

Heaps and Hops

Soutenance de thèse

Jules Villard

LSV, ENS Cachan, CNRS

Shift towards Concurrency

Moore's Law

The number of transistors one can put on a chip doubles every two years

Shift towards Concurrency

Moore's Law

The number of transistors one can put on a chip doubles every two years

Moore's law until recently

The frequency of processors doubles every two years

Shift towards Concurrency

Moore's Law

The number of transistors one can put on a chip doubles every two years

Moore's law until recently

The frequency of processors doubles every two years

Moore's law nowadays

- The frequency of processors is reaching limits
- Augment the number of processors on a chip!

Shift towards Concurrency

Moore's Law

The number of transistors one can put on a chip doubles every two years

Moore's law until recently

The frequency of processors doubles every two years

Moore's law nowadays

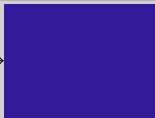
- The frequency of processors is reaching limits
 - Augment the number of processors on a chip!
-
- Concurrent programs are more needed than ever
 - They are hard to write **correctly** and **efficiently**

Message Passing in Multicore Systems

- New paradigm: message passing over a shared memory
- Leads to efficient, copyless message passing
- May be more error-prone

Copyful

data →



```
send(struct, e, data);
```

d

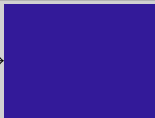
```
d = receive(struct, f);
```

- (e, f): channel
- data points to a big struct
- struct: type of message

To Copy or not to Copy?

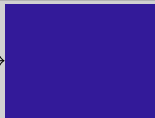
Copyful

data →



```
send(struct, e, data);
```

d →



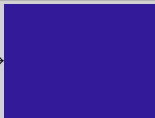
```
d = receive(struct, f);
```

- (e, f) : channel
- data points to a big struct
- struct: type of message

To Copy or not to Copy?

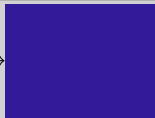
Copyful

data →



```
send(struct, e, data);
```

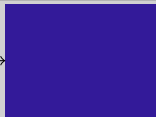
d →



```
d = receive(struct, f);
```

Copyless

data →



```
send(pointer, e, data);
```

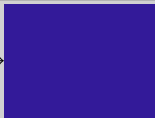
d

```
d = receive(pointer, f);
```

To Copy or not to Copy?

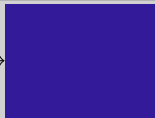
Copyful

data →



```
send(struct, e, data);
```

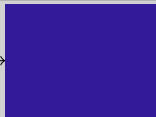
d →



```
d = receive(struct, f);
```

Copyless

data →



← d

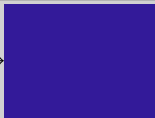
```
send(pointer, e, data);
```

```
d = receive(pointer, f);
```

To Copy or not to Copy?

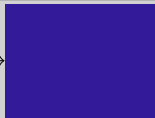
Copyful

data →



```
send(struct, e, data);
```

d →

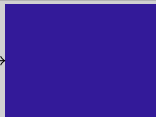


```
d = receive(struct, f);
```

Copyless

Race!

data →



← d

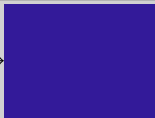
```
send(pointer, e, data);  
dispose(data);
```

```
d = receive(pointer, f);  
dispose(d);
```

To Copy or not to Copy?

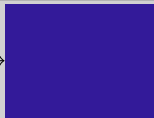
Copyful

data →



```
send(struct, e, data);
```

d →



```
d = receive(struct, f);
```

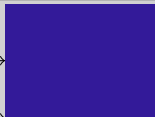
Copyless

Race!

data



d →



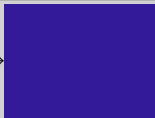
```
send(pointer, e, data);  
dispose(data);
```

```
d = receive(pointer, f);  
dispose(d);
```

To Copy or not to Copy?

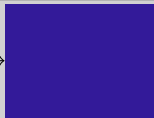
Copyful

data →



```
send(struct, e, data);
```

d →



```
d = receive(struct, f);
```

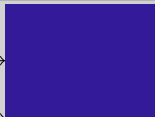
Copyless

No race

data



d →

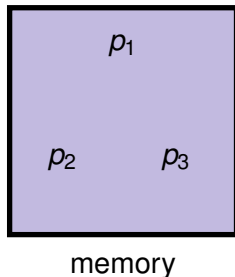


```
send(pointer, e, data);
```

```
d = receive(pointer, f);  
dispose(d);
```

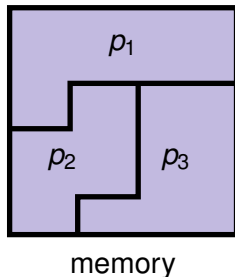
Singularity: a research project and an operating system.

- No hardware memory protection
- Sing \sharp language
- Isolation is verified at compile time
- Invariant: each memory cell is owned by at most one thread
- No shared resources
- Copyless message passing



Singularity: a research project and an operating system.

- No hardware memory protection
- Sing_# language
- Isolation is verified at compile time
- Invariant: each memory cell is owned by at most one thread
- No shared resources
- Copyless message passing



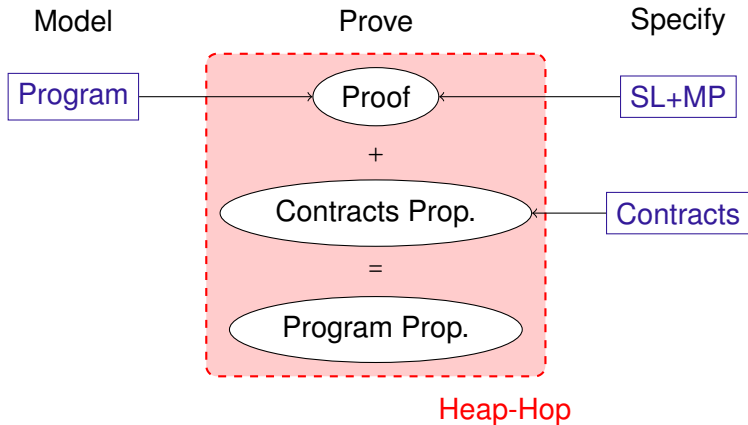
[Fähndrich et al. '06]

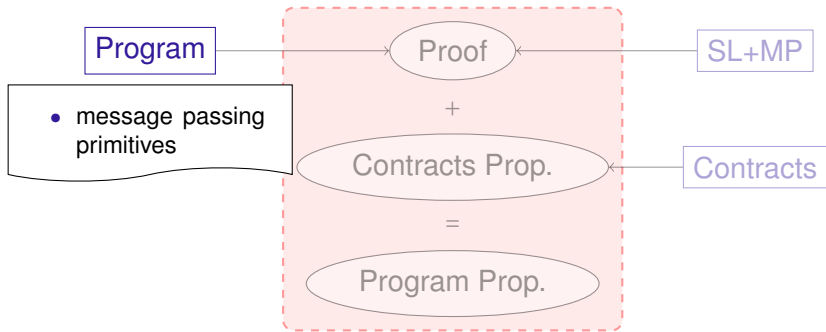
- Channels are **bidirectional** and **asynchronous**
channel = pair of FIFO queues
 - Channels are made of two **endpoints**
similar to the socket model
 - Endpoints can be allocated, disposed of, and
communicated through channels
similar to the π -calculus
 - Communications are ruled by user-defined **contracts**
similar to session types
- ⊖ No formalisation
- How to ensure the absence of bugs?

