Formalising Observer Theory for Environment-Sensitive Bisimulation

Jeremy Dawson and Alwen Tiu

Logic and Computation Group
School of Computer Science
College of Engineering and Computer Science
Australian National University
Canberra ACT 0200, Australia
http://users.rsise.anu.edu.au/~jeremy/
http://users.rsise.anu.edu.au/~tiu/

August 7, 2009

Outline

Observer theories

We consider a formalisation of a notion of observer theories.

- used in various "environment-sensitive" bisimulation for process calculi, e.g., the spi-calculus.
- describes the knowledge and capabilities of an observer
- given a formal account using deductive systems

Two critical notions:

- decidability of message deduction by the observer
- consistency of a given theory

We formalise a theory in Isabelle/HOL, encoding observer theories as pairs of symbolic traces.

Messages

Messages are formed from

- "names" (or flexible names), a, x, y : like variables
- rigid names, **a**, **b** : like constants
- pairs of messages, $\langle M, N \rangle$,
- symmetric encryption, $\{M\}_K$, (key K, message M)

```
datatype msg = Name nat
             | Rigid nat
             | Mpair msg msg
             | Enc msg msg
```

Message indistinguishability

Can an observer differentiate two processes based on the messages output by the processes?

With encryption, indistinguishability is not just syntactic equality (one encrypted message looks "just like" another).

For Γ an observer theory (a finite set of pairs of messages), $\Gamma \vdash M \leftrightarrow N$ means the observer cannot distinguish between M and N, given the *indistinguishability assumption* Γ

In Isabelle, $(\Gamma, M, N) \in$ `indist`

Data structures involving pairs of messages can be projected to the first (or second) component. Thus $\pi_i(X)$, $i = 1, 2$, for X a pair, theory, bi-trace, sequent, etc.

Proof system for message indistinguishability

$$\frac{x \in \mathcal{N}}{\Gamma \vdash x \leftrightarrow x}\ (var) \qquad \frac{(M, N) \in \Gamma}{\Gamma \vdash M \leftrightarrow N}\ (id)$$

$$\frac{\Gamma \vdash M_a \leftrightarrow N_a \quad \Gamma \vdash M_b \leftrightarrow N_b}{\Gamma \vdash \langle M_a, M_b\rangle \leftrightarrow \langle N_a, N_b\rangle}\ (pr)$$

$$\frac{\Gamma \vdash M_p \leftrightarrow N_p \quad \Gamma \vdash M_k \leftrightarrow N_k}{\Gamma \vdash \{M_p\}_{M_k} \leftrightarrow \{N_p\}_{N_k}}\ (er)$$

$$\frac{\Gamma, (M_a, N_a), (M_b, N_b) \vdash M \leftrightarrow N}{\Gamma, (\langle M_a, M_b\rangle, \langle N_a, N_b\rangle) \vdash M \leftrightarrow N}\ (pl)$$

$$\frac{\Gamma \vdash M_k \leftrightarrow N_k \quad \Gamma, (M_p, N_p), (M_k, N_k) \vdash M \leftrightarrow N}{\Gamma, (\{M_p\}_{M_k}, \{N_p\}_{N_k}) \vdash M \leftrightarrow N}\ (el)$$

Cut-admissibility and some invertibility results hold

Observer theory consistency

Intuitively, is the theory plausibly a set of pairs of messages which the observer would see as indistinguishable ?

For example, $\{(\{\mathbf{a}\}_{\mathbf{b}}, \{\mathbf{c}\}_{\mathbf{d}}), (\mathbf{b}, \mathbf{c})\}$ is not consistent, since

- one can decrypt $\{\mathbf{a}\}_{\mathbf{b}}$ using $\mathbf{b}$, but
- one cannot decrypt $\{\mathbf{c}\}_{\mathbf{d}}$ using $\mathbf{c}$

Definition (consistent)

A theory Γ is *consistent* if for every M and N, if $\Gamma \vdash M \leftrightarrow N$ then

- M and N are of the same type of expressions, i.e., M is a pair (an encrypted message, a (rigid) name) if and only if N is.
- If $M = \{M_p\}_{M_k}$ and $N = \{N_p\}_{N_k}$ then $\pi_1(\Gamma) \vdash M_k$ implies $\Gamma \vdash M_k \leftrightarrow N_k$ and $\pi_2(\Gamma) \vdash N_k$ implies $\Gamma \vdash M_k \leftrightarrow N_k$.
- For any R, $\Gamma \vdash M \leftrightarrow R$ implies $R = N$ and $\Gamma \vdash R \leftrightarrow N$ implies $R = M$.

