Precise Dataflow Analysis of Event-Driven Applications

Ming-Ho Yee, Ayaz Badouraly, Ondřej Lhoták, Frank Tip, Jan Vitek
January 23, 2020

```
var fs = require('fs');
var sum;
fs.readdir('.', function f(err, filenames) {
  if (err) throw err;
  sum = 0;
  filenames.forEach(function g(fn) {
    fs.stat('./' + fn, function h(err, stats) {
      if (err) throw err;
      var size = stats.size;
      sum += size;
      console.log(fn + ': ' + size);
      console.log('sum: ' + sum);
   });
 });
});
console.log('done');
```

```
var fs = require('fs');
var sum;
fs.readdir('.', function f(err, filenames) {
  if (err) throw err;
  sum = 0;
  filenames.forEach(function g(fn) {
    fs.stat('./' + fn, function h(err, stats) {
      if (err) throw err;
      var size = stats.size;
      sum += size;
      console.log(fn + ': ' + size);
      console.log('sum: ' + sum);
console.log('done');
```

```
var fs = require('fs');
var sum;
fs.readdir('.', function f(err, filenames) {
  if (err) throw err;
  sum = 0;
  filenames.forEach(function g(fn) {
    fs.stat('./' + fn, function h(err, stats) {
      if (err) throw err;
      var size = stats.size;
      sum += size;
      console.log(fn + ': ' + size);
      console.log('sum: ' + sum);
console.log('done');
```

```
var fs = require('fs');
var sum;
fs.readdir('.', function f(err, filenames) {
  if (err) throw err;
  sum = 0;
  filenames.forEach(function g(fn) {
    fs.stat('./' + fn, function h(err, stats) {
      if (err) throw err;
      var size = stats.size;
      sum += size;
      console.log(fn + ': ' + size);
      console.log('sum: ' + sum);
console.log('done');
```

```
var fs = require('fs');
var sum;
fs.readdir('.', function f(err, filenames) {
  if (err) throw err;
  sum = 0;
  filenames.forEach(function g(fn) {
    fs.stat('./' + fn, function h(err, stats) {
      if (err) throw err;
      var size = stats.size;
      sum += size;
      console.log(fn + ': ' + size);
      console.log('sum: ' + sum);
console.log('done');
```

 $\langle e, f \rangle \in M$ — map of events to functions $f \in Q$ — queue of functions

```
\langle e, f \rangle \in M – map of events to functions f \in Q – queue of functions
```

- Register function f on event e
 - Add $\langle e, f \rangle$ to M

 $\langle e, f \rangle \in M$ — map of events to functions $f \in Q$ — queue of functions

- Register function f on event e
 - Add $\langle e, f \rangle$ to M
- **Emit** event *e*
 - Look up $\langle e, f \rangle$ in M, add f to Q

 $\langle e, f \rangle \in M$ – map of events to functions $f \in Q$ – queue of functions

- Register function f on event e
 - Add $\langle e, f \rangle$ to M
- **Emit** event *e*
 - Look up $\langle e, f \rangle$ in M, add f to Q
- **Invoke** function *f*
 - When the call stack is empty, remove f from Q and invoke f

IFDS and IDE Frameworks

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \sqcap \rangle$$

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \sqcap \rangle$$

• $G^* = \langle N^*, E^* \rangle$ is the supergraph

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \sqcap \rangle$$

- $G^* = \langle N^*, E^* \rangle$ is the supergraph
- *D* is a finite set of dataflow facts

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \sqcap \rangle$$

- $G^* = \langle N^*, E^* \rangle$ is the supergraph
- D is a finite set of dataflow facts
- $F \subseteq 2^D \to 2^D$ is a set of distributive dataflow functions

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \Pi \rangle$$

- $G^* = \langle N^*, E^* \rangle$ is the supergraph
- D is a finite set of dataflow facts
- $F \subseteq 2^D \to 2^D$ is a set of distributive dataflow functions
- $M_F: E^* \to F$ assigns dataflow functions to supergraph edges

Interprocedural Finite Distributive Subset

$$P = \langle G^*, D, F, M_F, \Pi \rangle$$

- $G^* = \langle N^*, E^* \rangle$ is the supergraph
- D is a finite set of dataflow facts
- $F \subseteq 2^D \to 2^D$ is a set of distributive dataflow functions
- $M_F: E^* \to F$ assigns dataflow functions to supergraph edges
- □ is the meet operator