- **Model** of the program
- **Specify** a correctness criterion in a mathematical language
- **Prove** a theorem which links the two

Main Contributions of the Thesis

- **Model** of the program
 - Semantics of copyless message passing programs
- **Specify** a correctness criterion in a mathematical language
 - Hoare triples: separation logic for channels in the heap
 - Contracts
- **Prove** a theorem which links the two
 - Automatic tool: **Heap-Hop**
 - Extend the proof system of separation logic
 - Properties of contracts rub off on programs





Heap-Hop

Message Passing Primitives

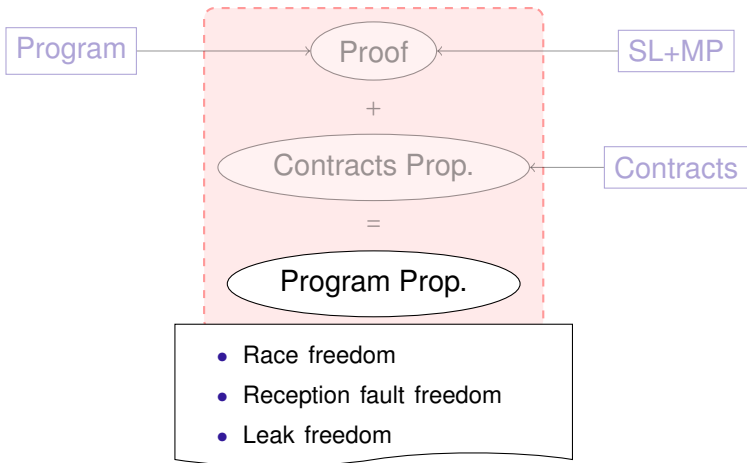
- $(e, f) = \text{open}()$ Creates a bidirectional channel between endpoints e and f
- $\text{close}(e, f)$ Closes the channel (e, f)
- $\text{send}(a, e, x)$ Sends message starting with value x on endpoint e . The message has type/tag a
- $x = \text{receive}(a, e)$ Receives message of type a on endpoint e and stores its value in x

```
1 set_to_ten(x) {  
2   local e, f;  
3   (e, f) = open();  
4   send(integer, e, 10);  
5   x = receive(integer, f);  
6   close(e, f);  
7 }
```

- `switch receive` selects a receive branch depending on availability of messages

```
if( x ) {  
    send(cell,e,x);  
} else {  
    send(integer,e,0);  
}
```

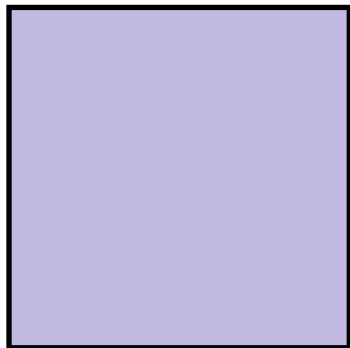
```
switch receive {  
    y = receive(cell,f): {dispose(y);}  
    z = receive(integer,f): {}  
}
```



Separation property

At each point in the execution, the state can be **partitioned** into what is owned by each program and each message in transit.

- Programs access only what they own.
- Prevents races.

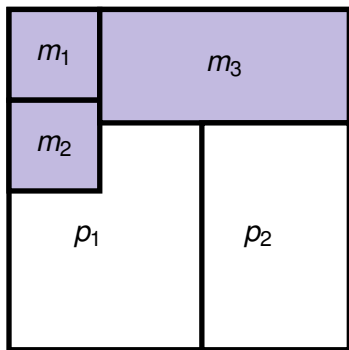


memory

Separation property

At each point in the execution, the state can be **partitioned** into what is owned by each program and each message in transit.

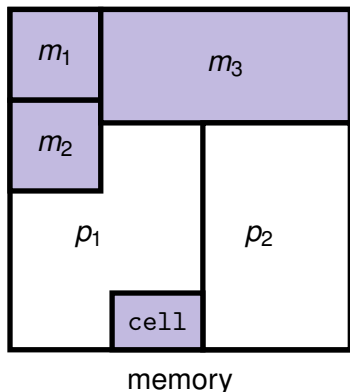
- Programs access only what they own.
- Prevents races.



Separation property

At each point in the execution, the state can be **partitioned** into what is owned by each program and each message in transit.

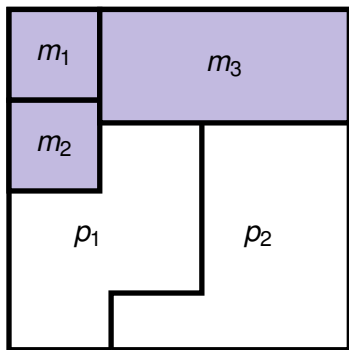
- Programs access only what they own.
- Prevents races.



Separation property

At each point in the execution, the state can be **partitioned** into what is owned by each program and each message in transit.

- Programs access only what they own.
- Prevents races.



Separation property

Invalid receptions freedom

`switch receive` are exhaustive.

```
...  
switch receive {  
  y = receive(a,f): { ... }  
  z = receive(b,f): { ... }  
}  
...
```

```
...  
send(c,e,x);  
...
```

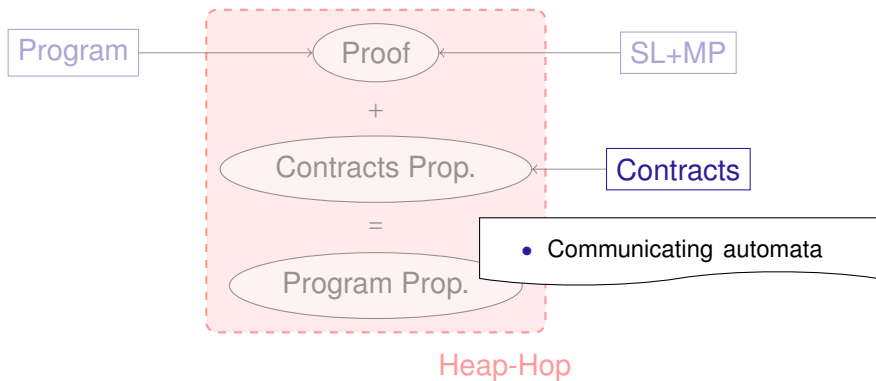
Separation property

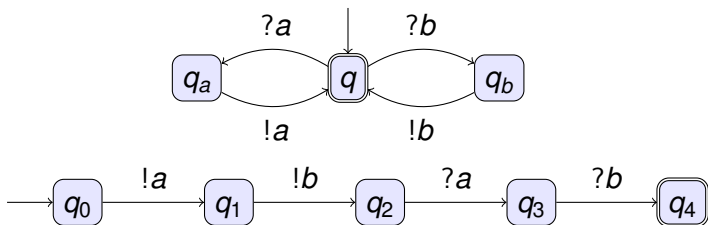
Invalid receptions freedom

Leak freedom

The program does not leak memory.

