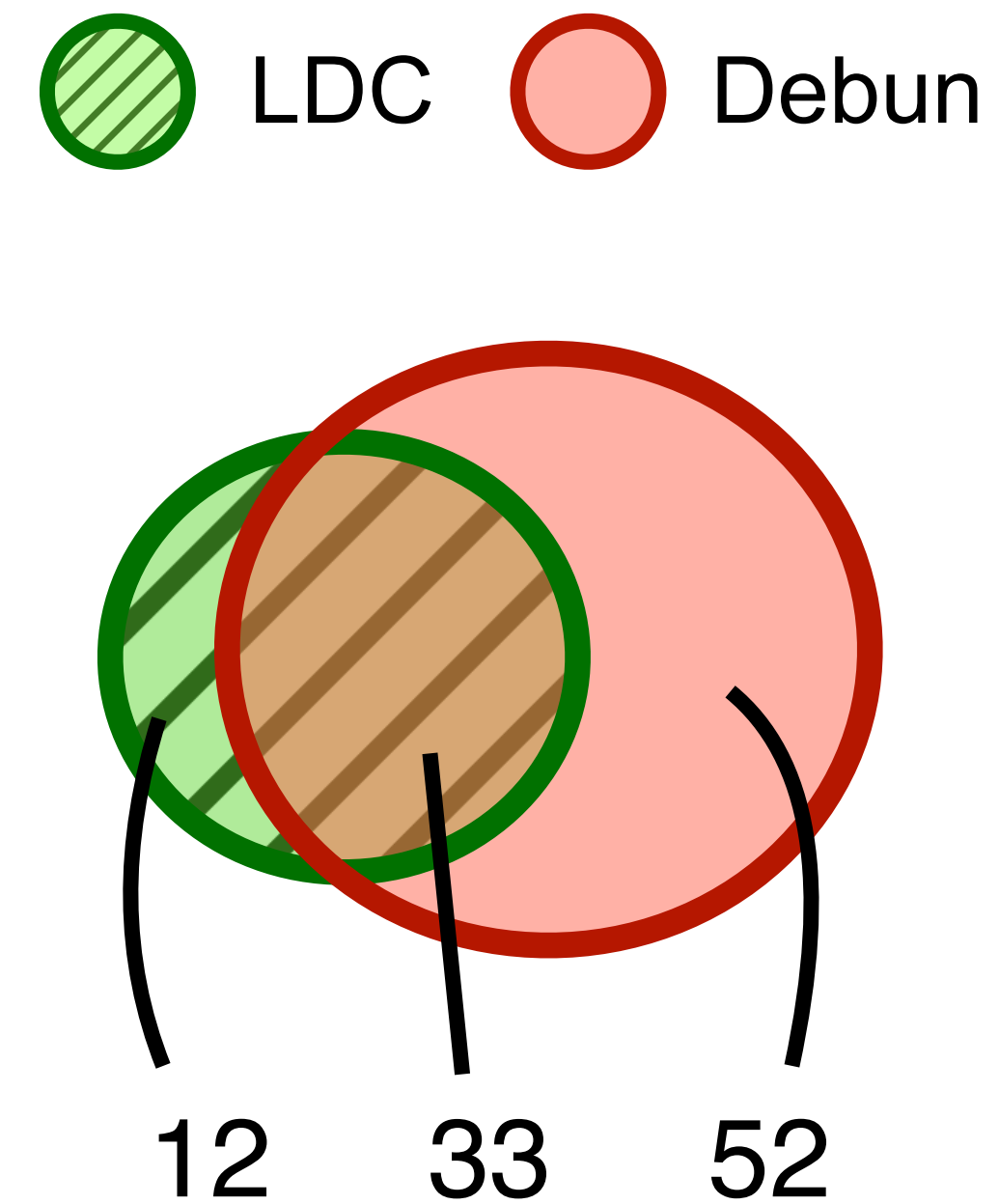
Metric	LDC	PTDETECTOR	DEBUN
TP	111	82	195
FP	3	9	7
FN	112	141	28
Precision	97.37%	90.11%	95.53%
Recall	49.78%	36.77%	87.44%
F1-score	65.88%	52.23%	91.76%



RQ2) Library Version Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	DEBUN
TP	45	85
FP	0	16
FN	60	20
Precision	100.00%	84.16%
Recall	42.86%	80.95%
F1 score	60.00%	82.52%