Decidability of $\vdash$

want consistency to be decidable — involves deciding $\Gamma \vdash M \leftrightarrow N$

- Naive approach to testing $\Gamma \vdash M \leftrightarrow N$ can loop: in (el) rule, left premise can equal the conclusion
- Finiteness argument shows decidability since backwards proof only introduces sub-messages of messages in conclusion
- but we want a more focussed procedure than exhaustive search
- theory reduction: we "reduce" a theory to its "simplest" form

Theory reduction

As originally defined (Tiu, 2007)

$$\begin{array}{rcl} \Gamma, (\langle M_a, M_b\rangle, \langle N_a, N_b\rangle) & \longrightarrow & \Gamma, (M_a, N_a), (M_b, N_b) \\ \Gamma, (\{M_p\}_{M_k}, \{N_p\}_{N_k}) & \longrightarrow & \Gamma, (M_p, N_p), (M_k, N_k) \\ & & \text{if } \color{red}{\Gamma, (\{M_p\}_{M_k}, \{N_p\}_{N_k}) \vdash M_k \leftrightarrow N_k} \end{array}$$

(assume $(\langle M_a, M_b\rangle, \langle N_a, N_b\rangle) \notin \Gamma$; $(\{M_p\}_{M_k}, \{N_p\}_{N_k}) \notin \Gamma$)
This involves deciding whether $\Gamma, (\{M_p\}_{M_k}, \{N_p\}_{N_k}) \vdash M_k \leftrightarrow N_k$
Alternative definition:

$$\begin{array}{rcl} \Gamma, (\langle M_a, M_b\rangle, \langle N_a, N_b\rangle) & \longrightarrow' & \Gamma, (M_a, N_a), (M_b, N_b) \\ \Gamma, (\{M_p\}_{M_k}, \{N_p\}_{N_k}) & \longrightarrow' & \Gamma, (M_p, N_p), (M_k, N_k) \quad \text{if } \color{red}{\Gamma \vdash M_k \leftrightarrow N_k} \end{array}$$

This involves deciding $\Gamma \vdash M_k \leftrightarrow N_k$ (a *smaller* theory Γ).

Results about theory reduction

Lemma

1. *If* $\Gamma \longrightarrow \Gamma'$ *then* $\Gamma \vdash M \leftrightarrow N$ *if and only if* $\Gamma' \vdash M \leftrightarrow N$
2. $\longrightarrow$ *is well-founded (total size reduces)*
3. *As* $\longrightarrow' \subseteq \longrightarrow$*, the above hold for* $\longrightarrow'$ *also.*
4. $\longrightarrow$ *is confluent*

Proof of (4): By (1), side condition $\Gamma \vdash M_k \leftrightarrow N_k$ iff $\Gamma' \vdash M_k \leftrightarrow N_k$ (Isabelle proof difficult, number of cases explodes).

Theorem

- Γ *has a* $\longrightarrow$*-normal form* $\Gamma\Downarrow$
- $\Gamma \vdash M \leftrightarrow N$ *if and only if* $\Gamma\Downarrow \vdash M \leftrightarrow N$

Theorem

$\Gamma\Downarrow \vdash M \leftrightarrow N$ *if and only if* $\Gamma\Downarrow \vdash_R M \leftrightarrow N$

Use of theory reduction

We define a right derivation $\vdash_R$

$$\frac{x \in \mathcal{N}}{\Gamma \vdash_R x \leftrightarrow x}\ (var) \qquad \frac{(M, N) \in \Gamma}{\Gamma \vdash_R M \leftrightarrow N}\ (id)$$

$$\frac{\Gamma \vdash_R M_a \leftrightarrow N_a \quad \Gamma \vdash_R M_b \leftrightarrow N_b}{\Gamma \vdash_R \langle M_a, M_b \rangle \leftrightarrow \langle N_a, N_b \rangle}\ (pr)$$

$$\frac{\Gamma \vdash_R M_p \leftrightarrow N_p \quad \Gamma \vdash_R M_k \leftrightarrow N_k}{\Gamma \vdash_R \{M_p\}_{M_k} \leftrightarrow \{N_p\}_{N_k}}\ (er)$$

Now $\Gamma \vdash_R M \leftrightarrow N$ is obviously decidable (just keep decomposing the right-hand side, and testing for (id) rule).