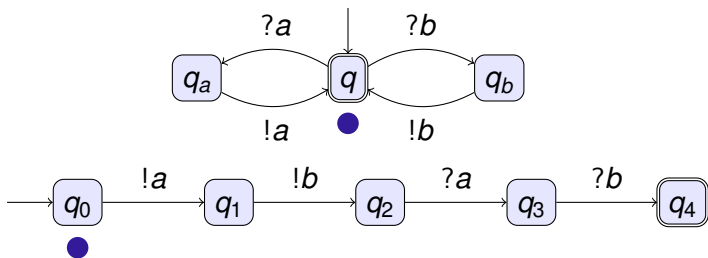
```
1 main() {  
2     local x,e,f;  
3  
4     x = new();  
5     (e,f) = open();  
6     send(cell,e,x);  
7     close(e,f);  
8 }
```





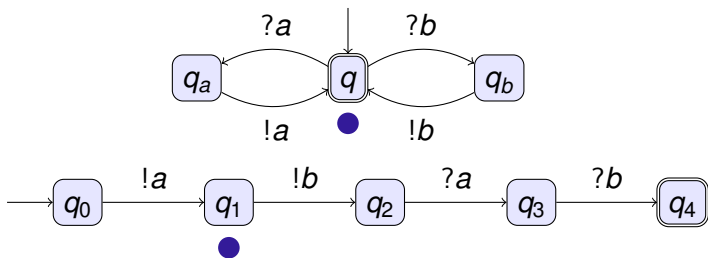
- Sending transitions: $!a$
- Receiving transitions: $?a$
- Two buffers: one in each direction
- Configuration: $\langle q, q', w, w' \rangle$

A Dialogue System



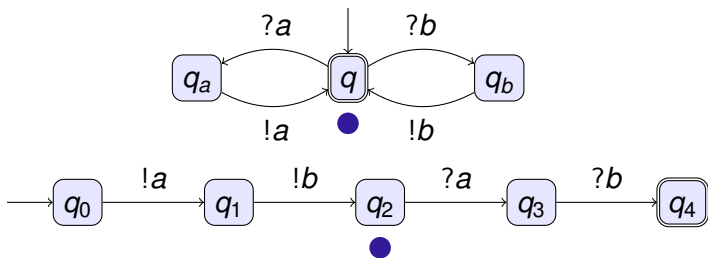
$\langle q, q_0, \varepsilon, \varepsilon \rangle$

A Dialogue System



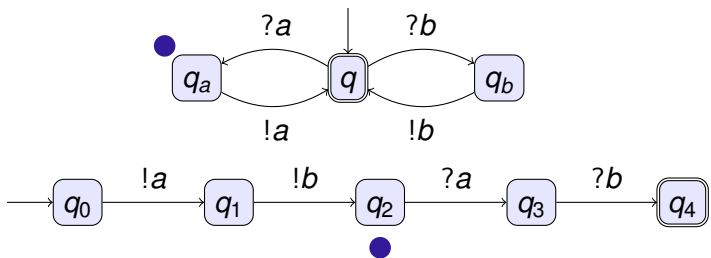
$\langle q, q_1, a, \varepsilon \rangle$

A Dialogue System



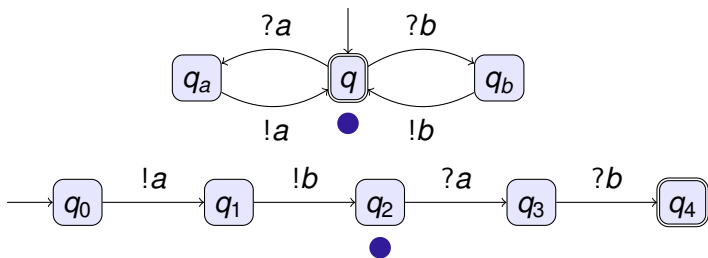
$\langle q, q_2, ab, \varepsilon \rangle$

A Dialogue System



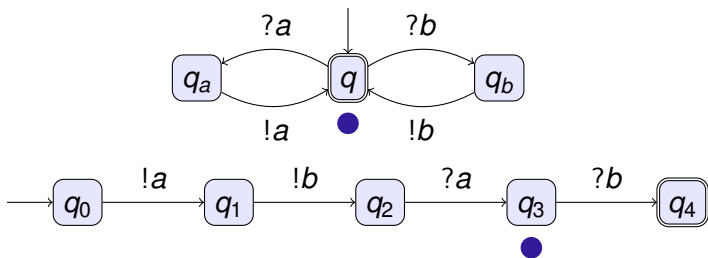
$\langle q_a, q_2, b, \varepsilon \rangle$

A Dialogue System



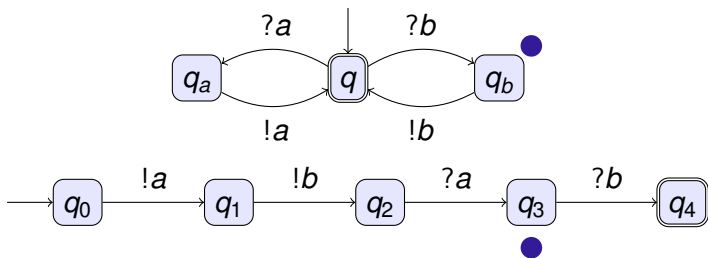
$\langle q, q_2, b, a \rangle$

A Dialogue System



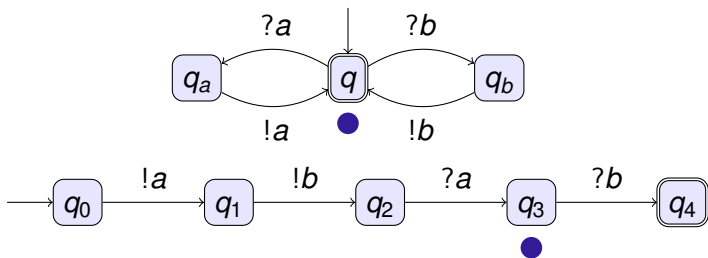
$\langle q, q_3, b, \varepsilon \rangle$