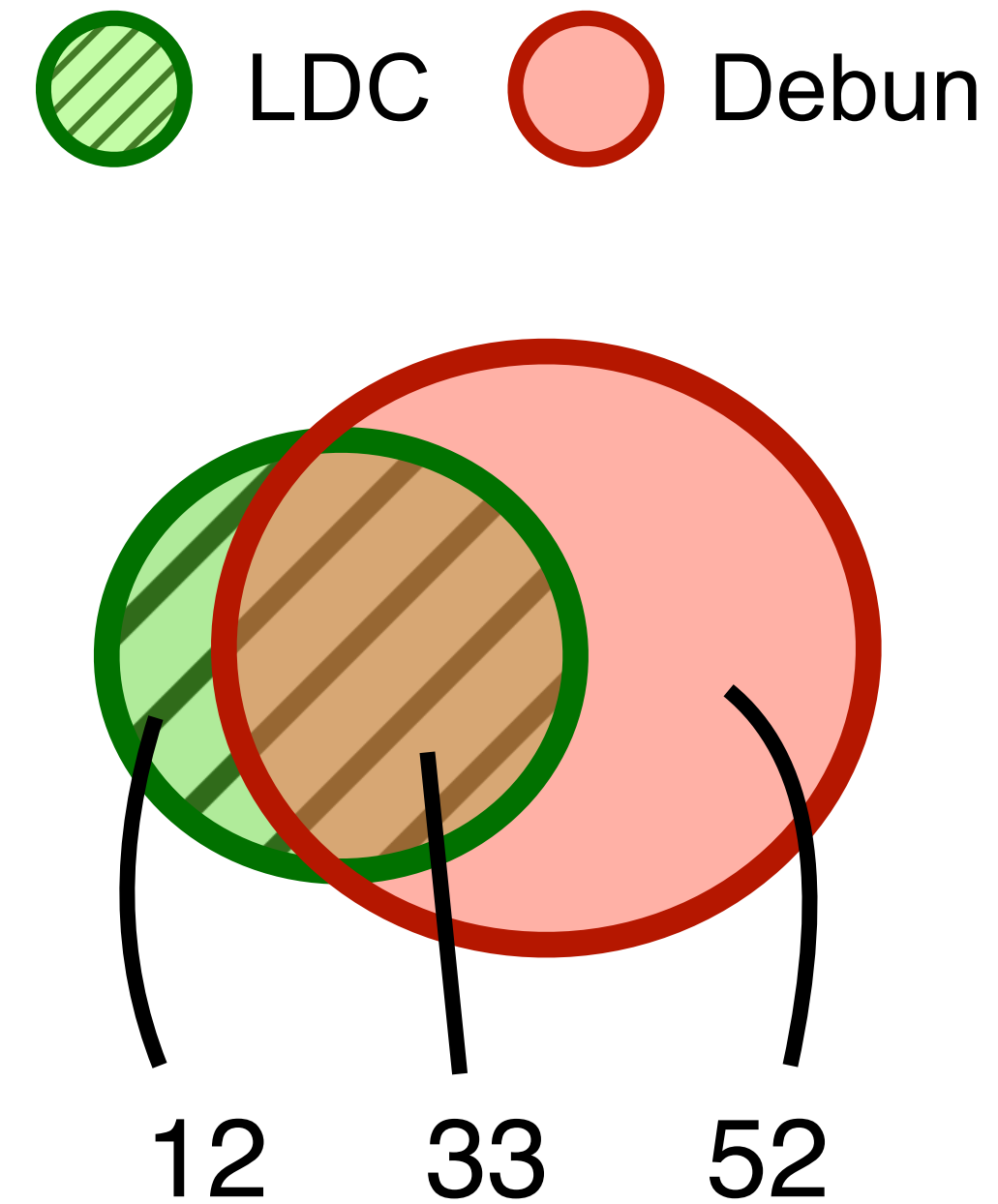


RQ2) Library Version Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	DEBUN
TP	45	85
FP	0	16
FN	60	20
Precision	100.00%	84.16%
Recall	42.86%	80.95%
F1 score	60.00%	82.52%