IFDS algorithm computes a meet-over-valid-paths solution:

$$MVP_{IFDS}(P) = \lambda n. \sqcap_{p \in VP(n)} M_F(p)(\emptyset)$$

IFDS algorithm computes a meet-over-valid-paths solution:

$$MVP_{IFDS}(P) = \lambda n. \sqcap_{p \in VP(n)} M_F(p)(\emptyset)$$

IFDS algorithm computes a meet-over-valid-paths solution:

$$MVP_{IFDS}(P) = \lambda n. \sqcap_{p \in VP(n)} M_F(p)(\emptyset)$$

IFDS algorithm computes a meet-over-valid-paths solution:

$$MVP_{IFDS}(P) = \lambda n. \sqcap_{p \in VP(n)} M_F(p)(\emptyset)$$

IFDS algorithm computes a meet-over-valid-paths solution:

$$MVP_{IFDS}(P) = \lambda n. \square_{p \in VP(n)} M_F(p)(\emptyset)$$

IFDS – Representation Relation

Distributive dataflow function ⇔ representation relation

0

$$f = \lambda S \cdot \text{if } y \in S \lor z \in S$$

$$\text{then } S \cup \{x\}$$

$$\text{else } S \setminus \{x\}$$

$$R_f = \begin{cases} \begin{cases} x \\ y \\ y \\ y \end{cases} \end{cases}$$

IFDS – Exploded Supergraph

Stitch all bipartite graphs to get the exploded supergraph:

$$G_P^\# = \langle N^\#, E^\# \rangle$$

IFDS – Exploded Supergraph

Stitch all bipartite graphs to get the exploded supergraph:

$$G_P^\# = \langle N^\#, E^\# \rangle$$

 $P = \langle G^*, D, F, M_F, \Pi \rangle$ is encoded by $G_P^{\#}$

IFDS – Exploded Supergraph

Stitch all bipartite graphs to get the exploded supergraph:

$$G_P^\# = \langle N^\#, E^\# \rangle$$

 $P = \langle G^*, D, F, M_F, \Pi \rangle$ is encoded by $G_P^\#$

 $d \in MVP_{IFDS}(P)(n) \Leftrightarrow \langle n, d \rangle$ is reachable from start node

Interprocedural Distributive Environment

Interprocedural Distributive Environment

L is a finite-height lattice used for the analysis

Interprocedural Distributive Environment

L is a finite-height lattice used for the analysis

- Environment $D \rightarrow L$
 - Dataflow set D

Interprocedural Distributive Environment

L is a finite-height lattice used for the analysis

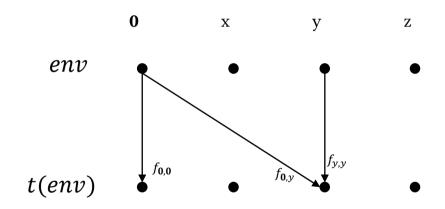
- Environment $D \rightarrow L$
 - Dataflow set D
- Distributive environment transformer $(D \to L) \to (D \to L)$
 - Distributive dataflow function $D \rightarrow D$

IDE – Formal Definition

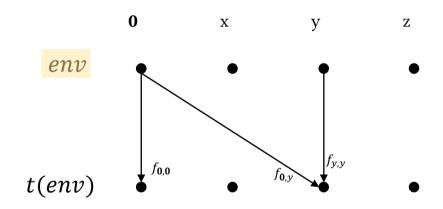
$$P = \langle G^*, D, L, M_{Env} \rangle$$

Meet-over-valid-paths solution:

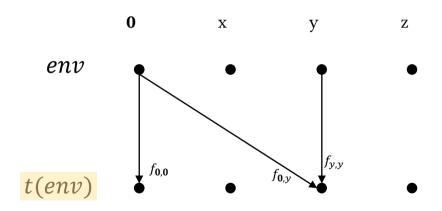
$$MVP_{IDE}(P) = \lambda n. \sqcap_{p \in VP(n)} M_{Env}(p) (\top_{Env})$$



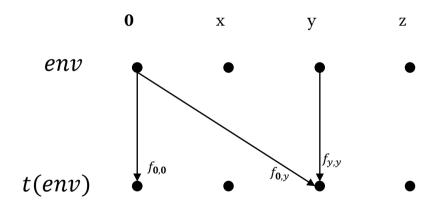
$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{d' \in D} f_{d',y}(env(d')))$$



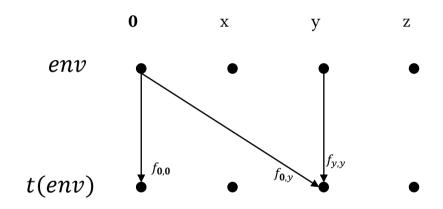
$$t(\underline{env})(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{d' \in D} f_{d',y}(env(d')))$$



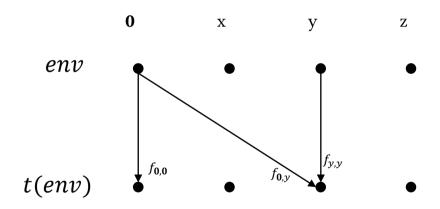
$$t(env)(y) = f_{\mathbf{0},y}(T) \sqcap (\prod_{d' \in D} f_{d',y}(env(d')))$$