A Dialogue System



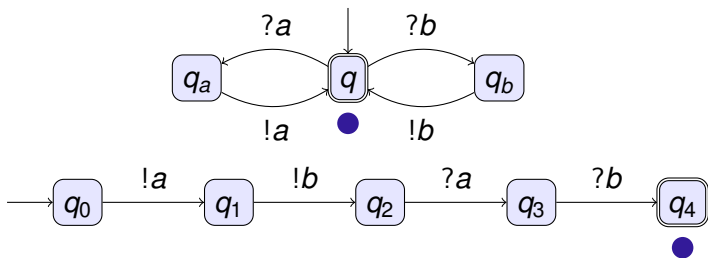
$\langle q_b, q_3, \varepsilon, \varepsilon \rangle$

A Dialogue System



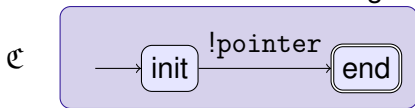
$\langle q, q_3, \varepsilon, b \rangle$

A Dialogue System

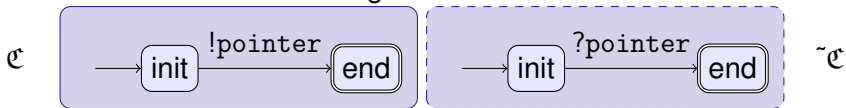


$\langle q, q_4, \varepsilon, \varepsilon \rangle$

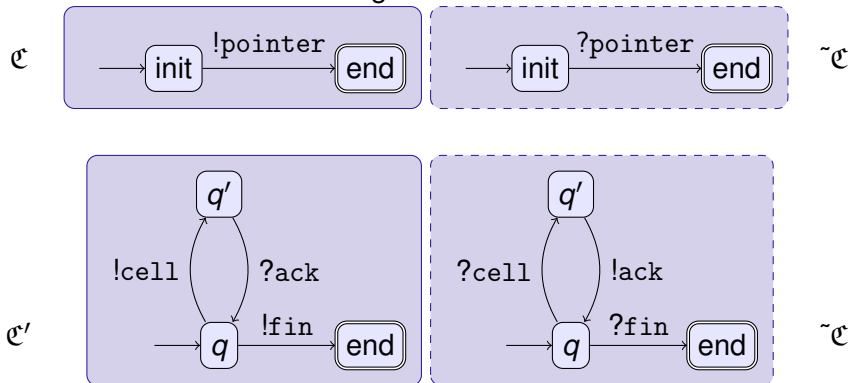
Describe dual communicating finite state machines



Describe dual communicating finite state machines

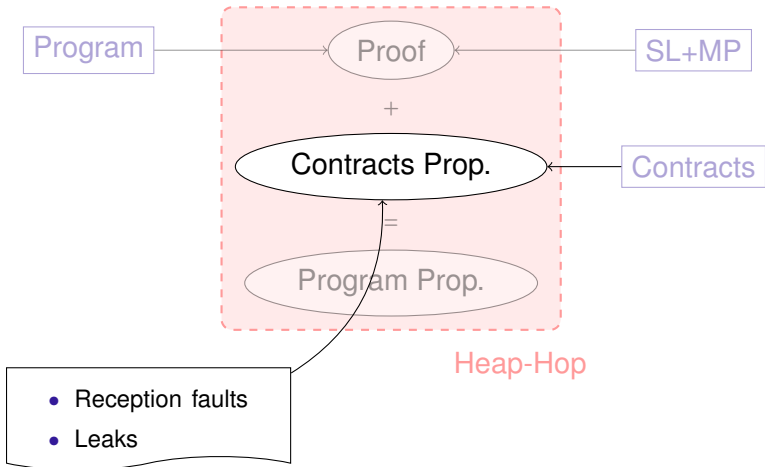


Describe dual communicating finite state machines



Contracts as Protocol Specifications

- $(e, f) = \text{open}(\mathcal{C})$: initialise endpoints in the initial state of the contract
- $\text{send}(a, e, x)$: becomes a $!a$ transition
- $y = \text{receive}(a, f)$: becomes a $?a$ transition
- $\text{closed}(e, f)$ only when both endpoints are in the same **final** state.

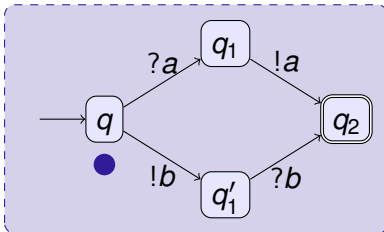
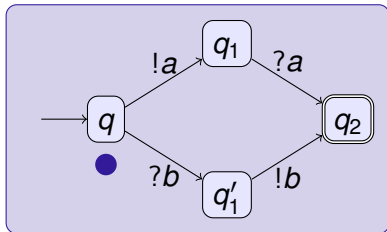


Definition

Reception fault

$\langle q_1, q_2, a \cdot w_1, w_2 \rangle$ is a **reception fault** if

- $q_1 \xrightarrow{?b} q$ for some b and q and
- $\forall b, q. q_1 \xrightarrow{?b} q$ implies $b \neq a$



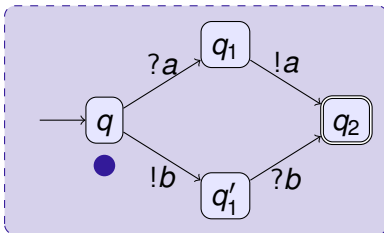
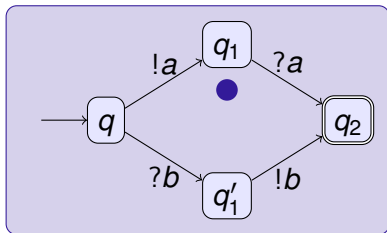
$\langle q, q, \varepsilon, \varepsilon \rangle$

Definition

Reception fault

$\langle q_1, q_2, a \cdot w_1, w_2 \rangle$ is a **reception fault** if

- $q_1 \xrightarrow{?b} q$ for some b and q and
- $\forall b, q. q_1 \xrightarrow{?b} q$ implies $b \neq a$



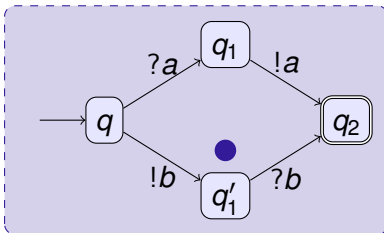
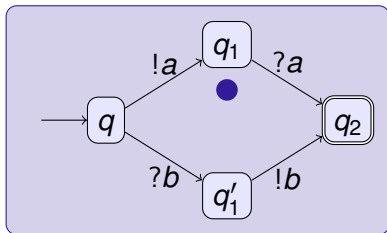
$\langle q_1, q, a, \varepsilon \rangle$

Definition

Reception fault

$\langle q_1, q_2, a \cdot w_1, w_2 \rangle$ is a **reception fault** if

- $q_1 \xrightarrow{?b} q$ for some b and q and
- $\forall b, q. q_1 \xrightarrow{?b} q$ implies $b \neq a$