Theorem

$\Gamma\Downarrow\ \vdash M \leftrightarrow N$ *if and only if* $\Gamma\Downarrow\ \vdash_R M \leftrightarrow N$

The alternative theory reduction $\longrightarrow'$

Theorem

- *If Γ is $\longrightarrow$-reducible, then it is $\longrightarrow'$-reducible, (though the same reduction may not be available) (long proof in paper)*
- *Thus (as $\longrightarrow' \subseteq \longrightarrow$) $\Gamma\Downarrow$ is also the $\longrightarrow'$-normal form of Γ*

Theorem

$\Gamma \vdash M \leftrightarrow N$ *is decidable*

Procedure: calculate $\Gamma\Downarrow$ and determine whether $\Gamma\Downarrow \vdash_R M \leftrightarrow N$.

Calculating $\Gamma\Downarrow$ (*using* $\longrightarrow'$) requires deciding questions of the form $\Gamma' \vdash M_k \leftrightarrow N_k$, where Γ' is *smaller* than Γ (because a pair $(\{M_p\}_{M_k}, \{N_p\}_{N_k})$ is omitted).

Thus this procedure terminates.

Theory reduction and consistency (Tiu, 2007)

Lemma

- *If $\Gamma \longrightarrow \Gamma'$ then Γ is consistent if and only if Γ' is consistent*
- *Γ is consistent if and only if $\Gamma{\Downarrow}$ is consistent*

Lemma

There is a simpler, finitely checkable, characterisation of consistency for a reduced theory Γ:

using *"for every $(M, N) \in \Gamma$ ..."*

not *"for every M and N, if $\Gamma \vdash M \leftrightarrow N$ then ..."*

Thus theory consistency is decidable

Proving decidability or computability

To prove computability formally requires modelling the computation process, but we can prove it "semi-formally":

We have a definition of $\Gamma \vdash_R M \leftrightarrow N$ in Isabelle as an inductively defined set (rules above); we gave another corresponding definition as a recursive function (here, $\vdash_f$), eg

$$\Gamma \vdash_f \langle M_a, M_b \rangle \leftrightarrow \langle N_a, N_b \rangle \iff (\langle M_a, M_b \rangle, \langle N_a, N_b \rangle) \in \Gamma \ \vee \ (\Gamma \vdash_f M_a \leftrightarrow N_a \wedge \Gamma \vdash_f M_b \leftrightarrow N_b)$$

- Isabelle makes us prove that the recursive definition of $\vdash_f$ terminates.
- We can inspect to see the absence of any further "infinite" features (eg testing for membership of an infinite set, quantification over an infinite set)
- We then proved $\Gamma \vdash_R M \leftrightarrow N$ iff $\Gamma \vdash_f M \leftrightarrow N$

Decidability of reduction

For theory reduction using $\longrightarrow'$,

- reducing Γ (calculating $\Gamma\Downarrow$) required deciding $\Gamma' \vdash M_k \leftrightarrow N_k$, for some Γ' smaller than Γ,
- deciding $\Gamma' \vdash M_k \leftrightarrow N_k$ required calculating $\Gamma'\Downarrow$ (and then testing $\Gamma'\Downarrow \vdash_R M_k \leftrightarrow N_k$)

Definition of reduction as a function is further complicated by the fact that the single-step reduction relation is not deterministic.

We defined a function `reduce` which chooses a possible reduction, performs it, and then reduces the result.

Isabelle wasn't able to prove the termination conditions automatically, so we had to use `recdef (permissive)`.

The reduction function — using `recdef` (permissive)

We had to prove that the measure function gets smaller, and thereby simplify the simplification rules produced by Isabelle

That is, with a definition (measure function m) :
`reduce` $S = \ldots$ if $m(F(S)) < m(S)$ then $F(S)$ else arbitrary $\ldots$
we had to prove that the arbitrary clause never applied

Finally we proved that `reduce` gives the $\longrightarrow'$-normal form.