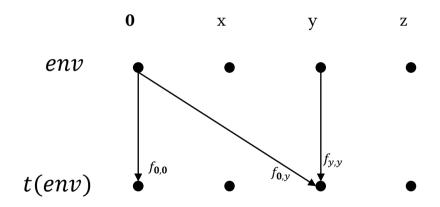
$$t(env)(y) = f_{0,y}(\top) \sqcap (\sqcap_{d' \in D} f_{d',y}(env(d')))$$



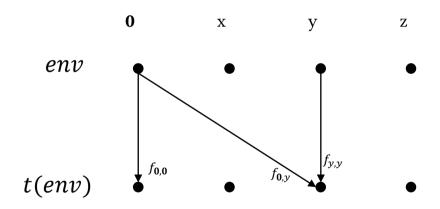
$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{\underline{d'} \in D} f_{\underline{d'},y}(env(d')))$$



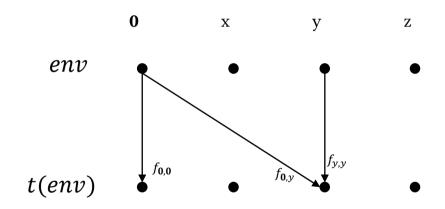
$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{d' \in D} f_{d',y}(env(d')))$$



$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{\underline{d'} \in D} f_{\underline{d'},y}(env(d')))$$



$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \, \square \, (\square_{d' \in D} f_{d',y}(env(d')))$$



$$t(env)(y) = f_{\mathbf{0},y}(\mathsf{T}) \sqcap (\sqcap_{d' \in D} f_{d',y}(env(d')))$$

IDE – Labelled Exploded Supergraph

- Like IFDS exploded supergraph
 - But each edge is labelled with a micro-function

 $\langle G^{\#}, EdgeFn \rangle$

IDE – Labelled Exploded Supergraph

- Like IFDS exploded supergraph
 - But each edge is labelled with a micro-function

$$\langle G^{\#}, EdgeFn \rangle$$

$$P = \langle G^*, D, L, M_{Env} \rangle$$
 is encoded by $\langle G_P^\#, EdgeFn_P \rangle$

IFDS to IDE Transformation

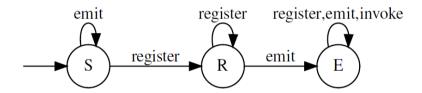
Transformation Overview

Transform IFDS problem instance to IDE problem instance

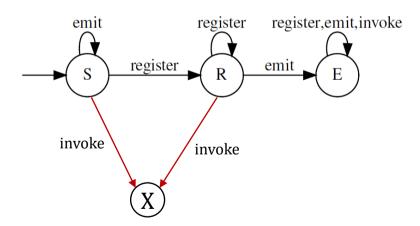
$$T: G^{\#} \rightarrow \langle G^{\#}, EdgeFn \rangle$$

Assign micro-functions to edges of the exploded supergraph

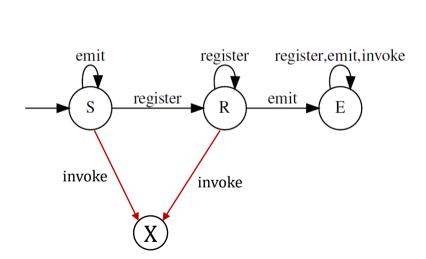
Event Handler State – Model

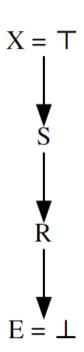


Event Handler State – Model



Event Handler State – Model





- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

```
register(X) = X
register(S) = R
register(R) = R
register(E) = E
```

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X
register(S) = R
register(R) = R
register(R) = R

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X emit(X) = X register(S) = R emit(S) = S register(R) = R emit(R) = E register(E) = E emit(E) = E

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X emit(X) = X register(S) = R emit(S) = S register(R) = R emit(R) = E register(E) = E emit(E) = E

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X emit(X) = X invoke(X) = X register(S) = R emit(S) = S invoke(S) = X register(R) = R emit(R) = E invoke(R) = X register(E) = E invoke(E) = E

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X emit(X) = X invoke(X) = X register(S) = R emit(S) = S invoke(S) = X register(R) = R emit(R) = E invoke(R) = X register(E) = E invoke(E) = E