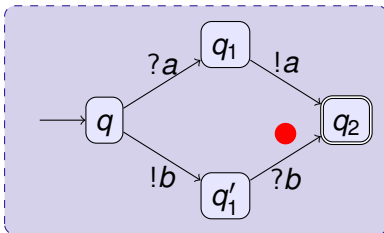
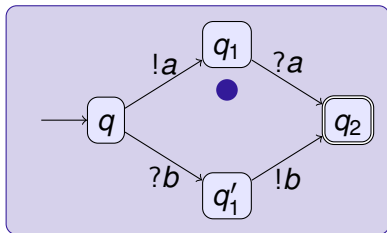
$\langle q_1, q'_1, a, b \rangle$

Definition

Reception fault

$\langle q_1, q_2, a \cdot w_1, w_2 \rangle$ is a **reception fault** if

- $q_1 \xrightarrow{?b} q$ for some b and q and
- $\forall b, q. q_1 \xrightarrow{?b} q$ implies $b \neq a$



$\langle q_1, q'_1, a, b \rangle \xrightarrow{?b}_2$ **error**

Definition

Reception fault

$\langle q_1, q_2, a \cdot w_1, w_2 \rangle$ is a **reception fault** if

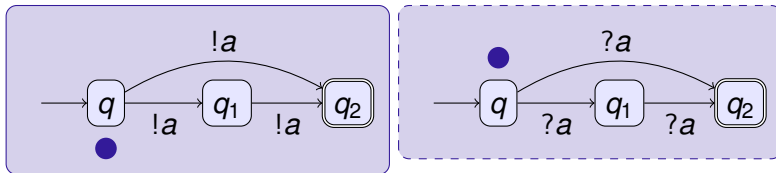
- $q_1 \xrightarrow{?b} q$ for some b and q and
- $\forall b, q. q_1 \xrightarrow{?b} q$ implies $b \neq a$

- A contract is **reception fault-free** if it cannot reach a reception fault.

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.

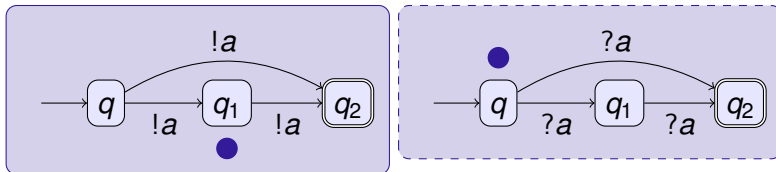


$\langle q, q, \varepsilon, \varepsilon \rangle$

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.

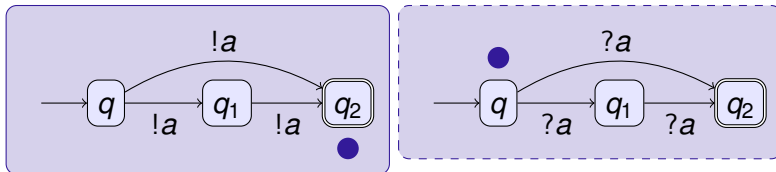


$\langle q_1, q, a, \varepsilon \rangle$

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.

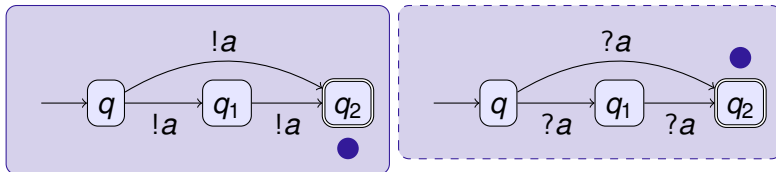


$\langle q_2, q, aa, \varepsilon \rangle$

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.

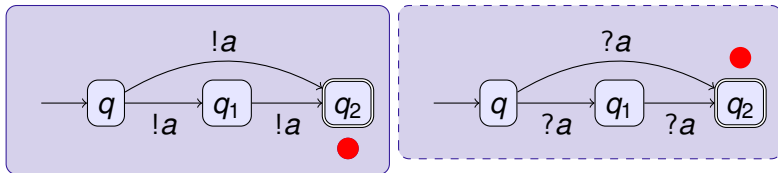


$\langle q_2, q_2, a, \varepsilon \rangle$

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.



$\langle q_2, q_2, a, \varepsilon \rangle$

Definition

Leak

$\langle q_f, q_f, w_1, w_2 \rangle$ is a **leak** if $w_1 \cdot w_2 \neq \varepsilon$ and q_f is final.

- A contract is **leak free** if it cannot reach a leak.
- A contract is **safe** if it is reception fault free and leak free.

- ⊖ Safety of communicating systems is undecidable in general
Channel's buffer \approx Turing machine's tape

- ⊖ Safety of communicating systems is undecidable in general
Channel's buffer \approx Turing machine's tape
- ⊕ Contracts are restricted (dual systems)

- ⊖ Safety of communicating systems is undecidable in general
 - Channel's buffer \approx Turing machine's tape*
- Contracts are restricted (dual systems)
- ⊖ Contracts can encode Turing machines as well

Theorem

Safety is undecidable for contracts.

- ⊖ Safety of communicating systems is undecidable in general
 - Channel's buffer \approx Turing machine's tape*
- Contracts are restricted (dual systems)
- ⊖ Contracts can encode Turing machines as well

Theorem

Safety is undecidable for contracts.

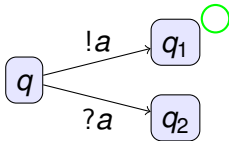
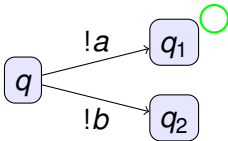
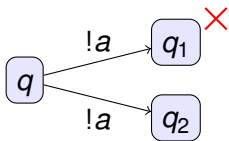
- We give **sufficient conditions** for safety.

Sufficient Conditions for Reception Safety

Definition

Deterministic contract

Two distinct edges in a contract must be labelled by different messages.



Sufficient Conditions for Reception Safety

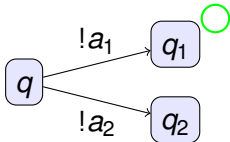
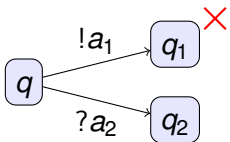
Definition

Deterministic contract

Definition

Positional contracts

All outgoing edges from a same state in a contract must be either all sends or all receives.