Then, by inspection of the text of the definition, we asserted that there was no part of it whose computation would not be finite.

Bi-traces and respectful substitutions

Definition

- A *bi-trace* is an *ordered* set (a list) of message pairs, each marked as *i* (input) or *o* (output), where any free name first appears in an input pair.
- A *substitution pair* is a pair of mappings $\vec{\theta} = (\theta_1, \theta_2)$ from free names to messages, where $\theta_1(\theta_2)$ applies to the first (second) message of any pair.
- a *respectful* substitution is (roughly) a substitution where for each variable x in an input pair, $\Gamma\vec{\theta} \vdash x\theta_1 \leftrightarrow x\theta_2$ where Γ is the set of previous pairs (that is, input messages are only those which an outsider is capable of creating)

Consistent bi-traces

Definition (Consistent bi-trace)

A bi-trace is *consistent* if
... for *every* respectful substitution pair $\vec{\theta}$, $\Gamma\vec{\theta}$ is a consistent theory

The quantification makes deciding this difficult (work in progress).
But note, the definition of respectful substitution involves the order of pairs; deciding whether a theory is consistent involves reducing it, which requires an unordered theory.
So we defined a variant of respectfulness:

Definition (`thy_strl_resp`)

A substitution pair $\vec{\theta}$ satisfies `thy_strl_resp` for Γ and p if, for each x in Γ, $(\Gamma|p(x))\ \vec{\theta} \vdash x\theta_1 \leftrightarrow x\theta_2$ where $\Gamma|p(x)$ is got by removing message pairs containing free names other than those in $p(x)$ from Γ

Theory reduction and `thy_strl_resp`

Lemma

For given θ and p (see below), if Γ satisfies $\mathtt{thy\_strl\_resp}$

- *if $\Gamma \longrightarrow \Gamma'$, then Γ' satisfies $\mathtt{thy\_strl\_resp}$*
- *$\Gamma\Downarrow$ satisfies $\mathtt{thy\_strl\_resp}$*

We use this result where $p(x)$ is the set of free names which appeared prior to x in the bi-trace from which Γ was obtained.

This definition and result enabled us to combine the ideas of reduction of an unordered theory with the respectfulness of a substitution pair with respect to an ordered bi-trace.

Unique Completion of a Respectful Substitution

In analysing bi-trace consistency ("for *all* respectful substitutions"), the following result is useful.

Theorem

Given a consistent bi-trace h whose projections to a single message trace are s_1 and s_2, and a substitution θ_1 which respects s_1, there exists θ_2 such that $\vec{\theta} = (\theta_1, \theta_2)$ respects h, and θ_2 is "unique" in the sense that any two such θ_2 act the same on names in $\pi_2(h)$

Given θ_1 we want to compute θ_2.

Computing the Unique Completion

First we defined a function `match_rc1` which, given a theory Γ and a message M, "attempts" to determine a message N such that $\Gamma \vdash M \leftrightarrow N$. ($N$ is unique if Γ is consistent).

Theorem

If Γ is consistent, then
$\Gamma \vdash M \leftrightarrow N$ *iff* $\mathit{match\_rc1}\ \Gamma{\Downarrow}\ M = \mathit{Some}\ N$

Then we defined a function `second_sub`, using `match_rc1`, to find the appropriate value of $x\theta_2$ for each new x in the bi-trace

Theorem

If h is a consistent bi-trace, and θ_1 satisfies the respectfulness condition for $\pi_1(h)$, and $\theta_2 = \mathit{second\_sub}\ h\ \theta_1$, then (θ_1, θ_2) respects h

Informal arguments show that `match_rc1` and `second_sub` are finitely computable.

Conclusion : value of the formalisation

- theories very intricate, with low-level detail
- details can be overlooked in paper proofs, we have found bugs
- symbolic decision procedures (on-going work) are often very technical and complicated; no-one has verified any symbolic techniques for process calculi as far as we know: this is a first attempt
- it has helped us find better proofs about reduction and respectful substitutions
- theorem proving system helps keep track of results proved (numerous when several different definitions of reduction and sets of rules for deriving $\Gamma \vdash M \leftrightarrow N$)