+22.52%p

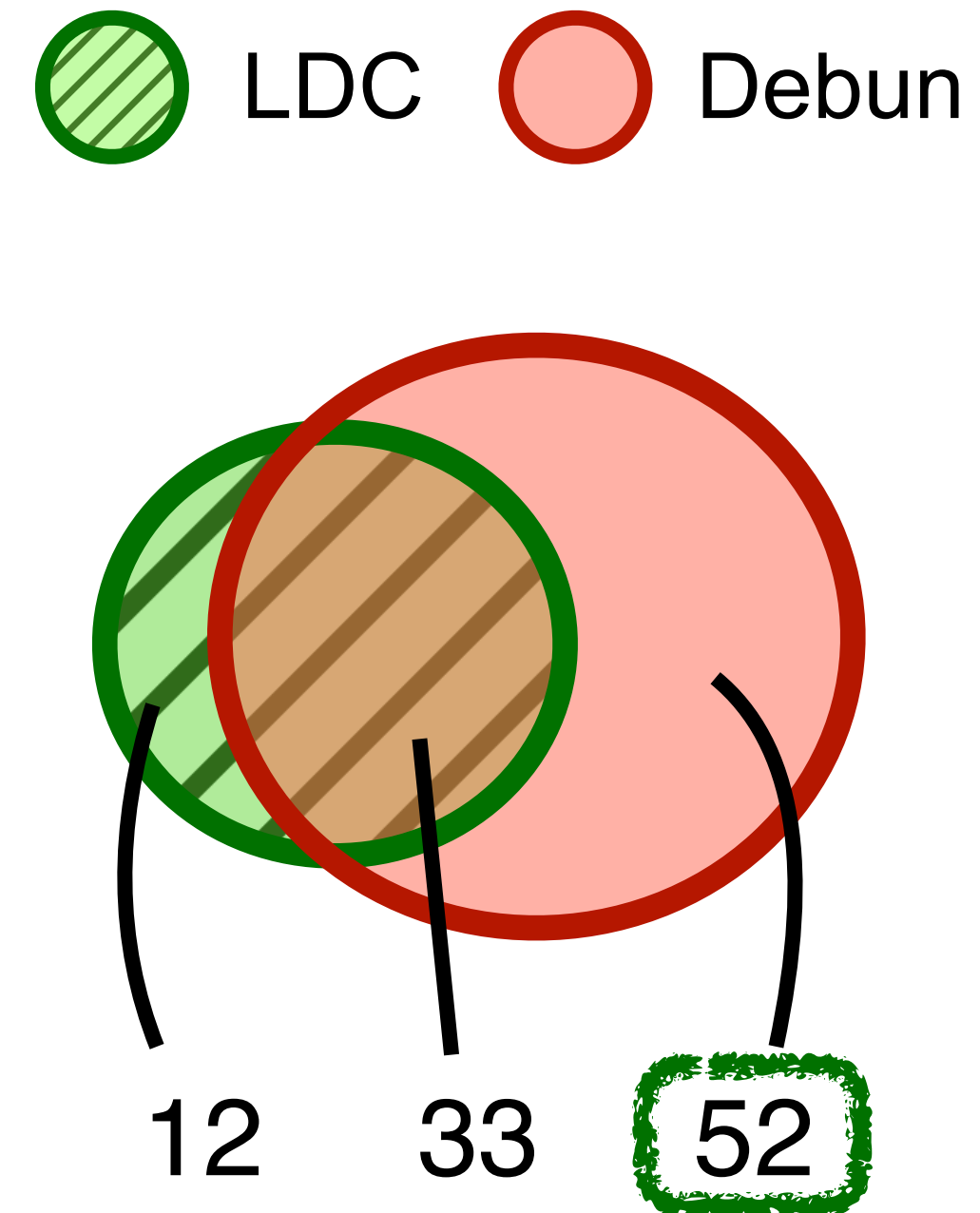


RQ2) Library Version Detection

- Average detection time: **1,009 ms per website**

Metric	LDC	DEBUN
TP	45	85
FP	0	16
FN	60	20
Precision	100.00%	84.16%
Recall	42.86%	80.95%
F1 score	60.00%	82.52%

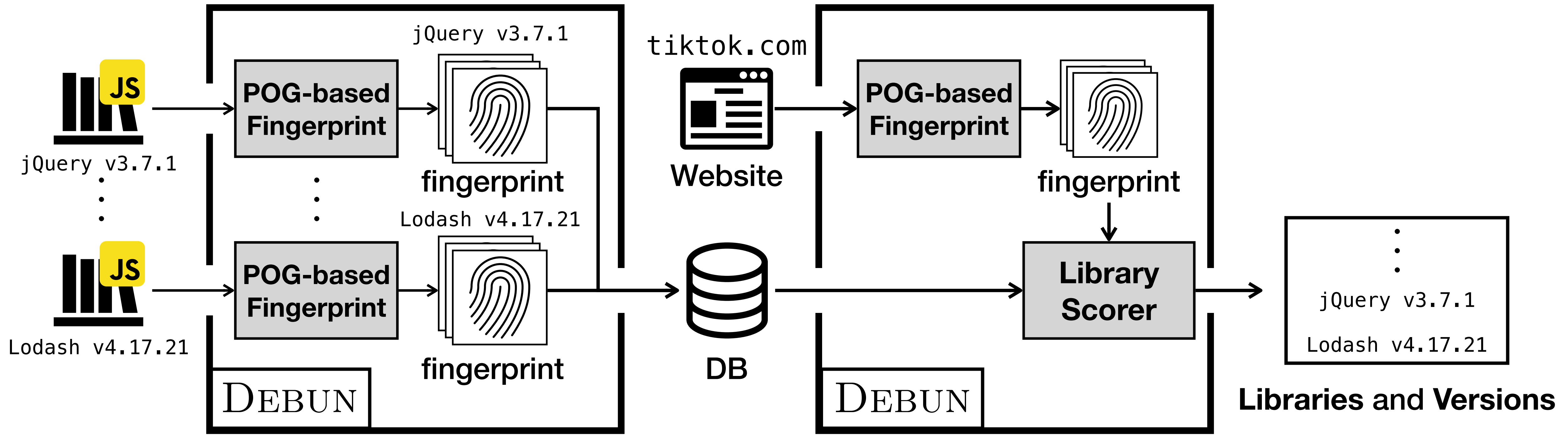
+22.52%p



Debun (Artifact)



Video



1. Library Fingerprint Database Construction

2. Library and Version Detection