Sufficient Conditions for Reception Safety

Definition

Deterministic contract

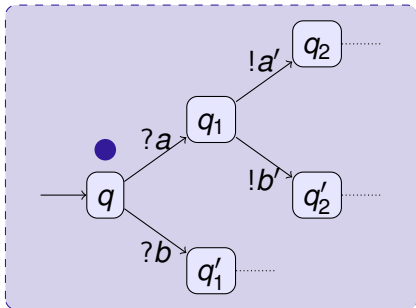
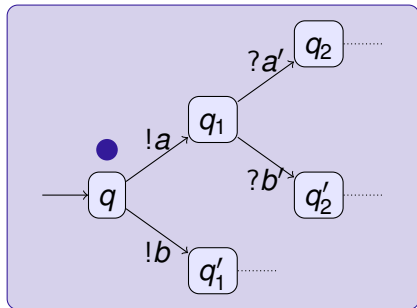
Definition

Positional contracts

Theorem

[Stengel & Bultan'09] • [V., Lozes & Calcagno '09]

*Deterministic positional contracts are **reception fault free**.*



Sufficient Conditions for Reception Safety

Definition

Deterministic contract

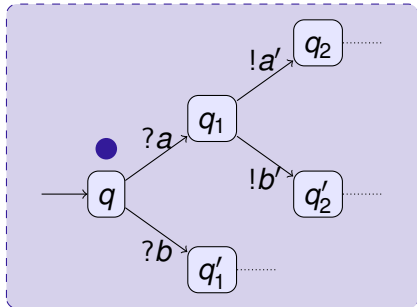
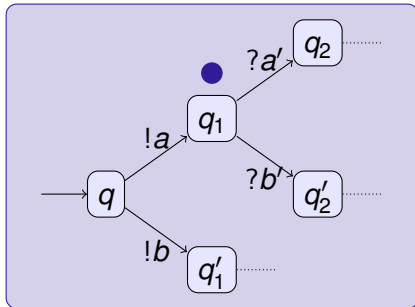
Definition

Positional contracts

Theorem

[Stengel & Bultan'09] • [V., Lozes & Calcagno '09]

*Deterministic positional contracts are **reception fault free**.*



Sufficient Conditions for Reception Safety

Definition

Deterministic contract

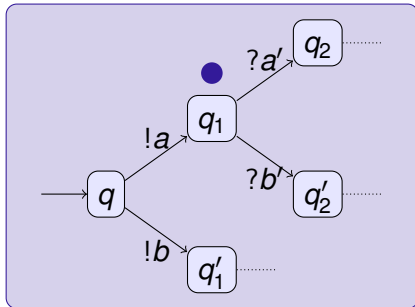
Definition

Positional contracts

Theorem

[Stengel & Bultan'09] • [V., Lozes & Calcagno '09]

*Deterministic positional contracts are **reception fault free**.*



Sufficient Conditions for Reception Safety

Definition

Deterministic contract

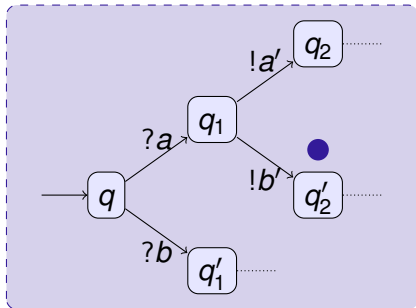
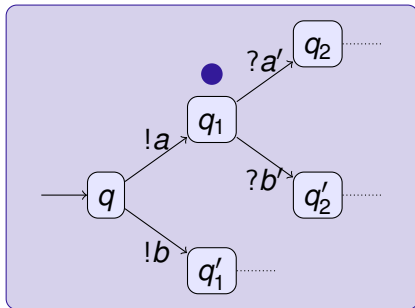
Definition

Positional contracts

Theorem

[Stengel & Bultan'09] • [V., Lozes & Calcagno '09]

*Deterministic positional contracts are **reception fault free**.*



Sufficient Conditions for Reception Safety

Definition

Deterministic contract

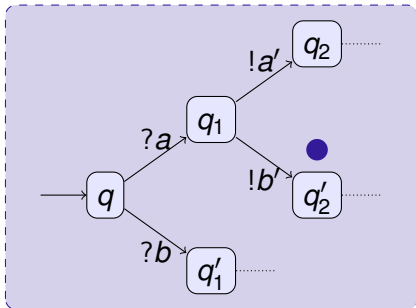
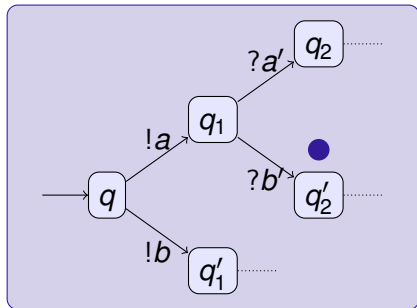
Definition

Positional contracts

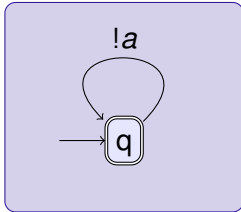
Theorem

[Stengel & Bultan'09] • [V., Lozes & Calcagno '09]

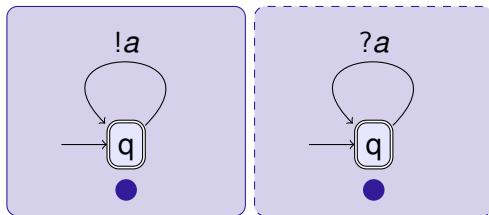
*Deterministic positional contracts are **reception fault free**.*



Another Source of Leaks

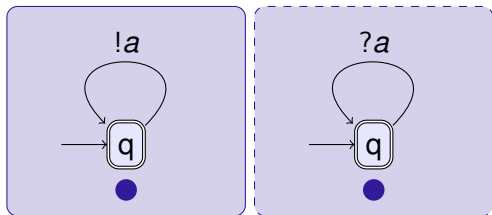


Another Source of Leaks



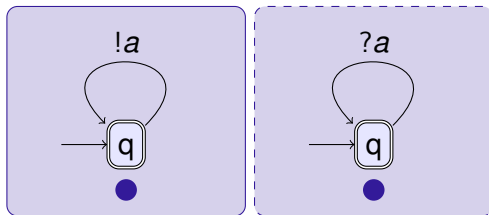
$\langle q, q, \varepsilon, \varepsilon \rangle$

Another Source of Leaks



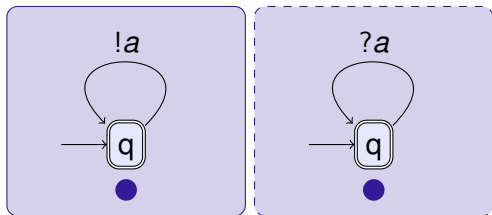
$\langle q, q, a, \varepsilon \rangle$

Another Source of Leaks



$\langle q, q, aa, \varepsilon \rangle$

Another Source of Leaks



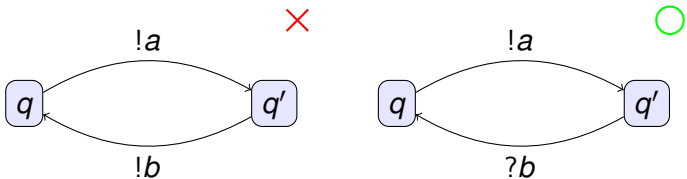
$\langle q, q, aaa, \varepsilon \rangle$

Synchronising Contracts

Definition

Synchronising state

A state s is synchronising if every cycle that goes through it contains at least one send and one receive.