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Three basic micro-functions, plus identity
 - Most edges are labelled with the identity micro-function

register(
$$X$$
) = X emit(X) = X invoke(X) = X register(S) = R emit(S) = S invoke(S) = X register(R) = R emit(R) = E invoke(R) = X register(E) = E invoke(E) = E

$$EdgeFn(e) = \begin{cases} register & \text{if edge } e \text{ registers the handler} \\ emit & \text{if edge } e \text{ emits an event for the handler} \\ invoke & \text{if edge } e \text{ invokes the handler from the event loop} \\ id & \text{otherwise} \end{cases}$$

- Define the IDE lattice $L' = H \rightarrow L$
 - *H* is the set of event handlers in the program
 - *L* is the event handler state lattice

- Define the IDE lattice $L' = H \rightarrow L$
 - *H* is the set of event handlers in the program
 - L is the event handler state lattice
- IDE computes environments: $D \to (H \to L)$
 - For each node n and fact d, we have a map of handlers to states

- Define the IDE lattice $L' = H \rightarrow L$
 - *H* is the set of event handlers in the program
 - L is the event handler state lattice
- IDE computes environments: $D \rightarrow (H \rightarrow L)$
 - For each node n and fact d, we have a map of handlers to states
- Micro-functions: $(H \to L) \to (H \to L)$
 - Alternate representation: $H \to (L \to L)$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

Soundness

Result computed along concrete execution path

→ Result computed by our technique

Soundness

Result computed along concrete execution path

→ Result computed by our technique

Precision

Result from our technique ⊆ Result computed by IFDS

Soundness

Result computed along concrete execution path

→ Result computed by our technique

Precision

Result from our technique ⊆ Result computed by IFDS

Formal statements and proofs in the paper

Conclusion

 Problem: static analysis of event-driven programs does not respect event handler ordering

- Our approach: transform an existing IFDS problem to an IDE problem
 - IDE problem maintains information about event handler state
- Transformation is sound and precise
 - Formal statements and proofs in paper

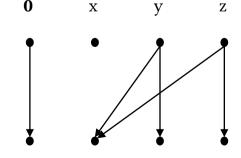
Extra Slides

IFDS – Representation Relation

Distributive dataflow function ⇔ representation relation

$$\begin{split} R_g &= \{\langle \mathbf{0}, \mathbf{0} \rangle\} \cup \\ &\{\langle \mathbf{0}, d \rangle \mid d \in g(\emptyset)\} \cup \\ &\{\langle d_1, d_2 \rangle \mid d_2 \in g(\{d_1\}) \land d_2 \notin g(\emptyset)\} \end{split}$$

$$f = \lambda S \cdot \text{if } y \in S \lor z \in S$$
 then $S \cup \{x\}$ else $S \setminus \{x\}$
$$R_f = \{\langle \mathbf{0}, \mathbf{0} \rangle, \langle y, x \rangle, \langle y, y \rangle, \langle z, x \rangle, \langle z, z \rangle\}$$



IDE – Lattices

If L is a lattice with top element T, the pair $L \times L$ is a lattice:

- Top: ⟨T, T⟩
- Meet: $\langle x_1, y_1 \rangle \sqcap \langle x_2, y_2 \rangle = \langle x_1 \sqcap x_2, y_1 \sqcap y_2 \rangle$

The map $D \rightarrow L$ is also a lattice:

- Top: $T_{Env} = \lambda d$. T
- Meet: $m_1 \sqcap m_2 = \lambda d. (env_1(d) \sqcap env_2(d))$

For a result, if any handler is in state X, discard that result

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

"Untransform" function U applied to IDE result R:

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

"Untransform" function U applied to IDE result R:

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n \{d \mid \forall h \in H . R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H . R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

For a result, if any handler is in state X, discard that result

IFDS result: $N^* \rightarrow D$

IDE result: $N^* \rightarrow (D \rightarrow (H \rightarrow L))$

$$U(R) = \lambda n. \{d \mid \forall h \in H. R(n)(d)(h) \neq X \}$$

$$d \in M_F(p)(\emptyset) \Longrightarrow d \in U(MVP_{IDE}(T(P)))(n)$$

$$d \in M_F(p)(\emptyset) \Longrightarrow d \in U(MVP_{IDE}(T(P)))(n)$$

$$d \in M_F(p)(\emptyset) \Longrightarrow d \in U(MVP_{IDE}(T(P)))(n)$$

$$d \in M_F(p)(\emptyset) \implies d \in U(MVP_{IDE}(T(P)))(n)$$

$$d \in M_F(p)(\emptyset) \Longrightarrow d \in U(MVP_{IDE}(T(P)))(n)$$

$$U\left(MVP_{IDE}(T(P))\right)(n) \subseteq MVP_{IFDS}(P)(n)$$

$$U\left(MVP_{IDE}\left(T(P)\right)\right)(n) \subseteq MVP_{IFDS}(P)(n)$$

$$U\left(MVP_{IDE}(T(P))\right)(n) \subseteq MVP_{IFDS}(P)(n)$$