Synchronising Contracts

Definition

Synchronising state

A state s is synchronising if every cycle that goes through it contains at least one send and one receive.

Definition

Synchronising contract

A contract is synchronising if all its final states are.

Synchronising Contracts

Definition

Synchronising state

A state s is synchronising if every cycle that goes through it contains at least one send and one receive.

Definition

Synchronising contract

A contract is synchronising if all its final states are.

Theorem

[V., Lozes & Calcagno '09]

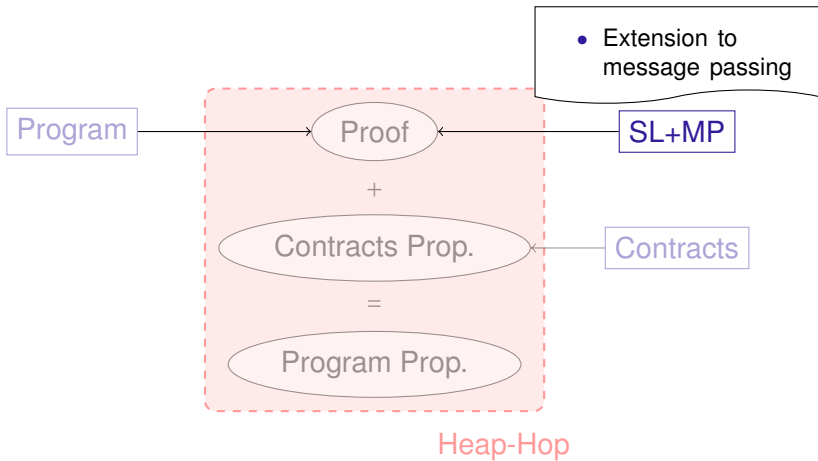
*Deterministic, positional and synchronising contracts are **safe** (fault and leak free).*

Definition

Singularity contract

Singularity contracts are deterministic and **all** their states are synchronising.

- This is missing the positional condition!
- Does not guarantee reception fault freedom
- In fact, we proved that safety is still **undecidable** for deterministic or positional contracts.
- Positional Singularity contracts are **safe** and **bounded**.



Separation Logic

[Reynolds 02, O'Hearn 01, ...]

- An **assertion language** to describe states
- A **proof system** for Hoare triples

- Local reasoning for heap-manipulating programs
- Naturally describes ownership transfers
- Has been extended to storable locks [Gotsman et al. 07]

Syntax

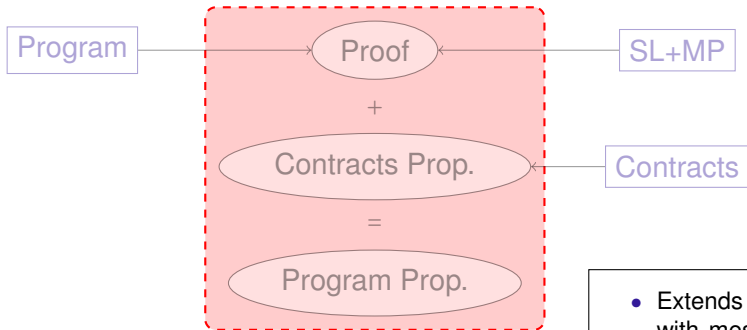
E	$::=$	$x \mid n \in \{0, 1, 2, \dots\} \mid \dots$	expressions
ϕ	$::=$	$E_1 = E_2 \mid E_1 \neq E_2$	stack predicates
		$\mid \text{emp} \mid E_1 \mapsto E_2$	heap predicates
		$\mid \exists x. \phi \mid \phi_1 \wedge \phi_2 \mid \neg \phi \mid \phi_1 * \phi_2$	formulas

Syntax (continued)

$$\phi ::= \dots$$
$$| E \mapsto (\mathcal{C}\{q\}, E') \quad \text{endpoint predicate}$$

Intuitively $E \mapsto (\mathcal{C}\{q\}, E')$ means:

- E is an allocated endpoint
- it is ruled by contract \mathcal{C}
- it is currently in the control state q of \mathcal{C}
- its peer is E'

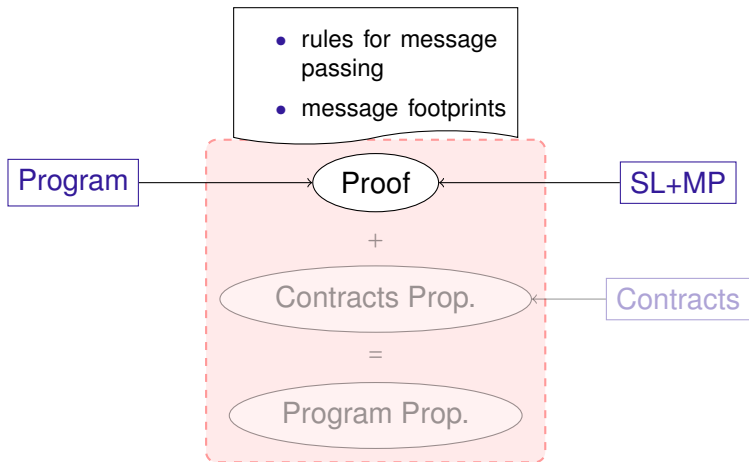


Heap-Hop

- Extends Smallfoot with message passing
- Written in OCaml
- Open source

HEAD HOP

[V., Lozes & Calcagno TACAS'10]



Heap-Hop

Memory States σ

A memory state σ has three components

- A variable valuation (stack)
- A heap for memory cells
- Buffers for endpoints

Semantics of programs

Small-step interleaving operational semantics for programs p :

$p, \sigma \rightarrow^* p', \sigma'$ (intermediate state)

$p, \sigma \rightarrow^* \sigma'$ (final state)

$p, \sigma \rightarrow^* \mathbf{error}$ (error state)

$\{\phi\} p \{\psi\}$: Hoare triple

- ϕ, ψ : formulas
- p : program

Fault-free interpretation of Hoare triples

If $\{\phi\} p \{\psi\}$ is provable, then for all state $\sigma \models \phi$,

1. p has no race or memory faults from σ
2. p implements its contracts
3. if $p, \sigma \rightarrow^* \sigma'$ then $\sigma' \models \psi$

Proof system

Derivation rules to **prove** Hoare triples.

Rules of the Proof System

SKIP ASSUME ASSIGN LOOKUP MUTATE
NEW DISPOSE SEQUENCE PARALLEL CHOICE
STAR LOCAL FRAME WEAKENING
CONJUNCTION DISJUNCTION EXISTENTIAL OPEN
CLOSE SEND CHANNELDISPATCH EXTCHOICE

OPEN

$$\frac{i = \text{init}(\mathcal{C})}{\{\text{emp}\} (e, f) = \text{open}(\mathcal{C}) \{e \mapsto (\mathcal{C}\{i\}, f) * f \mapsto (\mathcal{C}\{i\}, e)\}}$$

CLOSE

$$\frac{q \in \text{finals}(\mathcal{C})}{\{e \mapsto (\mathcal{C}\{q\}, f) * f \mapsto (\sim\mathcal{C}\{q\}, e)\} \text{close}(e, f) \{\text{emp}\}}$$

SEND

$$\frac{q \xrightarrow{!a} q' \in \mathcal{C} \quad e \mapsto (\mathcal{C}\{q'\}, -) * \phi \Rightarrow \gamma_a(e, x) * \phi'}{\{e \mapsto (\mathcal{C}\{q\}, -) * \phi\} \text{send}(a, e, x) \{\phi'\}}$$

RECEIVE

$$\frac{q \xrightarrow{?a} q' \in \mathcal{C}}{\{e \mapsto (\mathcal{C}\{q\}, X')\} x = \text{receive}(a, e) \{e \mapsto (\mathcal{C}\{q'\}, X') * \gamma_a(X', x)\}}$$

CLOSE

$$\frac{q \in \text{finals}(\mathcal{C})}{\{e \mapsto (\mathcal{C}\{q\}, f) * f \mapsto (\sim\mathcal{C}\{q\}, e)\} \text{close}(e, f) \{\text{emp}\}}$$

RECEIVE

$$\frac{q \xrightarrow{?a} q' \in \mathcal{C}}{\{e \mapsto (\mathcal{C}\{q\}, X')\} x = \text{receive}(a, e) \{e \mapsto (\mathcal{C}\{q'\}, X') * \gamma_a(X', x)\}}$$

CLOSE

$$\frac{q \in \text{finals}(\mathcal{C})}{\{e \mapsto (\mathcal{C}\{q\}, f) * f \mapsto (\sim\mathcal{C}\{q\}, e)\} \text{close}(e, f) \{\text{emp}\}}$$

CLOSE

$$\frac{q \in \text{finals}(\mathcal{C})}{\{e \mapsto (\mathcal{C}\{q\}, f) * f \mapsto (\sim\mathcal{C}\{q\}, e)\} \text{close}(e, f) \{\text{emp}\}}$$



General Rule for Receive

RECEIVE

$$q \xrightarrow{?a} q' \in \mathcal{C}$$

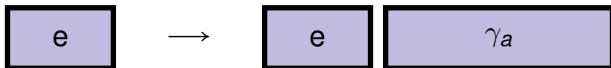
$$\frac{}{\{e \mapsto (\mathcal{C}\{q\}, X')\} x = \text{receive}(a, e) \{e \mapsto (\mathcal{C}\{q'\}, X') * \gamma_a(X', x)\}}$$

General Rule for Receive

RECEIVE

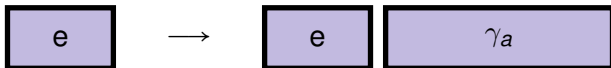
$$q \xrightarrow{?a} q' \in \mathcal{C}$$

$$\frac{}{\{e \mapsto (\mathcal{C}\{q\}, X')\} x = \text{receive}(a, e) \{e \mapsto (\mathcal{C}\{q'\}, X') * \gamma_a(X', x)\}}$$



RECEIVE

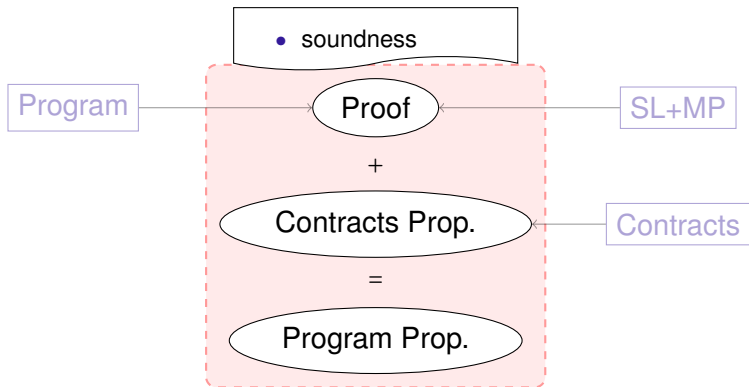
$$\frac{q \xrightarrow{?a} q' \in \mathcal{C}}{\{e \mapsto (\mathcal{C}\{q\}, X')\} x = \text{receive}(a, e) \{e \mapsto (\mathcal{C}\{q'\}, X') * \gamma_a(X', x)\}}$$



Can be instantiated for each example:

$$\gamma_{\text{cell}}(\text{src}, \text{val}) \triangleq \text{val} \mapsto -$$

$$\gamma_{\text{ep}}(\text{src}, \text{val}) \triangleq \text{val} \mapsto (\mathcal{C}\{\text{end}\}, -) \wedge \text{val} = \text{src}$$



Heap-Hop

Definition

Program validity

$\{\phi\} p \{\psi\}$ is valid if, for all $\sigma \models \phi$

- p has **no race or memory fault** starting from σ
- p has **no reception faults** starting from σ
- if $p, \sigma \rightarrow^* \sigma'$ then $\sigma' \models \psi$

Definition

Leak free programs

p is **leak free** if for all σ

$p, \sigma \rightarrow^* \sigma'$ implies that the heap and buffers of σ' are empty

Properties of Proved Programs

Theorem

Soundness

If $\{\phi\} p \{\psi\}$ is provable with **reception fault free** contracts then $\{\phi\} p \{\psi\}$ is valid.

Theorem

Leak freedom

If $\{\phi\} p \{\text{emp}\}$ is provable with **leak free** contracts then p is leak free.

Conclusion

Contracts

- Formalisation of contracts
- Automatic verification of contract properties

Program analysis

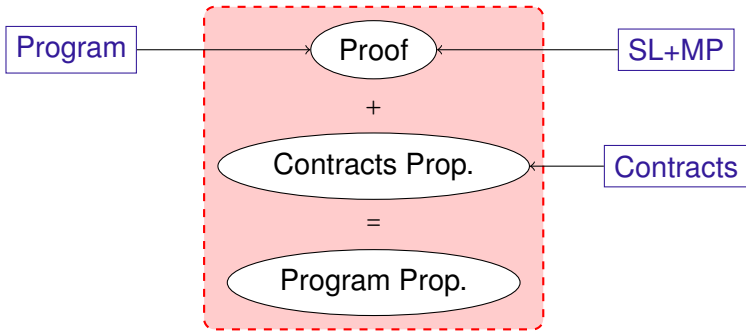
- First extension of separation logic to message passing
- Formalisation of heap-manipulating, message passing programs with contracts
- Contracts and proofs collaborate to prove freedom from reception errors and leaks
- Tool that integrates this analysis: **Heap-Hop**

Contracts

- Prove progress for programs
- Extend to the multiparty case
- Enrich contracts with counters, non determinism, . . .

Automatic program verification

- Discover specs and message footprints
- Discover contracts
- Fully automated tool



Heap-Hop